MOLECULAR AND CELLULAR BIOLOGY, July 1985, p. 1735–1742
0270-7306/85/071735-08$02.00/0

Vol. 5, No. 7

Construction of a cDNA to the Hamster CAD Gene and Its Application Toward Defining the Domain for Aspartate Transcarbamylase

KATSUYA SHIGESADA,[1,2] GEORGE R. STARK,[1,3] JULIE A. MALEY,[4] LEE A. NISWANDER,[5] AND JEFFREY N. DAVIDSON[5,6]*

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305[1]; Institute for Virus Research, Kyoto University, Sakyo-Ku, Japan[2]; Imperial Cancer Research Fund, London, England[3]; Department of Biochemistry, Biophysics and Genetics, University of Colorado Health Sciences Center,[4] and Eleanor Roosevelt Institute for Cancer Research, Denver, Colorado 80262[5]; and Department of Medical Microbiology and Immunology, University of Kentucky Medical Center, Lexington, Kentucky 40536[6]

Received 8 January 1985/Accepted 10 April 1985

cDNA complementary to hamster mRNA encoding the CAD protein, a multifunctional protein which carries the first three enzymes of pyrimidine biosynthesis, was constructed. The longest of these recombinants (pCAD142) covers 82% of the 7.9-kilobase mRNA. Portions of the cDNA were excised and replaced by a *lac* promoter-operator-initiation codon segment. The resultant plasmids were transfected into an *Escherichia coli* mutant defective in aspartate transcarbamylase, the second enzyme of the pathway. Complementation of the bacterial defect was observed with as little as 2.2 kilobases of cDNA sequence, corresponding to the 3′ region of the mRNA. DNA sequencing in this region of the hamster cDNA reveals stretches which are highly homologous to the *E. coli* gene for the catalytic subunit of aspartate transcarbamylase; other stretches show no homology. The highly conserved regions probably reflect areas of protein structure critical to catalysis, while the nonconserved regions may reflect differences between the quaternary structures of *E. coli* and mammalian aspartate transcarbamylases, one such difference being that the bacterial enzyme in its native form is allosterically regulated and the mammalian enzyme is not.

The six-step pathway of pyrimidine biosynthesis is common to organisms ranging from bacteria to mammals. In *Escherichia coli*, each step is catalyzed by a separate protein or protein complex, and aspartate transcarbamylase (ATCase; EC 2.1.3.2), the second enzyme, is the major point of allosteric control of the pathway (24). In mammals, the six enzymes are associated with only three proteins, and carbamoylphosphate synthetase (CPSase; EC 6.3.5.5), the first rather than the second enzyme of the pathway, is affected through allosteric regulation (20). The first three enzymes, i.e., CPSase, ATCase, and dihydroorotase (DHOase; EC 3.5.2.3), are found on a single trifunctional protein designated CAD. The CAD protein of hamster occurs as an oligomer of identical monomers of M_r 220,000 (8, 10, 25). Within each monomer, the protein appears to be organized into a series of independent functional domains separated from one another by proteolysis-sensitive peptide bridges (11, 25, 39). Each of the three enzyme activities is associated with a different domain.

The CAD monomer of Syrian hamsters is encoded by an mRNA of 7.9 kilobases (kb; 45). The gene, on the other hand, spans more than 25 kb and is interrupted by over 30 introns (35). Previously, Wahl and coworkers (45) constructed a plasmid clone (pCAD41) containing a 2.3-kb insert complementary to the 3′ region of CAD mRNA. However, for purposes of analyzing the organization of the CAD gene, a more complete CAD cDNA would be invaluable. In this paper we describe the construction of much larger cDNA inserts by the highly efficient cloning technique developed by Okayama and Berg (32). This paper also shows how the CAD cDNA can be used in conjunction with a bacterial mutant defective in pyrimidine biosynthesis as a powerful expression system (9) that can define the portion of cDNA sequence required to code for the hamster ATCase domain. The validity of this approach is verified by sequencing a portion of the hamster CAD cDNA and comparing it with the coding sequence for the catalytic subunit of the *E. coli* ATCase.

MATERIALS AND METHODS

Cells and cell culture. The Syrian hamster cell line, 165-28, which overproduces CAD protein and mRNA 120-fold relative to normal cells (34, 45) was used as a source of mRNA. Cells were grown in Dulbecco modified Eagle medium supplemented with 10% (vol/vol) calf serum as described by Wahl et al. (45).

Isolation of CAD mRNA. Cells grown in roller bottles to approximately 75% confluence were fed with fresh medium 3 to 4 h before harvest. RNA was extracted from the cells with a guanidinium thiocyanate solution by a modification of the method of Chirgwin et al. (7). Poly(A)$^+$ RNA was isolated by chromatography on oligo(dT)-cellulose (3) and further fractionated on a 10 to 40% linear sucrose gradient as described by Wahl et al. (45). Fractions containing the 7.9-kb CAD mRNA (34) were pooled, and RNA was recovered by ethanol precipitation and stored in 50% ethanol at −20°C.

Synthesis and cloning of cDNA. Avian myoblastosis virus reverse transcriptase (kindly provided by Joseph Beard, Life Sciences, Inc.) was further purified by gel filtration on a Sephacryl S-200 (Pharmacia Fine Chemicals, Inc.) column to remove contaminating RNase activity (45). cDNA was synthesized and cloned as described by Okayama and Berg (32). The dT-tailed vector primer DNA and dG-tailed linker DNA were generously provided by H. Okayama. The plasmid

* Corresponding author.

cDNA-mRNA derivative, having been dC tailed and digested with *Hin*dIII, was subjected to centrifugation in a 5 to 20% sucrose gradient in a Beckman SW60 Ti rotor at 350,000 × *g* for 3 h. The fractions containing cDNA products larger than 3 kb were pooled and used in the remaining steps of cloning. The closed circular DNA formed was transfected into *E. coli* χ1776 by the protocol of Maniatis et al. (26) as described by Okayama and Berg (33). The transformed cells were pooled, added to 1 liter of χ-broth (26) containing 25 μg of ampicillin per ml, and grown at 37°C for 24 h. The stationary culture was adjusted to 7% dimethyl sulfoxide and stored in 2-ml samples at −70°C.

