Vol. 215, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 4, 1995 Pages 106-113

MOLECULAR CLONING, SEQUENCE, AND EXPRESSION PATTERNS OF THE HUMAN GENE ENCODING CCAAT/ ENHANCER BINDING PROTEIN α (C/EBPα)[1]

Per Antonson and Kleanthis G. Xanthopoulos*,†

Karolinska Institute, Department of Bioscience at NOVUM, Center for Biotechnology, NOVUM, S-141 57 Huddinge, Sweden

Received August 23, 1995

The human gene encoding the transcription factor C/EBPα was isolated from an umbilical cord genomic library screened by low stringency hybridization. Two overlapping clones were characterized by restriction enzyme analysis and included 13.2 kb of the C/EBPα locus. The entire gene and 471 bp of the promoter were sequenced. The human C/EBPα gene is 2783 bp long and encodes a 356 amino acid long protein, which is the same in length as for rat C/EBPα. Compared to rat C/EBPα, there are two insertions of two amino acids and one deletion of four. The amino acid similarity between the two proteins is over 92%. The human C/EBPα gene was found to be expressed at the highest levels in placenta. High expression was also found in liver, lung, skeletal muscle, pancreas, small intestine, colon and in peripheral blood leukocytes. However, the expression was undetectable or very low in brain, kidney, thymus, testis and ovary. These results show that the human C/EBPα gene is expressed in a tissue restricted manner.

CCAAT/enhancer binding protein alpha (C/EBPα) is a transcription factor and the prototype of the basic region-leucine zipper (bZIP) class of DNA binding proteins. The bZIP transcription factors are characterized by a region rich in basic amino acids which binds DNA and a leucine zipper domain that is used for dimerization (1). This class of transcription factors binds to DNA as either heterodimers or homodimers (1), and can be divided into several subgroups depending on their DNA binding and dimerization specificities. C/EBPα selectively dimerizes with factors belonging to the C/EBP family (2). To date, at least six members of this family have been identified (2-10). These proteins are believed to bind to the same nucleotide sequence, with the exception of C/EBPξ (CHOP) which not is believed to bind DNA (10).

Genes encoding C/EBPα have been cloned from several species and they show a high degree of evolutionary conservation, suggesting that it has a critical function. So far the C/EBPα gene sequences for the rat (3), mouse (4), chicken (11) and for the amphibians, *Xenopus* (12) and *Rana Catesbeiana* (13) have been reported. A homologue to the C/EBP family has also been cloned from *Drosophila* (14).

[1]The nucleotide sequence reported in this paper has been deposited in the EMBL/GenBank data base libraries under accession number U34070.

*To whom correspondence should be addressed. Fax: +46-8-774 5538.

†Present address: NIH, NCHGR, CGTB Bldg 49, Room 2A03, Bethesda, MD 20892.

ABBREVIATIONS: C/EBPα, CCAAT/enhancer binding protein α; USF, upstream stimulating factor; bZIP, basic region-leucine zipper.

0006-291X/95 $12.00

Studies of the expression patterns in rat and mouse have shown that C/EBPα is expressed in a rather tissue-restricted manner. High expression has been reported in liver, white fat, brown fat, placenta, intestine, lung and in myelomonocytic cells (4, 9, 15, 16). Within a given tissue, expression is often highest in terminally differentiated cells. Thus, C/EBPα is highly expressed in differentiated adipocytes (2, 15, 17-20) as well as the more differentiated cells of the liver and gut epithelial cells (21). However, during macrophage differentiation the opposite is true with high C/EBPα expression in proliferating cells and lower in terminally differentiated cells (16).

It has been reported that the rat and mouse C/EBPα mRNA can be translated into two major proteins with molecular weights of 42 and 30 kDa, respectively (22, 23). This is due to a mechanism in which ribosomes scan the C/EBPα mRNA and can start translation at a second start codon resulting in the shorter protein. The ratio between the products seems not to be subject to regulation which is in contrast to the regulated expression of the corresponding products translated from the C/EBPβ mRNA (24). The shorter gene products have been shown to act as repressor molecules in both cases (22-24).

Studies of the regulation of the mouse C/EBPα gene has shown that the gene is autoregulated, indicating that a tight control of its expression is needed (25, 26). A USF/MYC binding site, important for high levels of expression, is also found in the promoter. Recently, the human C/EBPα promoter was cloned and characterized (27). It was seen that the human C/EBPα gene is autoregulated through a mechanism different from that operating on the mouse gene due to the fact that it lacks C/EBP binding sites. Autoregulation of the human gene is achieved by enhanced expression of the USF transcription factor by C/EBPα. USF has been shown to positively regulate the C/EBPα gene.

