

## 3'-End processing of pre-mRNA in eukaryotes

Elmar Wahle<sup>a,\*</sup>, Ursula Rügsegger<sup>b,1</sup>

<sup>a</sup> Institut für Biochemie, Martin-Luther-Universität Halle-Wittenberg, Kurt-Mothes-Strasse 3, 06120 Halle, Germany

<sup>b</sup> Biozentrum, Universität Basel, Klingelbergstr. 70, 4056 Basel, Switzerland

Received 30 October 1998; revised 21 December 1998; accepted 21 December 1998

### Abstract

3'-Ends of almost all eukaryotic mRNAs are generated by endonucleolytic cleavage and addition of a poly(A) tail. In mammalian cells, the reaction depends on the sequence AAUAAA upstream of the cleavage site, a degenerate GU-rich sequence element downstream of the cleavage site and stimulatory sequences upstream of AAUAAA. Six factors have been identified that carry out the two reactions. With a single exception, they have been purified to homogeneity and cDNAs for 11 subunits have been cloned. Some of the cooperative RNA-protein and protein-protein interactions within the processing complex have been analyzed, but many details, including the identity of the endonuclease, remain unknown. Several examples of regulated polyadenylation are being analyzed at the molecular level. In the yeast *Saccharomyces cerevisiae*, sequences directing cleavage and polyadenylation are more degenerate than in metazoans, and a downstream element has not been identified. The list of processing factors may be complete now with approximately a dozen polypeptides, but their functions in the reaction are largely unknown. 3'-Processing is known to be coupled to transcription. This connection is thought to involve interactions of processing factors with the mRNA cap as well as with RNA polymerase II. © 1999 Published by Elsevier Science B.V. All rights reserved.

**Keywords:** RNA processing; Poly(A) tail; Polyadenylation

### Contents

1. Introduction . . . . .	278
2. Cleavage and polyadenylation in mammalian cells . . . . .	278
2.1. Cleavage and polyadenylation signals . . . . .	278
2.2. Cleavage and polyadenylation factors . . . . .	279
2.3. Complex assembly and reaction mechanism . . . . .	282
2.4. Regulatory aspects . . . . .	285
3. Cleavage and polyadenylation in <i>S. cerevisiae</i> . . . . .	287
3.1. Cleavage and polyadenylation signals . . . . .	287

\* Corresponding author. Tel.: +49 (345) 5524920; Fax: +49 (345) 5527014; E-mail: ewahle@biochemtech.uni-halle.de

<sup>1</sup> Present address: University of California at San Francisco, Department of Biochemistry and Biophysics, 513 Parnassus Ave., San Francisco, CA 94143-0448, USA.

3.2. Cleavage and polyadenylation factors . . . . .	287
3.3. Reaction mechanism . . . . .	290
4. The relationship between 3'-end processing and transcription . . . . .	291
References . . . . .	292

## 1. Introduction

Mature eukaryotic messenger RNAs are generated from larger precursors, the primary products of RNA polymerase II, by a series of processing steps, all of which take place in the cell nucleus. 3'-End processing is one of them. All pre-mRNAs in all eukaryotic organisms first lose sequences from their 3'-ends by an endonucleolytic cleavage. Almost all RNAs then receive a polyadenylate tail (Fig. 1). The only known exceptions to this rule are the major histone mRNAs in metazoan cells. While they are also cleaved endonucleolytically, these RNAs are not polyadenylated, and the factors responsible for cleavage differ from the ones acting on all other pre-mRNAs.

The poly(A) tails at the 3'-end of the mRNA function in the initiation of translation. A protein-mediated interaction with the cap at the 5'-end of the RNA helps to load the 40S ribosomal subunit onto the message [1]. In addition, the poly(A) tails play a role in the regulation of mRNA stability as deadenylation is frequently the first and rate-limiting step of mRNA decay [2]. Finally, the poly(A) tail is thought to be involved in the export of mRNA from the nucleus into the cytoplasm [3,4].

Poly(A) tails are also found on prokaryotic

mRNAs. However, whereas polyadenylation is an essential step in the production of functional mRNA in eukaryotes, it appears to be a prelude to degradation in bacteria (see elsewhere in this volume).

Progress in the understanding of mRNA 3'-end formation has been summarized in numerous reviews [5–12]. This overview will therefore be selective rather than comprehensive, focusing on recent advances in the biochemical understanding of the mechanism of cleavage and polyadenylation in mammalian cells and in *Saccharomyces cerevisiae* and on selected regulatory aspects.

The basic 3'-end processing reaction is identical from yeast to man. However, the RNA sequences directing the process and the proteins catalyzing it are sufficiently distinct that the two systems will be discussed separately.

## 2. Cleavage and polyadenylation in mammalian cells

### 2.1. Cleavage and polyadenylation signals

Sequence requirements for 3'-end formation have been reviewed [10,13] and will be treated briefly. The reaction depends on three sequences in the pre-mRNA. Among these, the hexanucleotide AAUAAA, located 10–30 nucleotides upstream of the cleavage site, stands out by its high degree of conservation: most mRNAs have a perfect hexanucleotide; AUUAAA is the only relatively common variant, and all other deviations from the consensus are extremely rare. When these rare deviations are introduced by directed mutagenesis, they essentially inactivate a wild-type signal.

A second sequence, the so-called downstream element, is usually located roughly within the first 30 nucleotides of the cleavage site, but can also function further downstream [14]. Although several sequence motifs have been proposed based on sequence comparisons and mutagenesis studies (reviewed in [15]), a

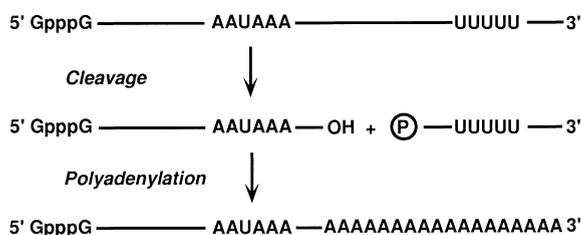


Fig. 1. The two steps in 3'-processing. Endonucleolytic cleavage between the AAUAAA sequence and the downstream element (represented by the oligo(U) stretch) generates an upstream fragment terminating with 3'-OH and a downstream fragment starting with 5'-phosphate. The upstream fragment is polyadenylated, the downstream fragment is degraded. Reproduced from [9].

convincing consensus for the downstream element was not achieved, and the element is usually described vaguely as ‘U-rich’ or ‘GU-rich’. The diffuse nature of the downstream element is also reflected in the fact that it is relatively insensitive to point mutations; sometimes extended deletions are necessary for a complete inactivation. More recently, sequence requirements for the downstream element have been analyzed by RNA selection procedures with the purified protein responsible for the recognition of this sequence. These results will be discussed below.

There is no strict requirement for the precise sequence at which cleavage occurs, but the phosphodiester bond on the 3′-side of an A is preferred, and frequently a CA dinucleotide is found at the site of cleavage. The spacing of AAUAAA and the downstream element, together with such a local sequence preference, is thought to determine the precise cleavage site [16].

Many viral and at least some cellular genes also contain sequences upstream of AAUAAA that increase the efficiency of 3′-processing. These sequences are generally U-rich but have no obvious further similarity. One particular type of upstream sequences is essential for a cytoplasmic poly(A) tail extension reaction that occurs under developmental control during oocyte maturation and early embryogenesis in a wide variety of metazoans [17] (earlier work reviewed in [10,18]). These sequences, called cytoplasmic polyadenylation elements (CPE), are usually located immediately upstream of the AAUAAA hexanucleotide and are, in contrast to other upstream elements, well defined: those active during oocyte maturation are related to UUUUUUAU whereas those active during early embryogenesis consist of more extended runs of Us.

The ability of very short AAUAAA-containing oligoribonucleotides to bind polyadenylation factors and be polyadenylated [19] suggests that this sequence is recognized in a single-stranded form, and this conclusion is supported by RNA structure prediction and probing and by mutagenesis [20–22]. The optimal conformation of the region surrounding AAUAAA is less certain. Sequence comparisons and mutagenesis indicate that a stem-loop structure of low stability with the hexanucleotide in the loop may have a moderate stimulatory effect on processing [20,23]. Another study concluded from selection

experiments that a larger unstructured region around AAUAAA favors the recognition of the hexanucleotide [22].

## 2.2. Cleavage and polyadenylation factors

Short RNAs containing the sequences discussed above are specifically cleaved and polyadenylated in crude nuclear extracts of HeLa cells. Reconstitution of the reactions with fractions derived from these extracts or from calf thymus extracts has been the basis for purification of the cleavage and polyadenylation factors.

Cleavage and polyadenylation specificity factor (CPSF) can be regarded as the central player in 3′-processing. CPSF binds directly to the cleavage and polyadenylation signal AAUAAA and, like this sequence, is required for both partial reactions. The protein contains four subunits, apparently one copy of each. Subunit association is tight as demonstrated by co-purification and co-precipitation with antibodies [24–27]. The largest polypeptide, with a molecular mass of 160 kDa (CPSF-160), can be cross-linked to AAUAAA by UV irradiation [26,28], and the recombinant subunit on its own binds this sequence, albeit with limited specificity [29]. A possible similarity of a part of the protein to a well-characterized RNA binding domain, the so-called RNP domain (reviewed by [30,31]), has not yet been tested experimentally [29]. The smallest subunit of CPSF (30 kDa) can also be cross-linked to the polyadenylation hexanucleotide [26,28]. The protein contains five zinc fingers and one zinc knuckle and binds RNA as an isolated polypeptide [27]. CPSF-30 prefers poly(U) among the homopolymers, but SELEX experiments [32] did not define a specific binding site [33]. RNA binding is partially dependent on the zinc knuckle motif. The 30-kDa subunit has been undetectable in stained SDS gels in two published CPSF preparations, suggesting that the polypeptide is held in the complex less tightly and may not be essential for CPSF function [25,34]. A requirement for mammalian CPSF-30 has not been rigorously tested. Studies on the *Drosophila* homologue of CPSF-30, *clipper*, have revealed a developmental regulation of its expression. Specifically, the protein was not detectable throughout embryogenesis [33], although transcription and, thus, RNA processing in the embryo starts

at the cellular blastoderm stage. While the limit of detection in these experiments was not defined, the results suggest that the protein might be dispensable for CPSF function. In contrast, the gene encoding the yeast protein corresponding to CPSF-30 is essential for viability (see Section 3.2). Mammalian CPSF-30 is bound by the influenza virus NS1 protein. NS1 inhibits cleavage and polyadenylation by blocking the RNA binding activity of CPSF and thus contributes to a shut-down of host gene expression [35]. The two remaining CPSF subunits of 100 and 73 kDa have also been cloned [36,37]. Their functions are unknown, and their sequences have been uninformative except for the surprising fact that they are similar to each other.

