Unique organization of the human BCR gene promoter

Qin-Shi Zhu, Nora Heisterkamp and John Groffen*
Section of Molecular Diagnosis, Department of Pathology, Childrens Hospital of Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA 90027, USA

Received April 13, 1990; Revised and Accepted August 1, 1990

ABSTRACT

The promoter of the human BCR gene, regulating the transcription of the chimeric BCR/ABL mRNA in leukemia, has been isolated and characterized. A region of 1.1 kb immediately 5′ to the transcription start site was analyzed in detail by sequencing, DNase1 footprinting, gel retardation and functional studies. These experiments localized a minimal promoter to a 650 bp sequence, composed of 270 bp of 5′ flanking sequences and 380 bp of exon 1 transcribed sequences. The promoter region includes a TTTAA box, one Sp1 site and a novel protein-binding sequence absolutely necessary for efficient transcription *in vivo*. Six additional protein-binding regions were identified more to the 5′. Of these, one is found in an inverted repeat in the 3′ coding and splice donor region of BCR exon 1.

INTRODUCTION

The Philadelphia (Ph) chromosome is a derived chromosome 22 found specifically in the leukemic cells of chronic myelocytic leukemia (CML) patients and in 17–25% of patients with acute lymphoblastic leukemia (ALL) (1–4). The Ph chromosome consists of chromosome 22 sequences up to band q11 fused to the distal part of chromosome 9 (9q34-qter); on a genomic level, 5′ sequences of the BCR gene including its promoter and coding sequences are fused to 3′ sequences of the ABL oncogene in a head-to-tail fashion (5,6). A chimeric BCR/ABL mRNA transcribed from this fusion gene is initiated on the BCR gene promoter (7–9). The BCR/ABL fusion proteins produced from the chimeric gene are strongly associated with Ph-positive leukemia and are directly or indirectly causative for this malignancy (10,11). The proteins exhibit an abnormal tyrosine-specific protein kinase activity from the ABL moiety of the protein (12–15).

The entire BCR gene has been molecularly cloned; it encompasses a region of approximately 130 kb on chromosome 22 and contains a minimum of 21 exons (16). In man, the gene produces two transcripts of 7.0 and 4.3 kb; the 4.3 kb species has been cloned and sequenced (17). Despite considerable research efforts the organization of the 7.0 kb mRNA remains unclear. However, all sequences present in the 4.3 kb species are also present in the 7.0 kb mRNA (18). The BCR transcripts are found in a wide variety of cells, where they are usually produced in relatively low abundance. The highest level of expression is in brain (19).

The product of the BCR gene is a phosphoprotein of 160,000 molecular weight (20) and is cytoplasmic (21,22); in *vitro*, the protein exhibits an associated kinase activity and is phosphorylated on serine residues (20–22).

Although the role of the ABL moiety of the BCR/ABL protein has been well investigated, the contribution of the BCR amino acid residues is unclear. However, it is the BCR gene promoter which drives the production of the aberrant BCR/ABL mRNAs found in CML and Ph- positive ALL and its contribution must be crucial in these types of leukemia. To date, the 5′ region of this gene has not been sequenced, nor has the promoter been defined. In the present study, we have examined this region in detail by nucleic acid sequencing, DNase1 footprinting assays, gel retardation assays and functional assays *in vivo*. In addition, we have cloned and compared the human and mouse BCR promoter sequences. Our results show that the human BCR gene promoter has several unique features which distinguish it from its murine counterpart as well as from other promoters.

MATERIALS AND METHODS

Molecular cloning

A 6.2 kb SalI human BCR genomic DNA fragment containing the 1.2 kb exon 1 and 5 kb of 5′ flanking sequence was subcloned from a BCR genomic phage clone into pUC 8. An 0.95 kb BamHI fragment was subcloned into pUC 8 from a human BCR genomic cosmid clone (16); this fragment contains 0.8 kb of the 3′ end of exon 1 and 0.15 kb of intron 1. Two mouse genomic clones containing the murine BCR exon 1 and 5′ and 3′ flanking sequences were isolated from a mouse genomic phage library in lambda gt10 (Clontech) and subcloned into pUC 18. A 1.8 kb HindIII/EcoRI fragment containing 1.6 kb of coding sequences and 0.3 kb of 5′ flanking region as well as a 1.3 kb BamHI/HindIII fragment containing the additional 5′ sequences were subcloned into pSK (Stratagene). Sequencing was performed on both strands of DNA fragments subcloned into M13mp18 or 19 (BRL) using ^{35}S dATP.

Promoter activity assay

DNA fragments (1.5 kb PstI/SmaI, 1.04 kb ApaI/SmaI, 0.65 kb PstI/SmaI, 0.49 kb TaqI/SmaI and 0.16 kb PstI/TaqI

* To whom correspondence should be addressed

fragments, see Figure 8) for promoter activity measurement were isolated from the BCR genomic clone. After modifying their end restriction sites, these DNA fragments were inserted in the promoterless human growth hormone expression vector pOGH (Nichols Institute Diagnostics). Plasmids were transfected into NIH 3T3 cells using the low pH calcium phosphate precipitation method (23). All DNA preparations used for transfection were purified through cesium chloride gradient centrifugation. 24 hours after the transfections, the human growth hormone produced was assayed with an immunoassay kit (Nichols Institute Diagnostics) according to the manufacturer's instructions. Each fragment was assayed at least twice (2−5 times) and the average was taken as the results given in Figure 8.

