Nucleic Acids Research, 1994, Vol. 22, No. 16 3365–3372

Identification of a novel HIV-1 TAR RNA bulge binding protein

Bernadette Baker, Martina Muckenthaler[1], Eric Vives[2], Andy Blanchard[2], Martin Braddock[2], W.Nacken[3], Alan J.Kingsman[2] and Susan M.Kingsman[2,*]

Glaxo Group Research and Development Ltd, Greenford Road, Greenford, Middlesex UB6 OHE, UK, [1]EMBL, Meyerhofstrasse, D-6900 Heidelberg, Germany, [2]Retrovirus Molecular Biology Group, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK and [3]Institut für Experimentelle Dermatologie, Universitat Munster, 48149 Munster, Germany

Received May 19, 1994; Revised and Accepted July 18, 1994

ABSTRACT

The Tat protein binds to TAR RNA to stimulate the expression of the human immunodeficiency virus type 1 (HIV-1) genome. Tat is an 86 amino acid protein that contains a short region of basic residues (aa49 – aa57) that are required for RNA binding and TAR is a 59 nucleotide stem-loop with a tripyrimidine bulge in the upper stem. TAR is located at the 5′ end of all viral RNAs. *In vitro*, Tat specifically interacts with TAR by recognising the sequence of the bulge and upper stem, with no requirement for the loop. However, *in vivo* the loop sequence is critical for activation, implying a requirement for accessory cellular TAR RNA binding factors. A number of TAR binding cellular factors have been identified in cell extracts and various models for the function of these factors have been suggested, including roles as coactivators and inhibitors. We have now identified a novel 38 kD cellular factor that has little general, single-stranded or double-stranded RNA binding activity, but that specifically recognises the bulge and upper stem region of TAR. The protein, referred to as BBP (bulge binding protein), is conserved in mammalian and amphibian cells and in *Schizosaccharomyces pombe* but is not found in *Saccharomyces cerevisiae*. BBP is an effective competitive inhibitor of Tat binding to TAR *in vitro*. Our data suggest that the bulge-stem recognition motif in TAR is used to mediate cellular factor/RNA interactions and indicates that Tat action might be inhibited by such competing reactions *in vivo*.

INTRODUCTION

The expression of human immunodeficiency virus type 1 (HIV-1) genes is dependent upon the activity of the virally encoded Tat protein. Tat binds to an RNA sequence called TAR, that is located at the immediate 5′ end of all viral mRNAs [1]. The interaction between Tat and TAR results in a marked stimulation of transcription and consequent increase in virus production. Tat action increases both the initiation of transcription and processivity of polII [2]. There is evidence from *in vitro* studies that Tat contacts TFIID [3] and SP1 [4], however a number of other Tat binding proteins have also been identified [5, 6] and the precise mechanism whereby Tat activates transcription is not yet clear. In addition, Tat can stimulate translation of TAR containing RNA in *Xenopus* oocytes [7] and there is some evidence for a post-transcriptional activity in mammalian cells [8].

The Tat/TAR interaction *in vitro* has been well characterised. TAR forms a partially base paired stem-loop with a tripyrimidine bulge on the 5′ side of the loop. Chemical and 2-D NMR analyses provide no strong evidence that the nucleotides in the loop are paired or stacked although the data do not preclude the formation of non Watson–Crick base pairs [9–11]. Bulges are known to induce pronounced kinks in RNA [12] and it appears that the bulge in TAR serves to deform the A-form helix, resulting in a widening of the major groove to reveal the functional groups on the Watson–Crick base pairs [13, 14]. Tat binding depends upon sequences in the bulge and upper stem. The critical residues are U23 in the bulge, G26 in the stem and a base paired purine at position A27, and it has been proposed that U23 in the bulge can form a base triple interaction with the A27 [15–19].

The binding domain of Tat contains a cluster of basic amino acids (residues 49 to 57) which can bind to TAR in isolation [20], but full specificity requires additional flanking amino acids [21]. The precise contacts between Tat and TAR have not yet been characterised, although the finding that a single arginine can bind to G26 combined with *in vivo* mutagenesis studies [20, 22], has led to the suggestion that Arg 52 and 53 in Tat play a key role, however contact with other residues outside the basic region, such as Lys41 may also be a possibility [14]. Tat therefore appears to belong to a class of proteins that recognise RNA via an arginine

*To whom correspondence should be addressed

rich domain, these include λN protein and the R17 and Qβ phage coat proteins [23].

The putative hexanucleotide loop in TAR is not critical for Tat binding *in vitro* [16, 21]. However, *in vivo,* the loop sequence is essential. Single or multiple base substitutions in the loop dramatically reduce Tat activation [24, 25]. Given that the loop has structural flexibility *in vitro* and there is little evidence that it contacts other regions of TAR [14], the *in vivo* requirement is unlikely to reflect a simple structural constraint. It appears that one or more loop binding cellular factors are required to achieve activation. Two models have been proposed to account for the requirement for loop binding proteins. It has been suggested that Tat interacts with a loop binding protein in solution to form a binding complex. This binding complex interacts with TAR via the combined interaction of Tat with the bulge and the cellular factor with the loop sequence. In this model, neither Tat nor the loop factors bind TAR independently *in vivo* [26]. A second model has been proposed, based on genetic analysis of Tat activation of translation in *Xenopus* oocytes [27]. In this system, TAR loop sequences are also required for Tat activation, but the requirement is conditional. Tat specific activation of a loop mutated TAR element can be obtained if excess of a short TAR RNA containing the loop mutation is present. It appears that the loop mutated TAR sequence in excess is removing a negative factor. One interpretation is that the role of the loop sequence is to bind a protein that will prevent the negative factor from interfering with Tat binding. Mutation analysis indicates that the negative factor requires the bulge region of TAR. Biochemical support for these models requires the purification and characterisation of TAR RNA binding proteins. Model 1 predicts the presence of a loop binding factor that is dependent upon Tat for binding and does not require any additional factors such as a bulge binding protein. Model 2 predicts the existence of loop binding proteins and bulge binding proteins which can bind independently to TAR.

