J Mol Evol (1989) 28:442–450

Journal of
Molecular Evolution

Molecular Evolution of Enzyme Structure: Construction of a Hybrid Hamster/*Escherichia coli* Aspartate Transcarbamoylase

Joseph G. Major, Jr.,[1] Melinda E. Wales,[1] John E. Houghton,[1]* Julie A. Maley,[2]
Jeffrey N. Davidson,[2] and James R. Wild[1]

[1] Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843, USA
[2] Department of Microbiology and Immunology, University of Kentucky, Albert B. Chandler Medical Center, Lexington, Kentucky 40536, USA

Summary. Aspartate transcarbamoylase (ATCase, EC 2.1.3.2) is the first unique enzyme common to de novo pyrimidine biosynthesis and is involved in a variety of structural patterns in different organisms. In *Escherichia coli,* ATCase is a functionally independent, oligomeric enzyme; in hamster, it is part of a trifunctional protein complex, designated CAD, that includes the preceding and subsequent enzymes of the biosynthetic pathway (carbamoyl phosphate synthetase and dihydroorotase). The complete complementary DNA (cDNA) nucleotide sequence of the ATCase-encoding portion of the hamster CAD gene is reported here. A comparison of the deduced amino acid sequences of the hamster and *E. coli* catalytic peptides revealed an overall 44% amino acid similarity, substantial conservation of predicted secondary structure, and complete conservation of all the amino acids implicated in the active site of the *E. coli* enzyme. These observations led to the construction of a functional hybrid ATCase formed by intragenic fusion based on the known tertiary structure of the bacterial enzyme. In this fusion, the amino terminal half (the "polar domain") of the fusion protein was provided by a hamster ATCase cDNA subclone, and the carboxyl terminal portion (the "equatorial domain") was derived from a cloned *pyrBI* operon of *E. coli* K-12. The recombinant plasmid bearing the hybrid ATCase was shown to satisfy growth requirements of transformed *E. coli pyrB*$^-$ cells. The functionality of this *E. coli*–hamster hybrid enzyme confirms conservation of essential structure–function relationships between evolutionarily distant and structurally divergent ATCases.

Key words: Gene fusion — Protein evolution — Multifunctional enzyme complexes

Introduction

Aspartate transcarbamoylase (carbamoylphosphate: L-aspartate carbamoyl-transferase, ATCase, EC 2.1.3.2) is a ubiquitous enzyme of pyrimidine biosynthesis and exists in a variety of oligomeric aggregates and multifunctional enzyme complexes in various organisms (Fig. 1; for reviews, see Makoff and Radford 1978; Jones 1980; Keppler and Holstege 1982). The simplest organization is a nonregulated trimer (c_3, 3 × 35,000–40,000 daltons) reported in Gram-positive bacteria (Bethell et al. 1968; Chang and Jones 1974; Brabson et al. 1985). The enteric bacteria possess a structurally complex, allosterically regulated enzyme containing two interactive catalytic trimers associated with three regulatory dimers, $2(c_3):3(r_2)$ (Gerhart and Schachman 1965; Wild et al. 1980; Foltermann et al. 1981, 1986). *Pseudomonas fluorescens* contains both regulatory and catalytic ATCase functions on identical polypeptide dimers (2 × 180,000 daltons) (Bethell and Jones 1969; Adair and Jones 1972), whereas plants appear to possess UMP-regulated trimers (Ong and Jackson 1972; Lovatt and Cheng 1984). Other eukaryotic organisms express ATCases in multifunc-

Offprint requests to: J.R. Wild
* *Current address:* Department of Biology, Yale University, New Haven, Connecticut 06511, USA

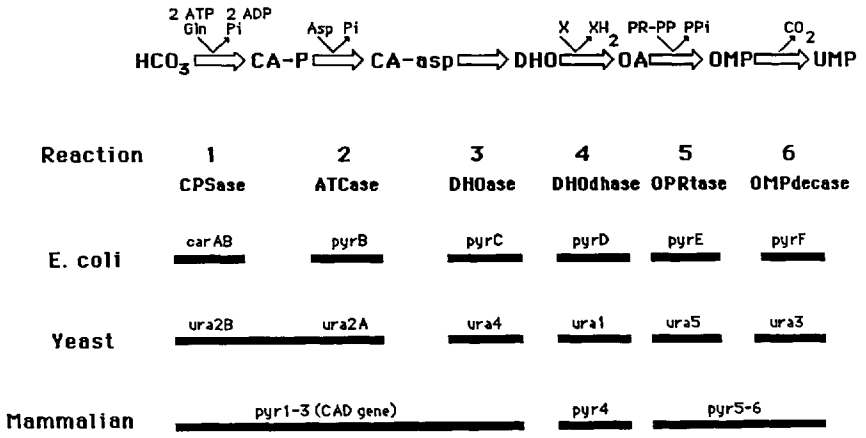

Fig. 1. Schematic representation of the de novo pyrimidine biosynthetic pathway and comparison of the genetic organization of this pathway in representative species. Continuous lines indicate a multifunctional gene product where the corresponding enzymes are contiguous on a single polypeptide chain. CA-P: carbamoyl phosphate; CA-asp: carbamoyl aspartate; DHO: dihydroorotate; OA: orotate; OMP: orotidylate; UMP: uridylate; PR-PP: 5-phosphoribosyl-1-pyrophosphate.

tional enzyme complexes. The fungi, yeast, and *Neurospora* have evolved bifunctional enzyme complexes ("CA," 800,000 daltons) containing both ATCase and carbamoyl phosphate synthetase (CPSase, EC 6.3.5.5), the preceding enzyme in the pathway (Lue and Kaplan 1969; William et al. 1970; Makoff and Radford 1978). Metazoa, including mammals, possess a UTP-regulated trifunctional enzyme complex ("CAD," 800,000–1,200,000 daltons) that includes ATCase, CPSase, and dihydroorotase (DHOase, EC 3.5.2.3), the first three enzymes in pyrimidine biosynthesis (Coleman et al. 1977; Davidson and Patterson 1979). A number of hypotheses have been proposed to explain the possible metabolic advantages to the evolution of multifunctional proteins, including substrate channeling (Jones 1980) and coordinate regulation (Stark 1977; Davidson and Patterson 1984).

