BBA 91518

# Comparison of the 5′ regions of human and mouse carbonic anhydrase II genes and identification of possible regulatory elements

Patrick J. Venta, Jeffry C. Montgomery, David Hewett-Emmett * and Richard E. Tashian

*Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109 (U.S.A.)*

(Received July 31st, 1985)

Key words: Carbonic anhydrase II gene; Nucleotide sequence homology; Regulatory sequence; 5′ Flanking region; (Human, Mouse)

**The nucleotide sequence of the 5′ region of the human carbonic anhydrase II gene has been determined. This sequence begins 643 base pairs upstream from the ATG start site and continues through exon 1, intron 1, exon 2 and the adjoining 125 nucleotides of intron 2. The human sequence is compared with homologous regions of the mouse (YBR strain) carbonic anhydrase II gene by aligning the two sequences for optimal homology. In addition to a TATA box and a putative CCAAT box (CCACC in human and CCACT in mouse), three conserved tandem-repeat elements in mouse and two in human (consensus: cCNGTCACCTC-CgC) are located 15 and 22 base pairs upstream, respectively, from the CCAAT boxes in the human and mouse sequences. This repeat element is similar to a tandem repeat sequence located at about the same position in mammalian β-globin genes, and may represent regulatory elements common to both the carbonic anhydrase and β-globin genes. The regions surrounding exon 1 are extremely G + C-rich in both human and mouse genes. In addition, severel CCGCCC or GGGCGG sequences which may be important for transcriptional efficiency are found in the 5′ flanking regions of the human and mouse genes.**

## Introduction

The mechanisms responsible for the tissue-specific regulation of genes in eukaryotes are still largely unknown. Several general regulatory elements have been identified such as the TATA box, the CCAAT ·box, and GC-rich elements which occur upstream from the transcriptional initiation site [1–3]. In several systems, tissue-specific regulatory elements have also been located within a few hundred base pairs (bp) of the transcription initiation site [4–8], although some regulatory sequences are found at distances of up to about 2 kb (kilobase pairs) 5′ from the initiation site [5,9].

The carbonic anhydrase (CA) multigene family is well suited for the study of tissue-specific gene regulation. This gene family is made up of at least five genes, and the products of three of these genes (termed CA I, CA II and CA III) have been structurally characterized from a number of vertebrate species [10]. One of the other two genes is expressed in liver mitochondria [11,12] and another encodes a membrane-bound form in kidney [13] and lung [14]. The structure of the mouse CA II gene has been determined [15], and the nucleotide sequences of mouse CA II cDNA [16] and the rabbit CA I cDNA [17] have been reported. The CA II gene in mammals is expressed in many different tissues, but in only a limited subset of cells within each tissue [18], and thus

* Present address: Center for Demographic and Population Genetics, University of Texas Health Science Center at Houston, Houston, TX 77225, U.S.A.

0167-4781/85/$03.30 © 1985 Elsevier Science Publishers B.V. (Biomedical Division)

may be expected to have a variety of controlling elements. For example, although erythrocytes, osteoclasts and granulocytes are all presumably derived from stem cells in the bone marrow [19], CA II is expressed only in erythrocytes and osteoclasts [18,20,21]. In addition, CA II is expressed in glial cells but not in neurons [18,22], although both of these brain cell types presumably originate from cells of the neural crest [23]. This pattern of expression is in contrast to that of the so-called 'housekeeping' genes [24] which are expressed in all cells, and other genes, such as the globins which are primarily expressed in only one, or a few, cell types. If each tissue has a different system to regulate a subset of genes, the CA II gene might be expected to have multiple regulatory regions, each for a different tissue.

As a first step in determining the location of the regulatory sequences of the CA II genes, we have compared the nucleotide sequences surrounding the first and second exons of the human and mouse CA II genes. Underlying this approach is the assumption that important regulatory sequences common to mammalian tissue-specific gene expression will have been conserved in evolution.

## Materials and Methods

*Origin of the clones.* The *Eco*RI subclone, H25-3.8, containing the first two exons of the human CA II gene was prepared from a previously isolated λ clone, and used to detect a polymorphism 1 kb upstream from the human CA II gene [25,26]. The *Eco*RI subclone, A6-2.7, containing the first two exons of the mouse CA II gene, was prepared from a previously isolated cosmid clone [15,25].

