The full length coding sequence of rat liver androsterone UDP-glucuronyltransferase cDNA and comparison with other members of this gene family

Michael R.Jackson and Brian Burchell

Department of Biochemistry, The University, Dundee DD1 4HN, UK

Received 16 October 1985; Revised and Accepted 12 December 1985

ABSTRACT

Cloned cDNAs coding for hepatic UDP-glucuronyltransferase (UDPGT) have been isolated from a rat liver cDNA library in the expression vector bacteriophage λ gt11 using anti-UDPGT antibodies. Four different mRNAs have been identified by sequencing of 15 UDPGT cDNA clones. The sequences of the four classes of cDNA were determined to be 85-95% homologous. Restriction fragments were isolated from the cDNA in each class and used as class specific probes. Hybridisation of these probes to northern blots of total RNA prepared from the livers of normal and genetically deficient Wistar rats identified the cDNA in class 4 with androsterone UDPGT. Translation of the cDNA sequence of clone rlug 23, the longest member of class 4, allowed determination of the complete amino acid sequence of androsterone UDPGT.

INTRODUCTION

Hepatic microsomal UDP-glucuronyltransferase (UDPGT) is of major importance in the conjugation and safe elimination of potentially endotoxic compounds such as bilirubin and steroids as well as thousands of xenobiotic drugs and carcinogens (1). Overwhelming evidence now indicates that in rat liver the glucuronidation of the numerous varied compounds is catalysed by a family of isoenzymes, which exhibit different aglycone substrate specificity (2). Recent protein purification work has established the existence of several isoenzymes exhibiting different subunit molecular weights (50-56 kDa) when examined by SDS polyacrylamide gel electrophoresis (3-6). Indirect evidence suggests that several more isoenzymes have not yet been isolated and characterised (7).

Our major interest is to obtain a complete molecular characterisation of the UDPGTs in rat and human liver which catalyse the glucuronidation of endogenous compounds such as testosterone, bilirubin and androsterone and study the biogenesis, development and genetic deficiency of these enzymes.

Recently, we have prepared a polyclonal antibody which recognises several UDPGTs (6,8,9). This radioiodinated, affinity-purified, anti-UDPGT

Nucleic Acids Research

antibody has been used to isolate cDNAs coding for UDPGTs from a rat liver cDNA library cloned in the expression vector, bacteriophage λ gt11 (10). The identity of the cDNAs was confirmed by the hybrid-select translation and further immunochemical analyses (10). However, the relationship between an individual isoenzyme and its cDNA clone was not established, as protein sequences have not yet been obtained (7).

Here, we report that a full-coding length cDNA has been isolated, completely sequenced and identified with androsterone UDPGT. Further comparison of the nucleic acid sequences of 15 cDNAs has established the existence of a gene family of at least four different members which are 85-95% homologous.

MATERIALS AND METHODS

Restriction endonucleases, T4 DNA ligase, DNA polymerase and DNA polymerase Klenow fragment, were purchased from Anglian Biotechnology Ltd., Essex, (U.K.). DNA sequencing reagents and M13mp18 and 19 were obtained from BRL/Gibco Biocult Ltd., Paisley, (Scotland). $[\alpha - S]$ dATP (400 Ci/mol) and $[\alpha - P]$ dCTP (3,000 Ci/mol) were from Amersham International plc., Amersham, Bucks. [9,11-H(N)]-androsterone was from New England Nuclear, Du Pont, (U.K.) Ltd., Southampton, and androsterone from Sigma Chemical Co., Poole, Dorset, (U.K.). β -napthoflavone was obtained from Aldrich Chemical Co. Ltd., Dorset. Clofibrate was from Fluorochem Ltd., Glossop, Derbyshire and Phenobarbital was from British Drug Houses Ltd, Poole, (U.K.). Zeta Probe was obtained from Bio-Rad, Watford, Herts (U.K.).BRIJ 58 was from Atlas, Essen, (F.R.G.).

Pretreatment of Animals

Gunn and Wistar rats were from the colonies maintained within the Institute. Male Wistar rats (6-8 weeks old) were given clofibrate intraperitoneally twice daily (200 mg/kg in corn oil) for 4 days or a single intraperitoneal injection of β -naththoflavone (100 mg/kg) in olive oil for 3 days. Phenobarbital (2g/l) was given for 5 days in drinking water. Enzyme assays

Liver microsomes were prepared (11) and assayed by the methods previously described for the different substrates: androsterone (5), bilirubin (12), testosterone (13), 2-aminophenol and 4-nitrophenol (14) in the absence and presence of 0.25 mg BRIJ 58/mg microsomal protein. Protein concentrations were determined by the method of Lowry <u>et al</u> (15).

Isolation of UDPGT cDNAs

A cDNA library prepared in the bacteriophage λ gt11 using mRNA from untreated rat livers was screened using an affinity purified radioiodinated anti UDPGT IgG as previously described (10).

Analysis of fusion proteins

Temperature sensitive lysogens were prepared for each clone by infection of <u>E. coli</u> Y1089. The IPTG inducible fusion proteins were then analysed by immunoblotting as previously described (10). Large scale preparation of bacteriophage DNA

Recombinant λ rlug phage DNA was isolated from lysogens prepared in <u>E.coli</u> W3110 by the procedure described by Maniatis <u>et al</u> (18). <u>Subcloning of cDNA into pBR325</u>

cDNA inserts in the positive recombinants (λ rlug 1-34) were sized by electrophoresis of an <u>Eco</u>RI restriction digest of the purified phage, on a 0.8% agarose gel containing 0.5 µg/ml ethidium bromide, and visualisation by U.V. light. cDNA inserts were purified by excision of the required <u>Eco</u>R1 restriction fragment from the agarose gel followed by electroelution in the presence of <u>E.coli</u> tRNA (18). The cDNA was then subcloned into the <u>Eco</u>R1 site of pBR325 (19) for large scale preparation of cDNA and restriction mapping.

Sequencing of cDNA

DNA sequencing was carried out using the M13 cloning protocols of Messing <u>et al</u> (20) and the dideoxy techniques of Sanger <u>et al</u> (21). cDNA inserts (purified by electroelution as above) were subcloned into the <u>Eco</u>R1 site of M13mp18.

Further subcloning into M13mp18 or 19 to allow directional sequencing of specific restriction fragments of cDNAs was either done as described above or by restriction of the recombinant M13 replicative form using endonucleases which cuts in both the polylinker and the cDNA insert followed by religation of the truncated recombinant M13 replicative form.

Dideoxy sequencing reactions were performed as detailed in the Amersham International M13 Sequencing kit instructions with an $[\alpha - S]$ dATP label and analysis of reaction products on 6% acrylamide/urea gels with and without a buffer gradient.

Computer analysis of DNA sequence

Single stranded sequence from both strands was aligned with the aid of the Staden program (22).

Northern Blot Analysis

RNA was prepared from 2 g of liver from each of the above sources by the method of Chirgwin (23). The RNA was then electrophoresed on a formaldehyde denaturing agarose gel (2,4,18) and electroblotted onto Zeta Probe as described by the manufacturer's instructions. After transfer blots were baked for 2 hrs at 80° C under vacuum and then prehybridised in 5 x SSPE, 50% (v/v) formamide 5 x Denhardt's solution, 0.5% (w/v) SDS, 10% w/v dextran sulphate, 100 ug/ml denatured salmon sperm DNA at 42°C for 12 hrs (18). Hybridisation was carried out using the same conditions but including the required preboiled ³² P labelled probe. Hybridisation was continued for 16 hrs and the blots were then washed in 2 x SSPE 0.5% SDS at room temperature, followed by further washes of increased stringency, with a final wash of 0.1 x SSPE 0.1% SDS at 65°C. Autoradiography was for 2-5 days at -80°C with intensifying screens.

Dot blots

Plasmid DNA including the required cDNA was applied to the Zeta Probe filter under vacuum in the presence of 5 x SSPE. The filters were baked at $^{\circ}$ 80 C and probed with $^{\circ}$ P labelled cDNA with the hybridisation and washing condition described above.

Preparation of radioactively labelled cDNA

Restriction fragments of the cloned cDNAs were purified by agarose gel electrophoresis followed by electroelution (17). Nick translation of the DNA was performed as described by Rigby <u>et al</u> (25). Probes used were for: class 1 cDNA (Probe1), the <u>SacI/EcoR1</u> fragment (bases 1520-1760); class 2 cDNA (Probe2), the <u>SacI/AccI</u> fragment (bases 1520-1885); class 3 cDNA (Probe4), the <u>SacI/EcoR1</u> fragment (bases 1520-1728) for class 4 cDNA (Probe23) the <u>SacI/EcoR1</u> fragment (bases 1520-1885) and for actin mRNA a cDNA containing most of the coding region of mouse γ actin processed 4 gene.

RESULTS AND DISCUSSION

Isolation of UDPGT cDNA clones

Anti-UDPGT antibodies were isolated by affinity chromatography using testosterone-UDPGT Sepharose and bilirubin-UDPGT Sepharose (9). Plaque screening of 500,000 independent rat liver cDNA recombinants with the radioiodinated affinity purified antibodies identified 34 putative UDPGT cDNA clones. These rat liver UDPGT cDNA clones (λ rlug 1-34) were plaque-purified (17) and lysogens of the recombinants were prepared in <u>E.coli</u> Y1089 to identify the production of fusion proteins of translated polypeptides from



 \underline{Fig} 1. Characterisation of β -galactosidase / UDPGT fusion proteins. Bacterial lysates were prepared from lysogens of $\underline{E.coli}$: Y1089(λ gt11) lane 1, Y1089(λ rlug1) lane 2, Y1089(λ rlug2) lane 3, Y1089(λ rlug4) lane 4 and Y1089(λ rlug23) lane 5. Samples (100 μ g of protein) were subjected to electrophoresis on a 6% SDS-polyacrylamide gel. (A) Coomasie blue stained gel, (B) Autoradiogram of an identical gel western blotted onto nitrocellulose and probed with 125 I labelled affinity purified anti UDPGT antibodies. Molecular weights (kDa.) of protein standards lane S, are indicated on the left margin.

<u>lacZ</u>-rlug cDNA genes. Induction of all the fusion proteins was dependent on the addition of the gratuitous inducer IPTG and could be observed by SDS-polyacrylamide gel electrophoresis of cell lysates (Fig. 1A). Immunoblotting of identical gels with affinity purified anti-UDPGT antibodies indicates the presence of UDPGT antigenic determinants only in the large fusion proteins and their breakdown products (Fig. 1B).

The fusion proteins coded by λ rlug 1 and 2 are approximately the same size consisting of approximately 34 kDa of cDNA-encoded protein additional to that coded by the <u>lacZ</u> DNA (114 kDa). λ rlug 23 produced a fusion protein of approximately 170 kDa (Fig. 1b) which is large enough to code for a complete UDPGT of up to 56 kDa (7). λ rlug 4 did not produce a stable fusion protein (see below). All of the putative UDPGT cDNAs could be used to hybrid-select translatable UDPGT mRNA as previously described(10).



