Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 3698–3702, June 1988
Biochemistry

The C-terminal repeat domain of RNA polymerase II largest subunit is essential *in vivo* but is not required for accurate transcription initiation *in vitro*

(heptapeptide repeat/promoter recognition/subunit II_b/*in vitro* transcription)

W. A. ZEHRING, J. M. LEE, J. R. WEEKS, R. S. JOKERST†, AND A. L. GREENLEAF

Department of Biochemistry, Duke University Medical Center, Durham, NC 27710

Communicated by Irwin Fridovich, January 15, 1988

ABSTRACT **DNA sequence analysis of *RpII215*, the gene that encodes the M_r 215,000 subunit of RNA polymerase II (EC 2.7.7.6) in *Drosophila melanogaster*, reveals that the 3′-terminal exon includes a region encoding a C-terminal domain composed of 42 repeats of a seven-residue amino acid consensus sequence, Tyr-Ser-Pro-Thr-Ser-Pro-Ser. A hemi- and homozygous lethal *P*-element insertion into the coding sequence of this domain causes premature translation termination and therefore truncation of the protein, leaving only 20 heptamer repeats. While loss of approximately 50% of the repeat structure in this mutant is a lethal event *in vivo*, enzyme containing the truncated subunit remains capable of accurate initiation at promoters *in vitro*. Moreover, treatment of purified intact RNA polymerase II with protease, to remove the entire repeat domain, does not eliminate the enzyme's ability to initiate accurately *in vitro*. Possible *in vivo* functions for this unusual protein domain are considered in light of these results.**

The C-terminal sequence of the largest subunit of RNA polymerase II (EC 2.7.7.6) consists of multiple repeats of a seven-residue amino acid sequence with the consensus Tyr-Ser-Pro-Thr-Ser-Pro-Ser. This heptamer unit is tandemly repeated 26 times in yeast and 52 times in mammalian largest subunits (1–3). In *Drosophila melanogaster* we find that the C-terminal domain (CTD) is composed of 42 heptamer repeats, albeit a greater proportion of the repeats deviate from the consensus than has been observed in either yeast or mammals (see below). Since the same CTD repeat structure is also found in RNA polymerase II from plants (M. Dietrich and T. Guilfoyle, personal communication), it is clear that the C-terminal repeat structure is a highly conserved feature of this transcriptase. Because the CTD is not found in largest subunits of eukaryotic RNA polymerases I (4) or III (1), prokaryotic RNA polymerase (5), or a viral RNA polymerase (6), we surmise that it is unique to RNA polymerase II and therefore performs a function uniquely important to this enzyme.

We have initiated an investigation of CTD function in *Drosophila* by analyzing the properties of enzymes harboring a largest subunit either partially or completely lacking the CTD. Partial truncation of the CTD occurs in individuals bearing a previously isolated genetic variation of the locus for the largest subunit, *RpII215* (7, 8). We have also investigated purified enzyme from which the entire CTD was removed proteolytically. With the development of a fractionated RNA polymerase II-dependent *in vitro* transcription system from *Drosophila melanogaster* (9), it is possible to assess the ability of genetically or enzymatically altered RNA polymerase II to initiate transcription accurately at promoters *in vitro*. Taken together, the genetic and biochemical data presented here suggest that although the unusual repeat structure found at the C terminus of the largest subunit is essential *in vivo*, partial truncation or complete removal of the structure does not affect overall polymerase activity or eliminate accurate initiation *in vitro*.

MATERIALS AND METHODS

Genetic Variants of *RpII215*. All stocks were maintained on a standard cornmeal/sugar/agar medium or on modified instant *Drosophila* medium (10) at 25°C. Origins of wild-type and mutant alleles of *RpII215* used in this study have been described previously (7, 11). The P2 (wild-type), *C4* homozygous, and *W81*/FM7 stocks were used for embryo collection for purification of RNA polymerase II (see below). The FM7 chromosome harbors a wild-type (amanitin-sensitive) allele of *RpII215*.

Sequence Analysis of Subclones of *RpII215*. All sequencing was carried out by using the dideoxy chain termination method (12) on nested deletions of subclones inserted into M13mp18 or mp19 vectors (13). The subclone containing the *W81 P*-element insertion was the kind gift of Lillie Searles. The complete sequence of the locus has been determined for both DNA strands (14, 15). The details of this analysis will be presented elsewhere (R.S.J., J.R.W., W.A.Z., and A.L.G., unpublished data).