CPSF may also be involved in the recognition of upstream sequence elements. In *in vitro* experiments, purified CPSF was able to discriminate between an RNA containing the cytoplasmic polyadenylation element mentioned above and an RNA lacking this sequence; the former was preferentially bound and polyadenylated [38,39]. Similar results have been obtained with upstream elements of viral poly(A) sites [34,40]. Direct binding of CPSF to the upstream element of the HIV poly(A) site was demonstrated by a protein-RNA cross-link mapped to this region [34]. The upstream region of the HIV poly(A) site has been randomized and functional variants promoting binding of CPSF have been selected. They had in common only that they were unstructured [22]. Although this is not what one would have expected for a direct binding of CPSF to sequences upstream of AAUAAA, a structural effect does not necessarily contradict a direct interaction either. Evidence has also been presented that upstream elements can be recognized by proteins distinct from CPSF: the CPE binding protein from *Xenopus* oocytes [41–43] binds cytoplasmic polyadenylation elements, and its depletion from extracts inhibits CPE-dependent polyadenylation. The protein has not been tested in reconstitution experiments with purified polyadenylation factors. The A protein of the U1 snRNP has been suggested to bind the upstream element of the SV40 late polyadenylation signal [44,45]. Although the protein was shown to stimulate polyadenylation in a reconstituted reaction, this effect was independent of the upstream element. Recently, the polypyrimidine tract binding protein (PTB) has been shown to

bind the upstream element of the C2 complement gene, and reconstitution experiments demonstrated a stimulation of 3'-cleavage dependent on the presence of the upstream element [46].

A second well-studied protein required for cleavage is the cleavage stimulation factor (CstF) [47]. CstF binds the downstream element of the polyadenylation signal, and the protein, like its binding site, is essential for cleavage but not required for polyadenylation. The protein has three subunits of 77, 64 and 50 kDa in a 1:1:1 ratio, all of which are essential for activity [47,48]. The 77-kDa subunit connects the two others. The 64-kDa subunit can be cross-linked to substrate RNAs [47], and the cross-link has been mapped to the downstream element [49]. A clear match to the RNP domain is found in this protein [50].

Since no consensus sequence could be deduced from natural downstream elements (see above), SELEX experiments [32] have been used to determine which exact sequence is recognized by CstF. In one study, SELEX experiments were carried out with the isolated RNP domain of CstF-64 [14]. RNA molecules with repeated GU dinucleotides were selected from the random pool, in agreement with the GU-rich nature of naturally occurring downstream elements. The RNA sequences selected bound not only the isolated RNP domain but also intact CstF, and they functioned as downstream elements in cleavage assays. In an independent study, the complete CstF heterotrimer was used for selection [15]. In this case, three extended consensus sequences were selected: AUGCGUCCUCGUCC and the related sequences YGUGUYN<sub>0–4</sub>UUYAYUGYGU and UUGYUN<sub>0–4</sub>AUUUACUGN<sub>0–2</sub>YCU. All of these sequences were functional downstream elements in processing assays. The basis for the discrepancy between the two SELEX experiments is not clear. It is possible that the isolated RNP domain does not have the full sequence specificity of CstF – it certainly does not have the full affinity [14]. The results of Beyer et al. [15] may have been influenced by the fact that the 64-kDa subunit of CstF is heterogeneous [15,47]. While the basis of the heterogeneity is unknown, it is unlikely to be simply an artifact of proteolysis during purification. Thus, it is conceivable that different forms of CstF differ in their sequence preferences.

The results of the selection experiments demonstrate that the variability of the downstream element is not due to a lack of sequence specificity of CstF. Instead, the variability is probably related to the cooperative nature of substrate recognition by the processing complex: interactions among the protein components of the complex (see Section 2.3) are as important as their interactions with specific RNA sequences, so that variable RNA sequences can be accommodated in the context of a cleavage and polyadenylation signal.

In addition to the RNA binding domain, the three subunits of CstF contain other interesting sequence motifs. These include transducin repeats in the 50-kDa subunit [51] and tetratricopeptide repeats in the 77-kDa subunit, both of which are thought to mediate protein-protein interactions (reviewed by [11]). Their roles in complex assembly or processing are not yet known.

CstF-77 has strong sequence similarity to the *Drosophila* gene *suppressor of forked (su(f))* [48,52]. The phenotype of the *su(f)* mutations is consistent with a role of the protein in mRNA 3'-end formation [48,53]. Like CPSF-30, CstF-77 does not show a uniform expression in fly tissues [54]. CstF-77 autoregulates its own synthesis by promoting the use of an internal cleavage/polyadenylation site that leads to a truncated mRNA [55].

Although CstF binds the downstream polyadenylation element and is not essential for pre-mRNA polyadenylation, it can also stimulate the polyadenylation of the C2 complement mRNA by binding to the upstream polyadenylation element. As discussed above, a different factor, PTB, is thought to use the same sequence for stimulating the first step in the 3'-processing reaction [46].

Cleavage factor I (CF I<sub>m</sub>; the abbreviation distinguishes the protein from yeast CF I (CF I<sub>y</sub>; see Section 3.2), which is not related in protein structure) is a third protein required for cleavage. Like CstF, it is dispensable for polyadenylation. CF I<sub>m</sub> has been purified as a mixture of polypeptides of 72, 68, 59 and 25 kDa [56,57]. In vitro reconstitution from recombinant polypeptides demonstrated that the 68- and 25-kDa subunits are sufficient to reconstitute functional CF I<sub>m</sub> [57]. Partial cDNA sequences and antibody reactivity revealed that the 72- and 59-kDa polypeptides are related to the 68-kDa protein. Presumably,

the 25-kDa subunit can associate with any one of the large subunits, generating different forms of CF I<sub>m</sub>. The factor binds RNA in gel shift and filter binding experiments, and at least the 68-, 59- and 25-kDa polypeptides can be UV cross-linked to RNA [56]. Although a preference of CF I<sub>m</sub> for binding polyadenylation substrates was reported, the RNA binding specificity of the protein remains to be investigated more thoroughly. It is also a matter of speculation whether the different forms of CF I<sub>m</sub> have different substrate preferences. Interestingly, the sequence of CF I-68 shows the characteristic features of the so-called spliceosomal SR proteins: an N-terminal RNP domain and a C-terminal domain containing a number of serine-arginine (SR) dipeptides as well as arginines alternating with glutamate or aspartate. The protein is recognized by a monoclonal antibody diagnostic for SR proteins. In CF I-68, the RNP domain and the SR domain are connected by an extremely proline-rich region. With the exception of CF I-68, all known SR proteins play essential and/or regulatory roles in pre-mRNA splicing. In particular, they bind the RNA substrate and recruit other processing factors, thus committing the RNA to the splicing event (reviewed by [58–60]). A possibly similar role of CF I<sub>m</sub> in 3'-end processing will be discussed in Section 2.3.

An additional factor, cleavage factor II (CF II<sub>m</sub>), is also essential for cleavage [61], but it has not been purified to homogeneity, and its function is unknown.

Poly(A) polymerase is the enzyme catalyzing the addition of the poly(A) tail, but it is also involved in the cleavage reaction. In reconstituted reactions, the enzyme appears to be essential for the cleavage of some substrates and non-essential but stimulatory for at least one RNA. Poly(A) polymerase is a single monomeric polypeptide that occurs in different splice forms [62–67]. The catalytically active forms have molecular masses around 80 kDa. They differ in amino acid sequences at the C-terminus, which are irrelevant for the catalytic function. Several forms of approximately 43 kDa have also been found by cDNA cloning [63,65,67]. They are catalytically inactive (but see [65]), and, while the mRNAs are present in cells, protein expression is below the limit of detection. As the shortened mRNAs are generated by the utilization of cleavage and polyadenylation

sites within introns, they may be the product of an autoregulatory mechanism [67]. The catalytic center of poly(A) polymerase is related to those of other nucleotidyl transferases [68]: from sequence alignments and site-directed mutagenesis it has been concluded that three aspartate residues bind two divalent metal ions that are directly involved in catalyzing an in-line attack of the primer's 3'-OH group on the  $\alpha$ -phosphate of the incoming ATP. Inversion of configuration at the  $\alpha$ -phosphate is consistent with this mechanism [69]. A site for binding the RNA primer is thought to overlap with the nuclear localization signal [68,70]. Indirect evidence suggests that the closely related poly(A) polymerase from yeast may bind the RNA with two distinguishable sites [71]. A C-terminal domain of about 20 kDa is dispensable for the catalytic activity, but multiple phosphorylations in this domain can regulate the activity of the polymerase (see Section 2.4).

The minimal set of proteins directly involved in 3'-end formation is completed by the poly(A) binding protein II (PABP2; the original name PAB II was revised based on human genetic nomenclature regu-

lations), which binds poly(A) and plays a role during the elongation of poly(A) tails [72,73]. PABP2 has a monomeric molecular mass of 33 kDa. It forms various ill-defined oligomers in solution, the functional significance of which is uncertain. The protein has an N-terminus rich in glutamate residues, followed by an RNP-type RNA binding domain and an arginine-rich C-terminus [74]. PABP2 has recently received some attention because mutations leading to oculopharyngeal muscular dystrophy have been mapped to the PABP2 gene [75]. The protein carries at its N-terminus the sequence Met-Ala<sub>10</sub>-Gly-Ala<sub>2</sub>. Minor extensions of the polyalanine sequence – a recessive allele has only one additional alanine – lead to disease, most likely by causing the protein to precipitate in the cell nucleus. It remains to be elucidated how this damages the cells and why certain cell types are more affected than others.

The mammalian 3'-processing factors are summarized in Table 1.

