Cloning and characterization of a naturally occurring antisense RNA to human thymidylate synthase mRNA

Bruce J.Dolnick

Department of Experimental Therapeutics and The Grace Cancer Drug Center, Roswell Park Cancer Institute, Buffalo, NY 14263, USA

Received February 9, 1993; Accepted March 11, 1993

EMBL accession no. X67098

ABSTRACT

Based upon reverse transcription and polymerase chain reaction results with human KB cell RNA, a cDNA (i.e., 3'rTS1, 1557 nt) with complementarity to thymidylate synthase mRNA was cloned and sequenced. Northern blot analysis showed that 3'rTS1 corresponded to a cytoplasmic 1.8 kb RNA found in several tumor cell lines. The remaining 5' region of this antisense RNA was cloned by a RACE (Rapid Amplification of cDNA Ends) procedure. A full length cDNA (i.e., rTS, 1811 nt) was generated by splicing 3'rTS1 with RACE-generated cDNA. rTS RNA is likely a mRNA that contains four open reading frames. Based upon sequence analysis of the RACE cDNAs and the rTS cDNA, rTS RNA is likely processed from a gene containing at least six introns. Northern blot analysis indicates rTS RNA is expressed in a variety of human tumor cell lines and an aberrant form is expressed in a methotrexate-cell line.

INTRODUCTION

Thymidylate synthase (TS) is the enzyme responsible for the *de novo* synthesis of deoxythymidylate. TS is present in proliferating cell types and is considered an important target in cancer chemotherapy. During the establishment of a quantitative polymerase chain reaction (PCR) assay for human TS mRNA, preliminary evidence indicated the existence of RNA complementary to TS mRNA (1-3). In recent years a number of antisense transcripts to known genes have been described (4-15). However, there are few instances in which an antisense RNA has been associated with the features common to mRNAs such as polyadenylation, splicing and translation (4,8,11,14,15). This report describes the cloning and characterization of an antisense RNA from human cells that arises from DNA organized in a canonical gene structure. Analysis of sequence and expression data suggests this RNA it is a functional mRNA.

METHODS

Tissue culture and RNA manipulations

KB cells (ATCC CCL 17, human oral epidermoid carcinoma), were maintained in RPMI 1640 culture medium supplemented with

5% dialyzed horse serum at 37° in a humidified atmosphere of 95:5% air.CO2. K562 cells (ATCC CCL 243, human chronic myelogenous leukemia were maintained as suspension cultures in RPMI 1640 culture medium supplemented with 10% fetal bovine serum. B1A is a methotrexate-resistant subclone of K562 that overproduces dihydrofolate reductase approximately 200-fold that of the parental K562 cell line was obtained by dilution cloning from cells provided by Dr Bertino (16). FaDu (ATCC HTB 43, human squamous cell carcinoma), A-253 (ATCC HTB 41, human epidermoid carcinoma), MCF7 (ATCC HTB 22, human breast adenocarcinoma), SE (human renal cell carcinoma, ref 13), A2780 (human ovarian carcinoma) and HCT-8 (ATCC CCL 244, human ileocecal adenocarcinoma) were provided by Drs. McGuire and Rustum of this Department. Cells were extracted for RNA when approximately 40-70% confluent. RNA was prepared either from whole cells using guanidinium thiocyanate extraction or from cytoplasmic or nuclear enriched subcellular fractions (17). All RNAs were subjected to a final step of CsCl gradient purification as previously described (17). The DNA content of RNA preparations was determined by measuring the fluorescence resulting from the DNA intercalating agent Hoechst 33258 using a TKO fluorometer (Hoeffer Instruments, San Francisco, CA.) as recommended by the manufacturer. The DNA content of different RNA preparations ranged from 0 to 0.025 µg DNA per μ g RNA (average 1.08 ± 1.23%). Denaturing gel electrophoresis in formaldehyde-containing agarose gels, Northern transfers to nylon membranes (Gene Screen, Amersham, Arlington Heights, IL) and hybridization were performed by standard procedures unless otherwise indicated (18). RNA samples for gel electrophoresis were denatured in the presence of ethidium bromide to allow inspection of rRNA integrity and an estimation of gel loading (19). Gels for Northern blotting were treated with NaOH prior to transfer as described (18).

