The EMBO Journal vol.6 no.1 pp.115–119, 1987

cDNA sequence for human *bcr*, the gene that translocates to the *abl* oncogene in chronic myeloid leukaemia

Iswar K.Hariharan and Jerry M.Adams

The Walter and Eliza Hall Institute of Medical Research, Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

Communicated by E.Klein

The hallmark of human chronic myeloid leukaemia is a 9;22 chromosome translocation that fuses most of the c-*abl* oncogene to the 5′ portion of the breakpoint cluster region (*bcr*) gene, such that a hybrid *bcr*–*abl* mRNA and polypeptide are generated. To clarify further the nature of this translocation, we have analysed the structure of normal human *bcr* mRNA by isolating large cDNA clones that collectively span the entire coding region and extend 2.6 kb upstream of those previously described. The 3150-bp nucleotide sequence reported here includes 534 bp of a GC-rich 5′ non-coding segment and indicates, in conjunction with published sequences, that the *bcr* polypeptide comprises 1271 amino acid residues. The predicted polypeptide is unrelated to serine or tyrosine kinases, or indeed to any previously published sequence; its structure provides no evidence of a transmembrane region. Since probes from throughout the 4.8-kb cloned region hybridized to both the 4.5 and 6.7 kb normal *bcr* transcripts, both RNAs contain most or all of that region.

Key words: oncogene activation/tyrosine kinase/myeloid leukaemia/9;22 chromosome translocation/*bcr*–*abl* oncogene

Introduction

Chronic myeloid leukaemia (CML) is a clonal disease resulting from neoplastic transformation of a pluripotent haemopoietic stem cell. The chronic phase (lasting 3–5 years) is dominated by an expansion of committed myeloid progenitors, leading to the accumulation of mature granulocytes in the bone marrow and peripheral blood. Most patients then develop a fatal acute leukaemia (blast crisis) characterized by the accumulation of immature myeloid or lymphoid blast cells (Champlin and Golde, 1985).

A striking feature of CML is that the leukaemic cells from >90% of patients display a shortened chromosome 22 — the Philadelphia chromosome — due to a reciprocal translocation involving chromosomes 9 and 22 (Rowley, 1973). The c-*abl* proto-oncogene from chromosome 9 is thereby fused to the breakpoint cluster region (*bcr*) on chromosome 22 to generate a hybrid *bcr*–*abl* gene (Shtivelman *et al.*, 1985; Grosveld *et al.*, 1986). Transcription of *bcr*–*abl* presumably initiates at the normal *bcr* promoter(s) and spans the chromosomal breakpoint; *bcr* sequences are then spliced in frame to those from the second exon of c-*abl* to form an 8.7-kb *bcr*–*abl* mRNA (Collins *et al.*, 1984; Shtivelman *et al.*, 1985). Hence the N-terminal portion of the resulting chimeric *bcr*–*abl* polypeptide (p210) derives from *bcr*, while the C-terminal segment comprises most of the c-*abl* sequence, including its entire tyrosine kinase domain. In contrast to the p145 encoded by c-*abl*, the p210 of *bcr*–*abl* has readily demonstrable tyrosine kinase activity *in vitro* (Konopka *et al.*, 1984; Kloetzer *et al.*, 1985).

A critical step towards determining the role of *bcr*–*abl* in the development of CML is to clarify the structure and function of the *bcr* gene. The normal *bcr* transcripts are known to be 4.5 and 6.7 kb long and the sequence of the 3′ terminal portion of the *bcr* coding region has been published (Heisterkamp *et al.*, 1985). Here we describe the cloning and characterization of cDNAs that collectively span the entire *bcr* coding region. Their nucleotide sequence indicates that the *bcr* polypeptide comprises 1271 amino acid residues. Probes from the cloned region have also allowed us to analyse the relationship of the two *bcr* transcripts.

Results

To identify a good source of *bcr* mRNA, the level in a number of human cell lines was assessed using an oligonucleotide probe based on the sequence of Heisterkamp *et al.* (1985). As the Burkitt's lymphoma cell line JOY (Bernard *et al.*, 1983) had a significant level, poly(A)$^+$ RNA from this line was used to construct a cDNA library. Approximately 2×10^5 recombinant plaques were screened with the oligonucleotide probe, and 30 hybridizing clones were detected. The four longest *bcr* inserts are shown in Figure 1. Their restriction maps indicate that they collectively span 4.8 kb. The longest extends 2.6 kb upstream from the 2.2 kb region characterized previously. The nucleotide sequence of the 5′ 3.1 kb was determined (see Figure 1). This sequence and its predicted polypeptide product are shown in Figure 2.

