Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 2916–2920, April 1990
Genetics

Cloning of three human multifunctional *de novo* purine biosynthetic genes by functional complementation of yeast mutations

(HepG2 cDNA library/phosphoribosylglycinamide synthetase/phosphoribosylaminoimidazole synthetase/phosphoribosylaminoimidazole carboxylase/tetrahydrofolate metabolism)

DAVID SCHILD*†, ANTHONY J. BRAKE‡, MICHAEL C. KIEFER‡, DARU YOUNG‡, AND PHILIP J. BARR‡

*Cell and Molecular Biology Division, Lawrence Berkeley Laboratory, Berkeley, CA 94720; and ‡Chiron Corporation, Emeryville, CA 94608

Communicated by Stanley M. Gartler, January 16, 1990 (received for review November 16, 1989)

ABSTRACT **Functional complementation of mutations in the yeast *Saccharomyces cerevisiae* has been used to clone three multifunctional human genes involved in *de novo* purine biosynthesis. A HepG2 cDNA library constructed in a yeast expression vector was used to transform yeast strains with mutations in adenine biosynthetic genes. Clones were isolated that complement mutations in the yeast *ADE2*, *ADE3*, and *ADE8* genes. The cDNA that complemented the *ade8* (phosphoribosylglycinamide formyltransferase, GART) mutation, also complemented the *ade5* (phosphoribosylglycinamide synthetase) and *ade7* [phosphoribosylaminoimidazole synthetase (AIRS; also known as PAIS)] mutations, indicating that it is the human trifunctional GART gene. Supporting data include homology between the AIRS and GART domains of this gene and the published sequence of these domains from other organisms, and localization of the cloned gene to human chromosome 21, where the GART gene has been shown to map. The cDNA that complemented *ade2* (phosphoribosylaminoimidazole carboxylase) also complemented *ade1* (phosphoribosylaminoimidazole succinocarboxamide synthetase), supporting earlier data suggesting that in some organisms these functions are part of a bifunctional protein. The cDNA that complemented *ade3* (formyltetrahydrofolate synthetase) is different from the recently isolated human cDNA encoding this enzyme and instead appears to encode a related mitochondrial enzyme.**

Although many human genes have already been cloned by a variety of techniques, there is a need for additional methods for cloning human genes. Here we present a method for cloning cDNA copies of human genes by using functional complementation of mutations in *Saccharomyces cerevisiae*. Interspecific complementation has already been used to clone genes from a variety of organisms, and the first yeast genes were cloned by complementing auxotrophic mutations in *Escherichia coli* (1). McKnight and McConaughy (2) cloned a number of yeast genes by using a cDNA library in a yeast expression vector and they suggested that this method could be extended to cloning heterologous genes by constructing suitable cDNA libraries. There have been several cases in which heterologous genes were cloned by complementing yeast mutations. Using a genomic *Drosophila* library in a yeast vector, Henikoff and co-workers (3) cloned part of the *Drosophila melanogaster* phosphoribosylglycinamide formyltransferase gene (GART) by complementing the yeast *ade8* mutation, and a human gene was cloned by complementation of a cell cycle mutation of the yeast *Schizosaccharomyces pombe* (4). In addition, heterologous cDNA libraries in yeast expression vectors have already been used to clone an alcohol dehydrogenase gene from *Aspergillus* (5), three intermediary metabolism genes from *Ustilago* (6), and a rat brain cAMP phosphodiesterase gene (7).

We have concentrated on cloning human cDNAs involved in *de novo* purine biosynthesis.§ These cDNAs are well suited for cloning by complementation, since they encode enzymes that have similar functions in yeast and mammals (Fig. 1) (8). This pathway has both biological and medical significance, and it is particularly interesting because many functions are encoded by multifunctional genes that differ between organisms (9). Although molecular analyses of the mammalian genes involved in the purine salvage and the pyrimidine biosynthetic pathways have been done, no such analysis exists for any of the mammalian *de novo* purine biosynthetic genes, except for two genes encoding folate interconverting enzymes (10, 11) and the IMP dehydrogenase gene (12). Patterson and co-workers have genetically characterized this pathway in mammals by isolating CHO cell lines with mutations in different steps (13) and have determined on which human chromosome each of the genes resides (14). Of the human purine biosynthetic genes, the GART gene is of special interest. It has been shown biochemically to contain three enzymatic activities in purine biosynthesis (15), and since the GART domain of this protein utilizes a folate cofactor, it is being studied as a target for new chemotherapeutic agents (16). This gene has also been localized to chromosome 21 and mapped to the third of this chromosome associated with Down syndrome (reviewed in ref. 17). Trisomy of this locus has been postulated to play a role in the excess of purines observed in individuals with this syndrome, perhaps contributing to its pathology (18).

