JOURNAL OF VIROLOGY, Dec. 1988, p. 4774–4777
0022-538X/88/124774-04$02.00/0

Vol. 62, No. 12

# Identification of the Coding Sequence for Herpes Simplex Virus Uracil-DNA Glycosylase

DIANE M. WORRAD[1] AND SAL CARADONNA[1,2]*

*Department of Biochemistry, School of Osteopathic Medicine,[1]* and the Robert Wood Johnson Medical School,[2] University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854*

Received 13 June 1988/Accepted 26 August 1988

**We have recently isolated a herpes simplex virus (HSV) type 2 (strain 333)-specific cDNA that encodes uracil-DNA glycosylase. This cDNA lies between 0.065 and 0.08 map units on the HSV genome. Within this region there are five overlapping transcripts which encompass three open reading frames. We have determined that the second open reading frame, UL-2, codes for glycosylase. In vitro transcription of the UL-2 region and subsequent translation yielded uracil-DNA glycosylase activity. Sequence analysis of the UL-2 open reading frame indicated a coding capacity of 295 amino acids. Comparison to the HSV type 1 (strain 17) sequence indicated that there is 74% amino acid homology between the two strains, with most of the conservation occurring in the middle and the 3′ end. The 5′ end, however, has diverged considerably.**

Uracil-DNA glycosylase is a DNA repair enzyme that removes uracil residues from DNA (8). The misincorporation of uracil can arise from incorporation of dUTP during replication or from deamination of cytosine (9). If left uncorrected, deaminated cytosine residues can lead to GC → AT transition mutations (8). We have recently isolated a herpes simplex virus (HSV)-specific cDNA that can hybrid select a message which encodes a uracil-DNA glycosylase species (2). This cDNA was mapped between 0.065 and 0.08 map units on the prototypic arrangement of the HSV genome. Within this region of the genome there are five overlapping transcripts of 3.4, 2.8, 2.4, 1.7, and 1.0 kilobases (kb) (Fig. 1B). Recently, three open reading frames were identified within this region of HSV type 1 (strain 17) (10). We wanted to determine which one of these open reading frames codes for the uracil-DNA glycosylase.

The original cDNA isolate, 184, used to hybrid select HSV glycosylase-specific mRNA, is 1.25 kb in size and contains most of UL-3 and about half of the 3′ end of UL-2 (Fig. 1C). This isolate was used to screen an HSV type 2 (strain 333) cDNA library constructed as previously described (2), and several cDNA isolates were selected. One cDNA isolate (3.4 kb; Fig. 1B), which was identified by screening, contained all three open reading frames. According to the HSV type 1 DNA sequence, open reading frames UL-1, UL-2, and UL-3 have the capacity to encode proteins with 224, 334, and 235 amino acid residues, respectively. HSV glycosylase has a monomeric molecular weight of between 35,000 and 40,000 (2). We surmised that UL-2 was the only open reading frame sufficiently large enough to encode the glycosylase function. Therefore, the 3.4-kb cDNA isolate was subcloned to yield a 1.5-kb fragment which contained all of UL-2 (1.5-kb *Pvu*II fragment, Fig. 1C). This 1.5-kb fragment was cloned into the pGEM vector (Promega Biotec), and orientation was determined. This construct was then linearized and transcribed in vitro by the procedure of Krieg and Melton with T7 RNA polymerase (7). Transcription was performed in the absence and presence of the RNA cap structure analog m7G(5′)ppp(5′)G (6) (New England BioLabs, Inc.). Full-length transcripts were verified by gel electrophoresis as described previously (2). A single species of 1.5-kb RNA was produced in both the absence and the presence of the cap structure (Fig. 2). This RNA was then purified and translated in a rabbit reticulocyte translation system. Following translation, the lysate was analyzed for glycosylase activity as described previously (2).

