1662–1668 *Nucleic Acids Research, 1996, Vol. 24, No. 9*

Centromeric polymerase III transcription units in *Chironomus pallidivittatus*

Carlos Rovira[+] and Jan-Erik Edström*

Department of Genetics, Lund University, Sölvegatan 29, S-22362 Lund, Sweden

Received January 26, 1996; Revised and Accepted March 19, 1996

ABSTRACT

Cp1 is a polymorphic short interspersed repeat (SINE) which is distributed over the whole genome of the dipteran *Chironomus pallidivittatus*, and is particularly abundant in the centromeres. It contains two different sequence modules, one of which, the B module, has a polymerase III internal control region (ICR) typical for tRNA genes (A and B box). Such sequence motifs are common in SINEs and assumed to function in RNA-mediated transposition. In the present case, however, several structural features speak for another role. An investigation of the transcription of the B module shows that it encodes a 99 nt RNA species *in vivo*, Cp1-RNA, terminating within the module. The transcription unit is likely to have evolved from a pre-tRNA gene and the transcript has sequence similarities to non-processed pre-tRNA. Most of the *in vitro* transcription is eliminated by deletion or substitution mutation of an upstream TATA box, present within the B module, as well as by changing either the A or B box. The properties of the transcript suggest that it does not have a role in transposition but may have some other function, perhaps in the centromere.

INTRODUCTION

SINEs are transposable elements derived from RNA polymerase III (pol III) transcribed small structural RNAs which often carry functional internal promoters (ICR) for their own transposition. The *Alu* family in primates is the best characterized member of this group. It consists of an ~280 bp long dimeric DNA element derived from a 7SL RNA gene (1) and constitutes ~5% of the human genome (2). Most SINEs are, however, derived from tRNAs (reviewed in 3) and have been seen in many species, like mammals (4–6), fish (7) and plants (8).

We have described a new type of SINE, termed Cp1, in chironomid insects (9). This element is interspersed between arrays of centromeric tandem repeats (10) and is also present at extracentromeric sites. Cp1 is polymorphic and modular in design. It consists of two modules, the SCA and the B module, each ~200 bp long. The latter contains ICR consensus sequences (A and B box) and termination signals typical for pol III genes, and it has other sequence similarities with tRNA genes. Most SINEs only have internal transcription control elements and probably use stop signals downstream of the element for transposition. The unit in the B module has, however, properties suggesting transcription for other purposes than transposition. There is a TATA box within the SINE, upstream of the ICR region and there are typical pol III termination signals within the element. Centromeric tRNA genes have previously been observed in *Schizosaccharomyces pombe* (11,12). We were therefore interested in learning whether this component of the B module is transcriptionally active *in vivo* and *in vitro* and, if so, how *cis*-acting control elements can be defined in relation to the transcription unit.

Here we show that the B module is indeed able to direct pol III transcription *in vitro* and *in vivo*. Transcription *in vitro* requires the upstream TATA box in addition to both boxes in the ICR. The *in vitro* transcript is only 4 nucleotides (nt) larger than the *in vivo* form and some processing is therefore not excluded. The transcript is, however, more similar to pre-tRNA than tRNA and starts 3 nt upstream of the 5′-end and extends 23 nt beyond the 3′-end of a processed tRNA. A predictive computer program suggests that the product differs from tRNA in secondary structure and adopts a hairpin structure. The transcript is unlikely to be of importance for transposition of Cp1 since it only represents a minor part of the element. The well defined transcriptional control elements which have survived evolution from a tRNA gene indicate some function for this transcription unit. We propose that this role is exerted in the centromere.

MATERIALS AND METHODS

RNA isolation, sequencing and Northern analysis

Total RNA was extracted from *C.pallidivittatus* embryos and salivary glands by standard methods (13). For RNA sequencing 10 µg total RNA were annealed to 10 pmol 5′-end-labelled primer (5′-AATGACTCTTCCCGAGC-3′). The cDNA strand was polymerized with 3 U AMV reverse transcriptase (Boehringer Mannheim) per 20 µl buffer at 42°C in a mixture of deoxy/dideoxynucleotides.

For Northern analysis total RNA was used as well as the poly(A)$^+$ fraction, isolated with the Dynabeads biomagnetic separation system (Dynal). The RNAs were fractionated in a 6% polyacrylamide denaturing gel (19:1) containing 8 M urea. The RNA was then electrophoretically transferred (Trans Blot Cell,

* To whom correspondence should be addressed

[+]Present address: Max-Planck Institute for Molecular Genetics, D-14195 Berlin (Dahlem), Germany

BioRad) onto Hybond$^+$ nylon membranes (Amersham) in 0.5× Tris–borate buffer at 20 V/cm overnight in the cold. The filters were hybridized with the clone pCp254 (10) and rehybridized with the salivary gland secretory protein gene sp12 (14).

Primer extension

Total RNA (10 µg) was mixed with 5 pmol 5′-end-labelled primer described above in hybridization buffer (0.15 M KCl, 10 mM Tris–HCl pH 8.3, 1 mM EDTA). The mixture was incubated in boiling water for 15 min and then annealed at 48°C for 1 h.

The annealing reaction was precipitated and resuspended in 6 µl 5× AMV reverse transcriptase buffer (250 mM Tris–HCl, 40 mM $MgCl_2$, 150 mM KCl, 5 mM DTT; pH 8.5 at 20°C), 3 µl deoxynucleotides (5 mM each), 3 U AMV reverse transcriptase (Boehringer Mannheim) and water to 30 µl. The reaction was incubated at 42°C for 1 h. After precipitation cDNA was analyzed on a 8% polyacrylamide–7 M urea sequencing gel (19:1).

