# A conserved sequence block in murine and human T cell receptor (TCR) Jα region is a composite element that enhances TCR α enhancer activity and binds multiple nuclear factors

Chia-Lam Kuo*†, Mei-Ling Chen*, Kai Wang‡, Chuan-Kai Chou*, Bernard Vernooij‡, Donald Seto§, Ben F. Koop¶, and Leroy Hood‡

*Institute of Biochemical Sciences, College of Science, National Taiwan University, Taipei, Taiwan, Republic of China; ‡Department of Molecular Biotechnology, University of Washington School of Medicine, Seattle, WA 98195; §Division of Endocrinology, National Institutes of Health, Bethesda, MD 20892; and ¶Department of Biology–Center for Environmental Health, University of Victoria, Victoria, BC Canada, V8W 2Y2

**ABSTRACT**    A conserved sequence block (CSB) located in a noncoding region of the mouse and human TCR α/δ loci, showing six differences over 125 nucleotide positions (95% similar), was subjected to detailed analyses in this study. Transient transfection results showed that the CSB-containing element in conjunction with the TCR α enhancer up-regulated the α enhancer activity, whereas no enhancer activity was detected when CSB alone was assayed. *In vitro* occupancy analyses of CSB by nuclear factors reveal the existence of an unexpectedly intricate network of CSB–protein and protein–protein interactions. Lymphoid-specific as well as T-lineage-specific nuclear factors are involved in differentially form CSB-bound complexes in extracts of various tissues and cell lines. Liver was shown to contain factor(s) sequestering thymic CSB-binding factors. Furthermore, the putative binding sites for transcription factors known to be important for lymphoid-lineage development are present in CSB and are targeted by nuclear factors. On the basis of these results, we propose that the CSB element may play a role in shaping the chromatin structure by which the accessibility of TCR α/δ loci to the recombinase complex and/or to the transcriptional apparatus can be controlled.

To recognize and respond to foreign antigens specifically is one of the important features of the vertebrate lymphoid system. This function is subserved in large part by two major cell lineages—B cells and T cells. B cells express cell-surface immunoglobulin (Ig) recognizing soluble antigens, whereas T cells express cell-surface T cell receptor (TCR) molecules, which recognize antigenic peptides bound to major histocompatibility complex molecules on the surface of antigen-presenting cells (reviewed in refs. 1 and 2). At least two classes of T cells can be distinguished on the basis of their differential expression of distinct TCRs. The major subset of circulating peripheral blood T cells expresses the heterodimeric αβ TCR (2), while a minor population of circulating T cells expresses the heterodimeric γδ TCR (3).

The Ig and TCR genes are structurally related; each is composed of multiple variable (V), diversity (D) (in the case of Ig heavy chain and TCR β- and δ-chain genes), and joining (J) gene segments that undergo rearrangements during lymphocyte ontogeny (3, 4). Despite the fact that a common recombinase is responsible for both Ig and TCR gene rearrangements (5), the rearrangement and expression of these genes is tightly regulated and lineage-specific, with TCR gene expression limited to T cells and Ig rearrangement limited to B cells (6). Similarly, the rearrangement and expression of the

TCR α- and δ-chain genes are lineage-specific, with TCR δ-chain gene expression limited to γδ T cells and α-chain gene expression limited to αβ T cells. Experiments have suggested that rearrangement of the Ig and TCR genes may be regulated by their lineage-specific transcriptional activators as well as by negative regulatory elements or transcriptional silencers (reviewed in ref. 7). Thus, it is important to understand the elements involved in regulating the lineage-specific expression of these genes to help understand the mechanisms that control the tissue-specific pattern of immune receptor gene rearrangement.

An important first step in this regard is the identification of the cis-regulatory elements. We previously analyzed the organization and structure of 95 kb of DNA spanning the TCR Cα/Cδ loci in mice and humans (8). Within this region, several previously described regulatory elements were localized, and sequences were compared (9–13). The sequences of TCR α enhancer regions in human and in mouse are 86% similar (10–13). This similarity value is higher than the overall sequence similarity found between human and mouse noncoding region in this locus (≈71%) and is consistent with a functional role (8). Most interestingly, a more conserved sequence was identified by this extended comparison. This conserved sequence block (CSB), just 5′ to the Cα gene, shows six differences over 125 nucleotide positions (95% similar) and is conserved throughout many vertebrates (10). No transcript was shown to hybridize to the CSB region in different mouse tissues (10). We reasoned that this conserved sequence may play a role relevant to the conserved coordinate regulation of TCR gene expression. Here, we describe the studies of CSB with respect to its regulatory functions in conjunction with the known TCR α enhancer and its *in vitro* occupancy by nuclear factors as the first step toward deciphering the functional significance of CSB. Our studies reveal two interesting features of CSB. First, when analyzed in conjunction with TCR α enhancer it enhanced α enhancer activity by 2-fold, whereas no enhancer activity could be detected when CSB alone was assayed. Second, an unexpectedly complicated network of CSB–nuclear protein interactions was observed. These results distinguish CSB from the known TCR enhancer and suggest that it may, instead, play a role in shaping the chromatin structure to facilitate TCR α- and/or δ-chain gene rearrangements. These results also demonstrate the power of comparative genomic sequence analyses in identifying conserved noncoding regions across species that may have important functional roles such as regulation.

Abbreviations: TCR, T cell receptor; CSB, conserved sequence block; CAT, chloramphenicol acetyltransferase; SV40, simian virus 40; EMSA, electrophoretic mobility-shift assay.
†To whom reprint requests should be addressed at: National Taiwan University, P.O. Box 23-106, Taipei, Taiwan, Republic of China. e-mail: chialam@ccms.ntu.edu.tw.