Results from this experiment revealed that glycosylase activity was produced in lysates containing RNA transcribed from the UL-2 open reading frame. RNA transcribed in the absence of the cap structure produced 10.2 pmol of acid-soluble uracil per min per μg of RNA. Transcribed RNA containing the cap structure analog yielded three times the activity, 33 pmol of uracil released per min per μg of RNA. RNA quantitation was performed in parallel transcription reactions in the presence of radiolabeled CTP. Background levels of glycosylase activity from translations without any input RNA or RNA transcribed from the original cDNA 184, isolate, were less than 0.15 pmol of uracil released per lysate. It is interesting to note that the in vitro transcription and translation of the 3.4-kb cDNA did not yield measurable levels of glycosylase activity.

In vitro translation of the transcripts generated from the UL-2 subclone in the presence of [$^{35}$S]methionine and analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed three proteins with approximate molecular weights of 35,000, 32,000, and 28,000 (Fig. 3). The 35,000-molecular-weight protein is in the approximate range of the molecular weight of the purified glycosylase protein (39,000) and is similar to the predicted 37,700-molecular-weight protein, as determined by DNA sequence analysis. The 32,000- and 28,000-molecular-weight proteins appear to be truncated protein products of the 35,000-molecular-weight protein resulting from incomplete translation. Potential alternate start sites within the coding region of the 1.5-kb subclone do not yield proteins of these sizes. In addition, it was necessary to heat denature the in vitro transcripts before translation to recover glycosylase activity in the lysates. Translation of nondenatured in vitro transcripts in the presence of [$^{35}$S] methionine produced only the 32,000- and 28,000-molecular-weight bands. The 35,000-molecular-weight band was absent. These results lend further support to the idea that the lower-molecular-weight products are the result of incomplete translation. The observation that the 35,000-molecular-weight translation product was smaller than predicted may

* Corresponding author.

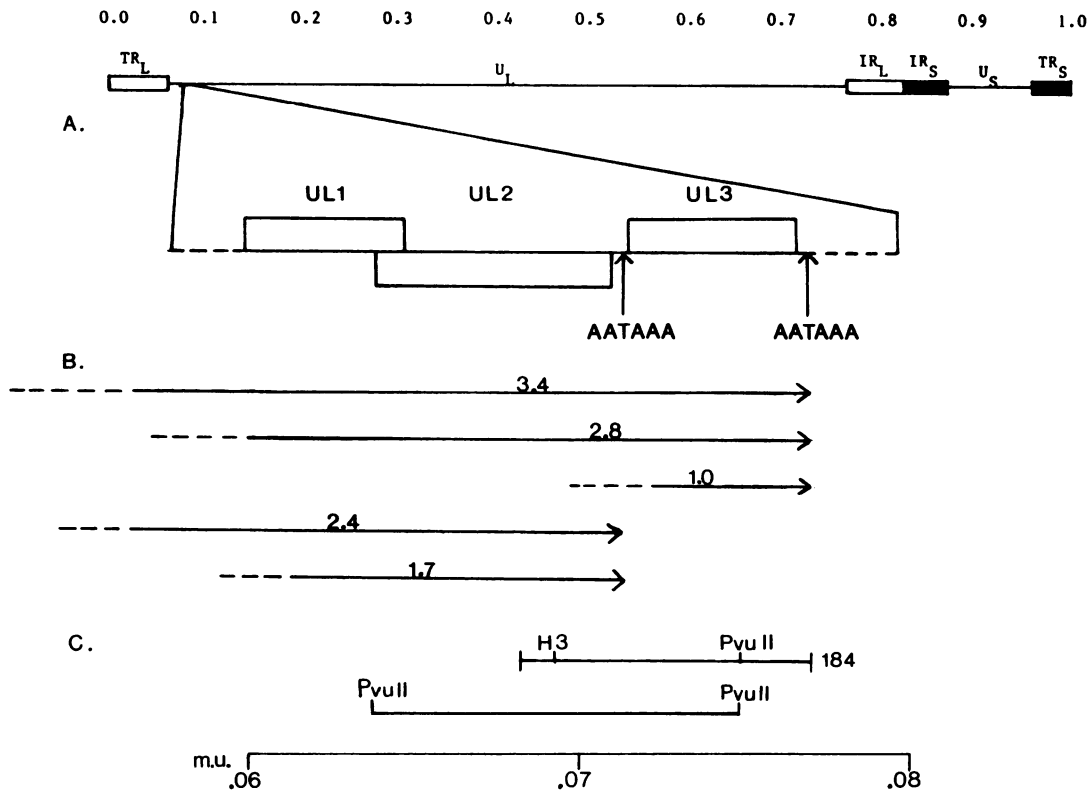
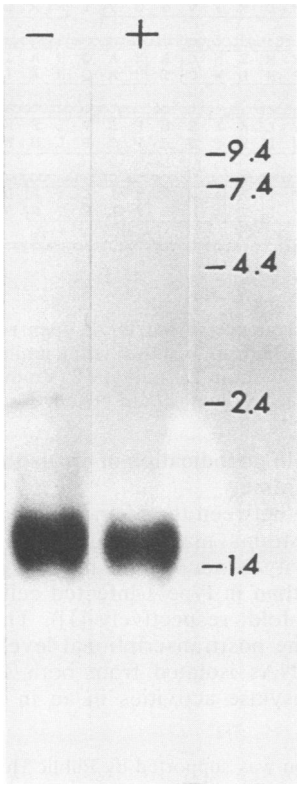



