RT step. This simplified the work and reduced the risk of contamination. Presumably enough DNA, perhaps unintegrated proviral DNA, was liberated from a few dying or dead cells in the culture at the time of virus production. We heated a fraction of stored supernatant at 60°C for 30 min to inactivate the virus. Viral supernatant (5 μl) was added to a PCR reaction. The V3 loop primers BRUV5 and BRUV3 amplified all samples well. A titration of a viral supernatant with the V3 primers showed that 2 to 10 μl gave efficient amplification. However, the SK122/123 *env* primers or the GP5/3 *gag-pol* primers worked less efficiently and sometimes hardly at all.

51. The primers and probes used in this study were the following: V3 loop primers were BRUV5 (5′-GAGGAATTCAGTCTAGCAGAAGAAGAGGT-3′) and BRUV3 (5′-GGCAAGCTTGTGCGTTA-CAATTTCTGGGT-3′). The hybridization probe was BRUVP (5′-GGACCAGGGAGAGCATT-TGTTACAATAGGA-3′). A hybridization probe (SPBRU) specific for the QRGPG motif was used as a guide before sequence data was available. Its sequence was 5′-GTATCCAGAG(A/G)GGAC-CAG-3′ where (A/G) denotes a 50:50 mixture of A and G. The *gag-pol* primers were GP5 (5′-GCAG-GAATTCTTTAGGGAAGATCTGGCCTT-3′) and GP3 (5′-GACGAAGCTTGGGTCGTTGCCAA-AGAGTGAT-3′). The hybridization probe was GPP (5′-TTTCTTCAGAGCAGACCAGAGCC-AACAGCCC-3′). The V1/2 *env*-specific primers SK122 and SK123 and probe SK129, as well as the *tat*-specific primers (T1 and T2) and probes (S1, S2, and S3) have been described (*44, 45*). To eliminate the possibility of PCR contamination, DNA was extracted by colleagues working in laboratories in which an HIV plasmid had never been used. PCR reactions were carried out in a special plexiglass hood in our PCR room which housed in a separate building. Only the workers' forearms enter the hood. No HIV plasmid had ever before been manipulated in this room, nor had any tube containing amplified material ever been opened there. Tubes were opened and analyzed in the main molecular biology laboratory. Negative controls were run each time. All were negative. Reactions contained 2.5 mM $MgCl_2$, 50 mM KCl, 10 mM tris-HCl (pH 8.3), 50 μM of each deoxynucleotide triphosphate, 100 pmol of each primer, and 5 U of *Taq* polymerase (Perkin-Elmer Cetus). Reaction volumes were 100 μl, including 1 to 2 μg of DNA or 5 μl of viral supernatant; 50 μl of mineral oil was laid on top. Thermal cycling parameters were as follows: 80°C, 5 min; 45 cycles of (95°C, 30 s; 55°C, 30 s; 72°C, 30 s); 72°C 10 min. The PCR products were purified from 5% acrylamide gels, treated with kinase and ligated into Sma I–cleaved and dephosphorylated M13mp18 replicative form DNA. The ligated products were transformed into *Escherichia coli* TG1. Plaques were screened in situ with ^{32}P-labeled oligonucleotide probes. Approximately 20 positive clones from each sample were grown and sequenced by the dideoxy method. A total of 217 V3 clones were sequenced. All sequences were read twice and double-checked from the autoradiographs. Two cDNA clones, pLAV75 and pLAV82, isolated in April 1984 from the B-LAV isolate have been described (*18*). Their Pst I inserts were subcloned into the Pst I site of M13mp18 RF DNA. Recombinant clones were sequenced as described above. All the unique nucleotide sequences shown in Figs. 1 and 2 have been deposited with GenBank under accession numbers M64178 to M64223 and M64406 to M64417.
52. J. Palca, *Science* **251**, 1306 (1991).
53. H. W. Kestler III *et al.*, *Nature* **331**, 619 (1988).
54. M. Essex and P. Kanki, *ibid.*, p. 621.
55. C. Mulder, *ibid.*, p. 562.
56. E. I. Golub, G. Li, D. J. Volsky, *J. Virol.* **64**, 3654 (1990).
57. M. Stephenson *et al.*, *ibid.*, p. 3792.
58. S. Wain-Hobson and G. Myers, *Nature* **347**, 18 (1990).
59. D. Concar and C. Anderson, *ibid.*, p. 3.
60. R. A. Weiss, *ibid.* **349**, 374 (1991).
61. P. Aldous, *ibid.*, p. 359.
62. B. Åsjö *et al.*, *Lancet* **ii**, 660 (1986).
63. L. Montagnier *et al.*, *Science* **225**, 63 (1984).
64. Reference to the three cDNA clones derived from B-LAV (HIV-1 Lai) can be found in M. Alizon's notebook dated 3 May 1984. We have sequenced two of these cDNA clones, pLAV75 and pLAV82. They differ from the published sequence of LAV (*19*) by 1.5% (10/649 bases) and 2.2% (7/312 bases), respectively. In addition, the laboratory of M. Martin (National Institute of Allergy and Infectious Diseases, Bethesda, MD) molecularly cloned LAV from a sample denoted C6Tx/LAV. This virus passage from the Institut Pasteur can be traced back to the M2T−/B sample. M. Martin *et al.* sequenced ~4 kb from the 3′ end of the genome. Their sequence and the French LAV J19 sequence were 98% identical (*40*).
65. The single letter amino acid code is as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine.
66. L. Ratner *et al.*, *AIDS Res. Hum. Retroviruses* **3**, 57 (1987).
67. A. Meyerhans, J. P. Vartanian, S. Wain-Hobson, *Nucleic Acids Res.* **18**, 1687 (1990).
68. The *Taq* polymerase error rate may be estimated to be <0.8 bases per 20 V3 loop sequences or <0.6 amino acid substitutions per 20 sequences (*44, 45*). There may be <6 minor forms, presumably, due to *Taq* polymerase error. Care must be exercised in interpreting in biological terms the sequences of minor forms. The homogeneity of the JBB/LAV July 1983 group of sequences is a good indication of the low level of *Taq* polymerase error.
69. As HIV-1 DL (IDAV-2) was cultured at the same time as HIV-1 Bru and HIV-1 Lai from 20 July 1983 onwards, the V3 loop region was amplified from a viral supernatant sample that had been deposited in the CNCM on 15 September 1983. Twenty recombinant M13 clones were sequenced. The V3 loop sequence was CTRPNNNTRERL-SIGPGRPFYATRRIIGDIRQAHC for 95% of the clones. Among 20 protein sequences a variant existed encoding a single R → K amino acid substitution at position 26. This sequence is different from those of authentic HIV-1 Bru and HIV-1 Lai and does not resemble any of the V3 sequences published to date.
70. Samples of HIV-1 Bru, HIV-1 Lai, and B-LAV have been deposited with the CNCM. Samples of each still remain. Additional vials of the M2T−/B sample are available. All samples used for amplification in this study, as well as the recombinant M13 clones, are available for corroboration if necessary.
71. We thank the following colleagues for extracting lysed DNA in their HIV-free laboratories or for performing oligonucleotide synthesis (often with little or no advance warning): M. Buckingham, S. Cole, C. Gouyette, J. Igolen, G. Langsley, B. Robert, and I. Tardieux. We thank M. Alizon for the cDNA clones and G. Myers and J. Goudsmit who educated us in the fine arts of V3 loop sequences. M. Chahine kept the laboratory running while the rest were working. Supported by grants from Institut Pasteur and l'Agence Nationale de Recherches sur le SIDA.