The *E. coli* holoenzyme is an allosterically regulated dodecamer and its structure has been described in some detail (Monaco et al. 1978; Honzatko and Lipscomb 1982; Honzatko et al. 1982; Volz et al. 1986; Krause et al. 1987). Individual catalytic trimers are capable of unregulated Michaelian catalysis with its substrates, aspartate and carbamoyl phosphate (Gerhart and Schachman 1965). The holoenzyme demonstrates cooperative binding of aspartate, and it is allosterically inhibited by the ultimate end product of the pathway, cytidine-5′-triphosphate (CTP). In addition, adenosine-5′-triphosphate (ATP) is an allosteric activator of catalysis (Bethell et al. 1968). Each active site is formed at the interface between adjacent catalytic polypeptides in the functional trimers (Honzatko and Lipscomb 1982; Volz et al. 1986; Krause et al. 1987) and each catalytic polypeptide is structurally divided into an amino terminal "polar domain," implicated in binding carbamoyl phosphate, and a carboxy terminal "equatorial domain" primarily involved in aspartate binding (Honzatko et al. 1982).

In hamster, a single polypeptide of 200–240 kilodaltons (kd) provides the first three enzymatic steps of pyrimidine biosynthesis: CPSase, ATCase, and DHOase (Fig. 1). The native protein aggregates are arranged as trimeric or hexameric complexes (Coleman et al. 1977; Davidson and Patterson 1981). In the hamster enzyme, only the CPSase activity appears to be allosterically regulated. Several studies, including proteolytic cleavage of purified CAD protein, have demonstrated that the CAD multifunctional complex is organized into three structurally and functionally distinct regions (Davidson et al. 1981; Mally et al. 1981; Rumsby et al. 1984; Grayson et al. 1985). In both *Drosophila melanogaster* and hamster it appears that the catalytic domains are linearly ordered with CPSase located at the amino terminus, DHOase in the central portion of the polypeptide, and ATCase at the carboxy terminus (Freund and Jarry 1987; Carrey and Hardie 1988). Active hamster ATCase, produced by limited proteolytic cleavage of CAD, has been shown to aggregate in trimeric subunits. In both intact and proteolytically cleaved CAD complexes, ATCase does not appear to be allosterically regulated (Grayson and Evans 1983). Structure–function relationships within the ATCase or other individual catalytic domains of the CAD protein have not been determined.

An independent hamster ATCase superdomain has been produced in *E. coli* from a 3′ segment of CAD complementary DNA (cDNA) (Davidson and Niswander 1983; Maley et al. 1984). This truncated cDNA construct contained approximately 1000 bp of the hamster coding sequence and produced a polypeptide similar in size to the *E. coli* ATCase catalytic polypeptide (Maley and Davidson 1988). The expressed hamster gene was able to satisfy pyrimidine auxotrophy in *E. coli* (Davidson and Niswander 1983). This paper details the sequence analysis of this gene and describes the genetic construction of an active hybrid enzyme formed from the fusion of the polar domain of the hamster enzyme and the

equatorial domain of the *E. coli* ATCase. The structure–function similarities between the two enzymes suggested by comparative sequence analysis are verified experimentally by this enzymatically active construction. Thus provides direct support for the general argument that functional structure similarities can be implied by deduced sequence similarities.

Materials and Methods

Plasmid Constructions. Plasmids encoding ATCases from both *E. coli* and hamster have been constructed as described elsewhere. pPBh105-Ec contains an intact *E. coli pyrBI* operon (Roof et al. 1982) and pAL4 contains a 1.35-kb portion of a 6.5-kb hamster CAD cDNA (Davidson and Niswander 1983; Maley et al. 1984). A truncated cDNA plasmid construct containing approximately 1000 bp of coding sequence was expressed from the *lac* promoter/operator producing an active ATCase of approximately 40,000 daltons (Maley and Davidson 1988). DNA restriction fragments containing hybrid ATCase genes were introduced into M13 series vectors according to established procedures (Zinder and Boeke 1982).

Cell Strains. All plasmids used in this study were transformed into HB101-4442, a *pyrB*$^-$ *E. coli* K-12 cell line produced by Mu-cts insertional activation (Casadaban and Cohen 1979) of *pyrB* (Foltermann et al. 1984). JM103 is an *E. coli* strain suitable for growth of single-stranded phage M13 and its recombinants (Messing and Seeburg 1981).

DNA Sequencing. The complete DNA sequence of the hamster CAD ATCase gene was determined from both DNA strands using dideoxynucleotide plasmid sequencing (Sanger et al. 1977; Chen and Seeburg 1985). A series of sequencing primers was synthesized with an Applied Biosystems Oligonucleotide Synthesizer using phosphoramidite chemistry. Appropriate primers were determined and prepared through successive stages of sequencing in the TAES Oligonucleotide Services Facility. DNA sequencing of the chimeric ATCase was performed on M13 phage recombinants (Zinder and Boeke 1982) also using dideoxynucleotide sequencing.

Enzyme Characterization. Confirmation of the synthesis and determination of the apparent molecular weight of the pHEc270 hybrid protein were achieved through polyacrylamide gel electrophoresis of radioactively labeled *E. coli* CSR603 maxicells (Sancar et al. 1979) transformed with pAL4 and pHEc270 and electrophoresed with appropriate size standards. The native molecular weight of the hybrid enzyme was determined through Western blot analysis of fractions from a Sephadex G-200 gel filtration chromatography using rabbit polyclonal anti-*E. coli* ATCase. All enzymatic assays employed the techniques of Prescott and Jones (1969). Kinetic analyses were performed with saturating levels of carbamoylphosphate (4.8 mM) and varying aspartate.

