Nucleic Acids Research, 1993, Vol. 21, No. 22 5067–5073

Molecular cloning of a zinc finger protein which binds to the heptamer of the signal sequence for V(D)J recombination

Lai-Chu Wu[1,2,3], Chi-Ho Mak[3], Neil Dear[1], Thomas Boehm[1,+], Letizia Foroni[1,§] and Terence H.Rabbitts[1,*]

[1]Medical Research Council, Laboratory of Molecular Biology, Cambridge CB2 2QH, UK, [2]Departments of Medical Biochemistry and Internal Medicine and [3]The Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210, USA

Received August 9, 1993; Revised and Accepted October 6, 1993 GenBank accession no. L07911

ABSTRACT

The somatic V(D)J recombination for the assembly of the Ig and TCR genes is mediated by the recombination signal sequences (Rss) and the V(D)J recombinase. A cDNA clone was isolated from a λgt11 expression library made from mouse thymocyte poly(A)$^+$ RNA, using the Rss as a ligand. The deduced amino acid sequence of the putative protein, designated *R*ecognition *c*omponent (Rc), reveals a pair of Cys_2-His_2 zinc fingers followed by a Glu- and Asp-rich acidic domain. In addition, there are five copies of the Ser/Thr-Pro-X-Arg/Lys sequence, which are putative DNA binding units. The zinc finger-acidic domain structures present in Rc are also found in several enhancer binding proteins, such as those for the κB motif of the Ig κ light chain enhancer or related sequences. Bacterial fusion proteins for Rc bind preferentially to the Rss heptamer and to the κB motif. The dual affinities of Rc for the Rss heptamer and the κB motif suggest a possible link between Ig transcription and somatic recombination. The formation of multiple 'gel-shifted' DNA–protein complexes for Rc and its DNA ligand suggests that these complexes tend to multimerize.

INTRODUCTION

Activation and diversification of antigen receptors in vertebrates are generated by the somatic recombination of the V, (D) and J gene segments of the Ig and the TCR genes (1,2). Somatic recombination events occur in early lineages of B and T lymphocytes (3,4), mainly in the spleen and the thymus (5). This specific recombination process is performed by the V(D)J recombinase. The sequence specificities of these cleavage and joining processes are dictated by the recombination signal sequences (Rss) flanking the V, D and J gene segments (6). An Rss consists of a conserved heptamer sequence and a conserved nonamer sequence separated by a non-conserved spacer sequence of 12 bp (i.e. 12 bp-Rss) or 23 bp (i.e. 23 bp-Rss).

The Rss has been employed as a substrate or a ligand to determine the components of the V(D)J recombinase. RAG-1 and RAG-2 were cloned based on their capability to activate rearrangement of DNA substrates with the Rss (7,8). Jκ RBP (9) and T160 (10) were obtained based on their abilities to bind the 23 bp-Rss and the 12 bp-Rss, respectively. At least three other proteins have been shown biochemically to bind the Rss (11–13) but their genes are yet to be cloned. The composition of the V(D)J recombinase is still largely undefined.

In the present study, we have isolated cDNA clones encoding lymphoid-specific Rss binding proteins. We used thymocyte poly(A)$^+$ RNA isolated from 3-week-old mice to construct a λgt11 cDNA library. By screening this expression library with a DNA probe harboring alternating Rss heptamers and nonamers, we were able to isolate a cDNA clone whose putative protein is named *R*ecognition *c*omponent (Rc) in light of its binding affinity to the Rss heptamer. Sequence analyses show that Rc is related to several enhancer binding proteins for the κB motif or related sequences (14–17). The κB motif was first defined as the B site, for its restricted B cell occurrence, in the Ig κ light chain gene enhancer within the $J_κ$-$C_κ$ intron (18). Subsequently, this motif was found in several cellular or viral gene enhancers (19,20). We have shown that Rc (produced in bacteria) not only binds the Rss heptamer, but also the κB motif.

MATERIALS AND METHODS

Isolation of cDNA clones for Rss binding proteins

A λgt11 cDNA expression library was constructed using thymocyte poly(A)$^{+}$ RNA isolated from 3-week-old BALB/c mice, random hexanucleotide primers for the synthesis of the first cDNA strands, and a cDNA Synthesis System Plus cloning

* To whom correspondence should be addressed

Present addresses: +Klinikum der Albert-Ludwigs-Universitat Freiburg, Abteilung Innere Medizin I, Hugstetterstrasse 55, D-7800 Freiburg, Germany and §Department of Hematology, Royal Free Hospital, London, NW3 2QG, UK

kit (Amersham), per manufacturer's instructions. Fusion proteins were generated and protein filters processed and screened with a Rss-containing probe, *R* (described below), according to the method of Vinson *et al.* (21). Hybridization probes were generated from the clone λT1 to screen DNA filters prepared from a thymocyte cDNA library (this work) and a brain cDNA library purchased from Clontech. The cDNA inserts of positive clones were subcloned into pBluescript (Stratagene) for restriction mapping and for sequencing.

Plasmids, DNA fragments, and probes

Plasmids

pBS79-1 a 145 bp Rss oligonucleotide with alternating heptamer and nonamer stretches cloned into the *Pst*I and *Eco*RI sites of pBluescript.

pUR-R 1.9 kb *Nde*I-*Hind*III fragment of λT1 subcloned into *Nde*I-*Hind*III sites of pUR (22).

pκB κB motif (18): 5′-CCGGGGGGGACTTTCCG-CTCCAC-3′ cloned into the *Eco*RV site of pBluescript.

pHN8 Rss heptamer (6) : 5′-GGATCCCACACTGCTCC-AGGATCC-3′ cloned into the *Sma*I site of pBluescript.