9 April 1991; accepted 26 April 1991

Similarity of Human Mitochondrial Transcription Factor 1 to High Mobility Group Proteins

MELISSA A. PARISI AND DAVID A. CLAYTON

Human mitochondrial transcription factor 1 (mtTF1) has been sequenced and is a nucleus-encoded DNA binding protein of 204 amino acids (24,400 daltons). Expression of human mtTF1 in bacteria yields a protein with correct physical properties and the ability to activate mitochondrial DNA promoters. Analysis of the protein's sequence reveals no similarities to any other DNA binding proteins except for the existence of two domains that are characteristic of high mobility group (HMG) proteins. Human mtTF1 is most closely related to a DNA binding HMG-box region in hUBF, a human protein known to be important for transcription by RNA polymerase I.

TRANSCRIPTION OF HUMAN MITOchondrial DNA (mtDNA) proceeds in opposite directions from a promoter on each strand. These promoters, the light-strand promoter (LSP) and heavy-strand promoter (HSP), are both localized to the major regulatory region for mtDNA transcription and replication (*1*). The use of an in vitro transcriptional assay of human mitochondrial extracts (*2*) coupled with deletional (*3*) and linker substitution (*4*) analyses has revealed the bipartite nature of each promoter; one domain of ~15 bp encompasses the transcriptional start site, and the other is an upstream domain of ~30 bp that is bound by the transcriptional activator protein mtTF1 (*5, 6*). This promoter structure is conserved in both human and murine mitochondria despite the relatively unconserved (<50% identity) nucleotide sequence of the mtTF1 binding sites in the two species (*6*).

Biochemical fractionation of mitochondrial transcription extracts has revealed an absolute requirement for a minimum of two proteins: a mitochondrial RNA (mtRNA) polymerase and mtTF1, which has been isolated from both human and mouse mitochondria (*6–8*). The species specificity of mitochondrial transcription appears to reside in the polymerase-containing fraction; human mtTF1 can substitute for its murine counterpart on the heterologous promoter, but only if mouse extracts containing polymerase activity are provided. Purification of

Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305.

human mtTF1 was facilitated by its abundance in mitochondria and its ability to retain the properties of DNA binding and transcriptional stimulation after denaturation and subsequent renaturation (*9*).

We determined the 21 NH_2-terminal amino acids of purified mtTF1 and the sequences of four tryptic peptides. Degenerate oligonucleotide probes corresponding to two of these peptides hybridized with a single ~540-bp clone from a human lymphocyte cDNA library (*10*). This incomplete clone was used to rescreen the same library and ultimately identified four unique overlapping clones encoding 1936 bp (GenBank M62810). The cDNA sequence revealed an open reading frame of 246 or 240 amino acids starting with an ATG codon at nucleotide position 133 or 151 (Fig. 1). The first ATG is a better match to the Kozak consensus sequence for vertebrate translation initiation (*11*), but the initial methionine of the protein as a whole has not been definitively determined. The open reading frame corresponding to the 25-kD polypeptide as isolated from human mitochondria, identified on the basis of the NH_2-terminal peptide sequence obtained from purified protein, encodes 204 amino acids, with a predicted relative molecular mass (M_r) of 24,400 (in agreement with its migration on protein gels) and an amino acid composition nearly identical to that obtained from the purified protein (*12*). The initial 42 (or 36) amino acids presumably encode a mitochondrial presequence that would be cleaved within the mitochondrial matrix. The total length of the mtTF1 cDNA clone was 1936 bp, and the 5′ end was confirmed by nuclease S1 analysis of mRNA transcripts (Fig. 2) and primer extension (*12*, *13*).