Results

The complete hamster ATCase nucleotide and amino acid sequences are shown in Fig. 2. The first seven codons corresponded with the expected β-galactosidase amino terminus of the cloning vector (Silhavy et al. 1984) and the similarity with ATCase began after 36 amino acids designated the "CAD linker" (Fig. 2). When the hamster sequence was aligned with other reported ATCase sequences (Fig. 3), the deduced *E. coli* and hamster ATCases sequences had an overall 44% amino acid similarity, equally distributed between the two domains of the protein. Allowing for conservative substitutions, there was a similarity of 65% between the two enzymes. In contrast, there was only 34% positional similarity between the *E. coli* and *Bacillus subtilis* enzymes, 31% between the reported sequences for *Drosophila* and *E. coli*, and 45% between the hamster and *Drosophila* enzymes. It should be noted that 5–6 nucleotide changes (insertions or deletions) within the *Drosophila* sequence would significantly improve the positional similarity with hamster to over 60%. Even without changes in the *Drosophila* sequence, there was 15% positional similarity between the four enzymes. In the hamster and *B. subtilis* enzymes, all 12 of the active site residues that have been directly implicated in *E. coli* in the binding of the bisubstrate analogue N-phosphonacetyl-L-aspartate (Krause 1985; Volz et al. 1986; Krause et al. 1987) are conserved. The reported *Drosophila* sequence did not contain 2 of the 12 active site residues; however, a simple +1 insertion/deletion would restore those residues (Table 1). The predicted secondary structure for the hamster ATCase is very similar to the actual *E. coli* catalytic monomer secondary structure as determined by x-ray crystallography (Fig. 3). Using the Chou–Fasman algorithm as modified by Nishikawa (1983), a substantial portion of the predicted secondary structures in the hamster enzyme was positionally conserved relative to the *E. coli* enzyme, particularly within the polar domain of both enzymes. In addition, both of the enzymes were recognized by Western blots with polyclonal antibody against the *E. coli* ATCase holoenzyme (J.G. Major and M.E. Wales, unpublished).

A recombinant plasmid, pHEc270, was constructed to form a hybrid cistron containing the promoter-proximal half of the hamster cDNA gene and the distal portion of the *E. coli* gene. This construction also included the *pyrI* cistron that encodes the bacterial regulatory polypeptide. A conserved *BstE*II restriction endonuclease site (GGTGACC), located near the *E. coli* sequence encoding GLY161–ASP162, provided a convenient site for the desired in-frame genetic fusion. A 1.5-kb restriction fragment (*Pst*I:*BstE*II) coding for the polar domain of the hamster ATCase from pAL4 was ligated into a 5.1-kb fragment of pPBh105 restricted with the same two enzymes (Fig. 4). The hybrid plasmid was transformed into HB101-4442, a *pyrB* auxotroph of *E. coli* and *pyrB*$^+$, ampicillin-resistant colonies were

```
          {— β-GALACTOSIDASE------} {------------ CAD  LINKER  ------------------------------
ACA GCT ATG ACC ATG ATT ACG GAT TCA CTG GTG GGG ACC CCC GAC GGT CCT GCT ACC CTG CAC CGC CGC CCT AGA
            Thr Met Ile Thr Asp Ser Leu Val Gly Thr Pro Asp Gly Pro Ala Thr Leu His Arg Arg Pro Arg

----------------------------------------------------------------------------------} {--ATCASE--➡
CAG GCA TCA CCT CAG AAC CTG GGC TCT TCT GGC CTA CTG CAC CCA CAG ACT TCA CCC CTG CTG CAC TCC TTA GTG   5
Gln Ala Ser Pro Gln Asn Leu Gly Ser Ser Gly Leu Leu His Pro Gln Thr Ser Pro Leu Leu His Ser Leu Val

GGC CAA CAC ATC CTG TCT GTC AAG CAG TTC ACT AAG GAT CAG ATG TCT CAT CTG TTC AAC GTC GCG CAC ACA CTA  30
Gly Gln His Ile Leu Ser Val Lys Gln Phe Thr Lys Asp Gln MET Ser His Leu Phe Asn Val Ala His Thr Leu

CGG ATG ATG GTG CAG AAA GAG CGG AGC CTT GAC ATC CTA AAG GGC AAG GTC ATG GCC TCC ATG TTC TAC GAG GTG  55
Arg MET MET Val Gln Lys Glu Arg Ser Leu Asp Ile Leu Lys Gly Lys Val MET Ala Ser MET Phe Tyr Glu Val

AGC ACC CGC ACC AGT AGC TCC TTT GCA GCA GCC ATG GCC CGG CTC GGG GGC GCT GTC CTC AGC TTT TCA GAA GCC  80
Ser Thr Arg Thr Ser Ser Ser Phe Ala Ala Ala MET Ala Arg Leu Gly Gly Ala Val Leu Ser Phe Ser Glu Ala

ACG TCC TCC GTC CAG AAG GGG GAA TCC CTT GCC GAC TCT GTG CAG ACC ATG AGT TCG TAC GCT GAT GTC GTT GTG 105
Thr Ser Ser Val Gln Lys Gly Glu Ser Leu Ala Asp Ser Val Gln Thr MET Ser Ser Tyr Ala Asp Val Val Val

CTC CGG CAC CCT CAG CCT GGA GCT GTG GAG CTG GCA GCC AAA CAC TGT CGC AGA CCA GTG ATC AAT GCT GGG GAT 130
Leu Arg His Pro Gln Pro Gly Ala Val Glu Leu Ala Ala Lys His Cys Arg Arg Pro Val Ile Asn Ala Gly Asp

GGA GTC GGA GAG CAC CCT ACT CAG GCC CTG CTG GAC ATC TTC ACT ATC CGG GAA GAG CTG GGG ACT GTC AAT GGC 155
Gly Val Gly Glu His Pro Thr Gln Ala Leu Leu Asp Ile Phe Thr Ile Arg Glu Glu Leu Gly Thr Val Asn Gly

ATG ACG ATC ACC ATG GTA GGT GAC CTG AAG CAT GGG CGC ACA GTG CAC TCC CTG GCC TGC CTG CTC ACC CAG TAC 180
MET Thr Ile Thr MET Val Gly Asp Leu Lys His Gly Arg Thr Val His Ser Leu Ala Cys Leu Leu Thr Gln Tyr

CGT GTG AGC CTA CGC TAC GTG GCA CCT CCC AGC CTG CGC ATG CCA CCT AGC GTG TGG GAC TTT GTG GCT TCC CGG 205
Arg Val Ser Leu Arg Tyr Val Ala Pro Pro Ser Leu Arg MET Pro Pro Ser Val Trp Asp Phe Val Ala Ser Arg

GGC ACC AAA CAG GAG GAG TTT GAG AGC ATT GAG GAG GCG CTG CCC GAC ACG GAC GTG CTC TAC ATG ACT CGC ATC 230
Gly Thr Lys Gln Glu Glu Phe Glu Ser Ile Glu Glu Ala Leu Pro Asp Thr Asp Val Leu Tyr MET Thr Arg Ile

CAG AAA GAG CGA TTT GGC TCC ACC CAG GAA TAC GAA GCT TGC TTT GGT CAG TTC ATC CTC ACT CCC CAC ATC ATG 255
Gln Lys Glu Arg Phe Gly Ser Thr Gln Glu Tyr Glu Ala Cys Phe Gly Gln Phe Ile Leu Thr Pro His Ile MET

ACC CGG GCC AAG AAG AAG ATG GTG GTG ATG CAT CCG ATG CCC CGA GTC AAT GAG ATA AGC GTG GAG GTG GAC TCA 280
Thr Arg Ala Lys Lys Lys MET Val Val MET His Pro MET Pro Arg Val Asn Glu Ile Ser Val Glu Val Asp Ser

GAC CCC CGA GCA GCC TAC TTC CGC CAA GCT GAG AAC GGC ATG TAC ATC CGC ATG GCG CTT CTT GCC ACC GTG CTA 305
Asp Pro Arg Ala Ala Tyr Phe Arg Gln Ala Glu Asn Gly MET Tyr Ile Arg MET Ala Leu Leu Ala Thr Val Leu

GGC CGT TTC TAG GCC AGC CTC CTG CTG GCC TGG TCT
Gly Arg Phe  ·
```

