Genomic Structure and Transcriptional Regulation of the Early B Cell Gene *chB1*

Ryo Goitsuka, Hiroshi Mamada, Daisuke Kitamura, Max D. Cooper and Chen-lo H. Chen

*J Immunol* 2001; 167:1454-1460; doi: 10.4049/jimmunol.167.3.1454

http://www.jimmunol.org/content/167/3/1454
Genomic Structure and Transcriptional Regulation of the Early B Cell Gene chB1

Ryo Goitsuka,† Hiroshi Mamada,§ Daisuke Kitamura,† Max D. Cooper,‡§ and Chen-lo H. Chen

The avian B cell differentiation Ag chB1 is a membrane glycoprotein relative of the mammalian B cell differentiation Ag CD72. Unlike CD72, this C-type lectin is expressed in relatively high levels on immature B cells in the bursa of Fabricius and is down-regulated on mature B cells in the periphery. An immunoreceptor tyrosine-based inhibitory motif in the chB1 cytoplasmic tail suggests a potential regulatory role in intrabursal B cell development. To gain further insight into the selective expression and function of chB1, we determined the genomic organization of chB1 and examined the mechanism of its transcriptional regulation.

The 8-exon chB1 gene proved to have very similar organization to that of mouse CD72, further supporting the idea that chB1 is a CD72 relative. As for mouse CD72, the chB1 promoter region lacks a TATA box but contains a conserved initiator element. The 131-bp region (−161 to −30) proximal to the transcriptional start site, which contains a potential early B cell factor binding site, is essential for the B lineage stage-specific transcription of chB1, whereas PU.1 and B cell-specific activator protein/Pax5 have been shown to play important roles in CD72 promoter activity and cell-type specificity. This analysis suggests that differences in transcriptional regulation of these phylogenetically related genes may determine the differences in expression pattern and, therefore, the function of avian chB1 and mammalian CD72 during B cell development. The Journal of Immunology, 2001, 167: 1454–1460.

E arly B cell development, including cell surface IgM expression, replication, and repertoire diversification by gene conversion, takes place in the avian bursa of Fabricius (1–3), although the regulatory elements for these intrabursal processes are largely unknown. To explore the unique interaction between the emerging B cells and their intrabursal microenvironment, we have begun to examine the cell surface molecules expressed by bursal lymphocytes (4). The gene encoding one of these, the chB1 Ag, has been characterized as a member of the C-type lectin family. The chB1 Ag is a type II transmembrane homodimeric glycoprotein consisting of disulfide-linked 52-kDa chains. This cell surface molecule is expressed on intrabursal B cells and on the DT40 B cell line (5), which are distinctive in their expression profiles of antibodies and cell surface molecules. The intracellular domain of chB1 features a consensus immunoreceptor tyrosine-based inhibitory motif (ITIM) characteristic of inhibitory receptors (8). The ligation of inhibitory receptors leads to phosphorylation of the tyrosine residue of the ITIM and recruitment of Src homology (SH2)-containing phosphatases with resultant inhibition of cell activation pathways (8–10).

The ligand for chB1 and the physiological role of chB1 in intrabursal B cell development are still conjectural. Sequence analysis indicates that chB1 is most homologous to mammalian CD72, another type II transmembrane glycoprotein of the C-type lectin family (5, 11, 12). However, CD72 is expressed at all stages of B lineage differentiation except the mature plasma cell stage (13). Although functional studies suggest that CD72 serves as a costimulatory molecule for B cell activation, inhibitory effects have been shown in response to CD72 cross-linkage (13). CD72 contains two ITIMs in the cytoplasmic domain, of which the N-terminal one serves as an in vivo substrate of the protein tyrosine phosphatase SH2 domain-containing phosphatase-1 (14, 15). The other ITIM recruits Grb2 that associates with B cell linker protein/SH2 domain containing leukocyte protein-76/B cell adaptor containing SH2 domain (16). Thus, CD72 may negatively regulate B cell Ag receptor (BCR)-mediated B cell activation (17, 18). It may also transmit a stimulatory signal independent of BCR in that Ab-mediated ligation of CD72 activates lyn, blk, and btk in the absence of syk, which is essential for BCR signaling (19). The mouse CD72 ligand has been identified recently as CD100 (20).

