Molecular Biology Reports **17**: 17–20, 1993.

Nucleotide sequence of *bup*, an upstream gene in the *bmi*-1 proviral insertion locus

Y. Haupt, G. Barri & J.M. Adams
The Walter and Eliza Hall Institute of Medical Research, PO Box Royal Melbourne Hospital, Victoria 3050, Australia

Received 2 April 1992; accepted 28 April 1992

Key words: Moloney virus, bup gene, insertion locus

Abstract

The ability of Moloney murine leukemia virus to accelerate lymphomagenesis in Eμ-*myc* transgenic mice is frequently associated with proviral integration within a locus denoted *bmi*-1. This locus contains not only the *bmi*-1 gene implicated as a collaborator with *myc* in lymphomagenesis but also just upstream an unknown gene denoted *bup*. The nucleotide sequence reported here for *bup* cDNA and flanking genomic sequences reveals that this widely expressed gene comprises at least 7 exons and potentially encodes a polypeptide of 195 amino acid residues. Computer searches with this polypeptide sequence revealed no close homolog in the databases, nor any conserved motifs, and it is unrelated to the product of the *mel*-13 gene, which lies just upstream from the *bmi*-1 homolog *mel*-18.

Insertional mutagenesis with retroviruses has proven to be a powerful approach for identifying genetic loci involved in tumorigenesis [1, 2]. The rationale for this approach is that retroviruses that lack an oncogene, such as Moloney murine leukemia virus (Mo-MLV), promote tumor development primarily by inadvertent integration within or near a cellular oncogene, altering its structure or augmenting its transcription. The application of insertional mutagenesis to tumor-prone trangenic mice [3–6] is of particular interest, because it provides a means to identify genes that can collaborate with the transgene in tumorigenesis.

In a search for cellular genes that could collaborate with the *myc* gene in lymphomagenesis, Mo-MLV was used to infect Eμ-*myc* transgenic mice, which carry a transgene expressed throughout the B lymphoid lineage [7]. Development of pre-B and B lymphomas was markedly accelerated by Mo-MLV infection and half the resulting lymphomas contained proviral inserts in a novel locus denoted *bmi*-1 [4, 5].

The *bmi*-1 locus (Fig. 1) proved to be complex. We identified two genes, one denoted *bmi*-1 and another just upstream designated *bup* (*bmi*-1 upstream) [4]. The expression of the *bmi*-1 gene, which encodes a novel type of zinc finger protein, was markedly stimulated by proviral insertion (arrows in Fig. 1) and *bmi*-1 presumably represents the gene that collaborates with *myc* in lymphomagenesis, although no direct evidence for cooporation has yet been reported. The *bup* gene is widely expressed in normal tissues and hematopoietic cell lines as a 1.3 kb mRNA [4]. The level of *bup* expression was elevated in some

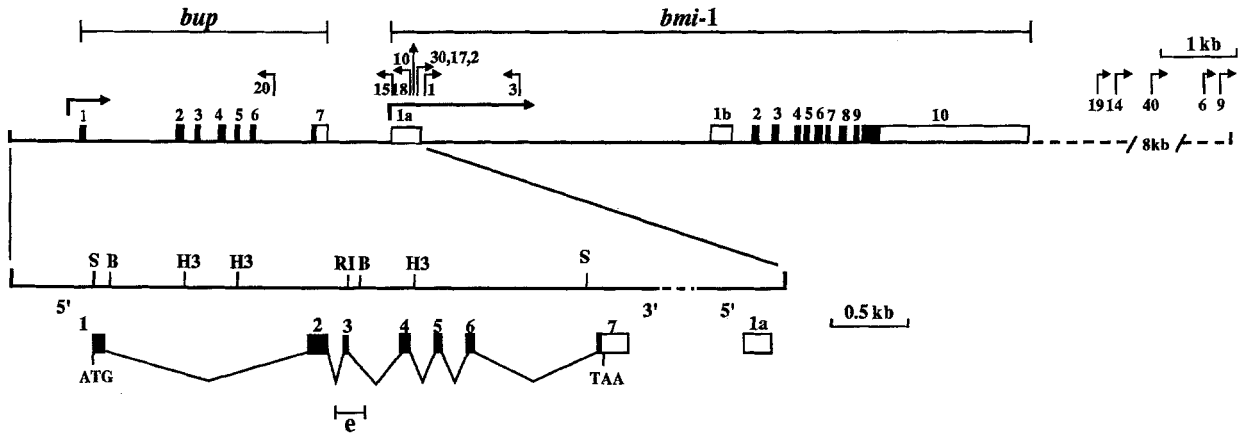

Fig. 1. Structure of the *bmi*-1 locus and the *bup* gene. On the map of the *bmi*-1 locus (top), a solid line delineates the cloned region and a broken line the region inferred from genomic blots. Arrows denote the position and (where known) orientation of proviral inserts. Boxes denote exons and filled areas indicate the coding region. On the enlarged map of the *bup* gene, the sequenced region is indicated by a solid line, probe e is marked and restriction sites are shown for *Hind* III (H3), *Eco* RI (RI), *Sac* I (S) and *Bam* HI (B).

