The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site

Mark P.Ashe, Louise H.Pearson and Nick J.Proudfoot¹

Sir William Dunn School of Pathology, Oxford University, South Parks Road, Oxford OX1 3RE, UK

¹Corresponding author

The inactivity of the 5' long terminal repeat (LTR) poly(A) site immediately downstream of the cap site maximizes the production of HIV-1 transcripts. In this paper, we demonstrate that this inactivity is mediated by the interaction of the U1 snRNP with the major splice donor site (MSD). The inhibition of the HIV-1 poly(A) site by U1 snRNP relies on a series of delicately balanced RNA processing signals. These include the poly(A) site, the major splice donor site and the splice acceptor sites. The inherent efficiency of the HIV-1 poly(A) site allows maximal activity where there is no donor site (in the 3' LTR) but full inhibition by the downstream MSD (in the 5' LTR). The MSD must interact efficiently with U1 snRNP to completely inhibit the 5' LTR poly(A) site, whereas the splice acceptor sites are inefficient, allowing full-length genomic RNA production.

Keywords: HIV/poly(A)/RNA processing/U1 snRNP

Introduction

Most eukaryotic RNA polymerase II transcripts are processed at their 5' and 3' ends by capping and cleavage/polyadenylation respectively. Both these processing reactions are thought to occur on nascent transcripts (Nevins and Darnell, 1978; Rasmussen and Lis, 1993). In addition, introns may be removed by either co- or post-transcriptional splicing (Nevins and Darnell, 1978; Beyer and Osheim, 1988; LeMaire and Thummel, 1990; Baurén and Wieslander, 1994; Zhang *et al.*, 1996). The 5' and 3' ends of introns are defined by *cis*-acting signals, the splice donor (SD) and acceptor (SA) sites respectively. These signals are recognized by a complex of proteins associated with five snRNA molecules which perform the splicing reaction. In particular, an early complex committed to splicing contains the U1 snRNP, which interacts at the SD (for reviews see Lamond, 1993; Sharp, 1994). The majority of mRNA 3' ends are formed by endonucleolytic cleavage followed by the addition of a poly(A) tail. Minimal *cis*-acting signals have been defined for mammalian poly(A) sites and consist of an AAUAAA motif with a downstream GU/U-rich element (Levitt *et al.*, 1989). A complex set of protein factors recognize these *cis*-acting poly(A) signals and drive the processing reaction (reviewed by Keller, 1995; Manley, 1995).

The HIV-1 primary transcript contains two identical

minimal poly(A) signals. The first is situated close to the 5' end of all viral RNAs and the second is found downstream of the viral genes (see Figure 1) (reviewed in Proudfoot, 1991). In HIV-1 proviral clones, the downstream poly(A) site is used to produce the viral transcripts. To maximize their production the poly(A) site near the cap is inactive. However, use of this HIV-1 poly(A) site is possible if it is replaced by a more efficient poly(A) site or if the downstream major splice donor (MSD) site is mutated (Ashe *et al.*, 1995). This latter observation is particularly interesting as it suggests a connection between the splicing and polyadenylation machinery.

Many studies have found connections between components of the capping, splicing and polyadenylation processes. For example, a capped transcript displays enhanced splicing of its first intron (Konarska *et al.*, 1984; Ohno *et al.*, 1987; Inoue *et al.*, 1989; Izaurralde *et al.*, 1994) and can also exhibit enhanced polyadenylation (Cooke and Alwine, 1996). In addition, splicing or splicing factors can either activate or inhibit the use of poly(A) sites (Levitt *et al.*, 1989; Niwa *et al.*, 1990; Boelens *et al.*, 1993; Liu and Mertz, 1993; Wassarman and Steitz, 1993; Furth *et al.*, 1994; Lutz and Alwine, 1994; Peterson *et al.*, 1994; Lou *et al.*, 1996; Scott and Imperiale, 1996). Furthermore, it has been shown that in multi-intronic transcripts the splicing of one intron influences the splicing of others (Nesic and Maquat, 1994). Therefore, it is conceivable that for any given transcript, interconnections occur between these various RNA processing reactions.

The range of RNA processing reactions for HIV-1 is remarkably complex (Figure 1), with more than 40 different mRNA species (including full-length unspliced RNA) being derived from a single transcription unit (Purcell and Martin, 1993 and references therein). The production of these mRNAs is temporally regulated with early viral proteins controlling both the amount and species of RNA. Specifically, *tat* increases the levels of RNA while *rev* facilitates nucleocytoplasmic transport of unspliced and partially spliced transcripts as well as inhibiting *tat/rev* intron splicing (reviewed by Kingsman and Kingsman, 1996).

This study focuses on the role of the MSD in the inhibition of the 5' LTR poly(A) site. The MSD is 195 nt downstream of the poly(A) site and its sequence conforms to the consensus at eight out of nine positions. It is the only splicing signal which is invariably used whenever the HIV-1 transcript is spliced (see Figure 1). The production of unspliced genomic and gag/pol RNAs relies more upon the weakness of the various acceptor sites, in combination with *rev* activity, than on any limitation to MSD function (Amendt *et al.*, 1994; O'Reilly *et al.*, 1995).

The experiments presented here analyse the contributions of the 5' LTR poly(A) site, the MSD and SAs to the inhibition of polyadenylation near the cap. A range of

Fig. 1. Splicing and poly(A) site use for HIV-1 transcripts. A diagram is shown representing an HIV-1 integrated provirus (top), nascent transcription (middle) and the possible alternative splicing events (bottom). HIV-1 LTRs and the site of transcription initiation are indicated on the provirus. The poly(A) (pA), splice donor ($\diamond$) and splice acceptor sites ($\triangle$) are depicted on the nascent transcript. The gene products of the alternatively spliced mRNAs have been used to label them and the approximate size of each is shown.

mutations to the MSD are all shown to relieve the 5' LTR poly(A) site inactivation. This inhibition of the poly(A) site relies on the U1 snRNP-SD interaction as a mutant U1 snRNA, engineered to bind near the mutated MSD, rescues the inhibition of the poly(A) site. A mutational analysis reveals that the HIV-1 poly(A) site is perfectly balanced, allowing maximal activity in the absence of a donor site and minimal activity in its presence. Replacement of the weak HIV-1 SAs with a strong SA gives constitutive inhibition of the poly(A) site by activating cryptic SDs which bypass the mutated MSD. This analysis emphasizes the fact that the HIV-1 RNA processing signals have evolved to allow maximal levels and diversity of transcripts.

Results

The experimental system

In order to study HIV-1 poly(A) site use, we have previously described an experimental system which faithfully reproduces HIV-1 infection, generating viral transcripts and infectious viral particles. Proviral clones were used to transiently transfect HeLa cells and cytoplasmic RNA was harvested and mapped using RNase protection analysis (Ashe *et al.*, 1995). In the course of these studies transcripts were detected from both HIV-1 LTR promoters. Therefore, to facilitate analysis of viral transcripts, the 3' LTR promoter was inactivated without affecting the production of viral transcripts from the 5' LTR promoter (Ashe *et al.*, 1995). The resulting plasmid pHAPB has been used throughout this study.

Splice donor mutations

Previously we have shown that a MSD mutation in pHAPB activates the 5' LTR poly(A) site (Ashe *et al.*, 1995). However, in these experiments splicing across the MSD was not analysed. To examine both poly(A) site and MSD

use, a riboprobe was designed which is complementary to the entire region from upstream of the HIV-1 promoter to downstream of the MSD. Figure 2A schematically shows this riboprobe and the RNase-protected fragments detected in assays using cytoplasmic RNA from HeLa cells transiently transfected with a variety of pHAPB-based plasmids. These are shown in Figure 2A and include three MSD mutants (pHAPBM1, pHAPBM2 and pHAPBM1,2) and a poly(A) site mutant (pHAPB5'S) where the HIV-1 poly(A) site has been exchanged for the efficient synthetic poly(A) site (SPA) (Levitt *et al.*, 1989).

