Volume 15 Number 2 1987 Nucleic Acids Research

The chicken carbonic anhydrase II gene: evidence for a recent shift in intron position

Corinne M.Yoshihara+, Jiing-Dwan Lee and Jerry B.Dodgson

Departments of Microbiology and Biochemistry and Genetics Program, Michigan State University, East Lansing, MI 48824, USA

Received September 19, 1986; Revised and Accepted December 5, 1986

ABSTRACT

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 chicken genomic library. The sequence of a nearly full length chicken CA II cDNA clone has also been obtained. The gene is approximately 17 kilobase pairs (kb) in size and codes for a protein that is comprised of 259 amino acid residues. The 5' flanking region contains consensus sequences commonly associated with eucaryotic genes transcribed by RNA polymerase II. Six introns ranging in size from 0.3 to 10.2 kb interrupt the gene. The number of introns as well as five of the six intron locations are conserved between the chicken and mouse CA II genes. The site of the fourth intron is shifted by 14 base pairs further 3' in the chicken and thus falls between codons 147 and 148 rather than within codon 143 as in the mouse gene. Measurements of CA II RNA levels in various cell types suggest that CA II RNA increases in parallel with globin RNA during erythropoiesis and exists only at low levels, if at all, in non-erythroid cells.

INTRODUCTION

The carbonic anhydrases constitute an ancient family of proteins. Five different carbonic anhydrase (CA) isozymes (1-9) (possibly more, 10-12) have so far been identified, each of which is believed to be coded for by a separate genetic locus. 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, other CA functions have 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 variable in its expression pattern. For example, CA II is the only isozyme expressed in avian red blood cells, whereas both CA I and CA II isozymes are expressed in most mammalian red blood cells (16).

Only recently has CA gene structure been examined at the DNA level. DNA sequence data of the mouse CA II gene (17,18) and a rabbit CA I cDNA

clone (19) have been reported. Analysis of the mouse CA II gene showed that it was composed of seven exons that were stretched over 16 kb of DNA (18). The human CA II gene has been partially sequenced, and a comparison of human and mouse CA II promoters has been made (20). We previously described the characterization of a partial chicken CA II cDNA clone (21). In this paper the isolation and structural analysis of the complete chicken CA II gene is given. The chicken gene shows interesting similarities to and differences from the analogous mouse gene. In addition, CA II mRNA levels in various cells and tissues have been measured.

MATERIALS AND METHODS

Isolation of cDNA and Genomic Clones

A λgt10 cDNA library prepared from chicken red cell poly(A)$^+$ mRNA (22) was screened at a 99% representation of the library. The plaques were transferred to nitrocellulose filters and processed as described (23). 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 (17, generous gift of Dr. P. Curtis, Wistar Institute). Probes were prepared by nick translation (24). CA II cDNA fragments were isolated from positive λgt10 clones by digestion at the Eco RI linker sites and inserted into pBR325 plasmid DNA. A fine structure restriction map was derived for the subclone with the largest insert (pCA-1.2) and used for the sequence analysis of the cDNA. A λCharon 4A chicken genomic DNA library (23) was also screened with the 5'-end chicken CA II cDNA clone as described above. Restriction maps of the unique CA II-containing phage, designated λcaIII and λcaXVI, were prepared by standard multiple restriction digestion (24). Subclones of restriction fragments of phage DNAs were prepared by standard techniques (23,24) following their isolation from agarose gels (25). The resultant plasmid DNAs were further analyzed by restriction enzyme digestion and blot hybridization (26).

DNA Sequence Analysis

Subcloned CA II DNA was restriction enzyme digested, gel fractionated and the appropriate fragment isolated (25). Fragments were treated with calf alkaline phosphatase, 5'-end labeled with (γ-^{32}P)ATP, recut with the appropriate secondary enzyme and the resultant singly-labeled fragments isolated. Chemical degradation and gel electrophoresis were as described previously (27-29).

S1 Analysis

Restriction fragments were isolated from restriction enzyme digested

pBBca-2.8 by cleavage within exon 1 (RsaI site at +57, Fig. 3) and upstream beyond the likely start site (RsaI, -600). This fragment was labeled with polynucleotide kinase as described above, recut with SinI (-133) and the resultant 190 base pair (bp) fragment isolated. The end-labeled fragment (40 ng at 5 x 10^6 dpm/µg) was hybridized to 50 µg of each RNA to be tested at 55° in 80% formamide hybridization buffer (30) for 15 hours. Samples were quenched by dilution into 0.3 ml of S1 digestion buffer (30 mM NaOAc, 250 mM NaCl, 4 mM Zn $(OAc)_2$, 200 µg/ml denatured salmon sperm DNA, pH 4.5). S1 nuclease (1200 U/ml) was added, digestion was allowed to proceed for 15 min at 37°, and the reaction was stopped by phenol/chloroform extraction. (S1 digestion at 300 U/ml gave equivalent results.) Labeled DNA was run on a sequencing gel as described (27).