### 2.3. Complex assembly and reaction mechanism

The proteins discussed above are sufficient to reconstitute 3'-cleavage and polyadenylation *in vitro*. As already alluded to, CPSF, CstF, CF I<sub>m</sub>, CF II<sub>m</sub> and poly(A) polymerase are involved in cleavage, whereas CPSF, poly(A) polymerase and PABP2 carry out polyadenylation. In reconstituted reactions as well as in crude extract, the two partial reactions are tightly coupled, i.e. cleavage is immediately followed by polyadenylation so that a 5'-cleavage fragment lacking poly(A) is not detectable. Coupling is presumably due to the fact that poly(A) polymerase is already involved in pre-mRNA cleavage, i.e. present in the RNA-protein complex carrying out this reaction and positioned to extend the newly created 3'-end. Coupling of cleavage and polyadenylation is, however, not obligatory: selective inhibition of polyadenylation by chain-terminating ATP analogs or by a low concentration of EDTA leads to an accumulation of the cleavage product (extended by a single nucleotide under the former condition). Also, in the case of the SV40 late polyadenylation site, cleavage occurs in the absence of poly(A) polymerase, albeit with reduced efficiency. So-called pre-cleaved RNAs, generated *in vitro* by run-off transcription so as to end at or near the natural cleavage

Table 1  
Mammalian cleavage and polyadenylation factors

Factor	Subunits (molecular mass in kDa)	Function
CPSF	160 100 73 30	cleavage and polyadenylation; binds AAUAAA; interacts with CstF, PAP and CF I <sub>m</sub>
CstF	77 64 50	cleavage; binds downstream element; interacts with CPSF
CF I <sub>m</sub>	72, 68, 59 25	cleavage; binds RNA; interacts with CPSF; functional CF I <sub>m</sub> is probably composed of one of the three large subunits and the 25-kDa subunit
CF II <sub>m</sub>	unknown	cleavage
PAP	82	cleavage and polyadenylation; catalyzes AMP polymerization; interacts with CPSF
PABP2	33	poly(A) extension; stimulates poly(A) polymerase; essential for length control

Reproduced from an article entitled 'mRNA formation: 3' end', Encyclopedia of Life Sciences, Macmillan Reference Ltd/Grove's Dictionaries Inc. (forthcoming).

site, are polyadenylated independently of the cleavage reaction. This cleavage-independent polyadenylation is AAUAAA-dependent. Cleavage leads to an upstream fragment ending with a 3'-hydroxyl and a downstream fragment with a 5'-phosphate. As this type of endonuclease reaction is usually catalyzed with the help of metal ions, the resistance of the cleavage reaction against EDTA is unusual. The downstream RNA fragment generated in the cleavage reaction is degraded unless nucleases have been removed by purification of processing factors or inactivated by sequestration of  $Mg^{2+}$  with EDTA.

Early experiments in crude extracts showed that ATP was required for the reaction not just as a substrate for poly(A) polymerase but already for the cleavage reaction and even for the assembly of the RNA-protein complex carrying out the processing reaction (reviewed in [5]). Hydrolysis of ATP, however, was not required. The ATP requirement has been re-examined in a reconstituted reaction with recombinant poly(A) polymerase and purified or partially purified CPSF, CstF, CF I<sub>m</sub> and II<sub>m</sub> [76]. The surprising result was that the reaction did not require ATP but creatine phosphate. This compound is traditionally added to replenish the ATP pool depleted by ATPases in a crude extract, but a number of controls excluded the possibility that creatine phosphate served in this function in the reconstituted reaction. Whereas ATP was dispensable with one substrate RNA it was strongly stimulatory with another. Hydrolysis of creatine phosphate or transfer of its phosphate group to proteins was not detected. Thus, creatine phosphate may have an allosteric effect. However, very high concentrations were required. Also, arginine phosphate was as efficient as creatine phosphate, and even simple inorganic phosphate substituted at detectable levels. Thus, as suggested by the authors, the phosphorylated compounds may serve as a substitute for the true physiological effector, possibly a phosphorylated protein, that remains to be identified. A possible candidate for this protein is discussed in Section 4.

Before cleavage takes place, a complex is formed between most or all of the processing factors and the substrate RNA. Initially identified in crude extracts, the 3'-processing complex has now been partially reconstituted with purified proteins. Although com-

plexes form rapidly, the cleavage reaction proceeds very slowly, even in reconstituted reactions, in which protein concentrations and reaction conditions have been optimized (e.g. [57]). Polyadenylation of a pre-cleaved substrate, in contrast, can be complete in a few seconds, at least under favorable conditions in a reconstituted reaction.

The cleavage complex is assembled in a cooperative manner: while CPSF and CstF, independently of each other, can bind the AAUAAA sequence and the downstream element, respectively (see above), they mutually stabilize their interaction with the RNA [25,77,78]. An interaction of CstF-77 with CPSF-160 has been demonstrated [29] and is presumably involved in complex assembly. The efficiency of a cleavage/polyadenylation site has been found to be reflected in the stability of its interaction with CPSF and CstF (discussed in [6]). Poly(A) polymerase on its own binds RNA very weakly and non-specifically, but it stabilizes the binding of CPSF to the RNA and thus is tethered to the RNA by CPSF [25,79]. Again, an interaction with CPSF-160 has been detected [29].

Based on the binding assays discussed above, the prevailing view was that the binding of CPSF and/or CstF are early events in the assembly of a 3'-processing complex, followed by the recruitment of other factors. However, when these assays were done, CF I<sub>m</sub> and CF II<sub>m</sub> were not available in a purified form to be tested in binding assays. Now CF I<sub>m</sub> is available and has been demonstrated to be capable of independent RNA binding (see above). Again, cooperative effects exist: CF I<sub>m</sub> stabilizes an RNA-CPSF complex in the presence or absence of CstF and poly(A) polymerase. In contrast, CF I<sub>m</sub> does not stabilize the interaction of CstF with RNA, and no factor except CPSF strengthens RNA binding of CF I<sub>m</sub>. This implies an interaction between CF I<sub>m</sub> and CPSF [56]. (CF II<sub>m</sub> has not been tested in this type of assay.) Moreover, a preincubation of the substrate RNA with CF I<sub>m</sub> has a readily detectable stimulatory effect on the cleavage efficiency, strongly suggesting that CF I<sub>m</sub> should contact the RNA first [57]. It is possible that CF I<sub>m</sub> in some manner prepares the RNA for the binding of the other proteins. For example, it has been observed that CstF prefers to bind short RNAs compared to long RNAs, suggesting that binding might be inhibited by RNA

structure [14]. Binding of CstF is also slow (K. Beyer, unpublished observations, cited in [57]).

The stimulatory effect of a preincubation of RNA and CF I<sub>m</sub> is reminiscent of the effects of SR proteins in splicing. A puzzling aspect is the following: a 30-min preincubation is sufficient to increase cleavage efficiency. However, the stimulation is due not so much to the shortening of the lag period preceding the cleavage reaction, but mostly to an increased cleavage rate that is maintained over the course of 2 h. In other words, the rate of the reaction without preincubation never catches up, even after a length of time exceeding that of the preincubation. In this context, it is a mystery why cleavage proceeds at such a slow rate. Of course, as long as CF II<sub>m</sub> has not been purified, one cannot exclude the possibility that its concentration limits the reaction rate. However, it is also possible that in the presence of saturating amounts of all proteins some intrinsically slow step, like a rearrangement of the complex, is limiting.

Another unsolved and important question concerns the identity of the nuclease cleaving the RNA. None of the polypeptides discussed above has so far been proven to be responsible for this activity. The 30-kDa subunit of CPSF has been suggested to be a ribonuclease [80]. However, as discussed above, it is possible that this subunit is not essential for CPSF activity. In addition, a yeast strain carrying a mutation in the gene encoding the homologous protein has a specific defect in polyad-

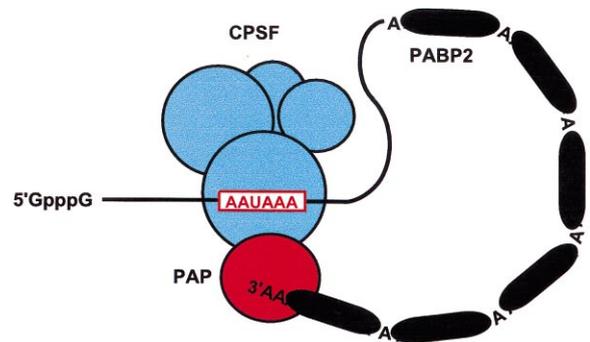


Fig. 3. A model of the polyadenylation complex. CPSF remains on the AAUAAA sequence whereas PABP2 binds the poly(A) tail. The extended structure of the tail may not be real. A direct interaction between poly(A) polymerase and PABP2 has not been proven. Reproduced from [9].

enylation, not in 3'-cleavage [27]. These data may have to be interpreted with caution, though, as discussed below.

A model of the cleavage complex is presented in Fig. 2.

Assembly of the polyadenylation complex, like that of the cleavage complex, also involves cooperative interactions. Poly(A) polymerase on its own is essentially inactive, apparently due to its weak affinity for RNA [81]. It can be activated in a non-physiological manner by Mn<sup>2+</sup> ions and will indiscriminately polyadenylate any RNA under these conditions. In the physiologically meaningful reaction, taking place in the presence of Mg<sup>2+</sup>, the enzyme interacts with CPSF on the RNA, as shown both by the stimulatory effect of CPSF on poly(A) synthesis and by gel retardation assays [25,79]. Specific binding of CPSF to AAUAAA-containing substrates is responsible for the specific polyadenylation of these RNAs. Provided that the RNA also contains an oligo(A) tail of at least 10 nucleotides, PABP2 can join the complex and further stabilize it, as demonstrated by measurements of complex half-life [79]. The effect of PABP2 on the polymerase is independent of CPSF: a strong stimulation of polymerase activity is observed on a simple poly(A) primer in the absence of CPSF [72]. The tethering of poly(A) polymerase to the RNA leads to an increased processivity of elongation: whereas the enzyme by itself is entirely distributive, i.e. dissociates after the incorporation of every single nucleotide,

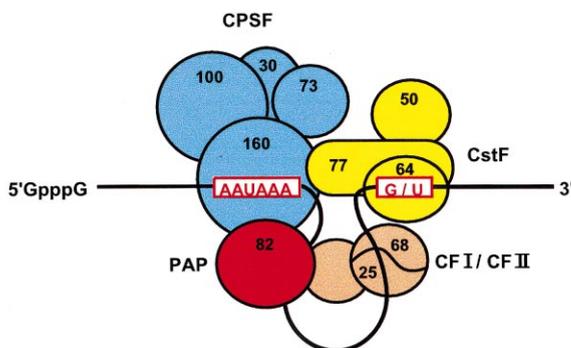


Fig. 2. A model of the cleavage complex. Whereas some of the interactions shown have an experimental basis (see text), others are hypothetical. The positions of CF I, CF II and poly(A) polymerase in the complex are not known. Reproduced from [9], with modifications.

either PABP2 or CPSF make it slightly processive. The combined activity of both stimulatory factors leads to a fully processive elongation: A complete poly(A) tail can be synthesized without dissociation of the polymerase [79].