Figure 2B shows the RNase protection analysis for these constructs (lanes 1–5). The protected fragments generated using pHAPB are shown in lane 2. The absence of a C-pA product confirms the conclusion that the 5' LTR poly(A) site is inactive. The two species which are detected correspond to transcripts which either use (C-S) or readthrough (C-RT) the MSD. In the absence of co-transfected tat (lane 7) the pattern and levels of protected fragments are identical showing that endogenous viral tat is produced and active.

The pHAPBM1 mutant (previously pHAPBMuMSD; Ashe *et al.*, 1995) has the critical GU (position +1 and +2) of the MSD altered to CA (see Figure 2A). The analysis using this mutant is shown in lane 3. As previously shown, this mutation leads to the production of a C-pA short transcript through activation of the poly(A) site in the 5' LTR. Two additional MSD mutants, pHAPBM2 and pHAPBM1,2 are also shown (Figure 2A). pHAPBM2 has a G to A change at position +5 which is known to inhibit SD activity (Zhuang and Weiner, 1986), while pHAPBM1,2 is a combination of the M1 and M2 mutations. The analysis of these new constructs (Figure 2B, lanes 4 and 5) shows a substantial activation of poly(A) site use similar to that seen for the original mutant. Therefore, we have extended the analysis of the MSD to show that three different MSD mutations all give activation of the

Fig. 2. Mutations to the MSD inhibit the promoter proximal poly(A) site. (A) The figure represents the 5' LTR and MSD of HIV-1 constructs used in this study. The replacement of the HIV-1 poly(A) site with SPA is depicted. The sequence of the MSD and M1, M2 and M1,2 mutations are expanded (mutations in bold). The cryptic SDs activated by the MSD mutations are also shown. The riboprobe is shown as a thick line and protected fragments are represented by thinner lines below. The abbreviations in fragment labelling are cap (C), poly(A) site (pA), MSD (S), cryptic SDs (C1/C2) and readthrough (RT). (B) Figure showing an RNase protection analysis using cytoplasmic RNA from transiently transfected HeLa cells. The constructs used are labelled above each lane and co-transfection of a tat plasmid (Adams *et al.*, 1988) is depicted by +. The protected fragments in (A) are labelled down the side of each panel. SPA generates a broad set of C-pA products due to overdigestion at sites of hybrid breathing (Weichs an der Glon *et al.*, 1993). The panel below shows an S1 nuclease protection assay measuring expression of the co-transfected human β globin gene which allows the effects of tat on HIV-1 promoter activity to be ascertained. The right panel shows a longer exposure of lanes 6–10. The signal obtained using the pHAPBM1,2 plasmid is consistently low throughout this whole series of experiments; the reason for this is unclear but may be due to a poor-quality plasmid preparation. This does not appear to be an effect of the mutation as other derivatives of M1,2 such as Δ G and AS (see Figures 3 and 7) give higher signals.

promoter-proximal poly(A) site. Furthermore, the M2 mutant analysis shows that even a single base change to a proviral construct can dramatically alter the pattern of 5' LTR poly(A) site use.

By the use of an extended riboprobe which covers the MSD, the effect of each of the MSD mutations on splicing can also be determined. The M1 mutation completely inactivates the MSD, but leads to the activation of two cryptic SDs, C1 and C2, which lie 4 nt and 50 nt downstream of the MSD, respectively (lane 3, data not shown). The C1 cryptic SD has been described before using a similar system (Purcell and Martin, 1993). The M2 mutation (lane 4) does not completely inactivate the MSD but still leads to the activation of the C2 cryptic SD. The C1 cryptic SD, however, is not activated—possibly because the nucleotide mutated would be the critical +1 G of the GU sequence for this cryptic SD. The combined mutation M1,2 (lane 5), as might be expected, completely inactivates both the MSD and the C1 SD, leaving only the C2 SD active. Cryptic SD sites have been described using many systems and are only used when the wild-type SD site is mutated (e.g. Wieringa *et al.*, 1983). These are very weak SDs in that they barely relate to the consensus SD site. It is assumed that they arise from a pressure to splice at or near the position of the wild-type donor site. This pressure is applied by other signals such as the SA and branch site and may be accounted for mechanistically by a network of protein

interactions connecting the 5' and 3' splice sites (Wu and Maniatis, 1993). Although HIV-1 has only weak 3' splicing signals which could apply this pressure (Amendt *et al.*, 1994; O'Reilly *et al.*, 1995), there are nine possible SAs which together might act to encourage use of the cryptic SDs. In addition, across the first exon a network of interactions may also link the cap to the first SD and this may contribute to the activation of cryptic SDs in this system (Lewis *et al.*, 1996; Colot *et al.*, 1996). When the level of spliced products [at the MSD (C-S) and cryptic SDs (C-C1, C-C2)] is compared with the unspliced products (C-RT), there is a small trend towards a decreased splicing/readthrough ratio with the more severe mutations (see Figure 2B, data not shown).

The viral tat protein, which *trans*-activates the HIV-1 promoter (reviewed in Jones and Peterlin, 1994) relies on splicing at the MSD for production of its mRNA (see Figure 1). Previously, co-transfected tat was required for the production of steady-state transcripts using the pHAPBM1 mutant. It was suggested that the mutation inactivates the MSD leading to no tat mRNA and therefore low HIV-1 promoter activity (Ashe *et al.*, 1995). However, in the more complete analysis shown here, steady-state transcripts are detected without added tat, although at greatly reduced levels (Figure 2B, lanes 8–10). In addition, as described above, the MSD mutations do not inactivate splicing but generate cryptic spliced mRNAs and these should still allow translation of tat protein. It is possible

that for the MSD mutants the activation of polyadenylation combined with the reduced splicing efficiency at cryptic sites is sufficient to decrease viral tat mRNA so that the activity of the HIV-1 promoter is dramatically decreased. In support of this, the poly(A) site mutant, pH Δ PB5'S, which gives efficient polyadenylation regardless of the MSD (Ashe *et al.*, 1995) causes a significant reduction in spliced RNA (C-S) in the presence of added tat (Figure 2B, compare lanes 1 and 2). This is reduced further when tat is not added (lane 6), suggesting that endogenous tat is limited by increased poly(A) site use. However, even in the absence of exogenous tat, the pattern of transcripts generated remains the same (Figure 2B, compare lanes 8–10 in the longer exposure with lanes 3–5). In particular, the MSD mutations all cause dramatic activation of the 5' LTR poly(A) site and the level of poly(A) site use is the same despite the absence of exogenous tat (data not shown).

It is possible that the MSD mutations also affect the accumulation of other viral proteins coded by spliced mRNAs. The viral protein rev increases the levels of unspliced and partially spliced mRNAs (the 9 kb and 4 kb transcripts) and may be relevant to the activation of polyadenylation seen when the MSD is mutated. Like tat, rev mRNA production relies on splicing at the MSD (see Figure 1). As readthrough products are detected in cytoplasmic RNA for all the mutant MSD constructs, rev activity must be present. In the absence of rev (using a proviral clone with a mutant rev start codon) only spliced mRNAs (C-S) are detected in cytoplasmic RNA (data not shown). In addition, when the tat and rev mRNAs were quantitated with riboprobes over the relevant SAs, rev mRNA was in great excess over tat mRNA (data not shown). Hence, although tat is limiting, rev may not be. However, the most compelling evidence that rev is not involved in poly(A) site activation is that the process can be recreated in mini-genes in the complete absence of rev (Figure 4).

