Isolation of genomic DNA encoding transcription factor TFIID from *Acanthamoeba castellanii*: characterization of the promoter

Jie Min Wong, Feng Liu and Erik Bateman*

Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT 05405-0068, USA

Received June 22, 1992; Revised and Accepted August 21, 1992

GenBank accession no. M93340

ABSTRACT

We have isolated a genomic clone encoding *Acanthamoeba castellanii* TFIID. The clone contains the entire TFIID gene, 300 bp of 5' promoter sequences and several hundred base pairs of 3' non-coding sequence. The coding region is interrupted by two short introns, but is otherwise identical to *Acanthamoeba* TFIID cDNA. Comparisons between forty four *Acanthamoeba* intron 5' and 3' boundaries suggest a 5' splice site consensus of GTACG(T/C) and a 3' consensus of CAG. We determined the position of the transcription initiation site used *in vivo*, and show that the same site is used *in vitro* by homologous nuclear extracts. Deletion analysis of the promoter region shows that the minimal promoter required for efficient expression *in vitro* is located between –97 and +4 relative to the transcription start site. Three regions within the promoter are important for transcription *in vitro*; sequences between –97 and –35, the TATAAA box and the initiation region. The initiation region is dispensable but appears to position the transcription start site relative to the TATAAA box. The TATAAA box is absolutely required for transcription initiation whereas the upstream region stimulates transcription approximately five-fold.

INTRODUCTION

TFIID is a cellular protein that is required for transcription of genes transcribed by RNA polymerases I, II and III (1–4). TFIID binds to the sequence TATAAA found upstream of many eukaryotic genes but will also bind to other sequences that are not always recognizable as TATA elements (5–9). TFIID is one of the general transcription factors and is essential for pre-initiation complex formation, that is, to establish a complex between DNA, TFIID, RNA polymerase II, TFIIB, E and F. The pre-initiation complex is relatively stable and transcriptionally competent. TFIID remains bound to the promoter after one or more rounds of transcription (10–12), whereas polymerase and the other factors may dissociate. Within the cell, TFIID is complexed with several other proteins (TAFs) that confer specificity and activity (4,13).

The role of TFIID, and perhaps other general factors involved in transcription initiation, has been conserved in eukaryotes. For example, a yeast or *Acanthamoeba* TFIID-containing fraction can function in human extracts (14–16). This conservation of function, albeit imperfect, has also been noted for other components of the transcription apparatus (8).

Within the pre-initiation complex, TFIID may form protein-protein contacts with other factors, and perhaps RNA polymerase II. For example, it has been suggested that a contact between TFIID and polymerase II is mediated by another factor, TFIIB (17). This suggestion is supported by experiments which substitute yeast TFIID for HeLa TFIID and show that the distance between the TATAAA box and the transcription start site is determined by a factor other than TFIID (14,15). This distance might also be dictated by interactions with polymerase. In either event, it is likely that TFIID has two or more functional domains; one involved in DNA interactions and others involved in protein-protein interactions.

Because of its central role in eukaryotic transcription, it might be expected that TFIID levels are tightly regulated to avoid non-specific DNA interactions or unbalanced interactions with other proteins. In *Acanthamoeba*, we have found that TFIID mRNA is rather rare, suggesting that TFIID genes are constitutively regulated. However, nothing is known about the regulatory or basal components of the TFIID gene promoter. Indeed, only TFIID genomic clones from *S.pombe* and *S.cerevisiae* have been previously isolated (18–20).

Acanthamoeba is a small, free-living amoeba that is commonly found in air, soil, fresh water and in the human nasopharyngeal tract as part of the normal flora (21–23). It is an opportunistic pathogen associated with *Acanthamoeba* keratitis and occasionally can cause granulomatous encephalitis or related conditions in immunosuppressed individuals (21). The lifecycle of *Acanthamoeba* is characterized by its ability to differentiate into a dormant cyst when exposed to adverse conditions such as starvation, or in response to a variety of chemical agents. Lab cultures of these amoeba can be grown axenically and differentiation is easily manipulated. The process of differentiation culminates in the formation of a cellulose cell wall and a cessation of cell growth, but the pathway can be easily reversed at any

