An RNA–protein complex links enhanced nuclear 3' processing with cytoplasmic mRNA stabilization

Xinjun Ji¹, Jian Kong¹ and Stephen A Liebhaber^{1,2,*}

¹Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA and ²Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Post-transcriptional controls are critical to gene regulation. These controls are frequently based on sequence-specific binding of *trans*-acting proteins to *cis*-acting motifs on target RNAs. Prior studies have revealed that the KH-domain protein, α CP, binds to a 3' UTR C-rich motif of *h α -globin* mRNA and contributes to its cytoplasmic stability. Here, we report that this 3' UTR α CP complex regulates the production of mature *α -globin* mRNA by enhancing 3' processing of the *h α -globin* transcript. We go on to demonstrate that this nuclear activity reflects enhancement of both the cleavage and the polyadenylation reactions and that α CP interacts *in vivo* with core components of the 3' processing complex. Consistent with its nuclear processing activity, our studies reveal that α CP assembles co-transcriptionally at the *h α -globin* chromatin locus and that this loading is selectively enriched at the 3' terminus of the gene. The demonstrated linkage of nuclear processing with cytoplasmic stabilization via a common RNA–protein complex establishes a basis for integration of sequential controls critical to robust and sustained expression of a target mRNA.

The EMBO Journal (2011) 30, 2622–2633. doi:10.1038/emboj.2011.171; Published online 27 May 2011

Subject Categories: RNA

Keywords: α CP; human *α -globin*; post-transcriptional control; 3' processing; 3' UTR

Introduction

Post-transcriptional controls have a major role in the regulation of eukaryotic gene expression (for reviews see Moore, 2005; Keene, 2010). These controls reflect interactions of RNA-binding proteins and/or non-coding RNAs with sequences and structures on target RNAs (Glisovic *et al*, 2008). In certain cases, the corresponding factors can shuttle between the nucleus and cytoplasm, establishing the potential for coordination of multiple steps in gene expression (Hentze and Kulozik, 1999; Chen *et al*, 2004). The extent to which such coordination of nuclear and cytoplasmic controls takes

place and the underlying mechanisms involved in these processes remains largely unknown.

The development and differentiation of the erythrocyte is intimately linked to high-level globin production. Globin gene expression is initiated by induction of transcription early in erythroid differentiation and is sustained in the terminal phases of erythroblast maturation by stabilization of translationally active globin mRNAs (Liebhaber, 1997). Prior studies have identified a 3' UTR RNA–protein (RNP) complex that enhances *h α -globin* mRNA stability in erythroid cells (Weiss and Liebhaber, 1994, 1995; Kiledjian *et al*, 1995; Ji *et al*, 2003; Kong *et al*, 2003; Kong and Liebhaber, 2007). This stability complex is comprised of the KH-domain RNA-binding protein, α CP (also known as polyC-binding protein (PCBP) and hnRNP E; reviewed in Makeyev and Liebhaber, 2002), bound to a C-rich determinant (Chkheidze *et al*, 1999; Thisted *et al*, 2001). In addition to its role in *h α -globin* mRNA stabilization, α CP associates with multiple other mRNAs (Waggoner and Liebhaber, 2003a) and is likely to constitute a widely distributed post-transcriptional determinant of gene regulation (Holcik and Liebhaber, 1997; Waggoner and Liebhaber, 2003a, b; Chaudhury *et al*, 2010).

The three most abundant α CP isoforms are α CP1, α CP2, and α CP2-KL. These proteins are present in both the nucleus and the cytoplasm (Gamarnik and Andino, 1997; Chkheidze and Liebhaber, 2003) and shuttle between the two compartments based on a set of non-canonical nuclear import determinants and a leucine-rich nuclear export signal (Chkheidze and Liebhaber, 2003). The potential for the nuclear-localized α CPs to impact on gene expression is supported by its binding to multiple sites on the *h α -globin* transcript and by its complex impact on the splicing reaction (Ji *et al*, 2007). Whether the assembly of a specific α CP complex at a unique site on an mRNA transcript can mediate both nuclear and cytoplasmic controls over the expression of a gene transcript to enhance its overall expression remains to be explored.

Alterations in the positioning or efficiency of 3' end processing can impact significantly on the expression of structural gene transcripts (Danckwardt *et al*, 2008). For non-histone transcripts, this process involves cleavage followed by polyadenylation (Moore and Sharp, 1985). These two reactions are tightly linked *in vivo* but can be studied independently under defined *in vitro* conditions (Moore and Sharp, 1985; Gilmartin, 1997). 3' processing requires the actions of multiple protein complexes including the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CFIm), cleavage factor II (CFIIm), poly(A) polymerase (PAP), and the scaffold protein, symplekin (Mandel *et al*, 2008). The assembly of some of these complexes on the nascent transcript may be facilitated by their interaction with the elongating Pol II (Buratowski, 2009).

The accuracy and efficiency of 3' processing is determined by two major *cis*-elements: the polyadenylation hexanucleotide, AAUAAA, positioned 10–30 nt upstream of the cleavage

*Corresponding author. Departments of Genetics and Medicine, University of Pennsylvania School of Medicine, Room 560A CRB, 415 Curie Boulevard, Philadelphia, PA 19104, USA. Tel.: +1 215 898 7834; Fax: +1 215 573 5157; E-mail: liebhaber@mail.med.upenn.edu

Received: 24 February 2011; accepted: 19 April 2011; published online: 27 May 2011

site, and a GU/U-rich region, or 'downstream sequence element' (DSE) located within 30 nt 3' of cleavage site. The CPSF complex associates with the AAUAAA element via sequence-specific RNA binding of CPSF-160 and the CstF complex interacts with DSE through the binding of CstF-64 to GU/U-rich element. Additional *cis*-elements, working together with their interacting protein factors, can modulate, sometimes dramatically, the efficiency of the 3' end processing reaction at a particular site. Two such determinants are the 'upstream sequence elements' (USEs) characteristically located 5' of the AAUAAA signal, and the 'auxiliary downstream sequence elements' (AuxDSEs) positioned 3' of the GU/U-rich element (Millevoi et al, 2009). The USEs are generally U-rich but a consensus sequence has not been established (Mandel et al, 2008). The few AuxDSEs that have been identified are G-rich, but they appear to lack a conserved sequence or distance from the cleavage site (Mandel et al, 2008). Multiple auxiliary elements can be present in a single transcript (Millevoi et al, 2009). These determinants are thought to function by enhancing, either directly or indirectly, the recruitment of the 3' processing complexes (Brackenridge and Proudfoot, 2000; Mandel et al, 2008).

In the present report, we test a model in which the 3' UTR C-rich determinant of the *h α -globin* mRNA acts in the nucleus as a USE to enhance 3' processing of the *h α -globin* transcript. The data support this model and further reveal that this C-rich USE enhances both the cleavage and the polyadenylation reactions. In support of these functional data, we demonstrate that α CP2 is recruited to the *h α -globin* transcript at the endogenous *h α -globin* chromatin locus and associates *in vivo* with core components of the 3' processing complex. These findings, in conjunction with prior studies, support a model in which α CP assembles co-transcriptionally on the 3' UTR of the nascent *h α -globin* transcript, setting the stage for a coordinated set of nuclear and cytoplasmic controls critical to the robust expression of the *h α -globin* gene in the differentiating erythroblast.

