Genomic organization and transcriptional analysis of the human L-glutaminase gene[1]

Cristina PÉREZ-GÓMEZ[2], José M. MATÉS[2], Pedro M. GÓMEZ-FABRE, Antonio DEL CASTILLO-OLIVARES, Francisco J. ALONSO and Javier MÁRQUEZ[3]

Departamento de Biología Molecular y Bioquímica, Laboratorio de Química de Proteínas, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain

In mammals, glutaminase (GA) is expressed in most tissues, but the regulation of organ-specific expression is largely unknown. Therefore, as an essential step towards studying the regulation of GA expression, the human liver-type GA (hLGA) gene has been characterized. LGA genomic sequences were isolated using the genome walking technique. Analysis and comparison of these sequences with two LGA cDNA clones and the Human Genome Project database, allowed the determination of the genomic organization of the LGA gene. The gene has 18 exons and is approx. 18 kb long. All exon/intron junction sequences conform to the GT/AG rule. Progressive deletion analysis of LGA promoter–luciferase constructs indicated that the core promoter is located between nt -141 and $+410$, with several potential regulatory elements: CAAT, GC, TATA-like, Ras-responsive element binding protein and specificity protein 1 (Sp1) sites. The minimal promoter was mapped within $+107$ and $+410$, where only an Sp1 binding site is present. Mutation experiments suggested that two CAAT recognition elements near the transcription-initiation site (-138 and -87), play a crucial role for optimal promoter activity. Electrophoretic mobility-shift assays confirmed the importance of CAAT- and TATA-like boxes to enhance basal transcription, and demonstrated that HNF-1 motif is a significant distal element for transcriptional regulation of the hLGA gene.

Key words: breast cancer cells, gene regulation, glutamine, hepatoma cancer cells, promoter analysis.

INTRODUCTION

Phosphate-activated glutaminase (GA; E.C. 3.5.1.2) initiates glutamine catabolism by hydrolysis of this amino acid to form glutamate and ammonium. In mammals, two isoenzymes were originally described as liver-type (LGA) and brain-type [1], although this last isoform was lately named kidney-type (KGA) enzyme [2]. Some of the main physiological functions of GA include renal ammoniagenesis, nitrogen supply for urea biosynthesis in the liver, synthesis of the excitatory neurotransmitter glutamate in the brain, and energy supply for the bioenergetics of many normal and transformed cell types [3,4].

The KGA isoform seems to be ubiquitous in all tissues with GA activity, with the exception of postnatal liver [3]; however, expression in liver endothelial cells has also been described [5]. In kidney, the enzyme is strongly induced by metabolic acidosis: a post-transcriptional mechanism accounts for its enhanced expression due to increased stability of the KGA mRNA [6]. The KGA gene has been shown to elicit many mRNA transcripts, employing multiple polyadenylation sites [7] or alternative splicing [8]. The alternatively spliced mRNA encodes a protein with a distinct C-terminal region to that of KGA; thus, it represents a new GA isoform, named GAC, first described in human kidney and colon cancer cells [8]. Recently, GAC mRNA has been shown to be expressed also in rat kidney and pig renal cells [9]. The rat (r)KGA promoter has been characterized; it lacks a TATA sequence but is GC-rich and contains two CAAT boxes and two specificity protein 1 (Sp1) sites [10].

The LGA isoenzyme was originally thought to be present only in adult liver tissue [3,11]. However, emerging evidence has now indicated clearly that expression also occurs in extrahepatic tissues like brain, pancreas and breast cancer cells [12]. Human (h)LGA has been extensively characterized in our laboratory. We first isolated a cDNA clone from ZR-75 breast cancer cells [12] and, surprisingly, the deduced amino acid sequence was highly similar to the rat liver enzyme, but with the addition of 67 extra amino acids at the N-terminal [12,13]. Two loci were then identified for human GA: the human (h)KGA gene was mapped to chromosome 2, whereas a second locus in chromosome 12 was assigned to hLGA [14,15]. The proximal promoter of the rat (r)LGA gene also lacks a functional TATA box, but contains recognition elements HNF-1, HNF-5 and CAAT-enhancer binding protein (C/EBP) that may be important for its basal expression in liver [13]. Increased expression of hepatic GA has been reported in diabetes, starvation, or on feeding a high-protein diet. An enhanced rate of gene transcription was shown to be the mechanism responsible for these adaptive changes [11,13].

Abbreviations used: C/EBP, CAAT-enhancer binding protein; CREB, cAMP-responsive element binding protein; EMSA, electrophoretic mobility-shift assay; ER, oestrogen receptor; GA, phosphate-activated glutaminase; HLF, hepatic leukaemia factor; KGA, kidney-type glutaminase; hKGA, human kidney-type glutaminase; rKGA, rat kidney-type glutaminase; LGA, liver-type glutaminase; hLGA, human liver-type glutaminase; rLGA, rat liver-type glutaminase; PDZ, PSD95/Dlg/ZO1 domains; RREB, Ras-responsive element binding protein; RT, reverse transcription; Sp1, specificity protein 1; UTR, untranslated region.

[1] This paper is dedicated to Professor I. Núñez de Castro on the occasion of his retirement. We are indebted to him for his example as a scientist and as a colleague.

[2] Both authors contributed equally to this work.

[3] To whom correspondence should be addressed (e-mail marquez@uma.es).

The nucleotide sequence data reported in this paper have been submitted to the DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under the accession number AF348119.

GA is thought to play important roles in tumour cell biology and its activity has been associated with malignancy and tumour growth rates [4]. Overexpression of GA is a phenotype characteristic shown by many experimental and human tumours [12,16]. Elucidation of the role of GA in the control of cell growth and proliferation, and of the intriguing question as to why two isoenzymes are expressed in the same organ, will require the characterization of GA gene expression and tissue-specific regulation. Some clues may be obtained by analysis of the transcriptional activity of GA genes. In this study we sought to elucidate the gene structure and molecular requirements of the basal transcriptional machinery controlling hLGA gene expression. We found that a core glutaminase promoter is located between nt -141 and $+410$ within a GC-rich region, and four elements were essential for optimal promoter activity: a TATA-like box, two CAAT elements, and an HNF-1 binding site. Furthermore, an Sp1 site at the 5′-UTR may account for a strong basal expression in hepatoma and breast cancer cells.

EXPERIMENTAL PROCEDURES

Cloning and sequencing

To clone the 5′-flanking region of the hLGA gene, we used the Genome Walker kit (Clontech, Palo Alto, CA, U.S.A.) according to manufacturer's instructions and employing glutaminase-specific primers deduced from the 5′ end of a cDNA LGA clone (GenBank accession number AF110330, [12]). PCR products were subcloned into pGEMT Easy vector (Promega, Madison, WI, U.S.A.) and sequenced using the M13 forward and reverse primers. By searching the Human Genome Project database with two putative intronic sequences present in another cDNA LGA clone (GenBank accession number AF110329, [12]), we identified a match with a Human Genome Project working draft sequence (accession number AC024884.00003) located on chromosome 12. Arrangement of this sequence using the cDNA LGA clones produced a genomic structure extending from exons 4 to 18. Another BLAST search was done using the 5′-flanking sequence cloned with the Genome Walker System: a significant alignment was obtained with a second working draft sequence of the Sanger Center Human Genome Project database (accession number AC024884.00008). Sequence analyses of this contig with cDNA and genomic LGA clones showed that it encompassed the first exon, part of the first intron and 2705 bp of 5′-untranslated region (UTR). However, the cDNA sequence between the second and fourth exon of clone AF110330 did not align with any sequence from the Human Genome Project database. Primers were then designed to obtain the genomic sequence between exons 2 and 4 and to amplify contigs comprising the whole LGA gene (Table 1). PCR was performed on 2 μl (400 ng) of genomic DNA isolated from ZR-75 cells [12] using specific pairs of primers for the LGA gene (Table 1). PCRs were hot started at 94 °C for 5 min and then continued for 30 cycles with denaturation for 30 s at 94 °C, annealing for 90 s at 55 °C and extension for 2 min at 72 °C, with a final extension for 10 min at 72 °C. To obtain the insert between introns 7 and 9, the annealing temperature was established at 52 °C.

Subcloned insert DNAs were sequenced by the dideoxy chain-termination method [17] using an LKB-ALF DNA sequenator (Amersham Biosciences) and the Thermo Sequenase dye-primer cycle kit (Amersham Biosciences). Sequencing was also performed using the ABIPRISM automated DNA Sequencer model 310 (Applied Biosystem) at the sequencing facility of the Centro de Investigaciones Biológicas (Madrid, Spain).

Primer extension analysis

To map the transcription start site we first did three different reverse transcription (RT)-PCRs, using total RNA of ZR-75 human breast cancer cell, essentially as described previously [12] with minor modifications. Total RNA (10 μg) was reverse transcribed with the RT primer, then, we amplified this cDNA for 40 cycles with three different sense primers: Prom20SacI, Prom19SacI and Prom18SacI, and oligonucleotide GSP2 9 as the common anti-sense primer (Table 1). Primer extension reactions were performed to localize the transcription start site using 0.5 and 2 μg of poly(A^+)mRNA prepared from ZR-75 cells [12]. Annealing and labelling of the reverse primers PEXT48 and PEXT42 (Table 1) and extension reactions were performed as described previously [12]. ΦX174 DNA/*Hin*fI dephosphorylated markers (Promega) were radiolabelled with [γ-^{32}P]dATP just before use. The samples were denatured for 10 min at 95 °C before loading on denaturing polyacrylamide analytical gels containing 8 % (w/v) polyacrylamide, 7 M urea and 1 × TBE (Tris/borate/EDTA) buffer. The gel was then vacuum dried and autoradiographed for 3–4 days at -80 °C.

Generation of glutaminase promoter–luciferase constructs

Genomic DNA was prepared from the human breast cancer cell line ZR-75 as described previously [12]. The high-fidelity cDNA Advantage polymerase mix (Clontech) was used for PCR. Two overlapping fragments of the hLGA promoter, ranging in size from 1.4 to 2.8 kb, were amplified by PCR using genomic DNA and a primer (PROM3R, Table 1) specific for the 5′ end of hLGA cDNA sequence [12], in combination with two upstream sense primers (PROMD1 and PROMD2, Table 1). Both products contain a segment at their 3′ end that matches the 5′-UTR sequence of the previous cDNA hLGA clone [12], but adding 189 new nucleotides at the 5′ end. A fragment of 2828 bp of the GA gene promoter was subcloned into pGEMT using primers PROMD1 and PROM3R (Table 1). Promoter DNA sequence was finally cloned at the *Sac*I/*Nhe*I sites of the pGL3 luciferase reporter vector (Promega, Madison, WI, U.S.A.). The nucleotide upstream of the transcription start site was numbered -1. A series of reporter constructs containing 5′ deletion mutations were prepared by PCR using specific 5′ primers and a common 3′ primer named Prom3R*Nhe*I (Table 1). The orientation of the inserts was verified by restriction digestion and/or DNA sequencing.

Mutation analysis of the glutaminase minimal promoter

Mutations of the constructs were performed using the Quick Change site-directed mutagenesis kit (Stratagene Europe, Amsterdam, The Netherlands) and the appropriate HPLC-purified mutated primer pairs. The presence of mutations was verified by DNA sequencing. Reporter luciferase vectors containing mutations were chosen for large-scale DNA preparation and used in transfection experiments. Oligonucleotides containing mutations were also annealed and used in electrophoretic mobility-shift assays (EMSAs) as competitors to verify the effect of mutations.

