Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 8020–8024, September 1992
Biochemistry

Human dUTP pyrophosphatase: cDNA sequence and potential biological importance of the enzyme

EVAN M. MCINTOSH*, DARLENE D. AGER, MICHAEL H. GADSDEN, AND ROBERT H. HAYNES†

Department of Biology, York University, Toronto, Ontario Canada M3J 1P3

Communicated by Richard B. Setlow, June 1, 1992

ABSTRACT **Two functional human dUTP pyrophosphatase (dUTPase; EC 3.6.1.23) cDNAs were isolated from a cDNA expression library by genetic complementation in *Escherichia coli*. These cDNAs differed in size but exhibited a common overlapping DNA sequence. Contained within this sequence was a single long open reading frame sufficient to encode a polypeptide of 141 amino acids with a calculated molecular mass of 16.6 kDa. The amino acid sequence of this protein exhibits 35% identity with the *E. coli* dUTPase and 53% identity with the *Saccharomyces cerevisiae* enzyme. The human dUTPase was found to contain five characteristic amino acid sequence motifs that are common to the dUTPases of *E. coli*, yeast, and herpesviruses and to dUTPase-like sequences encoded by some retrovirus *gag* and *pol* genes. A high degree of amino acid sequence identity (>60%) was also observed between the human dUTPase and the putative pseudoproteases of two poxviruses, indicating that these virus proteins are dUTPases. Northern hybridization analysis reveals that dUTPase is encoded by at least two species of poly(A)$^+$ mRNA and possibly a third, smaller species. All of these mRNAs are present in a variety of human tissues but their relative levels vary between tissues. Southern analysis indicates that the dUTPase gene has been conserved to some extent throughout vertebrate evolution; however, the gene may be very large, or its organization somewhat complex in some systems. We suggest that dUTPase may generally perform an essential role in DNA replication and therefore could serve as a target enzyme for the development of chemotherapeutic compounds.**

In most biological systems, uracil is not a normal base component of DNA. It can arise in DNA, however, via the spontaneous deamination of cytosine, an event that occurs at low but significant rates (1). If left unrepaired, DNA uracil can induce transition mutations during subsequent rounds of DNA replication (2). To avoid this problem, most organisms possess a repair system that acts specifically to remove uracil from the DNA. This process is initiated by the enzyme uracil glycosylase and involves base excision and repair polymerization of the DNA strand encompassing the uracil-substituted site (3, 4).

Apart from cytosine deamination, uracil can become incorporated into DNA through the utilization of dUTP by DNA polymerases. dUTP is an intermediate compound in the dTTP biosynthetic pathway and is formed in all dividing cells from UDP via ribonucleotide reductase and nucleoside diphosphate kinase. If present in substantial amounts, dUTP can facilitate the synthesis of highly uracil-substituted DNA during replication (5). In turn, such substituted DNA would induce extensive uracil glycosylase-mediated excision repair. It has been suggested that repair replication under these conditions results in a reiterative process of uracil misincorporation and excision repair that leads to DNA fragmentation and cell death (1, 6–8). This potentially destructive process does not occur under normal growth conditions because the enzyme dUTP pyrophosphatase (dUTPase; EC 3.6.1.23) degrades dUTP to dUMP and PP_i. The activity of this enzyme is believed to maintain extremely low levels of dUTP and thereby effectively excludes this compound as a substrate for DNA polymerases (9, 10).

Despite a study in *Escherichia coli* that has shown dUTPase to be essential for cell viability (11), the potential importance of the enzyme in DNA replication has not been widely appreciated. Recently, we have discovered that dUTPase is also necessary for viability of the yeast *Saccharomyces cerevisiae*. This was demonstrated by performing tetrad analysis of a diploid strain heterozygous for a dUTPase null mutation and observing 2:0 segregation for viability among the haploid spores (M.H.G., E.M.M., J. C. Game, and R.H.H., unpublished work). The discovery that dUTPase is of vital importance in two such widely divergent organisms suggests that it may be indispensable for all cells. In addition to these studies, however, another finding pointing to an important role of dUTPase in biological systems has come from the recent observation that dUTPases are encoded by several members of the lentivirus and oncovirus subfamilies of retroviruses (12). Furthermore, it has also been shown that a developmentally regulated inhibitor of dUTPase exists in *Drosophila melanogaster* (13). This discovery in particular has suggested that regulated inhibition of the enzyme could constitute a mechanism of programmed cell death (13). In view of the growing number of observations which indicate that dUTPase may perform a more significant role than that of a simple metabolic catalyst, we have undertaken a study aimed at assessing the importance of this enzyme in human cells. To facilitate molecular approaches to this problem, we have cloned and characterized functional human dUTPase cDNAs.‡

