Vol. 197, No. 1, 1993
November 30, 1993

CLONING AND CHARACTERIZATION OF THE MOUSE CLARA CELL SPECIFIC 10 KDA PROTEIN GENE: COMPARISON OF THE 5'-FLANKING REGION WITH THE HUMAN RAT AND RABBIT GENE

Manas K. Ray, Susan Magdaleno, Bert W. O'Malley and Francesco J. DeMayo[1]

Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030

Received October 1, 1993

The mouse Clara Cell 10 kiloDalton (kDa) protein (mCC10) cDNA was used to isolate a recombinant phage containing the mCC10 gene sequence in a 14 kilobase (kb) insert from a mouse genomic library. A total of 7.7 kb of this clone was sequenced. The sequenced region included: 3.3 kb of 5'-flanking region, 4.2 kb intragenic sequence and 0.2 kb of DNA flanking the 3' end of the gene. Computer assisted sequence analysis identified potential *cis* acting response elements for the glucocorticoid receptor, hepatocyte nuclear factor (HNF3) and octamer (Oct1) binding protein. The presence of B1 murine repetitive sequence also has been identified in a similar position reported in rat CC10 5'-flanking sequence. As with the rat CC10, the mCC10 5'-flanking region also contains deletions of a 2.1 kb and a 0.3 kb sequence present in the rabbit uteroglobin gene, these regions are reported to contain a cluster of glucocorticoid/progesterone receptor binding sites and estrogen receptor binding sites, respectively.

The cell specific 10 kDa protein (CC10) is the major secretory product of the Clara cell, a nonciliated secretory cell involved in the lining of bronchioles of the lung (1). CC10 has been studied in a wide variety of species including human, rabbit, rat and dog (2-4). Depending on the species in which the protein has been studied, it has been referred to by various names. In the rat, CC10 has been referred to as the polychlorobiphenyl binding protein and in the rabbit, the protein is known as uteroglobin (UG). While the major site of expression of CC10 in most species is the lung (5-8), UG is mainly expressed in uterus and to a lesser extent in the rabbit lung (9). CC10 has been reported to bind to polychlorinated biphenyl derivatives (10) but the physiological importance of this interaction is not known. An antiinflammatory role of CC10 also has been

[1]To whom correspondence should be addressed. Fax: (713) 790-1275.

Abbreviations: CC10, Clara cell 10 kDa protein; UG, uteroglobin; kb, kilobases; kDa, kildaltons; bp, basepairs; GRE, glucocorticoid response element; PRE, progesterone response element; ERE, estrogen response element; Oct 1, octamer binding protein; HNF 3, hepatocyte nuclear factor.

0006-291X/93 $4.00

proposed based on the inhibitory effect of UG on phospholipase A_2, but its physiological relevence to the lung has yet to be determined (11).

Human, mouse and rat CC10 are highly homologous to rabbit UG and thought to be derived from the same ancestral gene (12). cDNA sequences for human, rabbit, rat and mouse have been reported (3, 4, 13, 14). The deduced amino acid sequence of mouse CC10 shows 90%, 52% and 51% amino acid homology to rat and human CC10 and rabbit UG. Isolation of genomic clones has been reported for rabbit, human and rat (3, 15, 16). Like rabbit UG, rat CC10 gene also consists of three exons and two introns. The 5'-flanking region of rat CC10 gene shows sequence homology with rabbit UG 5'-flanking sequence (3), however, two major deletions have been found in rat CC10 5'-flanking region of approximately 2.1 kb and 0.3 kb in size. The larger fragment has been reported to contain a cluster of glucocorticoid/progesterone receptor binding sites (2, 17). Alignment of human and rabbit 5'-flanking sequences identified the progesterone response element (PRE) to be present between -1.77 to -2.51 kb upstream of transcription start site (18). The presence of a potential GRE/PRE in rat 5' regulatory region also has been reported but it lacks the estrogen responsive element (ERE) present in the proximal promoter region of rabbit UG (3). *Cis* acting response elements important for cell specific pulmonary expression of the rat CC10 gene have been reported, also (19).

Here we report the cloning of the mouse CC10 gene. We have sequenced 7.7 kb of the mCC10 gene that includes 3.3 kb of the 5'-flanking sequence. The 5'-flanking sequence of mCC10 gene was compared with rabbit UG, rat and human CC10 gene. The sequence was searched for potential transcription factor binding sites. mCC10 5'-flanking region is highly homologous to rat CC10 and partially homologous to human CC10 and rabbit UG 5'-regulatory region. Computer assisted analysis for potential transcription factor binding sites identified sites for the members of hepatocyte nuclear factor (HNF3) and octamer (Oct1) as well as binding sites for the glucocorticoid/progesterone receptors. The mCC10 gene was cloned because the mouse offers potential for genetic manipulation.

Materials and Methods

Gene isolation and mapping :

A genomic library prepared in λ Dash II (Stratagene, La Jolla, CA) using genomic DNA isolated from the 129/SvEv strain of mouse was screened with Pst I fragment of mCC10 cDNA which does not contain the first fortyfive nucleotides of the coding sequence (4). The average size of the insert in this library was 15 kb. *E. coli* NM538 cells were infected and the plaques were screened using the radiolabelled cDNA probe.