*Sequence analysis.* Sequencing was performed by the chemical cleavage method as previously described [15]. Storage and analysis of the data were performed using programs available for microcomputers [27].

## Results and Discussion

The strategy used to determine the human and mouse CA II DNA sequences reported here is shown in Fig. 1, and comparisons of the two gene sequences are shown in Fig. 2 and Table I. As shown in Table I, although the nucleotide sequences in the two exons have remained highly conserved (88 and 79% homology), the sequences in the introns have diverged to a greater extent (44–59% homology). When the complete 259 amino acid sequences of human and mouse CA II are compared, the amount of identical residues at

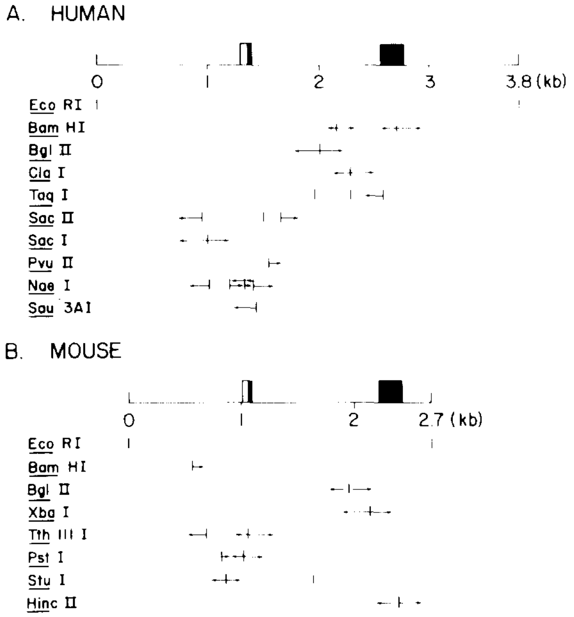

Fig. 1. Restriction maps of the 5′ regions of human and mouse (YBR strain) carbonic anhydrase II genes. Open boxes indicate untranslated portions of exons 1 and solid boxes represent coding regions of exons 1 and 2.

TABLE I

COMPARISON OF HOMOLOGOUS PORTIONS OF THE MOUSE AND HUMAN CA II GENES

For the number of positions compared, insertions and deletions were counted as one position.

| Sequences compared | No. of positions compared | Identical bases (%) |
|---|---|---|
| Region 5′ to TATA box | 243 | 60 |
| TATA box to start codon (complete) | 88 | 69 |
| Exon 1 (coding region; complete) | 34 | 88 |
| Intron 1 (5′ end) | 135 | 44 |
| Intron 1 (3′ end) | 425 | 59 |
| Exon 2 (complete) | 198 | 79 |
| Intron 2 (5′ end) | 124 | 59 |

homologous positions is 81% [16]. The predicted amino acid sequences for the coding regions of the first and second exons of the human and mouse CA II gene agree completely with the reported amino acid sequences of human CA II [28] and the derived amino acid sequence of mouse CA II [16], except for a single amino acid strain difference at position 38 (His/Gln) [15].

As can be seen in Fig. 2, a TATA box (TATAAAA) is located 99 and 86 bp, respectively, upstream from the coding regions in both human and mouse genes, and possible CCAAT boxes (CCACC in human and CCACT in mouse) are located about 42 bp upstream from the TATA boxes. The regions surrounding the TATA boxes, about 44 bp downstream and 42 bp upstream, have remained well conserved. Also, the regions 700–800 bp on either side of the first exon are extremely G + C-rich in both the human and mouse CA II genes. Such G + C-rich 'islands' have been shown to be associated with the 5′ regions of many mammalian genes [29,30]. About 90% of methylated cytosines in mammalian DNA are found in the CpG sequence [31], and this type of methylation has been implicated in the control of gene expression in higher eukaryotes [31].