MATERIALS AND METHODS

Strains, Plasmids, cDNA Library, and Enzyme Assay. The *E. coli* strain used for screening the library, BW286 [*dut-1*, Δ(*xth–pncA*)*90*], has been described (11) and was supplied by B. Weiss (Johns Hopkins University, Baltimore). *E. coli* strain JM101 (14) was used to isolate single-stranded phagemid DNA. Phagemid pTZ19R was obtained from Pharmacia. The human cDNA library (λYES-R) and phage λKC were supplied by S. Elledge (Baylor University, Houston) and have been described (15, 16). dUTPase enzyme assays using toluene-permeabilized *E. coli* cells were performed as described (17), with minor modifications.

DNA Sequencing. DNA sequencing was performed with a Sequenase kit (United States Biochemical). The two human dUTPase cDNAs isolated from the screening procedure were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ORF, open reading frame; dUTPase, dUTP pyrophosphatase.

*To whom reprint requests should be addressed.
†Present address: Annual Reviews, 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0897.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M89913).

recovered from the expression vector by digestion with *Eco*RI and then subcloned into the *Eco*RI site of phagemid pTZ19R. Plasmids isolated from this procedure containing the 0.5-kilobase (kb) and 0.8-kb inserts were then screened for insert orientation by isolating single-stranded DNA and sequencing it by using the reverse sequencing primer (Pharmacia). The DNA sequence of each of the two cDNAs was determined by sequentially synthesizing appropriate oligonucleotide primers based on the DNA sequence information obtained using the reverse sequencing primer first, and subsequently, the synthesized primers. Both strands of the human dUTPase open reading frame (ORF) were sequenced.

Hybridization Analyses. Southern and Northern hybridization analyses were performed using a zoo-blot and a multiple tissue northern blot, respectively, purchased from Clontech. Hybridization conditions were as recommended by the supplier. Hybridization probes were labeled with [α-^{32}P]dATP by using the random primer method (18) and a kit purchased from GIBCO/BRL. The DNA used to prepare the probe was made by PCR amplification of the human dUTPase ORF from plasmid pHUMDUT1 using oligonucleotides EM100 (5'CCGGTTGCTAGCCTCGAGATGCAGCTCCGCTTTGCCC-3') and EM101 (5'-CCGGTTGCTAGCCTCGAGTTAATTCTTTCCAGTGGAACCAA-3'). The resulting 0.45-kb PCR product was purified by gel electrophoresis prior to labeling with random primers. The probe for Northern hybridization was prepared as follows; the 0.45-kb PCR product described above was digested with *Nhe* I and cloned into the *Nhe* I site of plasmid pMAMneo (Clontech). The resulting plasmid was then digested with *Xho* I and the 0.45-kb fragment containing the dUTPase coding region was purified and labeled with [α-^{32}P]dATP as described above.

RESULTS

Isolation of Human dUTPase cDNA Clones by Complementation. To isolate a functional human dUTPase cDNA, we used the λ-based cDNA expression library described by Elledge and Spottswood (15). This library has been designed to allow for screening by genetic complementation of mutations in either *E. coli* or *S. cerevisiae*. We chose to screen the library by using BW286, a *dut-1, xth* strain of *E. coli* (11). The *dut-1* allele encodes a temperature-sensitive dUTPase and the *xth* deletion mutation eliminates exonuclease III activity. The combination of these two markers results in lethality at 37°C on rich medium. To screen the library, we first isolated a kanamycin-resistant derivative of strain BW286 by lysogenizing with the λKC phage. The resulting strain (BW286-λKC) was then used to screen the phage cDNA library as described (16). Infected cells were plated on YT medium containing kanamycin and ampicillin and were grown for 48 hr at 37°C. Plasmids were recovered from several survivors and tested for their ability to transform BW286 to temperature independence. Two plasmids characterized from this secondary screen were found to complement the temperature-sensitive phenotype of BW286 at high frequency (4×10^5 transformants per μg of plasmid DNA). The two plasmids, designated pHUMDUT1 and pHUMDUT2, were found to contain cDNA inserts of approximately 0.5 and 0.8 kb, respectively.