The effect of viral particle formation

A potential problem when analysing splicing at the MSD is that levels of readthrough RNA may be under-represented due to packaging of these RNAs into viral particles. To test this possibility a mutation to the start codon and myristoylated glycine of the gag coding sequence was made to give the Δ G constructs (specifically the sequence AUGGGU was changed to GUGCAC). Gag protein is essential for particle formation and even mutations to the myristoylated glycine residue alone are known to abolish this process (Göttlinger *et al.*, 1989). Figure 3A shows a Western blot analysis, using p24 gag and gp120 env antibodies, on whole-cell lysates of transfected HeLa cells. Lane 1 shows a mock transfection which allows assignment of non-specific bands (*). Lanes 2 and 3 show the positive controls, pH (Ashe *et al.*, 1995) and pNL4-3 (Adachi *et al.*, 1986), where constructs containing the entire HIV-1 provirus were transfected and both p55 gag and gp120 env proteins are detected. A negative control pHSPA+ (the proviral clone has been inactivated by insertion of SPA; Ashe *et al.*, 1995) was used in lane 4. Lanes 5 and 6 show the effect of the gag mutation. pH Δ PB gives both p55 and gp120 proteins whereas pH Δ PB Δ G gives gp120 with no p55, but does give a small amount of protein

(p35) initiated at the first internal methionine. This p35 protein does not contain the N-terminal matrix domain of gag which is known to be essential for viral particle formation (Göttlinger *et al.*, 1989; Fäcke *et al.*, 1993; Spearman *et al.*, 1994).

Figure 3B shows an RNase protection assay assessing the effect of this gag mutation on the levels of unspliced RNA. A comparison of spliced and readthrough products for the pH Δ PB construct and each of the MSD mutants reveals that the levels of splicing are the same with and without gag. The only difference is that the C2 cryptic SD is mutated by the Δ G mutation and this leads to use of another cryptic site C3. Thus, the analysis of splicing using whole viral constructs reflects the authentic levels of HIV-1 RNA species and surprisingly viral particle formation accounts for an undetectable amount of full-length viral RNAs.

Mini-gene system and splice acceptor requirements

A mini-gene construct, pHmg, was made to allow easier analysis of HIV-1 poly(A) site regulation (Figure 4A). In this construct the first 2 kb of HIV-1 were fused to a region in the middle of the virus which contains five SAs. This was subsequently joined to part of the human γ -globin last exon containing the γ -globin poly(A) site. The analysis of mRNAs from pHmg and a construct where the MSD has been mutated, pHmgM1 (Figure 4B, lanes 1 and 2), shows that the pattern obtained is very similar to that observed previously (Figure 2B). Once again, mutation of the MSD activates the 5' LTR poly(A) site. One difference to previous results is that the levels of readthrough products for these constructs are much lower than for the proviral construct. This may be due to the absence of rev protein which is required for the nucleocytoplasmic transport of these unspliced products (Fischer *et al.*, 1995).

Analysis of the role of SAs in the inhibition of polyadenylation in the provirus is complicated by the presence of nine weak SAs which can all be used in conjunction with the MSD. Furthermore, the region containing the SAs also contains four SDs (see Figure 1). As the mini-gene described above only contains five of the SAs without SDs and reproduces the effects seen in the full proviral clones, we examined the role played by SAs in the regulation of the promoter-proximal poly(A) site using this simpler system. The construct pHmg2 was designed so that the whole γ -globin final exon is fused to the first 2 kb of HIV-1 (see Figure 4A). This allows the comparison of the five inefficient HIV-1 SAs in pHmg with the more efficient γ globin SA in pHmg2. Figure 4B, lanes 3 and 4, show that for pHmg2, a mutation of the MSD to give pHmg2M1 does not lead to activation of the 5' LTR poly(A) site. This mutation does, however, inactivate the MSD and activate the two cryptic SDs C1 and C2.

These results show that a strong SA appears to activate the cryptic SDs, allowing them to inhibit the poly(A) site in the 5' LTR. The mechanism of this activation may involve a network of protein interactions involving the branch site and SA (Wu and Maniatis, 1993), which activates the normally weak cryptic SDs. This further suggests that the precise sequence of the donor site is not important and that any SD of sufficient activity will inhibit poly(A) site use.

Fig. 3. Analysis of the effect of viral RNA packaging on levels of readthrough RNA. **(A)** Western blot analysis on whole-cell lysates from transiently transfected HeLa cells using p24 gag and gp120 env antibodies. Constructs used to transfect cells are labelled above and immunoreactive proteins are labelled to the right of the panel. * represents bands which are detected in mock and negative control samples which represent cross-reacting proteins from HeLa cells. These serve as controls for the level of protein loaded onto the gels. **(B)** Figure showing an RNase protection assay using the ΔG constructs. The labelling is similar to that in Figure 2B, except C-C3 represents an extra cryptic SD which is activated by the ΔG mutation (which affects the C2 SD). An artefactual band (*) above C-C1, which is also seen in mock-transfected controls (data not shown) was found in this analysis due to the use of RNase I in the RNase protection. This band is not seen when RNase A and T1 are used.

The specificity of the poly(A) site

The replacement of the HIV-1 poly(A) site with an efficient poly(A) site, SPA, gives constitutive poly(A) site use when the MSD is wild-type, whereas the HIV-1 poly(A) site is inactive unless the MSD is mutated (Figure 2B; Ashe *et al.*, 1995). To examine this difference in more detail the mini-gene constructs pHmg and pHmg5'S [where the HIV-1 poly(A) site is replaced with SPA] were used for transfection and RNase protection analysis. As before, mutation of the MSD gives substantial poly(A) site activation for pHmg (Figure 5B, lanes 1 and 2), whereas using the construct pHmg5'S, SPA is active regardless of the MSD mutation (Figure 5B, lanes 9 and 10). The low basal polyadenylation in the presence of the MSD allows the HIV-1 poly(A) site to be highly activated on MSD mutation, whereas SPA is activated to a much lower extent from an already high level of basal poly(A) site use.

Therefore, in order to analyse this difference between the two poly(A) sites, mutations were made in the poly(A) sequences of pHmg and pHmg5'S. Alignment of the two poly(A) sequences shows that they are equal in length and relatively similar in sequence, although two major differences are apparent (Figure 5A). First, the SPA downstream element has two GU regions separated by a tract of Us, whereas the HIV-1 poly(A) site has a C-rich tract between two GU regions. Furthermore, SPA has a sequence U₃AU₄ between the A₂UA₃ and the cleavage site, whereas the HIV-1 poly(A) site lacks this sequence. Both a downstream U-rich element and a U₃AU₄ element, are thought to increase the efficiency of the poly(A) site

(Proudfoot, 1991; A.Moreira, personal communication). Therefore, a series of mutations (Figure 5A) were constructed, where a U₃AU₄ (H1), a U-rich downstream element (H2) and a combination (H1,2), were introduced into the HIV-1 poly(A) site in pHmg. Moreover, HIV-1 like sequences were used to replace the U₃AU₄ (S1), U-rich motif (S2) and a combination of both elements (S1S2) in pHmg5'S. This complement of poly(A) site mutations was then assayed for both MSD-dependent poly(A) site inhibition and poly(A) site strength.

To test the sensitivity of the mutant poly(A) sites to MSD-dependent inhibition, they were analysed upstream of both wild-type MSD (+) and mutant MSD (-) (Figure 5B and E). Mutation of the HIV-1 poly(A) site to H1 and H2 (Figure 5B, lanes 3 and 4; lanes 5 and 6) gives a lower activation effect on MSD mutation, primarily because of higher poly(A) site use in the presence of the MSD. The H1,2 double mutant shows a pattern of poly(A) site use more reminiscent of SPA than the HIV-1 poly(A) site, giving high poly(A) site use with both wild-type and mutant MSD (Figure 5B, lanes 7 and 8). For the SPA mutations, S1 has a small effect in that the activity of the poly(A) site is less with wild-type MSD (lanes 11 and 12). The S2 and S1,2 combined mutations have dramatic effects giving virtually no poly(A) site use in the presence of the wild-type MSD, in a manner more similar to the HIV-1 poly(A) site than SPA (Figure 5B, lanes 13 and 14, and lanes 15 and 16).