Purification of RNA Polymerase II from Embryos and Protein Gel Analysis. Embryos were collected from expanded stocks of P2, *C4*, and *W81*/FM7 at 12-hr intervals, dechorionated, and quick frozen to −85°C for later use. Five grams of frozen dechorionated embryos were ground in liquid nitrogen, and the powder was transferred to a glass Dounce homogenizer, resuspended in 25 ml of 25 mM Hepes, pH 7.6 (Calbiochem)/15% (vol/vol) glycerol/0.3 M ammonium sulfate/0.1 mM EDTA/1 mM dithiothreitol/0.1% phenylmethylsulfonyl fluoride (from a saturated solution in isopropyl alcohol)/5.0 μg of soybean trypsin inhibitor per ml (0.3 M HGAEDP) supplemented with 1 mM sodium bisulfite, and homogenized with 10–15 passages of the plunger (this and all subsequent steps were performed at 2°C unless otherwise indicated). After filtration through one layer of Miracloth (Calbiochem), the sample was centrifuged in a T865 rotor (Sorvall) for 1 hr at 140,000 × *g* (average). The supernatant was gently collected, diluted to 150 mM ammonium sulfate with HGEDP (identical to HGAEDP but without ammonium sulfate) and loaded onto a 30-ml bed volume DEAE-cellulose (DE52, Whatman) column previously equilibrated in 0.12 M HGAEDP. RNA polymerase was eluted with a 0.3 M HGAEDP step and assayed with calf thymus DNA as previously described except that 1 mM $MnCl_2$ replaced

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CTD, C-terminal domain; wt, wild type.

†Present address: Monsanto, 700 Chesterfield Village Parkway, Chesterfield, MO 63198.

$MgCl_2$ (9). Active fractions were pooled and diluted to 180 mM ammonium sulfate with HGEDP. This fraction was then loaded onto a 3-ml bed volume heparin-Sepharose 6B-CL column previously equilibrated in 0.18 M HGAEDP. Polymerase was eluted with 0.5 M HGAEDP step and active fractions were pooled. After dilution to 0.2 M ammonium sulfate with HGEDP, this fraction was further chromatographed on a 1.0-ml Mono-Q FPLC column (Pharmacia), with gradient elution from 0.2 to 1.0 M HGKEDP (identical to the buffer system described above except that KCl at the indicated concentrations replaces ammonium sulfate). RNA polymerase II activity eluting at approximately 0.45 M KCl was pooled and dialyzed against 0.05 M HGKEDP until the KCl concentration was reduced to 100 mM. Aliquots were quick frozen and stored at −85°C.

Protein gel electrophoresis and transfer to nitrocellulose for analysis with antibody were carried out as described (16). Gel-fractionated proteins were silver stained by the method of Morrissey (17).

Transcription Reconstructions with Fractionated Factors. Transcription reconstructions were carried out with partially purified transcription factors essentially as described (9). Factor 1 (DNase inhibitor) was a highly purified (Mono-Q FPLC) fraction (18). Factor 3 was separated from endogenous RNA polymerase II by chromatography on DEAE-cellulose and Mono-Q (FPLC). Additional required factors, including "TATA"-binding activity, were present in a crude 0.3–0.4 M KCl step of the K_c nuclear extract on phosphocellulose (ref. 9 and D. H. Price, A. E. Sluder, and A.L.G., unpublished results). Typically, 10 units (defined in ref. 9) of partially purified RNA polymerase II was added to the reaction mixture. Reconstructions were programmed with the *Pst* I-digested *Drosophila* actin 5C template previously described (9).

Protease Digestion of Purified RNA Polymerase II. Purified RNA polymerase II was treated at 25°C with chymotrypsin at 0.25 μg/ml in the presence of 50 mM Tris·HCl, pH 7.8/0.1 mM EDTA/25% (vol/vol) glycerol/0.5 mM dithithreitol. Reactions were terminated at the indicated times by addition of 0.6 vol of a mix of soybean trypsin inhibitor (50 μg/ml) and phenylmethylsulfonyl fluoride (1:100 dilution of saturated solution in isopropyl alcohol), followed by quick freezing in dry ice.