Plasmids and probes

The plasmid used for the generation of single stranded sense and antisense RNA probes, and for double stranded TS-specific cDNA probes was pTS-PCRmut1. This plasmid is comprised of a human TS mRNA region corresponding to bases 351-950 (with respect to the initiation codon), mutated to contain a Smal site and cloned into the EcoRI/PstI site of pGEM1 (2). Linearization with HindIII was used in the preparation of TS

^{*} To whom correspondence should be addressed

'sense' RNA transcripts with SP6 RNA polymerase. Linearization with EcoRI was used in the preparation of TS 'antisense' RNA transcripts with T7 RNA polymerase. Double stranded ³²P-probes were generated using either the PstI fragment of 3'rTS1 DNA (see below) or the EcoRI/PstI fragment of pTS-PCRmut1, after radiolabelling with a multiprime labelling kit as recommended from the manufacturer (Amersham, Arlington Heights, IL).

Oligonucleotides and PCR

Oligonucleotides were synthesized on an Applied Biosystems 381A, DNA synthesizer and purified by either butanol precipitation (20) or reverse phase HPLC (2). PCR and PCR subsequent to reverse transcription (RT-PCR) were performed essentially as described (2), unless otherwise indicated. Reverse transcription reactions contained 0.5 μ M oligo #258 when RT-PCR analysis was performed for 'antisense' TS RNA and oligo #259 when RT-PCR was performed for TS mRNA. PCR reactions for TS or 'antisense' TS contained 0.5 μ M each primer. Primers #258 and #259 span introns III-VI of the human TS gene and PCR of genomic DNA with these primers does not yield amplified cDNA products (2). Primers for PCR and RT-PCR have the composition: #258, TTTGGACAGCCTGGGATTTCTC (sense with respect to TS mRNA, position 351-371); #259, AAA-GCACCCTAAACAGCCATT (antisense with respect to TS mRNA, position 930-950).

cDNA isolation

An oligo(dT)-primed cDNA library was constructed in lambda phage from nuclear poly(A⁺) RNA using a ZAP-cDNA synthesis and cloning kit obtained from Stratagene (La Jolla, CA), as recommended by the manufacturer. Nuclear poly(A⁺)RNA was used as the RNA source based upon preliminary Northern blot analysis with sense and antisense TS RNA probes which suggested a relative enrichment of TS antisense RNAs in the nuclear fraction of KB cells (1). The protocol used employs an oligo(dT)-XhoI linker in the reverse transcription step to ensure the generation of a unidirectional cDNA library such that the poly(A⁺) tail of any mRNAs are positioned next to an XhoI site in the cloning vector. Approximately 10⁶ independent pfus were screened for homology to TS RNA with a ³²P-labelled human TS cDNA probe. Positive clones were isolated, TS homology verified by a second round of screening and cDNA-containing plasmids generated in the vector pBluescript SK- by in vivo excision as recommended by the manufacturer. DNA sequencing was performed with supercoiled plasmid DNA (isolated by banding in ethidium bromide-CsCl gradients generated by ultracentrifugation) with the Sequenase version 2.0 sequencing kit as recommended by the manufacturer (United States Biochemical, Cleveland, OH). Oligonucleotide primers used for sequencing (sequences available upon request) were prepared as described for reverse transcription and PCR.

RACE

A RACE (Rapid Amplification of cDNA Ends) procedure (21) to generate the 5'end of rTS cDNA (see below) was performed using a 5'RACE kit obtained from GIBCO/BRL and was used as recommended by the manufacturer. For the reverse transcription of cellular RNA, primer #342 (TTCAGCGCC-TGGCACA) was prepared based upon the structure of 3'rTS1 cDNA (position 441-446 of rTS cDNA, see Results). Control

reactions for the RACE procedure consisted of replicate RNA samples processed identically to those which yielded cDNA clones except reverse transcriptase was omitted. A second set of control reactions consisted of cDNAs which were not dC tailed but otherwise processed identically to samples which gave rise to the RACE clones described. After reverse transcription and dC tailing, cDNA was generated using the universal amplification primer supplied in the kit and primer #337 (GTATCC-TTGTGCCAGCATTT) which is complementary to the sense strand of rTS cDNA from position 362-381 (see Results). The specificity of the amplification reaction was then confirmed by Southern blot analysis of the PCR products using the nested primer #339 (ATTTCTAGGGCATCCTCCTCA). Control reactions yielded no cDNA products. The cDNAs generated by PCR were then cloned into pCRII and transformed into competent $INV\alpha F'$ using a TA cloning kit (Invitrogen, San Diego, CA) as recommended by the manufacturer.

Computerized data analysis

Nucleic acid and protein sequence analyses and database homology searches were performed using the DNAStar software program (Madison, WI). Database homology searches for proteins and nucleic acids were conducted using the DNAstar Geneman program. Protein pattern similarity scores and alignments were performed using the DNAstar Patscan program. Databases searched for sequence homologies include GenBank release 73, EMBL release 32, and the PIR protein sequence database release 34.0.