3′ RACE PCR

The 3′-end of the transcript was mapped according to Frohman *et al.* (15). Total embryonic RNA was treated with *E.coli* poly(A) polymerase (Pharmacia) in the presence of ATP. Polyadenylated RNA (~5 µg) was heated at 70°C and cooled on ice. The RNA was then added to a 20 µl reverse transcription mixture containing 2 pmol 3′-poly(dT) primer: 5′-AAGGATCCGTCGACATCGA-TAATACGACTCACTATAGGGAT$_{(17)}$-3′. After the addition of 10 U AMV reverse transcriptase the reaction was incubated for 2 h at 42°C. The resulting cDNA pool was diluted to 1 ml with water, 5 µl of which were used for PCR amplification with the following primer pair: (i) a 17mer primer present in the adapter poly(dT) primer (5′-AAGGATCCGTCGACATC-3′) containing a *Bam*HI restriction site, (ii) a 17mer internal to the B module and complementary to the 17mer used for primer extension and RNA sequencing (5′-GCTCGGGAAGAGTCATT-3′). The cDNA was then amplified by cycling at 94°C for 1 min, 50°C for 45 s and 72°C for 45 s, in 30 cycles. A single low molecular weight band was amplified. This product was isolated from agarose gel, blunt-ended with T4 polymerase and finally digested with *Bam*HI. The resulting fragment was cloned into a *Bam*HI/*Hin*dII digested pUC18 plasmid and sequenced by the dideoxy-termination method.

In vitro transcription

Chironomus tentans tissue cultured cells (16) were used for cytoplasmic cell extracts (17). All 40 µl transcription reactions contained 20 µl cell extract and were carried out in 80 mM KCl, 3 mM $MgCl_2$, 3 mM DTT, 30 mM HEPES–KOH pH 8.0, 0.5 mM each of unlabelled UTP, ATP and CTP, 0.050 mM [α-^{32}P]GTP (5 Ci/mmol), 8 mM creatine phosphate, 0.5 U creatine phosphokinase and 1 U/ml RNAsin (Promega). Supercoiled plasmid (500 ng) was used as template. All Cp1 constructs were cloned in pUC18 and most of them have previously been described: pCp627 and pCp254 (10), pCp116, pCp125, pCp413 and pCt2 (9). The pCt2.2 clone is the fragment of pCt2 between the left *Eco*RI site in pCt2 and the *Eco*RI site between the S and C fragments of the SCA module. The pCp1A clone is an unpublished PCR generated fragment where the B module is preceded by a SCA module and followed by a 155 bp centromeric repeat (10). The reactions were incubated at room temperature for 2 h, phenol/chloroform extracted and precipitated prior to analysis in 6% polyacrylamide–8 M urea gels (19:1).

Construction of 5′ deletions

The insert of the pCp627 clone (10) was isolated with *Eco*RI and treated with T4 polymerase. It was then cleaved with *Hin*dIII and a fragment containing the B module preceded by the A segment of the SCA module (which is the one present in pCp254, a subclone of pCp627) subcloned into the *Hin*dIII/*Hin*dII sites of pUC18. This plasmid was then linearized with *Xba*I/*Sac*I and digested with Exo III nuclease (Erase-a-base system, Promega) at 16°C. Mutants were recircularized, transformed and sequenced by the dideoxy-termination method.

Construction of substitution mutants

A series of plasmids with substitutions of TATA box, transcription start site, A and B boxes, respectively, were created by PCR with the clone pCp254 as template. The target regions were substituted by the introduction of restriction endonuclease sites with oligonucleotides containing the desired sequences in conjunction with primers flanking the cloning site of pUC18. The oligonucleotides used are shown in Table 1.

The PCR products were digested with *Spe*I for the TATA box mutant, *Cla*I for the transcription start site mutant and *Xho*I and *Sty*I for the A and B box mutants, respectively, and either *Eco*RI or *Hin*dIII and ligated into the polylinker of pUC18.

Table 1.

	Oligonucleotide
TATA box mutant	5′-CCTGC**ACTAGT**TTGGCATTCTGCCAAT-3′ and
	5′-CGATCG**ACTAGT**TGAACCACTCTTTTTTC-3′
Transcription start site mutant	5′-GATGCC**ATCGAT**AGCCCGATAGCTCAG-3′ and
	5′-CATAGC**ATCGAT**TTGGCATTGGCAGAATG-3′
A box	5′-CGATT**CTCGAG**TCTGAGCACTTGACCGGC-3′ and
	5′-TTTTC**CTCGAGTTATG**TCGGGCTTTGCCATTGGC-3′
B box	5′-TTAGCC**TCCTTGG**CGCTCGGGAAGAGTCATTGG-3′ and
	5′-GCTATT**CCAAGGACGTA**TCGCACCTCTCGATTGCC-3′

Substitutions shown in boldface.