RESULTS

DNA and Protein Sequence of the C-Terminal Repeat and of a Mutation Therein. The deduced amino acid sequence of the C-terminal repeat portion of the wild-type *RpII215* gene is presented in Fig. 1. Examination of the sequence reveals a repeating motif similar to the motifs previously reported in yeast and mouse (1–3), with a consensus heptamer sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (Y S P T S P S). Repeat no. 1 is defined as shown to maximize homology with the mouse sequence (3), for both the CTD proper (repeats 1–42) and the region immediately preceding it (pseudorepeats a–f). As depicted, the *Drosophila* C-terminal repeat domain comprises 42 units and is thus intermediate in length between the repeat domains of yeast and mammals. Besides differing in length, the *Drosophila* domain adheres to the consensus sequence noticeably less strictly than that of either yeast or mammals; in fact, there are only two repeats in *Drosophila* that conform completely to the consensus sequence, numbers 15 and 18 in Fig. 1. In addition to the wild-type allele, we have also sequenced the amanitin-resistant allele, *C4*, and find that it is identical to wild type in the region presented in Fig. 1 (R.S.J., J.R.W., W.A.Z., and A.L.G., unpublished data); amanitin resistance thus maps outside the CTD.

Three previously described mutations in the *RpII215* locus map to the 3′-most exon of the gene and therefore probably

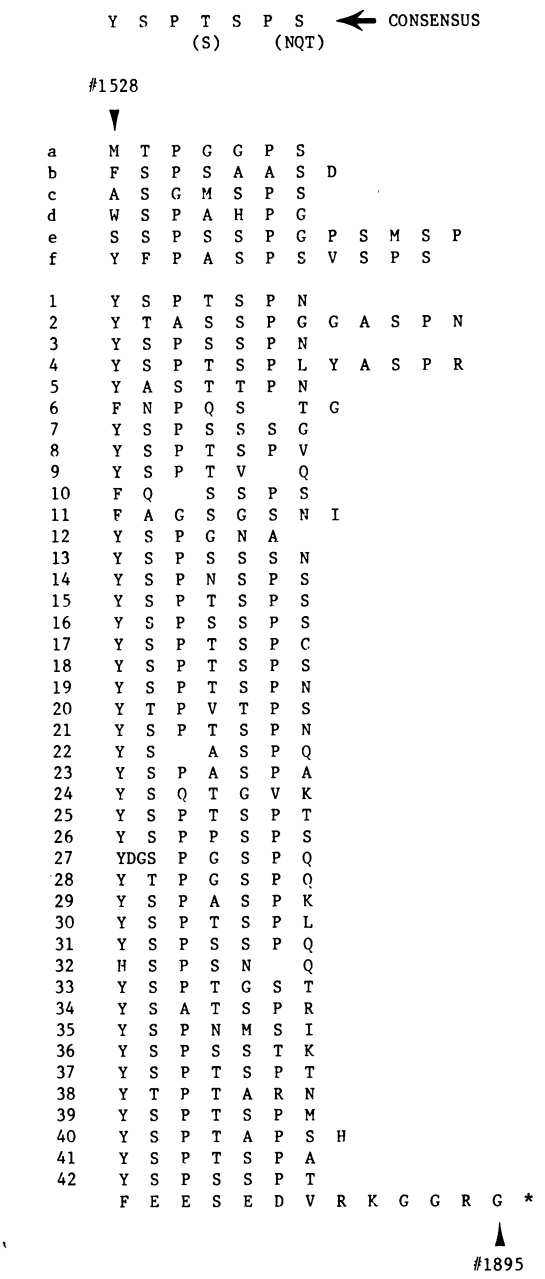

FIG. 1. Amino acid sequence of CTD of the largest subunit of wild-type *Drosophila* RNA polymerase II. The standard one-letter code is used. The sequence is aligned so as to emphasize this domain's repeating nature and its homology to the consensus sequence (1–3, 24) shown at the top (the most frequent substitutions found in the *Drosophila* sequence are shown in parentheses). Six pseudorepeats (a–f, see text) and 42 repeats are depicted; chain termination follows amino acid 1895.

alter the CTD. Mutation *H20* was induced in the wild-type *RpII215* gene by ethylnitrosourea mutagenesis and appears to be a small deletion [approximately 600 base pairs (bp)] within the *Sal* I/*Eco*RI fragment at coordinates −5.7 to −6.7 (19). Mutations *W81* and *W38* were generated in hybrid dysgenesis-inducing crosses and are *P*-element insertions into the analogous region of the α-amanitin-resistant *C4* allele (7, 8). Because these mutations are hemi- and homozygous lethals they must affect some essential feature of the *RpII215* gene or gene product.