Fig. 2. Nucleotide sequence of the hamster CAD ATCase domain and the predicted primary structure of the protein deduced from the DNA sequence of the cDNA. The numbering of the hamster ATCase sequence was based upon similarity alignment with the amino acid sequence of the catalytic subunit of *E. coli* ATCase (Hoover et al. 1983). With this alignment, the hamster ATCase coding region was comprised of 308 amino acids as compared with 310 amino acids for *E. coli* ATCase. The hamster ATCase protein included 7 amino acids of vector-encoded β-galactosidase (Silhavy et al. 1984) as well as 36 amino acids with no similarity to *E. coli* ATCase. This was presumed to be part of a protein linker bridging the ATCase domain to the rest of the CAD protein (Rumsby et al. 1984). The standard three-letter amino acid code was used with all sequences.

Table 1. Conservation of PALA binding residues between *E. coli*, hamster CAD, *Drosophila*, and *B. subtilis* ATCase sequences

E. coli	Function	Hamster	*Drosophila*	*Bacillus*
Ser 52	H bond to phosphonate	Ser 56	Ser 56	Ser 46
Thr 53	H bond to phosphonate	Thr 57	Thr 57	Thr 47
Arg 54	Salt link to phosphonate; H bond to phosphonate	Arg 58	Arg 58	Arg 48
Thr 55	H bond to phosphonate; H bond to carbonyl of peptide bond	Thr 59	Thr 59	Thr 49
Ser 80	H bond to phosphonate	Ser 82	Ser 82	Ser 73
Lys 84	H bond to phosphanate; H bond to α-carboxylate; H bond to β-carboxylate	Lys 86	Lys 86	Lys 76
Arg 105	Salt link to phosphonate; H bond to α-carboxylate; H bond to carbonyl of peptide bond	Arg 107	Arg 107	Arg 98
His 134	H bond to carbonyl of peptide bond	His 135	His 133	His 126
Arg 167	H bond to α-carboxylate; H bond to β-carboxylate	Arg 168	Arg 164	Arg 159
Arg 229	H bond to β-carboxylate	Arg 229	—[a]	Arg 210
Gln 231	H bond to β-carboxylate	Gln 231	—[a]	Gln 212
Leu 267	H bond to carbamyl nitrogen	Met 268	Ile 266	Ala 249

Residues implicated in substrate binding in *E. coli* ATCase (Krause 1985; Volz et al. 1986; Krause et al. 1987) are identified on the left. Proposed principal functions of the residues are listed in the center. Corresponding residues in the other sequences are indicated on the right

[a] Corresponding active site residues are missing from the reported *Drosophila* sequence of Freund and Jarry (1987). However, a careful examination of the nucleotide sequence (see text) reveals that a deletion of the first T of the CYS216 codon and subsequent insertion of an A immediately after the SER229 codon establishes seven new amino acid residue homologies to the *E.coli* sequence. Most importantly, this insertion/deletion creates both previously missing active site residues, arginine and glutamine, at amino acid positions 27 and 229, respectively, in the adjusted *Drosophila* sequence

Fig. 3. Amino acid sequence similarity between *E. coli* catalytic (Hoover et al. 1983), hamster CAD, *D. melanogaster* (Freund and Jarry 1987) and *B. subtilis* (Lerner and Switzer 1986) ATCases, and secondary structure comparison between *E. coli* and hamster ATCases. The alignment of *E. coli* and *B. subtilis* sequences was according to Lerner and Switzer (1986). The alignment of *E. coli* and *D. melanogaster* was according to Freund and Jarry (1987). The numbering of amino acids was based on the *E. coli* catalytic subunit sequence starting from the N-terminal alanine of the mature protein. Boxes indicate conservation of identical residues. The one-letter amino acid code is as follows: 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.

Reported secondary structure of the *E. coli* catalytic subunit (Honzatko et al. 1982) and Chou–Fasman prediction [Chou and Fasman (1974) as modified by Nishikawa (1983)] of the hamster CAD ATCase secondary structure are indicated immediately below the respective sequences using the following symbols: ~, alpha helix; >, beta sheet; – –, random coil. Amino acid residues common to all four sequences are in bold letters. Underlined residues indicate similarity to the respective amino acids within the *E. coli* sequence.