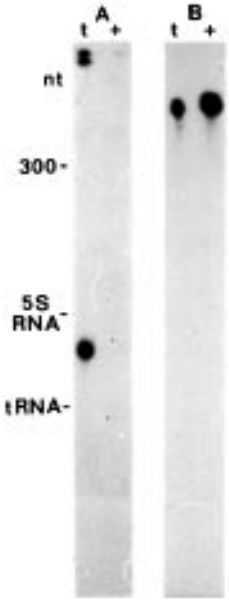

Figure 1. Northern blot of a total RNA extract from salivary glands (t) and the poly(A)$^+$ fraction of this extract (+) probed in (**A**) with pCp254 (10) and reprobed in (**B**) with sp12 salivary gland secretory protein gene (14). Markers were a mixture of tRNA, 5S RNA and RNA size marker III (Boehringer Mannheim).

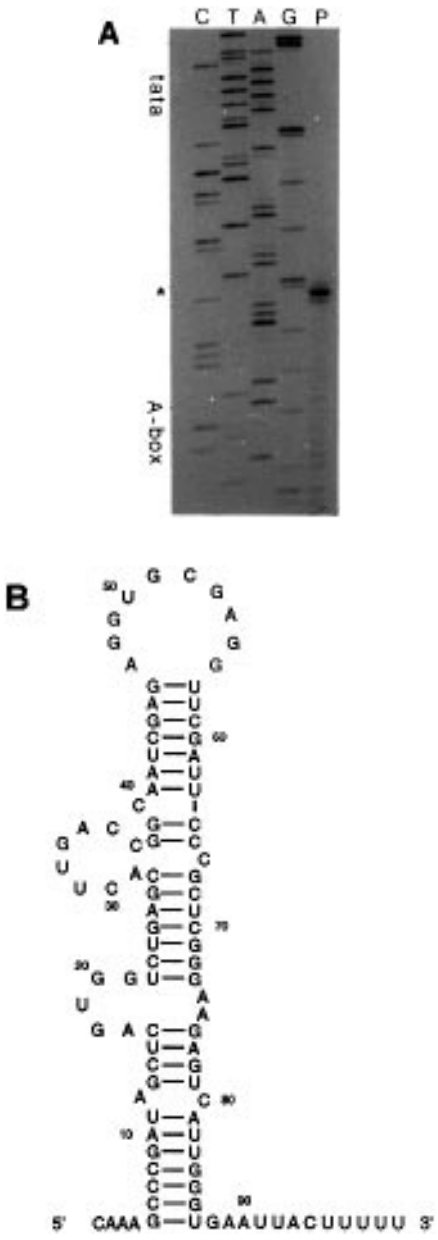

Figure 2. (**A**) Primer extension analysis of the start of the B module transcript. Lane P shows the band from the extended primer, position of which is indicated with an asterisk. (**B**) RNAFold plot of the 99 nt Cp1-RNA from module B.

RESULTS

The B module transcript

Northern analysis of RNA from embryos or salivary glands showed a low molecular weight RNA species in the 100 nt range in total RNA, absent in the poly(A)$^+$ fraction when pCp254 or a clone containing only the B module was used as probe (pCp254, with the whole A segment of the SCA module and the initial 29 bases of the B module deleted; Fig. 1). The product, designated Cp1-RNA, contained 99 nt as determined by sequencing. Its 5′-end was mapped by primer extension with help of a 17mer primer hybridizing downstream of the B box (positions 151–168 in the pCp254 sequence shown in Fig. 6). The main extension product corresponded to an 81 nt fragment, mapping the 5′-end of Cp1-RNA to 10 positions upstream of the first base pair of the A box (Fig. 2A). We sequenced the transcript in a total RNA mixture using the same 17mer as for primer extension. The 3′-end of Cp1-RNA was investigated with 3′-RACE PCR. A single electrophoretic band was obtained. The PCR product was cloned in pUC18 and sequenced. The RNA derived sequences agreed with those obtained directly from DNA within the stretch of 99 bp and are shown in Figure 2B (in predicted secondary structure form). The localization within the B module of DNA corresponding to the transcript is shown in Figure 3.

Genetic origin of transcription unit

Previously we found that much of the sequence can be aligned to tRNA genes from *D.melanogaster* (9). We have extended the search and found high scores in comparisons with several tRNA genes from different species, including rat tRNA^{Lys} and tRNA^{Thr} from prokaryotes (*Chlamydia* and *Micrococcus*). The identity ranges from 98% compared with the *Chlamydia* tRNA^{Thr} 5′-end (30 starting nt) to 86% with rat tRNA^{Lys} (for a stretch of 26 nt). However, the sequence did not give a typical tRNA-like cloverleaf with the RNAFold software but a hairpin (Fig. 2B).

On the basis of tRNA nearly invariant positions and the size of the different stems and loops (18) the stretch of 79 nt from the 5′-end can be arranged like a tRNA cloverleaf, although with significant differences (Fig. 4). Cp1-RNA starts 3 nt upstream of a processed tRNA and continues 23 nt beyond the downstream processing point. The cloverleaf structure, although unlikely to exist *in vivo*, strengthens the arguments for an evolutionary origin of the Cp1-RNA gene in a tRNA gene and it permits us to identify borders between regions that have an origin in mature tRNA and processed leader and trailer, respectively.

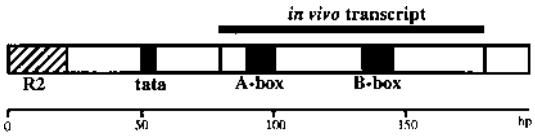

Figure 3. Schematic illustration of the B module and its various components. R2 stands for the initial 22 bp that have strong similarity to the integration site of the R2 non-LTR retrotransposon into the pre-ribosomal gene.