Cell culture and transient transfections

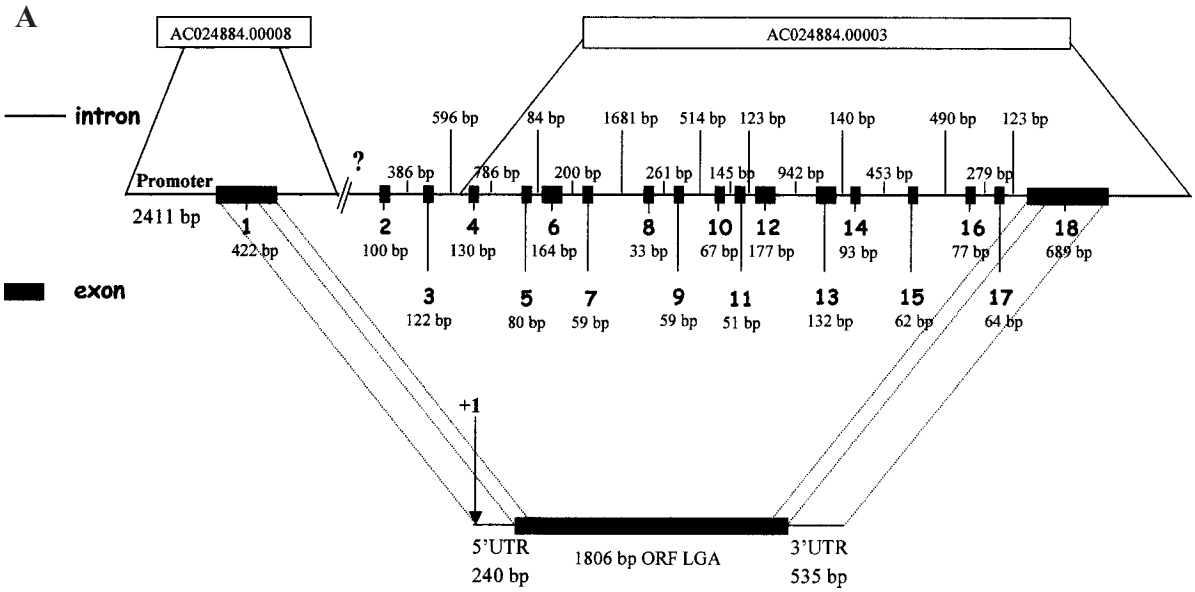
Human hepatoblastoma cells (HepG2, European Collection of Cell Cultures, Cambridge, U.K.) were maintained in Eagle's minimum essential medium with 1 % non-essential amino acids and 1 mM sodium pyruvate (BioWhittaker, Walkersville, MD,

Table 1 Oligonucleotides used for human L-glutaminase gene isolation, chimeric plasmid constructions, and primer extension analysis

Name	Sequence 5′→3′	nt
Genome Walker		GenBank sequence
GSP1 7 reverse	GCAGTCCAGTGGCCTTTAGTGC	7632/7653
GSP2 9 reverse	AGCAAATCACCCAGGCGGGACAG	7168/7190
AP1	GTAATACGACTCACTATAGGGC	Vector
AP2	ACTATAGGGCACGCGTGGT	Vector
Promoter isolation		Promoter sequence
PROMD1	GCCCTTGTCATAGTTGGCTAATAGT	−2411/−2387
PROMD2	CCTTATGCGAGAATTCTTAATTTCC	−1021/−997
PROM3R	TGGTGCTGCGGCTGGTGGCTGTGTGGCGTCTCTCT	382/416
Full LGA isolation		GenBank sequence
1E1I F	TCCATGAAGGCTCTGCAGAA	2761/2780
1E1I R	ACACAGGATTATAGTGATCAGA	5216/5237
2E4E F	CATCAGAAAGTGGCATGCTGTCC	7151/7173
2E4E R	GGTCAGGAGCACAATGTTGCTGCTC	8358/8382
4E7I F	CCAATCGTGTCTGGGATCCAGG	8331/8352
4E7I R	GCTAGTCTGACACCCTGTTCAAG	10312/10334
7I9I F	AGTTATTCACGTCACCTGGAAC	10279/10300
7I9I R	CATCTAGGAAAACTGGTGAAG	12177/12197
9I12I F	GGACTGACTGAAAGCATTGTGG	11927/11948
9I12I R	GGAAATGACATCTAAGCTGAGAC	13512/13534
U1D	CGGGGTGTGGATCCTGAATTGTGCT	13461/13485
U2R	CCGTGGGTCTAACTTCCGAGCAC	14262/14284
5F	GTGCTCGGAAGTTAGACCCACGG	14262/14284
0R	GGGCAAGCCATTAGGCTGTACC	16288/16309
18E18I F	ACCAAGCAGTCATTTGGTGAC	16152/16172
18E18I R	GATGTGAGCCAGGTCTCTGTGGT	18062/18084
Promoter constructions		Promoter sequence
PromD1SacI	GCCCTTGTCATAGTGAGCTCATAGT	−2411/−2387
Prom2SacI	CAAGTCACGAGCTCATGCTCCATC	−2246/−2223
Prom3SacI	AGTTAAGTCCAGAGCTCCAAGTG	−2189/−2168
Prom4SacI	CTCTAAGCAAAGCGCTCCCTGCTTCC	−2152/−2128
Prom5SacI	GAGAGTAAGGAGCTCATTAGA	−2067/−2047
Prom6SacI	GCTAATTCAAAGAGCTCGAAATAGG	−2039/−2015
Prom7SacI	CCTAACAGATGGTATTGAGCTCTTG	−1971/−1947
Prom8SacI	CTATGAAACGAGCTCAAGATTCGTC	−1913/−1889
Prom9SacI	GAGACCCAGAGCTCTGCCTCTTGG	−1839/−1816
Prom10SacI	GCCTGTAATCCGAGCTCTTGTG	−1525/−1504
Prom11SacI	GTGATACAGCCAGAGCTCTCACA	−1233/−1211
Prom12SacI	CTGGTTCTGGAGCTCACTAGCTG	−1049/−1027
Prom13SacI	GGAACACCATGAGCTCTCAATA	−912/−891
Prom14SacI	GGTGAAAGTTCGTGCTCCAAATG	−822/−800
Prom15SacI	GTGGGGAGCTCTTCATGGT	−698/−680
Prom16SacI	CTCCTGAACTGAGCTCACTCT	−376/−356
Prom17SacI	GTCCCACCGTCGAGCTCTCA	−157/−138
Prom18SacI	CAGTGATTGGAGCTCGGCATTG	−93/−72
Prom19SacI	CTGAGGGAGCTCCTAACCCTGGA	24/76
Prom20SacI	CACCCACCGAGCTCACATTCTC	94/115
Prom3RNheI	GCGGCTGCTAGCTGTGTGGCGTCTC	385/409
Primer extension		Promoter sequence
RT	CCGTGGGTCTAACTTCCGAGCAC	14262/14284
PEXT48	CCTGCCCCTTAAAGCTGGCTCCA	43/66
PEXT42	CCTTTAAAGCTGGCTCCAGGGTTA	37/60

U.S.A.), supplemented with 2 mM L-glutamine, 10 % (v/v) foetal calf serum (Harlan, Indianapolis, IN, U.S.A.), 100 units/ml penicillin, 100 μg/ml streptomycin and 2.5 μg/ml amphotericin (antibiotics were from Roche). Human breast cancer cell line MCF-7 (American Type Culture Collection, Rockville, MD, U.S.A.) were maintained in Dulbecco's modified medium with 4.5 g/litre glucose, supplemented with 10 % (v/v) foetal calf serum, 100 units/ml penicillin, 100 μg/ml streptomycin and 2.5 μg/ml amphotericin. ZR-75 breast cancer cell line was cultured as described previously [12].

Approx. 265000 cells were grown overnight at 37 °C under air/CO_2 (19:1), on 35-mm culture dishes. Before transfection, all assayed cell types were washed twice with PBS and pre-incubated under serum-free conditions for 20–30 min. Cells were transfected with 5 μl of Lipofectin reagent (Life Technologies, Paisley, Renfrewshire, Scotland, U.K.). Aliquots of 1.5 μg of pGL3-1 (the longest luciferase construct) and equimolar amounts for the other plasmids were used. HepG2 and MCF-7 cells were co-transfected with 50 ng of pCMV-βGal (a plasmid containing the human cytomegalovirus promoter upstream the β-galactosidase gene) to normalize for transfection efficiencies, and variable amounts of pBlueScriptII SK (−). After 7 h of incubation at 37 °C, the transfection solution was withdrawn and replaced with the complete mediums described above, and cultivated for



B

```
hLGA    --------------MRSMKALQKALSRAG-------------------------SHCGR 20
rLGA    ------------------------------------------------------------
hKGA    MMRLRGSGMLRDLLLRSPAGVSATLRRAQPLVTLCRRPRGGGRPAAGPAAAARLHPWWGG 60
                     :  . . :   :                                . .

hLGA    GGWGHP------SRSPL-----LGGGVRHH---LSEAAAQGRETPHSHQPQHQDHDSSE- 65
rLGA    ------------------------------------------------------------
hKGA    GGWPAEPLARGLSSSPSEILQELGKGSTHPQPGVSPPAAPAAPGPKDGPGETDAFGNSEG 120
        ..          : :.       . .        : .:: .   . .   . .  ..:.

hLGA    ------------SGMLSRLGDLLFYTIAEGQERIPIHKFTTALKATGLQTSDPRLRDCMS 113
rLGA    --------------MLPRLGDLLFYTIAEGQERIPIHKFTTALKATGLQTSDPRLQDCMS 46
hKGA    KELVASGENKIKQGLLPSLEDLLFYTIAEGQEKIPVHKFITALKSTGLRTSDPRLKECMD 180
                  ..:*. * ***********:**:*** ****:***:******::**.

hLGA    EMHRVVQESSSGGLLDRDLFRKCVSSNIVLLTQAFRKKFVIPDFEEFTGHVDRIFEDVKE 173
rLGA    KMQRMVQESSSGGLLDRELFQKCVSSNIVLLTQAFRKKFVIPDFEEFTGHVDRIFEDAKE 106
hKGA    MLRLTLQTTSDGVMLDKDLFKKCVQSNIVLLTQAFRRKLVIPDFMSFTSHIDELYESAKK 240
         ::  :* :*.* :**::**:***.***********:*:***** .**.*:*.::*..*:

hLGA    LTGGKVAAYIPQLAKSNPDLWGVSLCTVDGQRHSVGHTKIPFCLQSCVKPLTYAISISTL 233
rLGA    LTGGKVAAYIPHLAKSNPDLWGVSLCTVDGQRHSVGHTKIPFCLQSCVKPLTYAISVSTL 166
hKGA    QSGGKVADYIPQLAKFSPDLWGVSVCTVDGQRHSTGDTKVPFCLQSCVKPLKYAIAVNDL 300
         :***** ***:*** .*******:*********.*.*.**:***********.***::. *

hLGA    GTDYVHKFVGKEPSGLRYNKLSLNEEGIPHNPMVNAGAIVVSSLIKMDCNKAEKFDFVLQ 293
rLGA    GTDYVHKFVGKEPSGLRYNKLSLNEEGIPHNPMVNAGAIVVSSLIKMDCNKAEKFDFVLQ 226
hKGA    GTEYVHRYVGKEPSGLRFNKLFLNEDDKPHNPMVNAGAIVVTSLIKQGVNNAEKFDYVMQ 360
        **:***::**********:*** ***:. ************:****. *:*****:*:*

hLGA    YLNKMAGNEYMGFSNATFQSEKETGDRNYAIGYYLKEKKCFPKGVDMMAALDLYFQLCSV 353
rLGA    YLNKMAGNEFMGFSNATFQSEKETGDRNYAIGYYLKEKKCFPNPVDMMAALDLYFQLCSV 286
hKGA    FLNKMAGNEYVGFSNATFQSERESGDRNFAIGYYLKEKKCFPEGTDMVGILDFYFQLCSI 420
        :*********::**********:*:****:************: .**:. **:*******:

hLGA    EVTCESGSVMAATLANGGICPITGESVLSAEAVRNTLSLMHSCGMYDFSGQPAFHVGLPA 413
rLGA    EVTCESGSVMAATLANGGICPITGESVLSAEAVRNTLSLMHSCGMYDFSGQFAFHVGLPA 346
hKGA    EVTCESASVMAATLANGGFCPITGERVLSPEAVRNTLSLMHSCGMYDFSGQFAFHVGLPA 480
        ******.***********:****** ***.****************************

hLGA    KSAVSGAILLVVPNVMGMMCLSPPLDKLGNSHRGTSFCQKLVSLFNFHNYDNLRHCARKL 473
rLGA    KSAVSGAILLVVPNVMGMMCLSPPLDKLGNSHRGISFCQKLVSLFNFHNYDNLRHCARKL 406
hKGA    KSGVAGGILLVVPNVMGMMCWSPPLDKMGNSVKGIHFCHDLVSLCNFHNYDNLRHFAKKL 540
        **.*:*.************* ******:*** :*  **:.**** ********** *:**

hLGA    DPRREGAEIRNKTVVNLLFAAYSGDVSALRRFALSAMDMEQKDYDSRTALHVAAAEGHIE 533
rLGA    DPRREGGEVRNKTVVNLLFAAYSGDVSALRRFALSAVDMEQKDYDSRTALHVAAAEGHID 466
hKGA    DPRREGGDQRVKSVINLLFAAYTGDVSALRRFALSAMDMEQRDYDSRTALHVAAAEGHVE 600
        ******.: *.*:*:********:**************:****:****************::

hLGA    VVKFLIEACKVNPFAKDRWGNIPLDDAVQFNHLEVVKLLQDYQDSYTLSETQAEAAAEAL 593
rLGA    VVKFLIEACKVNPFVKDRWGNIPLDDAVQFNHLEVVKLLQDYHDSYMLSETQAEVAAETL 526
hKGA    VVKFLLEACKVNPFPKDRWNNTPMDEALHFGHHDVFKILQEYQVQYTPQGDSDNGKENQT 660
        ******:******** ****.* *:*:*::*.* :*.*:**:*: .*  .  . :   :

hLGA    SKENLESMV 602
rLGA    SKENLESMV 535
hKGA    VHKNLDGLL 669
         ::**:.::
```

