

## Organization of the Chick CDC37 Gene\*

(Received for publication, October 21, 1997)

Lei Huang, Nicholas Grammatikakis, and Bryan P. Toole‡

From the Department of Anatomy and Cellular Biology, Tufts University School of Medicine, Boston, Massachusetts 02111

**CDC37 and the chaperone protein, Hsp90, form a complex that binds to several kinases, resulting in stabilization and promotion of their activity. CDC37 also binds DNA and glycosaminoglycans in a sequence-specific manner. In this study, we further characterize chick CDC37 and examine the organization of the CDC37 gene. Chick CDC37 is a ~50-kDa protein encoded by an mRNA of ~1.7 kilobases. The CDC37 gene is ~8.5 kilobases and contains 8 exons and 7 introns of various sizes. The presumptive promoter and 5'-flanking regions contain an E2 box and consensus binding sites for SP1, for the S8 homeodomain protein, and for two zinc finger clusters within the myeloid progenitor transcription factor, MZF1. Particularly striking is a ~470-base pair region composed of a highly repetitive 10–11-base pair sequence, (T/C)gCTAT(A/G)GGG(A/T) (where g represents the additional G present in the 11-base pair sequence). This region includes 15 copies of the sequence, TAT-GGGGA, which conforms to the DNA consensus sequence recognized by one of the zinc finger clusters in MZF1. These findings emphasize the potential importance of CDC37 in regulation of cellular behavior during tissue development and reorganization.**

Genetic studies have shown CDC37 is essential for the START event in yeast (1, 2), but other studies have shown that CDC37 is probably required in G<sub>2</sub>/M as well as G<sub>1</sub> (3, 4). Vertebrate CDC37 has been cloned from several species (5–8), and, recently, it was shown to be identical to p50 (8, 9). Initially, p50 was characterized as a component of a ternary complex with the chaperone protein, Hsp90, and p60<sub>src</sub> kinase (10, 11). Based on these and other studies, it was suggested that the p50-Hsp90-p60<sub>src</sub> kinase complex may mediate trafficking of p60<sub>src</sub> kinase to the plasma membrane, and that similar events occur with several other kinases (11–14). Further work showed that the *Drosophila* homologs of CDC37 and Hsp90 are involved in signal transduction in the *sevenless* pathway (3). It is now known that CDC37 and Hsp90 form a complex whose interaction with various protein kinases, e.g. Cdk4, p60<sub>src</sub> kinase, casein kinase II, MPS1 kinase, and Raf-1, is required for optimal activity (8, 14–17). The CDC37-Hsp90 complex may stabilize these enzymes or enable their correct folding, although there is some evidence that CDC37 can act independently as a chaperone (16). Thus, CDC37 most likely

plays several important roles in intracellular signaling, especially with respect to the cell cycle.

In addition to the chaperone interactions described above, CDC37 binds DNA in a sequence-specific manner at a site that is also recognized by the retinoblastoma gene product, pRB<sup>1</sup> (18). CDC37 and pRB also bind specifically to each other, suggesting that CDC37-pRB-DNA interactions may be involved in cell cycle regulation (19).

We have shown that CDC37 contains two binding motifs for the polysaccharide, hyaluronan, and that CDC37 binds hyaluronan and related glycosaminoglycans: chondroitin sulfate, heparan sulfate, and heparin (5). Although these glycosaminoglycans are usually associated with extracellular matrices, they are also present at the cell surface (20) and within the cytoplasm and nucleus of several cell types (21–23). Although their intracellular functions are not yet established, it is noteworthy that a specific subpopulation of heparan sulfate is targeted to the nucleus of hepatoma cells (21, 24) and that heparin inhibits Fos- and Jun-induced transcription events *in vitro* and *in situ* within the nucleus (25). The latter events may occur via competition with DNA binding (18, 25). Another possibility is that nascent hyaluronan, attached to hyaluronan synthase at the inner face of the plasma membrane (26), interacts with CDC37 at G<sub>2</sub>/M since published evidence indicates that hyaluronan synthesis may be required for completion of mitosis (27).

Because of the apparent importance of CDC37 as a regulatory protein, we have further characterized chick CDC37 and analyzed the organization of its gene, the first CDC37 gene for which this has been done.

### EXPERIMENTAL PROCEDURES

**Northern Blot Analysis**—For RNA preparation, limb and brain tissue from chick embryos at various stages of development were dissected, frozen immediately on dry ice, and stored at –80 °C. Total RNA was extracted with Trizol reagent (Life Technologies, Inc.) according to the manufacturer's instructions. Poly(A)<sup>+</sup> RNA was prepared by oligo(dT)-cellulose chromatography. Samples of the poly(A)<sup>+</sup> preparations were subjected to electrophoresis in formaldehyde gels and transferred to nitrocellulose membranes. The probe used for hybridization was labeled with [<sup>32</sup>P]dCTP using a random priming DNA labeling kit (Boehringer Mannheim). Hybridization and washes were done under standard high stringency conditions (28).

**3'- and 5'-RACE**—Poly(A)<sup>+</sup> RNA, prepared from 6-day chick embryo limbs as in the previous section, was used for 3'- and 5'-RACE (29).

After reverse transcription, a Marathon cDNA Amplification Kit was used according to the manufacturer's instructions (CLONTECH) for 3'- and 5'-RACE. The primer, RS1, used for 3'-RACE was 5'-CTGCTAAT-TACCTGGTCATCTGG-3'; this primer corresponds to bases 314–336 in our partial cDNA for CDC37, NG-13 (5), and is *underlined* in Fig. 1 at residues 675–697. The primer, RAS1, used for 5'-RACE was 5'-GC-CAACTCCAGGATGAAGTGCAT-3'; this is the reverse complement of bases 403–425 of NG-13 (5) and bases 764–786, *underlined* in Fig. 1. Unique bands obtained by 3'- and 5'-RACE were purified and ligated

\* This work was supported by National Institutes of Health Grants DE05838 and HD23681 (to B. P. T.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF035530 and AF035769.