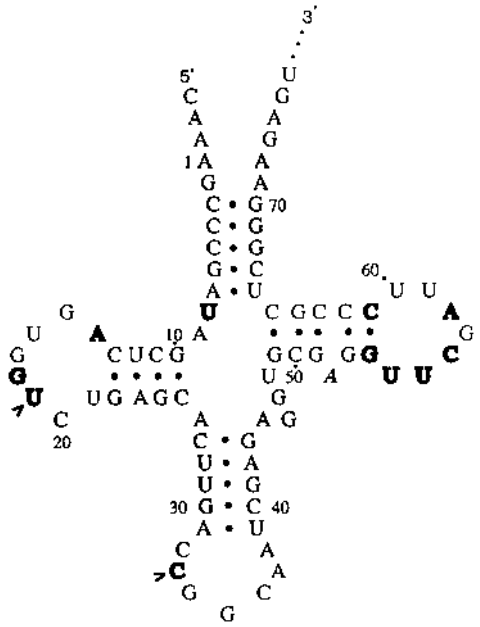

Figure 4. The initial 79 nt of Cp1-RNA arranged as a tRNA cloverleaf. Positions where tRNA is nearly invariant are shown in boldface, disagreements between Cp1-RNA and tRNA indicated with arrowheads. Compared to the tRNA consensus Cp1-RNA starts at –3 and has 5 rather than 7 bp in the acceptor stem. The invariant U-8 is present in Cp1-RNA like the 4 bp in the D-stem and A-14 and G-18. There is, however a U-19 instead of the tRNA G-19. It is also possible to reconstruct an anticodon stem with 5 bp but with a C-33 instead of a U-33 in the loop. What would be position 47 in a tRNA is missing in Cp1-RNA like in some tRNAs. The T-stem is disturbed by insertion of an *A* (italicized) between the tRNA positions 51 and 52 but other invariant features in the T-stem are retained like G-53, U-54, U-55 (Ψ-55 in tRNA), C-56, A-58 and C-61. Finally the 3′-end of Cp1-RNA is that of a pre-tRNA rather than a tRNA, in which the post-transcriptionally added CCA residues occupy positions 74–76. Cp1-RNA extends 23 nt beyond position 73 in the cloverleaf.

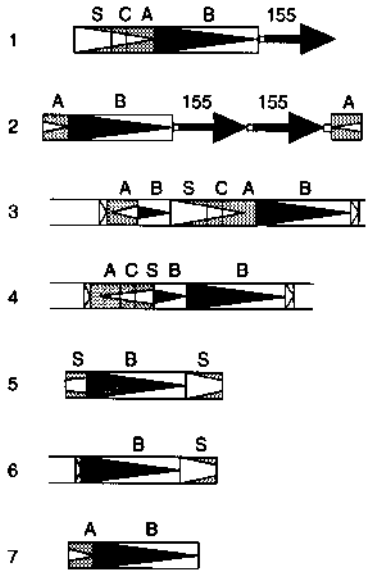

Figure 5. Different cloned Cp1 elements used in the *in vitro* transcription studies shown in Figure 7A: 1, pCp1A; 2, pCp627; 3, pCp116; 4, pCp413; 5, pCp125; 6, pCt 2.2 and 7, pCp254. The 155 bp centromeric tandem repeat is shown as a black arrow with the initial 14 bp as a white box, present on both sides of Cp1 elements that are interposed between 155 bp repeats. The B module is shown as a shadowed arrowhead against a white background and the SCA module as a white arrowhead against a shadowed background. The bases of the arrowheads are the 5′-ends and the tips the 3′-ends. The SCA module consists of three segments, S, C and A, with intervening *Eco*RI sites. Modules may be inverted and truncated to an extent that can be estimated from what is left of the arrowheads. No. 1 is an unpublished PCR generated intracentromeric fragment where the B module is preceded by a complete SCA module and followed by a centromeric 155 bp repeat. No. 2 represents parts of two different elements on either side of two centromeric repeats and no. 7 is the left part of no. 2. No. 3 and 4 are complete extracentromeric elements between target site duplications short unfilled arrowheads), embedded in flanking genomic host DNA. No. 5 is a fragment from an extracentromeric element. No. 6, pCt2.2 is the left part of pCt2, which is an extracentromeric element from *C.tentans*, pCt2.2 extends between an *Eco*RI site in the flanking DNA to the *Eco*RI site between the S and C segments in the SCA module. Nos 3–6 are described in references 9 and 2 and no. 7 in reference 10.

In vitro transcription of natural B modules

To examine transcriptional properties of the B module we tested different Cp1 containing plasmid clones in a cell free extract made from a *C.tentans* epithelial cell culture. These clones represent the B module in different naturally occurring DNA environments, schematically illustrated in Figure 5. The sequences of the B modules in these clones, which have some mutational differences, are given in Figure 6. It can be seen in Figure 7A that the three centromeric clones (no. 1, 2 and 7) and the extracentromeric transpositions (no. 3, 4 and 5) transcribed a 103 nt long RNA (exact size determined in sequencing gel) whereas the extracentromeric transposition no. 6 was inactive. A shorter 78 nt byproduct of the cell extract (used in the experiments shown in Fig. 7A and B) was also labelled, present also when no template was included (not shown). Consequently, there was a minor size difference between *in vitro* and *in vivo* transcribed RNA. We do not yet know the start point for the primary transcript which is likely to be that of the *in vitro* transcript and, consequently, within 4 nt of the start of the *in vivo* product.