In addition to the TATA and CCAAT boxes, a number of CCGCCC or GGGCGG sequences are

TABLE II

COMPARISON OF PRESUMED REGULATORY SINGLE OR REPEATED ELEMENTS IN THE 5′ FLANKING REGIONS OF SOME MAMMALIAN GENES

The positions are the number of nucleotides from T in AṪA box (T = −1) to A in CȦCC element. Consensus sequences: 50–80%, lower case letters; 80–100%, capital letters; R = purine; N = any base. References for the mammalian adult β-globin sequences and rat pancreatic enzyme sequences are from Collins and Weissman [39] and Swift et al. [42], respectively. Gaps have been introduced to optimize the sequence homologies.

| Source | Gene | Position | Sequence |
|---|---|---|---|
| Human | CA II | -93<br>-80 | GAGTTCȦCCTCCGC<br>CCCGTCȦCCTCCTC |
| Mouse | CA II | -109<br>-96<br>-83 | CCTGTCȦCCTCTGC<br>CCTGTCȦCCTCCGT<br>TCCGCCȦCCTCCAC |
| | CA II (consensus) | | cCNGTCȦCCTCCgC |
| Human | β-globin | -75<br>-60 | GACCTCȦCC-CTGT<br>AGCCACȦCC-CTAG |
| Mouse | $\beta^{maj}$-globin | -87<br>-61 | GTCCTCȦCC-GAAG<br>AGCCACȦCC-CTGG |
| Mouse | $\beta^{min}$-globin | -77<br>-62 | AGCCTCȦCC-CTGC<br>GGTAACȦCCCCTGG |
| Rabbit | β1-globin | -86<br>-75<br>-60 | GTCATCȦCC-CAGA<br>GACCTCȦCC-CTGC<br>AGCCACȦCC-CTGG |
| Goat | $\beta^{A}$-globin | -75<br>-60 | AGCCTCȦCC-CTGT<br>AACCACȦAC-TTGG |
| | β-globin (consensus) | | RgCCtCȦCC-CTGg |
| Rat | Elastase I | -83 | CATGTCȦCCTGTGC |
| Rat | Elastase II | -69 | -ATTCCȦC-TGGGC |
| Rat | Chymotrypsin B | -182 | CAGGGCȦCCTGTCC |
| Rat | Trypsin I | -151 | CTTGTCȦCCTGTAG |
| Rat | Trypsin II | -182<br>-130 | GTTTCCȦC-TGGTT<br>CTTATCȦC-TGACC |
| | Pancreatic exocrine genes (consensus) | | c$\overset{t}{a}$TgtCȦCcTGtNc |

```
                                                  -600
          .         .         .         .         .         .         .         .         .         .
HCAII  ACCCGCTTGCCGCCCCAGGCCGGGGACACCAGAGACTGAGCCCTTGCGCGCTGGAGACCCGGCGCGGGTGGGGCGGGAGGGCACCGGGGCCAAGAGACAG

                                                  -500
          .         .         .         .         .         .         .         .         .         .
HCAII  GTCACAATGGGGACGGGGACGAAGGCGCCGGGGGTCCGGGTCCCGAGCAGTCCCCGCCGCCGCCAGACTCCGCAGCCGGGAGGGGGATGGGGTCGCGGAG

                                                  -400
          .         .         .         .         .         .         .         .         .         .
HCAII  AGGCTCCCCGGCCCCGGGTTGGACGGAGGAGCCCAGGAGCCACCGCTGCCGCCGCACGCCCCGAGCTCCGAGCCCGCGCCAGCCCTGGCCCCCGGGGCCCT
                                                                                                          *
MCAII                                                                                                AAGTCG

                                                  -300
          .         .         .         .         .         .         .         .             .         .