The identity of mtTF1 cDNA was confirmed with the use of an *Escherichia coli* in vivo expression system. Polymerase chain reaction primers corresponding to the ends of the gene, as dictated by protein isolated from human mitochondria, were synthesized with restriction sites added to the 5′ ends appropriate for cloning into the inducible expression vector pT7-7 (*14*). Analysis of proteins from a whole-cell lysate of *E. coli* transformed with the mtTF1–pT7-7 construct revealed a species of ~25 kD that was not present in cells transformed with the pT7-7 plasmid alone (*12*). After extraction, full denaturation, renaturation, and concentration, mtTF1 protein expressed in *E. coli* was able to retard an end-labeled fragment containing the target LSP with a pattern characteristic of mtTF1 purified from human mitochondria (Fig. 3A). Comparable amounts of mtTF1 protein from both sources were then used in deoxyribonuclease I (DNase I) footprinting experiments (*5*); a characteristic footprint of ~30 nucleotides (nt) was present on the LSP in both cases but was absent in the control lanes (Fig. 3B). Finally, functional activity of expressed mtTF1 was demonstrated by its ability to stimulate accurate transcription initiation at both the LSP and HSP in the presence of a fraction of partially purified mtRNA polymerase (Fig. 3C); the ratios of runoff transcripts were identical for the two sources of protein. These data establish the identify of the translated open reading frame of the nuclear mtTF1 gene with the stimulatory activity isolated from human mitochondria.

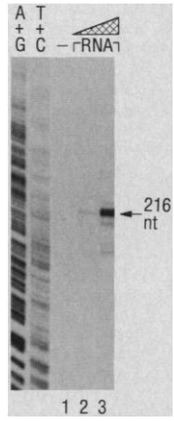
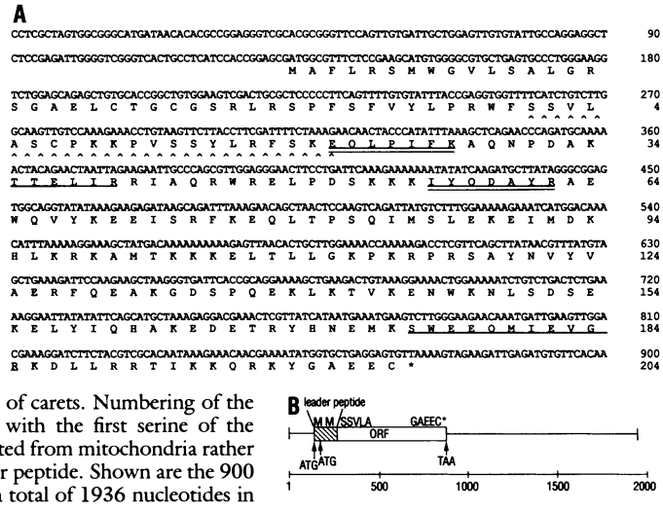


Fig. 2. Analysis of 5′ ends of mtTF1 mRNA transcripts. Nuclease S1 mapping of 5′ termini of mtTF1 mRNA with a double-stranded DNA probe. Template DNA was prepared by digestion of one of the overlapping clones with Sal I, which cleaves at position 216 on the antisense strand, 5′ end–labeling with [γ^{32}P]ATP, digestion with Eco RI, and gel purification of the 714-bp fragment as described (*5*). Approximately 0.15 μg of this 5′ end–labeled fragment were added to the reactions, each of which contained a different amount of RNA (*39*): lane 1, *E. coli* tRNA (100 μg); lane 2, total human RNA (50 μg); lane 3, total human RNA (100 μg). Lanes (A+G) and (T+C) indicate the Maxam and Gilbert sequencing ladders (*40*) produced by chemical cleavage of the same DNA fragment; these lanes were used for location of the endpoint of the major product (indicated by an arrow) protected from digestion at a position 216 nt upstream of the uniquely labeled 5′-phosphate on the Sal I–Eco RI fragment; this corresponds to position 1 of the mtTF1 sequence in Fig. 1A.

A computerized search for sequences similar to the mtTF1 amino acid sequence disclosed no similarities in the PROSITE data bank, which contains protein signatures and common motifs found in transcription factors (*15*). However, a FastA search (*16*) of the translated GenEMBL DNA database or Swiss-Prot protein database with the mtTF1 protein sequence revealed similarity to domains of a small, macronucleus-associated *Tetrahymena* basic protein (LG-1) (*17*), to two moderately abundant yeast nuclear proteins [nonhistone binding proteins 6A and 6B (Nhp6A and Nhp6B, respectively)] (*18*), and to a family of small, abundant nuclear HMG proteins (HMG1 and HMG2) (*19*). A recently cloned member of this group is a human nucleolar transcription factor, hUBF (upstream binding factor) (*20*). Jantzen and co-workers (*20*) proposed the existence of an ~80–amino acid DNA binding domain characteristic of HMG proteins and identified four putative HMG boxes in hUBF. Using an alignment program (FastDB) (*16, 21*) and two other postulated HMG-box consensus sequences (*20, 22*), we compared the mtTF1 protein to hUBF and aligned it with other related members of the HMG1 family (Fig. 4).

Examination of residues likely to be strongly conserved revealed that mtTF1 appears to contain two HMG boxes, although the second is more divergent, in amino acid positions 8 to 80 and 111 to 181; these have 26% identity and 36% overall similarity

Fig. 1. Nucleotide and amino acid sequence of mtTF1. (**A**) Sequence of mtTF1 (*10*), with the deduced amino acid translation of the largest open reading frame shown below in one-letter code. The sequences of tryptic peptides generated from digestion of protein purified from mitochondria (*9*) are indicated by underlines, whereas the NH_2-terminal sequence of mitochondrially purified protein is indicated by a string of carets. Numbering of the amino acid residues starts with the first serine of the NH_2-terminal peptide isolated from mitochondria rather than with the putative leader peptide. Shown are the 900 5′-end nucleotides, out of a total of 1936 nucleotides in GenBank M62810. (**B**) Diagram of the open reading frame of mtTF1 (ORF) and corresponding nucleotide positions. The hatched box denotes the proposed peptide leader sequence, and M indicates potential starting methionine residues corresponding to ATG codons indicated below the box; stop codons are indicated by an asterisk above the box and the codon TAA below the box. (Single-letter abbreviations for the amino acid residues are A, Ala; C, Cys; D. Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.)