S1 nuclease mapping :

S1 nuclease mapping was carried out following the procedure described by Weaver and Weissman (20) with the modifications of the Berk and Sharp procedure (21). A 271 nucleotides Pst I - Sac I fragment from the phage clone spanning the presumed transcription start site was isolated in a 1.5% low melt agarose gel and end-labelled with T4 polynucleotide kinase following the protocol described by Sambrook *et. al.* (22). Total RNA isolated from mouse lung was hybridized with 5' [^{32}P] labeled 271 nucleotide Pst I - Sac I fragment (40,000 CPM) in 40mM Pipes, pH 6.5, 0.4M NaCl, 1mM EDTA, 80% formamide (EM Science, Gibbstown, NJ) at 40°C

temperature for overnight. After 10-fold dilution with 0.25M NaCl, 0.03 M sodium acetate, pH 4.6, 0.001M $ZnSO_4$, denatured salmon sperm DNA, 20 µg/ml (Sigma Chem. Co. MO), S1 nuclease digestion was carried out under the conditions described in the figure legend.

DNA sequence analysis:

Nucleotide sequence was analyzed by the dideoxy method using USB-Sequenase Version 2.0, DNA Sequencing Kit (United States Biochemical Corporation, Cleveland, OH) (23). We used both double and single stranded DNA for sequencing. Sequencing of single stranded DNA was employed to overcome highly repetitive regions. Single stranded DNA was made using the modified single stranded rescue protocol from Stratagene using VCS M13 helper phage for XL-1 Blue cells (24) .

Results and Discussion

Mapping of the Gene for mCC10 Protein:

Our initial objective was to isolate the gene for the mCC10 protein and to characterize the regulatory sequences in the 5'-flanking region. The mCC10 cDNA was used as a probe to screen the mCC10 gene from the mouse genomic library. Three recombinant phage clones containing the mouse genomic sequence were isolated and subjected to restriction enzyme digestion and Southern blot analysis. Two clones were completely overlapping, one is approximately 2 kb shorter in length at the 3'-end and we ended up with two overlapping clones. The 5'-end of the clones were identified using Pst I fragment of mCC10 cDNA containing the first 45 nucleotides of the coding sequences and the 5'-untranslated region of the mCC10 mRNA. Figure 1 shows the restriction endonuclease digestion map of the two clones which contained non-overlapping regions. Oligonucleotides containing the coding sequences of mCC10 were used to identify the fragments containing exons of the gene. We restricted our entire study to the clone having approximately 6 kb 5'-flanking region.

mCC10 Gene Structure:

The sequence for mCC10 is shown in Figure 2. All putative restriction sites in mCC10 gene were verified by the sequence analysis. Approximately 7.7 kb of mCC10 was sequenced,

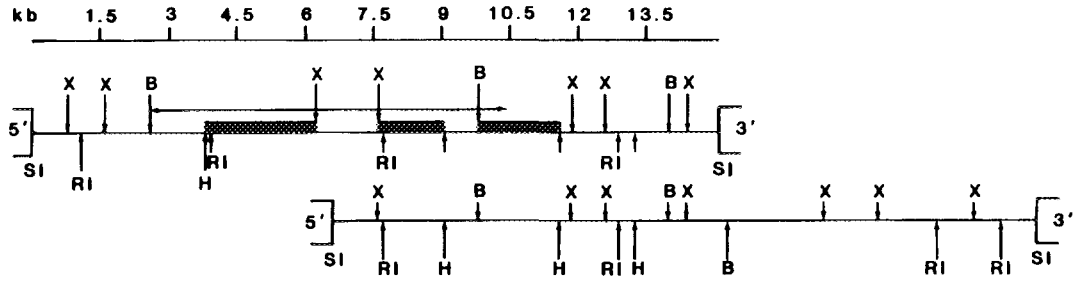

Figure 1. Restriction map of recombinant phage containing the mCC10 gene in an approximately 14 kb insert of Sal I restriction enzyme. Abbreviations for restriction enzymes are: S, Sal I; X, Xba I; R, Eco RI; B, Bam HI; H, Hind III. Black boxes indicate the restriction fragments that hybridized with mCC10 cDNA. The region which is sequenced is indicated by a double headed arrow.

