A distinct bipartite motif is required for the localization of inhibitory κ B-like ($\text{I}\kappa\text{BL}$) protein to nuclear speckles

Jennifer I. SEMPLE, Stephanie E. BROWN, Christopher M. SANDERSON and R. Duncan CAMPBELL¹

Functional Genomics Group, MRC UK HGMP Resource Centre, Hinxton, Cambridge CB10 1SB, U.K.

The inhibitory κ B ($\text{I}\kappa\text{B}$)-like ($\text{I}\kappa\text{BL}$) gene is located within the Class III region of the MHC on human chromosome 6. Previous analysis of the predicted amino acid sequence of the human $\text{I}\kappa\text{BL}$ protein revealed three putative functional domains; 2–3 ankyrin repeat sequences, which are similar to the second and third ankyrin repeats of the nuclear factor κ B (NF- κ B) protein; three PEST sequence motifs (a sequence that is rich in proline, serine, aspartic acid and threonine residues), which are also found in other $\text{I}\kappa\text{B}$ family members; and a C-terminal leucine zipper-like motif. In the present study we have identified a novel bipartite motif, which is required for nuclear localization of the $\text{I}\kappa\text{BL}$ protein. Analyses of $\text{I}\kappa\text{BL}$ -specific transcripts revealed the existence of a widely expressed spliced variant form of $\text{I}\kappa\text{BL}$ ($\text{I}\kappa\text{BLsv1}$), which lacks the amino acid sequence GELEDE-WQEVMMGRFE (where single-letter amino-acid notation has been used). Interestingly, translation of $\text{I}\kappa\text{BL}$ mRNA *in vivo* was

found to initiate predominantly from the second available methionine, thereby resulting in the disruption of the predicted N-terminal PEST sequence. Also, transient expression of T7 epitope-tagged $\text{I}\kappa\text{BL}$ and $\text{I}\kappa\text{BLsv1}$ proteins in mammalian cells showed that both proteins were targeted to the nucleus, where they accumulate in nuclear speckles. To define the protein domains required for nuclear import and subnuclear localization, a complementary set of deletion mutants and enhanced green fluorescent protein- $\text{I}\kappa\text{BL}$ domain fusions were expressed in mammalian cells. Data from these experiments show that a combination of the ankyrin-repeat region and an adjacent arginine-rich sequence are necessary and sufficient for both nuclear import and speckle localization.

Key words: ankyrin, $\text{I}\kappa\text{B}$, MHC, nuclear localization signal, splicing.

INTRODUCTION

The human inhibitory κ B ($\text{I}\kappa\text{B}$)-like ($\text{I}\kappa\text{BL}$) gene is located at the telomeric end of the MHC Class III region on chromosome 6 [1]. Recent studies have shown potential genetic association between $\text{I}\kappa\text{BL}$ and various autoimmune diseases. In particular, a 70 kb region telomeric of the tumour necrosis factor (TNF) genes, which contains the $\text{I}\kappa\text{BL}$ gene, was found to be associated with rheumatoid arthritis [2]. Also, a non-synonymous polymorphism in the coding sequence of $\text{I}\kappa\text{BL}$ has been associated with a propensity to develop extensive and severe ulcerative colitis [3]. In order to better understand the possible pathological implications of genetic variations occurring within the $\text{I}\kappa\text{BL}$ gene, it is essential that we develop a far more comprehensive understanding of $\text{I}\kappa\text{BL}$ function, in particular, which protein domains and amino acid motifs are required for different aspects of $\text{I}\kappa\text{BL}$ function.

Previous analysis of the predicted amino acid sequence of the $\text{I}\kappa\text{BL}$ protein revealed 2–3 ankyrin repeat sequences that most closely resemble the second and third ankyrin repeats of p100 nuclear factor κ B (NF- κ B). This sequence similarity led to the suggestion that $\text{I}\kappa\text{BL}$ may be a member of the $\text{I}\kappa\text{B}$ family of proteins [1]. Members of this family (for review see [4]) contain multiple ankyrin repeats which serve as protein–protein interaction domains, and are necessary for binding to the transcription factor NF- κ B [5]. In addition to ankyrin repeat regions, most $\text{I}\kappa\text{B}$ proteins also contain a C-terminal acidic sequence rich in proline, serine, aspartic acid and threonine residues (i.e. a PEST sequence), which can target these proteins to the 26 S proteasome for degradation [6]. Similar acidic PEST sequences

are also present at the N-terminal end of $\text{I}\kappa\text{BL}$ and downstream of the ankyrin repeats [1].

The $\text{I}\kappa\text{B}$ proteins bind to the Rel homology domains of NF- κ B dimers via their ankyrin repeat regions [7]. As a result of $\text{I}\kappa\text{B}$ binding, the nuclear localization signal (NLS) contained within the NF- κ B binding site is masked and NF- κ B is maintained in an inactive state within the cytoplasm [8]. In response to a variety of stimuli the $\text{I}\kappa\text{B}$ proteins are phosphorylated and subsequently degraded [9]. NF- κ B is then free to enter the nucleus and activate transcription of a number of genes, including the $\text{I}\kappa\text{B}$ genes [10]. The newly synthesized $\text{I}\kappa\text{B}$ proteins are transported into the nucleus where they bind once more to NF- κ B, and this complex is exported from the nucleus to the cytoplasm [11]. $\text{I}\kappa\text{B}$ proteins do not contain an NLS; instead, the nuclear import of $\text{I}\kappa\text{B}\alpha$ has been shown to depend on its ankyrin repeats [12]. However, although ankyrin repeat sequences from other proteins can also function as NLSs, this is not a universal property of all ankyrin repeats [12]. As such, the precise mechanism by which ankyrin domains mediate nuclear import remains unclear.