Cross-linkage of chB1 on bursal B cells and the DT40 B cell inhibits their proliferation without having a demonstrable effect on BCR signaling (Ref. 5 and our unpublished observations). To gain additional insight into the physiological role(s) of chB1, we have characterized the chB1 genomic organization and examined the transcriptional regulation of this gene.

Materials and Methods

Abs and cell lines

The M4 monoclonal anti-chicken μ-chain (21) and CB1 anti-chB1 Abs (4) were produced in our laboratory. The avian leukemia virus-transformed
chicken B cell lines DT40 and 249L4 and a Marek’s disease virus-transformed T cell line, MSB1 (22), were maintained in RPMI 1640 with 10% FBS at 40°C in a humidified atmosphere.

Isolation and sequence analysis of the chB1 genomic clone

A chicken genomic library constructed in the pWE15 cosmid vector (Clontech Laboratories, Palo Alto, CA) was screened with a full-length 5′-P-labeled chB1 cDNA probe according to the manufacturer’s protocol. A chβ1 clone was isolated, and an 11-kb EcoRI fragment of the insert that hybridized with the chB1 cDNA probe was subcloned into pBluescript vector (Stratagene, La Jolla, CA) for restriction endonuclease mapping. Enzyme-digested fragments were then subcloned and sequenced by the dideoxy-chain termination method with an automatic DNA sequencer (Applied Biosystems, Foster City, CA). A search for consensus binding sites for transcription factors in the chB1 gene was conducted using the TFMA-TRIX transcription factor binding site profile database and also by manual inspection.

Determination of the transcription initiation site

The transcriptional start site of the chB1 gene was determined by a modified 5′ RACE technique (23) using a Marathon cDNA amplification kit (Clontech Laboratories). Poly(A)+ RNA from the bursa of a 1-day-old chick was reverse transcribed using a chB1 cDNA-specific reverse primer (5′-TGCCCTCAGCGCTGGCTGGCCGTGCC-3′). Double-stranded cDNA was synthesized and ligated with a specific adapter according to the manufacturer’s protocol. 5′ RACE was conducted using a chB1 cDNA-specific reverse primer (5′-CCCTGCCCTTTGCGAACTCTCAGTC-3′) and a primer specific for the adapter. The resulting PCR fragments were subcloned into the pGEM-T Easy vector (Promega, Madison, WI) and sequenced.

Construction of the reporter vectors

A 13-kb BamHI fragment (~1309 to +66) upstream of the translation initiation codon (+81) of the chB1 gene was cloned by blunt-end ligation into the SmaI site of pGL2-Basic luciferase vector in a 5′ to 3′ orientation (Promega). Serial 5′ deletion mutants ending 3′ of base pair 30 were generated using exonuclease III digestions. All constructs were verified by restriction analysis and sequencing. pA-luci was also constructed by ligating the BamHI-HindIII fragment of chicken β-actin promoter into the pGL2-Basic vector (24). The pAct-βGal vector was kindly provided by Dr. T. Yagi (National Institute of Physical Sciences, Okazaki, Japan).

Transfection and luciferase assay

Cell lines were transfected by electroporation (Bio-Rad electroporation apparatus; Bio-Rad, Hercules, CA). The conditions for electroporation were 250 V and 975 mF for DT40 cells and 275 V and 975 mF for 249L4 and MSB1 cells. Cells (5 × 10⁶) were cotransfected with 10 μg test construct and 1 μg pAct-βGal. The pAct-βGal plasmid was used for normalizing the transfection efficiency. Each experiment included pA-luci and pGL-2-transfected samples. pGL-2 was used for subtracting the background luciferase activity. The pA-luci was used as an internal positive control to compare different experiments. The data are presented as the percentage of activity divided by the luciferase activity obtained from cells transfected with pA-luci. At 48 h after transfection, luciferase activity was determined in cell extracts according to the instructions of the luciferase assay kit (Promega). A β-galactosidase assay was conducted using a Galact-light kit (Tropix, Bedford, MA). Light emission was measured in a Lumat LB9501 luminometer (Berthold, Wildbad, Germany).