tumors bearing Moloney inserts in the *bmi*-1 locus, but overall the elevation did not correlate well with viral insertion or with the level of *bmi*-1 expression [4]. Nevertheless, the close proximity of *bup* and *bmi*-1 and their shared transcriptional orientation suggest that they might have a shared function or partly coordinated expression. Therefore we have determined the *bup* nucleotide sequence.

Fig. 1 shows that the *bup* gene comprises 7 exons. Its poly A addition site lies only about 0.7 kb from the start of *bmi*-1 transcription. The cloning of *bup* cDNAs using fragment e (Fig. 1) as a probe was described previously [4]. Here we present the nucleotide sequence of *bup* (2020 bp). The sequence includes the longest cDNA sequence we obtained (from nucleotide 16 to 882) and flanking 5′ and 3′ genomic sequences (Fig. 2). A consensus polyadenylation signal resides 23 bp 5′ to the poly A tract (nucleotide 882).

The *bup* sequence contains a single long open reading frame. Translation probably commences with the ATG at nucleotide 1, since its context fits well with the consensus initiation sequence [8] and there is an in-frame stop codon upstream (bolded in Fig. 2). The intervening region contains two non-ATG potential initiation codons, CTG at −349 and GTG at −244 (underlined in Fig. 2), but their context is likely to reduce their efficiency as initiation codons [8]. Assuming that the ATG initiates translation, the predicted *bup* polypeptide is 195 residues long. A search of the available data bases with the FASTA program [9] revealed no close homolog and did not highlight any structural or functional domains. Thus, at present little can be said about possible functions of the widely expressed *bup* gene.

The shared transcriptional orientation of *bup* and *bmi*-1 raise the possibility that transcripts initiated at a *bup* promoter might occasionally read through the normal *bup* polyadenylation signal and generate a *bup-bmi*-1 precursor RNA from which novel *bup-bmi*-1 mRNA might be spliced. Obviously such RNA potentially could encode *bup-bmi*-1 fusion polypeptides. A hint that such transcripts may occasionally arise is that we previously found one cDNA clone involving sequences from both genes; it involves a fusion from *bup* exon 6 to *bmi*-1 exon 2 and places the *bmi*-1 sequences out of frame. However, an in-frame fusion protein could be generated by other hypothetical splicing events, for example from *bup* exon 3 to *bmi*-1 exon 2, or by other cryptic splicing events. In any case, any fusion transcripts are likely to be rare. A screen

