How the messenger got its tail: addition of poly(A) in the nucleus

Marvin Wickens

Most mRNAs end in a poly(A) tail, the addition of which is catalysed by a poly(A) polymerase in conjunction with a distinct factor that provides specificity for mRNAs. The reaction is dynamic, involving separable initiation, elongation and termination phases. A companion article in next month's *TIBS* will review the regulation of poly(A) addition and removal during early animal development.

NEARLY ALL mRNAs in eukaryotic cells end in a homopolymer of 20–250 adenosine nucleotides. This poly(A) 'tail', which is added in the nucleus following transcription, is usually shortened once the mRNA enters the cytoplasm. The poly(A) tail both stimulates translation and stabilizes the mRNA. This review will outline the biochemistry of poly(A) addition in the nucleus, which has been clarified considerably through recent work in several laboratories. The mechanism of nuclear polyadenylation is important not only in its own right, but also because it provides a useful perspective from which to view regulated, mRNA-specific polyadenylation early in development. In a companion article in next month's issue of *TIBS*, we will concentrate on those regulated events.

Uncoupling polyadenylation from cleavage

Polyadenylation in the nucleus[1] and in a cell-free nuclear extract[2] immediately follows a cleavage reaction in which a phosphodiester bond in the mRNA precursor is broken (Fig. 1). *In vivo*, this cleavage reaction occurs while the precursor is still being synthesized, with the advancing RNA polymerase a considerable distance downstream[1]. Cleavage leaves a 3' hydroxyl and a 5' phosphate, but otherwise appears not to modify the RNA[3,4]. Adenosine nucleotides are then added to the new 3' hydroxyl group, one at a time[2–4].

From an experimental standpoint, cleavage complicates analysis of poly(A) addition: for example, features of the substrate that are critical for poly(A) addition can only be identified if they are different from those required for cleavage. To obviate these problems and to analyse poly(A) addition directly, synthetic RNAs that end at the natural poly(A) addition site have been used[5]. These substrates are equivalent to natural intermediates that have just been cleaved. Most of the work described here has been performed using substrates of this type incubated in a nuclear extract derived from a transformed human cell line (HeLa)[2].

The uncoupling of polyadenylation from cleavage, however convenient, is unnatural: *in vivo*, once an RNA is cut, poly(A) addition follows immediately. Here, however, we will focus on poly(A) addition *per se*, and discuss cleavage only in that context. The reader interested in the mechanism of cleavage is directed to excellent reviews of mRNA 3' end formation[6–8].

The substrate: AAUAAA and a minimum distance

Polyadenylation, like cleavage, requires the sequence AAUAAA[5,9] (M. D. Sheets, S. Ogg and M. Wickens, submitted). This hexanucleotide is typically located 15–25 nucleotides upstream of the poly(A) addition site of mRNAs and is very highly conserved (Fig. 2). Every base in this sequence is required, although the stringency of the requirement varies with the position and the base substituted (Fig. 2)[10] (Sheets, Ogg and Wickens, op. cit.). The only common natural variant, AUUAAA, is nearly as active as the canonical sequence[10] (Sheets, Ogg and Wickens, op. cit.).

AAUAAA is not only required for polyadenylation, but also is virtually sufficient: very short substrates – as short as 11 nucleotides – support polyadenylation, providing they contain AAUAAA and a minimum distance between AAUAAA and the 3' end. Polyadenylation of these short substrates is optimal with eight or more nucleotides downstream (3') of AAUAAA[11]. The identity of the downstream nucleotides appears to be relatively inconsequential, however[11,12].

The cleavage reaction that normally precedes polyadenylation requires both AAUAAA and additional sequences located downstream of the poly(A) addition site[6,7]. These 'downstream elements' are generally U- or UG-rich, but are less highly conserved than AAUAAA. Although quantitatively they are less significant than AAUAAA for cleavage, they nonetheless are critical. The failure of internal AAUAAA sequences (in introns, for example) to direct cleavage is due to their lack of a companion downstream element. Furthermore, downstream elements may also regulate the choice between alternative cleavage sites[13].