The transcriptionally inactive B module (Fig. 7A, lane 6) contains ICR and TATA boxes matching the consensus. Three insertions are, however, present upstream of the A box at –2, –1 and +6 relative to the start of the *in vivo* transcript (Fig. 6). In another B module the A box contains a duplication of the C in the fourth box position and has a C substituted for a T in the first position. Nevertheless, transcription of this clone was not significantly lower than that of other clones (Fig. 7A, lane 5).

Several gene classes transcribed by RNA pol III are known to either require or be influenced by both 5′- and 3′-flanking regions. In the experiment reviewed above the templates contain B modules flanked at their 5′- and 3′-ends by DNA sequences of different kinds. At the 5′-side transcriptionally active units border

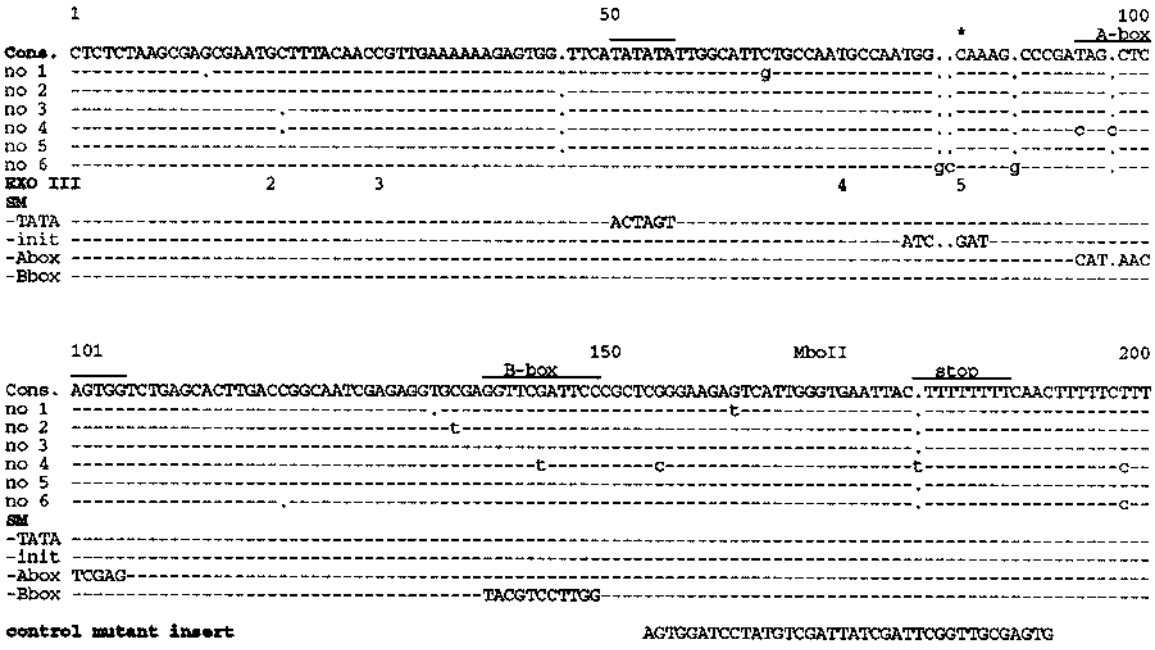

Figure 6. Top line shows consensus of B modules from clones used in experiments of Figure 7A. Asterisk shows first nucleotide of *in vivo* transcript. Below are sequences of B modules shown schematically in Figure 5 and used for experiments in Figure 7A. The B module in no. 2 (from pCp627) is identical to that of pCp254, i.e. no. 7 in Figure 5 and 7A. First bases in cloned fragments remaining after Exo III deletions from the 5′-end of pCp254 are indicated in subsequent line by numbers used in lanes for electrophoretic separations of Figure 7B. Further down, under SM, are substitution mutants, covering TATA box, initiator region, A and B box, respectively. Bottom line shows 38 bp fragment in the control mutant inserted into the *Mbo*II site between position 167 and 168 in pCp254.

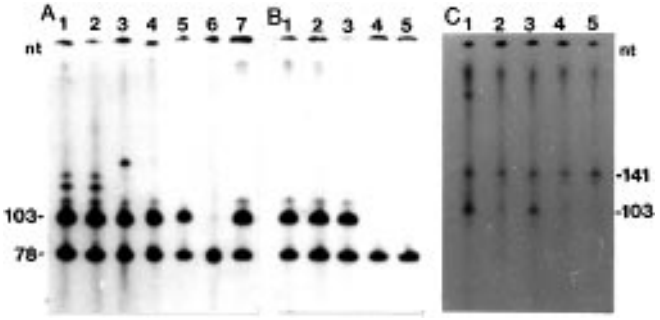

Figure 7. (**A**) Autoradiogram of polyacrylamide gel showing *in vitro* transcription products of Cp1 elements. The clones used as templates are shown in Figure 5 and have the same numbering. (**B**) Autoradiogram of polyacrylamide gel showing *in vitro* transcripts of deletion constructs, numbered as in the Exo III line of Figure 6, where first bases in cloned fragments after Exo III deletions from the 5′-end of the pCp254 insert are shown. The leftmost separation no. 1 refers to the intact B module. (**C**) *In vitro* transcription of substitution mutants shown in Figure 6. The templates in (C) are in turn and order 1, the non-mutated pCp254; 2, the TATA box mutant; 3, the initiator region mutant; 4, the A box mutant and 5, the B box mutant. In all separations the 103 nt band is due to transcription units in the B module, the 78 nt band in (A) and (B) is a byproduct of the extract used in these separations, the 141 nt band in (C) is from the control construct with 38 bp insertion into module B.

to a complete SCA module (nos 1 and 3), to the terminal A segment of such a module (nos 2 and 7), to an inverted, initial part of the SCA module (no. 5), or to the terminal part of another B module, in turn preceded by a large segment of an inverted SCA module (no. 4). At the 3′-side there are centromeric repeats for no. 1 and 2, the upstream part of the SCA module in no. 5, target site duplication with adjoining host genomic DNA in nos 3 and 4 and vector sequence in no. 7. These differences did not affect the transcriptional efficiency of the templates.

