Nucleotide sequence of the xanthine guanine phosphoribosyl transferase gene of E. coli

Katherine Koren Richardson, Jennifer Fostel and Thomas R. Skopek

Genetic Toxicology Department, Chemical Industry Institute of Toxicology, Research Triangle Park, NC 27709, USA

Received 16 September 1983; Revised and Accepted 7 November 1983

ABSTRACT

The nucleotide sequence of the gpt gene coding for the enzyme xanthine guanine phosphoribosyl transferase has been determined. The gene codes for a protein of molecular weight 16,950. The construction of deletions in the gpt gene which can be used for the genetic analysis of mutations in the gpt gene, is described.

INTRODUCTION

The ability to clone and sequence unique DNA fragments permits the detailed study of the types of nucleotide alterations that occur during spontaneous and induced mutagenesis. Analyses of the types of sequence changes following mutation of procaryotic DNA have been reported (1,2,3). A logical extension of the work being done in procaryotes is to carry out the analogous study in mammalian cells. An excellent target gene for this type of study is the pSV2-gpt gene complex constructed by Mulligan and Berg (4,5,6). This consists of an E. coli DNA fragment containing the gpt gene which codes for the enzyme xanthine guanine phosphoribosyl transferase (XGPRT) which is flanked by SV40 DNA carrying the sequences necessary for mRNA initiation and processing in mammalian cells. This complex is expressed both in mammalian cells and in E. coli, and selective conditions exist for the isolation of cells with and without gpt function. Thackar et al. (7) and Tindall et al. (8) have described mutation systems in Chinese hamster ovary cells using the gpt gene complex of Mulligan and Berg.

A necessary starting point in the analysis of induced mutations is knowledge of the wild type sequence of the gpt gene. Single-strand sequence data for a section of the fragment containing gpt has been obtained by Mulligan and Berg (6). We present here the complete nucleotide sequence of the E. coli fragment in pSV2-gpt, which contains the gpt coding region, and describe the construction of deletion mutants in the gene.
MATERIALS AND METHODS

Bacterial strains

E. coli strain 33694 (leu B6, pro A2, rec A13, thi-1, ara-14, lac Y1, gal K2, xyl-5, mtl-1, rps L20, lambda-, sup E44, hsd S20), strain 33572 (supE, supF, thyA, metE), and strain 33526 (met Bl, lac-3 or lac Y1, gal K2, gal T22, lambda-, sup E44, hsd R2) were obtained from American Type Culture Collection (Rockville, MD). E. coli strain GP120 (pur E, delta (lac pro gpt)) was the generous gift of J. Gots (University of Pennsylvania, Philadelphia, PA). Bacteria were grown routinely in brain-heart infusion (Difco). The gpt activity of plasmid derivatives in GP120 was tested by growing the bacteria in medium containing 0.2 mg/ml MgSO_4·7H_2O, 2.0 mg/ml citric acid·H_2O, 10 mg/ml K_2HPO_4, 3.5 mg/ml NaNH_4PO_4·4H_2O, 20 mg/ml glucose, 0.1mg/ml thiamine, 10 mg/ml casamino acids, and 25 μg/ml guanine.

Nucleic Acid

Plasmid pSV2-gpt (Figure 1) was the generous gift of R. Mulligan (MIT, Cambridge, MA). Plasmid gpt2Eco (Figure 1), a derivative of SV2-gpt, was constructed by first protecting the existing EcoRl site with EcoRl methylase, cutting with PvuII, adding EcoRl linkers, and transfecting into E. coli 33694. In gpt2Eco, the gpt gene and surrounding SV40 sequences, and derivatives of this complex, can be excised by cutting with EcoRl for future transfer to other vehicles. EcoRl, HindIII, and BamHI linkers were purchased from Bethesda Research Laboratories (Gaithersburg, MD).

Enzymes

EcoRl, HindIII, BglII, KpnI, BamHI, PvuII, EcoRl methylase, T4 polynucleotide kinase, T4 DNA polymerase, T4 DNA ligase, exonuclease III, bacterial alkaline phosphatase and S1 nuclease, were purchased from Bethesda Research Laboratories (Gaithersburg, MD). T4 DNA ligase used for 32p labelling was purchased from New England Nuclear (Boston, MA). E. coli DNA polymerase I was purchased from Boehringer Mannheim (Indianapolis, IN). Reaction buffers used were those specified by the vendor.