Gel retardation and footprinting assays

Preparation of nuclear extracts from K562 (a CML cell line) and A498 (a kidney carcinoma cell line) as well as gel retardation and DNase1 footprinting analyses were performed as previously described (24). For gel retardation assays, 2.5 μg of poly(dAdT) (Sigma) was preincubated with 5 μg nuclear protein for 10 min. at room temperature prior to the addition of end-labelled DNA fragments (approximately 10 fmol). After incubation of another 30 min. at room temperature, the samples were electrophoresed through a 5.5% nondenaturing acrylamide (acrylamide: bisacrylamide = 29:1) gel. For footprinting analysis, 2.5 μg of poly(dAdT) was preincubated with 11 μg nuclear extract. After the addition of 1−5 fmol end-labelled DNA (one end), the mixture was incubated (room temperature) for another 20 min. before DNase 1 digestion. Partially digested DNA was phenolized and ethanol precipitated. Samples were heated for 2 min. at 90°C in the presence of loading buffer and loaded on a 7.5% acrylamide-8M urea gel (acrylamide:bisacrylamide = 40:1).

Transgenic mice and Northern blot analysis

Transgenic mice were made as described by (25). The construct used for microinjection contained approximately 3.5 kb of 5′ flanking sequences joined to a 4.3 kb BCR mRNA isolated from a human kidney carcinoma cDNA library in lambda gtwes (unpublished results). Total RNA was isolated using guanidine thioisocyanate (26). Northern blots were run as described (27).

RESULTS

The human BCR gene transcribes two mRNAs, of 4.3 and 7.0 kb. To date, all probes from the 4.3 kb species have been found to hybridize to the 7.0 kb species. It is possible, that the 4.3 kb mRNA does not have its own promoter and is derived by RNA processing from the 7.0 kb species. To examine this, we tested an extended region 5′ to the gene *in vivo* for promoter activity. In initial experiments, we made constructs containing approximately 3.5 kb of sequences 5′ to the 5′ end of a 4.3 kb BCR cDNA joined to the cDNA and introduced these into transgenic mice. The construct produced the human mRNA of 4.3 kb in three independent transgenic lines (Figure 1, lanes 1−3); virtually no hybridization signal is visible in the non-transgenic control (lane 4). Most likely no human 7.0 kb mRNA is produced, as compared to mouse and human mRNA controls (Figure 1, lanes 4−5); the stronger intensity of the 7.0 kb band seen in lanes 2 and 3 is due to the slightly larger amount and/or superior quality of RNA loaded in these lanes. In mouse, the 7.0 kb species is much more prominent than the approximately 4.3 kb species (lane 4). The level of expression of the human

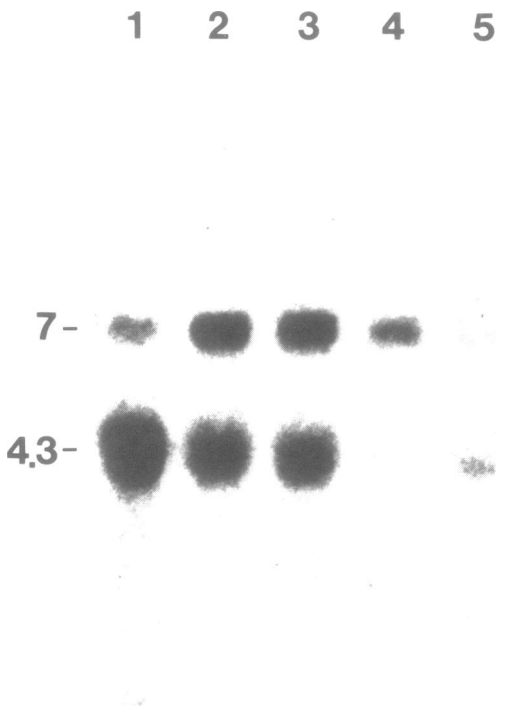

Figure 1. Northern blot analysis of transgenic BCR gene expression. Each lane contains approximately 10 ug of total RNA; as probe a cDNA fragment for human BCR gene exon 2 was used. Hybridization of this blot to an unrelated probe showed that lanes 1,4 and 5 contain approximately 1.5−2 times less RNA than lanes 3 and 4. Post-hybridization washing was at 0.3×SSC at 65°C. Lanes 1−3; brain RNA from transgenic mice 22−0, 112−0 and 120−0. Lane 4; brain RNA from non-transgenic control mouse. Lane 5; RNA from the human glioblastoma cell line A172.

4.3 kb transcript was highest in brain and substantial in kidney (about half the quantity observed in brain), whereas expression was low in fibroblasts (hybridization signal just above background [not shown]). This is in concordance with levels of expression found in chicken, mouse and human tissues and cell lines (19, unpublished observations) and indicates that at least some tissue-specific controlling elements are present immediately 5′ to the BCR gene as opposed to a location within intron sequences. In addition, these data confirm that the 4.3 kb mRNA is initiated from DNA sequences in relative proximity to the 5′ coding region of the gene; indeed, an 0.42 kb BglII/ApaI probe encompassing sequences starting 643 nucleotides upstream of the generally accepted initiation site (Figure 8, nucleotide 981) did not hybridize to either 4.3 kb or 7.0 kb mRNA species on a Northern blot (not shown). This initiation site has been designated by various investigators (7,17,18,30) who reported that this region coincides with the most 5′ end of BCR and BCR/ABL cDNAs. A second putative initiation site may be present approximately 100 nucleotides upstream of the initiation codon; multiple cDNAs starting at this position have been isolated (17,30, our unpublished results).

