Volume 14 Number 2 1986 Nucleic Acids Research

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 λgt11 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

antibody has been used to isolate cDNAs coding for UDPGTs from a rat liver cDNA library cloned in the expression vector, bacteriophage λgt11 (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). [α-^{35}S] dATP (400 Ci/mol) and [α-^{32}P] dCTP (3,000 Ci/mol) were from Amersham International plc., Amersham, Bucks. [9,11-^{3}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 et al (15).

Isolation of UDPGT cDNAs

A cDNA library prepared in the bacteriophage λgt11 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 E. coli Y1089. The IPTG inducible fusion proteins were then analysed by immunoblotting as previously described (10).

Large scale preparation of bacteriophage DNA

Recombinant λrlug phage DNA was isolated from lysogens prepared in E.coli W3110$^-$ by the procedure described by Maniatis et al (18).

Subcloning of cDNA into pBR325

cDNA inserts in the positive recombinants (λrlug 1-34) were sized by electrophoresis of an EcoRI 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 EcoR1 restriction fragment from the agarose gel followed by electroelution in the presence of E.coli tRNA (18). The cDNA was then subcloned into the EcoR1 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 et al (20) and the dideoxy techniques of Sanger et al (21). cDNA inserts (purified by electroelution as above) were subcloned into the EcoR1 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 [α-^{35}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 80°C and probed with ^{32}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 et al (25). Probes used were for: class 1 cDNA (Probe1), the SacI/EcoR1 fragment (bases 1520-1760); class 2 cDNA (Probe2), the SacI/AccI fragment (bases 1520-1885); class 3 cDNA (Probe4), the SacI/EcoR1 fragment (bases 1520-1728) for class 4 cDNA (Probe23) the SacI/EcoR1 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 (λrlug 1-34) were plaque-purified (17) and lysogens of the recombinants were prepared in E.coli Y1089 to identify the production of fusion proteins of translated polypeptides from

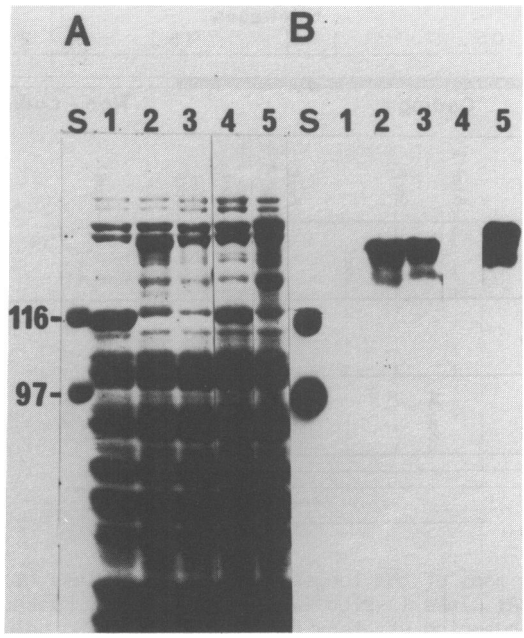


Fig 1. Characterisation of β-galactosidase / UDPGT fusion proteins. Bacterial lysates were prepared from lysogens of E.coli: Y1089(λgt11) lane 1, Y1089(λrlug1) lane 2, Y1089(λrlug2) lane 3, Y1089(λrlug4) lane 4 and Y1089(λrlug23) lane 5. Samples (100 μg of protein) were subjected to electrophoresis on 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.

lacZ-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 λrlug 1 and 2 are approximately the same size consisting of approximately 34 kDa of cDNA-encoded protein additional to that coded by the lacZ DNA (114 kDa). λrlug 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). λrlug 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).