Since the C/EBPα gene is often expressed in terminally differentiated cells and is involved in terminal differentiation of certain cell types (18, 28) suggests that it is an attractive molecule to study in terms of carcinogenesis and neoplasm formation. In an effort to further characterize members of the C/EBP gene family, we have isolated the human homolog to the C/EBPα gene. Here we report the nucleotide sequence of the entire gene and its promoter region. We also show the mRNA expression patterns in several human tissues.

MATERIAL AND METHODS

Isolation of genomic clones containing the human C/EBPα gene

A human genomic Lambda DASH II library derived from human umbilical vein endothelial cells from a pool of sixteen umbilical cords (Stratagene) was screened under low hybridization conditions with probes corresponding to the bZIP region of the mouse C/EBPα gene (a 400bp PstI/SstI fragment) (4) and a 1 kb PstI fragment from the rat C/EBPδ gene including almost the entire coding region (2). Probes were radioactively labeled by the random primer method (Amersham). The hybridization was performed in 43% formamide, 5x Denhart's solution, 5x SSC, 0.5% SDS and 100 μg/ml salmon sperm DNA at 37°C with a mixture of both probes.

Southern blot analysis and phage DNA restriction mapping

Phage DNA was prepared as described (29). Southern blot analysis was performed with Hybond N filters according to the manufacturers recommendations (Amersham). For restriction mapping of the phage DNA the inserts were excised by NotI digestion. The DNA was then partially cut with either EcoRI, BamHI or HindIII and Southern blot was performed according to standard procedures(30). Hybridization was in 6xSSPE, 3x Denhart's solution, 1% SDS with T3 and T7 primers, which hybridizes to the phage vector, labeled by phosphorylation with kinase and [γ-^{32}P]dATP as probes.

Subcloning and sequencing

The phage clone containing the longest insert was chosen for DNA sequence determination. From this phage a 3.7 kb EcoRI fragment was subcloned into the pBluescript KS- vector making the plasmid pB28E3.7 and a 5 kb BamHI fragment was subcloned into the pGEM3zf+ vector making the plasmid pG28B5.0. Sequencing was performed by dideoxynucleotide chain termination (31), using Sequenase version 2.0 (USB, Cleveland, OH) and fluorescent dye-labeled terminators (ABI) on the ABI 373A automated sequencer.

Northern blot analysis

Human multiple-tissue Northern blots (Clontech) containing approximately 2 µg poly(A)+ RNA were used to determine the expression of the C/EBPα gene. The blots were prehybridized and then hybridized at 42°C in 50% formamide/ 10x Denhardt's solution/ 5x SSPE/ 2% SDS containing 100 µg/ml sonicated salmon sperm DNA. The probes, a 700 bp Eco RI/ Hind III fragment from the 3´ untranslated region of the human C/EBPα gene and the human β-actin cDNA, were labeled as described above. The blots were washed twice in 2xSSC/ 0.05% SDS at room temperature and twice in 0.1xSSC/ 0.1% SDS at 50°C which was followed by autoradiography.

RESULTS

Isolation and restriction enzyme analysis of two overlapping human C/EBPα genomic clones

Screening of approximately 9 x 10^5 plack forming units from a human genomic library under low hybridization conditions with probes that included the mouse C/EBPα and the rat C/EBPδ bZIP regions yielded two clones that contained the C/EBPα gene. These clones were mapped using several restriction enzymes, creating a 13.2 kb long restriction map of the locus. In figure 1, a restriction map of the insert of the longest clone, is presented. In addition three clones containing the human C/EBPδ gene were found. However, these clones were not characterized further, since the human C/EBPδ gene has recently been identified (32, 33).