To examine the promoter in detail, a region of 1.1 kb 5′ to the 5′ end of the cDNA was subcloned and mapped in detail (see Figure 2A). To identify regions of regulatory importance, fragments from the 5′ region (Figure 2C) were tested for protein binding using nuclear extracts from the cell line K562, which expresses BCR gene mRNA. A 5′ 0.54 kb PstI/NcoI fragment showed protein binding (Figure 3, lane 1) and two DNA-protein

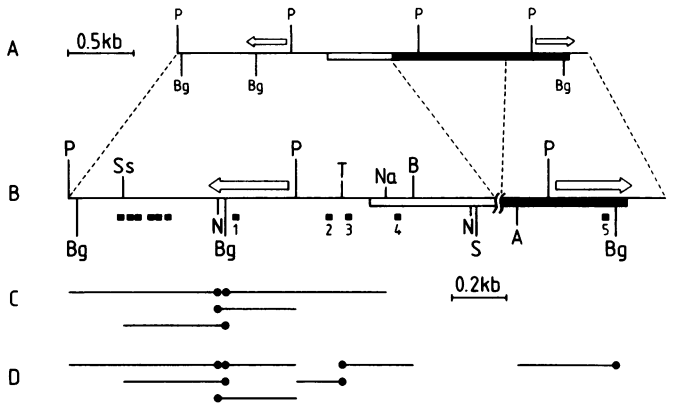

Figure 2. Restriction enzyme map of BCR gene exon 1 and promoter region and strategy used for gel retardation and DNase1 footprinting experiments. The open double line in A represents the 5′ untranslated region and the closed bar indicates the coding region of exon 1 of the BCR gene. The repeat sequences are indicated by arrows directly above the map. B is an enlarged part of A. DNA sequences protected by nuclear extracts in DNase1 footprinting experiments are shown as solid boxes below the map; sequences 1−5 are numbered. In C, lines indicate DNA fragments used in gel retardation assays; the circled end represents the labelled end. In D, DNA fragments used in DNase1 footprinting experiments are shown, with the circled end representing the labelled end. Enzymes used include: ApaI (A), BalI (B), BglII (Bg), NaeI (Na), NcoI (N), PstI (P), SmaI (S), SstI (Ss), and TaqI (T). Not all TaqI, BalI and NcoI sites within the 3′ end of the exon are shown.

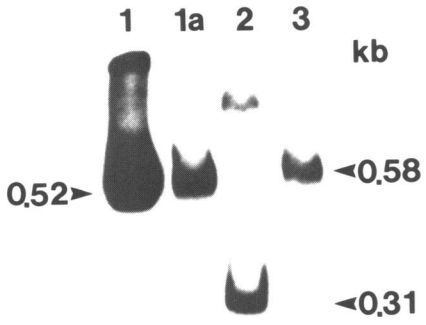

Figure 3. Gel retardation assay of the BCR promoter region. Lanes 1, 2 and 3 show the binding of K562 nuclear proteins to the 0.54 kb PstI/NcoI, 0.29 kb NcoI/PstI and 0.58 kb BglII/NaeI fragments, respectively. 1a is a lighter exposure of lane 1. The position of DNA fragments (no protein bound) is indicated.

bands are visible. This suggests the presence of more than one protein binding sequence in that region. An 0.29 kb NcoI/PstI fragment immediately to the 3′ also shifted in mobility but showed only a single band (Fig. 3, lane 2), suggesting a single protein binding site. Surprisingly, an 0.58 kb BglII/NaeI fragment which partly overlaps with the 0.29 kb NcoI/PstI fragment and contains all sequences 5′ of and including the putative transcription start site (also see below) did not clearly bind protein (Figure 3, lane 3);possibly, this can be explained by the relatively large size of this fragment.

Next, the DNA sequence of the entire region was determined, to identify known recognition sites for transcription factors as well as provide a basis for DNase1 footprinting experiments (Figure 4). The cDNA containing the longest 5′ region isolated to date (7) has its 5′ terminus 200 bp 3′ to a PstI site (see Figure 4, arrow at nucleotide 981). A TTTAA sequence is located 50 bp upstream of this site (see Figure 4, boxed). Not unlikely, this sequence may constitute a functional TATA box. The sequence is similar to that found in the ABL 1a and ADA promoter (TTAA, [23,28]) and in human the thymidine kinase promoter (TTTAAA, [29]). In addition, there is a cap site consensus-like sequence (actual sequence:TCATCG; consensus sequence PyCATTCPu) present 19 bp 3′ to the TTTAA sequence. One potential Sp1 binding site was found 23 bp 5′ to the TTTAA sequence. A putative TCCAATT ('CAT') sequence and a reversed CCAATA sequence are present 120 bp and 200 bp 5′ to the TTTAA sequence (Figure 4, overlined) but these CAT sequences showed no evidence of functioning *in vitro* or *in vivo*. A reversed CAATTG sequence 670 bp 5′ of the TATA box did not appear to be functional either (also see below). No other protein binding sequences containing consensus sequences known to bind transcription factors were found further upstream.

Previous unpublished experiments aimed at cloning the first exon of the BCR gene had revealed an interesting feature of this region: a repeat-free probe from the 3′ end of exon 1 was found to hybridize to sequences 5′ of the coding region. The nucleotide sequence analysis identified this repeated sequence: the 3′ coding and splice donor region of exon 1 is present in the 5′ region of the gene in an inverted orientation (see Figure 2, arrows, and Figure 5 and 6). The two areas have 91% homology and are 277 bp in length (from nucleotide 2468 to 2744 in the 3′ region). The homology at the 3′ end terminates after the splice donor site of exon 1. Interestingly, the 5′ repeat is situated on a DNA fragment that tested positive in the gel retardation assay (the 0.29 NcoI/PstI fragment, see Figure 3).