Miscellaneous

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. Enzymes were used according to manufacturers specifications. Other materials and bacterial strains were as previously described (23,24,30). RNA was isolated as described (23,24). Manipulations of recombinant DNA were done according to the appropriate NIH Guidelines.

RESULTS

Isolation of Larger Chicken CA II cDNA Clones

The partial chicken CA II cDNA clone previously described (21) was used to obtain more nearly complete clones from a λgt10 red cell poly(A)$^+$ cDNA library (22). The largest CA II cDNA insert (1.2 kb) subcloned into pBR325 is termed pCA-1.2. The cDNA clone was restriction mapped and its complete sequence was determined. It was found to code for sequences from codon 9 to a point about 440 bp 3' to the TAA stop codon. The cDNA sequence will be described in more detail below in comparison to the genomic sequence.

Isolation of the Chicken CA II Gene

A λCharon 4A chicken genomic library (23) was screened with the chicken CA II cDNA clone (21) containing the 5'-end of the cDNA. Two of the resulting clones, λcaIII and λcaXVI, have been mapped and characterized in detail. The restriction maps of the two phage are shown in Fig. 1. Subclones of appropriate phage restriction fragments were constructed, and their fine structure restriction maps are given in Fig. 2

Sequence of the Chicken CA II Gene

The subclone maps were used to develop a sequencing strategy for each

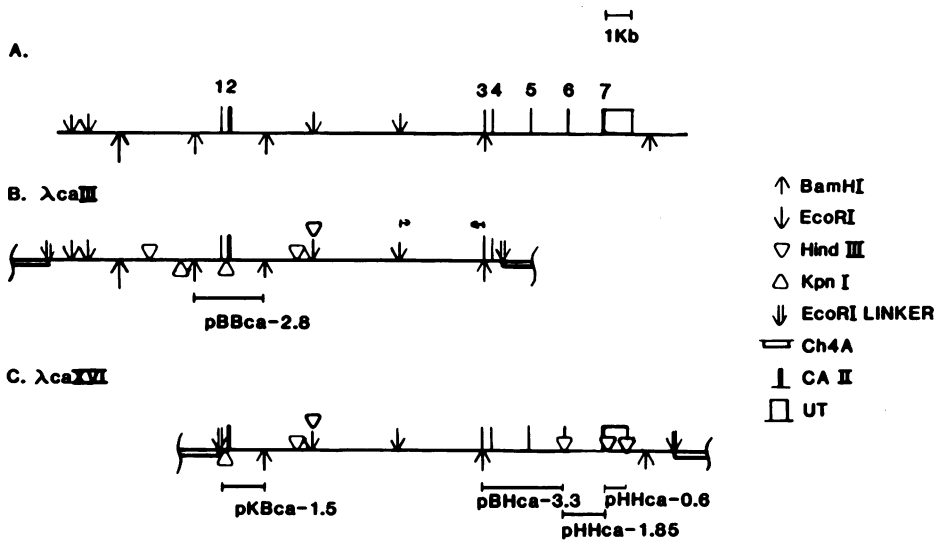

Figure 1. Restriction map of the chicken CA II gene locus. (A) Restriction map of chromosomal DNA contained within clones λcaIII and λcaXVI. (B) Restriction map of clone λcaIII which contains exons 1 to 4. (C) Restriction map of clone λcaXVI which contains exons 2 to 7. The solid boxes represent the 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. Only the BamHI and EcoRI sites are shown in line A.

of the CA II exons. The direction and extent of the regions sequenced are indicated by the arrows in Figs. 1 and 2. Except for that sequence 5' to codon 9, all the coding sequences were also sequenced on one or both strands of the cDNA clone, pCA-1.2.