Figure 1 Organization of hLGA gene

(**A**) Structure of the hLGA gene. Exons are indicated by bold numbers and depicted as black boxes. Below each box, the fragment size is shown (bp). Intron sequences are depicted as black horizontal lines and their lengths are also shown. At the bottom of the Figure, the 5′-flanking region, 5′-UTR, 3′-UTR and open reading frame (ORF) sequences of the hLGA gene are displayed. (**B**) Comparison of the exon/intron organization of the hLGA gene with those of rLGA (GenBank numbers J05499, AC109891.1, AC095733.3 and AC111940.2) and hKGA [9] genes. Arrowheads indicate the positions at which introns interrupt exons.

Table 2 Exon/intron organization of the human L-glutaminase gene

Exon sequences and intron sequences are in uppercase and lowercase letters respectively.

Exon	5′ splice donor	3′ splice acceptor	Amino acid interrupted
1	CAGGATCAgtaaggctcc Q D H	gcccccaaacagTGATTCATC D S S	H61
2	TTCACCACTgtaagttgccc F T T	tttcctctaaagGCACTAAAG A L K	T94
3	TTCCGAAAgtgagggccc F R K	tctgggatccagGTGTGTGAGC C V S	K135
4	GGAGGCAAAgtgagagtcagg G G K	actcatgtccagGTGGCAGCC V A A	K178
5	GGTCAACGgtgagatgctgg G Q R	gtctgcctccagGCACTCTGTG H S V	R205
6	AATGAGGAAGgtgagcactacc N E E G	tatttccccagGAATCCCCCAT I P H	G260
7	CTGATCAAGgtcagtgccacc L I K	tttctctgcagATGGACTGT M D C	K279
8	TTTGATTTTgtaagttccctt F D F	ttctcttccagGTGTTGCAG V L Q	F290
9	AGCAATGCCACgtaagtcccta S N A T	gtttgcttcagATTCCAGTCA F Q S	T310
10	GAAAAGAAGgtaaccagaag E K K	tcttccttgcagTGCTTTCCT C F P	K332
11	TACTTCCAGgtaagaaaacac Y F Q	tgttttcatagCTGTGTTCT L C S	Q349
12	GCCTTCCACgtgagtttctgg A F H	ttctgttccccagGTGGGCCTG V G L	H408
13	TTCTGCCAGgtgagatgctttg F C Q	ttacaattctagAAGTTGGTG K L V	Q452
14	GAAATTCGGgtaaggaaaact E I R	atttgtgttccagAACAAGACT N K T	R483
15	CTTCGAAGgtaatgctactgt L R R	ccactaccctcagGTTTGCCTTG F A L	R504
16	GCAGCTGAAGgtactgaaggg A A E G	ccatattcctagGACACATCGAA H I E	G530
17	AAGGACAGgtgaggaatag K D R	tattcctctcagGTGGGGCAAC W G N	R551

an additional period of 48 h at 37 °C. Transfections were performed in duplicate, and repeated at least twice.

Luciferase assay

Luciferase and β-galactosidase assays were performed with a kit from Tropix (Bedford, MA, U.S.A.), according to the manufacturer's protocol. Briefly, cells were washed twice with PBS, solubilized with 100 μl of lysis buffer supplemented with 0.5 mM dithiothreitol and detached with a cell scraper. Luciferase activity was determined in duplicate using 10 μl aliquots of supernatants (kept at −80 °C) in a Lumat Luminometer (LB9501, Berthold, Bundoora, Australia), and reported as light units per mg of protein. Alternatively, luciferase activity was divided by β-galactosidase activity to correct for transfection efficiency and the reporter gene expression was also expressed as relative luciferase activity. As a positive control, cells transfected with pGL3 basic, without any promoter sequence, and with pCMV-βGal and pBlueScriptII SK (−) were used. Luciferase activity of untransfected cells was determined as a negative control.

Nuclear extracts and EMSA

Approximately 1.5×10^8 cells were trypsinized, collected and washed once in 5 ml of ice-cold PBS. The cell pellet was resuspended in a 50 ml Falcon tube for 30 min at 4 °C with 40 μl/10^6 cells of lysis buffer containing 0.6 % Nonidet P40, 0.15 M NaCl, 10 mM Tris/HCl, pH 7.9, 1 mM EDTA, 0.5 mM PMSF, 0.1 mM benzamidine, and 1 μg/ml leupeptin. Nuclei were pelleted at 2000 ***g*** for 5 min and washed once with the same buffer but without detergent. Nuclear proteins were extracted from the final pellet with one volume (packed volume of cell pellet) of high salt extraction buffer [10 mM Hepes, pH 7.9, 0.42 M NaCl, 0.1 mM EDTA, 0.1 mM EGTA, 1.5 mM $MgCl_2$, 0.5 mM dithiothreitol, 25 % (v/v) glycerol, 0.5 mM PMSF, 0.1 mM benzamidine, and 1 μg/ml leupeptin] for 30 min at 4 °C with occasional shaking. Nuclear debris was removed by centrifugation at 10000 ***g*** for 10 min at 4 °C. The protein concentration was estimated by the Bio-Rad Bradford reagent kit and was set at 3 μg/μl in all preparations. The nuclear proteins were kept at −80 °C and used in band shift assays.

DNA-binding reactions were performed in 10 mM Tris/HCl, pH 7.6, 0.1 mM EDTA, 50 mM NaCl, 0.5 mM dithiothreitol, 10 % (v/v) glycerol, 0.2 % Nonidet P40 and 1 μg of poly[d(I–C)], using 5 μg of nuclear protein and the [γ-^{32}P]ATP-labelled probe (Amersham Biosciences, 200000 c.p.m./reaction) in a final volume of 20 μl. All probes were adjusted to the same specific activity. Competing unlabelled oligonucleotides were added to the binding mixtures along with the radiolabelled oligonucleotide. A 100-fold molar excess of non-radiolabelled mutated competitor DNA was used. For the competition assays the probe and competitor were added at the same time. Samples were incubated on ice for 30 min and subjected to electrophoresis by non-denaturing PAGE (4.5 % gel) in TBE buffer (100 mM Tris, pH 8.0/100 mM boric acid/1 mM EDTA). Electrophoresis was carried out using a refrigerated system at 6 V/cm for 2 h. Gels were vacuum-dried and autoradiographed overnight at −80 °C.

For supershift analyses, reactions were performed in the presence or absence of 2 μl of the antibody of interest [anti-(NF-1), anti-TFIID (TBP), and anti-(C/EBP β)], from Santa Cruz Biotechnology, La Jolla, CA, U.S.A.). For all CAAT constructs and the TATA-like box in construct pGL3-19, nuclear extracts were pre-incubated with the antibody for 20 min at 20 °C before addition of the radiolabelled probe (200000 c.p.m.) and continued incubation for 25 min at 20 °C. For NF1 and TATA boxes included at pGL3-15, the radiolabelled probe was incubated with the nuclear extract before adding the antibody. Samples were analysed by non-denaturing PAGE as above.