covering the mCC10 gene with 3,274 bp of 5'-flanking sequence and 200 bp of 3'-flanking DNA. The structure of the mCC10 was similar to that observed in other species. The first exon was located between 61 to 114, first intron between 115 to 2549, second exon between 2550 to 2737,

```
-3274GGATCCCAGG CAAGCACACG GGACAAACAT TATGTAAAAG CAGAGCACCC ACACACATAA
-3214 ATTTTTTTTA AAAATTTAAA TTATAAAAAG AAAAGTTGGA AGGATCTAGG CATCATCCCT
-3154 GCTAAGCAAA ACAACCAAGT GCTGAGAAGG CGCCAGGCCT TCCTAGGAGG GCTCTAGGTC
-3094 AATGCCGGAG TCACCTTGGG TTCCTAGGAG GGCTCTAGGT CAATGCAGGA GTCAGCCTTG
-3034 GGGCAGAAGA AAGTGAACCA AACTCATGAC GGACAATCTT CTGTTTGGCA GCTATTTCTT
-2974 GAGAGCATGC TCAGAGCTGG CCACTGGGGG AGGACACAGA GACTGGCCTC TACGTTTGCC
-2914 CTCACAGTTG ACCTTGTCAG CTGGCCCCAA CCTGTCGTCT CACCAGTCCT TCTGTGGCCC
-2854 AGCAAGATCA CTAGGCTCAG AATGACCCAC CCAGAGCCTC CAACCTGCTT TCTGGGCCGT
-2794 GCTGGGACTT GTTTACCAGC TTCTGCCTTT TGGGATAGGC TAAGCACTGA AGCCAGGGGA
-2734 TAAACATCTG GTCTTAGGCA ATAGCTCCTG GGGTACCTGG TGGAGGCCAC TAGAGGCTTT
-2674 CACCCCTCCC TCCTGTCATT GCCATCTACT ATTCATGAGA GAACAAGGAA AGGGCTCACT
-2614 AGGAGACAAT GAACATGGCC CAACAACCTG GCACCCACAT GCTAATCACG TTATTCACAC
-2554 ACACACAGGC ACACACACGC ACACACACGC ACACACACAC TGTGTGGTGT TGGGGAGGGT
-2494 CTGCTGACCT AACGACTTAG GAATCTAAAT ACCAGGTGAA TGTCAGGCAG TCAATTCCCC
-2434 TCAGGCTGTG AGTGAGACTC TTCCTGACCT TGTCTGAAAC CAAGCCTGAA GTTATTTATA
-2374 TCACTTCAGA AGACTCCTAG CCCAGACCTC AAAGTCTCCA TCAACCTAGG AATCAGGACT
-2314 CTTCAAAACC AGTCACAGCA GATTGCAGGG AGTAAGGAGG GTCATGGGAC TAGTACGTGA
-2254 TGTGCCAATT CAATACTCGT GACACCATAG TGCTAGCAAT CAAGTGGCTT CCTTTCCCA
-2194 CTTGGAGTCT GTCTTTACTA TATACCTGCC CCAGTGCAGC CAGAGAGAGC AGTCCATAGC
-2134 CCCCCCGCAG GTGCTAAGCA AGCTTAGGGG TCTATCCTGG ACGTCAAAGT GTAGGGTGGG
-2074 GGGGTGCTTT AGGGAAGACT GCTTAGGGAA GAATTCAAGG CACATCTAGC TTTTTTTTT
-2014 TTTTTTTTTT TTTTTTTTT ATGGAATGCT TGAAATTCAG TACCTCATAC TTATCCAATA
-1954 GGCACGGTAT ACCCACTTGA AGCCACATCC CCAGATTCAT TCATTCATTC ATTCATTCAT
-1894 TCATTCATTC ATTTAGGTTT GGCTTGGTTT GGTTTTAGTT TTGTTTGTTT GTTTGTTTTG
-1834 TTTTGTTTTA AAGACAGGGT TTCTCTCTGT AGCCCTGGCT ATCCTAAAAC TTTCTTCCAT
-1774 CTTTCCTTCC TTCCTTCCTT CCTTCCTTCC TTCCTTCCTT CCTTCCTTTA AGTTTGGTTT
-1714 GGGGTTTTGG TGGTTTTTTT TTCATGGCAG GGTTTCACTG TGTAGACCTG GCTGCCCTGG
-1654 AACTCATTCT ATAGCCAGGT TGGCTTGAAC TCAGAAATCT GCCTGCCTCT GCCTCCTGAG
-1594 TGCTAGAATT AAAGGCGTGT GCCACCACAG CTAGGCACTA CATTCATTTT AAAGGCTCTT
-1534 GTCTCTTTAA AATGACACTC ATACTAGTTC TGTACACAGC CACTGGGCGT TAATCACTAA
-1474 GTGGCCAGAC ACAGGTCCAG GAAGGTTTCA CTATTTAAAA GTTGGTTTGG AAATGGGAAA