Despite the similarities that exist between $\text{I}\kappa\text{BL}$ and members of the $\text{I}\kappa\text{B}$ family, there are also a number of potentially important differences. First, $\text{I}\kappa\text{BL}$ contains only 2–3 ankyrin repeats, whereas ‘classical’ $\text{I}\kappa\text{B}$ proteins contain 6–7 ankyrin repeats [5]. Secondly, at the C-terminus of $\text{I}\kappa\text{BL}$ there are three leucine residues and one valine residue that occur with a periodicity of every seven amino acids, and as such may constitute a type of leucine zipper domain [1]. No such domain has yet been described in other members of the $\text{I}\kappa\text{B}$ family of proteins.

Abbreviations used: EGFP, enhanced green fluorescent protein; EST, expressed sequence tag; FBS, fetal bovine serum; $\text{I}\kappa\text{B}$, inhibitory κ B; $\text{I}\kappa\text{BL}$, $\text{I}\kappa\text{B}$ -like; $\text{I}\kappa\text{BLsv1}$, spliced variant form of $\text{I}\kappa\text{BL}$; NF- κ B, nuclear factor κ B; NLS, nuclear localization signal; RT, reverse transcriptase; TNF, tumour necrosis factor; UTR, untranslated region.

¹ To whom correspondence should be addressed (e-mail rcampbel@hgmp.mrc.ac.uk).

In order to establish the functional significance of these similarities and differences between the $\text{I}\kappa\text{BL}$ and $\text{I}\kappa\text{B}$ proteins we have analysed $\text{I}\kappa\text{BL}$ mRNA expression, the site of *in vivo* translational initiation and the protein domains required for correct subcellular targeting of the $\text{I}\kappa\text{BL}$ protein.

EXPERIMENTAL

Tissue culture

The human cell lines U937, Raji, K562, Jurkat and Molt 4 were maintained in suspension culture in RPMI 1640 medium supplemented with 10% (v/v) fetal bovine serum (FBS), 100 units/ml penicillin and 100 $\mu\text{g}/\text{ml}$ streptomycin. The adherent cell lines Cos-7 (monkey), HeLa (human) and Raw264 (mouse), were maintained in Dulbecco's modified Eagle's medium containing 10% (v/v) FBS, 100 units/ml penicillin and 100 $\mu\text{g}/\text{ml}$ streptomycin.

RNA isolation and reverse transcriptase (RT)-PCR

Approx. 10^8 cells were used to prepare total cellular RNA using the *RNAce* Total Pure kit (Bioline, London, U.K.). cDNA was synthesized using a Reverse Transcription System (Promega). Two rounds of nested PCR were carried out for $\text{I}\kappa\text{BL}$. The PCR reaction contained 0.2 mM dNTPs, 1.5 mM magnesium chloride, PCR buffer, 2.5 units of *Taq* polymerase and distilled water up to a total volume of 25 μl . The first round reaction also contained 0.05 vol. of the cDNA synthesis reaction and 0.2 μM F1 (5'-ATGAGTAACCCCTCCCCCAG-3') and 0.2 μM R1 (5'-CTTGAGGGCCTCTGCATGG CG-3') primers. The second round of PCR contained 2.5 μl of the first round reaction as a template and used primers F2 (5'-GCCATGGGAATAAAGAATAAGG-3') and R2 (5'-CAGGATCCCTCTGCTTCTCG-3'). The PCR reactions were performed on a MJ-Tetrad PCR machine; components were first incubated for 2 min at 95 °C, before being amplified for 30 cycles consisting of 1 min at 94 °C, 1 min at 60 °C and 1 min at 72 °C. In order to improve visualization of the two splice variants of $\text{I}\kappa\text{BL}$ [$\text{I}\kappa\text{BL}$ and the spliced variant form of $\text{I}\kappa\text{BL}$ ($\text{I}\kappa\text{BLsv1}$)] and minimize heteroduplex formation, 5 μl of each sample was heated to 95 °C for 3 min before being gradually cooled to 26 °C, prior to loading on to a 3% (w/v) agarose gel.

Expression constructs

All expression constructs that were amplified by PCR were created using 1.25 units of the proof-reading enzyme Pfu DNA polymerase (Stratagene), 0.2 mM dNTPs, PCR buffer, 50 ng of template DNA and distilled water up to a total volume of 25 μl . The PCR reactions were heated to 95 °C for 5 min before being taken through 15 cycles of 94 °C for 1 min, 64 °C for 1 min and 72 °C for 4 min. All PCR-generated constructs were bi-directionally sequenced to confirm nucleotide fidelity prior to expression.

The predicted full-length coding sequence of $\text{I}\kappa\text{BL}$ [nt 69–1211; numbering according to NM_005007 (GenBank® accession number)] was amplified from the $\text{I}\kappa\text{BL}$ -pGEX-2T vector [13], and was inserted into the *EcoRI*/*NheI* sites of the T7-pBluescript KS⁺ vector [14]. The $\text{I}\kappa\text{BL}$ -T7-pcDNA3 construct was created by ligating an *EcoRI*-*NotI* fragment of the $\text{I}\kappa\text{BL}$ -T7-pBluescript KS⁺ vector, encoding an in-frame fusion of $\text{I}\kappa\text{BL}$ with the T7 epitope tag at the C-terminal end of the protein into the *EcoRI*/*NotI* sites of the mammalian expression vector pcDNA3 (Invitrogen). The T7- $\text{I}\kappa\text{BL}$ -pcDNA3 construct, coding for an in-frame fusion of $\text{I}\kappa\text{BL}$ with the T7 epitope tag at the N-terminal

end of the protein, was created by ligating an *EcoRI* fragment of $\text{I}\kappa\text{BL}$ -pGEX-2T into *EcoRI*-digested pBluescript KS⁺. The insert was cut out of this vector with *EcoRV* and *XbaI*, end-filled, and then ligated into T7-pBluescript KS⁺ that had been digested with *BamHI* and then end-filled prior to ligation. This T7- $\text{I}\kappa\text{BL}$ -pBluescript KS⁺ construct was digested with *HindIII*/*NotI* and ligated into *HindIII*/*NotI*-digested pcDNA3, giving the T7- $\text{I}\kappa\text{BL}$ -pcDNA3 construct. T7- $\text{I}\kappa\text{BLsv1}$ -pcDNA3 was created by replacing a *PflMI*-*BamHI* fragment of T7- $\text{I}\kappa\text{BL}$ -pBluescript, corresponding to nt 488–736 of $\text{I}\kappa\text{BL}$, with a *PflMI*-*BamHI* fragment of $\text{I}\kappa\text{BLsv1}$ (lacking nt 580–624 of the full-length transcript) that was amplified by nested RT-PCR from cDNA derived from Jurkat cells. The sequence coding for the T7 tag fused to $\text{I}\kappa\text{BLsv1}$ (T7- $\text{I}\kappa\text{BLsv1}$) was excised with *HindIII*/*NotI*, and ligated into *HindIII*/*NotI*-digested pcDNA3.