A cartoon of the polyadenylation complex is shown in Fig. 3.

Interestingly, the processive elongation reaction terminates when approximately 250 nucleotides have been added. The further elongation is distributive and, therefore, slow. Thus, the reconstituted reaction makes poly(A) tails of the same length that is generated in vivo [82]. This length control relies on a true measurement of poly(A) tail length, as substrates with different lengths of pre-made poly(A) tails always receive the correct number of adenylate residues to result in a total length of about 250.

How the cleavage complex makes the transition to the polyadenylation complex is unknown. As discussed previously [10], there is reason to believe that the downstream fragment and the cleavage factors may not be released until a tail of 10 adenylate residues has been added to the upstream RNA fragment and PABP2 binds the growing tail. The final transition of the polyadenylation complex that leads to the termination of processive synthesis also remains to be analyzed.

#### 2.4. Regulatory aspects

Regulation of 3'-end formation has been reviewed [10,11]. Therefore, only three selected systems will be discussed in which significant progress has been made recently.

As mentioned above, poly(A) polymerase has a serine/threonine-rich C-terminal domain dispensable for catalytic activity and for interaction with other components of the cleavage and polyadenylation complexes. This domain is phosphorylated in vivo [64,66,83]. The kinase that is most likely responsible is the *cdc2/cyclin B* complex, also known as mitosis promoting factor (MPF), which is responsible for driving the cell cycle from G2 into mitosis. The evidence is as follows. The C-terminal domain of poly(A) polymerase contains several consensus sites for the kinase as well as 'non-consensus' sites that differ from the consensus sites in a single amino acid.

MPF can phosphorylate the polymerase in vivo and in vitro, full phosphorylation is dependent on the aforementioned sites, and the polymerase is most strongly phosphorylated in vivo at the time when MPF activity is at its peak [84,85]. Poly(A) polymerase activity is inhibited by phosphorylation. The effect is moderate when the steady-state population of different phosphorylated forms is compared to the unphosphorylated enzyme [69], but quite dramatic when the most highly phosphorylated form is assayed [84]. Both specific (AAUAAA- and CPSF-dependent) and non-specific ( $Mn^{2+}$ -dependent) activities are reduced. Thus, phosphorylation must affect either the active center or the interaction with the RNA primer. Phosphatase treatment activates the enzyme [84]. It is thought that the inhibition of polyadenylation contributes to a general downregulation of gene expression during mitosis.

The poly(A) polymerase is also the target in an unrelated regulatory mechanism: the U1A protein is a component of the U1 snRNP which recognizes 5'-splice sites. A specific sequence of seven nucleotides in a stem-loop structure of the U1 snRNA is required for binding of U1A. Two molecules of U1A protein also bind two conserved stem-loop structures containing the essential heptanucleotide in its own mRNA and autoregulates its expression through inhibition of 3'-processing [86]. U1A protein inhibits the polyadenylation of a substrate RNA containing the two binding sites even in the non-specific,  $Mn^{2+}$ -dependent polyadenylation reaction in which no 3'-processing factor except poly(A) polymerase is involved. U1A protein also binds to immobilized poly(A) polymerase, but only in the presence of an RNA containing two U1A binding sites.

The U1A protein binds to the RNA by means of its N-terminal RNP domain plus a few flanking amino acid residues [87,88]. The interaction with poly(A) polymerase is mediated by approximately a dozen amino acids directly adjacent to the RNA binding domain [89]. While a monomeric synthetic peptide containing these amino acids has no effect on poly(A) polymerase activity, the enzyme is strongly inhibited when the peptide is conjugated to a carrier protein at high density. Thus, the only function of the RNA with its two U1A binding sites is to effect a dimerization of the bound proteins.

Poly(A) polymerase is sensitive to U1A due to an

amino acid sequence at its extreme C-terminus: fusion of the last 20 amino acids of mammalian poly(A) polymerase confers U1A sensitivity to yeast poly(A) polymerase, which is normally insensitive, and an inert carrier protein to which this amino acid sequence has been attached binds RNA-U1A complexes [89]. Although a U1A dimer is required for binding and inhibiting poly(A) polymerase, the enzyme contains only one copy of the interacting C-terminal amino acid sequence, which also does not have any obvious repeat structure. Thus, it is likely that the two U1A proteins interact in such a way that they present a single surface for recognition by poly(A) polymerase. As in the case of the C-terminal phosphorylation of poly(A) polymerase, it will be interesting to determine how the U1A interaction at the non-essential C-terminus is transmitted to the catalytic core of the enzyme.

The significance of the regulatory mechanism just described goes beyond the regulation of U1 snRNP production because it affords insights into other regulatory phenomena and possibly also into the more general problem of the connection between splicing and cleavage/polyadenylation. Several genes have been described in which splice sites or sequences resembling them regulate cleavage and polyadenylation. In one case, found in papillomaviruses, an isolated 5'-splice site inhibits 3'-processing at a downstream cleavage/polyadenylation site [90]. The effect has been reconstituted in a simple, highly purified *in vitro* system: purified U1 snRNP inhibits the polyadenylation of a substrate RNA containing an upstream 5'-splice site. As with U1A autoregulation, inhibition is seen even in the non-specific,  $Mn^{2+}$ -dependent reaction which contains no protein except recombinant poly(A) polymerase and the inhibitory snRNP [91]. However, the inhibitor in this case is not the U1A protein, presumably because it is present in the U1 snRNP as a monomer. Instead, the U1 70 K protein is the active ingredient. This protein contains four different motifs resembling the U1A amino acid sequence responsible for inhibition of poly(A) polymerase. Simultaneous alanine replacements in all four motifs abolished the inhibitory effect on polyadenylation [91].

3'-Processing is believed to be coupled to the splicing of 3'-terminal introns. Coupling has been postulated on theoretical grounds in the context of the

'exon definition model' [92], but can also be observed experimentally: *in vitro* and *in vivo*, splicing and cleavage/polyadenylation occurring on the same RNA molecule stimulate each other [93,94]. This stimulation is disrupted by the U1A peptide-BSA conjugate mentioned above [89]. It remains to be determined which interaction exactly is sensitive to the peptide and how this stimulatory effect of splicing can be reconciled with the inhibitory effect in the case discussed above and in others.

The best-studied example of alternative polyadenylation is the synthesis of IgM heavy chains. Alternative polyadenylation is associated with alternative splicing: the upstream poly(A) site ( $\mu$ s site) is located in an intron. Cleavage and polyadenylation at this site remove two downstream exons and lead to the synthesis of secreted IgM. If this poly(A) site is not used, it is removed by splicing of the intron so that two additional exons are included in the mature mRNA, a poly(A) site (the  $\mu$ m site) further downstream is used, and the membrane-bound form of the antibody is generated.

The preference for one or the other type of RNA processing is regulated during B cell differentiation: early stages use the downstream  $\mu$ m poly(A) site and produce mostly membrane-bound IgM, whereas fully differentiated cells use the upstream  $\mu$ s poly(A) site and generate secreted antibodies. Two types of experiments strongly suggested that the poly(A) site switch does not require any gene-specific factor. In one experiment it was shown that the switch operated only when the two poly(A) sites competed in the same pre-mRNA but not when they were located in different transcription units [95,96]. In the second type, a gene was built that contained a similar arrangement of alternative splice and poly(A) sites as the IgM heavy chain gene, but was composed of different sequences. Provided that the strengths of the poly(A) sites were matched, this pre-mRNA showed the same poly(A) site switch during B cell development as the authentic gene [97]. The interpretation of these experiments, supported by additional evidence [98], was that pre-B cells are inefficient in cleavage/polyadenylation and thus prefer the stronger  $\mu$ m site. Differentiation to plasma cells entails an increase in 3'-processing efficiency; thus, the weaker  $\mu$ s site is used due to its upstream location.

CstF was suggested to be the factor limiting in

pre-B cells and upregulated during differentiation [99], and this suggestion is strongly supported by recent experiments [100]. Western blots showed that CstF-64 is expressed at least 10-fold more strongly upon induction of B-cell differentiation, whereas CstF-77 is not upregulated. Increased synthesis of CstF-64 leads to a higher abundance of heterotrimeric CstF. Artificial overexpression of CstF-64 in a cell line was sufficient to induce a preferential use of the  $\mu$ s site. As predicted by the model, the  $\mu$ m site was found to bind CstF more efficiently than the  $\mu$ s site.

### 3. Cleavage and polyadenylation in *S. cerevisiae*

#### 3.1. Cleavage and polyadenylation signals

Analysis of 3'-end formation signals in *S. cerevisiae* has been much more complicated than in mammalian cells because no highly conserved sequence motif equivalent to AAUAAA was apparent in sequence comparisons. In their degeneracy, the yeast signals are comparable to the upstream and downstream elements of mammalian poly(A) sites. Nevertheless, a large number of mutagenesis studies have finally led to the consensus that a cleavage and polyadenylation signal in *S. cerevisiae* is composed of three blocks of sequences: the nucleotides at the poly(A) addition site itself, a 'positioning element' upstream of this site and an 'efficiency element' still further upstream [101–103].