AAUAAA-specificity factor and poly(A) polymerase

Polyadenylation of RNAs that end at the cleavage site requires at least two separable activities: an AAUAAA-specificity factor and an enzyme that polymerizes ATP into poly(A) (Fig. 3)[14–20]. Although neither activity has yet been purified to homogeneity from the HeLa extract, important features of each activity have emerged.

A poly(A) polymerizing activity was first detected 30 years ago in crude cellu-

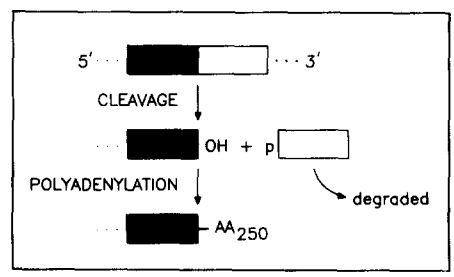


Figure 1
Polyadenylation is normally coupled to cleavage. In this and all subsequent figures, a black box indicates the portion of the RNA upstream of the poly(A) addition site, while a white box indicates the portion downstream. The poly(A) addition site is at the junction between the two. Cleavage generates an upstream half-molecule carrying a 3'OH, to which approximately 250 adenosine nucleotides [poly(A)] are then added.

M. Wickens is at the Department of Biochemistry, The Graduate School, College of Agriculture and Life Sciences, University of Wisconsin–Madison, Madison, WI 53796, USA.

 0376–5067/90/$02.00

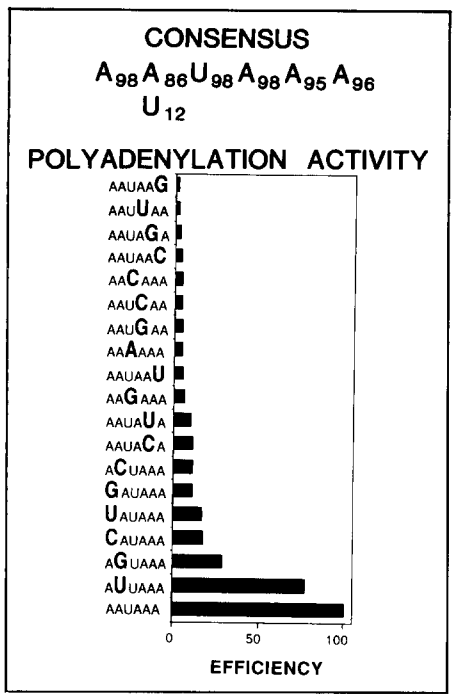


Figure 2

AAUAAA: conservation and analysis of point mutations. The consensus sequence was compiled by comparison of 269 cDNA sequences from vertebrates in the EMBL database (M. D. Sheets, S. Ogg and M. Wickens, submitted). Sequences from non-vertebrates were not included. Subscript numerals indicate the percentage frequency of occurrence. The only common natural variant sequence is AUUAAA; it appears in 12% of the mRNAs compared. Polyadenylation activity was determined by incubating a labeled SV40 late RNA that ends at the poly(A) addition in a HeLa cell nuclear extract (Sheets, Ogg and Wickens, op. cit.). Analysis of cleavage, using substrates that contain a downstream element, yields a very similar rank order (Sheets, Ogg and Wickens, op. cit.).

lar extracts and was partially purified in the mid-1970s (reviewed in Refs 21, 22). The biological function of this activity was uncertain, since it would add poly(A) to decidedly unnatural substrates such as tRNA. It is now clear that these classical poly(A) polymerases are responsible for mRNA-specific polyadenylation. The activity partially purified from HeLa nuclear extracts exhibits the general biochemical properties of classical poly(A) polymerases[21,22], including its behavior during biochemical fractionation, apparent molecular weight (40–60 kDa)[13,16,18] and antigenicity[20]. Furthermore, a classical poly(A) polymerase, purified from calf thymus by virtue of its activity on a non-mRNA substrate, becomes specific for AAUAAA-containing RNAs when mixed with a distinct 'specificity factor' fraction from HeLa cells (Fig. 4)[19]. Thus, with hindsight, it is clear that poly(A) polymerase was the first enzyme involved in mRNA processing to have been identified and characterized.