FIG. 1. (A) Location of the region encoding the HSV glycosylase on the prototypic arrangement of the HSV genome. UL-1, UL-2, and UL-3 are the three open reading frames determined by McGeoch et al. (10). (B) Locations of the five overlapping transcripts. Broken lines indicate indeterminate 5′ boundaries for the transcripts. (C) Location of the original cDNA isolate, 184 (2), with respect to this region of the genome. Also shown is the 1.5-kb *Pvu*II fragment encompassing UL-2. Open reading frames in panel A, transcripts in panel B, and cloned fragments in panel C are presented in a colinear fashion and correspond to the fractional map units (m.u.) presented at the bottom of the figure.




indicate that this in vitro transcription-translation product is also truncated. Sequence analysis of this clone indicated the potential for significant secondary structure within the transcript. This observation, along with the necessity of denaturing the in vitro transcripts, as mentioned above, may indicate that we did not obtain full-length translation. However, uracil-DNA glycosylase activity and the 35,000-molecular-weight band were both recovered from translation of heat-denatured transcripts, indicating that UL-2 is the open reading frame for the HSV glycosylase.

DNA sequence analysis was performed on the 1.5-kb subclone. This fragment of DNA was concatenated by ligation and sonicated, and fragments of 500 to 1,000 base pairs were gel purified and cloned into the pGEM vector. The shotgun cloning procedures used here were developed by Deininger (3). Plasmid sequencing was performed as described by Promega Biotec with the GemSeq system and Klenow polymerase. The nucleotide sequence was determined at least twice for each strand. Analysis of sequence data was performed with a VAX computer and programs developed by Devereux et al. (4).

Figure 4 shows the sequence data for HSV type 2 glycosy-

FIG. 2. Analysis of in vitro transcription of the 1.5-kb *Pvu*II fragment by formaldehyde-agarose gel electrophoresis. Shown are transcription products synthesized in the absence (−) or presence (+) of the RNA cap structure analog. Molecular size markers are shown in kilobases.