The DNA sequence of the chicken CA II gene is given in Fig. 3. The genomic sequence shown encodes 259 amino acid residues and approximately 0.8 kb of untranslated region. The coding sequence of the gene was identified by comparison to the chicken CA II cDNA (pCA-1.2) sequence. The amino acid sequence predicted from the nucleotide sequence is identical to that predicted from the cDNA clone sequence. However, there are five differences in the cDNA and genomic sequence (Fig. 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 three 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;

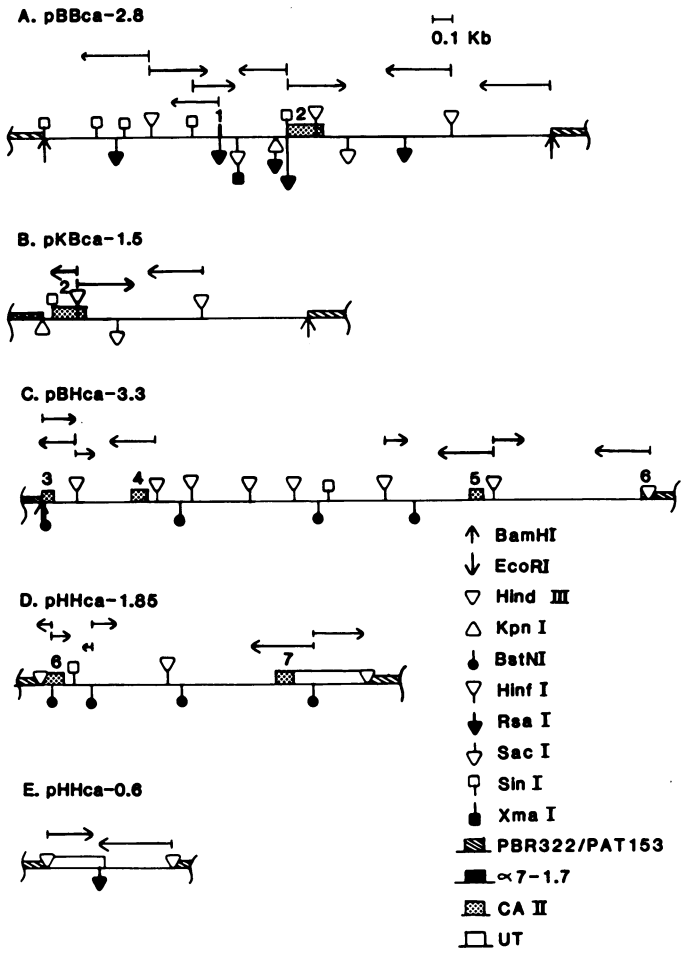

Figure 2. Restriction maps of the subcloned DNA fragments of the chicken CA II gene. Partial restriction maps of (A) subclone pBBca-2.8 which contains exons 1 and 2, (B) subclone pKBca-1.5 which overlaps with pBBca-2.8 and also contains exon 2, (C) subclone pBHca-3.3 which contains exons 3,4,5, and a 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. Open boxes show 3' untranslated sequences. The region designated as α7-1.7 (B) is a portion of chicken globin gene DNA used to provide the KpnI site for this subclone. The arrows above the boxes show the direction and extent of DNA sequence analysis.