With these mutants available we were in a position, first, to test the hypothesis that lethality could be attributed to an alteration of the CTD and, second, to assess the consequences of a lethal alteration of the CTD on the *in vitro*

transcriptional properties of RNA polymerase II. For this second goal, the lethality of the mutations presents a problem in that mutant enzyme must be purified from heterozygotes and can therefore be obtained only as a mixture of mutant and wild-type polymerases. This problem can, however, be overcome if the polymerase with an altered CTD is also amanitin resistant, since the contaminating wild-type enzyme can be inhibited by addition of amanitin to the *in vitro* assays (9, 11). For this reason we have focused our initial investigations on *W81*, whose parent was the amanitin-resistant mutant *C4*.

We sequenced the 3′ portion of the *W81* gene (the *Eco*RI/*Sal* I fragment from coordinates −5.7 to −6.7) and found that a 631-bp incomplete *P* element had inserted into DNA encoding repeat 20, as defined in Fig. 1. Sequences surrounding the upstream junction of *P* element and *RpII215* DNA are presented in Fig. 2. Examination of the coding potential of the mutant sequence reveals that in the *RpII215* reading frame the *P*-element DNA encodes five novel amino acids followed by a termination codon. Translation termination at this point would eliminate approximately 50% of the repeat structure (see Fig. 1). Consistent with previous characterization of *P*-element insertions, an 8-bp host sequence is duplicated adjacent to the mobile element's own terminal repeats at both this and the downstream junction (W.A.Z., unpublished results), though it has only slight similarity with a consensus target sequence (20).

***W81* Mutant Enzyme Contains a Truncated Large Subunit.** The protein truncation predicted by the sequence data was confirmed by $NaDodSO_4$/polyacrylamide gel analysis of purified mutant and wild-type enzymes (Fig. 3*A*). The only two polymerase subunits shown on the gel are the largest (II_a) and second-largest (II_c), with apparent molecular weights of 215,000 and 140,000, respectively. Note that these rapidly purified polymerases (see *Materials and Methods*) contain only small amounts of subunit form II_b ($M_r \approx 180{,}000$), a subunit form (or forms) often produced in variable amounts by proteolytic removal of the CTD from subunit IIa during purification (1, 2, 21, 22).

Subunit II_a migrates normally in both the wild-type and *C4* enzyme preparations, as expected. However, in the *W81*/wt heterozygote preparation, there is a novel band migrating faster than the II_a subunit and present in amounts nearly equimolar to the normal II_a species. This polypeptide is a shortened version of the largest subunit, since it reacts with antibodies directed specifically against the largest subunit (W.A.Z. and J.M.L., unpublished data). Moreover, the RNA synthetic activity of this preparation, measured with denatured DNA as a template in the presence of increasing concentrations of amanitin (*Materials and Methods* and ref. 11), was found to be 30–40% amanitin resistant. This fraction of amanitin-resistant enzyme is in close proportion to the representation of the *W81* allele in the embryos used as starting material (see *Materials and Methods*).

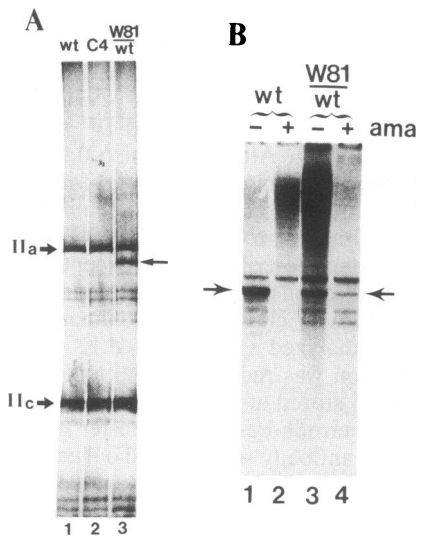

FIG. 3. (*A*) $NaDodSO_4$/polyacrylamide gel analysis of purified wild-type and mutant embryo RNA polymerase II. Equivalent amounts of purified embryonic enzyme were analyzed in three lanes of a 5% gel. Lane 1, wild type; lane 2, *C4* (amanitin-resistant); lane 3, *W81*/wild type (wt). Electrophoresis was carried out at constant power for 150% of the time required for a bromphenol blue dye marker to reach the bottom of the gel. Silver staining was carried out as described (17). (*B*) Polyacrylamide/urea gel analysis of run-off transcripts produced by wild-type and mutant RNA polymerases II, with and without α-amanitin. Transcription reconstructions were carried out with the actin 5C template (9). Lane 1, wild-type enzyme; lane 2, wild-type enzyme plus α-amanitin at 0.8 μg/ml; lane 3, *W81*/wt enzyme; lane 4, *W81*/wt enzyme plus α-amanitin at 0.8 μg/ml. The arrows indicate the expected run-off transcript of 450 bases (9).