To test the poly(A) site strength of each mutant, an established assay system was used (Moreira *et al.*, 1995). The poly(A) site is placed downstream of an inactive

Fig. 4. A mini-gene system and the role of splice acceptor site. (A) Diagram showing the HIV-1 provirus and two mini-gene derivatives, pHmg and pHmg2. The LTRs, cap sites and poly(A) sites are labelled for these constructs. In addition SDs and SAs are represented by $\diamond$ above and $\triangle$ below the construct diagram respectively (the MSD is indicated by $\blacklozenge$). Dashed lines represent the regions of the provirus (the first 2 kb and a region from the centre of the genome containing five SAs) used to construct the mini-genes. For pHmg, the hatched box depicts the 3' end of the human γ globin gene whereas for pHmg2 a larger fragment was used including the final γ globin splice acceptor site. (B) Figure showing an RNase protection assay for the constructs pHmg (lanes 1 and 2) and pHmg2 (lanes 3 and 4). Both are shown either with the wild-type MSD (lanes 1 and 3) or bearing the M1 MSD mutation (lanes 2 and 4). Constructs are labelled above and protected fragments on the left of the panel.

mutant α -globin poly(A) site and upstream of the efficient SPA (see Figure 5C) allowing the strength to be measured versus SPA. A 189 nt fragment containing each of the poly(A) sites described above was transferred into this vector. Figure 5D shows an S1 nuclease protection assay with homologous end-labelled probes using cytoplasmic RNA from transfected HeLa cells. The strengths of the poly(A) sites relative to SPA were measured as percentages and plotted next to the MSD data in Figure 5E. From this graph, the relative strengths of the poly(A) sites correlate reasonably well with the levels of MSD mutation-dependent activation, with strong poly(A) sites giving high activity regardless of the MSD and weaker poly(A) sites giving high levels of activation due to low basal poly(A) site use with wild-type MSD. The only exception to this is the H1 mutation where the poly(A) site strength relative to SPA is almost the same as the wild-type HIV-1 poly(A) site, whereas in the MSD assay the poly(A) site is quite active even where the MSD is wild-type. This may be due to some specific feature of this mutation in either of the assays. The H1 mutation may allow a greater degree of secondary structure as the mutant U_3AU_4 can form a hairpin with the A_2UA_3 , which might influence either of the assays. However, the main focus of the experiment is the HIV-1 poly(A) site and it is clear that it is delicately poised so that it is an efficient signal yet remains inactive when upstream of the wild-type MSD.

An analysis of spacing constraints

To test the range over which the MSD inhibits the 5' LTR poly(A) site, a set of linkers was inserted between the poly(A) site and the MSD in pHAPB (Figure 6A). The normal distance of 195 nt between the poly(A) site and MSD was increased by 174, 348, 587 and 1174 nt to give

the constructs pHAPB1–L4. Figure 6B shows an RNase protection analysis of cytoplasmic RNA from transfected HeLa cells. However, the protected fragments for this analysis are complicated by a mismatch at the site of linker insertion, giving rise to the RNA species C-*, *-S and *-RT (Figure 6A). A riboprobe was used which contains a heterologous linker at the same site as pHAPB1–L4, thus ensuring the mismatch is recognized for these constructs and allowing comparison with the constructs without linker sequences. From this analysis, it can be seen that the increased distance between the poly(A) site and MSD gives at least some poly(A) site use for all the linkers used (Figure 6B, lanes 3–6, the C-pA product). All the mRNAs which do not use the poly(A) site will generate a C-* product. Therefore, the level of poly(A) site use can be measured by comparison of C-pA with C-* and the percentage of polyadenylated transcripts can be calculated (Figure 6C). From these data it appears that there is a gradual increase in poly(A) site use with increasing distance. With over 1 kb of linker sequence, the level of poly(A) site use is the same as when the MSD is mutated (Figure 6B, lane 2).

The precise size of the *-S band in Figure 6B varies slightly due to both the use of the C1 cryptic SD for pHAPBM1 and a slight variation in the degree of homology at the mismatch for the linker constructs. When the relative proportion of spliced (*-S) to readthrough (*-RT) products is measured it remains similar with increasing spacing (Figure 6D). As shown in Figures 2 and 6D, mutations to the MSD cause HIV-1 poly(A) site activation but also decrease the levels of spliced products. Whereas, the spacing experiment shows that activation of the poly(A) site can occur independently of changes to the MSD and the ratio of spliced mRNA. This suggests that the inhibition

Fig. 5. Analysis of HIV-1 poly(A) site specificity. **(A)** An alignment of the HIV-1 poly(A) and SPA sites. The regions shown are divergent between the constructs pHmg and pHmg5'S. A₂UA₃ and the GU/U elements are shown in bold and arrows depict the sites of cleavage. Highly divergent sequences are underlined and the mutations introduced either singly or in combination are shown above the sequences. **(B)** Figure showing an RNase protection analysis where the poly(A) sites above were tested in mini-gene, pHmg-based constructs. The labelling above the panel represents the poly(A) sites which were analysed upstream of either wild-type MSD (+) or the M1 MSD mutant (-). Protected fragments are shown on the left of the panel. Mutant constructs which contain AU-rich sequences just upstream of the cleavage site give a broad set of C-pA products. This is presumably due to overdigestion at sites of RNA hybrid breathing and has been described previously using SPA (Figure 2B; Weichs an der Glon *et al.*, 1993). **(C)** Diagram showing the construct used to test each poly(A) site in competition with downstream SPA. The human $\alpha 2$ globin gene is shown with hatched areas representing exons and unfilled areas introns. The α poly(A) site is inactive due to a mutation and test poly(A) sites are placed upstream of SPA. Levels of use of each test poly(A) site are measured versus SPA using homologous end-labelled DNA probes and S1 nuclease protection analysis. Protected fragments are represented below the construct diagram. **(D)** Figure showing an S1 nuclease protection assay to test the strength of each of the poly(A) sites using constructs described above. The poly(A) sites tested are labelled above and the identity of protected fragments are shown to the left of the panel. TPA and SPA represent use of the test poly(A) site and SPA respectively, whereas R β CoT represents the level of a co-transfected rabbit β globin mRNA 3' end. **(E)** Graph summarizing the analysis above. The percentage poly(A) site use versus the wild-type MSD, the M1 MSD mutation or SPA is shown for each poly(A) site and was calculated from the amount of poly(A) site use (all bands representing C-pA were included) over the total initiated transcripts. Calculations for the RNase protection assays take into account the radiolabel present in each band, roughly three times less radiolabel is present in C-pA as in C-S, C1, C2 and RT products.

of the poly(A) site in wild-type HIV-1 does not involve the process of splicing but does rely on the MSD.