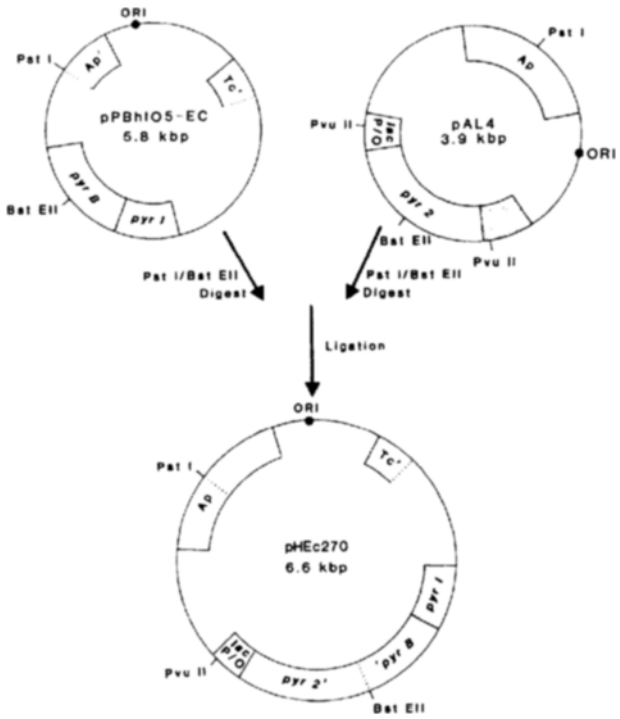

Fig. 4. Formation of pHEc270 containing a chimeric ATCase gene. Plasmid pPBh105 contains the intact *pyrBI* operon of *E. coli* (Roof et al. 1982). Plasmid pAL4 contains the productive cDNA clone of the 3′-fragment of the CAD mRNA that produces a functional ATCase under *lac* control (Maley and Davidson 1988). Both plasmids were restricted with *Bst*EII and *Pst*I and religated to form pHEc270 with chimeric ATCase under lac control.

screened for proper plasmid size and characterized by appropriate restriction endonuclease digestion following rapid plasmid isolation. DNA sequencing through the gene fusion junction (data not shown) verified that the hybrid protein resulted from an in-frame fusion between GLY161 and ASP162. In order to confirm the production and molecular weight of the chimeric polypeptide, *E. coli* maxicells (Sancar et al. 1979) were transformed with pAL4 and pHEc270. ^{35}S-methionine-labeled proteins were separated by electrophoresis under denaturing condition on a 6% SDS-acrylamide gel. The plasmid-borne gene produced a polypeptide of approximately 40 kd (Fig. 5), consistent with the expected molecular weight for the hybrid enzyme.

Escherichia coli HB101-4442, a *pyrB* auxotroph, was transformed with pHEc270 and grew slowly on minimal medium plates without pyrimidine supplement. Under similar conditions, without pHEc270, HB101-4442 cells did not show growth. The original and reisolated plasmids were transformed repeatedly into competent *pyrB* auxotrophs and shown to provide the essential ATCase activity as described. The enzyme was labile upon isolation from cell-free extracts and activity decreased over 2–3 days. The catalytic activity was not protected

by buffers containing serine protease inhibitors, such as phenylmethylsulfonyl fluoride, substrates, 20% glycerol, or dimethylsulfoxide. Other unsuccessful stabilizing procedures included gentler cell lysis protocols (Felix 1982) and expression in protease-deficient host strains (Chung and Goldberg 1981; Simon et al. 1983). The chimeric enzyme aggregated to molecular weights of 100,000–250,000 daltons, as demonstrated by Western blot analysis following Sephadex G-200 column chromatography. These molecular weight sizes are consistent with trimeric (C_3 = 120,000 daltons) and hexameric (C_6 = 240,000) aggregation. The hybrid demonstrated a K_m for aspartate within the range (2–10 mM) of the parental enzymes (data not shown). Even though the pHEc270 plasmid construction included the regulatory cistron of *E. coli*, the chimeric ATCase was not affected by millimolar concentrations of nucleotide effectors, UTP, CTP, or ATP, nor did it appear to associate with the regulatory subunit, as indicated by failure of the hybrid enzyme aggregates to cross-react with polyclonal antisera specific for the *E. coli* regulatory subunit (data not shown).

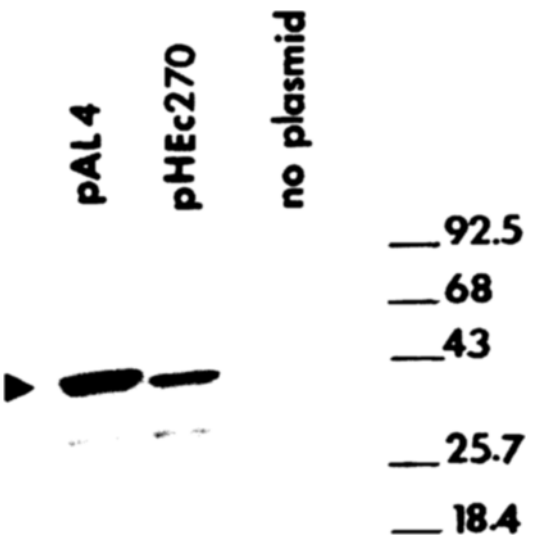

Fig. 5. Autoradiography of ^{35}S-methionine-labeled maxicells containing plasmid constructions. Labeled proteins were separated by electrophoresis under denaturing conditions on a 6% SDS-acrylamide gel. The contents of each lane are indicated on the figure. A lower band migrating below the heavy band in the first two lanes is the *bla* gene product or β-lactamase with a molecular weight of 28,000 daltons. The locations of unlabeled molecular weight markers are indicated on the figure.