MATERIALS AND METHODS

Construction of HepG2 cDNA Library. Total RNA was isolated from the human hepatoma cell line HepG2 by the guanidinium isothiocyanate/CsCl method (19). Poly(A)$^+$ RNA was isolated by fractionation over oligo(dT)-cellulose (20). First strand cDNA was synthesized as described (21) using 10 μg of poly(A)$^+$ RNA and 2.5 μg of oligo(dT) primer. Second strand cDNA was synthesized and the double-stranded cDNA was ligated to *Bst*XI adapters (22):

5′-CGAATTCAGATCTGCACA-3′

3′-GCTTAAGTCTAGAC-5′

The resulting cDNA was fractionated on Biogel A-15 and the DNA >400 base pairs (bp) was further size-selected by electrophoresis in 1% agarose. cDNA molecules larger than

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: GART, phosphoribosylglycinamide formyltransferase; PCR, polymerase chain reaction; GARS, phosphoribosylglycinamide synthetase; AIRS, phosphoribosylaminoimidazole synthetase; THF, tetrahydrofolate.

†To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M32082).

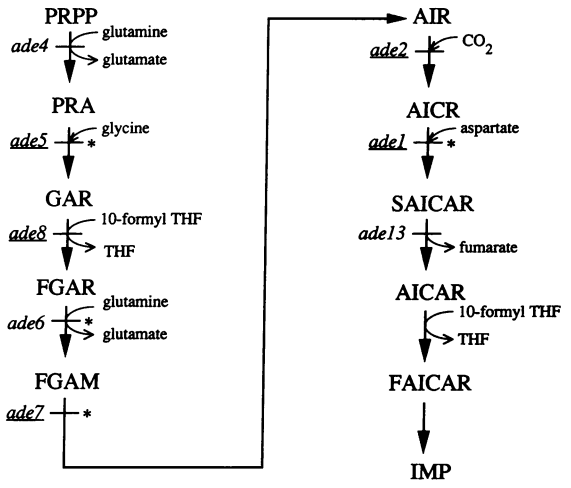

FIG. 1. Pathway of *de novo* purine biosynthesis to IMP. IMP is an intermediate in the synthesis of both AMP and GMP. *S. cerevisiae* mutations that block different steps in this pathway are shown on the left (8), and the underlined mutations are those that have been complemented in this study. In yeast, *ADE5* and *ADE7* are part of a bifunctional gene. Asterisks indicate ATP being converted to ADP and P_i. PRPP, phosphoribosylpyrophosphate; PRA, phosphoribosylamine; GAR, phosphoribosylglycinamide; FGAR, phosphoribosylformylglycinamide; FGAM, phosphoribosylformylglycinamidine; AIR, phosphoribosylaminoimidazole; AICR, phosphoribosylaminoimidazole carboxylate; SAICAR, phosphoribosylaminoimidazole succinocarboxamide; AICAR, phosphoribosylaminoimidazole carboxamide; FAICAR, phosphoribosylformaminoimidazole carboxamide.

≈2 kbp were excised, purified on an Elutip-d column (Schleicher & Schuell), and ligated to *Bst*XI-digested pAB23BX. Transformation of *E. coli* SCS-1 cells (Stratagene) resulted in a library of ≈33,000 randomly oriented clones. Plasmid pAB23BX was constructed by the insertion of a 48-bp polylinker into pAB23 (23). Questions concerning the HepG2 library should be addressed to A.J.B.

Yeast Strains, Media, and Transformation. The following strains of *S. cerevisiae* were used in this study: XS947-2A (*MATα ade1 ura3 his3 leu2 trp1*), YNN281 (*MATa ade2-1 ura3 his3 lys2 trp1*), XS964-1C (*MATα ade2-2 ura3 arg8 his3 hom3 leu2 lys2 trp1*), XS968-12B (*MATα ade3 ura3 leu2*), KSY1 (*MATα ade3-130 ura3 ser1*), XS945-4B (*MATα ade5 ura3 his3 leu2 met1 trp1*), XS927-1A (*MATa ade7 ura3 leu2*), and XS926-4B (*MATa ade8-18 ura3 his3 leu2 trp1*). Strain KSY1 was kindly supplied by Karen Shannon (24), YNN281 was obtained from the Yeast Genetic Stock Center (Berkeley, CA), and all other strains were constructed for this study by using strains from the Yeast Genetic Stock Center. Standard yeast synthetic complete media lacking uracil (SC-ura) and lacking adenine (SC-ade) were used for selecting and screening transformants. A modification of the lithium acetate method was used for yeast transformation (25).

Characterization of Plasmids and DNA Sequencing. Total DNA was isolated from yeast transformants (26) and used to transform *E. coli* strain DH5α, selecting for ampicillin resistance. Dideoxynucleotide sequencing was performed using the Sequenase system (United States Biochemical). Both double-stranded and single-stranded sequencing from 4- or 6-base restriction sites were done with subclones in pUC19 and M13. The sequence (see Fig. 3) was confirmed by sequencing both strands.

Polymerase Chain Reaction (PCR) Technique. Amplification by PCR (27) was done with the Perkin–Elmer/Cetus GeneAmp DNA kit and thermal cycler and the protocols included. Amplification was carried out for 30 cycles of 60 sec at 94°C, 70 sec at 60°C, and 30 sec at 72°C, and the products were run on a 3% NuSieve GTG agarose gel (FMC).