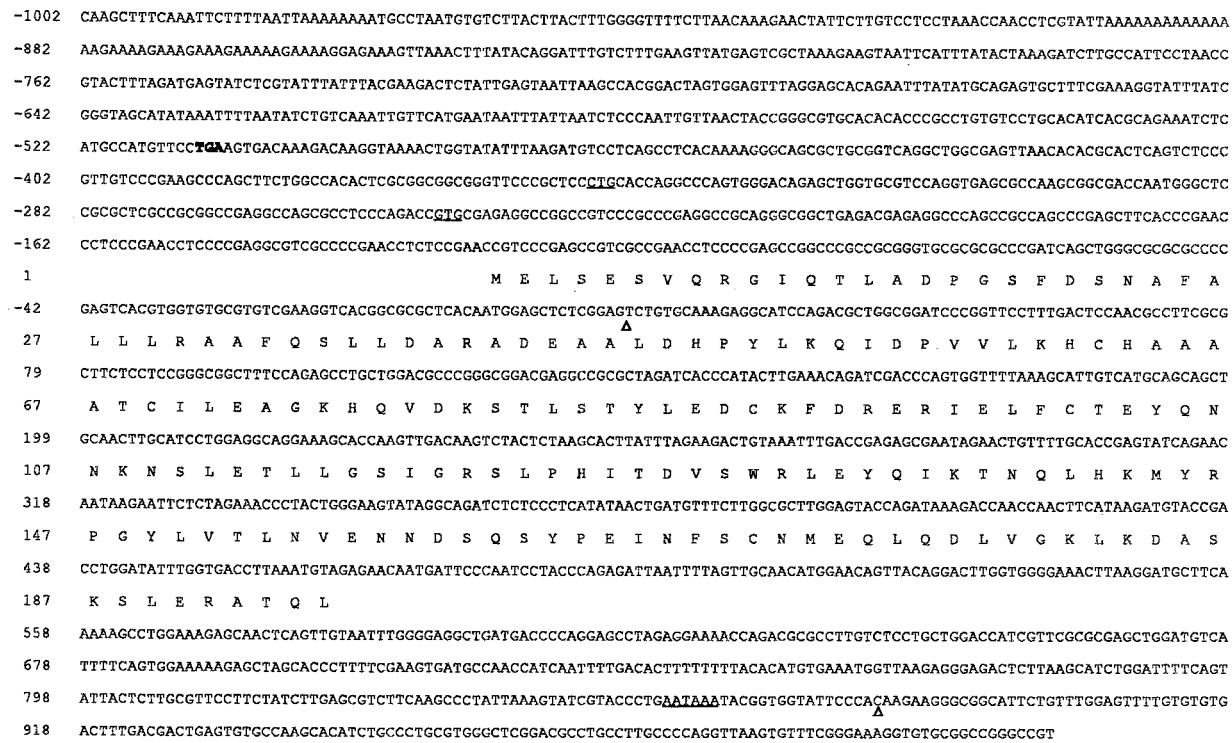

Fig. 2. Nucleotide and predicted polypeptide sequences of *bup*. The beginning and end of the cDNA sequence is marked by triangles. In the 5′ untranslated region, the inframe stop codon is bolded, and two potential alternative initiation codons are underlined. The polyadenylation signal in the 3′ untranslated region is also underlined.

of 60 cDNA clones from a plasmacytoma cDNA library revealed no fusion clones [4]. Moreover, northern hybridization experiments with mRNA from various tissues and cell lines has not revealed any prominent species that is likely to contain both *bup* and *bmi*-1 sequences.

Recently it has been noted [10] that *bmi*-1 is closely related to another murine gene, *mel*-18, which is highly expressed in mouse melanoma and other tumor cells [11]. Within 2 kb upstream of the *mel*-18 gene [11] a second small gene was found, which was denoted *mel*-13 (Professor M. Taniguchi, personal communication). The striking similarity of the genomic organization between *bup*/*bmi*-1 and *mel*-13/*mel*-18 suggested that these loci were derived from a common ancestor. We therefore compared the *bup* sequence with that of *mel*-13 (kindly provided by Professor Taniguchi). This analysis revealed no obvious similarity between *bup* and *mel*-13. Hence these two genes probably are not related, despite the close similarity between their respective neighbors *bmi*-1 and *mel*-18.

Acknowledgements

We thank Professor Taniguchi for exchanging unpublished data. This work was supported by the NH&MRC (Canberra) and the U.S. National Cancer Institute (CA43450).

References

1. van Lohuizen M & Berns A (1990) Biochim. Biophys. Acta. 1032: 213–235
2. Peters G (1990) Cell Growth & Diff. 1: 503–510
3. van Lohuizen M, Verbeek S, Krimpenfort P, Domen J, Saris C, Rudaszkiewcz T & Berns A (1989) Cell 56: 673–682
4. Haupt Y, Alexander WS, Barri G, Klinken SP & Adams JM (1991) Cell 65: 753–763
5. van Lohuizen MP, Verbeek S, Scheijen B, Weintjenz E,

van der Gulden H & Berns A (1991) Cell 65: 737–752
6. Haupt Y, Harris AW & Adams JM (1992) Oncogene 7: 981–986
7. Adams JM, Harris AW, Pinkert CA, Corcoran LM, Alexander WS, Cory S, Palmiter RD & Brinster RL (1985) Nature 318: 533–538
8. Kozak M (1986) Cell 44: 283–292
9. Pearson WR & Lipman DJ (1988) Proc. Natl. Acad. Sci. USA 85: 2444–2448
10. Goebl G (1990) Cell 66: 623
11. Tagawa M, Sakamoto T, Shigemoto K, Matsubara H, Tamura Y, Ito T, Nakamura I, Okitsu A, Imai K & Taniguchi M (1990) J. Biol. Chem. 265: 20021–20026