

Title: The human gC1q-R/p32 gene (*C1qBP*): Genomic Organization and Promoter Analysis.

Keywords: C1q, , SF2, complement, coagulation, kininogen, HIV, internalin B.

Angela J. Tye^a, Berhane Ghebrehiwet^b, Ning Guo^c, Kedarnath N. Sastry^c, Billy K. C. Chow^a,
Ellinor. I. B. Peerschke^d and Boon-Leong Lim^{a,*}

^a*Department of Zoology, University of Hong Kong, Pokfulam, Hong Kong, China*

^b*Department of Medicine, State University of New York, Stony Brook, N.Y., 11794-8161, USA*

^c*Department of Pathology, Weill College of Medicine of Cornell University, New York, N.Y.
10021, USA*

^d*Department of Pathology and Laboratory Medicine, Boston University School of Medicine,
Boston, MA, USA*

*To whom correspondence should be addressed: Department of Zoology, the University of
Hong Kong, Pokfulam, Hong Kong.

Tel: 852-22990826

Fax: 852-25599114

Email: blim@hkucc.hku.hk

Abbreviations:

AP, Transcription factor activator protein; bp, base pair(s); cDNA, complementary DNA;
DNA, Deoxyribonucleic acid; HSF, Heat-shock factor; NFAT, nuclear factor of activated T
cells; PCR, polymerase chain reaction.

Abstract

gC1qR is an ubiquitously expressed cell protein that interacts with the globular heads of C1q (gC1q) and many other ligands. In this study, the 7.8 kb human gC1q-R/p32 (*ClqBP*) gene was cloned and found to consist of 6 exons and 5 introns. Analysis of a 1.3 kb DNA fragment at the 5' flanking region of this gene revealed the presence of multiple TATA, CCAAT and Sp1 binding sites. Luciferase reporter assays performed in different human cell lines demonstrated that the reporter gene was ubiquitously driven by this 1.3 kb fragment. Subsequent 5' and 3' deletion of this fragment confined promoter elements to within 400 bp upstream of the translational start site. Since the removal of the 8-bp consensus TATATATA at -399/-406 and CCAAT at -410/414 did not significantly affect the transcription efficiency of the promoter, GC rich sequences between this TATA box and the translation start site may be very important for the promoter activity of the *ClqBP* gene. One out of seven GC rich sequences in this region bind specifically to PANC-1 nuclear extracts, and the transcription factor Sp1 was shown to bind to this GC rich sequence by the supershift assay. Primer extension analysis mapped three major transcription start regions. The farthest transcription start site is 49bp upstream of the ATG translation initiation codon and is in close proximity of the specific SP1 binding site.

1. Introduction

gC1q-R is a biologically important, widely distributed, multiligand binding and multifunctional protein [1]. Numerous reports have claimed that gC1qR and its homologue could be isolated or identified in various cellular compartments, including plasma membrane, cytoplasm, mitochondria and nucleus. gC1qR isolated from the plasma membrane was originally characterized as a high affinity C1q-binding protein [2], and later many reports showed that gC1q-R could interact with several proteins of the intrinsic coagulation/bradykinin forming cascade, including high molecular weight kininogen[3], Factor XII [4], fibrinogen [5] and multimeric vitronectin [6]. Conversely, intracellular gC1qR was shown to interact and subsequently down-regulate the surface expression of the α 1b-adrenergic receptor [7] as well as bind to the kinase domain of protein kinase C μ and thus prevent its substrate-phosphorylation activity [8]. In addition, gC1qR was also reported to bind to a nuclear splicing factor SF2 and to many viral proteins, including HIV-1 Tat[9] and Rev[10], core protein V of adenovirus[11], EBNA-1 of Epstein-Barr virus[12] and open reading frame P of herpes simplex virus[13], implying that gC1qR may play a role in virus-host interaction.

The full length cDNA of gC1qR encodes a pre-pro-protein of 282 residues from which a 73-residue-long N-terminal segment is removed by site-specific cleavage to generate the mature gC1q-R [2]. It was shown that the fusion of the residues 1-81 or 1-33 of the pre-pro-protein to the N-terminus of the green fluorescent protein directed the fusion protein to mitochondria [14]. However, the findings of Dedina et al [14] do not exclude the possibility that gC1qR, like many other proteins, could be exported from the mitochondria by an unknown mechanism [15]. This is supported by a recent report showing that anti-gC1qR monoclonal antibody can reverse the anti- proliferation effects of Hepatitis virus C core

antigen on activated T cells[16].

The human *C1qBP* gene was assigned to human chromosome 17q13.3 [17]. High degree of amino acid identity exists between the human, rat and mouse gC1qR cDNA sequences [18]. In this study, the full length gene of human gC1qR was cloned and its exon-intron boundaries were revealed. Furthermore, the transcription start site and its promoter elements were also mapped and characterized.

2. Materials and Methods

2.1. Screening of human genomic library

A human genomic library in bacteriophage EMBL3 was purchased from Clontech, California, USA (Cat. No. HL1067J). The human cDNA of the *C1qBP* gene was used as the probe for library screening [2]. The cDNA insert was released from its vector and gel-purified before radiolabeling with ³²P-dATP by the random priming method. Positive plaques were picked, replated and rescreened until single positive plaques were picked. The insert in a positive plaque was digested with various restriction enzymes and the positive fragments were mapped by southern blot analysis. Six overlapping subclones (Figure 1) were obtained by inserting the positive fragments into the plasmid pBluescript. Each subclone was sequenced by primer walking method using the ABI Prism 310 Genetic Analyzer (Applied Biosystems Hong Kong Limited, Hong Kong).

2.2. Sequence assembly and analysis

Sequence data were assembled and analyzed by DNA processing softwares including MAC DNASIS (Hitachi, Japan) and DNA Strider (Christian Marck, Service de Biochimie, Department de Biologie, Institut de Recherche Fondamentale, CEA, France). Promoter

analysis was performed by the MatInspector Software (V 2.2) [19] on the world wide web (<http://bioinformatics.weizmann.ac.il/transfac/>). A transcription factor database (TRANSFAC V 4.0) was employed for the search of promoter elements [20].

2.3. Construction of expression plasmids for promoter assay

Various DNA fragments of the 5' upstream region of the human gC1qR gene (*C1qBP*) were obtained by PCR using subclone gI as the template. Restriction sites BglIII or SacI were added to the 5' end of the primers (Table 1A) so that the PCR products can be easily subcloned into the corresponding restriction sites in the Dual-Luciferase reporter vector pGL3-Basic (Promega, Hong Kong). A nested family of 5' and 3' deletion clones (Table 2) were generated in this manner (Figure 4).

2.4. Cell culture, transient transfection and promoter assays

The human cell lines PANC-1 (ATCC no. CRL-1469), MDA-MB-231 (ATCC no. HTB-26), SVG P12 (ATCC no. CRL-8621), HuTu80 (ATCC no. HTB-40) and 293-EBNA (Invitrogen, no. R620-07) were selected for transient transfection studies. SVG P12 and HuTu80 cells were cultured in MEM, (Life Technologies, Hong Kong); PANC-1, MDA-MB-231 and 293-EBNA cells were cultured in DMEM, high glucose (Life Technologies, Hong Kong). All culture mediums were supplemented with 10%(v/v) fetal bovine serum (50 ml FBS in 500 ml medium) and 1X antibiotic-antimycotic mixture (Life Technologies, Hong Kong) at 37°C with 5% CO₂. Plasmid DNA used for transfection was prepared using the Quantum prep Kit (BIORAD, Hong Kong), and treated with phenol chloroform before transfection. For transient transfection, cells were seeded onto 12-well plates (Costar, Corning Inc., USA) at a density of 1.0 X 10⁵ cells /well. Co-transfection was performed 19 hours later using the LipofectAMINE PLUS[®] Reagent (Life Technologies,

Hong Kong), with procedures performed according to the manufacturer's protocol. The molar ratio of the recombinant plasmid to be assayed and the internal control pRL-SV40 plasmid (Promega, Hong Kong) was kept at 1:1 in all transfection. Immediately after transfection, the cells were incubated in the medium without serum and antibiotics at 37°C with 5% CO₂ for 3.5 hr before changing into the complete medium. The cells were lysed at 39 hr post-transfection by washing the cells twice with 1X Phosphate Buffer Saline (PBS), followed by the addition of reporter Passive Lysis Buffer (Promega, Hong Kong). Dual-Luciferase reporter assay was performed according to the manufacturer's protocol (Dual-Luciferase Reported Assay kit, Promega, Hong Kong) using a luminometer (Lumat LB 9507, EG & G, Berthold) as the measuring apparatus.