Results

The *h α -globin* 3' UTR stability determinant enhances 3' processing of the *h α -globin* transcript

Previous studies have identified a C-rich stability determinant in the 3' UTR of the *h α -globin* mRNA (Weiss and Liebhaber, 1995; Ji et al, 2003; Kong et al, 2003). This determinant recruits the KH-domain protein, α CP (α^{WT} ; Figure 1A), and the resultant RNP ' α -complex' stabilizes cytoplasmic *h α -globin* mRNA in erythroid cells. (Note: the major α CP isoforms are collectively referred to in the text as α CP in settings in which isoform-specific functions have not been demonstrated). Replacement of the 42-nt C-rich region (α CP protected region; PR) in the WT *h α -globin* mRNA (α^{WT}) with a 'neutral' fragment of identical size but unrelated sequence (α^{Neut} ; Figure 1A) destabilizes the mRNA from a half-life of 11 to 7 h (Kong et al, 2003). An even greater destabilization, to a half-life of 3 h, is caused by the naturally occurring anti-termination mutation, $\alpha^{Constant Spring}$ (UAA to CAA) (Clegg et al, 1971) (α^{CS} ; Figure 1A). The anti-terminated ribosome on the α^{CS} mRNA dislodges α CP from the C-rich 3' UTR stability motif during translation and terminates at the in-frame UAA of the polyA signal (AAUAAA) (Kong and Liebhaber, 2007). The respective mRNA half-lives of the α^{WT} , α^{Neut} , and α^{CS}

mRNAs of 11, 7, and 3 h would predict steady-state mRNA accumulations with a corresponding rank order.

To directly test the prediction that the steady-state levels of the α^{WT} , α^{Neut} , and α^{CS} mRNAs corresponded directly with the respective cytoplasmic half-lives, the levels of *h α -globin* mRNA generated by the three corresponding genes were quantified in transfected mouse erythroleukaemia (MEL) cells. Contrary to the initial prediction, we consistently observed that the level of α^{Neut} mRNA was significantly lower than that of α^{CS} mRNA (Figure 1B). This observation suggested that the 3' UTR C-rich region was contributing to *h α -globin* gene expression in a more complicated manner than the simple enhancement of the cytoplasmic mRNA stability. Since the displacement of the α -complex from the α^{CS} mRNA is translation dependent, the impact of the α -complex on α^{CS} mRNA expression should be limited to its cytoplasmic actions. Hence, we considered that the greater negative impact of the α^{Neut} mutation as compared with the α^{CS} mutation might reflect the loss of a critical nuclear function of the α -complex. Based on the close proximity and positioning of the C-rich determinant to the polyA signal, we hypothesized that the α CP complex was acting as an enhancer of 3' end processing.

To determine whether the 3' UTR α CP complex has an impact on 3' processing of the *h α -globin* transcript, we compared the processing of the α^{WT} , α^{Neut} , and α^{CS} transcripts by a pulse-chase analysis. The α^{WT} , α^{Neut} , and α^{CS} genes, each placed under the control of the 'Tet-off' promoter, were separately transfected into MEL/tTA cells (MEL cell based 'Tet-off' system; Kong et al, 2003). Transcription was allowed to proceed for 4 h in tet(-) media. Tetracycline was then added to the media to block further transcription and total RNA was prepared from cells at 0, 2, 4, and 8 h post-tet addition. The ratio of transcripts retaining 3' flanking sequences ('uncleaved') to total spliced *h α -globin* transcripts (cleaved + uncleaved) was determined for each sample (Figure 1C). For the α^{WT} and α^{CS} genes, the analysis revealed only minimal levels of precursor (uncleaved transcript) at the beginning of the transcriptional chase ($t=0$). In contrast, the analysis of the α^{Neut} gene revealed substantial levels of uncleaved transcripts at $t=0$ with residual levels visible up to 2 h into the chase period (Figure 1D). These data, confirmed by real-time PCR analysis (Supplementary Figure S1) are consistent with a model in which the 3' UTR α CP complex enhances 3' processing.

We have previously reported the nuclear assembly of an α CP complex within intron I of human *h α -globin* transcripts (Ji et al, 2007). This intronic complex works jointly with the 3' UTR α CP complex to regulate splicing of the *h α -globin* transcript. To further isolate and define the impact of 3' UTR α -complex on 3' end processing, the α CP-binding site within intron I was eliminated by base substitutions (Figure 2A; Ji et al, 2007). The *h α -globin* transcripts with the mutated intron I α CP-binding site ('mut I'), with or without the 3' UTR α -complex site, were transfected into MEL cells, and the efficiency of 3' end formation was determined (as in Figure 1C). The absence of the α CP-binding site in the 3' UTR resulted in a marked increase in the levels of unprocessed transcripts (Figure 2A). This result confirmed that the 3' UTR α CP-binding site has a major positive impact on 3' processing of the *h α -globin* transcript. The results of a parallel set of studies in which the entire intron I was deleted from

Figure 1 The C-rich determinant within the α -globin 3' UTR links mRNA expression with enhancement of 3' processing of the α -globin transcript. (A) α^{WT} , α^{CS} , and α^{Neut} mRNAs are shown in diagrammatic format. The cap (circle), C-rich α CP-binding site (dark rectangle), and α CP protein (shaded oval) are shown. α CP is bound to the C-rich determinant of the α^{WT} and α^{CS} in the nucleus and α CP is displaced from the α^{CS} mRNA in the cytoplasm as the elongating ribosome enters the 3' UTR (shown as the dotted extension arrow). In contrast, α CP does not assemble on the α^{Neut} mRNA (Kong *et al*, 2003). The half-lives of each mRNA species in MEL cells, as previously reported (Kong *et al*, 2003), are shown on the right. (B) The α^{WT} , α^{CS} , and α^{Neut} genes, each under 'tet-off' transcriptional control (shown in diagrammatic format), were individually transfected along with a fixed amount of a β -globin gene plasmid into MEL/tTA cells and transcription was induced in tet(-) media for 24 h. Levels of α - and β -globin mRNAs were then determined by a quantitative RNA protection assay (RPA). A representative gel is shown. Lanes 1 and 2 represent duplicate transfections for each indicated gene. The histogram represents the α -globin: β -globin mRNA ratios generated in these studies ($n=2$). The error bars represent the corresponding standard deviations ($\pm$ s.d.). (P) values are as follows: WT versus CS = 0.062, WT versus Neut = 0.004, CS versus Neut = 0.022. (C) Analysis of 3' processing of the α -globin transcripts. Position of primers a, b, and c are indicated below the α -globin diagram. Primers a and b correspond to sites within exon 2 and exon 3, respectively. Primer c corresponds to a site 53 nt 3' of the normal cleavage site. The positions of the C-rich PR motif (grey box), the bound α CP (grey oval), and the G/U-rich downstream element (white box) are all shown. The replacement of the PR motif with a 'neutral' sequence of identical size that does not bind α CP is indicated by the open box. The arrow indicates the site of α -globin transcript 3' cleavage and poly(A) addition. The amplicon sets were positioned to bracket an intron in order to distinguish signals reflecting residual genomic DNA and unspliced transcripts from those reflecting spliced RNA transcripts. (D) *In vivo* assessment of α -globin transcript 3' end processing. α^{WT} , α^{CS} , and α^{Neut} genes were transfected into MEL/tTA cells and transcriptionally activated for 4 h. Tet was then added to the media to block further transcription and cells were collected at 0, 2, 4, and 8 h of the chase period. Total cellular RNAs were extracted and aliquots were DNaseI treated and reverse transcribed from primers b or c. cDNAs were then amplified after addition of primer a to each reaction. The [a + b] amplicon includes both 3' cleaved ('mature') and intact ('uncleaved') RNAs (total spliced). The [a + c] amplicon is specific to uncleaved RNA. Mouse GAPDH was included as a RT-PCR internal control (Ji *et al*, 2007).

the α -globin gene was in full agreement with these observations and conclusions (unpublished data). These data support a model in which the 3' UTR α -complex enhances 3' end processing of the α -globin transcript.