‡ To whom correspondence should be addressed: Dept. of Anatomy and Cellular Biology, Tufts University School of Medicine, 136 Harrison Ave., Boston, MA 02111. Fax: 617-636-0380; E-mail: btoole@infonet.tufts.edu.

<sup>1</sup> The abbreviations used are: pRB, retinoblastoma gene product; RACE, random amplification of cDNA ends; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase PCR; UTR, untranslated region; bp, base pair(s); kb, kilobase(s); PAGE, polyacrylamide gel electrophoresis.

FIG. 1. **Sequence of chick CDC37 cDNA and deduced protein.** Nucleotide and deduced amino acid sequences are shown; nucleotides are numbered on the left and amino acids on the right. The initiation and stop codons flanking the open reading frame are shown in **bold**. An in-frame stop codon within the 5'-UTR and the polyadenylation signal in the 3'-UTR are double underlined. The two hyaluronan-binding motifs are double underlined and **bold**. Labels on the right refer to the primers used for 3'- and 5'-RACE (RS1, RAS1, and PAS1), for inverse PCR (PS1 and PAS1), and for cloning the CDC37 gene (GS1 and GAS1). RS1, RAS1, PS1, and PAS1 are single underlined; GS1 and GAS1 are shaded.

```

1 CAAATCGCGACCGGAAGTGG
21 GGGCCCGTGCCTGGAGGTTTCCAGAAGCGGCCGAGACGGCGCGGCGCGGGGGGG
81 CGGTTTATGGGGTCCGCTTCGCTATGGTGGATTACAGCGCTCTGGGACCATCGAGGTGT GS1 & PAS1
M V D Y S V W D H I E V S 13
141 CGGACGATGAGGACGAGACGACACCCCAACATCGACACCGCTCCCTCTTCGATGGCGGC PS1
D D E D E T H P N I D T A S L F R W R H 33
201 ACCAGGCCCGCTGGAGAGGATGGAGCAGTTCAGAAAAGAGAAGAGAGCTGGACAAGG
Q A R V E R M E Q F Q K E K E E L D K G 53
261 GCTGCAGGGAATGCAAGAGGAACTCGCCGAGTGCCAGAAAGAACTGAAAGAGTTGGAGG
C R E C K R K L A E C Q K K L K E L E V 73
321 TGGCGGAGCCCGGGGGGGATCCGGGGGGGGCGGGGGGAGAGGAGAGCTGCAGCGCG
A E P G G G S G G G R G E R E R L Q A E 93
381 AGGCGCAGCAGCTGCCCCACGAGGAGAGGAAGTGGGAGAGCAAAATGGAGGAGCTGCACA
A Q Q L R H E E R N W E S K M E E L R K 113
441 AGAAGGAGAAGATATGCCCTGGAACGTGCACACGCTGAGCAAGGACGGCTTCAGCAAGA
K E K N M P W N V H T L S K D G F S K S 133
501 GCGTTTTCACCTGTAAGGAGAGGAGAGGAGGAGACGAGGAGCAGAAGGAGCAGAGG
V P N V K A E E K E E T E E Q K E Q K H 153
561 ACAAAACCTTCGTGGAGCGCCACGAGAAGCAGATCAAGCATTTCGGCATGTCGCGCGCTW
K T F V E R H E K O I K H F G M L R R W 173
621 GGGACGACGCCAGAAGTACCTCTCGGACAACCCACCTCGTTTGGCAGGAGAGCTGCTA RS1
D D S Q K Y L S D N P H L V C E E T A N 193
681 ATTACCTGGTCTCTGGTGCATCGACCTGGAGGTGAGGAGAAGCAGGCTCTGATGGAGC
Y L V I W C I D L E V E E K Q A L M E Q 213
741 AGTTCGCCCCACGACCATCTGTCATGCAGTTTCCTCGGAGTTGGCCAAAGAGCTCAAGG RAS1
V A H Q T I V M Q F I L E L A K S L K V 233
801 TGGATCCCCGCGCTGCTTCGTCAGTTCTTCAACAAATCAAGACCGCCAGCAGCAGT
D P R A C R F Q F F T K I K T A D Q Q Y 253
861 ACATGGAGGGCTTCAACGATGAGTTGGAGGCTTCAAGGAGCGGTCGCGCGCGGCCA
M E G F N D E L E A F K E R V R G R A K 273
921 AGGCGCGCATCGAGAGGGCCATGCGGGAGTACGAGGAGGAGGAGCGGCGAGCGCTCG
A R I E R A M R E Y E E E E R Q K R L G 293
981 GGCCCGCGGCGCTCGACCCCGTCGACGCTACGAGTCCCTCCACCCGAGCTGCAGAAAT
P G G L D P V D V Y E S L P P E L Q K C 313
1041 GCTTTGACGCCAAAGACGTGCAGATGCTGCAGGACACCATCAGCCGAATGGACCCCATG
F D A K D V Q M L Q D T I S R M D P T E 333
1101 AGGCCAAATACACATGCAGCGCTGCATCGATTGCGGGCTGTGGTCCCCAACGCCAAAG
A K Y H M Q R C I D S G L W V P N A K A 353
1161 CGGCAGCTGAGGGTGGGGGCGAGGGGGGCGCACGCCAGCCGGGGGGGGCGACTCTG
A A E G G G Q G G A H G Q P G G A D S E 373
1221 AAGCACTCTATGAGAAATCCCCAAGAGAGTGGGAGGAAGAGCGGGGGAGGGGAAGG
A L Y E E I P K E S G E E E G G K A 393
1281 CTTAGGGGGCCCCACACCCCTTCAGTGACCCCATCCCCACTACCCCCCCCCCTCC
*
1341 CCCCCTCCCGACCCCGGGGGGGGGGACACACGCGCTCTCTTACCCCCCCCCCCCCC
1401 CATGCTTTGCCCACTTTTGGGGGGACACACACGCTGTGGGAGGGGGGGGGGAGATG
1461 GGTGGCTCTTTTGGGGGTGGGACCCCCCAGTTGTGGGTTCGCGGGTGGGGGTCCCCA
1521 GGGCTGCTGTGCCCCACCCCCCTCCCAACCCCACTGACGACCAATAAGTGA GAS1
1581 TGGAAATG-polyA

```

into the pCRII vector (Invitrogen) for transformation, cloning, and sequencing. The 3'- and 5'-RACE products were fused by PCR, and the resulting full-length cDNA was cloned into the same vector for sequencing. To obtain as much of the 5'-UTR sequence as possible, the primer PAS1 (Fig. 1) was also used in 5'-RACE. The size of 5'-RACE products was determined by agarose gel electrophoresis, transfer to nitrocellulose, and hybridization with the probe, GS1 (Fig. 1). Another sample was then electrophoresed on agarose for cloning and sequencing of the products.

