Identification of a poxvirus gene encoding a uracil DNA glycosylase
(Shope fibroma virus/vaccinia virus/DNA repair/ethidium bromide fluorescence assay)

CHRIS UPTON, DAVID T. STUART*, AND GRANT MCFADDEN†

Department of Biochemistry, 576 Medical Sciences Building, University of Alberta, Edmonton, Alberta T6G 2H7 Canada

Communicated by W. K. Joklik, February 19, 1993

ABSTRACT An open reading frame, BamHI D6R, from the central highly conserved region of the Shope fibroma virus (SFV) genome was sequenced and found to have significant homology to that of uracil DNA glycosylases from a number of organisms. Uracil DNA glycosylase catalyzes the initial step in the repair pathway that removes potentially mutagenic uracil from duplex DNA. The D6R polypeptide was expressed in reticulocyte lysates programed with RNA transcribed from an expression vector containing the T7 RNA polymerase promoter. A highly specific ethidium bromide fluorescence assay of the in vitro translation product determined that the encoded protein does indeed possess uracil DNA glycosylase activity. Open reading frames from other poxviruses, including vaccinia virus (HindIII DMR) and fowlpox (D4), are highly homologous to D6R of SFV and are predicted to encode uracil DNA glycosylases. Identification of the SFV uracil DNA glycosylase provides evidence that this poxviral protein is involved in the repair of the viral DNA genome. Since this enzyme performs only the initial step required for the removal of uracil from DNA, creating an apurinic site, we suggest that other, possibly virus-encoded, repair activities must be present in the cytoplasm of infected cells to complete the uracil excision repair pathway.

The poxviridae are a family of large double-stranded DNA viruses distinguished by their high degree of autonomy from host cell functions (1). These viruses possess linear, hairpin-terminated genomes that range in size from 150 kb to >300 kb and encode several hundred gene products (2). The site of poxvirus replication within cytoplasmic viral factories is consistent with the fact that they encode essentially all of the proteins required for viral gene expression and DNA synthesis, indicating the high degree to which these complex viruses are independent of many of the host functions that other viruses must parasite (3). From the identified open reading frames (ORFs) of vaccinia virus (VV) (4) and what is known about the genomic structure of other poxviruses, it appears that the central region of the archetype poxvirus genome primarily consists of essential viral genes that function in virus replication, transcription, and progeny virion assembly. Conversely, genes that map toward either termini of the linear genome tend to be nonessential for growth in vitro but frequently have specific roles in determining host range, tissue specificity, or virulence of the virus. And large, poxviral genomes from different genera tend to manifest a high degree of sequence conservation in the genes mapping to the central region of the genome. For example, Shope fibroma virus (SFV), a tumorigenic poxvirus of rabbits (5), is known to encode DNA topoisomerases (6) and mRNA capping enzymes (7), which are essential virus genes and are closely related to homologues in other poxviruses.

One can also describe another category of poxvirus genes that encode proteins shown to be dispensable for virus replication in tissue culture, which, though not the classical virulence genes previously described, probably play important roles in enhancing the efficient propagation of the virus in infected animals. The members of this group provide functions associated with the autonomy of viral DNA replication from host nucleotide pool levels and S-phase-specific proteins. Known examples of viral genes that contribute to S-phase independence of poxviruses include thymidine kinase (8, 9), DNA ligase (10, 3), ribonucleotide reductase (12), thymidylate kinase (13), and deoxyuridine 5′-triphosphatase (dUTPase) (14).

In this paper we present the DNA sequence and predicted primary structure of a uracil DNA glycosylase gene from the Leporipoxvirus SFV and show that the expressed protein catalyzes the removal of dU from double-stranded DNA to leave an apurinic site. This is the first step in the pathway that eliminates potentially mutagenic uracil from DNA and provides evidence that poxviruses can repair DNA as part of their replicative cycle.⁡

EXPERIMENTAL PROCEDURES

DNA Sequencing. Using the chair termination method (15) with Sequenase (United States Biochemical) the viral DNA sequence was determined in both orientations from sets of overlapping deletions (16). These were created from overlapping subclones from the 6.3-kb BamHI-HindIII subfragment of the 15.5-kb SVF BamHI D fragment (7). DNA sequences were analyzed using the IG-Suite programs (IntelliGenetics) at the Molecular Mechanisms in Growth Control Computer Facility, University of Alberta. The Protein Identification Resource data base was searched using NW-ALIGN (an alignment program option of the package SEQSEE; developed by D. Wishart and R. Boyko, University of Alberta). Editing of the aligned sequences was performed with the MASE program (17).