Taken together, these data strongly support the conclusion that the *W81* insertion results in a premature termination of translation within the C-terminal repeat structure of the largest subunit of RNA polymerase II, but the resulting protein is still capable of forming an active RNA polymerase, functional in a nonspecific transcription assay. We conclude that the lethality of the *W81* allele is not due to elimination of catalytic activity, or to some negative influence on expression of *RpII215* or on stability of its gene product, but must be due to some other adverse effect on RNA polymerase II resulting from truncation of the C-terminal repeat.

Accurate Initiation by Mutant Enzyme *in Vitro*. The recent development in our laboratory of an enzyme-dependent promoter-driven transcription reconstruction system (9) permits us to test the ability of this purified mutant enzyme, in the presence of the required transcription factors, to recognize promoters and initiate transcription accurately. As was stated above, the mutant enzyme is of necessity purified along with the complementing wild-type enzyme. However,

```
                #1716
                  ↓
Wild Type:
                  P  T  S  P  N  Y  T  P  V  T  P  S  Y  S  P  T  S  P  N
          ... CCA ACG AGT CCG AAC TAC ACA CCC GTA ACA CCC TCA TAC TCG CCG ACA AGT CCG AAC ...
repeat# ⟶          19          |          20          |          21
                                |                     |
Mutant W81:                     |                     |
                  P  T  S  P  N  Y  T  P  H  D  Q  I  T  *
          ... CCA ACT AGT CCT AAC TAC ACA CCC CAT GAT GAA ATA ACA TAA GGTGGTCCCGTCG ...
                                |Host 8bp|
```

FIG. 2. DNA and derived protein sequences of analogous domains of wild-type and mutant *W81* largest subunit genes, showing the insertion of *P*-element DNA (underlined) into repeat 20 of the CTD (see Fig. 1). The five novel amino acids encoded by the *P*-element DNA are italicized. Both sequences begin with the codon for amino acid 1716. The 8 bp of target DNA duplicated in the process of *P*-element DNA insertion are indicated (see text).

we can inactivate the wild-type enzyme and test specifically the ability of the mutant polymerase to initiate accurately by adding α-amanitin to the reconstruction assays. The results of such a reconstruction are presented in Fig. 3*B*.

Two purified enzymes were assayed in the reconstruction system: wild type and that purified from the *W81*/wt heterozygote. The assays were performed with partially purified transcription factors and were programmed with the *Pst* I-digested *Eco*RI/*Pst* I subclone of the *Drosophila* actin 5C gene. Accurate initiation at the TATA box-containing promoter in this subclone produces a 450-base RNA run-off transcript (arrow; see also ref. 9). The wild-type enzyme is completely inhibited in the presence of α-amanitin at 0.8 μg/ml, as expected (lane 2). Most importantly, however, the specific run-off transcript is produced by the *W81*/wt enzyme mixture both in the absence (lane 3) and in the presence (lane 4) of α-amanitin, indicating that the truncated (and amanitin-resistant) enzyme is capable of accurate initiation at the actin 5C promoter. The reconstruction level in lane 4 appears to be somewhat lower than the expected 30–40% of that in lane 3 (see above). We find that low levels of reconstruction with such mixed enzyme preparations often occur in the presence of α-amanitin, most likely due to interference by the inhibited sensitive enzyme. Other bands represent RNA polymerase I/III transcripts or nonspecific RNA polymerase II transcripts and vary in amount, depending primarily on the complementing fractions used for reconstruction (e.g., see Fig. 4 and ref. 9).