In the simplest view, selectivity for AAUAAA-containing RNAs is the result of a direct contact between specificity factor bound to AAUAAA and poly(A) polymerase. This contact, which has yet to be demonstrated directly, probably decreases the effective dissociation constant (K_d) of the polymerase for AAUAAA-containing substrates. At sufficiently high substrate or enzyme concentrations, as in classical purifications, poly(A) polymerase adds poly(A) to any RNA[21,22]. At substrate concentrations 1000-fold lower, however, it is inert unless complemented with specificity factor, in which case it acts only on RNAs with AAUAAA[13,14,18]. Specificity factor may also decrease the non-specific activity observed with high levels of polymerase[16]. The requirement for a certain distance beyond AAUAAA may arise because the polymerase cannot contact both an AAUAAA-bound specificity factor and the 3' end of the RNA if they are too close together[11].

In transcription and replication, segregation of catalytic and specificity functions into separate molecules is very common. It enables the same polymerase to be targeted to different sequences, as with alternate σ factors in bacterial transcription. With respect to polyadenylation, very similar poly(A) polymerases have been identified in nuclear and cytoplasmic fractions (for discussion, see Refs 16, 21 and 22). Perhaps the cytoplasmic polyadenylation events that occur early in development and require distinct, non-AAUAAA sequences (to be reviewed in next month's *TIBS*) will turn out to rely on distinct specificity factors that target very similar poly(A) polymerases to specific mRNAs.

Specificity factor: snRNP or protein?

Specificity factor has an apparent molecular weight of at least 250–300 kDa[13,15,17] and may contain a small nuclear ribonucleoprotein (snRNP)[17,23]. Since a snRNP is known to be essential for cleavage at histone mRNA 3' termini[24], by analogy one might expect a snRNP to also be required for cleavage and polyadenylation. U11 snRNA[17,25] is the most abundant RNA species in partially purified specificity factor fractions[15,17]. Its proposed participation in cleavage and polyadenylation[17] is controversial, however, in part because the RNA component of U11 snRNP is refractory to inactivation[25]. The observation that U11 and specificity factor activity do not precisely cofractionate argues against U11 being a critical component[15], but until the activity is pure, the data will be inconclusive.

Neither U11 nor any other known cellular snRNA contains a sequence complementary to AAUAAA[25], which might be expected if such an RNA was responsible for cleavage. Nevertheless, a known snRNP may be involved in polyadenylation, much as the U4/U5/U6 snRNP is required for splicing but does not base pair to the pre-mRNA. A protein may make at least one contact with AAUAAA, since the 2' hydroxyl group of the U in AAUAAA (which is not expected to affect RNA–RNA base pairing) is essential for binding of specificity factor[11]. At present, it remains possible that, unlike splicing and the formation of histone mRNA 3' termini, cleavage and polyadenylation may be catalysed entirely by protein[15].

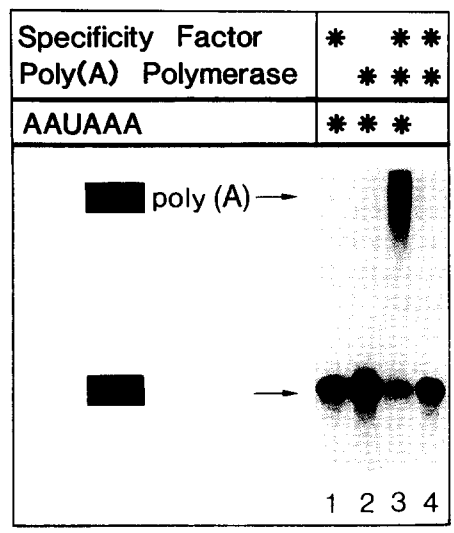


Figure 3

Two separate fractions derived from the HeLa extract are required for poly(A) addition. As demonstrated originally by Takagaki *et al.*[14] and later confirmed in several laboratories[15–20], these two fractions contain an enzyme that polymerizes ATP into poly(A) [poly(A) polymerase], and an activity that gives that enzyme specificity for AAUAAA-containing substrates (specificity factor). In the experiment shown, a labeled substrate that ends at the poly(A) addition site is incubated with fractions containing either specificity factor (lane 1), poly(A) polymerase (lane 2), or a mixture of the two (lane 3). RNA containing AAGAAA does not receive poly(A) when incubated with the combined fractions (lane 4).