Results

Genomic organization and nucleotide sequence of the chB1 gene

Sequence analysis of the chB1 gene indicates ~4 kb of genomic DNA from the putative transcription start site to the polyadenylation signal sequence (Fig. 1). Comparison of the chB1 cDNA (5) with this genomic sequence revealed three nucleotide differences within the 3′ untranslated region. Because the cDNA and DNA libraries were derived from chickens of different strains, these differences suggest chB1 gene polymorphism.

The chB1 gene consists of eight exons bordering seven introns: one exon for the 3′ untranslated region and seven exons for the coding region (Figs. 1 and 2). The first exon encodes the 5′ untranslated region and the last 28 amino acid residues of the cyto-

**FIGURE 1.** Nucleotide sequence of the chB1 gene. Exons (boxed) were identified by sequence comparison of the genomic and cDNA sequences. The 5′ boundary of exon 1 was determined by the modified 5′ RACE, and the 3′ end of exon 8 corresponds to the location of the poly(A) tail of the cDNA sequence. Numbers on the right indicate positions in the nucleotide sequence. The potential polyadenylation signal (AATAAA) is in bold. The nucleotide sequence has been deposited in the DDBJ/GenBank/EMBL database (GenBank accession number AB052935).
plasmic region that contains the tyrosine residue of the ITIM. The second exon includes most of the remainder of the cytoplasmic domain. The remaining eight amino acid residues of the cytoplasmic domain and the transmembrane domain are encoded by the third exon. The fourth exon encodes the entire stalk region that separates the transmembrane domain from the potential carbohydrate ligand-binding domain. Exons 5–7 encode the 146 COOH-terminal residues comprising the carbohydrate-binding domain. Most of the 3’ untranslated region is encoded in exon 8. All of the introns begin with the dinucleotide GT and end with the dinucleotide AG, which conforms to the eukaryotic splice consensus sequence (25).

**Structural comparison of the chB1 gene with other C-type lectin genes**

Comparison of chB1 with other C-type lectin genes indicates that the gene organization of the chB1 gene is most homologous with the mouse CD72 gene (Fig. 2). The sequences and locations of the exon/intron boundaries for chB1 and CD72 were found to be similar throughout the coding region, except that the chB1 gene lacks exon 3 of the CD72 gene (Figs. 2 and 3A and Ref. 26).

In the carbohydrate-recognition domain, the similarity of exon/intron boundaries of gene organization extends to other C-type lectins, including CD23 (27, 28), rat asialoglycoprotein receptor type 1 (29), hepatic lectins (30, 31), and Kupffer cell receptors (32, Fig. 2, and data not shown). The carbohydrate-binding domains of the C-type lectins are encoded by three exons, the last exon of which, in most C-type lectin family genes such as rASGPR1 and human CD23, encodes both the COOH-terminal carbohydrate-binding domain and the 3’ untranslated region (Fig. 2). However, the 3’ untranslated regions of the chB1 and CD72 genes are coded by separate terminal exons.

The conservation of the exon/intron structure in chB1 and CD72 includes the codon splitting patterns. All the introns upstream of the exons coding the carbohydrate-recognition domain of chB1 and mouse CD72 genes are located at position 1 within the amino acid codons (Fig. 3A), whereas two introns dividing the carbohydrate-recognition domain are located at positions 0 and 2 within...
the amino acid codons. This pattern of codon splitting is also highly conserved among other C-type lectin receptors (Fig. 3B). These findings suggest that the chb1 and CD72 genes represent a subfamily encoding C-type lectins expressed on lymphocytes that may have diverged from a common ancestor for other C-type lectins.

Identification of the chb1 gene transcriptional initiation site and potential transcription factor binding sites in the 5'-flanking region