Proteolytic Removal of the CTD. Because the *W81* enzyme retains approximately 50% of the terminal repeat structure, we wanted to test the effects of more complete removal of this domain. Previous results suggested that protease treatment of RNA polymerase II would remove the CTD (1, 2, 21, 22) without detectably degrading other subunits of the enzyme (ref. 23 and J.M.L., unpublished data). We therefore treated purified *C4* (amanitin-resistant) enzyme with chymotrypsin to remove the CTD. Recall that the sequence of *C4* is identical to that of wild type in the C-terminal region. Also note that chymotrypsin cleaves at aromatic residues and thus each heptameric repeat contains a potential cleavage site. A time course of chymotrypsin treatment of purified *C4* RNA polymerase II is shown in Fig. 4*A*, where the large subunits of RNA polymerase II are resolved on a 5% polyacrylamide/$NaDodSO_4$ gel and stained with silver. This gel reveals that the conversion of the II_a subunit to a faster-migrating species we here call II_b^* is nearly complete after only 5 min of incubation at room temperature. After 20 min there is no longer any detectable II_a subunit; it has been fully converted to II_b^*, which appears to be resistant to further digestion. Protein immunoblotting demonstrates that the limit proteolysis product (II_b^*) contains no sites that react with affinity-purified antibody against the CTD (W.A.Z. and J.M.L., unpublished results). Note that, in contrast to subunit II_a, subunit II_c ($M_r \approx 140{,}000$) was not degraded by the protease. In fact, analysis of the digested enzyme on gels that resolve all the subunits indicates that none of the subunits other than II_a was discernibly affected by the chymotrypsin treatment (ref. 23 and J.M.L., unpublished data). When samples of RNA polymerase II from these protease digestions were assayed with denatured calf thymus DNA as template no loss of RNA synthetic activity was observed (W.A.Z., unpublished data).

Aliquots of RNA polymerase II from the 0-, 5- and 20-min time points were then assayed in the transcription reconstruction system to determine their ability to synthesize the run-off transcript diagnostic of accurate initiation at the actin 5C promoter. The results of this experiment are presented in Fig. 4*B* and show that in this assay system the proteolysis of the largest subunit had no significant effect on accurate initiation. These assays were carried out in the presence of

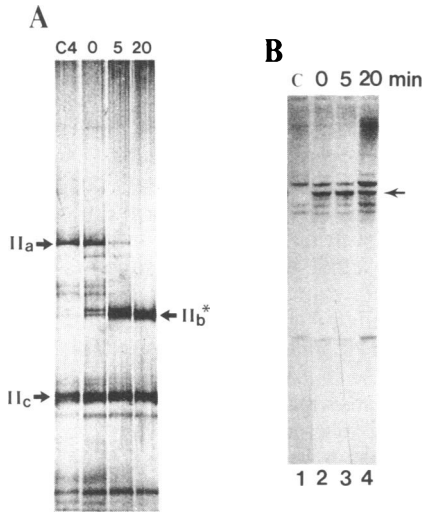

FIG. 4. (*A*) $NaDodSO_4$/polyacrylamide gel analysis of chymotrypsin-treated purified amanitin-resistant (*C4*) RNA polymerase II. Chymotrypsin treatment (0, 5, and 20 min) was carried out as described in *Materials and Methods*. Electrophoresis was carried out as described for Fig. 3*A*. (*B*) Polyacrylamide/urea gel analysis of run-off transcripts produced by chymotrypsin-treated amanitin-resistant (*C4*) RNA polymerase II. Transcription reconstructions were carried out in the presence of α-amanitin at 0.8 μg/ml as described for Fig. 3*B*. Lane 1, no added enzyme; lanes 2, 3, and 4, enzyme treated with chymotrypsin for 0, 5, or 20 min, respectively. The arrow indicates the expected run-off transcript of 450 bases.

α-amanitin at 0.8 μg/ml to ensure that production of the observed transcript was due to the input amanitin-resistant enzyme and not to low levels of sensitive enzyme, which can sometimes contaminate transcription factor fractions (9). A similar test of the chymotrypsin-treated enzyme with a template containing the *Drosophila* histone H3 and H4 promoters showed that the truncation had no effect on accurate initiation at either of these TATA box-containing promoters (W.A.Z., unpublished data). Proteolysis of purified wild-type enzyme was also carried out with trypsin, with similar results. A 60-min trypsin digestion, sufficient to remove most of the CTD, had no effect on the ability of the RNA polymerase II to recognize and initiate at the actin 5C promoter in this *in vitro* assay system (J.M.L., unpublished data).

