Nucleic Acids Research, 1994, Vol. 22, No. 22 4599–4606

Heterogeneity in the 5′ untranslated region of mouse cytochrome c_T mRNAs leads to altered translational status of the mRNAs

Gary K.Yiu, Wei Gu and Norman B.Hecht*
Department of Biology, Tufts University, Medford, MA 02155, USA

Received August 18, 1994; Revised and Accepted September 29, 1994

ABSTRACT

Previous studies have shown that the differential regulation of mouse somatic cytochrome c (cyt c_S) and testicular cytochrome c (cyt c_T) during spermatogenesis is accompanied by changes in mRNA length [Hake *et al.* (1990) *Development,* 110, 249–257]. When analyzed by polysomal gradient sedimentation, cytochrome c_T sediments in two broad size classes: non-polysomal mRNAs are about 0.6 to 0.75 kb and polysomal mRNAs range from 0.7 to 0.9 kb. Both classes of mRNAs shorten to about 0.5 kb following deadenylation. Oligonucleotide-directed cleavage of the cytochrome c_T RNAs by RNase H reveals that the size heterogeneity of cytochrome c_T mRNAs resides in the 5′ untranslated regions (UTRs). Ribonuclease protection assays reveal that multiple cytochrome c_T mRNAs are transcribed from six different transcriptional start sites spanning a region of 59 nucleotides in the 5′UTR from +1 to +59. Transcripts derived from the first and second transcriptional initiation sites are not loaded onto polysomes as efficiently as those transcripts initiated from the other start sites. Each of the longer mRNAs has an upstream open reading frame, which starts at +8 and ends at +136 in the 5′UTR of the cytochrome c_T transcript. Computer analysis suggests that the lengthened 5′ UTR sequences allow additional hairpin structures to be formed. Since the upstream open reading frame and the additional stem loop structure are absent in the 5′ UTRs of the cytochrome c_T mRNAs initiated from the four downstream start sites, we suggest that these sequences in the two longest cytochrome c_T transcripts hinder their loading onto polysomes.

INTRODUCTION

Spermatogenesis is a complex sequence of events in which spermatogonia and spermatocytes undergo cellular differentiation and ultimately develop into haploid spermatozoa. During these programmed events, many genes are under translational control leading to their temporal and tissue-specific expression (1,2). The translational status of a specific mRNA among others within a population may be controlled by changes in its stability, primary structure, polyadenylation, as well as by protein–mRNA interactions (3–6). Fluctuations in the amount or activity of specific *trans*-acting factors in accordance with changes in metabolic state can modulate the levels of translation of pre-existing mRNAs. The 5′ and 3′ UTRs are recognized as important elements in the post-transcriptional regulation of gene expression. For example, the intracellular iron concentration controls the translation of ferritin and transferrin receptor (7), as well as 5-aminolevulinate synthase (8) through the interaction of iron-responsive element binding proteins with the iron-responsive elements present in the 5′ or 3′ UTRs of the respective mRNAs.

Cytochrome c is a nuclear-encoded mitochondrial protein which serves as an electron carrier in the respiratory chain. In mouse testis, there exists two distinct forms of cytochrome c which differ in 14 of 104 amino acids (9,10). The mouse cytochrome c_T and c_S genes have been mapped to chromosome 2 and chromosome 6, respectively (11). The somatic form, cytochrome c_S, is ubiquitously expressed in all somatic tissues as well as in germ cells in their early stages of differentiation; whereas cytochrome c_T is present only in later stages of germ cells from mid-meiotic spermatocytes to spermatozoa. Cytochrome c_T mRNAs of 0.6–0.9 kb are detected in pachytene spermatocytes of mouse testis, with the size class of 0.7–0.9 kb being primarily polysomal while the 0.6–0.75 kb mRNAs are mostly non-polysomal. In round spermatids, the mRNA sizes become less heterogenous and converge to about 0.5 kb (4). The fact that polysomal cyt c_T mRNAs have longer poly(A)$^+$ tails than non-polysomal ones contributes to these differences in lengths.

As translation of mRNAs can be influenced by minor variations in their structure (12–14), alterations in 5′ UTRs and/or 3′ UTRs as well as poly(A)$^+$ tail lengths could lead to a differential polysomal distribution of transcripts. According to the scanning model for initiation of translation in eukaryotes, the 40S ribosomal subunits carrying Met-tRNA$_i$Met and various initiation factors scan from the 5′ end of mRNAs until they reach the first AUG codon in a favorable context for translation (15,16). The presence

*To whom correspondence should be addressed

of upstream open reading frames and secondary structures in the 5'UTRs of mRNAs affect the efficiency of translational initiation. Downstream secondary structures presumably slow down ribosomal scanning, thereby increasing the fidelity of initiation (17); while the presence of hairpin structures upstream from the initiator codons impairs translation in accordance with the strength and position of the secondary structures (18,19). In the cytochrome c_T gene of the rat, three major transcriptional start sites have been identified. The transcript initiated at the most 5' site has an additional upstream open reading frame (ORF) consisting of 129 nucleotides which terminate 91 nucleotides upstream of the ORF that encodes the cytochrome c_T protein (20). The functional significance of this extra ORF in the post-transcriptional regulation of cytochrome c_T mRNAs has not been determined.

In this study, using RNase H digestion analysis, Northern blotting and ribonuclease protection assays of mouse testicular RNAs, we have determined the sources of the size heterogeneity of the mouse cytochrome c_T transcripts. Moreover, changes in the 5' UTR lengths of the cyt c_T mRNAs lead to secondary structure differences that correlate with translational status.

MATERIALS AND METHODS

Construction of cDNA and genomic subclones of the mouse cytochrome c_T

The cDNA coding for cytochrome c_T was isolated from a λgt-10 adult CD-1 mouse testis cDNA library as described elsewhere (4). The cDNA clone was inserted into the *Eco*RI site of M13 mp18 vector to generate the bacteriophage clone 1e. The primary cytochrome c_T subclone (pGt.W) was produced by inserting the 566 bp *Eco*RI–*Hin*dIII insert of clone 1e into pGEM 4Z plasmid. The subclone pGt.5' was derived from it (Fig. 1).

