
Unusual C-terminal domain of the largest subunit of RNA polymerase II of *Crithidia fasciculata*

Raymond Evers, Andrea Hammer and Albert W.C.A.Cornelissen*

Max-Planck-Institut für Biologie, Molecular Parasitology Unit, Spemannstrasse 34, 7400 Tübingen, FRG

Received February 16, 1989; Revised and Accepted April 10, 1989

EMBL accession no. X13489

ABSTRACT

The C-terminal domain of the largest subunit of RNA polymerase II in higher eukaryotes is present in the protozoan parasite *Trypanosoma brucei* in a strongly modified form. To determine whether this is a general feature of the Kinetoplastida and to determine the role of this domain in RNA polymerase II transcription, we have analysed the C-terminal domain of the distantly related species *Crithidia fasciculata*. No positional identity of amino acid residues between the C-termini of *C.fasciculata* and *T.brucei* can be found. Moreover, both domains lack the heptapeptide repeat structure present in higher eukaryotes. The two domains are, however, very similar in amino acid composition, being rich in acidic residues as well as serine and tyrosine. The latter observation is compatible with the concept that in vivo phosphorylation of the C-terminus activates RNA polymerase II.

INTRODUCTION

Eukaryotic RNA polymerases are characterised by their complex multi-subunit structure (1). The largest subunit genes of the three classes of RNA polymerases, named I (= A), II (= B), and III (= C), from a number of eukaryotic species have been identified, cloned and sequenced (reviewed in (2,3)). The largest subunit of RNA polymerase II of all species analysed to date, with the exception of trypanosomes, contains a C-terminal extension, which is characterised by a heptapeptide tandem repeat sequence with the consensus: Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The consensus sequence was found to be repeated 26 or 27 times in yeast (4,5), 44 times in *Drosophila* (6,7) and 52 times in mouse (8,9) and hamster (6).

A number of functions, based primarily on the tandem repeat structure, have been proposed for the C-terminal domain of RNA polymerase II, eg. a receptor domain for transcription factors, a nuclear anchor, or a domain that must be phosphorylated to generate the active enzyme (reviewed in (2)).

In *Trypanosoma brucei*, the C-terminal domain was found to be unusual, lacking the tandemly repeated consensus sequence. The only resemblance between the C-terminal extensions of this protozoan and other eukaryotes is a relatively high proportion of serine and tyrosine residues. We therefore postulated that the presence of potential phosphorylation sites determines the basic function of the C-terminus (10).

To substantiate our hypothesis, we have isolated the gene encoding the largest subunit of RNA polymerase II of *Crithidia fasciculata*. Within the protozoan order of the Kinetoplastida, this species is the least related to *Trypanosoma brucei* based on a phylogenetic tree constructed from a comparison of the mitochondrial 9S and 12S rRNA gene sequences (11,12). A direct comparison of the primary sequence of such distantly related species is most likely to reveal the sequences which are conserved due to a functional constraint.

RESULTS

Isolation of the Crithidia RNA polymerase II largest subunit gene

We used a gene-internal probe from the RNA polymerase II largest subunit gene of *T. brucei* to identify the corresponding gene in *C. fasciculata*. At reduced stringency conditions, $1\times\text{SSC}$ at 65°C , this probe recognised a 6 kb HindIII fragment in *C. fasciculata* genomic DNA, which was subsequently isolated from a genomic DNA recombinant phage library (Figure 1, panel A). The *Crithidia* probes cross-hybridised to other sequences, resulting in the faint 20 kb HindIII fragment. At present we are trying to isolate this fragment from the *Crithidia* genomic bank in order to determine whether this fragment contains sequences corresponding to the largest subunits of either RNA polymerase I or RNA polymerase III.

To localise the approximate position of the coding sequence in the putative RNA polymerase II locus, several restriction fragments were used as hybridisation probes against *Crithidia* poly(A)⁺ RNA (Figure 1, panel B). Probe P₁ (Figure 2) recognises a transcript of approximately 5.5 kb, which is in the size range of other eukaryotic RNA polymerase II largest subunit mRNAs (2,10,13). Probe P₂ (Figure 2) hybridised to the same RNA species; this probe contains the C-terminal domain (see below). These and other hybridisation data (not shown) resulted in the restriction map of the putative RNA polymerase II locus of *Crithidia* (Figure 2).