* To whom correspondence should be addressed

A

```

      10      20      30      40      50      60      70      80      90     100     110
      *      *      *      *      *      *      *      *      *      *      *
GGATCCCATC TTCGACCAGC TCGCGCGTAG ATCAGGCCCTT CCATGACGAA GTCTTCCTCG AACCATATCG GCTCTTCATC GACGTTGCCC TTCACCGTGC CCACCGCGAA

      120     130     140     150     160     170     180     190     200     210     220
      *      *      *      *      *      *      *      *      *      *      *
TAACGCAAGG CCGCACAGGA CCAGCGCCAG GCAAAGGGCG AGCCTCATCT TCGTGAAGCT GAAGCTGGGT TGTGCGGGA AACGACGCTT TGCAACAAGC TGAGAAAAAA

      230     240     250     260     270     280     290     300     310     320     330
      *      *      *      *      *      *      *      *      *      *      *
CCAGGATCGG CAAGGAAGGA TTTCAACGG AAACGTTGGC ATCACCGGGT ATAAAAAGGG GCCAATTTTT TTGTTGATTI GTTGCAGGAA TTCTTGCTTT CGGCATCGAA

      340     350     360     370     380     390     400     410     420
      *      *      *      *      *      *      *      *      *
TTCAAGGGAG AAGGAGTCGA TTCACACATA CAACAAG ATG AGC GGG ATT ACT CTA CCG AGC TTG ACG AAT GTC CTT CAG AGC GCG GGC ATG GCG GTG
      Met Ser Gly Ile Thr Leu Pro Ser Leu Thr Asn Val Leu Gln Ser Ala Gly Met Ala Val

      430     440     450     460     470     480     490     500     510
      *      *      *      *      *      *      *      *      *
CAC GGC CAC CCG TCG GCA CCC GGG AGC ACC CAG CTA CCC CCA CTT CAT CAA TTG AAC ATC TCG TCA CAG CCC TCC TCG CAG CCT CCT CAG
      His Gly His Pro Ser Ala Pro Gly Ser Thr Gln Leu Pro Pro Leu His Gln Leu Asn Ile Ser Ser Gln Pro Ser Ser Gln Pro Pro Gln

      520     530     540     550     560     570     580     590     600
      *      *      *      *      *      *      *      *      *
CCT TCA CTG CAG TAC TCT GAG CCC GCG CAA TCG ACT GCT GCC AGC GAC GAT ATG GAC AGC GAC GTG GAT CGC ACC AAG CAC CCG TCG GGC
      Pro Ser Leu Gln Tyr Ser Glu Pro Ala Gln Ser Thr Ala Ala Ser Asp Asp Met Asp Ser Asp Val Asp Arg Thr Lys His Pro Ser Gly

      610     620     630     640     650     660     670     680     690     700     710
      *      *      *      *      *      *      *      *      *      *      *
ATT GTC CCT ACT CTG CA GTACGT AGTCATTGAG TTTCTGGGGG GGGTCAAAT TCAAACCTCG TTACTCAGAT TGTCTCTTC TTCAAAGTCC ATTCTCATGA
      Ile Val Pro Thr Leu Gln

      720     730     740     750     760     770     780     790     800
      *      *      *      *      *      *      *      *      *
TGTGTGATGAT TCTGCAACAG A AAC ATC GTC TCC ACG GTG AAT TTG GGG TGC AAG CTC GAC CTC AAG AAC ATC GCA CTG CAT GCA CGT AAC GCC
      Asn Ile Val Ser Thr Val Asn Leu Gly Cys Lys Leu Asp Leu Lys Asn Ile Ala Leu His Ala Arg Asn Ala

      810     820     830     840     850     860     870     880     890
      *      *      *      *      *      *      *      *      *
GAG TAC AAC CCG AAG CGT TTT GCT GCC GTC ATC ATG AGA ATT CGC GAG CCG AAG ACG ACC GCG CTC ATC TTT GCA TCG GGC AAG ATG GTG
      Glu Tyr Asn Pro Lys Arg Phe Ala Ala Val Ile Met Arg Ile Arg Glu Pro Lys Thr Thr Ala Leu Ile Phe Ala Ser Gly Lys Met Val

      900     910     920     930     940     950     960     970     980
      *      *      *      *      *      *      *      *      *
TGT ACG GGC GCC AAG AGC GAA GAG GCA TCT CGT CTG GCT GCC AGG AAG TAC GCT CGC ATC ATC CAG AAG CTC GGA TTC GCC GCC AAG TTC
      Cys Thr Gly Ala Lys Ser Glu Glu Ala Ser Arg Leu Ala Ala Arg Lys Tyr Ala Arg Ile Ile Gln Lys Leu Gly Phe Ala Ala Lys Phe

      990     1000     1010     1020     1030     1040     1050     1060     1070
      *      *      *      *      *      *      *      *      *
CTC GAT TTC AAG ATT CAG AAC ATC GTC GGC TCG TGC GAT GTG CGA TTC CCC ATT CGT CTC GAG GGT CTC GCC TTT GCC CAC AAC CAC TAC
      Leu Asp Phe Lys Ile Gln Asn Ile Val Gly Ser Cys Asp Val Arg Phe Pro Ile Arg Leu Glu Gly Leu Ala Phe Ala His Asn His Tyr

      1080     1090     1100     1110     1120     1130     1140     1150     1160     1170     1180
      *      *      *      *      *      *      *      *      *      *      *
TGC AGC G TACGTCTCTC ATCCCCCTCC CCCGCTCTCT ACTTGCTCAT CTGCCAGAGG CAGCTCGCCT AATCTCCATG CTGTGCTGTG CATGCACACC ACGTGAG TAC
      Cys Ser Tyr

      1190     1200     1210     1220     1230     1240     1250     1260     1270
      *      *      *      *      *      *      *      *      *
GAG CCA GAG CTG TTC CCG GGT CTC ATC TAC CGC ATG GTG CAG CCC AAG ATC GTG CTG CTC ATC TTC GTG TCG GGA AAG ATC GTG CTC ACG
      Glu Pro Glu Leu Phe Pro Gly Leu Ile Tyr Arg Met Val Gln Pro Lys Ile Val Leu Leu Ile Phe Val Ser Gly Lys Ile Val Leu Thr

      1280     1290     1300     1310     1320     1330     1340     1350     1360
      *      *      *      *      *      *      *      *      *
GGC GCC AAG GTG CGA GAG GAG ATC TAC GAG GCC TTC GAG AAC ATC TAC CCG GTG TTG ACC GAG TAC AAG AAG ACC TAA G CCATCCACCT
      Gly Ala Lys Val Arg Glu Glu Ile Tyr Glu Ala Phe Glu Asn Ile Tyr Pro Val Leu Thr Glu Tyr Lys Lys Thr End

      1370     1380     1390     1400     1410     1420     1430     1440     1450     1460     1470
      *      *      *      *      *      *      *      *      *      *      *
CACCTGTACA AACTACCAAC CCCCCCAACC ATCGCCGCGC TCACCTGGAT CTACTCCCTC CCTCCCGCCA ACCAACTACT CAGCCCAACG CGGTGTGCCA CCGGCGCGTG

      1480     1490     1500     1510     1520     1530     1540     1550     1560     1570     1580
      *      *      *      *      *      *      *      *      *      *      *
CACGAATGTA CAATCACCAG GGGGCGTTGA GAACAAAAAC ATACACCGGT TTTTITAACA ACATGTGTG TGTCTATGTT ATGTTGGTGA AGACACGTCA TGTGGGCTAT

      1590     1600     1610     1620     1630     1640     1650     1660     1670     1680     1690
      *      *      *      *      *      *      *      *      *      *      *
GTGGTGTGTA TGAGGGTCCA GAGAGGCTTG TAGGCTGCTT GTTTGTCAAG TCTACAAAAG AGTGCTAACA AAATTG6CCA ATCGTCAGTC TGGTTCGAAC CCCGTCTGG

      1700     1710     1720     1730     1740     1750     1760     1770
      *      *      *      *      *      *      *      *
TGATTGGCTT GCGACAGAGA GGAGGAGAAG AGGGGAGTA CAGGTGTAGG TAGGGTGTAG AGCACATCAT CAGCTGATCA GTT