RESULTS

Cloning and characterization of 3'rTS1 cDNA

A quantitative RT-PCR assay for human TS mRNA was established that employed an oligonucleotide (i.e., #259) complementary to TS mRNA in the reverse transcription step (2). As a control reaction, the upstream (i.e., sense, #258) oligonucleotide primer was employed as the primer for reverse transcription. The subsequent PCR reaction when both sense and antisense primers were present consistently yielded a discrete cDNA which had TS mRNA homology as ascertained by Southern blot analysis (1). Preliminary Northern blot analysis of KB RNA, using single stranded ³²P-RNA probes indicated the presence of RNA complementary to TS mRNA (1). Since Northern blot data also suggested an abundance of TS complementary RNA in the nuclear RNA fraction (data not shown), a nuclear cDNA library was prepared in an attempt to optimize the antisense:sense TS cDNA ratio in the resulting library.

A unidirectional cDNA library was prepared from KB nuclear poly(A^+)RNA as described. Twenty seven clones were identified as having TS cDNA homology by filter hybridization of phage plaques. cDNA clones corresponding to an antisense orientation with respect to TS mRNA were identified by PCR utilizing an oligonucleotide complementary to the T7 RNA polymerase promoter region located downstream of the oligo(dT)-XhoI cloning site and oligonucleotides with either a sense or antisense orientation with respect to TS mRNA (i.e. #258 and #259, respectively). One cDNA clone appeared to be in the 'antisense' orientation and this cDNA was chosen for further characterization.

The antisense cDNA clone containing TS cDNA homology was sequenced and found to contain 1557 nt and a poly (A) tail



Figure 1. Relation of 3'rTS1 to the TS gene. The relation of 3'rTS1 to the TS gene is shown schematically. The upper portion of the figure shows the 3' end of the TS gene including the last intron (VI) and exon 7. The termination site for TS mRNA translation is indicated (TAG) as are the two known polyadenylation signals (inverted arrows). The sense strand regions of the TS gene corresponding to the predicted antisense introns are indicated by dashed lines and the dinucleotides on each side of the predicted splice junctions. The 3'rTS1 transcript is presented in the opposite orientation at the bottom of the figure with the polyadenylation signal indicated by an arrow and the splice junction dinucleotide sequences for each predicted excised intron. The lengths of all predicted exons and excised introns are indicated numerically.

(see below). A search of the NIH GeneBank and EMBL Data Library for related sequences revealed homology only to the antisense strand of the human TS gene and its 3' flanking region (22). Analysis of this homology by dot plot indicated that the 3'end of the antisense cDNA clone was essentially identical to the antisense strand of the 3'end of the human TS gene (data not shown). A schematic representation of the relation of this cDNA to the TS gene is shown in Fig 1. To signify the homology of the antisense strand to the 3' region of the human TS gene, this clone has been designated 3'rTS1. While the sequence of 3'rTS1 is unrelated to the TS gene 5' of base 708, the 3'-most 849 bases of 3'rTS1 cDNA are 99.5% identical to the antisense strand of the 3' end of the human TS gene, with the exception of two discontinuities. The region of complementarity to the TS gene includes all of TS exon 7 (633 bases, including 138 bases of the protein coding region) and extends 303 bases into intron 6. In the region of complementarity to the TS gene, the 3'rTS1 cDNA displays two discontinuities between nucleotides 789-790 and 1036-1037 (Fig 1). Examination of the antisense strand regions of the TS gene that correspond to each discontinuity suggests consensus splice donor and acceptor sites and the polypyrimidine tracts typical of eukaryotic introns were removed to generate the RNA species corresponding to 3'rTS1 (23). 3'rTS1 contains the polyadenylation signal ATTAAA that is found in approximately 12% of eukaryotic mRNAs (24, see Fig 4). This, and the fact that 3'rTS1 cDNA arose from an oligo(dT)primed RNA, suggest that 3'rTS1 cDNA derives from a spliced and polyadenylated product of a transcript initiating downstream of the TS gene and terminating within the last intron of the TS gene.