Preparation of single-stranded, G-overhang and double-stranded oligomers

DNA oligonucleotides (see Figure 6A for oligonucleotides sequences) were purchased from Amersham Biosciences and gel-purified before 5′ end-labelling. Each oligonucleotide was separated from residual species by purification by native PAGE (16 % gel), and analysis on a fluorescent TLC plate (Merck, Darmstadt, Germany). After elution and purification of DNA fragments, spectrophotometric quantification was carried out. Double-stranded oligonucleotides were prepared by annealing. Partial duplex 3′G rich overhang DNA was obtained by mixing equimolar amounts of forward and reverse oligonucleotides, followed by heating at 95 °C and slow cooling at 20 °C. For annealing, 1 nmol of each oligonucleotide was used in arrays covering promoter regions and corresponding to the DNA-binding motifs. Annealed oligonucleotides (40 pmol) were end-labelled for 10 min at 37 °C using 1 μl of T4 polynucleotide kinase, 2 μl of [γ-^{32}P]ATP (> 5000 Ci/mmol) and the appropriate kinase buffer. Reactions were stopped by adding

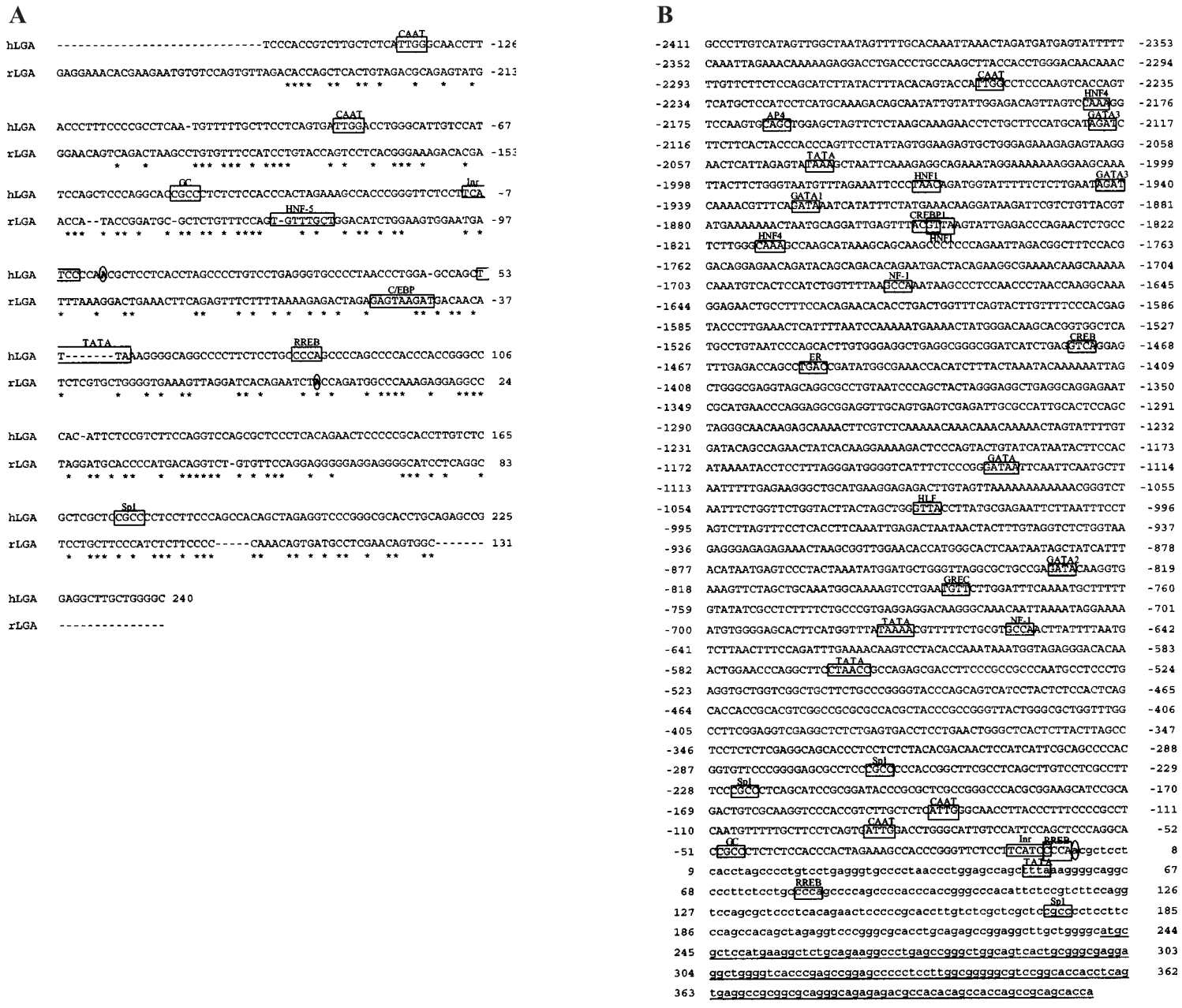


Figure 2 For legend, see facing page

2 μl of 0.5 M EDTA. The fully double-stranded DNAs were quantified in a beta counter.

RESULTS

Genomic organization of the hLGA gene

Recently the gene coding for hLGA was mapped to chromosome 12 [14]. Figure 1(A) summarizes the genomic organization of the hLGA gene; the relevant sequence information is available under GenBank accession number AF348119. By a combination of genomic walking, PCR cloning and restriction analysis, the hLGA gene was determined to span more than 18.1 kb and consists of 18 exons and 17 introns. DNA exonic coding sequences perfectly matched the previous cDNA sequence cloned from breast cancer cells [12], except at six positions, namely nt 311, 470, 820, 1034, 1036, and 1167, which result in amino acid discrepancies at residues 87, 140, 257, 328, and 329. The exon/intron organization of the hLGA gene was determined using the sequence of two human cDNA LGA clones [12]. Exon/intron boundaries fit consensus motifs; each intron beginning with GT at the 5′-splice donor site and ending with an AG at the 3′-splice acceptor site (Table 2). The initial exon contains 422 bp, including 240 bp of 5′-UTR and the sequence encoding the initial 61 amino acids. Exon 18 encodes the last 51 amino acids and has 535 bp of 3′-UTR including a polyadenylation signal. Four additional poly(A) signals are present in the 1.5 kb fragment following exon 18.

It is interesting to note that three *Alu* sequences are located on introns 7, 14, and on the 3′ flanking sequence of the gene after exon 18. Intron 1 is the largest one with more than 4.2 kb and a fragment of unknown sequence, which was shown to be highly refractory to sequencing. Figure 1(B) compares the length of exonic coding sequences, translated to amino acids, in the hLGA gene, the human KGA transcript [9], and those deduced for the rat (r)LGA gene employing the rat genome project draft sequences and the cDNA of rat liver enzyme [13]. Out of 17 boundaries, 16 are identical in the three genes and, furthermore, the amino acids interrupted are almost identical. The lengths of exons 3–17 within the hLGA and hKGA transcripts, and exons 2–16 for the rLGA transcript, are the same. The nucleotide and amino acid sequences of these exons have identities ranging from 69 % to 95 %. In sharp contrast, the N- and C-terminal sequences encoded by exons 1 and 18 of the hLGA gene have only 62.5 % and 29.4 % similarity with hKGA protein, respectively. Furthermore, exon 1 and the first 6 amino acids coded by exon 2 are missing in the rLGA gene, accounting for a rat liver protein 67 amino acids shorter than hLGA.

Isolation of the hLGA 5′-flanking sequence and determination of the transcription start site

A combined strategy of PCR and restriction digestion was used to identify and clone 2827 bp of the hLGA 5′-flanking sequence, including the translation initiation codon. The 5′-UTR is mainly composed of GC-rich domains that are potential binding sites for Sp-1, but also for the C/EBP. Alignment of the hLGA sequence with approximately 1 kb of 5′-flanking sequence currently available for the rLGA gene [13], showed no significant similarity around the transcription start site (Figure 2A).

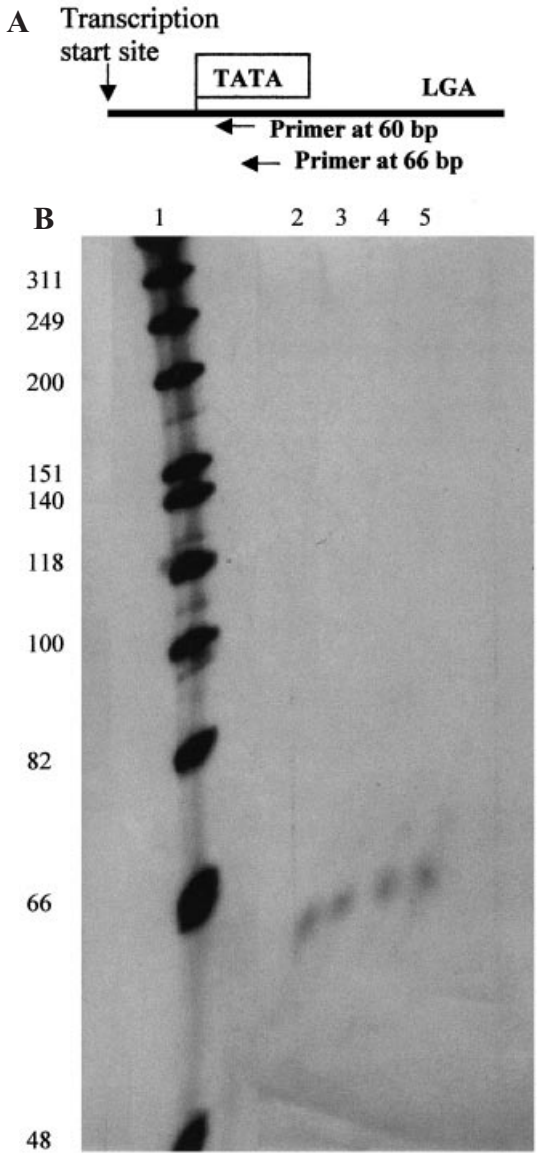


Figure 3 Determination of the hLGA transcription start site

(**A**) Schematic diagram of the transcription start site and positions of the reverse primers used in primer-extension analysis. (**B**) Primer extension of 0.5 and 2 μg of poly(A$^+$) mRNA from ZR-75 cells. Duplicate extension products of 60 bp (lanes 2 and 3) were obtained by using the PEXT42 reverse primer. Duplicate extension products of 66 bp (lanes 4 and 5) were obtained by using the PEXT48 reverse primer. Both primers located downstream of the TATA box (see **A**). ϕX174 DNA/*Hin*fI dephosphorylated markers, previously radiolabelled and denatured, are indicated in lane 1. The transcription start site for hLGA can thus be assigned to the adenine nucleotide situated 240 bp upstream from the initiation codon.

Figure 2 Organization of the hLGA promoter

(**A**) Nucleotide sequence and binding motifs of the proximal 5′-flanking region of the hLGA and rLGA genes. Transcription start sites are indicated by circled and bold letters. Putative protein binding sites, determined by computer analysis, are identified by boxes and are named above the sequence. Numbering is based on the distance from the transcription start sites at A^{+1}. The sequence from nt −141 to nt +256 for hLGA is represented. The sequence of 404 bp at the 5′-flanking region of the rat LGA gene (GenBank number L76175) is also shown. (**B**) Nucleotide sequence of the hLGA gene promoter region from nt −2411 to +416. Numbering of nucleotides is based on the transcription start site. The 5′-UTR is shown in lower case letters and the coding region, beginning at nt +241, is underlined. Actual CAAT protein binding sites determined by EMSA are in bold boxes.