The positioning element resembles the mammalian AAUAAA sequence in at least two respects: AAUAAA is one of several related sequences that function as positioning elements, and the sequence leads to pre-mRNA cleavage approximately 20 nucleotides downstream. Cleavage occurs preferentially at PyA<sub>n</sub> sequences. It is unknown and cannot be deduced from a comparison of genomic sequences and cDNAs which phosphodiester bond exactly is cleaved, but in analogy with the mammalian system, it is likely to be on the 3'-side of an A residue. Cleavage is enhanced by the efficiency element, which is optimally located 10–20 nucleotides upstream of the positioning element, but can also function further upstream. Efficiency elements are also variable, but are related to the sequence UAUAUA.

The somewhat unexpected summary of all this is that the polyadenylation signals in yeast may not be so different from the mammalian sequences after all: the sequences at the cleavage sites are similar, the mammalian AAUAAA sequence appears to correspond to the positioning element, and the mammalian upstream elements might correspond to the yeast efficiency elements [103]. The major difference between the two types of signals is the apparent absence of a downstream element in *S. cerevisiae*. In *Schizosaccharomyces pombe*, however, a downstream element has been identified [104]. One should also keep in mind that the elements of the mammalian polyadenylation signals are defined not just by their nucleotide sequences and their positions with respect to the cleavage site, but also by the proteins which bind them. Although, as described in the following section, a number of the polypeptides involved in cleavage and polyadenylation in yeast have mammalian counterparts, including those responsible for sequence-specific RNA binding, the way in which the yeast substrate RNA is recognized is essentially unknown.

#### 3.2. Cleavage and polyadenylation factors

An in vitro system for 3'-end processing derived from *S. cerevisiae* has been available for some time [105] and has been essential for the purification of processing factors. As one would expect from an experimental system like yeast, genetic screens have also made major contributions. Perhaps more importantly, genetic experiments provide a stringent assay for testing the biological relevance of at least some of the observations made in vitro.

The nomenclature of yeast 3'-processing factors is based on an initial fractionation of crude extracts by ion exchange chromatography [106]: cleavage factors I and II (CF I<sub>y</sub> and II<sub>y</sub>; note that these proteins are not related to the mammalian factors bearing the same names) are necessary and sufficient to reconstitute cleavage. CF I<sub>y</sub>, polyadenylation factor I (PF I) and poly(A) polymerase catalyze polyadenylation of a pre-cleaved RNA. This scheme has recently led to some confusion and debate, because the association and co-purification of components of these factors turned out to be variable to some extent. While it seems likely that the inventory of polypeptides is

now complete, variable associations have made it difficult to assign functions to them.

The first 3'-processing factor from yeast to be purified and cloned was poly(A) polymerase [107,108]. The protein is encoded by an essential gene, PAPI. Pap1p is 47% identical with the mammalian enzyme over its 400 N-terminal amino acids, and the catalytic properties are also similar. In particular, like the mammalian polymerase, the yeast enzyme does not discriminate between different RNA substrates. As a striking example of the variability of association, poly(A) polymerase was initially purified as an isolated polypeptide and included as such in the scheme outlined above. However, more recently, the enzyme has been identified as a subunit of a more complex assembly, PF I (see below).

Upon purification, CF I<sub>y</sub> was separated into CF IA and CF IB. The former consists of four polypeptides [109,110]. Two subunits of 76 and 38 kDa are the products of the RNA14 and RNA15 genes, respectively, which had been previously identified in a genetic screen and shown to be components of the 3'-processing machinery by biochemical experiments [111]. The Rna15 protein contains an RNP domain and binds RNA *in vitro*. The Rna14 and Rna15 proteins have limited sequence similarity to the 77-kDa and 64-kDa subunits, respectively, of mammalian CstF (see Section 2.2). A third CF I<sub>y</sub> subunit of 70 kDa is the product of the PCF11 gene, also identified in a genetic screen [112]. Rna14, Rna15 and Pcf11 exemplify what has been found for all polypeptides involved in 3'-end processing that have been analyzed so far: the genes encoding the proteins are essential, the phenotypes of conditional-lethal mutations confirm their involvement in 3'-end formation *in vivo*, and extracts made from these mutants recapitulate the phenotype *in vitro*. A fourth CF I<sub>y</sub> subunit of 50 kDa has been termed Clp1 [110]. Mammalian sequences homologous to Clp1 have been found in the data base [12], but they do not correspond to any of the known components of the mammalian cleavage and polyadenylation factors.

An additional polypeptide found associated with CF IA is the poly(A) binding protein of yeast, encoded by the PAB1 gene. This protein, not to be confused with mammalian PABP2 (see Section 2.2), has four RNP domains and binds specifically to poly(A). It covers the poly(A) tails of mRNAs in

the cytoplasm and mediates the poly(A) tail's effects in translation and probably mRNA decay (see Introduction). The significance of its association with CF I<sub>y</sub> will be discussed below (see Section 3.3). Although mammals have a poly(A) binding protein very similar to yeast Pab1p in structure and probably at least partially equivalent in its cytoplasmic function, the mammalian protein is not known to be involved in polyadenylation.

CF IB is a single polypeptide of 73 kDa, the product of the HRP1 or NAB4 gene [113]. Initial experiments suggested that Hrp1p/Nab4p, like CF IA, was required both for cleavage and for polyadenylation. More recently, however, it has been demonstrated that Hrp1p/Nab4p is not required for cleavage *per se*, but for the selection of the correct cleavage site: in the absence of the protein, several cleavage sites upstream of the correct one were used [114]. Although an effect on polyadenylation was found in reconstitution experiments *in vitro*, no polyadenylation defect was seen *in vivo* [114]. Genetic experiments showed direct interactions of Hrp1/Nab4 with subunits of CF IA. The protein contains two RNP domains and can be UV cross-linked to cleavage/polyadenylation substrates. Hrp1p/Nab4p is a shuttling protein, i.e. it is exported from the nucleus into the cytoplasm and transported back. This is interesting in connection with the observation that polyadenylation is essential for mRNA export. Whether Hrp1p/Nab4p has indeed a direct role in the transport of mRNA remains to be elucidated. It is also interesting that Hrp1p/Nab4p has features of an hnRNP protein. None of the numerous mammalian hnRNP proteins has been shown to be involved in mRNA 3'-end formation so far.

CF II<sub>y</sub> has been purified as a heterotetramer [115]. The largest subunit (150 kDa) is the product of the CFT1 [116] or YHH1 [117] gene and homologous to the largest subunit of mammalian CPSF. The 105-kDa subunit, encoded by the CFT2 [115] or YDH1 [117] gene, is homologous to the 100-kDa subunit of CPSF. This protein could be UV cross-linked to RNA in an ATP-dependent manner in the context of the tetramer [115]. The 100-kDa subunit, product of the BRR5 [118] or YSH1 [37] gene, is related to CPSF-73. A fourth polypeptide of 90 kDa was not identified.

PF I has been purified as an assembly of at least

Table 2  
Cleavage and polyadenylation factors in yeast

Factor	Subunit (kDa)	Gene	Mammalian homologue <sup>a</sup>
CF IA <sup>b</sup>	76	RNA14	CstF-77
	72	PCF11	-
	50	CLP1	-
	38	RNA15	CstF-64
CF IB	73	HRP1	-
CF II <sub>y</sub> <sup>c</sup>	150	YHH1/CFT1	CPSF-160
	105	YDH1	CPSF-100
	100	YSH1/BRR5	CPSF-73
PF I <sup>c</sup>	90	unknown	-
	150	YHH1/CFT1	CPSF-160
	105	YDH1	CPSF-100
	100	YSH1/BRR5	CPS-73
	85	PTA1	-
	64	PAP1	poly(A) polymerase
	58	PFS1	-
	55	FIP1	-
	53	PFS2	-
26	YTH1	CPSF-30	

Reproduced from an article entitled 'mRNA formation: 3' end', Encyclopedia of Life Sciences, Macmillan Reference Ltd/Grove's Dictionaries Inc. (forthcoming).

<sup>a</sup>Mammalian homologues have been listed only as far as they are known 3'-processing factors. For some other subunits of the yeast factors, homologous mammalian sequences have been found in the data base, but it is unknown whether the polypeptides play a role in 3'-processing.

<sup>b</sup>Since PAB I is not a stoichiometric subunit of CF I<sub>y</sub> it is not listed here.

<sup>c</sup>At least three subunits of CF II are also present in PF I. Therefore, it is possible that PF I actually contains all of CF II.

nine polypeptides [117]. Surprisingly, three of them correspond to the three genetically assigned subunits of CF II<sub>y</sub> just described and, thus, the three largest subunits of CPSF. A small PF I subunit of 26 kDa, encoded by the YTH1 gene, is homologous to the 30-kDa subunit of CPSF [27]. The apparent absence of Yth1p in the CF II<sub>y</sub> preparation just discussed may well have been due to a sensitivity problem: the protein was detectable in PF I only by antibody reactivity, not by silver staining [117], and CF II<sub>y</sub> was not tested with this antibody. Poly(A) polymerase is also associated with PF I as shown by identification of the polypeptide in the preparation and by poly(A) polymerase activity of purified PF I. A subunit of 85 kDa is the product of the PTA1 gene. Although a mutation in this gene, *pta1-1*, was initially found because it accumulates unspliced tRNA

precursors, splicing endonuclease activity in extracts is normal [119]. Extracts do, however, have a defect in polyadenylation whereas cleavage of pre-mRNA is normal [117]. A subunit of 55 kDa is encoded by a gene, FIP1, that was first identified in a two-hybrid screen with poly(A) polymerase as the bait [120]. Fip1p by itself binds tightly to poly(A) polymerase [120], but, surprisingly, it inhibits the enzyme [71] while the complete PF I assembly stimulates (see Section 3.3). Two additional subunits of PF I of 58 and 53 kDa are encoded by genes named PFS1 and PFS2, respectively (P. Preker, unpublished data cited in [12]).

The 3'-processing factors of *S. cerevisiae* and their relationships to the mammalian proteins are summarized in Table 2.

With respect to the list of processing factors, the published data leave open two related questions. The first concerns the identity of CF II<sub>y</sub>. It seems likely that the CF II<sub>y</sub> described by Moore and colleagues [115] is completely contained in the PF I preparation of Preker et al. [117]. This makes the untested prediction that PF I together with CF IA and CF IB should suffice for the entire 3'-processing reaction, cleavage and polyadenylation. However, it is also possible that the CPSF homologues form a core complex that associates variably with additional polypeptides, resulting in CF II<sub>y</sub> or PF I. The second question concerns the polypeptides' roles in cleavage and polyadenylation. Since the CF/PF classification implies functions in one or the other of the two partial reactions, the assignment of subunits to the three factors is not merely a question of nomenclature. Both questions have not been resolved for the simple reason that so far the factors purified to homogeneity have only been assayed in crude systems, either by complementation of extracts prepared from mutant strains or in reconstitution assays in which at least one of the other components was not pure.