2.5. RNA isolation and primer extension analysis

Total cellular RNA was isolated from human cell line MDA-MB-231 by TRIzol reagent (Life Technologies, Hong Kong). Poly(A)⁺ RNA was extracted from 2 mg of total RNA using the PolyAtract mRNA isolation kit (Promega, USA). Labeling of primer and primer extension reaction was performed using 1µg Poly(A)⁺ RNA with 10pmole antisense primer *G1A2Bgl* following the instruction of the Primer Extension System (Promega, USA). 4 µl of the primer extension reaction product was analyzed on 6%(w/v) acrylamide/7M urea sequencing gel. The size of the primer extension product was determined by comparison with a DNA sequence ladder generated with the same oligonucleotide primer using ERES3 plasmid as a template.

2.6. Gel mobility shift assays

The human cell line PANC-1 was selected for gel shift assays. Cells were cultured as mentioned in section 2.4 to exponential phase and harvested. End-labeling of

double-stranded oligos (Table 1B) was done by the Ready-To-Go T4 polynucleotide kinase labeling kit (Pharmacia) and [γ - 32 P]ATP (5000 Ci/nmol) (Amersham, Arlington Heights, IL). Nuclear extracts preparation and gel shift assays were performed as earlier described [21], [22].

3. Results

3.1. Genomic organization of the human *ClqBP* gene

Figure 1 is a schematic representation of the genomic organization of the gene, which, including its 5' and 3' flanking regions, spans about 7.8 kb. From the first codon of the initiation methionine till the stop codon of the gene, the gene spans 6055 bp (Figure 2). By alignment of the cDNA and the genomic sequences, intron-exon boundaries were defined. There are 6 exons and 5 introns in the *ClqBP* gene. The sizes of exons range from 94 bp (exon 3) to 232 bp (exon 1), and that of the introns range from 128 bp (intron 5) to 3156 bp (intron 2). Amino acid codons are split by introns 1 and 2 at the junctions of their adjacent exons. (Table 3) A poly A signal is located at 369 bp from the stop codon. The entire genomic sequence has been deposited in the Genbank under Accession No: AF338439.

3.2. Characterization of the 5'-flanking region of the human *ClqBP* gene

A 1.3 kb nucleotide sequence located upstream of the ATG initiation codon of the gene was analyzed (Figure 3) using the TRANSFAC transcription factor database. Putative promoter and enhancer elements including TATA boxes, CCAAT boxes, Octamers, Sp1 binding sites, GATA sequences, E boxes and AP elements were identified by the software MatInspector. There are four putative TATA boxes (-806/-811, -614/-617, -446/-449 and -399/-406) and three putative CCAAT boxes (-1033/-1037, -460/-463 and -410/-414). The 8-bp consensus TATA box with sequence TATATATA located at -399/-406 is the longest

TATA element found in the region, and it is in close proximity to a CCAAT box located at -410/414. Sp1 binding GC-rich motifs are also found throughout the 5'-flanking region of the gene, at positions -1309/-1314, -959/-965, -535/-540, -519/-524, -337/-342, -177/-182, -148/-154, -139/-144, -111/-118, -83/-90 and -51/-61.

A variety of consensus elements for transcription factors are also present, including an octamer site for the homeobox domain factor Oct-1 (-28/-34), several recognition sites for the putative zinc transcription factor GATA, which is expressed in high levels in the pancreas and gut-derived cells [23]. Other putative sites for enhancer elements, AP1, AP2, AP3 and AP4, heat-shock factor (HSF) [24], nuclear factor of activated T cells (NFAT) [25], AML transcription factor [26] and c-Ets-1 transcription factor [27] are also identified in the 5' flanking region.

3.2 Determination of the transcription start sites

A primer extension analysis was performed to determine the transcription start site of the gene. As shown in figure 4, there are three regions detected with stronger intensities, each region was found with a number of individual start sites. These regions are the major transcription start sites. The farthest extension band was observed at 49bp upstream of the ATG translation initiation codon and is an adenine residue. This nucleotide is positioned just downstream to an SP1-binding site found in the promoter region. All these regions lie within 50bp upstream to the translation initiation codon.

3.3. Mapping of the human gC1qR gene (*ClqBP*) promoter element in PANC-1 cells

To localize the essential promoter region for human gC1q-R gene expression, several 5' and 3' deletion clones were constructed by cloning various restriction fragments produced by PCR into a reporter vector pGL3-Basic. Since these fragments were cloned upstream to a

luciferase reporter gene, transcriptional activities of the promoter-luciferase cartridge could be studied by transfecting various clones to the human cell line PANC-1 (Figure 5). It was found that the essential region for promoting transcription is very close to the start site of translation. By comparing luciferase activities between constructs ERES 2 to ERES 7 and that of constructs ERES 7 with ERES 2H and 2P, the essential promoter region could be mapped to a region spanning from the translation start site to -364 bp. This observation is supported by the fact that no significant difference in the luciferase activities was observed upon deletion of the 5' sequence from -364 to -1319 (ERES 7). However, in the two constructs in which the 3' region was deleted, the luciferase activities were completely abolished. Since no 5' deletion studies were carried out beyond -364, we cannot conclude and finely map the essential promoter elements in the -1 to -365 region upstream of the translation start site. Experimental data also showed that inverting the 1.3 kb 5'flanking region (ERES 1) greatly decreased the transcription activities.

3.4 Deletion analysis on the other human cell lines

Three constructs, ERES 2, ERES 7 and ERES 2P were transfected to the other human cell lines to test the cell specificity of the promoter element as well as to map the element. All the four cell lines MDA-MB-231, SVG P12, 293-EBNA and HuTu80 showed enhanced level of luciferase activities for constructs ERES 2 and ERES 7, ranging from tens to hundreds folds depending on cell lines (Figure 6). However, similar to the results obtained in PANC-1, deletion of the 3' region of the promoter (ERES 2P) completely abolished the promoter-reporter luciferase activities in all of these four cell lines. It was surprising that the removal of the TATA box at -399/-406 and the CCAAT box at -410/414 in ERES 7 did not significantly affect the transcription efficiency of the promoter. This implied that there are strong promoter elements in the -1 to -364 region of the 5' flanking region.

3.5 Interaction between DNA-specific nuclear protein factors and promoter elements

In this -1 to -364 regions, seven GC rich sequences with high homology to the consensus SP1 binding site were identified (Figure 3). In order to test if there is any nuclear factor interacting with these GC rich sequences, gel shift assays (Figure 7A) were performed with nuclear extracts of PANC-1 cells. Seven pairs of synthetic oligonucleotides in both sense and antisense orientations (Table 1B) were designed based on these GC-rich motifs. Out of the eight oligo pairs (including a negative control that is not GC-rich), only the L6 probe formed DNA-protein complexes with the nuclear extract. In the competitive assay, decreased intensity of the hybridization bands was observed (Figure 7B) when the concentrations of the unlabeled L6 probe were 10 to 100 folds to the labeled L6 probe. When an unrelated DNA probe were used for the competition assay (Figure 7B), one (complex 1) out of the four bands also showed a reduced signal, indicating that complex 1 is a non-specific signal. To reveal the identity of the complexes, monoclonal anti-Sp1 antibodies (Research Diagnostics Inc., New Jersey) and BSA were mixed with the complexes in a super shift assay. It was observed that the DNA-protein complex 4 was only shifted by anti-Sp1 antibodies but not by BSA, indicating that complex 4 is a complexes between oligonucleotide L6 and Sp1 (Figure 8).