A number of different cellular factors that bind to TAR independently of Tat have been identified. Some of these factors are general RNA binding proteins such as the p68 DI kinase [28]. Some recognise general features of RNA with some specificities for TAR such as TRBP1 [29] which recognise double stranded RNA with a G.C preference and therefore binds to the TAR upper stem, and p140, which recognises dsRNA and binds to the lower stem [30]. Others display a marked preference for binding to TAR. p185 (TRP1) [31, 32] and p68 [33] specifically recognise the TAR loop sequence and TRP2 (70−110kD) specifically recognises the TAR bulge. To date no Tat-dependent TAR RNA binding proteins have been identified. The only TAR binding protein for which a clone is available is TRBP1, originally identified as a 36 kD protein based on the original cDNA [29] it is now known to be 54kD (A.Gatignol pers. comm.) and to be related to the RNP class of RNA binding proteins [34]. In addition, a number of TAR binding proteins can be induced by phorbol esters [35] or UV treatment [36] but the specificities of these have yet to be characterised. In this paper we report the identification of a novel TAR bulge-specific RNA binding protein which we call bulge binding protein, BBP. It has a marked specificity for the sequence of the bulge and shows little general binding to single or double stranded RNA. It is approximately 38kD and, by several criteria, appears to be distinct from any of the previously identified TAR RNA binding proteins. It is conserved in all mammalian cells tested but although present in *Schizosaccharomyces pombe* it was not detected in the unrelated yeast *Saccharomyces cerevisiae*. Interestingly, the protein functions as an effective competitive inhibitor of Tat binding to TAR *in vitro*. Our data and those of Sheline *et al.* [31] therefore indicate that cells contain at least two proteins, TRP2 and BBP, which by virtue of their interaction with the bulge stem region of TAR, could function as inhibitors of Tat action. The data indicate that the TAR bulge-stem motif in RNA may also be used to mediate specific RNA−protein interactions in the cell.

MATERIALS AND METHODS

Plasmid construction and labelling of RNAs

The plasmids used to produce short +18 to +44 TAR RNA *in vitro* [37] were constructed by subcloning an oligonucleotide containing the wild type (WT) HIV-1 TAR sequence 5′-ATT-CCAGATCTGAGCCTGGGAGCTCTCTGGA-3′, or the corresponding mutant sequences into the *Eco*RI and *Hin*dIII site of pGEM3Zf+ (Figure 1). Short TAR RNAs had the sequences shown in Figure 1 flanked by the polylinker sequence 5′-CTT-ATGAGTTCGAA//TAR//G-3′. The identity of the plasmids was confirmed by dideoxy sequencing [38]. Sense RNA from wild type and mutant constructs was prepared from *Hin*dIII linearized templates using T7 polymerase (BRL). All labelled and non-labelled RNAs were synthesized according to the Promega manual. The labelled RNAs were purified from a denaturing 6% polyacrylamide gel, eluted overnight into 0.5M sodium acetate (pH7.0), 1mM EDTA and 0.5% SDS and precipitated with 0.3M NaCl and 2.5 volumes of ethanol. The RNAs studied were wild type TAR (WT), antisense TAR (AS), a triple base substitution in the bulge, UCU to CUC (BS1), a triple base substitution in the bulge, UCU to AAG (BS2), a single base substitution in the bulge, U23 to G23 (BS23), a triple base substitution in the loop, CUGGGA to AGGGUA (LS), an upper stem mutation, A27.U38 to U27.A38 (S1), an upper stem mutation, G26.C39 to C26.G39 (S2), an upper stem mutation with an additional three G.C pairs inserted between A27.U38 and C29.G36 (S3), a double mutation combining the bulge substitution, BS1 with the loop substitution, LS (DM1), a double mutation combining the bulge substitution BS1 and the stem mutation S2 (DM2), a mutant in which the tripyrimidine bulge, UCU, was moved 3′ of the loop (RHB) and a bulge deletion mutant (BD).

Cell lines and yeast strains

HeLa S3 (human cervical carcinoma epithelial cells) were grown in suspension in SMEM (Gibco), 1% non-MEM amino acids, 5% NCS, 1% sodium pyruvate. H9 cells (human CD4+ T cell line) and K562 (human myeloid leukaemia cells) were grown in RPMI 1640, 10% FCS. CHO (Chinese hamster ovary) cells were maintained in DMEM (Gibco), 10%FCS, 1% glutamine. *Saccharomyces cerevisiae* (DBY 745 [MAT α, *ade*1-100, *leu*2-3, *leu*2-112, *ura*3-52]) was grown in YEPD and *Schizosaccharomyces pombe* (972h$^-$) in YES media at 30° C to a density of 5×10^7 cells/ml. Whole yeast extracts were prepared by the method of Mellor *et al.* [39].

Partial purification of TAR binding cellular factors

Nuclear extracts from mammalian cells were prepared as described previously by Dignam *et al* [40]. Nuclear extract was either assayed directly or applied to a heparin agarose (Promega) column (0.5cm×3cm), which was previously equilibrated in TM buffer (50mM Tris−HCl (pH7.9), 12.5mM $MgCl_2$, 1mM

EDTA, 20% glycerol, 0.1mM PMSF, 1mM DTT) containing 0.1M KCl. The column was washed with 5 volumes of TM 0.1M and successive step elutions were obtained using TM 0.2M KCl to TM 0.7M KCl. Fractions eluted from a heparin agarose column were directly analysed by mobility shift assay. The fractions containing specific protein were pooled, concentrated by Centricon-30 (Amicon) centrifugation, dialysed against RNA binding buffer (10mM Tris−HCl (pH7.9), 2mM $MgCl_2$, 0.1mM EDTA, 4% glycerol, and 50mM KCl) and then tested against various RNA probes. In competition experiments cellular factors were further purified to remove non-specific proteins (E.Vives *et al.* manuscript in preparation).

RNA binding reactions and mobility shift assays

RNA binding was performed on ice for 20min in 10mM Tris−HCl (pH7.9), 2mM $MgCl_2$, 0.1mM EDTA, 4% glycerol, and 50mM KCl in a final volume of 20μl, using 5μg of total HeLa nuclear extract or 80ng of heparin agarose fractionated protein, 10μg of *S.cerevisiae* whole extract or 5μg of *S.pombe* whole extract, and α-^{32}P-UTP (Amersham) labelled +18 to +44 WT-TAR or mutant TAR (10^5 cpm per reaction, 10^8 cpm/μg) and 1μg of total *Xenopus* oocyte RNA or double stranded poly I.C (Pharmacia) as a nonspecific competitor. Competition studies were done using 1μl of Dignam extract and standard BBP binding conditions, except that reactions were made to 100mM KCl, Triton-X-100 0.5% and contained 0.2 ng labelled TAR, and either poly(U) (Pharmacia) or TAR at 10, 100 or 1000 fold excess. RNAs were premixed, added to the extract and incubated at 4°C for 10 min. The reactions were directly applied to a 8% non-denaturing polyacrylamide gel, containing 0.5×TBE and 5% glycerol and run at 15V/cm at 4°C. Protein−RNA complexes were quantified by phosphorimaging (Molecular Dynamics, Sunnyvale, Calif.) without file conversion [41] after overnight exposure.