HCAII  GCCGTGGGCCGAGGGGAGCCGGGCGGCGGGAGAGGTGCCCCGGTGCCCCGGTGCCCCGGTGCCCCGCCGGCGCCCAGCC----GAGAGGGCGTTTTCCCC
       *   *       ***  *  * *  * * *        *  * *** * *   *   *    *  *   **** ****      **** * **   * *
MCAII  GTGCTCACAATGGGGCTGGGGAGTAGGGCGCCGACCCGCGAGCTGCGCAGCGCCGGGGCGAGCTGGAGAGCGCGCAGCTGCTGCAGAGCGGGTCGGCGCA

                                                      -200
              .         .         .         .         .
HCAII  GCCCCCAGGAGACTCGCCCGCCCCGGCCGCCCCGCCCCGAGCGCGGGAGT--------------------------------------------TCACCT
       *      *****   *** *     **************  *  **  ** *                                             ******
MCAII  G--AAGCGGAGAGCAGCCGG-AGGGCCGCCCCGCCCCTATAGCAAGGTGCGGGGGCCCTCGAACGAGAGGCCTTCCGCCTGTCACCTCTGCCTGTCACCT

        .                            .         .         .         .         .         .         .
HCAII  CCGCCCGTCACCTCC---------TCCCCTTGTCGCCTAGGTCCACCCGAGCCCCTCCCCCGGGCCGCCCCCGAGCACGAAGTTGGCGGGAGCCTATAA
       *** *** *******         ********   *  ******** *  *  ******** *********  *****  ********************
MCAII  CCGTCCGCCACCTCCACGGTCTCCTCCCCTTG---CTCAGGTCCACTC--GGTCCCTCCCCTGGGCCGCCC-AGAGCAGCAAGTTGGCGGGAGCCTATAA

     -100
       .         .         .         .         .         .         .         .         .         .
HCAII  AAGCGGGCCGGCGCGACCCCGGACACACAGTGCAGGCGGCCCAAGCCGCCGGCCGCCAGATCGGTGCCGATTCCTGCCCTGCCCCGACCGCCAGCGCGACC
       **** ** *** ** ******* ******* ******  * * **                *  * * * * **** ******* * * ** *** ****
MCAII  AAGCCGGACGGTGCCACCCGCG-ACACACACTGCAGGGCCGCTAG------------ACCGGACGACAACTTCTGCTCTGCCCCAATGACCGGCGTGACC

                     exon 1               .                             intron 1
       I                                                                                                 100
       MetSerHisHisTrpGlyTyrGlyLysHisAsnG     .         .         .         .         .         .         .
HCAII  ATGTCCCATCACTGGGGGTACGGCAAACACAACGGTGAGTGCCGGCGACGGCCAGCGCGGGGGCGCCCCGATCCCCGATCCCCGATCCCCGATCCCCGAT
       ******** ******** *** **** **************  *  * **   *    ****     ****   * **   *  *  **    ** *
MCAII  ATGTCCCACCACTGGGGATACAGCAAGCACAACGGTGAGTGGGGAGGCCGGCGCGCTCCGGGCCGCGCCGGGTCTCGGCTCAGGCACCTGCCTCGGT--C
       MetSerHisHisTrpGlyTyrSerLysHisAsnG

                                                                                                         200
                .         .         .         .         .         .         .         .         .         .
HCAII  CCCCGATCCCCGATCCCGGTCGCCGGCCCGGGGCGCAGCGCCCGCACATGCTGTTTACCGCGGCCGCGGTGGTGCTGGAGGCTCAGGTGCGGCCCCGGGG
       * * ****** ** * * *  *   *** **    *              ***    ***   *    *
MCAII  CGCAGATCCCGGAGCGGCAGCTGGATGCCGGGTCTCCGCTCTGGTACGTGCCTGGCGCCCGACTCGAGAGC

                                                                                                         300