To determine whether the isolated plasmids carried a functional dUTPase gene, we performed dUTPase enzyme assays on permeabilized cells of BW286 and BW286 transformed with pHUMDUT1 and pHUMDUT2. Both transformants exhibited 9-fold more dUTPase activity than the untransformed control (data not shown). Since dUTPases are generally highly specific for dUTP as substrate, we also tested for hydrolysis of dTTP with the permeabilized cell preparations to determine whether the observed activity was actually some form of nonspecific phosphatase or 5'-nucleotidase activity. In each case, dTTP was not hydrolyzed (data not shown), indicating that the two plasmids recovered from the screening procedure did, in fact, encode dUTPase.

Characterization of dUTPase cDNAs. The two positive clones identified from complementation screening and enzyme assays were then further characterized by DNA sequence analysis. Both clones contained a common sequence of 526 bp (Fig. 1). Within this sequence is an ORF extending 141 codons from a common translation initiation codon. The predicted molecular mass of the protein encoded by this ORF is 16.6 kDa. An alignment of the amino acid sequence of this protein with the amino acid sequences of dUTPases of *S. cerevisiae* (M.H.G., E.M.M., J. C. Game, and R.H.H., unpublished data) and *E. coli* (19) is shown in Fig. 2. The protein encoded by the ORF is similar in length (141 residues) to both the yeast (147 residues) and *E. coli* (150 residues) enzymes and exhibits significant amino acid sequence identity with both proteins (53% with *S. cerevisiae*, 35% with *E. coli*). The percent identity exhibited between these proteins is also reflected approximately at the DNA sequence level (55% with *S. cerevisiae*, 44% with *E. coli*). On the basis of the degree of identity shared by these proteins, combined with the results described above, we concluded that the cDNAs carried on plasmids pHUMDUT1 and pHUMDUT2 encoded functional dUTPase.

The difference in the size of the two cDNAs isolated arises from variation in the length of both the 5' and 3' flanking sequences. The 5' leader sequence of the pHUMDUT1 cDNA is 42 base pairs (bp) longer than that of pHUMDUT2, but the 3' flanking sequence is 291 bp shorter. Inspection of the 5' flanking sequences of the two cDNAs reveals a nearly perfect match to consensus sequence for the *E. coli* ribosome binding site, AGGA, centered 10 bp upstream of the dUTPase translation initiation codon. The fortuitous presence of this sequence is most likely responsible for the ability of the cDNAs to be expressed in *E. coli*.

```
                                                    CTCGCCTTCTGGCTCTGCATGCCC
                                  *
TGCCTCTGAAGAGACACCCGCCATTTCACCCAGTAAGCGGGCCCGGCCTGCGAGGTGGGCGGC

 M   Q   L   R   F   A   R   L   S   E   H   A   T   A   P   T
ATG CAG CTC CGC TTT GCC CGG CTC TCC GAG CAC GCC ACG GCC CCC ACC

 R   G   S   A   R   A   A   G   Y   D   L   Y   S   A   Y   D
CGG GGC TCC GCG CGC GCC GCG GGC TAC GAC CTG TAC AGT GCC TAT GAT

 Y   T   I   P   P   M   E   K   A   V   V   K   T   D   I   Q
TAC ACA ATA CCA CCT ATG GAG AAA GCT GTT GTG AAA ACG GAC ATT CAG

 I   A   L   P   S   G   C   Y   G   R   V   A   P   R   S   G
ATA GCG CTC CCT TCT GGG TGT TAT GGA AGA GTG GCT CCA CGG TCA GGC

 L   A   A   K   H   F   I   D   V   G   A   G   V   I   D   E
TTG GCT GCA AAA CAC TTT ATT GAT GTA GGA GCT GGT GTC ATA GAT GAA

 D   Y   R   G   N   V   G   V   V   L   F   N   F   G   K   E
GAT TAT AGA GGA AAT GTT GGT GTT GTA CTG TTT AAT TTT GGC AAA GAA

 K   F   E   V   K   K   G   D   R   I   A   Q   L   I   C   E
AAG TTT GAA GTC AAA AAA GGT GAT CGA ATT GCA CAG CTC ATT TGC GAA

 R   I   F   Y   P   E   I   E   E   A   Q   A   L   D   D   T
CGG ATT TTT TAT CCA GAA ATA GAA GAA GCT CAA GCC TTG GAT GAC ACC

 E   R   G   S   G   G   F   G   S   T   G   K   N
GAA AGG GGT TCA GGA GGT TTT GGT TCC ACT GGA AAG AAT TAA

                                  *
AATTTATGCCAAGAACAGAAAACAAGAAGTCATACCTTTTTCTTAAAAAAAAAAAAAAAGTTT

TTGCTTCAAGTGTTTTGGTGTTTTGCACTTCTGTAAACTTACTAGCTTTACCTTCTAAAAGTA

GTGCATTTTTTCTTTTTTTTATGATCAAGGAAAAGATCATTAAAAAAAAACACAAAGAAGTTT

TTCTTTGTGTTTGGATCAAAAAGAAACTTTGTTTTTCCGCAATTGAAGGTTGTATGAAATCTG

CTTTGTGGTGACCTGATGTAAACAGTGTCTTCTTAAAATCAAATGTAAATCAATTACAGATTA

AAAAAAAAA
```