UTR1- $\text{I}\kappa\text{BL}$ (nt 15–1211; where UTR stands for untranslated region), ATG1- $\text{I}\kappa\text{BL}$ (nt 59–1211), UTR2- $\text{I}\kappa\text{BL}$ (nt 83–1211) and ATG2- $\text{I}\kappa\text{BL}$ (nt 129–1211) expression constructs for the analysis of the translation initiation site of $\text{I}\kappa\text{BL}$ were created by PCR, and ligated into *EcoRI*/*NheI*-digested $\text{I}\kappa\text{BL}$ -T7-pcDNA3. $\text{I}\kappa\text{BL}$ - ΔANK , $\text{I}\kappa\text{BL}$ - ΔNLS and $\text{I}\kappa\text{BL}$ - ΔLeuZ constructs coding for the T7 epitope tag fused to $\text{I}\kappa\text{BL}$ lacking the ankyrin repeats (nt 189–455), the NLS (nt 153–188), or the leucine zipper (nt 1125–1211) respectively were created using 'splicing by overlap extension' PCR [15], and were then ligated into the *HindIII* and *NotI* sites of pcDNA3. $\text{I}\kappa\text{BL}$ (nt 69–1211), NLS (nt 69–189), ankyrin (nt 186–479) and NLS-ankyrin (nt 69–479) constructs coding for domains of $\text{I}\kappa\text{BL}$ were created by PCR using T7- $\text{I}\kappa\text{BL}$ -pcDNA3 as a template, and the resulting fragments were ligated into the *EcoRI* site of the pEGFP-C3 vector (ClonTech) to form N-terminal fusions with enhanced green fluorescent protein (EGFP).

Transient transfections and Western-blot analysis

Transient transfection of Cos-7 or HeLa cells was carried out by the DEAE dextran method as described previously [14]. After 2 days the cells were harvested in Laemmli's SDS sample buffer [16]. The samples were sonicated and then boiled before being resolved by gel electrophoresis.

Nuclear-cytoplasmic fractionation of cells was performed with a detergent-based method as described by Spector et al. [17]. The nuclear and cytoplasmic fractions were made up to the same final volume in Laemmli's SDS sample buffer, sonicated briefly and then boiled for 5 min. pcDNA3-G11Z [18], pcDNA3-G9-Tag-1 [14] and pcDNA3G15Tag [19] were used as markers for different subcellular fractions.

The samples were resolved on SDS/12.5% (w/v) polyacrylamide gels [16], and the proteins were transferred on to Hybond^{TM} ECL® nitrocellulose membranes (Amersham Pharmacia Biotech). After blocking the membrane for 30 min in PBS containing 5% (w/v) skimmed milk powder and 0.1% Tween 20, the blot was first incubated for 1 h in a 1:10000 dilution of mouse anti-(T7 tag) monoclonal antibody (Novagen) in blocking solution, washed, and then incubated for 1 h in a 1:3000 dilution of peroxidase-conjugated sheep anti-mouse IgG antibody (Sigma) in blocking solution. Bound antibodies were detected by ECL® (Amersham Pharmacia Biotech).

Immunofluorescence

Transient transfection of Cos-7 or HeLa cells was carried out as described previously [14]. Two days following transfection, cells were fixed for 15 min at room temperature (20–25 °C) with 4% (v/v) formaldehyde in PBS. Following fixation the cells were permeabilized for 5 min with PBS containing 0.1% (v/v) Triton

X-100. The cover slips were blocked for 10 min in PBS containing 10% (v/v) FBS, and then incubated for 1 h with a 1:3000 dilution of mouse anti-(T7 tag) antibody (Novagen) in blocking solution, followed by a 1 h incubation in a 1:300 dilution of FITC-conjugated goat anti-mouse IgG antibody (Sigma) in blocking solution. Fluorescence was analysed using a Nikon Eclipse E800 microscope in conjunction with a Radiance Plus confocal scanning system (Bio-Rad).

Co-localization experiments were carried out using the blocking solution described by Dostie et al. [20]. The cover slips were incubated for 1 h with a mixture of a 1:3000 dilution of mouse anti-(T7 tag) antibody and a 1:500 dilution of human anti-Sm serum (ANA human serum #5, Lot 960005L; Centers for Disease Control, Atlanta, GA, U.S.A.) in blocking solution. This was followed by a 1 h incubation with a mixture of a 1:100 dilution of Cy3-conjugated AffinityPure goat anti-mouse IgG F(ab')₂ fragment (Jackson ImmunoResearch Laboratories) and a 1:100 dilution of FITC-conjugated AffinityPure donkey anti-human IgG (H+L) antibody (Jackson ImmunoResearch Laboratories) in blocking solution.

Cells transfected with constructs in the pEGFP vector were fixed as described above and were then directly mounted on to glass slides for analysis by confocal microscopy.

RESULTS

Before attempting to characterize functional domains of the κ BL protein, it was necessary to determine in what form the human gene is expressed. In particular, we were interested to know if variations in mRNA splicing or translation could result in the expression of different forms of the κ BL protein.