-1414 ATATGTAAGG ACAATCTCAC TCTAGGACTC AAAGACAGAG ATAAAATGTT CCTTACAAAC
-1354 CTATTCTGGG GGAACAACAG CATCAGAACT CTTTCTATAT CAGCCTCTGC ACTAGGGTCA
-1294 GGCCAGAGAC ACAGAAATGA GCTAAGCCTA CAATGAGAAC GGAGGAACCA TGGGAATGTG
-1234 CATCCTTAGG GCCCAGGGCA CTAGGAAAAG GCAACCCAAA TGCCAGGACT TTAAATCCCC
-1174 TTCCTGTTGT GTCATTAGAT GCAAACAGCC CTGGAAAGGA TGTGACCATA AGCAAGGCAA
-1114 GGTTCTGTAG CGGGGGCAGA TCCTGAAGTT ACTAGAAGCT GAACGTTCTA GGCTGACCAT
-1054 GCTTCTCACA GTCACTAGAT AGGATGCTCA TGAAATGAAC TCTTGAAATG CCTCTCCAGG
-994  TCTTCCCCAA TTTCCAGTCC CACCAGCACC ATAGTAGGGA CTGGGCATCT ATTATCTGGG
-934  TGGATGAACA TTTGAGACAA CCTGGAAGTT TGAATAGGAT TTGTGGAAAT GGAGAAGGTG
-874  TTGGTCCTAA GACACAAGAG CACTGAAGAA GTAAGAGTTA ACACAAACAT GGCCAAGCTA
-814  AGGAGACTAA GGTAAGGCCT GGGAATGGCT AACTACTTGC AGAACTGTCA ACATCGTGAA
-754  AGAATAACGA ACAGACATGA CCAAGGAAAG CAAAAGCAGG AAAGAGCTAA GCGTGGGAGA
-694  GTCCTGGAGA GAATGGAGCA AGAAGGGGGA GGTTATAGGG TAAAGGCCAG GGAGAGGTCA
-634  GGTCAGGTCA GATGAAGACT GATGGTGCTT TCAATTGGCA GGTACTCAAG GGCTGCTCTG
-574  CGTAGGAACA GCCTAAGCCT GCCTGATCTA GGCCCTGGGT CTCTGATGTG TACTATGGAG
-514  AAGTCTTTCT ATGTTCACGT CTACTGTATG TAGGATCGAG CCTGTCTAAC AATGCCCAAG
-454  AATCGAGTGA CCTTGTGGCT TGAAGTCTAG CCACGTTCGT TGGAGGGAGG CAATAGAAGG
-394  AGTCTAGTGA CATCTCAGAG TCCTGATGTC TTTGTCCTTC CCTGTGATTC CTGAAGGGTC
-334  TCCGGCCTCT GGTTCTCCAG GGTTGGCAAG TCTACAGTTC GTTCCTGGAA CCTGGAGTGC
-274  TCAGTGCTTG ACTTCCAAGA GAGGACACAG TTGTCTTCTA CAGTTCCACG ACCTCTGACT
-214  TGGGTCCTCC ACTGCCTGAA TACTCACAAG TGGCCTATTG TGTGAGTGAG CTCAGTTTCA
-154  ATGGGAACAG AAACTGGGTT TATGAAAAGA GATTATTTGC TTATTCCACG GAGAAGATGA
-94   CAAGTAAATA ATGCAATCTC CTAAGTGGAG CGCAATCACT GCCCTCTACC TCTTGTGGGC
-34   TGCAAGGAAC ATATAAAAAG CCACACACCC ACACATACCC ACACATTACA ACATCACCCC
27   ACATCTACAG ACACCAAAGC CTCCAACCTC TACCATGAAG ATCGCCATCA CAATCACTGT
87   GGTCATGCTG TCCATCTGCT GCAGCTCAGG TGAGTGCCCA GGGTCTCTTT CCCAAGGTCA
147  GGAGGAGGAA GTGTCCAGTC CCTCTGTGGC TGCCTGGAAC GGCTGAATAT GCTGACTCTC
207  AGAAGCATTG CGTCGTATAA TCGACGTTAT TTTGTGATCT TGTGGAAAGC CTTGAACTTC
267  TGCACTCCAG AGACTAGGGA AGCCTTGTGG GGACGCTGGA GAGAGCCCAT CTAGACCAGC
327  ACATCTGTCC AAAACTGGAC TAGGAATAAG GAGCCCTTTG AAGCCTGGAC TAGGACTGGA
387  TTAAGATGCT AAGAGATTTT GATACAGTCA CCAGGTTGAG GGGTGGTGGA TTCCCACCCA
447  ACTCTGACAT TTAGTCTGGC AGAAGAAGTA CAGAGACTCA TGTCCTCAAA CCCAAGGCAG
507  GGTCTCTTGT TTTGGTAGGA GAGAAATATT CCAGAAAAAG ATCTTCCCTT TGGAAAACTG
567  CCAACTTTGG GTGTAGTTTT CTAAACATTC TGTCCTCTGG AAAAAGAATT AAAGGCTCTG
627  GATAGTTGTT GCTATCACTG TGCATAAAGG AATGGGAATG AGAGAGAGGG AGGGAAGGAT
687  GGGTGGACAG AGAGAGGGAG GGAAGGATGG ATGGACAGAG AGAGGGAGGG AAGGATGGAT
```