FIG. 1. DNA sequence of the two human dUTPase cDNAs isolated. The start of the pHUMDUT2 cDNA and the end of the pHUMDUT1 cDNA are marked by asterisks. Underlined sequences are described in the text. The predicted amino acid sequence of the human enzyme derived from the ORF is shown in single-letter code.

```
                                          1
                         * *  ** ** *** ***** *    * **  * *    *
Yeast      MTATSDKVLNIQLR-SASATVPTKGSATAAGYDIYASQ-DITI--PAMGQGMV
Human             MQLRFARLSEHATAPTRGSARAAGYDLYSAY-DYTI--PPMEKAVV
E.coli      MKKIDVKILD-PRVGKEFPLPTYATSGSAGLDLRACLNDAVELAPGDTTLVP
                     *   *       **      ** **     *     *     *

                              2                    3
             ***    * * *** ******* *  *  *  *** * ** * * ** **
Yeast       STDISFTVPVGTYGRIAPRSGLAVKNGIQTG--AGVVDRDYTGEVKVVVFNHS
Human       KTDIQIALPSGCYGRVAPRSGLAAKHFIDVG--AGVIDEDYRGNVGVVLFNFG
E.coli     TGLAIHIADPSL-AAMMLPRSGLGHKHGIVLGNLVGLIDSDYQGQLMISVWNRG
               * ** **       *****  ** *  *   * ** ** *       * *

                      4                              5
              *  ***** **** * *     *     *    ** ********
Yeast      QRDFAIKKGDRVAQLILEKIVDDAQIVVVDSLEESARGRGGFGSTGK
Human      KEKFEVKKGDRIAQLICERIFYPE-IEEAQALDDTERGSGGFGSTGKN
E.coli     QDSFTIEPGERIAQMIFVPVVQAE-FNLVEDFDATDRGEGGFGHSGRQ
              *    * **** *       *        * * ** ****  *
```

FIG. 2. Alignment of the amino acid sequences of the human, *S. cerevisiae*, and *E. coli* dUTPases. Asterisks above and below the aligned sequences indicate, respectively, residues that are identical between the *S. cerevisiae* and human enzymes and between the *E. coli* and human enzymes. Gaps were introduced in the alignment, where indicated (dashes), to achieve maximum identity. Amino acid sequence motifs that are conserved among these enzymes and also with the herpesvirus dUTPases (20) are underlined and identified numerically (nos. 1–5).

The 3′ flanking sequence of the pHUMDUT1 cDNA does not exhibit a poly(A)$^+$ tail. Since the λYES-R cDNA library was constructed from poly(A)$^+$ mRNA (15), it is likely that this particular cDNA resulted from a cloning artifact such as partial degradation of the mRNA prior to cDNA synthesis. The 3′ flanking sequence of the pHUMDUT2 clone is considerably longer and contains a short poly(A)$^+$ tail. The polyadenylylation signal sequence AATAAA is not present within this region, although there are two direct repeats of a similar sequence (AATCAA) within 26 bp of the poly(A)$^+$ stretch.