A search of the expressed sequence tag (EST) database with the κ BL cDNA sequence (GenBank® accession number NM_005007) identified seven ESTs (GenBank® accession num-

bers: T05372, BF312108, BE273054, AL552315, BE313264, BE312881 and AL575644) that lacked the last 45 nt of exon III of the predicted full-length κ BL cDNA (Figure 1A). The new alternative splice site in exon III conforms to the GT rule for consensus 5' splice sites [21] and the reading frame in exon IV is maintained. Therefore the two splice variants of κ BL are identical except for the absence of the amino acid sequence GELEDEWQEVMMGRFE (where single-letter amino-acid notation has been used) from the κ BLsv1 protein. To confirm the expression of κ BL and κ BLsv1 transcripts in human cells, nested RT-PCR analysis was performed using primers F1 and R1 in the first round reactions and primers F2 and R2 in the second round reactions (Figure 1A). PCR products corresponding to full-length κ BL (293 bp) and κ BLsv1 (248 bp) were detected in all the cell lines tested (Figure 1B). Each of the PCR products was sequenced to confirm nucleotide specificity. In addition to the κ BL and κ BLsv1 PCR products a third, more slowly migrating band was also observed (Figure 1B). Sequence analyses of this DNA implied that it corresponded to a heteroduplex of the κ BL and κ BLsv1 PCR products. This was confirmed by denaturing and slowly reannealing the PCR products prior to electrophoresis, which led to a decrease in the intensity of the heteroduplex band and a corresponding increase in the intensity of the two lower bands (results not shown).

Translation initiation of κ BL

In accordance with the 'ribosome scanning' model of translation initiation [22], the translation initiation site of κ BL was predicted to be at the first ATG codon in the cDNA, which occurs at nt 69. However, the sequence surrounding this codon (CTGatgA) conforms poorly to either the strong (RNNatgG, where R = G or A, and N = G, A, T or C) or adequate (RNNatgY or YNNatgG, where Y = C or T) Kozak consensus sequence for

Figure 1 Identification of a spliced variant of human κ BL and two sites of translational initiation

(A) A schematic representation of the two alternative mRNA transcripts of the κ BL gene. Grey boxes represent exons I–IV, which are common to both κ BL transcripts, and the striped box represents the 3' end of exon III that is absent in the splice variant (κ BLsv1). Arrows indicate the positions of the primers used for nested RT-PCR analysis. (B) Nested RT-PCR analysis was performed on total RNA isolated from Jurkat (J), K562 (K), Molt 4 (M), Raji (R), U937 (U) or HeLa (H) cells, using the primers shown in (A). The bands corresponding to the two transcripts of κ BL (κ BL and κ BLsv1), as well as a heteroduplex formed by these two cDNA species, are indicated on the left of the gel. The sizes of the molecular mass markers (M) are shown in kb. (C) A schematic representation of the constructs used to analyse the site of translation initiation of human κ BL. Grey boxes represent κ BL-coding sequence, and black boxes represent the T7 epitope tag. Kozak sequences surrounding the initiation codons (atg) are shown in capital letters. The G-rich and C-rich sequences located upstream and downstream respectively of the first ATG codon are also represented. The numbers above the diagrams show the distance in nucleotides from the first ATG. (D) The constructs described in (C) were transiently expressed in Cos-7 cells, and cell lysates were analysed by Western blotting with a 1:10000 dilution of a mouse anti-(T7 tag) monoclonal antibody, followed by a 1:3000 dilution of a peroxidase-conjugated anti-mouse antibody. ATG1, ATG2, UTR1 and UTR2 correspond to constructs described in (C). The sizes of the molecular mass markers (M) in kDa are shown to the right of the blot.

Figure 2 Nuclear localization of IκBL

(A–C) Cos-7 cells transfected with the pcDNA3 vector encoding either IκBL (A), G11Z (B) or G9 (C) cDNAs fused to the T7 epitope tag. The cells were immunostained with a 1:3000 dilution of a mouse anti-T7 tag monoclonal antibody (Novagen), followed by a 1:100 dilution of an FITC-conjugated anti-mouse antibody (Sigma). (D and E) Cos-7 cells transiently transfected with pcDNA3 constructs coding for the T7 epitope tag fused to either the C-terminus of IκBL (Iκ-T7), the N-terminus of IκBL (T7-Iκ), the N-terminus of IκBLsv1 (T7-Iκsv1), or the C-termini of the nuclear protein G11Z, or the non-nuclear proteins G9 and G15. Following transfection cells were fractionated into nuclear (Nuc) and cytoplasmic/non-nuclear-organelle (Cyt) fractions. Equal volumes of the nuclear (D) and the cytoplasmic (E) fractions were analysed by Western blotting with a 1:10000 dilution of a mouse anti-T7 tag monoclonal antibody (Novagen), followed by a 1:3000 dilution of a peroxidase-conjugated anti-mouse antibody (Sigma). The sizes of the molecular mass markers (M) in kDa are shown to the right of each blot.

translation initiation [23]. However, at nt 139 there is a second ATG that has an adequate Kozak sequence (TCCatgG). Furthermore, the first ATG (nt 69) is flanked by G-rich (nt 48–60) and C-rich (nt 77–87) stretches, which may inhibit translation initiation by formation of a hairpin loop.

To determine if translation can initiate from either of the first two ATGs *in vivo* the ATG1-pcDNA3 and ATG2-pcDNA3 vectors were constructed (Figure 1C), which encode the IκBL protein starting at the first and second ATG codons respectively, together with a short fragment (9–10 bp) of the sequence upstream of these ATGs. To assess the effect of the G-rich and C-rich stretches on translation initiation, we designed the UTR1-pcDNA3 and UTR2-pcDNA3 constructs (Figure 1C), which encode the IκBL protein starting at the first and second ATG codons respectively, together with a long fragment (54–57 bp) of the sequence upstream of these ATGs.