Discussion

A comparison of *E. coli* and hamster ATCases has proven valuable in probing the structure–function similarities within evolutionarily divergent bacterial and hamster enzyme complexes. The two enzymes are functionally similar and both can satisfy pyrimidine auxotrophy in either prokaryotic or eukaryotic cell lines (Davidson and Niswander 1983; Ruiz and Wahl 1986). Despite wide divergence of quaternary structure and regulatory characteristics, the basic catalytic trimer motif is conserved in both enzymes. In addition, both active ATCase catalytic trimers have similar Stokes radii, sedimentation coefficients (Evans 1986), and cross-react with antisera to *E. coli* ATCase holoenzyme (M.E. Wales and J.R. Wild, unpublished). A primary sequence analysis of hamster and *E. coli* ATCases (Fig. 3) revealed similar coding regions of 308 and 310 amino acids, respectively. A partial sequence comprising residues 165–308 was reported earlier by Shigesada et al. (1985). The complete sequence reported here possesses some third base differences from this earlier sequence but shows no difference in deduced amino acid sequence for this region. The hamster protein possessed 44% overall amino acid positional similarity and 65% functional similarity with the *E. coli* sequence. A comparison of ATCases from *E. coli*, *B. subtilis*, *D. melanogaster*, and hamster revealed that the hamster appeared to be the most similar to the *E. coli* ATCase catalytic polypeptide sequence (44%) and most dissimilar to the *B. subtilis* enzyme (31%). The widespread conservation of the 12 active site residues in *E. coli*, *B. subtilis*, and hamster but absence of corresponding residues ARG229 and GLN231 in *Drosophila* prompted us to examine the *Drosophila* sequence more closely. If the reported nucleotide sequence is adjusted with a single base insertion/deletion surrounding the region CYS216–SER229, both active site residues at positions ARG227 and GLN229 within the *Drosophila* sequence would appear and five additional similarities would be created in this region. Indeed, a very small number of insertions/deletions within the *Drosophila* sequence could bring the similarity with hamster to over 60%, instead of the present similarity of 45%. These corrections may be appropriate and would seem logical when considering the comparatively short evolutionary period separating hamster and *Drosophila*, relative to the longer divergence time from a bacterial progenitor. The significant amount of primary structure and active site residue similarity between hamster and *E. coli* enzymes was accompanied by substantial conservation of predicted secondary structure, particularly within the polar domain of both enzymes. When considered with cross-reactivity of hamster ATCase with anti-*E. coli* ATCase antisera, these data argue for substantial conservation of ATCase structure and function.

In order to verify the actual significance of these conservations, a hybrid enzyme was constructed by an intragenic fusion that successfully united the polar domain (amino acids 1–162) of the hamster enzyme to the equatorial domain (amino acids 163–310) of the *E. coli* enzyme. This hybrid enzyme proved difficult to recover in a stable, enzymatically active form from cell-free extracts and there may be a number of contributing factors. The native *E. coli* enzyme is relatively stable in cell-free extracts; however, the pAL4 hamster ATCase (the catalytic superdomain produced from the CAD cDNA) is much more labile (J.G. Major and J.R. Wild, unpublished). This has also been reported for hamster ATCase superdomains enzymatically released from CAD protein (Grayson and Evans 1983). To limit possible proteolysis within cell free extracts, protease inhibitors were included; however, the stability of the protein was not dramatically improved. Other potential problems might involve the loss of some important amino acid interactions between the polar and equatorial domains in the hybrid formation. Therefore, the cell disruption and processing of the cell-free extract might have been too harsh for a structurally weakened hybrid enzyme and resulted in substantial loss of activity. Instabilities with artificially created fusion proteins have been observed with other proteins, such as β-galactosidase, amino-transferases, or tryptophan synthase fusions (Kania and Mueller-Hill 1980; Swamy and Goldberg 1981).

Although it was not possible to thoroughly analyze the hybrid kinetically in purified enzyme preparations, substrate saturation studies indicated that the hybrid enzyme exhibited a K_m for aspartate within the range of the parental *E. coli* catalytic trimer and hamster trimeric ATCases. Because the active hybrid enzyme must be able to bind carbamoyl phosphate for catalysis, the *E. coli* and the hamster ATCases must have functionally similar structural domains. Evans and coworkers (Evans 1986) have shown that extensive proteolytic digestion of the 40-kd CAD fragment containing active ATCase produced distinct 29-kd and 11-kd protein subspecies that were inactive. Similar work in progress with the pAL4 and *E. coli* ATCases in our laboratory has also indicated the presence of discrete, similarly sized domains within both enzymes. The hybrid catalytic species did not appear to associate with the bacterial regulatory subunit and exhibited native molecular weight species 100,000–250,000 daltons that suggested trimeric and hexameric aggregates of catalytic protomers. These structures are more reminiscent of the type of aggregation patterns found in the hamster enzyme than its bacterial counterpart. The amino terminal half of the hamster ATCase including the linker region connecting the ATCase to the rest of CAD has been implicated in the novel oligomerization of the CAD protein (Davidson and Patterson 1979; Davidson et al. 1979) and the hybrid ATCase aggregation pattern observed in these studies supports that contention. The successful formation of the hamster–bacterial fusion ATCase verified the structural and functional conservation of ATCase catalytic domains between extremely divergent, isofunctional enzymes.

Many other hybrid proteins have been formed utilizing fusions with β-galactosidase or alkaline phosphatase and have supported the concept that diverse domains may fold independently to form functional proteins (Kania and Mueller-Hill 1980; Hoffman and Wright 1985). All of these observations are consistent with the concept of modular protein evolution in which genetic cassettes corresponding to functionally discrete protein domains could randomly assort to provide new hybrid and perhaps functional proteins (Gilbert et al. 1986). Support for this concept of interchangeable genetic modules has been extended by the observations that exons often code for structurally and/or functionally distinct domains or secondary structural units (Rashin 1981; Blake 1983; Traut 1986). The ongoing sequencing of the human CAD gene has revealed a number of introns within the portion of the CAD sequence encoding the ATCase superdomain; none of the introns has interrupted the coding sequences of the highly conserved, secondary, and supersecondary structural units of the protein (J.N. Davidson, unpublished). Recent studies in our laboratory have succeeded in forming intragenic fusions between other diverged ATCases or ornithine transcarbamoylase catalytic domains to form functional hybrid bacterial ATCases (Houghton 1986; Kedzie 1987). In addition, studies in other laboratories have resulted in the formation of hybrid enzymes in various bacterial tryptophan synthases (Schneider et al. 1981) as well as yeast and human glycerol kinases (Mas et al. 1986). The research reported in this paper demonstrates a much more dramatic evolutionary conservation of structural and catalytic components within a superdomain of a mammalian multifunctional protein and a bacterial counterpart. Future research along this line may help to elucidate functional consequences of specific regions of sequence divergence and similarity.

Acknowledgments. This research has been supported by research grants to J.R.W. by the National Institutes of Health (NIGMS, GM33191), the National Science Foundation (DMB-8703732), and the Texas A&M University System; and to J.N.D. from the National Institutes of Health (NIGMS, GM34336).