Southern Hybridization Analysis. To examine the genomic organization of the human dUTPase gene and to gain some insight as to the conservation of this gene between various species, we performed a Southern hybridization of a zoo-blot (Fig. 3). All of the genomic DNAs represented on the blot were digested with *Eco*RI. The results of this analysis revealed numerous strong hybridization bands for both human and rhesus monkey genomic DNA. Reasonably strong signals from multiple bands were also seen for bovine DNA. With the exception of yeast, the human dUTPase probe hybridized with specific DNA fragments from all other species represented on the blot. However, for each of these species, the number of bands hybridizing to the probe and the intensity of the signal were less than those observed between human and rhesus monkey DNA. These results indicate that the dUTPase gene sequence is well conserved between primates but that there is less conservation of the sequence between primate and nonprimate vertebrates. The number of human and monkey *Eco*RI fragments hybridizing to the cDNA probe also indicate that, in these species, (*i*) the dUTPase gene may be very large with multiple introns, or (*ii*) there may be multiple copies of the gene exhibiting restriction fragment length polymorphisms, or (*iii*) dUTPase pseudogenes exist.

Northern Hybridization Analysis. The size and distribution of dUTPase mRNA in various human tissues were analyzed by Northern hybridization. The RNA gel, obtained from a commercial source (Clontech), contained poly(A)$^+$ mRNA (2 μg per lane) isolated from various tissues. The probe, encompassing the coding region of the dUTPase cDNA, hybridized to at least two distinct species of mRNA (2.3 and 1.5 kb) and possibly a third, smaller species (1.0 kb) (Fig. 4). Each of these mRNA species is present in all of the tissues represented on the blot; however, the level of these transcripts varies among tissues. Also, the abundance of the transcripts relative to each other differs in various tissues. For example, the 2.3-kb mRNA appears to be more abundant in the placenta than the heart, whereas the opposite is true of the smaller, 1.5-kb transcript.

Several reasons could account for the existence of these different mRNA species. First, they might reflect differential splicing of a primary transcript, which, in turn, could generate isoforms of the enzyme. This would account for discrepancies in the reported molecular mass of the active human holoenzyme [45 and 62.5 kDa in HeLa cells (21, 22) and 68 kDa in lymphoid cells (23)] and the difference between the previously reported size of the HeLa dUTPase monomer [21 kDa (23)] and that calculated (16.6 kDa) for the dUTPase characterized here. Alternatively, two or possibly three dUTPase genes exhibiting differences in the extent of their coding regions could also account for these transcripts and various molecular forms of the dUTPase in human cells. It is tempting to speculate that one of these transcripts might encode a mitochondrial enzyme; however, evidence of such an enzyme has not been reported.

DISCUSSION

During the course of our work on the yeast dUTPase, we discovered a striking similarity between the amino acid

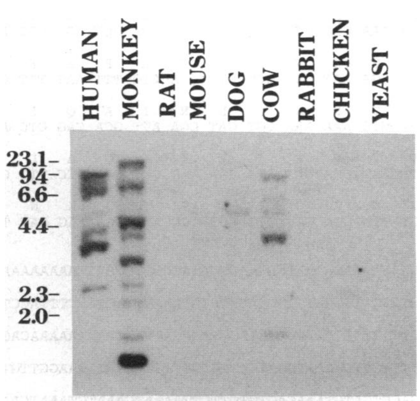

FIG. 3. Interspecies Southern hybridization analysis. A ^{32}P-labeled probe encompassing the human dUTPase coding region was hybridized to a dried gel containing *Eco*RI-digested DNAs prepared from various species. The gel was obtained commercially. Each lane contained 5 μg of genomic DNA. The species represented in each lane is indicated. The positions of molecular size standards (values in kilobases) are indicated to the left.