Sequence of the Crithidia RNA polymerase II largest subunit gene

The nucleotide sequences and derived amino acid sequences of the pCrp10 and pCrp22 inserts are presented as a composite in Figure 3. Figure 4 shows a comparison of the amino acid sequences of *Crithidia* with the sequences of other eukaryotic RNA polymerase largest subunit genes. It is clear that the *Crithidia* sequence contains the predicted conserved domains F and H (Figure 2) that are found in all eukaryotic RNA polymerase largest subunit

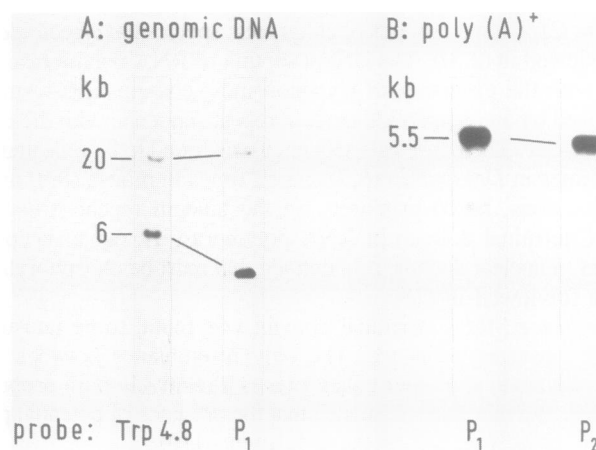


Fig. 1. Identification and isolation of the largest subunit gene of RNA polymerase II of *C. fasciculata*. Four μg of *C. fasciculata* DNA was digested with HindIII, size fractionated on a 0.5% agarose gel and transferred to nitrocellulose. The blot was hybridised with a nick-translated 2.4 kb EcoRI/HindIII fragment of pTrp4.8, encoding part of the RNA polymerase II largest subunit of *T. brucei* (10). The 6 kb fragment that strongly cross-hybridises with the *T. brucei* probe was isolated from a genomic bank (probe P₁, Figure 2) and used on an identical blot under high stringency conditions ($0.1\times\text{SSC}$ at 65°C ; panel A). To localise the approximate position of the coding sequence, probes P₁ and P₂ (Figure 2) were used as hybridisation probes against 6 μg poly(A)⁺ RNA from *C. fasciculata* (panel B).

genes (2,3), indicating that the gene is a homologue of an RNA polymerase largest subunit gene. As expected on the basis of the heterologous hybridisation with the *T.brucei* RNA polymerase II probe, the highest degree of conservation is observed between the deduced amino acid sequences of *C.fasciculata* and the *T.brucei* RNA polymerase II genes. Of the parts analysed, the amino acid positional identity between both representatives of the Kinetoplastida is 88% of 140 residues (block F) and 85% of 54 residues (block H)(Figure 4). Since block H is followed by a C-terminal extension, a unique characteristic of all eukaryotic RNA polymerase II largest subunit genes (see below), these results strongly suggest that the isolated *Crithidia* locus encodes the largest subunit of RNA polymerase II.

The C-terminal domain of *Crithidia*

The *Crithidia* and *T.brucei* domains F and H can be aligned rather easily, indicating that no major deletions or insertions have occurred during the evolutionary separation of these species. The amino acid sequence conservation is, however, abruptly lost at the start of the *Crithidia* C-terminal domain. To our surprise, we were unable to detect any significant positional homology of amino acid residues in the *T.brucei* and *C.fasciculata* C-terminal domains. However, the overall amino acid composition of both domains is strikingly similar. Both domains are characterised by a high content of acidic residues and are rich in serine and tyrosine residues. Moreover, the C-terminal domain of *T.brucei* is 79 amino acids longer than that of *Crithidia* (Table 1). Analysis of the sequences of the C-terminal domain revealed that the typical tandemly repeated heptapeptide sequence of eukaryotic RNA polymerases II was absent (Figure 5).

Finally, we noticed that most of the acidic residues are found dispersed throughout the C-terminal domain. However, the 3' most amino acid residues are characterised by an excess of acidic residues, creating a terminal acidic region (Figure 6). This acidic region is not only present in *C.fasciculata* and *T.brucei*, but is a general feature of all eukaryotic RNA polymerases II analysed to date (Figure 6). Computer analysis indicated that this region cannot form a putative amphipatic alpha-helix (data not shown) as was observed in the yeast transcription activator GAL4 (reviewed in (14)).

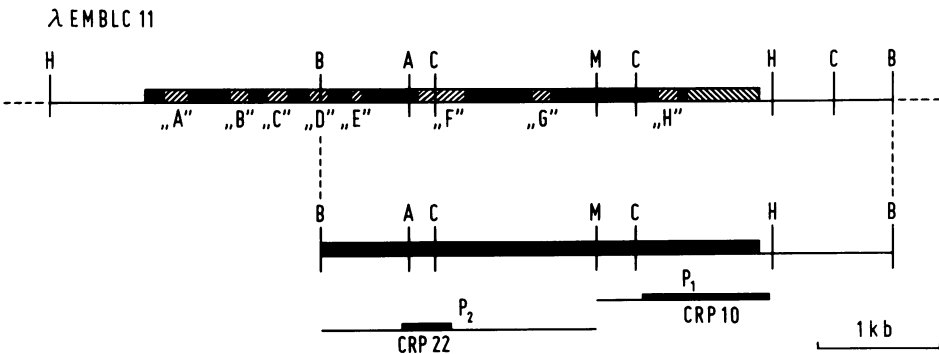


Fig. 2. Molecular analysis of the putative RNA polymerase II locus of *C.fasciculata*. The restriction map of the coding region is indicated in the upper panel. Probes originating from the genomic phage clone EMBLC11 are indicated by capital letters below the map. The putative structure and the conserved regions of the gene are indicated, based on the structure of other eukaryotic RNA polymerase II genes (2,3). Fragments predicted to contain homology block F (pCrp22) and homology block H and the C-terminus (pCrp10) are indicated below the diagram (lower panel). Abbreviations of restriction enzyme sites: A = SacII, B = BglII, C = SacI, H = HindIII and M SmaI.