References

Adair LB, Jones ME (1972) Purification and characteristics of aspartate transcarbamylase from *Pseudomonas fluorescens*. J Biol Chem 247:2308–2315

Bethell MA, Smith KE, White JS, Jones ME (1968) Carbamyl phosphate: an allosteric substrate for aspartate transcarbamylase of *Escherichia coli*. Proc Natl Acad Sci USA 60:1442–1449

Blake C (1983) Exons and the evolution of proteins. Trends Biochem Sci 8:11–13

Brabson JS, Maurizi MR, Switzer RL (1985) Aspartate transcarbamylase from *Bacillus subtilis*. Methods Enzymol 113: 627–635

Carrey EA, Hardie DG (1988) Mapping of catalytic domains and phosphorylation sites in the multifunctional pyrimidine-biosynthetic protein CAD. Eur J Biochem 171:583–588

Casadaban MJ, Cohen SN (1979) Lactose genes fused to exogenous promoter in one step using Mu-lac bacteriophage. *In vivo* probe for transcription control sequences. Proc Natl Acad Sci USA 76:4530–4540

Chang T, Jones ME (1974) Aspartate transcarbamylase from *Streptococcus faecalis*. Purification, properties and nature of an allosteric activator site. Biochemistry 13:629–637

Chen EJ, Seeburg PH (1985) Supercoil sequencing—a fast and simple method for sequencing plasmid DNA. DNA 4:165–170

Chou PY, Fasman GD (1974) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry 13:211–245

Chung CH, Goldberg AL (1981) The product of the *lon* (*capR*) gene in *Escherichia coli* is the ATP-dependent protease, protease La. Proc Natl Acad Sci USA 78:4931–4935

Coleman PF, Suttle DP, Stark GR (1977) Purification from hamster cells of the multifunctional protein that initiates *de novo* synthesis of pyrimidine nucleotides. J Biol Chem 252: 6379–6385

Davidson JN, Niswander LA (1983) Partial cDNA sequence to a hamster gene corrects defect in *Escherichia coli pyr B* mutant. Proc Natl Acad Sci USA 80:6897–6901

Davidson JN, Patterson D (1979) Alteration in structure of multifunctional protein from Chinese hamster ovary cells defective in pyrimidine biosynthesis. Proc Natl Acad Sci USA 76:1731–1735

Davidson JN, Patterson D (1984) Genes encoding multifunctional proteins—the equivalent of bacterial operons. DNA 3:78

Davidson JN, Carnright DV, Patterson D (1979) Biochemical genetic analysis of pyrimidine biosynthesis in mammalian cell: III. Association of carbamyl phosphate synthtase, aspartate transcarbamoylase, and dihydroorotase in mutants of cultured Chinese hamster cells. Somatic Cell Genet 5:175–191

Davidson JN, Rumsby PC, Tamaren J (1981) Organization of a multifunctional protein in pyrimidine biosynthesis. J Biol Chem 256:5220–5225

Evans DR (1986) CAD, a chimeric protein that initiates *de novo* pyrimidine biosynthesis in higher eucaryotes. In: Hardie DG, Coggins JR (eds) Multidomain proteins—structure and evolution. Elsevier, Amsterdam, pp 283–331

Felix H (1982) Permeabilized cells. Anal Biochem 120:211–234

Foltermann KF, Shanley MS, Wild JR (1984) Assembly of the aspartate transcarbamoylase holoenzyme from transcriptionally independent catalytic and regulatory cistrons. J Bacteriol 157:891–898

Foltermann KF, Wild JR, Zink DL, O'Donovan GA (1981) Regulatory variance of aspartate transcarbamoylase among strains of *Yersinia enterocolitica* and *Yersinia enterocolitica*-like organisms. Curr Microbiol 6:43–47

Foltermann KF, Beck DA, Wild JR (1986) *In vivo* formation of hybrid aspartate transcarbamoylase from native subunits of divergent members of the family *Enterobacteriaceae*. J Bacteriol 167:285–290

Freund JN, Jarry BP (1987) The *rudimentary* gene of *Drosophila melanogaster* encodes four enzymic functions. J Mol Biol 193:1–13

Gerhart JC, Schachman HK (1965) Distinct subunits for the regulation and catalytic activity of aspartate transcarbamoylase. Biochemistry 4:1054–1062

Gilbert W, Marchionni M, McKnight G (1986) On the antiquity of introns. Nature 271:501

Grayson DR, Evans DR (1983) The isolation and characterization of the aspartate transcarbamoylase domain of the multifunctional protein, CAD. J Biol Chem 258:4123–4129

Grayson DR, Lee L, Evans DR (1985) Immunochemical analysis of the domain structure of CAD, the multifunctional protein that initiates pyrimidine biosynthesis in mammalian cells. J Biol Chem 260:15840–15849

Hoffman CS, Wright A (1985) Fusions of secreted proteins to alkaline phosphatase: an approach for studying secretion. Proc Natl Acad Sci USA 82:5107–5111

Honzatko RB, Lipscomb WN (1982) Interactions of phosphate ligands of *Escherichia coli* aspartate carbamoyltransferase in the crystalline state. J Mol Biol 160:264–286

Honzatko RB, Crawford JL, Monaco HL, Ladner JE, Edwards BFP, Evans DR, Warren SF, Wiley DC, Ladner RC, Lipscomb WN (1982) Crystal and molecular structures of native and ATP-liganded aspartate carbamoyltransferase from *Escherichia coli*. J Mol Biol 160:219–263

Hoover TA, Roof WD, Foltermann KF, O'Donovan GA, Bencini DA, Wild JR (1983) Nucleotide sequence of the structural gene (*pyrB*) that encodes the catalytic polypeptide of aspartate transcarbamoylase of *Escherichia coli*. Proc Natl Acad Sci USA 80:2462–2466

Houghton JE (1986) Structural and functional comparisons between OTCase and ATCase, leading to the formation and characterization of an active "domain fusion" between *argI* (OTCase) and *pyrB* (ATCase). Dissertation, Texas A&M University, College Station TX