Construction of SFV Uracil DNA Glycosylase Expression Vector. Two oligonucleotide primers were synthesized (using an ABI 392 DNA synthesizer) that were complementary to the 5′ and 3′ ends of the SFV D6R ORF. Unique Nde I and BamHI sites were incorporated into the 5′ (5′-GGCATAT-GAAGGCGGTATT-3′) and 3′ (5′-GGGATCCATT-GATTTAAAAAC-3′) primers and the D6R ORF was amplified by 25 cycles of 96°C for 30 sec, 55°C for 45 sec, and 72°C for 45 sec using VENT polymerase (New England Biolabs) from SFV genomic DNA in a Tyler Research Instruments thermal cycler (Edmonton, Alberta). The 660-bp

Abbreviations: AP, apurinic site; ccc, covalently closed circular; DMSO, dimethyl sulfoxide; dUTPase, deoxyuridine 5′-triphosphatase; FPV, fowlpox virus; HSV, herpes simplex virus; oc, open circular; ORF, open reading frame; SFV, Shope fibroma virus; VV, vaccinia virus; VZV, varicella zoster virus.

*Present address: Scripps Research Institute, Department of Molecular Biology, 10666 North Torrey Pines Road, La Jolla, CA 92037.

†To whom reprint requests should be addressed.

The sequence reported in this paper has been deposited in the GenBank data base (accession no. L01417).
product was purified, digested with Nde I and BamHI, and cloned into the similarly restricted pET3a vector (18) to create clone pXSUG-1. This process resulted in the insertion of the SFV uracil DNA glycosylase gene downstream from a bacteriophage T7 promoter.

**In Vitro Expression of SFV Uracil DNA Glycosylase.** One microgram of plasmid pXSUG-1 was linearized with BamHI, which cleaves uniquely at the 3' end of the coding sequence. After phenol extraction and ethanol precipitation, the DNA was resuspended in 5 μl of 10 mM Tris-HCl, pH 7.5/1 mM EDTA. In vitro transcription reactions were performed at 37°C for 1 hr in 40 mM Tris-HCl, pH 7.5/6 mM MgCl₂/1 mM spermidine in the presence of 500 μM (each) ATP, CTP, GTP, and UTP, 50 μM GTP, and 500 μM Mg(5')ppp(5')G cap analog. Four units of T7 RNA polymerase (Pharmacia) was added and the reaction mixtures were incubated at 37°C for 1 hr. The target DNA was then degraded by the addition of 1 unit of RNase-free DNase (Promega) and the RNA was purified by extraction with phenol/chloroform (1:1) and chloroform, followed by precipitation and resuspension in water. This RNA was used to program a nuclease-treated rabbit reticulocyte lysate (Promega) and the translation products were assayed directly for uracil DNA glycosylase activity with unprotonated lysate as control.