Complex assembly prior to cleavage

Specificity factor associates with the mRNA precursor prior to cleavage, and remains associated afterward to participate in polyadenylation. The conclusion that the same factor confers AAUAAA-specificity on cleavage and on polyadenylation was first drawn from competition experiments[26], and has since been demonstrated by biochemical fractionation[13,15,18]. The observation that cleavage and polyadenylation are similarly affected by a range of mutations in AAUAAA further suggests that the same factor is involved in both reactions (Sheets, Ogg and Wickens, op. cit.).

Surprisingly, poly(A) polymerase is also required for cleavage of most RNAs[13,14,17,18,20], and so almost certainly is present in the complex that catalyses that reaction. Thus, even before the mRNA precursor is cut, both activities required for poly(A) addition are in place.

Cleavage requires at least three additional components[13,15]. In principle, all five components could either preassemble before interacting with the substrate, or could bind to the substrate sequentially. *In vitro*, the partially purified components bind to the RNA through an ordered pathway[14], but *in vivo*, as in the unfractionated extract, this important issue is unresolved.

The final complex within which cleavage occurs is massive (500–1000 kDa)[8]. The electrophoretic mobility of the complex changes upon cleavage[27]. Thus, although the specificity factor and polymerase remain associated with the cleaved RNA, the other components may be released.

Two phases of poly(A) addition

Addition of the poly(A) tail proceeds in two stages[28]. At first, the addition of each of the first few adenosine nucleotides is dependent on AAUAAA. Once an oligo(A) primer of 10 or more nucleotides is synthesized, it replaces the requirement for AAUAAA (Fig. 5).

It has been suggested that the same polymerase is responsible for the addition of each adenosine nucleotide in the tail[28], although neither specificity factor nor polymerase are yet pure enough to permit definitive conclusions. Specificity factor is required not only to initiate poly(A) synthesis, but also to stimulate extension of oligo(A) tails[13,28]. Perhaps, even after specificity factor has left AAUAAA, it still contacts the advancing polymerase. This is supported by analysis of complexes with RNAs bearing prematurely terminated poly(A) tails of different lengths[29]. A partially purified poly(A) polymerase shows no marked preference for oligo(A) as a substrate[30], suggesting that another factor, possibly either specificity factor or poly(A) binding protein, is required for the second phase of polyadenylation. The specificity factor may disengage from AAUAAA at the transition between phases[29].

Polyadenylation may be compared to transcription in *E. coli*, with AAUAAA-specificity factor and poly(A) polymerase being analogs of σ factor and RNA polymerase, respectively[29]. Sigma factor, which imparts promoter specificity to RNA polymerase, is released after the transcription of about 10 nucleotides, probably as a result of a conformational change in the holoenzyme. Similarly, in polyadenylation, specificity factor imparts AAUAAA-specificity to an otherwise non-specific poly(A) polymerase. After about 10 nucleotides have been added, the reaction undergoes a transition from AAUAAA-dependence to independence. Specificity factor, like σ factor, may disengage from its cognate sequence at that time. Unlike σ factor, however, specificity factor may continue to be associated with polymerase after the transition.