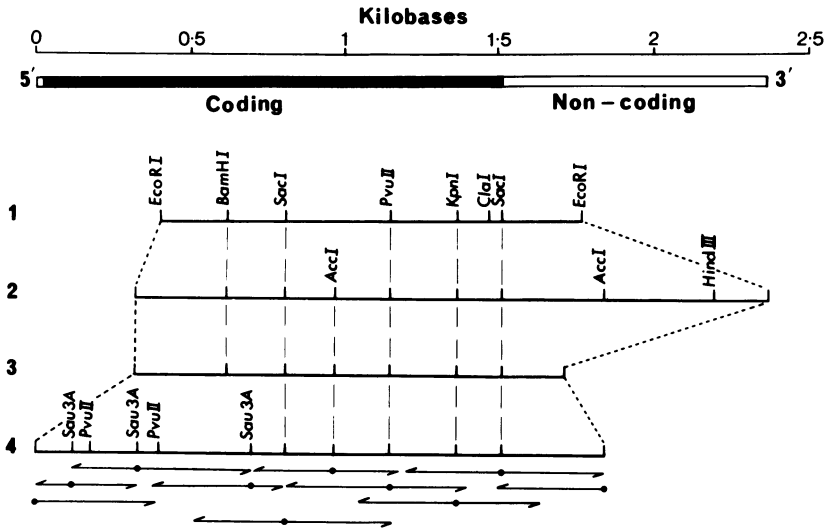


Fig 2. Restriction maps of the longest cDNA in each class (indicated on the left) of UDPGT cDNA: Class 1 -rlug 1, Class 2 -rlug 2, Class 3 -rlug 4, Class 4 -rlug 23. The sequencing strategy for rlug 23 is shown below its restriction map, where the length and direction of the horizontal arrows indicate 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 EcoRI to release the cDNA inserts which were determined to be 0.3-2.1 kb in size. A double digest of λrlug recombinants with KpnI and BamHI 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

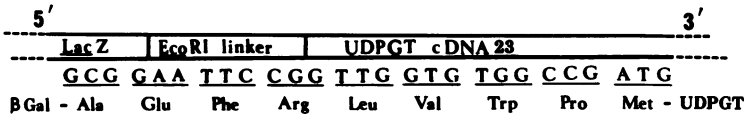


Fig 3. The nucleotide sequence of the junction between the lacZ gene and the UDPGT cDNA in λrlug 23. The figure shows the corresponding translation and illustrates how the reading frame of the cDNA can be determined.

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 lacZ-cDNA junction (Fig. 3) showed this to be in reading frame with the lacZ 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
      GTTGGTGTGGCCG 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
                30                                  40                                  50
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
                60                                  70                                  80
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 Ile 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 TTT TCT TAT TTT TAT CTA AGT ATT TGT AAA GAC GCT GTT TCA AAC AAG 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
                120                                 130                                 140
TTT 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
                150                                 160                                 170
TCT 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
                180                                 190                                 200
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
                210                                 220                                 230
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
                240                                 250                                 260
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
                270                                 280                                 290
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 Gln 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
AAG TGG CTC CCG CAG AAT GAC ATC CTG GGT CAC CCA AAA ACC AAA GCC TTT GTA ACT CAT GGT GGA GCC AAT GGC CTC TAT GAG GCA ATC
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
                360                                 370                                 380
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
                390                                 400                                 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
                420                                 430                                 440
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
                480                                 490                                 500
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