Nucleic Acids Research

A

```

CGC GCC AAG AAG GCG CTG AGC AAT CGC CGC ACG AAC AGC TTC AAG GTG ATG ATT 57
A A K K A L S N R R T N S F K V M I E GAG
GCC GGC AGC AAG GGC AGT GAT CTT AAC ATT TGC CAG ATC CCG GTC TTC GTC GGC 114
A G S K G S D L N I C Q I P V F V G Q CAG
CAG AAC GTC GCG GGC AGC CGC ATC CCC TTC TTC CGC CGC CGC ACC GTG CCG CAC TTC 171
Q N V A G S R I P F F R R R T V P H F
ATG TTG GAC GAC TAC GGC GAG ACG TCG CGT GGC ATG GCG ACC CGC GGC TAC GTG GAG 228
M L D D Y G E T S R G M A T R G Y V E
GGG CTG CAG CCG CAC GAG TTC TAC TTC CAC ACC ATG GCC GGT CGT GAG GGC CTC ATC 285
G L Q P H E F Y F H T M A G R E G L I
GAC ACA GCC GTG AAG ACC GCC GAT ACG GGC TAC CTG CAG CGC AAG CTG GTG AAG CGC 342
D T A V K T A D T G Y L Q R K L V K A
CTG GAG GAC GTG CAC GCC GCG TAC GAC GGG ACG GTG CGC AAC GCC AAC CAG GAG CTC 399
L E V H A A Y D G N A N Q G E L
ATC CAG CTG GCG TAT GGC GAG GAC GGG CTG GAT GGG GCA CGC ATT GAG GGC AAC CAG 456
I Q L A Y G E D G L D G A R I E G N Q
ACC TTC CGA TCC
T F R S
    
```

B

```

TAC ACC ATC CTC GTC GAC ACC ATG TTG CCA CGC GGC TAC TTG ATG GCG GTG AGC 57
Y T I L V D T M L P R G Y L M A V S R CGC
ACG GGC ATC AAC CGC AGC GAG ACG TCC GGC CCG CTC ATG CGC AGC TCC TTC GAG 114
T G I N R S E T S G P L M R S S F G E E GAG
ACG GTG AAG GTG CTC ATG ACG GCG GCT GCC TTT GGT GAG AAG GAC CCG GTG CCG GGC 171
T V K V L M T A A A F G E K D P V R GGC
GTC TCG GCG AAC CTC GTG CTG GGC AAC CAG GCC CGC ATC GGC ACC GGG CTC TTC GAC 228
V S A N L V L G N Q A R I G T G L F D
CTC ATG CTT GAC ATG AGC AAG CTG CAG CAC GTC GTG CAC AAG GCG ACC GAG 285
L M L D M S K L Q H V V P L D K A C T E
ACG CGC ACC TCG AAC GTG TAC CAC ACC GAT GCC TCG GTG GCG CCG GGG TCG TCG 342
T R T S N V Y H T D A S V A P G S S L CTG
CAG GGC CTG CAC AGC GAC CTG CCG CCG TCC ACC GTG CAC GAG AAC AGC TCC CTC 399
Q G L H S D L P P S T V H E N S S L GGT
CTC GGC GCT TCT TCG GTG TAC CCG GCC ACG GCG GTG GTG GCG GTG GTG GCC TCT 456
L G A S S V Y P A T A V V A V V A S T ACG
CCG GCA TGG CTA TCG AGG CGA GTC AGG TGC ACT TCA GCA GCG TCG CCC ACC CAC 513
P A W L S R R V R C T S A A S P T H GAT
GGG GGC CGC CCT CGC CGC GTC CAA CAC GAG CGA CTA CCA GAG CAG CCA GCG CCA 570
G G R P R R V Q H E R L P E Q CCA GCG CCA CTC
CGC CGC GTC GAG TTA CGT GGC GTC GTC GAA CCT GGT CAG CGC GAG CCG CTG GAC 627
R R V E L R G V V E P G Q R E P L D A GCG
GAG CTG TCC TCG TAC CAC CTG CAG TCC GTC GCT CCC TCC GCG TAT CCG GCA GCC 684
E L S S Y H L Q S V A P S A Y P A A GGC
TAC GGG GCG GCG CTG TCA GCC ATG CCG ATG CCA CCC GGT CGC GAG TCC GGT GCG 741
Y G A A L S A M P M P P G A S F G A TTG
GCC GGC CAG CCG CCG TAC CCG TTT GAG GCC TCC ACC ACG GCA GCG GGG GTG CTG 798
A G Q P Y P F E A S N T A A G V L S AGC
GCG GCG GGC AGC TCG CGT GCG AGT GCG CCG TAC GAC CCC AAC CAG CAG TCG CAG 855
A A G S S R A S A P Y D P N Q Q S Q D
TTC TCG CCT ACT GAG GAA CAG GAG GAG CCA TAGGCGGAGGAAAGGACAGAGTTGTGTCTTCGCC 920
F S P T E E Q E E P *
    
```

1024
CACAGATATGAAAAGAAAAATGCGTGAGG

DISCUSSION

Structure of the C-terminal domain of the largest subunit of C.fasciculata RNA polymerase II
We have described the cloning of the gene encoding the largest subunit of RNA polymerase II of *C.fasciculata* (Figures 1, 3 and 4). We show here that the *C.fasciculata* gene also encodes a C-terminal extension, which is present in the largest subunit of all RNA polymerases II analysed so far (Table 1, Figure 5).

