

This is to certify that the

thesis entitled

THE NUCLEOTIDE SEQUENCE
OF THE CHICKEN
CARBONIC ANHYDRASE II GENE

presented by

Corinne Misae Yoshihara

has been accepted towards fulfillment
of the requirements for

Doctor of Philosophy degree in Genetics

A handwritten signature in cursive script, reading "Jerry Dodge", written over a horizontal line.

Major professor

Date 2/21/86

THE NUCLEOTIDE SEQUENCE
OF THE CHICKEN
CARBONIC ANHYDRASE II GENE

By
Corinne Misae Yoshihara

A DISSERTATION
Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY

Genetics Interdisciplinary Program

1986

Copyright by
CORINNE MISAE YOSHIHARA
1986

ABSTRACT

THE NUCLEOTIDE SEQUENCE OF THE CHICKEN CARBONIC ANHYDRASE II GENE

By

Corinne Misae Yoshihara

The complete nucleotide sequence of the coding region of the chicken carbonic anhydrase II (CA II) gene has been determined from clones isolated from a λ Charon 4A chicken genomic library. The gene is approximately 17 kilobase pairs (kb) in size and codes for a protein that is comprised of 259 amino acid residues. Six introns ranging in size from 0.3-10.3 kb interrupt the gene. The 5'-flanking region contains the ATA and CCAAT consensus sequences commonly associated with eucaryotic genes transcribed by RNA Polymerase II. The 5'-flanking region farther upstream is extremely GC-rich and characteristic short stretches of GC sequences are found around the CCAAT sequence. The 3'-flanking region contains long stretches of the nucleotides A and T. There is a relatively long untranslated region with 5 possible sites of polyadenylation, the most probable site being the one closest to the stop codon.

A comparison of the derived amino acid sequence of chicken CA II to that of mouse CA II shows a 65% amino acid sequence homology. The number of introns as well as 5 of the 6 intron locations are conserved between the two homologous genes. The site of the fourth intron is

Corinne Misae Yoshihara

shifted by $4 \frac{2}{3}$ codons further 3' in the chicken and now falls between the codons 147/148 rather than within the codon 143 as in the mouse gene. A comparison of the 5'-flanking regions of mouse and chicken CA II genes shows that the genes are quite homologous in a region up to 80 bp 5' to the first nucleotide coding for mRNA (Cap site), although they are less similar than the mouse and human CA II genes in this region.

To My Mother and Father



ACKNOWLEDGEMENTS

I would like to thank several people for their support during my career as a graduate student.

First I would like to thank Dr. Jerry Dodgson for the opportunity to be in science and for the guidance and training he provided.

I would next like to thank Dr. Richard Tashian and Dr. Pat Venta for the informative conversations, materials, and helpful suggestions they offered from afar.

I would also like to thank my committee members for their patience and suggestions through it all.

The lab. would not be the same without its members. I would like to acknowledge the people in the lab. for the help and friendship they offered.

I would like to thank my friend S. Decker for his patience and understanding through the difficult times.

Last of all I would especially like to thank my parents and brother for their long distance support and for their belief in my abilities.

TABLE OF CONTENTS

	Page
List of Tables	vi
List of Figures	vii
INTRODUCTION	1
CHAPTER 1. LITERATURE REVIEW	2
The Carbonic Anhydrase Protein	2
General Properties	2
Mammalian Isozyme Comparison	3
CA Isozymes	3
Comparison of CA I, CA II, and CA III	3
CA IV	9
Mitochondrial CA	10
Tissue Distribution and Function	10
CA I - CA V	10
Transport Epithelia	11
Macromolecule-secreting Epithelia	11
Nonepithelial Cells	12
Unique CA's	13
Expression During Development	14
Expression in Erythrocytes	16
Variability in Expression of Isozymes	16
Changes in Isozyme Levels During Red Cell Maturation	18
Quantitative Genetic Variation	18
Qualitative Genetic Variation	21
Hormonal Control	21
Evolution of Carbonic Anhydrases	22
The Carbonic Anhydrase Gene	23
Mouse CA II Gene	23
Isolation of Mouse CA II Gene	24
Rabbit CA I cDNA	26

	Page
CHAPTER 1. (CONTINUED)	
Comparison of Mouse and Human CA II Genes	27
References	29
CHAPTER 2. ISOLATION OF THE CHICKEN CARBONIC ANYHDRASE II GENE .	36
Acknowledgements	39
References	39
CHAPTER 3. THE NUCLEOTIDE SEQUENCE OF THE CHICKEN CARBONIC ANYHDRASE II GENE	40
Introduction	40
Experimental procedures	42
Isolation of cDNA	42
Isolation of λ CA III and λ CA XVI	42
Preparation of Subclones Containing Fragments of the CA II Gene	43
Restriction Enzyme Digestion and Southern Blotting . .	43
DNA Sequence Analysis	43
Nuclease Protection/End Analysis	44
Materials	45
Results	45
Isolation of Larger Chicken CA II cDNA Clones	45
Isolation of the Chicken CA II Gene	45
DNA Sequencing Strategy of the Chicken CA II Gene . . .	49
Sequence of the Chicken CA II Gene	52
Chicken CA II Gene Intron/Exon Organization	63
5' and 3' Ends of the Chicken CA II mRNA	66
5' and 3' Flanking Sequences of the Chicken CA II Gene.	69
Discussion	76
References	82
Appendix	86

LIST OF TABLES

TABLE		PAGE
<u>CHAPTER 1</u>		
I	Residues located, or postulated to occur, in the active site regions of the carbonic anhydrase isozymes, CA I, CA II and CA III	8
<u>CHAPTER 3</u>		
I	Nucleotide and amino acid divergence of chicken CA II gene	61
II	Unique and invariant residues for the mammalian CA isozymes and the homologous residues in the chicken CA II isozyme	62
III	DNA sequence of intron donor and acceptor sites	64
IV	Intron size comparison of the chicken and mouse CA II genes	65

LIST OF FIGURES

FIGURE		PAGE
<u>CHAPTER 1</u>		
1	Location of amino acid substitutions [o] for genetic variants of the human CA I, CA II and CA III isozymes on a stylized version of the alpha-carbon chain of human CA I and CA II molecules, which have very similar three-dimensional structures.	6
2	Restriction endonuclease map and exon positions of the mouse carbonic anhydrase II gene in cosmid cosMCAII-2... .	25
<u>CHAPTER 2</u>		
1	cDNA clone of carbonic anhydrase II	37
2	Nucleotide sequence of pPE5-0.3	38
<u>CHAPTER 3</u>		
1	Restriction map of the chicken CA II gene locus	48
2	Restriction maps of the subcloned DNA fragments of the chicken CA II gene.	51
3	DNA sequence of the chicken CA II gene	54
4	Amino acid sequence comparison of chicken and mouse CA II genes	59
5	5' end analysis of chicken CA II cytoplasmic mRNA	68
6	DNA sequence of putative signal sequences flanking the CA II gene.	71
7	Comparison of the 5' flanking region of the chicken, mouse, and human CA II genes.	74
8	Comparison of intron 4 location in chicken and mouse CA II genes.	79
9	Nucleotide sequence of 5'-flanking and intron regions of the chicken CA II gene	86



INTRODUCTION

Carbonic anhydrase (CA) is a zinc metalloenzyme that functions as a general acid-base catalyst. It is because it has such a basic role that CA is fundamental to nearly all organisms.

CA was first found in equine blood as the enzyme responsible for catalyzing CO_2 hydration. It was not until chromatographic separation allowed the purification of individual proteins that red blood cell CA was found to be two isozymes, CA I and CA II. Since then, there have been several other isozymes of CA that have been identified from cells other than erythrocytes. Genetic studies as well as protein structure data provided evidence that the isozymes were coded for by separate genes.

This review will first discuss what is known about CA protein, its structure and function. In the following section the review will cover the more recent findings on CA gene structure and its role as a model system in which to study eucaryotic gene expression.

CHAPTER 1

LITERATURE REVIEW

THE CARBONIC ANHYDRASE PROTEIN

General Properties

Since its initial identification in equine blood CA has been found in a wide variety of organisms, including bacteria, algae, yeast, mosses, plants and numerous vertebrate and invertebrate species. CA is thought to have such diverse physiological roles as CO₂ fixation (photosynthesis); calcification (shell formation); respiration, where the specific catalytic role of the enzyme is the interconversion of CO₂ and HCO₃⁻ (the hydration of CO₂ released in body tissues and the dehydration of HCO₃⁻ in the lungs); and the transport and accumulation of H⁺ or HCO₃⁻ in organs of secretion such as stomach and kidney (for a review see 1,2).

The primary reaction catalyzed by CA is the reversible hydration of CO₂; i.e., $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+$. However, CA also utilizes substrates other than CO₂ and HCO₃⁻. It is able to hydrolyze certain esters of carboxylic acids (3,4,5,6), carbonic acids (7), and cyclic sulfonates (8), and hydrate a variety of aldehydes (9,10) and pyruvic acid (11).

Mammalian Isozyme Comparison

CA Isozymes

Five different CA isozymes have been identified in mammalian species. Three soluble isozymes of CA; CA I, CA II, and CA III, are known to exist in amniotes (birds, reptiles, and mammals) (12). A fourth isozyme, a membrane-bound CA, has been found in mammalian brain (13), lung (14,15), and kidney (16,17,18). A fifth isozyme, a soluble mammalian CA in mitochondria (19,20) that has been implicated in the urea cycle, gluconeogenesis and fatty acid synthesis (21), has recently been characterized. A form that is synthesized in the parotid accinar cells and exported to saliva (22) is possibly another isozyme.

Comparison of CA I, CA II, and CA III

The isozymes that have been well characterized (I,II,III) vary in terms of catalytic efficiency, amino acid composition, appearance during development, sulfonamide inhibition, tissue distribution, and sometimes in molecular weight.

The amino acid sequence of human CA was first determined in 1972 (23). Since then the amino acid sequences of several CA I, II, and III isozymes from different species have been determined (33). However, since CA I and CA II are the best characterized isozymes, most of the isozyme comparisons are made between these two. Comparisons made with the newer isozymes are limited.

CA I and CA II are similar in amino acid composition. Both isozymes contain a high proportion of proline residues, acidic and basic residues, and few sulfur-containing residues. The high activity enzyme, CA II, contains more basic residues (histidine, lysine, and

arginine) than CA I. CA I, on the other hand, has a high serine content. In mammals, CA I possesses one and a half times as many serine residues as CA II (24). Despite these differences there is 60% homology between the amino acid sequences of CA I and CA II in humans (25).

In contrast, the amino acid composition of CA III is distinct from that of I and II. CA III has been found to have a higher basic amino acid content (26).

CA I and CA II isozymes also differ in their immunological properties. The antigenic sites of CA are independent of the active sites. The enzyme-antibody complex thus retains some CA activity and the presence of sulfonamide inhibitor (which binds to the active site) does not affect enzyme-antibody complex formation. The antigenic sites differ for CA I and CA II. Antisera specific for the human CA I enzyme will not cross-react with CA II and vice versa. In fact, there is more homology between the same isozymes between species than between CA I and CA II of the same species. CA I and CA II from species closely related to man cross-react with antisera prepared against the homologous human isoenzyme (2).

CA I and CA II isozymes differ in their sulfonamide binding affinities. Sulfonamides inhibit CA activity by binding to the zinc ion in the active site of the CA molecule. CA II isozymes bind most sulfonamides more readily than the other CA isozymes (27) and this difference in binding reflects the difference among the isozymes at the active site. For example, the isozyme with the least homology among the three isozymes, CA III, is the least sensitive to inhibition (26)

in chicken.

This type of variation in isozyme inhibition also holds true for anions that bind to the active site. The CO_2 hydrase activity of human CA I is strongly inhibited by physiological levels of chloride and bicarbonate ions found in human erythrocytes whereas the activity of CA II is not (28).

Although all the CA's catalyze the reversible reaction $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+$, the three isozymes vary considerably in their catalytic efficiencies. CA II has a higher specific CO_2 hydrase activity than CA I, and CA III has the lowest activity. The specific activities of chicken CA isozymes toward $\text{CO}_2/\text{H}_2\text{CO}_3^-$ differ according to the ratio III:I:II = 1:6:47 (26).

The x-ray diffraction study of crystals of human erythrocyte CA II was initiated by Strandberg in 1962 and by 1975 the three-dimensional structures of both human CA I (29,30) and CA II (31,32) had been determined. The tertiary structures of CA I and CA II are very similar. Human CA I and CA II are ellipsoidal in shape and measure 41 x 42 x 55 Å (Figure 1). Approximately 20% of the CA protein exists as helical structures and another 35% of the protein comprises a large beta-structure which is made up of 10 segments. The beta-structure is the dominating feature of the protein and passes through the whole molecule dividing the molecule in half. Each half contains 6 pieces of polypeptide chain that run perpendicular to the beta-structure. The active site cavity is conical, approximately 12 Å deep and contains a single zinc ion at the bottom. The zinc is liganded by His 94, His 96,

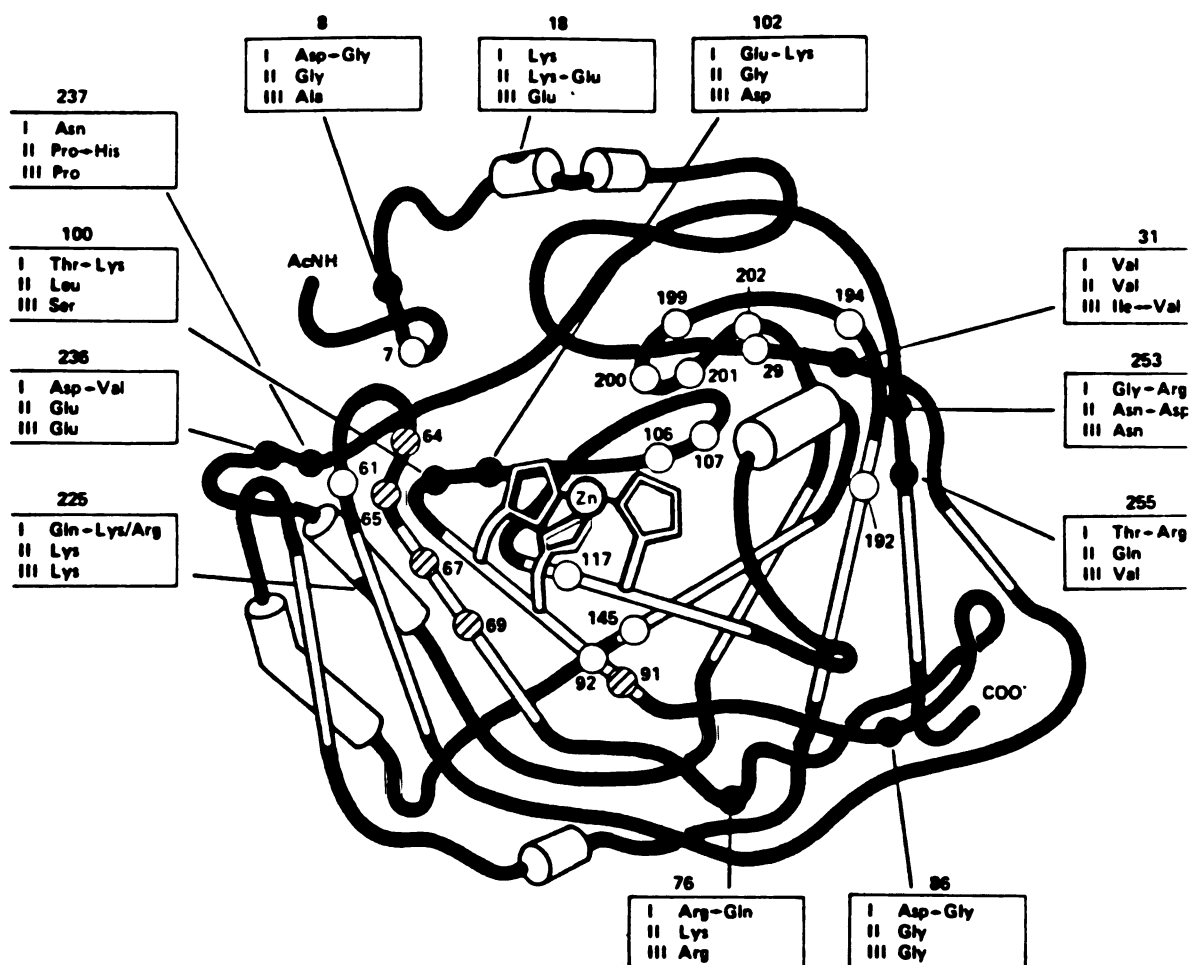


Figure 1. Location of amino acid substitutions [o] for genetic variants of the human CA I, CA II and CA III isozymes on a stylized version of the alpha-carbon chain of human CA I and CA II molecules, which have very similar three-dimensional structures. Boxes include the original substitution and homologous residues at this position for the other human isozymes. Beta-Structure segments are unshaded and helical structures are indicated by cylinders. The zinc ion is shown liganded to histidine residues 94, 96, and 119. Active site residues unique to human CA III [o] are Lys-64, Thr-65, Arg-67, Val-69, and Arg 91. (Source: *Biology and Chemistry of the Carbonic Anhydrases*. 1984. R.E. Tashian and D. Hewett-Emmet, eds.)

and His 119. One-half of the active site is composed of hydrophobic side chains and the other half is composed of hydrophilic residues which are exposed at the surface of the site (2).

Five types of residues of CA I and CA II determined from the tertiary structures appear to be important in maintaining stability and activity of the molecules (27). These residues include 3 aromatic clusters, 23 active site residues, 30 hydrophobic core residues, and 8 residues hydrogen-bonded to the zinc ligands, His 94 and His 119, or to the zinc-bound solvent molecule.