**Ethidium Bromide Fluorescence Assay of Uracil DNA Glycosylase Activity.** Uracil DNA glycosylase activity was measured using an established assay that is dependent on the enhanced fluorescence of ethidium bromide when intercalated into double-stranded DNA at pH 12 (19). Substrate DNA, covalently closed circular (ccc) uracil containing plasmid DNA, was prepared by propagating pUC19 (20) in *Escherichia coli* strain CJ236, which is deficient in dUTPase (dut⁻) and uracil DNA glycosylase (ung⁻) (21). Control DNA was propagated in *E. coli* HB101. Plasmids were prepared by banding in CsCl and tested for purity of the ccc species by the fluorescence assay. Reaction conditions for uracil DNA glycosylase activity were 37°C, 50 mM Tris-HCl, pH 7.5/20 mM EDTA/100 μg of heat-denatured gelatin per ml, and 1 μg of substrate in a volume of 50 μl. For each time point, a 10-μl sample was taken and added to the ethidium bromide buffer (0.5 μg of ethidium bromide per ml/0.5 mM EDTA/20 mM K₃PO₄, pH 12), the sample was excited with light at 520 nm, and the fluorescence at 600 nm was measured with a Gilson Spectra/glu fluorometer. This mixture was then heated to 95°C for 8 min and cooled and the fluorescence was remeasured. The removal of uracil residues from the substrate DNA by uracil DNA glycosylase results in AP sites, but there is no change from the pretreated fluorescence level and the level of intercalated ethidium bromide is unaltered since the phosphodiester backbone of the DNA is not broken. However, heat treatment under these assay conditions at pH 12 results in hydrolysis of the AP sites and irreversible denaturation of the substrate DNA molecules containing AP sites, whereas unreacted substrate (ccc DNA molecules) rapidly reanneals and regains most of the initial ethidium bromide fluorescence. Since intramolecular base-pairing is inhibited at pH 12, the loss of ethidium bromide fluorescence observed in uracil-containing substrate is a direct function of the introduction of AP sites by DNA glycosylase activity. Any loss of fluorescence due to non-specific nicks or either the uracil-containing or control DNA substrates was monitored by adding 2 volumes of deionized dimethyl sulfoxide (DMSO) to the samples prior to dilution into pH 12 buffer, which causes denaturation of nicked open circular (oc) DNA but has no effect upon DNA containing AP sites (22). The presence of 20 mM EDTA in the reaction mixture was shown to be sufficient to inhibit almost all non-specific nuclease activities in the reticulocyte lysates, although a low level of reticulocyte-specific uracil DNA glycosylase activity was detectable (see Fig. 4C). Thus, the loss of fluorescence at pH 12 of uracil-containing substrate DNA caused by heating, and not by DMSO treatment, is a direct reflection of the production of AP sites by uracil DNA glycosylase activity.

**RESULTS**

**DNA Sequence of SFV ORF D6R.** As part of a larger study to examine the genome organization of poxviruses from different genera, portions of the central region of SFV have been cloned and sequenced for comparison to the central region of VV. Fig. 1 indicates six closely grouped ORFs that map in a region of the viral genome known to be important for virus viability and that are conserved between SFV and VV. The percentage amino acid identity between individual ORF pairs appears to reflect the essential nature of individual proteins: ORF D1R of SFV encodes a DNA topoisomerase and is 61% identical to ORF H6R of VV (6), whereas ORFs D4L and D5R of SFV are only 36% and 33% identical to the VV ORFs D2L and D3R, which have recently been identified as core proteins (23).

The DNA sequence and translated protein sequence of SFV ORF BamHI D6R are presented in Fig. 2. The relation-

---

**Fig. 1.** Organization of viral ORFs around the uracil DNA glycosylase gene (boxed) in SFV and VV. SFV genes (BamHI map) D1R, D2R, D3R, D4L, D5R, and D6R and VV (strain Copenhagen) genes (HindIII map) H6R, H7R, D1R, D2L, D3R, and D4R have been identified as follows: DNA topoisomerase, unknown, large subunit of the mRNA capping enzyme, core protein, core protein, and uracil DNA glycosylase, respectively. TIR, terminal inverted repeat.
Fig. 2. DNA sequence and predicted amino acid sequence of the SFV uracil DNA glycosylase gene. Partial ORFs, 5' and 3' of the SFV uracil DNA glycosylase, are shown in lowercase letters.