Mouse cytochrome c_T genomic clones were obtained by screening a mouse (ICR Swiss) genomic library in λGEM-11 (Promega, Madison, WI). The library was plated and screened with an [α-^{32}P]-labeled mouse cytochrome c_T cDNA (4). Approximately 10^6 plaques were plated, and after several rounds of plaque purification, 12 clones were isolated. These clones were then counter-screened with a probe which encodes part of the 5' UTR of the cytochrome c_T mRNA. Eight of the clones were found to contain the 5'UTR of the cytochrome c_T cDNA. From one of these clones, a 6.5 kb *Bam*HI fragment was subcloned into pGem11Zf(+) (Promega, Madison, WI) to generate plasmid pKC2-4. This plasmid contains 1.75 kb of 5' flanking sequence, exons I and II, and introns I and II of the mouse cytochrome c_T gene. The nucleotide sequence of the 5' flanking sequence, exon I, and a portion of intron I (pKC2-4A3) was determined completely on both strands by the dideoxy chain termination method (21) (Fig. 1).

Oligonucleotide directed RNase H cleavage of cytochrome c_T mRNAs

RNA samples were prepared from total testis or gradient fractionated testicular post-mitochondrial extracts as described previously (4). Oligonucleotide ST, with a sequence of 5'-CCITTTTCCACIGTGTGGCACTG-3' which is complementary to bases +243 to +265 of the cytochrome c_T sequence (see Fig. 4) was synthesized by Operon Technologies, Inc. (Alameda, CA). RNAs (15 μg) were annealed with 250 pmol of oligo-nucleotide ST and/or 500 pmol of oligo $d(T)_{12-18}$ (Collaborative Research, Bedford, MA) by first heating the samples at 65°C in 10 μl of 100 mM KCl and 0.1 mM EDTA for 2 min, followed by slow cooling to room temperature. The hybrids formed between oligonucleotide ST and cytochrome c_T mRNA and between oligo dT and the poly A tail were incubated in digestion buffer (one unit of RNase H, 50 mM Tris, pH 7.5, 10 mM $MgCl_2$, 60 mM KCl, 1 mM DTT, and 0.5 mg/ml BSA) (22) at 37°C for 30 min. After digestion, the RNA samples were extracted with phenol–chloroform, ethanol precipitated and analyzed by Northern blotting.

Northern blot analysis

To resolve the small size differences among the cytochrome c_T transcripts, a modification of the procedure for analyzing RNA on polyacrylamide gels as described by Childs *et al.*, (23) was employed. Briefly, RNA samples in 80% deionized formamide, 1×TBE (90 mM Tris–borate, pH 8.5, 2 mM EDTA) and glycerol dye mix (0.6% bromophenol blue, 0.6% xylene cyanol, 50% glycerol) were denatured at 65°C for 5 min and then chilled quickly on ice. The RNAs were electrophoresed through a 5% polyacrylamide gel containing 7 M urea in 1×TBE until the xylene cyanol dye front reached the bottom of the gel and then electroblotted onto Genescreen membrane (New England Nuclear, Boston, MA) in 1×electroblotting buffer (10 mM Tris, pH 7.8, 5 mM sodium acetate and 0.5 mM EDTA) at 4°C.

Radiolabeled antisense RNA probes were synthesized from linearized cytochrome c_T cDNA subclones in pGEM4Z by using T7 RNA polymerase. Hybridization was carried out following the procedure of Church and Gilbert (24). Prehybridization was performed at 65°C for 2 h in hybridization buffer containing 1% crystalline BSA, 1 mM EDTA, 0.5 M sodium phosphate, pH 7.2, 7% SDS and 100 μg/ml of denatured sheared salmon sperm DNA. Hybridization was performed in fresh hybridization solution containing $1-10\times10^7$ cpm/ml of probe at 65°C for 12–16 h. Membranes were then washed twice in washing buffer (40 mM sodium phosphate, pH 7.0, 1 mM EDTA, and 1% SDS) at 23°C each for 15 min, followed by two more washes at 65°C each for 30 min. In order to minimize non-specific background hybridization, the membranes were treated with RNase A at 37°C for 15 min in digestion buffer containing 20 μg/μl RNase A, 20 mM Tris, pH 7.5, 0.6 M NaCl and 1 mM EDTA. Autoradiographs were developed with an intensifying screen for 1–7 days.

Ribonuclease protection assay

For RNase protection assays, a 301-bp genomic DNA fragment from −133 to +168 was subcloned into pGEM3zf(+) (Promega, Madison, WI) to generate the plasmid pKC2-4A3. It contains 133 bp of 5' flanking sequence, the entire first exon of 141 bp and 27 bp of the first intron. The plasmid pGt.5', which contains a cDNA fragment from +82 to +220, was also used. RNase protection assays were performed using a modification of the procedure of Krieg and Melton (25). A Riboprobe Gemini System II kit (Promega, Madison, WI) was employed to produce radiolabeled full length antisense RNA transcripts from linearized plasmids (pGt.5' and pKC2-4A3) by using [α-^{32}P]CTP (800 Ci/mmol) (Amersham Corp., Arlington Heights, IL) and either T7 or SP6 RNA polymerase. The RNA probes were gel purified by electrophoresis through 4% denaturing polyacrylamide gels and recovered from the gel slices by shaking at 37°C in 400 μl of elution buffer (0.3 M sodium acetate, pH 5.2, 0.5% SDS and

2 mM EDTA) for 2–4 h, followed by ethanol precipitation. RNA samples (10–20 μg) were coprecipitated with $2-5\times10^5$ c.p.m. of antisense probe, washed once with 70% ethanol and resuspended in 30 μl of annealing buffer (80% deionized formamide, 0.4 M NaCl, 40 mM PIPES, pH 6.4 and 1 mM EDTA). Annealing was performed at 45°C for 12–16 h after heating the annealing mix at 85°C for 5 min. After hybridization, the samples were incubated in 350 μl of 0.5 M NaCl, 10 mM Tris, pH 7.5, 5 mM EDTA, 50 μg/ml RNase A and 2 μg/ml RNase T1 at 30°C for 30 min. Following the RNase digestion, the samples were incubated with 50 μg of proteinase K and 10 μl of 20% SDS at 37°C for 15 min, extracted twice with phenol–chloroform, and ethanol precipitated in the presence of 10 μg of yeast or *E.coli* tRNA. The samples were then resuspended in 8 μl of RNA loading buffer (5 mM EDTA, pH 8.0, 89% deionized formamide, 0.1% bromophenol blue, 0.1% xylene cyanol), heated at 90°C for 3 min, and analyzed in 5% denaturing polyacrylamide gels.