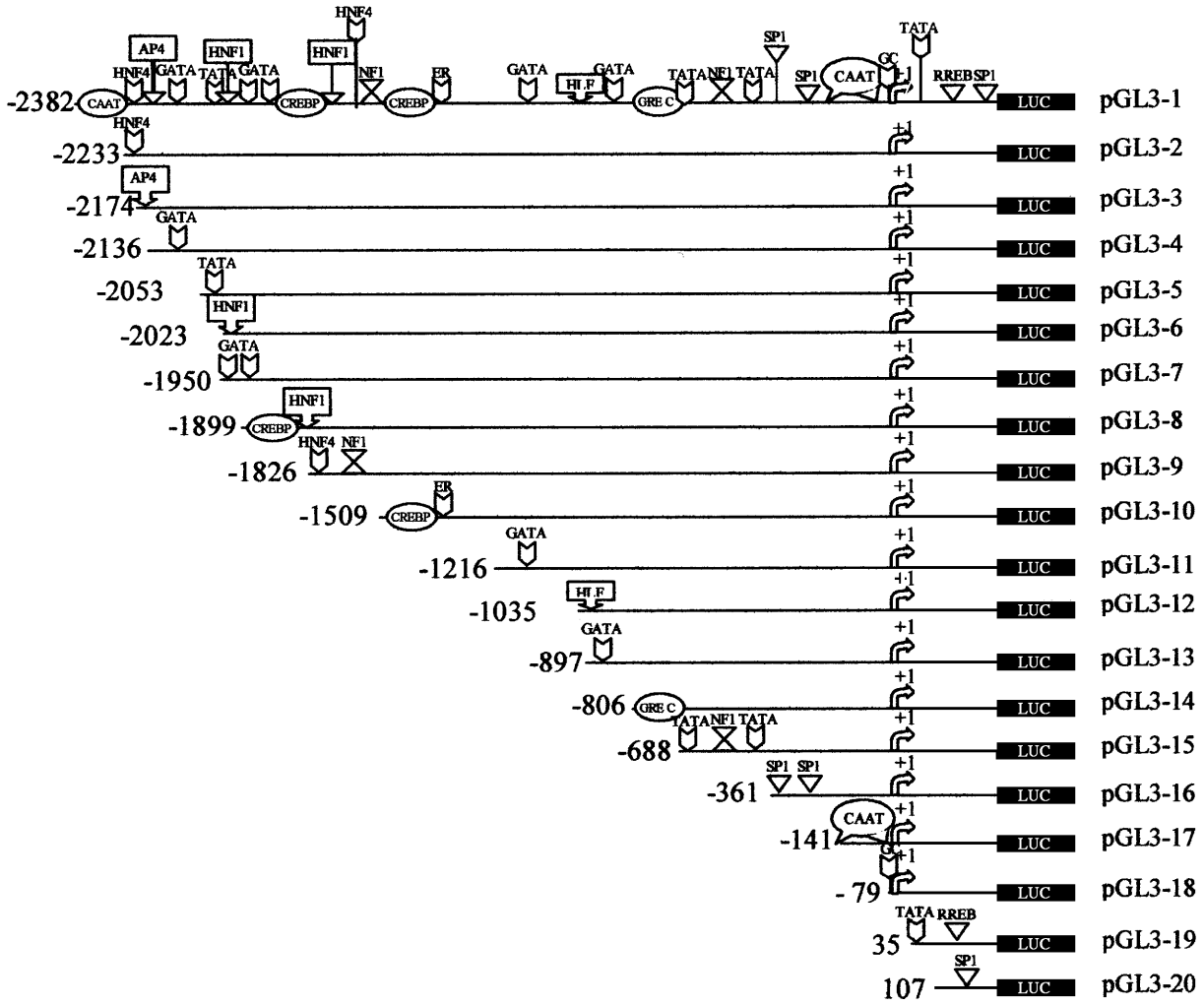


Figure 4 Schematic representation of the different constructions employed in promoter analysis

The positions of putative regulatory elements are indicated. Numbering for each construction is based on the transcription start site (+1) and is shown on the left-hand side of the Figure.

The transcription initiation site of the hLGA gene was determined by primer-extension analysis, using two specific antisense primers (Table 1) flanking the 5′ end of available cDNA sequences. Primer PEXT42, complementary to nt +37 to +60, produced one single band of 60 nt on the autoradiogram (Figure 3B). When primer PEXT48 (complementary to nt +43 to nt +66) was employed, another single band of 66 nucleotides in size was amplified (Figure 3B). Therefore, both extension reactions stopped at the same position; no other specific LGA products were obtained in the primer extension analyses. A unique transcription start site for hLGA can thus be assigned to the A nucleotide located 240 bp upstream from the initiation codon. All positions within the human 5′-flanking sequence have been indexed −1, −2, etc., relative to this +1 transcription start site.

Sequence analysis of the 5′-flanking region of the hLGA gene

The 2.4 kb DNA sequence of the 5′-flanking region of the hLGA gene is shown in Figure 2(B). The Transfac 4.0 database (http://transfac.gbf.de/TRANSFAC/) was searched to localize putative transcription factor binding sites, using MatInspector V2.2 software [18]. There are three TATA-box-like elements at −2043, −676 and −565, two cAMP-responsive element binding proteins (CREB and CREBP1) at −1851 and −1475, an NF-1 sequence at −658, a GATA-C binding site at −1133, and a hepatic leukaemia factor (HLF) at −1025. Other common regulatory elements included two putative binding sites for liver-enriched factor HNF-1 at −1969 and −1849, an oestrogen receptor (ER) site at −1454, and a glucocorticoid response element at −785. Two consensus binding sites for CAAT (−138 and −87) are present close to the transcription start site, suggesting potentially important roles. In this region, we also identified a GC-box element at −50, two Ras-responsive element binding proteins (RREBs) at −4 and +79, and a muscle TATA box (TTTAAA) at +53. Several binding sites for Sp-1 (−265, −225 and +174), HNF-4 (−2181), GATA-3 (−2121 and −1943), GATA-1 (−1927) and GATA-2 (−829) were also noted.

5′-Deletion analysis of the hLGA promoter

As a first approach to determine the important DNA elements regulating hLGA gene expression, a series of eighteen 5′-

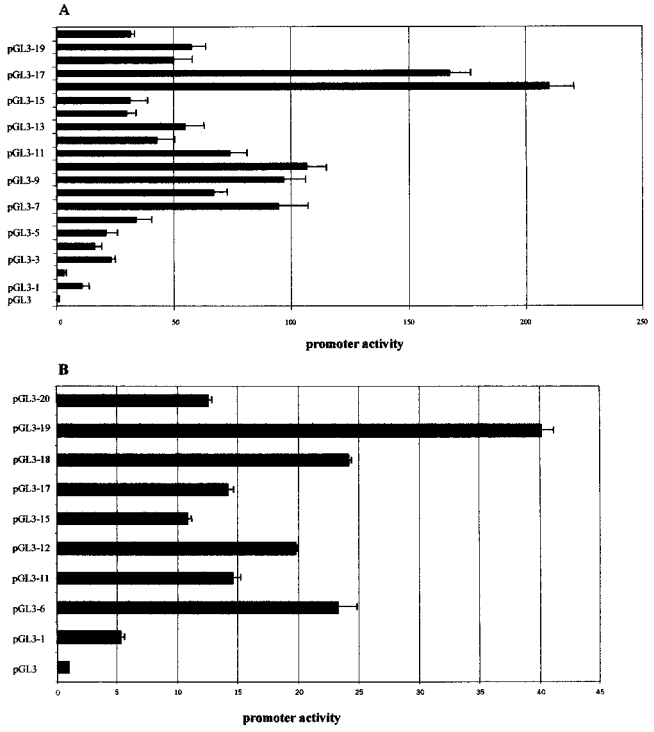


Figure 5 Functional analysis of the hLGA promoter in HepG2 and ZR-75 cells

Promoter activity of the 5′ region of the hLGA gene in HepG2 (**A**) and ZR-75 (**B**) cells. Transcriptional activity of deletion constructs was studied in the human hepatoma HepG2 cell line and in the ZR-75 breast cancer cell line. Cells were lysed and assayed for luciferase and β-galactosidase activities. Luciferase activity was divided by β-galactosidase activity to correct for transfection efficiency and is expressed as relative luciferase activity (promoter activity). The background activity of the pGL3 basic promoterless vector is also shown. The results are expressed as means $\pm$ S.D. from at least three different experiments, in duplicate for each construct.

truncated promoter fragments (pGL3-1 to pGL3-18), starting from nt -2382 to -79 of the 5′-upstream sequence and ending at nt $+410$, were prepared (Figure 4). In addition, two more constructs were assayed: pGL3-19 from nt $+35$ to $+410$ (containing TATA and RREB elements), and pGL3-20 encompassing nt $+107$ to $+410$ (containing a Sp-1 binding site). They were cloned into a promoterless luciferase reporter vector, pGL3, and the fusion promoter–luciferase constructs were transiently transfected into human HepG2 hepatoma cells (Figure 5A) and ZR-75 breast cancer cells (Figure 5B). The luciferase activity was normalized with β-galactosidase activity for transfection efficiency. Plasmid pGL3-1 from nt -2380 to $+410$ was able to provide a minimal promoter activity, while deletion of a distal CAAT element (plasmid pGL3-2) decreased its activity, suggesting a putative enhancer role for this CAAT box.

The deletions from nt -2233 to -2023 (Figure 5, constructs pGL3-3 to pGL3-6) yielded a 2- to 3.5-fold increase in luciferase activity relative to the pGL3-1 construct, suggesting the presence of silencer elements in a region where HNF-4, AP-4, GATA-3 and TATA-like motifs were identified (Figure 4). A further deletion of a 73 bp fragment (pGL3-7) increased 9-fold the luciferase activity, pointing at the HNF-1 element removed as another negative regulatory element, whereas GATA sequences at -1943 and -1927 behaved as putative positive elements, as judged by the lower activity (6-fold) shown by the next construct (pGL3-8).

A remarkable change in luciferase activity occurs after deletion to -1.5 kb (pGL3-10). This construct shows a more than 10-fold increase in promoter activity compared with pGL3-1, hinting at CREBP1 and/or HNF-1, and at HNF-4 and/or NF1 as very important silencer elements. Luciferase activity was dramatically enhanced to approx. 20-fold by deletion to -361 bp; in fact, pGL3-16 shows the highest promoter activity in HepG2 hepatoma cells. TATA-like elements at nt -676 and -565, as well as the NF-1 box, located at -658, appear as the most important silencer *cis*-acting elements within the hLGA promoter. In contrast, removal of two Sp-1 boxes at nt -265 and -225 (pGL3-17) left essentially unaltered luciferase expression. Further 5′ deletion to position -79 (pGL3-18) abolished a significant part of promoter activity in HepG2 cells, indicating that a core promoter is located between nt -141 and $+410$; two CAAT boxes are present here and would be important positive *cis*-acting elements. In contrast, a different behaviour was found for ZR-75 breast cancer cells: a stepwise increase was observed by progressive deletions from nt -141 (pGL3-17) to -79 (pGL3-18) and $+35$ (pGL3-19). In fact, this last construct had the maximum activity (8-fold with regard to pGL3-1), and further deletion to $+107$ (pGL3-20) significantly reduced the promoter activity. These results indicate that the region between nt -141 and $+107$ is important for optimal promoter activity in ZR-75 cells. In this region, GC sequences, TATA-like box and an RREB element were identified.

Mutagenesis of the core, proximal and distal enhancer elements of the hLGA promoter inhibit luciferase activity

In an attempt to identify transcription regulation elements, additional reporter constructs were prepared in which target boxes were mutated (Figure 6A). All plasmid inserts were analysed by sequencing the recombinant plasmids. Four bases within each CAAT-box, included in the pGL3-17 plasmid, were simultaneously mutated to produce two reporter constructs that dramatically reduced promoter activity in HepG2 (Figure 6B) and MCF-7 cells (Figure 6C). These boxes are thus necessary to confer on the basal promoter the high response of enhancement in hLGA gene transcription. Mutations at the proximal TATA-like element and Sp-1 box (pGL3-19 and pGL3-20 constructs, respectively) also yielded significant impairment of luciferase activity. However, mutagenesis at GC sequences in the pGL3-18 plasmid lowered the activity, but not as radically. Cells transfected with the promoter engineered to contain a mutated cassette of a TATA-like element (-676) or NF-1 box (-658), both at pGL3-15, exhibited low levels of basal transcription and, in particular, mutations at the TATA box decreased luciferase activity to values of 10% (HepG2) and 40% (MCF-7) of the wild-type promoter (Figures 6B and 6C, respectively). To investigate distal enhancer elements, we made site-specific mutations at the HLF site (pGL3-12) and the GATA C site (pGL3-11). A resulting large loss in promoter activity ranging from 16 to 34% of wild-type promoter activity was seen in both cell types. Modification of very distal enhancer elements, such as the HNF-1 sequence at -2023 (pGL3-6) or the CAAT cassette placed at -2255 (pGL3-1), decreased significantly the expression compared with the wild-type promoter, but only in HepG2 cells.