Molecular weight estimation of TAR binding proteins

A standard mobility shift binding reaction containing WT probe and HeLa nuclear extract was exposed to UV in a Stratalinker (Stratagene) (4000μW/cm^2 at a distance of 7cm) for 10 min before separation on a 6% nondenaturing polyacrylamide gel. The band corresponding to the BBP−RNA complex was excised, digested with RNAse A (Sigma) *in situ*, boiled in sample buffer and run on a 12.5% SDS polyacrylamide gel. The size was estimated in comparison to a prestained molecular weight marker (BRL).

Tat binding

Tat was produced as an N-terminal hexahistidine tagged protein as described previously (46). Tat was bound to TAR at 4°C in Tat-binding buffer (10mM Tris−HCl (pH 7.5), 1mM EDTA, 50mM NaCl, 0.5U/μl RNasin, 0.04μg/μl BSA and 0.05% glycerol) in a final volume of 20μl.

Binding competion assays

Competition assays were done in Tat binding buffer. A probe concentration (about 0.1ng per reaction) at which greater than 90% of the probe was bound by 50ng of Tat or by 200ng of partially purified BBP extract was used. Reactions were set up such that the probes were preincubated for 8mins with a range of concentrations of Tat from 5ng to 500ng per 20μl reaction. 200ng of BBP extract was then added to each reaction and after a further 20 minutes the shifted products were analysed in an 8% non-denaturing polyacrylamide gel.

Factor Xa cleavage of BBP

Semi-purified BBP was digested with 1μg of Factor Xa (Promega) in a total volume of 20μl of RNA binding buffer, at room temperature for 15 minutes. WT TAR probe (0.2ng) was then added and incubation continued for a further 15 minutes. In the converse experiment semi-purified BBP and WT TAR probe were incubated in RNA binding buffer for 15 minutes at room temperature and then 1μg of Factor Xa was added for a further 15 minutes.

RESULTS

Identification of TAR RNA binding factors in Hela cell nuclear extracts

Nuclear extracts prepared by the method of Dignam *et al* [40], were tested for the presence of TAR binding products in an electrophoretic mobility shift assay (EMSA), against different TAR related and unrelated RNAs, (Figure 1). As shown in Figure 2, a number of proteins bound to TAR but only two of these, labelled LBF and BBP showed any specificity. LBF failed to bind

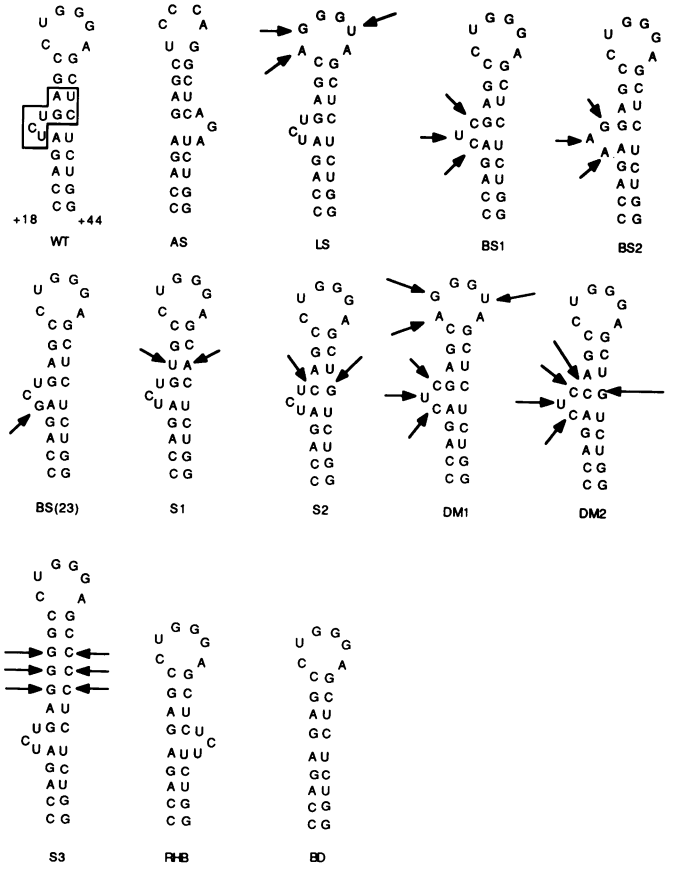


Figure 1. Structures of TAR RNAs. Sequences of WT and mutant TAR RNAs used in gel mobility shift assays. Arrows indicate base changes in comparison to the wild type molecule. Nucleotide numbers correspond to +1 in HIV-1 transcripts. Lettering beneath RNA structure indicates mutation as described in text. The key nucleotides for Tat binding to WT TAR are boxed.

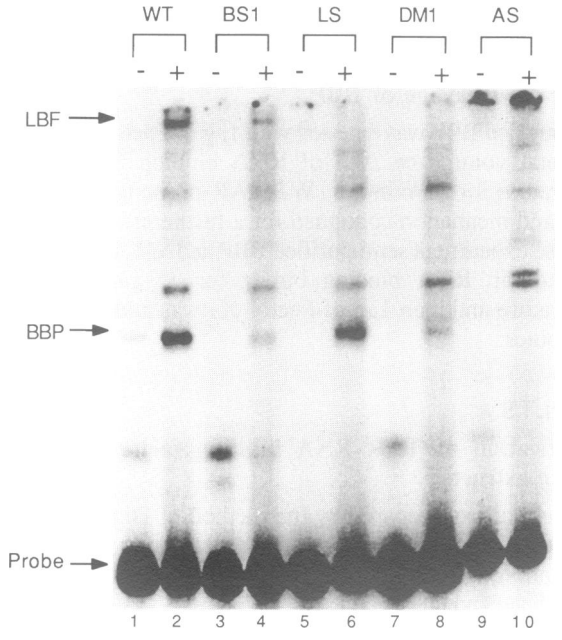


Figure 2. Identification of two TAR specific binding proteins in HeLa cell nuclear extracts. A mobility shift assay using whole nuclear extract with wild type (WT), antisense (AS) and mutant (BS1, LS, DM1) TAR probes is shown. BBP refers to a band shift that displays a specificity for the bulge region of TAR, and LBF refers to a band that displays a specificity for the loop region of TAR.