## MATERIALS AND METHODS

**Plasmids, Transfections, and Chloramphenicol Acetyltransferase (CAT) Assays.** Restriction fragments containing CAT gene, the CSB element, and TCR $\alpha$ enhancer were cloned in pBluescript (Stratagene) as follows. The basic CAT vector includes a 1.6-kb *Bgl*II–*Bam*HI fragment, containing the partial simian virus 40 (SV40) promoter, the CAT gene, and the SV40 splice sites and poly(A) site in the *Sma*I site of pBluescript KS(−). The CSB-containing element was obtained by *Eco*RI (5′) and *Sst*I (3′) restriction enzyme digestions of mouse genomic DNA and was subsequently inserted into the CAT vector in both orientations in the *Bam*HI site (pCT44B) or in the *Cla*I site (pCT45B and pCT49A). A 235-bp *Pvu*II–*Bgl*II fragment containing previously identified TCR $\alpha$ enhancer, marked C alpha enhancer in Fig. 1*A*, was used for these plasmid constructions. The EL4 $\alpha\beta$ T-cell line was transfected with CsCl-purified plasmid DNA by electroporation as previously described (14). Acetylation of [$^{14}$C]chloramphenicol (New England Nuclear) was assayed as described (11) and was quantified by a Molecular Dynamics PhosphorImager.

**Nuclear Extract Preparations.** Nuclear extracts from mouse tissues and cultured cell lines were prepared by using slight modifications of the procedure of Schreiber *et al.* (15). In brief, cells were lysed by incubation in buffer A (10 mM Hepes, pH 7.9/10 mM KCl/1.5 mM MgCl$_2$/0.1 mM EDTA/1 mM dithiothreitol). The cells were allowed to swell on ice for 15 min, after which Nonidet P-40 (Fluka) was added to a final concentration of 0.5%. Nuclei were then pelleted by centrifugation and resuspended in buffer B [20 mM Hepes, pH 7.9/0.4 mM NaCl/1.5 mM MgCl$_2$/1 mM EDTA/1 mM dithiothreitol/20% (vol/vol) glycerol]. After incubation on ice for 1 hr, lysates were centrifuged and supernatants containing the nuclear proteins (nuclear extracts) were stored at −70°C until use. Tissue samples were first minced manually before the lysis procedure. To minimize proteolysis, all manipulations after dissection were conducted at 4°C and protease inhibitors (0.5 mM phenylmethanesulfonyl fluoride, 1 $\mu$g/ml antipain, 1 $\mu$g/ml leupeptin, 1 $\mu$g/ml pepstatin A, 0.1 mg/ml trypsin inhibitor, and 10 mM iodoacetamide) were included in all buffers.

**DNA Binding Substrates and Electrophoretic Mobility-Shift Assay (EMSA).** The CSB substrate for EMSA was derived by PCR amplification with specific primers (MCB51: 5′-AGTGCTTTCAAGCGAGGCTG-3′; MCB31: 5′-TGCT-GCAATAAAAGAGAATT-3′). Resultant PCR products were analyzed on agarose gels, and the 125-bp fragment was purified from a 1.5% agarose gel and used in mobility shift assays. The fragment was also subcloned into the *Sma*I site of pUC19 (pUC-CSB) and sequenced to confirm the sequences. Purified CSB as well as annealed complementary oligonucleotides with sequences shown in Fig. 3*A* were end-labeled with [$\gamma$-$^{32}$P]ATP by T4 polynucleotide kinase. The labeled DNA substrate was then incubated with 5–10 $\mu$g of nuclear extract as indicated in figure legends in the presence of 1 $\mu$g of sonicated double-stranded salmon sperm DNA (15). Where indicated, the nonlabeled competitor DNA was added before the protein extracts. Incubations were for 30 min at 4°C. Complexes were resolved in nondenaturing 0.5× TBE/6% polyacrylamide gels (1× TBE is 90 mM Tris/90 mM boric acid/2 mM EDTA, pH 8.3). After electrophoresis at 150 V at room temperature, the gels were dried and exposed to x-ray films.

**DNase I Footprint Analysis.** CSB-containing fragment used for footprint analysis was derived by PCR amplification with universal forward and reverse primers and pUC-CSB as template. One of the primers was labeled with [$\gamma$-$^{32}$P]ATP by T4 polynucleotide kinase before being added to the PCR to produce CSB probe labeled at one end only. The probe was

then purified on nondenaturing polyacrylamide gels. The binding reaction was carried out by the same procedure as for EMSA except it was scaled up 2-fold in the footprint analysis. After incubation on ice for 30 min, an equal volume of a solution (5 mM CaCl$_2$/10 mM MgCl$_2$) was added to the reaction, followed by the addition of 1–5 $\mu$l of freshly diluted DNase I (Promega) to perform the digestion at room temperature for 30–90 sec. Reactions were stopped by the addition of 2 vol of stop solution (200 mM NaCl/30 mM EDTA/1% SDS/100 ng/ml yeast RNA), followed by extraction with phenol/chloroform and precipitation with ethanol. Samples were analyzed on 6% polyacrylamide/7 M urea sequencing gels.