Protein binding to the HNF-1, HLF and Sp-1 boxes

To characterize the regulatory elements that show the greatest effect on luciferase activity, EMSA experiments were carried out using HepG2 nuclear extracts and specific double-stranded

A

Name	Sequence 5′ → 3′	Original sequence	nt
mut1CAAT D	CAGTACCA**CGAA**CCTCCCAAG	TTGG	-2263/-2243
mut6HNF1 D	GAAATTCCC**GGCA**AGATGGTATTTTTC	TAAC	-1978/-1952
mut11GATA D	CATTTCTCCCGG**TCCCC**TTCAATTCAATG	GATAA	-1145/-1117
mut12HLF D	CTTACTAGCTGG**AGGC**CCTTATGCGAG	GTTA	-1037/-1011
mut15/1NF1 D	GTTTTTCTGCGT**AATC**ACTTATTTTAATG	GCCA	-670/-642
mut15/2TATA D	CATGGTTTA**TGGA**ACGTTTTTCTGCG	TAAA	-685/-660
mut17/1CAAT D	CTTGCTCTCA**CGAA**GCAACCTTAC	TTGG	-147/-124
mut17/2CAAT D	CCTCAGTGA**CGAAA**CCTGGGCATTG	TTGG	-96/-72
mut18Sp1 D	CTCCCAGGCAC**TATT**CTCTCTCC	CGCC	-61/-39
mut19TATA D	CTGGAGCCAGCT**CC**AAAGGGGCAG	TTTA	42/65
mut20Sp1	CTCGCTCGCTC**TATT**CCTCCTTC	CGCC	163/185

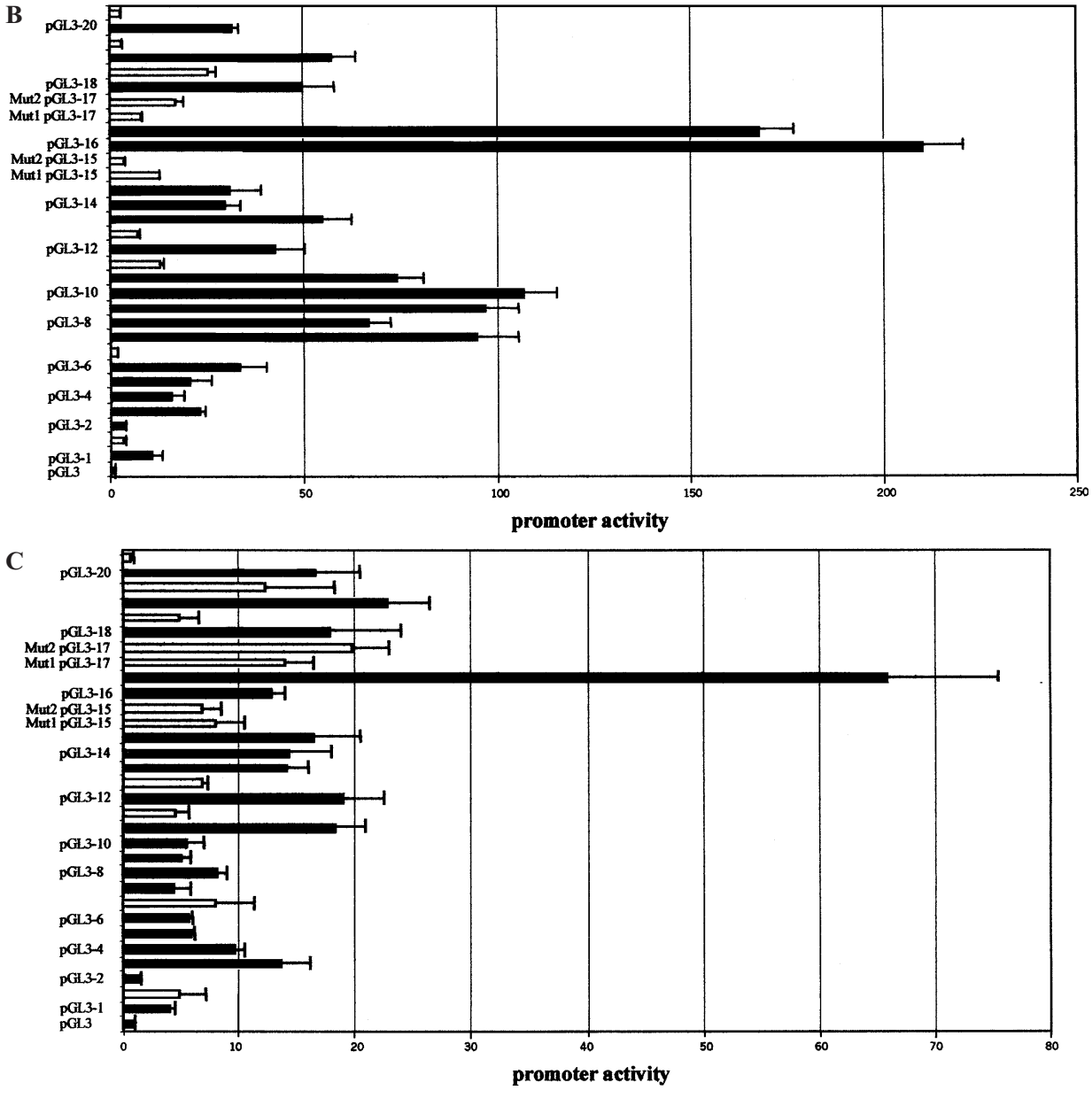


Figure 6 Effect of mutations on hLGA promoter activity

(**A**) Name and schematic representation of sequence and location of primers used for mutagenesis. Only sense strands are shown. The numbering refers to the transcription start site designated as +1. 5′-Deletion and mutagenesis analyses of the hLGA gene in HepG2 cells (**B**) and in MCF-7 cells (**C**) are shown. Transcriptional activity of deletion mutants is displayed (open bars) next to its respective standard deletion construction (black bars). Cells were lysed and assayed for luciferase and β-galactosidase activities. Luciferase activity is expressed as indicated in Figure 5. The background activity of pGL3 basic promoterless vector is also shown. The results are means ± S.D. and represent more than three different experiments, in duplicate for each construct.

A

Name	Sequence 5′ → 3′	nt
E1F	CACAGTACCATTGGCCTCCCAAGTCA	-2265/-2240
E1R	GTGACTTGGGAGGCCAATGGTACTGT	-2264/-2239
E6F	GTTTAGAAATTCCCTAACAGATGGTATTTTT	-1983/-1953
E6R	GAAAAATACCATCTGTTAGGGAATTTCTAAA	-1982/-1952
E12F	GGTACTTACTAGCTGGGTTACCTTATGCGAGAATT	-1041/-1007
E12R	GAATTCTCGCATAAGGTAACCCAGCTAGTAAGTA	-1039/-1006
E15.1F	GTTTTTCTGCGTGCCAACTTATTTTAATGT	-670/-641
E15.1R	GACATTAAAATAAGTTGGCACGCAGAAAAA	-669/-640
E15.2F	GCACTTCATGGTTTATAAAACGTTTTTCTG	-691/-662
E15.2R	CGCAGAAAAACGTTTTATAAACCATGAAGT	-689/-660
E17.1F	GTCTTGCTCTCATTGGGCAACCTTAC	-149/-124
E17.1R	GGGTAAGGTTGCCCAATGAGAGCAAGA	-148/-122
E17.2F	GCTTCCTCAGTGATTGGACCTGGGCATTGT	-100/-71
E17.2R	GGACAATGCCCAGGTCCAATCACTGAGGAAG	-147/-69
E19F	CCTGGAGCCAGCTTTAAAGGGGCAGG	41/66
E19R	GGCCTGCCCCTTTAAAGCTGGCTCCA	43/68
E20F	GTCTCGCTCGCTCCGCCCCTCCTTCCCAG	161/189
E20R	GGCTGGGAAGGAGGGGCGGAGCGAGCGAGA	162/191
Competitor		
1 CAAT D	CAGTACCA**CGAA**CCTCCCAAG	-2263/-2243
6 HNF1 D	GAAATTCCC**GGCA**AGATGGTATTTTC	-1978/-1952
12 HLF	CTTACTAGCTGG**AGGC**CCTTATGCGAG	-1037/-1011
15 NF1	GTTTTTCTGCGT**AATC**ACTTATTTTAATG	-670/-642
15 TATA	CATGGTTTA**TGGA**ACGTTTTTCTGCG	-685/-660
17/1 CAAT	CTTGCTCTCA**CGAA**GCAACCTTAC	-147/-124
17/2 CAAT	CCTCAGTGA**CGAA**ACCTGGGCATTG	-96/-72
19 TATA	CTGGAGCCAGCT**CC**AAAGGGGCAG	42/65
20 Sp1	CTCGCTCGCTC**TATT**CCTCCTTC	163/185

B

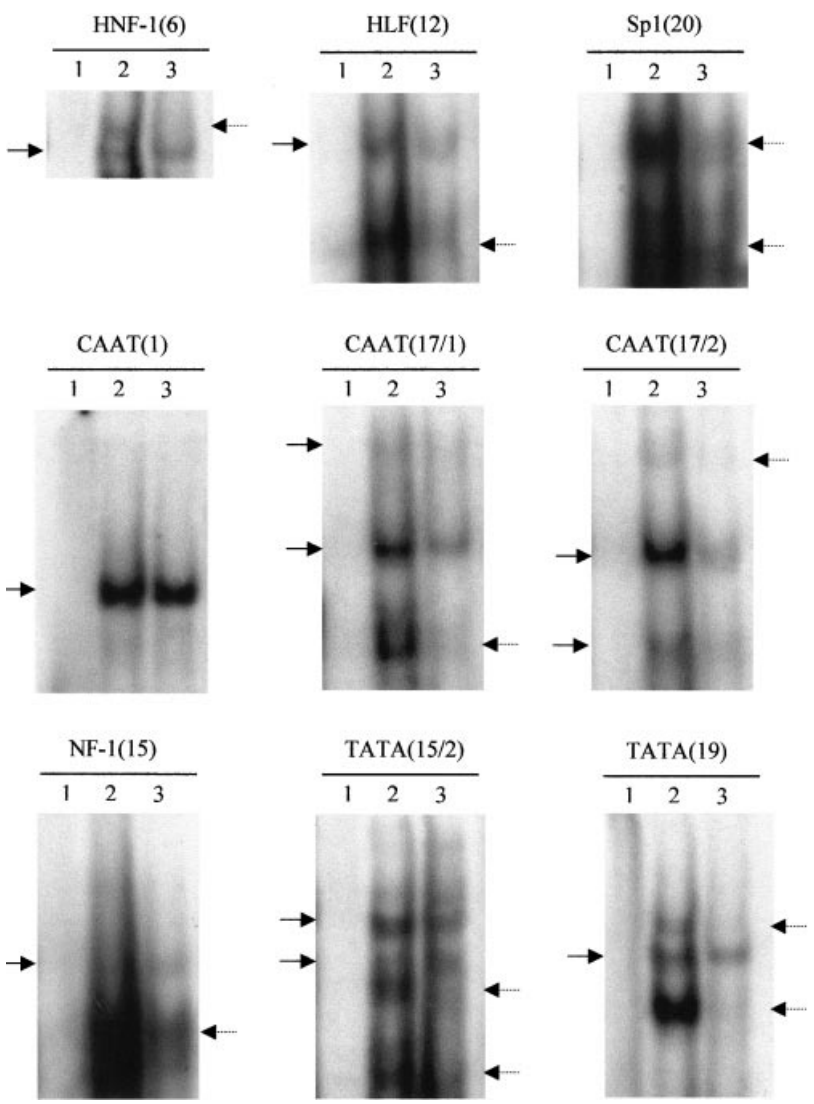


Figure 7 Formation of specific DNA–protein complexes at the hLGA promoter

(**A**) List of double stranded oligonucleotides used in these assays. Consensus sites are underlined. Sense strands of competitor mutated oligonucleotides, used in a 100-fold molar excess, are shown in bold and underlined. The numbering refers to the transcriptional start site designated as +1. (**B**) EMSA using HepG2 nuclear extracts demonstrates protein specificity (black arrows on the left-hand sides of the panels) in binding complexes to wild-type (lanes 2) and mutated (lanes 3) double-stranded oligonucleotides containing HNF-1, HLF, CAAT, NF-1, and TATA elements as indicated. The construct number is shown in brackets (when more than one binding motif is present in a construct, it is indicated after the construct number). Dashed arrows on the right-hand sides of the panels denote non-specific binding complexes. As a control, free probes without nuclear extracts were also assayed (lanes 1). For clarity, the free probes are not included at the bottom of the panels.