                .         .         .         .         .         .         .         .         .         .
HCAII  CGCTCCGCTCGCGGCTCCGCGCGCCGGGGATGTCCCCCTTGCCCCAGCTGCGAGGCCACTGTGGAGGAATCCCCGCGCCGCCGGAGGAGGGCCCGAGGGA

                                                                                                         400
                .         .         .         .         .         .         .         .         .         .
HCAII  GGGGAAGGCGCGGCCGACCGCGGGACCCGAGGACAGTCCCTCCCGGGTCCCGACCTGGGGATCATTTTAACCGGACCTAGGAGGAGGAGGCGGGAAAGGG

                                                                                                         500
                .         .         .         .         .         .         .         .         .         .
HCAII  TTGTAACGGAAAATTCTAGTTGTTGATCGCAGAGAAATTAAGAGACTCCCCTTCCCCCCTTCCCCCACCTTCCACCCCCACCCCACCCCTCCAGCTTCAG

                                                                                                         600
                .         .         .         .         .         .         .         .         .         .
HCAII  CACCACCTGTGGACTAAGGCGCTCAGCACGAACTGTCCCGGGGCATTTTCCAGTGCTGCTTTGAATCCATGGCTCTGATTTCCGAGTTTCCCCTTCATCT

                                                                                                         700
                .         .         .         .         .         .         .         .         .         .
HCAII  CTCGACTTCTAATGTTAGGGGGTCGGACATCAGGAATCGGCTTCTTGCCAGATCTGGTTCGGAGCCAGCGGAGCGAGGAGCATGCGTCTGGCGCACCTAG

                                                                                                         800
                .         .         .         .         .         .         .         .         .         .
HCAII  CGCTCTTTGGAGGGTGTGGGGCTTCCCAGGTAGTGGGGAACCCTGACGGTTAAAGGTGGGGTGGGCCCGGGCCGGGCAGTGAGGAAAGGATCCAGACCTC
                                                           *  ****         *  *   *   **    ** *    ***
MCAII                                                GCACCTAGTGTGGCAGCACGAGGTTCTCAGACCTGTTATAATGCCTTAGAATCT

                   .         .          .         .         .         .         .         .         .
HCAII  ---CTTGAATGTCTTAAGTGAGCTTGCATATC-CCAAAATCGCAACCACAAGCCCTGACATTAGTGTCTGCCCGATTTCAGTTGCTGAATTTCAGTAAAA
           *  **  **       **   ** ***  **  ***** ******      *   * **** **** *   * * ***  * ****** *** ***
MCAII  ATGTTCCAAACTCAGTCACAAGAGGGCTTATATCCCGAATCG-AACCACTCTGTTTTCTAGTAGT-TCTGTCTAGTATTAGTCACCGAATTTGAGTGAAA
```

```
         900
          .         .         .         .         .         .         .         .         .         .
HCAII   CGACCTTAAAATAGCTAATATTTATATAGCACTCAGTGATCTAAGAGCTTTACATATATCGATTCGAATTCTTACAGCGACATCTATGAGGTAGATTT--
         ** ** ***** *****    ** * *  ** *        ***** *  * ***** ** ** **** *       * * * *********** **
MCAII   AGATCTCAAAAT-GCTAACGCTTGTCTGACAAT------GTTAAGAAATAGATGTATATTGACTCTAATTTTCGTGTTGGCCT-TATGAGGTAGACTTGG