The role of U1 snRNP

U1 snRNP interacts directly with the SD and we have investigated how this association with the RNA affects the inhibition of the HIV-1 poly(A) site. The processes of splicing and SD recognition can be uncoupled *in vivo* by the use of shift-U1 snRNAs (S-U1s) (Cohen *et al.*, 1994;

Hwang and Cohen, 1996). S-U1s can be engineered to bind at specified distances from mutated SDs and can suppress the SD mutations. Figure 7A shows the S-U1 used in these experiments which was constructed so that the 5' end of the U1 snRNA (which normally interacts with the SD) interacts with a sequence 14 nt downstream of the MSD in HIV-1. This S-U1 construct was co-transfected with pHAPB and each of the MSD mutants (M1, M2 and M1,2). Expression of S-U1 snRNAs was

Fig. 6. The analysis of spacing between the poly(A) site and the MSD. (A) Diagram showing the 5' end of HIV-1 proviral based constructs where the spacing between the poly(A) site and the MSD was varied by insertion of linkers (L1–L4). The site of linker insertion is 90 nt downstream of the cleavage site and the size in nucleotides of each linker is shown in boxes. All other abbreviations, etc. are the same as Figure 2A except in the protected fragments; * represents an end created by mismatch at the site of linker insertion. (B) Figure showing an RNase protection assay with the spacing constructs pHAPBL1–L4. In addition, the controls pHAPB [no poly(A) site use] and pHAPBM1 [active poly(A) site use] are shown. The riboprobes used here differ from others in this study in that they contain a non-homologous linker in the same position as the linkers in pHAPBL1–L4. This means the protected fragments (labelled on the left of the panel) give mismatch bands (denoted by *) at the site of linker insertion [see (A)]. (C) Graph showing the percentage poly(A) site use for the spacer and control constructs. (D) Graph showing the ratio of spliced (*-S) to readthrough (*-RT) protected fragments for the spacer and control constructs. Values were calculated taking account of the radiolabel present in each protected fragment.

confirmed using a specific riboprobe (Figure 7B, lower panel). The co-transfection of a plasmid producing wild-type U1 snRNA did not alter the pattern of poly(A) site activation brought about by the MSD mutations (Figure 7B, lanes 1–4 compared with Figure 2B, lanes 2–5). However, co-transfection of S-U1 causes a significant decrease in the levels of poly(A) site use for each mutant MSD construct (Figure 7B, lanes 5–8). S-U1 also has small effects on splicing; for the M1 mutation, S-U1 activates the C1 SD and inhibits the C2 SD (lane 6). For the M2 mutation, S-U1 activates the MSD but inhibits the C2 SD (lane 7) and for the combined M1,2 mutant an inhibition at the C2 SD can be seen (Figure 7B, lane 8 and 7C, lanes 1–5). Therefore, inhibition of polyadenylation is consistently observed when the S-U1 is added, even though there is a variety of effects on splicing. More specifically, for pHAPBM1,2 the presence of S-U1 actually decreases the level of spliced RNAs (Figure 7B, compare lanes 4 and 8, a 2.1-fold decrease in the ratio of spliced to readthrough RNAs on quantitation). This means that increased splicing does not correlate with the inhibition of the poly(A) site and suggests that the interaction of U1 snRNP is the critical determinant of poly(A) site use.

To demonstrate further that U1 snRNP interaction is responsible for the inhibition of the 5' LTR HIV-1 poly(A) site, an experiment was performed where the S-U1 plasmid was titrated. Figure 7C, lanes 1–5, shows that using the

pHAPBM1,2 mutant and increasing amounts of S-U1 (Figure 7C, lower panel), the levels of 5' LTR HIV-1 poly(A) site use decrease. This analysis shows that the level of S-U1 correlates directly with the inhibition of the poly(A) site. This is shown quantitatively in Figure 7D.

As a further control, the region of pHAPBM1,2 where S-U1 interacts, was mutated to give pHAPBM1,2 AS, and this was transfected with either S-U1 or anti-shift-U1 (AS-U1) plasmids. Figure 7A shows the AS mutation which can only interact with AS-U1. Figure 7C, lanes 6 and 7, shows that the AS mutation restores poly(A) site use with S-U1, but with the complementary AS-U1, inhibition of poly(A) site use is evident.

The levels of S-U1 and AS-U1 (lower panels of Figure 7B and C) correspond to 1.8% and 7% of endogenous U1 respectively (5 µg lanes). Firstly, it is difficult to explain the difference between S-U1 and AS-U1, but the mutations may have effects on the transcription or stability of the snRNAs. Furthermore, the levels seem too low to account for the complete inhibition of the poly(A) site which can be obtained using the AS-U1 in particular (Figure 7C, lane 7). However, <10% of the cells will be transfected by the plasmid, whereas the endogenous signal is generated from every cell. In addition, the U1 snRNP is the most abundant snRNP and therefore a low level versus endogenous U1 does not necessarily mean the transfected U1 is limiting in any one cell.

Fig. 7. Analysis of the role of U1 snRNP. **(A)** Diagram showing the HIV-1 sequence over the MSD. The site of MSD cleavage and the changes caused by the M1, M2 and M1,2 mutations are shown. The complementary 5' end of U1 snRNA is depicted interacting with the MSD. In addition, the shift U1 snRNA, which has been engineered to bind to the HIV-1 sequence 14 nt downstream of the MSD, is shown interacting with this site. A mutation which disrupts this interaction [anti-shift (AS)] is represented in bold below the HIV-1 sequence. An Anti-shift (AS) U1 snRNA which has been engineered to bind to the 14 nt region of the AS construct is also shown. **(B)** Upper panel: RNase protection assay using pHAPB and the MSD mutants. The construct and the presence (3 µg) of co-transfected wild-type U1 (WT-U1) or S-U1 is depicted above each lane. The identity of protected fragments (Figure 2A) is represented to the left of the panel. Lower panel: separate RNase protection assay using a S-U1 snRNA-specific riboprobe for each RNA sample. Protected fragments from wild-type and shift U1 snRNAs are labelled. **(C)** Upper panel: RNase protection assay where the level of S-U1 was titrated (lane 1, no S-U1; lane 2, 0.625 µg; lane 3, 1.25 µg; lane 4, 2.5 µg; lane 5, 5 µg) with constant pHAPBM1,2. In addition, for the anti-shift (AS) version of M1,2 either 5 µg of S-U1 (lane 6) or AS-U1 (lane 7) was co-transfected. Lower panel: level of U1 snRNAs using either S-U1 (lanes 1–5) or AS-U1 (lanes 6–7) -specific riboprobes. Note that the S-U1 product detected in lane 6 is a smaller RNA due to mismatch with the AS-U1 riboprobe. Protected fragments for each riboprobe are shown to the left and right of this panel respectively. The level of AS-U1 expression is higher than S-U1 and this is probably an effect of the individual mutations in the U1 gene. **(D)** Graph depicting quantitation of data in (C). The construct used, pHAPBM1,2 (M1,2) or pHAPBM1,2 AS (M1,2 AS) and the amount in µg of either S-U1 (S) or AS-U1 (AS) are indicated below the graph.

Discussion

Through a combination of different experiments, the role of the MSD in HIV-1 5' LTR poly(A) site use has been analysed. Several lines of evidence suggest that an interaction at the donor site is critical for the process of poly(A) site inhibition. Mutations to the donor site increase the level of poly(A) site use. Although these mutations affect both the level of spliced RNAs and the interaction of U1 snRNP at the MSD, the variable nature of the splicing effect for each of the mutants in conjunction with the consistency of the poly(A) site effect, suggests that splicing is not directly involved in poly(A) site inhibition. In addition, increasing the spacing between the poly(A) site and the MSD leads to activation of the poly(A) site without decreasing the ratio of spliced to readthrough RNAs. Finally, shift U1 snRNAs (S-U1s) engineered to bind 14 nt from the MSD, cause a specific inhibition of poly(A) site use for each of the mutant MSD constructs while having a variable effect on splicing (either increasing

or decreasing levels of spliced RNA). Therefore, from these experiments we conclude that the poly(A) site in the 5' LTR of HIV-1 is inhibited by the interaction of U1 snRNP at the major splice donor site.

A recent study where the HIV-1 poly(A) site was investigated downstream of the adenovirus major late promoter, found that short C-pA products were unstable and only found in nuclear RNA (Scott and Imperiale, 1997). The discrepancy between these results and our own can be explained if the TAR sequence at the 5' end of HIV-1 viral transcripts stabilizes the C-pA product. TAR is known to stabilize short transcripts produced by tat minus transcription. Therefore, use of proviral constructs has facilitated the detection of the C-pA product and allowed the identification of the mechanism for poly(A) site inhibition.