to antisense TAR RNA (Figure 2, track 10) and showed little binding to TAR RNA that had been mutated in the loop sequences (LS) (Figure 2, track 6). Binding was not significantly reduced by a 3bp substitution mutation in the bulge (Figure 2, track 4). This indicated that it recognised the loop sequences of TAR and consequently we referred to it initially as loop binding factor (LBF). BBP showed no binding to antisense TAR (Figure 2, track 10) and showed markedly reduced binding to TAR RNA that had a 3 base substitution in the bulge (BS1) (Figure 2, track 4). Binding was not reduced by the mutation in the loop (Figure 2, track 6). These data indicated that the protein was specific for the bulge region of TAR and we therefore refer to it as bulge binding protein (BBP). No other TAR specific proteins were obvious in these whole nuclear extracts. We noted that the properties of BBP varied between extracts. In some cases, BBP appeared to have reduced specificity and in others the levels were extremely low (unpublished data). The full specificity was always evident when excess poly I. C. was included in the binding assays presumably because non-specific interactions were eliminated. These data indicate that levels of 'active' BBP can be very variable, which might explain why BBP has not been observed previously.

Dignam extracts were subjected to a preliminary fractionation by binding to heparin agarose and elution with increasing concentration of KCl (Figure 3). We noted that under these conditions BBP appeared in the 500mM fraction and LBF appeared in the 200mM fraction (Figure 3a.). The elution profile and the size of the complex confirmed from the mobility suggest that LBF is in fact TRP1/p185, and UV cross linking confirmed a molecular weight of about 200kD (data not shown).

In addition, another TAR specific protein was detected in the 300mM eluate (Figure 3b). This protein failed to bind antisense

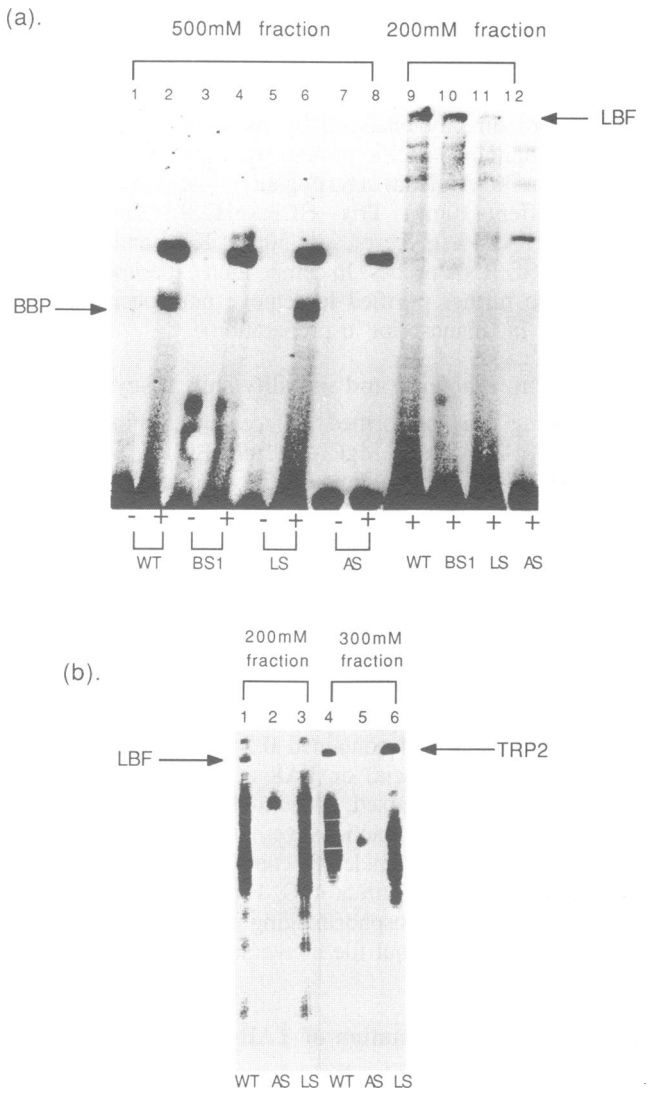


Figure 3. Separation of TAR RNA binding proteins by heparin agarose fractionation. Mobility shift assays with three KCl elution fractions (500mM, 200mM and 300mM) from heparin agarose are shown. **a.** Fractions from 200 and 500mM KCl eluates were tested against wild type (WT, lane 2 and 9), antisense (AS, lane 8 and 12) and mutant (LS, lane 6 and 11; BS1, lane 4 and 10) TAR probes. The positions of LBF and BBP are indicated. **b.** Fractions from 200 and 300mM KCl eluates were tested (5% native gel) against wild type (WT, lanes 1 and 4), antisense (AS, lanes 2 and 5) and loop substitution (LS, lanes 3 and 6) TAR probes. The position of LBF and the possible TRP2 protein are indicated.

TAR but still bound TAR with a loop mutation, indicating that it was TAR specific but loop independent. It is possible that this second potential bulge binding factor is TRP2 given its apparent high molecular weight [31]. As BBP appeared to be a previously undescribed protein, we analyzed it further. The protein was UV cross linked to radio-labelled TAR in a standard binding reaction and after treatment of the complex with RNase the protein was sized by SDS−PAGE. As shown in Figure 4 the approximate molecular weight of BBP is 38kD.