The coding sequence

The sequence reveals a single long open reading frame which commences with the ATG at position 535. The surrounding sequence (GCCATGG) is well matched to the Kozak consensus (PuNNATGG) for the initiation of translation (Kozak, 1984). Upstream from this codon there are two other ATG codons (at positions 80 and 428), but they are followed by terminators 14 and 18 codons downstream respectively. Since at least the ATG sequence at position 428 fits the Kozak rule and that at 80 is a near fit, it is conceivable that the mRNA generates a peptide of 18 or 14 amino acids in addition to the *bcr* polypeptide. The long open reading frame continues through the rest of the region sequenced here and that published (Heisterkamp *et al.*, 1985) for the 3′ segment (see Figure 1). Taken together, the sequences define an open reading frame of 3813 nucleotides that encodes a polypeptide of 1271 amino acid residues.

The predicted mol. wt of the *bcr* polypeptide is 142 645 daltons. It is a relatively hydrophilic protein with an excess of basic residues (202 Lys and Arg residues and 24 His residues) over acidic ones (175 Asp and Glu residues). Hydropathy analysis, using the method of Kyte and Doolittle (1982), revealed none of the regions of marked hydrophobicity that are characteristic of transmembrane domains. Analysis of the amino acid sequence, using the DIAGON program (Staden, 1982a), revealed no internal repeats.

To determine whether the *bcr* cDNA is similar to any previously characterized sequence, the entire 4.8-kb *bcr* sequence

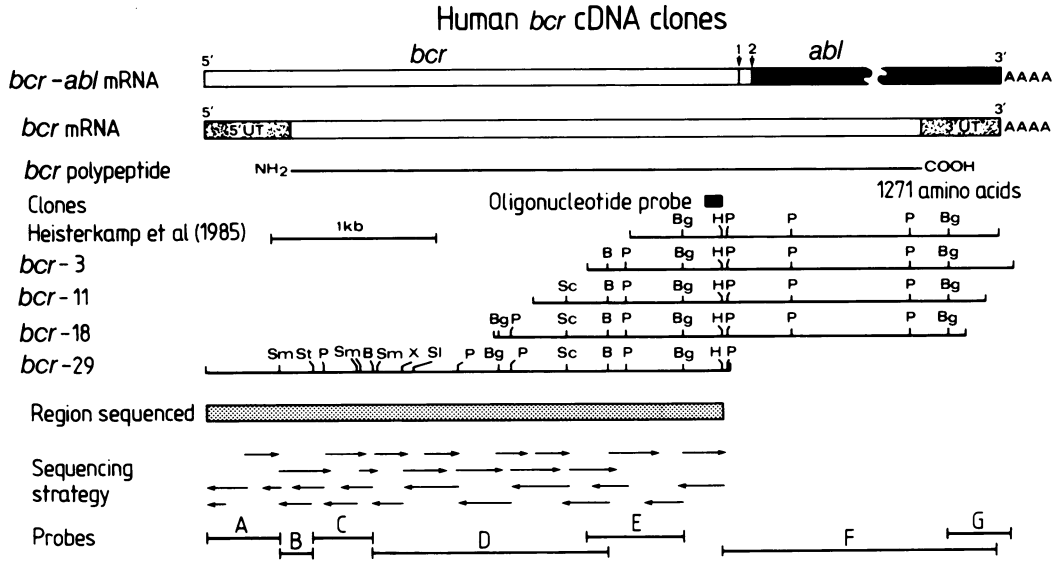

Fig. 1. Human *bcr* cDNA clones. The indicated oligonucleotide probe was used to isolate the *bcr* cDNA clones. Restriction maps were constructed for the enzymes *Bam*HI (B), *Bgl*II (Bg), *Hin*dIII (H), *Pst*I (P), *Sac*I (Sc), *Sal*I (Sl), *Sma*I (Sm), *Stu*I (St) and *Xho*I (X). The stippled region was sequenced by the strategy indicated by the arrows. Probes used for Northern hybridizations are also shown. The inferred structures of the *bcr* and *bcr–abl* mRNAs and the *bcr* polypeptide are illustrated. In the *bcr–abl* mRNA, the sequences between the two arrows (1 and 2) are included in the hybrid message only when the break occurs between the *bcr* exons designated 3 and 4 (Heisterkamp *et al.*, 1985).