Enhancement of α -globin 3' processing by the C-rich region is cell-type restricted

Prior studies of α -globin mRNA expression in a variety of cell lines suggest that the stabilizing role of the 3' UTR α -complex on cytoplasmic α -globin mRNA is erythroid restricted (Weiss and Liebhaber, 1995; Kong *et al*, 2003).

With this in mind, we compared the impact of this complex on 3' end processing in erythroid (MEL) and non-erythroid (HeLa) cell lines. α^{WT} and α^{Neut} plasmids were individually transfected into MEL/tTA and HeLa/tTA cells (Kong *et al*, 2003), and the ratios of unprocessed: total transcripts were determined after a 24-h transcriptional pulse in tet(-) media. Consistent with the prior results (Figure 1D), deletion of the C-rich region (α^{Neut}) resulted in a marked increase in the proportion of unprocessed transcript in the MEL/tTA cells (Figure 2B, left). In contrast, there was no significant impact of the 3' UTR stability motif on processing of the α -globin

Figure 2 Enhancement of *hα-globin* 3' processing by the 3' UTR α -complex is independent of the α -complex in intron I and is erythroid specific. (A) The α CP-binding site in intron I of the *hα-globin* transcript (Ji *et al*, 2007) was inactivated by previously validated base substitutions in pTet/ α^{WT} and pTet/ α^{Neut} genes. The $\alpha^{WT}/mut I$ and the $\alpha^{Neut}/mut I$ constructs generated in this manner were separately transfected into MEL/tTA cells and the efficiency of *hα-globin* 3' processing was determined (as in Figure 1C). A representative gel analysis is shown and the composite results from three independent studies ($n = 3$) are represented in the histogram with $\pm$ s.d. and P value. The data confirm a marked enhancement of 3' processing by the 3' UTR C-rich determinant. (B) The 3' UTR C-rich determinant enhances 3' processing of the *hα-globin* transcript in MEL cells but not HeLa cells. α^{WT} and α^{Neut} gene constructs were transfected into MEL/tTA cells or HeLa/tTA cells (Kong *et al*, 2003). RNA was prepared 24 h post-transcription and 3' processing was assessed by determining the ratios of 'uncleaved' to 'total spliced' *hα-globin* mRNA in each sample (as in Figure 1C). The corresponding histograms represent the data from four independent transfections ($n = 4$). The error bars represent $\pm$ s.d. and the levels of significance are indicated.

transcript in HeLa cells (Figure 2B, right). These results reveal that the 3' UTR α -complex demonstrates the same erythroid specificity in 3' end processing of the *hα-globin* transcript in the nucleus as previously described for stabilization of *hα-globin* mRNA in the cytoplasm.

A competition assay confirms the role of the 3' UTR C-rich region of the *hα-globin* mRNA as an enhancer of 3' processing

The impact of the α CP complex on 3' processing of the *hα-globin* transcript was further tested in a competition assay. α^{WT} and α^{Neut} genes were each inserted into the pcDNA3 vector 5' to the bovine growth hormone (*bGH*) 3' processing element (Figure 3A). These *hα-globin* genes retained 87 bp of native 3' flanking sequences that encompasses a G/U-rich element thought to be involved in the native 3' processing reaction (Levitt *et al*, 1989). The resultant recombinant transcript thus places the *hα-globin* 3' processing signals, with or without the C-rich determinant, in direct competition with the processing determinants in the contiguous *bGH* 3' end processing element. The efficient activity of the strong downstream *bGH* poly(A) site was directly confirmed by a mapping study (Supplementary Figure S2). The insertion of the *bGH* 3' element downstream to the *hα-globin* gene increased the relative abundance of unprocessed WT *hα-globin* transcripts in the transfected MEL cells from 3.7 to 31% (compare with Figure 2B). This increase is consistent with a strong competition effect by the *bGH* 3' processing signals (Levitt *et al*, 1989). Even at this high baseline, the loss of the C-rich region resulted in a significant doubling in the level of uncleaved transcripts (Figure 3B). Taken together, these data further support the role of the C-rich region as an enhancer of 3' end processing of the *hα-globin* transcript.

The major C-rich α CP-binding site within the *hα-globin* 3' UTR enhances 3' cleavage of the *hα-globin* transcript

The foregoing studies demonstrate that the 3' UTR C-rich elements enhance 3' end formation on the *hα-globin* tran-

script *in vivo*. As such, this region conforms to the definition of a USE for RNA 3' end processing (Danckwardt *et al*, 2008). USEs can mediate their effects by enhancing 3' cleavage and/or by augmenting the polyadenylation (Danckwardt *et al*, 2008). To define the mechanistic basis for the enhancing activity of the C-rich USE, we determined its impact on each of these two steps in 3' end formation in a series of *in vitro* assays. To increase the resolution of our subsequent analyses and directly link this activity with α CP binding, we introduced sets of C to G substitutions at each of the three C-rich patches that constitute the 42 nt C-rich PR region in the *hα-globin* 3' UTR (segments 1, 2, and 3; Figure 4A). These substitutions were introduced into the *hα-globin* gene both individually and in combination. As a control on the subsequent functional studies, we also introduced a set of base substitutions that inactivate the 3' end processing functions of the poly(A) signal (AAUAAA to ACUAAG) (Whitelaw and Proudfoot, 1986) (Figure 4A).

RNA substrates for analysis of the 3' cleavage reaction were generated by *in vitro* transcription of *hα-globin* gene templates that encompass the entire 3' UTR and extend to 87 bases of contiguous 3' sequences ('*in vitro* cleavage Rx 32 P-substrate'; Figure 4A). The processing of these RNA templates was then assessed *in vitro* in cell extracts. To directly link α CP binding at the C-rich region to the assay results, α CP binding to each substrate was re-mapped in nuclear extract under *in vitro* cleavage assay conditions by a 32 P-label-transfer assay (Figure 4B, left panel). The identity of the α CP2-RNA complex (α -complex) was confirmed by immunoprecipitation with an α CP2-specific antisera (Figure 4B, right panel). The effects of each substitution on α complex formation were also assessed by RNA EMSA using recombinant α CP2 protein (Supplementary Figure S3). In both analyses, the mutations on the C-patches 1 or 2 were able to assemble the α complex, although at reduced efficiency, while mutation in the third C-patch alone or in combination with the other two C-patches totally disrupted α complex assembly. The two sets of concordant mapping data indicated that

Figure 3 Enhancement of *hα-globin* 3' processing by the 3' UTR C-rich determinant is confirmed by a competition assay. (A) The α^{WT} and α^{Neut} genes were inserted 5' to the bGH 3' processing unit in pcDNA3 (188 bp, grey box). The α -globin genes contained the native 87 bp of contiguous 3' end flanking sequences. The resultant α /bGH constructs have two competing 3' end processing units, that of the *hα-globin* gene (α^{WT} or α^{Neut}) and that of the bGH gene. (B) The α^{WT} /bGH and the α^{Neut} /bGH constructs were separately transfected into MEL cells, and total RNAs were prepared after 24 h. The efficiency of *hα-globin* 3' processing in the presence of the competing bGH 3' elements was determined (as in Figure 1). In this study, the amplified fragment from the (a + c) reaction represent transcripts that are 'uncleaved' at the *hα-globin* polyA site. It is assumed that the bulk of these transcripts are cleaved at the strong competing bGH polyA site (also see text and Supplementary Figure S2). A representative gel analysis is shown and the composite results from three independent studies ($n = 3$) are represented in the histogram with $\pm$ s.d. and level of significance.

mutation of the third C-patch fully blocks $\alpha\text{CP}2$ binding to the *hα-globin* 3' UTR while individual mutations of the other sites (including mutation of the AAUAAA) had at most a partial adverse effect.