BBP binding is dependent upon a 5'-tripyrimidine bulge

There are numerous general RNA binding proteins in the cell which recognise either single stranded or double stranded RNA

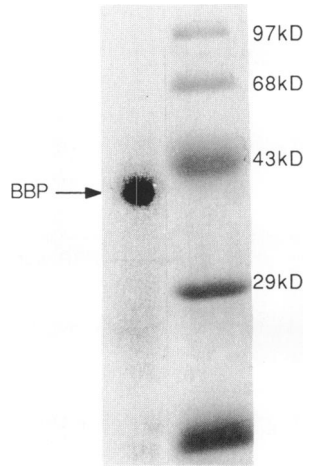


Figure 4. Molecular weight determination of BBP. SDS/PAGE analysis of the RNA–protein complexes after covalent coupling induced by UV is shown. The molecular weight was assessed relative to prestained, molecular weight markers as shown.

with no other unique sequence or conformational features. We wished to establish the extent of the specificity of BBP for TAR. As shown in Figure 5a, the BBP/TAR complex could be disrupted by competition with a 10–100 fold excess of unlabelled TAR (lanes 9,10). However a 1000 fold excess of single stranded RNA could not disrupt the complex (lane 6). Furthermore, when binding reactions were conducted under buffer conditions that were optimal for hnRNP binding to single stranded RNA [42], no competition was observed (Figure 5a, lane 8). These data suggest that BBP is not a single stranded RNA binding protein. To determine the importance of the bulge we next tested the ability of BBP to bind to TAR RNA that had the bulge deleted (BD) or which had the bulge deleted and then a tripyrimidine bulge inserted at the same location, but 3′ to the loop (RHB). As shown in Figure 5b., there was no detectable BBP binding to either the RHB or BD mutated TAR RNAs. This, together with the fact that there is substantially reduced binding to an antisense (AS) TAR, indicate that BBP is unlikely to be a general double stranded RNA binding protein such as TRBP1 [34]. Furthermore, the requirement for the bulge was not simply a general requirement for a distorted A-form helix, nor just a sequence specific requirement for three unpaired pyrimidines. It appears that BBP binding required the correct tripyrimidine sequence in the bulge in the correct sequence context.

Optimum BBP binding requires a conserved configuration of the upper stem

The ability of BBP to bind to a range of mutated TAR RNAs was tested. As shown in Figure 5c., mutation at the G26.C39 pair (S2) markedly reduced binding either alone (lane 3) or in combination with the bulge substitution (DM2, lane 2 and 10). Similarly a three base G.C extension of the upper stem alone (S3, lane 11) also virtually abolished BBP binding. We also confirmed the importance of the sequence of the bulge. A substitution of the bulge with purines (BS2, lane 7) had the same deleterious effect as the pyrimidine substitution (BS1, lane 6). These data confirm that in order for BBP to bind to TAR, the sequence and location of the tripyrimidine bulge must be

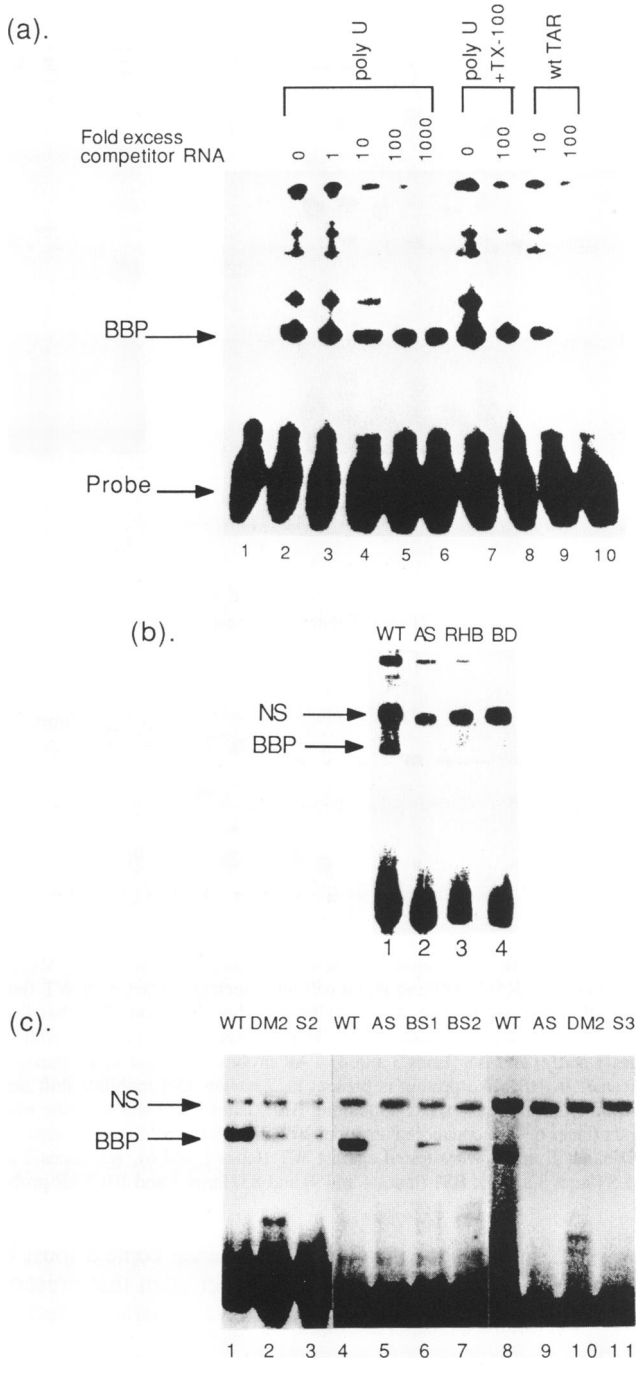
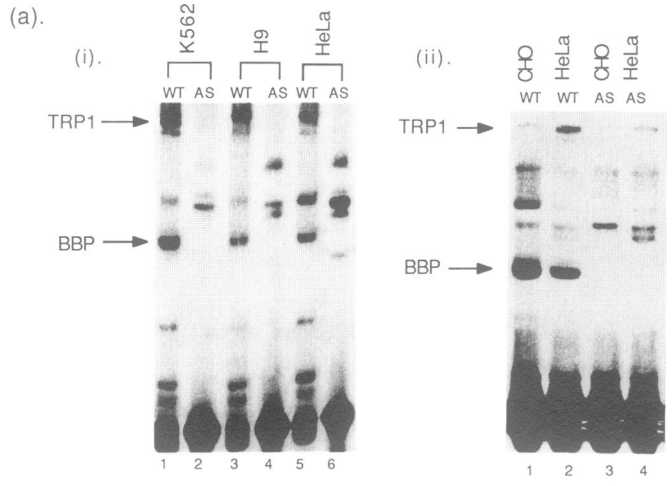


