Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 31–35, January 1982
Biochemistry

Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of the gene

(signal peptide/intervening sequences/functional domains of proteins/amino acid sequence homology)

CARMEN QUINTO*, MARGARITA QUIROGA, WILLIAM F. SWAIN, WILLIAM C. NIKOVITS, JR., DAVID N. STANDRING, RAYMOND L. PICTET, PABLO VALENZUELA, AND WILLIAM J. RUTTER†

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Communicated by Hans Neurath, August 24, 1981

ABSTRACT **Rat carboxypeptidase A cDNA clones have been isolated from a cDNA library prepared from pancreatic mRNA. An almost complete mRNA sequence has been deduced that predicts a polypeptide having 78% amino acid sequence homology with bovine carboxypeptidase A. The amino acid sequence of the activation and signal peptides of the carboxypeptidase A precursor were inferred from the nucleotide sequence. The cDNA was used as a probe to identify DNA fragments containing carboxypeptidase A sequences in a bacteriophage λ library of rat genomic DNA. Heteroduplexes revealed that the DNA coding sequence occupies 5.5 kilobases and is interrupted by nine intervening sequences. The nucleotide sequence of the 5′ end of the gene and the adjacent flanking region provides information on the site of initiation of transcription and the putative control regions. There is no evident relationship between the localization of intervening sequences in the gene and functional/structural domains of the protein.**

Carboxypeptidase A (peptidyl-L-amino-acid hydrolase, EC 3.4.17.1) is a pancreatic exopeptidase that degrades polypeptides in a sequential fashion from their COOH terminus. This molecule is well characterized: the amino acid sequence of the bovine enzyme has been determined by Neurath and collaborators (1), and the three-dimensional structure has been elucidated by x-ray crystallographic analysis by Lipscomb's group (2). The mechanism of action of the enzyme and its chemical properties have been studied extensively (e.g., ref. 3).

Like most digestive proteases, carboxypeptidase A is formed from an inactive precursor, procarboxypeptidase A. Activation involves the loss of a large peptide whose structure has not been reported. It is presumed that procarboxypeptidase A requires a signal peptide which targets the molecule for secretion (4). Blobel and colleagues (5) have obtained preliminary evidence suggesting that the NH_2-terminal peptides of exocrine pancreatic protein precursors might be similar or identical. Because most signal peptides exhibit considerable variation in sequence, this might suggest a cell- or tissue-specific secretory mechanism.

In this paper we report the isolation and nucleotide sequence of a DNA complementary to rat carboxypeptidase A mRNA and the isolation and partial characterization of a rat genomic DNA clone that contains the carboxypeptidase A gene. The nucleotide sequence predicts the primary translation product, preprocarboxypeptidase A, that contains a putative signal peptide at the NH_2 terminus. Like many other eukaryotic genes, the carboxypeptidase A gene contains intervening sequences (introns). Because of the extensive structural information available about the carboxypeptidase A protein, the relationship between the structures of the gene and the protein may be used to analyze the hypothesis that coding sequences (exons) correspond to functional/structural domains in eukaryotic proteins.

MATERIALS AND METHODS

Library Screening. A cDNA library, constructed from rat pancreatic poly(A)$^+$RNA (unpublished data), was screened according to the method of Grunstein and Hogness (6), except that Whatman 541 filter paper replaced nitrocellulose filters.

A library of rat genomic DNA (7) was screened (8) by using nick-translated (9) pCQ1260 as a probe. Positive plaques were purified (three cycles) and used to prepare DNA (10).

Restriction Mapping and Nucleotide Sequence Analysis. To construct a restriction map, single and double restriction enzyme digestions were carried out. The resulting fragments were analyzed by electrophoresis on agarose or acrylamide gels. DNA fragments were labeled and their sequence was determined by the procedure of Maxam and Gilbert (11). Most of the data presented were obtained by analyzing only one strand of DNA. Appropriate fragments of genomic clones were subcloned into pBR328 (12) prior to sequence determination.