**Determination of Gene Organization by PCR**—Primers GS1 and GAS1 (shaded in Fig. 1), which correspond to the extreme 5' and 3' ends, respectively, of the cDNA extended by 5'- and 3'-RACE, were used in PCR with chick genomic DNA (CLONTECH) as template. Elongase (Life Technologies, Inc.) was employed according to manufacturer's instructions to enable accurate amplification of long PCR products. The amplified genomic DNA product was cloned into the pCR-Script vector (Stratagene). Boundaries between exons and introns were determined by PCR and sequencing, using primers complementary to sequences at various positions along the cDNA.

To determine genomic sequences upstream of the cDNA sequence, inverse PCR (30) was used. Aliquots of chick genomic DNA were digested with *Bam*HI, *Eco*RV, *Eco*RI, or *Hinc*II restriction enzymes. Each digest was diluted to 2 ng DNA/ml and ligated with DNA ligase (Life Technologies, Inc.). The circularized DNA digests were then used as templates for PCR with two primers (PAS1 and PS1, underlined in Fig. 1) corresponding to sequences in the first exon of the CDC37 gene.

**Sequencing**—After ligating into the pCRII or pCR-Script vector and transforming bacteria, the nucleotide sequences of selected cloned inserts were determined by the double-stranded DNA/dideoxy chain termination method (31) using a Sequenase 2.0 kit (U. S. Biochemical Corp.).

**Antibody to CDC37 Fusion Protein**—The partial length cDNA for CDC37, NG-13 (5), was ligated into the pGEX-2T vector (Pharmacia Biotech Inc.) and used to transform competent *Escherichia coli* DH  $\alpha$ 5; the transformed cells were grown and induced with 0.2 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside at 37 °C. Fusion protein was purified on an affinity column of glutathione-Sepharose following the manufacturer's instructions (Pharmacia), followed by preparative SDS-PAGE and electro-elution of the protein from the gel. Polyclonal antibody was raised in rabbits (HTI Bioproducts, Inc.), and the antiserum was purified by antigen affinity chromatography.

**Western Blot Analysis of Chick Embryo Fibroblasts**—Chick embryo fibroblasts (line CL-29 from American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium plus 10% fetal bovine serum and 1% antibiotics. At confluence, the cells were washed with PBS and lysed with extraction buffer (0.05 M Tris-HCl, pH 7.5, 0.5 M NaCl, 1% Nonidet P-40, 0.1% SDS, 5  $\mu$ M leupeptin, 5  $\mu$ M pepstatin, 5 mg/ml aprotinin, 2.5 mM phenylmethylsulfonyl fluoride). Insoluble material was removed by centrifugation, and the lysate was subjected to SDS-PAGE, transferred to polyvinylidene difluoride membrane, and probed with the purified polyclonal antibody or with the monoclonal antibody, IVd4, used for immunoscreening (5).

## RESULTS AND DISCUSSION

**Cloning of Full-length cDNA for Chick CDC37**—In a previous study (5) we obtained a partial cDNA for CDC37, termed NG-13, by immunoscreening of a chick embryo heart cDNA library. By Northern blot analysis, the size of the chick CDC37 mRNA was found to be ~1.7 kb (5). Using 3'- and 5'-RACE with primers RS1 and RAS1 (Fig. 1), followed by PCR-mediated fusion of the two products, we have now obtained a larger chick

	1				50
Mcdc37	MVDYSVVDHI	EVSDDEDETH	PNIDTASLFR	WRHQARVERM	EQFQKEKEEL
Hcdc37	MVDYSVVDHI	EVSDDEDETH	PNIDTASLFR	WRHQARVERM	EQFQKEKEEL
Ccdc37	MVDYSVVDHI	EVSDDEDETH	PNIDTASLFR	WRHQARVERM	EQFQKEKEEL
	51				100
Mcdc37	DRGCRECKRK	VAEQCKLKE	LEVAE....	SDGQVEIERI	RAEAQQLRKE
Hcdc37	DRGCRECKRK	VAEQCKLKE	LEVAE....	.GGKAEIERI	QAEAQQLRKE
Ccdc37	DKGCRECKRK	LABCQKLLKE	LEVAEPGGGS	GGGRGERERI	QAEAQQLRHE
	101				150
Mcdc37	ERSWEQKLED	MRKKEKNMPW	NVDTLKSDGF	SKSMVNTKPE	KAEEDSEEAR
Hcdc37	ERSWEQKLEE	MRKKEKSPW	NVDTLKSDGF	SKSMVNTKPE	KTEEDSEVR
Ccdc37	ERNWESKMEE	LKKKEKNMPW	NVHTLSKSDGF	SKSVFNKAE	EKEE.TEEQK
	151				200
Mcdc37	BQKHKTVEEK	YEKQIKHFGM	LHRWDDSQKY	LSDNVHLVCE	ETANYLVIWC
Hcdc37	BQKHKTVEEK	YEKQIKHFGM	LRRWDDSQKY	LSDNVHLVCE	ETANYLVIWC
Ccdc37	BQKHKTVEER	HBKQFKHFGM	LRRWDDSQKY	LSDNVHLVCE	ETANYLVIWC
	201				250
Mcdc37	IDLEVEEKCA	LMEQVAHQTM	VMQFILELAK	SLKVDPRACE	RQFFTKIKTA
Hcdc37	IDLEVEEKCA	LMEQVAHQTI	VMQFILELAK	SLKVDPRACE	RQFFTKIKTA
Ccdc37	IDLEVEESQA	LMEQVAHQTI	VMQFILELAK	SLKVDPRACE	RQFFTKIKTA
	251				300
Mcdc37	DHQMMEGFKY	ELEAFKERVN	GRAKLRIEKA	MKEYEEERK	KRLGPGGLDP
Hcdc37	DHQMMEGFND	ELEAFKERVN	GRAKLRIEKA	MKEYEEERK	KRLGPGGLDP
Ccdc37	DQYMMEGFND	ELEAFKERVN	GRAKLRIEKA	MREYEEERQ	KRLGPGGLDP
	301				350
Mcdc37	VEVYESLPPE	LQKCFDVKDV	QMLQDAISK	DPTDAKYHMQ	RCIDSGWLVP
Hcdc37	VEVYESLPPE	LQKCFDVKDV	QMLQDAISK	DPTDAKYHMQ	RCIDSGWLVP
Ccdc37	VDVYESLPPE	LQKCFDAKDV	QMLQDTISRM	DPTDAKYHMQ	RCIDSGWLVP
	351				394
Mcdc37	NSKSGEAKEG	EEAGPGDE..	.....LEAV	PKAGNEKDVS	A...
Hcdc37	NSKASEAKEG	EEAGPGDE..	.....LEAV	PKTGDEKDVS	V...
Ccdc37	NAKAAAEGGG	QGAHGGQEGG	ADSEALYEEI	PKESGGEEGG	EGKA