Expression of 3'rTS1 RNA

PstI restriction of 3'rTS1 cDNA liberates a 456 bp fragment (nt 326-782 of rTS, see below) which does not overlap the transcribed region of the TS gene. Northern analysis was performed with this fragment and the pattern of 3'rTS1 and TS RNAs were compared in the poly(A⁺)and poly(A⁻)RNA from fractionated KB cells (Fig 2). The 3'rTS1 RNA specific probe



Figure 2. Northern blot analysis of 3'rTS1-homologous RNA in KB cells. RNA preparation, gel electrophoresis (1.2% agarose) and Northern blotting were performed as described (see Materials and Methods). Samples were loaded in duplicate wells for later probing with either 3'rTS or TS RNA. Lanes contained the following: lanes 1 and 10, 30 μ g cytoplasmic RNA, lanes 2 and 11, 30 μ g cytoplasmic poly(A⁻) RNA, lanes 3 and 12, approximately 1 μ g cytoplasmic poly(A⁺) RNA, lanes 4 and 13, 30 μ g nuclear RNA, lanes 5 and 14, 30 μ g nuclear poly(A⁺) RNA, lanes 6 and 15, approximately 1 μ g nuclear poly(A⁺) RNA, lanes 6 and 15, approximately 1 μ g nuclear poly(A⁺) RNA, lanes 7 and 16 are blank and lanes 8 and 17 contain ³²P-lambda DNA digested with HindIII as gel transfer and molecular weight markers (sizes in kbp are indicated at right). Brackets indicate areas of interest for 3'rTS1-homologous RNA. Prior to probing, the filter was cut longitudinally and the side corresponding to lanes 1–8 probed with TS ³²P-cDNA (see text for details).

reveals the presence of a cytoplasmic 3'rTS1 RNA approximately 1.8 kb in length. Nuclear RNA contains a discrete RNA species approximately the same size as 28s rRNA (6.3 kb) and an abundance of higher molecular weight RNA with a heterogeneous size distribution. An examination of 3'rTS1 RNA expression in several other cell lines (Fig 3) indicates similar patterns of expression to that found in KB cells, although the abundance of the high molecular weight RNA appears to vary widely between cell lines (e.g., Fig 3, compare results from FaDu and K562 cells). The methotrexate-resistant cell line K562 B1A shows elevated levels of the nuclear 6.6 kb RNA species with 3'rTS1 homology as well as a species of approximately 600 nt in length. This pattern of elevated and altered expression was not found in another methotrexate-resistant cell line (KB1BT) which overexpresses dihydrofolate reductase approximately 250-fold (2,17 and data not shown). Levels of RNA species with 3'rTS1 homology do not appear to correlate with levels of the (sense) TS mRNA in different cell lines (e.g., Fig 3, compare SE and FaDu). 3'rTS1 RNA has also been found in normal human skin cell lines as determined by nuclease mapping (A. Black, personal communication).

RACE and characterization of rTS RNA

A comparison of the 3'rTS1 sequence data (1557 nt) with the Northern blot data (Figs 2,3) indicated 200-300 bases from the 5'-end of the 1.8 kb cytoplasmic RNA species was absent from the 3'rTS1 cDNA clone. A RACE procedure was performed in an attempt to isolate cDNA corresponding to this 5' region. FaDu RNA was used as a template for RT-PCR due to the abundance of 3'rTS1 homologous RNA (Fig 3). RACE was performed as

described and cDNA clones identified by PCR and Southern blotting of cDNA inserts using a nested oligonucleotide probe (see Methods). The cDNAs obtained varied in size, but ranged from smaller to larger than expected for the remainder of the 1.8 kb cytoplasmic rTS RNA (data not shown, see text). While



Figure 3. Expression of 3'rTS1-homologous RNA in various tumor cell lines. Cell lines from various tumors of origin were extracted for RNA as described. Duplicate wells of a denaturing agarose gel were loaded with 15 μ g RNA from the indicated cell lines. After electrophoresis and blotting the gel was sliced longitudinally and probed for either 3'rTS1-homologous RNA (panel A) or TS RNA (panel B) as described. Only certain samples of the original gel are shown.

smaller than expected cDNAs would be expected to arise from premature termination of cDNA synthesis during reverse transcription, the reason for longer cDNAs was not obvious. Six cDNA clones of various sizes were chosen for further characterization. Two clones (i.e., #1 and 6) were found to be identical and to correspond in size to the missing 5'end of the 3'rTS1 cDNA (see Figs. 4 and 5). Based upon the identity of these two clones, their identity with clone #2 which appears to have arisen from a prematurely terminated reverse transcription reaction, and the overlap of all three clones with 3'rTS, it was concluded these likely arose from the 1.8 kb cytoplasmic RNA species detected by Northern blot analysis. One of these two identical cDNAs (clone #6) was spliced to the 3'rTS1 cDNA at a unique EcoN1 site (nt 297, Fig 4) to generate a cDNA of the same length as the cytoplasmic rTS RNA. This cDNA is designated 'rTS' to signify that it corresponds in size to the full length cytoplasmic RNA species. Three other clones (i.e., #8, 13 and 18) contained regions not present in rTS. The composition of these rTS divergent regions have characteristics associated with introns (see below). Because of the presence of apparently spliced rTS species present in the cytoplasm, and other characteristic features of mRNAs (i.e., polyadenylation, large open reading frames, see below), it is concluded rTS is likely a mRNA. Part of the rTS cDNA sequence, indicating a number of significant features is presented in Fig 4.