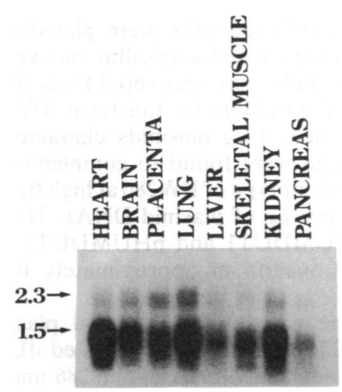

FIG. 4. Northern blot analysis of human dUTPase mRNA. The RNA blot was obtained from a commercial source. Each lane contained 2 μg of poly(A)$^+$ mRNA isolated from various tissues as indicated. The blot was hybridized with a ^{32}P-labeled probe encompassing only the coding region of the human dUTPase cDNA. The sizes of the transcripts hybridizing with the probe are indicated to the left (values in kilobases).

sequence of this enzyme and the pseudoprotease domains described previously for some retrovirus and poxvirus genes. This similarity was also reported by McGeoch (20) during his analysis of the herpes simplex virus 1 dUTPase gene. Although the dUTPase of this virus and the putative dUTPases of two other herpesviruses (Epstein–Barr and varicella–zoster) are highly divergent from the *E. coli* enzyme, five distinct amino acid sequence motifs are common to these dUTPases and also to the pseudoproteases of retroviruses and poxviruses (20). As shown in Fig. 2, these same five motifs are found in both the yeast and human dUTPase enzymes. McGeoch (20) has observed that a short amino acid sequence highly similar to motif 3 (GVIDXDXXG) is common to several phosphofructokinases, and he has suggested that this motif may be involved in recognition of the sugar and/or phosphate component of the substrate. We have noted that a portion of motif 5 (GFGSTG), located at the extreme carboxyl termini of the yeast and human dUTPase monomers, conforms to the P-loop consensus sequence (GXGXXG) of many nucleotide binding proteins (24). In the case of the $p21^{Ha\text{-}ras}$ GTP-hydrolyzing protein, structural analyses have shown that this P-loop motif hooks around the triphosphate moiety of the nucleotide between the β and γ phosphates (25). By analogy, it is likely that this motif performs a similar role in the dUTPase enzymes and therefore may be critically involved in substrate recognition and PP_i hydrolysis.

Our observation that the five amino acid sequence motifs described by McGeoch (20) are also conserved in both the yeast and human dUTPase protein sequences further supported the hypothesis that the pseudoprotease domains of some retroviruses encode dUTPases. This notion was confirmed by Elder *et al.* (12), who demonstrated that the pseudoprotease domain of the feline immunodeficiency virus *pol* gene encodes a dUTPase enzyme. They also reported evidence indicating that three other pseudoprotease-containing retroviruses encode dUTPases. The extensive amino acid sequence similarity exhibited between the human dUTPase and the pseudoproteases (26, 27) of the vaccinia (63% identity) and orf (70% identity) poxviruses (Fig. 5) strongly suggests that these proteins are also dUTPases. The presence of a dUTPase gene in poxviruses is not surprising, since these viruses also encode other dNTP-metabolizing enzymes such as ribonucleotide reductase and thymidine and thymidylate kinases (28). However, the presence of dUTPases within the small genomes of retroviruses is remarkable and could suggest a vital role for this enzyme in the replication strategy of these viruses. This might be expected if, for some reason, these particular viruses are exposed to high levels of dUTP in their natural host cell environments. Attempted genome replication under these conditions could lead to uracil

```
                 *   ***  ** * ***  *** ** ****  **      ** **  *
orf       MEFCHTETLQVVRLSQNATIPARGSPGAAGLDLCSAYDCVIPSHCSRVVFTDLLIK
Human           MQLRFARLSEHATAPTRGSARAAGYDLYSAYDYTIPPMEKAVVKTDIQIA
Vaccinia   MNINSPVRFVKETNRAKSPTRQSPYAAGYDLYSAYDYTIFPGERQLIKTDISMS
                  **     *  *** *  ************** * *    ****


          ******* ******* ***************************   *********
orf       PPSGCYGRIAPRSGLAVKHFIDVGAGVIDEDYRGNVGVVLFNFGNSGFEVKKGDRI
Human     LPSGCYGRVAPRSGLAAKHFIDVGAGVIDEDYRGNVGVVLFNFGKEKFEVKKGDRI
Vaccinia MPKGCYGRIAPRSGLSLKG-IDIGGGVIDEDYRGNIGVILINNGKCTFNVNTGDRI
          * ***** ******  *  ** * ********** ** * * **  * *  ****


          ********  *   **  ** * **  ******
orf       AQLICERISCPAVQEVNCLDNTDRGDSGFGSTGSGACGGRDTAWYIS
Human     AQLICERIFYPEIEEVQALDDTERGSGGFGSTGKN
Vaccinia AQLIYQRIYYPELEEVQSLDSTNRGDQGFGSTGLR
          ****  ** *** **** ** * **  ******
```