The amino acid residues that are linked to the metal ligands by a hydrogen-bond network are invariant in all sequenced CA's. However, there are certain characteristic differences between the isozymes with respect to polar amino acid residues in their active sites (Table 1) (34). Residues 64 and 200 appear to critically modify catalytic behavior. His is present at position 64 and Thr or Asn at position 200 in all sequenced CA II's. Histidine residues appear in both positions in all sequenced CA I's. The presence of His at position 200 is the only feature that distinguishes the active sites of CA I from those of CA II. CA III appears to have Thr at position 200, Lys at position 64 and additional basic arginine residues at positions 67 and 91 in their active sites. The active site residues that are constant for each isozyme type but differ between isozymes (67 His/Asn, 69 Asn/Glu, 131 Leu/Phe, 204 Tyr/Leu, and 211 Ile/Val) may be responsible for the differences in the activities of the two isozymes, and in their inhibition patterns by sulfonamides and anions (27). The active sites also can be indirectly influenced by variable residues not in the

Table I. Residues Located, or Postulated to Occur, in the Active Site Regions of the Carbonic Anhydrase Isozymes, CA I, CA II, and CA III^a

Isozymes	Residue Number																
	7	29	61	64	65	67	69	91	92	94	96	106	107	117	119		
Human I	Tyr	Ser	Asn	His	Ser	His	Asn	Phe	Gln	His	His	Glu	His	Glu	His		
Rhesus I	Tyr	Ser	Asn	His	Ser	His	Asn	Phe	Gln	His	His	Glu	His	Glu	His		
Ox I	Tyr	Ser	Asn	His	Ser	His	Asn	Phe	Gln	His	His	Glu	His	Glu	His		
Horse I	Tyr	Ser	Asn	His	Ser	Gln	Lys	Val	Gln	His	His	Glu	His	Glu	His		
Turtle I	Tyr	Ser	Asn	His	Ser	His	Asn	His	Gln	His	His	Glu	His	Glu	His		
Human II	Tyr	Ser	Asn	His	Ala	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His		
Ox II	Tyr	Ser	Asn	His	Ser	Asn	Glu	Val	Gln	His	His	Glu	His	Glu	His		
Horse II	Tyr	Ser	Asx	His	Ser	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His		
Rabbit II	Tyr	Ser	Asn	His	Ser	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His		
Mouse II	Tyr	Ser	Asn	His	Ser	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His		
Chicken II	Ser	Ser	Asn	His	Ser	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His		
Human III	Tyr	Ser	Asn	Lys	Thr	Arg	Val	Arg	Gln	His	His	Glu	His	Glu	His		
Ox III	Tyr	Ser	Asn	Lys	Thr	Arg	Val	Arg	Gln	His	His	Glu	His	Glu	His		
Horse III	Tyr	Ser	Asn			X	Val	X	Gln	His	His	Glu	His	Glu	His		
Chicken III	121	131	141	143	145	192	194	199	200	201	202	204	206	207	211		
Human I	Ala	Leu	Leu	Val	Gly	Trp	Tyr	Thr	His	Pro	Pro	Tyr	Ser	Val	Ile		
Rhesus I	Val	Leu	Leu	Val	Gly	Trp	Tyr	Thr	His	Pro	Pro	Tyr	Ser	Val	Ile		
Ox I	Val	Phe	Leu	Leu	Gly	Trp	Tyr	Thr	His	Pro	Pro	Leu	Ser	Val	Ile		
Horse I	Val	Phe	Leu	Ile	Gly	Trp	Tyr	Thr	His	Pro	Pro	Tyr	Ser	Val	Ile		
Turtle I	Val	Ile	Leu	Leu	Gly	Trp	Tyr	Thr	His	Pro	Pro	Ser	Ser	Val	Ile		
Human II	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val		
Ox II	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Ser	Val	Val		
Horse II	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val		
Rabbit II	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val		
Mouse II	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val		
Human III	Val	Phe	Ile	Val	Gly	TRP	Tyr	THR	THR	Pro	Pro	Cys	Cys	Ile	Leu		
Ox III	Val	Ile				Trp	Tyr	Thr	Thr	Pro	Pro	Cys	Cys	Ile	Leu		
Horse III	Val	Tyr				Trp	Tyr	Thr	Thr	Pro	Pro	Glu	Cys	Ile	Leu		

^a Residues are numbered according to CA I. Boxed residues are either invariant in all three isozymes or are unique to and invariant in a particular isozyme (e.g., Val-69 in CA III or His-200 in CA I).

Source: Biology and Chemistry of the Carbonic Anhydrases. 1984. R.E. Tashian and D. Hewett-Emmett, eds.

active site regions.

CA IV

There is a membrane-bound form of CA designated CA IV (14). Histochemical and immunohistochemical techniques have shown that membrane-bound CA is present in most electrolyte-secreting epithelia, in skeletal muscle (35,36), and in the endothelium of lung capillaries (37). CA activity is found in plasma membranes, luminal and/or anti-luminal, and in cell organelles. Membrane-bound CA has been best studied in the gastrointestinal tract (38,39) and epithelia of the kidney (40,41).

The amino acid composition of human CA IV differs from that of the cytoplasmic enzymes in that CA IV has more glutamic acid, methionine, and arginine, and less proline and lysine than CA I and CA II (18). The membrane-bound enzymes of bovine lung (14), rat saliva (42), and rat brain myelin (43) show the same pattern when compared to the cytoplasmic forms of the respective species.

The specific activity of purified human CA IV is lower than that of CA II (18). The inhibition by sulfanilamide is similar to that of CA II. CA IV of bovine lung (14) and rat brain myelin (13) have inhibitor and substrate kinetics similar to those of CA II.

CA IV is antigenically different from the cytoplasmic enzymes (18). This difference is seen for CA IV of human kidney (44) as well as for the CA IV of bovine lung (14).

Mitochondrial CA

Mitochondrial CA (CA_{mit}) has been reported in the liver of guinea pig (45). The CA_{mit} is in the matrix, it is soluble, and its activity is very dependent on pH. It is believed to be involved in citrulline synthesis in the mitochondrial matrix.

CA_{mit} is different from CA I and CA II in guinea pig in being less negatively charged but it is the same as CA I and CA II in molecular weight. CA_{mit} has approximately one third the catalytic efficiency of erythrocyte CA II, is more dependent on pH, more prone to inhibition by SO_4^- than by Cl^- and is unaffected up to .05M by the anions phosphate, citrate, and succinate at pH 7.4. The only resemblance it has to CA II is its insensitivity to chloride. CA_{mit} makes up 0.14% of the mitochondrial protein. CA_{mit} appears to be a different isozyme that may be coded for by a separate genetic locus (46).

Tissue Distribution and Function

CA I - CA V

Carbonic anhydrases have considerable variability in their patterns of tissue distribution and functional diversity in different organisms. By immunohistochemical techniques, human CA I and CA II were first found in erythrocytes (47-50). Human CA II has since been localized in normal or pathological organs of submandibular gland (51), stomach (49,51,52), appendix (49,52), liver (51), gall bladder (51), pancreas (51,53), kidney (51,52,54), trachea and bronchi (51), brain (48,55,56), retina (56,57), ciliary process (58), bone and cartilage (90), salivary gland (69), and sweat glands (51,52). CA I is found at higher levels and is more limited in its distribution. It has been

localized to appendix (51) and colon (51) in addition to erythrocytes.

CA III is primarily found in red skeletal muscle of sheep, cat and chicken (26,59), rabbit (60,61), human (62,63) and ox (27,64) but has also been found in rodent liver (65), rabbit liver and sheep lung (59), and human smooth cardiac muscle, liver, and lung (66).

Transport Epithelia. CA is found in the epithelial cells of ducts, tubules or surfaces which specialize in ion transport and lack the capacity for secretion of macromolecules. An example of such a cell is the gastric parietal cell. Immunocytochemical staining shows CA to be present in the cytoplasm (51,67) and histochemical staining shows CA to be present in the membrane (38) of the parietal cell. CA is thought to generate protons which the parietal cell secretes across the apical plasmalemma into the lumen. The cell releases the bicarbonate ions, also generated by CA, across the basolateral plasmalemma alkalinizing the interstitial fluid at the surface of the epithelium (68).

Macromolecule-Secreting Epithelia. CA has been immunocytochemically located in epithelial cells which contain granules for storing macromolecules which the cells secrete by a merocrine secretory mechanism (68). Such CA rich merocrine cells include cells of the tracheobronchial and submandibular glands in man (51), parotid, submandibular, and exorbital lachrymal glands in mouse and rat (69). Since CA has actually been found in the saliva of certain species such as sheep (70), there is a possibility that CA may be located in the secretory granule and secreted along with the other molecules.

Moreover, a glycosylated CA has been reported in rat salivary gland; the carbohydrate association expected of a secretory product.

CA has also been found in glycoprotein-secreting cells such as the goblet cells of the guinea pig intestine (68). In the secretory cells CA may have a role in dehydrating the secretory product.

Nonepithelial Cells. Immunocytochemistry shows that CA is distributed throughout the cytoplasm in erythrocytes, adipocytes, oligodendroglia, and skeletal muscle (48,67,71,72). In erythrocytes CA correlates with the specialization of the cell in respiratory gas exchange. Its role in adipocytes is uncertain; it may influence intracellular pH (68).

There are high levels of CA in oligodendrocytes and myelin sheath of brain. CA's role may involve ion fluxes which are fundamental to neural reactivity. Since impulse conduction by axons is dependent on depolarization mediated by Na^+ influx, CA in myelin and oligodendroglia sheathing the axon may be involved in maintaining ion concentration at the outer surface of the axolemma. CA has been found both in bound and soluble forms in the central nervous system (73,74).

Cobalt-trapping experiments show that CA is in the sarcolemma of all and in the sarcoplasm of some muscle fibers (35). Immunostaining shows CA II to be present in slow twitch fibers (75). CA is thought to provide buffering action in the site of rapid CO_2 production.

Unique CA's

CA's that are distinct from the isozymes CA I, CA II, CA III, membrane-bound CA, and mitochondrial CA have greater molecular weights. These newly identified CA's include ovine parotid gland CA, rat saliva CA, rabbit erythrocyte CA and plant CA.

a. As mentioned in the previous section, the ovine parotid gland which secretes bicarbonate in saliva has been shown to contain a cytoplasmic high molecular weight CA (49) in addition to CA II. The native molecular weight of the new enzyme is 540,000 and its subunit molecular weight is 45,000. The enzyme is similar to CA II in that it has zinc and is inhibited by sulfonamide but its amino acid composition shows no homology to CA II; it has 3 cysteine residues per subunit whereas cysteine is usually absent from CA's. It is also immunologically distinct.

b. Rat saliva CA has a higher molecular weight than rat erythrocyte CA. Up to 35% of the saliva CA is composed of oligosaccharide components. Its amino acid sequence shows more homology to bovine lung membrane-bound CA than to erythrocyte CA (22).

c. A 54,000 molecular weight CA has been found in rabbit erythrocytes (77). The CA appears to be composed of the 30,000 molecular weight CA I (soluble CA I being at a reduced level) and a 24,000 molecular weight binding protein. The bound CA appears to be associated with the erythrocyte membrane where it may be involved in increasing the diffusion rate of CO_2 or HCO_3^- across the membrane.

d. Plant CA's differ in a number of ways from animal CA's (78). Plant CA's have a wide range of molecular weights from 42K-250K, usually 180K

(78). The plant CA's have a similar amino acid content, subunit size, zinc content, to the CA's in animals (78). However, the plant CA's have no esterase activity and as a whole are not strongly inhibited by sulfonamides. The CA's can be separated by gradient polyacrylamide gel electrophoresis into two separate isozymes, a chloroplast and a cytosol form. The roles of CA in plants may be 1) to facilitate diffusion of CO_2 to the site of photosynthetic fixation, 2) regulate pH, 3) provide bicarbonate in underwater plants, or 4) concentrate inorganic carbon within the chloroplast (78).

Expression During Development

The levels of the CA isozymes change during development. CA I and CA II are both present in human fetal blood cells as early as the 22nd week of gestation (79). CA I reaches adult levels more slowly when compared to the rise of hemoglobin (Hb) A levels after birth. Hb A reaches adult levels at age 6 months whereas CA I has 30% the activity of adult levels at that time. CA I activity is 60% of adult levels after 1 year and reaches adult levels after 5-6 years of life.

CA III appears in muscle extracts 8 days after birth (59). CA III activity is 50% of adult levels 21 days after birth.

In cats, CA II is the only isozyme present in a mid-term fetus (26). CA III appears in muscle in the late stages and CA I appears in cecum within 2 days of birth.

The location of CA II in cell types changes during the development of the vertebrate retina (80). By immunocytochemical staining, CA II is found very early in all retinoblast cells. Later, CA becomes restricted to certain cell types with the cell type depending on the

vertebrate class. In adult mammals CA II is found in a specific neuronal cell type as well as in a type of satellite glial cell. As one goes down the vertebrate classes CA becomes confined to the glial cells and then to the primitive horizontal cells. The variability of CA II localization during development may indicate a varied role of CA in tissue functions. The early generalized CA II expression in chick and mouse retinas may indicate a CA role in eye morphogenesis by influencing intraocular pressure.

CA has been found to be involved in formation of avian eggshell (81). During shell formation blood bicarbonate is transformed into insoluble calcium carbonate and carbonic acid. The carbonic acid is rapidly broken down by CA to H_2O and CO_2 to prevent the eggshell from redissolving by the reversible reaction. CA has been found to be involved in calcification reactions in other species such as the mollusks (82), fresh water snails (83), barnacles (84) and reef corals (85).

CA has also been found to be involved in calcium transport and dissolution of avian eggshell (86). During chick embryonic development calcium is supplied to the embryo by two sources. During the first half of gestation (up to day 9 or 10) the yolk supplies calcium. During the second half of gestation (days 11-12) shell calcium ($CaCO_3$) is mobilized and embryonic calcium content increases. CA activity has been found to parallel calcium uptake by the choriallantoic membrane (CAM). Moreover, sulfonamide inhibits calcium uptake as well as CA activity in CAM (87,88,89). Chick erythrocyte CA and CAM CA have the same molecular weight but are antigenically different and are different

in CO₂ hydration. CAM CA is thought to be CA I.

CA also appears to be involved in bone resorption (90). The carbonic acid produced by CA may be the acid secreted by osteoclasts to dissolve mineral. Bone resorption by osteoclasts is accompanied by H⁺ release and greater CA activity. Further support of CA involvement is a deficiency of CA II which is associated with osteopetrosis where bone resorption is impaired (91).

CA II has been identified in the mouse brain (glia) by immunological, electrophoretic, and kinetic studies (73,92,93). Over 50% of the CA in the central nervous system (CNS) glia is membrane-bound (94,95) and may be an intrinsic membrane protein. Both soluble and membrane-bound CA were purified from mouse brain (13). Both forms had the same specific activity, esterase activity, the same molecular weight and isoelectric point which was that of a CA II isozyme. However, pulse chase experiments of CA biosynthesis showed that one form was not a modification of the other.

CA activity has been found in myelin (73,74,98) in several species and at several developmental stages. The highest CA activities in myelin occur in the mouse. In rat, myelin from spinal cord has lower CA activity than brain myelin (95,97). Immunocytochemical staining is consistent with the biochemical data that the primary sites of CA in the CNS is in oligodendroglia and myelin.

Expression in Erythrocytes

Variability in Expression of Isozymes

CA is the second most abundant protein in the red blood cell (RBC) next to hemoglobin (98). There are 1-2 grams of CA in a liter of

mammalian RBC (1,99,100).

In mammalian erythrocytes there are two isozymes of CA present-- the low activity isozyme CA I and the high activity isozyme CA II which has 5-10 times the activity of CA I (27). It has been shown that both isozymes contribute substantially to the overall activity (101). The physiological or metabolic significance of the presence of the two isozymes in erythrocytes is not known since they appear to perform a straightforward role.

There is considerable variation in the CA I and CA II protein levels in mammalian red cells and each species has a characteristic level. In humans and many other primates the CA I level is greater than CA II, whereas in some inbred mouse strains the CA II level is greater. In primates the ratio of CA I:CA II can range from 2.6:1 in macaque monkeys to 27:1 in orangutan. The ratio in humans is 6:1 (27).

CA I is absent from the RBC of some mammalian species. Only CA II is present in the red cell in two groups of mammals, ruminants (ox, sheep, goat, elk, deer) and felids (cat, tiger, lion, jaguar, leopard) (27). Dog, dolphin, chicken, and shark RBC also lack CA I (24). Species whose erythrocytes contain both CA I and CA II are primates (man, green monkey, pig-tailed macaque, baboon, spider monkey, rhesus monkey), deer mouse, pig, rabbit, rat, guinea pig, and horse (24). Although CA I is absent in blood it is not necessarily absent in other tissues. CA I is absent from bovine blood but is present in very high concentration in bovine rumen epithelium (24).

There are inherited deficiencies of red cell CA I in humans, pigtail macaques, and chinchilla (98,102) where there are no detectable

defects associated with the absence of CA I in mature red cells. Therefore, although CA I may have alternate functions in other tissues, it does not appear to be essential for adequate rates of CO₂ exchange (27). CA II appears to have been more strongly selected for in mammalian red cells.

Changes in Isozyme Levels During Red Cell Maturation

CA activity increases as cells mature from erythroblasts to reticulocytes during erythroid development in rabbit marrow (103). When lysates from the three major developing cell types (basophilic, polychromatic, and orthochromatic erythroblasts) and reticulocytes in rabbit marrow are measured for the levels of CA I and CA II by immunoassay, CA II levels remain unchanged in erythroblasts but increase in reticulocytes. CA I shows a progressive increase in the developing cells. The increased activity of CA during erythroid development cannot be accounted for on the basis of the isozyme levels alone since CA II has a 20-fold higher specific hydrase activity than CA I. Immunoreactive precursors of CA II may be inactive, for example, because they lack zinc which is at lower levels in early erythroblasts (104). CA accumulates earlier than hemoglobin in erythroid cells and at a lower rate. The increase in CA activity continues after the late polychromatic stage after termination of RNA synthesis so the half-life of CA mRNA appears to be fairly long.