Screening of CAD cDNA clones. The cDNA library was screened by the Grunstein and Hogness colony hybridization method (17). The following three kinds of probes were used for screening (see Fig. 2A): a 1.2-kb *Pvu*II fragment from pCAD41 (45), a 0.9-kb *Pvu*II-*Bal*I fragment from pCAD103 isolated in this work, and a 2.2-kb *Bam*HI fragment from p206, a plasmid carrying a partial genomic CAD DNA sequence including sequences which correspond to the 5′-terminal region of CAD mRNA (35). These fragments were purified by agarose gel electrophoresis and labeled with ^{32}P by nick translation (36). The individual clones positively responding to these probes were grown in χ-broth containing 25 μg of ampicillin per ml, and their plasmid DNAs were isolated by the lysozyme-Triton procedure (12). The sizes of cDNA inserts were determined by cleavage of recombinant plasmids with *Hin*dIII and electrophoresis on a 1% agarose gel. Clones carrying inserts longer than 3 kb were examined for cleavage patterns with several other restriction endonucleases with conditions recommended by the supplier.

Northern blot analysis. Poly(A)$^+$ RNA 165-28 cells was glyoxylated, electrophoresed in 1% agarose gel, and transferred to diazobenzyloxymethyl-paper (1) or nitrocellulose paper (44). The filters were probed with the ^{32}P-labeled cDNA inserts cloned as described above.

Construction of pCL plasmids. The plasmid with the longest insert (pCAD142) was purified from *E. coli* HB101 (29) and digested to completion with *Eco*RI (Boehringer-Mannheim Biochemicals). This DNA was then partially digested with *Pvu*II (New England BioLabs, Inc.). The fragments of interest should have had a *Pvu*II blunt end and an *Eco*RI staggered end. The staggered end was converted to a blunt end by a fill-in reaction with the Klenow fragment of *E. coli* polymerase I, dATP, and TTP (26). After the DNA fragments were separated by agarose gel electrophoresis through an 0.5% gel, fragments in the range of 4.5 to 8.8 kb were electroeluted onto DE81 paper (13). DNA was eluted from the paper and further purified by passage through an Elutip column (Schleicher & Schuell, Inc.). The DNA was precipitated in ethanol and suspended in 20 mM Tris-hydrochloride (pH 8.0). A 203-base-pair (bp) *Hae*III fragment from *lac*UV5, containing the *E. coli lac* promoter-operator-initiation codon and $7^1/_3$ additional codons of the *lac*Z gene, was generously provided by the late John R. Sadler (University of Colorado Health Sciences Center). This blunt-end fragment, designated *lac*-203, was mixed with the cDNA fragments isolated as described above and ligated with T4 DNA ligase by the method of Maniatis et al. (26). The resulting "pCL" plasmids were transfected into an *E. coli pyr*B (ATCase$^-$) mutant as described previously (9). The *pyr*B strain 4517, kindly provided by Barbara Bachmann, is a uracil-requiring F$^-$ strain originally isolated by A. L. Taylor (AT2535). Ampicillin-resistant transformants were tested for the presence of the *lac*-203 sequence by growth on medium containing 5-bromo-4-chloro-3-indolyl-β-D-galactoside and for complementation of the bacterial ATCase defect by plating on minimal M9 medium (30) without uracil. Plasmid DNA was isolated from transformants by the method of Meagher et al. (29) and analyzed by restriction digestion and electrophoresis on horizontal agarose gels (0.75 to 1.5 g/100 ml) in Tris-acetate buffer (26).

DNA sequencing. Portions of the cDNA corresponding to the 3′ end of CAD mRNA were cloned into M13 phage vectors MP8 and MP9 (2, 18, 40) and sequenced by the Sanger dideoxy method (41) with the gradient gel and ^{35}S-labeling modifications described by Biggin et al. (5). Similar portions were sequenced by the Maxam-Gilbert protocol (27).

RESULTS

Synthesis of full-length CAD cDNA. To clone as much of the 7.9-kb CAD mRNA as possible, we directed our efforts toward optimizing the conditions for synthesis of full-length, single-stranded cDNA. First, CAD mRNA was prepared from a mutant hamster cell line (165-28) which has approximately 120 times the wild-type level of this RNA. To minimize degradation of RNA due to endogenous RNase and also to ensure effective removal of RNA-binding proteins, the cells were lysed with a potent protein-denaturing agent, guanidinium thiocyanate (7), instead of the nonionic detergent Nonidet P-40 used previously (45). Then the CAD mRNA was partially purified by sucrose gradient centrifugation. The resultant RNA preparation contained a substantial amount of intact 7.9-kb CAD mRNA, which was detectable

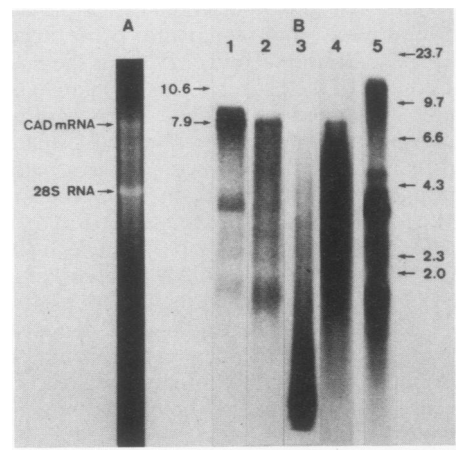

FIG. 1. Synthesis of a full-length cDNA to CAD mRNA. (A) Electrophoretic pattern of CAD mRNA used as template. RNA prepared from 165-28 cells by the method of Chirgwin et al. (7) and purified on a sucrose gradient was electrophoresed on methylmercury-agarose gel (1%) and stained with ethidium bromide as described by Bailey and Davidson (4). (B) Autoradiograph of ^{32}P-labeled cDNA products as analyzed on a 1% alkaline agarose gel (28). The reverse transcriptase reaction was performed in a 10-μl volume as described in the text, with the following variations. Reverse transcriptase: lanes 1, 3, 4, and 5, purified; lane 2, unpurified. Template: lanes 1, 2, 3, and 5, extracted with guanidium thiocyanate; lane 4, extracted with Nonidet P-40. Primer: lanes 1 through 4, oligo(dT); lane 5, dT-tailed vector DNA. dNTP concentration: lanes 1, 2, 4, and 5, 2 mM; lane 3, 0.2 mM. The numbers with arrows indicate the length in kbs of complete CAD transcripts and molecular size markers.