The tissue distribution pattern of the chB1 Ag suggests B lineage-specific and differentiation stage-specific transcription regulation of the chb1 gene. Therefore, we examined the DNA sequence from the 5' end to the transcriptional start site to gain insight into the transcriptional regulation of the gene. As a first step, the transcription start site of the chb1 gene was determined by a PCR-based approach. Sequence analysis of clones amplified by 5' RACE indicated that the major transcription initiation site corresponds to the cytosine residue located 80-bp upstream of the transcription initiation codon (Fig. 4). This nucleotide was assigned the +1 position. A 1.3-kb DNA sequence from the 5' end to the putative transcription start site was searched for consensus binding sites for transcription factors (Fig. 4). The chb1 gene lacks a TATA box typical of eukaryotic type II promoters but contains a sequence motif (-3+7) that is similar to the ribosomal protein initiator sequence (CTTCCCTTTTCC) encompassing the transcription start site (33). The 5' 1.3-kb sequence contains several putative binding sites for transcriptional factors, including two consensus binding sites for E2A (positions -1214 to -1208 and -477 to -471). A NF-kB-binding motif is identified between the two E2A-binding sites (-926 to -917), and the region downstream of these elements contains a potential binding site for the early B cell factor (EBF; -158 to -145). Other potential binding sites include those for Aiolos, the CAMP-responsive element (-1123 to -1118), Sp1 (-1025 to -1017), NF-IL-6 (-965 to -951 and -23 to -15), c-Myc (-688 to -683), AP-1 (-564 to -556), Ets-1 (-391 to -384), GATA3 (-369 to -361) and AP-2 (-79 to -70).

Regulatory region for the chb1 promoter

To determine the region responsible for the tissue-specific expression of the chb1 gene, the 5' 1.3-kb DNA fragment was cloned upstream to a luciferase reporter gene and transfected into two B cell lines, DT40 and 249L4, and a T cell line, MSB1. Luciferase expression plasmids with or without the chicken β-actin promoter were also transfected into the cell lines to evaluate relative promoter activity. The entire 1.3-kb 5'-flanking region of chb1 promoter-reporter luciferase activity to a level ~20-fold above background when transfected into the DT40 cells that express the endogenous chb1 gene (Fig. 5). The same construct yielded no detectable luciferase activity when transfected into the MSB1 T cell line that does not express endogenous chb1. Thus, the promoter elements responsible for the tissue-specific expression of chb1 appear to reside in this 1.3-kb 5'-flanking region. Interestingly, luciferase activity could not be detected when the construct was transfected into the mature B cell line 249L4, which does not express endogenous chb1, indicating that the regulatory elements responsible for B cell stage expression are also located in this region.

To further analyze the regulatory region for chb1 promoter activity, we constructed a series of 5'-end deletion mutants of the chb1 gene (Fig. 5A). Each mutant was fused to a luciferase reporter gene and transfected into the same cell lines used in the experiments described above. Progressive deletion of 5' sequences up to position -661 had no significant effect on promoter activity. However, further deletion of upstream sequence up to position -355 diminished promoter activity, thereby suggesting the presence of positive regulatory elements in this region. Additional truncation of the 5' sequence up to position -161 restored promoter activity, and the construct retaining 161 bp of chb1 promoter region (del 10) exhibited the highest level of reporter gene activity, whereas sequence removal to position -30 abolished promoter activity. These findings suggest that a 194-bp region between positions -355 and -161 contains a negative element for promoter activity, whereas the 131-bp region between positions -161 to -30 possesses strong promoter activity. Notably, transfection of the del 10 construct into both 249L4 and MSB1 cell lines yielded 100-fold less activity in the DT40 cell line, indicating that the tissue-specificity region of the chb1 promoter lies within the first 161 bp of the 5' flanking region. Because one motif in this region (TGCCCAGGGGGC) matches closely with the consensus EBF binding site (TWCCCCNNGGAGT) (34), the results infer that an interaction with EBF at this site may be an important determinant of the tissue specificity of chb1 expression.

Discussion

These studies characterize the chb1 gene, the C-type lectin product of which is expressed on B cells in the bursa of Fabricius wherein the avian B cell repertoire is generated. C-type lectins can be divided into five groups on the basis of overall protein structure, the
The chB1 protein belongs to group 2 of the C-type lectin family. The members in this group share an α-helical coiled-coil stalk region separating the ligand-binding extracellular domain from the transmembrane domain (35). The present analysis indicates that, among the group 2 members, the chB1 gene is most similar to CD72 in gene organization and exon/intron structure, differing mainly in that chB1 lacks the third exon encoding the membrane proximal portion of the CD72 cytoplasmic region (26). Interestingly, alternative splicing may yield a CD72 mRNA lacking exon 3, the cDNA for which yields a cell surface protein product in transfected L cells (26).