The rTS mRNA contains four large open reading frames (ORFs). The first two open reading frames ORF1 (position 18-290, 90 aa codons) and ORF1A (position 67-432, 121 aa codons) overlap extensively (Fig 4). In addition there are two more large ORFs extending from positions 509-844 (ORF2, 111 aa codons) and from positions 844-1146 (ORF3, 100 aa codons). A search of the PIR protein database indicates the proteins potentially encoded by these ORFs have not been described. Preliminary data indicate rTS RNA generated by in

r	T	S

1	gccacggcgc	ggacgcc atg	cacacggacc	ctgactactc	ggctgcctat	gtcgtcatag	aaactg atg c	agaagatgga	80
81	atcaaggggt	gtggaattac	cttcactctg	ggaaaaggca	ctgaagttga	ttggtccaga	aaagggcgtg	gtgcacctgg	160
161	cgacagcggc	cgtcctaaac	gcggtgtggg	acttgtgggc	caagcaggag	ggaaagcctg	tctggaagtt	acttgtggac	240
241	atggatccca	ggatgctggt	atcctgcata	gatttcaggt	acatcac tga	tgtcctgact	gaggaggatg	ccctagaaat	320
321	actgcagaaa	ggtcaaattg	gtaaaaaaga	aagagagaag	caaatgctgg	cacaaggata	cctgcttaca	cgacatcgtg	400
401	cgcctggctg	gggtactcag	atgacacgtt	ga agcagcts	tgtgccaggc	gctgaaggat	ggctggacca	ggtttaaagt	480
481	aaaggtgggt	gctgatctcc	aggatgac at	gcgaagatgc	caaatcatcc	gagacatgat	tggaccggaa	aagactttga	560
561	tgatggatgc	caaccagcgc	tgggatgtgc	ctgaggcggt	ggagtggatg	tccaagctgg	ccaagttcaa	gccattgtgg	640
641	attgaggagc	caacctcccc	tgatgacatt	ctgggcacgc	caccatttcc	aaggcactgg	tccattagga	attgcattgc	720
721	cacaggagaa	cagtgccaca	atagagtgat	atttaagcaa	ctcctacagc	gaaggccctg	cagttcctcc	agattgacag	800
801	ttgcagactg	ggcagtgtca	atgagaacct	ctcagtattg	c tgatg gcca	aaaagtttga	aattcctgtt	tgcccccatg	880
881	ctggtggagt	tggcctctgt	gaactggtgc	agcacctgat	tatatttgac	tacatatcag	tttctgcaag	ccttgaaaat	960
961	agggtgtgtgtg	agtatgttga	ccacctgcat	gagcatttca	agtatcccgt	gatgatccag	cgggcttcct	acatgcctcc	1040
1041	<u>caaggatccc</u>	ggctactcaa	<u>cagaaatgaa</u>	ggaggaatct	gtaaagaaac	accagtatcc	agatggtgaa	gtttggaaga	1120
1121	<u>aactccttcc</u>	tgctcaagaa	<u>aattaagtgc</u>	tcagccccaa	caactttttt	ctttctgaag	tgaaagggct vī		1190
1311		<u>tcggcatcca</u>	gcccaacco	/ <u>********</u>	<u>KAAXAXXXXX</u>	/carrellext	<u>/XGGGGGGGGGGGGG</u>	•••	1370
1751		thtesatut	/aagtttttat	/sacementer	/ goessattes	/XXXXXXXXXX	/daatatedtt	<u>کم</u>	1812

Figure 4. Sequence of rTS. Both strands of rTS were sequenced in regions containing bases found to differ from that of TS (18). The region of overlap with the TS gene is signified by underlining. The region of rTS overlap with TS mRNA is signified by the crosshatched area. Start and stop codons for the four ORFs and the polyadenylation signal are highlighted by bold text. rTS cDNA has 57 A residues at the 3'end (not shown). The upper case letters (and numbers) indicate signals for the proposed excised introns and their deduced positions. 3'rTS1 cDNA is comprised of the region of rTS 3'-to nt 354.

vitro transcription and translated in either the rabbit reticulocyte lysate or wheat germ extract translation systems generates multiple translation products (B.J. Dolnick unpublished results).