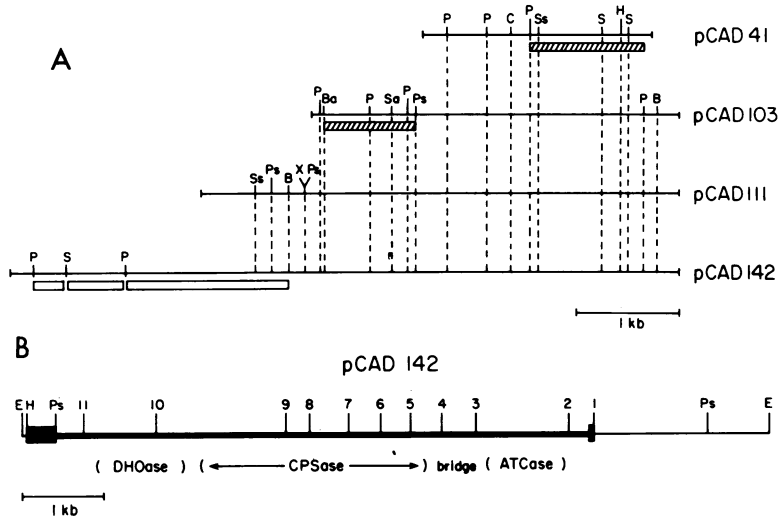

FIG. 2. (A) Restriction maps of the cDNA inserts of pCAD clones. The pCAD41 insert was previously isolated by Wahl et al. (45) and cloned into the *Pst*I site of pBR322 by way of dG-dC bridges. The cDNA inserts of pCAD103, pCAD111, and pCAD142 were primed from the dT tail of the Okayama and Berg vector as described in the text. Abbreviations and symbols: S, *Sma*I; P, *Pvu*II; B, *Bgl*II; H, *Hin*dIII; Ss, *Sst*I; C, *Cla*I; Ps, *Pst*I; Sa, *Sal*I; Ba, *Bal*I; X, *Xha*I; hatched boxes, probes for screening CAD cDNA clones; and open boxes, probes for Northern blot analysis. (B) Restriction map of pCAD142. Linear form generated by *Eco*RI digestion. Thick lines represent CAD cDNA, solid boxes represent the linkers derived from simian virus 40, and thin lines represent pBR322 DNA. *Pvu*II restriction sites are shown numbered from the 3′ [poly(A)] end of the cDNA. Abbreviations for other restriction sites are as in (A). Probable coding regions for enzyme domains are shown below the map.

as a prominent band on an agarose gel stained with ethidium bromide (Fig. 1A).

cDNA synthesis was carried out with avian myoblastosis virus reverse transcriptase that had been freed of RNase activity by gel filtration on a Sephacryl S-200 column as described by Wahl et al. (45). It has been noted that degradation of mRNA can occur during cDNA synthesis due to the RNase H activity inherently associated with reverse transcriptase, giving rise to anticomplementary DNA alongside uncompleted cDNA (31). However, the presence of high deoxynucleoside triphosphate (dNTP) concentration has been shown to effectively suppress RNase activity (21). Therefore, all four dNTPs were used in high concentrations (2 mM each) equimolar in total to that of magnesium ion (8 mM) as recommended by previous authors (32, 38). Under these conditions, full-length CAD cDNA of 7.9 kb was produced as a major fraction of all cDNA produced (Fig. 1B). The yield of full-sized products decreased drastically when the template was replaced with CAD mRNA prepared by previous methods (Fig. 1, lane B4; 45) or when the reaction conditions for reverse transcriptase were changed in any of the following ways: use of reverse transcriptase as supplied by the manufacturer without further purification (Fig. 1, lane B2), reduction of the concentration of dNTPs to 0.2 mM each (Fig. 1, lane B3), or increase in magnesium ion concentration.

Cloning and screening of CAD cDNA. We chose to construct the CAD cDNA library by the cloning procedure developed by Okayama and Berg (32), as this procedure allows the cloning of entire reverse transcripts with increased overall efficiency compared with that of the conventional procedure (14).

The first cDNA strand was synthesized with dT-tailed vector DNA (approximately 2.7 kb) as primer. The full-length cDNA product of expected size (7.9 plus 2.7 kb; total, 10.6 kb) was produced to an appreciable extent (Fig. 1, lane B5). Oligo(dC) tails were attached to the 3′ ends of the cDNA and vector DNA by the action of terminal transferase, and the dC tail on the vector DNA terminus was removed by cleavage at the unique *Hin*dIII restriction site near that end. At this stage, the vector cDNA-mRNA derivative was subjected to sucrose gradient centrifugation to remove short DNA fragments. The ensuing fractions enriched for large cDNA products were used for the subsequent steps as described by Okayama and Berg (32). The recombinant plasmids formed were transfected into χ1776 to yield a library containing approximately 2,000 independent transformants. This library was initially screened by colony hybridization with a ^{32}P-labeled probe from pCAD41, a previously isolated plasmid carrying a 2.3-kb sequence complementary to the 3′ region of CAD mRNA (Fig. 2A). Of 1,000 colonies screened, about 200 gave positive signals, among which a clone with an insert of 3.5 kb was found (pCAD103; Fig. 2A). From this clone, a new 0.9-kb probe complementary to a region further toward the 5′ end of the CAD mRNA (2.4 to 3.3 kb from the poly(A) tail) was prepared and used for screening the library a second time. About 300 of 10,000 screened colonies gave positive signals to the pCAD103 probe. In the hope of directly selecting clones that carry a full- or near-full-length CAD cDNA, a set of replicas of the colonies was screened with a probe prepared from p206, which contains a part of the genomic CAD DNA, including sequences encoding the 5′-terminal region of CAD mRNA (32). About 30 colonies gave weakly positive signals to this probe, but none of these colonies hybridized the labeled 0.9-kb probe derived from pCAD103.