DNA fragments

R Rss probe; *Hind*III-*Bam*HI fragment from *pBS79-1.*

p*C* 1.2 kb *Bam*HI-*Eco*RI fragment from λB-a (Fig. 1D).

p*D* 1.6 kb *Eco*RI fragment of λT1 subcloned into M13mp8 (Fig. 1D).

H Rss heptamer; *Hind*III-*Xba*I fragment from *pHN8.*

N Rss nonamer; similar to *H* but core sequence is 5′-GGTTTTTGT-3′.

H′ heptamer variant; similar to *H* but core sequence is 5′-CAGAGTG-3′.

κB *Hind*III-*Xba*I fragment from *pκB.*

polylinker A 42 bp *Hind*III-*Xba*I fragment from pBluescript

The sequence of the chemically synthesized 145 bp Rss oligonucleotide, with the heptamers and nonamers underlined, is as follows: 5′CTGCAGCACAGTGATTAAGACCTAACCT-ACACCACAACAAAAACCCCCACAGTGCGGGATAGGTGC-ACAAAAACCCACAGTGTTAAGACCTAACCTCACCACA-ACAAAAACCCACAGTGCGGGATAGGTGCACAAAAACC-GAATTC-3′.

Table 1. Comparative DNA binding of β-gal/Rc to the Rss heptamer or nonamer

DNA fragment	Experiment I (cpm)	Experiment I c.f. Rss heptamer binding	Experiment II (cpm)	Experiment II c.f. Rss heptamer binding
H, Rss *h*eptamer -CACAGTG-	40,033	1	104,170	1
N, Rss *n*onamer -GGTTTTTGT-	2,883	0.072	15,229	0.146
H′ heptamer mutant -CAGAGTG-	4,655	0.12	12,776	0.123
polylinker	5,326	0.133	29,580	0.28

Proteins from 5 ml of IPTG induced bacteria containing β-gal/Rc protein were resolved by preparative SDS-PAGE and electroblotted to nitrocellulose membranes. The membrane was cut into 1 cm wide vertical strips and hybridized with short DNA fragments: *H, N, H′*, or polylinker DNA. Following incubation with a probe, the membrane strips were washed with 25 mM NaCl (Expt. II) or 125 mM NaCl (Expt. 1) in buffer D for low stringency or high stringency washing, respectively, prior to autoradiography. For determination of the actual radioactivity, the radioactive regions were excised and counted by Cerenkov radiation.

Probes

DNA fragments R, H, N, H′, κB, and polylinker were labeled by 3′ end filling with [α-^{32}P]-dCTP and Klenow, and gel purified. Hybridization probes were prepared using a random-primed DNA labeling kit (US Biochemical Corp.) and [α-^{32}P]-dCTP.

DNA sequence analysis

Nucleotide sequence was determined using Sequenase 2.0 DNA sequencing kits (US Biochemical Corp.), and ^{35}S-dATP. All sequences were determined from both strands. Gel readings were assembled with Staden's DNA analysis software (24). Sequence analysis was accomplished with software from the Wisconsin Genetic Computer Group (25). A non-redundant search of protein databases was carried out using the BLAST Network Service at the National Center for Biotechnology Information (NCBI) (26).

Electrophoretic mobility shift assay (EMSA) and methylation interference analysis with Mbp/Rc$_{490}$

EMSA employing the Mbp/Rc$_{490}$ fusion protein produced in bacteria was performed as described (27). Probe *H* or probe *κB* (0.2 ng, ~5,000 cpm) were used as DNA ligands. For competition EMSA, binding reactions using 0.2 μg of Mbp/Rc$_{490}$ were supplemented with a 100-fold excess (20 ng) of unlabeled substrate or *Msp*I digested pBR322 DNA. For methylation interference analysis, end-labeled DNA fragments (~1 ng, 25,000 cpm) were partially methylated with dimethylsulfate (28), included in preparative binding reactions with 0.5 μg Mbp/Rc$_{490}$, and electrophoresed through a 5% polyacrylamide gel. Bound and free DNA were recovered and methylated residues cleaved by piperidine (28). Cleaved DNA was electrophoresed on 8% sequencing gels.

Southwestern blot analysis with β-gal/Rc

Southwestern blot analysis was performed according to Singh *et al.* (29). Proteins from IPTG induced bacterial cultures (5 ml) harboring *pUR-Rc* were resolved by SDS-PAGE and electrophoretically transferred to a nitrocellulose membrane (Hybond-C extra, Amersham). The membrane was cut into 1 cm wide vertical strips and each membrane strip was hybridized with *H, N, H′*, or polylinker probe (~10^5 cpm/ml). After autoradiography, the radioactive regions on the membranes were excised, and the radioactivity determined by scintillation counting.