RESULTS

Complementation of *ade2*, *ade3*, and *ade8*. The HepG2 cDNA library was constructed in the yeast expression vector pAB23BX, which contains the strong constitutive glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter, the yeast 2-μm replication origin, and the selectable yeast *URA3* gene (Fig. 2*A*). Lithium acetate (LiAc) transformable yeast strains containing both the *ura3* mutation and an adenine mutation were constructed. DNA from the HepG2 library was transformed into strains YNN281, XS968-12B, and XS926-4B, and transformants were selected on SC-ura plates. The resulting colonies were replica plated to SC-ade plates to screen for transformants that might contain an *ade*-complementing plasmid. Ade^+ transformants were further tested for whether they contained a complementing plasmid, rather than an Ade^+ revertant containing a random plasmid. Since yeast plasmids are frequently lost during nonselective growth, the coordinate appearance of some Ade^- and Ura^- segregants from an Ade^+ Ura^+ transformant was evidence that it contained a plasmid complementing the *ade* mutations. All of the Ade^+ transformants of the *ade8* strain contained complementing plasmids as expected, since the *ade8-18* allele is a deletion (28). For *ade8*, the frequency

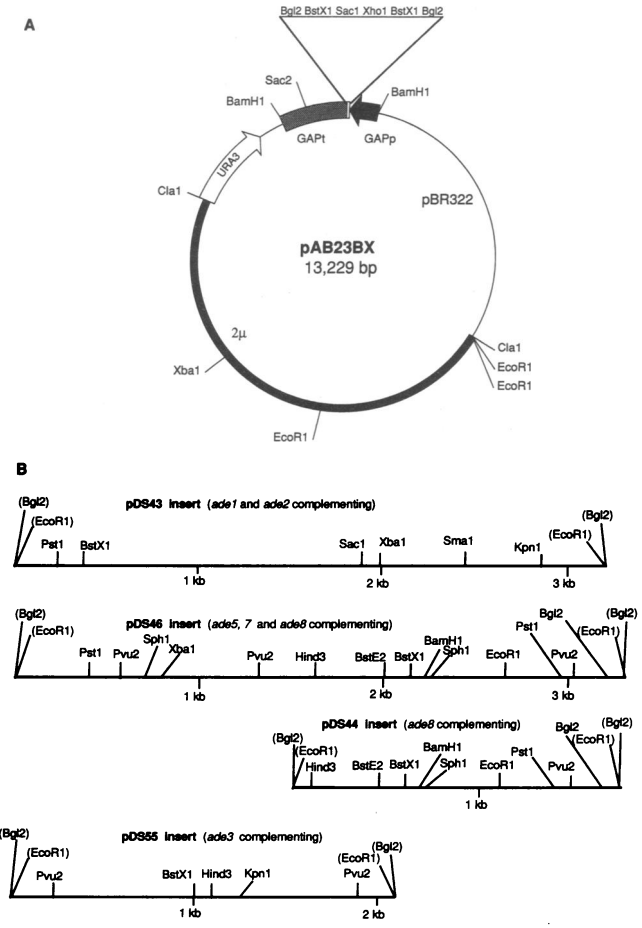

FIG. 2. Restriction maps. (*A*) The pAB23BX yeast expression vector. (*B*) The human cDNAs cloned. *A* and *B* are not drawn to the same scale. In *B*, the yeast promoter is on the left side of the inserts shown, and the flanking *Bgl* II and *Eco*RI sites are not encoded by the cDNAs. The pDS44 insert has been positioned under the 3′ end of the pDS46 insert. The pDS55 insert also contains several *Pst* I sites that have not been mapped.

of Ade$^+$ transformants was ≈1 in 4000 Ura$^+$ transformants; the frequencies for *ade3* and *ade2* were similar but more difficult to estimate since not all Ade$^+$ colonies were tested for the presence of complementing plasmids. We were unsuccessful in complementing the yeast *ade4, ade6, ade12,* and *gua1* mutations with this cDNA library.

Characterization of the Cloned cDNAs. For each *ade* mutation complemented, one to three plasmids were isolated and the size of the cDNA inserts was determined. Two insert sizes were found for *ade8*-complementing plasmids (1.7 and 3.3 kb) and for *ade3*-complementing plasmids (1.6 and 2.1 kb). Detailed restriction maps for each of these cDNA inserts were determined (Fig. 2*B*), except for the pDS56 plasmid (1.6-kb insert), which complemented *ade3* only weakly and failed to complement an *ade3* deletion. The restriction map of the pDS44 insert is identical to the 3′ end of the pDS46 insert.