Transient expression of the ATG2-pcDNA3 and UTR2-pcDNA3 constructs in Cos-7 cells resulted in the expression of a single, 41 kDa IκBL-T7 protein (Figure 1D). In contrast, the ATG1-pcDNA3 construct expressed two epitope-tagged proteins: one that co-migrated with the band from ATG2-pcDNA3 transfected cells, and also a 43 kDa band corresponding to IκBL translated from its first ATG (Figure 1D). These data show that translation can initiate from both of the first two ATG codons. Interestingly, the UTR1-pcDNA3 construct that contained the same coding sequence as ATG1-pcDNA3, but also included the G-rich stretch upstream of the first ATG, expressed only a single epitope-tagged protein that co-migrated with the band from ATG2-pcDNA3-transfected cells (Figure 1D). This indicates that expression of the longer open reading frame was

suppressed by the inclusion of an additional 44 bp of the endogenous IκBL 5' non-coding sequence.

Nuclear localization of IκBL

Having established the expression of IκBL and IκBLsv1 mRNAs in human cells we then analysed the subcellular localization of the encoded protein products. Localization experiments using confocal microscopy showed that transiently expressed IκBL-T7 was predominantly localized in the cell nucleus (Figure 2A). However, in a minority of Cos-7 cells IκBL-T7 was observed only in the cytoplasm (results not shown). These data suggested that nuclear localization of the IκBL protein could be conditional.

To determine the average subcellular distribution of IκBL within the whole population of transfected cells a detergent-based method of cell fractionation was used to separate the nuclei from the rest of the cell contents ('cytoplasmic' fraction). To monitor the efficiency of cell fractionation T7-tagged G11Z [18] was used as a nuclear marker (Figures 2B and 2D), and T7-tagged G9 [14] and G15 [19] proteins were used as markers for non-nuclear fractions (Figures 2C and 2E). Both IκBL and IκBLsv1 were found to be predominantly concentrated in the nucleus of transiently transfected Cos-7 cells (Figures 2D and 2E). Also, the nuclear localization of IκBL was unaffected by the N- or C-terminal position of the T7 epitope tag (Figure 2D). However, when C-terminally tagged IκBL was expressed, two IκBL products were observed due to the absence of a strong or adequate Kozak consensus sequence around the first ATG [23]. Significantly, both proteins showed the same subcellular localiza-

Figure 3 $I\kappa$ BL is present in nuclear speckles

HeLa cells were transfected with pcDNA3 vectors encoding T7 epitope-tagged versions of either full-length $I\kappa$ BL (**A**, **D**, **E** and **F**), the $I\kappa$ BL-ATG2 construct (**B**) or $I\kappa$ BLsv1 (**C**). The cells were immunostained with either a 1:3000 dilution of a mouse anti-T7 tag monoclonal antibody (Novagen) alone (**A**, **B** and **C**) or with a combination of anti-T7 tag antibody (Novagen) and a 1:500 dilution of anti-Sm polyclonal antibody (**D**, **E** and **F**). The cells were then incubated with a 1:100 dilution of an FITC-conjugated anti-mouse antibody (Sigma) (**A**, **B** and **C**), or a mixture of a 1:100 dilution of Cy3-conjugated anti-mouse monoclonal antibody (Jackson) and a 1:100 dilution of FITC-conjugated anti-human monoclonal antibody (Jackson) (**D**, **E** and **F**). In the co-localization experiment, $I\kappa$ BL staining (**D**) and Sm staining (**E**) are shown independently, and as a merged image (**F**).

tion. In addition, $I\kappa$ BL and $I\kappa$ BLsv1 both exhibited a nuclear distribution following transient expression in the human cell line HeLa (results not shown).

Subnuclear localization of $I\kappa$ BL

Closer examination of the nuclei of HeLa cells transfected with $I\kappa$ BL revealed that, in addition to a diffuse nuclear expression pattern, $I\kappa$ BL also concentrated within punctate foci (Figure 3A). Significantly, both $I\kappa$ BL-ATG2 and $I\kappa$ BLsv1 proteins demonstrated a similar punctate subnuclear distribution (Figures 3B and 3C). As the observed punctate distribution of $I\kappa$ BL proteins was reminiscent of that observed for 'nuclear speckles' [24], co-localization experiments were carried out with an antibody against the splicing factor Sm, which is present in nuclear speckles [25]. $I\kappa$ BL (Figure 3D) and Sm (Figure 3E) were found to be co-localized to the same subnuclear domains (yellow areas in Figure 3F), though in some foci $I\kappa$ BL was present in the absence of Sm (red areas in Figure 3F). These data show that $I\kappa$ BL is localized to areas of the nucleus that we know to be enriched in proteins involved in pre-mRNA splicing.

Identification of domains responsible for $I\kappa$ BL nuclear localization

Analysis of the $I\kappa$ BL protein sequence with the PredictNLS program [26] identified the RQRRER motif (amino acids 30–35) as a potential NLS by similarity to part of the NLS of CCAAT/enhancer-binding proteins [27]. Alternatively, $I\kappa$ BL nuclear import could be mediated by its ankyrin repeats, as is the

case for $I\kappa$ B α [12]. To identify domains responsible for $I\kappa$ BL nuclear targeting, we created epitope-tagged deletion constructs of $I\kappa$ BL that lacked either the basic region (amino acids 29–39) containing the putative NLS, the ankyrin repeats (amino acids 41–129) or the C-terminal leucine zipper domain (amino acids 359–381) (Figure 4A). Immunofluorescence of Cos-7 and HeLa cells transiently expressing these constructs showed that deletion of either the NLS (Figures 4C and 4G) or the ankyrin-repeat domain (Figures 4D and 4H) was sufficient to impair nuclear import and disrupt the localization of $I\kappa$ BL to nuclear speckles. In contrast, deletion of the leucine zipper domain had no effect on nuclear localization of the $I\kappa$ BL protein (Figures 4E and 4I), and did not affect the subnuclear localization of $I\kappa$ BL to nuclear speckles in HeLa cells (compare Figures 4F and 4I).