CACTAACTGAAATGAAATTTCAGCGTCATTCTAATTTATGAACCACCTTCTAAAAATTACTAATTTTTTATCAAGGTAGATAACCTTTGTAGGAAGACATATAACTCCGTGAATACTGA

TATGTACTCAAAAATCCATCATTTTTAAATTTTAAAACCACTTAATGTAAAAGTTACATTGTAGAAAAATGTGCAGAATAAAATTATTCTTGATAGAGTCCAAATAATCAAGTATTAAC

CTTAAAATATTTGAATAGTGCCATTAGCTTCTTTGTCTAACTGTAGCTTTCATACAATAAAATGTAGATAACTTGTA
```

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

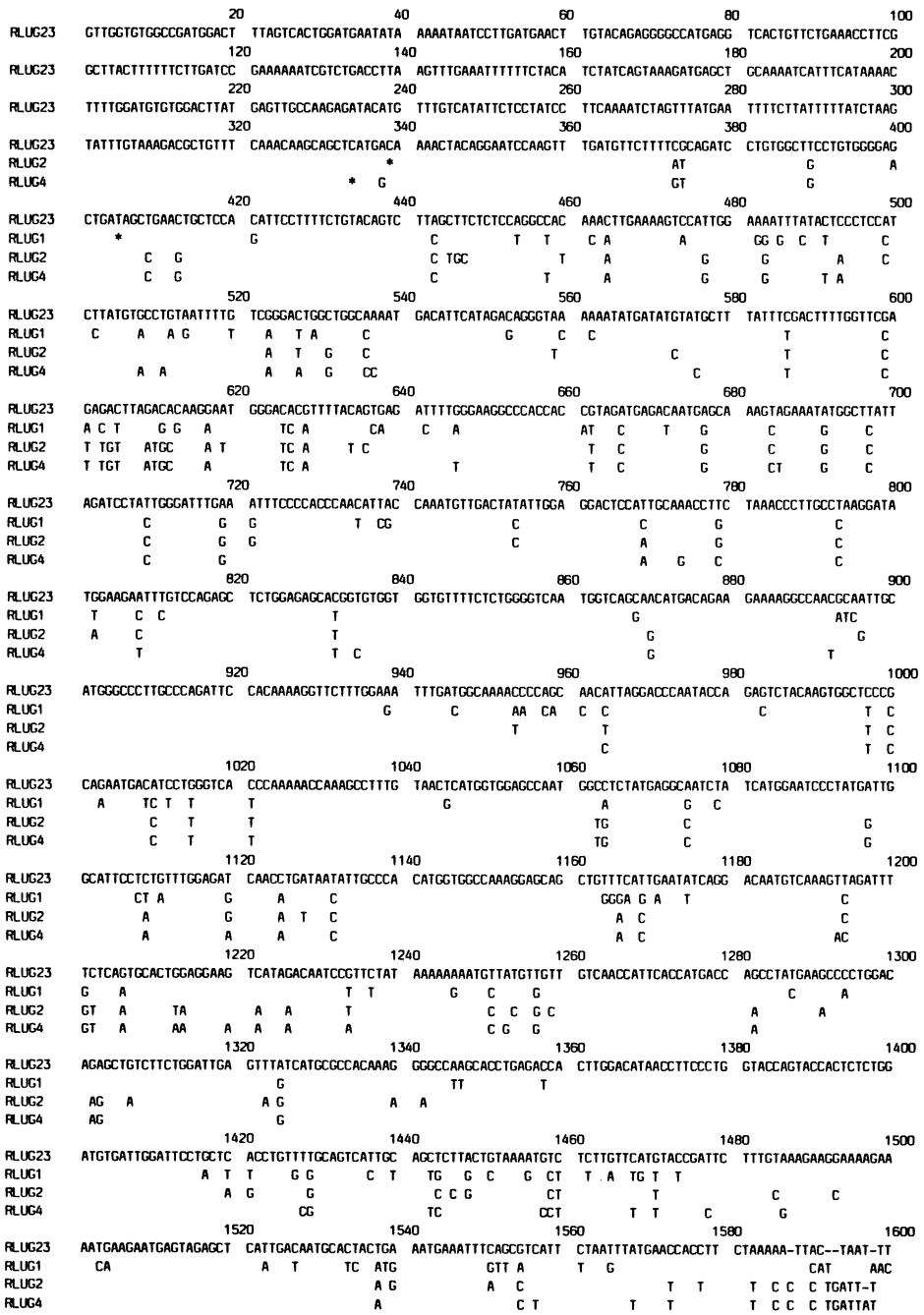
Table 1 Amino acid composition of androsterone UDP-glucuronyltransferase

Residue	Number of residues
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
Threonine	21
Tryptophan	11
Tyrosine	18
Valine	31
Total	500

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 lacZ (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).