4. Discussion

Like the mouse gene, the human *gClqBP* gene contains six exons and five introns and the sizes of each exon do not differ much except the exon 6 [28]. The exon 6 in human *ClqBP* gene is about 100 bp larger than the mouse gene, due to a longer 3' untranslated region defined by the polyadenylation sites. Even though the sizes of exons between the human and mouse genes are almost the same, three of the introns of the human *ClqBP* gene have

expanded sizes when compared to their mouse counterparts. The sizes of intron 2 in human and mouse genes are 3156 bp and 2071 bp, respectively, which makes the human gene approximately one kb larger than the mouse gene.

Promoter elements of the human gene were identified by computer analysis and transfection studies. Several TATA boxes and CCAAT boxes were found on the 1.3 kb 5' flanking region. TATA box is a crucial positioning component of the core promoter, and is usually located about 25 bp upstream of the transcription start site and it constitutes the only upstream promoter element that has a relatively fixed location with respect to the transcription start site. CCAAT boxes are often located close to the TATA boxes, but it can function at distances that vary considerably from the transcription start site in either orientation. Mutation studies have suggested that the CCAAT box play a strong role in determining the efficiency of the promoter; while its inclusion increases promoter strength.

Transcription start sites are usually located downstream in close proximity to TATA elements in the majority of promoters. However, the transcription start sites (-49 to -29) of the human C1qBP gene are around 350 bp from the closest TATA element (-399) found in the 5' flanking region of the gene. The 5' flanking region of the human gene contains an 8-bp consensus TATA box with sequence TATATATA at -399/-406 and a CCAAT box at -410/414, whereas in mouse gene, the TATA box closest to the ATG start codon is located approximately at -306/-309 [29]. However, in the promoter studies, the construct ERES7 which didn't carry any TATA nor CCAAT boxes still gave more than 90% transcription efficiency, indicating that none of the TATA boxes or CAAT boxes was essential for the transcription of Luciferase mRNA in the cell lines under test. Instead, deletion studies carried out in all cell lines showed complete abolishment in promoter-reporter activity when the 364 bp adjacent to the initiation methionine was deleted (ERES 2P). This 364 bp promoter sequence, similar to the large stretch of GC-rich sequence in the mouse gene, especially from

nucleotides -1 to -200 (75% G+C content) [28], contains seven GC-rich Sp1 sites, which are frequently linked to the transcriptional control of genes lacking a functional TATA box. As no TATA element was found to be essential for the activity of *gC1qBP* promoter, Sp1 sites in the upstream region close to the transcription start site are suggested to be very important for the promoter activity of the gene. In fact, one out of these seven GC-rich Sp1 sites (-96 to -76) was found to bind specifically to PANC-1 nuclear proteins in gel mobility shift assays, and one of these nuclear factors was further proved to be Sp1 binding factor in supershift assays employing anti-Sp1 antibodies. These results show that binding of Sp1 to the SP1 binding site located at around 80 bp upstream to the translation initiation codon may play an important role in transcription control in human *gC1qBP* gene.

With regards to the number of various transcription control elements in the 5'-flanking region, it may be possible that different transcription control elements are essential for the gene expression in different human cells.

The three dimensional structure of gC1qR was revealed by X-ray crystallography. The mature protein molecule has one N-terminal α -helix followed by seven consecutive antiparallel β -strands and two C-terminal α -helices [29]. Three molecules form a dough-nut shape quaternary structure with an internal channel of 10 Å diameter. By aligning the exon boundaries and the three dimensional structure, it is observed that: exon 1 encodes the first 77 amino acid residues containing the mitochondria-targeting sequence[14]; exon 2 encodes the N-terminal α -helix and the first two β -strands; exon 3 encodes β -strand 3; exon 4 encodes β -strands 4 and 5; exon 5 encodes β -strands 6 and 7 and the last exon encodes for the C-terminal α -helices. As gC1qR was shown to be a multi-ligand binding protein, it would be interesting to map the binding sites on gC1qR to various ligands.

More recently, gC1q-R was found to function as a receptor for the internalin B (InlB) invasion protein of *Listeria monocytogenes* [30], and to bind to protein A of *Staphylococcus*

aureus [31]. Another report showed that the hepatitis virus C core antigen can interact with gC1qR and the anti-proliferation effects of Hepatitis virus C core antigen on activated T cells can be reversed by anti-gC1qR monoclonal antibodies[16]. All these reports further suggest that gC1qR is a biologically important, widely distributed, multiligand binding and multifunctional protein [1].

Figure Legends

Figure 1. The genomic organization of the human C1qBP gene. Exons are highlighted as dark boxes. Dark, bold lines below the gene indicate the overlapping subclones. Arrows indicate the position of sequencing primers and the approximate length sequenced by each primer. Restriction enzyme sites are shown at locations above the gene.

Figure 2. DNA sequences of the human C1qBP gene. The first nucleotide of the translation start codon 'ATG' is assigned as 1. Exon-intron boundaries are shown. All intronic sequences, the 5' flanking region of the gene and the 3' flanking region as defined by the cDNA amino acid coding sequences, are shown as small letters. A poly A signal located 369bp from the stop codon was underlined.

Figure 3. Analysis of the 5' flanking region of the human C1qBP gene. The first nucleotide of the translation start codon 'ATG' (bolded) is assigned 1, and the preceding nucleotide is assigned -1. All potential transcription factor binding sites, promoters and elements that may be responsible in transcription control are underlined and labeled below the line.

Figure 4. Primer extension analysis of the human C1qBP gene promoter. Lanes 1 to 4 are sequencing ladders generated by the same primer (G1A2). Lane 5 is the primer extension reaction product. The position of the nucleotides marked on the left side of the gel was designated according to the translation initiation codon ATG, which was assigned as +1. Three regions of strong signal were identified. The farthest extension product was mapped at 49 bp upstream to the ATG codon.

Figure 5. Promoter deletion assay in PANC-1 cell. Schematic diagrams of various promoter-luciferase constructs were shown on the left. Arrows indicate the orientation of the sequence. Values indicated on top of arrows show respective position of the construct at the 5' -flanking region of the gene. The relative promoter activities of various constructs, after normalized by the activity of the pGL3-Basic vector (Basic), were shown on the right. This figure represents the mean of three individual transfection experiments.

Figure 6: Promoter assays in different human cell lines. The relative promoter activities of all cell lines were normalized by the pGL3-Basic activity. The figure represents the mean of two individual transfection experiments.

Figure 7: Gel-shift assays. 8 µg of PANC-1 nuclear extracts were used for each individual reaction. (A) Gel shift was only observed in the L6 DNA probe, corresponding to the sequence -96 to -76 of the human *ClqBP* gene. (B) Competition assay was performed using 8 µg of the PANC-1 nuclear extracts in the presence of an increasing concentration of unlabeled DNA probe (0, 10 and 100 fold). The formation of DNA-protein complexes was gradually inhibited only by the specific oligo but not by an unrelated oligo. In the negative control experiments (-ve) using an unrelated oligo, only the formation of DNA-protein complex 1 was inhibited but the formation of the other 3 complexes were not disturbed. This showed that DNA-protein complex 1 was an artifact.

Figure 8: Supershift assay was performed using the L6 oligo as a probe. When 4 µg of anti-Sp1 antibodies were added to the mixture, the band of DNA-protein complex 4 shifted upwards from its original position (as shown when no antibody was added) while the other

three bands were unaffected. DNA-protein complex 4 should be a complex of oligo L6 and the transcription factor Sp1. BSA was used as the negative control in this assay, which was unable to display the DNA-protein complex 4.