Analysis of the 1.3-kb promoter region indicates that chB1 lacks a TATA box, a characteristic of many lymphoid-specific genes including CD72 (36) and NK cell receptor-P1 (37). A sequence motif located at the start site of transcription itself often serves as the initiator element. The sequence CTTCCTTTT (−3 to +7) has only one nucleotide difference from the mammalian ribosomal protein initiator element (33) and is similar to the putative CD72 initiator (36).

The restriction of chB1 expression to the intrabursal phase of B cell differentiation implies the existence of lineage-specific and differentiation stage-specific transcription factors. Potential binding sites for E2A, NF-κB, and EBF, transcription factors that play important roles during mammalian B cell development, are found in the chB1 promoter region. NF-κB acts as a B cell-specific factor involved in the Ig κ-chain gene expression (38). E2A proteins E47 and E12 are essential for B lineage development in that they are required for expression of several important B lineage-specific genes, including Recombination-activating gene-1, CD19, VpreB, λ5, mb-1, and Igμ (34). EBF is a homodimeric transcription factor that is expressed at all stages of mammalian B cell development, with the exception of the terminal plasma cell stage (39). B cell differentiation in EBF-deficient mice is arrested at the pro-B cell stage; early B lineage cells with rearranged Dμ and Jμ gene segments are completely absent in these mice (40). In addition, EBF
and E2A have been shown to be targets for the regulation of B cell differentiation (41).

The chB1 promoter is active in immature chB1- DT40 B cells, but not in the chB1- 249L4 mature B cell line or the MSB-1 T cell line, indicating the presence of lineage-specific and differentiation stage-specific elements in the chB1 promoter region. Our deletion analysis of the chB1 promoter indicates that the minimal promoter elements directing high levels of tissue-specific reporter gene expression are localized in the 131-bp sequence (−161 to −30) proximal to the transcriptional start site (del 10). This region contains a potential binding site for EBF that is critical for development of committed B cell progenitors. λ5, VpreB, mb-1, and B29 are target genes for the EBF transcription factor in pre-B cells (34, 42), and our data suggests chB1 as another candidate target for EBF, although further analysis will be required to confirm that this transcription factor is indeed involved in chB1 expression. In contrast, binding of PU.1 is essential for the increase in CD72 promoter activity in mouse B cells (36) and B cell lineage-specific activator protein plays a critical role in determining the cell-type specificity of the CD72 promoter (43). The transcriptional regulation of these two genes by different transcription factors could contribute to the difference in expression patterns for avian chB1 and mammalian CD72 during B cell development.

Other potential DNA binding sites identified in the 1.3-kb chB1 promoter region included ones for the AP-1, AP-2, and NF-κB transcription factors that have been shown to modulate transcriptional activity in response to phorbol esters. (44–46). This is consistent with our unpublished observation that chB1 mRNA expression is rapidly induced by PMA stimulation of splenic B cells. Two NF-IL-6 binding sites are also recognizable in the chB1 promoter region. NF-IL-6, a C/EBP-related protein with a leucine zipper, can be rapidly induced by stimulation with cytokines such as IL-1 and IL-6 (47). Because these cytokines participate in the activation and proliferative response of B cells (48), chB1 expression in the bursa could potentially be induced by these cytokines through NF-IL-6. Deletional analysis of the chB1 promoter region also reveals the presence of a negative element between nt −355 to −161, suggesting that repressors or silencers may also play a role in the regulation of chB1 expression.

Although chB1 and CD72 share considerable structural similarity, the chB1 gene differs in that it lacks exon 3 and encodes for only one cytoplasmic ITIM. The putative transcription factors required for gene expression, cell expression profiles, and function also differ significantly between chB1 and CD72. These differences preclude the conclusion that chB1 is the true avian ortholog of the mammalian CD72. In fact, we have recently identified a chB1-related gene encoding a highly homologous protein that contains two ITIMs in the cytoplasmic region (our unpublished observations). This raises the possibility of a CD72 family of genes, only one of which has been identified in mammals. The differential disruption of chB1 and its close relative in DT40 immature B cells (49) will provide valuable models for future functional studies of this gene family.

Acknowledgments

We thank Akiko Umeda for technical assistance and Drs. Zhixin Zhang and Peter Burrows for helpful discussion and review of the manuscript.

References