DNA sequence analysis

DNA fragments from a representative phage clone were subcloned into plasmid vectors and sequenced. The nucleotide sequence of 471 bp of the promoter region and the entire gene, including 120 bp of the 5´ noncoding region and 1585 bp of the 3´ noncoding region as well as the deduced amino acid sequence is shown in Figure 2. The human C/EBPα gene is intron-less with an open reading frame of 1068 bp. The GC content of the coding region is very high, almost 75%. Compared to the rat C/EBPα gene, the nucleotide identity is 90% in the coding region. On amino acid level, the identity is over 92%. Figure 3 shows an alignment of the human and rat C/EBPα. The human C/EBPα has two insertions of two amino acids at positions 100 and 190. In addition,

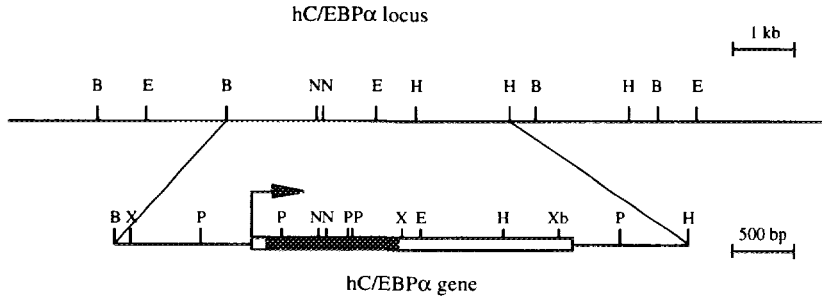

Figure 1. Restriction map of the human C/EBPα locus.
The longest genomic clone was mapped using restriction enzymes. The gene and surrounding area is shown in higher resolution. The open reading frame is shown as a dark box and the 5´- and 3´- untranslated regions as open boxes. The arrow indicates the start site and direction of transcription. Restriction enzymes are abbreviated as follows: B, BamHI; E, EcoRI; H, HindIII; N, NotI; P, PstI; X, XhoI and Xb, XbaI.

there is also a deletion of four amino acids at position 266. In total 18 amino acids differed in the human C/EBPα compared to the rat protein. However, in the bZIP region the rat and human proteins are identical.

The overall nucleotide identity of the first 390 bp between the mouse and the human promoter is 63%. The human promoter region contains a TATA box and initiator sequence that are identical with those of the mouse gene. Upstream of the TATA box binding sites for USF and several sites for SP1 and AP2 are found. The sequence between nucleotide -184 and -174 corresponds to the region where there is a C/EBP binding site in the mouse promoter. However, within this site there is one nucleotide that differ and it has been shown that C/EBPα can not bind to this region in the human promoter (27). We have also determined the chromosomal localization of the human C/EBPα to chromosome 19q13.1 gene by fluorescence *in situ* hybridization (not shown) This is in accordance with other reports (34, 35).

Expression of C/EBPa transcripts in human tissues

Northern blot analysis showed that the C/EBPα gene is expressed in a tissue specific manner (Figure 4). After normalization of the signal obtained with a β-actin probe, placenta appeared to express the highest steady-state levels of the approximately 2.7 kb C/EBPα mRNA. High expression levels was also detected in liver, lung, skeletal muscle, pancreas, small intestine, colon and in peripheral blood leukocytes. Intermediate expression was detected in heart, spleen and prostate. Brain, kidney, thymus, testis and ovary expressed very low or undetectable C/EBPα levels.