Acknowledgement

This work was supported in part by grants RGC HKU198/95M from the Hong Kong Research Grant Council (B.L.L.), and RPG-95068-04-CIM from the American Cancer Society (BG).

References

- [1] Ghebrehiwet, B. and Peerschke, E.I. (1998) *Immunobiology* **199**: 225-238
- [2] Ghebrehiwet, B., Lim, B.L., Peerschke, E.I., Willis, A.C. and Reid, K.B. (1994)
J.Exp.Med. **179**: 1809-1821
- [3] Joseph, K., Ghebrehiwet, B. and Kaplan, A.P. (1999) *Clin. Immunol.* **92**: 246-255.
- [4] Joseph, K., Ghebrehiwet, B., Peerschke, E.I., Reid, K.B. and Kaplan, A.P. (1996)
Proc.Natl.Acad.Sci.U.S A. **93**: 8552-8557
- [5] Lu, P.D., Galanakis, D.K., Ghebrehiwet, B. and Peerschke, E.I., (1999) *Clin.Immunol.* **90**:
360-367
- [6] Lim, B.L., Reid, K.B., Ghebrehiwet, B., Peerschke, E.I., Leigh, L.A. and Preissner, K.T.
(1996) *J.Biol.Chem.* **271**: 26739-26744
- [7] Xu, Z., Hirasawa, A., Shinoura, H. and Tsujimoto, G. (1999) *J. Biol. Chem.* **274**:
21149-21154
- [8] Storz, P., Hausser, A., Link, G., Dedio, J., Ghebrehiwet, B., Pfizenmaier, K. and Johannes,
F. (2000) *J. Biol. Chem.* **275**: 24601-24607.
- [9] Yu, L., Loewenstein, P.M., Zhang, Z. and Green, M. (1995) *J. Virol.* **69**: 3017-3023.
- [10] Luo, Y., Yu, H. and Peterlin, B.M. (1994) *J. Virol.* **68**: 3850-3856.
- [11] Mathews, D.A. and Russell, W.C. (1998) *J. Gen. Virol.* **79**: 1677-1685.
- [12] Wang, Y., Finnan, J.E., Middeldorp, J.M. and Hayward, S.D. (1997) *Virology* **236**: 18-29
- [13] Bruni, R. and Roizman, B. (1996) *Proc.Natl.Acad.Sci.U.S A.* **93**: 10423-10427

- [14] Dedio, J., Jahnen-Dechent, W., Bachmann, M. and Muller-Esterl, W. (1998) *J. Immunol.* **160**: 3534-3542
- [15] Soltys, B.J. and Gupta, R.S. (1999) *TIBS* **24**: 174-177
- [16] Kittlesen, D.J., Chianese-Bullock, K.A., Yao, Z.Q., Braciale, T.J. and Hahn, Y.S. (2000) *J.Clin. Invest.* **106**:1239-1249
- [17] Guo, N., Weremowicz, S., Lynch, N., Lim, B.L., Schwaeble, W., Peerschke, E.I., Morton, C.C., Reid, K.B., Ghebrehwet, B. and Sastry, K.N. (1997) *Cytogenet.Cell Genet.* **77**: 283-284
- [18] Lynch, N.J., Reid, K.B., van-den-Berg, R.H., Daha, M.R., Leigh, L.A., Ghebrehwet, B., Lim, W.B. and Schwaeble, W.J. (1997) *FEBS Lett.* **418**: 111-114
- [19] Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) *Nucleic Acids Res.* **23**: 4878-4884
- [20] Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., Podkolodny, N.L. and Kolchanov, N.A. (1998) *Nucleic Acids Res.* **26**: 364-370
- [21] Lee, K.A.W., Bindereif, A., and Green, M.R. (1988) *Gene anal. Tech.* **5**:22-31
- [22] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhi, K (1987) *Current Protocols in Molecular Biology*, Wiley, NY.
- [23] Huggon, I.C., Davies, A., Gove, C., Moscoso, G., Moniz, C., Foss, Y., Farzaneh, F. and Towner, P. (1997) *Biochim.Biophys.Acta* **1353**: 98-102
- [24] Kiang, J.G., Gist, I.D. and Tsokos, G.C. (2000) *Mol. Cell Biochem.* **204**: 169-78

- [25] Ranger, A.M., Gerstenfeld, L.C., Wang, J., Kon, T., Bae, H., Gravallesse, E.M., Glimcher, M.J. and Glimcher, L.H. (2000) *J.Exp.Med.* **191**: 9-22
- [26] Lutterbach, B. and Hiebert, S.W. (2000) *Gene* **245**: 223-235
- [27] Macias, D., Perez, P.J., Garcia, G.L., Carmona, R. and Munoz, C.R. (1998) *Anat.Embryol.(Berl.)* **198**: 307-315
- [28] Lim, B.L., White, R.A., Hummel, G.S., Schwaeble, W., Lynch, N.J., Peerschke, E.I., Reid, K.B. and Ghebrehiwet, B. (1998) *Gene* **209**: 229-237
- [29] Jiang, J., Zhang, Y., Krainer, A. R. and Xu, R. M. (1999) *Proc.Natl.Acad.Sci.U.S.A.* **96**: 3572-3577
- [30] Braun, L., Ghebrehiwet, B. and Cossart, P. (2000) *EMBO J.* **19**: 1458-1466
- [31] Nguyen, T., Ghebrehiwet, B. and Peerschke, E.I.B. (2000) *Infect. Immun.* **84**: 2061-2068

Table 1. Primers designed for construction of recombinant plasmids and gel shift assays.

(A) Primers designed for construction of recombinant plasmids.

Primers	Primer sequences (4 bases and RE sites added at 5')	Position in 5' flanking region
Pro0 <i>Sac</i>	5' attagagctc CTGCCCTTGAGGATG 3'	-1314 to -1300
Pro200 <i>Sac</i>	5' attagagctc AACAAATCAACACAGAAGACTT 3'	-1121 to -1101
Pro400 <i>Sac</i>	5' attagagctc CGACCAACCCGCTTCA 3'	-916 to -901
Pro600 <i>Sac</i>	5' attagagctc CAGGAACCTCTACG 3'	-713 to -702
Pro800 <i>Sac</i>	5' attagagctc GTTGTTCCTGAGGCCT 3'	-506 to -489
Pro954 <i>Sac</i>	5' attagagctc GTGGAAGGGTCTGGT 3'	-364 to -350
G1A2 <i>Bgl</i>	5' attagatctc GCGGAAACGACTGCGAA 3'	-18 to -1 (inverted)
Pro0 <i>Bgl</i>	5' attagatctc CTGCCCTTGAGGATG 3'	-1314 to -1300
G1A2 <i>Sac</i>	5' attagagctc GCGGAAACGACTGCGAA 3'	-18 to -1 (inverted)

(B) Primers for gel shift assays. All primers were designed from their GC context with respect to their location in the 5' flanking region of the human gC1qBP gene. "S" indicates sense primers and "A" is the antisense version of its sense primer. L8S and L8A, which do not contain GC-rich sequences, were designed as negative controls.

Primers	Primer sequences (4 bases and RE sites added at 5')	Position in 5' flanking region
L1S	5' AGAGCTGGCG GGAGACAA	5' -348 to -331
L1A	3' TTGTCTCCCG CCAGCTCT	3' -331 to -348
L2S	5' AGGCTTCCCG GGCTGCAC	5' -188 to -171
L2A	3' GTGCAGCCCG GGAAGCCT	3' -171 to -188
L3S	5' AGCAGAGGGC GCGATTCCG	5' -160 to -142
L3A	3' CGGAATCGCG CCCTCTGCT	3' -142 to -160
L4S	5' GCGATTCCCG CCGCTGGC	5' 150 to -133
L4A	3' GCCAGCGGGC GGAATCGC	3' -133 to -150
L5S	5' GAGCGCCGGC GCGGGAAGC	5' -120 to -102
L5A	3' GCTTCCCGCG CCGGCGCTC	3' -102 to -120
L6S	5' CGGGCAGGGG CGGGGCTTCC G	5' -96 to -76
L6A	3' CGGAAGCCCC GCCCCTGCCC C	3' -76 to -96
L7S	5' TCAGGTCGCG GGGCGCCTAG	5' -67 to -48
L7A	3' CTAGGCGCCC CGCGACCTGA	3' -48 to -67
L8S	5' GGCCACTAGA GGAATTAAT	5' -264 to -245
L8A	3' TTTAATTCCC TCTAGTGGCC	3' -245 to -264

Table 2. Expression plasmids for luciferase promoter assay.