Figure 5. a. BBP is not a single-stranded RNA binding protein. A mobility shift assay showing the effect of poly (U) on BBP binding to WT TAR probe is shown. Probe alone (lane 1), BBP alone (lane 2) or in the presence of 1 (lane 3), 10 (lane 4), 100 (lane 5) or 1000 (lane 6) fold excess of poly (U) under optimal BBP binding conditions. BBP alone (lane 7) or in the presence of 100 fold excess (lane 8) of poly (U) under optimal binding conditions for hnRNPs. BBP in the presence of 10 (lane 9) or 100 (lane 10) fold excess of WT TAR RNA under optimal BBP binding conditions. **b.** BBP binding is dependent on a 5′ tripyrimidine bulge. A mobility shift assay with HeLa nuclear extract to determine the specificity of BBP binding to WT (lane 1), AS (lane 2), RHB (lane 3) and BD (lane 4) TAR probes is shown. **c.** Optimum BBP binding requires sequences in the upper stem. Mobility shift assays using heparin agarose purified BBP with wild type and various mutant TAR probes. A prominent nonspecific band (NS) is indicated and can be used as an internal standard to compare the levels of BBP binding with different probes. BBP binding to wild type (WT, lanes 1, 4, and 8) , antisense (AS, lanes 5 and 9), DM2 (lanes 2 and 10), S2 (lane 3), BS1 (lane 6), BS2 (lane 7) and S3 (lane 11) is shown.



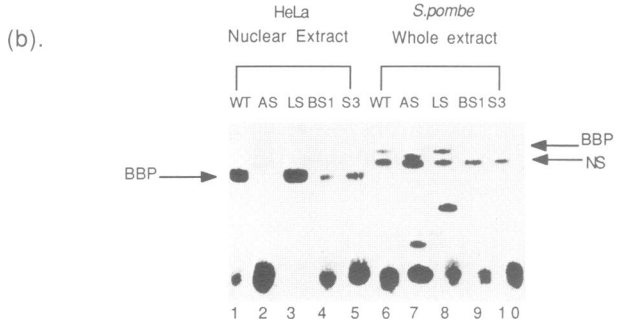


Figure 6. a. A BBP-like protein is present in human and rodent cells. i) Mobility shift assay with K562, H9 and HeLa cell total nuclear extract with WT (lanes 1, 3, and 5) and AS (lanes 2, 4, and 6) TAR probes. BBP and TRP1 bands are indicated. ii) Mobility shift assay with CHO and HeLa nuclear extracts with WT (lanes 1 and 2) and AS (lanes 3 and 4) TAR probes. BBP and TRP1 bands are indicated. **b.** BBP-like protein is present in *S.pombe*. Gel mobility shift assay showing HeLa nuclear extract containing BBP (lanes 1–5) and *S.pombe* whole extract (lanes 6–10) containing an upper BBP-like band and a lower nonspecific (NS) band. Extracts were tested against WT (lanes 1 and 6), AS (lanes 2 and 7), LS (lanes 3 and 8), BS1 (lanes 4 and 9) and S3 (lanes 5 and 10) TAR probes.

conserved. In addition, the flanking sequence context must be preserved, as subtle mutations in the upper stem that preserve base pairing but change the sequence have a marked effect on binding.

Species distribution of BBP-like proteins

Nuclear protein extracts prepared from a number of different mammalian cells were tested for the presence of BBP-like activity. As shown in Figure 6a., BBP was found in K562 (human erythroid), H9 (human T cell), Hela (human epithelial carcinoma) and in CHO (Chinese hamster ovary) cells. This latter observation contrasts with TRBP1 which is present at very low levels in rodent cells [29]. TRP1 (LBF) was detected in all three human cell lines but was not observed in CHO cells consistent with a previous report [31]. Some additional TAR binding proteins were seen in CHO cells, although these were not consistently observed, and the TAR specificity has not been characterised further. A BBP-like protein giving a identical TAR RNA mobility shift was also found in nuclear extracts of Xenopus oocytes (data not shown).

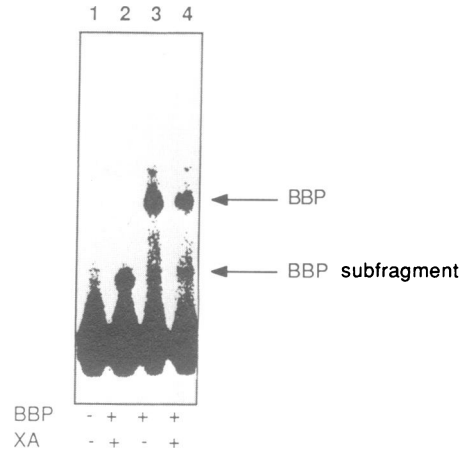


Figure 7. Cleavage of BBP with Factor Xa liberates an RNA binding fragment. Gel mobility shift assay showing binding of partially purified BBP before and after factor Xa digestion. WT TAR probe alone in the absence (−) of BBP (lane 1). WT TAR in the presence (+) of BBP (lane 3). WT TAR probe added to BBP after digestion with Factor Xa with appearance of smaller RNA binding fragment (lane 2). WT TAR probe added to BBP before digestion with Factor Xa (lane 4).

Extracts were also prepared from the two unrelated yeasts *S.pombe* and *S.cerevisiae*. These have both been shown to carry out many analogous processes to those of mammalian cells, although Tat does not appear to be active in either yeast [43, 44]. As shown in Figure 6b. a protein with BBP like character was detected in *S.pombe*. This protein migrated slightly slower than the Hela cell BBP, suggestive of a molecular weight greater than 38kD. The protein however displayed the same marked specificity for TAR RNA as Hela cell BBP as little binding was detected to the AS, BS1, S3 (Figure 6b.) or RHB (data not shown) TAR RNAs. No candidate for a TRP1 like protein was detected. TAR binding proteins were also detected in *S.cerevisiae* (data not shown), however none of these gave mobility shifts in the size range of either TRP1 or BBP and none of them distinguished between the TAR mutants, implying that they were general double stranded RNA binding proteins. These data show that many, but not all, eukaryotic cells contain a BBP-like protein.

Factor Xa cleaves BBP to produce a small RNA binding fragment

The size and binding characteristics of BBP appeared to distinguish it from other RNA binding proteins. In addition, we observed that the protein was susceptible to cleavage by the highly specific protease Factor Xa [45]. Partially purified (E.Vives, manuscript in preparation) BBP was treated with Factor Xa either before or after binding to TAR RNA. As shown in Figure 7, treatment before binding abolished the normal shifted complex and generated a new, much smaller protein RNA complex. Qualitatively the same result was obtained when a preformed TAR/BBP complex was treated, except that the cleavage was less efficient (Figure 7, lane 4). This suggests that the Xa cleavage site is less accessible in the bound complex, implying either that the protein conformation alters upon binding or that the TAR RNA masks the cleavage site. The data also suggest that it will be possible to produce a small, minimal TAR RNA binding protein.