DISCUSSION

Six members of the C/EBP family have been identified to date. However, the DNA sequences have only been reported for two of the human homologues. In this study genomic clones containing the

```
-471  CTGCAGCCTCCCCGGGACGCGGGTCCGGGACAGGCCTGGTTCTGGCTTTGAAAGAGAATCCGCGCC  -406
-405  CCAGCAGCTCAAGACCAAGACTCGCCCTCCGCCCCCCACCCCTACCCCGTGCAGCCTCGGGATACT  -340
-339  CCTGGGCTCCCGGCCGTGGCTGGATACGGGCGCCTAGGGCAGGCAGGAGGAGGGGGCCCCCGCTAC  -274
-273  CGACCACGTGGGCGCGGGGGCGACGGCCGGGCCGGGGGCGGAGCTTGGAGCGAGCGCCGCGGCTCT  -208
-207  GCTGGGCGCGCTGGAGGCGGTGGGCGTTGCGCCGCGGCCTGCCTGGGGAGCGCGGCGCTGTGCCGC  -142
-141  GTGGTTCGCCGCCCCATGCCGGCCGCGCGCTAGGACCCAGCAGGCGCCGCGCCGCCGCAGCCCGGG   -76
                                                    TATA-BOX
-75   GACAGAGGCCGCCTCGGACTCTAGGGGGCGACGCGGCCTGCCGGGTATAAAAGCTGGGCCGGCGCG   -10
           +1
-9    GGCCGGGCCATTCGCGACCCGGAGGTGCGCGGGCGCGGGCGAGCAGGGTCTCCGGGTGGGCGGCGG    57
1                                                                     M     1
58    CGACGCCCCGCGCAGGCTGGAGGCCGCCGAGGCTCGCCATGCCGGGAGAACTCTAACTCCCCCATG   123
2      E  S  A  D  F  Y  E  A  E  P  R  P  P  M  S  S  H  L  Q  S  P  P    23
124   GAGTCGGCCGACTTCTACGAGGCGGAGCCGCGGCCCCCGATGAGCAGCCACCTGCAGAGCCCCCCG   189
24     H  A  P  S  S  A  A  F  G  F  P  R  G  A  G  P  P  K  P  P  A  P    45
190   CACGCGCCCAGCAGCGCCGCCTTCGGCTTTCCCCGGGGCGCGGGCCCGCCGAAGCCTCCCGCCCCA   255
46     P  A  A  P  E  P  L  G  G  I  C  E  H  E  T  S  I  D  I  S  A  Y    67
256   CCTGCCGCCCCGGAGCCGCTGGGCGGCATCTGCGAGCACGAGACGTCCATCGACATCAGCGCCTAC   321
68     I  D  P  A  A  F  N  D  E  F  L  A  D  L  F  Q  H  S  R  Q  Q  E    89
322   ATCGACCCGGCCGCCTTCAACGACGAGTTCCTGGCCGACCTGTTCCAGCACAGCCGGCAGCAGGAG   387
90     K  A  K  A  A  V  G  P  T  G  G  G  G  G  G  D  F  D  Y  P  G  A   111
388   AAGGCCAAGGCGGCCGTGGGCCCCACGGGCGGCGGCGGCGGCGGCGACTTTGACTACCCGGGCGCG   455
112    P  A  G  P  G  G  A  V  M  P  G  G  A  H  G  P  P  P  G  Y  G  C   133
456   CCCGCGGGCCCCGGCGGCGCCGTCATGCCCGGGGGAGCGCACGGGCCCCCGCCCGGCTACGGCTGC   519
134    A  A  A  G  Y  L  D  G  R  L  E  P  L  Y  E  R  V  G  A  P  A  L   155
520   GCGGCCGCCGGCTACCTGGACGGCAGGCTGGAGCCCCTGTACGAGCGCGTCGGGGCGCCGGCGCTG   585
156    R  P  L  V  I  K  Q  E  P  R  E  E  D  E  A  K  Q  L  A  L  A  G   177
586   CGGCCGCTGGTGATCAAGCAGGAGCCCCGCGAGGAGGATGAAGCCAAGCAGCTGGCGCTGGCCGGC   651
178    L  F  P  Y  Q  P  P  P  P  P  P  P  S  H  P  H  P  H  P  P  P  A   199
652   CTCTTCCCTTACCAGCCGCCGCCGCCGCCGCCGCCCTCGCACCCGCACCCGCACCCGCCGCCCGCG   717
200    H  L  A  A  P  H  L  Q  F  Q  I  A  H  C  G  Q  T  T  M  H  L  Q   221
718   CACCTGGCCGCCCCGCACCTGCAGTTCCAGATCGCGCACTGCGGCCAGACCACCATGCACCTGCAG   783
222    P  G  H  P  T  P  P  P  T  P  V  P  S  P  H  P  A  P  A  L  G  A   233
784   CCCGGTCACCCCACGCCGCCGCCCACGCCCGTGCCCAGCCCGCACCCCGCGCCCCGCGCTCGGTGCC   849
244    A  G  L  P  G  P  G  S  A  L  K  G  L  G  A  A  H  P  D  L  R  A   265