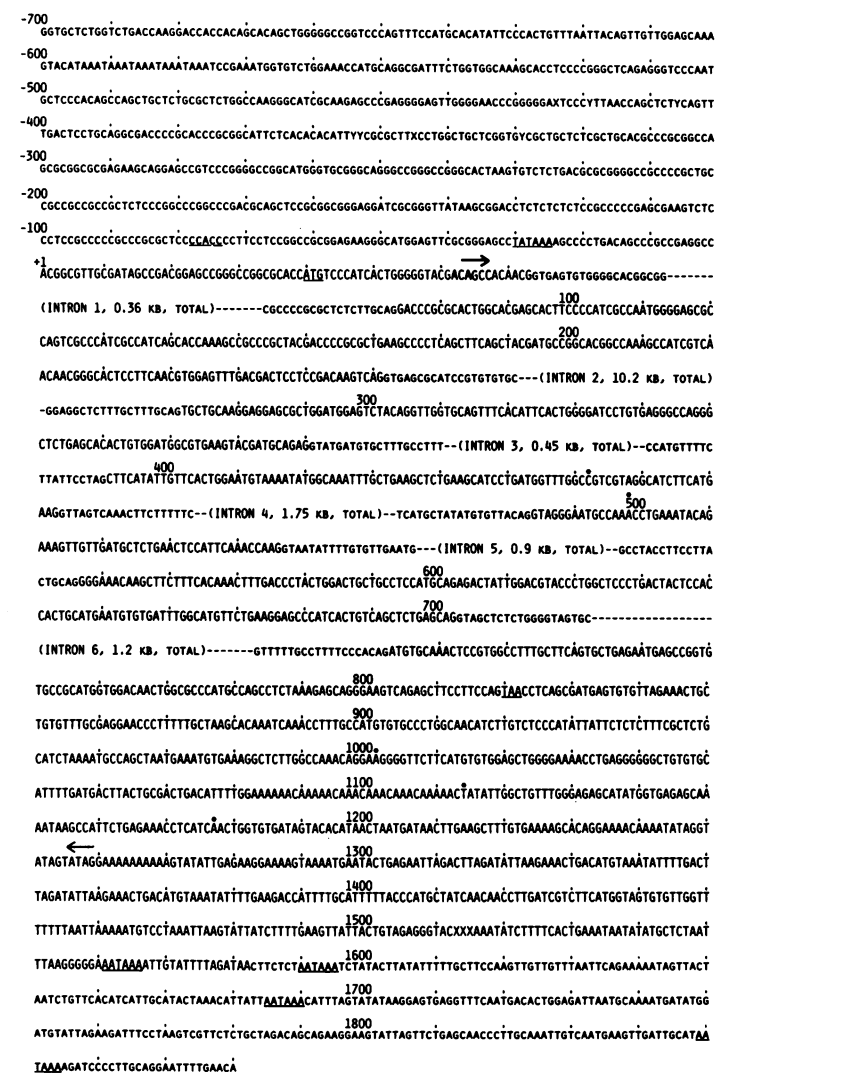

Figure 3. DNA sequence of the chicken CA II gene. The CCAAT, ATA, and AATAAA signal sequences as well as the initiation and stop codons are underlined. Numbers above the sequence indicate nucleotide numbering from the cap site (as +1). For convenience, intron sequences are ignored in the numbering. The intron sizes given include the partial intron sequences shown. The upper case letters indicate those sequences that are transcribed into mRNA and the lower case letters indicate those sequences that are processed or flanking. All the coding sequences present in exons within the left and right arrows above the sequence shown were also sequenced in the cDNA clone, pCA-1.2. Asterisks indicate sites where the cDNA and genomic clone sequences differ (see text). X indicates a nucleotide whose identity could not be resolved. Y indicates C or T.

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 (17,18). 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.

TABLE I. DNA sequence of intron donor and acceptor sites.

Intron	Donor	Acceptor
1	ACG/GTGAGT	CGCGCTCTCTTGCAG/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 (32)	$^{C}_{A}AG/GT^{A}_{G}AGT$	$(^{T}_{C})_{11}$ $N^{C}_{T}AG/G$

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. 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. (18) between YBR and BALB/c mouse strains.

The chicken CA II amino acid sequence (Fig. 4) has 65% homology to the mouse CA II sequence (18). There are 69 base changes that result in silent substitutions, and there are 164 base substitutions that result in amino acid changes. Overall, the nucleotide divergence between the mouse and chicken CA II genes is fairly evenly spread across all seven exons. The active site residues as well as the unique and invariant residues (31) are fairly well conserved. The chicken amino acid sequence is considered in more detail in the Discussion.

Chicken CA II Gene Intron/Exon Organization

The chicken CA II gene is interrupted by six 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 (Figs. 2 and 3). The locations of five of the six 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 human CA I and CA Z genes (12).

The 5' and 3' boundaries of the six introns of the chicken CA II gene

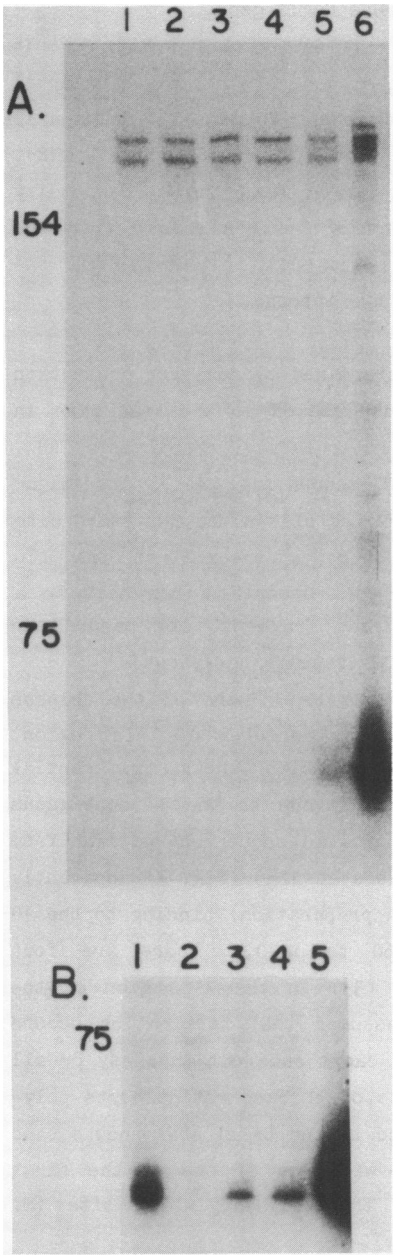

Figure 5. S1 analysis of CA II RNA levels. 50 µg of the various RNA samples were hybridized to the CA II 5' probe, digested, and the products electrophoresed as described in Materials and Methods. RNA samples were isolated from: breast muscle, lane 1; chicken embryo fibroblasts, lane 2; oviduct, lane 3; liver, lane 4; HD3 cells, lane 5; anemic red cell cytoplasm, lane 6. The dried gel was exposed for (A) 17 hr with no intensifying screen and (B) 40 hr with one intensifying screen. The position of co-electrophoresed labeled markers is shown by numbers to the left of the figure.