<u>Fig 2</u>. Restriction maps of the longest cDNA in each class (indicated on the left) of UDPGT cDNA: Class 1 -rlug 1, Class 2 -rlug 2, Class 3 -rlug 4, Class 4 -rlug 23. The sequencing strategy for rlug 23 is shown below its restriction map, where the length and direction of the horizontal arrows indicate the extent to which each strand was sequenced from a restriction site. The strategy for sequencing clones rlug 1,2 and 4 was similar, using the restriction sites shown.

Partial Restriction Maps of UDPGT cDNAs

DNA isolated from the recombinant bacteriophage was incubated with <u>Eco</u>RI to release the cDNA inserts which were determined to be 0.3-2.1 kb in size. A double digest of λ rlug recombinants with <u>KpnI</u> and <u>Bam</u>HI established the orientation of the cDNA inserts with respect to the 5'-3' transcript from lacZ (10).

Partial restriction maps of the 15 cDNA inserts obtained with a variety of restriction endonucleases allowed the segregation of the cDNAs into three classes 1, 2 and 4, the additional class 3, of which rlug 4 is a member was only established by sequence analysis (see below). The partial restriction maps of the longest clone from each class is illustrated (Fig. 2). These restriction maps indicate considerable similarities between the classes of cDNAs, they do not however correlate with maps of UDPGT cDNA from another study (26), however it is likely that they have isolated the cDNA for other UDPGT isoenzymes.

cDNA and amino acid sequence of rlug 23

The strategy for sequencing rlug 23 cDNA, the longest member of class 4

5	,									3′
	Lac Z	Eco	RI linke	tr	UDPC	GT cD	NA 23			
	GCG	GAA	TTC	CGG	TTG	<u>GTG</u>	TGG	CCG	ATG	
B Gal	- Ala	Glu	Phe	Arg	Leu	Val	Trp	Pro	Met -	UDPG1

<u>Fig</u> 3. The nucleotide sequence of the junction between the <u>lac</u>Z gene and the UDPGT cDNA in λ rlug 23. The figure shows the corresponding translation and illustrates how the reading frame of the cDNA can be determined.

is illustrated by the arrows in Fig. 2. cDNAs from the other classes were sequenced by a similar strategy using the restriction sites as indicated in Fig. 2, to generate the appropriate clones. Sequencing in both directions from restriction sites allowed full determination of the sequence of both DNA strands.

The nucleotide sequence of rlug 23 cDNA (Fig. 4) contained a continuous open reading frame of 1,518 bp; analysis of the <u>lacZ</u>-cDNA junction (Fig. 3) showed this to be in reading frame with the <u>lacZ</u> gene. An ATG at base positions 14-16 is flanked by sequence in reasonable agreement with the translation start site consensus sequence (27). Translation beginning at this site continued for 500 amino acid residues (Fig. 4) and is terminated by a TAG codon.

The amino acid composition (Table 1) of the predicted protein is in good agreement with that previously determined for purified UDP-glucuronyltransferase (2). The calculated molecular weight of the protein is 57,456 Da., which is slightly larger than that determined for the purified UDP-glucuronyltransferases by SDS polyacrylamide gel electrophoresis (Mr 52-56 kDa). However highly hydrophobic proteins and glycoproteins have been previously reported to exhibit abnormal behaviour during electrophoresis (28) and have been shown to be larger proteins that predicted by SDS polyacrylamide gel electrophoresis (29,30).

It is possible that certain of the UDPGT isoenzymes may be glycosylated and a single potential glycosylation site is shown in the sequence at residue 286 (31).

Comparison of the nucleotide sequence of the UDPGT cDNAs

Examination of the cDNA sequences of 15 rlug clones allowed their categorisation into four classes: Class 1 (rlug 1,3,16,24,31 & 32); Class 2 (rlug 2,8,30 & 33); Class 3 (rlug 4,34); Class 4 (rlug 17,22,23). These classes of cDNA code for four different mRNAs.

A comparison of their sequences (Fig.5) show that they exhibit extensive

10 20 GTTGGTGTGGGCCG ATG GAC TTT AGT CAC TGG ATG AAT ATA AAA ATA ATC CTT GAT GAA CTT GTA CAG AGG GGC CAT GAG GTC ACT GTT Met Asp Phe Ser His Trp Met Asn Ile Lys Ile Ile Leu Asp Glu Leu Val Gln Arg Gly His Glu Val Thr Val 50 30 40 CTG AAA CCT TCG GCT TAC TTT TTT CTT GAT CCG AAA AAA TCG TCT GAC CTT AAG TTT GAA ATT TTT TCT ACA TCT ATC AGT AAA GAT GAG Leu Lys Pro Ser Ala Tyr Phe Phe Leu Asp Pro Lys Lys Ser Ser Asp Leu Lys Phe Glu Ile Phe Ser Thr Ser Ile Ser Lys Asp Glu 80 60 70 CTG CAA AAT CAT TTC ATA AAA CTT TTG GAT GTG TGG ACT TAT GAG TTG CCA AGA GAT ACA TGT TTG TCA TAT TCT CCT ATC CTT CAA AAT Leu Gln Asn His Phe Tle Lys Leu Leu Asp Val Trp Thr Tyr Glu Leu Pro Arg Asp Thr Cys Leu Ser Tyr Ser Pro Ile Leu Gln Asn 90 100 110 CTA GTT TAT GAA TIT TCT TAT TAT CTA AGT ATT TGT AAA GAC GCT GTT TCA AAG CAG CAG CTC ATG ACA AAA CTA CAG GAA TCC AAG Leu Val Tyr Glu Phe Ser Tyr Phe Tyr Leu Ser Ile Cys Lys Asp Ala Val Ser Asn Lys Gln Leu Met Thr Lys Leu Gln Glu Ser Lys 130 140 120 TTI GAT GTT CTT TTC GCA GAT CCT GTG GCT TCC TGT GGG GAG CTG ATA GCT GAA CTG CTC CAC ATT CCT TTT CTG TAC AGT CTT AGC TTC Phe Asp Val Leu Phe Ala Asp Pro Val Ala Ser Cys Gly Glu Leu Ile Ala Glu Leu Leu His Ile Pro Phe Leu Tyr Ser Leu Ser Phe 160 170 150 TET CCA GGC CAC AAA CTT GAA AAG TCC ATT GGA AAA TTT ATA CTC CCT CCA TCT TAT GTG CCT GTA ATT TTG TCG GGA CTG GCT GGC AAA Ser Pro Gly His Lys Leu Glu Lys Ser Ile Gly Lys Phe Ile Leu Pro Pro Ser Tyr Val Pro Val Ile Leu Ser Gly Leu Ala Gly Lys 190 200 180 ATG ACA TTC ATA GAC AGG GTA AAA AAT ATG ATA TGT ATG CTT TAT TTC GAC TTT TGG TTC GAG AGA CTT AGA CAC AAG GAA TGG GAC ACG Met Thr Phe Ile Asp Arg Val Lys Asn Met Ile Cys Met Leu Tyr Phe Asp Phe Trp Phe Glu Arg Leu Arg His Lys Glu Trp Asp Thr 230 210 220 TTT TAC AGT GAG ATT TTG GGA AGG CCC ACC ACC GTA GAT GAG ACA ATG AGC AAA GTA GAA ATA TGG CTT ATT AGA TCC TAT TGG GAT TTG Phe Tyr Ser Glu Ile Leu Gly Arg Pro Thr Thr Val Asp Glu Thr Met Ser Lys Val Glu Ile Trp Leu Ile Arg Ser Tyr Trp Asp Leu 250 260 240 AAA TTT CCC CAC CCA ACA TTA CCA AAT GTT GAC TAT ATT GGA GGA CTC CAT TGC AAA CCT TCT AAA CCC TTG CCT AAG GAT ATG GAA GAA Lys Phe Pro His Pro Thr Leu Pro Asn Val Asp Tyr Ile Gly Gly Leu His Cys Lys Pro Ser Lys Pro Leu Pro Lys Asp Met Glu Glu 290 270 280 TTT GTC CAG AGC TCT GGA GAG CAC GGT GTG GTG GTG TTT TCT CTG GGG TCA ATG GTC AGC AAC ATG ACA GAA GAA AAG GCC AAC GCA ATT Phe Val Gin Ser Ser Gly Glu His Gly Val Val Val Phe Ser Leu Gly Ser Met Val Ser Asn Met Thr Glu Glu Lys Ala Asn Ala Ile 300 310 320 GCA TGG GCC CTT GCC CAG ATT CCA CAA AAG GTT CTT TGG AAA TTT GAT GGC AAA ACC CCA GCA ACA TTA GGA CCC ATT ACC AGA GTC TAC Ala Trp Ala Leu Ala Gln Ile Pro Gln Lys Val Leu Trp Lys Phe Asp Gly Lys Thr Pro Ala Thr Leu Gly Pro Asn Thr Arg Val Tyr 330 340 350 AND THE CALL OF CALL AND THE CALL OF CALL AND ACC AND ACC AND ACC AND ACC ANT GAL ACT CALL AND ACC ANT GAL ACC ANT Lys Trp Leu Pro Gln Asn Asp Ile Leu Gly His Pro Lys Thr Lys Ala Phe Val Thr His Gly Gly Ala Asn Gly Leu Tyr Glu Ala Ile 380 360 370 TAT CAT GGA ATC CCT ATG ATT GGC ATT CCT CTG TTT GGA GAT CAA CCT GAT AAT ATT GCC CAC ATG GTG GCC AAA GGA GCA GCT GTT TCA Tyr His Gly Ile Pro Met Ile Gly Ile Pro Leu Phe Gly Asp Gln Pro Asp Asn Ile Ala His Met Val Ala Lys Gly Ala Ala Val Ser 400 390 410 TTG AAT ATC AGG ACA ATG TCA AAG TTA GAT TTT CTC AGT GCA CTG GAG GAA GTC ATA GAC AAT CCG TTC TAT AAA AAA AAT GTT ATG TTG Leu Asn Ile Arg Thr Met Ser Lys Leu Asp Phe Leu Ser Ala Leu Glu Glu Val Ile Asp Asn Pro Phe Tyr Lys Lys Asn Val Met Leu 440 420 430 TTG TCA ACC ATT CAC CAT GAC CAG CCT ATG AAG CCC CTG GAC AGA GCT GTC TTC TGG ATT GAG TTT ATC ATG CGC CAC AAA GGG GCC AAG Leu Ser Thr Ile His His Asp Gln Pro Met Lys Pro Leu Asp Arg Ala Val Phe Trp Ile Glu Phe Ile Met Arg His Lys Gly Ala Lys 450 460 470 CAC CTG AGA CCA CTT GGA CAT AAC CTT CCC TGG TAC CAG TAC CAC TCT CTG GAT GTG ATT GGA TTC CTG CTC ACC TGT TTT GCA GTC ATT His Leu Arg Pro Leu Gly His Asn Leu Pro Trp Tyr Gln Tyr His Ser Leu Asp Val Ile Gly Phe Leu Leu Thr Cys Phe Ala Val Ile 500 480 490 GCA GCT CTT ACT GTA AAA TGT CTC TTG TTC ATG TAC CGA TTC TTT GTA AAG AAG GAA AAG AAA ATG AAG AAT GAG TAG AGCTCATTGACAATG Ala Ala Leu Thr Val Lys Cys Leu Leu Phe Met Tyr Arg Phe Phe Val Lys Lys Glu Lys Lys Met Lys Asn Glu End CACTAACTGAAATGAAATTTCAGCGTCATTCTAATTATGAACCACCTTCTAAAAATTACTAATTTTTTATCAAGGTAGATAACCTTTGTAGGAAGACATATAACTCCGTGAATACTGA TATGTACTCAAAAAATCCATCATTTTTAAAATTTTTAAAACCACTTAATGTAAAAGTTACATTGTAGAAAAATGTGCAGAATAAAATTATTCTTGATAGAGTCCAAATAATCAAGTATTAAC