```

Figure 1. *Acanthamoeba* genomic TFIID gene. **A**, DNA sequence of the *Acanthamoeba* TFIID genomic clone and flanking regions. The putative TATA box is in bold; the transcription start site is bold and underlined; intron consensus sequences are underlined, and putative poly(A) signals are double underlined. **B**, Diagram of the *Acanthamoeba* genomic TFIID gene emphasizing promoter and intron positions. The position of the primer used in start site determination is indicated by a short arrow. The Genbank Accession number for the TFIID sequence shown here is M93340.

stage (M.R.Paule, personal communication), providing a convenient lab system to study the underlying cellular events.

We recently reported that the *Acanthamoeba* TFIID gene is unique and encodes a single mRNA and we determined the complete cDNA sequence (16). The TFIID gene is thus the first protein-coding gene from *Acanthamoeba* for which complete cDNA and genomic DNA sequences are available, permitting an analysis of its transcription start site, promoter elements, intron splicing and polyadenylation signals.

Acanthamoeba promoters for RNA polymerase II have not previously been studied. Promoter elements from other eukaryotes fall into two broad categories; regulatory elements and basal elements. Regulatory elements, for example sequences that bind Sp1 or CTF, are members of a large family of DNA motifs that interact with an equally large number of regulatory proteins (see 6, 24, 25). The so-called basal elements seem to comprise only a TATAAA sequence and an initiator region (see 8, 26). Not all promoters contain a discernible TATAAA motif, and not all promoters have a requirement for an initiator, but one or both of these sequences are present in most or all genes studied (see 27, 28). The initiator, first described for the TdT gene in mammals is a poorly defined sequence that is characteristically pyrimidine-rich. However, because of the lack of a clear consensus, the presence (or absence) of such an element must be determined empirically.

In this report we describe the isolation and characterization of a genomic DNA clone encoding *Acanthamoeba* TFIID. We also characterized the *Acanthamoeba* TFIID gene promoter using a series of deletion mutants assayed in a homologous *in vitro* transcription system (29). These experiments were performed in order to identify promoter elements for this gene in particular, but also to provide an experimental framework for examining other *Acanthamoeba* genes transcribed by RNA polymerase II. The *Acanthamoeba* TFIID gene promoter is located between -97 and +4 relative to the transcription start site. The promoter is dependent upon a TATAAA box, but activity can be stimulated five-fold by one or more upstream sequences. The initiation region is dispensable but helps to position the transcription start site.