             1000
              .         .         .         .         .         .         .         .         .         .
HCAII   ---CTAATTATCCCATATTACAAATGTGGAAACTGAGGCACAGATTACGTGTTTTCCCAAAATTTAGCCCATTGTTAAGTGATGCTTCTAAAATTGGAAC
            * **** ***** ********      *********** *    ******** *                *  * ***   * **  **  * ****
MCAII   AATATTATTACCCCATGTTACAAATAAAAGGACTGAGGCACATTT----TGTTTTCCTA-------AATGTTAAGTTAGTATCACCTCCGAATGTTGAAC

             1100
              .         .         .         .         .         .         .         .         .    lyProGl
HCAII   TGAGCAGATTGGCTCCGGAATGATTGCTCTTCTCTAGGGGTCTGGGTGTACCTTTCCCCACAATGGGGATTCACATGTCTTCTTTCCCCCAGGACCTGA
        ******** ** **    *          ****    * ** * * ** * **** ** **   *  *        **  * ******     * * ****** **
MCAII   TGAGCAGACTGTCTAGAAACACCAGTGGCTTCC--GGTGGCCGGTGTCTCCCTTCCCACAGCAATGACCTAACAGCTCTCTTCTCATCTCTAGGACCAGA
                                                                                                    lyProGl

                                                    exon 2
             1200
        uHisTrpHisLysAspPheProIleAlaLysGlyGluArgGlnSerProValAspIleAspThrHisThrAlaLysTyrAspProSerLeuLysProLeu
HCAII   GCACTGGCATAAGGACTTCCCCATTGCCAAGGGAGAGCGCCAGTCCCCTGTTGACATCGACACTCATACAGCCAAGTATGACCCTTCCCTGAAGCCCCTG
        * ******* ******************* ***** ** *********** ***** *****     ** *** ** ******** ****  **** ***
MCAII   GAACTGGCACAAGGACTTCCCCATTGCCAATGGAGACCGGCAGTCCCCTGTGGACATTGACACAGCAACTGCCCAGCATGACCCTGCCCTACAGCCTCTG
        uAsnTrpHisLysAspPheProIleAlaAsnGlyAspArgGlnSerProValAspIleAspThrAlaThrAlaHisHisAspProAlaLeuGlnProLeu

             1300
        SerValSerTyrAspGlnAlaThrSerLeuArgIleLeuAsnAsnGlyHisAlaPheAsnValGluPheAspAspSerGlnAspLysAla         .
HCAII   TCTGTTTCCTATGATCAAGCAACTTCCCTGAGGATCCTCAACAATGGTCATGCTTTCAACGTGGAGTTTGATGACTCTCAGGACAAAGCAGGTCAGTGTT
           * ** ****** ****   * ***   *** **   ******* ** **   * ** ***** *********************** ************
MCAII   CTCATATCTTATGATAAAGCTGCGTCCAAGAGCATTGTCAACAACGGCCACTCCTTTAACGTTGAGTTTGATGACTCTCAGGACAATGCAGGTCAGTGTT
        LeuIleSerTyrAspLysAlaAlaSerLysSerIleValAsnAsnGlyHisSerPheAsnValGluPheAspAspSerGlnAspAsnAla

             1400
                                                    intron 2
              .         .         .         .         .         .         .         .         .
HCAII   TAGAAAATAA---CTTGTGTCTTTTAGCCAGTAGCTGTTTTCCGAGCTTAATGGAAGGAGCTAGGAACAGTGGCAAGGAACCCTCTTAATAATACAGTTT
        *** ***  *    * * *  ********* *** * ***** *  **********    *  ***      *    * ****       *  *** ***
MCAII   TAGGAAAACATCCATCGGGCTTTTTAGCCAATAGTTATTTTC-GGTCTTAATGGAAAATACAGGGACACCAGCAGGAGGTCCCT---GCCACCACAATTT

                1500
        .         .
HCAII   GTCTCAGGACTCAAGGATG
        ** ** * *  *  *
MCAII   GTGTCCGAAAGCTGGAGAT
```