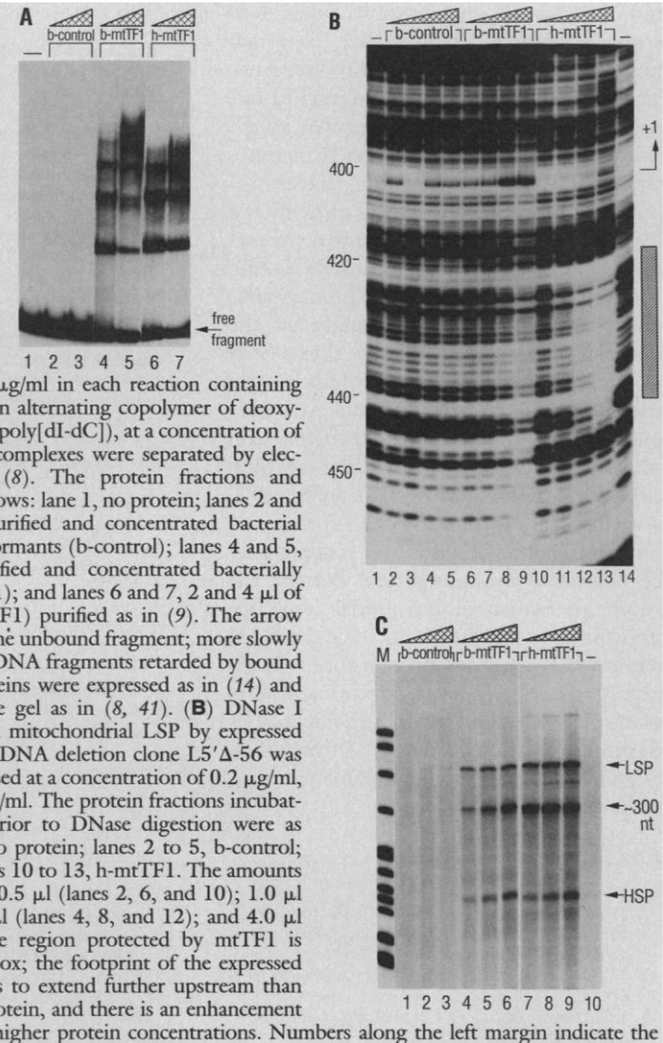
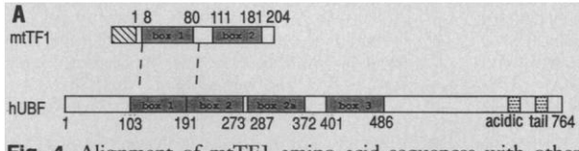


Fig. 3. Identification and functional analysis of mtTF1 produced in a bacterial expression system. (**A**) Nondenaturing gel electrophoresis of protein-DNA com-plexes formed by expressed and native mtTF1. The mtDNA deletion clone L5′Δ-56 containing an intact LSP (56 bp upstream of the transcriptional start site) (*3*) was 5′ end–labeled with ^{32}P (*5, 6*) and used at a concentration of 0.1 μg/ml in each reaction containing nonspecific carrier DNA, an alternating copolymer of deoxyinosine and deoxycytidine (poly[dI-dC]), at a concentration of 0.5 μg/ml. Protein-DNA complexes were separated by electrophoresis as described (*8*). The protein fractions and amounts added were as follows: lane 1, no protein; lanes 2 and 3, 0.5 and 1 μl of gel-purified and concentrated bacterial protein from pT7-7 transformants (b-control); lanes 4 and 5, 0.5 and 1 μl of gel-purified and concentrated bacterially expressed mtTF1 (b-mtTF1); and lanes 6 and 7, 2 and 4 μl of an mtTF1 fraction (h-mtTF1) purified as in (*9*). The arrow indicates the migration of the unbound fragment; more slowly migrating species indicate DNA fragments retarded by bound protein. The bacterial proteins were expressed as in (*14*) and isolated from a preparative gel as in (*8, 41*). (**B**) DNase I footprinting of the human mitochondrial LSP by expressed and native mtTF1. The mtDNA deletion clone L5′Δ-56 was end-labeled as in (A) and used at a concentration of 0.2 μg/ml, with poly(dI-dC) at 0.5 μg/ml. The protein fractions incubated with template DNA prior to DNase digestion were as follows: lanes 1 and 14, no protein; lanes 2 to 5, b-control; lanes 6 to 9, b-mtTF1; lanes 10 to 13, h-mtTF1. The amounts of the proteins added are: 0.5 μl (lanes 2, 6, and 10); 1.0 μl (lanes 3, 7, and 11); 2.0 μl (lanes 4, 8, and 12); and 4.0 μl (lanes 5, 9, and 13). The region protected by mtTF1 is indicated by the hatched box; the footprint of the expressed protein (b-mtTF1) appears to extend further upstream than that produced by native protein, and there is an enhancement of a hypersensitive site at higher protein concentrations. Numbers along the left margin indicate the position in the human mtDNA sequence (*42*). The arrow indicates the start site and direction of transcription (+1). (**C**) In vitro transcription of mtDNA by expressed and native mtTF1. In vitro transcription reactions were carried out essentially as described (*5*) with the use of template DNA containing both mitochondrial promoters and a chromatographically purified fraction of mtRNA polymerase (*8*). Transcription products generated by incorporation of [$\alpha^{32}P$]uridine triphosphate were separated by electrophoresis on denaturing gels and autoradiographed. Protein fractions and amounts added were as follows: lanes 1 to 3, 1, 2, and 4 μl of b-control; lanes 4 to 6, 1, 2, and 4 μl of b-mtTF1; lanes 7 to 9, 0.5, 1 and 2 μl of h-mtTF1; lane 10, no protein added. Lane M contains end-labeled pBR322 DNA markers. Arrows indicate the expected runoff products from the LSP (~416 nt) and HSP (~190 nt); the additional bands have been noted before (*5*) and probably represent processed or prematurely terminated transcripts. The prominent species at ~300 nt is commonly observed in this assay and is likely the product of a ribonuclease mtRNA processing cleavage event (*12, 43*).