Quantitative Genetic Variation

There are two mutations in humans that result in the absence of CA I or CA II in erythrocytes in individuals who are homozygous for the

defective gene. A deficiency gene of CA I has been found in a Greek family (102). Individuals homozygous for the deficiency gene show no clinical symptoms and have a normal level of CA II in the red cell (105). The CA I/CA II ratio in the red cell of heterozygous individuals are half those of normal individuals. Similarly, absence of CA I in the red cells of pigtail macaques homozygous for a CA I deficiency gene has no deleterious effect (98,106).

It has not yet been determined if CA I in a CA I-deficient individual is deficient in the other tissues in which it is normally present. Cultured lymphocytes from CA I-deficient humans show strong fluorescence using a fluorescent antibody technique (27). This result suggests that CA I-deficiency in red cells may be due to CA I protein or mRNA being labile and rapidly degraded during erythropoiesis. Degradation may not be as severe in different cells where the kinds and or levels of proteolytic enzymes or ribonucleases may be different (107).

Unlike the CA I deficiency, CA II deficiency individuals show an abnormality of bone, kidney, and brain. Individuals who are homozygous for a CA II-deficiency gene have the syndrome osteopetrosis with renal tubular acidosis and cerebral calcification (91). There have been a total of 11 families that have been diagnosed with this condition. Osteopetrosis was present in all the homozygous individuals but there was variation in the severity of renal tubular acidosis and in the degree of cerebral calcification (108). The CA I/CA II ratios of the heterozygotes (91,109) were twice that of normal individuals.

Interestingly, CA II is found in many other tissues yet only bone, kidney, and brain are affected in the CA II deficiency. If each CA isozyme is a product of a single, separate locus one would expect a defect in the gene to be present in every cell in which the isozyme is expressed. Furthermore, an organ such as the eye which contains mainly the CA II isozyme appears to lack CA II in the CA II deficiency yet be able to function without it (Sly, Whyte and Krupin, unpub. results). The reasons for the limitation are not known but possible explanations have been offered: 1) There may be alternate mechanisms that could substitute for CA II function in other tissues; CA I takes over the role of CA II in red cells. 2) If mutation caused the protein or mRNA to be susceptible to degradation, then the CA II turnover levels in a particular cell would depend on CA II turnover, or on the type of proteolytic enzymes (107) or ribonucleases in the cell. 3) There may be two or more CA II or CA II-like isozymes that are coded for by separate genes that are differentially expressed in different tissues. (There may be more than one CA II gene in the mouse (110).)

The level of CA I is slightly elevated in erythrocytes of a CA II-deficient individual and the CA III level is doubled (the relative levels of CA I, CA II and CA III isozymes in adult human red cells are 85:13:1, respectively)(111). However, since CA I and CA III have low specific activities and are thought to be inhibited by chloride ions in the red cell (112,113), they contribute only 10% of the total CA activity of the human red cell. Surprisingly, it appears that red cells of CA II-deficient individuals function normally. There appears to be a compensatory mechanism for respiratory function of CA under

unstressed conditions. (CA activity of homozygous individuals went from 22% of control level at 25°C to 55% of control level at 37°C.) Chloride inhibition of CA I may be relaxed in the CA II-deficient individual.

Qualitative Genetic Variation

There is considerable genetic variability for CA in mammals. In humans there are 25 CA I electrophoretic variants (105). The majority of these variants do not show any difference in CO₂ hydrase or specific esterase activity with the exception of one variant which is an activity variant rather than a mobility one. There are seven electrophoretic variants of human CA II (105). The amino acid substitutions of some of these variants have been determined and the substitutions as a whole are conservative in their effect in changing secondary structure. Electrophoretic and quantitative variation occur quite frequently at the CA locus in the higher apes other than humans (9), with greater variability occurring at the CA I locus than at the CA II locus (64).

Hormonal Control

A CA III isozyme which is immunochemically indistinguishable from rat muscle CA III is present in mature rat liver (65) and appears to be hormonally controlled (116). This isozyme is a major liver enzyme comprising 8% of cytosol protein (117). The concentration of CA III in homologous muscle from male and female rats was similar but the concentration of CA III in mature male rat liver was 30 times greater than in mature female rat livers by radioimmunoassay.

Sex hormone levels were manipulated in gonadectomized rats and CA III concentration was monitored by radioimmunoassay. Testosterone administration resulted in an increase in CA III level in female rats and estradiol administration resulted in a decrease in CA III levels in male rats.

When the amount of mRNA in rat liver was quantified by immunoprecipitation there was nine times more translatable CA III mRNA present in male rat liver than in female rat liver (116). These results indicated that the androgen-mediated difference in CA III expression may be at the level of translatable CA III mRNA.

CA III levels may also vary during rat development. CA III levels are low in prepuberty rats and increase 7-fold to become 1% of the total liver protein in adult rat (116). However, different control mechanisms may operate in mouse and human CA III levels since radioimmunoassay of human male and female livers show similar levels of CA III.

Evolution of Carbonic Anhydrases

The high degree of evolutionary homology between the structures of CA I and CA II isozymes indicates that CA I and CA II originated from a common ancestral gene (27). The greater enzymic, immunological, and chemical variability among mammalian CA I isozymes and the absence of CA I in some vertebrate blood suggests that CA II is an evolutionarily older molecule (118). CA I may have arisen by gene duplication.

It appears that there was an ancestral CA with an active site similar to that of CA II isozyme. After gene duplication both CA I and CA II lineages fixed important substitutions in their active sites that

are presumably responsible for their different properties. Since the mammalian radiation, CA I, and to a lesser extent CA III, have continued to fix substitutions in their active sites. If the molecule is considered as a whole, it appears that the CA III lineage evolved rapidly after duplication. Recently, since the mammalian radiation, CA III has been most conserved, and CA II the least conserved of the three CA isozymes (33).

THE CARBONIC ANHYDRASE GENE

Mouse CA II Gene

The first structural work at the nucleic acid level on CA II was done in mouse with the isolation of mouse CA II cDNA, pMCA II (119). The cDNA was prepared to size fractionated anemic spleen poly(A)⁺RNA and identified by hybridization selection. The predicted amino acid sequence from the cDNA sequence showed homology to known amino acid sequences of rabbit and human CA I and CA II. The cDNA clone, pMCA II, was 1500 bp and contained sequence from the coding region as well as 700 bp of sequence at the 3' non-coding region of the mRNA. To obtain the 5'-end of the cDNA, a library of mouse cDNA sequences derived from poly(A)⁺RNA of anemic spleen was screened using the mouse cDNA (pMCA II) as a probe (120). Two clones were selected that had more of the 5'-region but lacked sequence at the 3'-end.

The complete amino acid sequence of the mouse CA gene was determined from the nucleotide sequence data (120). The amino acid sequence was compared to residues which are invariant in a given isozyme but differ at this position in the other two isozymes. The

mouse CA had one residue in common with the 21 unique and invariant residues of the CA I isozyme, no residues in common with the 31 unique residues of CA III, and 16 of the 23 unique residues of CA II. pMCA II was thus confirmed to code for CA II.

Comparison of the predicted mouse amino acid sequence to that of human, sheep, ox, rabbit, and horse showed that 33 residues were unique to mouse (120). Based on the human CA II three-dimensional structure, all but 5 of the 33 residues are located on the surface of the human CA II molecule.

Isolation of Mouse CA II Gene

Cosmid and lambda libraries were screened using labeled pMCA II as the probe and four cosmid clones, A5, A6, 54, 103, and four lambda clones were isolated and studied (121). Cosmid clone A6 (cosMCAII) was found to contain a complete mouse CA II gene by comparing hybridizing bands of the cosmid with those seen on genomic blots. A6 contains the 16 kb-long gene, 12 kb of 5'-flanking region, and 10 kb of 3'-flanking region.

The mouse carbonic anhydrase II gene was found to have 7 exons and 6 introns (Figure 2) (121). The three introns A, B, and D interrupt the codons for Gly-11, Val-77, and Gly-143, respectively. Introns C, E, and F fall between the codons for Glu-116 and Leu-117, Lys-168 and Gly-169, and Gln-220 and Met-221, respectively. The exons range in size from 76 bp of exon 5 to 799 bp of exon 7. The introns range in size from 0.3 kb of intron C to 7.2 kb of intron B. There is a TATA box at position -92 bp and a putative "CAAT" box (CCACT) at position -140 bp upstream from the ATG initiation codon, similar to the location

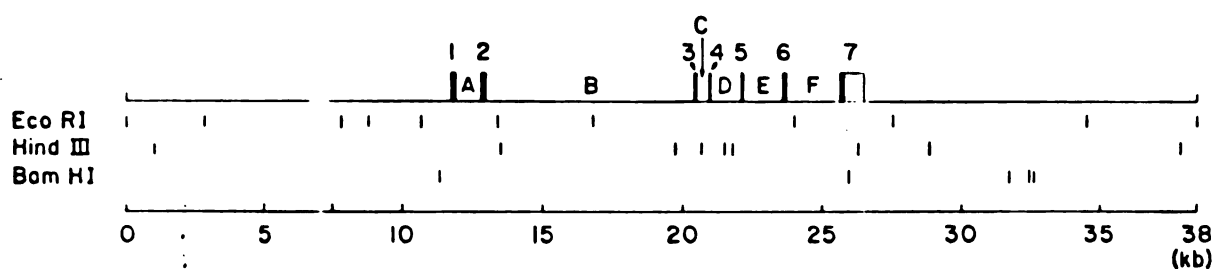


Figure 2. Restriction endonuclease map and exon positions of the mouse carbonic anhydrase II gene in cosmid cosMCAII-2. The solid boxes represent the coding regions and the open boxes are the untranslated regions of the exons. The exons are designated by numbers and the introns by capital letters. (From 121)

of consensus sequences common to eukaryotic genes particularly the chicken globin genes.

A comparison of the nucleotide sequence of the YBR mouse CA II gene exons and the cDNA sequence for Balb/c mouse CA II cDNA revealed strain differences represented by four point mutations (121). Two of the point mutations were silent. A third point mutation was in the 3'-untranslated region and a fourth at residue 38 resulted in an amino acid change (Gln/His). This change in amino acid would result in a change in electrophoretic mobility and may explain the electrophoretic variation of CA II in many inbred mouse strains.

Rabbit CA I cDNA

Low-abundance CA I mRNA from rabbit reticulocyte poly(A)⁺RNA was purified by immunoprecipitation of specific nascent polypeptides attached to rabbit reticulocyte polyribosomes. The purified mRNA was translated in a cell-free system and the protein product (³H-labeled) was fractionated on a polyacrylamide gel to confirm its identity by molecular weight (122).

The purified mRNA was used to identify homologous clones from size-selected cDNA libraries prepared from rabbit reticulocytes (122). The cDNA clones that hybridized to radiolabeled CA I probes but not to globin-specific probes were further studied and their identity verified in hybrid selection experiments. Identification was also performed by hybridizing nick translated CA I cDNA to RNA that had been gel fractionated, transferred to filters, and identifying the size of the RNA.

The cDNA insert was 300 bp and probably corresponds to the 3'-end of the mRNA. Therefore, it probably contains a noncoding region, and definitive proof of its identity could not be obtained by comparing amino acid sequence to known amino acid sequence. The cDNA clone was used to screen a new rabbit reticulocyte cDNA library in an attempt to isolate larger cDNA inserts (122).

Comparison of Mouse and Human CA II Gene

The mouse CA II gene has been examined at the DNA level (121). A comparison was made of mouse and human CA II (110). Since 81% of the amino acid residues of mouse and human CA II isozymes are identical at homologous positions and the three-dimensional structures of human CA I and CA II are very similar (60% amino acid sequence identity), it was assumed that the mouse and human CA II tertiary protein structures were identical (121). Also, mouse and human CA II genes have introns A and B in identical positions. (Part of the human CA II gene has been analyzed to date.) It was presumed that active site residues of human CA II occupied corresponding positions in the active site of mouse CA II.

The human CA I and CA II molecules have a large portion of β -structure which runs through the entire molecule (2). One-half of the active site cavity is formed from hydrophobic side chains. The other half of the active site is composed of exposed residues which are mainly hydrophilic.

The hydrophilic region is mostly coded by exons 2 and 3 and the hydrophobic region is mostly coded by exons 4 and 6. The three

histidine residues in the active site region which bind the zinc ion are encoded by exon 3 and exon 4. It appeared that there was no correspondence between exons and domain structure for CA (121).

Also, the three dimensional structure of human CA I and II proteins showed that the splice junction sites for the mouse CA II gene were located relative to the protein amino acid sequence on both the outside and inside of the molecule (121). Splice junctions for introns A, B, E, and F are located on the outside of the molecule and splice sites for introns C and D are located at the bottom of the active site cavity. These results are contrary to the idea that intron-exon junctions map to protein surfaces.

REFERENCES

1. Maren, T.H. 1967. *Physiol. Rev.* 47:595-781.
2. Lindskog, S., L.E. Henderson, K.K. Kannan, A. Liljas, P.O. Nyman and B. Strandberg. 1971. pp. 587-665. In *The Enzymes*. P.D. Boyer, ed., Academic Press, New York.
3. Tashian, R.E., C.C. Plato, T.B. Shows. 1963. *Science*. 140:53-54.
4. Tashian, R.E., D.P. Douglas and Y.-S.L. Yu. 1964. *Biochem. Biophys. Res. Commun.* 14:256-261.
5. Schneider, F. and M. Lieflander. 1963. *Z. Physiol. Chem.* 334:279-282.
6. Pocker, Y. and N. Watamori. 1973. *Biochemistry*. 12:2475-2482.
7. Pocker, Y. and L.J. Guilbert. 1974. *Biochemistry*. 13:70-78.
8. Kaiser, E.T. and K.W. Lo. 1969. *J. Am. Chem. Soc.* 91:4912-4918.
9. Pocker, Y. and J.E. Meany. 1965. *J. Am. Chem. Soc.* 87:1809-1811.
10. Pocker, Y. and J.E. Meany. 1967. *Biochemistry*. 6:239-246.
11. Pocker, Y., J.E. Meany and B.C. Davis. 1974. *Biochemistry* 13:1411-1416.
12. Tashian, R.E., D. Hewett-Emmett and M. Goodman. 1983. pp. 79-100. In *Isozymes: Current Topics in Biological and Medical Research*. M.C. Rattazzi, J.G. Scandalios and G.S. Whitt, eds., Alan R. Liss, Inc., New York.
13. Sapirstein, V.S., P. Strocchi and J.M. Gilbert. 1984. *Ann. N.Y. Acad. Sci.* 429:481-493.
14. Whitney, P.L. and T.V. Briggie. 1982. *J. Biol. Chem.* 257:12056-12059.
15. Henry, R.P. and J.N. Cameron. 1983. *J. Exp. Biol.* 103:205-223.

16. McKinley, D.N. and P.L. Whitney. 1976. *Biochem. Biophys. Acta.* 445:780-790.
17. Sanyal, G, N.I. Pessah and T.H. Maren. 1981. *Biochem. Biophys. Acta* 657:128-137.
18. Wistrand, P.J. 1984. *Ann. N.Y. Acad. Sci.* 429:195-206.
19. Dodgson, S.J., R.E. Forster, II, B.T. Storey and L. Mela. 1980. *Proc. Natl. Acad. Sci. USA* 77:5562-5566.
20. Vincent, S.H. and D.N. Silverman. 1982. *J. Biol. Chem.* 257:6850-6855.
21. Dodgson, S.J., J.M. Kamerling, S. Nioka and R.E. Forster, II. 1984. *Fed. Proc. USA* 43:300.
22. Feldstein, J.B., and D.N. Silverman. 1984. *Ann. N.Y. Acad. Sci.* 429:214-215.
23. Anderson, B., P.O. Nyman and L. Strid. 1972. *Biochem. Biophys. Res. Commun.* 48:670-677.
24. Carter, N. 1972. *Biol. Rev.* 47:465-513.
25. Tashian, R.E., M. Goodman, R.J. Tanis, R.E. Ferrell and W. Osborne. 1975. pp. 207-223. *In* *Isozymes*. C.L. Markert, ed. Academic Press, New York.
26. Holmes, R.S. 1977. *Eur. J. Biochem.* 78:511-520.
27. Tashian, R.E. 1977. pp. 21-62. *In* *Isozymes: Current Topics in Biological and Medical Research*. M.C. Rattazzi, J.G. Scandalios, and G. S. White, eds. Alan R. Liss, Inc., New York.
28. Maren, T.H., C.S. Rayburn and M.E. Liddell. 1976. *Science.* 191:469-472.
29. Kannan, K.K., B. Notstrand, K. Fridborg, S. Lovgren, A. Ohlsson and M. Petef. 1975. *Proc. Natl. Acad. Sci. U.S.A.* 72:51.
30. Notstrand, B., I. Vaara and K.K. Kannan. 1975. *In* *Isozymes*. C.L. Markert, ed. 1:575-599. Academic Press, New York.
31. Kannan, K.K., A. Liljas, I. Waara, P.C. Bergsten, S. Lovgren, B. Strandbert, U. Bengtsson, U. Carlbom, K. Fridborg, L. Jarup and M. Petef. 1971. *Cold Spring Harbor Symp. Quant. Biol.* 36:221.
32. Liljas, A., K.K. Kannan, P.C. Bergten, K. Fridborg, B. Strandberg, U. Carlbom, L. Jarup, S. Lovgren and M. Petef. 1972. *Nature (London) New Biol.* 235:131-137.