We noticed that the crithidial C-terminal domain is 79 amino acids shorter than the C-terminal domain of *T.brucei* (Table 1 and Figure 5). This suggests that part of this domain is not essential for *in vivo* RNA polymerase II transcription. This was clearly demonstrated in other eukaryotic species by elegant deletion mapping experiments in which it was shown that the presence of only half of the domain is required for viability (5,6,9,15). These experiments also showed that the repeats are not all functionally identical (9) and they exhibit species specificity (6). We are, unfortunately, unable to critically test these aspects for the *T.brucei* and *C.fasciculata* C-terminal domains, since a homologous transformation system is not available for these parasites.

In yeast, mouse and hamster, the structure of the C-terminal extension of the largest subunit of RNA polymerase II shows a high degree of conservation. In these three species the domain consists of a 26–52 tandemly repeated heptapeptide sequence with the consensus Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The sequence of the *Drosophila* domain has diverged considerably from this consensus sequence, but the basic structure observed in other higher eukaryotes is still present (reviewed in (2,10)). Three characteristics of the C-terminal domain in eukaryotes suggest that this region might form a unique secondary structure: (i) the amino acid composition, mainly residues with aliphatic hydroxyl side chains, (ii) the conservation of the consensus sequence and (iii) the repeat structure (see also (16)). It is, therefore, rather striking that no structural and consensus sequence homology was found between these domains and that of *T.brucei* and *C.fasciculata* (Figures 3 and 5). Not only is the repeat structure of the domain not conserved in either of the protozoans, but the strong reduction in the number of residues with aliphatic hydroxyl side chains and the presence of a relatively high number of acidic residues must have major implications for the secondary structure of the protozoan C-terminal domain.

For yeast and other higher eukaryotes, a functional, rather than a structural conservation of protein–protein interactions in RNA polymerase II transcription has been demonstrated. Transcription activators (reviewed in (14)), TATA-binding factors and other components of the transcription machinery function in heterologous transcription systems, both *in vitro* and *in vivo* (17–21). Because of its peculiar secondary structure, it has been suggested that the C-terminal domain plays a major role in binding RNA polymerase II-specific transcription factors, leading to accurate transcription and stabilisation of the transcription-initiation-complex (eg. (4,22)). The conservation of the structural features of the C-terminal domain among eukaryotes agrees nicely with the observed conservation of functional protein–protein interactions. Our data suggest, however, that the secondary structure of the C-terminal domain in Kinetoplastid flagellates is markedly different from that of higher eukaryotes. The aberrant structure of the C-terminal domain of Kinetoplastid species most

Fig. 3. Nucleotide and predicted amino acid sequences from the inserts of pCrp22 (panel A) and pCrp10 (panel B) of the *C.fasciculata* largest subunit gene of RNA polymerase II. The start of the C-terminal domain is indicated by an arrow and is in agreement with the C-terminal domain of *T.brucei*, based on the complete RNA polymerase II gene sequence (10). The sequence data have been submitted to the EMBL/GenBank Data Libraries under accession numbers X13489 (Crp10) and X13490 (Crp22).

Block F

C.facioluate pol II: M I E A G S K G S D L M I C H I P V F V G G H V A G S R I P F F . R R R T V P H F M L D D Y G E T S R G M A T R G Y V E G L O P M E F Y F N T M
T.brucel pol II : 749 T A G F L L M K F
yeast pol II : 746 . V M F I . A . M S A C . . S . E . K . A . G F Y D . L . . S K S P E . K . F V E M S . L R . T . G . F . F . A .
mouse pol II : 769 . V V S . A . . K I . S . V . A V E . K G F K H . L I K E P E F V E S S . L A . T . T . F . F . A .
Drosophila pol II : 761 . V V N I . S . V I A C K Y G F . K . L I K P F V E M S . L A . T . S A .
T.brucel pol III : 782 . V G S A A . M M A C T . S . K . L D A F O D . S L H R F E E A P A A F V A N S F Y S S . T F
yeast pol III : 794 . A T C T . V S . M V A V I I S . M . V . D G F O D . S L P K R S K T P Q . K . F V R M S F F P . L . L . A I
yeast pol I : 929 . A L S . A M V . V S M . L L A L E . R . V . V M V S G K . L . S . K P Y E T D A M A G . Y V K G R F Y S . I K . Q . Y C .