Northern blot analysis and *in situ* hybridization

Total RNA (~20 μg), isolated by the guanidine thiocyanate/cesium chloride method (30), was electrophoresed through 0.8% formaldehyde-agarose gels, transferred to nylon membranes (Hybond N, Amersham), and cross-linked to membranes by UV irradiation. Membranes were hybridized with probe p*C* at 42°C overnight at $1-2\times10^6$ cpm/ml in standard hybridization solution containing 50% formamide and denatured sonicated salmon sperm DNA (100 μg/ml). Following hybridization, membranes were washed with 0.2×SSC (30 mM NaCl, 3 mM sodium citrate, pH 7.0), 0.5% SDS at 65°C for 15 minutes twice, and autoradiographed at −80°C. Membranes were reprobed as above with ^{32}P-labeled glyceraldehyde 3-phosphate dehydrogenase (G3PDH) probes (Clontech). *In situ*

hybridization was performed with day 19 mouse embryo sections, using a sense or antisense p*D* probe as previously described (23).

RESULTS

Isolation of cDNA clones for Rss binding proteins

A λgt11 cDNA expression library was constructed using thymocyte poly(A)$^+$ RNA. The library had $\sim 8\times10^5$ recombinants and an average insert size of ~1.3 kb. Protein filters prepared from this library were hybridized to probe *R*, which harbors alternating Rss heptamer and nonamer stretches separated by spacers of various lengths. Three distinct positive clones binding to probe *R* were obtained. One of them, λT1, was chosen for further investigation because its expressed protein product bound to probe *H* (contains the Rss heptamer) ~4 to 7 fold more efficiently than to the polylinker sequence in Southwestern blot analyses (Table 1).

The DNA insert of λT1 was ~1.77 kb in size (Fig. 1C). Probes were generated from λT1 to isolate additional Rc cDNA clones from mouse thymocyte and brain cDNA libraries. Clones λT1-a, -b, -c, and -d, and λB-a and -b were obtained. Altogether they span ~3.2 kb of overlapping cDNA sequences (Fig. 1).

Rc contains structural domains for DNA binding and protein−protein interactions

A cDNA sequence of 2,304 bp containing a reading frame of 767 amino acid residues is presented in Figure 2. The deduced amino acid sequence of Rc contains several structural features characteristic of a DNA binding protein and/or a transcription factor (Fig. 1B). Amino acid residues 205 to 257 consist of two consecutive zinc finger motifs. The conserved Cys, His, Phe, and Leu residues in these putative zinc fingers are circled in Fig. 2. Zinc fingers are common nucleic acid binding protein structures and have been found in proteins with diverse regulatory functions (31). The presence of these motifs in Rc is suggestive of a specific DNA binding function. Immediately following the zinc fingers, at residues 268 to 326, is a region rich in acidic amino acid residues. Thirty-three of the 59 residues are either Glu or Asp. Transcription activation domains rich in acidic residues are commonly found in gene activators. These acidic domains may interact with basal transcription factors (32). Hence, the acidic domain of Rc may serve in a transcription regulatory capacity and/or in protein−protein interactions. Following the acidic domain are five copies of the Ser/Thr-Pro-X-Arg/Lys (SPXK) sequence (where X is a non-acidic residue, usually Lys or Arg) (33). These motifs may bind DNA (34) and are frequently found in DNA binding proteins that are involved in gene regulation (33).

When the putative Rc protein sequence was compared to sequences in the NCBI database, a stretch of 128 amino acid residues (residues 199 to 326, boxed in Fig. 2), corresponding to the zinc fingers and the acidic domain (Zn-A), showed striking similarities to six other DNA binding proteins. Based on the degree of sequence similarities in their Zn-A regions (Fig. 3), the Rc-related gene products can be categorized into three groups. Group I consists of mouse αA-CRYBP (35), rat ATBP2 (19), and human MBP1 (15−17). Group II includes rat ATBP1 (19) and human MBP2 (15,17). Group III consists of mouse Rc and human KBP1 (15).

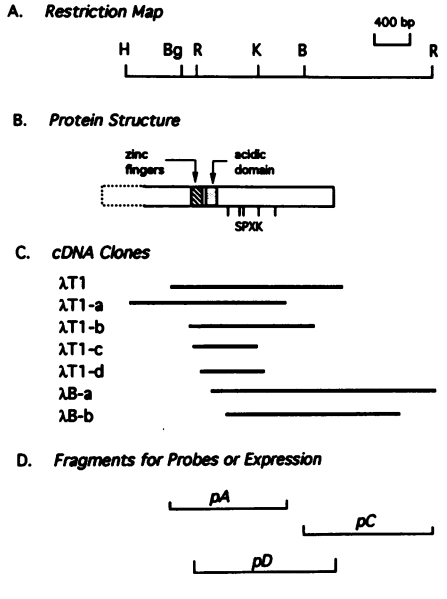

Figure 1. cDNA clones and predicted domain structure of Rc. (**A**) A restriction map of a ~3.2 kb Rc cDNA. H, *Hind*III; Bg, *Bgl*II; R, *Eco*RI; K, *Kpn*I; and B, *Bam*HI. (**B**) Protein domains of the predicted amino acid sequence. A vertical stroke represents a SPXK sequence. The N-terminus of the protein has not been defined and is represented with dots. (**C**) Overlapping Rc cDNA clones. The black bars correspond to the length of each cDNA clone and the name of each clone is indicated on the left of the figure. Clones isolated from the thymocyte or the brain cDNA libraries are initialed as λT or λB, respectively. (**D**) DNA fragments isolated from the Rc cDNA for expression (p*A*) or for generation of hybridization probes (p*C* and p*D*).