```
                    1620                1640                1660                1680                1700
RLUG23  TTTA-TCAAGG-TAGATAAC  CTTTGT---------- -AGGAAGACATATAACTCCG  TGAATACTGATATGTACTCA  AAAA-TCC-ATCATTTTT—
RLUG1     CAG   CA   —CTT       AATTTTGTATCATG  T AA          T  T   A GC T        GA      G A   C       A TT
RLUG2     GG GG  C-   C T  A                     A          G   T                  TA      ---- --- ---------
RLUG4     TG     C-   C T                       A              A   A              TA        T         CA
                    1720                1740                1760                1780                1800
RLUG23  AAATTTTAAAACCACTTAAT  GTAAAA—GTTACATTGT—  AGAAAAATGT-GCAGAATAA  AATTATTCTTGATAGAGTCC  AAATAATCAAAGTATTAACC
RLUG1   T      C    T T C          CAA  TGCC   GT         A  — T       GATCC*
RLUG2                T              AA   G                 TGA G A       G  - A       —  T  T  A    G
RLUG4              T G              AT*
                    1820                1840                1860                1880                1900
RLUG23  TTAAAATATTTGAATAGTGT  CCATTAGCTTCTTTGTCTAA  TACTGAATCTGTAGCTTTCA  TACAATAAAATGTAGATAAC  TTGTA*
RLUG2     C            T         GA C     C   G   G                                —   C          TACTATAATATGGAC
                    1920                1940                1960                1980                2000
RLUG2   ATATAATAGTTTTTCTGTAA  TAGTCTTAATTATTTGTAGT  CGGGGATAAAGTGTGGTTTG  GTTTGGATATTCATTTCAAA  GGGTAGGAATCTGTTGGCTA
                    2020                2040                2060                2080                2100
RLUG2   TTTTGTTCCTGTAACAAAAT  GTGCTGACCAAAAGCATCTC  CAGGGAAAAGCAGAGCAGTT  TATTTTGAGTTGTGCTTACA  GATCCTGAGAACGCAGGATA
                    2120                2140                2160                2180                2200
RLUG2   GATAGGAAGGCAGGGCAGCA  GTCAGCCAGATGACAAACTC  TCTCATTACATCTTAACCA  CACATAGAAAGGAAAAAGTG  TGACTATGGTGTGAACTTTCA
                    2220                2240                2260                2280                2300
RLUG2   AAGCTTGCTCCAGTGATATA  TTTCCTCCAAAAAGATTTAA  CCCCTTTAAATAATATTCCT  GTACCCCTGGAGTTGGGAGT  TTAGCTCAGTGGTAGAGCAT
                    2320                2340                2360                2380                2400
RLUG2   TTGCCTACCAAACACAAGGC  TCTGTGTTCAGTCCTCAGCT  CCGGGGGAAAAAAAAAGAAA  GATTCCATAACCTCAAAC