Figure 2. DNA sequence of mCC10 gene with 3.3 kb 5'-flanking region: First, second and third exons are underlined. The transcription start site is marked by an arrow. CAAT and TATA boxes are in bold print.

```
747   GGACAGAGAG AGGGAAGGAA AGATGGATGG ACAGAGAGAG GAGGGAGGAT AGATGGACAG
807   AGAGAGGGAG GGTTAAGGAT GGATGGACAG AGAGAGGGAG GGAAGGATAG ATGGACAGAG
867   AAAGGAAGGG GAGGATGGAT GGACAGAGAG AGGGAGGGAA GGAGGCATAG AGGGAGGATG
927   GCTCAATCTA TCACCGTCCG GGCATCTTGT GCTTTGAGAA CAGAGTTGGG GAAAACAATT
987   CTCTCATTAG CTTCTCCCAC TCCCTCCTGA TAGTGCTGTA TTATTTTCTA AAAACAAACC
1047  CTAAATAAGT ACTTGAAAAG GTTCATAATA CCTGTGAATA CAGGTTTTAT AAACCAAGGA
1107  GACTGAGCCC CCAACCAAGA GGCAGTGGAA CAGCTGTTTA ATCAAGCATT TTTTTTTCAT
1167  TTGTTCTCAG GGGGAAATAT CATTTTGAAT TAAAATGTTT ATTAATTAGA TAGCAATACT
1227  GTGCATAGAG GAGCCTAGCA AAACAGAGCT AAAACATGTG AAGTGGTGGA CTGGTAAGAT
1287  ATTTGCATAT CGGTGTGCTC TTGGATTTAA AAACTGTGAT AGTAAAACTC GTTAATGCCG
1347  CTTACAATGT GGTAGTCACT GTCCTCAGCC ATTTACATAC TATCCCTTCA TTTTCGAAAC
1407  AATCCAACGA AGTAGTGATA TTGTTCCCAT TTTGAAGTTT AGGAAACTGC CGGGTGATAT
1467  TATAGAAAGG GTTCAAGTTG GAGTGACCAT ACTTTTGGCT TTGAGGCAGG GTCTCATTCT
1527  GGTCTGAACT CATAACTATC CTGCCTCAGC CTCCCCGGTG CTGTGAGGTA TATAACAGGT
1587  GTGAGCTATC ATACCCAGCT AGAGTGGCCA TTAGCTGGAG TGGCCATTTG GAGTAGGAAT
1647  AGGGTCTAGA GTCAGAGCAG TGAATCCAAA ATTGCAAACA ATGCAGAGAA TTCCAAGGCA
1707  GAACATTTGA GACAAGGTGA GAAGTGGAGG CAAACAGTAG GAAATCATGT TCCCAAGCAG
1767  ACAGTACAAC CTAAATATCA CAGTCCTTGC AGGGAACATG GTGAGGTGGG AGGCTTGCCT
1827  GCAAACTGAC TGTGCCATCA GACTAGATCC CTTGTGTGTC TTCATCTTCC AGGCCAGGTT
1887  TGGGGATCAT GCTGGGCTTG GGTAAGGACT GTGTCACATG TTCCTAGTAA CGTGGCACAA
1947  ACCACTACTG GCCATGGGTG AGGGTATTCT GGTAGAAAGT AAGACTCCCA GAGGCCTCCT
2007  TCTTGTCCCT ATACGTCATA AAGCTGCCAG ACTGACAGCT TCAGTTACAC ACTCTGCTAG
2067  AATGATGGCA GAAACAATGG CCAACCCAAG CCAAGAATCC AAAGGCAGCT CAGGAAAGCC
2127  ACAAAAGCTG TGACCTAGAA AGTCCTCCCA GGGCCCCTTG GTATTGCAAA CTTTATATGC
2187  CCCAGTACAG GGGAATGCCA GGGCCAAGAA GCGGGAGTGG GTGGGTAGGG GAACAGAGCA
2247  GACGGAGAGT ATAGGGAACT TTCAGGATAG CATTTGAAAT GCAAATAAAG AAAATATCTA
2307  ATAAAAATAA TTAATTAACT AATTAAAATA AATTTAAAAA AAAAAAGAAG AAAGAAAGTT
2376  CTCCCAGGGT GTTAGATTTT CAGCTGCTTG GAGCCTCTCC TTAACCCCCA GAGCTTACAC
2427  TCTTCTCTAC CACCCAAATA CTGCTCCTCT TGCTGACCAG ACTGGCTGGT CAGCAGGATG
2487  CTGGCTGACC AGCTTTTAGA AAAGAAGCTG CTATCCTTCT GCTTTTCTTT GCCTGCTACT
2547  CCAGCTTCTT CGGACATCTG CCCAGGATTT CTTCAAGTCC TTGAGGCCCT CCTCATGGAA
2607  TCAGAGTCTG GTTATGTGGC ATCCCTGAAG CCTTTCAACC CTGGCTCAGA CCTGTAAAAT
2667  GCGGGCACCC AGCTGAAGAG ACTGGTGGAT ACCCTCCCAC AAGAGACCAG GATAAACATC
2727  ATGAAGCTCA CAGGTATGCA ACATCTGTCA CTCACATTGA TTCTAAGTGA TTTCCCAAGC
2787  ACAGTGTCAT ACCATTTTGG GGGTTACAAA GGCAAATGAA TATCTTGGCC ATGTACCCTG
2847  CCTTTTTATT TGTTTCTTTT GGGTTTTTGT TTTTATTGTT TGGTGGTGGT GGTGGTATGT
2907  ATGAGTGTTA GATGGGGATT CTGTGAGCTA GTCTCACTGT CCCCAAACTC CTGGGGTCAA
2967  GTGTTCTCCC CACCTCAGCC TCCTGAGTGG CTGGTACTAC AAGCTTGAAC CGTGCACACC
3027  CAACTCCACT GATTTTTTTT TCCATAATAA TTTTGATATT TTATCTACTT TGTGTGAGTG
3087  TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGAGACAC ATGTGGAGGT CAGAGGACAA
3147  TCTTTAAGGA GTTGGATCTT ACCTTCCACC ATGTGGTCCT GTGGATCAAA TTTAGGTAGT
3207  CTTTGAGAAA ACCATCCAAG GCCTTCTTTT TCTTTTCTTT TCTTTTCTTT CTTTCTTTCT
3267  TTCTCTCTTT CTTTCTTTCT TTCTTTCTTT CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT
3327  CTCTCTCTCT CTCTTCCTCT CTTTCTTTCT TTCTTTCTCT CTTTTTATAG ATATTCTTTA
3387  TTACATTTCA AATGTTATCT CTTTTCTAGT TTCCCCTCCA AAAATCCCCT ATCCACTCCC
3447  CCCTGCTCCC CAACCCACCC ACTCCCATTC CTGGTCTGGC ATTCCCCTAT ACTAGGGCAT
3507  AGAGCCTTCA CAGGACCAAT GGCCTCTCCT CCCATTGACA ACGGACTAGG CCATCCTCTG
3567  CTACAAATAT AGCTAGAGCC ACAAGTTCCA CCATGTGTTT TCTTTGATTG GTGGTATAGT
3627  TCCAAGGAGC TCTGAGGGTA CTGGTTTGTT CATATTGATG TTCCTTCTAT GGGTCTTCAG
3687  TCCCTTCAGC TCCCTGGGTA TTTCTTTGGC TCCTTCACTG GGGACCTTGT GCTCTGTCCA
3747  ATGGATGACT GTGAGCATCC ACTTCTGTAT TTGCCGAGCA CTTGCAGAGC CTTGCAGGAG
3807  ACAGCTATAT CAGGCTCCTG TCAGCAGGCT CTTGTTGATA TCTGCCTAGT GTCTGGGTTT
3867  GGCGGTTGTT TATGGGGTGG ATCCCAAGTG GGGCAGTCTC TGGATGGTCA TTCCTTCAGG
3927  CTCTGCTCCT AACTTTGTCT TTGTAACTCC ATCCATGAGT ATTTTGTTCC CCATCCAAGG
3987  CCAGTCTAAC CCTCAATTCC TCTCAGAATT TACTGTGGCT TTTGTCTGAT AAGTCTCTGC
4047  ACCACTACTG CATCTACTTC TGAGGTTCAC TCACTATTGG GGGTGTGGGG TGTTTCTTTA
4107  TTCTAGGAGA AAATCCTAAC AAGTCCTCTG TGTAAGCAAG ATTTAAGATT CTGAAGCCTC
4167  ACTGGATTCA GAGATATTCT ACTGCTAAAG CCTTGTCACT GCCCTGTGTC TCCTCGGCTC
4227  CTTCTCCCGC AATAAACTGC GAGCATCTCA CCTGCCCGCG CGCAATGGTT TTTTCTGTAC
4287  CTGCTGGAGC CAGAGCATTG GAGCTGGAGT GAGTGAAGGG AACCAGAAGT AAGGGAGGAC
4347  AGCTGAGTCC TGCCCTGGTG CCACACAAGA TCCCACAAGA TCCCACAAGT CCTCCTTCTC
4407  TACCTTAGTT ATTTGCCATC AGTAGACTGC AATGCCCTAC CTAGTAATAG AATGTTAAGT
4467  CCACCAT
```