850   GCCGGCCTGCCGGGCCCTGGCAGCGCGCTCAAGGGGCTGGGCGCCGCGCACCCCGACCTCCGCGCG   915
266    S  G  G  T  G  A  G  K  A  K  K  S  V  D  K  N  S  N  E  Y  R  V   287
916   AGTGGCGGCACGGGCGCGGGCAAGGCCAAGAAGTCGGTGGACAAGAACAGCAACGAGTACCGGGTG   981
288    R  R  E  R  N  N  I  A  V  R  K  S  R  D  K  A  K  Q  R  N  V  E   309
982   CGGCGCGAGCGCAACAACATCGCGGTGCGCAAGAGCCGCGACAAGGCCAAGCAGCGCAACGTGGAG  1047
310    T  Q  Q  K  V  L  E  L  T  S  D  N  D  R  L  R  K  R  V  E  Q  L   331
1048  ACGCAGCAGAAGGTGCTGGAGCTGACCAGTGACAATGACCGCCTGCGCAAGCGGGTGGAACAGCTG  1113
332    S  R  E  L  D  T  L  R  G  I  F  R  Q  L  P  E  S  S  L  V  K  A   353
1114  AGCCGCGAACTGGACACGCTGCGGGGCATCTTCCGCCAGCTGCCAGAGAGCTCCTTGGTCAAGGCC  1179
354    M  G  N  C  A  *                                                   358
1180  ATGGGCAACTGCGCGTGAGGCGCGCGGCTGTGGGACCGCCCTGGGCCAGCCTCCGGCGGGGACCCA  1245
1246  GGGAGTGGTTTGGGGTCGCCGGATCTCGAGGCTTGCCCGAGCCGTGCGAGCCAGGACTAGGAGATT  1311
1312  CCGGTGCCTCCTGAAAGCCTGGCCTGCTCCGCGTGTCCCCTCCCTTCCTCTGCGCCGGACTTGGTG  1377
1378  CGTCTAAGATGAGGGGGCCAGGCGGTGGCTTCTCCCTGCGAGGAGGGGAGAATTCTTGGGGCTGAG  1443
1444  CTGGGAGCCCGGCAACTCTAGTATTTAGGATAACCTTGTGCCTTGGAAATGCAAACTCACCGCTCC  1509
1510  AATGCCTACTGAGTAGGGGGAGCAAATCGTGCCTTGTCATTTTATTTGGAGGTTTCCTGCCTCCTT  1575
1576  CCCGAGGCTACAGCAGACCCCCATGAGAGAAGGAAGGGGAGCAGGCCCGTGGCAGGAGGAGGGCTC  1641
1642  AGGGAGCTGAGATCCCGACAAGCCCGCCAGCCCCAGCCGCTCCTCCACGCCTGTCCTTAGAAAGGG  1707
1708  GTGGAAACATAGGGACTTGGGGCTTGGAACCTAAGGTTGTTCCCCTAGTTCTACATGAAGGTGGAG  1773
1774  GGTCTCTAGTTCCACGCCTCTCCCACCTCCCTCCGCACACACCCCACCCCCAGCCTGCTATAGGCT  1839
1840  GGGCTTCCCTTGGGCGGAACTCACTGCGATGGGGGTCACCAGGTGACCAGTGGGAGCCCCCACCCC  1905
1906  GAGTCACACCAGAAAGCTAGGTCGTGGGTCAGCTCTGAGGATGTATACCCCTGGTGGGAGAGGGAG  1971
1972  ACCTAGAGATCTGGCTGTGGGGCGGGCATGGGGGGTGAAGGGCCACTGGGACCCTCAGCCTTGTTT  2037
2038  GTACTGTATGCCTTCAGCATTGCCTAGGAACACGAAGCACGATCAGTCCATCCCAGAGGGACCGGA  2103
2104  GTTATGACAAGCTTTCCAAATATTTTGCTTTATCAGCCGATATCAACACTTGTATCTGGCCTCTGT  2169
2170  GCCCCAGCAGTGCCTTGTGCAATGTGAATGTGCGCGTCTCTGCTAAACCACCATTTTATTTGGGTT  2235
2236  TTGTTTTGTTTTGGTTTTGCTCGGATACTTGCCAAAATGAGACTCTCCGTCGGCAGCTGGGGGAAG  2301
2302  GGTCTGAGACTCCCTTTCCTTTTGGTTTTGGGATTACTTTTGATCCTGGGGGACCAATGAGGTGAG  2367
2368  GGGGGTTCTCCTTTGCCCTCAGCTTTCCCCAGCCCCTCCGGCCTGGGCTGCCCACAAGGCTTGTCC  2433
2434  CCCAGAGGCCCTGGCTCCTGGTCGGGAAGGGAGGTGGCCTCCCGCCAACGCATCACTGGGGCTGGG  2499
2500  AGCAGGGAAGGACGGCTTGGTTCTCTTCTTTTGGGGAGAACGTAGAGTCTCACTCTAGATGTTTTA  2565
2566  TGTATTATATCTATAATATAAACATATCAAAGTCAATGTCGGTGTCTTTTTAAAACCAGAAAGAAG  2631
2632  CTACTTCCAAGGTTGTCTGTGGGCCAGGTCACATTTGTAAATAATACAGCATTTTCCCTGGCGGCA  2697
2698  ATCCTGACTTTCATGAGCTCTCCATCCATCCTGAGCCCCTCTTACCCTAAGGGGGTGACTTACTTC  2763
2764  CCCCAGGCAAGACAAATAAATAGCAGAGGACAAGGCTCCAAATGGAGTATGTCCAGAGCCTGAAGG  2829
2830  CAGTCTCTTGGCGTCAGG                                                  2847
```