FIG. 5. Alignment of the human dUTPase amino acid sequence with those of the putative pseudoproteases of the orf and vaccinia poxviruses. Asterisks above and below the alignments indicate, respectively, residues that are identical between the orf and human and between the vaccinia and human protein sequences.

substitution in both DNA strands during reverse transcription. Consequently, convergent excision repair of both DNA strands might result in the irreversible fragmentation of this vital replication intermediate.

The role of dUTPase in the life cycle of these retroviruses is of considerable biological interest. Should this enzyme prove to be essential for the viability of these viruses, it is curious that dUTPases are not encoded in all retrovirus genomes. Although several of the currently known lentiviruses exhibit dUTPase-like segments within the *pol* gene, the human, simian, and bovine immunodeficiency viruses do not encode this enzyme. A trivial explanation for this difference between the lentiviruses could be that these immunodeficiency viruses are simply not exposed to significant dUTP levels during replication in their natural hosts. However, we suggest two other possible explanations. First, since the feline immunodeficiency virus encapsidates its own dUTPase (12), the other lentiviruses may have evolved a mechanism to incorporate their host cell dUTPase within the mature virion. If this is true, then the human enzyme might be incorporated into mature human immunodeficiency virus particles. However, in view of the pseudoprotease/dUTPase relationship, another intriguing possibility exists, at least for the human immunodeficiency viruses. We base this second hypothesis on the occurrence within the human genome of the defective retrovirus HERV-K (29). This virus, which is estimated to be present at 50 copies per haploid genome, also encodes a pseudoprotease with similarity to dUTPases. If a functional copy of this gene were expressed, perhaps fortuitously in human T cells, then human immunodeficiency viruses might utilize this enzyme during replication. In this scenario, the viability of these immunodeficiency viruses in humans would be dependent upon genetic complementation by the endogenous HERV-K retrovirus dUTPase.

The essential nature of dUTPase in *E. coli* and yeast suggests that the enzyme may generally perform a vital role in DNA replication. If this is true, then dUTPase could serve as a useful target enzyme for chemotherapeutic drug design. dUTPase is a component of the thymidylate biosynthetic pathway, and inhibition of *de novo* dTTP synthesis has, for many years, been the rational basis for the development of new anticancer compounds. Although this approach has focused extensively on other enzymes involved in dTTP synthesis, primarily thymidylate synthase and dihydrofolate reductase, we suggest that dUTPase might also be exploited for the same purpose and perhaps yield a new class of inhibitory compounds.

The molecular mechanism of cell killing induced by antithymidylate drugs such as methotrexate is not fully understood. The predominant theory proposes that thymineless conditions are generally toxic because they induce elevated dUTP levels (6–8). As a consequence of this, cell death is believed to result from the following series of events. First, high dUTP levels facilitate extensive misincorporation of uracil into replicating DNA. The substituted nascent DNA is subsequently attacked by uracil glycosylase and excision repair enzymes. However, repair replication in the presence of elevated dUTP pools results in the synthesis of uracil-substituted DNA repair patches. These patches then become substrates for the same excision repair enzymes. Effectively, this sequence of events results in a self-inducing excision repair process that would severely inhibit the completion of daughter-strand synthesis, and thereby, lead to cell death. This theory is supported by previous studies in human cells, which have shown that methotrexate and other chemotherapeutic antifolates can induce highly elevated dUTP levels, promote uracil misincorporation into DNA, and induce DNA strand breakage and cell killing (1, 7, 30). Since dUTPase is the primary modulator of cellular dUTP levels, inhibitors of

this enzyme might also be expected to induce cell death by this mechanism.