33. Hewett-Emmett, D., P.J. Hopkins, R.E. Tashian and J. Czelusniak. 1984. *Ann. N.Y. Acad. Sci.* 429:338-358.
34. Lindskog, S., P. Engberg, C. Forsman, S.A. Ibrahim, B.H. Jonsson, I. Simonsson and L. Tibell. 1984. *Ann. N.Y. Acad. Sci.* 429:61-75.
35. Lonnerholm, G. 1980. *J. Histochem. Cytochem.* 28:427-433.
36. Ridderstrale, Y. 1979. *Acta Physiol. Scand.* 106:239-240.
37. Ryan, U.S., P.L. Whitney and J.W. Ryan. 1982. *J. Appl. Physiol.* 53:914-919.
38. Sugai, W. and S. Ito. 1980. *J. Histochem. Cytochem.* 28:511-525.
39. Lonnerholm, G. 1983. *Acta Physiol. Scand.* 117:273-279.
40. Ridderstrale, Y. 1976. *Acta Physiol. Scand.* 98:465-469.
41. Lonnerholm, G. and Y. Ridderstrale. 1980. *Kidney Int.* 17:162-174.
42. Feldstein, J.B. and D.N. Silverman. 1984. *Ann. N.Y. Acad. Sci.* 429:214-215.
43. Sapirstein, V.S. and M.B. Lees. 1978. *J. Neurochem.* 31:505-517.
44. Wistrand, P.J. 1980. *Uppsala J. Med. Sci.* 85:75.
45. Dodgson, S.J., R.E. Forster, II, B.T. Storey and L. Mela. 1980. *Proc. Natl. Acad. Sci. U.S.A.* 77:5562-5566.
46. Storey, B.T., S.J. Dodgson and R.E. Forster, II. 1984. *Ann. N.Y. Acad. Sci.* 429:210-211.
47. Hansson, H.P. 1965. *Life Sci.* 4:965-968.
48. Kumpulainen, T. and L.K. Korhonen. 1978. *Histochemistry* 58:183-192.
49. Kumpulainen, T. 1979. *Histochemistry.* 62:271-280.
50. Panero, C., G. Bonferraro, E.C. Biagioli and L. Burrioni. 1974. *Helv. Paediatr. Acta.* 29:157-166.
51. Spicer, S.S., M.A. Sens and R.E. Tashian. 1982. *J. Histochem. Cytochem.* 30:864-873.
52. Kumpulainen, T. 1981. *Histochemistry.* 72:425-431.

53. Kumpulainen, T. and P. Jalovaara. 1981. Gastroenterology. 80:796-799.
54. Jonas, L., L. Hoffman and D. Serfling. 1983. Z. Urol. Nephrol. 76:311-317.
55. Kumpulainen, T. and S.H.M. Nystrom. 1981. Brain Res. 220:220-225.
56. Kumpulainen, T., D. Dahl, L.K. Korhonen and S.H.M. Nystrom. 1983. J. Histochem. Cytochem. 31:879-886.
57. Kumpulainen, T. 1980. Acta Ophthalmol. (kbh) 58:397-405.
58. Kumpulainen, T. 1983. Histochemistry. 77:281-284.
59. Holmes, R.S. 1976. J. Exp. Zool. 197:289-295.
60. Koester, M.K., L.M. Pullman and E.A. Nottmann. 1981. Arch. Biochem. Biophys. 211:632-642.
61. Register, A.M., M.K. Koester and E.A. Nottmann. 1978. J. Biol. Chem. 253:4143-4152.
62. Carter, N., A. Shiels and R.E. Tashian. 1978. Biochem. Soc. Trans. 6:552-553.
63. Carter, N.D., S. Jeffrey, A. Shiels, Y. Edwards, T. Tipler, and D.A. Hopkinson. 1979. Biochem. Genet. 17:837-854.
64. Tashian, R.E., S.K. Stroup., Y.-S.L. Yu and D. Henricksson. 1978. Fed. Proc. 37:1797.
65. Carter, N.D., D. Hewett-Emmett, S. Jeffrey and R.E. Tashian. 1981. FEBS Lett. 128:114-118.
66. Jeffrey, S., Y. Edwards and N. Carter. 1980. Biochem. Genet. 18:843-849.
67. Spicer, S.S., P.J. Stoward and R.E. Tashian. 1979. J. Histochem. Cytochem. 27:820-831.
68. Spicer, S.S., M.A. Sens, R.A. Hennigar and P.J. Stoward. 1984. Ann. N.Y. Acad. Sci. 429:382-397.
69. Hennigar, R.A., B.A. Schulte and S.S. Spicer. 1983. Anat. Rec. 207:605-614.
70. Fernley, R.T., R.D. Wright and J.P. Coghlan. 1979. FEBS Lett. 105:299-302.

71. Roussel, G., J.-P. Delaunoy, J.-L. Nussbaum and P. Mandel. 1979. Brain Res. 160:47-55.
72. Vaananen, H.K., T. Kumpulainen and L.K. Korhonen. J. Histochem. Cytochem. 30:1109-1113.
73. Cammer, W.T., T. Fredman, A.L. Rose and W.T. Norton. 1976. J. Neurochem. 27:165-171.
74. Yandrasitz, J.R., S.A. Ernst and L. Salganicoff. 1976. J. Neurochem. 27:707-715.
75. Perlman, D. and H.O. Halvorson. 1981. Cell 25:525-536.
76. Fernley, R.T., M. Congiu, R.D. Wright and J.P. Coghlan. 1984. Ann. N.Y. Acad. Sci. 429:212-213.
77. Schafer, A. and P. Dietsch. 1984. Ann. N.Y. Acad. Sci. 429:241-242.
78. Graham, D., M.L. Reed, B.D. Patterson and D. Hockley. 1984. Ann. N.Y. Acad. Sci. 429:222-237.
79. Wehinger, H. 1973. Blut 27:172-185.
80. Linser, P. and A.A. Moscona. 1984. Ann. N.Y. Acad. Sci. 429:430-446.
81. Benesch, R.N., N.S. Barron and C.A. Mawson. 1944. Nature. 153:138.
82. Wilbur, K. and L. Jodrey. 1955. Biol. Bull. Woods Hole Mass. 108:359.
83. Freeman, J. 1960. Biol. Bull. Woods Hole Mass. 118:412.
84. Costlow, J.D. 1959. Physiol. Zool. 32:177.
85. Goreau, T. 1961. Endeavor. 20:32.
86. Tuan, R. 1984. Ann. N.Y. Acad. Sci. 429:459-472.
87. Tuan, R. and J. Zrike. 1978. Biochem. J. 176:67-74.
88. Crooks, R. and K. Simkiss. 1975. Q. J. Exp. Physiol. 60:55-63.
89. Crooks, R., C. Kyriakides and K. Simkiss. 1976. Q. J. Exp. Physiol. 61:265-274.
90. Gay, C., H. Schraer, R.E. Anderson and H. Cao. 1984. Ann. N.Y. Acad. Sci. 429:473-478.

91. Sly, W.S., D. Hewett-Emmett, M.P. Whyte, Y.-S. Yu and R.E. Tashian. 1983. Proc. Natl. Acad. Sci. U.S.A. 80:2752-2756.
92. Sapirstein, V.S., P. Strocchi, M. Wesolowski and J.M. Gilbert. 1983. J. Neurochem. 40:1251-1261.
93. Funakoshi, S. and H.F. Deutsch. 1971. J. Biol. Chem. 246:1088-1092.
94. Trachtenberg, M.C. and V.S. Sapirstein. 1980. Neurochem. Res. 5:573-577.
95. Sapirstein, V.S., M.C. Trachtenberg, M.B. Lees and O. Koul. 1978. Adv. Exp. Biol. 100:55-68.
96. Sapirstein, V.S., M.B. Lees and M.C. Trachtenberg. 1978. J. Neurochem. 31:283-287.
97. Cammer, W. and T.R. Zimmerman, Jr. 1983. Develop. Brain Res. 6:21-26.
98. Tashian, R.E. and N.D. Carter. 1976. pp. 1-56. In Advances in Human Genetics. H. Harris and K. Hirshorn, eds. Plenum Press, New York.
99. Keilin, D. and T. Mann. 1940. Biochem. J. 34:1163.
100. Rickli, E.E., S.A.S. Ghazanfar, B.H. Gibbons and J.T. Edsall. 1964. J. Biol. Chem. 239:1065.
101. Edsall, J.T. 1968. Ann. N.Y. Acad. Sci. 151:41-63.
102. Kendall, A.G. and R.E. Tashian. 1977. Science 197:471-472.
103. Denton, M.J., N. Spencer and H.R.V. Arnstein. 1975. Biochem. J. 146:205-211.
104. Spencer, N. and S. Peller. 1976. Biochem. Soc. Trans. 4:1153-1155.
105. Tashian, R.E., A.G. Kendall and N.D. Carter. 1980. Hemoglobin. 4:635-651.
106. Ferrell, R.E., W.R.A. Osborne and R.E. Tashian. 1981. Proc. Soc. Exp. Biol. Med. 168:155-158.
107. Buetler, E. 1983. Proc. Natl. Acad. Sci. U.S.A. 80:3767-3768.
108. Whyte, M.P., W.A. Murphy, M.D. Fallon, W.S. Sly, S.L. Teitelbaum, W.M. McAlister and L.V. Avioli. 1980. Am. J. Med. 69:64-74.

109. Hewett-Emmett, D. 1982. Fed. Proc. 41:1385.
110. Venta, P.J., J.C. Montgomery, K. Wiebauer, D. Hewett-Emmett and R.E. Tashian. 1984. Ann. N.Y. Acad. Sci. 429:309-323.
111. Carter, N.D., R. Heath, R.J. Welty, D. Hewett-Emmett, S. Jeffrey, A. Shiels and R.E. Tashian. 1984. Ann. N.Y. Acad. Sci. 429:284-286.
112. Wistrand, P.J. 1981. Acta Physiol. Scand. 113:417-426.
113. Maren, T.H. and E.O. Couto. 1979. Arch. Biochem. Biophys. 196:501-510.
114. Heath, R., N.D. Carter, D. Hewett-Emmett, E. Fincani, S. Jeffrey, A. Shiels and R.E. Tashian. 1983. Fed. Proc. 42:2180.
115. Hewett-Emmett, D., R.J. Welty and R.E. Tashian. 1983. Genetics. 105:409-420.
116. Carter, N.D., R. Heath, R.J. Welty, D. Hewett-Emmett, S. Jeffrey, A. Shiels and R.E. Tashian. 1984. Ann. N.Y. Acad. Sci. 429:287-301.
117. Shiels, A., S. Jeffrey, I.R. Phillips, E.A. Shephard, C.A. Wilson and N.D. Carter. 1983. Biosci. Rep. 3:475-478.
118. Tashian, R.E., D. C. Schreffler and T.B. Shows. 1968. Am. J. Hum. Genet. 17:259-72.
119. Curtis, P. 1983. J. Biol. Chem. 258:4459-4463.
120. Curtis, P.J., E. Withers, D. Demuth, R. Watt, P.J. Venta and R.E. Tashian. 1984. Gene 25:325-332.
121. Venta, P.J., J.C. Montgomery, D. Hewett-Emmett, K. Wiebauer and R.E. Tashian. 1985. J. Biol. Chem. 260:12130-12135.
122. Boyer, S.H., H. Ostrer, K.D. Smith, K.E. Young and A.N. Noyes. 1984. Ann. N.Y. Acad. Sci. 429:324-331.

CHAPTER 2

ISOLATION OF THE CHICKEN CARBONIC ANHYDRASE II GENE*

Corinne M. Yoshihara, Mark Federspiel, and
Jerry B. Dodgson

The carbonic anhydrase (CA) gene family displays considerable variation in its expression pattern. In amniotes (birds, reptiles, and mammals) the three genetic loci that have been identified that encode isozymes CA I, CA II, and CA III, vary in their expression between classes of amniotes and in tissues within a class. The genes for CA I and CA II are both expressed in most mammalian red blood cells (RBC), for example, while in avian species only the CA II gene is expressed in the RBC. In mammals, at least, the genes that encode CA I and CA II enzymes appear to be linked (1-3).

In order to study the relationship between the structure and organization of CA genes and CA gene expression, it is necessary to first isolate the genes. We describe here the initial steps toward isolation of the chicken CA II gene.

A chicken RBC cDNA library was prepared in the plasmid pBR322 with poly-(A)⁺ chicken anemic red cell cytoplasmic RNA by dG-dC tailing into the Pst I site (4). Bacterial colonies that were tet^Ramp^S were transferred to nitrocellulose filters, lysed, and the liberated DNA fixed to the filters. The filters were first screened for those

* Reprinted from Annals of the New York Academy of Sciences, Vol. 429, 1984.

colonies containing recombinant plasmids with globin DNA inserts by hybridizing the filters with and globin-specific probes. The nonglobin colonies were selected and screened with ^{32}P -labeled mouse CA II cDNA (5). The three colonies whose DNA gave a positive autoradiographic result were isolated. DNA sequence analysis demonstrated that one of the three clones isolated whose insert was approximately 300 bp was a bona fide chicken CA II cDNA clone.

The restriction map of the chicken CA II clone is shown in Figure 1 along with the strategy for sequence analysis. The amino acid sequence

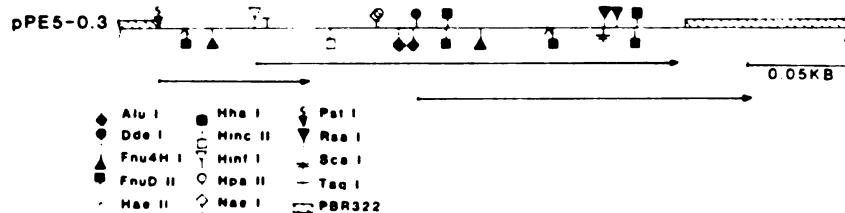


FIGURE 1. cDNA clone of carbonic anhydrase II. Restriction map of pPE5-0.3. Arrows indicate the direction and extent of the sequence determined by Maxam and Gilbert sequence analysis.

predicted from the nucleotide sequence shows extensive homology with the known amino acid sequences of human (65%), rabbit(6) (63%), and mouse(7) (60%) CA II (Fig. 2). The chicken cDNA clone contains sequence from the coding region at the 5'-end of the CA II mRNA from amino acids 7 to 86.

A comparison of amino acids 7 to 86 of chicken to the corresponding amino acids of mammalian CA I, II, III(8) indicates that chicken CA II is identical to 1 of the 12 invariant and unique residues of CA I; 1 of the 15 for CA III; and 5 of th 9 for CA II (residues 7, 26, 66, 68, and 75). Chicken CA II is identical with all of the 6 residues that are located in the active site regions of the CA isozymes (residues 28, 60, 63, 64, 66, and 68).

		GGG	AGY	CAC	AAC	GGA	CCC	GCG	CAC	TGG	YAC	GAG	YAC	TTC	CCC
					10										20
CHICKEN	CA II	Gly	Ser	His	Asn	Gly	Pro	Ala	His	Trp	Tyr	Glu	Tyr	Phe	Pro
HUMAN	CA II	--	Lys	--	--	--	--	Glu	--	--	His	Lys	Asp	--	--
RABBIT	CA II	--	Lys	--	--	--	--	Glu	--	--	His	Lys	Asp	--	--
MOUSE	CA II	Ser	Lys	--	--	--	--	Glu	Asn	--	His	Lys	Asp	--	--
		ATC	GCC	AAT	GGG	GAG	CGC	CAG	TCG	CCC	ATC	GCC	ATC	AGC	ACC
											30				
CHICKEN	CA II	Ile	Ala	Asn	Gly	Glu	Arg	Gln	Ser	Pro	Ile	Ala	Ile	Ser	Thr
HUMAN	CA II	--	--	Lys	--	--	--	--	--	--	Val	Asp	--	Asp	--
RABBIT	CA II	--	--	Asp	--	--	--	--	--	--	--	Asp	--	Asp	--
MOUSE	CA II	--	--	--	--	Asp	--	--	--	--	Val	Asp	--	Asp	--
		AAA	GCC	GCC	CGC	TAC	GAC	CCC	GCG	CYG	AAG	CCC	CTC	AGC	TTC
							40								
CHICKEN	CA II	Lys	Ala	Ala	Arg	Tyr	Asp	Pro	Ala	Leu	Lys	Pro	Leu	Ser	Phe
HUMAN	CA II	His	Thr	--	Lys	--	--	--	Ser	Leu	--	--	--	--	Val
RABBIT	CA II	Asp	--	--	Lys	His	--	--	Ser	Leu	--	--	--	Arg	Val
MOUSE	CA II	Ala	Thr	--	His	His	--	--	--	Leu	Gln	--	--	Leu	Ile
		AGC	TAC	GAT	GCC	GGC	ACG	GCC	AAA	GCC	ATC	GTC	AAC	AAC	GGG
			50										60		
CHICKEN	CA II	Ser	Tyr	Asp	Ala	Gly	Thr	Ala	Lys	Ala	Ile	Val	Asn	Asn	Gly
HUMAN	CA II	--	--	--	Gln	Ala	--	Ser	Leu	Arg	--	Leu	--	--	--
RABBIT	CA II	--	--	Glu	His	Pro	Ile	Ser	Arg	Arg	--	Ile	--	--	--
MOUSE	CA II	--	--	--	Lys	Ala	Ala	Ser	--	Ser	--	--	--	--	--
		CAC	TCC	TTC	AAC	GTG	GAG	TTC	GAC	GAC	TCC	TCC	GAC	AAG	TCA
									70						
CHICKEN	CA II	His	Ser	Phe	Asn	Val	Glu	Phe	Asp	Asp	Ser	Ser	Asp	Lys	Ser
HUMAN	CA II	--	Ala	--	--	--	--	--	--	--	--	Gln	--	--	Ala
RABBIT	CA II	--	--	--	--	--	--	--	--	--	--	His	--	--	--
MOUSE	CA II	--	--	--	--	--	--	--	--	--	--	Gln	--	Asn	Ala
		GTG	CYG	CAA	GGA	GGA	GCG	CTG	GAT	GCC	AGT				
					80										
CHICKEN	CA II	Val	Leu	Gln	Gly	Gly	Ala	Leu	Asp	Gly	Ser				
HUMAN	CA II	--	Leu	Lys	--	--	Pro	--	--	--	Thr				
RABBIT	CA II	--	Leu	Lys	Glu	--	Pro	--	Glu	--	Thr				
MOUSE	CA II	--	Leu	Lys	--	--	Pro	--	Ser	Asp	--				

FIGURE 2. Nucleotide sequence of pPE5-0.3. The predicted amino acid sequence of chicken carbonic anhydrase II is compared with homologous amino acid sequences of human, rabbit, and mouse carbonic anhydrase II. The amino acid sequence predicted by the nucleotide sequence is given below the coding regions along with its numbering. Y refers to C or T. Only those amino acids of the human, rabbit, and mouse CA II proteins that differ from the predicted chicken CA II sequence are shown. See reference 6 and citations therein for the human CA II sequence.