with each other (allowing one gap). The HMG boxes of mtTF1 demonstrate the most similarity to hUBF and the yeast and *Tetrahymena* members of the family as well as to the second HMG domain of the vertebrate HMG1 proteins. The greatest similarity in primary amino acid sequence was found between amino acid residues 8 to 94 of mtTF1 and residues 112 to 198 of hUBF [corresponding roughly to the previously identified HMG-box 1 of hUBF (Fig. 4A)], which are 30% identical; an additional 16 amino acids within this range represent conservative amino acid substitutions, for an overall similarity of 48%. A comparison of the second HMG domain of mtTF1 (mtTF1-box 2) with hUBF showed the most similarity between it and the third and most divergent HMG-box, 2a. In addition, this mtTF1-box 2 shares 28% identity with LG-1 (*17*); inclusion of the 12 conservative substitutions between the two sequences yields an overall similarity of 45%, implying an evolutionary relationship between them.

The HMG proteins are classified into two major groups on the basis of size; the class that includes HMG1 and HMG2 is composed of proteins with molecular masses in the range 24 to 29 kD and an amino acid composition with large amounts of acidic and basic residues (*23*). Three domains have been described in HMG1 and HMG2 (*24*); the first two are relatively basic in nature and represent the HMG boxes postulated to bind DNA, whereas the third is a very acidic COOH-terminal tail that has been proposed to interact with positively charged histones, implying a role for HMG1 and HMG2 in nucleosome formation or disassembly, or both (*25*). The HMG boxes located at the NH_2 end of the protein and in the middle of the protein within an HMG protein are different from each other, particularly at the NH_2-termini of each box; the middle HMG box better matches the consensus proposed (*20*) and is more similar to the hUBF boxes. As a consequence, only this middle HMG



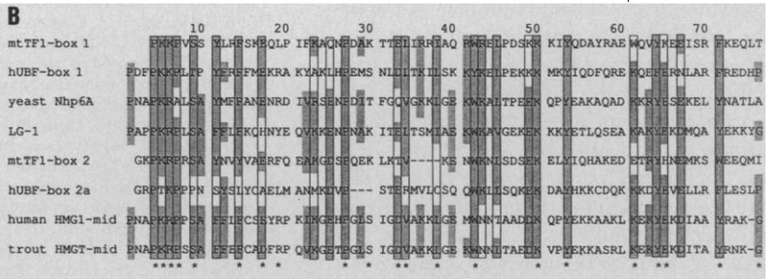


Fig. 4. Alignment of mtTF1 amino acid sequences with other members of the HMG-box family. (**A**) Diagrammatic comparison of mtFT1 and hUBF. The region with the greatest degree of amino acid sequence identity between the two proteins is connected by dashed lines. HMG boxes are represented by shaded boxes, and numbering indicates amino acid position. The proposed leader sequence of mtTF1 is illustrated by a hatched box; the acidic tail domains of hUBF (*20*) are indicated by dotted boxes. (**B**) Alignment of mtTF1 HMG box sequences with similar sequences. Shaded regions indicate residues that are similar in at least five of the sequences; outlined amino acid positions indicate amino acids identical or conserved in at least six of the sequences. Asterisks denote positions with strong conservation in past HMG alignments (*20, 22*). Sequences are as follows: line 1, mtTF1 residues 8 to 80; line 2, hUBF residues 109 to 184 (*20*); line 3, yeast Nhp6A residues 18 to 93 (*18*); line 4, LG-1 residues 9 to 98 (*17*); line 5, mtTF1 residues 111 to 181; line 6, hUBF residues 295 to 366 (*20*); line 7, human HMG1 residues 92 to 166 (*44*); line 8, trout HMGT residues 91 to 165 (*45*). Conserved amino acids are grouped as follows: (P,G), (S,T), (Q,N), (E,D), (K,R), (M,C), (V,L,I,A), (F,Y,W,H) (see Fig. 1 legend for amino acid one-letter code).

box is included in these alignments and Fig. 4B. The high abundance of these proteins (*26*) and their ability to unwind DNA (*27*) provide further support for their similarity to mtTF1 (*28*).