all fit the consensus donor and acceptor sequences (32) seen for most eucaryotic introns (Table I). 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 six chicken CA II introns are given in Fig. 3. In general, most of the intron sizes in the chicken gene are roughly similar to the corresponding mouse CA II introns (18). However, intron 1 of chicken is 0.35 kb which is approximately one-third the size of mouse intron 1 (18). Intron 2, on the other hand, is about 3 kb larger than the corresponding intron in mouse. The remaining four introns of chicken differ from the corresponding introns of mouse by anywhere from 0.1 to 0.8 kb but the difference is not as striking as in the first two introns.

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 (Fig. 5A, lane 6). A DNA fragment labeled at the RsaI site in exon 1 was hybridized to total chicken reticulocyte cytoplasmic RNA, digested with S1 nuclease and run on a 6% sequencing gel. The major protected fragment is about 58 bases in size which places the RNA start site (cap site) in the CCACG sequence about 39 bp upstream from the ATG initiation codon (Fig. 3). When the protected fragment is run next to a Maxam-Gilbert sequencing ladder of the S1 probe fragment, the major band corresponds to initiation at the A in the CCACG (results not shown).

We have been unable to definitively locate the 3'-end of the chicken CA II mRNA, apparently due to the very A:T rich character of this region. (S1 nuclease treatment of labeled DNA:RNA hybrids shows preferential cleavage at a site around 1460 in Fig. 3 probably due to the 17 contiguous A:T bp in this region as there are no poly(A) signal sequences upstream from this site.) The cDNA clone, pCA-1.2, terminates at 1259 (Fig. 3) apparently due to the oligo(dT) primer (used in the cDNA preparation) binding to the 10 contiguous transcribed A residues from 1260 to 1270. There are four potential poly(A) addition signal sequences (33) in the 3' region of the CA II gene (Fig. 3). We have arbitrarily assumed that termination occurs shortly 3' to the first of these although we can't rule out that any or all of the other three signals may be used. Blots of chicken reticulocyte poly-$(A)^+$ RNA run on denaturing gels and hybridized to the CA II cDNA show a band about 1650 nucleotides in size in agreement with use of one of the first three AATAAA signal sequences to position the polyadenylation site (M. Federspiel and J. Dodgson, unpublished results).

CA II RNA Levels

We have used the S1 nuclease protection assay to measure CA II RNA levels in several cell types. Levels of CA II RNA, as expected, are quite high in total cytoplasmic RNA isolated from anemic hen reticulocytes

(Fig. 5A, lane 6). A lower, but still significant, level of CA II RNA is observed in total cellular RNA isolated from uninduced HD3 cells (Fig. 5A, lane 5). The HD3 cell line is an erythroid progenitor cell line transformed by temperature-sensitive avian erythroblastosis virus, and its uninduced state has been shown to correspond roughly to the CFU-E (erythroid colony forming unit) stage of erythroid development (34-36). Comparison of several different exposure times of the gel shown in Fig. 5 indicates that anemic reticulocytes contain about 50-fold more CA II than uninduced HD3 cells. This difference is comparable to that seen for αA-, αD-, and β-globin RNAs (36; Wynne Lewis and J. Dodgson, unpublished results). The bands observed above 154 bases in Fig. 5 are probably due to undigested probe DNA, both free single strand and renatured double strand. The faint band at about 90 bases in all lanes is due to slight contamination of the 190 bp probe with the SinI/RsaI fragment from the other end of the originally labeled RsaI fragment (see Materials and Methods). A large excess of labeled probe was used in all lanes to insure that the observed signal was proportional to CA II RNA. We can't rule out that there is a small constant level of CA II RNA initiated at upstream sites in all samples tested. Clearly, however, the major CA II initiation site in reticulocytes is at the proposed cap site.

A longer exposure of this protection experiment (Fig. 5B) shows that low but measurable levels of CA II RNA are observed in total RNA isolated from adult liver (lane 4), oviduct (lane 3) and breast muscle (lane 1) whereas CA II RNA levels were not detectable in chicken embryo fibroblast RNA (lane 2) from cells grown in culture. The three adult tissues show about 1/10 the level of CA II RNA as do the HD3 cells. Control experiments which measure red cell contamination in these tissues by S1 protection of an αA-globin gene probe demonstrated that much, if not all, of the CA II RNA in liver may arise from red blood cell contamination (results not shown). Red cell contamination in the oviduct RNA was not detectable in this assay, and breast muscle RNA was not tested.