The impact of each mutation on cleavage was next assessed (see Materials and methods). An initial time-course study was carried out to assess the efficiency and accuracy of this *in vitro* cleavage reaction (Figure 4C, left panel). The *in vitro* cleavage analyses of the ^{32}P -labelled α^{WT} transcript generated two bands (Figure 4C, left panel) whose sizes (109 and 87 nt) correspond to those predicted for the 5' and 3' fragments generated by cleavage at the native polyadenylation site (16 bases 3' to the AAUAAA) (see Figure 4A). As expected, the intensity of these bands increased over time in the reaction (Figure 4C, left panel). The substrate lacking the C-rich region, α^{Neut} , markedly diminished this cleavage activity as shown by the cleavage efficiency (Figure 4C, middle panel). Several lines of evidence support the idea that the decreased intensity of cleaved bands in α^{Neut} truly reflects the decreased efficiency of cleavage, as opposed to an alteration in stability of the cleaved segments. First, our previous work (Ji *et al*, 2007) indicated that in *in vitro* reaction system, the stability of α^{WT} transcript is comparable to that of α^{Neut} . Furthermore, we calculated the ratio of processed signal + remaining substrate to the input substrate for the cleavage reactions of the α^{WT} and α^{Neut} transcripts and determined in both cases that mutation of the C-rich region (α^{Neut} transcript) does not decrease the overall signal of the RNA species (46% for both of α^{WT} and α^{Neut} after 90 min of incubation). Finally, the 3' cleavage product is identical for the α^{WT} and the

α^{Neut} transcripts, obviating any possible difference in their stabilities.

The efficiency of the cleavage reaction was calculated as the ratio of the cleavage products (3' + 5') over the initial input. Mutations at the first and second C-rich patches, which do not impact on α complex formation (Figure 4B), had no substantial adverse effect on cleavage efficiency (Figure 4C, right panel). (The origin of the doublet band at the position of the 5' cleavage product generated from the substrate with the mutation in the first C-patch, seen on multiple assays, was not further explored). In contrast, mutation of the third C-patch resulted in a substantial decrease of cleavage activity (90% decrease). The residual trace levels of cleavage products in this reaction were equivalent to that seen for the transcript containing the inactivated AAUAAA (70% decrease). These data lead us to conclude that an element co-mapping with the major C-rich binding site for αCP is essential for efficient 3' cleavage of the *hα-globin* transcript *in vitro*.

The C-rich 3' UTR determinant of *hα-globin* transcript enhances 3' polyadenylation

We next assessed the impact of the C-rich αCP -binding site on polyadenylation. Pre-cleaved substrates corresponding to the α^{WT} or a subset of derivative 3' UTRs were generated for the *in vitro* cleavage study (see Figure 4A) were assessed for $\alpha\text{CP}2$ binding in nuclear extract under the *in vitro* polyadenylation assay conditions (Figure 5A and B). These data demonstrated that mutation at the third C-patch prevents $\alpha\text{CP}2$ binding while mutations of the first or second C-patch

Figure 4 The 3' UTR α -complex enhances 3' cleavage of the *hα-globin* transcript; *in vitro* analysis. **(A)** Substrates for the *in vitro* cleavage assay. A schematic of the *hα-globin* gene, comprising 3 exons, is shown at the top of the figure. Transcripts corresponding to the full-length 109 nt *hα-globin* 3' UTR along with the 87-nt of contiguous 3' flanking sequence were generated *in vitro* and used as substrates in the cleavage reaction. The C-rich region (PR domain) encompassing the three individual C-rich patches is depicted by the grey box. The three C-rich subregions of the PR are shown in the expanded diagram in the second figure and mutations that interrupt each of these subdomains as well as mutation of the PA signal are shown below the corresponding regions. **(B)** The α^{WT} and each of the derivative substrates were assessed for α -complex assembly by a UV cross-linking label-transfer assay. Each ^{32}P -labelled RNA substrate was incubated with HeLa cell nuclear extract under cleavage conditions, UV cross-linked, RNase digested, and the residual labelled complex was visualized on a denaturing SDS-PAGE gel. The identity of the α -complex was confirmed by immunoprecipitation of the cross-linked products with antibody specific to α CP2/KL (the double-headed arrow). The designations above each lane indicate the identity of the 3' UTR sequence used in the reaction (see **A**). **(C)** The labelled RNA substrates were incubated with HeLa cell nuclear extract under cleavage conditions and the resultant products were separated on a 6% denaturing PAGE gel. The 5' and 3' cleavage fragments were of the predicted size and are indicated by arrows. (Left panel) Time course. The labelled α^{WT} cleavage RNA substrate was incubated with HeLa cell nuclear extract under cleavage conditions for indicated time period (0', 15', 30', 60', 90'), and the resultant products were separated on a 6% denaturing PAGE gel. The 5' and 3' cleavage fragments, indicated by arrows, first appear at 15' and accumulate over time. A 25-bp ladder size marker is shown (M). (Middle panel) Comparison of the *in vitro* cleavage on α^{WT} and α^{Neut} transcripts. (Right panel) Comparison of α^{WT} and its mutant variants. The efficiency of each cleavage reaction in the middle and right panels was calculated as the ratio of the cleavage products (3' + 5') over the initial input. These values were normalized ratio of products generated from the α^{WT} transcript (defined as 1.0) and are shown below the respective lanes. The time course (left panel) and the analyses of the mutant transcripts (middle and right panels) were run on different gels, accounting for the slight differences in fragment migration.

or the polyA signal had no appreciable adverse impact on α -complex assembly. A time-course analysis using the α^{WT} 3' UTR as substrate indicated that the polyadenylation increased in a linear manner over a 90-min time course (Figure 5C). The full set of *α-globin* 3' UTR transcripts was next assessed as substrates for polyadenylation. The substrates containing mutations at the first and second C-rich patches were polyadenylated as efficiently as the α^{WT} 3' UTR (Figure 5D). In contrast, mutation of the major (third) C-rich patch resulted in a marked decrease in activity. Remarkably, mutation of the polyA signal on its own had minimal adverse impact on polyA addition while addition of this mutation *in cis* to the mutation in the third C-rich patch resulted in a complete loss of activity. These findings were confirmed and extended in a second set of *in vitro* polyadenylation assays in which individual and combinations of mutations were tested using an independent preparation of nuclear extract (Figure 5E and unpublished data). Finally, we observed that

the addition of exogenous recombinant α CP2 increased the intensity of the extended polyA tail in a concentration-dependent manner while addition of a comparable amount of albumin (BSA) had no appreciable effect (Figure 5F). These data support a model in which the α CP complex assembles within the 3' UTR where it has a critical role in polyadenylation of the *hα-globin* transcript.