Chemicals

Chemicals were obtained from the following sources: deoxyadenosine 5' triphosphate (α-32p, 2000 Ci/mmmole) and adenosine 5' triphosphate (gamma 32p, 5000 Ci/mmmole)-- Amersham (Arlington Heights, IL); adenosine triphosphate, piperidine, formamide, ampicillin, agarose, ammonium persulfate, cesium chloride, guanine, hypoxanthine, agarose-- Sigma Chemical Co. (St. Louis, MO); deoxynucleotide triphosphates-- P-L Biochemicals, Inc. (Milwaukee, WI); hydrazine, Eastman Kodak Co.-- (Rochester, NY); dimethylsulfate, formic acid,
Figure 1. Restriction map of pSV2-gpt. SV40 sequence is indicated by open regions, E. coli sequence by solid regions, and pBR322 sequence by a single line. SV40 early promoter is shown as a dotted line. Plasmid gpt2Eco is identical to pSV2-gpt except the PvuII site has been destroyed and replaced by an EcoRI site.

pyridine—Aldrich Chemical Co. (Milwaukee, WI); acrylamide, bis-acrylamide—Bio-Rad Laboratories (Richmond, CA).

Construction of deletions

Deletion H10 (delta HindIII-BglII) was constructed by cleaving gpt2Eco with BglII and filling in the recessed 3' ends with E. coli polymerase I and the four deoxynucleotide triphosphates at 0°C. Phosphorylated HindIII linkers were added, the plasmid cut with HindIII to release the small HindIII-BglII region, and then closed with ligase.

Deletion H20 (delta HindIII-KpnI) was constructed by cleaving gpt2Eco with KpnI and using T4 DNA polymerase in the presence of the four deoxyribonucleotide triphosphates to digest the protruding 3' strand of the KpnI cut to produce a blunt end. HindIII linkers were added, the plasmid cut with HindIII to release the small HindIII-KpnI region, and then closed with ligase.

Further deletions were constructed in an analogous fashion to that described by McKnight and Gavis (9). H20 was cleaved with HindIII and then 6 μg in 100 μl was digested with 2.75 units of exonuclease III at 25°C for various lengths of time. Under these conditions approximately 5 nucleotides per minute were digested. The single stranded regions generated were cleaved with S1 nuclease (14 units in 100 μl) for 10 minutes at 25°C. The DNA was then extracted with phenol and precipitated with ethanol to remove the S1 nuclease. To insure clean blunt ends the DNA preparations were treated with T4 DNA
polymerase and the four deoxynucleotide triphosphates. HindIII linkers were added, cut with HindIII, and ligated to close the plasmid. Ampicillin resistant GP120 transfectants were selected and their plasmid DNA was isolated for analysis. The approximate amount of DNA deleted in the new derivatives was determined by cleaving the new plasmids with HindIII and BamHI and resolving the resulting fragments on an agarose gel. The gpt-containing HindIII-BamHI truncated fragments were then ligated onto the large HindIII-BamHI fragment of gptEco to create a set of homologous deleted plasmids.

**DNA Sequencing**

HindIII or BglIII cleaved DNA was 3' labelled with E. coli DNA polymerase I as previously described (2). 5' labelling was accomplished by first treating the cleaved DNA with phosphatase, extracting the DNA with phenol and ether, and then reacting with T4 kinase and gamma $^{32}$P ATP. Single end labelled fragments were obtained by secondary cuts with either BamHI, KpnI, or EcoRI. Fragments were resolved on agarose gels and eluted.

DNA was sequenced using the method of Maxam and Gilbert (10). Thin 20% and 8% acrylamide gels (0.05 x 20 x 40 cm) and 5.5% acrylamide gels (0.05 x 20 x 80 cm) were used to resolve reliably up to 250 bases from the labeled end (11).

The strategy used to sequence the E. coli fragment in gptEco using the parental plasmid and the constructed deletions described above is presented in figure 2.