Pre-mRNA is a significant component of cellular rTS RNA

As previously discussed, only three of the RACE-generated cDNAs are colinear with each other (i.e., clones #1, 6 and 2), while three of the RACE clones (i.e., #8, 13, and 18) were found to be alternatively divergent and coincident with rTS cDNA and with each other. Sequence analysis of the rTS divergent regions of clones #8, 13 and 18 suggests they correspond to unspliced introns, each having classical splice acceptor sites (i.e., a polypyrimidine tract and ag: motif at the 3'end of the intron). The sequences of the RACE cDNAs concluded to represent splice junctions and unspliced introns are presented in Fig 5. Since two of the three regions proposed to be unspliced introns are found in multiple clones (i.e. intron I in clones #8 and 18, intron IIIb in clones #13 and 18), it is likely these regions do not represent PCR-related artefacts. The region proposed as being the 3'end of intron II is only present in clone 13 but has a structure characteristic of a classical intron, with the flanking sequences on each side being identical to those in rTS and so has been concluded to be an intron. Intron III appears to have an internal splice acceptor site since only the 3'-most 113 nt of a 424 nt rTS-divergent region present in clone #13 (i.e., intron IIIb, Fig 5) are identical to the corresponding region of rTS divergence in clone #18, which contains ... GAG:gg... at the proposed exon 3: intron III junction. The same apparent intron in clone #13 contains the same sequence internally, but with another 313 nt at the 5' end that has a more conventional splice junction (i.e., ...,AG:gt...). The 113 nt sequence common to both clone #13 and 18 is not likely to arise from DNA contamination of the RNA used as a template in the RACE procedure as no cDNA clones were identified when the reverse transcription or cDNA tailing steps were omitted. Clone #18 may represent an alternatively spliced mRNA species or a misspliced RNA species. The presence of unspliced introns in half of the six RACE cDNAs examined likely reflects an abundance of unspliced or partially spliced RNAs in the cell line used to prepare RNA (i.e., FaDu,

rTS Exon 1 Exon 2 Exon 3GACAT: CCATCC CCCT: AGAA... Intron I Exon 2 Exon 3 Exon 4 CLONE 8 Intron I Exon 2 Exon 3 Exon 4 CLONE 18 Intron IIIb [:ggcc...t₄cacag:] 113 nt Intron II Exon 3 Exon 4 CLONE 13 (t),atttcttcc:AGAA....AGAG:[]:AGAAGCMMMTGKZTGKXMKMAGKXMK Intron IIIa Intron IIIb [:gtgg... ...(tg),cagggcc.. t,cacag:] 424 nt

Figure 5. Sequences of proposed intron:exon junctions from the 5'RACE clones. Proposed introns are aligned and are indicated by lower case letters and exons by upper case letters. Proposed splice junctions are indicated by colons (:). Exon regions found in clones are indicated by underlining. Base numbering is the same as rTS (Fig 4). The crosshatched bases indicate the position of oligo #337 utilized in the RACE protocol to generate the cDNA clones. see Fig 3). Based upon the published sequence of the human TS gene and sequence analysis of the RACE clones, it is concluded the rTS gene contains at least seven exons and six introns.

DISCUSSION

rTS is a previously unidentified gene which overlaps TS

In the present study, antisense RNAs complementary to TS mRNA, were cloned and characterized. A single cDNA clone, present in an antisense orientation relative to TS, was isolated and sequenced. This clone, 3'rTS1, has near identity (i.e., 99.5%) to the antisense strand at the 3'-end of the human TS gene (Fig 4), with the few isolated base descrepancies likely due to sequencing errors. When a fragment of this cDNA clone was used for Northern analysis, both discrete and heterogenous RNA species were detectable, including a cytoplasmic RNA of approximately 1.8 kb in length (Fig 2 and 3). The presence of both a 6.3 kb RNA and large amounts of heterogeneous high molecular weight rTS RNA species in the nuclear RNA fraction of several cell lines (Figs. 2 and 3) may be the result of inefficient pre-mRNA splicing for this gene contributing to the presence of multiple pre-mRNAs. It is highly unlikely the high molecular weight, heterogeneous rTS RNA detected by Northern analysis derives from DNA contamination of RNA preparations as the RNA used for these studies contains on the average only 1.08 \pm 1.23% DNA (see Methods), and DNA in this molecular weight region (i.e., ≥ 3 kb) would not be expected to transfer well under the conditions used for Northen blotting (i.e., gels were alklai treated to nick RNA but not processed to nick DNA). The interpretation that cells contain a substantial amount of rTS pre-mRNA is supported by the sequence data obtained from the RACE cDNA clones in which three of six clones sequenced contain structures consistent with unspliced introns (Fig 5). RACE cDNA was generated from RNA isolated from FaDu cells