DISCUSSION

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

Over 700 bp upstream from the CA II coding region have been sequenced. Putative signal sequences that are common to eucaryotic genes transcribed by RNA polymerase II are found in the chicken gene (Fig. 3). A Goldberg-Hogness (37) or ATA block is located 23 to 30 bp upstream from the cap site.

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

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

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

Figure 6. 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.

Although a sequence that corresponds accurately to the consensus sequence CCAAT (38) cannot be identified, a region (at -74 in Fig. 3) that has limited homology (CCACC) to the CCAAT site can be found. The absence of good consensus CCAAT sequences has also been noted for the adult chicken α-globin genes (39). Fig. 6 compares the -120 to +1 region of the chicken CA II gene to the corresponding regions of the mouse and human genes (18,40). It can be seen that the ATA and putative CCAAT sequences are very similar among the three genes both in terms of actual sequence and approximate spacing relative to the transcription start site. (The actual cap sites of the two mammalian CA II genes have not been determined experimentally, but are estimated from their sequence.) The 5' untranslated regions of the three CA II genes are different in both sequence and length being 39, 59, and 73 nucleotides long in chicken, mouse, and human, respectively.

As shown in Fig. 6, the region from 45 to 22 bp upstream of the mRNA start site (which includes the TATAAA sequence) is over 95% homologous between the chicken and mouse sequences. The homology in this 23 bp region between the CA II genes exceeds that seen between analogous chicken and mammalian globin gene promoter regions (30,39,41,42), suggesting that these

sequences may have an important role in the regulation of CA II gene expression. The position of this sequence block just 5' to and including the ATA region may indicate a role in controlling the initiation of transcription of these genes. The overall homology between the chicken and mammalian CA II genes in the -80 to +1 region is about 60% which is also unusually high in comparison with the same region in globin genes.

McKnight and Kingsbury (43) 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 (44). The core 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 CCACC box in the chicken CA II gene (Fig. 6). The mouse and human CA II genes contain several such sequences both 5' to CCAAT and between CCAAT and ATA (18,40). While no exact match to the consensus sequence is seen between CCACC 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. This latter GC-rich region (-69 to -59, Fig. 6) constitutes the one other region of high homology between the chicken and mammalian CA II gene promoters besides the CCACC and -45 to -22 regions discussed above.

An imperfect tandem repeat of 14 to 15 bp that is thought to function as an upstream promoter element and is found in five mammalian and one avian β-globin genes and in several rat pancreatic genes (40,45) is also found in the human and mouse CA II genes as CCNGTCACCTCCGC (40). In the β-globin genes this tandem repeat is 9 to 25 bp upstream from the CCAAT sequence in mammals and 55 bp upstream in chickens. In the human and mouse CA II genes these repeat elements have been found 15 and 22 bp upstream, respectively, from the CCAAT boxes. There are similar repeat elements 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, these two elements show a rather poor match to the consensus. The function of all 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.

Amino Acid Sequence of Chicken CA II

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 analogous active site residues of mammals most of the chicken residues are found to be conserved (Fig. 4). The 15 residues that are invariant in all of the CA I, II and III proteins of all species sequenced to date are also invariant in chicken. At position 90 the residue that is present in most mammals is Ile except for ox which has Val (31). Chicken is similar to ox in that it too has Val at this position. 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. When the overall chicken CA II amino acid sequence is compared to those of mouse and human there is 65% sequence homology to the mouse sequence and 70% homology to the human sequence.

Hewett-Emmett et al. (31) have compiled those amino acid residues that, to date, are invariant among the known examples of a specific CA isozyme but unique to that isozyme. The chicken CA II gene codes for 9 of the 15 previously unique and invariant residues for CA II including both of those (Asn at 66 and Glu at 68) in the active site. The gene possesses only 1 of 18 unique and invariant residues for CA I and 8 of the 39 for CA III. These results, along with the extensive nucleotide sequence homology of the chicken gene to the mouse CA II gene (Fig. 4) confirm its assignment as a CA II isozyme gene.