Block G

C.facioluate pol II: A G R E G L I D T A V K T A D T G Y L O R K L V K A L E D V H A A Y D G T V R N A M G E L L O L A Y G E D G L D G A R I E G N O T F R S
T.brucel pol II : S I I D F M G . L . P L 809
yeast pol II : G E . I . R I M V H . M . T . S L G M V F I M . A . H K Q S L D T I 806
mouse pol II : G . G . V E . I . R . I . S M . S . M V K . A S I M Q V V . R A . E S V . F O M L A T L 909
Drosophila pol II : G E . I . R . I . M . S . M V H S V G O R S . R R L V E F Q . M P T V K 900
T.brucel pol III : V E . I Y . R . M . M . M L S V R M T K . D V F G . D P L M G T P L 922
yeast pol III : S V E . M S . R . M . S L S C Q M T S A M G I V . F T G P L E M A D P V M 934
yeast pol I : S R S C . T . O G V S M S I . D . D G T . V . F M G . A I I T K E S H M T O . E F 1068

Block H

C.facioluate pol III: S G P L M R S S F E E T K V L M T A A F G E K D P P V R G V S A N L V L G H O A R I G T G L F D L M L D M S
T.brucel pol II : :1421 C A . S C V V V . M . A 1475
yeast pol II : :1394 A C E I F E . G . S A . L . D C E . V I D M . P A . V . I . E E 1446
mouse pol II : :1422 P K C D E . A . H . S . H K E . I M . Q L . P A C L . A E 1478
Drosophila pol II : :1415 G A . M R C D D M A . T . M E . I I M . Q L P K M C L . A E 1469
T.brucel pol III : :1434 F M V T M A R . T D N . Y H T O R V . R D L S D S I I V . K P V P L T S L . G . 1488
yeast pol III : :1381 D S V . O L A K . T D H . F D F Y H K . A . A E C I I O T M S S . R V V K G T M 1435
yeast pol I : :1607 T S S F . K M . Y T . C Q F . T K . V L D N . R E Q L D S P R I . V . K L M N V S V L A K V P 1651

likely has an effect on the structural requirements of the protein–protein interactions discussed above. Whether the absence of the repeat structure is confined to the Kinetoplastida or is a general feature of protozoans awaits analysis of other protozoan species.

The significance of the acidic tail, which is present in all eukaryotic C-terminal domains (Figure 6), is unclear. Deletion mapping experiments in yeast and mouse showed that mutants in which the acidic tail was removed were viable (5,9), indicating that removal of the acidic tail as such does not interfere with cell viability. This is probably due to the fact that the deletion of this structure, as well as part of the heptapeptide repeat structure, does not seem to affect the specificity of the initiation step of RNA polymerase II (7). These data are in contrast with two other observations. Firstly, the transfection efficiency of the two mouse deletion mutants (Del 50'-Ter and Del 36-Ter) in which a number of heptapeptide repeats as well as the acidic tail of the C-terminal domain were removed, resulting in a truncated C-terminus with a basic tail, was significantly reduced in comparison to similar deletion mutants in which the acidic tail was still present (9). Secondly, and more importantly, is the observation that all five yeast mutants (N2-B, N3-B, N4-B, N5-B and N10-B) in which the 15 residue acidic tail is replaced by a 39 amino acid basic tail containing a high proportion of hydrophobic residues (Figure 6) were non-viable (5). On the basis of the latter two observations, it cannot be excluded that the acidic tail of the C-terminal domain might have a function in RNA polymerase II transcription, a hypothesis supported by the observed conservation of this amino acid motif in all largest subunits analysed to date (Figure 6).

That an acidic region as such can mediate an important function is not without precedent (see (14) for review). For example, transcription of the structural genes for galactose metabolism in yeast is regulated by the transcription activator GAL4 (14). GAL4 contains two acidic regions, each of which can activate transcription (23), and mutation analysis showed that the rate of transcription is dependent on the acidity of these domains (24).

More extensive mutational analysis, coupled with a quantitative assay to determine RNA polymerase II activity, might lead to a better understanding of the role of the acidic tail in RNA polymerase II in transcription.

Phosphorylation

We have previously suggested that the most important function of the C-terminal domain in regulating RNA polymerase II transcription is determined by its *in vivo* phosphorylation. This was based on the documented role of phosphorylation in RNA polymerase II transcription (25–27) and the only similarity shared between the C-terminal domains of higher eukaryotes and *T.brucei*, namely the presence of potential phosphorylation sites (10). The present analysis of the distantly related species *C.fasciculata* provides further support for this proposed hypothesis. Although the positional homology of the C-terminal domain is low, the overall amino acid composition of this region is similar in both parasites (Table 1). Like in *T.brucei*, the C-terminus of *C.fasciculata* is characterised by the presence of potential phosphorylation sites. The strict conservation of these sites in the primary sequence of such distantly related species (11,12,28) strongly indicates a functional constraint. Therefore, the phosphorylation of the C-terminal domain is likely a decisive factor in RNA polymerase II function.