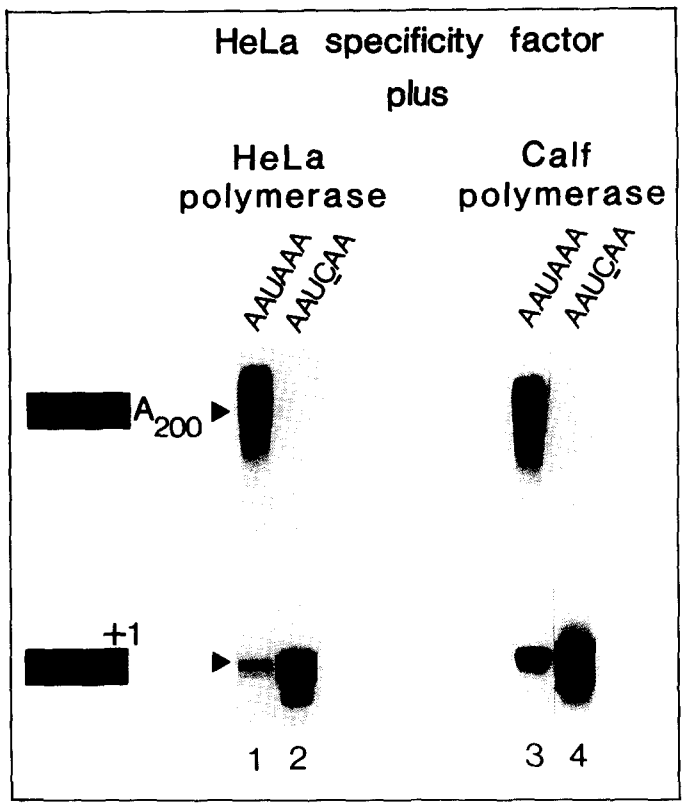


Figure 4

The enzyme that adds poly(A) is a classical poly(A) polymerase. A crude HeLa cell specificity factor fraction was mixed with either HeLa poly(A) polymerase (lanes 1 and 2), partially purified by complementation with the HeLa specificity factor, or with a calf thymus poly(A) polymerase (lanes 3 and 4), purified 15 years ago using the classical, 'non-specific' assay. These mixtures were then assayed for specific polyadenylation of an SV40 late RNA that ends at the poly(A) site and contains either AAUAAA (lanes 1 and 3) or AAUCAA (lanes 2 and 4). The calf and HeLa poly(A) polymerases behave essentially identically. (Taken from Ref. 19.)

Termination of poly(A) synthesis

Accurate termination of poly(A) synthesis apparently occurs in crude nuclear extracts. For example, in yeast extracts, 60–80 nucleotides are added, corresponding well with the *in vivo* length[31,32]. Likewise, in mammalian extracts, the length added *in vitro* (200–300 nucleotides) also corresponds well with that in nuclei. Extension beyond this length can occur, but is relatively slow[2]. Termination occurs over a defined range of lengths, rather than at a precise number of nucleotides.

Although it is not known what causes 'termination', it is clear that an activity present in crude nuclear extracts can discriminate the length of poly(A) present on an exogenous RNA: whereas poly(A) tails of less than 250 nucleotides are extended after addition to the extract, poly(A) tails longer than 250 are not[28]. This discrimination is not a property of the purified poly(A) polymerase. A 'termination factor', if it exists, might act by binding to the poly(A) tail and communicating its presence to the advancing polymerase. Although nuclear poly(A) binding protein is a reasonable candidate for such a termination factor, there is no evidence that this protein forms a unique structure with a 250-nucleotide poly(A) tail, or that it contacts poly(A) polymerase.

Prospects

A provisional model for polyadenylation is summarized in Fig. 6, in which the reaction is somewhat arbitrarily divided into five consecutive steps. This model is intended to provide a focus for discussion, and as such, makes certain assumptions.

Critical questions remain concerning the identity of specificity factor and the nature of its interactions with the substrate and polymerase. In none of the complexes diagrammed in Fig. 6 have the components indicated been detected by direct means. An important, immediate objective is to purify each complex, determine its constituents and their locations relative to each other and the RNA. Likewise, the precise identity of specificity factor must be determined. This should help resolve whether AAUAAA is recognized via RNA–RNA base pairing or by protein–RNA contacts. With pure specificity factor, it may be possible to assess whether it contains separable domains that interact with poly(A) polymerase, cleavage factors and AAUAAA.

Even without pure specificity factor, however, it is clear that its interactions with poly(A) polymerase and the substrate are dynamic. Initially, both the polymerase and specificity factor apparently are present in a macromolecular complex containing several other components. After cleavage, specificity factor remains associated with AAUAAA and with polymerase. After 10 adenosine nucleotides have been added, these interactions change such that the reaction becomes AAUAAA-independent. After 250 adenosine nucleotides have been added, polyadenylation ceases and the enzymes are released. What is the biochemical explanation for each of these transitions?