MATERIALS AND METHODS

Isolation of the TFIID gene

An *Acanthamoeba* genomic DNA library in λ EMBL3A was constructed using standard methods (30) and those recommended by the suppliers (Stratagene). Briefly, high molecular weight

Acanthamoeba DNA (31) was partially digested with Sau3A, and DNA fragments within the size range of 12–20 kb were purified by agarose gel electrophoresis and electroelution. These fragments were ligated to BamHI and EcoRI digested λ DNA. Ligated DNA was packaged *in vitro* and used to infect *E. coli* LE392 (P2 lysogen) for preparation of the primary library. Phage were then plated on *E. coli* LE392, and the library was screened for the TFIID gene using a labeled TFIID cDNA-derived probe as described elsewhere (16). The TFIID gene within one positive clone was mapped to a 6 kb BamHI fragment, which was subcloned into pSK (-) (32). The TFIID portion was sequenced using internal primers derived from the sequence of a TFIID cDNA clone. In some cases, deletions of portions of the TFIID genomic clone were sequenced using pSK vector primers (32).

Primer extension analysis of TFIID mRNA

Primer extension analysis of *Acanthamoeba* TFIID mRNA was performed as described elsewhere (30). 2 μ g of poly(A)⁺ RNA was used in each reaction. The extension reactions were done using AMV reverse transcriptase (USB) and [³²P] end-labeled primer.

Construction of Deletion Mutants

For promoter deletion from the 5' end, plasmid pAgTFIIDp, which contains a 530 bp PstI fragment at the PstI site in plasmid pSK-, was digested with KpnI and HincII. The linearized plasmid was then treated with exonuclease III as described by the suppliers (Promega and Stratagene) with the following modifications. Instead of using S1 nuclease, 4 μ l of mungbean nuclease (BRL) was used to trim the single stranded DNA region left by exonuclease III digestion. After addition of stop buffer, the reaction mixture was extracted with phenol and chloroform. DNA was ethanol precipitated, and redissolved in 20 μ l ligation buffer and ligated with T4 DNA ligase. Several deletion clones were selected and sequenced. Eight of these were used for the analysis of figure 3.

For deletion from the 3' end, plasmid pAGTFIIDp was digested with BamHI and SacI and was treated with exonuclease III and re-ligated as described above. Two additional 3' deletions; 3'dl-19a and 3'dl-19b were prepared by subcloning a HaeIII fragment from the TFIID promoter region into the SmaI site of pSK(-). Inserts in both possible orientations were obtained, as shown in Figure 4. All clones were sequenced prior to transcription analysis.

Figure 2. Determination of the *in vivo* and *in vitro* transcription start sites for TFIID mRNA. mRNA from *Acanthamoeba* or *in vitro* transcription products of the TFIID gene were analyzed using primer extension as described in materials and methods. The sequence of the start site region is shown to the right of the autoradiograms. Lanes marked A, C, G or T are from sequencing reactions; lanes marked R are the primer extension reactions.

In vitro transcription and primer extension

In vitro transcription was performed in a total volume of 50 μ l as described elsewhere (29). DNA fragments used in transcription reactions were prepared by digesting plasmids containing deleted TFIID genomic subclones with PvuII. Appropriate fragments were purified from 0.8% agarose gel. The amount of each DNA fragment was quantified by end-labeling with [α - 32 P] dCTP. Amounts of the 5' and 3' deletion-containing fragments were also estimated by spotting labeling reactions onto a DEAE 81 discs (Whatman), washing in phosphate as described elsewhere (33), and the counts determined by scintillation counting. In the 5' deletion series, where we anticipated quantitative changes in template activity, labeling reactions were also run on a 6% acrylamide gel and scanned on densitometer in order to quantify the amount of DNA more precisely.

About 100 fmol of DNA template was used in each transcription reaction. After incubating at 30°C for 60 min, transcription reactions were stopped and extracted once with phenol/chloroform and once with chloroform (33). Nucleic acids were precipitated with ethanol and redissolved in 30 μ l hybridization buffer containing 80% formamide, 0.4 M NaCl, 1 mM EDTA, and 40 mM PIPES (pH 6.6). 1–2 μ l of end labeled primer (50,000–100,000cpm) was added, mixed, and incubated at 30°C overnight. After annealing, DNA was ethanol precipitated, washed with 70% ethanol, and air-dried. Primer extension reactions were performed in a total volume of 25 μ l containing 50 mM Tris (pH 8.3), 50 mM KCl, 5 mM MgCl₂, 1 mM EDTA, 3.5 μ l of 2.5mM dNTP mix, 1mM DTT, 5U human placenta RNase inhibitor (BRL) and 5U AMV reverse transcriptase (USB). Reactions were at 42°C for 90 minutes and were subsequently processed by addition of 25 μ l H₂O and 50 μ l stop solution (0.6M potassium acetate [pH 5.2], 0.2% SDS, 200 μ g yeast tRNA/ml). Nucleic acids were precipitated with ethanol, dissolved in 6 μ l formamide loading buffer, and run on a 6% denaturing gel.