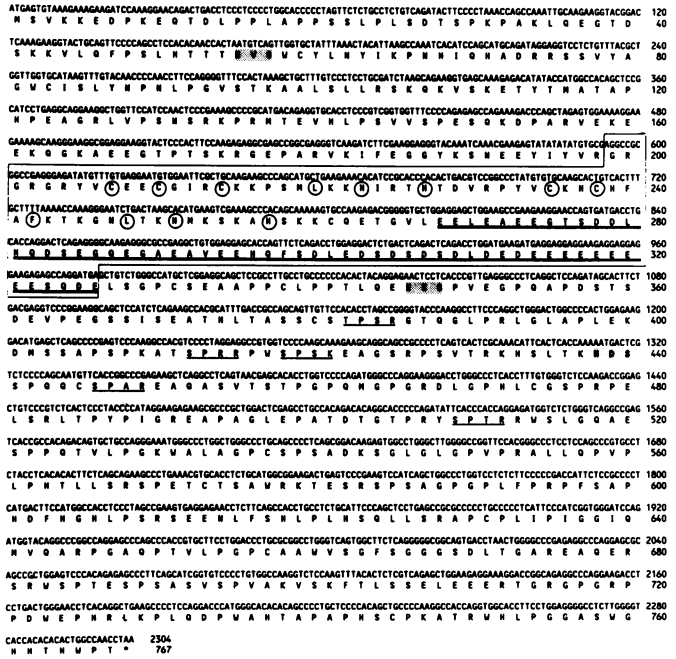

Figure 2. Nucleotide sequence and predicted amino acid sequence of Rc cDNA. The predicted amino acid sequence is shown below the DNA sequence. The translation termination codon is marked with an asterisk. The conserved residues in the putative zinc-fingers are circled. The acidic domain is doubly underlined. The SPXK sequences are underlined. The three potential asparagine glycosylation sites, N-X-S, are shaded. The region of the zinc fingers and the acidic domain aligned with the Rc-related proteins shown in Fig. 3 is boxed.

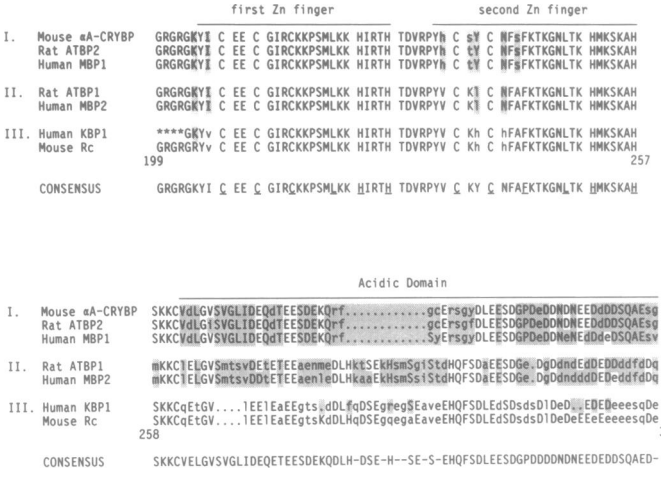

Figure 3. Amino acid sequence alignment of the zinc fingers and the acidic domain of Rc and Rc-related proteins. The sequences, derived from Ref. 14–17, 19 and 35, were compared with the software PILEUP and aligned by the program PRETTY from the Wisconsin Genetic Computer Group (25). Amino acid numberings are according to the Rc protein sequence shown in Fig. 2. The consensus sequence is shown beneath the alignment. Those residues that are identical to or different from the consensus sequence are shown in upper or lower case, respectively. The conserved Cys, His, Leu and Phe residues in the zinc fingers are underlined. Those residues that are different from Rc are shaded. A dot represents the sequence not present. An asterisk represents KBP1 sequence that is not available.

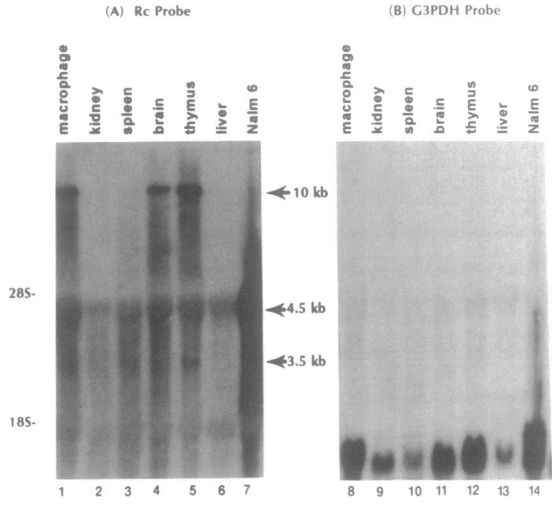

Figure 4. Expression of Rc in mouse tissues and in a human cell line. Total RNA (~20 μg) was electrophoresed through a 0.8% formaldehyde-agarose gel blotted, and hybridized with (**A**) a Rc specific probe p*C*, and (**B**) a glyceraldehyde 3-phosphate dehydrogenase (G3PDH) probe, after standard procedures (30).