The isolated plasmid DNAs were used to retransform the yeast strains from which they were originally isolated, selecting on SC-ura plates. In all cases, >99% of these Ura$^+$ transformants were also found to be Ade$^+$ (Table 1). Between 0.1% and 1% of the plasmids did not complement the *ade* mutation, possibly because of plasmid rearrangements in yeast resulting from interactions of the complete 2-μm sequences on pAB23BX with endogenous 2-μm sequences. A 100- to 1000-fold decrease in transformants recovered was observed when the transformation mixes were plated directly on SC-ade, indicating a phenotypic lag in expression of the heterologous cDNAs. Since the human GART enzyme has been shown to be part of a trifunctional enzyme (15), we tested the ability of plasmids pDS44 and pDS46 to complement the yeast *ade5* or *ade7* mutations in these other functions. Plasmid pDS46, but not pDS44, did complement both mutations, although the complementation was weaker than for *ade8* (Table 1). Since *ade2-1* is an ochre-suppressible allele, we confirmed that plasmid pDS43 also complemented the nonsuppressible allele *ade2-2*. The pDS43 plasmid was also tested for its ability to complement an *ade1* mutation, since it has been reported that in chickens the activities of both of these genes reside on a single polypeptide (29), and complementation of *ade1* was observed.

Characterization of the GART Domain Sequence. To confirm that a human GART cDNA had been cloned, part of the cDNA insert in plasmid pDS46 was sequenced for comparison to related genes from other organisms. Since the GART domain of the *Drosophila* GART gene is at the 3′ end, we sequenced the last third of this cDNA and found, as expected, that it encoded the GART domain (Fig. 3). Unexpectedly, the cDNA lacked a poly(A) tail; the sequence shown is followed by the linker sequence. Since the synthesis of the cDNAs in this library was primed by oligo(dT), the 3′ end of this cDNA was presumably degraded by contaminating exonucleases prior to ligation into the vector. A potential polyadenylylation consensus sequence (30) was found 11 bases from the end of this cDNA.

Table 1. Complementation of different yeast *ade* mutations by the isolated plasmids

Mutation	pDS43	pDS44	pDS46	pDS55
ade1	+++	−	−	
*ade2-1**	+++	−	−	
ade2-2	+++	−	−	
ade3				+++
ade3-130†				+++
ade5	−	−	+	
ade7	−	−	++	
ade8–18†	−	+++	+++	

The underlined results indicate confirmation, using isolated plasmids, of the original complementation. +++, Normal growth on SC-ade plates after 24 hr; ++, near normal growth on SC-ade plates after 48 hr; +, very slow growth on SC-ade plates, but some growth clearly visible by 72 hr; − no detectable growth on SC-ade plates by 72 hr.

*Suppressible by nonsense suppressors.

†Deletion mutations.