CTTAAAATATTTGAATAGTGCCATTAGCTTCTTTGTCTAACTGTAGCTTTCATACAATAAAATGTAGATAACTTGTA

 \underline{Fig} 4. The nucleotide and predicted amino acid sequence of clone rlug 23 UDPGT cDNA. The potential glycosylation site is indicated by the underlining around residue 287.

homology up to 95% within the coding regions (Fig. 2). Base changes do not seem to occur randomly, but distinct clustering of substitutions in particular areas is found (Fig. 5). The relatively highly conserved regions are also observed in the cDNA sequence of a human UDPGT (L. McCarthy et al

Residue	<u>Number of residues</u>
Alanine	25
Arginine	13
Aspartamine	17
Aspartic acid	26
Cysteine	7
Glutamine	12
Glutamic acid	26
Glycine	25
Histidine	19
Isoleucine	34
Leucine	58
Lysine	42
Methionine	18
Phenylalanine	33
Proline	30
Serine	34
Threenine	21
Tryptophan	11
Typosipe	10
Valipa	10
VATTIR	31 Tabal 500
	IOTAL SUU

Table 1 Amino acid composition of androsterone UDP-glucuronyltransferase

Molecular weight of unmodified chain = 57,456 Daltons

unpublished). All of the cDNA sequences except rlug 4 (which did not produce a strong antigenic response or stable fusion protein) were found to be in reading frame with <u>lac7</u> (Fig. 3) and with each other allowing reading frame to be continuously checked using clones which contain progressively less 5' coding sequence. Comparison of the translations of each cDNA class (data not shown) illustrates high homology with many of the substitutions being conservative. All translations end at the same TAG codon. This stop codon is relatively rarely used in rat genes but has been reported for rat cytochrome P_{.3} 450 cDNA (32). Spaces have to be introduced into the sequences after the stop codon to maximise homology, implicating the sequence is no longer coding. The predicted size of the translated UDPGT polypeptide from each clone correlates to the size of the UDPGT portion of corresponding fusion protein as seen (Fig. 1) by immunoblot analysis, only when a stop codon in the area indicated (Fig. 4) is used.

Members of a gene family whose cDNA sequences are highly homologous, but differ by many point mutations have been previously described in the cytochrome P450 gene family (33). The UDP-glucuronyltransferase family of cDNAs are likely to be even more similar (as we have found) explaining the difficulty in obtaining monospecific antibodies (7).