The first zinc fingers of the Rc-related proteins are nearly identical to that of Rc. There is only one difference, at position 204, between Rc and KBP1; while Rc has an Arg, KBP1 and all others have Lys. There is a total of five substitutions clustered around the two Cys among the second fingers. In general, each finger is rich in charged residues. There are two Glu, four Lys and four Arg in each first finger, and five to six Lys in each second finger. In contrast to the high degree of similarity of the zinc fingers, the acidic domains show more differences and are

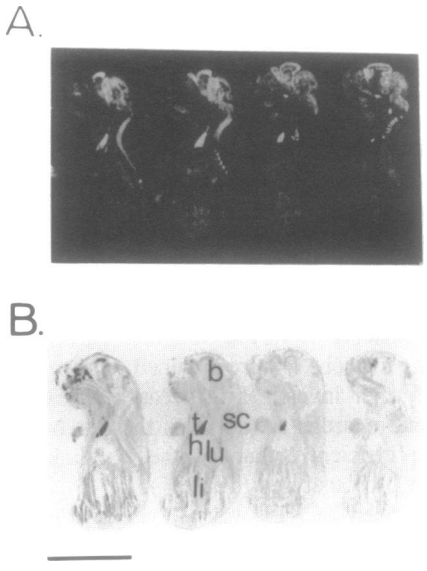

Figure 5. *In situ* hybridization showing the tissue expression of Rc transcripts in mouse embryos. Parasagittal sections of day 19 embryos of a BALB/c mouse were counterstained with thionin and autoradiographed after hybridization with the anti-sense Rc probe p*D*. (**A**) Dark-field exposure, and (**B**) Bright-field exposure. Abbreviations: b, brain; t, thymus; sc, spinal cord; h, heart; lu, lung, and li, liver. (Bar = 10 mm).

conserved only among members of the same group. If the acidic domains are involved in regulatory functions, the sequences of the Zn-A regions of these proteins suggest that, although these proteins may bind to similar DNA sites, they could trigger different physiological responses.

Rc transcripts are present in adult and fetal thymuses

The expression of Rss binding proteins may correlate with the V(D)J recombinase activities. The distribution of *Rc* transcripts among various tissues of a 4-week-old C3H mouse and a human pre-B cell line, Nalm 6, was investigated by Northern blot analysis using probe p*C* (Fig. 4A). A ~3.5 kb transcript was detected in RNA samples isolated from peritoneal macrophages and from the thymus. A ~4.5 kb message was detected at high levels mainly in Nalm 6. A hybridization signal of ~10 kb was observed in the RNA samples isolated from the macrophages, the brain and the thymus, but not in samples from the kidney, spleen, liver, or the Nalm 6 cell line. In essence, Rc transcripts of different size have been detected in different tissues/cells and some of them appear restricted to the V(D)J recombination competent tissues or cells. The sizes of the Rc transcripts suggest that the cDNA presented in Fig. 2 is not full length. As a control, the blot was hybridized with a glyceraldehyde 3-phosphate dehydrogenase (G3PDH) probe. Hybridization signals of ~2 kb for the G3PDH transcripts were observed in all samples, suggesting that these RNA samples were intact (Fig. 4B).

The distribution of *Rc* transcripts in mouse embryos was investigated by *in situ* hybridization. Embryonic day 19 mouse embryo sections were hybridized with probe p*D*. Strong hybridization signals were detected from the thymus, the brain, and the spinal cord (Fig. 5); the sense strand probe did not give detectable signals (data not shown). Weak hybridization signals were observed in the liver and the lung. No signal was detected

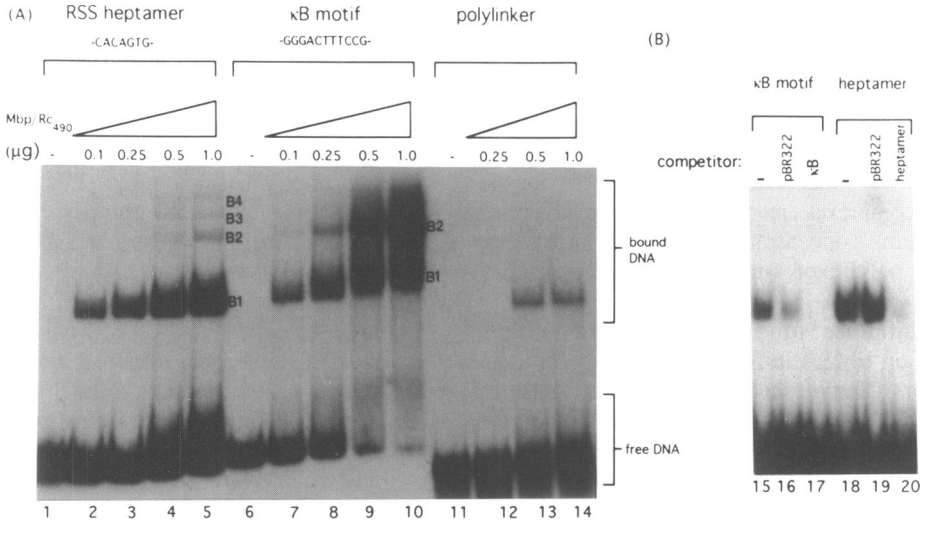

Figure 6. EMSA of Mbp/Rc_{490} with Rss heptamer or κB motif. (**A**) DNA fragments (0.2 ng, ~5,000 cpm) containing the Rss heptamer (probe *H*), the *κB* motif (probe *κB*), or the polylinker of pBluescript, were incubated with increasing concentrations (from 0.1 to 1 μg) of Mbp/Rc_{490}. DNA–protein complexes and free DNA were resolved on 5% non-denaturing polyacrylamide gels. (**B**) Competition EMSA. Probe *H* or probe *κB* (0.2 ng, ~5,000 cpm) were incubated with 0.2 μg of Mbp/Rc_{490} in the presence of a 100-fold excess (20 ng) of competitor DNA, as indicated at the top of each lane. (–), no competitor; pBR322 is *Msp*I digested pBR322 DNA.

from other organs, such as the heart. The detection of Rc transcripts in adult and embryonic thymuses suggests that Rc may be involved in T cell development.