The recent development of yeast (*Saccharomyces cerevisiae*) extracts that support cleavage and polyadenylation *in vitro*[31,32] should bring genetics to bear on these problems. Judging from work on other complex macromolecular assemblies (e.g. the spliceosome) genetics may prove invaluable. At the least, it should make it possible to connect biochemical activities *in vitro* with biological functions *in vivo*. We can expect the isolation of conditional mutants defective in poly(A) addition or cleavage, their complementation with cloned DNAs, and all the subsequent prestidigitations of modern yeast genetics. However, *S. cerevisiae* may differ from higher cells in the mechanism by which poly(A) is added, in that AAUAAA is not necessary for cleavage in this organism[32], and no other simple sequence appears to exist[33].

The famous dictum that 'one should not waste clean thoughts on dirty enzymes' obviously has not silenced discussion of polyadenylation, either here or elsewhere. It is clear, however, that the biochemical resolution of the questions raised here will require purified components, clones, overexpression systems, and a workshop full of all the molecular tools that a biochemist can acquire. Currently, equipping that invaluable workshop is a central objective in several laboratories. One hopes that the proper time for clean thoughts is nearly upon us.

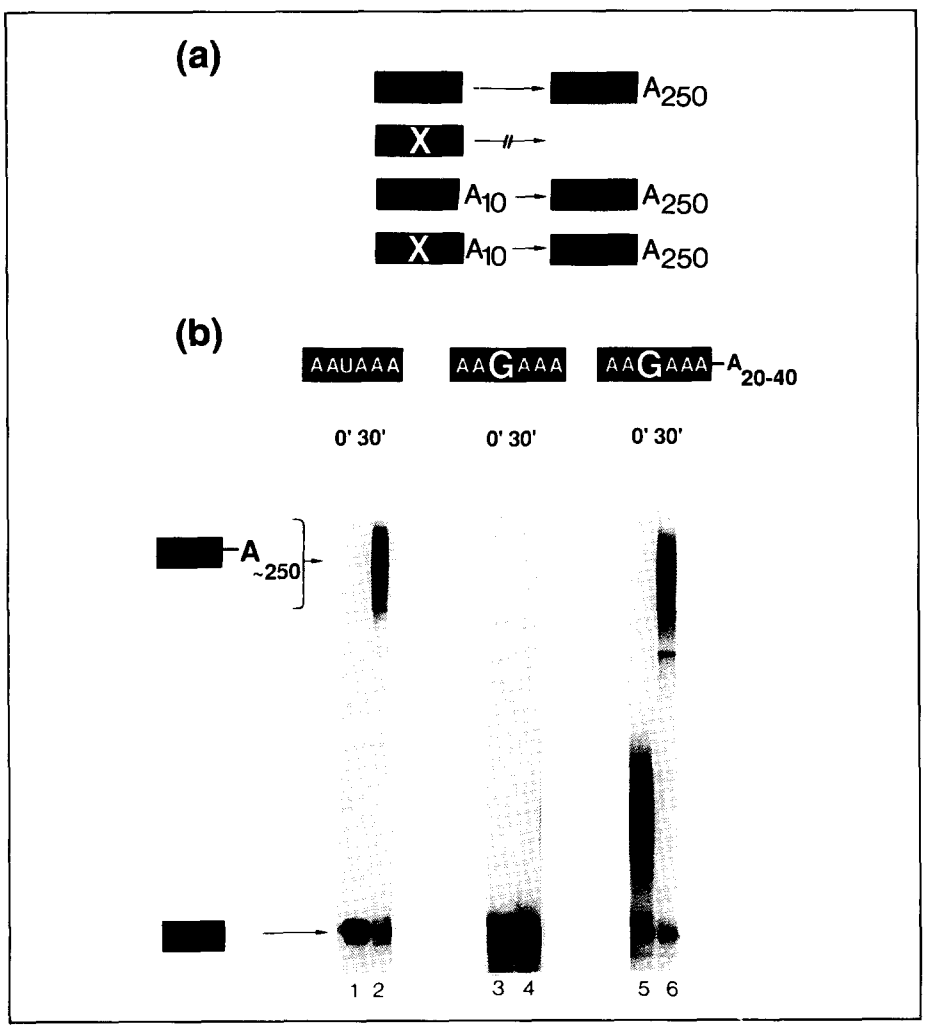


Figure 5

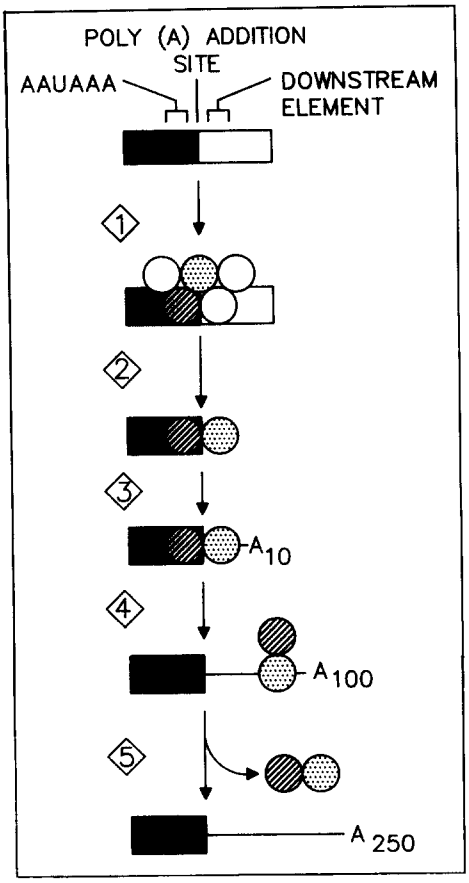
Two phases of poly(A) addition. (a) Diagram of experimental results that demonstrate that the need for AAUAAA is lost after 10 nucleotides have been added. The white X indicates AAGAAA in place of AAUAAA. This mutation prevents polyadenylation of an RNA without any oligo(A) tail, but has little effect on an RNA that ends in 10 adenosine nucleotides. (b) An example of the experimental results that demonstrate suppression of an AAUAAA mutation by an oligo(A) tail. Three different labeled RNAs were incubated in a HeLa cell nuclear extract. The first (lanes 1 and 2) contains AAUAAA and no oligo(A); it is fully active. The second RNA (lanes 3 and 4) contains AAGAAA and no oligo(A); it is inactive. The third RNA (lanes 5 and 6) is the same as the second, except that 20 to 40 nucleotides have been added to its 3' terminus using *E. coli* poly(A) polymerase. This substrate is efficiently extended in the extract even though it carries the mutation to AAGAAA. The minimum length of oligo(A) required for this effect is 10 nucleotides[28].