RESULTS

The genomic copy of the *Acanthamoeba* TFIID gene

We constructed a genomic library of *Acanthamoeba* genomic DNA in λ EMBL3. The library was screened using a probe derived from *Acanthamoeba* TFIID cDNA (16). Of several positive clones, one was mapped with various restriction enzymes and an approximately 6 kb fragment containing the entire TFIID gene was subcloned into the vector pSK(–) (32). The DNA sequence of the *Acanthamoeba* genomic TFIID gene as well as a summary diagram are shown in Fig. 1. The parental clone contains an additional 4000 bp downstream of the TFIID gene, which we have not sequenced.

With the exception of two short introns, the coding region of the TFIID gene is identical to a cDNA clone described previously (16). The cDNA encodes a functional protein that is active in transcription activation and DNA binding (16). The genomic sequence confirms that the cDNA clone is full-length with respect to coding sequences. The sequence colinearity allows us to identify the positions of intron boundaries and the polyadenylation site.

The introns within the TFIID gene are punctuated by sequences at the 5' and 3' splice sites resembling those found in other *Acanthamoeba* introns, as well as other eukaryotes. Interestingly, both introns interrupt the directly repeated amino acids found in the conserved C-180 domain of TFIID, and the second intron occurs in the same position as the equivalent *S.pombe* intron (20). We undertook a more detailed examination of DNA sequences within the TFIID introns by comparing them with 42 other *Acanthamoeba* introns found in the actin 1, myosin II heavy chain and myosin 1B genes (34–36). This comparison reveals a 5' consensus of G₄₄T₄₄A₃₉C₃₆G₃₃(T₂₀/C₁₇) and a 3' consensus of C₃₂A₄₄G₄₄. The 3' intron boundaries show additional preferences for certain nucleotides (not shown), but these preferences are less compelling than those seen at the 5' intron boundary. Both the 5' and 3' intron boundaries of the TFIID

Figure 3. Transcription from 5' deletions to the *Acanthamoeba* TFIID gene promoter. **A**, The endpoints of deletions used in each reaction are shown in the diagram, as are their activities and the corresponding lane numbers. The position of the primer used in RNA analysis is indicated. **B**, autoradiogram of RNA products analyzed by primer extension.

gene resemble the consensus splice sequences from *S.pombe*, but are nonetheless distinct. The two introns within the TFIID gene match the proposed splice consensus exactly.

At the 3' end of the gene, about 20 bp upstream of the polyadenylation site there is a putative polyadenylation signal (AACAAAA), that resembles a sequence found in several cDNA clones of *Acanthamoeba* genes (E.B. and D.Morgan, unpublished). The polyadenylation site is located 182 bp downstream from the stop codon in the TFIID gene (Fig. 1).

At the 5' end of the gene, within the putative promoter, there is a perfect TATAAA box 30 bp upstream of the transcription

start site (see below). Ten bp 3' of the TATAAA box, there is a perfect CCAAT box (Fig. 1). There are also two more CCAAT-like sequences, ten and thirty bp downstream of the transcription start site. In addition, there is an imperfect heat shock element at -140, relative to the transcription start site (7/9 match). However, we did not identify any other previously characterized promoter elements in the 300 bp of 5' sequence, nor is there significant resemblance of the region around the transcription start site that resembles an initiator element (27,28). The functional significance of these putative promoter elements was tested directly, as described below.

Determination of the TFIID *in vivo* transcription start site

The transcription start site used within the TFIID gene was determined using primer extension analysis of poly (A)⁺ RNA. This identified a major transcript originating from a T residue 30 bp downstream of the TATA box (Figure 2). An additional minor start site, not visible in Figure 2, was found three bases further downstream, also at a T residue. These experiments have been repeated several times using different mRNA preparations. The same start site was found using S1 nuclease mapping (results not shown).

Transcription *in vitro* from the *Acanthamoeba* TFIID gene promoter

We next tested the ability of the genomic copy of the *Acanthamoeba* TFIID gene to function in homologous nuclear extracts. We recently showed that *Acanthamoeba* nuclear extracts can accurately initiate transcription from the adenovirus major late promoter (29), but have not previously characterized transcription from a homologous *Acanthamoeba* gene. In the present experiment, a TFIID gene fragment containing the promoter region was transcribed by an *Acanthamoeba* nuclear extract and transcripts were analyzed by primer extension (Figure 2B). The major 5' end of the *in vitro* transcripts maps to the T residue found at the 5' end of mRNA, although one or two shorter products can also be seen. The transcription products are sensitive to inhibition by α -amanitin, and therefore transcribed by RNA polymerase II (not shown). We infer that the homologous TFIID gene promoter is efficiently utilized by our extracts.

Deletion analysis of the *Acanthamoeba* TFIID gene promoter

In order to identify which sequences within the *Acanthamoeba* TFIID gene contribute to its *in vitro* promoter activity, we constructed a series of 5' and 3' deletion mutations within the promoter region (Figures 3 and 4). The transcription activity of the deletions was assayed using *in vitro* transcription and primer extension. The TFIID promoter can be deleted from the 5' direction as far as -97 without effect. However, deletion to -35 reduces transcription activity by about five-fold (Figure 3, lane 5). This difference in activity between the -97 and -35 deletions was observed in four experiments using two different DNA preparations. The data shown was quantified by densitometry.