Help in assigning functions to polypeptides comes from the study of mutant extracts. However, care must be taken in the interpretation of these data. While mutations would be expected to be specific for single polypeptides, the genetic inactivation of one subunit may destabilize other components of a complex. Alternatively, an add-back experiment with an isolated polypeptide may be impossible because complex assembly can be so tight that the wild-type

polypeptide is not incorporated into the complex to replace the defective one (e.g. [117]). In addition, a point mutation may specifically affect only one of several functions of a protein.

The data currently available can be summarized as follows.

Point mutations in genes encoding three subunits of CF IA, RNA14, RNA15 and PCF11, lead to defects in both cleavage and polyadenylation in extracts [111,112]. This is in agreement with a requirement of CF IA for both reactions in reconstitution experiments. However, complementation of extracts by single polypeptides has not been achieved, so it remains to be determined if indeed each individual subunit plays a role in both reactions.

For the yeast homologues of the two largest CPSF subunits, CFT1/YHH1 and CFT2/YDH1, no conditional mutants are available. The situation is not entirely clear for BRR5/YSH1. Extracts made after genetic inactivation by two different means, including depletion of the polypeptide by transcriptional shut-off, had a clear polyadenylation defect, but cleavage was normal, suggesting that the Brr5/Ysh1 polypeptide is specifically required for polyadenylation [37,118]. However, a cold-sensitive *brr5* mutant appears to have a cleavage defect under restrictive conditions *in vivo* [118]. Thus, for the three identified subunits of CF II<sub>y</sub>, genetic evidence is incomplete. As far as it is available, it supports a role in polyadenylation, whereas a role in cleavage is uncertain.

For the smallest CPSF homologue, Yth1p, identified as a PF I subunit, *in vitro* assays with extracts from a point mutant showed a specific defect in polyadenylation, not cleavage [27]. The same is true for poly(A) polymerase: *pap1* mutants have defects in polyadenylation *in vivo* and *in vitro*, but cleavage is normal [111,121,122]. Similar experiments confirm that two other subunits of PF I, Pta1p and Fip1p are specifically involved in polyadenylation [117,120]. The available genetic information on PF I subunits is thus in good agreement with the biochemical assays, which show a requirement for PF I in polyadenylation but not in cleavage. No mutants are available for the other PF I subunits.

While the genetic data summarized above generally provide satisfying support for the biochemical experiments, reconstitution assays predict that there should be at least one polypeptide specifically re-

quired for cleavage. This should of course be a subunit of CF II<sub>y</sub> (or of PF I if this contains all of CF II<sub>y</sub>). This polypeptide remains to be identified and its role confirmed by the analysis of a mutant.

Obviously, there is a significant degree of conservation between the cleavage and polyadenylation factors of mammals and yeast [12]. Nevertheless, there are polypeptides in mammals for which no homologues have been found in the yeast genome (CstF-50, both subunits of CF I<sub>m</sub>). Likewise, several subunits of PF I have no known homologues in mammals. Other yeast polyadenylation factors (Clp1p, Pab1p) do have homologues (even well studied in the case of Pab1p), but they are not (yet) known to be involved in 3'-end formation of mRNA.

As discussed previously [12,27], the functions of sequence-related components of the 3'-end processing machineries of yeast and mammals do not appear to correspond: mammalian CstF is involved in cleavage, not polyadenylation, but the yeast counterparts of CstF subunits, RNA14 and RNA15, appear to be necessary for both reactions. Mammalian CPSF is involved in both reactions. Although the biochemical data for the yeast homologues are unclear, as these polypeptides have been found both in CF II and PF I, genetic data so far provide clear support only for a role in polyadenylation. While this apparent shift in function may seem unusual, one has to keep in mind that there are only two known catalytic functions involved in 3'-end formation, those of the anonymous endonuclease and of poly(A) polymerase. All other polypeptides have no known function but to build up the processing complex through a network of cooperative interactions. It is not so difficult to imagine that these interactions could change during the course of evolution.

### 3.3. Reaction mechanism

Like the mammalian reaction, cleavage in the yeast system is resistant to EDTA [106]. ATP stimulates cleavage, but other ribonucleoside triphosphates substitute [105,106]. The reaction also proceeds with low efficiency in the absence of added ATP [106]. Coupling between cleavage and polyadenylation is less tight than in the mammalian system: cleaved RNA lacking a poly(A) tail is readily detect-

able in an *in vitro* reaction even when no effort is made to block polyadenylation, and poly(A) polymerase is dispensable for cleavage (see Section 3.2).

Little is known about the way in which the factors described above bind their substrate RNA. Even the existence of a stable processing complex has not been demonstrated. Several polypeptides in the yeast cleavage and polyadenylation machinery are known to bind RNA (see Section 3.2), but the binding sites are mostly unknown. UV cross-linking of purified recombinant Hrp1p to RNA was dependent on the presence of the efficiency element (see Section 3.1). Hrp1p was thus suggested to be responsible for recognizing this sequence [113] (but see [114]).

As described above, purified PF I contains, in addition to poly(A) polymerase, at least eight polypeptides, including the homologues of the mammalian factor tethering poly(A) polymerase to AAUAAA-containing RNAs. The 'poly(A) polymerase holoenzyme' from yeast (= PF I) has much higher polymerase activity on a molar basis than the isolated Pap1p, possibly through an increased processivity. However, it appears that CF I<sub>y</sub> is still required for proper substrate recognition [117], although stringent specificity assays for this complex have not been published.

As mentioned above, purified CF I<sub>y</sub> contains some poly(A) binding protein 1 (Pab1p) [110]. Interactions between this protein and subunits of CF I<sub>y</sub> were also detected genetically [123]. Extracts made from *pab1* mutants have normal cleavage activity, and they also polyadenylate the upstream cleavage product. However, the poly(A) tails are approximately 40 nucleotides too long [110,123]. Addition of purified Pab1p restores normal length control to the extracts. Pab1p acts by activating the poly(A) nuclease (PAN) [124,125], which balances an excessive growth of poly(A) tails by trimming from the 3'-end [126]. Pab1p is also an inhibitor of poly(A) polymerase [107], and this might contribute to its effect on poly(A) tail length. In agreement with the biochemical data, both *pab1* and *pan* mutants have elongated poly(A) tails *in vivo* [126,127]. An additional protein, Pbp1p, affects poly(A) tail length *in vitro*, presumably through an interaction with Pab1p [128]. *In vivo*, *pbp1* mutants have normal poly(A) tail lengths. The mechanism of length control in *S. cerevisiae* appears to differ from the one described above for

the mammalian system (see Section 2.3): a different protein is involved (Pab1p in yeast versus PABP2 in mammalian cells), and the absence of the protein has opposite effects: poly(A) tail synthesis in the absence of PABP2 is inefficient, whereas the tails in the absence of Pab1p rapidly grow to an excessive length due to the lack of exonucleolytic shortening. Note however, that even in the absence of Pab1p or Pan the tails are of a relatively uniform length and only 40 nucleotides too long. Thus, an additional mechanism is likely involved in length control in yeast, which might be similar to the one in the mammalian system. The yeast genome contains an open reading frame with some sequence similarity to mammalian PABP2. The similarity is limited to the RNP-type RBD. Disruption of this open reading frame revealed that the gene is not essential, and the deletion strain has no polyadenylation defect (M. Sadowski, U. Kühn, L. Minvielle-Sebastia and E. Wahle, unpublished data).

#### 4. The relationship between 3'-end processing and transcription

Although 3'-end cleavage and polyadenylation can be studied as isolated phenomena *in vitro*, two types of evidence suggested several years ago that *in vivo* 3'-end formation is coupled to transcription. In one type of experiment, cleavage and polyadenylation were shown to be defective *in vivo* when the pre-mRNA was synthesized not by RNA polymerase II but by RNA polymerase I or III. In the second, transcription termination was shown to depend on proper 3'-end formation: mutational inactivation of the cleavage/polyadenylation signals also abolished termination (discussed in [5,6]). A biochemical basis for the connection between transcription and cleavage/polyadenylation is now beginning to emerge.

Coupling of 3'-processing to transcription is two-fold, with one indirect and one direct connection. The indirect connection is made by the 5'-modification distinguishing all mRNAs or mRNA precursors from almost all other RNAs, the m<sup>7</sup>G(5')ppp(5')G cap. Although the cap is not strictly required for 3'-end formation, it is strongly stimulatory [129,130]. The effect seems to be mediated by an interaction between the nuclear heterodimeric cap binding com-

plex (reviewed by [131]) and an as yet unidentified component of the 3'-processing complex [132]. The cap has been known for a long time to be added co-transcriptionally. After the vaccinia virus-encoded capping enzyme had first been found to associate with the viral RNA polymerase [133], capping enzymes from both yeast and mammalian cells have now been shown to bind the carboxy-terminal domain (CTD) of RNA polymerase II [134–136]. The CTD consists of a species-specific number of repeats of a heptapeptide. Binding of capping enzymes depends on phosphorylation of the repeat upon transcription initiation. Since the CTD is specific for RNA polymerase II, the association can account for the observed specificity of capping for RNA polymerase II transcripts, and the dependence of 3'-processing on capping effects its indirect dependence on RNA polymerase II transcription.

However, there is also a direct link between the polymerase and 3'-processing, and again it involves the CTD. When transcription *in vivo* is catalyzed by an RNA polymerase II from which most of the CTD has been deleted, cleavage and polyadenylation of the resulting RNA are impaired [137]. Although a failure to cap the RNA can account for some of the polyadenylation defect, even the proportion of RNA that does receive a cap is processed poorly [134]. This has been traced to an interaction between two cleavage/polyadenylation factors, CPSF and CstF, with the CTD [137]. Binding is independent of CTD phosphorylation in this case. More recently, it was reported that CPSF also binds subunits of the general transcription factor TFIID [138]. TFIID is part of the pre-initiation complex of RNA polymerase II. After the start of transcription, the association was no longer detectable, and CPSF was found to be associated with the RNA polymerase, presumably the CTD. It was thus suggested that at least some of the cleavage and polyadenylation factors are brought into the preinitiation complex of RNA polymerase II through an interaction with TFIID, handed over to the elongating polymerase after the start of transcription and then bind the nascent RNA once a proper binding site has been generated.