## RESULTS

**Contribution of CSB to TCR $\alpha$ Enhancer Activity.** We searched for functional roles within the CSB by generating a set of test plasmids in which a 0.6-kb *Eco*RI–*Sac*I fragment DNA containing mouse CSB was inserted into a basic CAT-expression vector with and without the mouse TCR $\alpha$ enhancer (Fig. 1*A*). The construct containing CSB alone was essentially inactive when transfected into the mouse $\alpha\beta$ T cell line EL4 (Fig. 1*B*, lanes 4 and 5). However, when TCR $\alpha$ enhancer element was included in the construct and was placed in front of both the CAT gene and the CSB, the CSB reproducibly increased the activity of TCR $\alpha$ enhancer by 2-fold (Fig. 1*B*, lane 9). Because the TCR $\alpha$ enhancer is required for the TCR $\alpha$-chain gene to be expressed specifically in T cells (11), it is suggested from this analysis that CSB-containing fragment plays a regulatory role together with the $\alpha$ enhancer in regulating expression of genes in TCR $\alpha/\delta$ loci. It also suggests the CSB does not function as a classical enhancer element.

**Identification of Specific Nuclear Trans-Acting Factors That Interact with CSB.** We next looked for the presence of CSB-binding proteins in extracts of mouse thymus, spleen, brain, and liver. Nuclear protein extract was analyzed by EMSA using radiolabeled CSB fragments. Complexes were observed in lymphoid organs but were barely detectable in other tissues. Two CSB–protein complexes, T1 and T2, were formed with the thymus extract, whereas distinct complexes were detected with extracts from spleen. All of these complexes were diminished by the addition of specific CSB competitor, whereas nonspecific DNA failed to compete for the formation of these complexes (Fig. 2*A* and data not shown).

Cells of T lineage in thymus are heterogeneous in terms of their respective developmental stages, whereas the spleen is an organ with mixed populations of mature B and T lymphocytes. It is, therefore, informative to identify CSB–protein complexes in cloned cell lines. Nuclear extracts from immortalized cell lines were tested for this purpose by EMSA. Interestingly, as shown in Fig. 2*B*, although the thymus extract produced T1 and T2 complexes, none of the T cell-derived cell lines formed both. Moreover, T cell lines expressing different type of TCR ($\alpha\beta$- or $\gamma\delta$-expressing T cells) display different patterns of CSB-bound complexes. While the nuclear extract from B lymphoma gave one CSB-factor(s) complex, none was detected with nuclear extract from HeLa cells (Fig. 2*B* and data not shown). These results suggest that there are thymus-specific as well as lymphoid-specific CSB-binding factors.

**Fine Mapping of CSB.** We then analyzed this 125-bp fragment for the subsite(s) responsible for the observed differential nuclear protein-binding activity to further explore the properties of CSB. A series of oligonucleotides, shown in Fig. 3*A*, spanning the CSB element were synthesized. Annealed double-stranded oligonucleotides were labeled and tested with various nuclear extracts by EMSAs.

Differential and very weak C1-binding activity was observed in thymus and spleen nuclear extracts. No specific C1-forming signal was obtained with nuclear extract from brain or liver
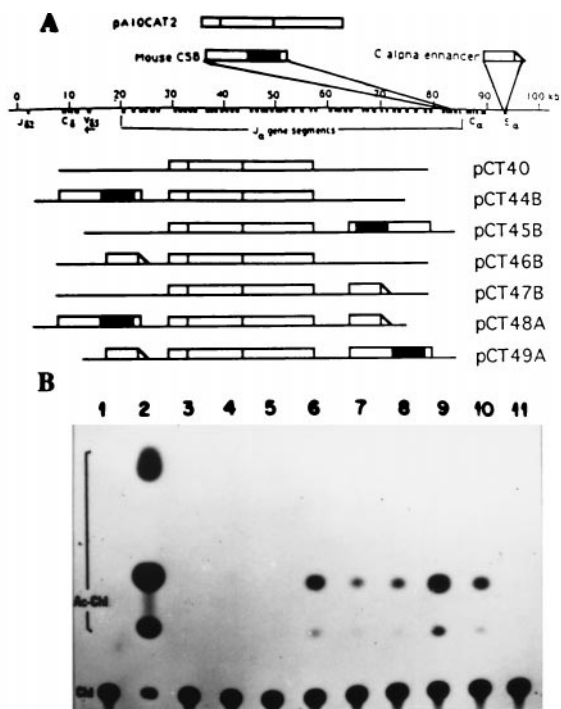
Fig. 1. Effects of CSB-containing element on the expression of the CAT gene. (*A*) Schematic diagram of constructs used for the assays. pA10CAT2 is a fragment containing the partial SV40 promoter, the CAT gene, and the poly(A) site. It was inserted into pBluescript KS(−), and the resultant plasmid (pCT40) was used as the basic CAT vector. Mouse CSB is a CSB-containing element obtained by restriction enzyme digestion of mouse genomic DNA. The solid bar indicates the actual location of CSB in this fragment DNA. C alpha enhancer corresponds to a *Pvu*II–*Bgl*II fragment containing the previously identified TCR α enhancer. The line below indicates the position of CSB and TCR α enhancer within the 95 kb of the mouse Cα/Cδ region. (*B*) Up-regulation of TCR α enhancer activity in αβ-expressing T cells. Constructs were electroporated into EL4 cells. The cells were harvested and samples were normalized with respect to total protein prior to performing the CAT assay. The acetylated chloramphenicol levels were measured qualitatively by autoradiography and quantitatively by a phosphoimager. The quantitative values are given in parenthesis below. Lane 1, assay blank (0); lane 2, *E. coli* CAT as a positive control (1,350); lane 3, baseline construct (pCT40) (16); lane 4, CSB in front of CAT gene (pCT44B) (12); lane 5, CSB in back of CAT gene (pCT45B) (0); lane 6, α enhancer in front (pCT46B) (176); lane 7, α enhancer in back (pCT47B) (74); lane 8, CSB in front, α enhancer in back (pCT48A) (90); lane 9, α enhancer in front, CSB in back (pCT49A) (333); lane 10, SV40 enhancer in front (138); and lane 11, mock transfection control (0). Chl indicates unacetylated [¹⁴C]chloramphenicol; Ac-Chl indicates acetylated [¹⁴C]chloramphenicol.