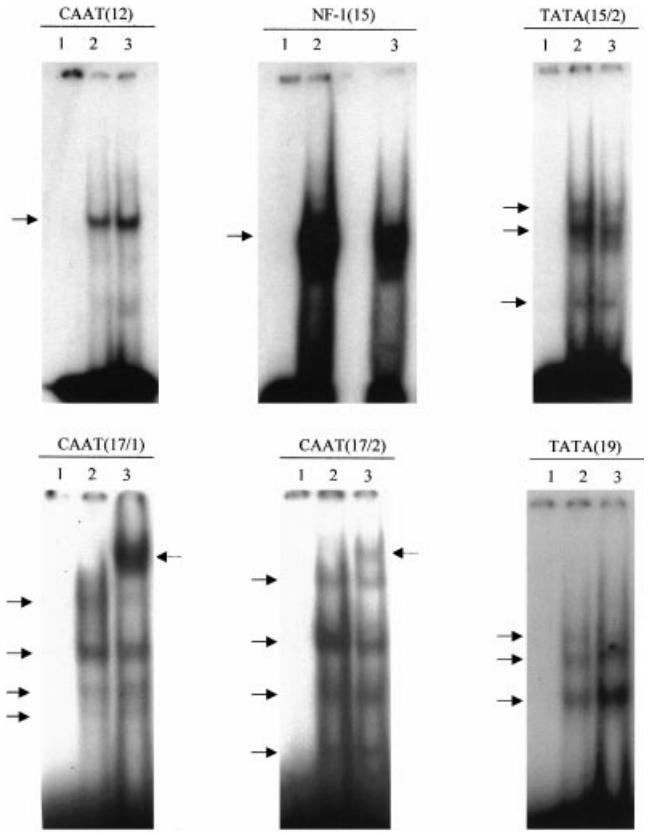


Figure 8 Antibody supershift analyses demonstrate the interaction of proximal elements with the hLGA promoter

EMSAs using HepG2 nuclear extracts depict specific binding complexes that shift and supershift ^{32}P-labelled oligonucleotides containing NF-1, CAAT, and TATA elements as indicated. The construct number is shown in brackets (when more than one binding motif is present in a construct, it is also indicated after the construct number). Black arrows on the left-hand sides of the panels show standard binding complexes (lanes 2), and dashed arrows on the right-hand sides of the panels denote supershift complexes consisting of DNA–protein complexes with anti-NF-1, anti-C/EBP, and anti-TFIID antibodies, respectively (lanes 3). As control, free probes without nuclear extracts were also assayed (lanes 1).

oligonucleotides for each target motif (Figure 7A). Controls with an excess of unlabelled (unmutated) oligonucleotides were also performed and they abolished the band shift observed with the respective labelled oligonucleotides (results not shown). The HNF-1 box at pGL3-6 shifted two binding complexes, but after incubation with the mutated double-stranded competitor oligonucleotide, a clear reduction in the intensity of the upper band was noted, whereas the lower band signal remained the same (Figure 7B). With regard to HLF, a less intense signal of specific binding was detected only for its upper band. On the other hand, although the Sp-1 element initially showed several shifted bands, their overall intensity significantly decreased after incubation with the mutated competitor oligonucleotide, as happened with the HLF lower band (Figure 7B). These results indicate that the HNF-1 motif at -2023 can be a significant distal element in the transcriptional regulation of hLGA.

CAAT, NF-1 and TATA-like elements constitute the gene cassette on the core promoter of hLGA

Additional EMSAs were carried out using oligonucleotides and antibodies specific for putative critical elements controlling expression of hLGA. These assays revealed mainly two CAAT boxes and, to a much lesser degree, a TATA-like element as the more important regulatory sequences. Experiments using specific probes for CAAT sites at pGL-17, NF-1 site at pGL-15, and TATA sites at pGL-15 and pGL-19 were assayed (Figure 7B). All of them were found to retard the band corresponding to the binding complexes, although the strongest bindings were obtained with CAAT and TATA elements. The interactions seem specific as they were inhibited by the unlabelled (unmutated) oligonucleotide and not by the unlabelled mutated oligonucleotide. Although incubation with mutated competitors produced the disappearance of some binding complexes and/or reduction of the overall band intensity, formation of complexes with the CAAT, TATA and NF-1 elements were not competed out by their respective mutated oligonucleotides. Thus, significant shifted signals still remained after competition, indicating the specificity of these bindings (Figure 7B). Antibody supershift analyses were employed in an attempt to identify the proteins producing the retarded bands. Anti-C/EBP antibody was found to consistently decrease the intensity of the binding complexes and dramatically supershift the band of the CAAT boxes corresponding to pGL-17 (Figure 8). Moreover, the first of the two CAAT boxes in this region showed a more intense signal (Figure 8). With the TATA-like element at pGL-19, we observed a decrease in the intensity of the two major bands and the clear disappearance of the upper faint band when anti-TFIID antibody (specific for TATA binding protein) was used for supershift experiments (Figure 8). In contrast, antibodies specific to the TATA-like element or NF-1 box at pGL-15, showed a less marked effect (Figure 8). As a control, we show a supershift experiment using the CAAT element at pGL3-1, in this case, C/EBP antibodies were without effect (Figure 8, top left-hand panel). These results suggest that CAAT elements at pGL-17 and TATA box at pGL-19 are necessary to enhance basal transcription of the hLGA promoter. They also indicate that both CAAT elements at pGL-17 recruit the transcription factor, but the first of the two CAAT boxes, placed at -138, is more accessible to antibody binding.

DISCUSSION

In addition to fulfilling key tissue-specific physiological roles, GA is also overexpressed in a variety of malignancies such as human hepatomas [19,20], breast, colon and leukaemia cancer cells [12,21], and experimental tumours [16]. Furthermore, new therapeutic strategies, aimed at inhibition of GA expression by antisense technology, have shown promising results in animal tumour models [22]. However, the mechanisms by which GA expression is regulated in tumours remain unknown, making it difficult to ascertain if GA overexpression is a cause or an effect of the malignant transformation. In ZR-75 breast cancer cells, LGA has a regulated pattern of expression showing maximum mRNA level and specific activity at the exponential phase of tumour growth [12]. In the present study, the genomic organization and promoter characterization of the hLGA gene is presented. The gene, located on chromosome 12, is approx. 18 kb long and split into 18 exons. All exon/intron boundaries conform to known consensus splice donor and splice acceptor rules [23]. A second glutaminase gene is present in mammals, the KGA gene, located on human chromosome 2 [14]. Its genomic sequence spans 82 kb and is distributed in four separate contigs in the Human Genome Project database. By comparison with available hKGA cDNAs, the gene was split into 19 exons [9]. At least two different transcripts arise from this gene: the KGA mRNA formed by joining exons 1–14 and 16–19, and the

alternative spliced transcript GAC mRNA which uses only the first 15 exons, omitting exons 16–19 [9]. This GAC isoform was first isolated from a human colon carcinoma cDNA library [8], but recently it has been reported to be present also in rat and pig [9].

Apart from the additional exon present in the hKGA gene, the main differences in the coding sequences of both genes are located at exons 1 and 18. Exon 1 shares 62.5% similarity, but it codes for 129 amino acids in KGA and only for 61 amino acids in LGA, accounting for the 67 extra amino acids of KGA protein at the N-terminal. The sequences encoded by exon 1 contain the signals involved in mitochondrial targeting and translocation processes [7,12]. Interestingly, exon 1 also contains an LXXLL signature motif for LGA, which might explain the nuclear localization recently demonstrated for LGA in mammalian brain [24]. Likewise, exon 18, which codes for the C-terminal region of both proteins, shows the lowest sequence similarity (29.4%). This region of the hLGA protein has been demonstrated recently to be involved in the recognition of PDZ (PSD95/Dlg/ZO1 domains)-interaction modules [25]. Therefore, the most significant differences between hLGA and hKGA exons are located in regions involved with organelle targeting and protein–protein interactions, which may help to explain their differential function and regulation. Conversely, exons 3–17 of both mRNA transcripts have the same length and show a high sequence similarity. Altogether, these data suggest that GA genes may have evolved from a common ancestral gene, arising by gene duplication and divergent evolution, as has been previously noted for the rat genes [9].

The full sequences of other mammalian LGA genes are still unknown. In spite of this, the exon/intron boundaries for the rLGA gene may be predicted by using genomic clones in the rat genome database and the cDNA previously cloned for the rat liver gene [13]. Nucleotide and amino acid sequences of hLGA and rLGA share a remarkably high level of identity (89% and 96%, respectively). The positions interrupted in the human and rat LGA genes differ slightly, thus the exon/intron organization of both genes closely resemble each other (Figure 1B). Nevertheless, it is noteworthy that rLGA is 67 amino acids shorter than hLGA at the N-terminal; the rLGA protein lacks all the amino acids encoded by exon 1 and the first 6 amino acids of exon 2. Therefore the existence of alternative spliced forms of the LGA gene seems very likely, as happens with the KGA gene. Although mRNA transcripts of different lengths have not been isolated yet, data on LGA protein expression in mouse tissues (J. A. Campos, J. C. Aledo, J. A. Segura, F. J. Alonso, P. M. Gómez-Fabre, I. Núñez de Castro and J. Márquez, unpublished work), as well as the sequence of several mouse expressed-sequence tags and mRNAs deposited on the database, strongly suggest the existence of LGA splicing forms.

Another interesting aspect of the LGA gene is the presence of two *Alu* sequences at introns 7 and 14. The *Alu* cassette has 230 nucleotides with 90% identity with the *Alu*-Sc and *Alu*-Sb human *Alu* subfamilies, when compared to the *Alu* database using the BLASTN program. The sequence at intron 14 also contains the typical adenine-rich linker in the middle and a poly(A) tail at the 3′ end. *Alu* family members are short interspersed repetitive DNA elements that are believed to have spread throughout primate genomes by retroposition [26]. In practice, *Alu* elements are found in the introns of almost all known protein-coding genes and, less frequently, they also appear in the coding regions of mRNA, accounting for a source of protein variability [27]. Its presence in the LGA gene may have some other functional relevance that remains to be established. For example, a new subclass of *Alu* DNA repeats that can function as an ER-dependent transcriptional enhancer has been found in oestrogen-positive human breast cancer MCF-7 cells [28]. Interestingly, the promoter region of the hLGA gene does contain such a sequence and the region from nt −1310 to −1478 has 83% identity with this *Alu* repeat, discovered as a novel ER-like enhancer.