Heteroduplex Mapping. λ11-CQ DNA was hybridized with rat pancreatic poly(A)$^+$RNA according to the method of Fergusson and Davis (13) except that purified DNA was used. Hybrid molecules were visualized with a Philips 300 electron microscope. Dihybrids (13) among λ11-CQ, pCQ1260, and wild-type Charon 4A were used to determine the orientation of the gene and to delimit its flanking regions.

RESULTS

Isolation and Identification of the Rat Carboxypeptidase A cDNA Clone. Rat pancreatic carboxypeptidase A cDNA clones were isolated from a cDNA library prepared from adult rat poly(A)$^+$RNA after elimination of α-amylase (14) and elastase cDNA clones (unpublished data) by screening with their respective probes. From the remaining clones, recombinants were selected at random and sized (15), and partial sequences were determined. One such clone, pCQ500, contained a 500-base-pair insert in which the amino acid sequence predicted from one reading frame of the nucleic acid was 78% homologous with bovine carboxypeptidase A. The cDNA library was rescreened with pCQ500 as a probe, and a larger recombinant plasmid, pCQ1260, containing an insert of approximately 1260 base pairs, was obtained. The sequencing strategy used for both carboxypeptidase A cDNA inserts is presented in Fig. 1. The nucleotide sequence of rat preprocarboxypeptidase A mRNA and the corresponding amino acid sequence, presented in Fig. 2, are derived from both recombinant plasmids and partly from

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

* Present address: Centro de Fijacion de Nitrogeno, UNAM, Apartado Postal 565-A, Cuernavaca, Morelos, Mexico.
† To whom reprint requests should be addressed.