Exon/Intron Organization

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 (46). The change in intron 4 position between the two CA II genes shifts this intron 14 bp toward the 3' end in the chicken relative to the mouse CA II gene (Fig. 7). This "new" intron position has also recently been found in the human CA I and CA Z genes (12). The similarity of the chicken CA II intron position to that of the human CA I and CA Z introns suggests that the chicken CA II gene structure is the more ancient form. 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

```
          141                                                    145        148
          VAL LEU G                                          LY TYR PHE LEU LYS ILE GLY
MOUSE     GTT TTG G/GT ATT TTT TC. ... TGC CCT GCA G/GC TAT TTT TTG AAG ATT GGA

CHICKEN   GTC GTA GGC  ATC TTC ATG AAG/GTT AGT ... CTA  TAT GTG TTA CAG/GTA GGG
          VAL VAL GLY  ILE PHE MET LYS                                VAL GLY
          141          145                                           148
```

Figure 7. Comparison of intron 4 location in chicken and mouse CA II genes. The upper half of the figure shows the nucleotide and amino acid sequence of the mouse CA II gene around intron 4 (18). The corresponding region of the chicken CA II gene is shown on the lower half of the figure. Slashes indicate the boundaries of the intron region in both genes.

most mammalian/avian homologues. The mutational shift may have occurred after divergence of birds and mammals. Fig. 7 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 to 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 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 to 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*.

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, intron junctions are fairly well conserved. There are some differences in the size of the analogous introns, but the effect of these changes is likely to be minimal.

CA II RNA Levels

Our preliminary assays of CA II RNA levels present in different chicken tissues and cells (Fig. 5) show that CA II is induced over 100-fold during erythroid differentiation. Comparison of CA II and globin RNA levels in HD3 erythroid progenitor cells versus anemic hen reticulocytes suggests that CA II and the adult globin genes are induced approximately in parallel in agreement with the protein studies of Weil *et al*. (47) for human and mouse

erythroid progenitor cells (but not for mouse erythroleukemia cells). As mentioned previously, chickens differ from most mammals in that they appear to express only CA II in their red cells as opposed to both CA I and CA II (16). The late induction of CA II and globin RNAs in chick red cell maturation is in contrast to the red cell-specific H5 histone gene whose RNA levels in HD3 cells are similar to those in mature reticulocytes (34, Paul Boyer and J. Dodgson, unpublished results). Despite the possible requirement for carbonic anhydrase activity in several, if not all, non-erythroid tissues, the levels of CA II RNA in liver, oviduct and muscle appear to be no more than 0.2% that seen in anemic reticulocytes, and CA II RNA was undetectable in chicken embryo fibroblasts grown in culture. Further, more sensitive, experiments will be required to accurately assess RNA levels in adult (and embryonic) tissue of CA II and other CA isozymes.

ACKNOWLEDGMENTS

We are grateful to Dr. P. J. Curtis for providing the mouse carbonic anhydrase cDNA clone. We also thank Drs. R. E. Tashian, D. Hewett-Emmett, J. C. Montgomery, and P. J. Venta for providing a human CA II exon 1 clone, and for communication of results prior to publication. We thank Mark Federspiel and Wynne Lewis for some of the RNA samples, the cDNA library, and the αA-globin probe used. This research was supported by Grants GM28837 and a Research Career Development Award to J.B.D. from N.I.H. This is Journal Article 12103 from the Michigan Agricultural Experiment Station.

+Present address: Department of Biology, 16-820 Massachusetts Institute of Technology, Cambridge, MA 02139, USA

REFERENCES

1. Tashian, R.E., Hewett-Emmett, D. and Goodman, M. (1983) In Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (eds), Isozymes: Current Topics in Biological and Medical Research. Alan R. Liss, Inc., New York. pp. 79-100.
2. Sapirstein, V.S., Strocchi, P. and Gilbert, J.M. (1984) Ann. N.Y. Acad. Sci. 429, 481-493.
3. Whitney, P.L. and Briggle, T.V. (1982) J. Biol. Chem. 257, 12056-12059.
4. Henry, R.P. and Cameron, J.N. (1983) J. Exp. Biol. 103, 205-223.
5. McKinley, D.N. and Whitney, P.L. (1976) Biochem. Biophys. Acta 445, 780-790.
6. Sanyal, G., Pessah, N.I. and Maren, T.H. (1981) Biochem. Biophys. Acta 657, 128-137.
7. Wistrand, P.J. (1984) Ann. N.Y. Acad. Sci. 429, 195-206.
8. Dodgson, S. J., Forster, R.E., II, Storey, B.T. and Mela, L. (1980) Proc. Natl. Acad. Sci. USA 77, 5562-5566.