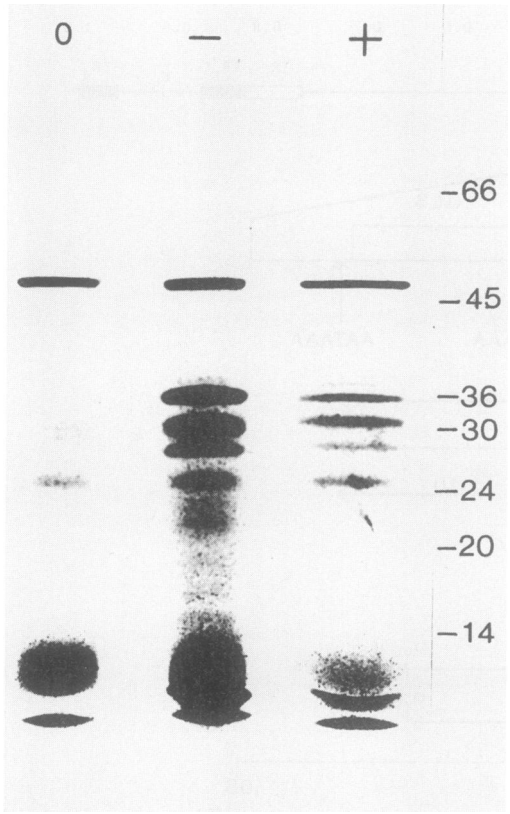



FIG. 3. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of the in vitro translation products of the transcripts shown in Fig. 2. Translation was performed in the presence of [$^{35}$S]methionine. Three protein products with molecular weights of 35,000, 32,000, and 28,000 were observed. Lanes: 0 background bands from lysates without RNA; −, translation products formed in the presence of noncapped RNA; +, translation products formed in the presence of capped RNA. Molecular size markers are in kilodaltons.

lase and also presents a comparison of the amino acid sequences of the type 1 (10) and type 2 HSV glycosylases. The type 2 glycosylase has an open reading frame of 295 amino acids, and the type 1 enzyme has an open reading frame of 335 amino acids. This results in a molecular weight of 37,700 for the type 2 enzyme and a molecular weight of 42,300 for the type 1 glycosylase. The type 2 results fit with our previous findings for the glycosylase molecular weight (2). We have not purified the type 1 glycosylase to verify whether these differences exist in infected cells. There is 74% amino acid homology between the glycosylases of the two strains of virus. The overall amino acid compositions of the two strains are predictably close. There is a twofold greater amount of basic amino acid residues relative to acidic residues. This is consistent with the basic isoelectric point of fractionated uracil-DNA glycosylase (pI, 8 to 10) (unpublished observation). Conservation of amino acid homology occurs in the central portion of the protein as well as in the 3′ end. The 5′ end, however, appears to have diverged quite significantly. In addition to verifying the sequence of the 5′ end several times, we have made a codon preference plot by using the program of Gribskov et al. (5). One of the uses for this type of plot is the detection of DNA sequencing errors resulting in the insertion or deletion of bases within the coding region. The results of this analysis, using codon usage tables from several HSV genes, reveal a continuous open