The chicken cDNA clone was used to isolate phage from a Charon 4A chicken genomic library (9). The phage that hybridized to the cDNA clone presumably contain the chicken CA II gene. These recombinant clones are presently being characterized by restriction enzyme analysis. Future experiments will involve detailed restriction enzyme analysis, subcloning, and DNA sequence analysis (particularly the 5' flanking region) of the clones and will provide preliminary data for the study of the chicken CA gene family.

Acknowledgments

We are grateful to Dr. Peter J. Curtis for providing the mouse carbonic anhydrase cDNA clone. We also thank Dr. Richard E. Tashian and Dr. David Hewett-Emmett for their advice and encouragement.

References

1. Carter, N.D. 1972. Carbonic anhydrase isozymes in *Cavia porcellus*, *Carvia apera* and their hybrids. *Comp. Biochem. Physiol.* 43B:743-747.
2. DeSimone, J., M. Linde & R. E. Tashian. 1973. Evidence for linkage of carbonic anhydrase isozyme genes in the pig-tailed macaque, Macaca nemestrina. *Nature (New Biol.)* 242:55-56.
3. Eicher, E. M., R. H. Stern, J.E. Womack, M. T. Davisson, T. H. Roderick & S. C. Reynolds. 1976. Evolution of mammalian carbonic anhydrase loci by tandem duplication: Close linkage of Car-1 and Car-2 to the centromere region of chromosome 3 of the mouse. *Biochem. Genet.* 14:651-660.
4. Maniatis, T., E. F. Fritsch & J. Sambrook. 1982. *Molecular cloning manual*. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
5. Curtis, P. J. 1983. Cloning of mouse carbonic anhydrase mRNA and its induction in mouse erythroleukemia cells. *J. Biol. Chem.* 258:4459-4463.
6. Ferrell, R. E., S. K. Stroup, R. J. Tanis & R. E. Tashian. 1978. Amino acid sequence of rabbit carbonic anhydrase II. *Biochim. Biophys. Acta* 533:1-11.
7. Curtis, P. J., E. Withers, D. Demuth, R. Watt, P. J. Venta & R. E. Tashian. 1984. The nucleotide sequence and derived amino acid sequence of cDNA coding for mouse carbonic anhydrase II. *Gene*. In press.
8. Tashian, R. E., D. Hewett-Emmett & M. Goodman. 1983. On the evolution and genetics of carbonic anhydrase I, II, and III. *Isozymes: Current Topics in Biological and Medical Research* 7:79-100.
9. Dodgson, J., J. Strommer & J. D. Engel. 1979. The organization of chicken globin genes. *Cell* 17:879-887.

CHAPTER 3
THE NUCLEOTIDE SEQUENCE OF THE CHICKEN
CARBONIC ANHYDRASE II GENE

INTRODUCTION

Carbonic anhydrase (CA) refers to an ancient family of proteins. Five different CA isozymes (1-9) (possibly more, 10, 11) have so far been identified that are believed to be coded for by separate genetic loci. Together, the CA isozymes are extensive in their distribution. They are found in practically all organisms and in most tissues of any higher organism. Although the most obvious and well-studied role of CA is the hydration of CO_2 in red blood cells (12), other CA functions have recently been elucidated. A few of these roles include involvement in ion fluxes in neurons (13), avian eggshell formation (14), and eye morphogenesis (15).

The CA gene family is quite variable in its expression pattern. Among the amniotes (birds, reptiles, mammals) CA II is particularly interesting because its tissue specificity varies with the class of organism considered. CA II is the only isozyme expressed in avian red blood cells (RBC) whereas both CA I and CA II isozymes are expressed in most mammalian RBC. The CA I and CA II genes of mammals are also known to be linked (16).

Although the CA protein has been thoroughly examined, it is only quite recently that CA structure at the DNA level has been examined.

DNA sequence data of the mouse CA II gene (17) and a rabbit CA I cDNA clone (18) have been reported. A closer analysis of the mouse CA II gene showed that it was composed of 7 exons that were stretched over 16 kb of DNA (17). Three of its introns interrupted codons and three introns fell between codons. There were also slight differences in the CA II gene sequence between two different mouse strains in the form of point mutations. Since the human CA II gene has been partially sequenced (19) a comparison of human and mouse CA II was made. There was considerable sequence homology in the 5'-flanking regions of the two genes in the presence and locations of common signal sequences.

In this paper we describe the isolation and primary structural analysis of the chicken CA II gene. The chicken gene provides another class of amniotes to be represented in the evolutionary analysis of the CA gene family. A detailed structural study of the CA genes as well as of their relationship to one another is imperative for the identification of structural regulatory elements controlling the expression of the CA gene family.

EXPERIMENTAL PROCEDURES

Isolation of cDNA

A λ gt10 cDNA library prepared from chicken red cell poly(A)⁺ mRNA was screened at a 99% representation of the library (4×10^4 phage). The plaques were transferred to nitrocellulose filters and processed as described (20). CA II cDNA was identified by hybridization to both a 5'-end chicken CA II cDNA clone probe (21) and a full length mouse CA II cDNA clone probe (22). Probes were prepared by nick translation (23). CA II cDNA fragments were isolated from positive λ gt10 clones by digestion at the EcoRI linker sites and inserted into pBR325 plasmid DNA. A fine structure restriction map was derived for the subclones with the largest inserts and used for the sequence analysis of the cDNA.

Isolation of λ caIII and λ caXVI

A λ Charon4A chicken genomic DNA library (24) was screened as described above. Twenty petri plates with approximately 5×10^4 pfu per plate were used. Filters were hybridized at 42° for 12 hours to nick translated 5'-end chicken CA II cDNA clone previously described (21). Filters were washed at 65° in 0.1 M Na⁺ and exposed. Duplicate positive plaques were purified through two platings at low plaque density and purified recombinants were grown as liquid cultures (25) for DNA preparation. Restriction maps of the unique CA II-containing phage, designated λ caIII and λ caXVI, were prepared by standard multiple restriction digestion analysis.

Preparation of Subclones Containing Fragments of the CA II Gene

λ caIII and λ caXVI were digested to completion with the appropriate restriction endonucleases (BamHI, EcoRI, HindIII, or KpnI). The fragments of interest were gel purified and ligated to plasmid pBR322 or pAT153 DNA cleaved with the same enzyme(s). T4 DNA ligase was added to the mixture and incubated at 15° overnight. E. coli HB101 was transformed with the ligated DNA. Transformants were selected by differential drug-resistance characteristics (ampicillin resistance and tetracycline sensitivity), hybridization to CA II probes and/or restriction digestion of DNA minipreps. Positive transformants were further analyzed by restriction enzyme digestion and blot hybridization.

Restriction Enzyme Digestion and Southern Blotting

Restriction digests were run according to the recommendations of the manufacturers. The products of digestion were run on agarose or acrylamide gels and visualized by ethidium bromide fluorescence. Where appropriate, the DNA in the gels was transferred to nitrocellulose (26) and the resultant Southern blots were hybridized to labeled CA II cDNAs as described (25).

DNA Sequence Analysis

Subcloned CA II DNA was restriction enzyme digested, gel fractionated and the appropriate fragment isolated (27). The chemical degradation technique of Maxam and Gilbert (28) as modified by Smith and Calvo (29) was employed for DNA sequencing. Fragments were treated with calf alkaline phosphatase, 5'-end labeled with γ -ATP, recut with the appropriate secondary enzyme and the resultant singly-labeled

fragments isolated. Chemical degradation and gel electrophoresis were as described previously (28,29,30).

Nuclease Protection / End Analysis

Restriction fragments were isolated from restriction enzyme digested pBBca-2.8 by cleavage within exon 1 (RsaI site at +19, Figure 2A) and upstream beyond the likely start site (SmaI at -168). The resultant 191 bp fragment was treated with calf intestine alkaline phosphatase and labeled with polynucleotide kinase. The end-labeled fragment (75 ng) was hybridized to 20 μ g of adult anemic chicken total red cell cytoplasmic RNA or to yeast tRNA at 60° in 80% formamide hybridization buffer for 12 hours. Samples were quenched by dilution into .3 ml of S1 digestion buffer (24). S1 nuclease (at concentrations of 1000, 2000, and 4000 units/ml) was added and the DNA-RNA hybrid processed as described (24).

For the 3'-end analysis, a 600 bp fragment was isolated from pHHca-0.6 digested with HindIII (Figure 2E). The fragment started 388 bp downstream from the stop codon and included 5 possible poly(A) addition signals. The fragment was labeled by filling the 5' overlapping end with AMV reverse transcriptase as described (29). The end-labeled fragment (70ng) was resuspended in hybridization buffer and hybridized to 20 μ g of chicken red cell RNA as above. Hybridization was performed at 50° for 12 hours. S1 nuclease was added as described above.

Materials

Restriction enzymes were obtained from International Biotechnologies, Inc., New England Biolabs, Inc. and Bethesda Research

Labs, Inc. Polynucleotide kinase was from Amersham or International Biotechnologies, Inc. Other materials and bacterial strains were as previously described (24).

RESULTS

Isolation of Larger Chicken CA II cDNA Clones

A λ gt10 chicken RBC cDNA library (31) was screened separately by hybridization to a complete mouse CA II cDNA, pMCA II, (22) and a 5'-end chicken CA II cDNA, pPE5-0.3 (21). The plaques that hybridized to both probes were purified through two more screenings.

Two of the recombinant clones, 4A and 2B, were shown to contain CA II cDNA-hybridizing inserts after digestion with EcoRI. Clone 4A contained an insert of 1.2 kb and clone 2B contained an insert of 0.4 kb. The cDNA insert of clone 4A was subcloned, (pEcD-1.2), and the restriction map and DNA sequence of the insert were determined. PEcD-1.2 was found to contain 1.2 kb of sequence that encodes amino acid 7 through to the 3'-untranslated region ending in a run of A:T base pairs that presumably result from the DNA complement of the 3' poly(A) portion of the CA II mRNA (Figure 3).

Isolation of the Chicken CA II Gene

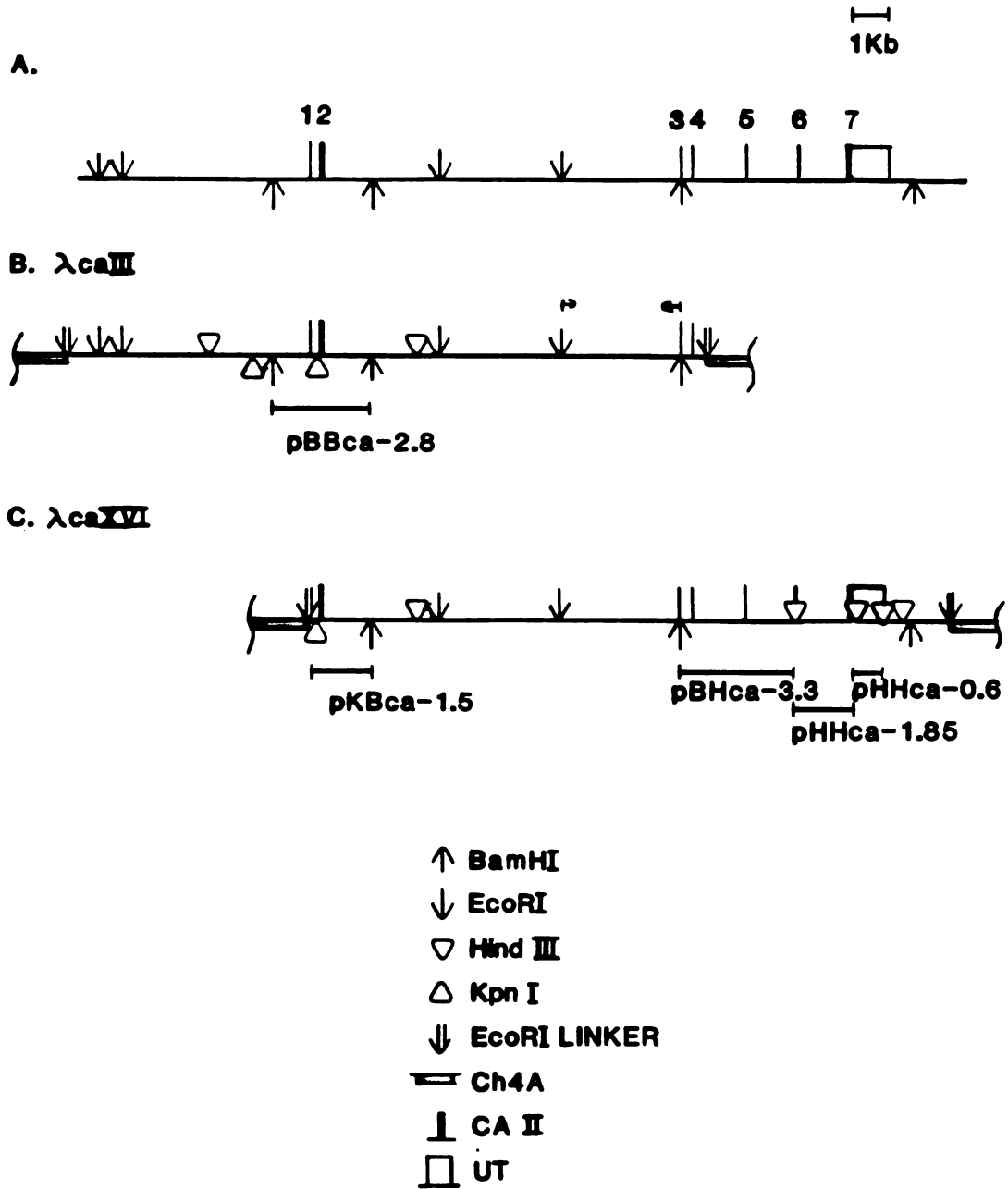
A λ Charon 4A chicken genomic library (24) was screened with ^{32} P-labeled CA II cDNA containing the 5'-end of the cDNA (pPE5-0.3, see chapter 2). Seventeen recombinant clones were isolated which upon restriction enzyme analysis proved to contain different recombinant types. Two of these clones, λ caIII and λ caXVI, have been mapped and characterized in detail. The restriction maps of the two phage are

shown in Figure 1.

Initially, a region was found in both clones which hybridized strongly to pPE5-0.3. Once the more nearly complete chicken cDNA clone, pEcD-1.2, was available, the phage digests were also hybridized to a middle and 3'-end probe as well. λ caXVI showed hybridizing bands with all three probes. λ caIII hybridized only to the 5'-end and middle probes. Both mapping and hybridization studies suggested that the chicken CA II coding region was interrupted by several introns, that most of the coding region was contained in λ caXVI, and that the 5'-coding region was contained in λ caIII. Since the two clones showed some common restriction fragments, it appeared that they overlapped. Figure 1 summarizes the results of the mapping and hybridization data. The two clones share bands when hybridized to the 5'-end and whole cDNA probes. However, λ caIII had additional hybridizing bands to the 5'-end probe and λ caXVI had additional hybridizing bands to the whole and 3'-end probes. λ caIII had no hybridizing bands to the 3'-end probe.

The chicken DNA inserts are approximately 16 kb in length. λ caXVI contained 4 regions that hybridized to chicken CA II cDNA. The first hybridizing region was a 1.5 kb KpnI, BamHI fragment at one end of the insert that hybridized strongly to the 5'-end probe. The next hybridizing region was a 2.8 kb BamHI, HindIII fragment approximately 10 kb from the first hybridizing region. A 1.63 kb HindIII fragment adjacent to the 2.8 kb fragment was the third hybridizing region and adjacent to this was a final, very weakly hybridizing HindIII fragment.

Figure 1. Restriction map of the chicken CA II gene locus. (A) Restriction map of chromosomal DNA contained within clones λ caIII and λ caXVI. (B) Restriction map of clone λ caIII which contains exons 1-4. (C) Restriction map of clone λ caXVI which contains exons 2-7. The solid boxes represent the coding regions and the open boxes represent the untranslated regions of the exons. The numbers above the boxes indicate the exons and the horizontal lines below the boxes represent the subclones. The arrows above the line indicate the direction and extent of DNA sequence determination.



The other clone, λ caIII, had a 2.8 kb BamHI fragment which hybridized strongly to the 5'-end probe and a second .72 kb BamHI, EcoRI fragment which hybridized to the complete cDNA. In between these two regions was a 3.5 kb BamHI, EcoRI region which hybridized weakly to both probes. The 3.5 kb BamHI, EcoRI fragment was also found in λ caXVI.