Although the sizes of hUBF and mtTF1 are disparate (~97 kD and ~25 kD, respectively), their similarities at the primary amino acid sequence level are compelling. The first HMG box of mtTF1 is most similar to the known DNA binding HMG box in hUBF (hUBF-box 1). However, the mtTF1-box 2 is more similar to hUBF-box 2a, and both of these domains are shorter and more divergent than other HMG boxes, requiring the insertion of gaps for optimal alignment. Nonetheless, the relationship between the mtTF1 and hUBF amino acid sequences provides a further parallel between nucleolar and mammalian mitochondrial transcription (*6*). Both types of DNA are present in multiple copies in their respective organelles, and there is a precise requirement for specific nucleotides at the transcriptional start sites in both systems (*29*). The mtRNA polymerase and nuclear RNA polymerase I produce polycistronic messages that are later processed to yield mature species; indeed, the HSP is responsible for transcription of the two mitochondrial ribosomal RNA (rRNA) genes. Most notably, the bipartite promoter organization of the two systems is similar (*4, 30*). Finally, in contrast to the case for RNA polymerase II transcription, the species specificity in both rRNA and mtRNA transcription is remarkably strong. SL1 is the human nucleolar protein responsible for this phenomenon (*31*), whereas the mtRNA polymerase (or an uncharacterized protein present in this fraction) performs the same function on a mitochondrial template (*6*). Thus, both hUBF and mtTF1 bind upstream heterologous promoter elements and activate transcription in the presence of other species-specifying elements, and they may do so in part by means of their HMG boxes.

The HMG boxes compose ~70% of the total mature protein and thus may be important for mtTF1 function. The putative leader sequence of 42 (or 36) amino acids exhibits properties consistent with an import function, including evenly spaced basic residues and a potential amphipathic helix (*32*). However, it is possible that mtTF1 exists stably in different molecular forms, or in multiple cellular locations, or both, and that this NH_2-terminal leader sequence is a reflection of that requirement. Finally, we note that mtTF1 lacks a hyperacidic tail at the COOH-terminus (much like the *Tetrahymena* and yeast members of the HMG1 family but unlike hUBF), consistent with the fact that mitochondria appear to lack histones (*33, 34*) and, possibly, other transcription proteins [such as SL1 (*31*)] that might interact with such a COOH-terminus. The general occurrence of an mtTF1-like protein in mitochondria is suggested by the existence of a similar protein in yeast mitochondria (*9, 28, 33, 35*); this protein's sequence is also related to HMG proteins and its presence is required for normal maintenance and expression of yeast mtDNA (*35*).

Given the wide variability in primary amino acid sequence of HMG proteins and the diversity of postulated roles for them, from general chromatin assembly proteins to specific transcriptional activators and even yeast mating-type proteins (*36*) and mammalian sex determination proteins [the recently cloned human and mouse SRY genes (*37*)], it seems likely that various evolutionary or functional subclasses of HMG proteins may be defined in the near future. Because the nature of human mitochondrial transcriptional initiation is known in some detail (*1*) and mtTF1 is a small protein that is easily expressed in *E. coli*, future functional studies of mtTF1's mode of action provide an opportunity to learn the role of HMG-box elements in a simple and streamlined transcriptional activator.