Constructs	Primers employed	RE sites at 5' end	RE sites at 3' end	Fragment size	deletion from
RES 1	Pro0 <i>Bgl</i> + G1A2 <i>Sac</i>	<i>Bgl</i> II	<i>Sac</i> I	1.3 kb	Nil (reverse)
RES 2	Pro0 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>Bgl</i> II	1.3 kb	Nil
RES 3	Pro200 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>Bgl</i> II	1.1 kb	5'
RES 4	Pro400 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>Bgl</i> II	0.9 kb	5'
RES 5	Pro600 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>Bgl</i> II	0.7 kb	5'
RES 6	Pro800 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>Bgl</i> II	0.5 kb	5'
RES 7	Pro954 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>Bgl</i> II	0.4 kb	5'
RES 2H	Pro0 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>Hind</i> III	0.67 kb	3'
RES 2P	Pro0 <i>Sac</i> + G1A2 <i>Bgl</i>	<i>Sac</i> I	<i>PsPA</i> I	1.1 kb	3'

Table 3. Exon-intron boundaries of the human C1qBP gene.

exon no.	location on cDNA	exon size	exon 5' seq	exon 3' seq	intron no.	intron 5' seq	intron 3' seq	intron size	amino acid interrupted
1	1-232	232bp		CACACCGACG	1	GTGAGGTTTA	TCTTTTCT AG	568bp	glycine-78
2	233-383	151bp	GAGACAAAGC	CCGGGGAAAA	2	GTAAGTACTG	GCCGTTTT AG	3156bp	lysine-128
3	384-477	94bp	AATCACGGTC	AGAACAGGAG	3	GTAAGCTAAT	GTTTCCCC AG	1104bp	NIL
4	478-576	99bp	CCTGAACTGA	AGAGGATGAG	4	GTATGCAGGA	TTCTTCT AG	253bp	NIL
5	577-699	123bp	GTTGGACAAG	CTTGGACTGG	5	GTGAGTGCTT	TTCTTTGC AG	128bp	NIL
6	700-845	516bp	GCCTTATATG						

The exon-intron boundaries follow the GT/AG rule (as indicated by bold letters).