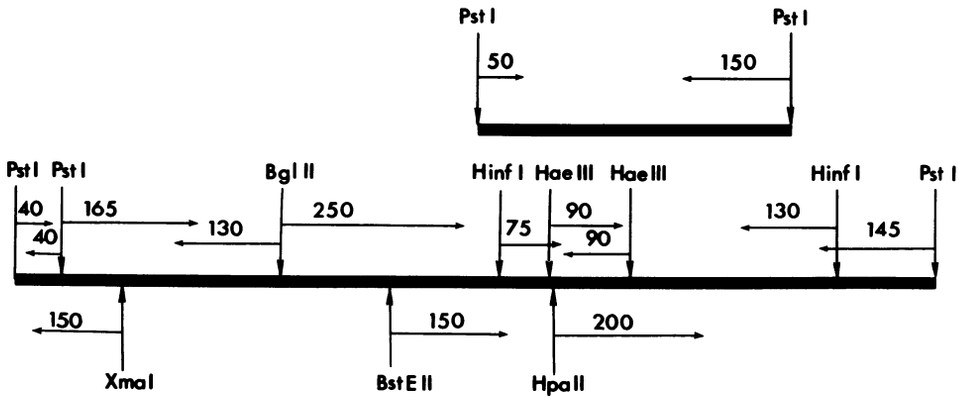

FIG. 1. Restriction endonuclease map of cloned carboxypeptidase A cDNAs and nucleotide sequencing strategy. pCQ500 (*Upper*) and pCQ1260 (*Lower*) inserts were analyzed. The direction and number of base pairs determined from each restriction site are indicated by the horizontal arrows. The terminal *Pst* I sites in each of the clones were generated by the cloning procedure.

```
          -110                                -100                                -90
          Met Lys Arg Leu Leu Ile Leu Ser Leu Leu Leu Glu Ala Val Cys Gly Asn Glu Asn Phe Val Gly His Gln Val Leu Arg
ACCTTCTCACC ATG AAG AGA CTA CTG ATT CTG AGT CTG CTG CTG GAA GCA GTC TGT GGC AAT GAG AAC TTT GTG GGA CAC CAG GTT CTC CGA
                                              ★
          -80                                 -70                                 -60
Ile Ser Ala Ala Asp Glu Ala Gln Val Gln Lys Val Lys Glu Leu Glu Asp Leu Glu His Leu Gln Leu Asp Phe Trp Arg Asp Ala Ala
ATC TCT GCA GCC GAT GAA GCC CAG GTG CAG AAA GTG AAG GAG TTG GAG GAC CTG GAG CAC TTG CAG TTG GAC TTC TGG CGG GAC GCT GCC

          -50                                 -40                                 -30
Arg Ala Gly Ile Pro Ile Asp Val Arg Val Pro Phe Pro Ser Ile Gln Ser Val Lys Ala Phe Leu Glu Tyr His Gly Ile Ser Tyr Glu
CGG GCC GGT ATC CCC ATT GAT GTC AGA GTG CCC TTC CCC AGC ATC CAA TCT GTG AAA GCA TTC TTG GAA TAT CAT GGT ATT AGC TAT GAG

          -20                                 -10                                  1
Ile Met Ile Glu Asp Val Gln Leu Leu Leu Asp Glu Glu Lys Gln Gln Met Ser Ala Phe Gln Ala Arg Ala Leu Ser Thr Asp Ser Phe
ATC ATG ATT GAA GAT GTG CAG TTA CTG CTC GAT GAG GAG AAA CAG CAG ATG TCT GCC TTC CAG GCC AGG GCC TTG TCC ACT GAC TCT TTC

          10                                  20                                  30
Asn Tyr Ala Thr Tyr His Thr Leu Asp Glu Ile Tyr Glu Phe Met Asp Leu Leu Val Ala Glu His Pro Gln Leu Val Ser Lys Ile Gln
AAT TAT GCC ACC TAT CAT ACG CTG GAT GAG ATC TAC GAA TTC ATG GAC CTG CTG GTG GCT GAG CAC CCA CAG CTT GTG AGC AAG ATC CAG

          40                                  50                                  60
Ile Gly Asn Thr Phe Glu Gly Arg Pro Ile His Val Leu Lys Phe Ser Thr Gly Gly Thr Asn Arg Pro Ala Ile Trp Ile Asp Thr Gly
ATC GGC AAC ACC TTT GAA GGT CGC CCC ATC CAT GTG CTG AAG TTC AGC ACT GGA GGG ACC AAT CGC CCA GCA ATC TGG ATC GAC ACT GGC

          70                                  80                                  90
Ile His Ser Arg Glu Trp Val Thr Gln Ala Ser Gly Val Trp Phe Ala Lys Lys Val Thr Lys Asp Tyr Gly Gln Asp Pro Thr Phe Thr
ATC CAT TCT AGG GAG TGG GTC ACC CAG GCT AGT GGG GTC TGG TTT GCA AAG AAG GTC ACC AAA GAC TAT GGC CAG GAC CCC ACC TTC ACA

          100                                 110                                 120
Ala Val Leu Asp Asn Met Asp Ile Phe Leu Glu Ile Val Thr Asn Pro Asp Gly Phe Ala Tyr Thr His Lys Thr Asn Arg Met Trp Arg
GCC GTT CTT GAC AAC ATG GAC ATC TTC TTG GAG ATT GTC ACC AAC CCT GAT GGT TTT GCC TAC ACC CAC AAA ACG AAT CGC ATG TGG CGC

          130                                 140                                 150
Lys Thr Arg Ser His Thr Gln Gly Ser Leu Cys Val Gly Val Asp Pro Asn Arg Asn Trp Asp Ala Gly Leu Gly Lys Ala Gly Ala Ser
AAG ACT CGA TCA CAC ACC CAG GGC TCC CTC TGT GTT GGT GTG GAC CCC AAC AGG AAC TGG GAC GCT GGC TTA GGG AAG GCC GGA GCA AGT

          160                                 170                                 180
Ser Asn Pro Cys Ser Glu Thr Tyr Arg Gly Lys Phe Pro Asn Ser Glu Val Glu Val Lys Ser Ile Val Asp Phe Val Thr Ser His Gly
AGC AAC CCC TGC TCG GAA ACT TAC CGA GGC AAA TTT CCC AAC TCT GAG GTG GAG GTC AAG TCC ATC GTG GAC TTT GTG ACG AGC CAT GGG

          190                                 200                                 210
Asn Ile Lys Ala Phe Ile Ser Ile His Ser Tyr Ser Gln Leu Leu Leu Tyr Pro Tyr Gly Tyr Thr Ser Glu Pro Ala Pro Asp Gln Ala
AAC ATC AAG GCC TTC ATC TCC ATC CAC AGC TAT TCC CAG CTC CTG CTC TAC CCC TAC GGC TAC ACG TCG GAA CCA GCC CCT GAC CAG GCA

          220                                 230                                 240
Glu Leu Asp Gln Leu Ala Lys Ser Ala Val Thr Ala Leu Thr Ser Leu His Gly Thr Glu Phe Lys Tyr Gly Ser Ile Ile Asp Thr Ile
GAG CTG GAT CAG CTA GCT AAG TCT GCT GTG ACA GCC TTG ACG TCT CTA CAC GGG ACC GAG TTC AAG TAT GGC AGC ATC ATC GAT ACA ATC

          250                                 260                                 270
Tyr Gln Ala Ser Gly Ser Thr Ile Asp Trp Thr Tyr Ser Gln Gly Ile Lys Tyr Ser Phe Thr Phe Glu Leu Arg Asp Thr Gly Leu Arg
TAT CAA GCC AGT GGG AGC ACT ATC GAT TGG ACC TAC AGC CAG GGC ATC AAG TAC TCT TTC ACT TTT GAA CTG AGG GAC ACT GGG CTT AGA

          280                                 290                                 300
Gly Phe Leu Leu Pro Ala Ser Gln Ile Ile Pro Thr Ala Glu Glu Thr Trp Leu Ala Leu Leu Thr Ile Met Asp His Thr Val Lys His
GGA TTC CTG CTG CCT GCC TCC CAG ATC ATC CCT ACG GCG GAG GAG ACA TGG CTG GCC CTT TTG ACC ATC ATG GAC CAC ACA GTC AAA CAC

    309
Pro Tyr OP
CCC TAC TGA AAGCAGCCCTTATCCCCCACCTCTTTCTTCTCCTTAGTC
```