Jones ME (1980) Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. Annu Rev Biochem 49:253–279

Kania J, Mueller-Hill B (1980) Artificial protein chimeras. In: Bisswanger H, Schmincke-Ott E (eds) Multifunctional proteins. Wiley, New York, pp 31–47

Kedzie KM (1987) Characterization of the *pyrB* gene of *Serratia marcescens* and hybrid gene formation with the *pyrB* gene of *Escherichia coli*, leading to the production of chimeric ATCases. Dissertation, Texas A&M University, College Station TX

Keppler D, Holstege A (1982) Pyrimidine nucleotide metabolism and its compartmentation. In: Sies H (ed) Metabolic compartmentation. Academic Press, London, pp 149–203

Krause KL, Volz KW, Lipscomb WN (1987) 2.5 A structure of aspartate carbamoyltransferase complexed with the bisubstrate analogue N-(phosphonacetyl)-L-aspartate. J Mol Biol 193:527–553

Lerner CG, Switzer RL (1986) Cloning and structure of the *Bacillus subtilis* aspartate transcarbamoylase gene (*pyrB*). J Biol Chem 261:11156–11165

Lovatt CJ, Cheng AH (1984) Aspartate transcarbamoylase. Site of end product inhibition of the orotate pathway in intact cells of *Cucurbita pepo*. Plant Physiol 75:511–515

Lue PF, Kaplan JG (1969) The aspartate transcarbamoylase and carbamoyl phosphate synthetase of yeast: a multifunctional enzyme complex. Biochem Biophys Res Commun 34: 426–433

Makoff AJ, Radford A (1978) Genetics and biochemistry of carbamoyl phosphate biosynthesis and its utilization in the pyrimidine biosynthetic pathway. Microbiol Rev 42:307–328

Maley JA, Davidson JN (1988) The aspartate transcarbamoylase domain of a mammalian multifunctional protein expressed as an independent enzyme in *Escherichia coli*. Mol Gen Genet 213:278–284

Maley JA, Niswander LA, Flower AM, Davidson JN (1984) Dissecting the organization of the hamster CAD gene by expression of cDNA pieces in *E. coli* mutants. DNA 3:93

Mally MI, Grayson DR, Evans DR (1981) Controlled proteolysis of the multifunctional protein that initiates pyrimidine biosynthesis in mammalian cells: evidence for discrete domains. Proc Natl Acad Sci USA 78:6647–6651

Mas MT, Chen CY, Hitzeman RA, Riggs AD (1986) Active human–yeast chimeric phosphoglycerate kinases engineered by domain interchange. Science 233:788–790

Messing JRC, Seeburg PH (1981) A system for shot-gun DNA sequencing. Nucleic Acids Res 9:309–321

Monaco HL, Crawford JL, Lipscomb WN (1978) Three-dimensional structure of aspartate carbamyltransferase from *Escherichia coli* and of its complex with cytidine triphosphate. Proc Natl Acad Sci USA 75:5276–5280

Nishikawa K (1983) Assessment of secondary structure prediction of protein comparison of computerized Chou–Fasman method with others. Biochim Biophys Acta 748:285–299

Ong BL, Jackson JF (1972) Pyrimidine nucleotide biosynthesis in *Phaseolus aureus*. Biochem J 129:583–593

Prescott LM, Jones ME (1969) Modified methods for the determination of carbamyl phosphate. Analyt Biochem 32:408–419

Rashin AA (1981) Location of domains in globular proteins. Nature 291:85–87

Roof WD, Foltermann KF, Wild JR (1982) The *pyrB* operon of *E. coli* possesses a *rho* independent attenuator sequence. Mol Gen Genet 187:391–400

Ruiz JC, Wahl GM (1986) *Escherichia coli* aspartate transcarbamoylase: a novel marker for studies of gene amplification and expression in mammalian cells. Mol Cell Biol 6:3050–3058

Rumsby PC, Campbell PC, Niswander LA, Davidson JN (1984) Organization of a multifunctional protein in pyrimidine biosynthesis. Biochem J 217:435–440

Sancar A, Hack AM, Rupp WD (1979) Simple method for identification of plasmid-coded proteins. J Bacteriol 137:692–693

Sanger FS, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schneider WP, Nichols BP, Yanofsky C (1981) Procedure for production of hybrid genes and proteins and its use in assessing significance of amino acid differences in homologous tryptophan synthetase alpha polypeptides. Proc Natl Acad Sci USA 78:2169–2173

Shigesada K, Stark GR, Maley JA, Niswander LA, Davidson JN (1985) Construction of a cDNA to the hamster CAD gene and its application toward defining the domain for aspartate transcarbamoylase. Mol Cell Biol 5:1735–1742

Silhavy TS, Berman ML, Enquist LW (eds) (1984) Experiments with gene fusions. Cold Spring Harbor Press, Cold Spring Harbor NY, pp 273–281

Simon LD, Randolph B, Irwin N, Binkowski G (1983) Stabilization of proteins by a bacteriophage T4 gene cloned in *Escherichia coli*. Proc Natl Acad Sci USA 80:2059–2062

Stark GS (1977) Multifunctional proteins: one gene—more than one enzyme. Trends Biochem Sci 2:64–66

Swamy KHS, Goldberg AL (1981) *E. coli* contains eight soluble proteolytic activities, one being ATP dependent. Nature 292: 652–654

Traut TW (1986) Are proteins made of modules? Mol Cell Biochem 70:3–10

Volz KW, Krause KL, Lipscomb WN (1986) The binding of N-(phosphonacetyl)-L-aspartate to aspartate carbamoyltransferase of *Escherichia coli*. Biochem Biophys Res Commun 136:822–836

Wild JR, Foltermann KF, O'Donovan GA (1980) Regulatory divergence of aspartate transcarbamoylases within the Enterobacteriaceae. Arch Biochem Biophys 201:506–517

William LG, Bernhardt S, Davis RH (1970) Copurification of pyrimidine specific carbamyl phosphate synthetase and aspartate transcarbamylase of *Neurospora crassa*. Biochemistry 9:4329–4335

Zinder ND, Boeke JD (1982) The filamentous phage as vectors for recombinant DNA—a review. Gene 19:1–10

Received June 1, 1988/Revised August 3, 1988