was compared with the EMBL and Genbank databases. The computer-assisted search utilized the Los Alamos/NIH SEQF program, which uses the algorithm of Wilbur and Lipman (1983). The analysis failed to detect any sequence with notable homology. Moreover, the entire amino acid sequence (1271 residues) was compared with the contents of the NBRF, Doolittle and PGtrans databases using the Dayhoff SEARCH program (Dayhoff *et al.*, 1983) and a mutation matrix. Again no significant homology was evident.

There is some recent evidence that the *bcr* protein may either have serine kinase activity or be the substrate for an associated serine kinase (Davis *et al.*, 1985; Ben Neriah *et al.*, 1986). Hence detailed comparisons were made between the *bcr* amino acid sequence and that of several proteins that phosphorylate serine or threonine residues. The *bcr* polypeptide did not display any homology with cAMP-dependent kinase, cGMP-dependent kinase, phosphorylase kinase, protein kinase C, v-*mos*, v-*raf* or v-*mil*. Hence the structural data do not reveal a serine kinase domain. Moreover, although the *bcr–abl* protein is in many ways analogous to the viral *gag–abl* protein (see Discussion), no homology was found between the primary sequence of *bcr* and that of the *gag* region.

The 5′ untranslated region

The cDNA clone has a 534-nucleotide long 5′ untranslated region. This region begins with a run of 46 A residues, which might represent a cloning artefact. The rest of the 5′ untranslated region has a high content of G and C nucleotides (81%). This composition predisposes to extensive secondary structure, as became evident during the sequencing of this region. Indeed 10 repeats of GGC occur between positions 139 and 178 and five repeats of the complementary CCG between positions 406 and 420. These regions can pair to form a stable secondary structure having an 18-nucleotide stem and a 228-nucleotide loop with ΔG of -65.9 kcal mol^{-1}, as calculated by the method of Tinoco *et al.* (1973).

Structure of bcr transcripts

It is not established how the normal 4.5- and 6.7-kb *bcr* transcripts are related to each other, or to the 8.7-kb *bcr–abl* mRNA. To investigate their relationship, we have hybridized various *bcr* probes to Northern blots of RNA from the JOY Burkitt's lymphoma, which contains the normal transcripts, and from the CML-derived cell line K562, which contains predominantly the 8.7-kb hybrid mRNA (Grosveld *et al.*, 1986). The probes used span the entire cloned *bcr* region (Figure 1); they include a probe from the 5′ untranslated region (A), one that spans the translation initiation codon (B), several from the coding region (C, D, E, F) and one from the 3′ untranslated region (G). Significantly, each of these probes hybridized to both normal *bcr* transcripts. Some of these results are shown in Figure 3. The data suggest that most or all of the cloned region, which spans 4.8 kb, is present in both normal *bcr* transcripts. All the probes that include sequences upstream from the K562 breakpoint (A–E) hybridized to the 8.7-kb *bcr–abl* transcript in K562. Hence the chimeric RNA as well as both *bcr* transcripts include that region. As would be expected, the chimeric RNA was also detected by a v-*abl* probe (data not shown).

Probes B and C also hybridize to a 1.2-kb transcript, the nature of which is not known. Furthermore, at moderate stringency, the probe from the 5′ end of the *bcr* clone (probe A) also hybridized strongly to rRNA, presumably due to the high GC content of that region. This problem was largely overcome by performing the hybridization and wash at very high stringency (see Materials and methods).

Discussion

Formation of the *bcr–abl* gene is thought to be a crucial event in the pathogenesis of CML. To delineate the structure of the normal *bcr* mRNA, we have characterized cDNA clones that cover the entire coding region. These clones reveal a single long open reading frame that extends for 3813 nucleotides and encodes a polypeptide of 1271 amino acids. As the predicted *bcr* amino acid sequence does not share notable homology with the serine or tyrosine kinases, or indeed with any other protein characterized previously, the normal function of *bcr* remains to be determined.