FIG. 2. Comparison of amino acid sequences for mouse, human, and chick CDC37. Amino acids identical in all three species are shaded. The tyrosine phosphorylation (residues 173–180) and pRB-binding (187–191) motifs are in bold; the two hyaluronan-binding motifs (163–173 and 268–276) are underlined. The human and mouse sequences were taken from Dai *et al.* (7) and Stepanova *et al.* (8).

CDC37 cDNA of ~1.6 kb. To extend the 5'-UTR sequence as far as possible, we also performed 5'-RACE with a primer, PAS1, situated at the beginning of the open reading frame. After electrophoresis through agarose, the products of this reaction were transferred to nitrocellulose membrane then hybridized with a probe just upstream of the PAS1 primer, *i.e.* GS1 (Fig. 1), to identify all products. A single band of ~130 bases was obtained. On cloning and sequencing, this band was found to contain several products of similar size but differing in length by a few nucleotides at their 5' ends; the sequence of the longest product is given as the 5' terminus of the cDNA in Fig. 1. Because of the above approaches, we conclude that this sequence includes virtually all of the 5'-UTR.

The full-length cDNA described above contains an open reading frame of 1179 bases, encoding a 393-amino acid polypeptide with a predicted molecular mass of ~45 kDa (Fig. 1). The first ATG codon conforms to the Kozak consensus for the translation initiation site; in addition, the presence of an in-frame stop codon immediately upstream of this ATG (Fig. 1) is consistent with this conclusion. The 3'-UTR contains the consensus polyadenylation signal, AATAAA.

**Analysis of Primary Structure of Chick CDC37**—The amino acid sequence of chick CDC37 has 84% identity and 91% similarity to that of mouse (7, 8), and 82% identity and 90% similarity to that of mouse (7, 8) (Fig. 2). These are the vertebrate homologues of yeast and *Drosophila* CDC37, previously characterized by genetic methods (2, 3).

A novel observation previously made in our laboratory was that the amino acid sequence of chick CDC37 includes two consensus motifs for hyaluronan binding (5). We also demonstrated directly that recombinant chick CDC37 binds hyaluro-

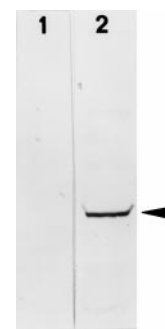


FIG. 3. Western blot of CDC37 from chick embryo fibroblasts. Extracts of chick embryo fibroblasts were separated by SDS-PAGE and probed with preimmune serum (lane 1) or with immune serum prepared against recombinant CDC37 (lane 2). A single band at ~50 kDa was obtained in the latter (arrowhead). A similar result was obtained with the monoclonal antibody, Ivd4, used for immunoscreening (5).

nan and other related glycosaminoglycans (5). Comparison with the human and mouse sequences reveals that these hyaluronan-binding motifs are conserved among vertebrate species (amino acid residues 163–173 and 268–276 in Fig. 2). Of these motifs, the latter is the classic B(X<sub>7</sub>)B motif (where B is arginine or lysine and X is any non-acidic amino acid); the former is B(X<sub>8</sub>)BB, which would be expected to have equivalent activity (32).

Other motifs are also found in human and mouse, as well as in the chick amino acid sequence. At least one putative tyrosine phosphorylation site is present in CDC37 and is highly conserved across species (amino acid residues 173–180 in Fig. 2). This is consistent with the finding that CDC37 is phosphorylated (33, 34). The pRB-binding motif (LVCEE) noted previously (19) is also highly conserved (residues 187–191 in Fig. 2).

**Expression of CDC37 in Chick**—We performed Western blots on chick embryo fibroblast extracts with both the monoclonal antibody, Ivd4 (5), and a polyclonal antibody raised against bacterially expressed recombinant CDC37 protein. In both cases, the major protein recognized by this antibody is ~50 kDa in size (Fig. 3), agreeing well with the size of the calculated product of the open reading frame. Expression of the recombinant protein in bacteria or in vertebrate cells transfected with the full-length CDC37 cDNA also yields a protein of ~50 kDa (data not shown).

We used the full-length composite cDNA as a probe in Northern analyses of mRNA obtained from different stages of chick limb and brain development, loading a constant amount of mRNA from each stage. In both tissues, maximum expression of the ~1.7-kb mRNA was reached at about 6 days of development, followed by a gradual decrease until hatching (Fig. 4). As expected, *in situ* hybridization and immunohistochemistry revealed wide distribution of CDC37 in morphogenetically active tissues throughout the chick embryo (data not shown).