RESULTS

Subcellular distribution of the multiple cytochrome c_T mRNAs

The sizes of the non-polysomal and polysomal cyt c_T mRNAs differ. The non-polysomal cytochrome c_T mRNAs range from 0.6–0.75 kb, whereas the polysomal cytochrome c_T mRNAs are longer and range from 0.7–0.9 kb (4). When these mRNAs are deadenylated and analyzed by Northern blotting, both the non-polysomal and polysomal cyt c_T mRNAs shorten to about 0.5 kb. To reveal any additional small differences in mRNA length which would be masked by the limited resolution of the agarose system previously used (4), a polyacrylamide gel system was employed to resolve the multiple forms of cytochrome c_T transcripts (23). To determine whether the size heterogeneity of the cytochrome c_T mRNAs resides in their 5′ or 3′ regions , they were cleaved specifically with RNase H after being hybridized to an oligonucleotide, oligo ST, which is complementary to the nucleotides encoding amino acids 17 through 23 of mouse cytochrome c_T. The multiple fragments obtained were detected by Northern blotting with cDNA probes Gt.5′ or Gt.W (Fig. 1).

When testicular mRNAs were hybridized with an antisense probe generated from pGt.W that recognizes both the 5′ and 3′ ends of the cytochrome c_T transcripts, a heterogeneous group of fragments ranging in size from about 650 to 900 nucleotides (nt) was detected (Fig. 2A). The non-polysomal fraction of testicular mRNA contains the entire population of cytochrome c_T mRNAs (Fig. 2A, lanes T and NP), whereas only the upper half of the broad band is detected in the polysomal fraction (Fig. 2A, lane P). After deadenylation, multiple bands of varying intensities ranging from 680 to about 390 nt are detected in the total testis, non-polysomal, and polysomal samples (Fig. 2A, panel dT). The most abundant transcript is about 680 nt long. Moreover, a minor fragment of approximately 390 nt long is seen. This is likely the result of spurious annealing of oligo dT to an adenine rich region at +350 to +362 of the cytochrome c_T mRNA. This additional band was also detected with Gt.5′, a probe specific for the 5′UTR of cytochrome c_T mRNA (Fig. 2B, panel dT).

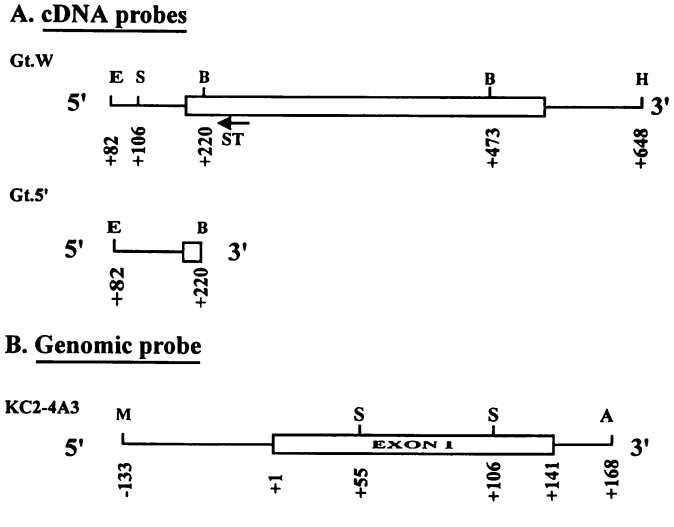

Figure 1. Schematic diagram of cytochrome c_T cDNAs and a cytochrome c_T genomic subclone. **(A)** The position of oligonucleotide ST in cytochrome c_T mRNA is shown with an arrow indicating its 5′ to 3′ orientation. Thin lines denote the untranslated regions and the open box represents the coding region of cytochrome c_T. Plasmid pGt.W, containing the primary cytochrome c_T cDNA subclone, was generated by inserting the 566-bp *Eco*RI–*Hin*dIII restriction fragment of clone 1e (4) into a pGEM4-Z plasmid. The subclone Gt.5′ was derived from Gt.W and is specific for the 5′ terminus of mouse cyt c_T. Restriction enzyme sites are denoted as: E, *Eco*RI; S, *Sty*I; B, *Bgl*II; H, *Hin*dIII. The numbers indicate the nucleotide positions in the cytochrome c_T cDNA sequence by assigning +1 to the first transcriptional start site. **(B)** The genomic subclone KC2-4A3 contains exon I, 27 bp of the first intron, and 133 bp of the 5′ flanking sequence of the mouse cyt c_T gene. The numbers represent the nucleotide positions in the genomic sequence. Restriction enzyme sites are denoted as:M, *Sma*I; S, *Sty*I; A, *Acc*I.

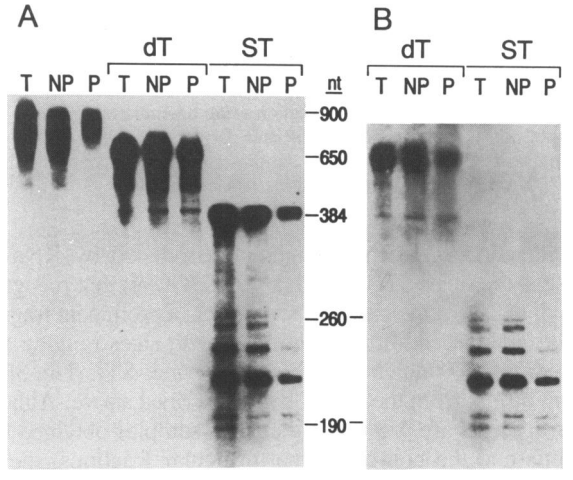

Figure 2. Polyacrylamide gel Northern blot analysis of cytochrome c_T RNA after deadenylation and cleavage with RNase H. Intact, deadenylated and internally cleaved samples (15 μg) prepared from total testis (T), non-polysomal (NP) and polysomal (P) testicular RNAs, were electrophoresed through 5% polyacrylamide gels and blotted onto nylon membranes. Autoradiograms were obtained after hybridization with Gt.W, a probe recognizing the entire cytochrome c_T mRNA (panel **A**) or Gt.5′, a probe that specifically recognizes the region upstream to oligo ST (panel **B**). dT above the panels indicates that the samples were deadenylated, while ST indicates that the deadenylated samples were further cleaved into 5′ and 3′ fragments by oligo ST-directed RNase H digestion. The estimated lengths of the RNA fragments in nucleotides (nt) are indicated.