αCPs interact with core components of 3' end processing machinery

The *in vitro* studies (Figures 4 and 5) support a model in which the α CP complex enhances both the cleavage and the polyadenylation steps in 3' processing of the *hα-globin* transcript. To explore how α CP enhances 3' processing, we asked whether it has the capacity to recruit components of the 3' processing machinery. The core components of the 3' processing complex comprise CPSF-73, CPSF-100, and symplekin (Sullivan *et al*, 2009). CPSF-73 and CPSF-100 are

Figure 5 3' UTR α -complex enhances polyadenylation of the *h* α -globin transcript, *in vitro* analysis. (A) The pre-cleaved and 32P internally labelled substrates for polyadenylation assay, comprising the full-length (109 nt) α^{WT} 3' UTR and its mutant derivatives, were synthesized *in vitro*. (B) The ability of each of the RNA substrate to assemble the α -complex in HeLa nuclear extracts was assessed under the *in vitro* polyadenylation conditions by the 32P-label-transfer assay (as in Figure 4B). The double-headed arrow marks the position of the α -complex as confirmed by the α CP-IP analysis (right panel). The designations above each lane indicate the identity of the 3' UTR sequence used in the reaction (see Figure 4A). 3-PA designates a combination of mutations of C-rich patch 3 and poly(A) site. (C) Time-course study of *in vitro* polyadenylation assay of pre-cleaved 32P-labelled α^{WT} substrates in HeLa nuclear extract. The polyadenylation reactions were carried out as described in Materials and methods and were terminated at the indicated time points. The resultant RNAs were separated by 6% denaturing PAGE. A 25-bp ladder size marker is shown (M). A polyA tail is first seen at 15 min and steadily accumulates further during the 90-min time course of the reaction. (D) *In vitro* polyadenylation assay of pre-cleaved 32P-labelled substrates in HeLa nuclear extract. The left panel displays each of the *in vitro* transcribed RNAs prior incubation in extract ('input'). All reactions were carried out for 90 min, at which time the reaction was terminated and the resultant RNAs were separated by 6% denaturing PAGE. The efficiency of the polyadenylation reaction was calculated as the ratio of the polyadenylated signal over the initial input with α^{WT} arbitrarily set as 1.0. (E) 3' UTR α CP complex can stimulate polyadenylation independent of the AAUAAA hexanucleotide. Pre-cleaved substrates (WT, 3, PA, 3-PA, 1/2/3, 1/2/3-PA) were incubated with HeLa nuclear extract under *in vitro* polyadenylation conditions. The reaction was stopped and the reaction products were analysed by 6% denaturing PAGE. The efficiency of the polyadenylation reaction is indicated below the respective lanes. (F) α CP proteins are rate limiting for *in vitro* polyadenylation. The α^{WT} 3' UTR was incubated with HeLa cell nuclear extract. Equal amount of BSA or recombinant α CP2 were added into the reactions. The reactions were terminated after 90 min of incubation and the resultant RNAs were analysed onto 6% denaturing PAGE. All lanes, including 'input', are from the same gel and shown at the same exposure. The conversion of the input substrate to a polyadenylated product was quantified and is shown below the respective lanes (the activity at the lowest level of BSA is defined as 1.0).

Figure 6 α CP interacts with the core components of the 3' processing complex. HeLa cell nuclear extracts were subjected to immunoprecipitations with anti- α CP2 antibody or pre-immune serum. One aliquot of the extract was treated with RNase A (100 μ g/ml) before the α CP immunoprecipitation. The input and the dissolved immunoprecipitation pellets were separated onto 10% SDS-PAGE and western blots were probed with antibodies to each of the indicated proteins (see Materials and methods). The 'input' represents 10% of the equivalent extract represented in the IP lanes.

components of the CPSF complex and symplekin is a scaffold protein involved in a number of nuclear processing events (Mandel *et al*, 2008). HeLa cell nuclear extracts were immunoprecipitated with an antibody to α CP2 or with pre-immune serum. The input extract and aliquots of each immunoprecipitate were separated onto 10% SDS-PAGE and analysed by western blotting (Figure 6). A parallel set of immunoprecipitations was carried out on nuclear extracts pre-treated with RNase to detect interactions that were dependent on RNA-dependent tethering.

Consistent with previous reports, the western analysis of the α CP2 co-IP pellets revealed that α CP2 self associates (Kim *et al*, 2000). The α CP2 immunoprecipitation also enriched for all three components of the core 3' processing complex (CPSF-73, CPSF-100, and symplekin) as well as the AAUAAA-binding subunit of the CPSF complex, CPSF-160. The co-IP also revealed interaction of α CP2 with nuclear poly(A) polymerase, PAP α (Figure 6; Supplementary Figure S4). These interactions were maintained when carried out in extracts treated with RNase A or RNase One (Figure 6; Supplemental Figure S4). The analysis failed to reveal interactions of α CP2 with CstF-64, a protein that interacts with the G/U-rich 'downstream element' that is proposed to enhance 3' processing of certain transcripts (Mandel *et al*, 2008). Finally, the hnRNP L protein, a multi-functional RNA-binding protein,

was enriched in the α CP2 immunoprecipitation but the association was RNA dependent. These co-immunoprecipitation studies support a model in which α CP enhances 3' end processing by recruitment of the 3' processing complex to the *h α -globin* 3' UTR via one or more direct protein-protein interactions.

α CP is recruited to the *h α -globin* chromatin locus

Current models suggest that components of 3' end processing machinery are enriched at sites of transcription and that for some complexes, this enrichment may be facilitated by associations with elongating Pol II (Richard and Manley, 2009). To extend this model to the current study, we assessed the enrichment of α CP and Pol II at the *h α -globin* chromatin locus. Chromatin immunoprecipitation (ChIP) assay analysis was performed on K562 cells, a human erythroid cell line that expresses abundant levels of *h α -globin* mRNA (Miller *et al*, 1984). A parallel set of ChIP assays were performed using HeLa cells as a non-erythroid cell control. Chromatin extracts were incubated with antibodies to α CP2, to Pol II (N-terminal epitope; recognizing all isoforms), or with pre-immune serum. Enrichment of chromatin DNA was assessed by semi-quantitative PCR at three sites corresponding to the *h α -globin* gene promoter (α Pro), the 3' UTR (α Ex3), and a site 2 kb 3' to the native 3' processing site (3' α) (Figure 7A). The amplification signals were plotted after normalization to an internal control and to the signals obtained in a parallel IP with pre-immune serum (see Materials and methods) (Figure 7B). The analyses revealed that Pol II was associated most strongly with the 5' end, somewhat less to the exon 3 region, and markedly less enriched at 2 kb 3' to the native 3' processing site in K562 cells (Figure 7B). In contrast, α CP2 associated preferentially with the 3' end of the *h α -globin* chromatin locus in K562 cells. These findings were confirmed using a second independent chromatin preparation and a distinct set of amplification primers (data not shown). In the non-erythroid HeLa cells, α CP2 failed to interact with either 5' end or 3' end of *h α -globin* gene (Figure 7C), consistent with the fact that the *h α -globin* gene is not expressed in these cells. There was no evidence of enrichment of α CP2 or Pol II at the transcriptionally silent growth hormone locus in either K562 cells or HeLa cells. While the resolution of the ChIP mapping analysis across the *h α -globin* gene is limited by the relatively small size of the gene (0.9 kb), it is evident that α CP is recruited to the active *h α -globin* chromatin locus and that the association is most pronounced towards the 3' end of the gene. These data suggest that enrichment of α CP at the *h α -globin* chromatin locus occurs in parallel with transcription through the 3' UTR region rather than via pre-loading on Pol II at the promoter.