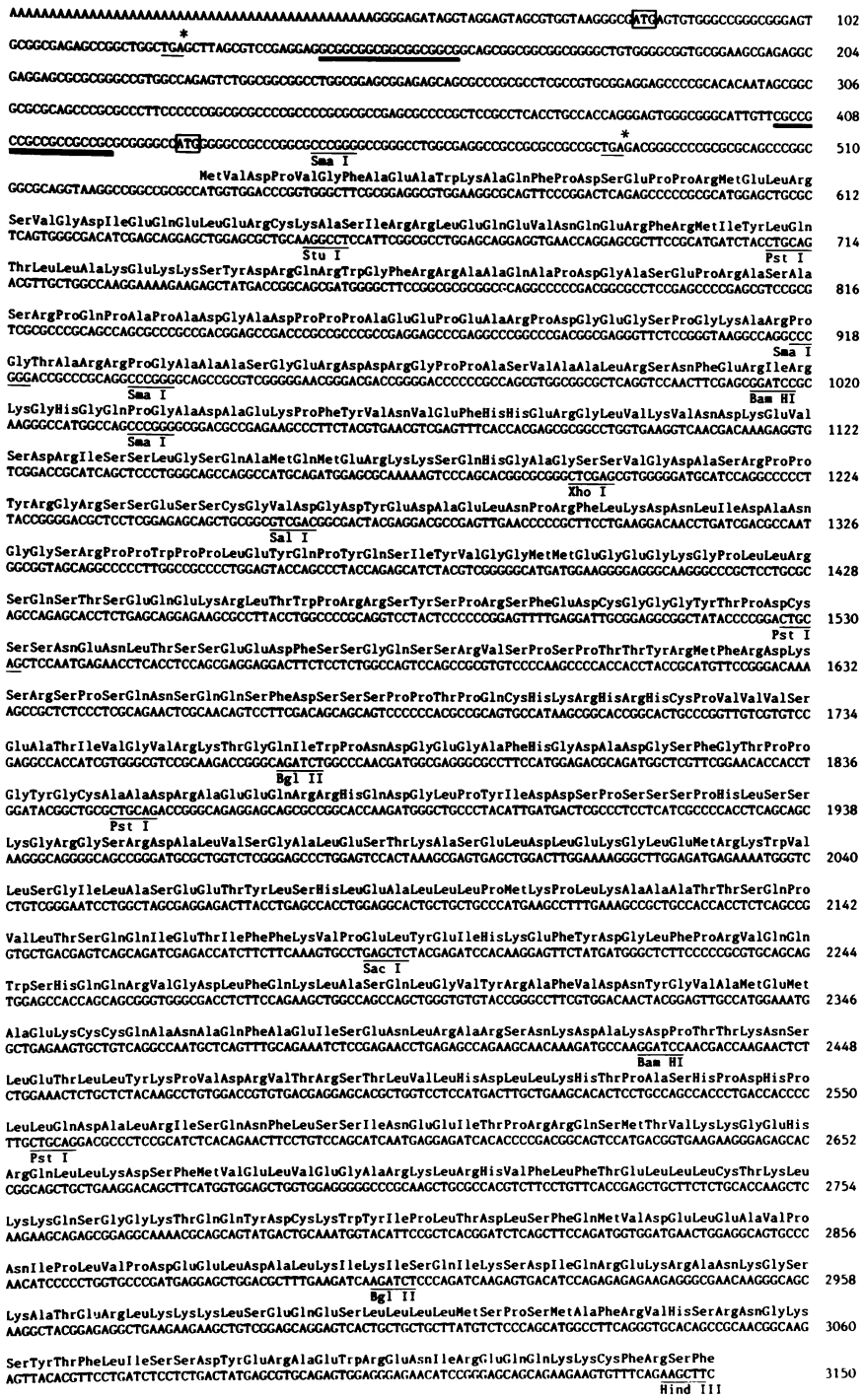

Fig. 2. Nucleotide sequence of *bcr*. The nucleotide and predicted amino acid sequences of the 5′ 3150 nucleotides of the *bcr* cDNA clones are shown. The sequence from position 2579 through to the poly(A) tail has been published previously (Heisterkamp *et al.*, 1985). Restriction sites are indicated. The two ATG codons (boxed) that occur 5′ to the long open reading frame are followed by in-frame stop codons (asterisks). The two complementary regions that can form a stable stem are indicated (bold underline).