These data suggest that the NLS and the ankyrin repeats are both necessary for nuclear localization. To determine if these domains are sufficient to promote nuclear localization we created fusion constructs in which different domains of $I\kappa$ BL were fused to the C-terminal end of EGFP. Cos-7 and HeLa cells transfected with the empty pEGFP vector expressed EGFP throughout the cytoplasm and nucleus (Figures 5A and 5F). EGFP fused to the full-length $I\kappa$ BL protein was efficiently concentrated within the nucleus as expected (Figures 5B and 5G). EGFP fused to the N-terminus of $I\kappa$ BL carrying the putative NLS (amino acids 1–40) was detected in the cytoplasm as well as the nucleus (Figures 5C and 5H), although this fusion protein was present at a higher concentration in the nucleus than unmodified EGFP, and unlike either EGFP or EGFP- $I\kappa$ BL, it accumulated at very high levels in the nucleoli. EGFP fused to the ankyrin domain

Figure 4 Domains necessary for nuclear localization of IκBL

(A) A schematic representation of the main putative domains of IκBL. The white box represents the arginine-rich putative NLS, shown in detail above the box. The checked box and striped box represent the ankyrin repeats (ANK) and leucine zipper domain (LeuZ) respectively. Amino acids are numbered from the first methionine coded by the mRNA, and the beginning and end of each domain is shown. Cos-7 (B–E) or HeLa (F–I) cells were transfected with either full-length IκBL (B and F), IκBL lacking the arginine-rich NLS (ΔNLS) (C and G), IκBL lacking the ankyrin repeats (ΔANK) (D and H) or IκBL lacking the leucine zipper domain (ΔLeuZ) (E and I). All of these constructs were fused at their N-terminus to the T7 epitope tag. The cells were immunostained with a 1/3000 dilution of a mouse anti-T7 tag monoclonal antibody (Novagen) followed by a 1:100 dilution of an FITC-conjugated anti-mouse antibody (Sigma).

(amino acids 40–137) was detected in both the cytoplasm and the nucleus, though it had a more even distribution between these compartments than EGFP–NLS, and it was excluded from the nucleoli (Figures 5D and 5J). Consequently, it would appear that neither the ankyrin repeats nor the putative NLS were individually sufficient to induce nuclear localization of EGFP. However, when both domains (amino acids 1–137) were fused to EGFP, complete nuclear localization was achieved (Figures 5E and 5J). Furthermore, these domains conferred a speckled distribution on EGFP in the nucleus (Figures 5E and 5J), like that exhibited by full-length IκBL (Figures 5B and 5G).

DISCUSSION

In the present study we present three novel observations. First, we have identified a variant form of IκBL (IκBLsv1), which is widely expressed in human cell lines. Secondly, we show that translation of human IκBL mRNA can occur at both of the first two methionines. Finally, we have identified a distinct bipartite protein-targeting motif, which is necessary and sufficient for targeting IκBL to nuclear speckles. With respect to the identification of a human IκBL splice variant, deletion of the GELED-EWQEVMGRFE amino acid motif does not alter any of the known functional domains of the IκBL protein, and it does not affect its subcellular localization. As such, the functional significance of this splice variant remains to be determined.

In addition to IκBL isoforms that result from splicing variation it is also possible that further changes in protein sequence may occur at the level of translation. Specifically, the formation of two distinct protein products from the ATG1-IκBL cDNA clone implies that both of the first methionine residues are capable of initiating translation of the IκBL protein *in vivo*. The preferential initiation at the second methionine in the UTR1-IκBL cDNA clone raises the possibility that variable usage of cryptic promoters, cryptic splice sites or the degree of mRNA secondary structure could differentially affect the balance of expression in different tissues or disease states. Whatever the mechanism involved, the fact that two IκBL proteins that differ at their N-termini can be produced from IκBL cDNA *in vivo* is of interest as amino acids 1–18 of IκBL are predicted to form a PEST sequence that could affect the post-translational regulation of IκBL expression. However, detailed peptide analysis will be required to define the precise N-terminal sequence of endogenously expressed IκBL proteins.

Significantly, all of the protein isoforms of IκBL analysed in the present paper (N-terminal and splice variants) were localized to nuclear speckles. We have shown that when both the NLS and the ankyrin repeats of IκBL are fused to a heterologous protein they can efficiently target it to nuclear speckles. A variety of different domains have been implicated in targeting proteins to nuclear speckles: the RNA recognition motifs of PTB-associated splicing factor ('PSF') and serine/arginine-rich ('SR') splicing

Figure 5 Domains of $\text{I}\kappa\text{BL}$ sufficient for nuclear localization of EGFP

Cos-7 (A–E) or HeLa (F–J) cells were grown on glass cover slips and transfected with either EGFP (GFP) alone (A and F), full-length $\text{I}\kappa\text{BL}$ fused to EGFP (B and G), the arginine-rich $\text{I}\kappa\text{BL}$ NLS fused to EGFP (C and H), the $\text{I}\kappa\text{BL}$ ankyrin repeats (ANK) fused to EGFP (D and I) or both the $\text{I}\kappa\text{BL}$ NLS and the ankyrin repeats fused to EGFP (E and J).

factors [28,29]; the serine/arginine-rich domains of serine/arginine-rich splicing factors [29,30]; the Sm core domain of the U-rich small nuclear ribonucleoprotein 1 ('U1 snRNP') [31]; the forkhead-associated domain of nuclear inhibitory subunit of protein phosphatase 1 ('NIPPI1') [32]; and a threonine/proline-rich domain in splicing factor SF3b [33]. To our knowledge this is the first time an ankyrin domain has been implicated in the targeting of proteins to nuclear speckles. In some cases the

known speckle localization domains are distinct from the domains necessary for nuclear localization [28,32], whereas in other proteins, including $\text{I}\kappa\text{BL}$, these domains overlap with, or are identical with, the domains necessary for nuclear localization [30,31,33]. Biochemical purification of nuclear speckles showed that they contain at least 75 different proteins, 19 of which were identified as splicing factors previously shown to localize in this site [25]. Upon stimulation of transcription, the splicing factors are recruited by RNA polymerase II to sites of active transcription where they are involved in the processing of nascent transcripts [34,35]. Consequently, the observed concentration of $\text{I}\kappa\text{BL}$ proteins within nuclear speckles may imply that these proteins are involved in the regulation of splicing. However, until the full complement of speckle proteins has been functionally analysed it is not possible to say with impunity that localization to these structures is diagnostic of a role for the protein in mRNA splicing.