Figure 2. DNA and deduced amino acid sequence of the C/EBPα locus.

The nucleotide sequence is numbered relative to the transcription start site (+1). The transcription start site was adapted from Timchenko et al. (27). The deduced amino acid sequence is shown using the single letter code and are numbered beginning with the first ATG. The nucleotide sequence was deposited in the EMBL/GenBank data base, accession no. U34070.

C/EBPα locus were isolated. The human C/EBPα gene turned out to be very similar to the rat gene with an amino acid similarity of over 92 %. In the bZIP region the proteins are identical, indicating that the human protein has the same DNA binding and dimerization specifity as has been reported for the rat C/EBPα. A methionine at position 120 that has been found to serve as an second

```
            10         20         30         40         50         60
Human  MESADFYEAEPRPPMSSHLQSPPHAPSSAAFGFPRGAGPPKPPAPPAAPEPLGGICEHET
       ||||||||||||||||||||||||||||:||||||||||: ||||||||||||||||||||
Rat    MESADFYEAEPRPPMSSHLQSPPHAPSNAAFGFPRGAGPAPPPAPPAAPEPLGGICEHET

            70         80         90        100        110        120
Human  SIDISAYIDPAAFNDEFLADLFQHSRQQEKAKAAVGPTGGGGGGDFDYPGAPAGPGGAVM
       ||||||||||||||;|||||||||||||||||:|| :|||||||||||||||||||||||
Rat    SIDISAYIDPAAFNDEFLADLFQHSRQQEKAKAAAGP--AGGGGDFDYPGAPAGPGGAVM

           130        140        150        160        170        180
Human  PGGAHGPPPGYGCAAAGYLDGRLEPLYERVGAPALRPLVIKQEPREEDEAKQLALAGLFP
       ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Rat    SAGAHGPPPGYGCAAAGYLDGRLEPLYERVGAPALRPLVIKQEPREEDEAKQLALAGLFP

           190        200        210        220        230        240
Human  YQPPPPPPPSHPHPHPPPAHLAAPHLQFQIAHCGQTTMHLQPGHPTPPPTPVPSPHPAPA
       |||||||||  ||||::|||||||||||||||||||||||||||||||||||||||||||
Rat    YQPPPPPPP--PHPHASPAHLAAPHLQFQIAHCGQTTMHLQPGHPTPPPTPVPSPHPAPA

           250        260            270        280        290
Human  LGAAGLPGPGSALKGLGAAHPDLR----ASGGTGAGKAKKSVDKNSNEYRVRRERNNIAV
       :|||||||||::|||:::|||||     ::||:|||||||||||||||||||||||||||
Rat    MGAAGLPGPGGSLKGLAGPHPDLRTGGGGGGGAGAGKAKKSVDKNSNEYRVRRERNNIAV

            300        310        320        330        340        350
Human  RKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRGIFRQLPESSLVKAMGNCA
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Rat    RKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRGIFRQLPESSLVKAMGNCA
```

Figure 3. Comparison of the deduced amino acid sequence of the human C/EBPα with rat C/EBPα.

The derived amino acid sequence of the human C/EBPα is compared to the sequence reported for the rat C/EBPα (3).