Discussion

We report that the 3' UTR C-rich region of the *h α -globin* mRNA, previously identified as a cytoplasmic stabilization determinant (Weiss and Liebhaver, 1995), also enhances 3' end processing of the nuclear *h α -globin* transcript as a USE (Moreira *et al*, 1998; Danckwardt *et al*, 2008). The data are consistent with the conclusion that both activities are dependent upon the RNA-binding protein, α CP, and appear to be cell-type restricted when assessed *in vivo*. The enhancement

Figure 7 α CP is recruited to the *h* α -globin chromatin locus. (A) Chromatin immunoprecipitation (ChIP) studies were carried out on K562 chromatin. The IP products were analysed by semi-quantitative PCR using amplimers that correspond to the *h* α -globin gene promoter, exon 3, and a region 2 kb 3' to the 3' processing site. (B) K562 cell chromatin extracts were incubated with antibody to α CP2, Pol II, or pre-immune serum (P.I.). DNAs from input samples or from the IP pellets were purified and used as templates for PCR assay (as in A). The non-expressing human growth hormone (*hGH*) gene is analysed in parallel. Serial dilutions of the input chromatin (1:100, 1:300, and 1:900 dilutions) confirmed that the assay was semi-quantitative (not shown). Histograms represent ratios of the PCR signal from the specific IP of K562 cells to that of the pre-immune incubation with normalization of both values to the signal generated by a reference amplimer (GAPDH; Anguita *et al*, 2004). Error bars correspond to the $\pm$ s.d. from at least two independent experiments. *P*-value for binary comparisons of the α CP2 ChIP data are shown. (C) HeLa cell ChIP. The study was carried out as described in (B).

of 3' processing by the C-rich determinant reflects enhancement of both steps of the 3' processing pathway, cleavage and polyadenylation. The assembly of the α -complex at the *h* α -globin chromatin locus and the evidence for *in vivo* interaction of α CP with core components of the 3' processing complex further support this model.

As in the case of *h* α -globin mRNA, the 3' UTR of *h* β -globin mRNAs has also been reported to interact with α CP proteins to form a stabilization complex (Yu and Russell, 2001). Interestingly, a recent study has identified an *h* β -globin gene USE element that interacts with the polypyrimidine tract binding (PTB) protein (Millevoi *et al*, 2009), which has been implicated in 3' end processing of several genes through the USE-mediated mechanism (Moreira *et al*, 1998; Hall-Pogar *et al*, 2007). It will be interesting to determine whether there is 'cross-talk' between α CP and PTB and how such interactions affect the expression of β -globin mRNAs.

A significant role for the functional interplay between nuclear processing and cytoplasmic fate of mRNAs mediated by RNA-binding protein has emerged from the recent literature. For example, the stability and decay of numerous eukaryotic mRNAs is controlled by adenosine/uridine-rich elements (AREs) located in their 3' UTRs. Zinc-finger protein, Cth2, binds to the ARE element and mediates mRNA decay. Interestingly, Cth2 protein is localized in the nucleus and is also involved in normal 3' end processing of ARE-containing mRNAs. These observations reveal that ARE-binding protein may link mRNA 3' end processing and mRNA destabilization (Prouteau *et al*, 2008).

The analysis of USE activity in the *h* α -globin 3' UTR identifies a single major C-rich patch that enhances both the 3' endocleavage of the transcript as well as the subsequent step of polyadenylation (Figures 4 and 5). Mutation of the major C-rich patch results in pronounced inhibition of cleavage comparable to that observed with inactivation of the AAUAAA (Figure 4C). Thus, it appears that the C-rich USE and the canonical AAUAAA are both required for efficient 3' cleavage of the *h* α -globin transcript, and that neither alone can fully compensate for the loss of the other. Mutation of the major C-rich region also results in a significant reduction in polyadenylation of the pre-cleaved *h* α -globin RNA substrate (Figure 5D). In contrast to the cleavage reaction, the C-rich USE of the *h* α -globin transcript can compensate (at least *in vitro*) at a significant level for the loss of the AAUAAA in the stimulation of polyA polymerase activity (Figure 5D and E).

It is most likely that the observed C-rich USE actions are mediated by enhanced recruitment of the 3' end processing machinery. Recent study revealed that the core of 3' end formation apparatus comprises three proteins, CPSF-100, CPSF-73, and symplekin (Sullivan *et al*, 2009). This core complex can be recruited to mRNAs via interaction with CPSF-160 bound to the AAUAAA, or in the case of cell-cycle-dependent histone mRNAs, by interaction with the stem-loop binding protein bound to the 3' terminal stem-loop (Sullivan *et al*, 2009). The co-immunoprecipitation studies reveal that α CP2 interacts with the all three proteins of the core complex as well as with CPSF-160 (Figure 6). These data are consistent with, and extend a model in which, α CP facilitates assembly of a macromolecular 3' processing

complex that bridges the C-rich USE with the AAUAAA element (see below).

3' end processing of both polyadenylated mRNAs and non-adenylated histone mRNAs is closely coupled to Pol II transcription (Richard and Manley, 2009; Sullivan *et al*, 2009). Once Pol II transcribes through the AAUAAA element, the CTD-associated 3' end processing factors may transfer to the nascent transcript, possibly mediated by the interaction of CPSF-160 with the AAUAAA element. The ability of α CP to interact with the C-rich elements of *h α -globin* gene and with the protein factors in the 3' processing complex, in particular with the AAUAAA-binding subunit, CPSF-160, suggest that the α CP bound to the C-rich USE of *h α -globin* gene may act as an anchor for the enhanced loading of the 3' end processing apparatus. The ChIP studies (Figure 7) suggest that α CPs may be recruited to the *h α -globin* chromatin locus as Pol II passes through the 3' region. Alternatively, it is formally possible that α CP may interact directly with 3' UTR-encoding region of the *h α -globin* gene with subsequent transfer to the nascent RNA. In either any case, the data suggest that α CP has a direct recognition and access to the 3' UTR of the nascent transcript.

It has been postulated that an important role for USEs may be to increase the efficiency of the coupled 3' end processing and transcription termination so as to minimize transcriptional interference with closely spaced downstream gene(s) (Proudfoot, 1986; Moreira *et al*, 1995). This model may be of particular relevance to the *h α 2-globin* gene given the short distance (3 kb) between its 3' processing site and the adjacent human *α 1-globin* gene promoter. The *h α 2-globin* gene has been shown to terminate transcription in close proximity to the poly(A) signal site (Whitelaw and Proudfoot, 1986) and this termination occurs more efficiently in erythroid than in non-erythroid cells (Enriquez-Harris *et al*, 1991). It would be interesting in future studies to directly test whether the presence of the C-rich USE is required to minimize interference between the two *α -globin* loci and sustain normal levels of *α -globin* gene expression.

Despite the *in vivo* erythroid cell-type specificity of the USE function (Figure 2B), we are able to model the α CP-dependent USE activity in HeLa cell extracts. Thus, the basis for this cell-type restriction of USE action in intact cells does not appear to reflect the action of an erythroid-specific factor(s). Instead, it may reflect a cell-type regulation of access and/or assembly of α CP and/or recruited 3' processing complexes to the *h α -globin* chromatin locus. This view is supported by the fact that α CP is highly enriched at the 3' end of the *h α -globin* chromatin locus in K562 cells but not at the corresponding locus in HeLa cells (Figure 7).

The present report reveals that the previously identified *h α -globin* mRNA stabilization determinant has a critical role on the 3' end processing of the *h α -globin* transcript. The likely effector α CP complex appears to assemble on the target RNA co-transcriptionally. The current report and prior studies further support a model in which the nuclear-assembled α -complex is retained on the mRNA post-processing and is co-exported along with the mature mRNA to the cytoplasm. By linking nuclear transcript processing and cytoplasmic mRNA stabilization in this manner, the α CP complex on the 3' UTR is well positioned to contribute to the robust and sustained *h α -globin* gene expression in the differentiating erythroblast.