FIG. 2. Nucleotide sequence of rat carboxypeptidase A mRNA and the corresponding amino acid sequence. Star, 5′ limit of the sequence determined from cDNA clones; arrows, probable processing sites for the pre and pro peptides.

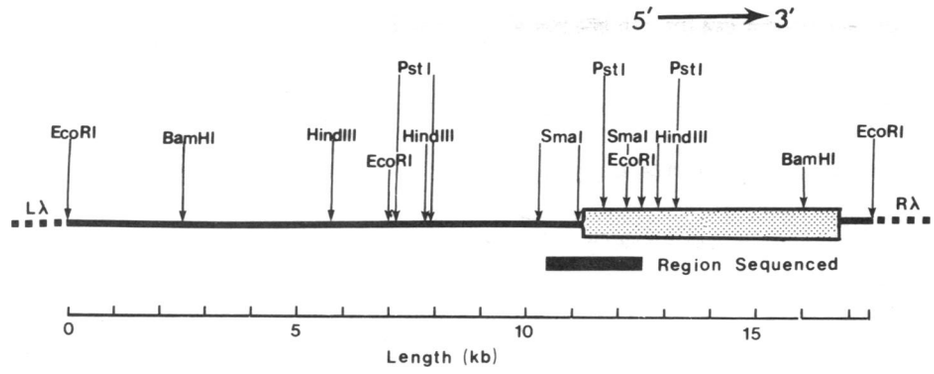

FIG. 3. Restriction endonuclease map of the cloned rat carboxypeptidase A gene. Stippled box, length of the carboxypeptidase A gene within the *Eco*RI fragment; black box, region at the 5′ end of gene whose sequence has been determined; broken lines, left arm (Lλ) and right arm (Rλ) of λ Charon 4A, respectively. kb, Kilobases.

the sequence of a genomic clone (see below). The size of the carboxypeptidase A mRNA sequences in rat pancreas poly(A)$^+$mRNA was determined, by the reverse Southern blotting method, to be 1450 ± 75 bases (data not shown). Thus, pCQ1260 contains a nearly full-length cDNA insert.