Fig. 2. Comparison of the nucleotide sequences of the 5′ regions of the human and mouse (YBR strain) carbonic anhydrase II genes. The proposed TATA and CCAAT sequences are boxed. The CACC tandem-repeat sequences are indicated by solid lines, and the CCGCCC/GGGCGG sequences by dashed lines. Regions which have been previously published are: mouse exons 1 and 2 and the 5′ non-coding region extending 150 bp upstream from the ATG start codon [15] and a small portion of human exon 2 [25].

found in the 5′ flanking regions of the human and mouse CA II genes (see Fig. 2). These sequences were first found in the 21 base pair repeats near the origin of replication of simian virus 40 (SV40) and have been shown to be important for the transcriptional efficiency of the early and late genes [32]. They have since been reported in a number of genes, e.g., c-*myc* [33], α-globin [34], interleukin 3 [35], hydroxymethylglutaryl-CoA reductase [36], and may represent elements controlling transcriptional efficiency. These CCGCCC/GGGCGG sequences have been shown to bind to a transcription factor, Spl, which potentiates the transcription of some genes [37,38]. Interestingly, nine of these sequences are found 5′ to each of the two human α-globin genes, α1 and α2 [34], but none are found 5′ to the β-globin genes [39].

In both human and mouse, tandemly repeated elements ($\mathrm{CC\overset{GAGT}{C}GTCACCTC\overset{G}{C}TC}$ in human and $\mathrm{\overset{T}{C}C\overset{C}{T}\overset{C}{G}TCACCT\overset{TAT}{CCG}C}$ in mouse) have been identified 15 and 22 bp upstream, respectively, from the putative CCAAT boxes (Fig. 2 and Table II). These distal elements appear to be structurally similar to repeated sequences (consensus = RgCCtCACC-CTGg) found in similar positions in five mammalian adult β-globin genes (see Table II); this element has been shown to be important for transcriptional efficiency of the rabbit β-globin gene in vitro [3]. In addition, two mutations (CACC → CAC<u>G</u> and CA<u>T</u>C) which produce $\beta^+$- and $\beta^0$-thalassemia, respectively [40,41] have been described in the human $\beta$-globin gene in the first distal CACC sequence (position −60 in Table II). The mouse interleukin 3 gene has 12 repeats of a similar sequence in its first intervening sequence [35]. In addition, the CA II repeat elements are also similar to an element found in a number of rat pancreatic genes (Table II). It has been suggested that these elements may confer pancreatic tissue specificity [42]. However, the CA II gene is expressed only in exocrine ductal cells in the pancreas [43], whereas pancreatic specific genes (e.g., trypsin, chymotrypsin, elastase) are only expressed in acinar exocrine cells [44].

Carbonic anhydrase activity, presumably due to CA II, has been reported to be increased in the endometrium of rabbit [45] and human [46] in response to progresterone. Since glucocorticoid receptor binding elements have been found at least 1 kb upstream from the transcription initiation site of the prolactin gene [9], we searched for such possible receptor binding elements. Although none were found, it is still possible that such regions are present further upstream.

Because of the increased homology between the human and mouse CA II genes near the TATA box (Fig. 2), it appears that some of the *cis*-acting regulatory elements of the CA II genes probably lie within 150 bp of the first exon. If this is so, then a number of *trans*-acting factors, some of which in turn may be controlled in a tissue-specific manner, will bind to sequences within this small region.

The functional significance of these conserved elements still remain to be elucidated. In order to determine which of these regions are important for the regulation of the CA II genes, we have constructed several mouse and human CA II 'minigenes', and are presently performing experiments on the in vitro expression of the CA II gene constructs.

## Acknowledgements

We thank Mrs. Ya-Shiou L. Yu for her excellent research assistance. This work was supported by Grant GM 24681 from the U.S. Public Health Service.

## References

1 Mathis, D.J. and Chambon, P. (1981) Nature 290, 309–315

2 Grosveld, G.C., Rosenthal, A. and Flavell, R.A. (1982) Nucleic Acids Res. 10, 4951–4971

3 Dierks, P., Van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weissman, C. (1983) Cell 32, 695–706

4 Walker, M.D., Edlund, T., Boulet, A.M. and Rutter, W.J. (1983) Nature 306, 557–561

5 Scott, R.W., Vogt, T.F., Croke, M.E. and Tilghman, S.M. (1984) Nature 310, 562–567

6 Ott, M.-O., Sperling, L., Herbomel, P., Yaniv, M. and Weiss, M.C. (1984) EMBO J. 3, 2505–2510

7 Ornitz, D.M., Palmiter, R.D., Hammer, R.E., Brinster, R.L., Swift, G.H. and MacDonald, R.J. (1985) Nature 313, 600–602

8 Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E. and Constantini, F. (1985) Nature 314, 377–380

9 Mauer, R.A. (1985) DNA 4, 1–9

10 Tashian, R.E., Hewett-Emmett, D. and Goodman, M. (1983) in Isozymes: Current Topics in Biological and Medical Research (Siciliano, M.J., Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S., eds.), Vol. 7, pp. 79–100, A.R. Liss, New York

11 Dodgson, S.J., Forster, R.E., Storey, B.T. and Mela, L. (1980) Proc. Natl. Acad. Sci. USA 77, 5562–5566

12 Vincent, S.H. and Silverman, D.N. (1982) J. Biol. Chem. 257, 6850–6855

13 McKinley, D.N. and Whitney, P.L. (1976) Biochim. Biophys. Acta 445, 780–790

14 Whitney, P.L. and Briggle, T.V. (1982) J. Biol. Chem. 257, 1256–1259

15 Venta, P.J., Montgomery, J.C., Hewett-Emmett, D., Wiebauer, K. and Tashian, R.E. (1985) J. Biol. Chem. 260, 12130–12135