Rc binds to the Rss heptamer and to the κB motif

The specific interactions of Rc with DNA ligands have been investigated by four different approaches: Southwestern blot analysis with normal and mutagenized Rss heptamer DNA, EMSA, competition assay and methylation interference/ footprinting experiments. Bacterial fusion proteins for Rc were used for these experiments.

The specific sequence within the Rss which interacts with the Rc-fusion proteins was initially defined by Southwestern blot analysis. Western blots containing a β-gal/Rc fusion protein, produced by subcloning the insert of λT1 into pUR (22), were hybridized with designated short double-stranded DNA fragments. The results of two such experiments are shown in Table 1. Probe *H*, containing the Rss *h*eptamer, bound ~10 times more efficiently to β-gal/Rc than the *n*onamer probe *N*. Binding of probe *N* to the fusion protein was similar to the presumed background binding of the polylinker. The binding specificity of Rc to the Rss heptamer was further indicated by a >7 fold decrease in binding when probe *H'*, in which the heptamer sequence CACAGTG was mutagenized to CAGAGTG, was used. This decrease in binding affinity is in keeping with previous results, in which this mutation of the heptamer decreased the efficiency of recombination (4), as well as the binding affinities of the $V_{κ21}$ binding protein (13) and T160 (10).

The specificities of Rc to the Rss heptamer were further demonstrated by using probe *H* for direct binding or competition by electrophoretic mobility shift assays (EMSA). Since β-gal/Rc fusion protein was insoluble, Rc cDNA was expressed with the vector pMAL-c to produce Mbp/Rc_{490} in bacteria (27). Mbp/Rc_{490} contains 490 Rc residues including the zinc fingers, the acidic domain, and the SPXK sequences at the carboxyl end of the maltose binding protein (Mbp). When Mbp/Rc_{490} (0.1 μg) and probe *H* were employed in EMSA, a DNA–protein complex labeled B1 was observed (Fig. 6A, lane 2). The amount of B1 increased with increasing protein concentrations from 0.1 μg to 1.0 μg. In addition, a ladder of slower migrating complexes, B2, B3 and B4, was observed when 0.5 μg or 1 μg of proteins was used (Fig. 6A, lanes 4 & 5).

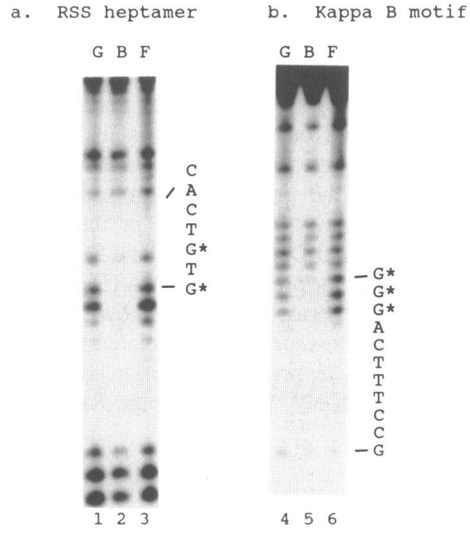

Figure 7. Methylation interference analysis. Labeled and partially methylated DNA (fragment *H* or *κB*, ~1 ng, 25,000 cpm) was incubated with 0.5 μg of Mbp/Rc_{490}. After incubation, bound (B) and free (F) DNAs were resolved in 5% non-denaturing polyacrylamide gels and recovered. The DNA samples were cleaved at sites of methylation with piperidine, and the products analyzed with sequencing gels. The G lanes show the pattern obtained by chemically cleaving the probes at G-residues prior to EMSA. The locations and the nucleotide sequences of the Rss heptamer and *κB* motifs are shown. The positions of the hypomethylated G-residues are indicated by asterisks.

Since Rc shows striking amino acid sequence similarity to κB binding proteins, EMSA was performed with Mbp/Rc_{490} using probe *κB* as a ligand so as to investigate whether Rc also binds to the *κB* motif. Protein−DNA complexes designated as B1 and B2 were observed (Fig. 6B, lanes 6 to 10). These results suggest that Mbp/Rc_{490} binds to DNA sequences containing the κB motif as well as the Rss heptamer. Moreover, the Rc-Rss or the Rc-κB DNA−protein complexes can form multimeric structures with increased protein concentrations. For comparison, Mbp/Rc_{490} bound the polylinker with much lower efficiency (Fig. 6A, lanes 11−14).

DNA binding specificities of Mbp/Rc_{490} were further illustrated by competition EMSA in the presence of specific or non-specific competitors. The presence of a 100-fold excess of the unlabeled heptamer DNA *H* competed away the binding of Mbp/Rc_{490} to probe (radioactive) *H*, while pBR322 DNA competed insignificantly (Fig 6B, lanes 18 to 20). In analogous binding reactions using probe (radioactive) κB as a ligand, unlabeled fragment κB efficiently competed against the binding of Mbp/Rc_{490} to probe *κB*, while ~50% binding was retained when a similar amount of pBR322 DNA was used (Fig. 6B, lanes 15−17).