This sequence did not hybridize to additional sequences in the human genome after washing to high stringency indicating that it is unique to the BCR gene. A probe prepared from exon 1 encompassing part of this region detected three TaqI restriction enzyme fragments of similar size in man, gorilla and chimpanzee; in man, these fragments represent the 5′ repeat, an exon 1 coding fragment and the 3′ repeat. In mouse DNA, only two fragments were detected, suggesting that the repeat sequence is a relatively novel addition to the gene in evolutionary terms (not shown).

To determine conclusively, if the repeat was absent from the mouse BCR gene, a genomic fragment containing the murine BCR gene exon 1 and flanking sequence was cloned from a mouse genomic library. The mouse exon 1 coding sequence (unpublished data) and 5′ flanking sequences (see Figure 4) was determined; this showed that the sequence was present in the 3′ coding and splice donor regions and was highly conserved between man and mouse. However, it was absent from the 5′ region of the mouse BCR gene promoter (Figure 4).

Different fragments (see Figure 2D) were subsequently tested for the binding of protein using DNase1 footprinting. Extracts were prepared from two different cell types (kidney fibroblast and myeloid); the cell lines used, A498 and K562, both express a relatively high level of BCR gene mRNA. Eleven discrete protein-binding sequences were identified with this method (Figure 2B and Figure 6); six sites are situated on the 5′ 0.54 kb PstI/NcoI fragment which had shown activity in the gel retardation assay. The actual sequence of these six regions (Figure 4, underlined) indicates that they do not correspond to factor-binding sequences identified previously.

The inverted 5′ repeat appears to have some function in the BCR promoter: within it, the complement of the sequence 5′ CACGATGGTGGCCTC*T*GACACGA 3′ (binding site 1 in Figures 2B and 4) is protected by protein factors in both K562 and A498 extracts against DNase1 digestion (Figure 6c); interestingly, the sense strand of the 3′ repeat located in the 3′