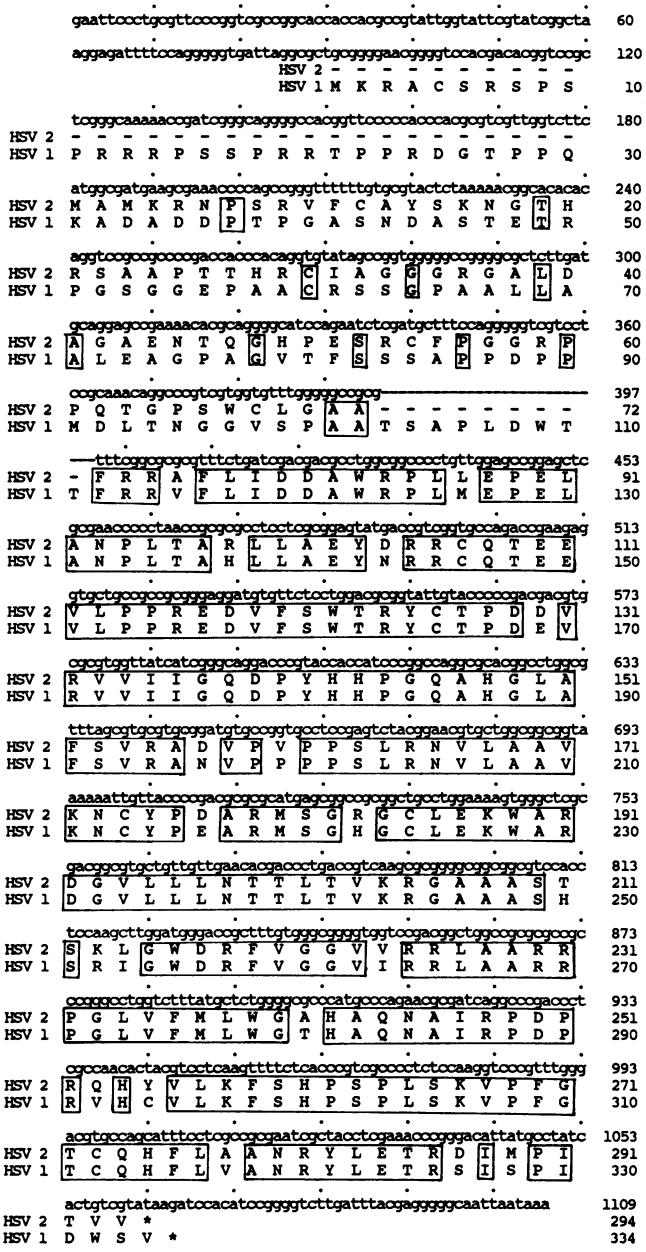


FIG. 4. DNA sequence of the UL-2 open reading frame from HSV type 2 (strain 333). Also shown is the amino acid comparison between HSV type 1 (strain 17) and type 2 (strain 333) glycosylases. Regions of amino acid homology are boxed.

reading frame with no indication of errors that would show a shift in reading frame.

The difference between the 5′ ends of the glycosylases of the two virus strains may explain the observation that glycosylase activity is induced to a much higher level in type 2-infected cells than in type 1-infected cells (10- to 15-fold versus 1.5- to 2-fold, respectively [1]). The difference appears to be at the posttranscriptional level within infected cells, since mRNAs isolated from both strains produced equivalent glycosylase activities in an in vitro translation assay (2).

This investigation was supported by Public Health Service grant CA-42605 from the National Cancer Institute.

We thank Prescott Deininger for assistance in the sequencing of the UL-2 subclone and Duncan McGeoch and co-workers for generously allowing us to use the HSV type 1 DNA sequence data before publication.

## LITERATURE CITED

1. **Caradonna, S., and Y. C. Cheng.** 1981. Induction of uracil-DNA glycosylase and dUTP nucleotidohydrolase activity in herpes simplex virus infected human cells. J. Biol. Chem. **256:**9834–9837.
2. **Caradonna, S., D. Worrad, and R. Lirette.** 1987. Isolation of a herpes simplex virus cDNA encoding the DNA repair enzyme uracil-DNA glycosylase. J. Virol. **61:**3040–3047.
3. **Deininger, P.** 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. **129:**216–223.
4. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
5. **Gribskov, M., J. Devereux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12:**539–549.
6. **Konarska, M. M., R. A. Padgett, and P. A. Sharp.** 1984. Recognition of cap structure in splicing *in vitro* of mRNA precursors. Cell **38:**731–736.
7. **Krieg, P. A., and D. A. Melton.** 1984. Functional messenger RNAs are produced by SP6 *in vitro* transcription of cloned cDNAs. Nucleic Acids Res. **12:**7057–7070.
8. **Lindahl, T.** 1982. DNA repair enzymes. Annu. Rev. Biochem. **51:**61–78.
9. **Lindahl, T., and B. Nyberg.** 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry **13:**3405–3410.
10. **McGeoch, D. J., M. A. Dalrymple, A. J. Davison, A. Dolan, M. C. Frame, D. McNab, L. J. Perry, J. E. Scott, and P. Taylor.** 1988. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. **69:**1531–1574.