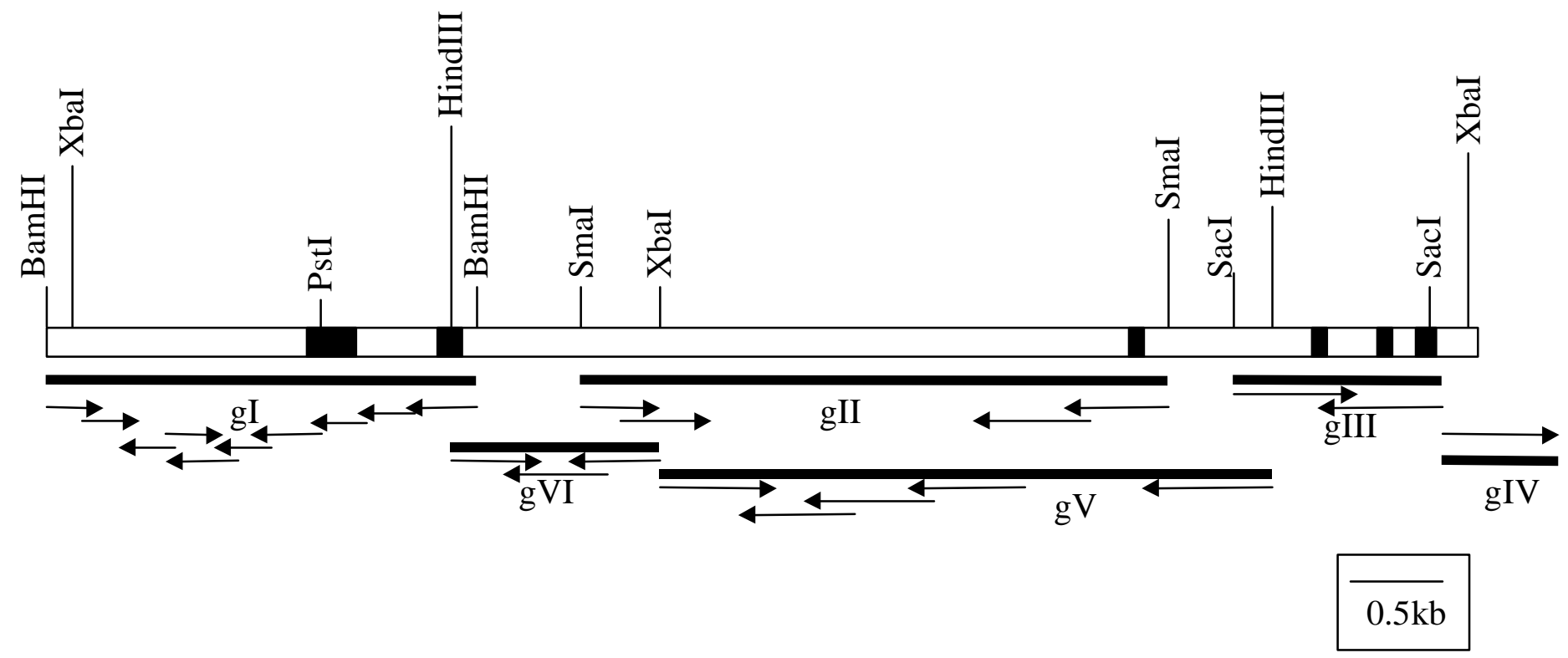


Figure 1


```

-1311 tcctgcctt gaggatgtca cagtctggtt gggttta .....5' flanking region
      ..... catctgcacg tgttcgcagt cgtttccgcy
1   ATGCTGCCTC TGCTGCGCTG CGTGCCCCGT GTGCTGGGCT CCTCCGTCGC CGGCCTCCGC
61  GCTGCCGCGC CCGCCTCGCC TTTCCGGCAG CTCCTGCAGC CGGCACCCCG GCTGTGCACC
121 CGGCCCTTCG GGCTGCTCAG CGTGCGCGCA GGTTCGAGC GGCGGCCGGG CCTCTGCGG
181 CCTCGCGGAC CCTGCGCCTG TGGCTGTGGC TGCGGCTCGC TGCACACCGA CGgtgaggtt
241 tagcgggccc ..... intron 1 (568bp) .....
      ..... tatctggatt tcttttctag
801 GAGACAAAGC TTTTGTGAT TTCCTGAGTG ATGAAATTA GGAGGAAAGA AAAATTCAGA
861 AGCATAAAAC CCTCCCTAAG ATGTCTGGAG GTTGGGAGCT GGAACGAAT GGGACAGAAG
921 CGAAATTAGT GCGGAAAGTT GCCGGGAAA Agtaagtact ggtctgggga .....
      .....intron 2 (3156bp)..... cttgtctggt ttccttggc
4101 gtttttagAAT CACGGTCACT TTCAACATTA ACAACAGCAT CCCACCAACA TTTGATGGTG
4161 AGGAGGAACC CTCGCAAGGG CAGAAGTTG AAGAACAGGA Ggtaagctaa tactataatc
      .....intron 3 (1104bp) ..... ttcttgttcc cccagCCTGA
5311 ACTGACATCA ACTCCCAAT TCGTGGTTGA AGTTATAAAG AATGATGATG GCAAGAAGGC
5371 CCTTGTGTTG GACTGTCATT ATCCAGAGGA TGAGgtatgc aggatggagt .....
      ..... intron 4 (253bp)..... ccca atattgttc ttctcagGTT
5661 GGACAAGAAG ACGAGGCTGA GAGTGACATC TTCTCTATCA GGAAGTTAG CTTTCAGTCC
5721 ACTGGCGAGT CTGAATGGAA GGATACTAAT TATACTCA ACACAGATTC CTTGGACTGG
5781 gtgagtgett gataaggtgg ..... intron 5 (128bp).....
      ..... tttactagt aaatattggt ctttgcagGC
5911 CTTATATGAC CACCTAATGG ATTTCTTGC CGACCGAGGG GTGGACAACA CTTTTCAGCA
5971 TGAGCTGGTG GAGCTCAGCA CAGCCCTGGA GCACCAGGAG TACATTACTT TTCTTGAAGA
6031 CCTCAAGAGT TTTGTCAAGA GCCAGTAGAG CAGACAGATG CTGAAAGCCA TAGTTTCATG
6091 GCAGGCTTTG GCCAGTGAAC AAATCTACT CTGAAGCTAG ACATGTGCTT TGAAATGATT
6151 ATCATCCTAA TATCATGGGG GAAAAAATAC CAAATTTAAA TTATATGTTT TGTGTCTCA
6211 TTTATTATCA TTTTTTCTG TACAAATCTA TTATTCTAG ATTTTGTAT AACATGATAG
6271 ACATAAAAT TGGTTATCTC CTCCAAGGCA GTTTGTCTTT TTCTATTCCT CCCCCTCAA
6331 CCTGCGTCAC AAAAGACCAA GAACAGATGT CGGAAAAGTT TTTTTTCTT CAGTATTGTT
6391 TAAAAGTTTC AATACAAAAT AAGTTATAAA TAAAaggctt gtatgtacaa ggctcctcag
6451 agggaatgag ttgtcttcaa ccccatagaa tgatgtgagt c