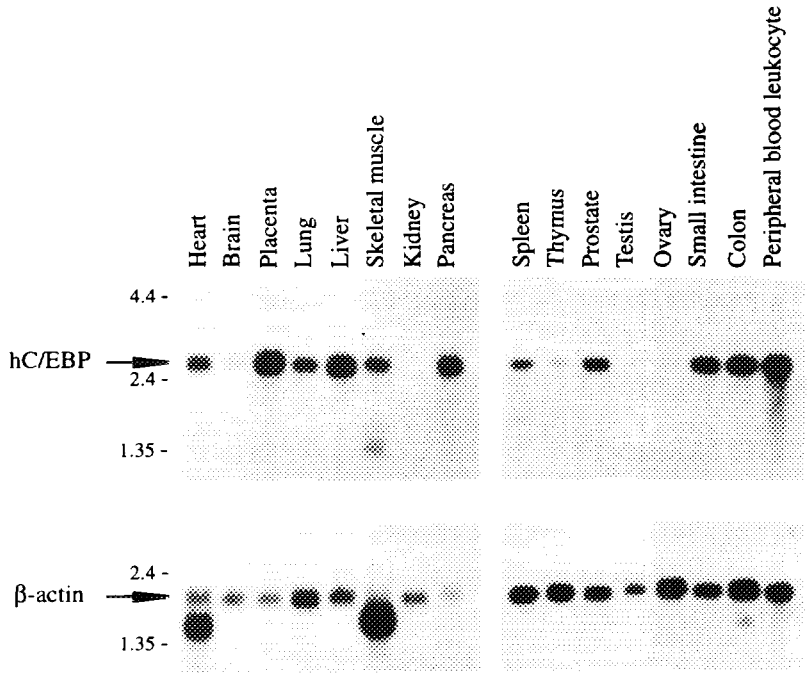

Figure 4. Expression of mRNA C/EBPα in various human tissues.

Northern blot analysis of C/EBPα. Each lane contains 2 μg of poly(A)+ RNA. C/EBPα gene expression was determined by hybridization, using a 700 bp Eco RI/ Hind III fragment from the untranslated region of the human C/EBPα gene as a probe. A β-actin probe was used to detect the β-actin mRNA on the same blots. The human C/EBPα and β-actin transcripts are indicated by arrows. The numbers at the left refer to the sizes of the RNA, given in kb.

translation start site in rat (22, 23) is conserved, suggesting that it serves the same function in the human gene. Two amino acid insertions and one deletion were found in the human protein. Interestingly, these are associated with repeated nucleotide sequences. The insertion at amino acid 100 and the deletion at 266 are associated with the trinucleotide repeat GGC and the insertion at 190 with the hexarepeat GCACCC. It has been reported that such events often occur during replication of repeat sequences (36), which might be the cause the cause of the insertions and the deletion in the human gene.

The promoter region is less conserved than the coding region. Several regions that have been shown to interact with DNA binding proteins in the mouse C/EBPα promoter (25, 26) were not conserved. However, the Myc/USF site that has been shown to be very important for regulation of the mouse C/EBPα gene, is also present in the human promoter. A detailed study of the human C/EBPα promoter was recently reported (27) in which it was shown that the difference of one nucleotide in the C/EBP binding site in the promoter (nucleotide -174, figure 2) makes C/EBPα unable to bind to this site. However, C/EBPα has been show to stimulate the expression of USF which in turn upregulates the expression of C/EBPα gene.

The expression pattern of C/EBPα showed that it is expressed in a tissue restricted manner with similar expression patterns described for the rat and mouse C/EBPα genes (2, 4, 9, 15), although the human gene is regulated by a different mechanism than the mouse (27). In similarity to the situation in the rat, highest expression was seen in placenta (15). As expected, expression is high in the liver. The high expression in peripheral blood leukocytes may be a result of the presence of C/EBPα in myelomonocytic cells for which C/EBPα expression has been described (16). The relatively high expression in pancreas and prostate has not been described before and the physiological significance is not clear at the moment.

Having determined the nucleotide sequence of the human C/EBPα locus and the expression patterns of the gene, the basis for further studies on the function of this gene is established.

ACKNOWLEDGMENTS

We thank Steven McKnight for the plasmid encoding C/EBPδ and Anna Berghard for critical reading of the manuscript. We also thank Susanne Müller for technical advice during the library screening. This work was supported by grants from the Swedish Cancer Society and Stockholm City County.