Analysis of the DNA sequence revealed only one long open reading frame and this was encoded in the strand that is presumed to be transcribed in yeast because of the orientation of the cDNA insert in the yeast expression vector. Comparison of the amino acid sequence of this open reading frame with the sequence of part of the *Drosophila* GART gene (31, 32) and with the last part of the yeast *ADE5,7* (32) and all of the *ADE8* (28) genes revealed significant similarity (Fig. 4). Much of this homology is also shared by both the *E. coli* and *Bacillus subtilis* GART-encoding genes (33, 34). The GART domains have been aligned to coincide with the second amino acid of the yeast GART-encoding protein, since in yeast, as in most eukaryotes, the initial methionine residue is frequently cleaved from the protein (35). The 5′ start of the pDS46 cDNA insert was also sequenced in one direction (data not shown); the first ATG codon occurs at position 73 and the predicted amino acid sequence of the start of this protein is MAARVLIIGSGGREHTLAWKLAQSH-

```
Ile Pro Arg Val Phe Ser Trp Leu Gln Gln Glu Gly His Leu Ser
ATC CCC AGG GTC TTC TCA TGG TTG CAG CAG GAA GGA CAC CTC TCT   45

Glu Glu Glu Met Ala Arg Thr Phe Asn Cys Gly Val Gly Ala Val
GAG GAA GAG ATG GCC AGA ACA TTT AAC TGT GGG GTT GGC GCT GTC   90

Leu Val Val Ser Lys Glu Gln Thr Glu Gln Ile Leu Arg Gly Ile
CTT GTG GTA TCA AAG GAG CAG ACA GAG CAG ATT CTG AGG GGT ATC  135

Gln Gln His Lys Glu Glu Ala Trp Val Ile Gly Ser Val Val Ala
CAG CAG CAC AAG GAA GAA GCC TGG GTG ATT GGC AGT GTG GTT GCA  180

Arg Ala Glu Gly Ser Pro Arg Val Lys Val Lys Asn Leu Ile Glu
CGA GCT GAA GGT TCC CCA CGT GTG AAA GTC AAG AAT CTG ATT GAA  225

Ser Met Gln Ile Asn Gly Ser Val Leu Lys Asn Gly Ser Leu Thr
AGC ATG CAA ATA AAT GGG TCA GTG TTG AAG AAT GGC TCC CTG ACA  270

Asn His Phe Ser Phe Glu Lys Lys Lys Ala Arg Val Ala Val Leu
AAT CAT TTC TCT TTT GAA AAA AAA AAG GCC AGA GTG GCT GTC TTA  315

Ile Ser Gly Thr Gly Ser Asn Leu Gln Ala Leu Ile Asp Ser Thr
ATA TCT GGA ACA GGA TCG AAC CTG CAA GCA CTT ATA GAC AGT ACT  360

Arg Glu Pro Asn Ser Ser Ala Gln Ile Asp Ile Val Ile Ser Asn
CGG GAA CCA AAT AGC TCT GCA CAA ATT GAT ATT GTT ATC TCC AAC  405

Lys Ala Ala Val Ala Gly Leu Asp Lys Ala Glu Arg Ala Gly Ile
AAA GCC GCA GTA GCT GGG TTA GAT AAA GCG GAA AGA GCT GGT ATT  450

Pro Thr Arg Val Ile Asn His Lys Leu Tyr Lys Asn Arg Val Glu
CCC ACT AGA GTA ATT AAT CAT AAA CTG TAT AAA AAT CGT GTA GAA  495

Phe Asp Ser Ala Ile Asp Leu Val Leu Glu Glu Phe Ser Ile Asp
TTT GAC AGT GCA ATT GAC CTA GTC CTT GAA GAG TTC TCC ATA GAC  540

Ile Val Cys Leu Ala Gly Phe Met Arg Ile Leu Ser Gly Pro Phe
ATA GTC TGT CTT GCA GGA TTC ATG AGA ATT CTT TCT GGC CCC TTT  585

Val Gln Lys Trp Asn Gly Lys Met Leu Asn Ile His Pro Ser Leu
GTC CAA AAG TGG AAT GGA AAA ATG CTC AAT ATC CAC CCA TCC TTG  630

Leu Pro Ser Phe Lys Gly Ser Asn Ala His Glu Gln Ala Leu Glu
CTC CCT TCT TTT AAG GGT TCA AAT GCC CAT GAG CAA GCC CTG GAA  675

Thr Gly Val Thr Val Thr Gly Cys Thr Val His Phe Val Ala Glu
ACC GGA GTC ACA GTT ACT GGG TGC ACT GTA CAC TTT GTA GCT GAA  720

Asp Val Asp Ala Gly Gln Ile Ile Leu Gln Glu Ala Val Pro Val
GAT GTG GAT GCT GGA CAG ATT ATT TTG CAA GAA GCT GTT CCC GTG  765

Lys Arg Gly Asp Thr Val Ala Thr Leu Ser Glu Arg Val Lys Leu
AAG AGG GGT GAT ACT GTC GCA ACT CTT TCT GAA AGA GTA AAA TTA  810

Ala Glu His Lys Ile Phe Pro Ala Ala Leu Gln Leu Val Ala Ser
GCA GAA CAT AAA ATA TTT CCT GCA GCC CTT CAG CTG GTG GCC AGT  855

Gly Thr Val Gln Leu Gly Glu Asn Gly Lys Ile Cys Trp Val Lys
GGA ACT GTA CAG CTT GGA GAA AAT GGC AAG ATC TGT TGG GTT AAA  900

Glu Glu  *
GAG GAA TGA AGCCTTTTAATTCAGAAATGGGGCCAGTTTAGAAAGAATTATTTGCT  956

GTTTGCATGGTGGTTTTTTATCATGGACTTGGCCCAAAAGAAAAACTGCTAAAAGACAA 1015

AAAAGACCTCACCCTTACTTCATCTATTTTTTTAATAAATAGAGACTCAC          1065
```

FIG. 3. DNA sequence of part of the AIRS domain and all of the GART domain of the human GART gene. Sequence from the third base of the *Bam*HI site to the 3′ end of the cDNA insert in pDS46. Homology to the GART sequences from other organisms starts at base ≈297 (see Fig. 4). A string of 10 consecutive adenine bases (287–296) adjacent to the start of the GART domain have been underlined. The underlined bases 712–731 and the complement of bases 894–913 were used as PCR primers. The underlined AATAAA sequence may represent the polyadenylylation signal, although the poly(A) tail is absent.

A. Comparison of part of AIRS domain:

```
D.m. LPPVFAWLAAAGNISSTELQRTYNCGLGMVLVVAPTEVEDVLKELRYPQ-RAAVVGEVVARKDPKKSQVVVQNFEASLARTQKMLSQ----------RR
H.s. IPRVFSWLQQEGHLSEEEMARTFNCGVGAVLVVSKEQTEQILRGIQQHKEEAWVIGSVVARAEGSPR-VKVKNLIESMQINGSVLKNGSLTNHFSFEKKK
S.c. VPRVFKWFGQAGNVPHDDILRTFNMGVGMVLIVKRENVKAVCDSLTEEGEIIWELGSLQERPKDAPGCVIENGTKLY
```