An indirect competition model for poly(A) site use

As described above, the limitation to poly(A) site use is enforced by the U1 snRNP interaction at the MSD. One

interpretation is that the MSD–U1 snRNP interaction is in equilibrium with the poly(A) site–poly(A) factor interactions. The experiments above suggest that when both signals are wild-type the equilibrium favours the MSD–U1 snRNP interaction leading to poly(A) site inhibition. However, any mutation which increases the strength of the poly(A) site might allow greater interaction of poly(A) factors, both before and after the MSD is transcribed, giving increased poly(A) site use. Similarly, MSD mutation activates poly(A) site use. However, in this case U1 snRNP can bind to cryptic SDs which are also activated by the mutations. We hypothesize that the weak cryptic SDs lead to decreased interaction of U1 snRNP which allows the equilibrium to tip in favour of the poly(A) site. In fact, any change to this subtle regulation may alter the balance allowing poly(A) site use.

Timing of the processing reactions and the effect of increased spacing

Potential clues as to the relative timing of the processing reactions involved come from the splice acceptor analysis. Replacement of the weak HIV-1 SAs with a strong SA inhibits use of the poly(A) site even when the MSD is mutated. This result can be explained if the increased affinity for splicing factors at the SA and branch site leads to a concomitant increase in U1 snRNP binding to the cryptic SDs. Therefore, the equilibrium favours cryptic SD–U1 snRNP interaction and effectively rescues poly(A) site inhibition. A connection between factors at the SA and those at the SD has been previously suggested and a network of interactions involving U1 snRNP has been described to explain this (Wu and Maniatis, 1993). Transcription of the SA is therefore a prerequisite for this effect. However, the SA lies 1.5 kb downstream of the MSD in this construct (Figure 4A). As this SA can affect the use of a far-upstream poly(A) site, the poly(A) site cannot normally be used until the transcribing polymerase has elongated a considerable distance further downstream.

In connection with this, an increase in spacing between the poly(A) site and the MSD gives increased poly(A) site use without affecting the efficiency of splicing. This result can be interpreted in many ways. Firstly, the increased distance might give the poly(A) site an advantage because of the increased time before the MSD is transcribed. In addition, the increase in spacing may decrease the rate of U1 snRNP–MSD interaction in some other way, or it might disrupt poly(A) site inhibition such that the poly(A) site is used even when U1 snRNP is bound to the MSD. The SA result explained above effectively rules out the explanation where the spacing allows more time for poly(A) site use as the poly(A) site is not used until after the RNA polymerase has transcribed at least a further 1.5 kb. In addition, the splicing/readthrough ratio for HIV-1 mRNAs does not change with increased spacing (Figure 6D) which shows that no other limitation to U1 snRNP–MSD interaction is brought about by the arrangement in these constructs. Therefore, the evidence suggests that the increase in the distance between the poly(A) site and the MSD causes the interaction of U1 snRNP at the MSD to no longer inhibit polyadenylation. This might come about via an altered structural and spatial arrangement of RNA and factors.

The mechanism of U1 snRNP-dependent inhibition of the HIV-1 poly(A) site

U1 snRNP could inhibit the use of the poly(A) site by several possible mechanisms. Firstly, U1 snRNP may contact the RNA over the poly(A) site, thus directly blocking the access of poly(A) factors. Experiments using two competing SDs where U1 interactions were mapped suggest that the two SDs can only mutually interfere leading to a general decrease in splicing when they are <45 nt apart (Eperon *et al.*, 1993). Since the HIV-1 MSD lies 195 nt downstream of the poly(A) site this would seem to be too far to inhibit polyadenylation in a similar fashion. However, U1 snRNP may interact with a non-specific RNA binding factor which would in turn block access of poly(A) factors to the poly(A) site. Another possibility is that a complex of factors which links the cap site to the U1 snRNP (Colot *et al.*, 1996; Lewis *et al.*, 1996) might indirectly block the access of poly(A) factors.

Alternatively, U1 snRNP may interact with factors which affect polyadenylation. These interactions may be direct or indirect and could have a number of consequences. They could disable poly(A) factors by preventing correct recognition of the poly(A) sequences or disrupting interactions with other poly(A) factors. U1 snRNP–poly(A) factor associations could also inhibit some enzymatic step in the process of cleavage and polyadenylation.

A number of studies have identified associations between poly(A) factors and components of the U1 snRNP, in particular the U1A protein. U1A protein has been shown to interact with both poly(A) polymerase (PAP) and the 160 kDa subunit of the cleavage–polyadenylation specificity factor (CPSF). The interaction with PAP occurs in a regulatory negative feedback loop for the U1A gene which is responsible for inhibition of U1A transcript polyadenylation (Boelens *et al.*, 1993). This relies on an RNA structure bound by two U1A proteins in a specific structural conformation (Gunderson *et al.*, 1997). This regulatory mechanism is therefore highly specific to the U1A locus and so may not be relevant to HIV-1 poly(A) site use. The interaction with CPSF can occur while the U1A protein is part of the U1 snRNP and has been shown to activate polyadenylation at the downstream SV40 late poly(A) site (Lutz and Alwine, 1994; Lutz *et al.*, 1996). This interaction may well be relevant to HIV-1 poly(A) site use, even though in this case the interaction would have to be inhibitory to poly(A) site use. It is possible that the arrangement of the HIV-1 poly(A) site with a downstream MSD may facilitate this, although studies on the papillomavirus have shown that a cryptic splice donor upstream of the late poly(A) site is inhibitory (Furth *et al.*, 1994).

The analysis of the mechanism of HIV-1 poly(A) site inhibition by U1 snRNP may require the development of a cell-free system which faithfully reproduces the regulation observed here. Previous studies using conventional nuclear extracts have shown that the HIV-1 poly(A) site is practically inactive in the absence of upstream sequences which are absent from the 5' LTR HIV-1 poly(A) site (data not shown; Gilmartin *et al.*, 1992; Valsamakis *et al.*, 1992). Therefore, the reconstruction of this regulation to allow molecular dissection will be particularly challenging.

In this study, we have used proviral HIV-1 clones *in vivo*

to identify the mechanism of HIV-1 poly(A) site selection. The results described here show that U1 snRNP interaction at the MSD inhibits the use of the upstream HIV-1 5' LTR poly(A) site. This has led to invaluable additional understanding in the complexity and subtlety of the RNA processing reactions of HIV-1 and may lead to a greater understanding of more general limitations to poly(A) site use.

Materials and methods

Plasmid construction

The plasmids pH, pHAPB, pHAPB5'S and pHAPBM1 (pHAPBMuMSD) have been described previously (Ashe *et al.*, 1995). The mutations to the major splice donor site (pHAPBM2 and pHAPBM1,2), gag initiation codon (Δ G series of plasmids) and the AS mutation were made using mutant oligonucleotides and the megaprimer PCR mutagenesis technique (Picard *et al.*, 1994). The linker clones pHAPBL1–4 were made by inserting the 174 nt and 587 nt *Hae*III fragments from pUC119 as either monomers or dimers (the linkers in the monomer and dimer constructs are all in the same orientation) into the *Ehe*I site (639, all HIV-1 numbering relates to the NL4-3 sequence; Adachi *et al.*, 1986).