(Fig. 3*B*). As shown in Fig. 3*C*, thymus, spleen, and brain appeared to contain nuclear factors to differentially form respective C2-binding complexes. Liver nuclear extract gave a weak and diffuse C2-binding signal only after longer exposure (Fig. 3*C* and data not shown). For fragment C3, the spleen appeared to be the sole organ containing strong and specific C3-binding activity among the tested tissues (C3S-1, Fig. 3*D*). In contrast to C3, all four tissues appeared to contain C5-binding activity. A C5-forming complex with similar mobility, marked by the arrow in Fig. 3*F*, was observed in spleen, brain, and thymus. Again, liver nuclear extract produced a weak C5-binding signal observed after longer exposure.

For fragments C4 and C6, the major thymic C4- and C6-forming complexes—i.e., C4T-1 and C6T-1—appeared to contain similar, if not identical, factor(s) because C4 could efficiently decrease the amount of shifted radioactive C6-forming complex and *vice versa* (Fig. 3 *E* and *G* and data not shown). Moreover, octamer sequences (ATGCAAAT) effi-
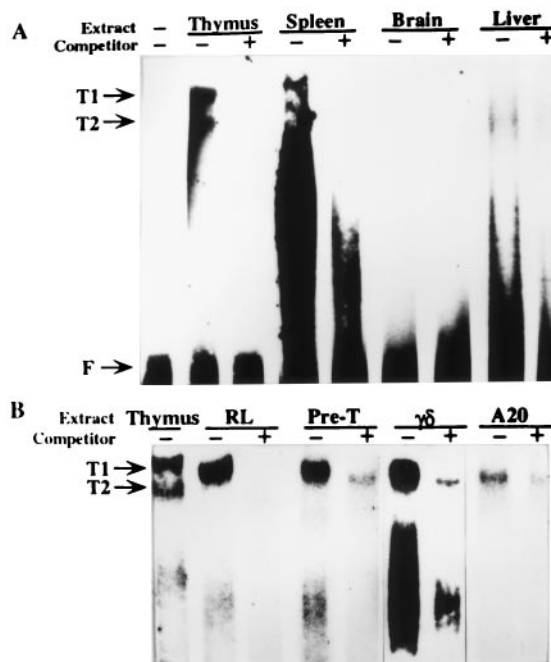


Fig. 2. Binding of nuclear factors to the 125-bp CSB element. (*A*) Tissue-specific binding to CSB. ³²P-labeled CSB was incubated with 5 μg of nuclear extracts prepared from adult mouse thymus, spleen, liver, or brain to perform EMSAs. No extract was added to the substrate in the first lane. Free probe (F) and CSB–factor complexes were resolved electrophoretically; the resulting autoradiogram is shown. The shifted complexes seen with the thymus extract are designated T1 and T2. Assays were done in the absence (−) or presence (+) of 50-fold excess of unlabeled CSB as specific competitor. (*B*) Cell type-specific binding to CSB. EMSA was performed as detailed for *A* except ³²P-labeled CSB was incubated with nuclear extract prepared from various cultured lines of T (RL, a leukemia T cell line; Pre-T, pre-T cell line SCI.ET.29F; γδ, a TCR γδ-expressing hybridoma GP39) and B (A20, a B cell lymphoma) cells.

ciently competed for the formation of C4- or C6-forming complexes (data not shown). Spleen nuclear extracts contained spleen-specific C4- and C6-binding factor(s), different from C4T-1 and C6T-1, respectively. Brain nuclear extracts appeared to contain C4- and C6-binding complexes, different from those formed in thymus or spleen. No C4- or C6-binding activity was detectable in liver nuclear extracts.

As also shown in Fig. 3 *E* and *F*, nuclear extracts of different cell lines produced different C4-forming complexes, whereas similar C5-forming complexes were observed among the assayed cell lines. The results suggest that cell type-specific as well as more commonly expressed factors are involved in interacting with subregions of CSB.

Analysis of the CSB sequence revealed several putative inverted repetitive sequences (Fig. 3*A*, INV1 to INV4). Because INV2 was unintentionally separated into the C3 and C4 fragments, a new fragment, C34, containing both INV2 and INV3 was synthesized and analyzed for its interaction with nuclear factor(s). In addition, because similar, if not identical, trans-acting factors were shown to recognize C4 and C6, a fragment C456 was synthesized and used in EMSA to determine how the flanking region, C5, may affect the factor(s) in recognizing C4 and C6.