Figure 2. - Continued

second intron between 2738 to 4111, third exon between 4112 to 4157 with polyadenylation site between 4237 to 4242. This structure was verified by PCR analysis using oligonucleotide sequences derived from the mCC10 cDNA (data not shown). From this map, mCC10 shows identical intron-exon boundaries to that of rat CC10 and rabbit UG gene (3, 15).

To determine precisely the transcription start site of the mCC10 gene, S1 nuclease protection assay was carried out. A 271 nucleotide long Pst I-Sac I fragment, thought to be

covering the transcription start site was used as a probe. As shown in Fig. 3, a group of bands in the range of 60-62 nucleotides upstream of the translational start site were protected and migrated according to their electrophoretic mobility. We considered A as the primary starting nucleotide because of its strongest intensity among the protected bands in the sequencing ladder (25). Sequence analysis of the genomic DNA indicates the TATA box is 23 bp upstream of transcription start site (Fig. 2). In cases of rat, human and rabbit the homologous sequence is located between 27 and 34 nucleotides upstream of the transcription start site (3,15,16).

Computer Analysis:

The 5'-flanking region of mCC10 gene is highly homologous to rat CC10 gene and partially homologous to rabbit UG sequences. The *Analysis Software Package* from the Genetic

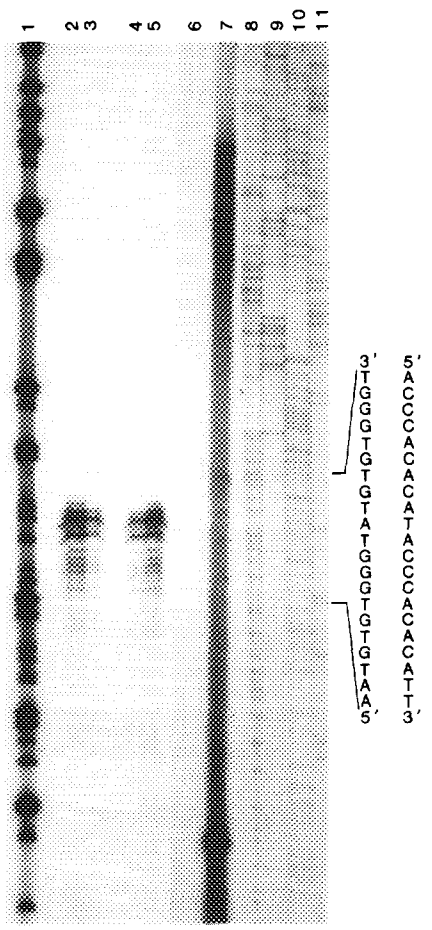

Figure 3. S1 nuclease protection assay: Reaction products were analyzed in the same gel but some of the lanes were not shown. Radiolabelled 271 nucleotides long Sac I - Pst I fragment were used for hybridizing total RNA (40 μg) from mouse lung, lanes (2 to 5) and kidney, lane 6. Lane 7 does not contain any RNA. An overnight autoradiograph of 6.5% sequencing gel is shown. Reaction mixture containing the RNA-DNA hybrids were treated as follows: lane 2: 100 U, 30° C. lane 3 : 400 U, 30° C. lane 4: 400 U, 37° C . lane 5: 100 U, 37° C. lane 6 : 100 U, 30° C. lane 7 : 400 U, 37° C. Lane 1 : labelled fragments of pBR322, digested with Hpa II. Lanes 8, 9, 10, and 11 corresponds to G, C, T and A in the reaction mixture.