The plasmid pHmg2 was made by replacing the entire 3' end of the provirus using *Apa*I (2011) and *Bam*HI (downstream of the 3' LTR) with a *Dra*I–*Dra*I fragment containing the γ^A -globin final exon and poly(A) site. The plasmid pHmg was made by replacing the 3' end of the provirus in pH [*Sst*I (6003)–*Bam*HI (downstream of the 3' LTR)] with the 3' end of the γ^A -globin gene (*Bst*EII–*Dra*I). The region between *Apa*I (2011) and *Eco*RI (5744) was then deleted. In pHmg5'S the *Ava*I (295)–*Kas*I (637) fragment from pHAPB5'S was inserted into pHmg. The pHmgM1 and pHmg5'SM1 constructs were made by replacing the *Nar*I (638)–*Spe*I (1507) fragment of pHmg and pHmg5'S with a similar fragment from pHAPBM1. The mutants H1, H2, H1,2, S1, S2 and S1,2 were made using mutant oligonucleotides and the megaprimer PCR mutagenesis technique (Picard *et al.*, 1994).

*Pvu*II (435)–*Asp*700 (842) fragments from most of the proviral–mini-gene constructs were placed in pGem4 using *Xba*I and the resulting construct linearized with *Bam*HI for the production of homologous riboprobes. Riboprobes used in the spacer analysis were the same except that they contained the 174 nt *Hae*III fragment from pUC119 in the opposite orientation to that found in pHAPBL1 and L2.

The parental α globin competition clone used here to test the strength of poly(A) sites has been described previously (Moreira *et al.*, 1995). Poly(A) site fragments [cap site (453) to downstream of the 5' LTR (642)] from each of the mutants were isolated using PCR and placed in the vector at a *Pvu*II site between the mutated α poly(A) site and SPA.

The parental U1 snRNA clone was pG3U1, a derivative of pHU1-1D (Lund and Dahlberg, 1994) and a gift from C.Smith, Cambridge University, UK. The mutant variants S-U1 and AS-U1 were made by replacing the sequence between the sites *Bcl*I and *Bgl*III with mutant oligonucleotides. *Bgl*III–*Bsp*EI fragments were placed in the *Xba*I site of pGem4 and the resulting construct was linearized using *Hind*III for homologous riboprobe production.

Human β globin, tat and rabbit β globin plasmids have been described previously (Grosveld *et al.*, 1982; Adams *et al.*, 1988; Proudfoot *et al.*, 1992).

Cell culture, lysate production and cytoplasmic RNA purification

Subconfluent HeLa cells were transiently transfected with 5–10 μ g of proviral or mini-gene plasmids, 2–3 μ g of co-transfection plasmid, 2 μ g of tat plasmid and 0.625–5 μ g of U1 plasmids. After 48 h, whole-cell lysates for protein analysis were made by treating cells with 85°C SDS sample buffer [50 mM Tris–HCl, pH 6.8, 100 mM DTT, 2% (w/v) SDS, 0.1% (w/v) bromophenol blue and 10% (v/v) glycerol] or cytoplasmic RNA was purified as described in Eggermont and Proudfoot (1993).

Western blot analysis of viral gag and env products

Boiled cell lysates were separated by SDS–polyacrylamide electrophoresis and transferred to Hybond-C pure nitrocellulose membrane (Amersham). Membranes were immunoblotted using anti-env gp120 (a gift from Ian Jones) and anti-p24 (a gift from William James) by standard procedures (Maniatis *et al.*, 1982).

RNase protection analysis

Riboprobes were made using either T7 or Sp6 RNA polymerase (according to the manufacturer's instructions) and the number of prematurely-terminated species made purification from a denaturing polyacrylamide gel necessary. RNA was eluted from gel slices in buffer [0.5 M ammonium acetate, 0.1% (w/v) SDS, 1 mM EDTA] for 1 h, extracted with phenol/chloroform and ethanol-precipitated. 500 c.p.s. of riboprobe was hybridized with 20–40 μ g (depending on the amount of RNA available) of cytoplasmic RNA. RNase protections were performed as described previously (Ashe *et al.*, 1995). Where RNase I was used the RNase protection buffer was different, i.e. 10 mM Tris–HCl, pH 7.5, 20 mM EDTA, 33 units/ml RNase I (Boehringer Mannheim). Electrophoresis gels were quantitated using a PhosphorImager and protected fragments assigned on the basis of size versus DNA markers and previous analyses using a variety of different riboprobes (Ashe *et al.*, 1995).

S1 nuclease protection analysis

Co-transfected human β globin and rabbit β globin were detected using *Eco*RI end-labelled DNA probes to the 3' ends of these genes. This was performed for all the experiments in this study (data not shown). The homologous probes to the 3' end of the α globin poly(A) site competition constructs were made by end-labelling at *Bsp*120I sites. S1 nuclease protection was performed as described by Ashe *et al.* (1995).

Acknowledgements

We thank H.Ashe, M.Dye and C.Birse for critical reading of the manuscript and helpful discussions. We thank C.Smith, I.Jones and J.Cohen for their advice, and M.Yonah and S.Murphy for their help with the Western blots. In addition, we thank C.Smith, W.James, I.Jones and S.Kingsman for gifts of plasmids and reagents. This work has been funded by a Wellcome programme (0327731/z/95) and a Medical Research council (MRC) project grant (G9603876) to N.J.P.

References

- Adachi,A., Gendelman,H.E., Koenig,S., Folks,T., Willey,R., Rabson,A. and Martin,M.A. (1986) Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.*, **59**, 284–291.
- Adams,S.E., Johnson,I.D., Braddock,M., Kingsman,A.J. and Kingsman,S.M. (1988) Synthesis of a gene for the HIV transactivation protein tat by a novel single stranded approach involving *in vivo* gap repair. *Nucleic Acids Res.*, **16**, 4278–4297.
- Amendt,B.A., Hesslein,D., Chang,L.-J. and Stoltzfus,C.M. (1994) Presence of negative and positive *cis*-acting RNA splicing elements within and flanking the first tat coding exon of Human Immunodeficiency Virus Type 1. *Mol. Cell. Biol.*, **14**, 3960–3970.
- Ashe,M.P., Griffin,P., James,W. and Proudfoot,N.J. (1995) Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev.*, **9**, 3008–3025.
- Baurén,G. and Wieslander,L. (1994) Splicing of balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell*, **76**, 183–192.
- Beyer,A.L. and Osheim,Y.N. (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.*, **2**, 754–765.
- Boelens,W.C., Jansen,E.J.R., van Venrooji,W.J., Striepecke,R., Mattaj,I.W. and Gunderson,S.I. (1993) The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, **72**, 881–892.
- Cohen,J.B., Snow,J.E., Spencer,S.D. and Levinson,A.D. (1994) Suppression of mammalian 5' splice-site defects by U1 small nuclear RNAs from a distance. *Proc. Natl Acad. Sci. USA*, **91**, 10470–10474.
- Colot,H.V., Stutz,F. and Rosbash,M. (1996) The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev.*, **10**, 1699–1708.
- Cooke,C. and Alwine,J.C. (1996) The cap and the 3' splice site similarly affect polyadenylation efficiency. *Mol. Cell. Biol.*, **16**, 2579–2584.
- Eggermont,J. and Proudfoot,N.J. (1993) Poly(A) signals and transcriptional pause sites combine to prevent interference between RNA polymerase II promoters. *EMBO J.*, **12**, 2539–2548.
- Eperon,I.C., Ireland,D.C., Smith,R.A., Mayeda,A. and Krainer,A.R. (1993) Pathways for selection of 5' splice site by U1 snRNPs and SF2/ASF. *EMBO J.*, **12**, 3607–3617.
- Fäcke,M., Janetzko,A., Shoeman,R.L. and Kräusslich,H.-G. (1993) A large deletion in the matrix domain of the Human Immunodeficiency