The relationship of the 4.5- and 6.7-kb normal *bcr* transcripts remains unclear, but we have found that every probe from the cloned 4.8-kb region, including one from the 5′ untranslated region, hybridized to both transcripts. Hence most or all of the entire cloned region is probably contained within both transcripts. The additional sequences in the 6.7 kb transcript might occur at its 5′ end, internally, or at its 3′ end. However, the existence of a single *bcr*−*abl* transcript in cells bearing the translocation suggests that the extra sequences in the 6.7-kb transcript are not part of the hybrid transcript and hence lie 3′ to the breakpoint. The 6.7-kb *bcr* transcript may be generated by a more distal polyadenylation signal or an alternative splice in the 3′ region.

The *bcr* 5′ untranslated region has a high GC content and is likely to display considerable secondary structure, including a prominent stem and loop structure. Such structures may be involved in post-transcriptional regulation, since increasing the secondary structure of a 5′ non-coding region has been shown to reduce the translational efficiency of eukaryotic mRNA, both *in vitro* and *in vivo* (Pelletier and Sonenberg, 1985). The *bcr* mRNA also has two short open reading frames (14 and 18

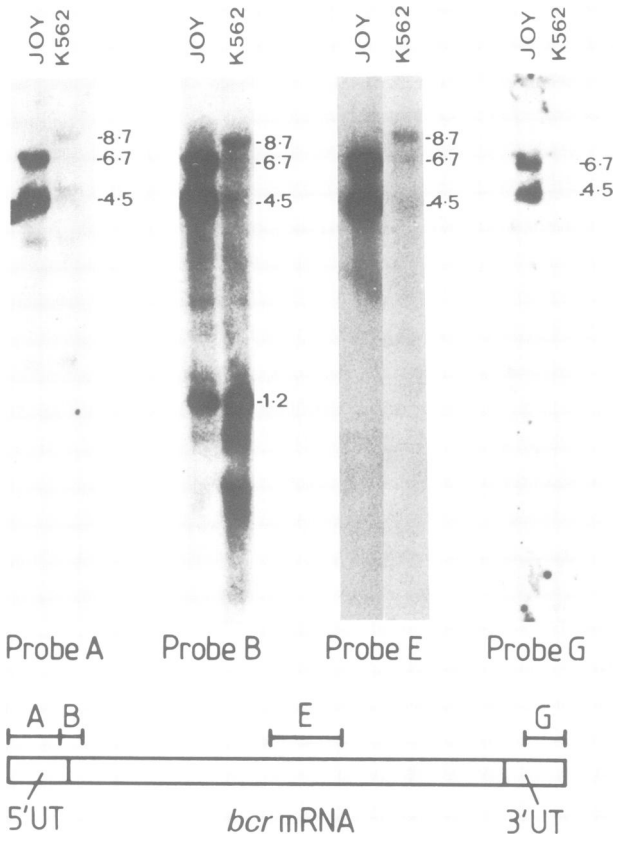

Fig. 3. Analysis of *bcr* transcripts. 2−10 μg of poly(A)$^+$ RNA was fractionated electrophoretically on a formaldehyde gel, transferred to nitrocellulose and hybridized to radioactive probes as described in Materials and methods. Probe A extends from the 5′ end of clone *bcr*-29 (see Figure 1) to the *Sma*I site; probe B is a *Sma*I−*Stu*I fragment from *bcr*-29; probe E extends from the 5′ end of *bcr*-3 to the *Bgl*II site; probe G extends from the 3′ *Blg*II site in *bcr*-3 to its 3′ end.

codons) in the 5′ non-coding region. These open reading frames may modulate the level of expression of the *bcr* and *bcr*−*abl* proteins at the translational level. A precedent exists in the case of the yeast GCN4 gene, which has four short open reading frames in the 5′ untranslated region; GCN4 is subject to translational regulation and deletion of the short open reading frames leads to greatly increased translation of the message (Mueller and Hinnebusch, 1986; Fink, 1986).

The manner in which *bcr*-coded sequences activate the c-*abl* gene remains unclear. Since *bcr* breakpoints in the t(9;22) occur within either of two adjacent introns (Heisterkamp *et al.*, 1985), *bcr* encodes either 902 or 927 amino acids in the chimeric protein. Thus the large *bcr* moiety may perturb the normal function of the c-*abl* polypeptide, as is suggested by the enhanced autophosphorylation of the p210 kinase (Konopka *et al.*, 1984). Since the translocation also subjects c-*abl* to the control of the *bcr* promoter elements, it is also possible that *bcr*−*abl* is expressed in a cell type that does not normally express c-*abl*. This possibility seems less likely, because c-*abl* is expressed in diverse haemopoietic cells, including CML cells (Westin *et al.*, 1982; Gale and Canaani, 1984). The oncogenic effect of *bcr*−*abl*, however, may be largely confined to the stem cells within a CML cell population; whether that minor fraction of cells expresses c-*abl* is not known.