Nucleic Acids Research

				~												
rlug23	GTTGGT	GTGGC	CGATGG	20 ACT 12	TTAC	TCACT	GGATGA	4 Atata 1	D AAAATAAT	CCTTGATGAAC	90 T tgtac 190	AGAGGGGGC	80 Catgagg 18	TCACTG	TCTGAAAC	
RLUG23	GCTTAC		TCTTGA	100	GAAA	AAATC	GTCTGA	CCTTA	AGTTTGAA	ATTTTTTCTAC	TCTAT	CAGTAAAG	ATGAGET	GCAAAA	CATTICAT	AAAAC
RLUG23	TTTTGC	ATGTG	TGGACT	TAT	GAGT	TGCCA	AGAGA1	ACATG	TTIGICAT	ATTCTCCTATC	C TTCAA	AATCTAGT	TATGAA		ГАТТТТТАТ	CTAAG
RLUG23	TATTO	TAAAG	ACGCTG	III	CAAA	CAAGC	AGCTCA	ITGACA	AAACTACA	GGAATCCAAGT	TGATG	TICTITIC	CAGATC	CTGTGG	TICCTOTO	GGGAG
RLUG2												A			G	A
HLUG4				47	_		•	6				G		-	G	
PL UC23	CTCATA	CCTCA	ACTOCT	42	U CATT	COTT	TOTOTA	44 CACTC	TTACCTTC	TOTOCACCOCA		TRAAACTO	48	0	ATACTOCO	500
RUG1	*	00104		CCA	6		101017	CHOIC	r	T T	, 111101 C A	IONANAOTO	100 N	CC C	r t	r
RLUG2		С	G		0				C TOC	· · .		,	, ,	000		r
RLUG4		Ċ	G						c	т	A		6 0	a a	ТА	U
				52	D			54	10		60		58	٥Ŭ		600
RLUG23	CTTATG	TGCCT	STAATT	TTG	TCGG	GACTG	GCTGGC	AAAAT	GACATTCA	TAGACAGGGTA		ATGATATG	ATRCTT	TATTIC		TTCGA
RLUG1	С	A	AG	т	A	TA		С		G (с . С			T		с.
RLUG2	-				A	T	G	c C		Ţ		c		Ť		r
RLUG4		A A			A	A	G	cc				•	С	Ť		č
				62	D			64	0	e	60		- 68	0		700
RLUG23	GAGACT	TAGACI	ACAAGG	AAT	GGGA	CACGT	TTTACA	GTGAG	ATTTTGGG	AAGGCCCACCAC	CGTAG	ATGAGACAA	TGAGCA	AAGTAGA	AATATGGC	TTATT
RLUG1	ACT	G	G A			TC A		CA	C A		AT	с т	G	С	G	С
RLUG2	T TGT	ATG	C A	T		TC A	T	С			т	C	G	С	G	С
RLUG4	T TGT	ATG	C A			TC A			T		T	С	G	CT	G	C
				72	3			74	0	7	60		78	0		800
RLUG23	AGATCC	TATTG	GGATTT	GAA	ATTT	CCCCA	CCCAAC	ATTAC	CAAATGTT	GACTATATTGGA	GGACT	CCATTGCAA	ACCITC	TAAACCO	TTGCCTAA	GGATA
RLUG1		С		G	G		T	CG		С		С	G		С	
rlug2		С		G	G					С		Α	G		С	
RLUG4		С		G								A G	C		С	
B 11003				820) 			84	0		60		88	0		900
RLUG23	TGGAAG		JULAG	AGC	ICIG	GAGAG	LACGUI	61661	GGIGITTI	CTCTGGGGTCA	TGGTC	AGCANCATO	ACAGAA	GAAAAGO	CCAACGCA	ATTGC
RLUGI		L L					1					6			AIC	
RLUG2	A	L T					+ -					G			- '	نا
NL004												1.				
				920	h			۵/	n		60	Ũ	08	n	•	1000
RLUG23	ATGGGC	CCTTG	CCAGA	920 TTC) Caca	AAAGG	пстт	94 666444	O TTTGATGG		60 AACAT	TAGGATTCA	98 Atarca		CAAGTGGC	1000
RLUG23 RLUG1	ATGGGC	CCTTGC	CCAGA	920 TTC) Caca	AAAGG	TICTIT	94 Ggaaa G	io TTTGATGGI C		60 AACAT	TAGGACECA	981 Atacca	D Gagteta	Caagtggc	1000 TCCCG T C
rlug23 Rlug1 Rlug2	ATGGGC	CCTTGO	CCAGA	920 TTC) Caca	aaagg"	TICTIT	94 Ggaaa G	O TTTGATGGI C	CAAAACCCCAGO AA CA T	60 AACAT CC T	TAGGACCCA	981 Atacca	D Gagtcta C	CAAGTGGC	1000 TCCCG T C T C
RLUG23 RLUG1 RLUG2 RLUG4	ATGGGC	CCTTGC	CCAGA	920 TTC) Caca	aaagg	тстт	94 Ggaaa G	O TTTGATGGI C	S Caaaaccccago Aa ca T	160 AACAT C C T C	TAGGACECA	981 Atacca	D GAGTCTA C	CAAGTGGC	1000 TCCCG T C T C T C T C
RLUG23 RLUG1 RLUG2 RLUG4	ATGGGC	CCTTGC	CCAGA	920 TTC 102) CACA 20	AAAGG	TICTIT	94 Ggaaa G	io TTTGATGGI C 140	g Caaaaccccago Aa ca T 1	160 AACAT C C T C 060	TAGGACCCA	981 Atacca 106	D GAGTCTA C 90	Caagtggc	1000 TCCCG T C T C T C 1100
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23	atgggc Cagaat	CCTTGC	CCCAGA	92(TTC 102 TCA	CACA CACA 20 CCCA	AAAGG AAAACI	TTCTTT	94 Ggaaa G 10 Ctttg	10 TTTGATGGI C K40 TAACTCATI	S CAAAACCCCAGC AA CA T 1 SGTGGAGCCAA1	IGO AACAT C C T C DGO GGCCTI	TAGGACCCA	981 Atacca 104 Aatcta	D GAGTCTA C 90 TCATGGA	CAAGTGGC	1000 TCCCG T C T C T C 1100 GATTG
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1	ATGGGC Cagaat A	CCTTGC GACATC TC 1	CCAGA CCTGGG T	92(TTC 102 TCA	CACA CACA 20 CCCA T	aaagg Aaaaci	TTCTTT	94 Ggaaa G 10 Ctttg	D TTTGATGG C K4O TAACTCATI G	S CAAAACCCCAGC AA CA T 1 Ggtggagccaat	IGO AACAT C C T C DGO GGCCTI A	TAGGACCCA	981 Atacca 106 Aatcta g c	D GAGTCTA C 90 TCATGGA	CAAGTGGC	1000 TCCCG T C T C T C 1100 GATTG
RLUG23 RLUG1 RLUG2 RLUG2 RLUG23 RLUG1 RLUG2	ATGGGC Cagaat A	CCTTGC GACATO TC 1 C	CCCAGA CCTGGG T T T	920 TTC 102 TCA	CACA CACA 20 CCCA T T	aaagg Aaaaci	TTCTTT	94 Ggaaa G 10 Ctttg	io Tttgatggi C 40 Taactcati G	S CAAAACCCCAGC AA CA T 1 Ggtggagccaat	IGO AACAT C C T C OGO GGCCTI A TG	TAGGACECA CTATGAGGC	980 Atacca 100 Aatcta G C C	D GAGTCTA C 90 TCATGGA	CAAGTGGC ATCCCTATI	1000 TCCCG T C T C T C 1100 GATTG G
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG2 RLUG4	atgggc Cagaat A	CCTTGC GACATC TC 1 C C	CCCAGA CCTGGG T T T	920 TTC 102 TCA	CACA CACA 20 CCCA T T T	aaagg Aaaaci	TTCTTT	94 Ggaaa G 10 Ctttg	io Tttgatggi C 40 Taactcati G	CAAAACCCCAGO AA CA T 3gtggagccaat	IGO AACAT C C T C C DGO GGCCTI A TG TG	TAGGACECA CTATGAGGC	980 ATACCA 100 AATCTA G C C C	D GAGTCTA C 90 TCATGGA	CAAGTGGC	1000 TCCCG T C T C T C 1100 GATTG G
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG2 RLUG4	ATGGGC CAGAAT A	GACATO TC 1 C	CCCAGA CCTGGG T T T T	920 TTC 102 TCA 112	20 CACA 20 CCCA T T T 20	AAAGG AAAACI	TTCTTT	94 Ggaaa G 10 Ctttg 11	O TTTGATGGI C K4O TAACTCATI G 40	S CAAAACCCCAGC AA CA T 1 SGTGGAGCCAA1	60 AACAT C C T C 060 GGCCTI A TG TG 160	TAGGACECA CTATGAGGC	98 ATACCA 106 AATCTA G C C C T18	D GAGTCTA C 90 TCATGGA 80	CAAGTGGC	1000 TCCCG T C T C T C 1100 GATTG G G 1200
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG2 RLUG23	ATGCGC CAGAAT A GCATTC	GACATO TC 1 C C CTCTG1	CCCAGA CCTGGG T T T T T	920 TTC 102 TCA 112 GAT	CACA CACA 20 CCCA T T T 20 CAACI	AAAGG AAAACI CTGAT/	TTCTTT	94 GGAAA G 10 CTTTG 11 GCCCA	O TTTGATGG C 40 TAACTCATI G 40 CATGGTGGG	SCAAAACCCCAGC AA CA T 3GTGGAGCCAAT 1 CCAAAGGAGCAG	IGO AACAT C C T C OGO GGCCTI A TG TG 160 CTGTT	TAGGACECA	981 Atacca 104 Aatcta G C C C 111 Atcacg	D Gagtcta C 90 TCATGGA 90 Acaatgt	CAAGTGGC	1000 TCCCG T C T C T C 1100 GATTG G 1200 GATTT
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2	ATGGGC CAGAAT A GCATTC	GACATO TC 1 C C C C C C C C C C C C C	CCCAGA CCTGGG T T T T T	920 TTC 102 TCA 112 GAT G	CACA CACA 20 CCCA T T T 20 CAACI	AAAGG AAAACI CTGAT/ A	TTCTTT CAAAGC AATATT	94 GGAAA G 10 CTTTG 11 GCCCA	O TTTGATGG C 40 CATGGTGGG	CAAAACCCCAGC AA CA T 1 GGTGGAGCCAA1 1 CCAAAGGAGCAG	GO AACAT C C T C OGO GGCCTI A TG TG 160 CTGTT GG	TAGGACECA ETATGAGGC ICATTGAAT GA G A	981 Atacca 108 Aatcta g c c c t atcagg t	D GAGTCTA C 90 TCATGGA 90 Acaatgt	CAAGTGGC ATCCCTATI CAAAGTTA(C	1000 TCCCG T C T C T C T C 1100 GATTG G G 1200 GATTT
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 PULG4	ATGGGC CAGAAT A GCATTC	GACATO GACATO TC 1 C C CTCTG1 A A	CCCAGA CCTGGG T T T T T	920 TTC 102 TCA 112 GAT G	20 CCCA T T CCCA T CCCA	AAAGG AAAAACI CTGAT/ A A T	TTCTTT CAAAGC AATATT C C	94 GGAAA G 10 CTTTG 11 GCCCA	40 C 440 TAACTCATI G 40 CATGGTGGG	CAAAACCCCAGO AA CA T 3 GGTGGAGCCAAT 1 CCAAAGGAGCAG	60 AACAT C C T C 060 GGCCTI A TG TG 160 CTGTT GGI	TAGGACCCA CTATGAGGC ICATTGAGGC JA G A A C	981 Atacca 108 Aatcta g C c c t atcagg t	D GAGTCTA C 90 TCATGGA 90 Acaatgt	CAAGTGGC ATCCCTATI CAAAGTTAU C	1000 TECCG T C T C T C 1100 GATTG G G 1200 GATTT
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTC	GACATO TC 1 C C CTCTG1 CT A A A	CCTGGG T T T T T T T T T T	920 TTC 102 TCA 112 GAT G A 122	20 CCCA T T CCCA T CCCA	AAAGG AAAAACI CTGAT/ A A T A	TTCTTT CAAAGC AATATT C C C	94 GGAAA G 10 CTTTG 11 GCCCA	D TTTGATGG C K40 TAACTCATI G 40 CATGGTGGG	CAAAACCCCAGC AA CA T 1 CGTGGAGCCAAT 1 CCAAAGGAGCAG	60 AACAT C C T C 060 GGCCTI A TG TG 160 CTGTT GGI	TAGGACECA CTATGAGGC ICATTGAAT GA G A A C A C	980 ATACCA 100 AATCTA G C C C ATCAGG T	D GAGTCTA C 30 TCATGGA 80 ACAATGT	CAAGTGGC ATCCCTATI CAAAGTTAI C C AC	1000 TCCCG T C T C T C 1100 GATTG G G G 1200 GATTT
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23	ATGGGC CAGAAT A GCATTO	GACATO TC 1 C C CTCTG1 CT A A A TGCACT	CCAGA CTGGG T T T T T T GGAGG	920 TTC 102 TCA 112 GAT G A 122 AAG		AAAAGG AAAAACI CTGAT/ A A T A AGACA/	TTCTTT CAAAGC AATATT C C C	94 GGAAA G 10 CTTTG 11 GCCCA 12 ICTAT	0 TTTGATGGI C 40 CATGGTGGI 40 CATGGTGGI	CAAAAACCCCAGC AA CA T SGTGGAGCCAAT 1 SCAAAGGAGCAG 1 1 IGLTATGTIGTI	60 AACAT C C T C 060 GGCCTI A TG TG 160 CTGTT GG0 A 260 GTCAAA	TAGGACECA CTATGAGGC ICATTGAAT GA G A A C A C EATTGACT	981 ATACCA 106 AATCTA G C C C ATCAGG T 126 ATCAGC	D GAGTCTA C 30 TCATGGA 80 ACAATGT 80	CAAGTGGC ATCCCTATI CAAAGTTAI C AC GAAGCCCC	1000 TCCCG T C T C T C 1100 GATTG G G 1200 SATTT 1300
RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG3 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG2 RLUG23 RLUG3	ATGEGE CAGAAT A GCATTO TCTCAG G A	GACATO TC 1 C C CTCTG1 CT A A A TGCAC1	CCCAGA CCTGGG TT T T T T T T T T T T T T T T T	920 TTC 102 TCA 112 GAT G A 122 AAG	CACA CACA 20 CCCCA T T T T CCAACI 20 CAACI	AAAGG CTGAT/ A A T A AGACA/	TTCTTT CAAAGC AATATT C C C ATCCGT	94 GGAAA G 10 CTTTG 11 GCCCA 12 TCTAT T	10 TTTGATGGG C 40 CATGCTCATI G 40 AAAAAAAAAA G	CAAAAACCCCAGC AA CA T 3GTGGAGCCAA1 3GTGGAGCCAA1 1 3CCAAAGGAGCAG 1 1GTTATGTTGTT C G	60 AACAT C C T C 060 GGCCTI A TG TG 160 CTGTT GGG 260 GTCAAU	TAGGACCCA CTATGAGGC CTATGAGGC GA G A A C A C A C CCATTCACC	98 ATACCA 100 AATCTA G C C C ATCAGG T 120 ATCAGC	D GAGTCTA C BO TCATGGA BO ACAATGT BO AGCCTAT	ATCCCTATI CAAAGTTAT C C AC GAAGCCCC	1000 TCCCG T C T C T C 1100 GATTG G G 1200 GATTT 1300 TGGAC
RLUG23 RLUG1 RLUG2 RLUG2 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG2	ATGEGE CAGAAT A GCATTO TCTCAG G A GT A	GACATO TC 1 C C CTCTG1 CT A A A TGCAC1	CCCAGA CCTGGG TT T T T T T T T GGAGG/	920 TTC 102 TCA 112 GAT G A 122 AAAG	20 CCCCA T T T CCCCA T CCAACI 1 1 1 20 CCAACI 1 20 CCAACI 1 20 CCCCA 7 T CCCCA 7 T 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	AAAAGG CTGAT/ A A T A AGACA/ A	TTCTTT CAAAGC AATATT C C C ATCCGT T T	94 GGAAA G 10 CTTTG 11 GCCCA 12 TCTAT T	ID TTTGATGGI C 40 TAACTCATI G 40 CATGGTGGG 40 AAAAAAAAA G	CAAAAACCCCAGC AA CA T 3GTGGAGCCAAT 1 CCAAAGGAGCCAAT 1 CCAAAGGAGCAG 1 CCAAAGGAGCAG C C C C C C C C C C C C C	60 AACAT C C C 060 GGCCTI A TG TG TG 160 CTGTT 260 GTCAAI	TAGGACCCA CTATGAGGC ICATTGAAT GA G A C 2CATTCACC	98 ATACCA 100 AATCTA G C C T ATCAGG T 120 ATCAGC	D GAGTCTA C BO TCATGGA BO ACAATGT BO AGCCTAT C A	CAAGTGGC ATCCCTATI CAAAGTTAG C C AC GAAGCCCC A	1000 TCCCG T C T C T C T C 1100 GATTG G G G 1200 GATTT 1300 TCGAC
RLUG23 RLUG1 RLUG2 RLUG2 RLUG4 RLUG3 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTC TCTCAG G A GT A GT A	Gacato TC 1 C C CTCTG1 CT A A A TGCAC1	CCCAGA CTGGGG T T T TTGGAG GGAGG TA AA	920 TTC 102 TCA 112 GAT G A 122 AAAG A	CACA CACAA CCCCAA T T T CCAACI A A A	AAAGG CTGAT/ A A T A A GAGACA/ A	TTCTTT CAAAGC AATATT C C C ATCCGT T T A	94 GGGAAA G 10 CTTTIG 11 GCCCA 12 TCTAT T	ID TTTGATGGI C 40 CATGCTCATI G CATGCTGGI 40 AAAAAAAAA G	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 CCAAAGGAGCAG 1 IGTTATGTTGTT C G C C G C C G G	60 AACAT C C T C 060 GGCCTI A TG TG 160 CTGTT GGG 260 GTCAAG	TAGGACCCA CTATGAGGC ICATTGAAT GA G A A C A C C CCATTCACC	98 ATACCA 100 AATCTA G C C C 111 ATCAGG T 120 ATGACC	D GAGTCTA C 30 TCATGGA BO ACAATGT AGCCTAT C A A	CAAGTGGC ATCCCTATI CAAAGTTAU C GAAGCCCC A A	1000 TCCCG T C T C T C 1100 GATTG G G 1200 GATTT 1300 TGGAC
RLUG23 RLUG1 RLUG2 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTC TCTCAG G A GT A GT A	Gacatic TC 1 C C CTCTG1 CT A A A TGCAC1	CCCAGA CCTGGG T T T T T T T GGAGGA TA AA	920 TTC 102 TCA 112 GAT G A A A A A A A A 132	20 CCCCA T T T CCCCA T CCAACI A A A A 20 CCAACI A A A 20 CCAACI A A A 20 CCCAACI A	AAAGG CTGAT/ A A T A A GAGACA/ A A	AAAAAGC C C C C C C C C C C C C C C C C C C C	94 GGAAA G 10 CTTTG 11 GCCCA 12 TCTAT T T 13	10 TTGATGGI C 140 TAACTCATI G 40 AAAAAAAAAA G 40	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 CCAAAGGAGCAG 1 CCAAAGGAGCAG C C C C C C C C C C C C C C C C	60 AACAT C C T C 060 GGCCTI A TG TG 160 CTGTT GGC J 260 GTCAAC	TAGGACCCA CTATGAGGC CTATGAGGC ICATTGAAT GA G A A C A C C CCATTCACC	98 ATACCA 106 AATCTA G C C C 111 ATCAGG T 126 ATGACC	D GAGTCTA C D TCATGGA BO ACAATGT AGCCTAT A A BO A A BO A BO	CAAGTGCC ATCCCTATI CAAAGTTAI C C AC GAAGCCCC A A	1000 TCCCG T C T C 1100 SATTG G G 1200 SATTT 1300 TCGAC
RLUG23 RLUG1 RLUG2 RLUG2 RLUG2 RLUG4 RLUG2 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23	ATGGGC CAGAAT A GCATTC TCTCAG G A GT A GT A AGAGCT	GACATIC TC 1 C C CTCTG1 CT A A A TGCAC1	CCCAGA TT T T TTGGAG TA AA CTGGAT1	920 TTC 102 TCA 112 GAT G A A A A A C A A C A A C A A C A C A C	CACA CACA 20 CCCA T T T CCAACI CCAACI CCAACI A A A A CO CCAACI A A A CO CCAACI CCAACI CCAACI CCAACI CCACA	AAAAGG AAAAACI A A T A A AGACAJ A A A A A A A TCAT(AATATTT C C C C C C C T T A SCCCCCA	94 GGAAA G 10 CTTTG 11 GCCCA 12 TCTAT T 13 CAAAG	10 TT TGATGGI C 40 40 CATGG TGGI 40 AAAAAAAAA G 40 GGGCCCAAGC	CAAAAACCCCAGC AA CA T 1 SGTGGAGCCAAT CCAAAGGAGCAG CCAAAGGAGCAG C C G C C C G C C C G C 1 CACCTGAGACCA	60 AACAT C C T C GGC GGCCTI A TG TG TG 160 CTGTT 260 GTCAAU	TAGGACCCA CTATGAGGC TCATTGAAGGC TCATTGAAGGC TCATTGAAG A C CCATTCACC	98 ATACCA 100 AATCTA G C C C 110 ATCACG T 120 ATCACG T 120 ATCACC 138 TCCCTG	D GAGTCTA C D TCATGGA BO ACAATGT A AGCCTAT C A GTACCAG GTACCAG	CAAGTGCC ATCCCTATI CAAAGTTAU C GAAGCCCC A A TACCACTC1	1000 TCCCG T C T C T C 1100 GATTG G G 1200 GATTT 1300 TGGAC
RLUG23 RLUG1 RLUG2 RLUG4 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG3 RLUG1 RLUG2 RLUG3 RLUG1 RLUG3 RLUG1 RLUG3 RLUG4 RLUG2 RLUG4 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG4 RLUG2 RLUG4 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLU	ATGGGC CAGAAT A GCATTCI GCATTCI GCA GCA GCA A GTA A GAGGCT	CCTTGC GACATIC TC 1 C C CTCTGI A A A TGCACI	CCCAGA CTGGG TT TTTGGA(TA AA CTGGAT)	920 TTC 102 TCA 112 GAT G A A 122 AAAG A 132 TGA	CACA CACA 20 CCCA T T T CCAACI CCAACI CCAACI A A A A CO CCAACI C CCAACI C CCAACI C CCAACI C CCCA C C CCCA C C CCCA C C CCCA C C CCCA C	AAAAGG AAAAACI A A T A A AGACAJ A A A A TCAT(3	AAAAGC CAAAGC C C C C C C C ATCCGT T T A SCGCCA	94 GGAAA G 10 CTTTG 11 GCCCA 12 ICTAT T 13 CAAAG	D TTTGATGG C 40 TAACTCATI G 40 AAAAAAAAA G GGCCCAACC TT	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 CCAAAGGAGCAG 1 CCAAAGGAGCAG C G G C G G 1 ACCTGAGGACCA T	60 AACAT C C T C GGCCTI A GGCCTI A TG TG CTGTT GGI CTGTT GGI GC A CTGCA CTTGG/	TAGGACCCA CTATGAGGC TCATTGAAG SA G A A C A C CCATTCACC	98 ATACCA 100 AATCTA G C C C 110 ATCACG T 120 ATCACG T 120 ATCACC	D GAGTCTA C HO TCATGGA HO ACAATGT A AGCCTAT C A A GTACCAG	CAAGTGC ATCCCTATI CAAAGTTAI C C AC GAAGCCCC A A TACCACTC1	1000 TCCCG T C T C T C 1100 GATTG G G 1200 GATTT 1300 rCGAC