- Virus *gag* gene redirects virus particle assembly from the plasma membrane to the endoplasmic reticulum. *J. Virol.*, **67**, 4972–4980.
- Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W. and Lührmann, R. (1995) The HIV-1 rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*, **82**, 475–483.
- Furth, P.A., Choe, W.-T., Rex, J.H., Byrne, J.C. and Baker, C.C. (1994) Sequences homologous to 5' splice sites are required for the inhibitory activity of papillomavirus late 3' untranslated regions. *Mol. Cell. Biol.*, **14**, 5278–5289.
- Gilmartin, G.M., Fleming, E.S. and Oetjen, J. (1992) Activation of HIV-1 pre-mRNA 3' processing *in vitro* requires both an upstream element and TAR. *EMBO J.*, **11**, 4419–4428.
- Göttlinger, H.G., Sodroski, J.G. and Haseltine, W.A. (1989) Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA*, **86**, 5781–5785.
- Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. (1982). DNA sequence necessary for transcription of the rabbit β globin gene *in vivo*. *Nature*, **295**, 120–126.
- Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M. and Mattaj, I.W. (1997) Involvement of the carboxy terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes Dev.*, **11**, 761–773.
- Hwang, D.-Y. and Cohen, J.B. (1996) Base pairing at the 5' splice site with U1 small nuclear RNA promotes splicing of the upstream intron but may be dispensable for splicing of the downstream intron. *Mol. Cell. Biol.*, **16**, 3012–3022.
- Inoue, K., Ohno, M., Sakamoto, H. and Shimura, Y. (1989) Effect of the cap structure on pre-mRNA splicing in *Xenopus* oocyte nuclei. *Genes Dev.*, **3**, 1472–1479.
- Izaurrealde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. and Mattaj, I.W. (1994) A nuclear cap binding complex involved in pre-mRNA splicing. *Cell*, **78**, 657–668.
- Jones, K.A. and Peterlin, B.M. (1994) Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.*, **63**, 717–743.
- Keller, W. (1995) No end yet to mRNA 3' processing. *Cell*, **81**, 829–832.
- Kingsman, S.M. and Kingsman, A.J. (1996) The regulation of human immunodeficiency virus type-1 gene expression. *Eur. J. Biochem.*, **240**, 491–507.
- Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1984) Recognition of cap structure in splicing *in vitro* of mRNA precursors. *Cell*, **38**, 731–736.
- Lamond, A.I. (1993) The spliceosome. *BioEssays*, **15**, 595–603.
- LeMaire, M.F. and Thummel, C.S. (1990) Splicing precedes polyadenylation during *Drosophila* *E74A* transcription. *Mol. Cell. Biol.*, **10**, 6059–6063.
- Levitt, N., Briggs, D., Gil, A. and Proudfoot, N.J. (1989) Definition of an efficient synthetic poly (A) site. *Genes Dev.*, **3**, 1019–1025.
- Lewis, J.D., Izaurrealde, E., Jarmolowski, A., McGuigan, C. and Mattaj, I.W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.*, **10**, 1683–1698.
- Liu, X. and Mertz, J.E. (1993) Polyadenylation site selection cannot occur *in vivo* after excision of the 3' terminal intron. *Nucleic Acids Res.*, **21**, 5256–5263.
- Lou, H., Gagel, R.F. and Berget, S.M. (1996) An intron enhancer recognised by splicing factors activates polyadenylation. *Genes Dev.*, **10**, 208–219.
- Lund, E. and Dahlberg, J.E. (1984) True genes for human U1 small nuclear RNA. *J. Biol. Chem.*, **259**, 2013–2021.
- Lutz, C.S. and Alwine, J.C. (1994) Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes Dev.*, **8**, 576–586.
- Lutz, C.S., Murthy, K.G.K., Schek, N., O'Connor, P., Manley, J.L. and Alwine, J.C. (1996) Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency *in vitro*. *Genes Dev.*, **10**, 325–337.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Manley, J. (1995) A complex protein assembly catalyses polyadenylation of mRNA precursors. *Curr. Opin. Genet. Dev.*, **5**, 222–228.
- Moreira, A., Wollerton, M., Monks, J. and Proudfoot, N.J. (1995) Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *EMBO J.*, **14**, 3809–3819.
- Nesic, D. and Maquat, L.E. (1994) Upstream introns influence the efficiency of final intron removal and RNA 3'-end formation. *Genes Dev.*, **8**, 363–375.
- Nevins, J.R. and Darnell, J.E., Jr (1978) Steps in the processing of Ad2mRNA: poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing. *Cell*, **15**, 1477–1493.
- Niwa, M., Rose, S.D. and Berget, S.M. (1990) *In vitro* polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev.*, **4**, 1552–1559.
- Ohno, M., Sakamoto, H. and Shimura, Y. (1987) Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proc. Natl Acad. Sci. USA*, **84**, 5187–5191.
- O'Reilly, M.M., McNally, M.T. and Beemon, K.L. (1995) Two strong 5' splice sites and competing, sub-optimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology*, **213**, 373–385.
- Peterson, M.L., Bryman, M.B., Peiter, M. and Cowan, C. (1994) Exon size affects competition between splicing and cleavage-polyadenylation in the immunoglobulin μ gene. *Mol. Cell. Biol.*, **14**, 77–86.
- Picard, V., Erdsdal-Badju, E., Lu, A. and Bock, S.C. (1994) A rapid and efficient one-tube PCR-based mutagenesis technique using *Pfu* DNA polymerase. *Nucleic Acids Res.*, **22**, 2587–2591.
- Proudfoot, N.J. (1991) Poly(A) signals. *Cell*, **64**, 671–674.
- Proudfoot, N.J., Lee, B.A. and Monks, J. (1992) Multiple Sp1 binding sites confer enhancer-independent, replication-activated transcription of HIV-1 and globin promoters. *New Biol.*, **4**, 369–381.
- Purcell, D.F.J. and Martin, M.A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication and infectivity. *J. Virol.*, **67**, 6365–6378.
- Rasmussen, E.B. and Lis, J.T. (1993) *In vivo* transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc. Natl Acad. Sci. USA*, **90**, 7923–7927.
- Scott, J.M. and Imperiale, M.J. (1996) Reciprocal effects of splicing and polyadenylation on Human Immunodeficiency Virus type 1 pre-mRNA processing. *Virology*, **224**, 498–509.
- Scott, J.M. and Imperiale, M.J. (1997) Promoter proximal poly(A) sites are processed efficiently, but the RNA products are unstable in the nucleus. *Mol. Cell. Biol.*, **17**, 2127–2135.
- Sharp, P.A. (1994) Split genes and RNA splicing (Nobel lecture). *Cell*, **77**, 805–815.
- Spearman, P., Wang, J.-J., Heyden, N.V. and Ratner, L. (1994) Identification of human immunodeficiency virus type 1 gag protein domains essential to membrane binding and particle assembly. *J. Virol.*, **68**, 3232–3242.
- Valsamakis, A., Schek, N. and Alwine, J.C. (1992) Elements upstream of the AAUAAA within the human immunodeficiency virus polyadenylation signal are required for efficient polyadenylation *in vitro*. *Mol. Cell. Biol.*, **12**, 3699–3705.
- Wasserman, K.M. and Steitz, J.A. (1993) Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP? *Genes Dev.*, **7**, 647–659.
- Weichs an der Glon, C., Ashe, M., Eggermont, J. and Proudfoot, N.J. (1993) Tat-dependent occlusion of the HIV poly(A) site. *EMBO J.*, **12**, 2119–2128.
- Weiringa, B., Meyer, F., Reiser, J. and Weissman, C. (1983) Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit beta-globin gene. *Nature*, **301**, 38–45.
- Wu, J.Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061–1070.
- Zhang, G., Zapp, M.L., Yan, G. and Green, M.R. (1996) Localisation of HIV-1 RNA in mammalian nuclei. *J. Cell Biol.*, **135**, 9–18.
- Zhuang, Y. and Weiner, A.M. (1986) A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell*, **46**, 827–835.

Received on April 28, 1997; revised on June 23, 1997