C34-binding complexes, different from those formed with C3 or C4, were observed in lymphoid tissues and cell lines. Barely detectable C34-forming complexes were seen when 3-fold excess of brain or liver extract was assayed. Surprisingly, neither C3 nor C4 could compete for the formation of C34-binding complexes, whereas octamer could diminish the
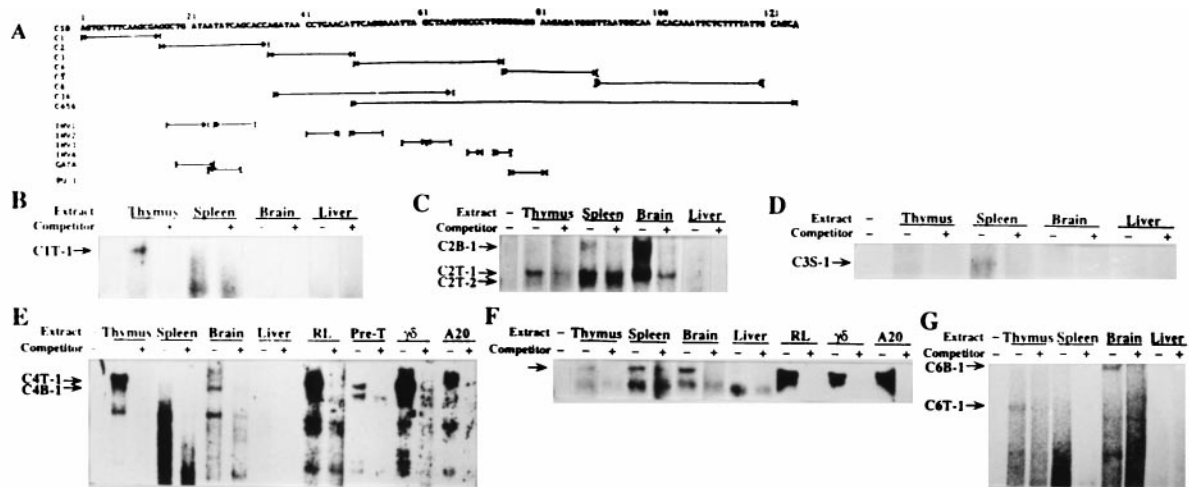
FIG. 3. Fine mapping of nuclear factor recognition sites within CSB. (*A*) The sequence of the CSB element (125 bp in length) is shown at the top. Below that are arrowhead-lines representing different subsites of CSB, including subregions assayed for bindings of nuclear factors by EMSA. They are C1, C2, C3, C4, C5, C6, C34, and C456. The inverted repeats (denoted as INV1 to INV4), the putative GATA-binding site (denoted as GATA), and PU.1-binding site (denoted as PU.1) are also presented (see text). (*B–G*) Nuclear extracts prepared from various adult mouse tissue and/or a series of cultured cell lines were incubated with $^{32}$P-labeled C1 (*B*), C2 (*C*), C3 (*D*), C4 (*E*), C5 (*F*), or C6 (*G*) to perform EMSAs. Assays were done in the absence (−) or presence (+) of 50-fold excess of unlabeled C1 (*B*), C2 (*C*), C3 (*D*), C4 (*E*), C5 (*F*), or C6 (*G*) as competitor. Free probe and CSB subregion–factor complexes were resolved by electrophoresis. Only the relevant resulting autoradiograms are shown. C1T-1, the major C1-binding complex in thymus; C2B-1, the slowest migrating C2-binding complex formed in brain; C2T-1 and C2T-2, the two major C2-binding complexes formed in thymus; C3S-1, the C3-binding complex formed in spleen; C4T-1 and C4B-1, the unique C4-binding complexes formed in thymus and brain, respectively; and C6B-1 and C6T-1, the major unique C6-binding complexes formed in brain and thymus, respectively.

binding signals to different extents depending on the types of cell lines or tissues assayed (data not shown).

The C456-forming complexes, different from those formed with C4, C5, or C6, were obtained in nuclear extracts of lymphoid tissues or cell lines. C4, C5, C6, C34, or Oct could differentially diminish part of C456-binding signals to different extents, depending on the tissues or cell types assayed (data not shown). These results suggest that the observed CSB-binding complexes are formed through intricate interactions between nuclear factors and this region of the CSB element.

**Liver Contained Factors Sequestering CSB-Bound Complexes in Thymus.** Liver nuclear extracts may contain factor(s) recognizing certain subregion(s) of CSB. The factor(s), however, may interact with other factor(s) to result in protein complexes with much lower affinity to CSB as observed by EMSA. Therefore, liver might contain specific factor(s) functioning to prevent the formation of thymic-specific CSB-binding complexes. To verify this possibility, experiments in which two different extracts were mixed and used for assays were performed.

Interestingly, the thymic CSB–protein complexes were sequestered when liver and thymus nuclear extracts were added together for reactions. Moreover, the nuclear extracts of B lymphoma A20 had a similar effect in sequestering the thymic CSB-complex-forming activity. Other nuclear extracts of T lineage did not affect the thymus T1-forming activity, whereas T2-forming activity was decreased (data not shown).

**CSB-Binding Specificity as Determined by DNase I Footprinting.** We further analyzed the sites within CSB that interact with nuclear proteins by DNase I footprinting. Very interestingly, two types of protection patterns could be observed reproducibly. As shown in Fig. 4, the patterns of DNase hyper-sensitive sites formed in spleen, liver, or A20 are similar to each other, but different from the type produced by thymus, brain, or cell lines of T lineage. Why did spleen and A20 produce a liver-like footprinting pattern despite the EMSA data indicating the CSB-binding complexes formed in the respective nuclear extract are different? These results could well be explained by proposing that both tissue-specific and commonly expressed factors are involved in interacting with

each other and with the CSB element. The resultant CSB-bound complexes detected by EMSAs are different among tissues or cell lines because of the participation of cell type-specific factors. However, the contacted nucleotide residues detected by footprinting assay may be similar, or identical, if contact is made by members belonging to the same family, or by an identical factor present in the assayed extracts. This explanation could explain the similar DNase protection patterns.