Restriction map. Plasmid DNA was isolated from each clone that responded positively to the pCAD103-derived probe and was examined for the size and restriction pattern of its insert. Wahl et al. (45) have previously indicated that CAD cDNA has a *Hin*dIII site about 500 bp from the 3′ end. Because the cDNA sequence cloned by the Okayama and Berg method extends from the poly(dT) tail abutting the vector-primer DNA, all clones derived from CAD mRNA should be cleaved by *Hin*dIII at a minimum of two sites, one within the insert (Fig. 2A) and the other on the vector DNA

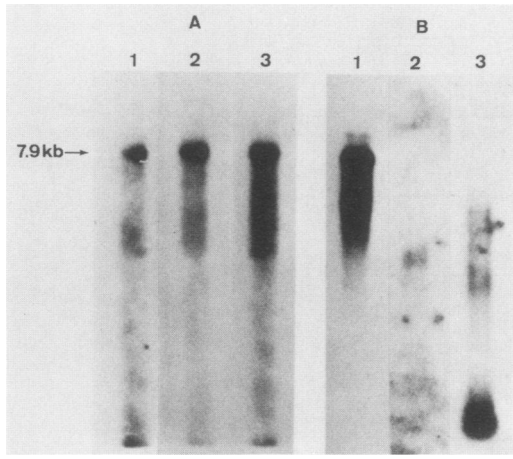

FIG. 3. Northern blot of poly(A)$^+$ RNA from 165-28 cells probed with cloned cDNA inserts. Poly(A)$^+$ RNA was electrophoresed on a 1% agarose gel after glyoxylation, transferred to nitrocellulose (A) or diazobenzyloxymethyl-paper (B), and hybridized with ^{32}P-labeled probes prepared as follows. (A) Segments of the cDNA insert in pCAD142 as indicated by open boxes in Fig. 2A were excised, purified by agarose gel electrophoresis, and labeled by nick translation. Lane 1, 0.3-kb *Pvu*II-*Sma*I fragment; lane 2, 0.5-kb *Sma*I-*Pvu*II fragment; lane 3, 1.2-kb *Pvu*II-*Bgl*II fragment. (B) The following plasmid DNA probes were generated by nick translation: lane 1, pCAD103; lanes 2 and 3, plasmids 6 and 16, respectively, selected by their hybridization to the p206-derived probe.

(Fig. 2B), to yield a pair of characteristic fragments. One fragment is expected to have a fixed length of about 2.8 kb (except for small variations in the dA-dT stretch), and the other fragment will vary in size depending on the extent of reverse transcription. A majority of the clones examined gave such patterns, and 20 of the clones had inserts longer than 3 kb. The longest insert obtained was 6.5 kb (pCAD142). Figure 2 shows a restriction map of pCAD142 and other representative clones. Their restriction maps overlap each other in a consistent manner as well as overlapping the map of the previously isolated clone pCAD41.

We also analyzed restriction maps of several clones that hybridized to the p206-derived probe. These clones had inserts ranging between 1 and 4 kb and showed various cleavage patterns, none of which overlapped that of pCAD142 (data not shown). These results raised doubts that these clones were actually derived from CAD mRNA (see below for further evidence).

Homology between cloned cDNA and CAD mRNA. To confirm that the pCAD clones carry cDNA copies of CAD mRNA, poly(A)$^+$ RNA from 165-28 cells was separated on an agarose gel, transferred to nitrocellulose or diazobenzyloxymethyl-paper, and hybridized with ^{32}P-labeled probes derived from pCAD142 and pCAD103 (Fig. 2A). Combined, these probes correspond to more than 90% of the 6.5-kb cDNA on pCAD142. All the probes clearly hybridized to the band of 7.9-kb CAD mRNA (Fig. 3, lanes A1 through 3 and B1), suggesting that the 6.5-kb insert is complementary to CAD mRNA over the entire length of the insert.

Although the restriction maps of clones screened with the p206-derived probe show no overlap of the map of pCAD142, it was possible that some of the clones were carrying the 5′ region of CAD mRNA not present in the 6.5-kb insert of pCAD142. To test this possibility, we prepared probes from four such clones and performed Northern blot analysis as described above. These probes hybridized to miscellaneous RNA components smaller than and apparently unrelated to CAD mRNA. The results obtained with two kinds of probes are shown in Fig. 3, lanes B2 and 3. These clones may represent RNA species which have some homology with intervening or repetitive sequences presumed to be present on p206 (35).

pCL plasmids. Recently, pCAD142 has been shown to complement the defect of *E. coli pyr*B (ATCase$^-$) mutants (9), indicating that ATCase activity is encoded by a partial CAD cDNA. Based on the sizes of proteolytic fragments (11, 16, 39) and amino acid composition of CAD (8), only 1 to 1.2 kb of the 6.5-kb cDNA insert in pCAD142 should be necessary to encode the ATCase domain of CAD. Because evidence (9) strongly suggests that the ATCase domain is the carboxyl-terminal domain of the CAD protein and is therefore encoded by the 3′ region of the mRNA, we undertook the deletion, by partial digestion with *Pvu*II, of pieces of cDNA from pCAD142 to create a series of plasmids having variously sized cDNA fragments corresponding to the 3′ end of the mRNA. In addition, the pBR322 sequences from the *Eco*RI site to the *Hin*dIII site (Fig. 2B) are thought to contain a portion of the *tet* promoter (43) and were removed in these constructions. The deleted segments were replaced by a *lac* promoter-operator-initiation codon segment (*lac*-203), and resultant plasmids were designated pCL. *E. coli pyr*B mutants transformed with these plasmids were selected for

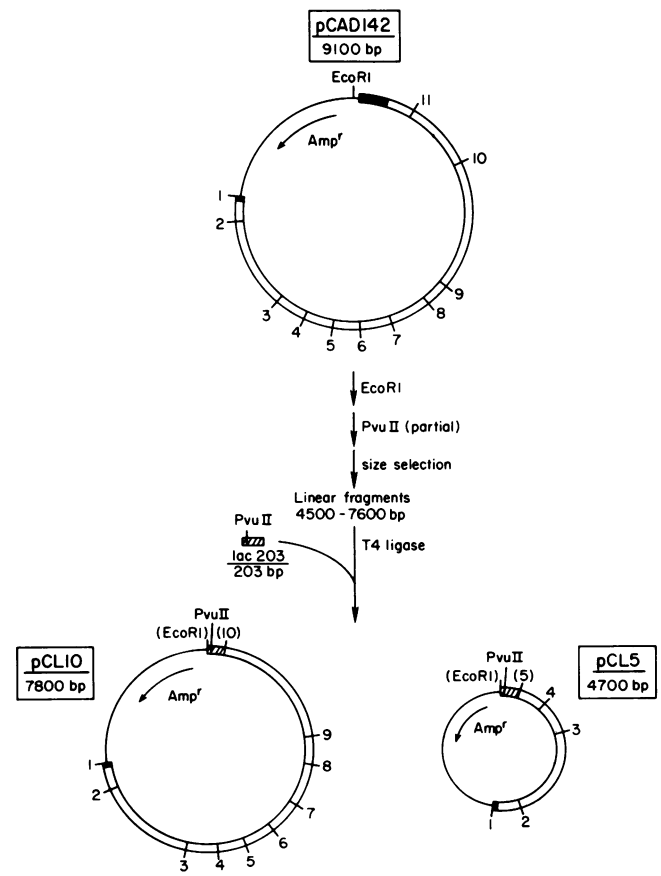

FIG. 4. Construction of pCL10 and pCL5. Symbols: closed boxes, simian virus 40 linker DNA; hatched box, *lac*-203 DNA; open box, hamster CAD cDNA; line, pBR322 DNA; Amp^r, ampicillin resistance. Numbered restriction sites are *Pvu*II sites; parentheses denote restriction sites destroyed in the constructions.