Isolation and Characterization of Carboxypeptidase A Rat Gene. Radiolabeled pCQ1260 was used to screen a rat genomic library of 10^6 recombinant phages (≈3 rat genome equivalents). Fifteen phages were independently isolated, and restriction endonuclease analysis suggested that all 15 represent overlapping fragments from a single carboxypeptidase A gene (data not shown). The restriction map of one genomic clone, λ11-CQ, which contains the entire carboxypeptidase A coding sequence, is shown in Fig. 3.

The structure of the carboxypeptidase A gene has been determined by electron microscopic examination of heteroduplexes between the genomic fragment λ11-CQ and rat pancreatic poly(A)$^+$RNA. A representative electron micrograph, the corresponding interpretation, and a schematic drawing of the carboxypeptidase A gene are shown in Fig. 4. Carboxypeptidase A encoding segments of approximately 80–200 bases lie within a 5.5-kilobase region of genomic DNA and are separated by nine intervening sequences ranging in length from 130 to 1380 bases. The estimated total length of the exons (1253 ± 150 bases) agrees well with the size of the mRNA (1450 ± 75). Dihybrids among λ11-CQ, *Sal* I-linearized pCQ1260, and wild-type Charon 4A revealed the same pattern of nine intervening sequences (data not shown), indicating that there are no extra introns at the 5′ or 3′ ends of the mRNA.

The exon–intron boundaries of three of the nine intervening sequences have been determined. These sequences (data not shown) are consistent with the consensus sequence compiled by Benoist *et al.* (16). The nucleotide sequence of the 5′ flanking end of the carboxypeptidase gene is shown in Fig. 5 which delineates the putative control regions for this gene.

The Rat Genome Contains One Carboxypeptidase A Gene. High molecular weight DNA from Sprague–Dawley rats was cleaved in separate experiments with *Eco*RI, *Hin*dIII, *Bam*HI, and *Pst* I and analyzed by Southern blotting using radiolabeled pCQ1260 as a probe. The hybridization pattern (data not shown) was similar to the restriction map of the genomic clone λ11-CQ (Fig. 3). Thus, the relevant sequences of λ11-CQ can account for all of the carboxypeptidase A fragments observed in the rat genome by this test. We conclude that the rat genome contains a single carboxypeptidase A gene. This idea is supported by the similarity of the restriction maps of the 15 independently derived genome clones.

DISCUSSION

In the present work we identified the rat carboxypeptidase A gene by two criteria. First, the predicted amino acid sequence