RLUG23 RLUG1 RLUG2 RLUG2 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG4 RLUG23 RLUG4 RLUG	ATGGGC CAGAAT A GCATTCI G A GT A AGT A	CCTTGC GACATIC TC 1 C C CTCTG1 A A A TGCAC1 GTCTTC	CCCAGA CCTGGG TT T T TTTGGA(G TA AA	92(TTC 102 TCA 112 GA G G A 122 AAAG A 132 IGA) CACA T T T T CAACI A A A CAACI A A CAACI A A A	AAAAGG CTGAT/ A A T A AGACA/ A A A A TCAT(3 3 3	AATATTT C C C C C C C ATCCGT T T A C C C C C C C C C C C C C C C C	94 GGAAA G 10 CTTTG 11 GCCCA 12 TCTAT T 13 CAAAG A	ID TTTGATGGI C 40 CATGCTCATI G CATGCTGGI 40 GGCGCCAAGQ TT A	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 CCAAAGGAGCAG 1 CCAAAGGAGCAG C C G C C G G 1 CACCTGAGACCA T	60 AACAT C C T C 060 GGCCTI A TG TG 160 CTGTT GG0 GTCAAI 360 CTTGG/	TAGGACCCA CTATGAGGC TCATTGAAGGC GA G A A C A C C CCATTGACC	98 ATACCA 100 AATCTA G C C C 110 ATCACG T 120 ATCACG T 120 ATCACC	D GAGTCTA C 30 TCATGGA 80 ACAATGT AGCCTAT C A A 30 GTACCAG	CAAAGTGGC ATCCCTATI CAAAGTTAT C C AC GAAGCCCC A A A TACCACTC1	1000 TCCCG T C T C T C T C 1100 G G 1200 SATTG 1300 TCGAC 1400 CTCG
RLUG23 RLUG1 RLUG2 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTC GCATTC GCATTC GCA GTA GTA AGAGCT AG	GACATIC TC 1 C C CTCTG1 A A TGCAC1 GTCTTC	CCCAGA T T T T TTCGA(TA AA	92(TTC 102 TCA 112 GA G A 122 AAAG A 132 IGA) CACA CCCAA T T T CCAACI CCAACI CCAACI A A CO CCAACI A A CO CCCAA C C CAACI A A C C CCCAA C C C CCCAA C T T T C C CCCAA C C C C	AAAAGG AAAAACI A A T A AGAACAI A ATCAT(3 3 3 3 3	TTETTT CAAAGE C C C C C C C ATCCGT T T A SCCGCCA	94 GGAAA G 100 CTTTG 111 GCCCA 12 TCTAT T 13 CAAAG A	D TTTGATGGI C K40 TAACTCATH G CATGCTCATH CATGCTGGI 40 AAAAAAAAAAA G GGGCCCAAGC TT A	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 GGTGGAGCGAGCAAT 1 CCAAAGGAGCAG C C G C C G G C C G G C 1 CACCTGAGACCA T	60 AACAT C C T C GGCCTI A GGCCTI A TG TG CTGTT 260 GCCAAI 360 CTTGG/	TAGGACCCA CTATGAGGC CTATGAGGC ICATTGAAT GA G A A C A C A C C CCATTGACC	984 ATACCA 106 AATCTA G C C C ATCAGG T 126 ATCAGG T 126 ATGACC 136 TCCCTG	D GAGTCTA C 30 TCATGGA 80 AGCCTAT A AGCCTAT C A GTACCAG	CAAGTGGC ATCECTATI CAAAGTTAC C C AC GAAGECECC A A TACCACTC1	1000 TCCCG T C T C T C T C 1100 G G G 1200 GATTT 1300 TCGGAC
RLUG23 RLUG1 RLUG2 RLUG2 RLUG2 RLUG4 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG3 RLUG3 RLUG4 RLUG3 RLUG4 RLUG	ATGGGC CAGAAT A GCATTCI G A GT A AGAGCTI AG AG	GACATIC TC 1 C C CTCTGG C TGCACT G TGCACT	CCCAGA T T T T T T T T T T T TA AA	92(TTC 100 TCA 110 GAT G A 122 GAAG A 132 IGA) CACA 20 CCCA T T T CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCCAAC CCCAAC CCCAAC CCCAAC T T CCCAAC CCCAAC T T CCCCAAC T CCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC T CCCCAAC CCCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC T CCCAAC C CCCAAC C CCCAAC C CCCAAC C CCCAAC C CCCAAC C C CCCAC C C CCCAAC C C CCCAAC C C CCCAAC C C C C C C C C C C C C C C C C C C C	AAAAGG AAAAACI A A A A A A A A A A A A A A A A	TTETTT CAAAGC CAAAGC C C C C C C C C C C C C C	92 GGAAA G 10 CTTTG 11 GCCCA 12 ICTAT T 13 GCCAAAG A A 444	D TTTGATGGI C 40 AAACTCATI G 40 AAAAAAAAAA G 40 GGGCCAAGC TT A	CAAAAACCCCAGC AA CA T 1 SGTGGAGCCAAT 1 SGTGGAGGAGCAA 1 CCAAAGGAGCAA C G G C G G C C G G C C G G C 1 CACCTGAGACCA T	60 AACAT C C T C C 060 GGCCTI A TG TG 160 CTGTIT GGC GCCAA CTGTI 360 CTTGG/	TAGGACCCA CTATGAGGC TCATTGAAGGC TCATTGAAGGC A C A C A C A C A C A C A C A C A C A	984 ATACCA 106 AATCTA G C C C 111 ATCACG T 126 ATGACC 136 TCCCTG	D GAGTCTA C 30 TCATGGA 80 ACCATGT A A GTACCAG 80	ICAAGTGGC ATCCCTATI CAAAGTTAI C GAAGCCCC A A TACCACTC1	1000 TCCCG T C T C 1100 GATTG G G 1200 GATTT 1300 TCGGAC 1400 TCTGG
RLUG23 RLUG1 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTCA G A G A A GT A A GT A A AGAGCTI A G A A G A A G C A C G A C	GACATIC TC 1 C C CTCTG1 A A TGCAC1 GTCTTC A	CCCAGA CCCGGG T T T TTTGGAG TGGAGG TGGAGG	92(TTC 102 TCA 112 GAT G G A 122 AAG A 132 IGA	CACA CCCA T T T CCAACI CCAACI CCAACI A A CO CCAACI C C C C C C C C C C C C C C C C	AAAAGG CTGAT/ A A T A AGACA/ A A A ATCAT(3 3 3 3 3 3 3 3 3 3 3 3 3	AAAAAGC CAAAAGC C C C C C C AACCCGT T T A SCCGCCA	94 GGAAA G 10 CTTTG 11 GCCCA 12 ICTAT T 13 CAAAG A 14 A 14	10 TT TGATGGI C 40 40 CATGC TGGI 40 CATGC TGGI 40 GGGCCCAAQC TT A 40 AGC TCT TTAC	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 GGTGGAGGCGAAC 1 CCAAAGGAGCAG C G G C G G C G G C G G C G G C G G 1 ACCTGAGAACCAG T	60 AACAT C C T C C 060 GGCCTT GGC C CGCCT GGC C CTGG C C CTGG C C CTGG C C C C C C C C C C C C C C C C C C C	TAGGACCCA CTATGAGGC CTATGAGGC TCATTGAGGC A C CCATTCACC CCATTCACC	984 ATACCA 100 AATCTA G C C C ATCACG T 120 ATCACG T 120 ATCACC 130 TCCCTG	D GAGTCTA C BO TCATGGA BO ACCATGT A A C A C A C A C A C C C C C C C C C	CAAGTGGC ATCCCTATI CAAAGTTAI C C C GAAGCCCC A A TACCACTC1	1000 TCCCG T C T C T C T C 1100 SATTG G 1200 SATTT 1300 TCGAC 1400 TCTCG
RLUG23 RLUG1 RLUG2 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG3 RLUG3 RLUG3 RLUG3 RLUG3 RLUG4 RLUG2 RLUG4 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTCI G A GT A GT A AGACCTI AG ATGTGA	GACATIC TC 1 C C CTCTG1 C A A TGCAC1 GTCTTC A	CCCAGA CTGGG T T T TTGGAG TA AA CTGGATT TCCTGC A	92(TTC 100 TCA 1112 GAT G G A 122 AAG A 132 IGA 142 CTC T) CACA 20 CCCA T T T CAACI CAACI A A CO CAACI C A C C C A C C T C C C C C C C C C C C	AAAAGG CTGAT/ A A T A A A A A A A A A A A A A A A	AATATTT C C C C C ATCCGT T T A SCCGCCA	92 GGAAA G 10 CTTTG 11 12 12 12 12 12 13 CAAAG A 14 ATTGC C T	0 TTTGATGGI C 40 AAACTCATI G 40 AAAAAAAAAAAAAAA C 40 GGGCCCAACC TT A 40 ACCTCTTAC TC C	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 GGTGGAGCGAGCAAT 1 CCAAAGGAGCAG C C G C C G G 1 1 CACCTGAGACCA T 1 CTGTAAAATGTC C G C T	60 AACAT C C T C C OGO GCCCTT A TC C GCCCTT GCC I TC C C TCTTCG/ C C TCCCC C C C C C C C C C C C C C C	TAGGACCCA CTATGAGGC CTATGAGGC TGATGAGG A C A C CCATTCACC ACATAACCT TGCATGTAC TG_T_T	984 ATACCA 106 AATCTA G C C C T ATCAGG T 126 ATGACC 136 TCCCTG 146 CGATTC	D GAGTCTA C D D TCATGGA D D ACCATGT A A C A G TACCAG G TACCAG G TTTGTAA	CAAGTGC ATCCCTATI CAAAGTTAC C C AC GAAGCCCC A A A TACCACTC1	1000 TCCCG T C T C T C T C 1100 G G G 1200 GATTT 1300 TGGAC 1400 CTGG
RLUG23 RLUG1 RLUG2 RLUG2 RLUG2 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG2 RLUG4 RLUG4 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG2 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTCI G A GT A AGAGCTI AGG ATGTGA	GACATIC TC 1 C C CTCTGI CT A A A TGCAC1	CCCAGA CTGGG TT T TGGAGGI TA AA TGGAT1 TCCTGC A	92(TTC 100 TCA 112 GAT G G A 122 AAG A 132 IGA A 132 IGA 142 CTC T A) CACA CCCA T T T CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAAC CCAA CCAAC CCCAAC CCCAA CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAAC CCCAACC	AAAAGG CTGATI A A A A A A A A A A A A A A A A A A	TTETTT CAAAGE CAAAGE C C C C C C C C C C C C C C C C C C C	94 GGAAA G 10 CTTTG 11 GCCCA 12 TCTAT T 13 CAAAG A 14 ATTGC C T	D TTGATGGI C C C C C C C C C C C C C C C C C C	CAAAAACCCCAGC AA CA T T SGTGGAGCCAAT SGTGGAGCGAGCAG CGAAAGGAGCAG CGGG CGGG CGG	60 AACAT C C T C C 060 GGCCTT G GGC T GGC GGC G GGC G G G G G	TAGGACCCA CTATGAGGC CTATGAGGC TCATTGAGG A C A C A C A C A C A C A C A C A C A C	984 ATACCA 106 AATCTA G C C C 111 ATCAGG T ATGACC 136 TCCCTG 146 CGATTC	D GAGTCTA C 30 TCATGGA 80 AGCCTAT A A GTACCAG GTACCAG C C	CAAGTGGC ATCCCTATI CAAAGTTAI C GAAGCCCC A A TACCACTC1 AGAAGGAA/ C	1000 TCCCG T C T C T C T C 1100 G G 1200 GATTT 1300 TCGGAC 1400 TCTGG 1500 VAGAA
RLUG23 RLUG1 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4 RLUG4	ATGGGC CAGAAT A GCATTC G G A GT A GT A G	GACATIC TC 1 C C CTCTGG TGCAC1 A A TTGCAC1	CCCAGA CTGGG T T T T T GGAGGA TA AA TCCTGC A	92(TTC 102 TCA 112 GAT GA GA 122 AAAG A 132 IGA 142 CTC T A 152) CACA CCCA T T T CCCA CCCA T CCAAC CCAAC C CCAAC C CCAAC C CCAAC C CCAAC C CCAAC C C CCAAC C C CCCA C C CCCA C C CCCA T T C CCCA C C CCCA T T T C C CCCA T T T C C CCCA T T T C C CCCA T T T C C CCCA T T T C CCCAA C C CCCA T T T C CCCAA C C CCCAA T T T C CCCAA C C CCCAA C C CCCAA C C CCCAA C C CCCAA C C CCCAA C C CCCAA C	AAAAGG CTGAT/ A A A A A A A A A A A A A A A A A A	TTETTT CAAAGC CAAAGC C C C C ATCCGT T T A SCGCCA	92 GGAAA G 10 CTTTG 11 GCCCA 12 ICTAT T 13 CAAAG A 14 ATTGC C T 15	0 TTGATGGI C 40 40 CATGCTCATI 40 CATGCTGGI 40 GGGCCCAAGX TT A 40 AGCTCCTTAG C C G 40 C C C G 40	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 GGTGGAGCCAAT 1 CCAAAAGGAGCAG C G G C G G C G G C G G 1 CACCTGAGACCAG T 1 CTGTAAAAATGTC C G C T 1 CTGTAAAAATGTC C G C T 1 CTGTAAAAATGTC C G C T 1 CTGTAAAAATGTC C G C T 1 C	60 AACAT C C T C C 060 GOCCTT GOC T T GOC T C C C C C C C C C C C C C C C C C C	TAGGACCCA CTATGAGGC CTATGAGGC TGATGAGGC A C CCATGCACC A C CCATGCACC TG T T T T	984 ATACCA 100 AATCTA G C C T ATCACG T ATCACG T 120 ATCACC 130 TCCCTG	D GAGTCTA C BO TCATGGA BO ACAATGT A A A G GTACCAG G TTTGTAA C G	CAAGTGC ATCCCTATI CAAAGTTAI C GAAGCCCC A A TACCACTC1 AGAAGGAA/ C	1000 TCCCG T C T C T C 1100 SATTG G 1200 SATTT 1300 TCGAC 1400 TCTGG
RLUG23 RLUG1 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG23 RLUG1 RLUG2 RLUG4 RLUG23 RLUG1 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4 RLUG23 RLUG4	ATGGGC CAGAAT A GCATTCA G A G A G A G A G A G A G A G A G A G	GACATCI TC 1 C C CTCTGT TC A A A TGCACT GTCTTC A C C C C C C C C C C C C C C C C	CCCAGA CTGGGC TT T TTGGAG CGGAGGA TA AA CTGGAT1 TCCTCC A	92(TTC TTC TCA TCA TCA TCA TCA TCA TCA TCA) CACA CCCA T T T CCACA T T CAACI CAACI C C C C C C C C C C C C C C	AAAAGG CTGAT/I A A T A AGACAI G G G TTTTI(G G G G G G G G G G G G G G G G G G G	TICITI CAAAGC CAAAGC C C C C C C C C C C C C C	92 GGAAA G 10 CTTTG 11 GCCCA 12 TCTAT T 13 CAAAG A 14 ATTGC C T 15 SACTGA	0 TT TGA TGG C 40 40 CATGC TGG 40 GGCCCCAAGC TT A 40 GGCCCCAAGC TT A 40 CC CC CC CC CC CC CC 40 CATGC TGG CC 40 CATGC TGG 40 CATGC TGG 40 CC 40 CC 40 CC 40 CC 40 CC 40 CC 40 CC CC CC 40 CC CC 40 CC CC 40 CC CC 40 CC CC 40 CC CC 40 CC CC 40 CC CC 40 CC 40 CC CC 40 CC CC CC CC CC CC CC CC CC C	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 GGTGGAGCCAAT 1 CCAAAAGGAGCAG C G G C G G C G G C G G C G G C G G C G C	60 AACAT C C T C C 060 GGCCTT GGC 160 CTGTT GGC A CTGTT CGC C CTGTT CGC C CTGGT T A S60 CTTGGT T A S550 CTATT	TAGGACCCA CTATGAGGC CTATGAGGC CTATGAGGC GA G A A C C CATTCACC CATTCACC TCATGTAC T T T T	984 ATACCA 106 AATCTA G C C C ATCACG T 126 ATGACC 136 C C C C C C C C C C C C C C C C C C C	D GAGTCTA C BO TCATGGA BO ACCATGT A A GTACCAG G TTTGTAA G G TADAAAA-	CAAGTGGC ATCCCTATI CAAAGTTAT C C GAAGCCCC A A A TACCACTC1 AGAAGGAA/ C	1000 TCCCG T C T C T C 1100 SATTG G 1200 SATTT 1300 TCGAC 1400 TCTGG 1500 VAGAA
RLUG23 RLUG1 RLUG2 RLUG2 RLUG4 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTCI G A GT A GT A GT A AGACCTI AG ATGTGA AATGAAA	GACATC TC 1 C C CTCTG1 CT A A A TGCACT GTCTTC A TTGGAT	CCCAGA CTGGG TT T TT CGGAGG/ A CTGGAT T CCCTGC A A CTGGAT	92(TTC TTC TCA 112 GAT G G A 122 AAG A 132 IGA 1422 CTC T A 152 SCT) CACA 20 CCCAA T T T CCAACU CCAACU A CCAACU C A CCATTC A C CCATTC A	AAAAGG CTGAT/I A A T A A AGACAI G G G G G G G G G G G G G G G G G G G	TTETTT CAAAGC CAAAGC C C C C C C C C C C C C C	92 GGAAA G 10 CTTTG 11 11 GCCCA 12 TCTAT T 13 CAAAG A 14 ATTGC C T 15 ATG ATG ATG	0 TTTGATGGI C 40 AAACTCATI G 40 AAAAAAAAAAAAAAA C 40 GGGCCCAACC TT A 40 ACCTCTTAC TC C 40 AACCTCTTAC C 40 AAAAAAAAAAAAAAA 40 ACCTCTTAC C 40 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	CAAAAACCCCAGC AA CA T T CGTGGAGCCAAT CGTGGAGCCAAT CCAAAGGAGCAG C C G C C G G C C G C C G G C C G C C G G 1 CACCTGAGACCA T C	60 AACAT C C T T C 060 GCCCTT A T G GCCCTT GCG 150 CTGTT GCG GTCAAI 360 CTTCG/ T A S60 CTTCG/ T A	TAGGACCCA CTATGAGGC CTATGAGGC TGATGAGG A C A C A C CCATTCACC CCATTCACC TG T T T T T T T T	984 ATACCA 106 AATCTA G C C C T ATCAGG T 126 ATGACC 136 C C C C C C C C C C C C C C C C C C C	D GAGTCTA C D D TCATGGA D A A C A A C A A C G TACAAGA C G C TAAAAA-	CAAAGTGGC ATCCCTATI CAAAGTTAC C C GAAGCCCC A A A TACCACTC1 AGAAGGAAA C TTACTAA	1000 TCCCG T C T C T C T C 1100 SATTG G G 1200 SATTT 1300 TGGAC 1400 TCTGG 1500 VAGAA 1600 VT-TT AAC
RLUG23 RLUG1 RLUG2 RLUG2 RLUG2 RLUG4	ATGGGC CAGAAT A GCATTCI GCATTCI GCATTCI GCATTCI AG GT A AGAGCTI AG ATGAAI CA	GACATIC TC 1 C C CTCTG A A A TGCACT A A TGCACT	CCCAGA CTGGG TT T TTGGAGG CGGAGG CGGAGGA CTGGATT A A CTGCTCC A	92(TTC TTC TTC TTC TTC TTC TTC TTC TTC TT	CAACA CCCCAA T T T CCCCAA T T CCAACC CCAACC CCAACC C CCAACC C C C	AAAAGG CTGAT/ A A T A AGACA/ A A AGACA/ A A G G G G G G G G G G G G G G G G G	TTETTT CAAAGC CAAAGC C C C C T T A SCCCCA C T C SCCACTA TC	94 GGAAA G 10 CTTTG 11 CTTTG 11 CCCCA 12 CCCA 13 CAAAG A 14 4 A T 15 SACTGA A G A G	D TTGATGGI C C C C C C C C C C C C C C C C C C	CAAAAACCCCAGC AA CA T 1 GGTGGAGCCAAT 1 GGTGGAGCGAAT 1 CCAAAGGAGCAG C G G C G G C C G G C C G G C C G G C C G C C C C G C	60 AACAT C C T C C 050 GGCCTT C GGC CTC C C C C C C C C C C C	TAGGACCCA CTATGAGGC CTATGAGGC CTATGAGGC A C A C A C A C A C A C A C A C A C A	984 ATACCA 106 AATCTA G C C C 111 ATCAGG T ATGACC 1380 C C C C C C 1580 ACCTT C	D GAGTCTA C 30 TCATGGA 80 AGCCTAT A A GTACCAG G TTTGTAA C G C TAAAAAA- T C C	CAAGTGGC ATCCCTATI CAAAGTTAU C GAAGCCCC A A TACCACTC1 AGAAGGAAA C TTACTAP C TGAT1	1000 TCCCG T C T C T C 1100 G G 1200 G 1200 G 1200 G 1200 G 1200 G 1400 TCTGG 1500 VAGAA 1600 IT-TT AAC -T