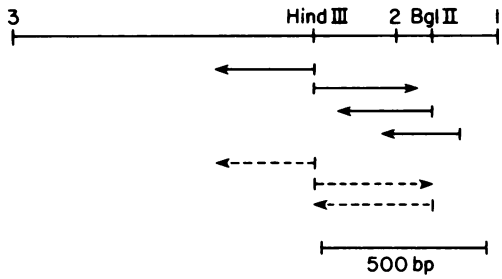

FIG. 5. Nucleotide sequencing strategy. Nucleotide sequence of subcloned fragments corresponding to the indicated segments of pCAD142 was determined by two methods. Solid lines indicate Maxam and Gilbert (27) sequencing of fragments derived from pCAD13 (35) or pCAD41. Dashed lines indicate fragments of pCAD142 subcloned into M13 and sequenced by the dideoxynucleotide method. *Pvu*II sites are numbered from the 3′ end of the cDNA insert as in Fig. 2B.

ampicillin resistance and tested for the presence of the *lac*-203 segment and uracil auxotrophy. Plasmid DNA was isolated from several clones which contained the *lac* operator and were able to grow in the absence of uracil. Restriction pattern analysis of plasmid DNA from such clones showed CAD inserts ranging from 2.2 to 5.3-kb in size. Plasmids representing the shortest and longest inserts found were designated pCL5 and pCL10, respectively (Fig. 4). pCL5 contains 2.2 kb of cDNA reaching from the 3′ end of the cDNA to the 5th *Pvu*II site from that end; pCL10 has 5.3 kb of cDNA reaching from the 3′ end to the 10th *Pvu*II site upstream. Both pCL5 and pCL10 are genetically active in complementing the defect in the *E. coli pyr*B mutant.

There is a *Pvu*II site located asymmetrically in the *lac*-203 segment, i.e., 17 bp from the 5′ end and 186 bp from the 3′ end. This cleavage site allowed us to determine the orientation of the *lac*-203 segment with respect to the cDNA sequence on the plasmid. *Pvu*II digestion and restriction

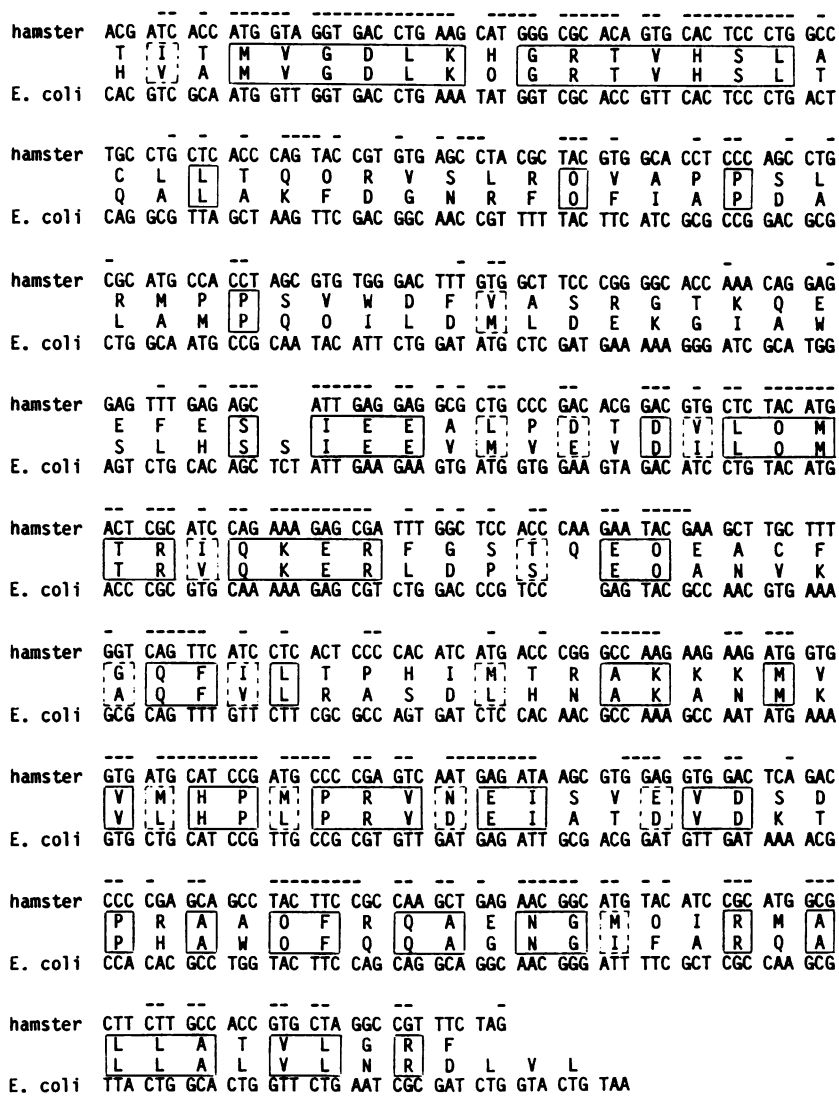

FIG. 6. Partial nucleotide sequence of pCAD142. Nucleotide sequence and deduced amino acid sequences from pCAD142 and sequences determined for *E. coli* ATCase catalytic subunit (19) are aligned for maximum homology. Dashes above nucleotides indicate positions of nucleotide identity between the sequences. Pairs of identical amino acids are enclosed by boxes with solid lines; pairs of amino acids with similar properties appear in boxes with dashed lines.