Some clues to the function of the p210 *bcr*−*abl* polypeptide can be gleaned by a comparison with v-*abl*, the p160 transforming protein of the Abelson murine leukaemia virus, in which the N-terminal portion of the viral *gag* gene has replaced that of c-*abl*. In addition to their structural similarity, the p210 and p160 proteins have similar *in vitro* tyrosine kinase activities (Konopka and Witte, 1985). Moreover, in contrast to the normal c-*abl* gene product (p145), both p210 and p160 are phosphorylated on tyrosine *in vivo* (Konopka and Witte, 1985). An intriguing possibility is that replacement of the N-terminal segment of c-*abl* by either *bcr*- or *gag*-coded sequences alters its substrate specificity.

Although *bcr*−*abl* and v-*abl* have not yet been studied in the same cells, there are hints that their biological properties differ. The Abelson virus can transform bone marrow cells to yield tumorigenic cell lines (Whitlock and Witte, 1985) and can render factor-dependent cell lines independent of exogenous factor (Cook *et al.*, 1985; Pierce *et al.*, 1985). In contrast, the development of CML appears to be a slow multi-step process, and CML cells that make the *bcr*−*abl* protein require exogenous factor for growth (Metcalf, 1985) and almost never form permanent cell lines. However, the greater susceptibility of murine cells to transformation may account for some of these differences.

Association with the plasma membrane appears to be crucial to the transforming ability of some of the tyrosine kinases (Hunter and Cooper, 1985). In v-*abl*, as in v-*src*, myristylation of the N-terminal glycine residue leads to association with the inner surface of the cell membrane (Kamps *et al.*, 1985; Mathey-Prevot and Baltimore, 1985). Disruption of myristylation removes this association and prevents cellular transformation. The *bcr* sequence, however, provides no indication that it would associate with the membrane; hydropathy analysis did not reveal a likely transmembrane domain, and the sequence does not predict an N-terminal glycine, which is thought to be necessary for myristylation. Hence membrane attachment may not be required for the function of *bcr*−*abl*. Alternatively, the chimeric polypeptide may associate with the membrane by interaction with a membrane protein.

The *bcr* gene is expressed in a wide variety of cell types, including fibroblasts (Heisterkamp *et al.*, 1985), diverse lymphoid and myeloid cell lines (our unpublished data), and HeLa cells (Ben Neriah *et al.*, 1986), which are of epithelial origin. Hence the *bcr*−*abl* gene presumably would be transcribed in many cell types. It is therefore remarkable that the 9;22 translocation is restricted to CML and some cases of acute lymphoblastic leukaemia (ALL). If, as seems likely, the translocation is a random event that could occur in any cell type, the marked specificity of this translocation for CML and ALL suggests that the substrates of the p210 kinase that mediate transformation are restricted to these lineages. Alternatively, production of the p210 protein may be incompatible with the viability of other cell types, in a manner analogous to the lethal effects of v-*abl* expression in certain fibroblast cells (Whitlock and Witte, 1985).

Following submission of this manuscript, we received a preprint of a paper by Mes-Masson *et al.* (1986) which describes the cloning and sequence determination of the hybrid *bcr*−*abl* cDNA from the cell line K562. That sequence differs from ours at seven nucleotides within the coding region. A review of our sequence data at those positions leaves us confident of our assignments. The most notable differences are one nucleotide insertions/deletions that generate a frameshift between 1992 and 2026, changing 12 amino acids. Two other amino acid differences occur at position 1982 A/T (Glu/Val) and positions 2527 and 2528 CC/GG (Pro/Gly). Interestingly, the sequences at the extreme 5′ end of the clones are completely different (positions 1−81).