REFERENCES AND NOTES

1. D. A. Clayton, *Trends Biochem. Sci.* **16**, 107 (1991).
2. M. W. Walberg and D. A. Clayton, *J. Biol. Chem.* **258**, 1268 (1983).
3. D. D. Chang and D. A. Clayton, *Cell* **36**, 635 (1984).
4. J. N. Topper and D. A. Clayton, *Mol. Cell. Biol.* **9**, 1200 (1989).
5. R. P. Fisher, J. N. Topper, D. A. Clayton, *Cell* **50**, 247 (1987).
6. R. P. Fisher, M. A. Parisi, D. A. Clayton, *Genes Dev.* **3**, 2202 (1989).
7. R. P. Fisher and D. A. Clayton, *J. Biol. Chem.* **260**, 11330 (1985).
8. ———, *Mol. Cell. Biol.* **8**, 3496 (1988).
9. R. P. Fisher, T. Lisowsky, G. A. M. Breen, D. A. Clayton, *J. Biol. Chem.*, in press.
10. The two peptide fragments indicated by double underlines were used to design the degenerate oligonucleotide probes NT-30 (5′-TT(G/A)AAIATI-GGIA(G/A)(T/C)TG(T/C)TC-3′), corresponding to the antisense sequence of residues 63 to 69, and NT-13 (5′-C(T/G)(G/A)TAIGC(G/A)TC(T/C)TG (G/A)TAIAT-3′), the antisense sequence of residues 98 to 104 (I, inosine). Oligonucleotide NT-30 was 5′ end–labeled with γ^{32}P-labeled adenosine triphosphate (ATP) and used to screen duplicate nitrocellulose filters from a human lymphocyte λ cDNA library [S. J. Elledge, J. T. Mulligan, S. W. Ramer, M. Spottswood, R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1731 (1991)] essentially as described in standard protocols (*38*). Hybridization conditions were 6× saline sodium citrate, 50 mM sodium phosphate, pH 7.0, 5× Denhardt's solution, and *E. coli* tRNA (200 μg/ml) for 1 to 2 days at 37°C, and, after hybridization, washes were in low stringency 6× SSC with several changes at room temperature for ~1 hour. Approximately 5×10^5 plaques were screened to obtain 33 positives, which, on subsequent screening with end-labeled oligonucleotide NT-13 under the same conditions, yielded exactly one clone that hybridized to both probes. After plaque purification, this clone was converted into a plasmid by means of a cre-lox recombination system, sequenced, and used to obtain the additional four clones that made up the complete sequence.
11. M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987).
12. M. A. Parisi and D. A. Clayton, unpublished observations.
13. Northern analysis of human RNA, with the 612-bp open reading frame as a probe, revealed that several mtTF1 RNA species in the range 1.5 to 2.0 kb were expressed in human cells, but we were unable to map their 3′ termini precisely. The longest transcript may extend even farther in the 3′ direction than the longest clone, which contains a polyadenylation signal at position 1916 (15 bp from its 3′ terminus) but lacks a polyadenylated tail. [This phenomenon has been noted in other cDNA clones isolated from this library (S. J. Elledge, personal communication).] Other potential polyadenylation sites can be found at positions 1516 and 834, but S1 analyses were insufficient to determine if these sites were utilized, perhaps because of alternatively spliced or processed mRNA species.
14. The original T7 expression system is described by S. Tabor and C. C. Richardson [*Proc. Natl. Acad. Sci. U.S.A.* **82**, 1074 (1985)]. The vectors and induction methods are described by S. Tabor in *Current Protocols in Molecular Biology* [F. M. Ausubel *et al.*, Eds. (Wiley, New York, 1987), suppl. 11, pp. 16.2 .1–16.2.11] and by F. W. Studier and B. A. Moffatt [*J. Mol. Biol.* **189**, 113 (1986)].
15. A. Bairoch, thesis, University of Geneva, Switzerland (1990).
16. W. R. Pearson and D. J. Lipman, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444 (1988).
17. I. G. Schulman, R. G. Cook, R. Richman, C. D. Allis, *J. Cell. Biol.* **104**, 1485 (1987); S. Y. Roth, I. G. Schulman, R. G. Cook, C. D. Allis, *Nucleic Acids Res.* **15**, 8112 (1987).
18. D. Kolodrubetz and A. Burgum, *J. Biol. Chem.* **265**, 3234 (1990).
19. HMG primary sequences are reviewed by J. M. Walker [in *The HMG Chromosomal Proteins*, E. W. Johns, Ed. (Academic Press, London, 1982), pp. 69–87].
20. H.-M. Jantzen, A. Admon, S. P. Bell, R. Tjian, *Nature* **344**, 830 (1990).
21. D. J. Lipman and W. R. Pearson, *Science* **277**, 1435 (1985); D. L. Brutlag, J.-P. Dautricourt, S. Maulik, J. Relph, *CABIOS* **6**, 237 (1990).
22. D. Kolodrubetz, *Nucleic Acids Res.* **18**, 5565 (1990).
23. Reviewed by E. W. Johns [in *The HMG Chromosomal Proteins*, E. W. Johns, Ed. (Academic Press, London, 1982), pp. 1–7] and L. Einck and M. Bustin [*Exp. Cell Res.* **156**, 295 (1985)].
24. G. R. Reeck, P. J. Isackson, D. C. Teller, *Nature* **300**, 76 (1982).
25. M. Carballo, P. Puigdomènech, J. Palau, *EMBO J.* **2**, 1759 (1983).
26. G. H. Goodwin and C. G. P. Mathew, in *The HMG Chromosomal Proteins*, E. W. Johns, Ed. (Academic Press, London, 1982), pp. 193–221.
27. K. Javaherian, L. R. Liu, J. C. Wang, *Science* **199**, 1345 (1978).
28. R. P. Fisher, T. Lisowsky, M. A. Parisi, D. A. Clayton, in preparation.
29. J. E. Hixson and D. A. Clayton, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2660 (1985); R. M. Learned, T. K. Learned, M. M. Haltiner, R. T. Tjian, *Cell* **45**, 847 (1986); P. Kownin, E. Bateman, M. R. Paule, *ibid.* **50**, 693 (1987); S. P. Bell, R. M. Learned, H.-M. Jantzen, R. Tjian, *Science* **241**, 1192 (1988).
30. S. T. Smale and R. Tjian, *Mol. Cell. Biol.* **5**, 352 (1985).
31. R. M. Learned, S. Cordes, R. Tjian, *ibid.* p. 1358; S. P. Bell, H.-M. Jantzen, R. Tjian, *Genes Dev.* **4**, 943 (1990).
32. M. G. Douglas, M. T. McCammon, A. Vassarotti, *Microbiol. Rev.* **50**, 166 (1986).
33. F. Caron, C. Jacq, J. Rouvière-Yaniv, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4265 (1979).
34. L. DeFrancesco and G. Attardi, *Nucleic Acids Res.* **9**, 6017 (1981).
35. J. F. X. Diffley and B. Stillman, personal communication.
36. M. Kelly, J. Burke, M. Smith, A. Klar, D. Beach, *EMBO J.* **7**, 1536 (1988).
37. A. H. Sinclair *et al.*, *Nature* **346**, 240 (1990); J. Gubbay *et al.*, *ibid.*, p. 245.

38. J. Sambrook, E. F. Fritsch, T. Maniatis, in *Molecular Cloning: A Laboratory Manual*, C. Nolan, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1989).
39. Hybridization was allowed to proceed for 7 hours at 55°C, a temperature calculated to maximize the stability of DNA-RNA hybrids. The samples were digested by nuclease S1 (5000 U/ml) for 1 hour at 37°C, stopped as described in (*38*), precipitated, separated by electrophoresis on a 5% polyacrylamide–7 M urea gel, and autoradiographed.
40. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
41. D. A. Hager and R. R. Burgess, *Anal. Biochem.* **109**, 76 (1980).
42. S. Anderson *et al.*, *Nature* **290**, 457 (1981).
43. D. D. Chang and D. A. Clayton, *EMBO J.* **6**, 409 (1987).
44. L. Wen *et al.*, *Nucleic Acids Res.* **17**, 1197 (1989).
45. B. T. Pentecost *et al.*, *ibid.* **13**, 4871 (1985).
46. M.A.P. thanks G. A. M. Breen for purification of protein and advice and support during sabbatical leave in this laboratory. We thank S. J. Elledge for the human cDNA library and D. L. Brutlag and H.-M. Jantzen for discussions of sequence alignments. We acknowledge the Protein and Nucleic Acid Facility, Beckman Center, for the synthetic oligonucleotides and amino acid analysis, and we thank R. P. Fisher and J. N. Doda for assistance; J. F. Hess, J. L. Bennett, L. L. Stohl, C. Szymkowiak, J. Poulton, and T. Lisowsky for reagents and advice; M. E. Schmitt and J. L. Bennett for comments and K. Shaw for help in preparation of the manuscript. M.A.P. is a Medical Scientist Training Program Trainee of the National Institute of General Medical Sciences (GM07365-15). Supported by grants from the National Institute of General Medical Sciences (GM33088-20) and from the American Cancer Society, Inc. (NP-9P).