B. Comparison of GART domain:

```
D.m. -KRVAVLISGTGSNLQALIDATRDSAQGIHADVVLVISNKPGVLGLQRATQAGIPSLVIS----------HKDFASREVYDAELTRNLKAARVDLICLAGFMRVLSAPFV
H.s. -ARVAVLISGTGSNLQALIDSTREPN--SSAQIDIVISNKAAVAGLDKAERAGIPTRVIN----------HKLYKNRVEFDSAIDLVLEEFSIDIVCLAGFMRILSGPFV
S.c. MARIVVLISGSGSNLQALIDAQKQGQLGEDAHIVSVISSSKKAYGLTRAADNNIPTKVCSLYPYTKGIAKEDKAARAKARSQFENDLAKLVLEEKPDVIICAGWLLILGSTFLS

D.m. REWRGRLVNIHPSLLPKYPG-LHVQKQALEAGEKE-----SGCTVHFVDEGVDTGAIIVQAAVPILPDDDEDSLTQ-RIHKAEHWAFPRALAMLVNGTALISPEVSSQ
H.s. QKWNGKMLNIHPSLLPSFKG-SNAHEQALETGVTV-----TGCTVHFVAEDVDAGQIILQEAVPVKRGDTVATLSE-RVKLAEHKIFPAALQLVASGTVQLGENGKICWVKEE
S.c. QLQSVPILNLHPALPGCFDGTTHAIEMAWRKCQDENKPLTAGCMVHYVIEEVDKGEPLVVKKLEIIPGEETLEQYEQRVHDAEHIAIVEATYKVLQQLHK
```

FIG. 4. Amino acid sequence comparisons (single-letter code) of part of the AIRS domain and all of the GART domain. The amino acid sequences are for parts of the *D. melanogaster* (*D.m.*) GART gene (*A* and *B*) and the *S. cerevisiae* (*S.c.*) *ADE5,7* gene (*A*) and *ADE8* gene (*B*), and the human (*H.s.*) sequence (*A* and *B*).

HVKQVLVAPGNAG. The underlined amino acids are identical to those of the *Drosophila* phosphoribosylglycinamide synthetase (GARS; also known as PRGS) domain.

Localization of the cDNA Insert in pDS46 to Chromosome 21. Since the GART gene has previously been shown to map on human chromosome 21, we tested whether the presumptive GART gene cloned in this study also mapped to this chromosome. Using our sequence of the GART domain, three sets of PCR primers were selected to amplify ≈200-bp regions and were used to amplify both pDS46 and total human genomic DNA. One set (Fig. 3) amplifies the same sized region on the cDNA and genomic DNA and was used to examine whether this cDNA is encoded on chromosome 21. As shown in Fig. 5, these primers amplify an ≈200-bp region on DNA from the mouse cell line HDm-15, which contains human chromosomes 4, 11, and 21 (36), as they do with human genomic DNA. We originally thought that HDm-15 contained only human chromosome 21, and we have recently repeated this observation with the SCC16-5 mouse cell line, which contains only human chromosome 21 (37) (data not shown). Since amplification with total mouse DNA does not result in an ≈200-bp product, these results indicate that the gene cloned, or at least a copy of it, resides on chromosome 21. Larger bands are observed for mouse, HDm-15, and SCC16-5 DNAs, which are absent from amplification of human DNA. These may represent amplification from the mouse GART gene or from some random region, as is frequently seen with PCR.

DISCUSSION

Three different human *de novo* purine biosynthetic genes have been cloned by functional complementation of mutations in the yeast *S. cerevisiae*, using a HepG2 cDNA library constructed in a yeast expression vector. The best characterized of these cDNAs is one that encodes the trifunctional GART locus. The protein containing the GART enzymatic activity in mammalian cells, including HeLa cells (15), has previously been shown to also contain the phosphoribosylaminoimidazole synthetase (GARS and AIRS; also known as PAIS) activities, and all three activities have been mapped to the same region on human chromosome 21 (17). It is assumed that these activities are encoded by a human trifunctional gene, as has been shown for *Drosophila* (32). Our work supports this model, since the cDNA isolated complements yeast mutations in each of these three domains. Independently, V. C. Sheffield and D. R. Cox (personal communication) have used published sequences to synthesize PCR primers for amplification of a region in the AIRS domain from various organisms and cloned part of the human GART gene by using these primers; they were also able to amplify the same sized fragment from our cDNA clone with these primers.

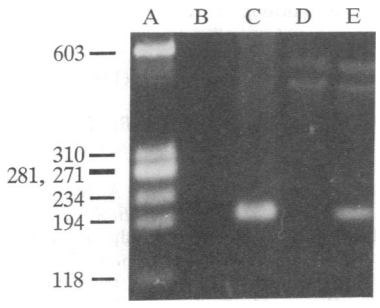

FIG. 5. PCR amplification of part of the human GART gene. Lanes: A, ϕX174 replicative form *Hae* III DNA size standards; B–E, two primers derived from the sequenced GART domain (see Fig. 3) were used for amplification with the following DNA samples. Lane B, no DNA control; lane C, human DNA; lane D, mouse DNA; lane E, HDm-15 (mouse cell line containing human chromosomes 4, 11, and 21).

The cloned GART cDNA has already been useful in determining some of the properties of this gene. The order of the three functions of the mammalian GART gene has not previously been definitively determined; our results indicate that the order is 5′-GARS-AIRS-GART-3′, the same as in the *Drosophila* gene (32). The size of the predicted open reading frame (≈3.1 kb) strongly suggests that the human gene encodes only one copy of the middle AIRS domain, unlike *Drosophila*, where this domain is present in two tandem copies (32). The size of the human protein (≈112 kDa) (15), as compared with that of *Drosophila* has also suggested a single copy of the AIRS domain in the human protein (9). In both *D. melanogaster* and *D. pseudoobscura* the GART gene encodes two different sized transcripts, one containing all three functions and one containing only the GARS domain, and a cuticle protein gene is encoded in an intron of the GART locus (38–40). It is not known whether the human GART gene encodes a second sized transcript or contains any nested gene. The shorter cDNA isolated, which complemented only an *ade8* (GART) mutation, may represent a second sized transcript, although it is likely that it is a cloning artifact.