ship between this ORF, VV HindIII D4R, and fowlpox virus (FPV) ORF D4 (24) (named for its homology to VV D4R) is readily apparent upon their alignment (Fig. 3, lines 1–3). Pairwise comparisons of the three poxviral ORFs have the following percentages of identical amino acids: SFV/VV = 68%, SFV/FPV = 57%, VV/FPV = 52%. The genetic organization of the genome surrounding this gene is also highly conserved between the Leporipoxviruses and Orthopoxviruses (Fig. 1). The order, orientation, and spacing of the seven ORFs 5' to SFV D6R are virtually identical to those in VV (4). Identification of SFV ORF D6R. Using NW-ALIGN, an exhaustive protein data base searching program, a significant relationship between the SFV ORF D6R and members of a previously well-defined family of uracil DNA glycosylases was established (Fig. 3). These enzymes constitute one of the most highly conserved families of proteins yet identified between prokaryotes and eukaryotes. Such an extent of conservation across phylogenetic boundaries suggests that it is essential for the replicative strategy of DNA-based organisms to remove potentially mutagenic dU from DNA, but it is also a reflection of the fact that the substrate for this enzyme, uracil residues in double-stranded DNA, is identical for each organism. The addition of this group of poxviruses to this group of uracil DNA glycosylases significantly reduces the number of highly conserved regions within this superfamily. In fact, within the 10 aligned sequences, there is only one triplet, two pairs, and 17 single amino acid residues that are completely conserved among all of the proteins.

Activity of the SFV Uracil DNA Glycosylase. Capped mRNA, synthesized in vitro from pXSSUG-1 using T7 RNA polymerase, was added to a reticulocyte lysate. The major mRNA-dependent translation product was shown to be 25 kDa by SDS/PAGE (not shown), similar to that predicted for the intact SFV D6R ORF. Control lysates and those containing the product of the SFV ORF D6R were tested for uracil DNA glycosylase activity by measuring the ability of the lysates to introduce AP sites into uracil-containing ccc plasmid DNA by an ethidium bromide fluorescence assay. In the mock reticulocyte lysate (Fig. 4C) a slight decrease in fluorescence is observed after heating but not after DMSO treatment of the uracil-containing substrate that is not seen with the control DNA (Fig. 4D). This demonstrates that nonspecific nicking of the control ccc DNA to form the oc plasmid DNA does not occur during the incubation; however, a low level of reticulocyte-specific uracil DNA glycosylase activity is detectable. In contrast, the reticulocyte lysate primed with the mRNA encoding the SFV D6R ORF shows a quite different result. During a 45-min incubation with the D6R-programed lysate (Fig. 4A) there is a quantitative loss of ccc substrate DNA as detected by fluorescence due to breakdown of the DNA backbone at the heat-sensitive AP site. Thus, the denatured and now separated strands of the substrate DNA cannot anneal in the low salt pH 12 buffer. This strand denaturation is not achieved with DMSO treatment (Fig. 4A), which melts nicked DNA substrates but does not DNA containing AP sites (22). Furthermore, neither DMSO nor heat is able to induce fluorescence changes in control ccc DNA (Fig. 4B), demonstrating that the observed fluorescence loss in Fig. 4A is due to the excision of uracil residues. We conclude that the SFV D6R protein specifically recognizes uracil residues in ccc template DNA and excises these bases to generate classical AP sites. A more detailed analysis of the enzymatic characteristics of the SFV uracil DNA glycosylase is presented elsewhere (40).

DISCUSSION

DNA glycosylases function to protect the integrity of the genetic information by initiating a repair pathway that removes abnormal or inappropriate bases from double-stranded DNA genomes (34). The role of uracil DNA glycosylase is to recognize and remove uracil from duplex DNA. The incorporation of dUTP into DNA is not mutagenic per se and dUTP and dTTP appear to be recognized equally well by most DNA polymerases (35). However, the spontaneous oxidative deamination of dC to dU in DNA is mutagenic since the dU that is produced by this process will result in a transversion mutation in one of the progeny genomes after the next round of DNA replication if it is not repaired. Since dU in DNA is not revealed by any perturbation of the helix, most organisms recognize dU in DNA as altered dC and therefore potentially mutagenic.

A number of DNA glycosylases, including T4 endonuclease V and E. coli endonuclease III, have been shown to possess AP endonuclease activity in addition to glycosylase activity (34). However, no uracil DNA glycosylase has yet been reported to possess an AP endonuclease function, and the SFV uracil DNA glycosylase we describe in this communication also does not break the deoxyribose-phosphate backbone under the conditions of our assay but leaves a typical heat-sensitive AP site. It is presently unknown whether a virus- or host-encoded protein performs the endonuclease step in the process of removing dU from poxvirus DNA.