```

Figure 2

-1374 GGATC

-1314 CTGCCCTTGA GGATGTCACA GTCTGGTTGG GTTTAACCTG GTTAAAGCAG CAACGCTAGG
 SP1

-1254 AGGTTCTCTGT TGAAAAGAGA ATGAAGCATA TGGGCTGGGT GCGCACAAAGC ATGAGTTCCG

-1194 AGCCGGTGGA AAAAAAGGTT CTAGAAGGCC TCTTTCAGGA AAACCCCTCAG ACTTTCCTTT
 NFAT STAT/NFAT

-1134 GCTCTCACAA CACAACAATC AACACAGAAG ACTTATGTGG CCAAAGGTTT GGGTTTTTCC

-1074 CCTCCACACT AAGCAGCAGA CACCAGCTGG GTGTCCTCCA ATTCAAATCCC AACAGTATCT
 AP4 CAAT box

-1014 ACCTGGGAGAT AGTGTCAGAT CCCATAGGTT GAGGGCTCAG TCCCCAAGAC CGCCCCCACC
 GATA AP2 SP1

-954 TTCAGACAGC AATGGCAAGT CTGGGCCTCT GAAACTTCCG ACCAACCCTG TTCAGTTGG
 HSF

-894 GGTCCCCACG ACCCCCCTCT TTGGATTATG TTAACACT GGAGCAGTTC ACAGAACTCA
 AP3

-834 GGGAAACACA TTCACCGGTT TCATATAAAG GATATGACCA AGGATACAGA TGAGATGAGC
 TATA box GATA

-774 AGGGTGGGGA ATGGGAAGA AGCTCCGAGG TTCTTGCTT ACCCTGGGCC ACCACCCTCC
 AP2 AP2

-714 AGGAACCTCT ACGTATTCAG CTGTCAGAGG CTCTGCAAAC CCTCTCCTCT TGGGTTTTTA

-654 TGGAAGCTC TGGACATCAG CATCCCTTCC CCCAGGGTAT ATGGTGGGAC CTTCTCTGGG
 HSF TATA box

-594 AAGGGACTTA AGACCCATAA TCAGAAATGT GGAAGGAGAT TAGATCCTG CTTTGGGGCA