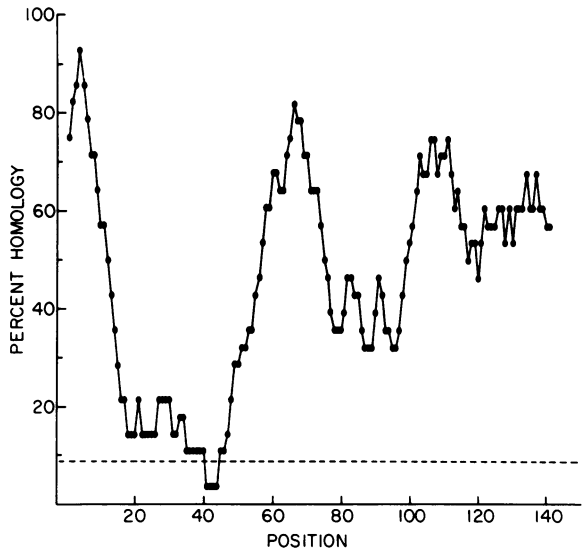

FIG. 7. Homology of deduced amino acid sequences for a portion of hamster ATCase and *E. coli* ATCase catalytic subunit. Homology was determined as follows. Positions at which the amino acids in the compared sequences are identical have a value of 1, amino acids with similar properties (leucine, isoleucine, valine, and methionine; lysine and arginine; glycine and alanine; threonine and serine; aspartate, asparagine, glutamate, and glutamine) have a value of 0.5. The sum of these values was determined for a window of 14 amino acids and converted to a percentage after being divided by 14. The percentage of homology was determined along the entire sequence by moving the 14-amino-acid window one position at a time.

fragment size analysis by agarose gel electrophoresis of pCL5 and pCL10 revealed that in each case, only one copy of the *lac*-203 segment is present, and the 186-bp fragment of the *lac*-203 segment is adjacent to the 5'-most *Pvu*II site of the cDNA fragment as diagrammed in Fig. 4 (data not shown). Other plasmids with multiple copies of the *lac*-203 segment were observed, but in every case where a plasmid could complement the *pyr*B mutant, at least one copy of the *lac*-203 segment was in the proper orientation for transcription of the CAD cDNA. Demonstration that the *lac*-203 segment had been inserted into pCL5 and pCL10 upstream from and in the same orientation as the CAD cDNA insert is consistent with its acting as the initiation site for the transcription and translation of the CAD sequences in these plasmids. Additional evidence comes from two facts. First, although pCAD142 does complement *E. coli pyr*B mutants (9), it does so only weakly; 3 days are required for large colonies to appear on medium without uracil. However, *pyr*B mutants containing pCL5 or pCL10 grow rapidly on medium without uracil, with very large colonies seen after 16 h. Second, pCAD41 (45), which contains more cDNA than pCL5 but lacks the *lac*-203 sequence, does not complement *E. coli pyr*B (J. N. Davidson and L. A. Niswander, unpublished results).

Nucleotide sequencing of pCAD142. The results with pCL5 indicate that the genetic information coding for ATCase lies near the 3' end of the cDNA. For this reason, the cDNA region between *Pvu*II sites 1 and 3 (Fig. 5) was sequenced.

Translating the nucleotide sequence into a polypeptide sequence yields an open reading frame of 153 amino acids (Fig. 6). Both the nucleotide and amino acid sequences align well with the carboxyl half of the catalytic subunit of *E. coli* ATCase. The overall nucleotide homology is 48%, while the amino acid homology is 42%. However, the homology is not evenly distributed throughout the sequence (Fig. 7). Instead, there are regions of high homology separated by regions with no or low homology. In this reading frame, the UAG termination codon is present precisely at the end of the ATCase coding domain. Approximately 330 nucleotides of 3' flanking sequence must separate the termination codon from the end of the mRNA. Besides locating the ATCase domain adjacent to the carboxyl end of the CAD protein, the sequencing data allows speculation regarding the locations of the other domains of the protein with respect to the cDNA sequence of pCAD142 (Fig. 2B, bottom).

If the nucleotide sequence data and the amino acid sequence deduced from those data are correct, then deletions at the 3' end of the cDNA in pCL5 should eliminate coding information essential for synthesizing an active ATCase. To test this prediction, a 354-bp *Hin*dIII-*Xba*I (a site located a few nucleotides 5' of the *Bgl*II site shown in Fig. 5) fragment was removed from pCL5. This removal should result in a protein lacking the final 67 amino acids shown in Fig. 6. When this deletion construct was transformed into *pyr*B cells, bacterial clones carrying the plasmid were selected. However, none of these could grow in the absence of uracil. Hence, the nucleotide sequence 3' to the *Hin*dIII site of pCL5 and found to be homologous to the *E. coli pyr*B gene does contain information critical to the synthesis of an active ATCase.

DISCUSSION

By using improvements in the preparation of cDNA, we have constructed a 6.5-kb cDNA complementary to the Syrian hamster mRNA coding for the CAD protein. This cDNA represents about 82% of the 7.9-kb CAD transcript. Construction of such cDNA clones was made possible primarily through the use of the cloning method of Okayama and Berg (32). Compared with other cDNA cloning procedures (14, 22), this method had two key features advantageous to the generation of large cDNA inserts. First, digestion with S1 nuclease is not needed. Second, the RNA replacement reaction proceeds distributively rather than processively over the length of the cDNA-mRNA hybrid. This reaction facilitates the completion of second-strand synthesis regardless of template length.

Preparation of mRNA and optimization of the reverse transcription reaction may also have contributed to our success in obtaining a much larger cDNA than the 2.3-kb cDNA obtained previously (45). Isolation of CAD mRNA with guanidinium thiocyanate (7) consistently gave far better results than did the extraction method which employs Nonidet P-40 for cell breakage and phenol for deproteinization. A possible contributing factor to this difference may be the extent of contamination with proteins bound to template RNA molecules, which could inhibit the elongation of reverse transcripts. Some RNA-bound proteins cannot be completely removed by phenol extraction alone, whereas guanidinium thiocyanate is known to be extremely powerful in denaturing proteins (7). Finally, the reverse transcription reaction was improved by the presence of high concentrations of dNTPs.

It would be interesting to compare the cDNA in pCAD142 to the Syrian hamster genomic sequence for CAD. Such a comparison might show just how much of the 25-kb genomic sequence is represented in this cDNA clone. However, the genomic clones isolated by Padgett et al. (35) have not yet been sufficiently characterized to allow significant conclusions to be drawn from such experiments.