DISCUSSION

While the function of the C-terminal repeat domain is not understood, its evolutionary conservation uniquely in RNA polymerase II suggests a critical role in some aspect of the physiology of this transcriptase. Although the *Drosophila* CTD repeats are more variable than those of yeast and mammals, the general structure of the repeat domain is undoubtedly very similar in these, and probably all, eukaryotes; this similarity of structure suggests a similarity of function. Our results prove that the *Drosophila* CTD does perform an essential function *in vivo*. These results parallel those obtained for the yeast and mouse repeat domains (24, 25). Additionally, our *in vitro* transcription studies begin to narrow the possible roles for this domain by showing that for *Drosophila* RNA polymerase II, *in vitro* promoter recognition and transcription initiation are not eliminated by genetic or enzymatic truncation of the repeat domain.

The *W81* mutation results in removal of slightly more than half of the heptamer repeats; RNA polymerase II containing this truncated subunit is stable *in vivo*, but functionally defective. First, *W81* hemi- or homozygotes are not viable. Second, *W81*/wt heterozygous males are fertile, in contrast

to *C4*/wt males, which are sterile (7); this restoration of male fertility is a common property of mutations that partially or completely inactivate the *C4* polymerase (7). These results are reminiscent of those for yeast and mouse, which also indicate the need for more than half of the repeats for cell viability or transformation of amanitin resistance, respectively (24, 25). Possibly in analogy with conditional lethality of yeast mutants carrying 10–12 complete repeats, *W81* is not a null mutation. In heterozygous combination with certain other *RpII215* lethal alleles, *W81* supports a low level of viability (W.A.Z., unpublished data). The molecular basis for this complementation has yet to be determined.

Because *W81* enzyme is amanitin resistant, we could purify RNA polymerase II from *W81*/wt heterozygotes and specifically test the transcriptional properties of the mutant enzyme in the presence of contaminating wild-type polymerase by carrying out assays in the presence of amanitin. This approach showed that the truncation did not eliminate the enzyme's ability to initiate accurately at the actin 5C promoter in our reconstituted *in vivo* transcription system.

It has been argued in the past that accurate initiation *in vitro*, catalyzed by RNA polymerase II containing apparently only subunit form II_b, was actually due to a small amount of intact polymerase present in the enzyme preparation (26). We can eliminate this possibility for the case of *W81*, since the only amanitin-resistant polymerase in our *W81*/wt preparations is genetically truncated yet is capable of accurate initiation *in vitro* in the presence of α-amanitin. We have also virtually eliminated this possibility for *C4* and wild type by showing that under our assay conditions the input enzyme is equally able to initiate transcription accurately at the three different promoters tested, whether it is untreated and contains intact subunit II_a or is proteolyzed to a form that contains no detectable intact subunit II_a. We conclude that, in our reconstituted *in vitro* system, factor-dependent promoter recognition and accurate initiation do not require an intact C-terminal repeat domain.

An extension of the above conclusion is that the CTD does not functionally interact with the transcription factors active in the reconstituted system. While all of these factors have not been completely characterized, the fractions used include *Drosophila* analogs of most of the so-called general transcription factors, including a TATA-binding activity (ref. 9 and D. Price and A. Sluder, personal communication). On the other hand, there may be additional regulatory factors that are either not present or not functional in the *in vitro* reconstitutions as performed here (see ref. 9). Such factors, which might quantitatively regulate promoter-dependent transcription, could have been removed during fractionation of the nuclear extract; thus, optimal or regulated promoter utilization might require an intact CTD. This proposition would be consistent with previous results, which suggested a role for the CTD in transcription initiation in crude extracts (26, 27) and isolated nuclei (28). Thus it is still a possibility that the CTD interacts with regulatory components to modulate promoter utilization.

Other possible roles have been proposed for the CTD, ranging from subnuclear localization to alteration of chromosome structure (1–3, 24, 25); these possibilities still need to be considered. In addition, because the CTD can exist in a highly phosphorylated state, in a subunit form called II_o (ref. 28 and references therein), it is likely that its functional properties are modulated by phosphorylation. Elucidating the physiological role of the repeat domain and the functional significance of its phosphorylation should be facilitated by *in vivo* investigations of specifically mutated *RpII215* genes and *in vitro* studies of the altered enzymes.

Note Added in Proof. The sequence of the *Drosophila* CTD was recently presented in a paper that appeared while this manuscript was under review (29).