The mouse CA II gene was known to be interrupted by 6 introns of which the second intron is very large (7.2 kb) (17,19). The fragment that hybridized strongly to pPE5-0.3 was subcloned and sequence analysis showed that it contained exon 2 (Figure 2). This then allowed us to estimate the location of other exon regions which were later confirmed and refined by DNA sequence analysis (Figures 2,3).

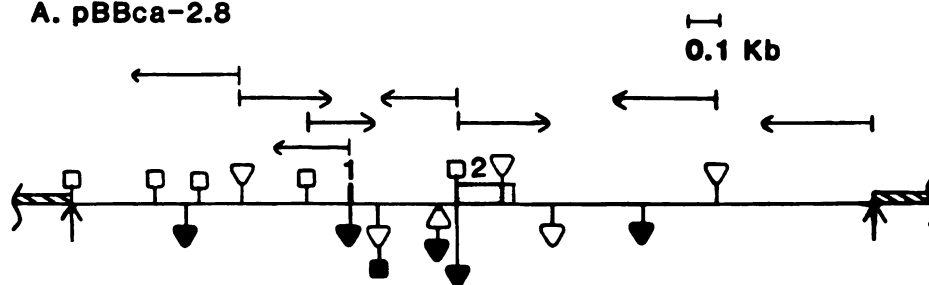
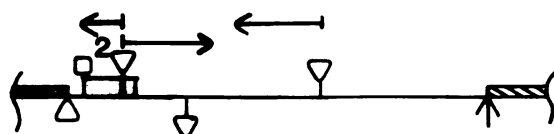
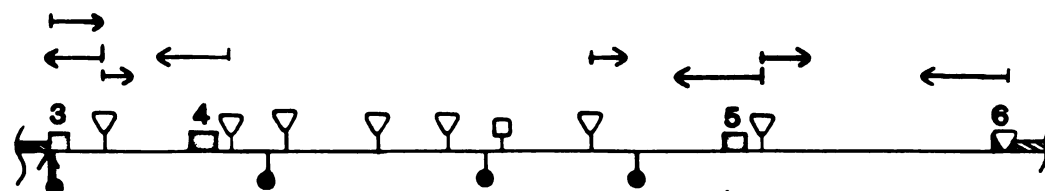
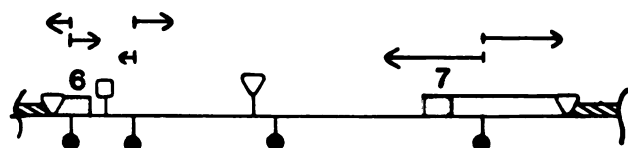
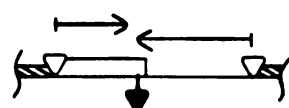
The restriction map and hybridization data obtained for the two recombinants were used to develop a subcloning strategy. The subclones generated, their fine structure restriction maps, and the sequencing strategy used in their analysis are given in Figure 2.

Since both chicken cDNA clones did not contain the entire exon 1 sequence (see chapter 1), a mouse genomic clone was used to locate exon 1. The mouse clone A6-2.7 (19), a generous gift of P. Venta and R. Tashian, Univ. of Michigan, was prepared such that the only coding region it contained was exon 1. The chicken exon 1 was approximately positioned by cross-hybridization to A6-2.7 and definitively located by DNA sequencing.

DNA Sequencing Strategy of the Chicken CA II Gene

The fine structure maps of each subclone were used to develop a sequencing strategy for each of the CA II exons. The direction and extent of the region sequenced are indicated by the arrows in Figures 1

Figure 2. Restriction maps of the subcloned DNA fragments of the chicken CA II gene. Partial restriction maps of (A) subclone pBBca-2.8 which contains exons 1 and 2, (B) subclone pKBca-1.5 which overlaps with pBBca-2.8 and also contains exon 2, (C) subclone pBHca-3.3 which contains exons 3, 4, 5, and a small portion of exon 6, (D) subclone pHHca-1.85 which contains the greater portion of exon 6, all of exon 7, and a portion of the 3'-untranslated region, (E) subclone pHHca-0.6 which contains the remainder of the 3'-untranslated region. The filled boxes represent the exons which are identified by numbers. The arrows above the boxes show the direction and extent of DNA sequence analysis.

A. pBBca-2.8**B. pKBca-1.5****C. pBHca-3.3****D. pHHca-1.85****E. pHHca-0.6**

↑ BamHI

↓ EcoRI

▽ Hinf I

△ Kpn I

● BstNI

▽ Hinf I

◆ Rsa I

◇ Sac I

□ Sin I

● Xma I

▨ PBR322/PAT153

▨ α 7-1.7

□ CA II

□ UT

and 2. The sequence of the coding exons has, in some cases, been determined on both the coding and noncoding strands of DNA. Except for that sequence 5' to codon 8, all the coding sequences were also sequenced on one or both strands in the cDNA clone, pEcD-1.2.

Once the sequence data was obtained the identification of the hybridizing regions as to exon number could be determined. Furthermore, the exact boundaries of all seven exons could be identified by comparison of the cDNA and genomic sequence. As Figure 2 indicates, subclone pBHca2.8 contains most of exon 3, all of exons 4 and 5, and a small portion of exon 6. Exon 6 is continued on subclone pHHca-1.63 which also contains exon 7 and the 3'untranslated region. Since the 3'untranslated region is relatively long, it is continued on the 0.7 kb subclone. Exons 2 and 3 are also on λ caIII. Exon 1 as well as the 5'flanking region is contained in the 2.8 kb BamHI subclone of λ caIII.

Exon 3 is interrupted by a BamHI site used in subcloning and the 2.8 kb BamHI, HindIII fragment containing the 5'end of exon 3 was not itself subcloned. Therefore, this portion of exon 3 was sequenced directly from a 3.5 kb BamHI, EcoRI fragment isolated from the λ caIII clone.

Sequence of the Chicken CA II Gene

The DNA sequence of the chicken CA II gene is given in Figure 3. The genomic sequence shown encodes 259 amino acid residues and approximately 1.5 kb of untranslated region. The coding sequence of the gene was identified by comparison to the chicken CA II cDNA (pEcD-1.2) sequence. The amino acid sequence predicted from the

Figure 3. DNA sequence of the chicken CA II gene. The CCAAT, ATA, and AATAAA or AACAAA signal sequences as well as the initiation and stop codons are underlined. Numbers above the sequence indicate nucleotide numbering from the Cap site. The upper case letters indicate those sequences that are transcribed into mRNA and the lower case letters indicate those sequences that are processed. All the sequences present in exons within the left and right arrows above the sequence shown were also sequenced in the cDNA clone, pEcD-1.2. Asterisks indicate sites where the cDNA and genomic clone sequences differ (see text). R refers to A or G. X refers to A, G, T, or C.

EXON 1

-200
CGCCGCCGCCGCTCTCCCGGCC

GGCCCGACGCAGCTCCGCGGCGGGAGGATCGCGGGTTATAAGCGGACCTCTCTCTCTCC

-100

GCCCCGAGCGAAGTCTCCCTCCGCCCCGCGCGCTCCCCACCCCTTCCTCCGGCCG

CGGAGAAGGGCATGGAGTTCGCGGGAGCCTATAAAAGCCCTGACAGCCCGCCGAGGCC

+1
ACGGCGTTGCGATAGCCGACGGAGCCGGGCGGCGCACCATGTCCCATCACTGGGGGTA

→

CGACAGCCACAACGGTGAGTGTGGGGCACGGCGG

EXON 2

CGCCCCGCGCTCTCTTGACAG

GACCCGCGCACTGGCACGAGCACTT100CCCATCGCCAATGGGGAGCGCCAGTCGCCCATC

GCCATCAGCACCAAAGCCGCCGCTACGACCCGCGCTGAAGCCCCTCAGCTTCAGCTA

200

CGATGCCGGCACGGCCAAAGCCATCGTCAACAACGGGCCTCCTTCAAAGTGGAGTTTG

ACGACTCCTCCGACAAGTCAGGTGAGCGCATCCGTGTGTGC

EXON 3

GGAGGCTCTTTGCTTTGCAG

TGCTGCAAGGAGGAGCGCTGGATGGAGTCTACAGTTGGTGCAGTTTCATTCACTGG

300

GGATCCTGTGAGGGCCAGGGCTCTGAGCACTGTGGATGGCGTGAAGTACGATGCAGA

GGTATGATGTGCTTTGCCTTT

EXON 4

CCATGTTTTCTTATTCCTAG

CTTCATATTGTTCACTGGAATGTAAATAATGGCAAATTTGCTGAAGCTCTGAAGCATCC

400

TGATGGTTTGCCGTCGTAGGCATCTTCATGAAGGTTAGTCAAACCTTCTTTTTTC

*

Exon 5

TCATGCTATATGTGTTACAG

GTAGGG⁵⁰⁰AATGCCAAACCTGAAATACAGAAAGTTGTTGATGCTCTGA⁶⁰⁰ACTCCATTCA⁷⁰⁰AAC
CAAGGTAATATTTTGTGTTGAATG

Exon 6

GCCTACCTTCCTTACTGCAG

GGG⁶⁰⁰AACAAGCTTCTTTACAAA⁷⁰⁰CTTTGACCCTACTGGACTGCTGCCTCCATGCAGAGA
CTATTGGACGTACCCTGGCTCCCTGACTACTCCA⁸⁰⁰CACTGCATGAATGTGTGATTGGC
ATGTTCTGAAGGAGCCATCACTGT⁹⁰⁰CAGCTCTGAGCAGGTA¹⁰⁰⁰CTCTGGGGTAGTGC

Exon 7

GTTTTGCCTTTTCCCACAG

ATGTGCA⁸⁰⁰AACTCCGTGGCCTTTGCTTCAGTGCTGAGAATGAGCCGGTGTGCCGCATGGT
GGACAACTGGCGCCCATGCCAGCCTCTAAAGAGCAGGGAAGTCAGAGCTTCCTTCCAGT
AACCTCAGCGATGAGTGTGTTAGAACTGCTGTGTTTGC⁹⁰⁰GAGGAACCCTTTTGCTAAG
CACAAATCAA¹⁰⁰⁰ACCTTTGCCATGTGTGCCCTGGCAACATCTTGTCTCCCATATTATTCTC
TCTTTGCTCTGCATCTAAAATGCCAGCTAATGAAATGTGA¹¹⁰⁰AAGGCTCTTGCCAAACA
¹²⁰⁰GGAAGGGTTCTTCATGTGTGGAGCTGGGGAAACCTGAGGGGGGCTGTGTGCATTTTG
ATGACTTACTGCGACTGACATTTTGGAAAAAC¹³⁰⁰AAAAACAAACAAACAA¹⁴⁰⁰AACTA
TATTGGCTGTTTGGGAGAGCATATGGTGAGAGCA¹⁵⁰⁰AAATAAGCCA¹⁶⁰⁰TTCTGAGAA¹⁷⁰⁰ACTCA
TCAACTGGTGTGATAGTACACATA¹⁸⁰⁰ACTAATGATA¹⁹⁰⁰ACTTGAAGCTTTGTGAAAGC²⁰⁰⁰ACAG
GAA²¹⁰⁰AC²²⁰⁰AAAATATAGGTATAGTATAG²³⁰⁰GA²⁴⁰⁰AAAA²⁵⁰⁰AAAA²⁶⁰⁰AGTATATTGA²⁷⁰⁰GAAGGAAAGTAA
AATGA²⁸⁰⁰ATACTGAGAATTAGACTTAGATATTAAGAAAC²⁹⁰⁰TGACATGTAAATATTTTGACTT

AGATATTAAGAACTGACATGTAAATATTTGAAGACCATTTTGCATTTTACCCATGC
 TATCAACAACCTTGATCGTCTTCATGGTAGTGTGTTGGTTTTTTTAATTAAAAATGTCC
 TAAATTAAGTATTATCTTTTGAAGTTATTACTGTAGAGGGTACXXXAAATATCTTTTCA
 CTGAAATAATATATGCTCTAATTTAAGGGGGAAATAAAATTGTATTTTAGATAACTTCT
 CTAAATAATCTATACTTATATTTTGTCTCCAAGTTGTTGTTTAATTCAGAAAATAGT
 TACTAATCTGTTCAATCATTGCATACTAAACAATTATTAATAAACATTTAGTATATAAG
 GAGTGAGGTTTCAATGACACTGGAGATTAATGCAAAATGATATGGATGTATTAGAAAGAT
 TTCCTAAGTCGTTCTCTGCTAGACAGCAGAAGGAAGTATTAGTTCTGAGCAACCCCTTGC
 AAATTGTCAATGAAGTTGATTGCATAATAAAAGATCCCCTTGCAGGAATTTTGAACARC

nucleotide sequence is identical to that of the cDNA clone. However, there are 5 confirmed differences in the cDNA and genomic sequence (Figure 3). These include the third nucleotide of codon 140 (C in the genomic and T in the cDNA) and codon 152 (A in the genomic and G in the cDNA). There are also 3 changes in the 3'untranslated region: at nucleotide 1001 a G in the cDNA is an A in the genomic DNA; at nucleotide 1177 a C in the cDNA is an A in the genomic DNA; and at nucleotide 1113 the cDNA contains an extra A that is not present in the genomic clone. These changes presumably reflect genetic diversity in the chickens used to prepare the cDNA and genomic libraries (both were White Leghorn; the cDNA sequence from a Michigan chicken and the genomic sequence from a Pennsylvania chicken). The extra A:T base pair in the cDNA clone with respect to the genomic clone comes in a long stretch of A:T base pairs interrupted by a few C:G base pairs. This mutation may have occurred due to slippage during DNA replication. It cannot be determined at present whether the changes seen are actually differences in the germ line of the chickens used or whether any or all of the changes could have occurred during the cloning procedures. Similar changes were seen by Venta et al. (19) between YBR and BALB/c mouse strains.

The chicken CA II amino acid sequence has 65% homology to mouse CA II sequence (17) (Figure 4). There are 69 base changes that result in silent substitutions (Table I). The greatest number of silent changes occur in exon 2. There are 164 base substitutions that result in an amino acid change and a majority of these occur in exons 2 and 7. The active site residues as well as the unique and invariant residues (Table II)(32) are fairly well conserved. There are only 2 changes

Figure 4. Amino acid sequence comparison of chicken and mouse CA II genes. The predicted amino acid sequence of chicken CA II is compared with homologous amino acid sequences of mouse CA II. The amino acid sequence predicted by the nucleotide sequence is given above the coding regions along with its numbering. Only those nucleotides in the mouse sequence that differ from that of the chicken sequence are given along with the resulting amino acid change, if any. Asterisks indicate those amino acid residues that are located in the active site region of the CA II protein.

	Ala	Ser	Phe	Gln	Stop
CCAII	GCT	TCC	TTC	CAG	TAA
MCAII	G		T	A	
				Lys	

TABLE I. Nucleotide and amino acid divergence of chicken CA II gene from mouse CA II gene.

Exon	Number of bp changes result- ing in an amino acid change	Number of silent nucleotide substitutions	Total number of amino acid residues	Number of bp	% Amino acid divergence	Total % nucleotide divergence
1	4	2	10.5	31	19	19
2	37	24	66	198	35	30
3	23	6	39.5	119	33	24
4	19	8	31	93	35	29
5	17	4	21	63	48	33
6	27	19	52	156	23	29
7	37	6	39	117	49	37

TABLE II. Unique and invariant residues for the mammalian CA isozymes and the homologous residues in the chicken CA II isozyme.

	1	3	4	8	14	18	19	26	27	32	33	34	36	38	54	56	58	62
MA CA I	Ala		Asp	Asp			Asp	Asn	Arg			Lys	Ser				Glu	Val
MA CA II	-	His	His													Ser		
MA CA III	-		Glu	Ala	Asp	Glu				Glu	Leu			Ile	Gly		Thr	
CH CA II	-	His	His	Asp	Ala	Glu	His	Glu	Arg	Ala	Ile	Ser	Lys	Ala	Gly	Ala	Ala	Asn
	*	*	*							*								
	65	67	69	71	73	74	75	78	80	87	91	95	112	125	126	129	130	135
MA CA I				Glu		Asp	Asn								Ala			Ser
MA CA II		Asn	Glu										Lys	Thr	-	Gly	Asp	
MA CA III	Thr	Arg	Val		Thr	Tyr		Met	Arg	Pro	Arg	Leu			-	Asn	Leu	
CH CA II	Ser	Asn	Glu	Asp	Ser	Ser	Asp	Val	Gln	Tyr	Val	Ile	Val	Val		Gly	Lys	Lys
	138	147	153	155	156	157	166	173	183	187	189	198	200	207	208	210	211	221
											*		*	*		*		
MA CA I		Leu								Ser			His					Gln
MA CA II	Pro						Ser											
MA CA III			Glu	Gly	Glu	Phe	Lys	Glu	Cys		Arg	Phe		Ile	Val	Leu	Leu	Asp
CH CA II	Pro	Phe	Ala	Pro	Glu	Leu	Ser	Gln	Gly	Pro	Arg	Leu	Thr	Val	Ile	His	Val	Glu
	226	230	231	232	235	241	255											
MA CA I							Thr											
MA CA II			Asn	Phe														
MA CA III	Leu			Ser	Asn	Leu												
CH CA II	Leu	Cys	Phe	Ser	Asn	Met	Glu											

- indicates residues not present in CA II or CA III. * indicates active-site residues.
Sequence numbering based on CA I. Table is revised from Hewett-Emmett et al., reference 32.

that involve active site residues and these occur at residue 90 and residue 202. Residue 90 which is an Ile in mouse changes to Val in chicken but remains nonpolar. Residue 202, however, is changed from a nonpolar Leu in mouse to a basic residue, His, in chicken.