Mixed nuclear extracts were also used in the footprinting assays. Consistent with EMSA findings, liver extracts were shown to contain factors able to sequester thymus CSB-binding factors because liver-like CSB footprinting pattern was produced in mixed extracts of liver and thymus. Similarly, spleen or A20 extract could convert the thymus-like into liver-like protection pattern when nuclear extract of spleen or A20 was mixed with that of thymus in the assays (Fig. 4).

## DISCUSSION

By analyzing the sequences of the murine and human T-cell receptor Cα/Cδ regions, we previously identified a highly conserved sequence block (CSB) located in the region between the Jα3 and Jα4 gene segments (8–10). This CSB shows six differences between murine and human sequences over 125 nucleotide positions, which is 95% similar. In this study, we performed intensive experiments to explore the functional significance of the evolutionarily conserved sequence block. The data of the CAT reporter gene-based assay showed this region in tandem with the TCR α enhancer increases CAT expression by 2-fold, but it has no effect on CAT expression by itself. This result suggests the existence of nuclear factor(s) capable of interacting with the CSB-containing element, leading to the observed stimulation of TCR α enhancer activity.

We then investigated the interactions between the CSB element and trans-acting factors in nuclear extracts from mouse tissues and cell lines. EMSA results suggest that there are tissue-specific factors binding to the CSB element. Dissections of CSB performed to localize subregion(s) responsible for the observed differential CSB-binding activity reveal the
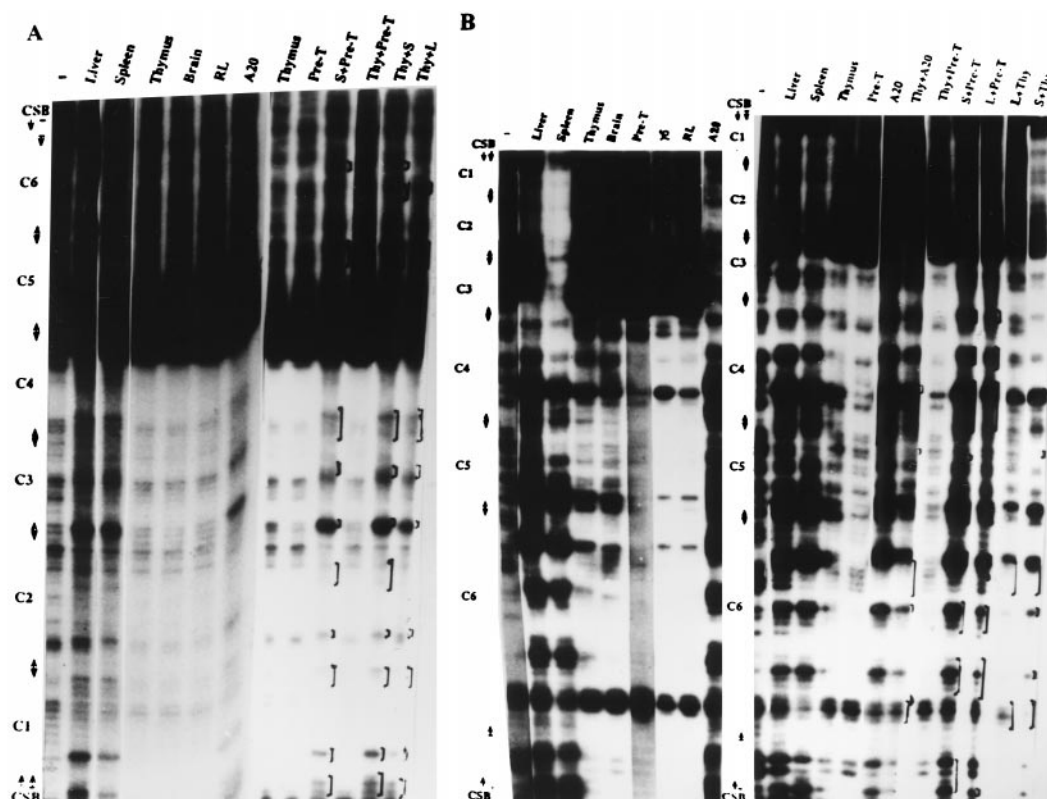
FIG. 4. Comparison of footprinting patterns established on CSB by nuclear extracts of various mouse tissues and cultured cell lines. A CSB-containing fragment was analyzed by DNase I footprinting, as described in the text. The autoradiograms of DNase I digestion of the top strand and the bottom strand are shown in (*A*) and (*B*), respectively. Nuclear extracts from various tissues and cell lines, and mixed nuclear extracts of spleen and pre-T (S+Pre-T), thymus and pre-T (Thy+Pre-T), thymus and spleen (Thy+S), thymus and liver (Thy+L), thymus and A20 (Thy+A20), or liver and pre-T (L+Pre-T) were used in the assays. −, No nuclear extract. The regions corresponding to subregions of CSB, shown in Fig. 3*A*, are indicated on the side of each gel. The bracketed regions indicate the regions with footprint pattern changes in mixed nuclear extracts when compared with the pattern produced by thymus, pre-T, or Thy+Pre-T.

existence of an unexpectedly intricate network of CSB–protein and protein–protein interactions. On the basis of these results, we discuss below known transcription factors and the proposed functional role served by CSB.