16 Curtis, P.J., Withers, E., Demuth, D., Watt, R., Venta, P.J. and Tashian, R.E. (1983) Gene 25, 325–332

17 Konialis, C.P., Barlow, J.H. and Butterworth, P.H. (1985) Proc. Natl. Acad. Sci. USA 82, 663–667

18 Tashian, R.E., Hewett-Emmett, D., Dodgson, S.J., Forster, R.E. and Sly, W.S. (1984) Ann. N.Y. Acad. Sci. 429, 262–275

19 Sultan, C., Goualt-Heilmann, M. and Imbert, M. (1985) Manual of Hematology, John Wiley and Sons, New York

20 Niels, B., Schwartz, J.H. and Tauber, A.I. (1984) J. Clin. Invest. 74, 455–459

21 Väänänen, H.K. (1984) Histochemistry 81, 485–487

22 Kumpulainen, T. (1984) Ann. N.Y. Acad. Sci. 429, 359–368

23 Ham, A.W. and Cormack, D.H. (1979) Histology, 8th Edn, Rippincott Co., Philadelphia

24 Stein, R., Sciaky-Gallini, N., Razin, A. and Cedar, H. (1983) Proc. Natl. Acad. Sci. USA 80, 2422–2426

25 Venta, P.J., Montgomery, J.C., Wiebauer, K., Hewett-Emmett, D. and Tashian, R.E. (1984) Ann. N.Y. Acad. Sci. 429, 309–323

26 Lee, B.L., Venta, P.J. and Tashian, R.E. (1985) Human Genet. 69, 337–339

27 Pustell, J. and Kafatos, F.C. (1982) Nucleic Acids Res. 10, 51–59

28 Henderson, L.E., Henriksson, D. and Nyman, P.O. (1976) J. Biol. Chem. 251, 5457–5463

29 Bird, A., Taggarrt, M., Frommer, M., Miller, O.J. and Macleod, D. (1985) Cell 40, 91–99

30 Wolf, S.F. and Migeon, B.R. (1985) Nature 314, 467–469

31 Razin, A. and Riggs, A.D. (1980) Science 210, 604–610

32 Everett, R.D., Batey, D. and Chambon, P. (1983) Nucleic Acids Res. 11, 2447–2464

33 Linial, M. and Groudine, M. (1985) Proc. Natl. Acad. Sci. USA 82, 53–57

34 Michelson, A.M. and Orkin, S.J. (1983) J. Biol. Chem. 258, 15245–15354

35 Miyatake, S., Yokota, T., Lee, F. and Arai, K.-I. (1985) Proc. Natl. Acad. Sci. USA 82, 316–320

36 Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) Cell 38, 275–285

37 Dynan, W.S. and Tjian, R. (1983) Cell 35, 79–87

38 Gidoni, D., Dynan, W.S. and Tjian, R. (1984) Nature 312, 409–413

39 Collins, F.S. and Weissman, S.M. (1984) Prog. Nucleic Acids Res. Mol. Biol. 31, 315–458

40 Orkin, S.H., Kazazian, H.H., Jr., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G. and Giardina, P.J.V. (1982) Nature 296, 627–631

41 Orkin, S.H., Antonarakis, S.e. and Kazazian, H.H., Jr. (1984) J. Biol. Chem. 259, 8679–8681

42 Swift, G.H., Craik, C.S., Stary, S.J., Quinto, C., Lahaie, R.S., Rutter, W.J. and MacDonald, R.J. (1984) J. Biol. Chem. 259, 14271–14278

43 Kumpulainen, T. and Jalovaara, P. (1981) Gastroenterology 80, 796–799

44 Howat, H.T. and Sarles, H. (eds.) (1979) The Exocrine Pancreas. W.B. Saunders, London

45 Hodgen, G.D. and Falk, R.J. (1971) Endocrinology 89, 859–864

46 Nicholls, R.A. and Board, J.A. (1967) Am. J. Obstet. Gynecol. 99, 829–832