Acknowledgements

I am grateful to numerous colleagues, including the members of my own laboratory, for helpful comments on the manuscript, and for critical discussions over the past several years. I am particularly grateful to Christine Guthrie who suggested that I should prepare this review. Work in my laboratory is supported by an NIH grant (GM31892) and

Figure 6

Working model: the region upstream of the poly(A) addition site is shown in a black box; the region downstream is in a white box. Poly(A) polymerase (◎), specificity factor (◍) and cleavage factors (○) are indicated. Other components required for cleavage are discussed in detail in Refs 14 and 16. The size of the circle does not reflect the size of the factors.



Step **1**. Assembly of the complex. Prior to cleavage, at least five molecules form a complex with the mRNA precursor. These include an AAUAAA-specificity factor and a poly(A) polymerase. Step **2**. Cleavage. The RNA is cleaved at or near the poly(A) site, generating 'half-molecules' bearing 5' phosphate and 3' hydroxyl termini. The downstream half-molecule is released and rapidly degraded. The upstream half-molecule is retained. Step **3**. Initiation phase of polyadenylation. To the upstream half-molecule, poly(A) is added, one nucleotide at a time. This reaction is catalysed by poly(A) polymerase in physical association with specificity factor bound to AAUAAA. In this early phase, addition of each adenosine is dependent on this interaction. Step **4**. Elongation phase of polyadenylation. After ~10 nucleotides have been added, specificity factor disengages from AAUAAA. Further poly(A) addition is independent of AAUAAA and depends instead on the oligo(A) segment already polymerized. Specificity factor may continue to contact polymerase even after this transition. Step **5**. Termination. After ~250 nucleotides have been added, polymerase and specificity factor are released and can then participate in cleavage and polyadenylation of another RNA.

an NIH Research Career Development Award (GM00521).