Fig. 4. Comparisons of homologous amino acid sequence motifs of domain F and H from all published eukaryotic RNA polymerase largest subunit genes. Amino acid positions are given at the beginning and end of each sequence. Residues identical to the *C.fasciculata* coding sequence are indicated by points.

Table 1. Amino acid composition of the C-terminal domains of the largest subunit of the putative RNA polymerases II from *C.fasciculata* and *T.brucei*.

<i>Crithidia fasciculata</i>¹					
Ala	29	(14.6)	Leu	15	(7.5)
Arg	13	(6.5)	Lys	0	(0.0)
Asn	4	(2.0)	Met	2	(1.0)
Asp	6	(3.0)	Phe	3	(1.5)
Cys	1	(0.5)	Pro	24	(12.1)
Gln	10	(5.0)	Ser	27	(13.6)
Glu	15	(7.5)	Thr	9	(4.5)
Gly	15	(7.5)	Trp	1	(0.5)
His	6	(3.0)	Tyr	7	(3.5)
Ile	0	(0.0)	Val	15	(7.5)
Acidic (Asp + Glu)				18	(9.0)
Basic (Arg + Lys)				13	(6.5)
Aromatic (Phe + Trp + Tyr)				11	(5.5)
Hydrophobic (Aromatic + Ile + Leu + Met + Val)				43	(21.6)
<i>Trypanosoma brucei</i>²					
Ala	31	(11.2)	Leu	11	(4.0)
Arg	14	(5.0)	Lys	2	(0.7)
Asn	9	(3.2)	Met	9	(3.2)
Asp	11	(4.0)	Phe	5	(1.8)
Cys	0	(0.0)	Pro	20	(7.2)
Gln	15	(5.4)	Ser	48	(17.3)
Glu	18	(6.5)	Thr	19	(6.8)
Gly	19	(6.8)	Trp	1	(0.4)
His	11	(4.0)	Tyr	15	(5.4)
Ile	1	(0.4)	Val	15	(5.4)
Acidic (Asp + Glu)				29	(10.4)
Basic (Arg + Lys)				16	(5.8)
Aromatic (Phe + Trp + Tyr)				21	(7.6)
Hydrophobic (Aromatic + Ile + Leu + Met + Val)				60	(21.6)

Footnotes to Table 1. The size of these domains is, respectively, 199 (1) and 278 (2) amino acid residues. The first column indicates the actual number of residues; the percentages are shown in the second column and are placed in brackets.

MATERIALS AND METHODS

Materials

Restriction enzymes and modifying enzymes were purchased from Boehringer-Mannheim and Pharmacia-LKB. ³²P-labeled nucleotides were purchased from Amersham. Sequence analysis was performed with the program described by Queen and Korn ((29); Microgenie™, Beckman Instruments).

Cells

C.fasciculata clone 1 (10) was grown in 3.7% (w/v) brain heart infusion broth supplemented with 20 mg/l hemin.

Molecular biology

Crithidia nuclear DNA was isolated according to standard procedures as described (30). Total RNA was isolated from *Crithidia* by LiCl-urea precipitation (31) and poly (A)⁺

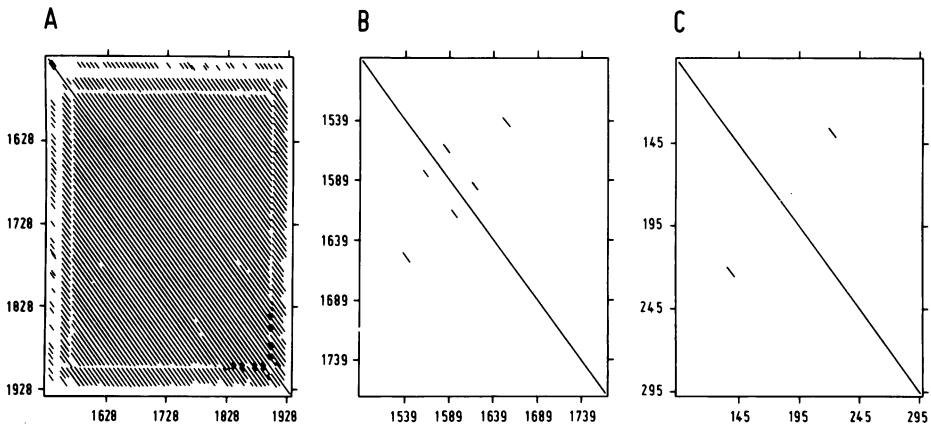


Fig. 5. Dot matrix analysis showing the absence of repetitions in the *C.fasciculata* C-terminal domain. The C-terminal domains of the genes encoding the largest subunit of RNA polymerase II of mouse ((8), panel A), *T.brucei* ((10), panel B) and *C.fasciculata* (panel C), were run against themselves. Dots were produced whenever 5 or more out of seven residues were identical (panel A). At this stringency no repeats were detected in *T.brucei* and *C.fasciculata*. A similar repeat structure as in panel A was seen when the yeast and *Drosophila* domains were analysed in this way (10).