A search of the SWISSPROT DATA has revealed no

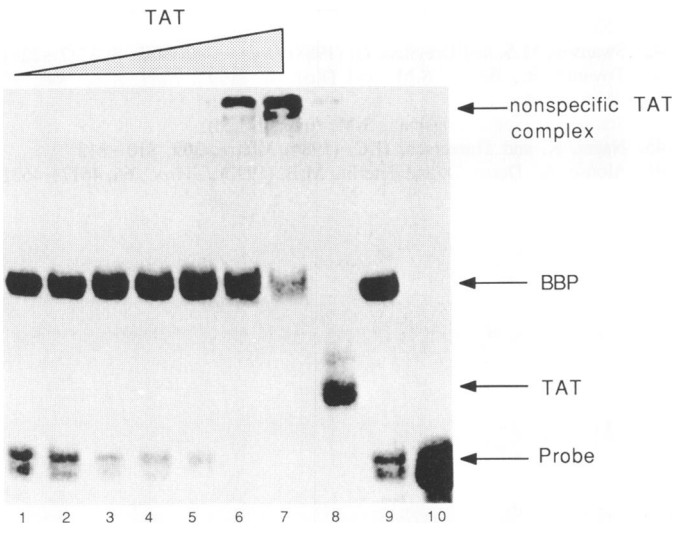


Figure 8. Competition between BBP and Tat for binding to TAR RNA *in vitro*. 0.1ng of WT TAR (lane 10) was incubated in Tat binding conditions with 200ng of partially purified BBP alone (lane 9), 50ng of Tat alone (lane 8) or with 200ng of partially purified BBP with increasing amounts of Tat, 5ng (lane 1), 10ng (lane 2), 20ng (lane 3), 50ng (lane 4), 100ng (lane 5), 250ng (lane 6) and 500ng (lane 7).

previously characterised RNA binding proteins containing the consensus (IEGR) cleavage recognition site. In fact, the site is rare, being found on less than 500 proteins in a database comprising more than 35,000 sequences. The only TAR RNA binding protein for which a clone is available is TRBP1 [29], and this lacks a Factor Xa cleavage site.

BBP is an effective competitive inhibitor of Tat binding to TAR *in vitro*

Given that the TAR binding sites for BBP and Tat are overlapping, we tested whether the proteins would compete for binding to TAR. Partially purified BBP was mixed with increasing amounts of Tat and then limiting TAR RNA was added. 200ng of semi-purified BBP and 50ng of pure Tat alone gave equivalent mobility shifts (Figure 8, lanes 9 and 8 respectively), and when these concentrations of Tat and BBP were mixed (lane 4), BBP binding predominated. This was true until much higher concentrations of Tat were used (lanes 6 and 7). At these concentrations Tat binds non-specifically to TAR to produce high molecular weight complexes. The presence of these complexes confirm that significant amounts of Tat are indeed present in the reaction, but that specific binding to TAR is inhibited by BBP. The same result was obtained if Tat was premixed with TAR, and BBP was added subsequently (data not shown). Although these data are preliminary in that the definition of association and dissociation constants for BBP compared to Tat await the availability of a homogenous BBP preparation, they provide a strong indication that BBP can function as a specific inhibitor of Tat binding *in vitro*. It remains to be seen if this is the case *in vivo*.

DISCUSSION

We have identified three proteins in Hela cell extracts that bind to HIV-1 TAR RNA in a sequence dependent fashion. One of these, that we referred to initially as loop binding factor (LBF) has a high molecular weight (ca. 200kD) and requires the loop sequence for binding. This strongly suggests that the protein is TRP1/p185 [31, 32]. A second high molecular weight protein co-migrated with TRP1 when whole nuclear extract was used in a mobility shift assay but was seperated from TRP1 by heparin agarose fractionation. This protein did not require the loop sequence for binding and was probably TRP2 [31, 32]. The third protein appeared to be unrelated to any previously identified TAR RNA binding proteins. It has an approximate molecular weight of 38 kD and shows marked specificity for binding to the upper bulge stem region of TAR, and therefore we have called it bulge binding protein (BBP). BBP binding appears to be most dependent upon specific sequences in TAR rather than on overall structure of a bulged stem-loop configuration. This was indicated by the fact that substitution mutations in the bulge and upper stem dramatically reduced binding. This sequence specificity is reminiscent of the Tat/TAR interaction and suggests that RNA/protein recognition via Watson–Crick base pairs exposed in a distorted A-form helix may prove to be a common strategy.

BBP appears to be conserved across a number of species, being found in human and rodent cells and in *Xenopus* oocytes. In addition, a slightly larger protein with the same binding characteristics was also found in *S.pombe* but *S.cerevisiae* appeared to contain no TAR specific proteins. The distribution of BBP is not correlated with Tat activation. Activation is inefficient in rodent cells [46] and not observed in either of the yeasts [43, 44]. In contrast, the loop binding protein showed a more restricted distribution, there being no candidate proteins found in either of the yeasts, or in CHO cells, confirming previous reports of lack of TRP1 in *S.cerevisiae* and CHO cells [31]. Cells that lack TRP1 fail to show significant Tat activity, suggesting a role for TRP1 in Tat activation.

At present we have no evidence that BBP plays any role in Tat activation *in vivo*. It may simply be fortuitous that we have detected a protein with a preference for a sequence specific bulge-stem loop. However, it is intriguing that defined nucleotides in TAR such as G26 and the pyrimidines in the bulge are recognised and that these are precisely those features recognised by Tat. Either this is a generic RNA binding motif or TAR has been selected to interact with specific cellular factors in addition to Tat.

The normal biological function of such a cellular factor is as yet unpredicted. The apparent conservation of the factors suggests that it will have an important role in the cell, although the absence of the factor in *S.cerevisiae* might indicate that it will not play an essential 'house-keeping' role. One possible function for proteins such as BBP during HIV infection would be as inhibitors of Tat action as we have suggested previously [27]. It is possible that the availability of TAR for Tat binding might be regulated by the intranuclear levels of TAR RNA binding proteins and this might contribute to determining the latent or active state.