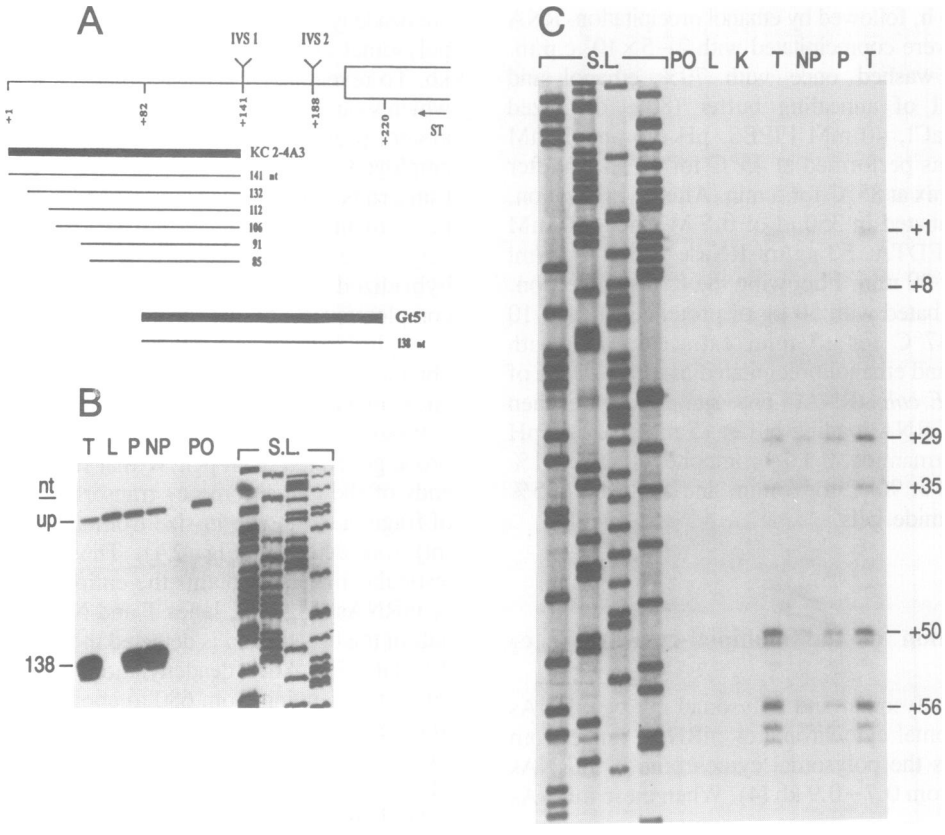

Figure 3. Ribonuclease protection analysis of mouse cytochrome c_T transcripts. RNA samples (15 or 20 μg) were annealed to radiolabeled antisense RNA probes, digested with RNases A and T1, and electrophoresed through 5% polyacrylamide sequencing gels. (**A**) Schematic representation of the various transcriptional start sites detected in the 5'UTR of mouse cytochrome c_T mRNAs. At the top of the figure, the line represents the 5' UTR whereas the open box denotes the coding region of cytochrome c_T. The site of oligonucleotide ST,the cytochrome c_T specific oligonucleotide, is denoted by ST. The numbers indicate the nucleotide positions of the cytochrome c_T transcript starting from the first major transcriptional start site. IVS 1 and IVS 2 indicate the positions of intron 1 and intron 2, respectively. The antisense RNA probes prepared from the cDNA subclone Gt.5' or the genomic subclone KC2-4A3 (only a portion of the probe from +1 to +141 is shown here) are indicated by thick lines while the protected fragments with their respective lengths in nucleotides (nt) are denoted by thin lines beneath each probe. (**B**) RNA samples were analyzed by RNase protection using an antisense RNA probe generated from Gt.5'. The RNA samples used are prepared from: testis (T), liver (L), polysomal testicular RNA (P) and non-polysomal testicular RNA (NP). The lane 'PO' represents a control in which no RNA samples were used in the annealing reaction. The undigested probe is indicated by 'UP'. The lanes S.L. represent the sequencing ladder of an unrelated DNA used for size determination. (**C**) RNA samples were analyzed by RNase protection using the antisense genomic probe KC2-4A3. The RNA samples used are prepared from: liver (L), kidney (K), testis (T), non-polysomal testicular RNA (NP) and polysomal testicular RNA (P). The lane 'PO' represents a control in which no RNA samples were added to anneal with the antisense probe. The sizes of the protected fragments are indicated in nucleotides besides the blots. They were determined by coelectrophoresis with an unrelated DNA sequence on the gel (lanes labeled as S.L.).

Hybridization of mRNA fragments produced by RNase H cleavage at the oligo ST site with a probe (Gt.W) that recognizes regions 5' and 3' of the cleavage site reveals a prominent fragment of 384 nt and six additional fragments with sizes ranging from about 190 to 260 nucleotides (Fig. 2A, panel ST). The 384 nt fragment differs from the 390 nt band described above. Although all six fragments are present in the RNA samples obtained from total testis and the non-polysomal testicular fractions, the two longest fragments cannot be detected in the polysomal testicular mRNAs (Fig. 2A and B, lane P of ST panel).

To determine whether the cytochrome c_T mRNA fragments are located 5' or 3' of the oligo ST site, the radioactivity was removed from the Northern blot previously hybridized with the Gt.W probe (panels dT and ST of Fig. 2A) and the blot was rehybridized with probe Gt.5' which specifically recognizes the 5' region of cytochrome c_T mRNA (see Fig. 1). The 384 nt fragment detected with the antisense coding region probe generated from pGt.W is not detected with the 5' region probe indicating that this RNA fragment is downstream of the oligo ST site (Fig. 2B). The six smaller fragments are all seen with this 5' region probe (Fig. 2B, panel ST) suggesting that the cytochrome c_T mRNAs contain 5' UTRs of variable lengths. Cleavage of cytochrome c_T transcripts at the oligo ST site reveals an approximately 70 nt difference in length between the shortest and the longest of the six cytochrome c_T mRNAs (the length difference is actually 59 nt as shown by RNase protection assay, see below). The mRNAs which have the four shorter 5' UTRs are detected in the total, non-polysomal, and polysomal fractions of testicular RNA (Fig. 2B, panel ST). In contrast, the two transcripts with the longest 5' UTR lengths are not loaded onto polysomes suggesting that some features of their longer 5' UTRs may hinder polysomal loading (see below).

