

Bg transposon transcription from both strands: two products similar to NFI and SET domain proteins may be involved in transcription and chromatin modulation

--Koterniak, VV

Bg transposon sequence and probable Bg-encoded proteins.

Analysis of the Bg transposon sequence suggests that this mobile element encodes several proteins (designated as PPBg1-PPBg3), described previously (MNL 79:32-35; MNL 80, submitted). The analysis also shows certain regions of Bg sequence may form Z-DNA and that Bg-encoded proteins have Z-DNA binding properties, indicating a possible autoregulation of this transposon at the transcriptional level (MNL 80, submitted). Structure of all above-mentioned proteins was deduced from the strand of Bg transposon containing the two longest ORFs. However, some mobile elements (e.g., the maize *MuDR* autonomous element) are transcribed from both strands (Hershberger et al., Genetics 140:1087-1098, 1995). Further analysis of the other strand for Bg indicates this mobile element may encode 2 further transcription and chromatin modulation proteins.

An 87 amino acid protein encoded by the second strand of Bg element is similar to Nuclear Factor I family of transcriptional

a)

```

PPBg4 1 MRQQLQWS-CAAWRQPHLP-----WRR-----TCWFWLSPSPSIS---- 35
NFI 109 MEEDVDTSPGGDYTSPNSPTSSSRNWTEDIEGGISSPVKKTMDKSPFNSPSPQDS PRLS 169

PPBg4 36 -----CSRGLA-TPRGI-----PQIDLHVNEVAVSWS-----LP--EPSSTLI- 47
NFI 170 SFTQHHRPVI AVHSGIARSPHTSALHFIAIPLPQTASTYFPHTAIRYPPHLNPODPLKD 230

PPBg4 48 IWEMELLRRVADGD----G 87
NFI 231 IVSLACDIAATQQPGPSWYLG 250
    
```

b)

```

NFI -SPTSPSIS
PolII YSPTSPS---
PPBg4 29 -SPSPSIS- 35
    
```

Figure 1. (a) Alignment of PPBg4 with the transcription factor NF I of *Mus musculus* (GenBank accession AAK21332.1). The T23-S35 sequence similar to the P-4 peptide (Fujii et al., 2003) is underlined. (b) Similarity of the S29-S35 region of PPBg78 with the SPTSPSYSP motif of NFI transcription factor (NFI) and with repeat YSPTSPS of the RNA polymerase II (PolII). Identical residues are shown in a black background, similar ones are in a grey background.

```

PPBg5 1 MSCTGLPCNVWIQSNELSTCLLIIVGPIICNNLNDILHPNLIINNHLNNT 48
CAG25109.1 3427 NNNMNMNMIIMNMIIMNMMNMIIMN-NEMNNMNMNI 3459

PPBg5 49 IINKFCILNNTS CIYRLVKKHPSTAIYHEINNAH HGRT 86
CAG25109.1 3460 INNNNIFNNDVSNNVDMQHKSDQICIFNS-NNIH 3492
    
```

Figure 2. Similarity of PPBg5 with a SET-domain protein of *Plasmodium falciparum* (GenBank accession CAG25109.1). Identical residues are shown in a black background, similar ones are in a grey background.

regulators. The ORF of the second strand of Bg, from position 724 to 460 (positions for both strands according to the first strand, the sequence of GenBank accession X56877.1), encodes an 87 amino acid protein, designated hereafter as PPBg4 (Fig. 1a). It is unusually rich in tryptophan (7 residues) and has several PS dipeptide residues.

This protein shows significant similarity with the transcription regulators of the nuclear factor I (NFI) family (Fig. 1a), using CLUSTALW analysis at the European Bioinformatics Institute (<http://www.ebi.ac.uk>) using default parameters. In the human genome, the promoter sites of NFI and Z-DNA forming regions (ZDRs) are near transcriptional start sites (Champ et al., Nucl. Acids Res. 32:6501-6510, 2004). In the case of the PPBg4 gene, possible ZDRs are located just downstream of the PPBg4 gene at positions 120 and 402. In addition, a perfect canonical NFI binding site (5'-TGG(N)₆GCCAA-3'; Zorbas et al., J. Biol. Chem. 267:8478-8484) is present at position 1775 of the Bg sequence; i.e., at -1051 bp upstream on the opposite strand in relation to the translation start site of PPBg4. The SP-rich stretch S29-S35 of PPBg4 (SPSPSTS, Fig. 1b) is similar to the SPTSPSYSP motif contained in the NFI transcriptional activation domain (Wendler et al., Nucl.