Further deletion of the promoter to -28, removes the first two bases of the TATAAA box, and completely abolishes transcription (Figure 3, lane 6). This result suggests that bases between -35 and -28 are of critical importance, and that the 5' -28 mutant is most likely inactive due to disruption of the TATA box. All deletions past -28 were also inactive (Figure 3, lanes 7 and 8).

Deletions to the TFIID gene promoter from the 3' side were made and assayed to determine the contribution of the initiation region to activity overall and to start site selection. The 3' side of the promoter region can be deleted as far as +4 without affecting the positioning of transcription initiation (Fig 4, lanes 1-7). However, there is a trend towards an increased level of transcription as the deletions approach the start site (see also below). While the levels of transcripts produced from the +93 and +59 deletions is lower than the parental template in the autoradiogram shown, these differences are not significant, based on several repetitions of the experiment.

Removal of the natural transcription start site in the 3' -5 deletion, however, results in efficient initiation at a new position

within the vector as well as at more minor sites corresponding spatially to the correct start site (Figure 4, lane 8). The major new transcription start is 35 bp from the TATAAA box, rather than 30 bp. Examination of the vector sequences that replaced the TFIID promoter sequences in the 3' mutant shows that the new initiation site is within the sequence 5'-TTTTGTTCCCTT-3', the underlined part of which is identical to a portion of the TFIID gene start site sequence. In a sense, this deletion represents a mutant that pushes apart portions of the start region by 5 base pairs.

Some initiation from the 3' -5 deletion occurs at the correct (30 bp) distance from the TATAAA box, suggesting that the sequence of the start site is not the sole determinant of initiation position (Figure 4, lane 8). We therefore constructed 3' deletions that completely remove the start site and place the 3' end in two different plasmid contexts. In both cases, initiation still occurs efficiently from the vector sequences, even though they are quite distinct in character from the wild type initiation region (Fig. 4, lanes 10 and 11). However, the initiation sites in these mutants are distributed among several nucleotides, rather than two major sites as seen in the wild type template. The rather evenly spaced start sites in 3'dl-19a span seven nucleotides, and those of 3'dl-19b span eleven nucleotides.

Although 3' mutants that remove the start site are fully active, deletion of the promoter region to -52 produces an inactive template (Fig. 4, lane 9). This result was expected since the 3' -52 deletion mutant lacks the TATA box and flanking sequences found to be important in assays of the 5' deletion mutants (above).

We infer from these experiments that in the *Acanthamoeba* TFIID gene the initiation region does not greatly affect the efficiency of transcription, but instead exerts a positioning effect. It is noteworthy that loss of the CCAAT motif (Figure 1) as in the 3' -19 deletion does not have a major effect on transcription efficiency. Indeed, in our experiments these mutants are consistently transcribed more efficiently than undeleted templates suggesting that sequences near the initiation site may exert a negative effect. Further experiments will be needed to establish the magnitude of such effects and to locate the relevant sequences.

DISCUSSION

Isolation of the genomic copy of the *Acanthamoeba* TFIID gene has permitted determination of its transcription start site, analysis of its promoter by *in vitro* transcription, and determination of consensus sequences for intron boundaries. It is the first protein-coding gene from *Acanthamoeba* for which complete cDNA and genomic DNA sequences are available, and the first *Acanthamoeba* promoter for an RNA polymerase II gene to be characterized. The gene itself, however, is unremarkable, in that coding regions match those from the cDNA exactly, with the exception of the two introns. The consensus sequences for intron splicing appear to be unique to *Acanthamoeba* but most closely resemble those from *S.pombe* (37).

Not all *Acanthamoeba* genes share sequence features described for the TFIID gene, such as the TATA box. While two actin genes have TATA boxes, none of the myosin genes sequenced to date have sequences resembling TATAAA, suggesting that myosin genes may instead operate through an initiator element (28). However, as pointed out elsewhere, the factor TFIID may bind sequences that have no superficial resemblance to a TATAAA element, partly as a consequence of its DNA-binding mechanism (38,39). We have obtained transcription *in vitro* from

Figure 4. Transcription from 3' deletions to the *Acanthamoeba* TFIID gene promoter. **A**, The endpoints of deletions used in each reaction are shown in the diagram, as are the corresponding lane numbers. 3' deletions to -19 were placed in two different plasmid contexts. The positions of primers used for primer extension are denoted by small arrows. **B**, autoradiogram of RNA products analyzed by primer extension.

the myosin heavy chain gene promoter and found that it can be deleted to as far as -15 with retention of full activity (F.L. and E.B. unpublished). Comparisons among the various *Acanthamoeba* protein coding genes did not reveal any additional sequence elements that could be regarded as a potential promoter or initiator element.

The TFIID gene is transcribed *in vitro* by homologous extracts roughly as efficiently as the adenovirus major late promoter (29). Comparisons between the adenovirus major late promoter and the *Acanthamoeba* TFIID gene promoter show some similarity in overall composition, particularly the G-rich regions that flank the TATA box in both promoters. These results, however present an enigma in that TFIID mRNA is quite rare in *Acanthamoeba*. It seems likely that the *Acanthamoeba* TFIID gene will be subject to negative regulation or to rapid mRNA turnover within the cell in order to prevent over-expression of TFIID. Alternatively, TFIID mRNA could be efficiently and rapidly transcribed *in vivo* during a particular phase of the cell cycle. Further characterization of mRNA synthesis and turnover will be required to discriminate among these possibilities.