To further establish the DNA binding specificities of Mbp/Rc_{490}, the contact nucleotides of Mbp/Rc_{490} with its DNA ligands were determined by methylation interference experiments. Comparing the patterns of partially methylated DNA bound (Fig. 7, lane 2), free (Fig. 7, lane 3), and prior to EMSA (Fig. 7, lane 1), shows that the binding of Mbp/Rc_{490} to its ligand is perturbed by methylation of the G residues within the Rss heptamer. Except for the two G-residues that are immediately 3′ to the heptamer, no other G residues in the probe show significant methylation interference. The nucleotide binding sites of Mbp/Rc_{490} to probe *κB* were also determined. The data show that the three consecutive G residues in the κB motif are hypomethylated in bound DNA (Fig. 7, lane 5) when compared to free (Fig. 7, lane 6), or pre-EMSA DNA (Fig.7, lane 4). The methylation of these G residues has been shown to interfere with protein binding using nuclear extracts isolated from B cell lines (18). Hence, these DNA−protein interaction analyses show that MBP/Rc_{490} binds precisely, not only to the Rss heptamer, but also to the κB motif.

DISCUSSION

Here we report the cloning of a murine gene, Rc, by screening a λgt11 expression library with the signal sequences for V(D)J recombination. The conserved residues in each Rc zinc finger are Tyr-X-Cys-X_2-Cys-X_3-Phe/Cys-X_5-Leu-X_2-His-X_{3-5}-His, assigning them to the Cys_2-His_2 class (36). The zinc fingers and acidic domain structures similar to that of the Rc protein are found in human *KBP1*, *MBP1* and *MBP2*, rat *ATBP1* and *ATBP2*, and mouse *αA-CRYBP* proteins (Fig.3). These genes were cloned either by screening expression libraries with specific enhancer sequences or by cross-hybridization with available cDNA probes. The structure of a Cys_2-His_2 finger consists of a N-terminal antiparallel β-sheet followed by about three turns of α-helix which forms the major DNA-binding surface (37). Since the helical regions of the Rc-related zinc fingers are nearly identical (Fig. 3), it is likely that these proteins will bind similar DNA motifs, such as the Rss heptamer, provided that the flanking protein sequences do not affect DNA binding.

Despite the fact that the zinc fingers of the Rc-related proteins are very similar, each protein appears to display distinct affinities to different sequences (15,19). One possible structure present in these proteins that may affect DNA binding is the SPXK sequences. This sequence has been shown to bind DNA at 'AT' rich sites (34). There are five copies of the SPXK sequence in the Rc protein presented in Fig. 2. The Rc-related proteins have been classified into three groups according to the amino acid sequences of the zinc fingers and the acidic domains (Fig. 3). Of the group I proteins, αA-CRYBP, ATBP2, and MBP1 contain two, zero, and two copies of the SPXK sequence, respectively. Of the group II proteins, ATBP1 has ten, and MBP2 has eleven copies of this sequence. Hence, it appears that in addition to the Zn-A domains, the numbers of the SPXK sequence are also more comparable among members of a group.

Southwestern blotting experiments, competition EMSA, and methylation interference analyses have shown that the Rc fusion proteins bind specifically to the Rss heptamer, and to the κB motif. In EMSA, Mbp/Rc_{490} showed similar binding affinities to DNA ligands harboring a 12 bp-spacer Rss, a 23 bp-spacer Rss, or the Rss heptamer (data not shown). Hence, we conclude that the heptamer binding of Rc is not affected by the presence or absence of the nonamer. The ability of Rc to bind to Rss heptamer alone conforms with the observation that V(D)J recombination can occur between Rss and a solitary heptamer (4,38). The formation of the multimeric DNA-protein complexes for Rc-Rss is intriguing and is the subject of an accompanying paper (27).

Apart from the Rss heptamer, it has also been shown that Rc also binds to the κB motif. The Rss and the κB are the regulatory sequences for Ig gene rearrangement and transcription, respectively. Rc is distinct from the transcription factor NF-κB, which is a heterodimer composed of 50 kDa and 65 kDa DNA-binding subunits (41). The binding activities for the κB motif are constitutively present only in those B cells of the appropriate developmental stage (18) and are crucial for κ enhancer function (39,40).

The relative binding affinities of Rc to the Rss heptamer and the κB motif remained to be determined. Mutational analysis of the Rc protein is needed to establish specific recognition domains within the molecule. A recent report has shown that deletion of the Ig κ chain enhancer abolished κ chain gene rearrangement (42). In addition, the activation of the Ig κ gene rearrangement correlates with the induction of germline κ gene transcription (43). The relevance of Rc's dual binding specificities to the above observation is under investigation. At this stage, it is of considerable interest that the Rc protein binds to both Rss heptamer and κB sequences.

ACKNOWLEDGEMENTS

We thank Drs C.Yung Yu and Philip Johnson for critically reviewing the manuscript, and A.Forster, I.Lavenir, and J.Strandtmann for expert technical assistance. This research was supported in part by The Ohio State University Seed Grant (#221396), the Croucher Foundation, The Bremer Foundation (#9106), and the American Cancer Society (#IRG-16-30 and #38618-5500) (L.C.W.). The nucleotide sequence reported in this paper has been submitted to the GenBank Data Bank with accession number L07911.