These data provide an additional plausible explanation for the dependence of 3'-end formation on RNA polymerase II transcription. They do not explain the dependence of transcription termination on

cleavage/polyadenylation, but they suggest that the elongating transcription complex undergoes a rearrangement by transfer of CPSF and possibly other factors to the RNA after the polyadenylation signal has been passed. It is conceivable that this change in the complex plays a role in causing the polymerase to terminate downstream of the poly(A) site. However, there is reason to believe that a change in the transcription complex by transfer of proteins is not sufficient to cause termination: In a survey of different yeast mutants deficient in 3'-processing, all mutations affecting cleavage had a termination defect whereas all mutants with a defect in polyadenylation had normal termination. From these and other data it was concluded that cleavage of the pre-mRNA may be essential to cause transcription termination [139].

It has been reported that in a crude nuclear extract, cleavage and polyadenylation of pre-mRNA generated *in situ* by RNA polymerase II is much more efficient than processing of exogenous RNA [140]. Similar experiments in which a reconstituted RNA polymerase II transcription system has been combined with a reconstituted cleavage/polyadenylation system have not yet been reported. However, recent results suggest that RNA polymerase II may have a role in 3'-end cleavage beyond delivering the proteins which carry out the job. In a reconstituted system that cleaved pre-made RNA precursors, independently of transcription, RNA polymerase II, and in particular its C-terminal domain, behaved like a cleavage factor [141]. Its depletion from nuclear extracts inhibited 3'-processing, and the efficiency of the reaction could be restored by the addition of purified enzyme. Although these experiments were prompted by the search for a phosphorylated protein that might replace creatine phosphate in the assay (see Section 2.3), phosphorylation of the CTD was not essential. Under some conditions, the phosphorylated form was more efficient, though. The exact role of RNA polymerase II and its CTD remains to be investigated.

## References

- [1] Sachs, A.B., Sarnow, P. and Hentze, M.W. (1997) *Cell* 89, 831–838.

- [2] Beelman, C.A. and Parker, R. (1995) *Cell* 81, 179–183.
- [3] Eckner, R., Ellmeier, W. and Birnstiel, M.L. (1991) *EMBO J.* 10, 3513–3522.
- [4] Huang, Y. and Carmichael, G.G. (1996) *Mol. Cell. Biol.* 16, 1534–1542.
- [5] Wahle, E. and Keller, W. (1992) *Annu. Rev. Biochem.* 61, 419–440.
- [6] Wahle, E. (1995) *Biochim. Biophys. Acta* 1261, 183–194.
- [7] Keller, W. (1995) *Cell* 81, 829–832.
- [8] Manley, J.L. (1995) *Curr. Opin. Genet. Dev.* 5, 222–228.
- [9] Wahle, E. and Keller, W. (1996) *Trends Biochem. Sci.* 21, 247–250.
- [10] Wahle, E. and Kühn, U. (1997) *Progr. Nucleic Acid Res. Mol. Biol.* 57, 41–71.
- [11] Colgan, D.F. and Manley, J.L. (1997) *Genes Dev.* 11, 2755–2766.
- [12] Keller, W. and Minvielle-Sebastia, L. (1997) *Curr. Opin. Cell Biol.* 9, 329–336.
- [13] Proudfoot, N. (1991) *Cell* 64, 671–674.
- [14] Takagaki, Y. and Manley, J.L. (1997) *Mol. Cell. Biol.* 17, 3907–3914.
- [15] Beyer, K., Dandekar, T. and Keller, W. (1997) *J. Biol. Chem.* 272, 26729–26779.
- [16] Chen, F., MacDonald, C.C. and Wilusz, J. (1995) *Nucleic Acids Res.* 23, 2614–2620.
- [17] Verrotti, A.C., Thompson, S.R., Wreden, C., Strickland, S. and Wickens, M. (1996) *Proc. Natl. Acad. Sci. USA* 93, 9027–9032.
- [18] Richter, J.D. (1996) in: *Translational Control* (Hershey, J., Sonenberg, N. and Mathews, M., Eds.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [19] Wigley, P.L., Sheets, M.D., Zarkower, D.A., Whitmer, M.E. and Wickens, M. (1990) *Mol. Cell. Biol.* 10, 1705–1713.
- [20] Sittler, A., Gallinaro, H. and Jacob, M. (1995) *J. Mol. Biol.* 248, 525–540.
- [21] Graveley, B.R., Fleming, E.S. and Gilmartin, G.M. (1996) *J. Biol. Chem.* 271, 33654–33663.
- [22] Graveley, B.R., Fleming, E.S. and Gilmartin, G.M. (1996) *Mol. Cell. Biol.* 16, 4942–4951.
- [23] Berkhout, B., Klaver, B. and Das, A.T. (1995) *Virology* 207, 276–281.
- [24] Bienroth, S., Wahle, E., Suter-Crazzolara, C. and Keller, W. (1991) *J. Biol. Chem.* 266, 19768–19776.
- [25] Murthy, K.G.K. and Manley, J.L. (1992) *J. Biol. Chem.* 267, 14804–14811.
- [26] Jenny, A., Hauri, H.-P. and Keller, W. (1994) *Mol. Cell. Biol.* 14, 8183–8190.
- [27] Barabino, S.M.L., Hübner, W., Jenny, A., Minvielle-Sebastia, L. and Keller, W. (1997) *Genes Dev.* 11, 1703–1716.
- [28] Keller, W., Bienroth, S., Lang, K.M. and Christofori, G. (1991) *EMBO J.* 10, 4241–4249.
- [29] Murthy, K.G.K. and Manley, J.L. (1995) *Genes Dev.* 9, 2672–2683.
- [30] Burd, C.G. and Dreyfuss, G. (1994) *Science* 265, 615–621.
- [31] Nagai, K. (1996) *Curr. Opin. Struct. Biol.* 6, 53–61.
- [32] Tuerk, C. and Gold, L. (1990) *Science* 249, 505–510.
- [33] Bai, C. and Tolias, P.P. (1998) *Nucleic Acids Res.* 26, 1597–1604.
- [34] Gilmartin, G.M., Fleming, E.S., Oetjen, J. and Graveley, B.R. (1995) *Genes Dev.* 9, 72–83.
- [35] Nemeroff, M.E., Barabino, S.M.L., Li, Y., Keller, W. and Krug, R.M. (1998) *Mol. Cell.* 1, 991–1000.
- [36] Jenny, A. and Keller, W. (1995) *Nucleic Acids Res.* 23, 2629–2635.
- [37] Jenny, A., Minvielle-Sebastia, L., Preker, P.J. and Keller, W. (1996) *Science* 274, 1514–1517.
- [38] Fox, C.A., Sheets, M.D., Wahle, E. and Wickens, M. (1992) *EMBO J.* 11, 5021–5032.
- [39] Bilger, A., Fox, C.A., Wahle, E. and Wickens, M. (1994) *Genes Dev.* 8, 1106–1116.
- [40] Graveley, B.R. and Gilmartin, G.M. (1996) *J. Virol.* 70, 1612–1617.
- [41] Hake, L.E. and Richter, J.D. (1994) *Cell* 79, 617–627.
- [42] Gebauer, F. and Richter, J.D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 14602–14607.
- [43] Stebbins-Boaz, B., Hake, L.E. and Richter, J.D. (1996) *EMBO J.* 15, 2582–2592.
- [44] Lutz, C.S. and Alwine, J.C. (1994) *Genes Dev.* 8, 576–586.
- [45] Lutz, C.S., Murthy, K.G.K., Schek, N., O'Connor, J.P., Manley, J.L. and Alwine, J.C. (1996) *Genes Dev.* 10, 325–337.
- [46] Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L. and Proudfoot, N.J. (1998) *Genes Dev.* 12, 2522–2534.
- [47] Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J. and Shenk, T. (1990) *Genes Dev.* 4, 2112–2120.
- [48] Takagaki, Y. and Manley, J.L. (1994) *Nature* 372, 471–474.
- [49] MacDonald, C.C., Wilusz, J. and Shenk, T. (1994) *Mol. Cell. Biol.* 14, 6647–6654.
- [50] Takagaki, Y. and Manley, J.L. (1992) *Proc. Natl. Acad. Sci. USA* 89, 1403–1407.
- [51] Takagaki, Y. and Manley, J.L. (1992) *J. Biol. Chem.* 267, 23471–23474.
- [52] Simonelig, M., Elliott, K., Mitchelson, A. and O'Hare, K. (1996) *Genetics* 142, 1225–1235.
- [53] Mitchelson, A., Simonelig, M., Williams, C. and O'Hare, K. (1993) *Genes Dev.* 7, 241–249.
- [54] Audibert, A., Juge, F. and Simonelig, M. (1998) *Mech. Dev.* 72, 53–63.
- [55] Audibert, A. and Simonelig, M. (1998) *Proc. Natl. Acad. Sci. USA* 95, 14302–14307.
- [56] Rügsegger, U., Beyer, K. and Keller, W. (1996) *J. Biol. Chem.* 271, 6107–6113.
- [57] Rügsegger, U., Blank, D. and Keller, W. (1998) *Mol. Cell* 1, 243–253.
- [58] Fu, X.-D. (1995) *RNA* 1, 663–680.
- [59] Manley, J.L. and Tacke, R. (1996) *Genes Dev.* 10, 1569–1579.
- [60] Valcárcel, J. and Green, M.R. (1996) *Trends Biochem. Sci.* 21, 296–301.
- [61] Takagaki, Y., Rhyner, L.C. and Manley, J.L. (1989) *Genes Dev.* 3, 1711–1724.
- [62] Raabe, T., Bollum, F.J. and Manley, J.L. (1991) *Nature* 353, 229–234.