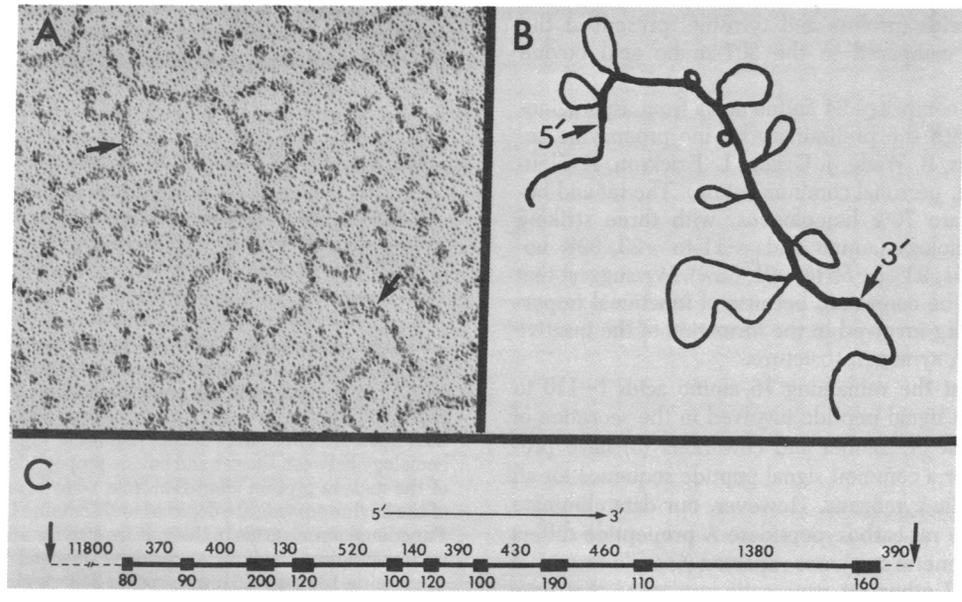

FIG. 4. Heteroduplex map of the rat carboxypeptidase A gene. (*A*) Electron micrograph of a representative heteroduplex formed between rat pancreatic poly(A)$^+$RNA and λ11-CQ genomic DNA. (×35,000.) (*B*) Interpretative drawing of the heteroduplex structure. Arrows delimit the region of homology between the genomic fragment and the RNA. (*C*) Schematic representation of the organization of carboxypeptidase A gene in the genomic DNA fragment. The lengths of the coding segments and introns are mean values determined from eight different hybrids.

```
   -290      -280      -270      -260      -250      -240      -230      -220      -210      -200
CGTCCTTGTTTCTGACATCGTGAGCTAACAATAAGGTATCTGCGGCTGTGGGAAACCGGATGAGTAGTCCCAGCTGGGAGACCTCAGGCGCAGCTGACCCCATG

-190      -180      -170      -160      -150      -140      -130      -120      -110      -100       -90
GTCAAGGGTAGAAGCCTGGCTTATCTCTCCACCTGCCTTGTTCCCCGATACTTTATCAGGAAGAGTGGAAAGGTGCCGCAGTTTGGAACTCCAGCCTCTCCCTC

      -80       -70       -60       -50       -40       -30       -20       -10        1        10 Met
CCCATGGGACTTGATCCAGATGTGAGGGGAACTGTCCCCGGGGACCCCTGTCAGCGTTTAAAAAGGCCTCAGATCTCAGTCTTGGCCTGACCTTCTCACCATG
```

FIG. 5. Nucleotide sequence of the 5′ flanking region of the carboxypeptidase A gene. The putative 5′ end of the mRNA (capping site) is indicated by an asterisk; the presumed T-A-T-A (Hogness–Goldberg) sequence is underlined; C-C-A-A-T box is underlined with a dashed line.

displays 78% homology with that of the mature bovine enzyme (1). Second, the invariant amino acids associated with biological function (2, 3) are present at the correct positions in the rat protein sequence: histidine-69, glutamate-72, and histidine-196 participate in zinc binding; arginine-71, arginine-145, tyrosine-198, isoleucine-247, tyrosine-265, and phenylalanine-279 are involved in substrate binding; tyrosine-248 may act as a proton donor and glutamate-270 as a nucleophile (or general base); the structure is maintained in part by a disulfide bond between cysteine-138 and -161.

The majority of the carboxypeptidase A mRNA sequence was determined from the sequence of two cDNA clones, but the untranslated leader sequence and the first 11 amino acids of the signal peptide are derived from the genomic DNA and, in principle, are uncertain. However, we feel confident of this assignment because (*i*) a heteroduplex analysis reveals no introns in this region, (*ii*) the predicted prepeptide displays characteristic structural features found in other signal sequences, and (*iii*) putative control regions are found in appropriate positions relative to the proposed start of the mRNA sequence (see below). The 1310 nucleotides of mRNA sequence presented in Fig. 2 account for most of the length of the carboxypeptidase A mRNA (1450 ± 75 bases). However, only 42 nucleotides have been determined beyond the UGA stop codon. This segment does not include the hexonucleotide A-A-U-A-A-A (17) or poly(A).

The mRNA nucleotide sequence predicts that preprocarboxypeptidase A is a protein containing 419 amino acids and demonstrates the presence of an NH_2-terminal signal peptide. The 309-amino acid mature enzyme [putting alanine-1 as the NH_2 terminus by analogy with the bovine enzyme (1)] contains two extra amino acids (proline and tyrosine) present at the COOH terminus, compared to the 307-amino acid bovine enzyme.