	16	320 16	40 166	50 1680	1700
RLUG23	TTTA-TCAAGG-TAGATAAC	CTTTGT	-AGGAAGACATATAACTCCG	TGAATACTGATATGTACTCA A	AAA-TCC-ATCATTTTT
RLUG1	CAG CACTT	AATTTTGTATCATG	ΤΑΑ ΤΤ	AGCT GA	GAC ATT
RLUG2	GGGG C- CT	A	A GT	TA -	
RLUG4	TG C- CT		A A	A TA	T CA
	17	720 17	40 17	60 1780	1900
RLUG23	AAATTTTAAAAACCACTTAAT	GTAAAAGTTACATTGT	AGAAAAATGT-GCAGAATAA	AATTATTCTTGATAGAGTCC A	AATAATCAAAGTATTAACC
RLUG1	T C TTC	CAA TGCC GT	A T	GATCC*	
RLUG2	т	AA G	TGA G A	G – A –– T T	A G
RLUG4	TG	AT*			
	16	320 18	40 186	50 1880	1900
RLUG23	TTAAAATATTTGAATAGTGT	CEATTAGETTETTTGTETAA	TACTGAATCTGTAGCTTTCA	TACAATAAAATGTAGATAAC T	TGTA*
RLUG2	C T	GAC C G	G	C	TACTATAATATGGAC
	19	320 19	40 19	60 1980	2000
RLUG2	ATATAATAGTTTTTCTGTAA	TAGTCTTAATTATTGTAGT	CGGGGATAAAGTGTGGTTTG	GTTTGGATATTCATTTCAAA G	GGTAGGAATCTGTTGGCTA
	20	320 20	40 20	60 2080	2100
RLUG2	TTTTGTTCCTGTAACAAAAT	GTGCTGACCAAAAGCATCTC	CAGGGAAAAGCAGAGCAGTT	TATTITGAGTTGTGCTTACA G	ATCCTGAGAACGCAGGATA
	21	120 21	40 21	60 2180	2200
RLUG2	GATAGGAAGGCAGGGCAGCA	GTCAGCCAGATGACAAACTC	TCTCATTACATCTTAACCA	CACATAGAAAGGAAAAAGTG TG	ACTATEGTETEAACTTTCA
	22	220 22	40 22	60 2280	2300
RLUG2	AAGCTTGCTCCAGTGATATA	TTTCCTCCAAAAAGATTTAA	CCCCTTTAAATAATATTCCT	GTACCCCTGGAGTTGGGAGT T	TAGCTCAGTGGTAGAGCAT
	23	320 23	40 23	60 2380	2400
RLUG2	TTGCCTACCAAACACAAGGC	TCTGTGTTCAGTCCTCAGCT	CCGGGGGAAAAAAAAAAAAAAAAA	GATTCCATAACCTCAAAC	