Ribonuclease protection analysis of the 5' UTRs of the cytochrome c_T transcripts

The heterogeneity in length of the 5'UTRs of the mouse cytochrome c_T mRNAs can be explained by differential processing and splicing of the cytochrome c_T mRNAs or by the

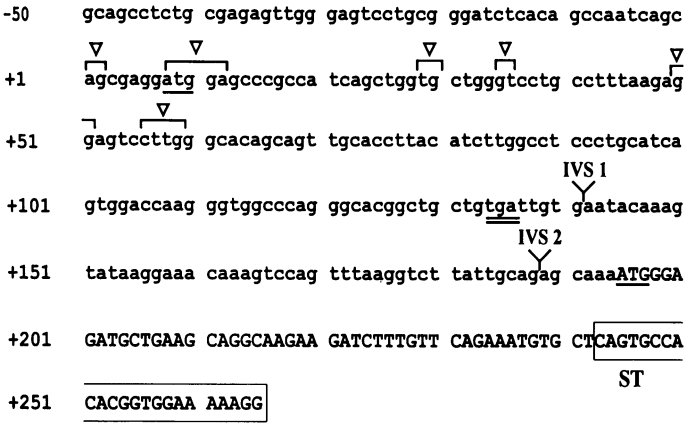

Figure 4. Nucleotide sequence of part of the mouse cytochrome c_T gene showing the multiple start sites of cytochrome c_T. The 5′ flanking sequence and 5′ UTR are shown in lower case letters while the coding region up to the site of oligonucleotide ST is shown in upper case letters. The positions of the transcriptional start sites detected by RNase protection assays are indicated by brackets and inverted triangles and numbers are assigned to the nucleotide positions by taking the first major initiation site as +1. The sequence complementary to oligonucleotides ST is boxed. The initiation codons which delimit the start of the upstream open reading frame (atg) and the coding region (ATG) are underlined, while the in-frame termination codon for the upstream ORF is double-underlined. IVS1 and IVS2 indicate the positions of intron 1 and intron 2, respectively.

utilization of multiple transcriptional start sites. To test the first possibility, the 138 nt probe Gt.5′ corresponding to the cytochrome c_T sequence from +82 to +220 which terminates 22 bp upstream of oligo ST (Fig. 1), was used in RNase protection assays. The region of cytochrome c_T mRNA covered by this probe includes the sites where the first two intervening sequences are spliced out of the primary transcript (Fig. 3, unpublished data). These introns could contain sequences which are alternatively spliced into cytochrome c_T mRNAs, thereby introducing size differences into the mRNA population.

The radiolabeled antisense RNA probe generated from linearized pGt.5′ was protected over its entire length of 138 nt by RNA samples from total testis, polysomal, and non-polysomal testicular RNAs (Fig. 3B, lanes T, P and NP). This indicates that the region between +82 and +220 is uninterrupted in the six cytochrome c_T transcripts. No protected fragment was detected in liver RNA, consistent with the lack of expression of cytochrome c_T in liver (Fig. 3B, lane L). We conclude from these data that alternative splicing in this region does not contribute to the size heterogeneity of the cytochrome c_T transcripts.

Cytochrome c_T utilizes multiple transcriptional start sites

To determine whether multiple transcriptional start sites contribute to mouse cytochrome c_T mRNA heterogeneity, we performed RNase protection assay with a radiolabeled antisense RNA probe generated from the genomic subclone KC2-4A3. This 301-nt probe contains sequences complementary to the entire first exon of 141 bp, 27 bp of the first intron, as well as 133 bp of the 5′ flanking region of the mouse cytochrome c_T gene.

The RNase protection assay of testicular RNAs revealed a heterogenous set of transcripts in six major clusters with estimated UTR sizes ranging from 85 to 141 nucleotides (Fig. 3A and C). This result confirms the multiple fragments detected by the RNase H digestion analysis of cytochrome c_T transcripts depicted above

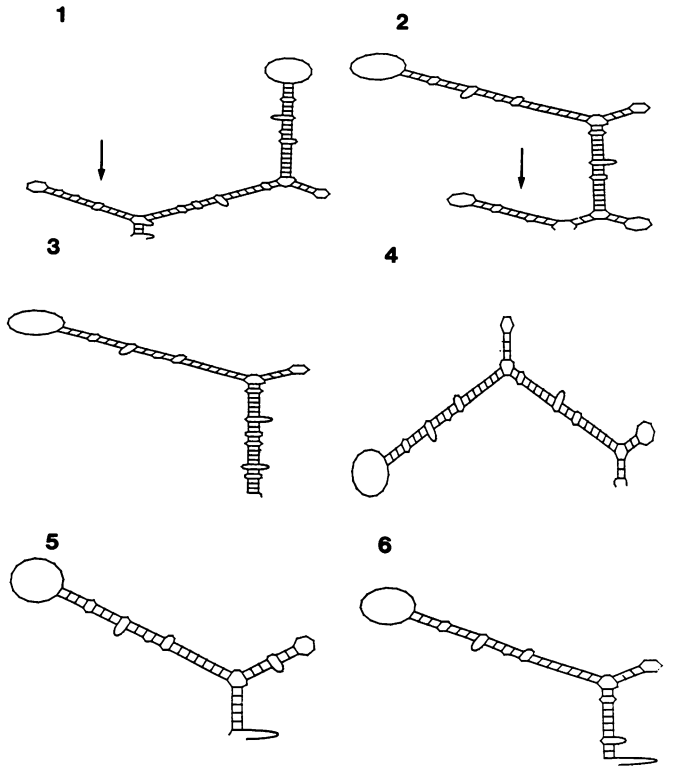

Figure 5. Predicted secondary structures of the 5′UTRs of the six mouse cytochrome c_T mRNAs. The structures predicted for the 5′UTRs initiated from the first to the sixth start sites of cytochrome c_T are numbered 1 to 6, respectively. These structures were determined by using the program FOLD designed by the University of Wisconsin Genetics Computer Group. Calculated free energy values for the structures 1 to 6 were −55.9, −53.4, −46.2, −44.6, −36.4 and −33.5 Kcal/mol, respectively. The arrows indicate the stem loop structures formed by the additional 5′UTR sequences present in the two longest cytochrome c_T transcripts.

(Fig. 2) and the broad band of cytochrome c_T mRNAs seen in Northern blot analysis of mouse testicular RNAs (4). No protected fragments are detected in the lanes containing liver and kidney RNA samples as cytochrome c_T is not transcribed in these tissues (Fig. 3C, lanes L and K). Comparison of the non-polysomal and polysomal fractions of testicular RNA (Fig. 3C, lanes NP and P) revealed that the cytochrome c_T mRNAs initiated at the first two start sites are primarily in the non-polysomal fraction. This confirms the results obtained in the RNase H digestion studies described earlier (Fig. 2). Initiation at the first two start sites (designated as +1 and +8 respectively, Fig. 3C) would produce cytochrome c_T transcripts with 5′UTRs that include an upstream initiator codon at +8 followed by an in-frame termination codon at +134 which is 58 nucleotides upstream of the open reading frame that encodes the cytochrome c_T protein. This upstream AUG is in a favorable context for initiation of translation, with an A three nucleotides before the AUG and a G immediately following the initiator codon, as predicted by the scanning model for translation (15). The multiple transcriptional start sites of cytochrome c_T are shown with the cytochrome c_T sequence in Figure 4.