The propeptide comprises 94 amino acids from asparagine-−94 by analogy with the preliminary bovine proenzyme sequence (M. G. Hass, R. Wade, J. Gagon, L. Erickson, H. Neurath, and K. Walsh, personal communication). The rat and bovine propeptides are 70% homologous, with three striking blocks of high homology (amino acids −11 to −23, 92% homology; −40 to −51, 92%; −58 to −91, 82%). We suggest that these regions may be conserved because of functional importance, perhaps being involved in the formation of the inactive carboxypeptidase A zymogen structure.

We propose that the remaining 16 amino acids (−110 to −93) comprise the signal peptide involved in the secretion of procarboxypeptidase A. Blobel and coworkers (5) have presented evidence for a common signal peptide sequence for all (or many) pancreatic zymogens. However, our data eliminate this possibility; the rat carboxypeptidase A prepeptide differs sharply from the general sequence reported (5) as do the signal peptides of several other rat pancreatic zymogens that have been determined to date (unpublished data).

Restriction analysis of independently isolated genomic clones and of rat genomic DNA indicates that the rat genome contains a single copy of the carboxypeptidase A gene (unpublished data). Approximately 2 kilobases of genomic DNA has been subjected to sequence analysis including 600 bases of DNA in the 5′ flanking putative control region. A possible capping site (18) is found 11 nucleotides upstream from the initiator methionine: we propose that the transcript initiates and is capped at the adenosine residue (18). The sequence T-T-T-A-A-A, a variant of the consensus Goldberg–Hogness (19) T-A-T-A-A-A sequence, occurs at position −30 with respect to the cap site. At −72 there is a sequence, C-C-A-G-A, that resembles the C-C-A-A-T box (20).

Heteroduplex analysis of the genomic clone against either the cDNA or the mRNA indicates that the gene contains at least nine introns. The first three have been localized precisely by nucleotide sequence analysis. It has been proposed (21) that exons delineate functional domains in proteins. Thus, the creation of new proteins in evolution may be facilitated by shuffling the exons. This argument has received support from studies of immunoglobulin (22) and globin (23) gene structure but has not been extensively applied to other proteins, especially those with catalytic activity (24). In Fig. 6 we present a correlation between the intron–exon structure, the three-dimensional structure of the enzyme, and the distribution of essential amino acids. There is no obvious overall correlation of exon boundaries with structural/functional features of the molecule: the first exon includes the signal peptide but also four propeptide residues. The second exon contains most of the conserved region of the propeptide, but this region extends two amino acids before and three amino acids after the exon boundaries. The third contains the remainder of the propeptide sequence, including two conserved regions, as well as 18 amino acids of the mature enzyme. Three possible structural domains (amino acids 1–127, 128–189, and 190–307) have been identified (25, 26) in the mature bovine

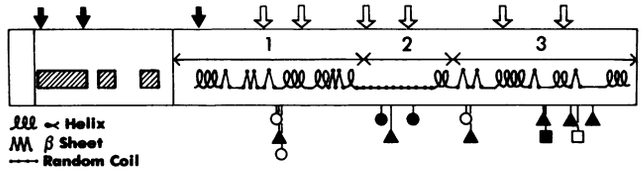

FIG. 6. Alignment of introns in the preprocarboxypeptidase A gene with features of the primary and tertiary structures of the protein. The box represents the preprocarboxypeptidase A protein which is divided into pre, pro, and mature regions (from left to right) by vertical lines. Hatched boxes denote regions of high amino acid sequence homology between the rat and bovine propeptides. Structural features of the mature protein identified from x-ray crystallographic analysis of the bovine enzyme are depicted and domains 1 through 3 are shown. Functional amino acids in the mature enzyme are indicated as follows: ○, zinc-binding residues; ▲, residues involved in substrate binding; ●, cysteine-138 and -161; ■, tyrosine-248; □, glutamate-270 (see *Discussion* for identity of residues). The positions at which introns interrupt the rat carboxypeptidase A coding sequence are depicted by vertical arrows above the protein. Solid arrows, introns that have been localized by DNA sequence; open arrows, introns localized by heteroduplex mapping.