The core promoter regions of both the human and rat LGA do not contain canonical TATA boxes, but do have G+C-rich domains. This characteristic is also found in the hKGA promoter [9]. The nucleotide sequence of the immediate upstream region of the hLGA gene contains several features characteristic of a liver-specific promoter (Figure 2B). The ubiquitously expressed CAAT, NF1, and Sp1 elements have been implicated in the regulation of several liver-specific genes [29,30]. Also in this region there are *cis*-acting elements such as a non-canonical TATA box, an RREB site and a GC box. Notwithstanding the high sequence identity (89%) between the coding regions of the human and rat LGA [12], their proximal promoter regions did not show significant sequence homology (Figure 2A). Unlike rLGA, hLGA has a canonical CAAT box at about −80 and RREBs, though it lacks a HNF-5 site present in rLGA.

By employing 5′-deletion analysis, the sequence from −141 to +410 nt was shown to direct maximum promoter activity in HepG2 and MCF-7 cells; consequently, this fragment was identified as the core promoter (Figures 4 and 5). It contains sites for CAAT, GC, TATA, RREB and Sp1. However, in ZR-75 breast cancer cells, the core promoter was assigned to a sequence from +35 to +410 nt, lacking CAAT and GC boxes, although transcriptional activities were considerably lower than in HepG2 cells. The minimal promoter was the same for all three cell types and corresponded to the sequence from nt +107 to +410, where a unique Sp1 site is found. It supplied approximately the same level of expression of luciferase reporter gene as that of 5′-flanking regions up to 2.4 kb in size. The minimal promoter directs expression in cells derived from human hepatoblastoma, HepG2, cells and in two epithelially derived breast cancer cell types, ZR-75 and MCF-7. HepG2 cells showed the greatest global expression of luciferase activity (Figures 5A and 6B), whereas MCF-7 cells had high, but less remarkable, luciferase values (Figure 6C). Finally, luciferase constructs were only moderately active in ZR-75 cells, which also exhibited the lowest transfection efficiency (Figure 5B).

Mutagenesis and transient transfections clearly demonstrated that two CAAT boxes at nt −137 and −87 (pGL3-17) play a crucial role in the transcriptional regulation of the hLGA gene, both in cells of liver origin (HepG2) and in MCF-7 breast cancer cells (Figure 6). In addition, other positive elements required for optimal transcription are located in the 5′-UTR: a TATA-like element (+53) and an Sp1 site (+174). These last elements, along with an RREB site (+80), seem to be sufficient to achieve maximum transcription in ZR-75 cells. EMSA experiments confirmed the involvement of CAAT- and TATA-like boxes in the regulation of the hLGA promoter (Figure 7). These transcription factors, or other proteins closely related immunologically with them, were able to bind to the mentioned elements (Figure 8). Therefore, these elements appeared to be suitable candidates for controlling promoter activity.

C/EBPs control cell growth and differentiation, causing growth arrest and inducing cellular differentiation in several adipocyte, granulocyte and keratinocyte lineages [31]. Furthermore, HepG2 hepatoma cells express significantly lower levels of C/EBPα and C/EBPβ than those found in normal terminally differentiated hepatocytes [32]. Thus conversion of hepatocytes into proliferating hepatoma cells might require strong down-regulation of C/EBPα and C/EBPβ expression. In future studies it will be interesting to examine the ability of C/EBPs to modulate

C/EBP-mediated cell growth arrest in various cellular contexts and its effect on LGA expression.

In summary, these studies provide evidence that multiple nuclear proteins may regulate the transcription of the hLGA gene promoter. Essential work is in progress to provide new insights into molecular mechanisms controlling hLGA expression, to understand its role in cell growth and proliferation. Furthermore, the recent striking finding of a nuclear localization of LGA in mammalian brain argues for the importance of this gene, suggesting its involvement in transcription regulation [24]. Interestingly, emerging evidence is also appearing linking glutamine with some signalling cascades related to cell differentiation and apoptosis [33]. Further studies will also be needed to ascertain the role of the glutamine/glutaminase coupling in the regulation of gene expression.

We thank Dr E. Múñoz-Blanco (University of Córdoba) for use of his luminometer, and Dr J. A. Campos and Ms C. Lobo for helpful suggestions during the course of this work. C.P.-G. was a predoctoral fellow from the Ministry of Education of Spain. This work was supported by Grant SAF 2001-1894 from the Ministry of Ciencia y Tecnología of Spain.

REFERENCES

1 Krebs, H. A. (1935) The synthesis of glutamine from glutamate and ammonia, and the enzymatic hydrolysis of glutamine in animal tissues. Biochem. J. **29**, 1951–1969
2 Kovacevic, Z. and McGivan, J. D. (1983) Mitochondrial metabolism of glutamine and glutamate and its physiological significance. Physiol. Rev. **63**, 547–605
3 Curthoys, N. P. and Watford, M. (1995) Regulation of glutaminase activity and glutamine metabolism. Annu. Rev. Nutr. **15**, 133–159
4 Medina, M. A., Sánchez-Jiménez, F., Márquez, J., Quesada, A. R. and Núñez de Castro, I. (1992) Relevance of glutamine metabolism to tumor cell growth. Mol. Cell. Biochem. **113**, 1–15
5 Lohmann, R., Souba, W. W. and Bode, B. P. (1999) Rat liver endothelial cell glutamine transporter and glutaminase expression contrast with parenchymal cells. Am. J. Physiol. **276**, G743–G750
6 Laterza, O. F., Hansen, W. R., Taylor, L. and Curthoys, N. P. (1997) Identification of an mRNA-binding protein and the specific elements that may mediate the pH-responsive induction of renal glutaminase mRNA. J. Biol. Chem. **272**, 22481–22488
7 Shapiro, R. A., Farrell, L., Srinivasan, M. and Curthoys, N. P. (1991) Isolation, characterization, and in vitro expression of a cDNA that encodes the kidney isoenzyme of the mitochondrial glutaminase. J. Biol. Chem. **266**, 18792–18796
8 Elgadi, K. M., Meguid, R. A., Qian, M., Souba, W. W. and Abcouwer, S. F. (1999) Cloning and analysis of unique human glutaminase isoforms generated by tissue-specific alternative splicing. Physiol. Genomics **1**, 51–62
9 Porter, L. D., Ibrahim, H., Taylor, L. and Curthoys, N. P. (2002) Complexity and species variation of the kidney-type glutaminase gene. Physiol. Genomics **9**, 57–66
10 Taylor, L., Liu, X., Newsome, W., Shapiro, R. A., Srinivasan, M. and Curthoys, N. P. (2001) Isolation and characterization of the promoter region of the rat kidney-type glutaminase gene. Biochim. Biophys. Acta **1518**, 132–136
11 Smith, E. M. and Watford, M. (1990) Molecular cloning of a cDNA for rat hepatic glutaminase. Sequence similarity to kidney-type glutaminase. J. Biol. Chem. **265**, 10631–10636
12 Gómez-Fabre, P. M., Aledo, J. C., del Castillo-Olivares, A., Alonso, F. J., Núñez de Castro, I., Campos, J. A. and Márquez, J. (2000) Molecular cloning, sequencing and expression studies of the human breast cancer cell glutaminase. Biochem. J. **345**, 365–375
13 Chung-Bok, M.-I., Vincent, N., Jhala, U. and Watford, M. (1997) Rat hepatic glutaminase: identification of the full coding sequence and characterization of a functional promoter. Biochem. J. **324**, 193–200
14 Aledo, J. C., Gómez-Fabre, P. M., Olalla, L. and Márquez, J. (2000) Identification of two human glutaminase loci and tissue-specific expression of the two related genes. Mammal. Genome **11**, 1107–1110
15 Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani, H., Nombra, N. and Ohara, O. (1998) Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. **5**, 355–364
16 Aledo, J. C., Segura, J. A., Medina, M. A., Alonso, F. J., Núñez de Castro, I. and Márquez, J. (1994) Phosphate-activated glutaminase expression during tumor development. FEBS Lett. **341**, 39–42
17 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
18 Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. **23**, 4878–4884
19 Matsuno, T. and Goto, I. (1992) Glutaminase and glutamine synthetase activities in human cirrhotic liver and hepatocellular carcinoma. Cancer Res. **52**, 1192–1194
20 Bode, B. P. and Souba, W. W. (1994) Modulation of cellular proliferation alters glutamine transport and metabolism in human hepatoma cells. Ann. Surg. **220**, 411–424
21 Colquhoun, A. and Newsholme, E. A. (1997) Aspects of glutamine metabolism in human tumour cells. Biochem. Mol. Biol. Int. **41**, 583–596
22 Lobo, C., Ruiz-Bellido, M. A., Aledo, J. C., Márquez, J., Núñez de Castro, I. and Alonso, F. J. (2000) Inhibition of glutaminase expression by antisense mRNA decreases growth and tumourigenicity of tumour cells. Biochem. J. **348**, 257–261
23 Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**, 349–383
24 Olalla, L., Gutiérrez, A., Campos, J. A., Khan, Z. U., Alonso, F., Segura, J. A., Márquez, J. and Aledo, J. C. (2002) Nuclear localization of L-glutaminase in mammalian brain. J. Biol. Chem. **277**, 38939–38944
25 Olalla, L., Aledo, J. C., Bannenberg, G. and Márquez, J. (2001) The C-terminus of human glutaminase L mediates association with PDZ domain-containing proteins. FEBS Lett. **488**, 116–122
26 Shen, M. R., Batzer, M. A. and Deininger, P. L. (1991) Evolution of the master *Alu* gene(s). J. Mol. Evol. **33**, 311–320
27 Makalowski, W., Mitchell, G. A. and Labuda, D. (1994) Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet. **10**, 188–193
28 Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., Iglehart, J. D., Deininger, P. L. and McDonnell, D. P. (1995) Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. J. Biol. Chem. **270**, 22777–22782
29 Boisclair, Y. R., Brown, A. L., Casola, S. and Rechler, M. M. (1993) Three clustered Sp1 sites are required for efficient transcription of the TATA-less promoter of the gene for insulin-like growth factor-binding protein-2 from the rat. J. Biol. Chem. **268**, 24892–24901
30 Park, E. A., Gurney, A. L., Nizielski, S. E., Hakimi, P., Cao, P., Moorman, A. F. M. and Hanson, R. W. (1993) Relative roles of CCAAT/enhancer-binding protein beta and cAMP regulatory element-binding protein in controlling transcription of the gene for phosphoenolpyruvate carboxykinase (GTP). J. Biol. Chem. **268**, 613–619
31 Darlington, G. J., Ross, S. E. and MacDougald, O. A. (1998) The role of C/EBP genes in adipocyte differentiation. J. Biol. Chem. **273**, 30057–30060
32 Friedman, A. D., Landschulz, W. H. and McKnight, S. L. (1989) CCAAT/enhancer binding protein activates the promoter of the serum albumin gene in cultured hepatoma cells. Genes Dev. **3**, 1314–1322
33 Matés, J. M., Pérez-Gómez, C., Núñez de Castro, I., Asenjo, M. and Márquez, J. (2002) Glutamine and its relationship with intracellular redox status, oxidative stress and cell proliferation/death. Int. J. Biochem. Cell. Biol. **34**, 439–458

Received 16 September 2002/25 October 2002; accepted 25 November 2002
Published as BJ Immediate Publication 25 November 2002, DOI 10.1042/BJ20021445