Further support that Cp1-RNA is a pol III product was obtained when α-amanitin was used *in vitro* during transcription of pCp254 (no. 7), which was resistant to the drug at the highest concentration tested, 500 μg/ml (results not shown), in agreement with previous observations of insensitivity of insect pol III *(Bombyx mori)* to the drug (19). A *KrYppel* pol II promoter construct (Stratagene) was completely inactive in the cell free extract from the *C.tentans* epithelial cell culture and, therefore, could not be used to control the efficiency of α-amanitin in this extract. In parallel experiments we could, however, show that the pol II construct transcribed in a *Drosophila* embryonic nuclear extract (Stratagene) and was completely sensitive to α-amanitin down to 5 μg/ml, as expected. Furthermore, this extract promoted weak transcriptional activity from pCp254 (tested in the presence of the *Krüppel* construct) and this transcription was completely insensitive to α-amantin even at the highest concentration tested, 250 μg/ml (results not shown). An additional important aspect of these experiments is that they show that our extract from the *C.tentans* epithelial cell culture is specific for pol III.

In vitro transcription of B modules progressively deleted from the 5′-end

Since transcription of the B module was independent of components outside of the unit, we focused our attention on possible regulatory elements within the module, in particular the region delimited by the transcription start and the 5′-end. Within this 80 bp long region there is a sequence that conforms with the TATA box consensus (positions –31 to –25 in relation to the first

nucleotide of the *in vivo* transcript). To determine whether upstream elements influence transcription of the B module, we created a series of partially deleted B modules by progressive digestions of the pCp254 insert (Fig. 5, no. 7) with Exo III nuclease. Deletion end points are shown in Figure 6 and results of *in vitro* transcription in Figure 7B. One mutant, extending to position –99 (in relation to the first nucleotide of the *in vivo* transcript) only eliminates part of the A segment of the SCA module upstream of the B module which is intact. The mutants used in lanes 2 and 3, with deletions to –61 and –50, affecting the 5′ part of the B module, directed transcription at the same rate as the intact module in lane 1. With larger deletions, passing the TATA box, transcription was completely abolished (lanes 4 and 5). No detectable activity was found even after long exposure times.

In vitro transcription of base substitution mutants

The B module of pCp254 was mutated by substitutions that eliminated the TATA box, the initiation region and the A and B boxes, respectively, as shown in Figure 6. DNA from the constructs was transcribed together with control DNA (Fig. 7C). The control contained a 38 bp insertion (Fig. 6) into the *Mbo*II site between the B box and the first stop signal (between position 167 and 168). Transcription was largely dependent on the integrity of the TATA box and the two boxes of the ICR. There were weak 103 nt signals for the TATA (lane 2) and A box (lane 4) constructs but incorporation directed by the box B mutant (lane 5) could be seen only after prolonged exposure. When, however, the initiation region was mutated (lane 3), there was no effect on transcription. The relative template efficiency of the control construct was somewhat lower than that of the non-mutated template, and it can therefore not be excluded that the region around the *Mbo*II site has some effect on transcription efficiency. The 78 nt byproduct was not seen with the extract used in these experiments.

DISCUSSION

Chironomus centromeres have accumulated a polymorphic SINE-like transposable element, Cp1, present also in extracentromeric positions (9). It contains two sequence modules, with variable arrangements in different elements. Both modules start with a 22 bp sequence similar to the integration site for the non-LTR retrotransposon, R2, in the pre-ribosomal gene (20). One part of the B module is likely to be tRNA derived, and contains consensus sequences for pol III internal control regions. The origin of the remainder of Cp1, i.e. less than half of B and most of SCA has not been traced. ICRs with A and B boxes are common in SINEs, reflecting their origin, and do not necessarily imply cellular function (21). The B module has, however, structural features unusual for a SINE, possibly due to a cellular role other than transposition. Firstly the ICR box sequences, with one exception, always agree with established consensus, secondly distinct pol III stop signals are present within the element, which is not the case for SINEs in general, and thirdly the transcription unit is preceded by a TATA box. Furthermore, DNA motifs for pol III transcription are not localized to the 5′-part of the element, which is probably necessary for transposition. We therefore suspected that the B module contains a complete pol III transcription unit, entirely controlled within the module, probably unrelated to a role in transposition.