Materials and methods

Plasmid constructions

pTet-Splice vector (Life Technologies) based pTet- α 2^{WT} (wild type) pTet- α 2^{CS}, and pTet- α 2^{Neut} plasmids have been described (Kong *et al*, 2003). The polyA site endogenous to the pTet-Splice vector was deleted in these constructs so that only the termination signals from the *h α -globin* gene were present. Mutations in the three C-rich regions (alone or in combination) and/or in the poly(A) site were introduced into the parent pTet- α 2^{WT} plasmid by PCR-based splice-overlap extension. The *EcoRI*-*NotI* fragments encompassing the *h α -globin* gene along with its contiguous 3' flanking region were excised from the parent pTet-plasmids (Kong *et al*, 2003) and re-cloned into pcDNA3 vector to establish the pcDNA3-based plasmids with two competing poly(A) sites (Figure 3).

Cell culture, transfection, extract preparations, and immunoprecipitation

Culture of human erythroleukaemia (K562) cells, cultures, and transfections of MEL/tTA and HeLa/tTA cells were performed as described (Ji *et al*, 2003; Kong *et al*, 2003). Preparation of HeLa nuclear extracts was carried out as described (Ji *et al*, 2003). For immunoprecipitation assay, 60 μ l of anti-CP2/KL rabbit serum (Chkheidze *et al*, 1999) or pre-immune rabbit serum was incubated with 140 μ l of 50% protein A Sepharose slurry in 600 μ l PBS for 1 h at 4°C with gentle rocking. Nuclear extract (100 μ l) were diluted in 600 μ l IP-binding buffer (20 mM Hepes, pH 7.9; 150 mM NaCl; 0.05% Triton-X-100). One aliquot of the extract was treated with RNase A (100 μ g/ml) for 15 min at 4°C (Ji *et al*, 2003) before the immunoprecipitation. The diluted extract was incubated with antibody-coupled protein A Sepharose for 2 h at 4°C. The immunoprecipitated complexes were pelleted by a brief low-speed spin and washed five times with IP-binding buffer. After the addition of SDS-loading buffer, the samples were analysed by SDS-PAGE.

SDS-PAGE and western blot analyses

SDS-PAGE and western blot analysis were performed as described (Ji *et al*, 2003). Affinity-purified rabbit sera to the α CP isoforms have been previously detailed and characterized (Chkheidze *et al*, 1999). Antibody to symplekin was purchased from BD Transfection Laboratories. Antibody to nuclear PAP α (A301-010A) was purchased from Bethyl Laboratories, Inc. Mouse monoclonal anti-hnRNP L antibody was a gift from Dr Gideon Dreyfuss (University of Pennsylvania). Anti-CPSF-73 was a gift from Dr David Bentley (University of Colorado). Antibodies to CstF-64 (3A7) and CPSF-160 (R7006) were gifts from Dr Clinton MacDonald (Texas Tech University Health Sciences Center). Antibody to CPSF-100 was a gift from Dr James Manley (Columbia University).

In vitro cleavage and polyadenylation assays

In vitro 3' end processing assays were performed as described with minor modifications (Gilmartin, 1997). The *in vitro* cleavage substrates contained the full 3' UTR along with 87 bp downstream genomic sequences. All processing reactions with HeLa cell nuclear extracts were carried out in a 25- μ l reaction system at 30°C for the indicated time (Gilmartin, 1997). Briefly, for cleavage reaction, the 25- μ l reaction contains 60% of nuclear extract (15 μ l), 1 mM EDTA, and 1 mM α β -methylene ATP. For the polyadenylation assay, the 25- μ l reaction comprised 50% of nuclear extract (12.5 μ l), 0.8 mM MgCl₂, 1 mM ATP, 2% PVA, and 5 mM creatine phosphate. After incubation, the reaction was terminated by adding 175 μ l stop buffer (10 mM EDTA, 10 mM Tris (pH 8.0, 0.5% SDS)) and 5 μ l proteinase K (20 mg/ml) and incubated for 10 min at 37°C. The RNA was phenol:chloroform extracted and ethanol precipitated before analysis. The samples were resolved onto 6% denaturing UREA-polyacrylamide gel and quantified by phosphorimager.

For UV cross-linking assay, the *in vitro* reactions were assembled, irradiated for 10 min at 254 nm at 4°C, followed by treatment with RNase A (0.125 μ g/ μ l) for 1 h at 37°C. Immunoprecipitations were carried out as described above. After the addition of SDS-loading buffer, the samples were analysed by SDS-PAGE.

RT-PCR and RPA

RT-PCRs and RPA were performed as described (Ji *et al*, 2003, 2007). Primers used for RT-PCR are as follows:

hα-globin a primer: 5'-GTGGACGACATGCCCAACGC-3';
hα-globin b primer: 5'-CCCACTCAGACTTTATTCAG-3';
hα-globin c primer: 5'-CACCTCCATTGTTGGCACATT-3';
bgh primer: 5'-CTGCTATTGTCTTCCCAATCCTC-3'.

Radioactivity in bands of interest was quantified by Phosphor-Imager analysis (Storm 840; Molecular Dynamics).

Chromatin immunoprecipitation assay

Chromatin extracts from K562 cells and HeLa cells were prepared according to the established protocol (Kim *et al*, 2007). ChIP assays with anti-αCP antibody (FF3; Ji *et al*, 2003), anti-Pol II antibody (Ho *et al*, 2002), and pre-immune serum were performed as described (Kim *et al*, 2007). After DNA-protein complexes were eluted and cross-linking was reversed, the DNAs in the DNA-protein complexes were extracted and used as templates for PCR assay. Primers used for ChIP assay are as follows:

hα-globin promoter (F): 5'-GGACTCCCCTGCGGTCCAG-3';
hα-globin promoter (R): 5'-GAGTCTGTGGGGACCAAGAG-3';
hα-globin exon-3 (F): 5'-ACGCTCCCTGGACAAGTTC-3';
hα-globin exon-3 (R): 5'-AGGGCCGGTGCAAGGAG-3';
hα-globin 3' flanking region (F): 5'-CACCTCCATTCTCAACCACAG-3';
hα-globin 3' flanking region (R): 5'-TTCAAGGTGACCAGGTTAGTCTG-3';
hGH (F): 5'-GCTATGCTCCGCGCCATCGTCTGC-3';
hGH (R): 5'-GAAGGACCGCCACCAAGGTCT-3'.

The reference *GAPDH* primer is as described (Anguita *et al*, 2004). The band intensities of the PCR signals were quantified on a Storm phosphorimager. The ratios of the PCR signals detected in the immunoprecipitated samples to that in the corresponding input samples were calculated. Each ratio was normalized to the corresponding ratio of the internal reference *GAPDH* signal

(Anguita *et al*, 2004). The enrichment is expressed as the ratio of each specific ChIP signal, after normalization to *GAPDH*, to that of the pre-immune serum signal. The *P*-value is calculated in each comparison of two independent UV cross-linking chromatin samples and from at least two independent ChIP-PCR assays from each chromatin preparation.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We thank Dr Russ Carstens (University of Pennsylvania) for technical advice regarding the *in vitro* processing studies. We thank Dr Gideon Dreyfuss (University of Pennsylvania) for anti-hnRNP L antibody, Dr David Bentley (University of Colorado) for anti-CPSF-73 antibody, Dr Clinton C MacDonald (Texas Tech University Health Sciences Center) for antibodies to CstF-64 (3A7) and CPSF-160 (R7006), and Dr James Manley (Columbia University) for antibody to CPSF-100. The authors appreciate the generosity of laboratory members for sharing various reagents. We appreciate the critical reviews of the manuscript by Drs Melanie Vishnu and Hemant Kini. This work was supported by an NIH MERIT HL 65449 and CA72765 to SAL.