References

1 Nevins, J. R. and Darnell, J. E. (1978) *Cell* 15, 1477–1493
2 Moore, C. L. and Sharp, P. A. (1985) *Cell* 41, 845–855
3 Moore, C. L., Skolnick-David, H. and Sharp, P. A. (1986) *EMBO J.* 5, 1929–1938
4 Sheets, M. D., Stephenson, P. and Wickens, M. (1987) *Mol. Cell. Biol.* 7, 1518–1529
5 Zarkower, D., Stephenson, P., Sheets, N. and Wickens, M. (1986) *Mol. Cell. Biol.* 6, 2317–2323
6 Humphrey, T. and Proudfoot, N. J. (1988) *Trends Genet.* 4, 243–245
7 Manley, J. L. (1988) *Biochim. Biophys. Acta* 950, 1–12
8 Manley, J. L., Proudfoot, N. J. and Platt, T. (1989) *Genes Dev.* 3, 2218–2244
9 Manley, J. L., Yu, H. and Ryner, L. (1985) *Mol. Cell. Biol.* 5, 373–379
10 Wilusz, J. and Shenk, T. (1989) *Nucleic Acids Res.* 17, 3899–3908
11 Wigley, P. L., Sheets, M. D., Zarkower, D. A., Whitmer, M. E. and Wickens, M. (1990) *Mol. Cell. Biol.* 10, 1705–1713
12 Conway, L. J. and Wickens, M. (1988) *EMBO J.* 6, 4177–4184
13 Gilmartin, G. M. and Nevins, J. R. (1989) *Genes Dev.* 3, 2180–2189
14 Takagaki, Y., Ryner, L. C. and Manley, J. L. (1988) *Cell* 52, 731–742
15 Takagaki, Y., Ryner, L. C. and Manley, J. L. (1989) *Genes Dev.* 3, 1711–1724
16 Ryner, L. C., Takagaki, Y. and Manley, J. L. (1989) *Mol. Cell. Biol.* 9, 4229–4238
17 Christofori, G. and Keller, W. (1988) *Cell* 54, 875–889
18 Christofori, G. and Keller, W. (1989) *Mol. Cell. Biol.* 9, 193–203
19 Bardwell, V. J., Zarkower, D. A., Edmonds, M. and Wickens, M. (1990) *Mol. Cell. Biol.* 10, 846–849
20 Terns, M. P. and Jacob, S. T. (1989) *Mol. Cell. Biol.* 9, 1435–1444
21 Edmonds, M. (1989) in *Methods in Enzymology* (Dahlberg, J. and Abelson, J., eds), Vol. 181, pp. 161–170, Academic Press
22 Jacob, S. T. and Rose, K. M. (1983) in *Enzymes of Nucleic Acid Synthesis and Modification* (Jacob, S. T., ed.), Vol. 2, pp. 135–157, CRC Press
23 Hashimoto, C. and Steitz, J. A. (1986) *Cell* 45, 581–591
24 Schaufele, F., Gilmartin, G. M., Bannawarth, W. and Birnstiel, M. L. (1986) *Nature* 323, 777–781
25 Montzka, K. A. and Steitz, J. A. (1988) *Proc. Natl Acad. Sci. USA* 85, 8885–8889
26 Zarkower, D. A. and Wickens, M. (1987) *EMBO J.* 6, 177–142
27 Zarkower, D. A. and Wickens, M. (1987) *EMBO J.* 6, 4185–4182
28 Sheets, M. D. and Wickens, M. (1989) *Genes Dev.* 3, 1401–1412
29 Bardwell, V. J. and Wickens, M. (1990) *Mol. Cell. Biol.* 10, 295–302
30 Winters, M. A. and Edmonds, M. (1973) *J. Biol. Chem.* 248, 4763–4768
31 Butler, J. S. and Platt, T. (1989) *Science* 242, 1270–1274
32 Butler, J. S., Sadhale, P. and Platt, T. (1990) *Mol. Cell. Biol.* 10, 2599–2605
33 Osborne, B. I. and Guarente, L. (1989) *Proc. Natl Acad. Sci. USA* 86, 4097–4101