Transcription initiation within the TFIID gene at a T residue is somewhat unusual, in that most eukaryotic genes begin transcription at an A or, less frequently, a G residue (5). There are however, some precedents, including the *Acanthamoeba* actin 1, myosin IB and myosin IL genes (34,36,40).

We undertook the promoter analysis to determine whether the TATA element and initiation regions are, in fact, functional or important in *Acanthamoeba*. In addition we anticipated the possible identification of a negatively-acting element. The results suggest that the TATA element is indeed a critical promoter element since removal of the first two bases of the TATA box completely inactivates the promoter.

Somewhat surprisingly, our results suggest the presence of an upstream stimulatory element between -97 and -35 that has an approximately 5-fold effect on initiation. We do not yet know what sequence within this region is important for efficient initiation, but it should be possible to identify the putative factor binding site using DNase footprinting combined with finer deletion mapping. It could be argued that the region between -97 and -35 contributes to binding of TFIID to the promoter, and does not contain an additional *cis* acting element. In preliminary studies, we have found that the 5' -35 deletion can bind recombinant TFIID, whereas the -28 deletion cannot, indicating that the -35 deletion contains sufficient sequence information for TFIID recognition, and that its reduced activity is due to removal of sequences that affect a distinct protein-DNA interaction. We have also obtained clear footprints within the upstream region using partly fractionated nuclear extracts (E.B. and F.L., unpublished data).

The results of our analysis are reminiscent of those obtained for the *in vitro* analysis of the adenovirus 2 major late promoter, which revealed a positive effect on initiation by sequences located between -50 and -60 (41). In the major late promoter, this region is known to bind the factor USF (9). In addition, those experiments showed that the region immediately 5' to the TATA box could tolerate G to A substitutions (41). The stimulatory effect of upstream elements *in vitro* has also been observed for CCAAT factors, AP1 and AP2, Sp1 and derivatives of GAL4-VP16 (42-45).

These observations support our contention that the *Acanthamoeba* TFIID gene contains an upstream element between -35 and -97, and that the TATA box is a distinct necessary element. However, we note that analysis of β -globin genes *in vivo* suggested that at one residue at -37 might have a modest

effect on initiation (46), whereas studies on the HSV *tk* gene did not show any effect of the corresponding region on initiation (47).

Experiments with the 3' deletion mutants, that remove the initiation site, suggest that in *Acanthamoeba* both the start site sequence and the distance from the start site are important for determining the site of initiation. It is particularly interesting that initiation can occur 5 bases downstream of the normal start site, if a sequence resembling the natural start is present, as in the 3' -5 deletion. Initiation in this mutant therefore occurs 180° around the helix relative to the normal site. Similarly, initiation can occur at several sites within an eleven base window, if the start site is removed, as in the 3' -19 deletions. This situation is distinct from that found in vertebrate systems, where the position of the start site is determined primarily by its distance from the TATA box, although in *Acanthamoeba*, the window for initiation also lies between 25 and 35 bp from the TATA box. In yeast, the start point is determined primarily by the start site sequence (48), although yeast polymerase will also initiate efficiently at other sites within 40 - 100 bp from the TATA box. *Acanthamoeba* RNA polymerase II initiation thus has features in common with yeast and mammalian transcription initiation, but is distinct from both.

The use of multiple initiation sites by *Acanthamoeba* RNA polymerase II might indicate that the polymerase is not subject to any constraints on its rotational positioning relative to the TATA box and is presumably rather loosely tethered to the initiation complex. It is equally possible that the polymerase can form distinct complexes that differ with respect to their orientation relative to the TATA box and to DNA. It will be of interest to determine whether the initiation region affects binding by RNA polymerase or, as in the case of *Acanthamoeba* RNA polymerase I (49,50), it affects the ability of RNA polymerase to melt DNA and form the first phosphodiester bond.

ACKNOWLEDGEMENTS

J.-M.W. and F.L. made equal contributions to the work described here. We are grateful to Dr. David Pederson for comments on the manuscript. Supported in part by grant # EY 08706 from the national Eye Institute to E.B., NSF Vermont EPSCOR Grant RII8610679, and a grant from the Lucille B. Markey Charitable Trust to the Department of Microbiology and Molecular Genetics, University of Vermont.