RNA was purified by oligo(dT) cellulose chromatography (32). Plasmid DNA was isolated by the alkaline lysis procedure (33). Restriction endonuclease digestion, electrophoresis and Southern and northern transfers were performed as described (10,34). DNA and RNA blots were hybridised with nick-translated ^{32}P -labeled probes (35) as described (10). All post-hybridisation washes were to a final stringency of $0.1 \times \text{SSC}$ at 65°C or as indicated in the text ($1 \times \text{SSC} = 0.15\text{M NaCl}$, 0.015M sodium citrate, $\text{pH } 7.0$).

Isolation of the largest subunit gene of RNA polymerase II

We used a *Criethidia* genomic library in phage EMBL3, which was constructed and described by Swinkels et al. (36). This library was screened with a gene-internal 700 bp HindIII/PvuII fragment of pTrp5.9, encoding sequences of the largest subunit of RNA polymerase II of *T.brucei* (10).

DNA sequence analysis

Restriction fragments of the recombinant EMBL3 phage were subcloned into pEMBL8/9. The subclones were linearised at unique restriction sites in the plasmid polylinker and treated with Bal31 nuclease to obtain a series of progressively deleted clones. Cloning and preparation of template DNA was performed using standard protocols (37). DNA sequencing was performed using the dideoxy method (38), with modifications as described (39). Areas with high GC content were sequenced using 7-deaza dGTP (Boehringer Mannheim, FRG) as a substitute for dGTP. Part of the nucleotide structure was determined by the chemical degradation method (40). Both strands of the plasmids were sequenced.

ACKNOWLEDGEMENTS

We thank Drs. Piet Borst, Carol Gibbs and Peter Overath for constructive comments on the manuscript, Dr. Wolfgang Schuster for expert help with some of the experiments, Mr. Klaus Lamberty for artwork, Ms. Carmen Müller for photography and Ms. Brigitte

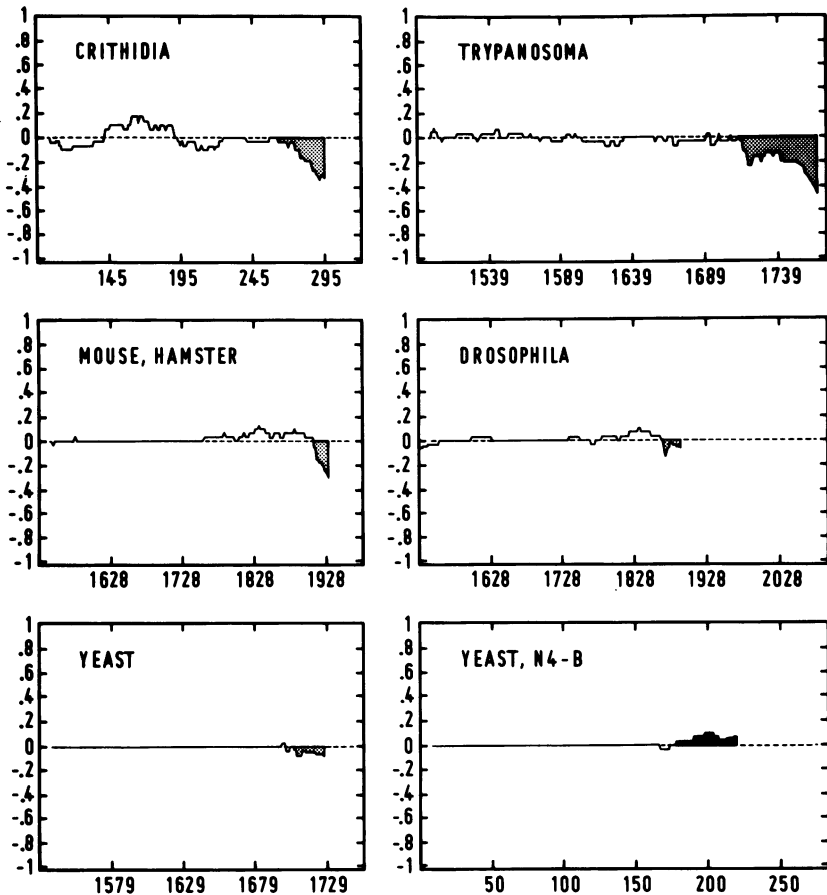


Fig. 6. Charge distribution of the C-terminal domains of the largest subunit of RNA polymerase II. The average net charges over 6 amino acid residues, measured at 1 amino acid intervals as described (41), is shown. The yeast mutant N4-B, in which the acidic tail sequence is replaced by a 39 amino acid tail is described by Nonet *et al.* (5).

Wall for editing the manuscript. This work was supported by Bundesministerium für Forschung und Technologie grant 0318885 and by a research award of the Erna and Victor Hasselblad Foundation to Albert W.C.A. Cornelissen.