**Are Members of the POU Family Involved in Forming CSB-Binding Complexes?** Octamer-containing sequences, known to play a critical role in regulating transcription of Ig genes in B cells (16), compete unexpectedly with the C4 and C6 fragments for nuclear binding factors (data not shown). To date, several octamer-binding factors and their cDNAs have been isolated. All of these octamer-binding proteins belong to the POU family of DNA-binding transcription factors (reviewed in ref. 17). Two of these have been well studied: Oct-1, a ubiquitously expressed transcription factor (reviewed in ref. 18), and Oct-2, a factor predominantly found in B cells (16). It was shown Oct-1 and Oct-2 could recognize, although with lower affinity, important cis-acting elements with little octamer motif similarity (19, 20). Previous studies also showed that both the POU-specific domain, contacting the 5′ half of the octamer (<u>ATGC</u>AAAT), and the POU homeodomain, contacting the 3′ half of the site (ATGC<u>AAAT</u>), could fold as independently stable structures (21, 22). Furthermore, a B cell-specific coactivator Bob1 was shown to interact specifically with Oct-1/Oct-2 and was the main determinant for B cell-specific activation of Ig promoters (23, 24). The AAAT segment is present in both C4 and C6, and our EMSA data also indicated that more than one factor is involved in forming C456-binding complexes. Taking these results together, it is tempting to speculate that some protein(s) capable of recognizing octamer sequence, most likely a POU homeodomain-containing factor(s), is involved in interacting with CSB and

with other nuclear factor(s) to differentially form CSB-binding complexes.

**CSB Contains Motifs Resembling Known Regulatory Elements: GATA and PU.1-Binding Sites.** A close inspection of the CSB nucleotide sequences reveals two interesting motifs. One is the consensus GATA binding site (WGATAR; W = T or A, R = G or A) (reviewed in ref. 7) present in mouse subregion C2 as inverted repeats. The other is the PU.1-binding site (GAGGAA) (25) in the subregion C5 as shown in Fig. 3*A*.

Functionally important GATA-binding sites have been identified in the TCR $\alpha$, $\beta$, and $\delta$ enhancers (26–28). These findings led to the cloning of a novel GATA family member, the zinc-finger transcription factor GATA-3 (26–28). GATA-3 was shown to be expressed in hematopoietic cells and in the developing kidney and nervous system (reviewed in ref. 7). Recently it was demonstrated that GATA-3 is required for development of the T-cell lineage through the studies of GATA-3$^{-/-}$-RAG-2$^{-/-}$ chimeric mice (29). In our studies, the nuclear extract from thymus, spleen, brain (Fig. 3*C*), or a T-cell line (data not shown) was shown to contain differential activity in forming C2-binding complexes. This result suggests that CSB may be a novel target for GATA family member(s).

PU.1, a member of the ets family of transcription factors, is expressed exclusively in cells of the hematopoietic lineage (30, 31). It was shown that development of B cells and macrophages in PU.1-null fetuses or PU.1-null newborn mice was undetectable (32, 33). In addition, the PU.1-binding site in the Ig $\kappa$ gene 3′ enhancer (E3′) was shown to be responsible for cell-type-specific V$\kappa$–J$\kappa$ joining, since DNA isolated from thymocytes of transgenic mice carrying a mutated PU.1-binding site had

undergone Vκ–Jκ joining (34). Our studies indicate that the perfect PU.1 binding-site, in C5 of mouse CSB, was recognized by a protein complex different from that recognizing the E3′ PU.1-binding site in the Ig κ gene. The activity in forming E3′ PU.1 site-binding complex was observed to be much stronger and more complex than that of forming C5-binding complex in spleen or in A20 nuclear extract (data not shown). The results are consistent with the importance of the entire sequence of CSB functioning to mediate interactions with multiple transcription factors.

**Proposed Functional Roles of CSB.** What are the components of the protein complexes interacting with the evolutionarily conserved sequence block, CSB, located in mouse and human TCR α/δ loci? Data presented here and our ongoing studies (M.-L.C. and C.-L.K., unpublished results) strongly suggest that both tissue-specific factor(s) whose expression is developmentally regulated and factors commonly expressed in many tissues are involved to form this CSB-binding complex. What, then is the function of CSB that apparently mandates the evolutionary conservation of the CSB nucleotide sequence?

We believe that there is an intricate, hierarchical network of protein–protein and protein–DNA interactions initiated by the specific binding of a factor(s) to a subregion of the CSB element. Following this process, some tissue-specific and/or developmentally regulated factor(s) is recruited through protein–protein interaction and/or protein–DNA recognition into this CSB–protein complex. This type of recruitment leads to the formation of cell type-specific CSB–protein complex. This DNA–protein complex may then participate in the formation of a highly ordered, "enhanceosome"-like structure (35), which may function as one of the elements in shaping TCR α/δ loci accessibility for the tightly regulated TCR gene rearrangement. The presence of binding sites in CSB putative for transcription factors important for lymphocyte development favors the proposed functional role of CSB.

Recently, it was demonstrated that chromatin structure determines *in vitro* the targeting of V(D)J recombinase activity (36). Interestingly, this work showed that a HeLa cell nuclear extract could not complement RAG-1 + RAG-2 activity to allow recombination signal sequence cleavage of permissive nuclear template (36). Our results suggest that there may be liver-specific factor(s) able to interact with and thus sequester factors that, otherwise, recognize and bind to CSB. It will be interesting to determine whether these two events are related. To define the factors involved in CSB-forming complexes will define the function of CSB and, we believe, will shed light on the coordinately regulated process of TCR gene expression. It may also help to understand the mechanism involved in forming cell type-specific chromatin structure suggested to play a role (36) in determining the targeting of V(D)J recombinase activity.