As mentioned previously, there are a number of similarities between $\text{I}\kappa\text{BL}$ and members of the $\text{I}\kappa\text{B}$ family of proteins, notably the sequence similarity in their ankyrin repeat domains and the presence of putative PEST sequences. In the present study we show that, like $\text{I}\kappa\text{B}\alpha$ [12], the $\text{I}\kappa\text{BL}$ protein depends on its ankyrin repeats for nuclear localization. However, unlike $\text{I}\kappa\text{B}\alpha$, the $\text{I}\kappa\text{BL}$ proteins also require an additional arginine-rich NLS. Previous studies have shown that the ankyrin repeat sequences of other proteins can mediate nuclear import [12]. However, nuclear import is clearly not a universal property of all ankyrin repeat sequences as the ankyrin domains of Notch 1 do not mediate nuclear import [12]. As such, there appear to be functionally distinct sets of ankyrin repeat domains. Precisely which structural motifs or key amino acid residues are required to confer a propensity for nuclear import remains unclear. Interestingly, key leucine residues, which are essential for the nuclear import activity of $\text{I}\kappa\text{B}\alpha$ [12], are not conserved in human $\text{I}\kappa\text{BL}$ proteins [1]. This discrepancy may well explain why the ankyrin domains of $\text{I}\kappa\text{BL}$ are not independently sufficient to promote nuclear import. One theory, which has been proposed to explain the ability of ankyrin repeat domains to mediate nuclear import suggests that ankyrin domains function as protein–protein interaction sites [36], which allow ankyrin repeat-containing proteins to travel into the nucleus in 'piggy back' with other proteins that themselves have a classical NLS [12,37,38]. If this model is correct, it is possible that different 'piggy back' partners may mediate nuclear import of different types of ankyrin-containing proteins. Consequently, the identification of different ankyrin-containing proteins with unique sequence requirements is an important step towards addressing these important fundamental questions.

The observation of cytoplasmic expression of $\text{I}\kappa\text{BL}$ in a minority of Cos-7 cells suggested that, like $\text{I}\kappa\text{B}\alpha$, nuclear localization of $\text{I}\kappa\text{BL}$ might be conditionally regulated. However, while $\text{I}\kappa\text{B}\alpha$ is rapidly degraded in response to a variety of stimuli, such as fetal calf serum, $\text{TNF}\alpha$, PMA and pervanadate, with subsequent nuclear import of newly synthesized $\text{I}\kappa\text{B}\alpha$ [9,39–41], $\text{I}\kappa\text{BL}$ showed no change in subcellular or subnuclear distribution following such stimuli (results not shown). Furthermore, the overall subcellular localization of $\text{I}\kappa\text{BL}$ differs quite dramatically from that reported for $\text{I}\kappa\text{B}\alpha$ [8]. The presence of $\text{I}\kappa\text{BL}$ in a nuclear splicing factor reservoir is an intriguing finding, which implies that the sequence similarity that exists between $\text{I}\kappa\text{BL}$ and $\text{I}\kappa\text{B}\alpha$ may reflect an evolutionary rather than a functional link between these proteins.

We are grateful to Dr Begona Aguado (HGMP Resource Centre, Cambridge, U.K.) for the pcDNA3-G11Z, pcDNA3-G9-Tag-1 and pcDNA3G15Tag constructs. This research was funded by the U.K. Medical Research Council.