Figure 6. Schematic of the RACE cDNA clones and rTS. A. The overall organization of the rTS RNA (1.8 kb species). B. the RACE clones and positions of predicted introns and ORFs are indicated and aligned relative to rTS. cDNAs shown in panel B terminate at oligo #337 at their 3'ends (see text for details). Solid boxes indicate regions corresponding to ORF1 and 1A, with triangles and roman numerals signifying the relative positions and numbering of introns, respectively. The open bracketed lines in B signify the 5'ends of the RACE clones do not contain splice donor sites, or rTS exon sequences and therefore likely correspond to incomplete introns. The open and hatched boxes represent the portion of rTS RNA complementary to intron IV and the protein coding portion of the TS gene, respectively. Not drawn to scale.

1752 Nucleic Acids Research, 1993, Vol. 21, No. 8

because of the high level of rTS RNA found in these cells, and this cell line also contains large amounts of high molecular weight RNA with rTS homology (Fig 3). While the RACE cDNA clones were generated by RT-PCR of RNA isolated from unfractionated cells, half of the cDNA clones sequenced contained regions divergent from rTS consistent with the presence of unexcised introns. A 113 nt region designated intron III was found to be present in one cDNA at a position corresponding to position 355 of rTS (i.e., #13) and as the 3' portion of a 426 nt insert present in the same site of another cDNA clone (i.e., 18). In contrast to known introns, this 113 nt fragment contains the sequence:gg at its 5'end. It is reasonable to conclude this non-canonical:gg splice donor is part of the reason for the apparent accumulation of unspliced rTS RNA found in many of the cell lines examined (Figs. 2 and 3). The presence of:gg at the 5' end of an intron would not be expected to allow efficient splicing, as deviations from the canonical: ag at this position are not well tolerated (23). This suggests that the 113 nt region in clone #18 may represent either a dead-end internal splice of intron III or an alternative splicing pathway. The presence of only part of intron I in RACE clones #8 and 18 also supports the notion of unspliced intron accumulation. These two RACE cDNAs likely arose from the cloning of cDNAs resulting from prematurely terminating reverse transcription reactions of pre-mRNAs.

Within the last seven years it has become evident that many eukaryotic genes such as erbA α , c- and N-myc, Xenopus laevis basic fibroblastic growth factor, rat gonadotropin-releasing hormone (GnRH), the human ERCC-1 DNA repair gene and others generate antisense transcripts (4-15). Although a number of antisense transcripts to eukaryoticgenes have been reported, only a few antisense transcripts which either code for proteins or exhibit the properties of protein- coding transcripts (e.g. polyadenylation, splicing and cytoplasmic transport of processed species) have been identified. To date only the rat GnRH, $erbA\alpha$, ear-1 and -7 antisense transcripts have been shown to undergo polyadenylation and intron excision (4, 14, 15). The human ERCC-1 antisense transcript, while not yet demonstrated to be spliced from an intron containing precursor, does appear to be polysome associated, suggesting that this antisense RNA has a protein coding function (11). Aside from the human ERCC-1. ear-1 and-7 antisense RNAs, rTS appears to be the only human antisense RNA, described to date, which is processed as many mRNAs are. The presence of several large ORFs and the ability of rTS mRNA generated in vitro to be translated in vitro (Dolnick unpublished result) suggest rTS is likely to code for protein in vivo.

Possible functions of rTS

Whether an rTS protein is produced in vivo has yet to be determined. The elevated expression of a 6.3 kb nuclear rTS RNA, as well as a novel 0.6 kb RNA in a methotrexate-resistant cell line (B1A), suggests rTS may have a function somehow related to the metabolic effects of methotrexate. The B1A cell line is resistant to methotrexate by virtue of amplification of the dihydrofolate reductase gene, and based upon expression of TS does not appear to have coamplified the TS gene (16). Thus it is unlikely the TS locus is coamplified with the dihydrofolate reductase locus in this cell line. The elevated rTS homologous RNA species in the B1A cell line also do not correspond to the 1.8 kb cytoplasmic rTS RNA. Since the 6.3 kb species is nuclear in the KB cell line (Fig 2), it is unclear whether the elevation of this species in B1A is coincidental or related to the appearance