![Figure 2](image-url)  
**Figure 2.** Location of deletions in gpt region and sequencing strategy used. The deletions extend from the existing HindIII site (-197) to the approximate nucleotide position given in parentheses below the various deletions (H10-H70). The HindIII linker inserted in the missing region of each deletion (see materials and methods) was cut with HindIII and either 3' or 5' labelled for sequencing. The existing HindIII and BglIII sites in the parental plasmid were also labelled for sequencing. Arrows indicate the region sequenced from each site.
RESULTS

The gpt gene activity of the constructed deletion plasmids was in agreement with that observed by Mulligan and Berg (6). H10 (delta HindIII-BglII) possessed only residual gpt activity; GP120 carrying H10 grew more slowly than GP120 carrying gpt2Eco when grown with guanine as a purine source. H20 (delta HindIII-KpnI) and more extensive deletions had no gpt activity and would not support the growth of GP120 with guanine. The deletions constructed with exonuclease III and S1 (H30, H40, H50, H60, H70) differed in length from one another by 100-200 nucleotides.

The nucleotide sequence of the gpt-containing fragment, obtained by sequencing both strands, is presented in figure 3. The sequence of the first 345 nucleotides from the HindIII site confirms the single-stranded sequence reported by Mulligan and Berg (6). The hybrid BamH1/BglII site located at nucleotides 859-864 marks the end of the E. coli sequence. The sequence beyond this site (865-1140) is of SV40 origin and corresponds to the nucleotides 4631-4326 of the SV40 genome (12). Methylated cytosines in the sequence 5'CCme(A or T)GG 3' are present at positions 24, 268, 399, and 472 in the antisense strand and at 26, 270, 401, and 474 in the sense strand, and produced the expected aberrant cleavage products in the sequencing reactions (13).

Two extended open triplet reading frames are present in the sequenced region. One extends from the ATG at nucleotide 1 to the TAA at nucleotide 457, the other from the ATG at nucleotide 551 to the TAA at nucleotide 932 located within the SV40 sequence. The first open reading frame is preceded by a perfect consensus sequence for the Pribnow box (TATAAT) at nucleotides -71 to -66; no promoter-like sequence precedes the second open reading frame.

To demonstrate that the gpt coding region corresponds only to the first open reading frame, a plasmid that only contained that segment of the sequence was constructed. Plasmid gpt2Eco was first digested with DdeI, and the 800 nucleotide DdeI-DdeI fragment extending from within the SV40 early promoter region (approximately 55 nucleotides from the HindIII site) to nucleotide 566 was isolated. This fragment was treated with DNA polymerase I and deoxy-nucleotide triphosphates at 0°C to create blunt ends. Phosphorylated BamH1 linkers were added and the fragment was then cleaved with BamH1 and HindIII. The gpt-containing HindIII-BamH1 fragment was then cloned into the large BamH1-HindIII fragment of gpt2Eco. The new plasmid possessed the same gpt activity as gpt2Eco since GP120 containing either plasmid grew equally well in minimal medium with guanine as the purine source.
Figure 3. Nucleotide sequence of E. coli HindIII-BamHI/BglII segment of plasmid gpt2Eco containing gpt coding region.
Liu and Milman (14) have reported that the native gpt protein has an apparent molecular weight of 55,000 and consists of three subunits of molecular weight 18,600. This is consistent with the calculated molecular weight of 16,950 predicted from the sequence in figure 3.

The mechanism of expression of the gptSV2 gene product in mammalian cells is still not understood in detail. Transcription of the gene complex is expected to begin approximately 65 nucleotides from the HindIII site at the early mRNA start site and proceed through the gene. This means, however, that ribosomes attempting to translate this mRNA will encounter two AUG codons (nucleotide position -169, -106) followed by several stop codons in the same reading frame (nucleotide position -103, -88, -70, -22) before reaching the coding region for XGPRT. It is believed that in general the first AUG of eucaryotic mRNA is used for translation initiation (15). It has been suggested that in mammalian cells translation of proteins may sometimes initiate at AUG codons other than the first one encountered (6). This would then allow the expression of the E. coli gpt gene in mammalian cells. Recently, Wasylyk et al. (16) have shown that the SV40 early promoter can stimulate mRNA initiation at TATAA sequences thousands of nucleotides away. Therefore, it is possible that in mammalian cells a species of mRNA is being produced by the gptSV2 plasmid that is initiated at the TATAA sequence at nucleotide -72. If this is the case, the first AUG nucleotide encountered would be the one at the start of the coding region at nucleotide position 1.

Since the gpt gene complex of gptSV2 can express in E. coli, rodent, and human cells, and since gpt- mutants can be selected in all three (4, 5, 7, 8), an excellent opportunity exists to compare the mutagenic specificity of various compounds in different species. The gpt deletions described here can be used in genetic crosses to localize mutations in the gpt gene in future mutation analyses.

REFERENCES