REFERENCES

- 1 Albertella, M. R. and Campbell, R. D. (1994) Characterization of a novel gene in the human major histocompatibility complex that encodes a potential new member of the I κ B family of proteins. *Hum. Mol. Genet.* **3**, 793–799
- 2 Ota, M., Katsuyama, Y., Kimura, A., Tsuchiya, K., Kondo, M., Naruse, T., Mizuki, N., Itoh, K., Sasazuki, T. and Inoko, H. (2001) A second susceptibility gene for developing rheumatoid arthritis in the human MHC is localized within a 70-kb interval telomeric of the TNF genes in the HLA class III region. *Genomics* **71**, 263–270
- 3 de la Concha, E. G., Fernandez-Arquero, M., Lopez-Nava, G., Martin, E., Allcock, R. J., Conejero, L., Paredes, J. G. and Diaz-Rubio, M. (2000) Susceptibility to severe ulcerative colitis is associated with polymorphism in the central MHC gene IKBL. *Gastroenterology* **119**, 1491–1495
- 4 Whiteside, S. T. and Israel, A. (1997) I κ B proteins: structure, function and regulation. *Semin. Cancer Biol.* **8**, 75–82
- 5 Hatada, E. N., Naumann, M. and Scheidereit, C. (1993) Common structural constituents confer I κ B activity to NF- κ B p105 and I κ B/MAD-3. *EMBO J.* **12**, 2781–2788
- 6 Lin, R., Beuparlant, P., Makris, C., Meloche, S. and Hiscott, J. (1996) Phosphorylation of I κ B α in the C-terminal PEST domain by casein kinase II affects intrinsic protein stability. *Mol. Cell. Biol.* **16**, 1401–1409
- 7 Jacobs, M. D. and Harrison, S. C. (1998) Structure of an I κ B α /NF- κ B complex. *Cell (Cambridge, Mass.)* **95**, 749–758
- 8 Zabel, U., Henkel, T., Silva, M. S. and Baeuerle, P. A. (1993) Nuclear uptake control of NF- κ B by MAD-3, an I κ B protein present in the nucleus. *EMBO J.* **12**, 201–211
- 9 Henkel, T., Machleidt, T., Alkalay, I., Kronke, M., Ben-Neriah, Y. and Baeuerle, P. A. (1993) Rapid proteolysis of I κ B- α is necessary for activation of transcription factor NF- κ B. *Nature (London)* **365**, 182–185
- 10 Brown, K., Park, S., Kanno, T., Franzoso, G. and Siebenlist, U. (1993) Mutual regulation of the transcriptional activator NF- κ B and its inhibitor, I κ B- α . *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2532–2536
- 11 Arenzana-Seisdedos, F., Turpin, P., Rodriguez, M., Thomas, D., Hay, R. T., Virelizier, J. L. and Dargemont, C. (1997) Nuclear localization of I κ B α promotes active transport of NF- κ B from the nucleus to the cytoplasm. *J. Cell Sci.* **110**, 369–378
- 12 Sachdev, S., Hoffmann, A. and Hannink, M. (1998) Nuclear localization of I κ B α is mediated by the second ankyrin repeat: the I κ B α ankyrin repeats define a novel class of *cis*-acting nuclear import sequences. *Mol. Cell. Biol.* **18**, 2524–2534
- 13 Albertella, M. (1997) The isolation and characterisation of new genes in the human major histocompatibility complex. Ph.D. Thesis, University of Oxford, Oxford
- 14 Milner, C. M., Smith, S. V., Carrillo, M. B., Taylor, G. L., Hollinshead, M. and Campbell, R. D. (1997) Identification of a sialidase encoded in the human major histocompatibility complex. *J. Biol. Chem.* **272**, 4549–4558
- 15 Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. and Pease, L. R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–68
- 16 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680–685
- 17 Spector, D. L., Goldman, R. D. and Leinwand, L. A. (1988) Preparation of nuclei from HeLa cell suspension cultures by osmotic swelling. In *Cell Biology – A Laboratory Manual*, vol. 1, pp. 43.10. Cold Spring Harbor Laboratory Press, Plainview, NY
- 18 Gomez-Escobar, N., Chou, C. F., Lin, W. W., Hsieh, S. L. and Campbell, R. D. (1998) The G11 gene located in the major histocompatibility complex encodes a novel nuclear serine/threonine protein kinase. *J. Biol. Chem.* **273**, 30954–30960
- 19 Aguado, B. and Campbell, R. D. (1998) Characterization of a human lysophosphatidic acid acyltransferase that is encoded by a gene located in the class III region of the human major histocompatibility complex. *J. Biol. Chem.* **273**, 4096–4105
- 20 Dostie, J., Lejbkowitz, F. and Sonenberg, N. (2000) Nuclear eukaryotic initiation factor 4E (eIF4E) colocalizes with splicing factors in speckles. *J. Cell Biol.* **148**, 239–247
- 21 Mount, S. M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459–472
- 22 Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208
- 23 Kozak, M. (1996) Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome* **7**, 563–574
- 24 Spector, D. L. (1996) Nuclear organization and gene expression. *Exp. Cell Res.* **229**, 189–197
- 25 Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S. and Spector, D. L. (1999) Purification and biochemical characterization of interchromatin granule clusters. *EMBO J.* **18**, 4308–4320
- 26 Cokol, M., Nair, R. and Rost, B. (2000) Finding nuclear localization signals. *EMBO Rep.* **1**, 411–415
- 27 Williams, S. C., Angerer, N. D. and Johnson, P. F. (1997) C/EBP proteins contain nuclear localization signals imbedded in their basic regions. *Gene Expression* **6**, 371–385
- 28 Dye, B. T. and Patton, J. G. (2001) An RNA recognition motif (RRM) is required for the IQD localization of PTB-associated splicing factor (PSF) to subnuclear speckles. *Exp. Cell Res.* **263**, 131–144
- 29 Caceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L. and Krainer, A. R. (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.* **138**, 225–238
- 30 Hedley, M. L., Amrein, H. and Maniatis, T. (1995) An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11524–11528
- 31 Malatesta, M., Fakan, S. and Fischer, U. (1999) The Sm core domain mediates targeting of U1 snRNP to subnuclear compartments involved in transcription and splicing. *Exp. Cell Res.* **249**, 189–198
- 32 Jagiello, I., Van Eynde, A., Vulsteke, V., Beullens, M., Boudrez, A., Keppens, S., Stalmans, W. and Bollen, M. (2000) Nuclear and subnuclear targeting sequences of the protein phosphatase-1 regulator NIPP1. *J. Cell Sci.* **113**, 3761–3768
- 33 Eilbracht, J. and Schmidt-Zachmann, M. S. (2001) Identification of a sequence element directing a protein to nuclear speckles. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3849–3854
- 34 Melcak, I., Cermanova, S., Jirsova, K., Koberna, K., Malinsky, J. and Raska, I. (2000) Nuclear pre-mRNA compartmentalization: trafficking of released transcripts to splicing factor reservoirs. *Mol. Biol. Cell* **11**, 497–510
- 35 Misteli, T., Caceres, J. F. and Spector, D. L. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature (London)* **387**, 523–527
- 36 Sedgwick, S. G. and Smerdon, S. J. (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316
- 37 Sachdev, S., Bagchi, S., Zhang, D. D., Mings, A. C. and Hannink, M. (2000) Nuclear import of I κ B α is accomplished by a ran-independent transport pathway. *Mol. Cell. Biol.* **20**, 1571–1582
- 38 Turpin, P., Hay, R. T. and Dargemont, C. (1999) Characterization of I κ B α nuclear import pathway. *J. Biol. Chem.* **274**, 6804–6812
- 39 Arenzana-Seisdedos, F., Thompson, J., Rodriguez, M. S., Bachelier, F., Thomas, D. and Hay, R. T. (1995) Inducible nuclear expression of newly synthesized I κ B α negatively regulates DNA-binding and transcriptional activities of NF- κ B. *Mol. Cell. Biol.* **15**, 2689–2696
- 40 Cressman, D. E. and Taub, R. (1993) I κ B α can localize in the nucleus but shows no direct transactivation potential. *Oncogene* **8**, 2567–2573
- 41 Singh, S. and Aggarwal, B. B. (1995) Protein-tyrosine phosphatase inhibitors block tumor necrosis factor-dependent activation of the nuclear transcription factor NF- κ B. *J. Biol. Chem.* **270**, 10631–10639

Received 5 June 2001/30 October 2001; accepted 23 November 2001