<u>Fig 5</u>. A comparison of the nucleotide sequences of the longest cDNA clone in each class. The clones rlug 1,2,4 and 23 representing classes 1,2,3 and 4 respectively. The sequence of rlug 23 is shown in full, whereas only those bases differing from this sequence in the other forms are shown. The beginning and the end of each cDNA sequence is indicated by a *.

Identification of rlug 23 as androsterone UDPGT cDNA

No sequence analysis of UDPGT proteins has been published. Whilst this work is underway in this laboratory, we have adopted a strategy to identify cloned UDPGT cDNA with individual isoenzymes based on the expression of each isoenzyme. Differential induction of specific isoenzymes by various xenobiotics has been shown to be due to increased levels of mRNA (7,34). Similarly a different spectrum of UDPGT isoenzyme and mRNAs is present during development, and also it is likely that specific genetic deficiencies affecting expression of certain UDPGTs are a result of the absence of the corresponding UDPGT mRNAs. The cDNA library was prepared from untreated rat liver mRNA, it is therefore expected to contain mainly cDNAs coding for constitutive isoenzymes and any assignment of a specific cDNA class to a particular isoenzyme should take this in to account.

Considerable cross-hybridisation of the different classes of cDNAs occurred when examined by Southern blotting (data not shown). This is not too surprising because of the extensive homologies of the different cDNA especially within the coding regions. Therefore a common restriction fragment was identified in the cDNAs whose sequences showed least homology. These 3' non-coding sequences (see methods for details) were isolated and nick translated and shown to exhibit selective hybridisation to self by dot



<u>Fig</u> <u>6</u>. Dot blot cross hybridisation analysis of the cDNA clones, rlug 1,2 4 and 23. DNA (100_n g) in 5xSSPE from the clones indicated was bound to Zeta Probe membrane under vacuum. Four identical strips were prepared and the DNA baked onto the filters. Each filter was hybridised (see experimental procedures for details) to a different labelled probe as indicated and further described in the text. After hybridisation filters were washed in 0.1 x SSPE, 0.1% SDS at 65°C for 2 hours followed by autoradiography.

blot analysis under stringent washing conditions, however some weak cross-hybridisation still occurs (Fig. 6).

We have used these probes in northern blot analysis of total RNA prepared from xenobiotic pretreated animals and Gunn rats. In these livers phenobarbital induced testosterone UDPGT activity 2 fold, β -naphthoflavone induced phenol UDPGT activity 2 fold and clofibrate induced bilirubin UDPGT activity 2.5 fold (I.Scragg and B.Burchell, unpublished). Bilirubin UDPGT and possibly phenol UDPGT protein are not expressed in Gunn rat livers (6).

These enzyme inductions are not very dramatic, but the effect of phenobarbital in altering the levels of UDPGT mRNA can be shown by northern blot analysis using probes for class 1, 2 or 3 cDNAs (Probes 1,2 and 4 respectively) (Fig. 7). These probes recognise a 2.5 kb UDPGT mRNA, and also a 3.8 kb RNA species in the case of class 2 which is also seen to be phenobarbital inducible, and is perhaps a partially processed RNA intermediate. The phenobarbital inducibility of the UDPGT mRNA recognised by these probes does not help with the immediate identification of these clones with individual UDPGT isoenzymes as more than one UDPGT isoenzyme is induced by phenobarbital (37).