Secondary structure analysis of mouse testicular cytochrome c

Cytochrome c_T transcripts initiated at the first and second transcriptional start sites are much more abundant in the non-

polysomal fraction than in the polysomal fraction of testicular RNA (Figs 2 and 3). Examination of the additional sequence present in these transcripts reveals that they contain an additional open reading frame (+8 to +134, Fig. 4). This type of upstream open reading frame has been described previously by Virbasius and Scarpulla (20) for both the rat and mouse cytochrome c_T mRNAs. From our genomic sequence and RNase protection assays, we have extended the sequence analysis of the 5′ UTRs of the multiple mouse cytochrome c_T transcripts and determined their predicted secondary structures and minimum free energies by the method of Zuker (26). We have found that apart from the similar structures shared by all the 5′ UTRs of cyt c_T mRNAs, the transcripts initiated from the first and second start sites each possesses an extra region with high G+C content that forms an additional putative hairpin structure (Fig. 5). The free energies of the two longest 5′ UTRs are predicted to be −55.9 and −53.4 Kcal/mol, respectively, producing more stable structures than the 5′ UTRs of the four smaller cytochrome c_T mRNAs.

DISCUSSION

The 5′ UTR sequences of cytochrome c_T mRNAs are heterogenous in length with a difference of 59 nt between the longest and the shortest of the six transcripts. Using RNase H digestion analysis and RNase protection assays of mouse cytochrome c_T RNAs with RNA probes generated from mouse cytochrome c_T cDNA and genomic subclones, we have established that the multiple transcripts result from the utilization of six different transcriptional start sites. This observation is consistent with the broad band detected in Northern blots of mouse testicular RNA (4). No evidence of differential splicing has been detected in the 5′UTR region of mouse cytochrome c_T mRNAs.

Neither the rat nor the mouse cytochrome c_T promoter contains a TATA box (20, our unpublished data), which is typically located about 25 nucleotides upstream from the transcriptional start site. The absence of a TATA box in a promoter is often correlated with multiple transcriptional start sites. In the majority of genes transcribed by RNA polymerase II, the TATA element binds TFIID complex containing TATA box-binding protein (TBP) and specifies the transcriptional initiation site. The promoters of genes such as farnesyl pyrophosphate synthetase (3), plasminogen activator (27), proenkephalin (28), androgen receptor (5) and ADP-ribosylation factor 3 (29) also lack the TATA sequence.

TATA-less promoters have been divided into two categories: promoters that are not G+C rich and initiate transcription from a single site; and promoters that are G+C rich and initiate transcription from multiple sites (30). A pyrimidine-rich initiator element which overlaps with the initiation site has been characterized for the first class of promoter (30). Eukaryotic genes which have the second type of promoter usually encode proteins with housekeeping functions. A tethering factor physically associated with TBP has been proposed to interact with activators, such as SP1, to recruit TFIID complex for efficient initiation of transcription (31). The cytochrome c_T promoter falls into the second class of promoters.

Multiple transcript sizes have been reported for several genes without TATA elements that are expressed in the testis. Rat farnesyl pyrophosphate (FPP) synthetase has different tissue-specific start sites in testis and liver that account for the extended 5′UTRs of testis FPP synthetase transcripts (3). Depending on the sites of transcriptional initiation, the 5′ UTR of FPP synthetase contains one or two upstream open reading frames. From S1 nuclease analysis, twelve major protected fragments were detected within the region −240 to −150 in the RNA from testis; while four major bands which spanned the region from −129 to −126 were seen in the RNA from liver. It has been suggested that FPP synthetase transcription is controlled by two different promoters, one of which functions in the testis, while the other functions in liver and other somatic tissues such as kidney, brain and adrenal gland. Likewise, pachytene spermatocyte transcripts of the mouse metallothionein-I gene are transcribed from a TATA-independent promoter utilizing different start sites spanning 160 bp from the first to the last (32). However, the predominant metallothionein-I transcripts synthesized in somatic tissues initiate with TATA-dependent transcription. The promoter of the human androgen receptor (hAR) gene also lacks a TATA box and utilizes two major initiation sites (5). It has a short GC box (−59 to −32), a long homopurine stretch (−117 to −60), and utilization of the two start sites is controlled by different promoter sequences. The ubiquitous transcription factor Sp1 interacts with the GC box and directs initiation from one start site. Utilization of the other initiation site is dependent upon sequences located between −5 and +57 in the hAR gene. The rat and mouse proenkephalin genes also have two distinct promoters, one of which is utilized exclusively in spermatogenic cells. The testis-specific transcripts are initiated downstream from the somatic promoter in the first somatic intron (intron A_S) where a consensus binding site for Sp1 was identified. The proenkephalin germ cell-specific promoter is G+C rich, does not have a TATA element, and has multiple start sites that span a region of about 30 bp (28). The transcriptional efficiency of the germ cell-specific mRNAs is lower than the somatic transcripts, a likely result of the presence of four short upstream ORFs within the 5′UTR (33). Similar to the mouse proenkephalin gene, the cytochrome c_S gene shows a differential utilization of two promoters during spermatogenesis in the mouse (34). Four somatic cytochrome c_S mRNAs of 1.3, 1.1, and 0.7−0.5 kb are transcribed from the somatic promoter in spermatogenic stem cells through early meiotic cells, while a larger transcript of 1.7 kb is initiated from an alternative putative testis-specific promoter which is located upstream of the somatic one. In contrast to the somatic transcripts, this 1.7 kb mRNA is primarily non-polysomal. It has over 1 kb of additional 5′UTR which is absent in the somatic transcripts of cytochrome c_S. The gene encoding β-galactosyltransferase similarly utilizes an alternative upstream promoter in the testis producing an apparently testis-specific transcript which contains an additional 650 nucleotides in its 5′ terminus (35).