*To whom correspondence should be addressed

REFERENCES

1. Sentenac, A. (1985) *CRC Crit. Rev. Biochem.* **18**, 31–90.
2. Cornelissen, A.W.C.A., Evers, R., and Köck, J. (1988) *Oxford Surveys Euk Gen* **5**, 91–131.
3. Köck, J., Evers, R. and Cornelissen, A.W.C.A. (1988) *Nucl. Acids Res.* **16**, 8753–8772.
4. Allison, L. A., Moyle, M., Shales, M., and Ingles, C. J. (1985) *Cell* **42**, 599–610.
5. Nonet, M., Sweetser, D., and Young, R.A. (1987) *Cell* **50**, 909–915.

6. Allison, L.A., Wong, J.K.-C., Fitzpatrick, V.D., Moyle, M., and Ingles, C.J. (1988) *Mol. Cell. Biol.* **8**, 321–329.
7. Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. (1988) *Proc. Natl. Acad. Sci. USA.* **85**, 3698–3702.
8. Corden, J. L., Cadena, D. L., Ahearn, J. M., and Dahmus, M. E. (1985) *Proc. Natl. Acad. Sci. USA.* **82**, 7934–7938.
9. Bartolomei, M. S., Halden, N. F., Cullen, C. R., and Corden, J. L. (1988) *Mol. Cell. Biol.* **8**, 330–339.
10. Evers, R., Hammer, A., Köck, J., Jess, W., Borst, P., Mémet, S., and Cornelissen, A.W.C.A. (1989) *Cell* **56**, in press.
11. Sloof, P., Van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R. and Noller, H. (1985) *Nucl. Acids Res.* **13**, 4171–4190.
12. Lake, J.A., de la Cruz, V.F., Ferreira, P.C.G., Morel, C., and Simpson, L. (1988) *Proc. Natl. Acad. Sci. USA.* **85**: 4779–4783.
13. Cornelissen, A.W.C.A., Evers, R., Grondal, E., Hammer, A., Jess, W. and Köck, J. (1989) *Nova Acta Leopoldina*, in press.
14. Ptashne, M. (1988) *Nature* **335**, 683–689.
15. Ahearn, J. M., Bartolomei, M. S., West, M. L., Cisek, L. J., and Corden, J. L. (1987) *J. Biol. Chem.* **262**, 10695–10705.
16. Sigler, P.B. (1988) *Nature* **333**: 210–212.
17. Buratowski, S., Hahn, S., Sharp, P.A., and Guarente, L. (1988) *Nature* **334**, 37–42.
18. Cavallini, B., Huet, J., Plassat, J.-L., Sentenac, A., Egly, J.-M., and Chambon, P. (1988) *Nature* **334**, 77–80.
19. Horikoshi, M., Carey, M.F., Kakidani, H., and Roeder, R.G. (1988) *Cell* **54**, 665–669.
20. Webster, N., Jin, J.R., Green, S., Hollis, M., and Chambon, P. (1988) *Cell* **52**, 169–178.
21. Webster, N.J.G., Green, S., Jin, J.R., and Chambon, P. (1988) *Cell* **54**: 199–207.
22. Ingles, C.J., Moyle, M., Allison, L.A., Wong, J. K.-C., Archambault, J. and Friesen, J.D. (1987) *UCLA Symp. Mol. Cell. Biol.* **52**, 383–393.
23. Ma, J., and Ptashne, M. (1987) *Cell* **48**, 847–853.
24. Gill, G., and Ptashne, M. (1987) *Cell* **51**, 121–126.
25. Bartholomew, B., Dahmus, M.E., and Meares, C.F. (1986) *J. Biol. Chem* **261**, 14226–14231.
26. Kim, W.-Y., and Dahmus, M.E. (1986) *J. Biol. Chem.* **261**, 14129–14225.
27. Cadena, D.L., and Dahmus, M.E. (1987) *J. Biol. Chem.* **262**, 12468–12474.
28. Sogin, M.L., Elwood, H.J., and Gunderson, J.H. (1986) *Proc. Natl. Acad. Sci. USA.* **83**, 1383–1387.
29. Queen, C., and Korn L.J. (1984) *Nucl. Acids Res.* **12**, 581–599.
30. Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H., and van Boom, J. (1982) *Nucl. Acids Res.* **10**, 3591–3604.
31. Auffray, C., and Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314.
32. Hoeijmakers, J.H.J., Borst, P., van den Burg, J., Weismann, C., and Cross, G.A.M. (1980) *Gene* **8**, 391–417.
33. Birnboim, M.C., and Doly, J. (1979) *Nucl. Acids Res.* **7**, 1513–1523.
34. Southern, E.M. (1975) *J. Mol. Biol.* **98**, 503–517.
35. Rigby, P.W., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
36. Swinkels, B.W., Evers, R., and Borst, P. (1988) *EMBO J.* **7**, 1159–1165.
37. Dente, L., Cesareni, G., and Cortese, R. (1983) *Nucl. Acids. Res.* **11**, 1645–1655.
38. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA.* **74**, 5463–5467.
39. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA.* **80**, 3963–3965.
40. Maxam, A.M., and Gilbert, W. (1980) *Meth Enzymol* **65**: 499–560.
41. Hopp, T.P., and Woods, K.R. (1981) *Proc. Natl. Acad. Sci. USA.* **78**, 3824–3828.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.