Nevertheless the RNA cross hybridising with the cDNA probes from classes 1-3 is not inducible by clofibrate or β -naphthoflavone and they are present in the Gunn rat. These results suggest that these probes are likely to be recognising phenobarbital inducible testosterone UDPGT mRNAs or other isoenzymes which are not well characterised (7).



<u>Fig 7</u>. Hybridisation of probe 4,2 or 1 to northern blots of total RNA prepared from the liver of untreated (U), or xenobiotic treated β -naphthoflavone (B), phenobarbital (P) or clofibrate (C) Wistar rats and from untreated Gunn rats (G). RNA (5 µg) from each source was electrophoresed on a formaldehyde 1% agarose gel, blotted onto Zeta Probe and hybridised with the radioactively labelled probes indicated on the figure. Hybridisation between the probes and mRNA was detected by autoradiography. Rat rRNA (18S and 28S) were used as gel markers and their positions are indicated on the left.

Hybridisation of rlug 23 (a class 4 cDNA) to the same RNA samples prepared from pretreated Wistar rats (as described above) produced a very different hybridisation signals to the other three probes following northern blot analysis (Fig. 8). The results show probe 23 strongly recognised an mRNA of approximately 2.7 kb in the RNA from control animals which is absent from the RNA in the other two samples. A smaller RNA (2.5 kb) is weakly identified in RNA from all three sources, however it appears to be slightly phenobarbital inducible, and is probably identified as a result of cross hybridisation to the messages hybridising to probes 1,2 and 4 (such cross hybridisation seems particularly likely as on much longer exposures of the autoradiogram the larger 3.8 kb RNA identified by probe 2 can also be seen). Reprobing of the blots with actin cDNA indicated that these results were not due to differences in loading or integrity of the RNA in the different samples (Fig. 8).

This strikingly different hybridisation signal observered with probe 23 was explained by examination of the UDPGT enzyme activities expressed by the microsomes prepared from the same livers as the RNA. The data from the assays



<u>Fig</u> 8. Hybridisation of probe 2 or 23 to northern blots of rat liver total RNA. The RNA samples analysed were as described in figure 7, but using probes 2 and 23. The specific activities of testosterone (nmol/min/mg protein x10) and androsterone (nmol/min /mg protein x100) UDPGT are indicated on the blots hybridised with probes 2 and 23 respectively. This hybridised DNA was stripped off the blots to allow reprobing with actin cDNA, the reprobed blots identifying actin mRNA are shown on the right. Rat rRNA markers are indicated on the far left of the figure.

of androsterone and testosterone UDPGT activity are indicated on figure 8. These data demonstrate the lack of androsterone UDPGT activity strongly correlates to the loss of hybridisation signal to the 2.7 kb mRNA by probe 23.

A genetic deficiency of androsterone UDPGT has been described in Wistar rats (36). Most laboratory colonies are a mixture of animals expressing high (HA) or low (LA) androsterone UDPGT activities. These activities are some 36-fold greater in HA rats than in LA rats (Table 2). We have also confirmed that UDPGT activities towards bilirubin, testosterone, oestradiol, 4-nitrophenol and 2-aminophenol are not affected by this deficiency as previously reported (36,37). Further, sulphation levels are unchanged between the two strains (37). Recently Green <u>et al</u> (38) have reported that very low levels of androsterone UDPGT are expressed in LA Wistar rat livers.

Thus, we have prepared RNA from HA and LA Wistar rat livers and analysed

Strain	BRIJ 58	Specific Activity (nmol/min/mgPr)	Activation (fold)
НА	-	0.258 <u>+</u> 0.03 (n=8)	_
	+	1.670 <u>+</u> 0.09 (n=8)	6.5
LA	-	0.020 <u>+</u> o.004 (n=7)	_
	+	0.046 <u>+</u> 0.010 (n=7)	2 . 3

<u>Table 2</u> BRIJ 58 was used in enzyme assays at 0.25 mg detergent/mg protein. The figures in parenthesis indicate the number of individual livers assayed. The data presented are the mean values and the S.E.M.

northern blots of this RNA with 32 P labelled probes from the different classes of cDNA. The results show that the probe prepared from rlug 23 (a class 4 cDNA) recognises an mRNA of approximately 2.7 kb in the RNA from HA rats which is not detectable in the RNA from LA rats, this could not be explained by differeces in loading or integrity of the RNA preparations as actin mRNA levels were approximately the same in all samples (Fig. 9). Probe 23 also cross hybridises with a smaller mRNA (as found previously) which is strongly recognised by the probes prepared from class 1-3 cDNA. In this



<u>Fig</u> 9. Hybridisation of probe 23 to northern blots of total RNA from the livers of rats expressing high (H) or Low (L) androsterone UDPT activity. HA and LA RNA was analysed on northern blots by hybridisation to Probe 23 and actin cDNA by the procedures described in (Fig 7 and 8).

experiment cross hybridisation is more noticeable (compared to that shown in Fig. 8), it is probable that the integrity of the probe was not fully maintained during nick translation resulting in an effectively shorter probe and thus allowing greater cross hybridisation. Obviously, oligonucleotide probes need to be prepared for totally specific identification of individual mRNAs. Nevertheless only the cDNA probe from rluq 23 (Class 4) specifically diagnoses the isoenzyme genetic deficiency of androsterone UDPGT. We therefore conclude that rlug 23 cDNA contains a full coding sequence of androsterone UDPGT.

ACKNOWLEDGEMENTS

We are especially grateful to Richard Hynes for supplying a rat liver cDNA library in λ qt11 and to David Leder for provision of the actin cDNA. We thank Lee McCarthy and Ian Hiles for help with DNA sequence analysis; Sheila Smith and Isobel Shepherd for help with enzymological studies. We also thank the Wellcome Trust for a grant supporting this work. BB is a Wellcome Trust Senior Lecturer.

REFERENCES

- Dutton, G.J. (1980) Glucuronidation of Drugs and Other Compounds. CRC 1. Press, Boca Raton, F1 (USA)
- Burchell, B. (1981) Rev.Biochem.Toxicol. 3, 1-32 2.
- Bock, K.W., Josting, D., Lillienblum, W. and Pfeil, H. (1979) 3. Eur.J.Biochem. 98, 19-26 Burchell, B. (1980) FEBS Lett. <u>111</u>, 131-135 Falany, C.N. and Tephly, T.R. (1983) Arch.Biochem.Biophys. <u>227</u>, 248-258
- 4.
- 5.
- Scragg, I., Celier, C. and Burchell, B. (1985) FEBS Lett. 183, 37-42 6.
- Burchell, B., Jackson, M.R., McCarthy, L.R. and Barr, G.C. (1985) in 7. Microsomes and Drug Oxidations, Boobis, A.R., Caldwell, J., De Matteis, F. and Elcombe, C.R., eds., p. 212-220. Taylor and Francis, London and Philadelphia
- Burchell, B., Kennedy, S.M.E., Jackson, M.R. and McCarthy, L.R. (1984) 8. Biochem.Soc.Trans. 12, 50-53
- Burchell, B., Jackson, M.R., McCarthy, L.R. and Barr, G.C. (1985) in 9. Advances in Glucuronide Conjugation, Matern, S., Bock, K.W. and Gerok, W. Eds., p. 117-124. MTP Press, Lancaster, U.K.
- Jackson, M.R., McCarthy, L.R., Corser, R.B., Barr, G.C. and Burchell, B. 10. (1985) Gene <u>34</u>, 147-153 Burchell, B. (1977) Biochem.J. <u>161</u>, 543-549
- 11.
- 12. Heirweigh, K.P.H., Van der Vijver, M. and Ferery, J. (1972) Biochem.J. <u>129</u>, 605-608
- Rao, G.S., Hauester, G., Rao, M.L. and Brener, H. (1976) 13. Anal.Biochem.<u>74</u>, 35-40 Wisnes, A. (1969) Biochim.Biophys.Acta. <u>191</u>, 279-291
- 14.
- Lowry, O.H., Rosebrough, N.J., Farr, A.J. and Randall, R.J. (1951) 15. J.Biol.Chem. 193, 265-275
- 16. Young, R.A. and Davis, R.W. (1983) Proc.Natl.Acad.Sci. USA 80, 1194-1198

- 17. Schwarzbauer, J.E., Tamkun, J.W., Lemische, I.R. and Hynes, R.O. (1983) Cell <u>35</u>, 421-431
- 18. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in Molecular
- Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York 19. Soberon, X., Covarrubias, L. and Bolivar, F. (1980) Gene <u>9</u>, 287-305
- 20. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl.Acids Res. <u>9</u>, 309-321
- 21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci. USA 74, 5463-5467
- 22. Staden, R. (1980) Nucl.Acids Res. 8, 3673-3694
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry <u>18</u>, 5194-5199
- Lehrach, M.D., Diamonds, J.M., Wozney, H. and Boedtker (1977) Biochemistry <u>16</u>, 4743-4751
- Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J.Mol.Biol. <u>113</u>, 237-251
- Mackenzie, P.I., Gonzalez, F.J. and Owens, S.I. (1984) J.Biol.Chem. <u>259</u>, 12153–12160
- 27. Kozak, M. (1983) Microbiol.Rev. <u>47</u>, 1
- 28. Leach, B.S., Collawn, J.F. and Fish, W.W. (1980) Biochem. <u>19</u>, 5734-5741
- Gonzalez, F.J., Nebert, D.W., Hardwick, J.P. and Kasper, C.B. (1985) J.Biol.Chem. <u>260</u>, 7435-7441
- Yabusaki, Y., Shimizu, M., Murakami, H., Nakamura, K., Oeda, K. and Ohkawa, H. (1984) Nucleic Acids Res. <u>12</u>, 2929-2938
- 31. Hanover, J.A. and Lennarz, W.J. (1981) Arch. Biochem. Biophys. 211, 1-19
- 32. Mizukami, Y., Soqawa, K., Suwa, Y., Muramatsu, M. and Fujii-Kuriyama, Y. (1983) Proc.Natl.Acad.Sci. USA <u>80</u>, 3958-3962
- Leighton, J.K., De Brunner-Yossbrinck, B.A. and Kemper, B. (1984) Biochemistry <u>23</u>, 204-210
- Mackenzie, P.I., Gonzalez, F.I. and Owens, S.I. (1984) Arch.Biochem.Biohys. <u>230</u>, 676-680
- Lillienblum, W., Walli, A.K. and Bock, K.W. (1982) Biochem.Pharmacol. <u>31</u>, 907-913
- 36. Matsui, M., Nagai, F. and Aoyagi, S. (1979) Biochem.J. <u>179</u>, 483-487
- 37. Matsui, M. and Watenbe, H.K. (1982) Biochem.J. 204, 441-447
- Green, M.D., Falany, C.N., Kirkpatrick, R.B. and Tephly, T.R. (1985) Biochem.J. <u>230</u>, 403-409