Computer Group of Wisconsin (UWGCG, Madison, WI) was used for comparison. Data obtained from the homology search was used for a dot matrix analysis plot. As shown in Fig. 4, the mCC10 5'-flanking region is highly homologous to the rat 5'-flanking sequences and partially homologous to rabbit UG and human CC10 sequences (data not shown). 3.4 kb mouse sequence (including first exon and 32 nucleotides from first intron) was used to compare 2.8 kb rat sequence (3), 3.482 kb human sequence and 4.028 kb rabbit sequences (4). Figure 4 shows that almost the entire sequence of mCC10 used in this comparison is highly homologous to rat CC10 regulatory region. mCC10, as rat CC10, lacks approximately the 2.1 kb and the 0.3 kb sequences upstream of the transcription start site which are present in rabbit UG . The larger fragment of rabbit UG contains a cluster of glucocorticoid/progesterone response elements and is believed to be involved in glucocorticoid induction of the UG gene (3,15).

The mCC10 gene along with 3.3 kb 5'-flanking sequences has been reported. As expected, the mouse sequence is highly homologous to both rat CC10 gene and 5'-flanking sequence. Also, there are significant homologies of the mCC10 5'-flanking region with the rabbit UG and human CC10 genes. In all species studied, CC10 is predominantly expressed in the Clara cells of the lung. The cellular specificity of the CC10 gene makes it ideal for the analysis of transcription factors regulating the expression of this gene. A computer assisted search was conducted to find putative regulatory elements. A TATA box is located at position -23 to -20, an appropriate position for the correct initiation of transcription by RNA polymerase II. We also have identified the homologous sequences to a CAAT box , located at -62/-59 , -81/-78 and -156/-153. Stripp and coworkers have demonstrated functional response elements for AP1, octamer (Oct1) and hepatocyte nuclear factors (HNF3) in the 5'-flanking region of rat CC10 gene (19). Consensus sequences for the binding sites of these transcription factors were used in the computer analysis

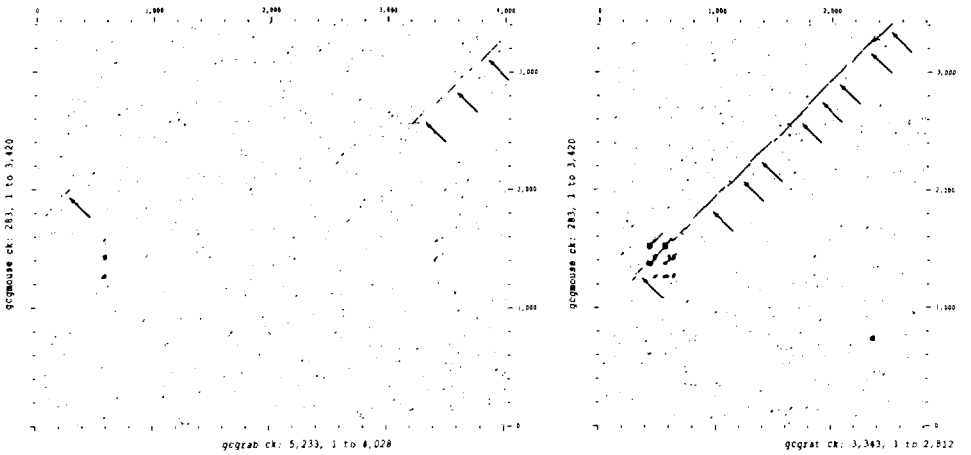

Figure 4. Similarity of 5'-flanking region between mouse/rat CC10 gene and mouse CC10/rabbit UG gene. Sequence comparison was done using dot matrix program. In the matrix each dot represents 14 identitical nucleotides in a total of 21 nucleotides. 3420 nucleotides (-3274 to -146) of mouse sequence were compared with 2785 nucleotides (- 2332 to 453) of rat cc10 , and 4028 nucleotides (-3920 to 108) of rabbit UG gene. Regions of homology are indicated by an arrow.

of the mCC10 flanking sequence. One hepatocyte nuclear factor (HNF3) binding site was located at position -121/-111. Oct1 binding sites were located at positions -120/-110 and -93/-87. The mCC10 is regulated by glucocorticoid hormones in the lung (26). Using the GRE consensus sequence present in tyrosine amino transferase gene upstream regulatory sequence (27), computer analysis identified three potential sites for glucocorticoid/progesterone response element. The GREs were found at positions -304/-290, -1039/-1025 and -2263/-2249. The positions of these sequences in mCC10 were similar to those positions identified in the rat CC10. Other *cis* acting elements identified were six CACCC boxes at positions -3229 , -2827 , -2674 , -2583 , -9 and 21. Finally a binding site for transcription factor E4TF2 (TGGGAAT) has been recognized at positions -795 and -1244. E4TF2 is involved in transcriptional regulation of the E4 gene in adenovirus (28). The functional significanse of these putative sequences remains to be verified.

The mCC10 sequence showed the presence of B1 murine repetitive sequence, located between -1545 and -1736. The presence of B1 repetitive sequence has been reported also for rat CC10 and rabbit UG 5'-flanking region. More than 105 copies of short interspersed repeats of B1 and B2 repetitive sequence are scattered throughout the mouse genome. The existence of sequences resembling the RNA polymerase III consensus promoter sequence in both B1 and B2, as well as the presence of short direct repeats at both ends, identify those sequences to be the transposable elements. There are numerous examples of mutational effects of transposon insertion such as inactivation of genes involved in Drosophila eye color (29), or activation of cellular oncogenes in neoplasia (30) or androgen responsiveness of a sex limited protein (SLP) gene by insertion of a regulatory element, later identified to be the 5 'long terminal repeat element (LTR) of a provirus (31). At this time it is not clear to us whether the presence of B1 repetitive sequence at the 5'-flanking region has any influence on mCC10 gene activity .