Chicken CA II Gene Intron/Exon Organization

The chicken CA II gene is interrupted by 6 introns. Introns 1 and 2 interrupt the codons for Gly-11 and Val-77, respectively. Introns 3, 4, 5, and 6 fall between the codons for Glu-116/Leu-117, Lys-147/Val-148, Lys-168/Gly-169, and Gln-220/Met-221, respectively (Figures 2 and 3). The location of 5 of the 6 introns relative to the amino acid sequence are conserved between the chicken and mouse CA II genes. Surprisingly, the location of one of the introns, intron 4, is different in the chicken gene, falling between codons 147 and 148 rather than within codon 143 as in the mouse. The chicken intron 4 location, however, is also observed in the mouse CA Z gene (11)(J. C. Montgomery, personal communication).

The 5' and 3' boundaries of the six introns of the chicken CA II gene all fit the consensus donor and acceptor sequences (33) seen for most eucaryotic introns (Table III). This fit to the consensus sequence holds true for the donor and acceptor sites for intron 4, the intron whose position is altered in the chicken gene relative to the mouse CA II gene.

The sizes of the 6 chicken CA II introns are given in Table IV, in comparison to the corresponding intron sizes in the mouse gene. It can be seen that some of the intron sizes are similar in both genes as, for example, are the intron sizes in the corresponding chicken and

TABLE III. DNA sequence of intron donor and acceptor sites. The sequences of intron donor and acceptor sites in the chicken CA II gene are given along with the intron donor and acceptor consensus sequences.

Intron	Donor	Acceptor
1	ACG/GTGAGT	CGCGCTCTCTTGACAG/G
2	CAG/GTGAGC	CTCTTTGCTTTGCAG/T
3	GAG/GTATGA	TTTTCTTATTCCTAG/C
4	AAG/GTTAGT	CTATATGTGTTACAG/G
5	AAG/GTAAT	CCTTCCTTACTGCAG/G
6	CAG/GTAGCT	TGCCTTTTCCCACAG/A
CONSENSUS	$\begin{smallmatrix} C \\ A \end{smallmatrix}AG/GT\begin{smallmatrix} A \\ G \end{smallmatrix}AGT$	$\begin{smallmatrix} (T) \\ C \end{smallmatrix}_{11}$ $N\begin{smallmatrix} C \\ T \end{smallmatrix}AG/G$

TABLE IV. Intron size comparison of the chicken and mouse CA II genes.

Intron	Intron Size (kb)	
	Chicken	Mouse
1	0.36	1.0
2	10.2	7.2
3	0.45	0.3
4	1.75	1.2
5	0.9	1.1
6	1.21	1.8

mammalian globin genes (34). However, other introns vary significantly in size between the two CA II genes. Intron 1 of chicken is 0.36 kb which is approximately one third the size of mouse intron 1 (17). Intron 2, on the other hand, is about 3 kb larger than the corresponding intron in mouse. The remaining 4 introns of chicken differ from the corresponding introns of mouse anywhere from 0.1 - 0.8 kb but the difference is not as striking as in the first two introns.

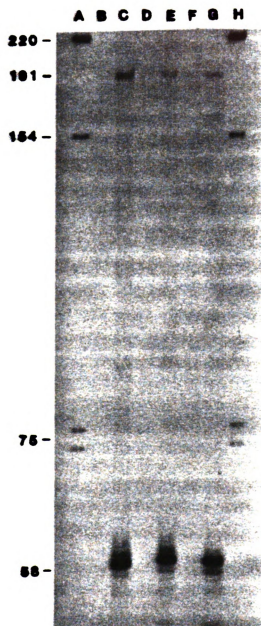
The portions of the chicken CA II gene introns that have so far been sequenced have a distinctive pyrimidine rich pattern (Appendix). These long stretches of C's and T's are apparently not characteristic of mouse CA II introns (35, P. Venta personal communication).

5' and 3' Ends of the Chicken CA II mRNA

The 5' end of the CA II mRNA has been determined by nuclease protection experiments (Figure 3). An end-labeled fragment extending from the RsaI site in exon 1 (codons 5 and 6) 5' to the SmaI site at -167 (Figure 2) was hybridized to total chicken reticulocyte mRNA, digested with S1 nuclease and run on an 8% sequencing gel. The major protected fragment is about 58 bases in size which places the mRNA start site (Cap site) in the CCACG sequence about 39 bp upstream from the ATG initiation codon (Figure 3). (When the fragment is run next to a Maxam-Gilbert sequencing ladder of the S1 probe fragment, the major protected band corresponds to initiation at the A in the CCACG, results not shown).

The 3' end of the CA II mRNA can be positioned by the cDNA clone (pEcD-1.2) sequence which indicates that polyadenylation of the CA II message occurs at the A at +1258 about 16 bp downstream of the sequence AAAACAAAA. 3' protection experiments (Figure 5) tend to confirm this

Figure 5. 5' end analysis of chicken CA II cytoplasmic mRNA. Protection of CA II gene fragment from S1 nuclease digestion by anemic adult red cell total cytoplasmic RNA is shown. End-labeled restriction fragments were hybridized to yeast RNA (lanes B, D, F) and red cell RNA (lanes C, E, G) as described (Experimental procedures) and then subjected to varying levels of S1 nuclease digestion. S1-treated fragments and marker DNA fragments are shown after electrophoresis on an 8% sequencing gel (9). The 5' end protection fragment was generated by hybridization of a 191 bp RsaI/SinI fragment to RNA followed by treatment with 1000 (lanes B, C), 2000 (lanes D, E), and 4000 (lanes F, G) units/ml S1 nuclease. Markers (lanes A, H) are polynucleotide kinase labeled restriction fragments of PBR322 cleaved with HinfI.



polyadenylation site, although the protected band is not distinct probably because of a string of 10 contiguous A:T bp downstream from the putative poly(A) addition site. The string of A:T bp could hybridize to the 3' poly(A) tail of the mRNA and exhibit partial S1 resistance. The AAAACAAAA sequence differs slightly from the AATAAA consensus polyadenylation signal (36). There are several AATAAA sequences in the 0.6 kb of DNA 3' to the putative 3' end of the mRNA. Although no bands were observed in the 3' protection experiment corresponding to the use of poly(A) sites near these AATAAA signals, we cannot rule out the possibility that one or more of them are used for at least some CA II mRNAs. Since this region of the genome contains several lengthy regions of 100% A:T base pairs which are known to be relatively S1 sensitive, there might be cleavage within a bona fide mRNA:DNA hybrid that extended to these more distal putative poly(A) sites, and thus these ends were not observed by S1 analysis. However, it is noteworthy that only one band was seen on a Northern blot of gel-fractionated red cell poly(A) RNA hybridized to a chicken CA II cDNA clone (M. Federspiel personal communication)

5' and 3' Flanking Sequences of the Chicken CA II Gene

A 783 bp region upstream from the CA II coding sequence has been sequenced. Putative signal sequences that are common to eucaryotic genes transcribed by RNA Polymerase II are found in the chicken gene (Figure 6). A Goldberg-Hogness (37) block "ATA" is located 23-30 bp upstream from the Cap site. Although a sequence that corresponds accurately to the consensus sequence "CCAAT" (38) cannot be identified, a region approximately 70-75 bp upstream from Cap (at -74 in Figure 3)

Figure 6.

DNA sequence of putative signal sequences flanking the CA II gene. (A) Conserved sequences 5' to the chicken CA II gene are compared to similar sequences in mouse and human CA II genes (12). Regions of CCAAT (13) homology, ATA boxes (14), CAP sites (13), and the initiation codon are shown. Numbers indicate the number of spacing nucleotides between the sequences shown. (B) GC-rich blocks 5' to the ATA boxes of the CA II genes are shown along with their position relative to their CCAAT boxes. The Spl binding sequences CCGCCC and GGGCGG (25) are indicated by straight lines above or below the sequences. The repeated sequences corresponding to the consensus CCNGTCACCTCCGC (17) are designated by dashed lines. (C) Polyadenylation signals and sites of the mouse and chicken CA II gene are shown relative to the termination codon. Arrows designate probable points at which poly(A) is added to the CA II mRNAs.

A.

CHICKEN	CCACC-43-TATAAAA-21-CCACG-36-ATG
MOUSE	CCACT-45-TATAAAA-21-CACAC-57-ATG
HUMAN	CCACC-48-TATAAAA-22-CACAC-71-ATG

B.

CHICKEN	<u>AGGACCACACAGC</u> -136-AAAGCACCTTCCC-422- <u>CCGCCCCCGCCC</u> -7-CCACC-9-GGCCGGG-26-TATAAAA
MOUSE	<u>CCGCCCCCGCCC</u> -54- <u>CGTGTACCTTGGTCCGCTCACCTCCAC</u> -22-CCACT-16- <u>CCGCCCC</u> --23-TATAAAA
HUMAN	<u>CCGCCCC</u> -13- <u>CCGCCCC</u> -2- <u>CCGCCCCCGCCC</u> -10- GAGTTCACCTCCGCTCCGTCACCTCCTC-16-CCACC-18- <u>CCGCCCC</u> --24-TATAAAA

C.

CHICKEN	TAA-414-AACAAAA-13-TAGGA ^Y ₁₀
MOUSE	TAA-662-AATAAAA-10-TATTCTG ^Y ₃

that has limited homology (CCACC) to the CCAAT site can be found. The Goldberg-Hogness box and the CCAAT box usually begin about 30 and 75 bp upstream of the Cap site, respectively. The absence of good consensus CCAAT sequences has also been noted for the adult chicken α -globin genes (39).

The ATA and CCAAT sequences are common to the mouse and human CA genes that have been sequenced. The ATA box for mouse and human are located at -99 and -86 bp from the ATG initiation codon, respectively. The CCAAT boxes for both are located 42 bp upstream from the ATA boxes.

When chicken CCAAT and ATA signal sequences are compared to that of human and mouse, the most obvious difference is the distance of the signal sequences from the initiation codon (Figure 7). Chicken ATA is located at -61 whereas mouse and human ATA boxes are located at -86 and -99. Although the mRNA initiation sites of mammalian CA II genes have not been mapped (see below), this presumably reflects a difference in the length of the 5' untranslated portion of the CA II mRNA. The chicken Cap site is found at 23 - 30 bp upstream of the transcription initiation site. The actual Cap site of the mouse and human CA II genes have not been experimentally determined but an alternating PyPu stretch (CACAC) at -70 bp from the initiation site in both mouse and human CA II genes is likely to be the Cap site (P. Venta personal communication).

Despite the difference in length and sequence of the 5' untranslated sequences of the chicken, the CA II gene shows a significant resemblance to the analogous mouse and human regions (35). This is shown in Figure 7. The region of greatest homology extends from the Cap site upstream for about 80 bp with approximately 60%

Figure 7. Comparison of the 5' flanking region of the chicken, mouse, and human CA II genes. The upper line gives the sequence of the chicken CA II gene for 120 bp upstream of the CAP site. The first mRNA nucleotide is numbered as +1 in this figure. The second line gives those nucleotides of the mouse gene that differ from the chicken gene and the third line gives those nucleotides of the human gene which differ from the chicken gene. Dashes indicate a deletion in one line relative to the others. The putative TATA and CCAAT sequences are indicated by a line over the chicken sequence.

CCAII ⁻¹²⁰ CCGCCCCGAGCGAAGTCTCCCTCCGCCCCCGCC ⁻¹⁰⁰
 MCAII A CT GTCC C CC CA GGT T T C T
 HCAII A CT G CC TC CC C----- T C T

CCAII ⁻⁸⁰ CG-CGC-TCCCCACCC---CTTCCTCC----GGCCGCGGAGAAGGGCAT ⁻⁶⁰
 MCAII T---T AGGT T --GG C CCTG CC--- A G
 HCAII TTCGCTAGGT GAG CC CCCG CC--CC A C

CCAII ⁻⁴⁰ GGAGTTCGCGGGAGCCTATAAAAGCCCCTGACAGCCCGCCGAGGCCACG ⁻²⁰ ⁺¹
 MCAII CA G GGACGGT AC C -A A
 HCAII A G GGGCCGGC GAC C A A

overall homology. The region from 45 to 22 bp upstream of the mRNA start site (which includes the TATAAA sequence 31 to 26 bp upstream from Cap) is over 95% homologous between the chicken and mouse sequences. The homology in this 80 bp region between the CA II genes exceeds that seen between analogous chicken and mammalian globin gene promoter regions (34,39,40,41), suggesting that these sequences may have an important role in the regulation of CA II gene expression.

McKnight and Kingsbury (42) identified GC rich sequences in the thymidine kinase gene of Herpes simplex virus that appear to be involved in efficient transcription, possibly by acting as binding sites for the Sp1 transcription factor (43). The consensus Sp1 transcription factor binding site is 5' GGGCGG 3' or its complement 5' CCGCCC 3'. Two such sequences exist 7 bp 5' to the CCAAT box in the chicken CA II gene (Figure 6B). The mouse and human CA II genes contain several such sequences both 5' to CCAAT and between CCAAT and ATA (Figure 6B). While no exact match to the consensus sequence is seen between CCAAT and ATA in chicken CA II, this gene does contain a partially homologous GC-rich sequence, GGCCGCGG, in the appropriate position which may function in the same manner.

An imperfect tandem repeat of 14 -15 bp that is thought to function as an upstream promoter element and is found in 5 mammalian and 1 avian β -globin genes and in several rat pancreatic genes (44) is found in the human and mouse CA II genes as "CCNGTCACCTCCGC" (35). In the β -globin genes this tandem repeat is found 9 -25 bp upstream from the CCAAT sequence in mammals and 55 bp upstream in chickens. In the human and mouse CA II gene these repeat elements have been found 15 and 22 bp upstream, respectively, from the CCAAT boxes (Figure 6B). There

are similar repeat elements found in the chicken CA II gene: CAAAGCACCTCCCC and AGGACCACCACAGC. However, these elements are located very far upstream of the CA II promoter at -427 and -572 in the chicken CA II gene, and they are not closely linked to each other. Furthermore the first of these two elements is a rather poor match to the consensus. The function of these elements, if any, remains unknown and whether the homologues near the chicken CA II gene are functional or merely coincidental is also unclear.

The chicken CA II gene 5' flanking region is very GC-rich as has also been noted for the mouse and human CA II genes (35). Also, in the case of the chicken gene, the DNA strand which acts as the mRNA template is very purine-rich in the region from -200 to +1 making the opposite strand (the strand shown in Figure 3) pyrimidine-rich. Lengthy polypyrimidine stretches appear in several of the sequenced intron regions of the chicken CA II gene as well. This does not appear to be the case for the analogous regions of the two mammalian genes. The significance of these trends, if any, is unknown.

The probable major poly(A) addition sites for the mouse and chicken CA II genes are shown in Figure 6C. Note that both genes contain a rather long 3' untranslated region. The unusual polyadenylation signal seen in the chicken CA II gene is discussed above in connection with its identification.

DISCUSSION

The chicken CA II gene has been isolated and its primary structure has been determined. Sequence data for the coding region, as well as for the 5'-flanking, 3'-flanking, and adjacent regions of the introns

is presented. The sequence of a nearly complete CA II cDNA clone has also been determined, allowing us to position intron sequences and the 3' end of the CA II mRNA.

There are 30 amino acid residues that are postulated to occur in the active site regions of CA isozymes (1). When chicken CA II amino acid residues are compared to the homologous active site residues of mammals most of the chicken residues are found to be conserved. The 16 residues that are invariant in all three isozymes across the species are also invariant in chicken. A similar conservation is seen in the two active site residues (Asn at residue 66 and Glu at residue 68) that are unique and invariant to CA II (Refer to Table II)(32); chicken has Asn and Glu at the appropriate positions. (Chicken CA II possesses 9 of the 15 residues in the molecule as a whole that are unique and invariant to CA II, 1 of the 18 residues that are unique and invariant to CA I, and 8 of the 39 residues that are unique and invariant to CA III). At position 90 of the active site region the residue that is present in most mammals is Ile except for ox which has Val. Chicken is similar to ox in that it too has Val at this position. Val at position 90 is just a slight divergence since both residues are nonpolar. Chicken does differ from mammals at active site residue 202. Mammalian CA IIs that have been sequenced all have Leu at this position whereas in chicken the residue is His. The amino acid change is from a nonpolar to a basic residue in this case.

When the chicken CA II amino acid sequence is compared to those of mouse and human there is greater than 60% homology to both mammalian sequences. Chicken CA II appears to be more homologous to human CA II than to mouse CA II. There is 65% sequence homology of the chicken

sequence to the mouse sequence and 70% homology of the chicken to the human sequence.

In comparing the exon/intron organization between the chicken and mouse CA II genes, the most surprising result is the different locations of intron 4 relative to the respective coding sequences. In almost all cases studied to date, intron positions within coding regions are conserved between homologous mammalian and avian genes. Even in genes such as the α - and β -actin genes where intron positions are known to vary considerably between different species, the rat and chicken genes retain identical intron locations (45). The change in intron position between the two CA II genes shifts this intron 4 2/3 codons toward the 3' end in the chicken relative to the mouse CA II gene (Figure 8). This "new" intron position has also very recently been found in the human CA Z (mitochondrial) (11) gene (J.C. Montgomery, personal communication). The similarity of the chicken CA II intron position to that of the mouse CA Z intron suggests that the chicken CA II gene structure is the more ancient form compared to the mouse CA II gene. Given the similarities observed in the rest of the two CA II genes, it appears that the intron shift occurred via a small number of mutational events as opposed to the possibility that the two genes are the product of two lines of CA II gene evolution that have been separate for longer than most mammalian/avian homologues. The mutational shift may have occurred after the divergence of birds and mammals. Figure 8 shows that the intron boundary sequence at the intron 4 acceptor site retains a considerable level of homology to the corresponding sequence of the mouse gene which in mouse is used to code for amino acids 144 -147. It seems likely that a mutation in the

mammalian evolutionary line resulted in a shift to previously cryptic splice donor and acceptor sites. For example, if the ancestral gene to the mouse CA II gene had the chicken arrangement, and the donor site at codon 147 was partially inactivated, a cryptic donor site around 14 bp upstream might have become active followed by a similar shift of the acceptor to the present intron 4 acceptor site in mouse. Note that the codon 147 - 148 junction in the mouse gene shows a good fit to the consensus intron acceptor sequence even though it is apparently not used in vivo. Of course, it is impossible to say exactly what the last common precursor to the mammalian and avian CA II genes looked like or what order of events might have led to the observed intron shift.