**Organization of the CDC37 Gene**—Southern analysis has shown that the CDC37 gene is present as a single copy in several vertebrate genomes (6), and we have confirmed this result with the chick. Since no further genomic analysis has been reported, we isolated a CDC37 genomic clone to characterize its organization. We obtained the clone by PCR, using chick genomic DNA with primers corresponding to the 5' and 3' ends of chick CDC37 cDNA (Fig. 1). The amplified product, ~8.5 kb in length, was then ligated into a plasmid vector and used for further analyses.

Using this genomic clone in PCRs with numerous primer pairs corresponding to progressively more 3' regions of the cDNA, we mapped the positions of intron/exon boundaries within the gene. These products were cloned and sequenced. Eight exons and seven introns were found (Fig. 5A); in each



case, the sequences at these boundaries complied with the AG...GT consensus sequences for splicing sites (Fig. 5B). The sizes of introns were determined by PCR using primers from exon regions flanking each intron, and were found to range from ~0.1 to ~2 kb (Fig. 5A).

Inverse PCR (30) was used to obtain sequences upstream of

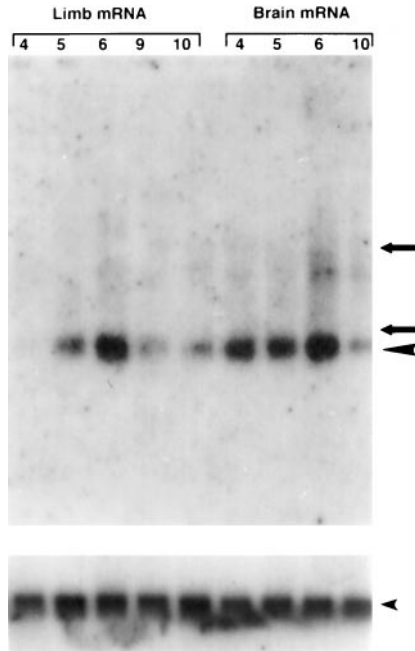


FIG. 4. Northern blot of RNA from chick embryo limb and brain. Lanes 1–5, mRNA from chick embryo brains at 4, 5, 6, 9, and 10 days of development, respectively; lanes 6–9, mRNA from chick embryo brains at 4, 5, 6, and 10 days of development, respectively. A band at ~1.7 kb (large arrowhead) was obtained in each case. Arrows indicate the positions of 18 S and 28 S ribosomal RNA. Glyceraldehyde-3-phosphate dehydrogenase (small arrowhead) was used as a measure of loading.

the cDNA sequence shown in Fig. 1; chick genomic DNA was used as template with a primer pair corresponding to sequences at the 5' end of the open reading frame of the chick CDC37 cDNA (Fig. 1). The upstream sequence of ~2 kb obtained in this way is shown in Fig. 6A. The accuracy of this sequence was confirmed with several primer pairs from within this sequence and exon 1, using genomic DNA as a template. The region between nucleotide residues -1115 and -645 upstream of the translation start codon (shaded in Fig. 6A) consists of a highly repetitive 10–11-bp sequence, (T/C)gCTAT(A/G)GGG(A/T) (where g represents the additional G present in the 11-bp sequence). Several other motifs, discussed later, are present in this upstream region.

The size of our full-length cDNA is ~1.6 kb (Fig. 1). Taking into account a poly(A) tail of ~100 bp, the size of this cDNA agrees very well with the size of mRNA observed in Northern blots, *i.e.* ~1.7 kb (Fig. 4). These observations imply that the major transcription initiation site is very close to the 5' end of the cDNA as shown in Fig. 1. As discussed above, this was confirmed by extending the 5' end of the cDNA as far as possible by 5'-RACE using the primer, PAS1, at the beginning of the open reading frame. The products were detected in Southern blots with a probe just 5' of the translation start codon, then cloned and sequenced to obtain the longest 5' sequence. Thus, we conclude that the 5' sequence shown in Fig. 1 is very close to the major transcription start site.

We investigated whether there might be additional initiation start sites. For this, we used RT-PCR with sense primers (S2, S3, and S4) corresponding to various sites upstream of the ATG start codon (and a control primer, S1, just 3' of the start codon) with a common antisense primer within exon 2 (primer AS1) (see Fig. 6, A and B); as a positive control, we also used genomic DNA as a PCR template with the same primers. The position of the antisense primer in exon 2 was chosen so that the expected RT-PCR products from mRNA would differ greatly in size from the products of any contaminating genomic DNA that might be in the mRNA preparation, since the ~1.5-kb intron 1 would then be included. As expected, we obtained a strong band using