Materials and methods

Preparation of a bcr oligonucleotide probe

Two oligonucleotides (60-mers) were synthesized using the Applied Biosystems Model 380A DNA Synthesizer. Their sequences were based on that published by Heisterkamp *et al.* (1985) (see Figure 1) and were derived from the two complementary strands with a 3′ overlap of 12 nucleotides. After heating for 10 min at 65°C, equimolar amounts of the 60-mers were annealed in 1 M NaCl. Approximately 400 ng of the annealed oligonucleotides were subsequently extended to yield double-stranded ^{32}P-labelled 108-mers in 50 mM Tris−HCl (pH 7.5), 50 mM $MgCl_2$, 150 mM NaCl, 1 mM dGTP, 1 mM dTTP, 150 μCi of both [α-^{32}P]dATP and [α-^{32}P]dCTP and 5 U of the Klenow fragment of DNA polymerase. After 15 min at 37°C, unlabelled dATP and dCTP were added to a final concentration of 1 mM and the reaction continued for a further 15 min. The probe was separated from the unincorporated dNTPs by ethanol precipitation after addition of ammonium acetate to 1 M. Hybridizations with oligonucleotide probes were carried out in 40% formamide/5 × SSC (SSC is 0.15 M NaCl, 15 mM sodium citrate pH 7.0) at 37°C and washing in 2 × SSC at 60°C. Probe was included at 5 × 10^6 c.p.m./ml for Northern hybridizations and 5 × 10^5/ml for plaque hybridizations.

Construction of the cDNA library

Poly(A)$^+$ total cellular RNA was isolated from ~10^9 cells by a method involving digestion with proteinase K and oligo(dT) cellulose chromatography (Gonda *et al.*, 1982). 5 μg of the RNA was used to prepare a cDNA library. First strand synthesis was performed at 37°C for 90 min in 50 mM Tris−HCl (pH 7.5), 75 mM KCl, 10 mM dithiothreitol (DTT), 3 mM $MgCl_2$, 50 μg/ml actinomycin D, 0.5 mM each of dATP, dCTP, dGTP and dTTP, 10 μg/ml oligo(dT) and 1000 U of M-MLV reverse transcriptase (BRL) (Kotewicz *et al.*, 1985). After extraction with phenol and precipitation with ethanol, the second strand was synthesized in 50 mM Tris−HCl pH 7.6, 10 mM $MgCl_2$, 10 mM DTT, 0.6 mM of each of the dNTPs, using 6 U RNase H (BRL) and 50 U of DNA polymerase I (Boehringer Mannheim) (Gubler and Hoffman, 1983). The ends of the cDNA were rendered blunt using T4 polymerase. After methylation and ligation of *Eco*RI linkers, the cDNA was size fractionated by electrophoresis on a low-melting point agarose gel. cDNA >1 kb in length was cloned into λgt10 (Huynh *et al.*, 1985). Recombinants that hybridized to the oligonucleotide probe were plaque purified, and their inserts were subcloned into the pUC19 plasmid for further analysis.

Sequencing

DNA sequencing was performed using the chain termination method (Sanger *et al.*, 1977). Specific oligonucleotide primers were used in regions where convenient restriction sites were unavailable. In high GC regions, reverse transcriptase and *Escherichia coli* single-stranded binding protein were used (Mason *et al.*, 1986). The sequence was analysed using the computer programs of Staden (1982b) as modified by Dr A.Kyne.

Northern hybridizations

Poly(A)$^+$ RNA (2−10 μg) was heated at 65°C for 5 min in buffered 2 M formaldehyde/50% formamide and fractionated electrophoretically on a 1.2% agarose/2 M formaldehyde gel and blotted onto nitrocellulose (Thomas, 1980). Probes were labelled by nick translation and included in the hybridization at 5 × 10^6 c.p.m./ml. Filters were routinely hybridized in 5 × SSC, 50% formamide at 42°C and washed in 0.2 × SSC at 65°C. When probe A (from the 5′ end of the *bcr* mRNA) was used, the conditions for hybridization (5 × SSC/50% formamide/50°C) and washing (0.1 × SSC, 75°C) were more stringent.

Acknowledgements

We are indebted to Dr Suzanne Cory for stimulating discussions, Dr A.Kyne for expert help with computing, Drs R.Simpson and D.Scanlon for oligonucleotides and Leonie Gibson for excellent technical assistance. I.H. is an NH and MRC postgraduate medical scholar. This work was funded in part by the National Health and Medical Research Council (Canberra), the US National Cancer Institute (CA 12421), the American Heart Association and the Drakensberg Trust.

References

Ben-Neriah,Y., Daley,G.Q., Mes-Masson,A.-M., Witte,O.N. and Baltimore,D. (1986) *Science*, **233**, 212−214.