An intriguing aspect of the DNA sequence is the presence of 10 adenine residues in the encoded strand only 1 bp from the start of the GART domain. Such a sequence is not present in the *Drosophila* GART gene (31, 32) and may only represent a chance occurrence, possibly combined with a requirement for basic amino acids in this region. Alternatively, this stretch of adenine residues may be a remnant from the evolution of this multifunctional gene, possibly representing a poly(A) tail of a GARS-AIRS bifunctional mRNA, which through reverse transcriptase was placed next to the GART domain. The favored model for multifunctional genes is that they evolved

from the fusion of adjacent genes in a polycistronic operon (9). If the internal poly(A) motif observed here is present in other multifunctional genes, it might indicate another evolutionary mechanism.

A second cDNA isolated complements mutations in both the yeast *ADE1* and *ADE2* genes and therefore probably encodes phosphoribosylaminoimidazole succinocarboxamide synthetase and phosphoribosylaminoimidazole carboxylase. In yeast, these two functions are encoded by genes on different chromosomes, but these enzymatic activities have been shown to copurify from chicken livers (29), and our clone suggests that a similar bifunctional enzyme does occur in humans. A third cDNA was isolated that complements mutations in the trifunctional formyltetrahydrofolate synthetase gene (*ADE3*), which interconverts different forms of tetrahydrofolate (THF), with the 10-formyl-THF synthesized acting as a cofactor in the GART reaction and in a later step of purine biosynthesis (Fig. 1). While our clone was being analyzed, the cloning and sequencing of the human gene encoding this enzyme was published (10). Surprisingly, both the size and the restriction map of our cDNA are clearly different than that of the published human clone. The restriction map of our cDNA was shown by R. E. MacKenzie to be the same as that of a human cDNA his group had recently independently isolated (ref. 41; personal communication), which encodes NAD-dependent methylene-THF dehydrogenase-methenyl-THF cyclohydrolase. Although this enzyme also produces 10-formyl-THF, it lacks one of the trifunctional activities of the formyl-THF synthetase gene, uses NAD instead of NADP, is found primarily in transformed cells (42), and is mitochondrially localized (43). Yeast has a separate mitochondrially localized form of the NADP-dependent trifunctional enzyme, but this gene (*MIS1*) fails to complement the *ade3* mutation even when this gene is present on a high copy number plasmid (24). This implies that the 10-formyl-THF produced in the mitochondria cannot substitute for the cytosolic form. Possible explanations of why the human gene complements the *ade3* mutation are that the human mitochondrial-targeting sequence may not function properly in yeast or that the human protein is present in very high levels because of the high copy number of the vector and its strong promoter.

The method of cloning by complementing yeast mutations using a heterologous cDNA library is a useful technique for a number of reasons. Complementation usually selects for complete coding regions during the initial cloning step, and once a cDNA has been cloned by this method, it can be studied in yeast without further subcloning. Studies examining different aspects of a gene are simplified by the ability to use yeast complementation as an assay for activity of *in vitro*-mutagenized cDNAs. In addition, the heterologous proteins should frequently already be overexpressed in yeast, facilitating their isolation and characterization.

Most of all we wish to acknowledge the encouragement and useful comments of Robert K. Mortimer, without whose support this research could not have been done. We also wish to thank Alfred J. McElroy (Sierra College, Rocklin, CA) and Paula Shah for help in restriction mapping; Maren Bell for excellent technical assistance at the beginning of this research; Keith A. Andrews and Edward M. Rubin for help with the PCR experiments; Robert K. MacKenzie, Val C. Sheffield, and David R. Cox for sharing unpublished data with us; and David Patterson for extremely useful comments on the manuscript. This work was supported by National Institutes of Health Biomedical Research Support Grant RR05918 funds (to D.S.) and Public Health Service Grant GM30990 (R. K. Mortimer, Principal Investigator), and by Lawrence Berkeley Laboratory institutional funds from the Office of Health and Environmental Research, Office of Energy Research, U.S. Department of Energy, under Contract DE-AC03-76SF00098.