 AP1 GATA SP1

-534 GGTAAAAGGA GGGCAGGAGA AGGGGAGAGT CTGTTTCTG AGGCCTAACA CACCCAACAT
 C/EBP SP1 NFAT

-474 TATCACAGAA CATTGTAACA AGGGCTATAG GAGTTACGAC CCAGGAACTG TGGATGAAAA
 CAAT box TATA box

-414 CCAATACATA TATATATCAT AACACCACAG CCTCCTAGAG GAAAGTGACC GTGGAAGGGT
 CAAT box TATA box AML NFAT

-354 CTGGTGAGAG CTGGCGGGAG ACAAGAGTGC ATTCCACGCA GCCCAAAGC CTTGCGTGCT
 SP1

-294 TTGAGAACAG CAGGTGGAAA GTGGGTAAGC GGCCACTAGA GGAATTAAA CAGCTTCGGA

-234 CAGCGGACTC GGCAGCGCAG CAAGAAAGCT CCGGAATCCG GATTGAGGC TTCCCGGGCT
 HSF SP1

-174 GCACCGGAGT CTCCAGCAGA GGGCGCGATT CGCCCCGCTG GCCTTAGGTC GTCAGAGCGC
 SP1 SP1 AP2

-114 CGGCGCGGGA AGCGCTTCCG GGCAGGGGCG GGGCTTCCGG CGGCGCCTCA GGTCCGCGGG
 SP1 c-Ets-1 SP1 c-Ets-1 SP1

-54 CGCCTAGGCC TGGGTTGTCC TTTGCACTG CACGTGTTCG CAGTCGTTTC CGCGATGCTG
 Oct-1

Figure 3

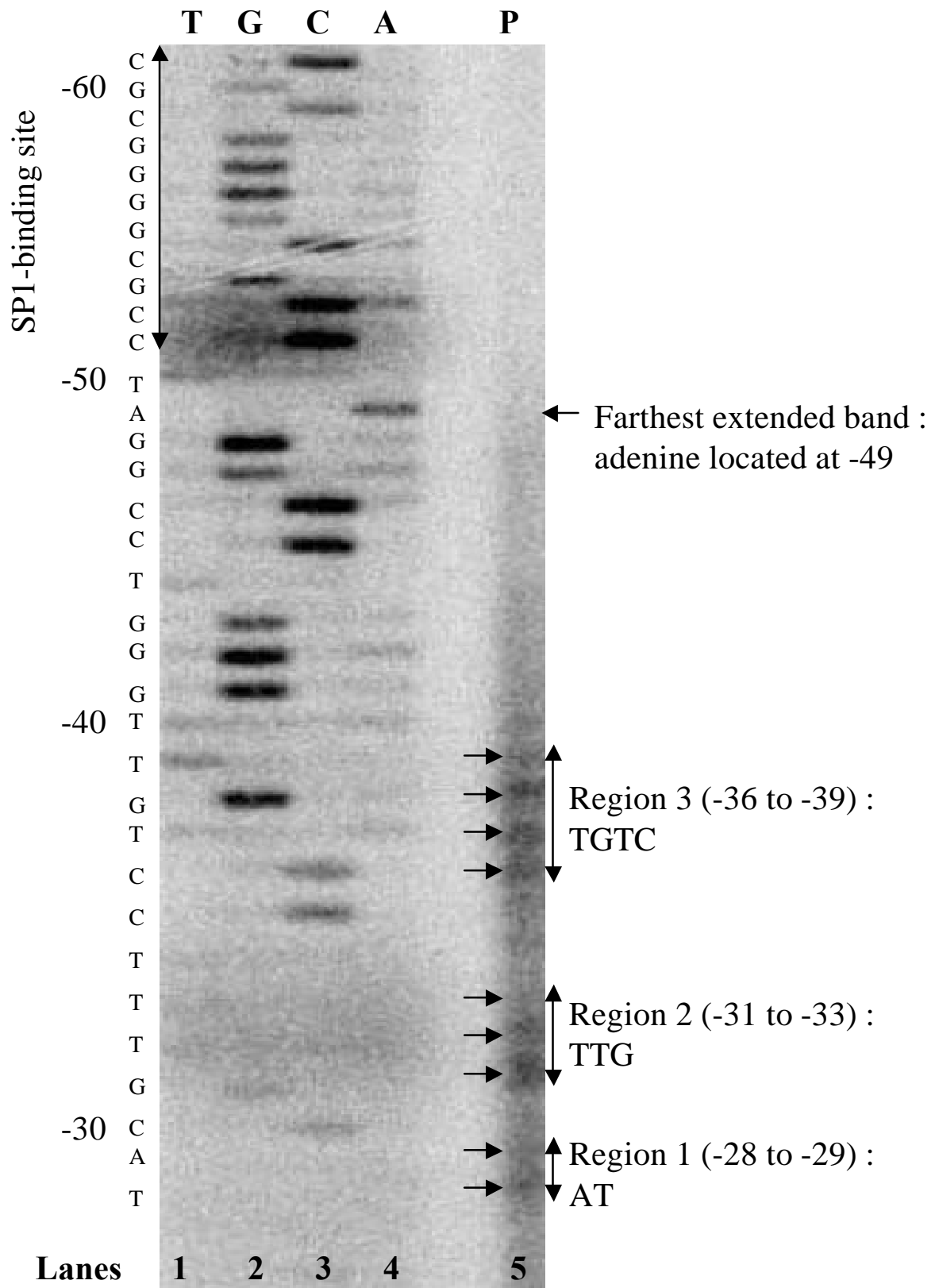


Figure 4

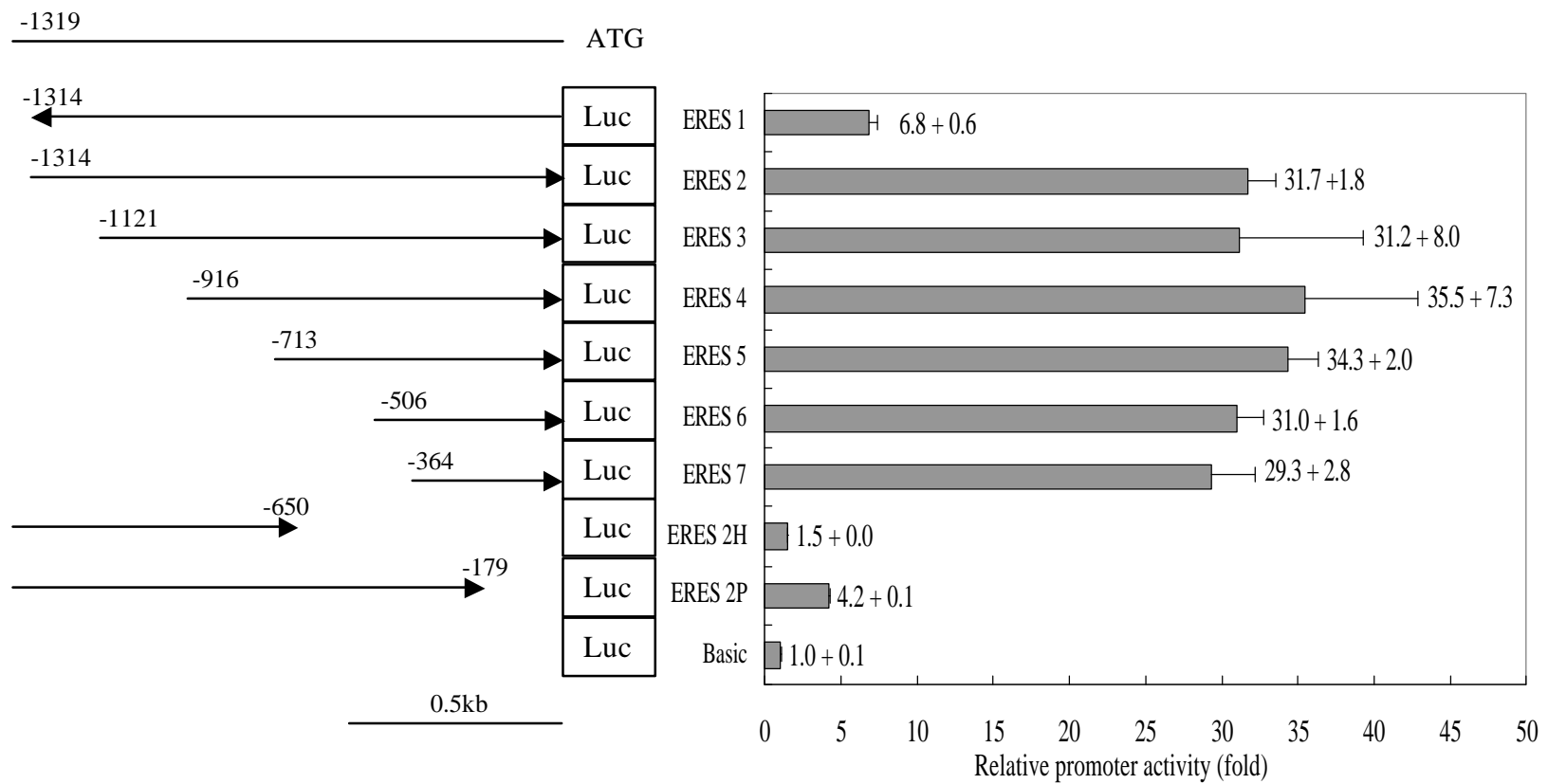


Figure 5

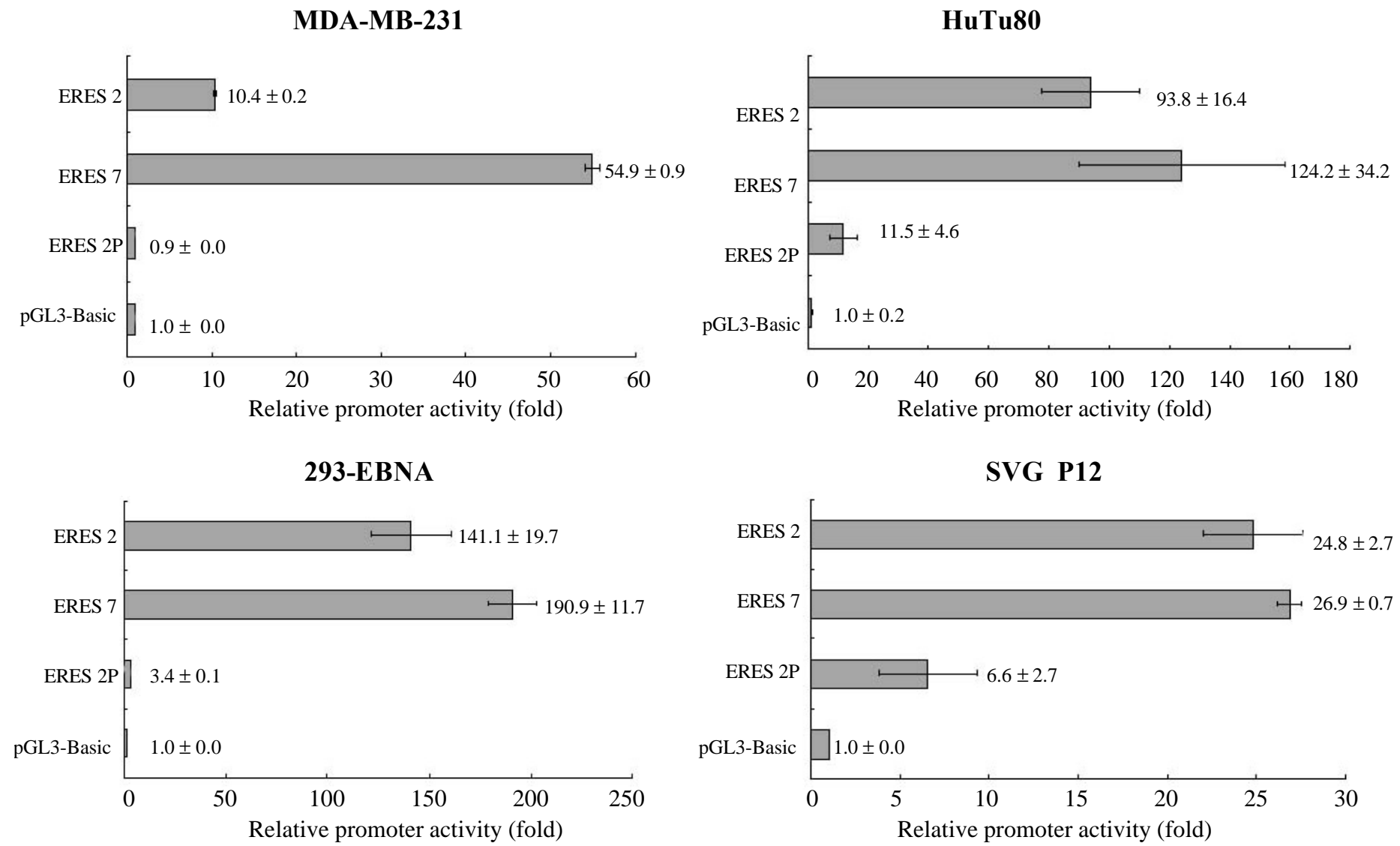


Figure 6

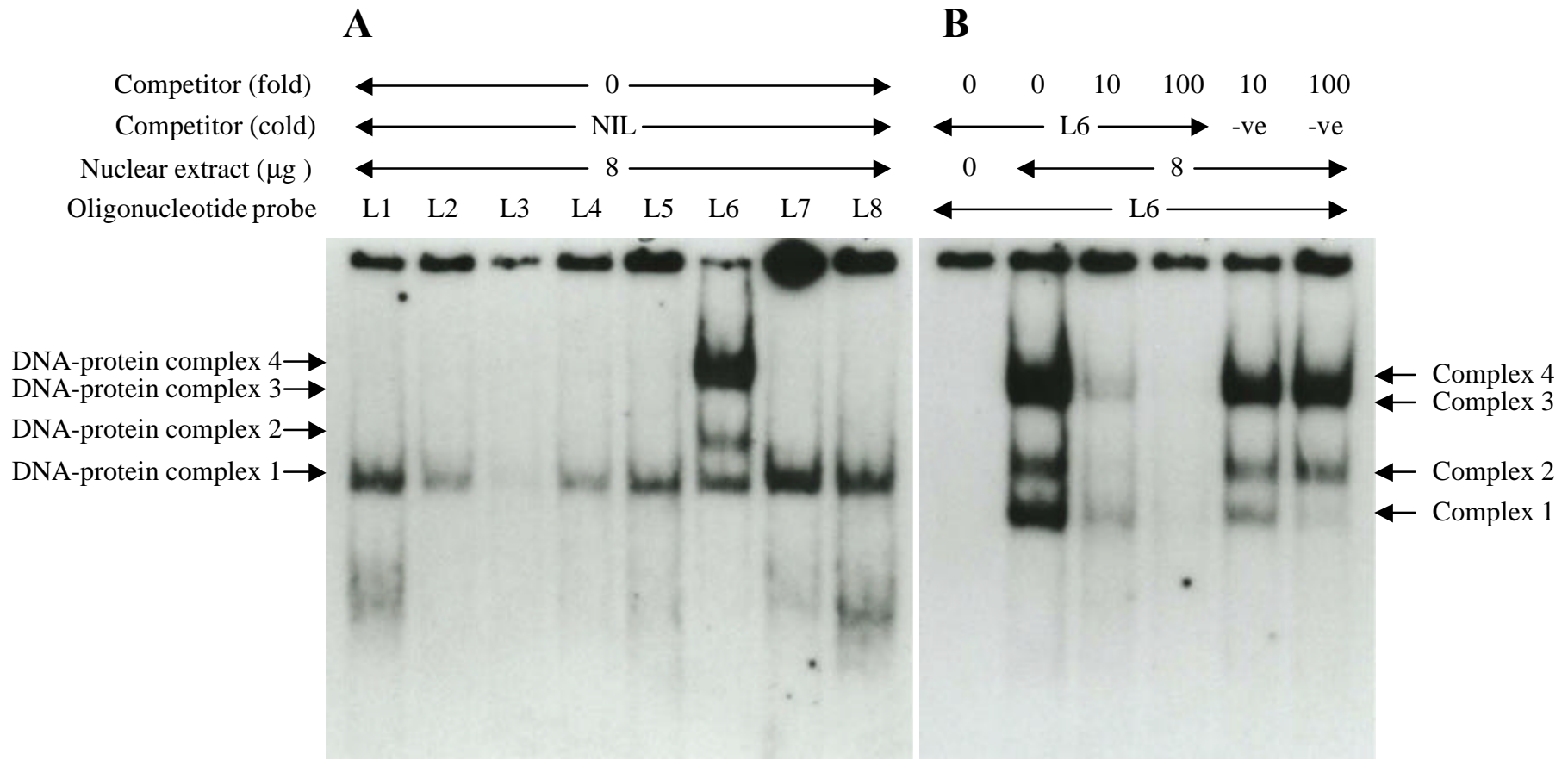


Figure 7

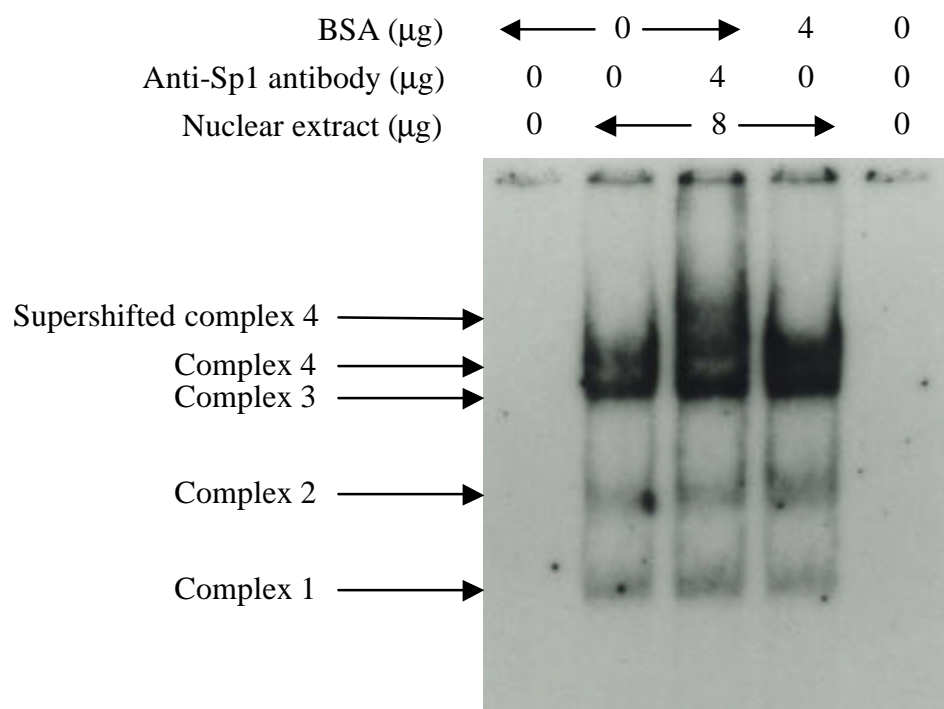


Figure 8