We thank J. Corden for providing results prior to publication, Lillie Searles for the subclone of the *W81* insertion, Ann Sluder and David Price for advice and reagents for transcription reconstructions, and Ann Sluder and Fred Schachat for critical reading of this manuscript. This work was supported by National Institutes of Health Grant GM 28078 to A.L.G.

1. Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985) *Cell* **42**, 599–610.
2. Corden, J. L., Cadena, D. L., Ahearn, J. M., Jr. & Dahmus, M. E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7934–7938.
3. Ahearn, J. M., Jr., Bartolomei, M. S., West, M. L., Cisek, L. J. & Corden, J. L. (1987) *J. Biol. Chem.* **262**, 10695–10705.
4. Buhler, J.-M., Riva, M., Mann, C., Thuriaux, P., Memet, S., Micouin, J. Y., Treich, I., Mariotte, S. & Sentenac, A. (1987) in *RNA Polymerase and the Regulation of Transcription*, eds. Reznikoff, W. S., Burgess, R. R., Dahlberg, J. E., Gross, C. A., Record, M. T., Jr., & Wickens, M. P., (Elsevier, New York), pp. 25–36.
5. Ovchinnikov, Y. A., Nomastryskaya, G. S., Gubanov, V. V., Guryev, S. O., Schroeder, E., Settlmeissl, G. & Streeck, R. E. (1982) *Nucleic Acids Res.* **10**, 4035–4044.
6. Broyles, S. S. & Moss, B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3141–3145.
7. Voelker, R. A., Wisely, G. B., Huang, S.-M. & Gyurkovics, H. (1985) *Mol. Gen. Genet.* **201**, 437–445.
8. Searles, L. L., Greenleaf, A. L., Kemp, W. E. & Voelker, R. A. (1986) *Mol. Cell. Biol.* **6**, 3312–3319.
9. Price, D. H., Sluder, A. E. & Greenleaf, A. L. (1987) *J. Biol. Chem.* **262**, 3244–3255.
10. Greenleaf, A. L., Borsett, L. M., Jiamachello, P. F. & Coulter, D. E. (1979) *Cell* **18**, 613–622.
11. Greenleaf, A. L., Weeks, J. R., Voelker, R. A., Ohnishi, S. & Dickson, B. (1980) *Cell* **21**, 785–792.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
13. Henikoff, S. (1984) *Gene* **28**, 351–359.
14. Biggs, J., Searles, L. L. & Greenleaf, A. L. (1985) *Cell* **42**, 611–621.
15. Jokerst, R. S. (1987) Thesis (Duke University, Durham, NC).
16. Weeks, J. R., Coulter, D. E. & Greenleaf, A. L. (1982) *J. Biol. Chem.* **257**, 5884–5891.
17. Morrissey, J. H. (1981) *Anal. Biochem.* **117**, 307–310.
18. Sluder, A. E., Price, D. H. & Greenleaf, A. L. (1987) *Biochimie* **69**, 1199–1205.
19. Lacy, L. R., Eisenberg, M. T. & Osgood, C. J. (1986) *Mutat. Res.* **162**, 47–54.
20. O'Hare, K. & Rubin, G. M. (1983) *Cell* **34**, 25–35.
21. Greenleaf, A. L., Haars, R. & Bautz, E. K. F. (1976) *FEBS Lett.* **71**, 205–208.
22. Dezélée, S., Wyers, F., Sentenac, A. & Fromageot, P. (1976) *Eur. J. Biochem.* **65**, 543–552.
23. Coulter, D. E. (1983) Thesis (Duke University, Durham, NC).
24. Nonet, N., Sweetser, D. & Young, R. A. (1987) *Cell* **50**, 909–915.
25. Bartolomei, M. S., Halden, N. F., Cullen, C. R. & Corden, J. L. (1988) *Mol. Cell. Biol.* **8**, 330–339.
26. Dahmus, M. E. & Kedinger, C. (1983) *J. Biol. Chem.* **258**, 2303–2307.
27. Bartholomew, B., Dahmus, M. E. & Meares, C. F. (1986) *J. Biol. Chem.* **261**, 14226–14231.
28. Cadena, D. L. & Dahmus, M. E. (1987) *J. Biol. Chem.* **262**, 12468–12474.
29. Allison, L. A., Wong, J. K.-C., Fitzpatrick, D., Moyle, M. & Ingles, C. J. (1988) *Mol. Cell. Biol.* **8**, 321–329.