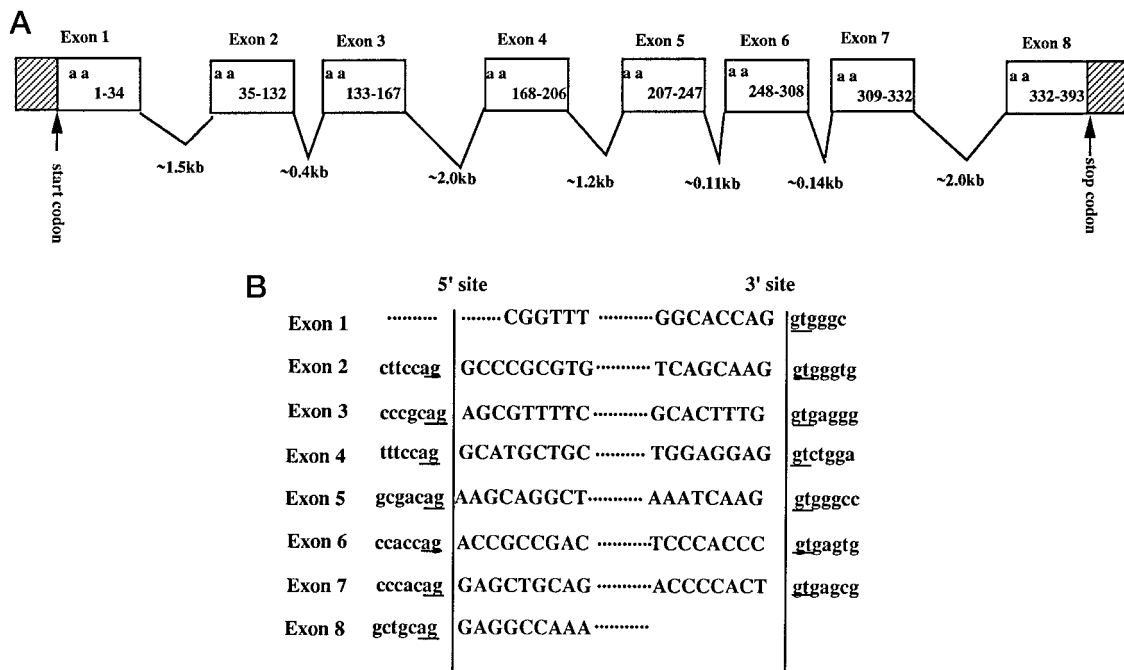


FIG. 5. Organization of exons and introns in the CDC37 gene. A, arrangement and sizes of exons and introns. B, nucleotide sequences immediately adjacent to each exon (exon sequences in *uppercase*; intron sequences in *lowercase*); the 5' ag and 3' gt consensus sequences for splicing are underlined. The precise sequence at the 5' end of exon 1 is not known; the 5' sequence shown is that of the most 5' primer used in these experiments and corresponds to nucleotides 81–86 in the cDNA sequence (Fig. 1).

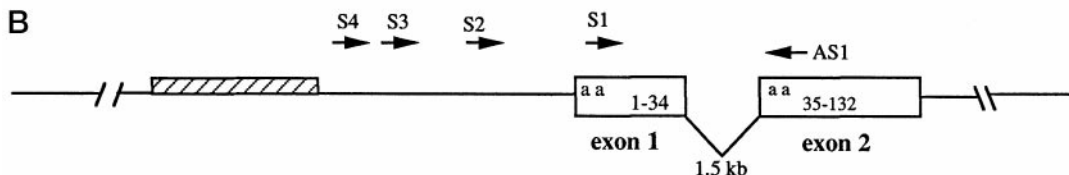
A

```

-2061 ACTCGCGGCG CCCCCCCCCT TGGCCGAGC CAATCAGAGC CCTCCGTCCC
-2011 ATCTCCATGG CGACCGCCGC ATCCTGGAGA GGAAGTACC GTCCCATATT S8
-1961 ATTAGCGCTA TTATTCGTAT TGCTAATTAG TAATAATCAT TTTCTTCTCC S8
-1911 CGGCGGGGGG GGTGGGCGCC CGGGGAGGAC GACAGAGTTT AACCCCTCTG
-1861 TGCCACGACG CACACCGGGG GGGGGTCCTA TGGGTAATA TTGGCCCCC
-1811 CCCAGAGTCC CCCCCCAATA ACCATCCCCT CTCCCCCCCC CCCAAAGGCC
-1761 ACACACGACG ACTGCACCCC CCCAGCCCGT GATGGGGGGG GGGGTCTCTA
-1711 TAGGTCTCTA TTAGCGCGCG CCCCCCCCCA AATAAGGGGG GGGGGAATTT MZF1
-1661 GACGACAGCC GACGATATGG GGGGGGGGCT TTGGGGTTT TGATGATGGG
-1611 GTTTTTATTT GGGGGGGGGG GGGGCGCATT TTTCCCCCCC TCTCTGTCT
-1561 TACCTGCCCC CACCATCCCA GCGCCCGGG CTGAGCGCAC GTCTCGCAGA
-1511 AGAAATCCAC CAGAGGCATT TTGTCGGCTT CGGAGGGGGG GGGGGGGGGG
-1461 CACACAGGAC GCCCTAAAA GCCTGTGTGT GTGTGTGGGG GGGGGGGGGT
-1411 CCCTCTTTCA CCCCCCACC CCTAATAAAT CAATACGACC CACCCTACG
-1361 GTGCCCTACG GCGTTTGGG GGGGGGCTCT GGGCATGTG GGAGCTTTGG
-1311 GGGGGGGGGT TAAATGCTGC TCCCCCCCAT TCTTCCCCC CCCCAAGCGA
-1261 AAAGAAAGAG CTGCTCCAC CTGAGGGGG GTCGCTCAC ATCGCTGCTG E2 BOX
-1211 TGCACACACC CCCACAGCT CTGCTTTGTG GGGGGGGGGG CAGTGGGGGG
-1161 ACCCTGGGGG CATTGAGGGG GGTAAAAGG GAGTGGGGGG AGGCCCTCTA
-1111 TGGGGTTCTA TAGGGTTCTA TGGGGCGCTA TGGGGTGAAG GAGGATGCTA
-1061 TGGGGTTGCT ATGGGGATCT GTAGGGTTGC TATGGGGATC TGTAGGGATG
-1011 CTATGGGGAT CTGTAGGGAT CTATAGGGAT GCTATGGGGA CCTATAGGGA
-961 TGCTATGGGG TCGCTATGGG GATCTATAGG GATCTATAGG GATGCTATGG
-911 GGTCACATAT GGGATCTATA GGGATGCTAT GGGGATCTAT AGGGATCTAT
-861 AGGGATGCTA TGGGGATCTA TAGGGATCTA TAGGGATCTG TAGGGATGCT
-811 ATGGGGTCGC TATGGGGATC TATAGGGATT TATAGGGATG CTATGGGGAC
-761 GCTATGGGGA TCTATAGGGT TGCTATGGGG ATCTATAGGG ATCTACAGGG
-711 TTGCTATGGG GATGCTCTGC GGTGCTATG GGGTTGCTAT GGGGATGCTA
-661 TGGGGATGCT ATAGGGGAGA GGGGGCCCGG GGGGGTCCCG GCGGCGGTGC
-611 GGGGCTGCGG GTGGGTCCGA CGCATGCGTG GGGCTGCGGG TTTTGCCTCC
-561 GCCCCCCCCC AAACCCCGTC CCACAGCTCC GCCACCCCC ATAATCCACC S4
-511 CCCCGCCCCC AACCCCCCCC CAAAAAAGC CCCTCCTGTT CGATCATGGA SP1
-461 AATAGGGGGG AGAGACCCCC ATGGGTCACG CCACAATGGG ATGAGGACAC S3
-411 GGCCCCCGCG CCCACGATGG CAGTGTCCCC ATACCCCAT ACCCCACATC
-361 CATACCACAT ACCCCAATA CCCCTACCC CATAACCATA CCCCACCATT
-311 GCACCCCAT ATTCCTACAT CCCACACCGC GACCCCACT CCCGTTGCTC
-261 CCCACATCCC TTTGCCCCA CAGCCCTCA CTCAATTTTC CGACTCCTGT S2
-211 CCCTTTAAAT CCGCCGCCAT CTTTCTCTCG CCTTCGCTCA CCGGGCGCTC
-161 CATTACGCGT GCGCGCACGC CCCCCCAAC ATCGTCGCAA CAGGAAGTGC
-111 GTCACAAATC GCGACCGGAA GTGGGCGCGC CGTGCGCTGG AGGTTTCCAG
-61 AAGCGGCCGA GAGCGGCGGA CGGGCGCGGG GGGGGCGGTT TAGGAGGGT
-11 GCGGTTCCGC TATGTTGGAT TACAGCGTCT GGGACCACAT CGAGGTGTGC
+40 GACGATGAGG ACGAGACGCA CCCCAACATC GACACCGCCT CCCTCTTCCG S1
+90 ATGGCGGCAC CTGgt...Intron 1(1.5kb)...agCCCCGCGT...ATGC
      CCTGGAACGTCGACGCTTGAGCAAGAACGGCTTCAGCAAGgt...Intron 2 AS1