**Role of Comparative Genomic Sequence Analyses in Identifying Functional Elements.** This study demonstrates the power of comparative genomic sequence analyses in identifying conserved noncoding regions that may encode important biological information (e.g., regulatory regions or regions carrying out chromosome-specific functions). As the Human Genome Project progresses, comparative genomic sequencing will be a powerful tool for identifying putative regulatory regions.

1. Tonegawa, S. (1983) *Nature (London)* **302,** 575–581.
2. Kronenberg, M., Siu, G., Hood, L. & Shastri, N. (1986) *Annu. Rev. Immunol.* **4,** 529–591.
3. Raulet, D. H. (1989) *Annu. Rev. Immunol.* **7,** 175–207.
4. Honjo, T. (1983) *Annu. Rev. Immunol.* **1,** 499–528.
5. Yancopoulos, G., Blackwell, H., Suh, H., Hood, L. & Alt, F. W. (1986) *Cell* **44,** 251–259.
6. Calame, K. & Eaton, S. (1988) *Adv. Immunol.* **43,** 235–275.
7. Leiden, J. N. (1993) *Annu. Rev. Immunol.* **11,** 539–570.
8. Koop, B. F. & Hood, L. (1994) *Nat. Genet.* **7,** 48–53.
9. Koop, B. F., Wilson, R. K., Wang, K., Vernooij, B., Zaller, D., Kuo, C.-L., Seto, D., Toda, M. & Hood, L. (1992) *Genomics* **13,** 1209–1230.
10. Koop, B. F., Rowen, L., Wang, K., Kuo, C.-L., Seto, D., Lenstra, J. A., Howard, S., Shan, W., Deshpande, P. & Hood, L. (1994) *Genomics* **19,** 478–493.
11. Winoto, A. & Baltimore, D. (1989) *EMBO J.* **8,** 729–733.
12. Ho, I.-C. & Leiden, J. M. (1990) *Mol. Cell. Biol.* **9,** 4720–4727.
13. Ho, I.-C. & Leiden, J. M. (1990) *J. Exp. Med.* **172,** 1443–1449.
14. Kuo, C.-L. & Hood, L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7614–7618.
15. Schreiber, E., Matthias, P., Muller, M. M. & Schaffner, W. (1989) *Nucleic Acids Res.* **17,** 6419.
16. Staudt, L. M., Clerc, R. G., Singh, H., LeBowitz, J. H., Sharp, P. A. & Baltimore, D. (1988) *Science* **241,** 577–580.
17. Herr, W. & Cleary, M. A. (1995) *Genes Dev.* **9,** 1679–1693.
18. Herr, W. (1992) in *Transcriptional Regulation,* eds. McKnight, S. & Yamamoto, K. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 1103–1135.
19. Baumruker, T., Sturm, R. & Herr, W. (1988) *Genes Dev.* **2,** 1400–1413.
20. Bendall, A. J., Sturm, R. A., Danoy, P. A. C. & Molly, P. L. (1993) *Eur. J. Biochem.* **217,** 799–811.
21. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. & Pabo, C. O. (1994) *Cell* **77,** 21–32.
22. Botfield, M. C., Jancso, A. & Weiss, M. A. (1992) *Biochemistry* **31,** 5841–5848.
23. Gstaiger, M., Georgiev, O., van Leeuwen, H., van der Vliet, P. & Schaffner, W. (1996) *EMBO J.* **15,** 2781–2790.
24. Luo, Y. & Roeder, R. G. (1995) *Mol. Cell. Biol.* **15,** 4115–4124.
25. Karim, M. J., (1990) *Genes Dev.* **4,** 1451–1453.
26. Ho, I.-C., Voorhees, P., Marin, N., Oakley, B. K., Tsai, S.-F., Orkin, S. H. & Leiden, J. M. (1991) *EMBO J.* **10,** 1187–1192.
27. Marine, J. & Winoto, A. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7284–7288.
28. Joulin, V., Bories, D., Eleouet, J.-F., Labastie, M.-C., Chretien, S., Mattei, M.-G. & Romeo, P.-H (1991) *EMBO J.* **10,** 1809–1816.
29. Ting, C.-N., Olson, M. C., Barton, K. P. & Leiden, J. M. (1996) *Nature (London)* **384,** 474–478.
30. Moreau-Gachelin, F. (1994) *Biochim. Biophys. Acta* **1198,** 149–163.
31. Klemsz, M., McKercher, S. R., Celada, A., Van Beveren, C. & Maki, R. A. (1990) *Cell* **61,** 113–124.
32. Scott, E. W., Simon, M. C., Anastasi, J. & Sigh, H. (1994) *Science* **265,** 1573–1577.
33. McKercher, S. R., Torbett, B. E., Anderson, K. L., Henkel, G. W., Vestal, D. J., Baribault, H., Klemsz, M., Feeney, A. J., Wu, G. E., Paige, C. J. & Maki, R. A. (1996) *EMBO J.* **15,** 5647–5658.
34. Hiramatsu, R., Akagi, K., Matsuoka, M., Sakumi, K., Nakamura, H., Kingsburg, L., David, C., Hardy, R. R., Yamamura, K.-I. & Sakano, H. (1995) *Cell* **83,** 1113–1123.
35. Thanos, D. & Maniatis, T. (1995) *Cell* **83,** 1091–1100.
36. Stanhope-Baker, P., Hudson, K. M, Shaffer, A. L., Constantinescu, A. & Schlissel, M. S. (1996) *Cell* **85,** 887–897.