During the last decade, the mouse has been used as a mammalian model system for the study of the genetic control of development. This is due to development of two potent technologies: the transfer of foreign genes via microinjection of DNA into the one-cell embryo (32) and the manipulation of the murine genome using pluripotent mouse embryonic stem cells (33). Isolation of this genomic clone with more than 6 kb of 5'-flanking region will allow us to identify the *cis* acting elements regulating the expression of the mCC10 gene. Once the sequences important for the tissue specific expression of mCC10 are identified, their interaction with nuclear proteins will be determined. Also, the mCC10 genomic secquences could be used for gene targeting in embryonic stem cells to identify the physiological significance of the mCC10 protein.

Acknowledgments

Acknowledgments are due to Philip Gideon and Osvaldo Trujillo for their technical expertise. This work is supported by an NIH grant, HL47620.

References

1. Reid, L. , and Jones, R. Federation Proc. (1979) 38: 191-196.
2. Cato, A. C. B., Geisse, S. Wenz, M., Westphal, H. M. and Beato, M. (1984) EMBO J. 3: 2771-2778.

3. Hagen, G., Wolf, M., Katyal, S. L., Singh, G., Beato, M. and Guntram, S. (1990) Nuclic Acids Res. 18: 2939-2946 .
4. Margraf, L. R., Finegold, M. J., Stanley, L. A., Majors, A., Hawkins, H. K. and DeMayo, F. J. (1993) Amer. J. of Respir. Cell and Mol. Biol. 9: 231-238.
5. Bedette, C. D. ,Singh, J., Singh, G., Katyal, S. L. and Wong-Chong, M. L. (1987) J. Histochem. and Cytochem. 35: 789-794 .
6. DeMayo, F. J., Damak, S., Hansen, T. and Bullock, D. W. (1991) Mol. Endocrin. 5: 311-318.
7. Aumuller, G., Seitz, J., Heyns, W. and Kirchner, C. (1985) Histochem. 83: 413-417.
8. Katyal, S. L., Singh, G . , Brown, W. E., Kennedy, A. L., Squeglia, N. and Wong Chong, M. (1990) Prog . Respir. Res . 25: 29-35.
9. Miele, L., Cordella-Miele, E., and A. B. Mukherjee. (1987) Endocr. Rev. 8:474-490.
10. Lund, J., Norlund, L.,and Gustaffson, J. A. (1988) Biochemistry 27: 7895-7901.
11. Meile, L., Cordella-Miele, E., Facchiano, A. and Mukerjee, A. B. (1988) Nature 335:736-730.
12. Miha , K. and Riedel, N. (1991) Am. J. Respir. Cell Mol. Biol. 5 : 1-3.
13. Singh, G., Katyal, S. L., Brown, W. E., Phillips, S., Kennedy, A. L., Anthony, J. and Squeglia, N. (1988) Biochem. Biophys. Acta. 950: 329- 337.
14. Menne, C., Suske, G., Arnemann, J., Wenz, M., Cato, A. C. B. and Beato, M. (1982) Proc. Natl. Acad. Sci. USA. 79: 4853-4857 .
15. Suske, G., Wenz, M., Cato, A. C. B. and Beato, M. (1983) Nuclic Acids Res. 11: 2257-2271.
16. Wolf, M., Klug, J., Hakenberg, R., Gessler, M., Grzeschik, K., Beato, M. and Suske, G. (1992) Human Molecular Genetics. 1: 371-378.
17. Jantzen, K., Fritton, H. P., Igo-Kemenes, T., Espel, E., Janich, S., Cato, A. C. B., Mugele, K. and Beato, M. (1987) Nuclic Acids Res. 15: 4535-4552.
18. Hayward, W. G., Neel, B. E. and Astrin, S. M. (1981) Nature. 210: 475-480.
19. Stripp, B.R., Sawaya, P.L., Luse, D.S., Wikenheiser, K.A., Wert, S.E., Huffman, J. A., Lattier, D.L., Singh, G., Katyal, S.L. and J.A. Whitsett. (1992) J. Biol. Chem. 261: 14703-14712.
20. Weaver, R. F., and Weissman, C. (1979) Nuclic Acids Res. 7: 1175-1193.
21. Berk, A., and Sharp, P. A. (1977) Cell. 12: 721-732 .
22. Sambrook, J., Maniatis, T. and Fritsch, E. F. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA. 74: 5463-5467.
24. Vieira, J., and Messing, J. (1987) Methods in Enzymology 153: 3-11.
25. Sollner-Webb, B. and Reeder R. H. (1979) Cell. 18: 485-499.
26. Lopez de Haro, M. S. and A. Nieto. (1985) Biochem. J. 225: 255-258.
27. Strahle, U. , Klock, G. and Schutz, G. (1987) Proc. Natl. Acad. Sci. USA. 84: 7871-7875.
28. Jones, N. C., Rigby, P. W. J. and Ziff, E. B. (1988) Genes Develop. 2:267-281.
29. Rubin, G. M., Kidwell, M. G., and Bingham, P. M. (1982) Cell. 29: 987-994.
30. Hayward, W.S., Neel, B.G. and Astrin, S.M. (1981) Nature 290: 475-480.
31. Stavenhagen, J. B., and Robins, D. M. Cell. (1988) 55: 247-254.
32. Palmiter, R. D and Brinster, R. L .(1986) Annu. Rev. Genet. 20: 465-499.
33. Capecchi, M. R. (1989) Science. 244: 1288-1292.