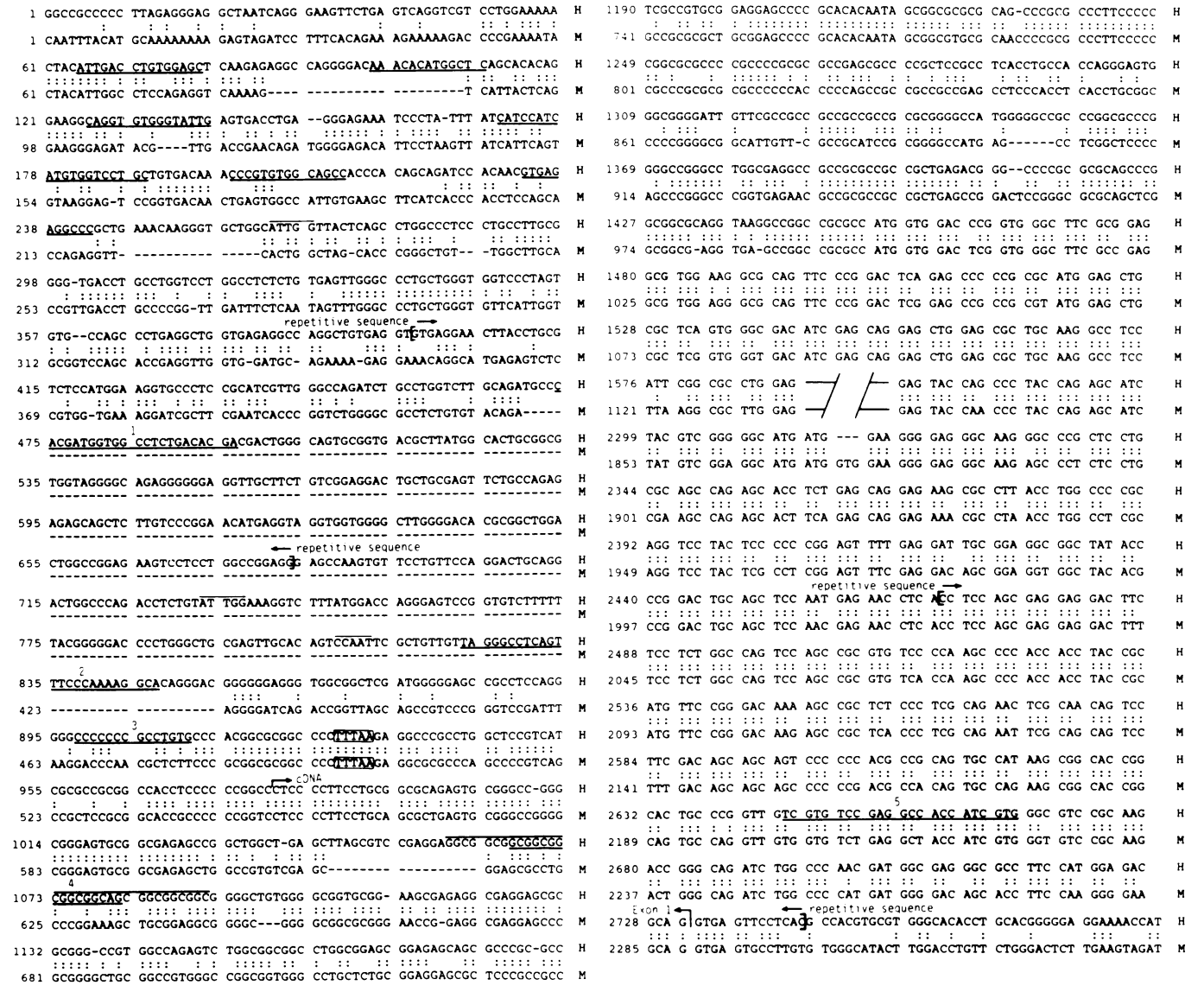

Figure 4. Nucleotide sequence of human and mouse BCR gene promoter region. Coding regions are indicated as triplets. The protein binding regions in the human DNA sequence are underlined and numbered as in Figure 2. The presumed TATA box in both human and mouse is boxed, and the start of the longest cDNA indicated with an arrow. The repeat sequences 5′ in the human BCR promoter and in the 3′ exon/splice donor region are bracketed. The 11 consecutive GGCs which contain protein binding site 4 are over-lined, as are CCAAT sequences in the human sequence.

coding and splice donor region of exon 1 (sequence: 5′ TCGTGTCCGAGGCCACCATCGTG 3′, binding site 5 in Figures 2 and 4) is also protected (not shown), although this sequence contains one basepair difference (underlined) compared to the 5′ repeat. In both areas, the digestion patterns are very similar; the protected areas are also similar except that in the 3′ repeat the protected area extends further to the 5′ side.

Three other sites were identified in this manner; site 3 and 4 are located immediately 5′ to and within the 5′ untranslated region of the mRNA (see Figure 2B). Site 3 is protected by K562 nuclear factors but not by A498 extracts (Figure 6b); it corresponds to an Sp1 site identified from the DNA sequence analysis (Figure 4), whereas site 4 contains a very GC-rich area consisting of 11 GGC's (with exception of the eighth, a AGC) in tandem; this region has the potential to form secondary structures (17). Very prominent protection was found by both K562 and A498 extracts of the sequence 5′ TAGGGCCTCAGTTTCCCAAAAGGCA 3′ corresponding to site 2 (Figure 6a). This sequence does not correspond to any previously identified factor-binding recognition sequence.

To investigate the activity of the promoter *in vivo*, different fragments from the 5′ region were inserted into human growth hormone (hGH) reporter gene vectors (see Figure 7). As recipient cell line, NIH 3T3 cells were used: this cell line is frequently used for promoter assays and high transfection efficiencies can be obtained. Moreover, detectable levels of mouse BCR mRNA are transcribed in this cell line; upon transfection of human BCR/ABL constructs under control of the human BCR promoter expression of P190 and P210 mRNA and protein can be easily detected (N. Heisterkamp, et al., unpublished results). After transfection, the production of hGH was determined using a radioimmunoassay. A 1.5 kb PstI/SmaI fragment encompassing the entire region under study had clear promoter activity; a 1.04 kb ApaI/SmaI fragment lacking sites the 5′ six protein binding sites had comparable activity, as did an 0.65 kb PstI/SmaI fragment also lacking site 1 (Figure 8). Thus, the removal of

```
2431 GGCTATACCC CGGACTGCAG CTCCAATGAG AACCTCACCT CCAGCGAGGA GGACTTCTCC
          :  :  ::: :  :   ::        ::: ::: :: :: :::: ::::::::::
 719 CCAGTCCTGC AGTCCTGGAA CAGGAACACT TGGCTC-CCT CCGGCCAGGA GGACTTCTCC

2491 TCTGGCCAGT CCAGCCGCGT GTCCCCAAGC CCCACCACCT ACCGCATGTT CCGGGACAAA
     ::::::: :::::::::: :::::::::: :::::::::: :::::::::: :::::::::
 663 ---GGCCAGT CCAGCCGCGT GTCCCCAAGC CCCACCACCT ACCTCATGTT CCGGGACAAG

2551 AGCCGCTCTC CCTCGCAGAA CTCGCAACAG TCCTTCGACA GCAGCAGTCC CC-------
     ::: ::::::  :: :::::: :::::: ::: :::: ::::: : ::::  :  ::
 603 AGCTGCTCTC TCTGGCAGAA CTCGCAGCAG TCCTCCGACA GAAGCAACCT CCCCCCTCTG

2603 ------CCAC GCCGCAGTGC CATAAGCGGC ACCGGCACTG CCCGGTTGTC GTGTCCGAGG
           :::: :::::::::: :::::::: : :::: ::::: ::: :: ::: ::::: ::::
 544 CCCCTACCAC GCCGCAGTGC CATAAGCGTC ACCG-CACTG CCCAGTCGTC GTGTCAGAGG

2657 CCACCATCGT GGGCGTCCGC AAGACCGGGC AGATCTGGCC CAACGATGGC GAGGGCGCCT
     :::::::::: :::: :: :: :::::: ::: :::::::::: ::::::::   ::::::::::
 484 CCACCATCGT GGGCATCTGC AAGACCAGGC AGATCTGGCC CAACGATGCG GAGGGCACCT
          exon1 ◄——► intron
2717 TCCATGGAGA CGCAGGTGAG TTCCTCACGC CACGT-GCGT GGGCACACCT GCACGGGGGA
     :::::::::: ::::::: :: :::::::: : : :   :: : :: : : :     :   :
 424 TCCATGGAGA CGCAGGTAAG TTCCTCACAC CTCACAGCCT GGCCTCTCAC CAGCCTCAGG
```

Figure 5. Comparison of 5′ and 3′ repeats of the BCR gene. The coding strand of the 3′ repeat and splice donor site is shown in the upper line; the antisense strand of the 5′ repeat is shown in a 3′-5′ orientation on the lower line. Protein-binding sequences in both fragments are underlined. The repeat sequence is defined as from basepair 2468−2744 in the 3′ region and from basepair 397−683 in the 5′ region; numbering of bases is as in Figure 4.