In conclusion, we have identified a novel 38 kD protein which binds specifically to the bulge-upper stem region of HIV-1 TAR. Further characterisation of this protein should provide insight into the molecular basis of RNA–protein interactions that are mediated by specific recognition of Watson–Crick base pairs in a distorted A-form helix. It will also be of interest to determine the normal cellular function of BBP and to define its natural target RNA. In addition, it may be possible to define a minimal TAR RNA binding fragment of BBP that could function as an effective inhibitor of HIV replication.

ACKNOWLEDGEMENTS

This research was funded by grants from Glaxo, MRC AIDS directed programme and SERC. M.Braddock is a Royal Society Research Fellow. W.Nacken was supported by the Deutsche Krebsforschungs-zentrum.

REFERENCES

1. Cullen, B.R. (1992) *Microb. Rev.*, **56**, 375–394.
2. Cullen, B.R. (1993) *Cell*, **73**, 417–420.
3. Kashanchi, F., Piras, G., Radonovich, M.F., Duvall, J.F., Fattaey, A., Chiang, C.-M., Roeder, R.G.and Brady, J.N. (1994) *Nature*, **367**, 295–296.
4. Jeang, K.-T., Chun, R., Lin, N.H., Gatignol, A., Glabe, C.G.and Fan, H. (1993) *J. Virol.*, **67**, 6224–6233.
5. Desai, K., Loewenstein, P.M.and Green, M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8875–8879.
6. Ohana, B., Moore, P.A., Ruben, S.M., Southgate, C.D., Green, M.R.and Rosen, C.A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 138–142.
7. Braddock, M., Chambers, W., Wilson, W., Esnouf, M.P., Adams, S.E., Kingsman, A.J. and Kingsman, S.M. (1989) *Cell*, **58**, 269–279.
8. Cullen, B.R. and Greene, W.C. (1989) *Cell*, **58**, 423–426.
9. Jaeger, J.A. and Tinoco, I.J. (1993) *Biochemistry*, **32**, 12522–12530.
10. Michnicka, M.J., Harper, W.and King, G.C. (1993) *Biochemistry*, **32**, 395–400.
11. Colvin, R.A. and Garcia-Blanco, M.A. (1992) *J. Virol.*, **66**, 930–935.
12. Riordan, F.A., Bhattacharyya, A., McAteer, S. and Lilley, D.M.J. (1992) *J. Mol. Biol.*, **226**, 305–310.
13. Weeks, K.M. and Crothers, D.M. (1991) *Cell*, **66**, 577–588.
14. Gait, M.J. and Karn, J. (1993) *TIBS*, **18**, 255–259.
15. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D.and Williamson, J.R. (1992) *Science*, **275**, 76.
16. Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Kam, J., Lowe, A.D., Singh, M. and Skinner, M.A. (1990) *EMBO*, **9**, 4145–4153.
17. Churcher, M.J., Lamont, C., Hamy, F., Dingwall, C., Green, S.M., Lowe, A.D., Butler, P.J.G., Gait, M.J. and Karn, J. (1993) *J. Mol. Biol.*, **230**, 90–110.
18. Delling, U., Reid, L.S., Barnett, R.W., Ma, M.Y.-X., Climie, S., Sumner-Smith, M. and Sonenberg, N. (1992) *J. Virol.*, **66**, 3018–3025.
19. Blanchard, A.D., Powell, R., Braddock, M., Kingsman, A.J.and Kingsman, S.M. (1992) *J. Virol.*, **66**, 6769–6772.
20. Frankel, A.D. (1992) *Protein Science*, **1**, 1539–1542.
21. Hamy, F., Asseline, U., Grasby, J., Shigenori, I., Pritchard, C., Slim, G., Butler, **64**, 1402–1406.
26. Madore, S.J. and Cullen, B.R. (1993) *J. Virol.*, **67**, 3703–3711.
27. Braddock, M., Powell, R., Blanchard, A.J., Kingsman, A.J.and Kingsman, S.M. (1993) *FASEB J.*, **7**, 214–222.
28. Gunnery, S., Greeen, S.R.and Mathews, M.B. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11556–11561.
29. Gatignol, A., Buckler-White, A., Berkhout, B.and Jeang, K.-T. (1991) *Science*, **251**, 1597–1600.
30. Rounsville, M.P. and Kuma, A. (1992) *J. Virol.*, **66**, 1688–1694.
31. Sheline, T., Milocco, L.H.and Jones, K.A. (1991) *Genes Dev.*, **5**, 2508–2520.
32. Wu, F.J., Garcia, D., Sigman, D.and Gaynor, R. (1991) *Genes Dev.*, **5**, 2128–2140.
33. Marciniak, R.A., Garcia-Blanc, M.A.and Sharp, P.A. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3624–3628.
34. Gatignol, A., Buckler, C.and Jeang, K.-T. (1993) *Molec. Cell. Biol.*, **13**, 2193–2202.
35. Masuda, T. and Harada, S. (1993) *Virol.*, **192**, 696–700.
36. Carrier, F., Gatignol, A., Hollander, M.C.and Jeang, K.-T. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 1554–1558.
37. Melton, D.A., Krieg, P.A., Rebagliatti, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucl. Acid. Res*, **12**, 7035–7036.
38. Sanger, F., Nicklen, S.and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
39. Mellor, J., Dobson, J.J., Roberts, N.A., Tuite, M.F., Emtage, J.S., White, S., Lowe, P.A., Patel, T., Kingsman, A.J. and Kingsman, S.M. (1983) *Gene*, **24**, 1–14.
40. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucl. Acid. Res*, **11**, 1475–1489.
41. Johnston, R.F., Pickett, S.C. and Barker, D.L. (1990) *Electrophoresis*, **11**, 353–360.
42. Swanson, M.S. and Dreyfuss, G. (1988) *Molec. Cell. Biol.*, **8**, 2237–2241.
43. Toyama, R., Bende, S.M. and Dhar, R. (1992) *Nucl. Acid. Res*, **20**, 2591–2596.
44. Stanway, C. and Kingsman, S.M. (unpublished).
45. Nagai, K. and Thøgersen, H.C. (1984) *Nature*, **309**, 810–812.
46. Alonso, A., Derse, D. and Peterlin, M.B. (1992) *J. Virol.*, **66**, 4617–4621.