There was pronounced transcription *in vivo* resulting in a product within the range of pre-tRNA, i.e. 99 nt, designated Cp1-RNA. *In vitro* an only slightly larger transcript was obtained, 103 nt. The small difference between *in vivo* and *in vitro* results could have alternative explanations, such as processing in the nuclear compartment, requiring factors not present in the cytoplasmic extract, readthrough at the stop signal *in vitro* and the possible presence of a master gene for a 99 nt transcript. There was no evidence for a processed product of a size more typical for mature tRNAs. The difference in size between a mature tRNA and Cp1-RNA is due to differences at both the 5′- and 3′-ends. Cp1-RNA starts 3 nt upstream and terminates 23 nt downstream of a typical tRNA. The similarities with pre $tRNA^{ala}_2$ of *Bombyx mori* are striking, where the processed parts of the 5′ leader and 3′ trailer contain 3 and 22 nt, respectively (22).

The transcription unit in the B module is thus likely to have evolved from a tRNA gene, during which mutations have taken place that result in loss of processing and may have changed the secondary structure from a cloverleaf to a hairpin.

Some transcriptional control features probably also have evolved. *Cis*-acting control elements were studied with natural B modules, experimental deletions and substitution mutants. We found that *cis* elements required for basal transcription are contained within a stretch extending to –50, still within the B-module. This region has a TATA box starting at –31. The TATA box is indeed an essential element, since a substitution mutation of this motif eliminated almost all transcription. Also each component of the ICR were, separately, essential for transcription as shown by other experimental mutations. A naturally occurring variant of the A box with only two base changes nevertheless transcribed. Our experiments do not exclude other control elements within the transcription unit and its upstream region extending to –50.

We found no evidence for regulatory sequences downstream of the B module (which terminates ~20 bp downstream of the transcription unit). In this respect transcription control may differ from that of some insect pol III transcribed genes which may be influenced by relatively large 3′ flanking regions (23,24).

Interestingly a natural variant which had three extra base pairs between the TATA box and the A box (pCt2.2), in the initiator region, was completely inactive *in vitro*. Without additional experimental support this result must be interpreted with caution. Since, however, a substitution mutation in the same region showed full activity, it may well be that it is the change in tilt of the DNA double helix between the two boxes due to the insertions, almost one third of a turn, that inactivates transcription. This could be because TFIIIC, binding the ICR, is interacting with TFIIIB and has a role in attaching it to the upstream control region (25). A more remote alternative is that it is the host genomic DNA upstream of the transposition that has an inhibitory influence.

In conclusion the Cp1-RNA gene shows a unique combination of pol III control elements. Although several tRNA genes have upstream regulatory motifs, some of which are AT rich (26–31) there is usually no TATA box, and the upstream elements are modulatory rather than essential. An exception is the *Xenopus laevis* $tRNA^{(Ser)Sec}$ gene with an essential upstream TATA-box in addition to other regulatory sequences (32,33). In that case the A box is, however, non-functional. The promoter of the U6 gene in *S.cerevisiae* (34) is similar to the Cp1-RNA gene in having a TATA box and an internal split promoter although there the B box has an unusual distal position. The U6 TATA box is, furthermore, modulatory rather than essential. In summary, thus, the Cp1-RNA

gene appears to be unique with its tripartite promoter in which all three components are nearly essential, at least *in vitro*.

Our results raise the following questions. First, if Cp1-RNA cannot have a role in transposition, what then promotes the mobility of Cp1? Secondly, is there any evidence that Cp1-RNA is functional at all?

Cp1 occurs both in centromeres and outside of them. In the former localization Cp1 is integrated between centromeric 155 bp tandem repeats such that each integration site is followed by a short partial duplication of the 155 bp sequence situated upstream of the element. Even if this could be a target site duplication, it is not necessarily of recent origin. This is because 155 bp sequences might be exposed to parallel evolution and the duplication thus conserved irrespective of age. It is possible, on the other hand, that extracentromeric elements transposed more recently, since they are surrounded by perfect target site duplications. Extracentromeric elements are highly polymorphic and may be functionless like the element cloned in pCt2.2. Furthermore, there are also large numbers of highly degenerate Cp1 fragments in the extracentromeric genome (9). Therefore it appears unlikely that Cp1 could have an important role in extracentromeric loci.

It is possible that the transposition of Cp1 has originally occurred into the centromeres from a site in the pre-ribosomal transcription unit judging by the nature of the DNA at the 5′-end of each of its sequence modules. Also extracentromeric elements might have this origin but another possibility is that they are continuously acquired by transpositions from intracentromeric units.

Cp1 might have a physiological role, but not necessarily in extracentromeric positions. This view is supported by the functional control elements in the Cp1-RNA gene and because a new well defined secondary structure has been obtained during evolution from a tRNA gene. It is also of considerable interest that the *S.pombe* centromeres contain clustered tRNA genes reported to be functionally active (11,12).

Cp1 may show interesting similarities with the *S.cerevisiae* centromere. When CEN 4 DNA from this organism was inserted into negatively supercoiled plasmid, it obtained properties suggesting that DNA unwinds to form RNA-like foldback structures, which may also exist *in vivo* (35). When individual strands from intracentromeric Cp1 are allowed to form secondary structures by the M-fold program V.8, they give highly base paired, relatively symmetrical structures with remarkably low free energy (our unpublished data). The transcription unit in the B module is at the right end not only of this module but of entire Cp1 elements (the transcriptionally inactive element in clone no. 6, Fig. 5, being an exception). This transcription is directed away from the element. Transcription might create negative superhelical tension in the region upstream of the transcription unit (36,37), i.e. within the larger part of Cp1. This in turn should favour transitions to single-strand structures. Cp1 is inserted into the AT-rich conserved region of the 155 bp repeat (38). These tandem repeats might function as spacers, periodically separating functionally important Cp1 insertions. AT-rich regions could facilitate structural transitions (39). We might speculate that structural–functional change in the centromere could be regulated in this way by control of pol III transcription.