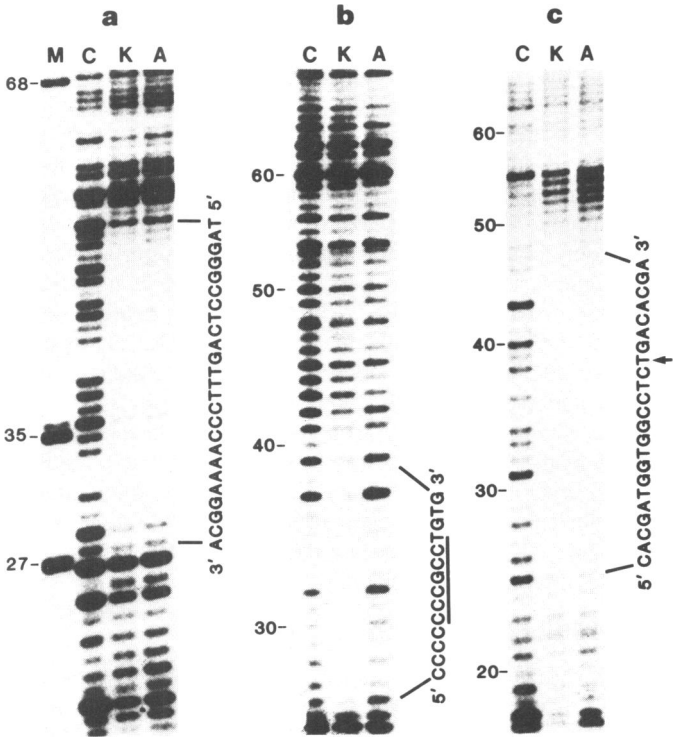

Figure 6. DNaseI footprinting analysis of the human BCR promoter. Panel a: 0.16 kb PstI/TaqI fragment (see Figure 2d) containing binding site **2**. The positive strand is labelled at the TaqI site. Panel b: 0.27 TaqI/BalI fragment containing the binding sequence **3**. The negative strand is labelled at the TaqI site. Panel c: 0.26 kb BglII/PstI fragment containing the binding sequence **1**. The negative strand is labelled at the BglII site. To the right side of each panel, the DNA sequence of the *positive* strand protected from DNase 1 digestion by nuclear extracts is shown for convenience, regardless of which strand was labelled. The Sp1 motif in Panel b is underlined. The number of basepairs from the labelled end is shown to the left of each panel; Lane M: pBR322/MspI marker. Lane C: partial digestion pattern by DNase 1 without nuclear proteins (control). Lane K: DNA digested in the presence of nuclear proteins isolated from K562 cells. Lane A: DNA digested in the presence of nuclear proteins from A498 cells. Note that the shown sequence represent the strongest protected nucleotides. In addition, in panel c an additional 7 nucleotides are hypersensitive bands.

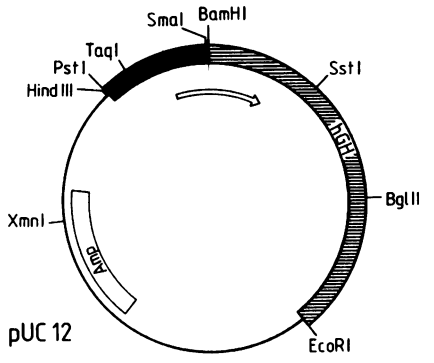

Figure 7. One of the constructs used in promoter activity measurement. A 0.65 kb PstI/SmaI fragment (solid bar) from the BCR promoter region (see Figure 8) was inserted into pSK (Stratagene); the fragment was excised with HindIII and BamHI and inserted into the promoterless expression vector pOGH (see Materials and Methods) digested with HindIII and BamHI. pOGH is a pUC12 (line)-based vector containing the entire genomic human growth hormone gene but lacking its promoter (hatched box) inserted into the polylinker through the restriction enzyme sites BamHI and EcoRI. The arrow indicates the predicted start site for transcription initiation. The restriction enzyme sites marked in this figure cut pOGH only once except PstI, TaqI and SmaI, which cut pOGH more than once.

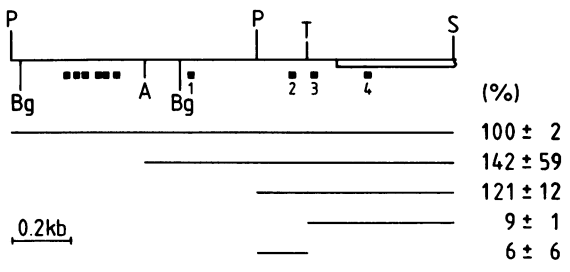

Figure 8. BCR gene promoter activity assay. DNA fragments indicated with lines under the restriction enzyme map were tested for promoter activity. For each fragment, the measurement was performed at least twice (2 to 5 times for different fragments). To the right of these fragments, numbers representing promoter activity relative to that of the longest fragment (1.5 kb PstI/SmaI), which was designated as 100%, are shown together with the standard deviation (in%). The open bar and the restriction enzyme sites are as in Figure 2; not all restriction enzyme sites are shown.

the six 5′ binding sites and as well as the 5′ repeat sequence does not seem to influence promoter activity in NIH 3T3 cells. However, a dramatic loss of activity was observed when site 2 was omitted from the construct by testing an 0.49 kb TaqI/SmaI fragment; site 2 on its own (on an 0.16 kb PstI/TaqI fragment) had no activity. Thus, the novel factor binding sequence identified with the DNaseI footprinting assays appears to be functionally active in this *in vivo* assay system.