REFERENCES

1. Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979) *Nature* **280**, 288–294.
2. Seidman, J.G., Nau, M.M., Norman, B., Kwan, S-P., Scheriff, M. and Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6022–6026.
3. Lieber, M.R., Hesse, J.E., Mizuuchi, K. and Gellert, M. (1987) *Genes Dev.* **1**, 751–761.
4. Hesse, J.E., Lieber, M.R., Mizuuchi, K. and Gellert, M. (1989) *Genes Dev.* **3**, 1053–1061.
5. Matsuoka, M., Nagawa, F., Okazaki, K., Kingsbury, L., Yoshida, K., Muller, U., Larue, D.T., Winer, J.A. and Sakano, H. (1991) *Science* **254**, 81–86.
6. Akira, S., Okazaki, K. and Sakano, H. (1987) *Science* **238**, 1134–1138.
7. Oettinger, M.A., Schatz, D.G., Gorka, C. and Baltimore, D. (1990) *Science* **248**, 1517–1523.
8. Schatz, D.G., Oettinger, M.A. and Baltimore, D. (1989) *Cell* **59**, 1035–1048.
9. Matsunami, N., Hamaguchi, Y., Yamamoto, Y., Kuze, K., Kangawa, K., Matsuo, H., Kawaichi, M. and Honjo, T. (1989) *Nature* **342**, 934–937.
10. Shirakata, M., Hüppi, K., Usuda, S., Okazaki, K., Yoshida, K. and Sakano H. (1991) *Mol. Cell. Biol.* **11**, 4528–4536.
11. Li, M., Morzychka-Wroblewska, E. and Desiderio, S.V. (1989) *Genes Dev.* **3**, 1801–1803.
12. Muegge, K., West, M. and Durum, S.K. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4151–4155.
13. Aguilera, R.J., Akira S., Okazaki, K. and Sakano, H. (1987). *Cell* **51**, 909–917.
14. Fan, C-M. and Maniatis, T. (1990) *Genes Dev.* **4**, 29–42.
15. Rustgi, A.K., Laura, J., van't Veer, L.J. and Bernards, R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8707–8710.
16. Maekawa, T., Sakura, H., Sudo, T. and Ishii, S. (1989) *J. Biol. Chem.* **264**, 14591–14593.
17. van't Veer, L.J., Lutz, P.M., Isselbacher, K.J. and Bernards, R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8971–8975.
18. Sen, R. and Baltimore, D. (1986) *Cell* **46**, 705–716.
19. Mitchelmore, C., Traboni, C. and Cortese, R. (1991) *Nucleic Acids Res.* **19**, 141–147.
20. Lenardo, M.J. and Baltimore, D. (1989) *Cell* **58**, 227–229.
21. Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and McKnight, S.L. (1988) *Genes Dev.* **2**, 801–806.
22. Rüther, U. and Müller-Hill, B. (1983) *EMBO J.* **2**, 1791–1794.
23. Boehm, T., Gonzalez-Sarmiento, R., Kennedy, M. and Rabbitts, T.H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3927–3931.
24. Staden, R. (1982) *Nucleic Acid. Res.* **10**, 4731–4751.
25. J. Devereux, P. Haeberli and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
26. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
27. Mak, C-H, Standtmann, J. and Wu, L.C. (submitted).
28. Maxam, A.M. and Gilbert, W. (1980) *Meth. Enzymol.* **65**, 499–560.
29. Singh, H., LeBowitz, J.H., Baldwin, A.S. and Sharp, P.A. (1988) *Cell* **52**, 415–423.
30. Chirgwin, J.M., Przybyla, A.G., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* **18**, 5294–5299.
31. Klug, A. and Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–467.
32. Ptashne, M. (1988) *Nature* **335**, 683–689.
33. Suzuki, M. (1989) *J. Mol. Biol.* **207**, 61–84.
34. Churchill, M.E.A. and Suzuki, M. (1989) *EMBO J.* **8**, 4189–4195
35. Nakamura, T., Donovan, D.M., Hamada, K., Sax, C.M., Norman, B., Flanagan, J.R., Ozato, K., Westphal, H. and Piatigorsky, J. (1990) *Mol. Cell Biol.* **10**, 3700–3708.
36. Evans, R.M. and Hollenberg, S.G. (1988) *Cell* **52**, 1–3.
37. Coleman, J.E. (1992) *Annu. Rev. Biochem.* **61**, 897–946.
38. Kleinfield, R., Hardy, R.R., Tarlinton, D., Dangl, J., Herzenberg, L.A. and Weigert, M. (1986) *Nature* **322**, 843–846.
39. Lenardo, M.J., Kuang, A., Gifford, A. and Baltimore, D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8825–8829.
40. Atchison, M. and Perry, R.P. (1987) *Cell* **48**, 121–128.
41. Blank, V., Kourilsky, P. and Israδl, A. *Trend Biochem. Sci.* **17**, 135–140.
42. Takeda, S., Zou, Y-R., Bluethmann, H., Kitamura, D., Muller, U. and Rajewsky, K. (1993) *EMBO J.* **12**, 2329–2336.
43. Schlissel, M. and Baltimore, D. (1989) *Cell* **58**, 1001–1007.