```

B



C

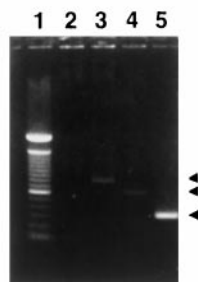


FIG. 6. Putative promoter region of the CDC37 gene. A, nucleotide sequence. The primers (S1–S4 and AS1) used for determining additional transcription start sites are single underlined. The translation initiation codon is in bold and single underlined. Several putative motifs of interest are double underlined: two S8 homeodomain motifs (S8), MZF1-binding motif for the C-terminal zinc finger cluster (MZF1), E2 box, and SP1 motif. The repeating motif, (T/C)gCTAT(A/G)GGG(A/T), is shaded; this motif includes 15 repeats of the MZF1-binding motif for the N-terminal zinc finger cluster: TATGGGGA. B, diagram indicating the relative positions of primers used for determining additional transcription start sites. The hatched box is the repetitive region shaded in A. C, PCR products obtained with the primers diagrammed in B. Lane 1, markers; lanes 2–5, primer AS1 with primers S4 (lane 2), S3 (lane 3), S2 (lane 4), and S1 (lane 5). Arrowheads indicate the single bands present in lanes 3, 4, and 5.

primer S1 (Fig. 6C, lane 5). However, we also obtained weak bands of expected sizes with primers S2 and S3, which are upstream of the major initiation site (Fig. 6C, lanes 3 and 4). No band was obtained by RT-PCR with primer S4 (Fig. 6C, lane 2). The most likely explanation of the latter finding is that an additional, minor transcription start site occurs between primers S3 and S4, *i.e.* at ~250 bp upstream of the major start site (Fig. 6A); however, no additional mRNA corresponding to this has been observed in Northern blots.

**The CDC37 Promoter Region**—As described above, we obtained a sequence of ~2 kb upstream of the translation start codon. This sequence includes a 5'-UTR of ~100 bp and an additional ~1900 bp upstream of the 5'-UTR (Fig. 6A). No TATA or CAAT boxes occur at appropriate positions but there is an SP1 site at residues -513 to -501.

An 11-bp DNA consensus sequence for binding the S8 homeodomain protein, a homeobox gene product expressed in specific regions of mesenchyme in the embryo, has been characterized. The DNA sequence to which the S8 homeodomain binds is AN(C/T)(C/T)AATTA(A/G)C, residues 3–9 being of particular importance (35). Two putative S8 homeodomain recognition sites are present in the CDC37 promoter region at residues -1966 to -1956 and -1941 to -1931. These two sequences are TATTAATTAGC and TGCTAATTAGT, respectively; both contain all critical components of the S8 consensus sequence, including the central nucleotides at positions 3–9 and the ATTA motif, which is the core sequence essential for DNA binding to most homeodomain proteins (36).

An E2 box, CACCTG, is present in CDC37 between residues -1244 and -1239. Several basic helix-loop-helix activator proteins bind to E2 box-containing sequences (37). However, E2 box repressors appear to act competitively by binding to sites that overlap with those of E2 activators and include at least part of the E2 box sequence. For example,  $\delta$ EF1, a zinc finger and homeodomain-containing protein thought to be important in developmental gene regulation, is a repressor of E2 box action (38). The binding site consensus sequence for one of the zinc finger clusters in  $\delta$ EF1 includes CACCT and consensus flanking sequences that include those found in CDC37, **TC-CCACCTGAG** (residues -1247 to -1237; flanking sequences in bold). Thus, the CDC37 promoter could conceivably include a site involved in E2 box activation or repression.

Of particular interest is the region of the CDC37 promoter between residues -1115 and -645, which has approximately 40 repeats of the 10–11-bp consensus sequence, **(T/C)gCTAT(A/G)GGG(A/T)** (bold letters indicate the most common alternative nucleotides). Within 15 of these repeats is the motif, TATGGGGA, which closely conforms to one of the DNA consensus sequences, AGTGGGGA (GGGGA being most critical), recognized by the myeloid zinc finger protein, MZF1 (39). MZF1 is a transcription factor that plays an important role in differentiation of myeloid progenitor cells. MZF1 contains two clusters of zinc fingers that bind independently to two DNA consensus sequences with G-rich cores. The N-terminal cluster binds to the GGGGA sequence (39). CDC37 also contains a *cis* sequence virtually identical to that used for binding the MZF1 C-terminal zinc finger cluster, *i.e.* GGNGAGGGGGAA (39). This second putative MZF1 motif lies between residues -1676 and -1665 of the CDC37 gene and has the sequence, GGGGGG-GGGGAA.