Finally, we compared the entire human promoter region to that of the mouse, to determine whether sequences implicated in the regulation of the human BCR gene promoter had been evolutionary conserved. It is clear, that both human and mouse have in common a TTTAA sequence which, in human, is located 50 bp upstream of the probable mRNA initiation site (Figure 4). In addition, the human and mouse sequences surrounding the possible TATA box are highly conserved. Based on this conserved TTTAA box and surrounding areas, it seems that the site of initiation may be controlled by a similar mechanism in human and mouse. Protein binding site 4 in human, consisting of 11 consecutive GGCs, is lacking in mouse. The murine

promoter also does not contain binding site 3, the Sp1 site, and site 2 mentioned above in human.

Further to the 5′, additional regions of limited homology between man and mouse are found. This homology is most pronounced from nucleotides 50–469 in the human sequence. However, none of six 5′ binding sites are precisely conserved. Further studies will therefore be needed to determine whether they are functional *in vivo* in the mouse.

DISCUSSION

In the current study, we have identified and characterized the promoter of the human BCR gene. Initially, using expression in transgenic mice, promoter activity was located to a region of up to 3.5 kb 5′ to the gene. Subsequently, a region of 1.1 kb was analyzed in detail by sequencing, DNase1 footprinting, gel retardation and functional studies. The combined results point to several unique and interesting features of this promoter and localize the minimal region for promoter activity to a 270 bp sequence.

Seven sequences (the six 5′ binding sequences and binding site 1) were identified, which are protected by nuclear proteins against DNase1 digestion by K562 and A498 extracts. The fragments, on which these sites are located, showed clear binding to protein factor(s) in gel retardation assays using K562 extracts. Therefore, it seems likely that they have some function in the regulation of BCR gene expression in these cell types. However, in a functional assay in NIH 3T3 fibroblasts, these fragments had little or no influence on the promoter activity. Since the BCR gene is normally expressed at a low level in mouse fibroblasts, this could indicate, that the sites may be involved in tissue-specific regulation: these factors would then be absent from this cell type. Indeed, in comparison with initial data obtained from experiments to determine the activity of the human ABL 1B promoter in NIH 3T3 cells (Q-S Zhu et al., unpublished) using the same growth hormone assay system, the BCR promoter has a relatively low overall activity. Therefore, the possibility of tissue-specific binding sites in the BCR gene promoter will be examined in future studies.

Of these protein binding sites, site 1 is the most enigmatic. Presently, we can only speculate as to its function. Since the repeat sequence and the binding site within it is conserved almost perfectly within the 3′ end of exon 1 and adjoining splice donor, it may be possible that factors bind primarily to the 3′ sequence, site 5. Our footprinting experiment indicates, that a protein factor protects both strands of DNA and thus appears to bind to double-stranded DNA. In the 5′ area the repeat is situated 5′ to the probable transcription start site and will not be included in the mRNA. Therefore, there is no evidence for a role in splicing or processing of RNA.

The repeat at the 5′ end seems to be a relatively recent addition in evolution: mouse genomic DNA lacks this 5′ repeated sequence. This indicates, that the 5′ repeat was derived from the 3′ BCR exon 1 repeat; the fact, that the homology between 5′ and 3′ repeat terminates after the splice donor site suggests a mechanism involving an RNA intermediate for the creation of the 5′ repeat. However, this could not have occurred through a simple retrotransposon mechanism of exon 1, since only 274 bp of exon 1 are included in the 5′ repeat.

In contrast to the six 5′ binding sites and site 1, three sequences (sites 2–4) are located on a DNA fragment which showed no activity in a gel retardation assay. This fragment is relatively large and is extremely GC-rich; perhaps the formation of secondary structures under the conditions of the *in vitro* gel retardation assay precluded the binding of protein. These results notwithstanding, sites 2–4 were protected by nuclear extracts against DNase1 digestion and the minimal promoter is located in this very region.

Within this area, the human BCR promoter contains a protein-binding GC-rich region (site 4) consisting of 11 tandem GGC's (17). The region is located within the 5′ untranslated region of exon 1. It has been suggested, that direct repeats and in particular this repeat will form secondary structures in mRNA; this affects the translation of BCR mRNA *in vitro* but not *in vivo* (30). However, the fact that this sequence in the form of double stranded DNA was found to bind nuclear protein suggests that it may have some role in transcription.

An Sp1 site (site 3) is also located here, about 20 bp 5′ of the TTTAA sequence. This Sp1 site very likely contributes to the promoter activity. However, the Sp1 site together with the TTTAA box can only sustain a very low level of transcription and it is clear that an additional sequence, site 2, is needed for efficient expression. Of all features identified in the BCR promoter region, this site exhibits the strongest protein binding. To date, this specific sequence has not been previously identified and our results suggest that the protein factor binding this sequence may represent a new transcription factor. The fact that site 2 on its own has no promoter activity may indicate that the TATA box is required for its activity. Indeed, the proximity of site 2 to the TTTAA box suggests that the factor binding this sequence may interact directly with RNA polymerase.