We have previously shown that the plasmid with 6.5 kb of

CAD cDNA (pCAD142) can complement *E. coli pyr*B mutants (9). By partial deletions of pCAD142 and the insertion of a *lac* promoter-operator-initiation codon, the region essential to complementing the bacteria defective in ATCase has been narrowed to 2.2 kb of cDNA. The expression of ATCase from such a small portion (28%) of the 7.9-kb mRNA supports the notion that the ATCase domain is encoded by a cassette of genetic information which is independent of the information encoding the CPSase and DHOase domains. Digestion of CAD protein with proteases (11, 25) clearly demonstrated that this multifunctional protein is organized into independent functional domains, each with a single enzyme activity. The experiments presented here indicate that the organization of the CAD gene may reflect the domain structure of the protein. We hypothesize that each enzyme domain of CAD is encoded by a separate cassette of genetic information in the gene. Such cassettes may be capable of moving within the genome to form genes encoding multidomain, multifunctional proteins and hence represent a vital force in the evolutionary process. The one cassette-one domain hypothesis is particularly plausible in the case of CAD, which appears to have evolved from at least three separate genes in bacteria via a stepwise genetic fusion through a bifunctional fungal intermediate in which CPSase and ATCase are fused (23, 46), to the trifunctional situation seen in mammals and *Drosophila melanogaster* (6). A possible fourth domain may be the allosteric regulatory region which affects the CPSase activity.

Hoover et al. (19) have shown that the catalytic subunit of the *E. coli* ATCase is encoded by about 1 kb of DNA. If the hamster ATCase domain is homologous to the *E. coli* enzyme, then even less than the 2.2 kb of cDNA in pCL5 should be required to code for active ATCase. In fact, recent preliminary results (J. A. Maley, L. A. Niswander, A. M. Flower, and J. N. Davidson, DNA 3:93, 1984) have shown that less than 1,350 bp of hamster CAD cDNA is required to complement an *E. coli* mutant defective in ATCase. Furthermore, the DNA sequencing presented here has identified a region of hamster cDNA which shows strong homology to the *E. coli* sequence coding for the catalytic subunit of ATCase. As expected, when a portion of the sequenced region was deleted, the modified hamster cDNA yielded no active ATCase.

Two unexpected findings were made by comparing the hamster and *E. coli* sequences. First, some regions show an extremely high degree of homology; regions of such high homology are probably crucial to the function of the enzyme. A recent report with *E. coli* has in fact shown that the highest homology area shown in Fig. 6, where 13 of 14 amino acids are identical between *E. coli* and hamster, is critical to the binding of the substrate aspartate (37). Second, intermingled with the regions of high homology are regions which show almost no homology. One plausible explanation for regions with such low homology is the difference in regulation in the *E. coli* and hamster enzymes. The native form of the bacterial ATCase is an aggregate of regulatory and catalytic subunits, whereas the mammalian enzyme has no analogous regulatory subunit. Therefore, the regions of the *E. coli* catalytic subunit that are involved in communicating the allosteric response from the regulatory subunit would have no corresponding role in the hamster sequence and would thereby display the least conservation in nucleotide or amino acid sequence.

Finally, the pCL series of deletion plasmids, the DNA sequencing, and the removal of the *Hin*dIII-*Xba*I fragment from pCL5 have revealed that the DNA sequence coding for the ATCase domain is located at the far 3′ end of the gene, in fact right up against the translation termination codon. This position places this region within 330 bp of the 3′ end of the 7,900-bp CAD mRNA. When our previous data (9, 39) are related to the findings for the CAD locus in *Drosophila melanogaster* (15), the following order of enzyme domains is deduced: NH_2-DHOase-CPSase-ATCase-COOH. This order has recently been confirmed in *D. melanogaster* by Segraves et al. (42). Based on the active proteolytic fragments of CAD (11, 16, 25, 39) and the amino acid composition of CAD (8), the proposed sizes of coding regions for the CAD domains are 1.2 kb for DHOase, 3.3 kb for CPSase, and 1 to 1.2 kb for ATCase. In addition, Rumsby et al. (39) have estimated that the peptide bridge between the CPSase and ATCase domains is at least 100 amino acids (or about 0.3 kb) and the bridge between the DHOase and CPSase domains is smaller (perhaps no more than 0.2 kb). Therefore, the total cDNA required to code for CAD (including 3′ nontranslated sequence) should be about 6.5 kb, which means that pCAD142 with its 6.5-kb insert may contain the entire coding sequence for this multifunctional protein. Although the predicted organization of these coding blocks is shown at the bottom of Fig. 2B, the validity of the prediction awaits further experiments.

ACKNOWLEDGMENTS

We are grateful to H. Okayama and P. Berg for the kind gift of pBR-SV40 vector plasmids and for valuable advice on the application of their cloning method. We are also indebted to R. A. Padgett for his kind technical guidance in the preparation of CAD mRNA and to D. Patterson and G. Wahl for their many helpful comments and suggestions on this work.

This work was supported by Public Health Service grant GM32456 from the National Institutes of Health and American Cancer Society grant FRA242 to J.N.D.

LITERATURE CITED

1. **Alwine, J. C., D. J. Kemp, and G. R. Stark.** 1977. Methods for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U.S.A. **74**:5350–5354.
2. **Anderson, S., M. J. Gait, L. Mayol, and I. G. Young.** 1980. A short primer for sequencing DNA cloned in the single-stranded phage vector M13 mp2. Nucleic Acids Res. **8**:1731–1743.
3. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408–1412.
4. **Bailey, J. M., and N. Davidson.** 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal. Biochem. **70**:75–85.
5. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. U.S.A. **80**:3963–3965.
6. **Brothers, V. M., S. I. Tsubota, S. E. Germeradd, and J. W. Fristrom.** 1978. Rudimentary locus of *Drosophila melanogaster*: partial purification of a carbamylphosphate synthase-aspartate transcarbamylase-dihydroorotase complex. Biochem. Genet. **16**:321–332.
7. **Chirgwin, J. M., A. E. Pryzybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**: 5294–5299.
8. **Coleman, P. F., D. P. Suttle, and G. R. Stark.** 1977. Purification from hamster cells of the multifunctional protein that initiates *de novo* synthesis of pyrimidine nucleotides. J. Biol. Chem. **252**:6379–6385.
9. **Davidson, J. N., and L. A. Niswander.** 1983. Partial cDNA sequence to a hamster gene corrects defect in *Escherichia coli pyr*B mutant. Proc. Natl. Acad. Sci. U.S.A. **80**:6879–6901.