Uracil DNA glycosylase is a nonessential gene in E. coli (21) and is similarly not required for the propagation of HSV in vitro (36). However, in contrast to these organisms, a DNA-negative conditional lethal mutation, ts4149 (37, 38), has recently been mapped in the D4R ORF of VV (A. M. DeLange, personal communication). This VV ORF is highly homologous (Fig. 3) to the SFV uracil DNA glycosylase and does indeed possess such activity (40). At the nonpermisive
Fig. 3. Alignment of representative uracil DNA glycosylases. Residues are boxed to indicate a minimum of 50% identity among all 10 proteins. A "+" above a residue indicates 100% identity between the three consensus proteins (101 residues). FPV D4, fowlpox virus ORF D4 (24); VV D4, VV ORF D4R (4); HSV 1, herpes simplex virus type 1 (25); HSV 2, HSV type 2 (26); VZV, varicella zoster virus (27); human (28); yeast (Saccharomyces cerevisiae) (29); E. coli (30); Strep. (Streptococcus pneumoniae) (31). The N-terminal regions of the HSV 1, HSV 2, VZV, human, and yeast proteins have not been aligned. It is notable that the human uracil DNA glycosylase (this figure) may be a mitochondrial enzyme and that two other human uracil DNA glycosylases (32, 33) that have been cloned and sequenced demonstrate no significant similarity to the proteins shown in this figure.
Fig. 4. Assay of SFV uracil DNA glycosylase translated in a reticulocyte lysate. c, Ethidium bromide fluorescence of the substrate ccc DNA after heating the samples at 96°C for 8 min; ♦, ethidium bromide fluorescence after treating the samples with 2 volumes of DMSO. The data at time 0 min are normalized to 100%. (A) Uracil-containing ccc DNA reaction with reticulocyte lysate programmed with RNA prepared from pXSUG-1. (B) Control ccc DNA reaction with reticulocyte lysate programmed with RNA prepared from pXSUG-1. (C) Uracil-containing ccc DNA reaction with control reticulocyte lysate. (D) Control ccc DNA reaction with control reticulocyte lysate.

temperature, the VV mutant ts4149 is specifically defective for DNA replication (38), indicating that this viral enzyme is essential for poxviral DNA replication and cannot be replaced by the host counterpart as is apparently the case for the VV DNA ligase (11).

It is possible that the poxviral uracil DNA glycosylase is only required to remove uracil from the replicated viral genomes. However, poxviruses also encode a dUTPase, which is not an essential gene (39), and one might expect the level of incorporation of dUTP into DNA in the absence of a removal mechanism should not abrogate bulk viral DNA synthesis. A more plausible explanation for the apparent requirement of this enzyme for viral DNA synthesis is that it performs another essential function or forms a structural component of the viral DNA replication/repair complex and is required for its activity. Although the precise role that uracil DNA glycosylase plays in poxviral DNA synthesis and genome maintenance remains to be clarified, the results of the experiments reported here suggest that DNA repair enzymes may play dual roles with the replication apparatus. Structure/function analysis of this viral protein should illuminate both processes.

Note Added in Proof. The SFV D6R sequence has recently been reported (41) and designated as D-4, but no homology or function was described for the gene.

We thank Dick Morgan for help with uracil DNA glycosylase assays, Rob Maranchuk for technical assistance, and Rita Whitford and Perry D’Obreman for DNA sequencing. Computer analysis was performed at the Molecular Mechanisms in Growth Control Computer Facility, University of Alberta. The sequence program was developed in the laboratory of B. D. Sykes, supported by the Medical Research Council Group in Protein Function, University of Alberta. This work was funded by the National Cancer Institute of Canada. G.M. is a Medical Scientist of the Alberta Heritage Foundation for Medical Research (AHFMR) and D.T.S. held an AHFMR studentship.