9. Vincent, S.H. and Silverman, D.N. (1982) J. Biol. Chem. 257, 6850-6855.
10. Feldstein, J.B. and Silverman, D.N. (1984) Ann. N.Y. Acad. Sci. 429, 214-215.
11. Montgomery, J.C., Venta, P.J. and Tashian, R.E. (1986) Isozyme Bull. 19, 12.
12. Venta, P.J., Montgomery, J.C. and Tashian, R.E. (1986) In Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (eds), Isozymes: Current Topics in Biological and Medical Research. Alan R. Liss, Inc., New York, in press.
13. Cammer, W.T., Fredman, T., Rose, A.L. and Norton, W.T. (1976) J. Neurochem. 27, 165-171.
14. Benesch, R.N., Barron, N.S. and Mawson, C.A. (1944) Nature 153, 138-139.
15. Linser, P. and Moscona, A.A. (1984) Ann. N.Y. Acad. Sci. 429, 430-446.
16. Tashian, R.E. (1977) In Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (eds) Isozymes: Current Topics in Biological and Medical Research. Alan R. Liss, Inc., New York. pp. 21-62.
17. Curtis, P.J., Withers, E., Demuth, D., Watt, R., Venta, P. J. and Tashian, R.E. (1983) Gene 25, 325-332.
18. Venta, P.J., Montgomery, J.C., Hewett-Emmett, D., Wiebauer, K. and Tashian, R.E. (1985) J. Biol. Chem. 260, 12130-12135.
19. Boyer, S.H., Ostrer, H., Smith, K.D., Young, K.E. and Noyes, A.N. (1984) Ann. N.Y. Acad. Sci. 429, 324-331.
20. Venta, P.J., Montgomery, J.C., Wiebauer, K., Hewett-Emmett, D. and Tashian, R.E. (1984) Ann. N.Y. Acad. Sci. 429, 309-323.
21. Yoshihara, C.M., Federspiel, M. and Dodgson, J. B. (1984) Ann. N.Y. Acad. Sci. 429, 332-334.
22. Yamamoto, M., Yew, N.S., Federspiel, M., Dodgson, J.B., Hayashi, N. and Engel, J.D. (1985) Proc. Natl. Acad. Sci. USA 82, 3702-3706.
23. Dodgson, J. B., Strommer, J. and Engel, J.D. (1979) Cell 17, 879-887.
24. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. Girvitz, S.C., Bacchetti, S., Rainbow, A.J. and Graham, F. L. (1980) Anal. Biochem. 106, 492-496.
26. Southern, E. (1975) J. Mol. Biol. 98, 503-518.
27. Maxam, A.M. and Gilbert W. (1980) Methods Enzymol. 65, 499-560.
28. Smith, D.R. and Calvo, J.M. (1980) Nucleic Acids Res. 8, 2255-2274.
29. Simoncsits, A. and Torok, I. (1982) Nucleic Acids Res. 10, 7959-7964.
30. Dodgson, J.D., Stadt, S.J., Choi, O.R., Dolan, M., Fischer, H.D. and Engel, J.D. (1983) J. Biol. Chem. 259, 12685-12692.
31. Hewett-Emmett, D., Hopkins, P.J. and Tashian, R.E. (1984) Ann. N.Y. Acad. Sci. 429, 338-350.
32. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
33. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
34. Beug, H., Palmieri, S., Freudenstein, C., Zentgraf, H. and Graf, T. (1982) Cell 28, 907-919.
35. Samarut, J. and Gazzolo, L. (1982) Cell 28, 921-929.
36. Weintraub, H., Beug, H., Groudine, M. and Graf, T. (1982) Cell 28, 931-940.
37. Goldberg, M. (1979) Ph.D. thesis, Stanford University, Palo Alto, California.
38. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R.M., O'Connell,

C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.
39. Dodgson, J.B. and Engel, J.D. (1983) J. Biol. Chem. 258, 4623-4629.
40. Venta, P.J., Montgomery, J.C., Hewett-Emmett, D. and Tashian, R.E. (1985) Biochem. Biophys. Acta 826, 195-201.
41. Dolan, M., Dodgson, J.B. and Engel, J.D. (1983) J. Biol. Chem. 258, 3983-3990.
42. Engel, J.D., Rusling, D.J., McCune, K.C. and Dodgson, J.B. (1983) Proc. Natl. Acad. Sci. USA 80, 1392-1396.
43. McKnight, S.L. and Kingsbury, R. (1982) Science 217, 316-324.
44. Gidoni, D., Dynan, W.S. and Tjian, R. (1984) Nature 312, 409-413.
45. Dierks, P., Van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weismann, C. (1983) Cell 32, 695-706.
46. Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) Nucleic Acids Res. 10, 3861-3876.
47. Weil, S.C., Walloch, J., Frankel, S.R. and Hirata, R.K. (1984) Ann. N.Y. Acad. Sci. 429, 335-337.