ACKNOWLEDGEMENTS

The investigation was supported by grants from the Swedish Cancer Society, the Philip-Sörensen Foundation and the Crafoord Foundation. We are indebted to Joakim Galli and Lars Wieslander, Karolinska Institutet, Stockholm for the sp12 DNA clone.

REFERENCES

1 Ullu,E. and Tschudi,C. (1984) *Nature*, **312**, 171–172.
2 Rinehart,F.P., Ritch,T.G., Deininger,P.L. and Schmid,C.W. (1981) *Biochemistry*, **20**, 3003–3010.
3 Okada,N. (1991) *Curr. Opin. Genet. Dev.*, **1**, 498–504.
4 Lawrence,C.B., McDonnell,D.P. and Ramsey,W.J. (1985) *Nucleic Acids Res.*, **13**, 4239–4252.
5 Daniels,G.R. and Deininger,P.L. (1985) *Nature*, **317**, 819–822.
6 Coltman,D.W. and Wright,J.M. (1994) *Nucleic Acids Res.*, **22**, 2726–2730.
7 Kido,Y., Aono,M., Yamaki,T., Matsumoto,K., Murata,S., Saneyoshi,M. and Okada,N. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 2326–2330.
8 Deragon,J.M., Landry,B.S., Pélissier,T., Tutois,S., Tourmente,S. and Picard,G. (1994) *J. Mol. Evol.*, **39**, 378–386.
9 He,H., Rovira,C., Recco-Pimentel,S., Liao,C. and Edström, J-E. (1995) *J. Mol. Biol.*, **245**, 34–42.
10 Rovira,C., Beermann,W. and Edström, J-E. (1993) *Nucleic Acids Res.*, **21**, 1775–1781.
11 Kuhn,R.M., Clarke,L. and Carbon,J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1306–1310.
12 Takahashi,K., Murakami,S., Chikashige,Y., Niwa,O. and Yanagida,M. (1991) *J. Mol. Biol.*, **218**, 13–18.
13 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Second edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14 Galli,J. and Wieslander,L. (1994) *J. Mol. Evol.*, **38**, 482–488.
15 Frohman,M.A., Dush,M.K. and Martin,G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9002.
16 Wyss,C. (1982) *Exp. Cell Res.*, **139**, 309–319.
17 Dingermann,T., Sharp,S., Appel,B., DeFranco,D., Mount,S., Heiermann,R., Pongs,O. and Söll,D. (1981) *Nucleic Acids Res.*, **9**, 3907–3918.
18 Sprinzl,M., Hartmann,T., Meissner,F., Moll,J. and Vorderwulbecke,T. (1987) *Nucleic Acids Res.*, **15** (suppl.), 53–188.
19 Sklar,V.E.F., Jaehning,J.A., Gage,L.P. and Roeder,R.G. (1976) *J. Biol. Chem.*, **251**, 3794–3800.
20 Roiha,H., Miller,J.R., Woods,L.C. and Glover,D.M. (1981) *Nature*, **290**, 749–753.
21 Deininger,P.L. (1989) In Berg,D.E. and Howe,M.M. (eds) *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 619–636.
22 Garber,R.L. and Gage,L.P. (1979) *Cell*, **18**, 817–828.
23 Wilson,E.T., Larson,D., Young,L.S. and Sprague,K.U. (1985) *J. Mol. Biol.*, **183**, 153–163.
24 Wilson,E.T., Condliffe,D.P. and Sprague,K.U. (1988) *Mol. Cell. Biol.*, **8**, 624–631.
25 Geiduschek,E.P. and Kassavetis,G.A. (1995) *Curr. Opin. Cell Biol.*, **7**, 344–351.
26 Sprague,K.U., Larson,D. and Morton,D. (1980) *Cell*, **22**, 171–178.
27 DeFranco,D., Schmidt,O. and Söll,D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3365–3368.
28 Dingermann,T., Burke,D.J., Sharp,S., Schaack,J. and Söll,D. (1982) *J. Biol. Chem.*, **257**, 14738–14744.
29 Larson,D., Bradford-Wilcox,J., Young,L.S. and Sprague,K.U. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3416–3420.
30 Young,L.S., Takahashi,N. and Sprague,K.U. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 374–378.
31 Taneja,R., Gopalkrishnan,R. and Gopinathan,K.P. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1070–1074.
32 Lee,B.J., Kang,S.G. and Hatfield,D. (1989) *J. Biol. Chem.*, **264**, 9696–9702.
33 Carbon,P. and Krol,A. (1991) *EMBO J.*, **10**, 599–606.
34 Eschenlauer,J.B., Kaiser,M.W., Gerlach,V.L and Brow,D.A. (1993) *Mol. Cell. Biol.*, **13**, 3015–3026.
35 Tal,M., Shimron,F. and Yagil,G. (1994) *J. Mol. Biol.*, **243**, 179–189.
36 Liu,L.F. and Wang,J.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7024–7027.
37 Freeman,L.A. and Garrard,W.T. (1992) *Crit. Rev. Euk. Gene Exp.*, **2**, 165–209.
38 Liao,C., Rovira,C. and Edström,J.-E. (1994) *J. Mol. Evol.*, **39**, 112–114.
39 Sullivan,K.M. and Lilley,D.M.J. (1986) *Cell*, **47**, 817–827.