enzyme. These domains do not correlate with the positions of introns. The essential amino acid residues are dispersed over five exons without any functional grouping. It seems possible that simple structural motifs (e.g., combinations of α-helix and β-sheet structures) that combine to build more complex structural domains could be the basis of coding units. The bovine carboxypeptidase A molecule is composed of an eight-strand array of β-sheets and α-helices; we find no obvious relationship between these structures and the location of the introns. The analysis of such an idea requires more precise resolution of intron–exon boundaries of this and other well-characterized proteins, especially those with catalytic activity. Meanwhile we are left with the impression that, in this gene, the position of the introns is governed by forces that disperse them throughout the gene structure without particular regard for the protein structure.

We thank Drs. R. J. MacDonald, G. I. Bell, C. Craik, and F. E. Sanchez for helpful discussions and Dr. G. M. Hass for sending us the amino acid sequence of bovine procarboxypeptidase A prior to publication. We also thank Leslie Spector for the preparation of the manuscript and Sonja Bock for assistance in the computer work for DNA sequence analysis. C.Q. was supported by Centro de Fijacion de Nitrogeno, Universidad Nacional Autonoma de Mexico. This work was supported by National Institutes of Health Grant AM 21344.

1. Bradshaw, R. A., Ericsson, L. H., Walsh, K. A. & Neurath, H. (1969) *Proc. Natl. Acad. Sci. USA* **63**, 1389–1394.
2. Rees, D. C., Lewis, M., Honzatko, R. B., Lipscomb, W. N. & Hardman, K. D. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3408–3412.
3. Vallee, B. L. (1977) *FEBS 11th Meeting*, Copenhagen, eds. Magnusson, S., Ottesen, M., Foltman, B., Davo, K. & Neurath, H. (Pergamon, New York), Vol. 47, pp. 57–67.
4. Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–851.
5. Devillers-Thiery, A., Kindt, T., Scheele, G. & Blobel, G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 5016–5020.
6. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
7. Sargent, T. D., Wu, J. R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A. & Bonner, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3256–3260.
8. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
9. Maniatis, T., Jeffreys, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
10. Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W. & Smithies, O. (1978) *Science* **202**, 1279–1284.
11. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
12. Soberon, X., Covarrubias, L. & Bolivar, F. (1980) *Gene* **9**, 287–305.
13. Fergusson, J. & Davis, R. W. (1978) in *Advanced Techniques in Biological Electron Microscopy*, ed. Koehler, J. K. (Springer, Berlin), pp. 123–171.
14. MacDonald, R. J., Crerar, M. M., Swain, W. F., Pictet, R. L., Thomas, G. & Rutter, W. J. (1980) *Nature (London)* **278**, 117–122.
15. Telford, J., Bosely, P., Schaffner, W. & Birnstiel, M. (1977) *Science* **195**, 391–392.
16. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8**, 127–142.
17. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
18. Corden, J., Wasylyk, A., Buchwalder, P., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209**, 1406–1413.
19. Goldberg, M. (1979) Dissertation (Stanford Univ., Stanford, CA).
20. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R., O'Conell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightman, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 653–668.
21. Gilbert, W. (1978) *Nature (London)* **271**, 501.
22. Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature (London)* **277**, 627–633.
23. Craik, C. S., Buchman, S. R. & Beychok, S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1384–1388.
24. Benyajati, C., Place, A. R., Powers, D. A. & Sofer, W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2717–2721.
25. Liljas, A. & Rossmann, M. G. (1974) *Annu. Rev. Biochem.* **43**, 475–507.
26. Lipscomb, W. N., Reeke, G. N., Jr., Hartsuck, J. A., Quiocho, R. A. & Bethge, P. H. (1970) *Philos. Trans. R. Soc. London Ser. B* **257**, 177–214.