Before dUTPase can be considered as a rational target for chemotherapeutic drug design, it will be necessary to determine the importance of the enzyme for human cell division. Since no effective inhibitors of the human dUTPase are currently known, it is not possible to resolve this question through inhibitor studies. However, characterization of a functional human dUTPase cDNA can allow for a molecular approach to this problem through various methods, such as the use of antisense oligonucleotides or by the construction of plasmids that can generate an inducible dUTPase ribozyme or antisense mRNA *in vivo*.

We thank Dr. Bernard Weiss for supplying *E. coli* strain BW286, Dr. Stephen Elledge for providing the cDNA library, and Dr. Brian Golding for helpful comments. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

1. Richards, R. G., Sowers, L. C., Laszlo, J. & Sedwick, W. D. (1986) *Adv. Enzyme Regul.* **22**, 157–185.
2. Impellizzeri, K. J., Anderson, B. & Burgers, P. J. M. (1991) *J. Bacteriol.* **173**, 6807–6810.
3. Lindahl, T. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 3649–3653.
4. Taylor, A. F. & Weiss, B. (1982) *J. Bacteriol.* **151**, 351–357.
5. Warner, H. R., Duncan, B. K., Garret, C. & Neuhard, J. (1978) *Nature (London)* **272**, 32–34.
6. Barclay, B. J., Kunz, B. A., Little, J. G. & Haynes, R. H. (1982) *Can. J. Biochem.* **60**, 172–194.
7. Goulian, M., Bleile, B. & Tseng, B. Y. (1980) *J. Biol. Chem.* **255**, 10630–10637.
8. Ingraham, H. A., Dickey, L. & Goulian, M. (1986) *Biochemistry* **25**, 3225–3230.
9. Shlomai, J. & Kornberg, A. (1978) *J. Biol. Chem.* **253**, 3305–3312.
10. Tye, B. & Lehman, I. R. (1977) *J. Mol. Biol.* **117**, 293–306.
11. El-Hajj, H., Zhang, H. & Weiss, B. (1988) *J. Bacteriol.* **170**, 1069–1075.
12. Elder, J. H., Lerner, D. L., Hasselkus-Light, C. S., Fontenot, D. J., Hunter, E., Luciw, P. A., Montelaro, R. C. & Phillips, T. R. (1992) *J. Virol.* **66**, 1791–1794.
13. Nation, M. D., Guzder, S. N., Giroir, L. E. & Deutsch, W. A. (1989) *Biochem. J.* **259**, 593–596.
14. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
15. Elledge, S. & Spottswood, M. R. (1991) *EMBO J.* **10**, 2653–2659.
16. Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M. & Davis, R. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1731–1735.
17. Lundberg, L. G., Karlstrom, O. H., Nyman, P. O. & Neuhard, J. (1983) *Gene* **22**, 115–126.
18. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
19. Lundberg, L. G., Thoresson, H., Karlstrom, O. H. & Nyman, P. O. (1983) *EMBO J.* **2**, 967–971.
20. McGeoch, D. J. (1990) *Nucleic Acids Res.* **18**, 4105–4110.
21. Caradonna, S. J. & Adamkiewicz, D. (1984) *J. Biol. Chem.* **259**, 5459–5464.
22. Mahagaokar, S., Rao, P. N. & Orengo, A. (1980) *Int. J. Biochem.* **11**, 415–421.
23. Ingraham, H. A. & Goulian, M. (1982) *Biochem. Biophys. Res. Commun.* **109**, 746–752.
24. Hanks, S. K., Quinn, A. M. & Hunter, T. (1988) *Science* **241**, 42–52.
25. Wittinghofer, A. & Pai, E. F. (1991) *Trends Biochem. Sci.* **16**, 382–387.
26. Slabaugh, M. B. & Roseman, N. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4152–4155.
27. Mercer, M. A., Fraser, K. M., Stockwell, P. A. & Robinson, A. J. (1989) *Virology* **172**, 665–668.
28. Hughes, S. J., Johnston, L. H., de Carlos, A. & Smith, G. (1991) *J. Biol. Chem.* **266**, 20103–20109.
29. Ono, M., Yasunaga, T., Miyata, T. & Ushikubo, H. (1986) *J. Virol.* **60**, 589–598.
30. Sedwick, W. D., Kutler, M. & Brown, O. E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 917–921.