Other than the difference in position of intron 4, the organization of the chicken and mouse CA II genes is quite similar. All other introns have identical locations within the coding sequence, both genes have relatively long 3' untranslated regions, and except for a few nucleotide differences in the donor/acceptor sites, the junctions are fairly well conserved. There are some differences in the size of the homologous introns, but the effect of these changes is likely to be minimal. Likewise, we have yet to understand the significance of the long CT stretches that appear to be characteristic of chicken CA II introns and 5' flanking region, or of the GC-rich character of the 5' flanking DNA.

Consensus sequences commonly found in eucaryotic promoter regions transcribed by RNA Polymerase II are present in the chicken CA II gene. These sequences have been described in detail in the Results section. Note that all these sequences appear to be components of a variety of eucaryotic promoters. We have yet to identify any sequences that might

confer tissue-specificity in or near the CA II gene promoter.

Preliminary results have proven, however, that the chicken CA II mRNA level is greater than 100-fold induced in anemic reticulocytes over that seen in liver and other adult tissues. Much of this induction occurs late in erythroid differentiation since the HD3 erythroid precursor cell line (transformed with avian erythroblastosis virus) has about 10 times as much CA II mRNA as does liver but 10-50 fold less than in terminally differentiated reticulocytes (J. Dodgson, personal communication).

REFERENCES

1. Tashian, R.E., D. Hewett-Emmett and M. Goodman. 1983. pp. 79-100. In Isozymes: Current Topics in Biological and Medical Research. M.C. Rattazzi, J.G. Scandalios and G.S. Whitt, eds. Alan R. Liss, Inc., New York.
2. Sapirstein, V.S., P. Strocchi and J.M. Gilbert. 1984. Ann. N.Y. Acad. Sci. 429:481-493.
3. Whitney, P.L. and T.V. Briggie. 1982. J. Biol. Chem. 257:12056-12059.
4. Henry, R.P. and J.N. Cameron. 1983. Exp. Biol. 103:205-223.
5. McKinley, D.N. and P.L. Whitney. 1976. Biochim. Biophys. Acta 445:780-790.
6. Sanyal, G., N.I. Pessah and T.H. Maren. 1981. Biochim. Biophys. Acta 657:128-137.
7. Wistrand, P.J. 1984. Ann. N.Y. Acad. Sci. 429:195-206.
8. Dodgson, S.J., R.E. Forster, II, B.T. Storey and L. Mela. 1980. Proc. Natl. Acad. Sci. USA 77:5562-5566.
9. Vincent, S.H. and D.N. Silverman. 1982. J. Biol. Chem. 257:6850-6855.
10. Feldstein, J.B. and D.N. Silverman. 1984. Ann. N.Y. Acad. Sci. 429:214-215.
11. Montgomery, J.C., P.J. Venta and R.E. Tashian. Isozyme Bull. 14 (in press).
12. Maren, T.H. 1967. Physiol. Rev. 47:595-781.
13. Cammer, W.T., T. Fredman, A.L. Rose and W.T. Norton. 1976. J. Neurochem. 27:165-171.
14. Benesch, R.N., N.S. Barron and C.A. Mawson. 1944. Nature 153:138.
15. Linser, P. and A.A. Moscona. 1984. N.Y. Acad. Sci. 429:430-446.
16. Tashian, R.E. 1977. pp. 21-62. In Isozymes: Current Topics in Biological and Medical Research. M.C. Rattazzi, J.G. Scandalios and G.S. White, eds. Alan R. Liss, Inc., New York.

17. Venta, P.J., J.C. Montgomery, D. Hewett-Emmett, K. Wiebauer and R.E. Tashian. 1985. *J. Biol. Chem.* 260:12130-12135.
18. Boyer, S.H., H. Ostrer, K.D. Smith, K.E. Young and A.N. Noyes. 1984. *Ann. N.Y. Acad. Sci.* 429:324-331.
19. Venta, P.J., J.C. Montgomery, K. Wiebauer, D. Hewett-Emmett and R.E. Tashian. 1984. *Ann. N.Y. Acad. Sci.* 429:309-323.
20. Benton, W.D. and R.W. Davis. 1977. *Science* 196:180-2.
21. Yoshihara, C.M., M. Federspiel and J.B. Dodgson. 1984. *Ann. N.Y. Acad. Sci.* 429:332-334.
22. Curtis, P., J.E. Withers, D. Demuth, R. Watt, P.J. Venta and R.E. Tashian. 1983. *Gene* 25:325-332.
23. Rigby, P.W.J., M. Diekmann, C. Rhodes, and P. Berg. 1977. *J. Mol. Biol.* 113:237.
24. Dodgson, J.B., J. Strommer and J.D. Engel. 1979. *Cell* 17:879-887.
25. Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. *Molecular Cloning*. Cold Spring Harbor Laboratory, New York.
26. Southern, E. 1975. *J. Mol. Biol.* 98:503.
27. Girvitz, S.C., S. Bacchetti, A.J. Rainbow and F.L. Graham. 1980. *Anal. Biochem.* 106:492.
28. Maxam, A.M. and W. Gilbert. 1980. *Methods Enzymol.* 65:499-560.
29. Smith, D.R. and J.M. Calvo. 1980. *Nucleic Acids Res.* 8:2255-2274.
30. Simoncsits, A. and I. Torok. 1982. *Nucleic Acids Res.* 10:7959-7964.
31. Yamamoto, M., N.S. Yew, M. Federspiel, J.B. Dodgson, N. Hayashi and J.D. Engel. *Proc. Natl. Acad. Sci. USA* 82:3702-3706.
32. Hewett-Emmett, D., P.J. Hopkins and R.E. Tashian. 1984. *Ann. N.Y. Acad. Sci.* 429:338-350.
33. Mount, S.M. 1982. *Nucleic Acids Res.* 10:459-472.
34. Dodgson, J.B., S.J. Stadt, O.-R. Choi, M. Dolan, H.D. Fischer and J.D. Engel. 1983. *J. Biol. Chem.* 258:12685-12692.
35. Venta, P.J., J.C. Montgomery, D. Hewett-Emmett and R.E. Tashian. 1985. *Biochim. Biophys. Acta* 826:195-201.
36. Proudfoot, N.J. and G.G. Brownlee. 1976. *Nature* 263:211-214.
37. Goldberg, M. 1979. Ph.D. Thesis, Stanford University.

38. Efstratiadis, A., J.W. Posakony, T. Maniatis, R.M. Lawn, C. O'Connell, R.A. Spritz, J.K. DeRiel, B.G. Forget, S.M. Weissman, J.L. Slightom, A.E. Blechl and O. Smithies. 1980. *Cell* 21:653-668.
39. Dodgson, J.B. and J.D. Engel. 1983. *J. Biol. Chem.* 258:4623-4629.
40. Dolan, M., Dodgson, J.B. and J.D. Engel. 1983. *J. Biol. Chem.* 258:3983-3990.
41. Engel, J.D., D.J. Rusling, K.C. McCune and J.B. Dodgson. 1983. *Proc. Natl. Acad. Sci. USA* 80:1392-1396.
42. McKnight, S.L. and Kingsbury, R. 1982. *Science* 217:316-324.
43. Gidoni, D., W.S. Dynan and R. Tijan. 1984. *Nature* 312:409-413.
44. Dierks, P., A. van Ooyen, M.D. Cochran, C. Dobkin, J. Reiser and C. Weismann. 1983. *Cell* 32:695-706.
45. Fornwald, J.A., G. Kuncio, I. Peng and C.P. Ordahl. 1982. *Nucleic Acids Res.* 10:3861-3876.

APPENDIX

APPENDIX

Figure 9. Nucleotide sequence of 5'-flanking and intron regions of the chicken CA II gene. (A) This stretch of sequence is a continuation of the sequence presented in Figure 3 (5'-flanking region and exon 1). The sequence given here extends the 5'-flanking region from -200 bp upstream to -770 bp upstream of the CAP site. (B - G) Partial sequence from introns 1 - 6 is presented. Exons mark the start of the intron sequence and the intron-exon junction is indicated by a slash line. The numbers in the intron sequence refer to the approximate distances that separate sequenced portions of the intron reading from a 5' to 3' direction. Refer to the arrows above the subcloned DNA fragments in Figure 2 for the sequencing strategy.

INTRON 1 EXON 1
ACG/GTGAGTGTGGGGCACGGCGGGACGCTTGCCTGCCGGCG_cA
AXXXCCCGGG- 30-CGGGGGCGCTTCGGTTTTGCTTCTGTAACYXTGAGATGYCACCTT
TTGAGTGGTGCGCCGAGCATCCTYTGCGTTCCGGCACGCGGGGATGGTGCACAGGGTTG
AACCCAGCCCGGAGGAACCCGGCTGGCGTTCCCTCACCTCGGGGACTGCCCCGCCTGC
TCCCCGAGCGGGTTTTTCCCTCAGCCGTACAGGTGCCGCGGACGTCACGACCGCAGAGC
GGTACCTCTCTGACCCCCGAGCCCTGAGCGCCCCGCGCYCYCRRRRRAG/GACC EXON 2

INTRON 2 CAG/GTGAGCGCATCCGTGTGTGCGTGCGGCCAGAGCGCGACGCTGT
CGTCAATGCTGCTATTAATGCTAAGCATCTACTGGAGCCGCAGCGTTTCTTTTTTATTA
AAAATACATATATGTTGTAAACTGAGCTCCTCTTAGAAGAACGCTATGGCTACCTGGG
CTGCCCTACTCTAAACTCAGTAGAAGTGAAGTCTACCTTTGCAGTTAATTAAGGAGGAT
TTGGGCATGGGCTTGGATGGGAAGGAGGAAGAAGGGAAAAAAAGATGGATTATGAGAA
CAGTTATAGGTATAGCTCATAGTAAGCTGCTGGAACAAACGGTTAACGAGGAGGCAAAA
GGCACCACRRGRGAAACGGGCTGTAGAATAACAGGRGATTACAGTTACCAACATGGAG
CAGAAATAGTTTXGAAATAGTTTCTTTGGTTACCAAATATTATTTTCATATTTXGAAG
CCAGTATGCTCCCAGGTTAGTACAGAAACAGAGAAATGATTAGGAAATACCTCTCTGTT
CTCCTGTAACCTAAACAAAAATCCCTGACGTTTCTTGACXGAAGGCTTTCTAGATAAAG
ACAGTGTATACCTGGCTCATTCCCACAAAGTTTGCAGGCAGGCAGCTGATCCGCCGTGT
CCCATGGCTCACTGTGCTGTGCAAGTGTGGTTCAGCACCGTGTCTGCAAACCCCGCAGT
AACCTGCTGGAGGGAAYGGCCCCATCGGCGGGAAAAGGGANTC-200-GCGCGATTAA
CTTGTGXGGCTAGTGCCCTGTTTTGTXGGGCGCTTCCTTAGCGAGTGTCACTTAGXAAT
GGTGCTGAAACGAAACGAAACCGATGTCTACCGAAAAAATGTGGGGTTTTTTTTGCGG
ATAATACTATATGAAAGGTATGTCAGACAGCACTGCACTGGGAATTTGTATTTCACTTG
TGATCTTTATACXRACAGCGTTAGTGGGTCAGCTAAGTGCTTCACAGCTGCATAGATAC
ACATAAGCCTATGTATAATTTATTCXGAACTTCTGTATTCATATCTGTATAGCTTTAAA
AAACGGAAAAGAAGGGAAAAAGGTCTTCTGAGTTCTTAGCATTATTGCTCXGAAGCAC
XGAGGAGCCGGAGGXGAAGTGCATGTGTTTTTAGCAGTGTTAGGAGGAGTAGATCCCAA
AGAGTATAAAGAGAGRGGGATCC-5500-TCCCTCTGTCTCTGGAAGCAATCTTATAGT

TCAGCCATTTTGATTACTTCTTTGACAAGTTTATTGCTTCAGTGAGATTTATTAGTCTC
 CTGCCCTGTTACATCCTTACACAGGGAGCTATGGTAATTGCACAGTATTCCCGCCTAAC
 TTAGGGCATGTGACTGCTGCGACTTCCTGAAGAAAAATACAAAGTAGTCCACAGAATAT
 CCTGTAAAGATGCCCTGTTCTGCATGGGAAAAGTAACTGCCCAGTGACAGCCTTTGGGA
 ACAGCAGTGAGGTTGTGTGGCCTGAGTGAGCAGCTGTCCCGGGTGCAGCACAGTGCCCC
 TTGGTTTCTG-3000-CTTTATTGCAGCAGCGTGCCTTXGAGAAAGCCTGGAGTAGCGTG
 GCAGCATGCAGAGCAGTAACTAAXXXCACTTGGGAGACTTGAGTGGATGTTCCAGTCT
 GAAGTGAGACTTGATAGCAAATGTTCCAGAGGTGCTAATGGAGGCTCTTTGCTTTGCAG
 EXON 3
 /TGC

INTRON 3 EXON 3
 GAG/GTATGATGTGCTTTGCCTTTCCAGTATTCTTACCTAAGAGTAT
 TTTGATTTTCACCTTCCACTCTGGTAGTAATAGCTCTACAAACCTTCACGGGGCAAGAG
 ACTGTGGGGGTTACCATGGAAGATTTTGGTGCAGTGCCATATTAGGATXXAGYCCCTTG
 TAAATTATAGANTC-100-GGCAAAACACCATCTGCTCTCAATGTCCAACAAATTACTT
 CCCTGCATCCTCATTGCAAAATATTATGGATXGAATATAATTTTCACAGTGCTCCATG
 EXON 4
 TTTTCTTATTCCTAG/CTT

INTRON 4 EXON 4
 AAG/GTTAGTCAAACCTCTTTTTCTATTATTTGTGTTAGTGTAG-
 1250-TXTXTXXGATTAGCCTAAXTTAACAACCTATTTTGATTTXCTGCTATTGTAAA
 TGTCCAXGAGCATTTCTATCGGCCAAATTGTCTTGCAGCTTCTTGAGGAAGGTGATTG
 TAACCCAACTGTGTACACTCTGXC-130-AAACGGACGGCTCACATCCAAAAATAARCC

CACATGCTGTAAACXGAAATAATTTTATTCTGACCAGTGTCTCAGCCGACCTATTTGT
 GCTGTCAACTGGCTXGATTTTTAGRGAGACAGTTGCAGTCACTTCTGCTTAATATTCCT
 GTTATGTAAXGAAATAATGCTCXGAAAATGTGCTACAAAATGTGTCCTAGAGAATTGTA
 ACAGAAAAC TAAGGTGTCGAGAACTGTGCCATTAAGTGCTTTTCATGCTATATGTGTT
 ACAG/GTA^{EXON 5}

INTRON 5^{EXON 5} AAG/GTAATATTTTGTGTXGAATGCCACCAGACACAAXCRYGTGCCC
 GAXXGGGAGATTCTCTATCTTGCATGTCCACTGTGTGTATGAGAGTAAAACCATATCGG
 CACAAATGCTCCTGTAATCGTTAAAGCCAAAGATGTTCAAAGTAAGTTAGGGTAGAGGG
 CAGGTAATTTGTATCTGTGACTGCATGTATTTGTCTTTCAGCTGCATTGAAACCTGTAA
 GGTAAGGATAATCTTCCAGACAAGAAGTTTGATCAGGATAAAAAC TGGGGAAAAAATTA
 ACCTTTTGTATTTTAAATACTGTATAAGGCATATATGGGATCCTAAC-375-CTGTATT
 CCCRXGTATACTGTGTGXGAAATCTCATTATTCCTTCXXGTAGGATCAAGTCATATCT
 GTAAXGAAAAGCCCATAAATTCAACATTXGAAATGCXGAAATGTCTTGACTGTCCTAAA
 AAAATAAXGAATATAGTGCAGACGTACATAATGCCTGTACTGTAAGAGGTACTAACCCT
 GTATGCCCAACCCAGTTTCTCCCTTCCTCCAAATATAGAGCTTCTGTAGAAGACTTCAC
 TGTGACGAGCCAGTATTTGTATGTTGTCACTTCCTTAACAATAGTCTGCCTACCTTCC
 TTA^{EXON 6}CTGCAG/GGG

INTRON 6^{EXON 6} CAG/GTAGCTCTCTGGGGTAGTGCCAATTGTATTGATTGCTTTAATT
 TAGT-70-AAACGGCTTGCCTTTTTCTGGCAAGCACTGCTTCTTTATGTGGAAGTAAAG

GTTGATGTAAAATGAAAGGTTGTCCTCCATCACAAA-17-ATTAATGAAATTACAAAGA
GGCAAGATTTTATXXTGAGGTTTGCAXCTATGTGCT-750-AGCCTCATGGATCGGGAT
GATGATACATAATTCAGATAGTCCTGCTGCAATGAGTATGCATTTGAGGATGTTTTTGC
CTTTTCCCACAG/ATG

EXON 7

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03175 2680