Bernard,O., Cory,S., Gerondakis,S., Webb,E. and Adams,J.M. (1983) *EMBO J.*, **2**, 2375−2383.

Champlin,R.E. and Golde,D.W. (1985) *Blood*, **65**, 1039−1047.

Collins,S.J., Kubonishi,I., Miyoshi,I. and Groudine,M.T. (1984) *Science*, **225**, 7274.

Cook,W.D., Metcalf,D., Nicola,N.A., Burgess,A.W. and Walker,F. (1985) *Cell*, **41**, 677−683.

Davis,R.L., Konopka,J.B. and Witte,O.N. (1985) *Mol. Cell. Biol.*, **5**, 204−213.

Dayhoff,M.O., Barker,W.C. and Hunt,L.T. (1983) *Methods Enzymol.*, **91**, 525−545.

Fink,G.R. (1986) *Cell*, **45**, 155−156.

Gale,R.P. and Canaani,E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5648−5652.

Gonda,T., Sheiness,D. and Bishop,J.M. (1982) *Mol. Cell. Biol.*, **2**, 617−624.

Grosveld,G., Verwoerd,T., van Agthoven,T., de Klein,A., Ramachandran,K.L., Heisterkamp,N., Stam,K. and Groffen,J. (1986) *Mol. Cell. Biol.*, **6**, 607−616.

Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263−269.

Heisterkamp,N., Stam,K., Groffen,J., de Klein,A. and Grosveld,G. (1985) *Nature*, **315**, 758−761.

Hunter,T. and Cooper,J.A. (1985) *Annu. Rev. Biochem.*, **54**, 897−930.

Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D. (ed.), *DNA Cloning: A Practical Approach*. IRL Press, Oxford, pp. 49−78.

Kamps,M.P., Buss,J.E. and Sefton,B.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4625−4628.

Kloetzer,W., Kurzrock,R., Smith,L., Talpaz,M., Spiller,M., Gutterman,J. and Arlinghaus,R. (1985) *Virology*, **140**, 230−238.

Konopka,J.B. and Witte,O.N. (1985) *Biochim. Biophys. Acta*, **823**, 1−17.

Konopka,J.B., Watanabe,S.M. and Witte,O.N. (1984) *Cell*, **37**, 1035−1042.

Kotewicz,M.L., D'Alessio,J.M., Driftmier,K.M., Blodgett,K.P. and Gerard,G.F. (1985) *Gene*, **35**, 249−258.

Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857−872.

Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105−132.

Mason,A.J., Niall,H.D. and Seeburg,P.H. (1986) *Biochem. Biophys. Res. Commun.*, **135**, 957−964.

Mes-Masson,A.-M., McLaughlin,J., Daley,G.Q., Paskind,M. and Witte,O.N. (1986) *Proc. Natl. Acad. Sci. USA*, in press.

Mathey-Prevot,B. and Baltimore,D. (1985) *EMBO J.*, **4**, 1769−1774.

Metcalf,D. (1985) *Science*, **229**, 16−22.

Mueller,P.P. and Hinnebusch,A.G. (1986) *Cell*, **45**, 201−207.

Pelletier,J. and Sonenberg,N. (1985) *Cell*, **40**, 515−526.

Pierce,J.H., Di Fiore,P.O., Aaronson,S.A., Potter,M., Pumphrey,J., Scott,A. and Ihle,J. (1985) *Cell*, **41**, 685−693.

Rowley,J.D. (1973) *Nature*, **243**, 290−293.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5467.

Shtivelman,E., Lifshitz,B., Gale,R.P. and Canaani,E. (1985) *Nature*, **315**, 550−554.

Staden,R. (1982a) *Nucleic Acids Res.*, **10**, 2951−2961.

Staden,R. (1982b) *Nucleic Acids Res.*, **10**, 4731−4751.

Thomas,P. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201−5205.

Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers, D.M. and Gralla,J. (1973) *Nature New Biol.*, **246**, 40−41.

Westin,E.H., Wong-Staal,F., Gelmann,E.P., Dalla Favera,R., Papas,T.S., Lautenberger,J.A., Eva,A., Reddy,E.P., Tronick,S.R., Aaronson,S.A. and Gallo,R.C. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2490−2494.

Whitlock,C.A. and Witte,O.N. (1985) *Adv. Immunol.*, **37**, 73−98.

Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 726−730.

Received on 7 October 1986