Genetic and biochemical studies have shown that CDC37 forms a complex with Hsp90 and that this complex stabilizes several protein kinases that are critical for signal transduction

and cell division (7–17), events that are central to embryonic development and tissue remodeling. The S8 homeodomain, E2 box, and MZF1 sequences discussed above are all important in regulating gene expression during tissue and organ development. Thus, the presence of putative binding sites for such transcription factors within the promoter region of the CDC37 gene further supports the idea that CDC37 is important in cellular behavior and in morphogenesis. Future promoter function studies will determine which of these sequences are active in regulation of CDC37 expression.

**Acknowledgments**—We thank Alik Grammatikakis for technical assistance, Dr. Marion Gordon for critical help with the manuscript, and Raymund Stefancsik for assistance with data base searches.

## REFERENCES

1. Reed, S. I. (1980) *Genetics* **95**, 561–577
2. Ferguson, J., Ho, J.-Y., Peterson, T. A., and Reed, S. I. (1986) *Nucleic Acids Res.* **14**, 6681–6697
3. Cutforth, T., and Rubin, G. M. (1994) *Cell* **77**, 1027–1036
4. Gerber, M. R., Farrell, A., Deshaies, R. J., Herskowitz, I., and Morgan, D. O. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 4651–4655
5. Grammatikakis, N., Grammatikakis, A., Yoneda, M., Yu, Q., Banerjee, S., and Toole, B. P. (1995) *J. Biol. Chem.* **270**, 16198–16205
6. Ozaki, T., Irie, K., and Sakiyama, S. (1995) *DNA Cell Biol.* **14**, 1017–1023
7. Dai, K., Kobayashi, R., and Beach, D. (1996) *J. Biol. Chem.* **271**, 22030–22034
8. Stepanova, L., Leng, X., Parker, S. B., and Harper, J. W. (1996) *Genes Dev.* **10**, 1491–1502
9. Perdew, G. H., Wiegand, H., Heuvel, J. P. V., Mitchell, C., and Singh, S. S. (1997) *Biochemistry* **36**, 3600–3607
10. Brugge, J. S., Erikson, E., and Erikson, R. L. (1981) *Cell* **25**, 363–372
11. Courtneidge, S. A., and Bishop, J. M. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 7117–7121
12. Brugge, J. S., Yonemoto, W., and Darrow, D. (1983) *Mol. Cell. Biol.* **3**, 9–19
13. Pratt, W. B. (1993) *J. Biol. Chem.* **268**, 21455–21458
14. Stancato, L. F., Chow, Y.-H., Hutchison, K. A., Perdew, G. H., Jove, R., and Pratt, W. B. (1993) *J. Biol. Chem.* **268**, 21711–21716
15. Dey, B., Lightbody, J. J., and Boschelli, F. (1996) *Mol. Biol. Cell* **7**, 1405–1417
16. Kimura, Y., Rutherford, S. L., Miyata, Y., Yahara, I., Freeman, B. C., Yue, L., Morimoto, R. I., and Lindquist, S. (1997) *Genes Dev.* **11**, 1775–1785
17. Schutz, A. R., Giddings, T. H., Steiner, E., and Winey, M. (1997) *J. Cell Biol.* **136**, 969–982
18. Ozaki, T., and Sakiyama, S. (1996) *Biochem. Biophys. Res. Commun.* **226**, 237–241
19. Ozaki, T., and Sakiyama, S. (1996) *DNA Cell Biol.* **15**, 975–979
20. Toole, B. P. (1991) in *Cell Biology of Extracellular Matrix* (Hay, E. D., ed.) 2nd Ed., pp. 305–341, Plenum Press, New York
21. Ishihara, M., Fedarko, N. S., and Conrad, H. E. (1986) *J. Biol. Chem.* **261**, 13575–13580
22. Ripellino, J. A., Margolis, R. U., and Margolis, R. K. (1989) *J. Cell Biol.* **108**, 1899–1907
23. Hiscock, D. R., Yanagishita, M., and Hascall, V. C. (1994) *J. Biol. Chem.* **269**, 4539–4546
24. Ishihara, M., and Conrad, H. E. (1989) *J. Cell. Physiol.* **138**, 467–476
25. Busch, S. J., Martin, G. A., Barnhart, R. L., Mano, M., Cardin, A. D., and Jackson, R. L. (1992) *J. Cell Biol.* **116**, 31–42
26. Weigel, P. H., Hascall, V. C., and Tammi, M. (1997) *J. Biol. Chem.* **272**, 13997–14000
27. Brecht, M., Mayer, U., Schlosser, E., and Prehm, P. (1986) *Biochem. J.* **239**, 445–450
28. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
29. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8998–9002
30. Ochman, H., Gerber, A. S., and Hartl, D. L. (1988) *Genetics* **120**, 621–625
31. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5465
32. Yang, B., Yang, B. L., Savani, R. C., and Turley, E. A. (1994) *EMBO J.* **13**, 286–296
33. Brugge, J. S., and Darrow, D. (1982) *Nature* **295**, 250–253
34. Perdew, G. H., and Whitelaw, M. L. (1991) *J. Biol. Chem.* **266**, 6708–6713
35. DeJong, R., Van der Heijden, J., and Meijlink, F. (1993) *Nucleic Acids Res.* **21**, 4711–4720
36. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell* **63**, 579–590
37. Murre, C., Voronova, A., and Baltimore, D. (1991) *Mol. Cell. Biol.* **11**, 1156–1160
38. Sekido, R., Murai, K., Funahashi, J.-I., Kamachi, Y., Fujisawa-Sehara, A., Nabeshima, Y.-I., and Kondoh, H. (1994) *Mol. Cell. Biol.* **14**, 5692–5700
39. Morris, J. F., Hromas, R., and Rauscher, F. J. (1994) *Mol. Cell. Biol.* **14**, 1786–1795