2 January 1991; accepted 5 March 1991

Retronphage φR73: An *E. coli* Phage That Contains a Retroelement and Integrates into a tRNA Gene

SUMIKO INOUYE, MELVIN G. SUNSHINE, ERICH W. SIX, MASAYORI INOUYE*

Some strains of *Escherichia coli* contain retroelements (retrons) that encode genes for reverse transcriptase and branched, multicopy, single-stranded DNA (msDNA) linked to RNA. However, the origin of retrons is unknown. A P4-like cryptic prophage was found that contains a retroelement (retron Ec73) for msDNA-Ec73 in an *E. coli* clinical strain. The entire genome of this prophage, named φR73, is 12.7 kilobase pairs and is flanked by 29–base pair direct repeats derived from the 3′ end of the selenocystyl transfer RNA gene (*selC*). P2 bacteriophage caused excision of the φR73 prophage and acted as a helper to package φR73 DNA into an infectious virion. The newly formed φR73 closely resembled P4 as a virion and in its lytic growth. Retronphage φR73 lysogenized a new host strain, reintegrating its genome into the *selC* gene of the host chromosome and enabling the newly formed lysogens to produce msDNA-Ec73. Hence, retron Ec73 can be transferred intercellularly as part of the genome of a helper-dependent retronphage.

A SATELLITE DNA, CALLED MSDNA, has been discovered in myxobacteria and some clinical strains of *E. coli* (*1*). These molecules consist of single-stranded DNA that is branched out from an internal guanosine residue by a 2′,5′ phosphodiester linkage. Reverse transcriptase (RT) is required for the synthesis of msDNA. We found an msDNA, called msDNA-Ec73, in an *E. coli* clinical strain (*2*). The msDNA-Ec73 is produced by a 2.4-kb retron (retron Ec73) that consists of the genes for DNA-linked RNA and msDNA, an open reading frame (ORF) that encodes a protein of unknown function (ORF 316), and the gene for an RT. Retron Ec73 is part of a 12.7-kb fragment of foreign DNA flanked by 29-bp direct repeats that was integrated into the 3′ end of the gene for selenocystyl tRNA (*selC*) at 82 min on the *E. coli* chromosome (*2*).

Except for the presence of the retron region, the 12.7-kb segment resembles the genome of satellite coliphage P4. Its ORFs show sequence similarity to genes and ORFs of P4 and have the same arrangement as on the P4 genome. This suggests that retron Ec73 exists in the genome of a prophage. This prophage, called a φR73, differs from P4 prophage in its location on the *E. coli* chromosome (P4 is integrated at 97 min) and in the flanking direct repeats (a different 20-bp sequence than for P4) (*3*).

We tested whether the φR73 prophage can be induced to generate an infectious retronphage. This would allow transfer of retron Ec73 by integration of the prophage into the *E. coli* chromosome. Phage P4 requires a helper genome, such as that of coliphage P2, to provide the late gene functions for its lytic growth (*3*). Because the φR73 genome resembles that of P4, we tested whether P2 could serve as helper for excision of φR73. For P2 to be a helper, φR73 DNA must contain a *cos* site, which allows P2 to package φR73 DNA into a phage head. Indeed, a segment of the φR73 DNA contains a 19-bp sequence that is similar to the consensus sequence of P2, P4, and phage 186 that exists between the sites that produce the ends of packaged DNA (*3*).

The plasmid pCl-23a (*4*) is about 18 kb in size; thus its dimer (36 kb) is only slightly larger than the P2 genome (33 kb). If the φR73 *cos* site is P2-compatible, P2 should be able to package pCl-23a as a dimer. To test this possibility, we infected *E. coli* C1a cells that carried pCl-23a with P2. We found that we could transduce pCl-23a into another *E. coli* strain (C366) that carries a P2 prophage (*5*). The transductant produced a low amount of phage that formed P4-like plaques with indicator strains lysogenic for P2. Phage isolated from one such plaque served as a source for the φR73 wild-type phage. That φR73, like P4, depended on a helper phage for its lytic growth was evident from its inability to form plaques on lawns of *E. coli* strains that lacked a P2 prophage.

Electron microscopic examination of a negatively stained preparation of retronphage φR73 indicated that its morphology was indistinguishable from that of P4. Both phages have icosahedral heads attached to tails with a contractile sheath (Fig. 1). We

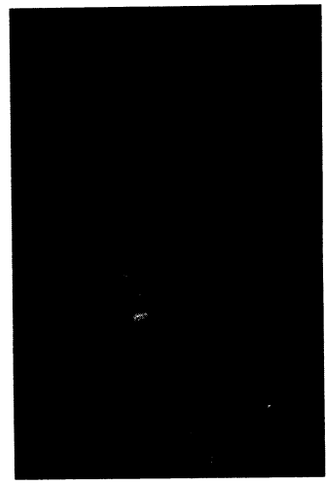


Fig. 1. Electron micrograph of P4-like retronphage φR73. Phage from φR73 lysate was negatively stained with phosphotungstate (*10*) and viewed under a Hitachi H-7000 electron microscope. The tail of the φR73 virion should be the same as that of its helper, phage P2, and may therefore serve as size reference (tail length, 135 nm).

S. Inouye and M. Inouye, Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey at Rutgers, Piscataway, NJ 08854.
M. G. Sunshine and E. W. Six, Department of Microbiology, University of Iowa, Iowa City, IA 52242.

*To whom correspondence should be addressed.