Post-transcriptional regulation of gene expression has become widely appreciated as an important step in controlling the level of protein synthesis. A scanning model for translation has been proposed which states that the 40S ribosomal subunit, carrying Met-$tRNA_i^{met}$ and various initiation factors, scans an mRNA from its 5′ end and stops at the first AUG codon situated in a favorable context where it initiates translation (15). Translational efficiency of mRNAs is heavily influenced by structural elements such as additional open reading frames and hairpin loop structures in their 5′ UTRs. The scanning hypothesis predicts that these features in the 5′UTR can impair mRNA translation by blocking the scanning process. There are numerous examples showing the importance of the 5′UTR in post-transcriptional regulation. In human transforming growth factor-β1 (TGF-β1) mRNA, a

portion of the 5′ UTR spanning the region +11 to +147 contains a stable stem-loop structure that inhibits translation by as much as 22-fold (36). Likewise, the expression of the Na,K-ATPase α and β subunits is under translational control (37). Deletion of the 5′UTR of the α_1 mRNA significantly enhances the efficiency of its translation. This region is G+C rich and forms a stable secondary structure. It was also shown to impair translation when attached to the coding sequence of the β_1 mRNA. Analysis of cellular iron homeostasis in mammalian cells shows a close relationship between mRNA secondary structure and post-transcriptional control of mRNA. The maintenance of iron levels in cells is controlled by the coordinate regulation of transferrin receptor and ferritin (38), proteins which function in the transport and storage of iron, respectively. When cells need more iron, the levels of transferrin receptor increase while ferritin levels decrease. Conversely, when cells have enough iron, more ferritin and less transferrin receptors are produced. Stem-loop structures known as iron-responsive elements (IREs) exist in the ferritin and transferrin receptor mRNAs. One copy of IRE is found near the 5′ end of the 5′UTR of all known ferritin mRNAs, whereas five IREs have been detected in the 3′UTRs of transferrin receptor transcripts (7). Regulation of ferritin translation occurs at the level of translation initiation. Binding of high affinity iron-responsive element binding protein (IRE-BP) to the IRE at the 5′UTR of ferritin mRNA inhibits its translation (39,40). It has been shown that efficient inhibition of translation of ferritin transcripts requires the IRE to be present within 50−60 nucleotides from the 5′ end of the mRNA (41). In contrast to their function in ferritin mRNAs, the IREs in transferrin receptor transcripts regulate their stability. Two distinct functions have been found within the 3′UTR: IREs and an iron independent instability determinant(s) (38,42). The binding of IRE-BP to the IREs renders the cells unable to detect or to utilize the transferrin receptor instability element, thus protecting the transferrin receptor transcripts from degradation.

In addition to forming highly stable RNA secondary structures, upstream ORFs in 5′UTRs also affect translational efficiency. In macaque erythrocyte carbonic anhydrase I mRNA, a C to G substitution twelve nucleotides downstream from the cap site creates an ORF that terminates 6 nucleotides before the normal translational start site. The presence of this upstream ORF greatly reduces the translational efficiency of carbonic anhydrase I mRNAs in erythrocytes (43). A similar phenomenon has also been demonstrated with transforming growth factor-β3 (TGF-β3). TGF-β3 mRNA has a single transcriptional start site and its 5′UTR is about 1.1 kb and contains 11 upstream ORFs (44). *In vitro* translation of the TGF-β3 precursor coding sequence as well as a construct with the TGF-β3 5′UTR placed upstream of a chloramphenicol acetyltransferase coding sequence, revealed that this 5′ UTR significantly inhibits translation. When the 640 nucleotides from the 5′ end including eight of the eleven upstream ORFs were deleted, much, but not all, of the inhibitory effect of the 5′UTR was relieved. Increased translation was also observed when the sequences of the two upstream initiation codons nearest to the translational start site of the TGF-β3 mRNA were altered.

Ornithine decarboxylase (ODC), a key enzyme involved in polyamine biosynthesis, is an excellent example of another mRNA whose 5′ UTR markedly affects translation. Based on the available values of stacking and loop-destabilizing energy of RNAs (26), the 130 nucleotides at the 5′ terminus of the rat ODC 5′UTR are predicted to form a stable stem-loop structure. ODC mRNAs containing this portion of the 5′UTR were found to be translated with a much reduced efficiency compared to constructs lacking this region. Moreover, the 160 nucleotides at the 3′ terminus of the 5′UTR of ODC also contain an upstream open reading frame which has been demonstrated to inhibit translation by 50−65% (12). Similar to rat ODC transcripts, the two transcripts of mouse cytochrome c_T transcribed from the first two start sites are primarily non-polysomal and each contains an upstream open reading frame (from +8 to +134) which ends 58 nucleotides upstream of the cytochrome c_T reading frame that begins at +195. Although this upstream ORF encodes a peptide of 42 amino acids, no significant homology was found between this amino acid sequence and that of other proteins in the National Biomedical Research Foundation protein database. Both the upstream initiator AUG and the start codon for the cytochrome c_T reading frame have an A three nucleotides before them and a G immediately following, making them strong start codons when compared to the consensus sequence for translational initiation codons (15). Although this additional upstream open reading frame may play a role in hindering the translation of the cytochrome c_T protein, the scanning hypothesis predicts that transcription does not merely initiate at it since there are fewer than 10 nucleotides between the upstream initiator codon and the 5′ terminus of the longest cytochrome c_T transcript (17). Ribosomes which scan from the 5′ end of the transcripts could translate this ORF and dissociate at its termination codon upstream to the start codon of the ORF that encodes the cytochrome c_T protein. This would hinder ribosomes from translating the functional protein. Based on the length of the 146 nt additional upstream ORF, one would expect one or two ribosomes to be engaged in translation at a time. The shorter mRNAs that are initiated from the other four start sites do not have this upstream ORF and appear to be translated. Moreover, extra stem loops that are G+C rich are predicted to form from the additional 5′UTR sequences which are present in the two longest cytochrome c_T transcripts but not the four shorter ones (Fig. 5). This hairpin structure could also interfere with translation by blocking ribosome scanning (15), or may interact with *trans*-acting factors which decrease the translational efficiency of the mRNA (45). In recent studies, a protein of about 58 Kd that bound specifically to a highly conserved region in the 5′UTR of rat ODC mRNA was suggested to regulate ODC mRNA translation (6), and a testicular protein of about 18 Kd represses the cell-free translation of protamine 2 mRNA by binding to its 3′UTR (46). At present, the role plays by *trans*-acting factors which interact with the 5′UTRs of cytochrome c_T mRNAs is unknown.

In summary, we have used RNase H digestion analysis combined with Northern blotting and ribonuclease protection assays to demonstrate that the heterogeneity of mouse cytochrome c_T mRNAs results from the transcription from six different initiation sites. Of the six transcripts, the cytochrome c_T mRNAs transcribed from the two most upstream initiation sites are primarily non-polysomal. The additional secondary structure and the presence of an open reading frame in the 5′UTRs of the two longest cytochrome c_T transcripts are likely to lead to reduced translational efficiencies of these mRNAs.