REFERENCES

1. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) Science 240, 1759-1764.
2. Cao, Z., Umek, R. M. & McKnight, S. L. (1991) Genes Dev. 5, 1538-1552.
3. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. (1988) Genes Dev. 2, 786-800.
4. Xanthopoulos, K. G., Mirkovitch, J., Decker, T., Kuo, C. F. & Darnell Jr ., J. E. (1989) Proc. Natl. Acad. Sci. USA 86, 4117-4121.
5. Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T.,

Hirano, T. & Kishimoto, T. (1990) EMBO J. 9, 1897-1906.
6. Roman, C., Platero, J. S., Shuman, J. & Calame, K. (1990) Genes Dev. 4, 1404-1415.
7. Descombes, P., Chojkier, M., Lichtsteiner, S., Falvey, E. & Schibler, U. (1990) Genes Dev. 4, 1541-1551.
8. Poli, V., Mancini, F. P. & Cortese, R. (1990) Cell 63, 643-653.
9. Williams, S. C., Cantwell, C. A. & Johnson, P. F. (1991) Genes Dev. 5, 1553-1567.
10. Ron, D. & Habener, J. F. (1992) Genes Dev. 6, 439-453.
11. Calkhoven, C. F., Ab, G. & Wijnholds, J. (1992) Nucleic Acids Research 20, 4093.
12. Xu, Q. & Tata, J. R. (1992) Mech. Dev. 38, 69-81.
13. Chen, Y., Hu, H. & Atkinson, B. G. (1994) Dev. Genet. 15, 366-377.
14. Montell, D. J., Rorth, P. & Spradling, A. C. (1992) Cell 71, 51-62.
15. Birkenmeier, E. H., Gwynn, B., Howard, S., Jerry, J., Gordon, J. I., Landschulz, W. H. & McKnight, S. L. (1989) Genes Dev. 3, 1146-1156.
16. Scott, L. M., Civin, C. I., Rorth, P. & Friedman, A. D. (1992) Blood 80, 1725-1735.
17. Herrera, R., Ro, H. S., Robinson, G. S., Xanthopoulos, K. G. & Spiegelman, B. M. (1989) Mol. Cell. Biol. 9, 5331-5339.
18. Umek, R. M., Friedman, A. D. & McKnight, S. L. (1991) Science 251, 288-292.
19. Samuelsson, L., Strömberg, K., Vikman, K., Bjursell, G. & Enerbäck, S. (1991) EMBO J. 10, 3787-3793.
20. Lin, F. T. & Lane, D. M. (1992) Genes Dev. 6, 533-544.
21. Chandrasekaran, C. a. G., J.I. (1993) Proc. Natl. Acad. Sci. 90, 8871-8875.
22. Ossipow, V., Descombes, P. & Schibler, U. (1993) Proc. Natl. Acad. Sci. USA 90, 8219-8223.
23. Lin, F. T., MacDougald, O. A., Diehl, A. M. & Lane, M. D. (1993) Proc. Natl. Acad. Sci. USA 90, 9606-9610.
24. Descombes, P. & Schibler, U. (1991) Cell 67, 569-579.
25. Christy, R. J., Kaestner, K. H., Geiman, D. E. & Lane, M. D. (1991) Proc. Natl. Acad. Sci. USA 88, 2593-2597.
26. Legraverend, C., Antonson, P., Flodby, P. & Xanthopoulos, K. G. (1993) Nucleic Acids Res. 21, 1735-1742.
27. Timchenko, N., Wilson, D. R., Taylor, L. R., Abdelsayed, S., Wilde, M., Sawadogo, M. & Darlington, G. J. (1995) Mol. Cell. Biol. 15, 1192-1202.
28. Tontonoz, P., Hu, E. D. & Spiegelman, B. M. (1994) Cell 79, 1147-1156.
29. Loenen, W. A. & Brammar, W. J. (1980) Gene , 249-259.
30. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Laboratory, Cold Spring Harbor, New York.
31. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
32. Kinoshita, S., Akira, S. & Kishimoto, T. (1992) Proc. Natl. Acad. Sci. USA 89, 1473-1476.
33. Cleutjens, C. B. J., van Eekelen C. C. E. M., van Dekken, H., Smit, E. M. E., Hagemeijer, A. & Wagner M. J., W. D. E. a. T., J. (1993) Genomics 16, 520-523.
34. Szpirer, C., Riviere, M., Cortese, R., Nakamura, T., Islam, M. Q., Levan, G. & Szpirer, J. (1992) Genomics 13, 293-300.
35. Hendricks-Taylor, L. R., Bachinski, L. L., Siciliano, M. J., Fertitta, A., Trask, B., DeJong, P. J., Ledbetter, D. H. & Darlington, G. J. (1992) Genomics 14, 12-17.
36. Richards, R. I. & Sutherland, G. R. (1994) Nature Genet. 6, 114-116.