Author contributions: XJ performed the majority of the experiments. JK contributed to the initial pulse-chase studies. XJ and SL conceived and designed the experiments and wrote the manuscript. SL supervised the current project.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Anguita E, Hughes J, Heyworth C, Blobel GA, Wood WG, Higgs DR (2004) Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J* **23**: 2841–2852
- Brackenridge S, Proudfoot NJ (2000) Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. *Mol Cell Biol* **20**: 2660–2669
- Buratowski S (2009) Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**: 541–546
- Chaudhury A, Chander P, Howe PH (2010) Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: focus on hnRNP E1's multifunctional regulatory roles. *RNA* **16**: 1449–1462
- Chen CY, Xu N, Zhu W, Shyu AB (2004) Functional dissection of hnRNP D suggests that nuclear import is required before hnRNP D can modulate mRNA turnover in the cytoplasm. *RNA* **10**: 669–680
- Chkheidze AN, Liebhaber SA (2003) A novel set of nuclear localization signals determine distributions of the alphaCP RNA-binding proteins. *Mol Cell Biol* **23**: 8405–8415
- Chkheidze AN, Lyakhov DL, Makeyev AV, Morales J, Kong J, Liebhaber SA (1999) Assembly of the alpha-globin mRNA stability complex reflects binary interaction between the pyrimidine-rich 3' untranslated region determinant and poly(C) binding protein alphaCP. *Mol Cell Biol* **19**: 4572–4581
- Clegg JB, Weatherall DJ, Milner PF (1971) Haemoglobin Constant Spring—a chain termination mutant? *Nature* **234**: 337–340
- Danckwardt S, Hentze MW, Kulozik AE (2008) 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J* **27**: 482–498
- Enriquez-Harris P, Levitt N, Briggs D, Proudfoot NJ (1991) A pause site for RNA polymerase II is associated with termination of transcription. *EMBO J* **10**: 1833–1842
- Gamarnik A, Andino R (1997) Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. *RNA* **3**: 882–892
- Gilmartin GM (1997) *In vitro* analysis of mammalian cell mRNA 3' processing. In *mRNA Formation and Function*, Richter JD (ed), pp 79–98. New York: Academic Press
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* **582**: 1977–1986
- Hall-Pogar T, Liang S, Hague LK, Lutz CS (2007) Specific transacting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *RNA* **13**: 1103–1115
- Hentze MW, Kulozik AE (1999) A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* **96**: 307–310
- Ho Y, Elefant F, Cooke N, Liebhaber SA (2002) A defined locus control region determinant links chromatin domain acetylation with long-range gene activation. *Mol Cell* **9**: 291–302
- Holcik M, Liebhaber SA (1997) Four highly stable eukaryotic mRNAs assemble 3' untranslated region RNA-protein complexes sharing cis and trans components. *Proc Natl Acad Sci USA* **94**: 2410–2414
- Ji X, Kong J, Carstens RP, Liebhaber SA (2007) The 3' untranslated region complex involved in stabilization of human alpha-globin mRNA assembles in the nucleus and serves an independent role as a splice enhancer. *Mol Cell Biol* **27**: 3290–3302
- Ji X, Kong J, Liebhaber SA (2003) *In vivo* association of the stability control protein alphaCP with actively translating mRNAs. *Mol Cell Biol* **23**: 899–907
- Keene JD (2010) Minireview: global regulation and dynamics of ribonucleic acid. *Endocrinology* **151**: 1391–1397
- Kiledjian M, Wang X, Liebhaber SA (1995) Identification of two KH domain proteins in the alpha-globin mRNP stability complex. *EMBO J* **14**: 4357–4364
- Kim TH, Barrera LO, Ren B (2007) ChIP-chip for genome-wide analysis of protein binding in mammalian cells. In *Current Protocols in Molecular Biology*, chapter 21: unit 21.13. New York: John Wiley & Sons, Inc
- Kim JH, Hahm B, Kim YK, Choi M, Jang SK (2000) Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J Mol Biol* **298**: 395–405

- Kong J, Liebhaber SA (2007) A cell type-restricted mRNA surveillance pathway triggered by ribosome extension into the 3' untranslated region. *Nat Struct Mol Biol* **14**: 670–676
- Kong J, Ji X, Liebhaber SA (2003) The KH-domain protein alphaCP has a direct role in mRNA stabilization independent of its cognate binding site. *Mol Cell Biol* **23**: 1125–1134
- Levitt N, Briggs D, Gil A, Proudfoot NJ (1989) Definition of an efficient synthetic poly(A) site. *Genes Dev* **3**: 1019–1025
- Liebhaber SA (1997) mRNA stability and the control of gene expression. *Nucleic Acids Symp Ser* **36**: 29–32
- Makeyev AV, Liebhaber SA (2002) The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA* **8**: 265–278
- Mandel CR, Bai Y, Tong L (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65**: 1099–1122
- Miller CW, Young K, Dumenil D, Alter BP, Schofield JM, Bank A (1984) Specific globin mRNAs in human erythroleukemia (K562) cells. *Blood* **63**: 195–200
- Millevoi S, Decorsière A, Loulergue C, Iacovoni J, Bernat S, Antoniou M, Vagner S (2009) A physical and functional link between splicing factors promotes pre-mRNA 3' end processing. *Nucleic Acids Res* **37**: 4672–8346
- Moore CL, Sharp PA (1985) Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell* **41**: 845–855
- Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**: 1514–1518
- Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev* **12**: 2522–2534
- Moreira A, Wollerton M, Monks J, Proudfoot NJ (1995) Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *EMBO J* **14**: 3809–3819
- Prouteau M, Daugeron MC, Séraphin B (2008) Regulation of ARE transcript 3' end processing by the yeast Cth2 mRNA decay factor. *EMBO J* **27**: 2966–2976
- Proudfoot NJ (1986) Transcriptional interference and termination between duplicated alpha-globin gene constructs suggests a novel mechanism for gene regulation. *Nature* **322**: 562–565
- Richard P, Manley JL (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev* **23**: 1247–1269
- Sullivan KD, Steiniger M, Marzluff WF (2009) A core complex of CPSF73, CPSF100, and symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. *Mol Cell* **34**: 322–332
- Thisted T, Lyakhov DL, Liebhaber SA (2001) Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and α CP-2KL, suggest distinct modes of RNA recognition. *J Biol Chem* **276**: 17484–17496
- Waggoner SA, Liebhaber SA (2003a) Identification of mRNAs associated with alphaCP2-containing RNP complexes. *Mol Cell Biol* **23**: 7055–7067
- Waggoner SA, Liebhaber SA (2003b) Regulation of alpha-globin mRNA stability. *Exp Biol Med* **228**: 387–395
- Weiss IM, Liebhaber SA (1994) Erythroid cell-specific determinants of α -globin mRNA stability. *Mol Cell Biol* **14**: 8123–8132
- Weiss IM, Liebhaber SA (1995) Erythroid cell-specific mRNA stability elements in the alpha 2-globin 3' nontranslated region. *Mol Cell Biol* **15**: 2457–2465
- Whitelaw E, Proudfoot NJ (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J* **5**: 2915–2929
- Yu J, Russell JE (2001) Structural and functional analysis of an mRNP complex that mediates the high stability of human beta-globin mRNA. *Mol Cell Biol* **21**: 5879–5888