ACKNOWLEDGEMENTS

This work was supported by NIH grant HD 11878. Figures 2 and 3B are modified from figures in the thesis of Laura E.Hake.

REFERENCES

1. Hecht, N.B. (1986) In Rossant, J. and Pedersen, R. (ed.), Experimental Approaches to Mammalian Embryonic Development. Cambridge University Press, New York, pp. 151–193.
2. Hecht, N.B. (1993) In Desjardins, C. and Ewing, L. L. (ed.), Cell and Molecular Biology of the Testis. Oxford University Press, New York, pp. 400–432.
3. Teruya, J.H., Kutsunai, S.Y., Spear, D.H., Edwards, P.A. and Clarke, C. F. (1990) *Mol. Cell Biol.*, **10**, 2315–2326.
4. Hake, L.E., Alcivar, A.A. and Hecht N.B. (1990) *Development*, **110**, 249–257.
5. Faber, P.W., van Rooij, H.C.J., Schipper, H.J., Brinkmann, A.O. and Trapman, J. (1993) *J. Biol. Chem.*, **268**, 9296–9301.
6. Manzella, J.M. and Blackshear, P.J. (1992) *J. Biol. Chem.*, **267**, 7077–7082.
7. Klausner, R.D., Rouault, T.A. and Harford, J.B. (1993) *Cell*, **72**, 19–28.
8. Cox, T.C., Bawden, M.J., Martin, A. and May B.K. (1991) *EMBO J.*, **10**, 1891–1902.
9. Hennig B. (1975) *Eur. J. Biochem.*, **55**, 167–183.
10. Goldberg, E., Sberna, D., Wheat, T.E., Urbanski, G.J. and Margoliash, E. (1977) *Science*, **196**, 1010–1012.
11. Hake, L.E., Kuemmerle, N., Hecht, N.B. and Kozak C.A. (1993) *Genomics*, **20**, 503–505.
12. Manzella, J.M. and Blackshear, P.J. (1990) *J. Biol. Chem.*, **265**, 11817–11822.
13. Brown, P.H., Tiley, L.S. and Cullen, B.R. (1991) *Genes & Dev.*, **5**, 1277–1284.
14. Gallie, D.R. (1991) *Genes & Dev.*, **5**, 2108–2116.
15. Kozak, M. (1989) *J. Cell Biol.*, **108**, 229–241.
16. Kozak, M. (1992) *Crit. Rev. Biochem. Mol. Biol.*, **27**, 385–402.
17. Kozak, M. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8301–8305.
18. Kozak, M. (1991) *J. Biol. Chem.*, **266**, 19867–19870.
19. Kozak, M. (1991) *J. Cell Biol.*, **115**, 887–903.
20. Virbasius, J.V. and Scarpulla, R.C. (1988) *J. Biol. Chem.*, **263**, 6791–6796.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
22. Sherman, L., Levanon, D., Lieman-Hurwitz, J., Dafni, N. and Groner, Y. (1984) *Nucleic Acids Res.*, **12**, 9349–9365.
23. Childs, G., Maxson, R. and Kedes, L.H. (1979) *Devel. Biol.*, **73**, 153–173.
24. Church, G.M. and Gilbert, W. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 1991–1995.
25. Krieg, P.A. and Melton, D.A. (1987) *Methods Enzymol.*, **155**, 397–415.
26. Zuker, M. and Stiegler, P. (1981) *Nucleic Acids Res.*, **9**, 133–148.
27. Feng, P., Ohlsson, M. and Ny, T. (1990) *J. Biol. Chem.*, **265**, 2022–2027.
28. Kilpatrick, D.L., Zinn, S.A., Fitzgerald, M., Higuchi, H., Sabol, S.L. and Meyerhardt, J. (1990) *Mol. Cell Biol.*, **10**, 3717–3726.
29. Haun, R.S., Moss, J. and Vaughan, M. (1993) *J. Biol. Chem.*, **268**, 8793–8800.
30. Smale, S.T. and Baltimore, D. (1989) *Cell*, **57**, 103–113.
31. Pugh, B.F. and Tjian, R. (1991) *Genes & Dev.*, **5**, 1935–1945.
32. Garrity, P.A. and Wold, B.J. (1990) *Mol. Cell Biol.*, **10**, 5646–5654.
33. Roa, S.M. and Howells, R.D. (1993) *J. Biol. Chem.*, **268**, 22164–22169.
34. Hake, L.E. and Hecht, N.B. (1993) *J. Biol. Chem.*, **268**, 4788–4797.
35. Shaper, N.L., Wright, W.W. and Shaper, J.H. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 791–795.
36. Kim , S.-J., Park, K., Koeller, D., Kim, K.Y., Wakefield, L.M., Sporn, M.B. and Roberts, A.B. (1992) *J. Biol. Chem.*, **267**, 13702–13707.
37. Devarajan, P., Gilmore-Hebert, M. and Benz, E.J. (1992) *J. Biol. Chem.*, **267**, 22435–22439.
38. Harford, J.B. and Klausner, R.D. (1990) *Enzyme*, **44**, 28–41.
39. Walden, W.E., Daniels-McQueen, S., Brown, P.H., Gaffield, L., Russell, D.A. Bielser, D., Bailey, L.C. and Thach, R.E. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 9503–9507.
40. Goossen, B., Caughman, S.W., Harford, J.B., Klausner, R.D. and Hentze, M.W. (1990) *EMBO J.*, **9**, 4127–4133.
41. Goossen, B. and Hentze, M.W. (1992) *Mol. Cell Biol.*, **12**, 1959–1966.
42. Casey, J.L., Koeller, D.M., Ramin, V.C., Klausner, R.D. and Harford, J.B. (1989) *EMBO J.*, **8**, 3693–3699.
43. Bergenhem, N.C.H., Venta, P.J., Hopkins, P.J., Kim, H.J. and Tashian, R.E. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8798–8802.
44. Arrick, B.A., Lee, A.L., Grendell, R. L. and Derynck, R. (1991) *Mol. Cell Biol.*, **11**, 4306–4313.
45. Wang, Y.-H., Sczekan, S.R. and Theil, E.C. (1990) *Nucleic Acids Res.*, **18**, 4463–4468.
46. Kwon, Y.K. and Hecht, N.B. (1993) *Mol. Cell Biol.*, **13**, 6547–6557.