10. **Davidson, J. N., and D. Patterson.** 1979. Alteration in structure of multifunctional protein from Chinese hamster ovary cells defective in pyrimidine biosynthesis. Proc. Natl. Acad. Sci. U.S.A. **76**:1731–1735.
11. **Davidson, J. N., P. C. Rumsby, and J. Tamaren.** 1981. Organization of a multifunctional protein in pyrimidine biosynthesis. J. Biol. Chem. **256**:5220–5225.
12. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Isolation of plasmid and bacterial DNA, p. 116–127. *In* R. W. Davis (ed.), Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Dretzen, G., M. Bellard, P. Sassone-Corsi, and P. Chambon.** 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. **112**:295–298.
14. **Efstratiadis, A., and L. Villa-Komaroff.** 1979. Cloning of double-stranded cDNA, p. 1–14. *In* J. K. Setlow and A. Hollander (ed.), Genetic engineering. Plenum Publishing Corp., New York.
15. **Fausto-Sterling, A.** 1977. Studies on the female sterile mutant *rudimentary* of *Drosophila melanogaster*. II. An analysis of aspartate transcarbamylase and dihydroorotase activities in wild-type and rudimentary strains. Biochem. Genet. **15**: 803–815.
16. **Grayson, D. R., and D. R. Evans.** 1983. The isolation and characterization of the aspartate transcarbamylase domain of the multifunctional protein, CAD. J. Biol. Chem. **258**:4123–4129.
17. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3965.
18. **Heidecker, G., J. Messing, and B. Gronenborn.** 1980. A versatile primer for DNA sequencing in the M13 mp2 cloning system. Gene **10**:68–73.
19. **Hoover, T. A., W. D. Roof, K. F. Foltermann, G. A. O'Donovan, D. A. Bencini, and J. R. Wild.** 1983. Nucleotide sequence of the structural gene (*pyr*B) that encodes the catalytic polypeptide of aspartate transcarbamoylase of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **80**:2462–2466.
20. **Jones, M. E.** 1980. Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. Annu. Rev. Biochem. **49**:253–279.
21. **Kacian, D. L., and J. C. Myers.** 1976. Synthesis of extensive, possibly complete, DNA copies of poliovirus RNA in high yields and at high specific activities. Proc. Natl. Acad. Sci. U.S.A. **73**:2191–2195.
22. **Land, H., M. Grez, H. Hauser, W. Lindenmaier, and G. Schutz.** 1981. 5'-Terminal sequences of eucaryotic mRNA can be cloned with high efficiency. Nucleic Acids Res. **9**:2251–2266.
23. **Lue, P. F., and J. G. Kaplan.** 1970. Heat-induced disaggregation of a multifunctional enzyme complex catalyzing the first steps in pyrimidine biosynthesis in baker's yeast. Can. J. Biochem. **48**:155–159.
24. **Makoff, A. J., and A. Radford.** 1978. Genetics and biochemistry of carbamoyl phosphate biosynthesis and its utilization in the pyrimidine biosynthetic pathway. Microbiol. Rev. **42**:307–328.
25. **Mally, M. I., D. R. Grayson, and D. R. Evans.** 1981. Controlled proteolysis of the multifunctional protein that initiates pyrimidine biosynthesis in mammalian cells: evidence for discreet structural domains. Proc. Natl. Acad. Sci. U.S.A. **78**:6647–6651.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 156, 254–255. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:497–559.
28. **McDonell, M. W., M. N. Simon, and F. W. Studier.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral alkaline gels. J. Mol. Biol. **110**:119–146.
29. **Meagher, R. B., R. C. Tait, M. Betlach, and H. W. Boyer.** 1977. Protein expression in *E. coli* minicells by recombinant plasmids. Cell **10**:521–536.
30. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 431–433. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Myers, J. C., and S. Spiegelman.** 1978. Sodium pyrophosphate inhibition of RNA:DNA hybrid degradation by reverse transcriptase. Proc. Natl. Acad. Sci. U.S.A. **75**:5329–5333.
32. **Okayama, H., and P. Berg.** 1982. High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2**:161–170.
33. **Okayama, H., and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. **3**:280–289.
34. **Padgett, R. A., G. M. Wahl, P. F. Coleman, and G. R. Stark.** 1979. *N*-(Phosphonacetyl)-L-aspartate-resistant hamster cells overaccumulate a single mRNA coding for the multifunctional protein that catalyzes the first steps of UMP synthesis. J. Biol. Chem. **254**:974–980.
35. **Padgett, R. A., G. M. Wahl, and G. R. Stark.** 1982. Structure of the gene for CAD, the multifunctional protein that initiates UMP synthesis in Syrian hamster cells. Mol. Cell. Biol. **2**:293–301.
36. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
37. **Robey, E. A., and H. K. Schachman.** 1984. Site-specific mutagenesis of aspartate transcarbamylase: replacement of tyrosine 165 in the catalytic chain by serine reduces enzymatic activity. J. Biol. Chem. **259**:11180–11183.
38. **Rothenberg, E., and D. Baltimore.** 1977. Increased length of DNA made by virions of murine leukemia virus at limiting magnesium ion concentration. J. Virol. **21**:168–178.
39. **Rumsby, P. C., P. C. Campbell, L. A. Niswander, and J. N. Davidson.** 1984. Organization of a multifunctional protein in pyrimidine biosynthesis. Biochem. J. **217**:435–440.
40. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
42. **Segraves, W. A., C. Louis, S. Tsubota, P. Schedl, J. M. Rawls, and B. P. Jarry.** 1984. The *rudimentary* locus of *Drosophila melanogaster*. J. Mol. Biol. **175**:1–17.
43. **Sutcliffe, J. G.** 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43**:77–90.
44. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. **77**:5201–5205.
45. **Wahl, G. M., R. A. Padgett, and G. R. Stark.** 1979. Gene amplification causes overproduction of the first three enzymes of UMP synthesis in *N*-(phosphonacetyl)-L-aspartate-resistant hamster cells. J. Biol. Chem. **254**:8679–8689.
46. **Williams, L. G., S. Bernhardt, and R. H. Davis.** 1970. Copurification of pyrimidine-specific carbamyl phosphate synthetase and aspartate transcarbamylase of *Neurospora crassa*. Biochemistry **9**:4329–4335.