- [63] Wahle, E., Martin, G., Schiltz, E. and Keller, W. (1991) *EMBO J.* 10, 4251–4257.
- [64] Thureson, A.-C., Aström, J., Aström, A., Grönvik, K.-O. and Virtanen, A. (1994) *Proc. Natl. Acad. Sci. USA* 91, 979–983.
- [65] Gebauer, F. and Richter, J.D. (1995) *Mol. Cell. Biol.* 15, 1422–1430.
- [66] Ballantyne, S., Bilger, A., Astrom, J., Virtanen, A. and Wickens, M. (1995) *RNA* 1, 64–78.
- [67] Zhao, W. and Manley, J.L. (1996) *Mol. Cell. Biol.* 16, 2378–2386.
- [68] Martin, G. and Keller, W. (1996) *EMBO J.* 15, 2593–2603.
- [69] Wittmann, T. and Wahle, E. (1997) *Biochim. Biophys. Acta* 1350, 293–305.
- [70] Zhelkovsky, A.M., Kessler, M.M. and Moore, C.L. (1995) *J. Biol. Chem.* 270, 26715–26720.
- [71] Zhelkovsky, A., Helmling, S. and Moore, C. (1998) *Mol. Cell. Biol.* 18, 5942–5951.
- [72] Wahle, E. (1991) *Cell* 66, 759–768.
- [73] Wahle, E., Lustig, A., Jenö, P. and Maurer, P. (1993) *J. Biol. Chem.* 268, 2937–2945.
- [74] Nemeth, A., Krause, S., Blank, D., Jenny, A., Jenö, P., Lustig, A. and Wahle, E. (1995) *Nucleic Acids Res.* 23, 4034–4041.
- [75] Brais, B. et al. (1998) *Nature Genet.* 18, 164–167.
- [76] Hirose, Y. and Manley, J.L. (1997) *J. Biol. Chem.* 272, 29636–29642.
- [77] Gilmartin, G.M. and Nevins, J.R. (1989) *Genes Dev.* 3, 2180–2189.
- [78] Wilusz, J., Shenk, T., Takagaki, Y. and Manley, J.L. (1990) *Mol. Cell. Biol.* 10, 1244–1248.
- [79] Bienroth, S., Keller, W. and Wahle, E. (1993) *EMBO J.* 12, 585–594.
- [80] Bai, C. and Tolia, P.P. (1996) *Mol. Cell. Biol.* 16, 6661–6667.
- [81] Wahle, E. (1991) *J. Biol. Chem.* 266, 3131–3139.
- [82] Wahle, E. (1995) *J. Biol. Chem.* 270, 2800–2808.
- [83] Raabe, T., Murthy, K.G.K. and Manley, J.L. (1994) *Mol. Cell. Biol.* 14, 2946–2957.
- [84] Colgan, D.F., Murthy, K.G.K., Prives, C. and Manley, J.L. (1996) *Nature* 384, 282–285.
- [85] Colgan, D.F., Murthy, K.G.K., Zhao, W., Prives, C. and Manley, J.L. (1998) *EMBO J.* 17, 1053–1062.
- [86] Gunderson, S.J., Beyer, K., Martin, G., Keller, W., Boelens, W.C. and Mattaj, I.W. (1994) *Cell* 76, 531–541.
- [87] Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H. and Nagai, K. (1994) *Nature* 372, 432–438.
- [88] Avis, J.M., Allain, F.H.-T., Howe, P.W.A., Varani, G., Nagai, K. and Neuhaus, D. (1996) *J. Mol. Biol.* 257, 398–411.
- [89] Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M. and Mattaj, I.W. (1997) *Genes Dev.* 11, 761–773.
- [90] Furth, P.A., Choe, W.-T., Rex, J.H., Byrne, J.C. and Baker, C.C. (1994) *Mol. Cell. Biol.* 14, 5278–5289.
- [91] Gunderson, S.I., Polycarpou-Schwarz, M. and Mattaj, I.W. (1998) *Mol. Cell* 1, 255–264.
- [92] Robberson, B.L., Cote, G. and Berget, S.M. (1990) *Mol. Cell. Biol.* 10, 84–94.
- [93] Niwa, M. and Berget, S.M. (1991) *Genes Dev.* 5, 2086–2095.
- [94] Niwa, M., Rose, S.D. and Berget, S.M. (1990) *Genes Dev.* 4, 1552–1559.
- [95] Peterson, M.L. and Perry, R.P. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8883–8887.
- [96] Galli, G., Guise, J.W., McDevitt, M.A., Tucker, P.W. and Nevins, J.R. (1987) *Genes Dev.* 1, 471–481.
- [97] Peterson, M.L. (1994) *Mol. Cell. Biol.* 14, 7891–7898.
- [98] Peterson, M.L., Gimmi, E.R. and Perry, R.P. (1991) *Mol. Cell. Biol.* 11, 2324–2327.
- [99] Edwalds-Gilbert, G. and Milcarek, C. (1995) *Mol. Cell. Biol.* 15, 6420–6429.
- [100] Takagaki, Y., Seipelt, R.L., Peterson, M.L. and Manley, J.L. (1996) *Cell* 87, 941–952.
- [101] Guo, Z. and Sherman, F. (1995) *Mol. Cell. Biol.* 15, 5983–5990.
- [102] Guo, Z. and Sherman, F. (1996) *Mol. Cell. Biol.* 16, 2772–2776.
- [103] Guo, Z. and Sherman, F. (1996) *Trends Biochem. Sci.* 21, 477–481.
- [104] Humphrey, T., Birse, C.E. and Proudfoot, N.J. (1994) *EMBO J.* 13, 2441–2451.
- [105] Butler, J.S. and Platt, T. (1988) *Science* 242, 1270–1274.
- [106] Chen, J. and Moore, C. (1992) *Mol. Cell. Biol.* 12, 3470–3481.
- [107] Lingner, J., Radtke, I., Wahle, E. and Keller, W. (1991) *J. Biol. Chem.* 266, 8741–8746.
- [108] Lingner, J., Kellermann, J. and Keller, W. (1991) *Nature* 354, 496–498.
- [109] Kessler, M.M., Zhao, J. and Moore, C.L. (1996) *J. Biol. Chem.* 271, 27167–27175.
- [110] Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y. and Keller, W. (1997) *Proc. Natl. Acad. Sci. USA* 94, 7897–7902.
- [111] Minvielle-Sebastia, L., Preker, P.J. and Keller, W. (1994) *Science* 266, 1702–1705.
- [112] Amrani, N., Minet, M., Wyers, F., Dufour, M.-E., Aggerbeck, L.P. and Lacroute, F. (1997) *Mol. Cell. Biol.* 17, 1102–1109.
- [113] Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A. and Moore, C.L. (1997) *Genes Dev.* 11, 2545–2556.
- [114] Minvielle-Sebastia, L., Beyer, K., Krecic, A.M., Hector, R.E., Swanson, M.S. and Keller, W. (1998) *EMBO J.* 17, 7454–7468.
- [115] Zhao, J., Kessler, M.M. and Moore, C.L. (1997) *J. Biol. Chem.* 272, 10831–10838.
- [116] Stumpf, G. and Domdey, H. (1996) *Science* 274, 1517–1520.
- [117] Preker, P.J., Ohnacker, M., Minvielle-Sebastia, L. and Keller, W. (1997) *EMBO J.* 16, 4727–4737.
- [118] Chanfreau, G., Noble, S.M. and Guthrie, C. (1996) *Science* 274, 1511–1514.
- [119] O'Connor, J.P. and Peebles, C.L. (1992) *Mol. Cell. Biol.* 12, 3843–3856.
- [120] Preker, P.J., Lingner, J., Minvielle-Sebastia, L. and Keller, W. (1995) *Cell* 81, 379–389.
- [121] Patel, D. and Butler, J.S. (1992) *Mol. Cell. Biol.* 12, 3297–3304.

- [122] Mandart, E. and Parker, R. (1995) *Mol. Cell. Biol.* 15, 6979–6986.
- [123] Amrani, N., Minet, M., Le Gouar, M., Lacroute, F. and Wyers, F. (1997) *Mol. Cell. Biol.* 17, 3694–3701.
- [124] Boeck, R., Tarun, S., Rieger, M., Deardorff, J.A., Müller-Auer, S. and Sachs, A.B. (1996) *J. Biol. Chem.* 271, 432–438.
- [125] Brown, C.E., Tarun, S.Z., Boeck, R. and Sachs, A.B. (1996) *Mol. Cell. Biol.* 16, 5744–5753.
- [126] Brown, C.E. and Sachs, A.B. (1998) *Mol. Cell. Biol.* 18, 6548–6559.
- [127] Sachs, A.B. and Davis, R.W. (1989) *Cell* 58, 857–867.
- [128] Mangus, D.A., Amrani, N. and Jacobson, A. (1998) *Mol. Cell. Biol.* 18, 7383–7396.
- [129] Hart, R.P., McDevitt, M.A. and Nevins, J.R. (1985) *Cell* 43, 677–683.
- [130] Cooke, C. and Alwine, J.C. (1996) *Mol. Cell. Biol.* 16, 2579–2584.
- [131] Lewis, J.D. and Izaurralde, E. (1997) *Eur. J. Biochem.* 247, 461–469.
- [132] Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W. and Gilmartin, G.M. (1997) *Proc. Natl. Acad. Sci. USA* 94, 11893–11898.
- [133] Shuman, S. (1995) *Prog. Nucleic Acids Res. Mol. Biol.* 50, 101–129.
- [134] McCracken, S. et al. (1997) *Genes Dev.* 11, 3306–3318.
- [135] Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S. (1997) *Genes Dev.* 11, 3319–3326.
- [136] Yue, Z., Maldonado, E., Pillutila, R., Cho, H., Reinberg, D. and Shatkin, A.J. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12898–12903.
- [137] McCracken, S. et al. (1997) *Nature* 385, 357–361.
- [138] Dantonel, J.-C., Murthy, K.G.K., Manley, J.L. and Tora, L. (1997) *Nature* 389, 399–402.
- [139] Birse, E.C., Minvielle-Sebastia, L., Lee, B.A., Keller, W. and Proudfoot, N.J. (1998) *Science* 280, 298–300.
- [140] Mifflin, R.C. and Kellems, R.E. (1991) *J. Biol. Chem.* 266, 19593–19598.
- [141] Hirose, Y. and Manley, J.L. (1998) *Nature* 395, 93–96.