of the novel 0.6 kb species in BIA. Of the antisense RNAs described to date which are processed in a manner similar to rTS, only the ear-1 and -7 have been shown to have possibly related functions (15). The unique rTS RNA profile in the methotrexate-resistant B1A cell line (Fig 3) suggests rTS may have a function related to a metabolic pathway perturbable by methotrexate. Since TS utilizes methylene tetrahydrofolate as a cofactor, it is also susceptible to effects of methotrexate. Thus it may be expected that rTS is metabolically if not functionally related to TS in some manner. A protein database scan for homology to the various ORFs of rTS has not yielded significant insights as to a potential function for the proteins potentially encoded by the rTS ORFs. Studies are currently underway to both characterize the in vitro translation products and characterize cellular expression further.

ACKNOWLEDGEMENTS

This work was supported by PHS grants CA34301, CA57634 and the institute CORE grant (CA16056) awarded by the National Cancer Institute.

REFERENCES

- 1. Dolnick, B.J. (1991) Proc. AACR 32, 1644.
- Dolnick, B.J., Zhang, Z-g., Hines, J.D., and Rustum, Y.M. (1992) Oncology Res. 4:65–72.
- 3. Dolnick, B.J. (1992) Proc. AACR 33, 2158.
- 4. Adelman, J.P., Bond, C.T., Douglass, J. and Herbert, E. (1987) Science 235, 1514-1517.
- Krystal, G.W., Armstrong, B.C. and Battey, J.F. (1990) Mol.Cell.Biol. 10, 4180–4191.
- 6. Spencer, C.A., Gietz, R.D. and Hodgetts, R.B. (1986) Nature 322, 279-281.
- 7. Nepveu, A. and Marcu, K.B. (1986) EMBO J. 5,2859-2865.
- 8. Williams, T. and Fried, M. (1986) Nature 322, 275-279.
- Farnham, P.J., Abrams, J.M. and Schimke, R.T. (1985) Proc. Natl. Acad. Sci. USA 82,3978-3982.
- 10. Stevens, J.G., Wagner, E.K., Devi-Rao, G.B., Cook, M.L., and Feldman, L.T. (1987) Science 235, 1056-1059.
- van Duin, M., van den Tol, J., Hoeijmakers, J.H.J., Bootsma, D., Rupp, I.P., Reynolds, P., Prakash, L. and Prakash, S. (1989) Mol.Cell.Biol. 9,1794-1798.
- 12. Heywood, S.M. (1986) Nuc. Acids Res. 14,6771.
- Celano, P., Berchtold, C.M., Kizer, D.L., Weeraratana, Nelkin, B.A., Baylin, S.B. and Casero, R.A.Jr. (1992) J.Biol.Chem. 267, 15092-15096.
- Lazar, M.A., Hodin, R.A., Darling, D.S. and Chin, W.W. (1989) Mol.Cell.Biol. 9,1128-1136.
- Miyajima, N., Horiuchi, R., Shibuya, Y., Fukushige, S-i., Matsubara, K-i., Toyoshima, K. and Yamamoto, T. (1989) Cell 57, 31-39.
- Srimatkandada,S., Medina,W.D., Cashmore,A.R., Whyte,W., Engel,D., Moroson,B.A., Franco,C.T., Dube,S.K. and Bertino,J.R. (1983) Biochemistry 22,5774-5781.
- 17. Will, C.L. and Dolnick, B.J. (1989) J.Biol.Chem. 264, 21413-21421.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Second Edition. Cold Spring Harbor University Press, Cold Spring Harbor.
- 19. Rosen, K.M. and Villa-Komaroff, L. (1990) Focus 12,23-24.
- 20. Sawadogo, M. and Van Dyke, M.W. (1991) Nucl. Acids Res. 19,674.
- Frohman, M.A. and Martin, G.R. (1990) in Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds.), PCR Protocols-A Guide to Methods and Applications. Academic Press, San Diego, CA, pp.228-236.
- Kaneda,S., Nalbantoglu,J., Takeishi, K., Shimizu, K., Gotoh, O., Seno, T. and Ayusawa, D. (1990) J. Biol. Chem. 265, 20277-20284.
- Padgett, R. A., Grabowski, P.J., Konarska, M.M., Seila, S., Sharp, P.A. (1986) Ann. Rev. Biochem. 55, 1119–1150.
- 24. Wickens, M. (1990) TIBS 15,277-281.