Most of the features identified in the human BCR gene promoter seem to have been acquired relatively recently in evolution, since they were not found in its murine counterpart; interestingly, although the coding sequences of the first exons are closely conserved, the 5′ regions diverge significantly and this suggests that the two genes may be subject to different controls in man and mouse. Whether this has bearing on the genesis of Ph-positive leukemia, a chromosomal analogue of which to date has not been identified in mouse, remains a matter of speculation. The identification of site 3 as a site specific for human, which is bound by protein from K562 nuclear extracts but not from kidney fibroblasts, indicates it may be possible to identify myeloid-specific factors. Thus, studying the regulation of the normal human BCR gene promoter may enable us to also investigate aspects relevant to the production of BCR/ABL mRNAs implicated in leukemia.

ACKNOWLEDGEMENTS

We wish to thank Guido Jenster for help with the transgenic BCR mice. This work was supported by PHS grant CA 47456 to JG, and PHS grant CA 47073 and ACS CD-421 to NH. Q-SZ is supported by a CHLA fellowship.

REFERENCES

1. Nowell, P.C. and Hungerford, D.A. (1960) Science, **132**, 1497.
2. Champlin, R.E. and Golde, D.W. (1985) Blood, **65**, 1039–1047.
3. LeBeau, M.M. and Rowley, J.R. (1984) Cancer Surveys, **3**, 371–394.
4. Chessels, J.M., Janossy, G., Lawler, S.D. and Secker Walker, L.M. (1979) Br. J. Haematol., **41**, 25.
5. Heisterkamp, N., Stephenson, J.R., Groffen, J., Hansen, P.F., de Klein, A. and Grosveld, G. (1983) Nature, **306**, 239–242.
6. Heisterkamp, N., Stam, K., Groffen, J., de Klein, A. and Grosveld, G. (1985) Nature, **316**, 758–761.
7. Mes-Masson, A.M., McLaughlin, J., Daley, G.Q., Paskind, M. and Witte, O.N. (1986) Proc. Natl. Acad. Sci. USA, **83**, 9768–9772.
8. Shtivelman, E., Lifshitz, B., Gale, R.B. and Canaani, E. (1985) Nature, **315**, 550–554

9. Grosveld, G., Verwoerd, T., van Agthoven, T., de Klein, A., Ramanchandran, K.L., Heisterkamp, N., Stam, K. and Groffen, J. (1986) Mol. Cell. Biol., **6**, 607–616
10. Heisterkamp, N., Jenster, G., ten Hoeve, J., Zovich, D., Pattengale, P.K. and Groffen, J. (1990) Nature, **344**, 251–254.
11. Daley, G.Q., Van Etten, R.A. and Baltimore, D. (1990) Science, **247**, 824–830.
12. Konopka, J.B., Watanabe, S.M. and Witte, O.N. (1984) Cell, **37**, 1035–1042.
13. Kloetzer, J.B., Kurzrock, R., Smith, L., Talpaz, M., Spiller, M., Gutterman, J. and Arlinghaus, R. (1985) Virol., **140**, 230–238.
14. Clark, S.S., Mc Laughlin, J., Crist, W.M., Champlin, R. and Witte, O.N. (1987) Science, **235**, 85–88.
15. Kurzrock, R., Shtalrid, M., Romero, P., Kloetzer, W.S., Talpas, M., Trujillo, J.M., Blick, M., Beran, M. and Gutterman, J.U. (1987) Nature, **325**, 631–635.
16. Heisterkamp, N., Knoppel, E. and Groffen, J. (1988) Nucleic Acids Res., **16**, 10069–10081.
17. Hariharan, I.K. and Adams, J.M. (1987) The EMBO J., **6**, 115–119.
18. Lifshitz, B., Fainstein, E., Marcelle, C., Shtivelman, E., Amson, R., Gale, R.P. and Canaani, E. (1988) Oncogene, **2**, 113–117.
19. Collins, S., Coleman, H. and Groudine, M. (1987) Mol. Cell. Biol., **7**, 2870–2876.
20. Stam, K., Heisterkamp, N., Reynolds, F. H., Jr. and Groffen, J. (1987) Mol Cell. Biol. **7**, 1955–1960.
21. Campbell, M.L., Li, W. and Arlinghaus, R.B. (1990) Oncogene, in press.
22. Dhut, S., Dorey, E.L., Horton, M.A., Ganesan, T.S. and Young, B.D. (1988) Oncogene, **383**, 561–566.
23. Chen, C. and Okayama, H. (1987) Mol. Cell. Biol., **7**, 2745–2752.
24. Zhu, Q-S., Heisterkamp, N. and Groffen, J. (1990) Oncogene, in press.
25. Hogan, B., Constantini, F. and Lacy, E. (1986) Manipulating the mouse embryo. Cold Spring Harbor University Press, Cold Spring Harbor.
26. Chomcznski, P. and Sacchi, N. (1987) Anal. Biochem., **162**, 156–159.
27. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
28. Valerio, D., Duyvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkveus, Th. M., van der Voorn, L., van Ormondt, H. and van der Eb, A.J. (1985) The EMBO J., **4**, 437–443.
29. Arcot, S.S., Flemington, E.K. and Deiningen, P.L. (1989) J. Biol. Chem., **264**, 2343–2349.
30. Muller, A.J. and Witte, O.N. (1989) Mol. Cell. Biol., **9**, 5234–5238.