REFERENCES

- Schultz, M.C., Reeder, R.H. & Hahn, S. 1992 *Cell* **69**, 697-702.
- Cormack, B.P. & Struhl, K. 1992 *Cell* **69**, 685-698.
- Sharp, P.A. 1992 *Cell* **68**, 819-821.
- Comai, L., Tanese, N. & Tjian, R. 1992 *Cell* **68**, 965-976.
- Breathnach, R. & Chambon, P. 1981 *Ann. Rev. Biochem.* **50**, 349-383.
- Lewin, B. 1990 *Cell* **61**, 1161-1164.
- Hahn, S., Buratowski, S., Sharp, P.A. & Guarente, L. 1989 *Proc. Natl. Acad. Sci. USA* **86**, 5718-5722.
- Sawadogo, M. & Sentenac, A. 1990 *Ann. Rev. Biochem.* **59**, 711-754.
- Sawadogo, M. & Roeder, R.G. 1985 *Cell* **43**, 165-175.
- Buratowski, S., Hahn, S., Guarente, L. & Sharp, P.A. 1989 *Cell* **56**, 549-561.
- Van Dyke, M.W., Sawadogo, M. & Roeder, R.G. 1989 *Molec. Cell. Biol.* **9**, 342-344.
- Van Dyke, M.W., Roeder, R.G. & Sawadogo, M. 1988 *Science* **241**, 1335-1338.
- Pugh, B.F. & Tjian, R. 1991 *Genes Dev.* **5**, 1935-1945.
- Buratowski, S., Hahn, S., Sharp, P.A. & Guarente, L. 1988 *Nature (London)* **334**, 37-42.
- Cavallini, B., Huet, J., Plassat, J.-L., Sentenac, A., Egly, J.-M. & Chambon, P. 1988 *Nature (London)* **334**, 77-80.
- Wong, J.M., Liu, F. & Bateman, E. 1992 *Gene* **117**, 91-97.
- Sharp, P.A. 1991 *Nature (London)* **351**, 16-18.
- Schmidt, M.C., Kao, C.C., Pei, R. & Berk, A.J. 1989 *Proc. Natl. Acad. Sci. USA* **86**, 7785-7789.
- Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A. & Roeder, R.G. 1989 *Nature (London)* **341**, 299-303.
- Hoffmann, A., Horikoshi, M., Wang, C.K., Schroeder, S., Weil, P.A. & Roeder, R.G. 1990 *Genes Dev.* **4**, 1141-1148.
- 1991 *Rev. Infect. Dis.* **13**.
- Byers, T.J. 1986 *International Review of Cytology* **99**, 311-341.
- Byers, T.J. 1979 *International Review of Cytology* **61**, 283-338.
- McKnight, S. & Tjian, R. 1986 *Cell* **46**, 795-805.
- Maniatis, T., Goodbourn, S. & Fischer, J.A. 1987 *Science* **236**, 1237-1245.
- Saltzman, A.G. & Weinmann, R. 1989 *FASEB J.* **3**, 1723-1733.
- Seto, E., Shi, Y. & Shenk, T. 1991 *Nature (London)* **354**, 241-245.
- Smale, S.T. & Baltimore, D. 1989 *Cell* **57**, 103-113.
- Liu, F. & Bateman, E. 1992 *Gene* **In press**.
- Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Hammer, J.A., Korn, E.D. & Paterson, B.M. 1986 *J. Biol. Chem.* **261**, 1949-1956.
- Short, J.M., Fernandez, J.M., Sorge, J.A. & Huse, W.D. 1988 *Nucleic Acids Res.* **16**, 7583-7600.
- Bateman, E. & Paule, M.R. 1986 *Cell* **47**, 445-450.
- Nellen, W. & Gallwitz, D. 1982 *J. Mol. Biol.* **159**, 1-18.
- Hammer, III., J.A., Bowers, B., Paterson, B.M. & Korn, E.D. 1987 *J. Cell Biol.* **105**, 913-925.
- Jung, G., Korn, E.D. & Hammer, J.A.III 1987 *Proc. Natl. Acad. Sci. USA* **84**, 6720-6724.
- Kaufer, N.F., Simanis, V. & Nurse, P. 1985 *Nature (London)* **318**, 78-80.
- Starr, D.B. & Hawley, D.K. 1991 *Cell* **67**, 1231-1240.
- Lee, D.K., Horikoshi, M. & Roeder, R.G. 1991 *Cell* **67**, 1241-1250.
- Jung, G., Schmidt, C.J. & Hammer, J.A.III 1989 *Gene* **82**, 269-280.
- Yu, Y.-T. & Manley, J.L. 1984 *Nucleic Acids Res.* **24**, 9309-9321.
- Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. & Tjian, R. 1987 *Cell* **48**, 79-89.
- Lee, W., Mitchell, P. & Tjian, R. 1987 *Cell* **49**, 741-752.
- Mitchell, P.J., Wang, C. & Tjian, R. 1987 *Cell* **50**, 847-861.
- Lin, Y.-S. & Green, M.R. 1991 *Cell* **64**, 971-981.
- Myers, R.M., Tilly, K. & Maniatis, T. 1986 *Science* **232**, 613-618.
- McKnight, S.L. 1982 *Science* **217**, 316-324.
- Struhl, K. 1987 *Cell* **49**, 295-297.
- Bateman, E. & Paule, M.R. 1988 *Molec. Cell Biol.* **8**, 1940-1946.
- Kownin, P., Bateman, E. & Paule, M.R. 1987 *Cell* **50**, 693-699.