1. Carbon, J., Ratzkin, B., Clarke, L. & Richardson, D. (1977) *Genet. Interact. Gene Transfer: Brookhaven Symp. Biol.* **29**, 277–296.
2. McKnight, G. L. & McConaughy, B. L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4412–4416.
3. Henikoff, S., Tatchell, K., Hall, B. D. & Nasmyth, K. A. (1981) *Nature (London)* **289**, 33–37.
4. Lee, M. G. & Nurse, P. (1987) *Nature (London)* **327**, 31–35.
5. McKnight, G. L., Kato, H., Upshall, A., Parker, M. D., Saari, G. & O'Hara, P. J. (1985) *EMBO J.* **4**, 2093–2099.
6. Kronstad, J. W., Wang, J., Covert, S. F., Holden, D. W., McKnight, G. L. & Leong, S. A. (1989) *Gene* **79**, 97–106.
7. Colicelli, J., Birchmeier, C., Michaeli, T., O'Neill, K., Riggs, M. & Wigler, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3599–3603.
8. Jones, E. W. & Fink, G. R. (1982) in *Molecular Biology of the Yeast Saccharomyces cerevisiae: Metabolism and Gene Expression*, eds. Strathern, J. N., Jones, E. W. & Broach, J. R. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 181–299.
9. Henikoff, S. (1987) *BioEssays* **6**, 8–13.
10. Hum, D. W., Bell, A. W., Rozen, R. & MacKenzie, R. E. (1988) *J. Biol. Chem.* **263**, 15946–15950.
11. Belanger, C. & MacKenzie, R. E. (1989) *J. Biol. Chem.* **264**, 4837–4843.
12. Collart, F. R. & Huberman, E. (1988) *J. Biol. Chem.* **263**, 15769–15772.
13. Patterson, D. (1985) in *Molecular Cell Genetics*, ed. Gottesman, M. M. (Wiley, New York), pp. 267–309.
14. Van Keuren, M. L., Hart, I. M., Kao, F. T., Neve, R. L., Bruns, G. A. P., Kurnit, D. M. & Patterson, D. (1987) *Cytogenet. Cell Genet.* **44**, 142–147.
15. Daubner, S. C., Young, M., Sammons, R. D., Courtney, L. F. & Benkovic, S. J. (1986) *Biochemistry* **25**, 2951–2957.
16. Fitzhugh, A. L., Chabner, B. A., Sholar, P. W., Baram, J. & Allegra, C. J. (1987) *Natl. Cancer Inst. Monogr.* **5**, 153–157.
17. Hards, R. G., Benkovic, S. J., Van Keuren, M. L., Graw, S. L., Drabkin, H. A. & Patterson, D. (1986) *Am. J. Hum. Genet.* **39**, 179–185.
18. Patterson, D., Graw, S. & Jones, C. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 405–409.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
20. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
21. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–289.
22. Aruffo, A. & Seed, B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8573–8577.
23. Fuller, R. S., Brake, A. & Thorner, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1434–1438.
24. Shannon, K. W. & Rabinowitz, J. C. (1988) *J. Biol. Chem.* **263**, 7717–7725.
25. Kramer, B., Kramer, W., Williamson, M. S. & Fogel, S. (1989) *Mol. Cell. Biol.* **9**, 4432–4440.
26. Lorincz, A. (1984) *Focus* **6:4**, 11.
27. Saiki, R. K., Gyllensten, U. B. & Erlich, H. A. (1988) in *Genome Analysis: A Practical Approach*, ed. Davies, K. E. (IRL, Oxford, U.K.), pp. 141–152.
28. White, J. H., DiMartino, J. F., Anderson, R. W., Lusnak, K., Hilbert, D. & Fogel, S. (1988) *Mol. Cell. Biol.* **8**, 1253–1258.
29. Patey, C. A. H. & Shaw, G. (1973) *Biochem. J.* **135**, 543–545.
30. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41**, 349–359.
31. Henikoff, S. & Furlong, C. E. (1983) *Nucleic Acids Res.* **11**, 789–800.
32. Henikoff, S. (1986) *J. Mol. Biol.* **190**, 519–528.
33. Smith, J. M. & Daum, H. A., III (1987) *J. Biol. Chem.* **262**, 10565–10569.
34. Ebbole, D. J. & Zalkin, H. (1987) *J. Biol. Chem.* **262**, 8274–8287.
35. Sherman, F., Stewart, J. W. & Tsunasawa, S. (1985) *BioEssays* **3**, 27–31.
36. Lugo, T. G., Handelin, B., Killary, A. M., Housman, D. E. & Fournier, R. E. K. (1987) *Mol. Cell. Biol.* **7**, 2814–2820.
37. Janssen, J. W. G., Collard, J. G., Tulp, A., Cox, D., Millington-Ward, A. & Pearson, P. (1986) *Cytometry* **7**, 411–417.
38. Henikoff, S., Sloan, J. S. & Kelly, J. D. (1983) *Cell* **34**, 405–414.
39. Henikoff, S., Keene, M. A., Fechtel, K. & Fristrom, J. W. (1986) *Cell* **44**, 33–42.
40. Henikoff, S. & Eghtedarzadeh, M. K. (1987) *Genetics* **117**, 711–725.
41. Peri, K. G., Belanger, C. & MacKenzie, R. E. (1989) *Nucleic Acids Res.* **17**, 8853.
42. Mejia, N. R. & MacKenzie, R. E. (1985) *J. Biol. Chem.* **260**, 14616–14620.
43. Mejia, N. R. & MacKenzie, R. E. (1988) *Biochem. Biophys. Res. Commun.* **155**, 1–6.