```

Fig 5. 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

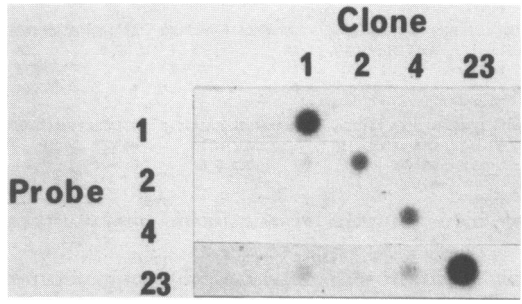


Fig 6. Dot blot cross hybridisation analysis of the cDNA clones, rlug 1,2 4 and 23. DNA (100 n g) in 5xSSPE from the clones indicated was bound to Zeta Probe membrane under vacuum. Four identical strips were prepared and the DNA baked onto the filters. Each filter was hybridised (see experimental procedures 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, ß-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 ß-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).

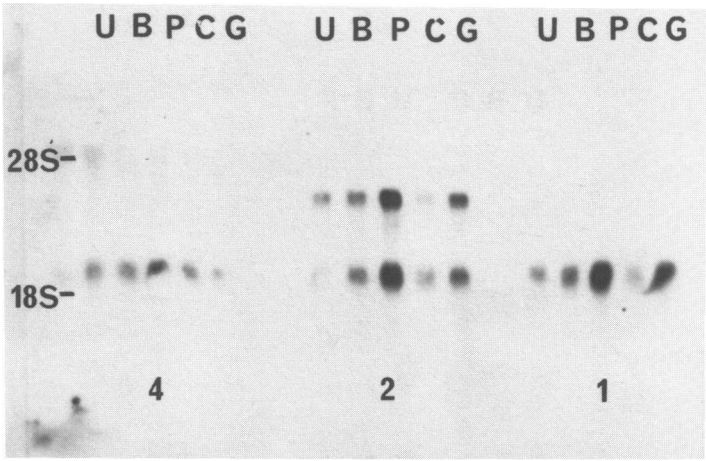


Fig 7. Hybridisation of probe 4,2 or 1 to northern blots of total RNA prepared from the liver of untreated (U), or xenobiotic treated β -naphthoflavone (B), phenobarbital (P) or clofibrate (C) Wistar rats and from untreated Gunn rats (G). RNA (5 µg) from each source was electrophoresed on a formaldehyde 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

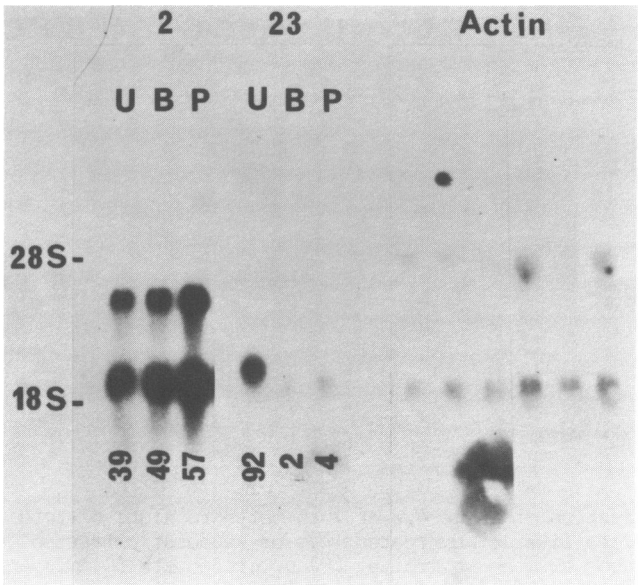


Fig 8. Hybridisation of probe 2 or 23 to northern blots of rat liver total RNA. The RNA samples analysed were as described in figure 7, but using probes 2 and 23. The specific activities of testosterone (nmol/min/mg protein x10) and androsterone (nmol/min /mg protein x100) UDPGT are indicated on the 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 et al (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

Table 2 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.

Strain	BRIJ 58	Specific Activity (nmol/min/mgPr)	Activation (fold)
HA	-	0.258 ± 0.03 (n=8)	-
	+	1.670 ± 0.09 (n=8)	6.5
LA	-	0.020 ± o.004 (n=7)	-
	+	0.046 ± 0.010 (n=7)	2.3

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

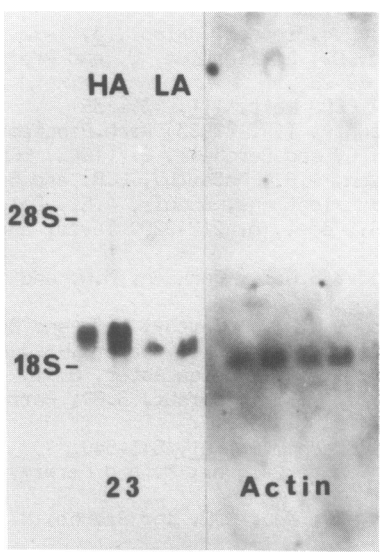


Fig 9. Hybridisation of probe 23 to northern blots of total RNA from the livers of rats expressing high (H) or Low (L) androsterone UDPT activity. HA and LA RNA was analysed on northern blots by hybridisation to Probe 23 and actin cDNA by the procedures described in (Fig 7 and 8).

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 rlug 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 λgt11 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

1. Dutton, G.J. (1980) Glucuronidation of Drugs and Other Compounds. CRC Press, Boca Raton, Fl (USA)
2. Burchell, B. (1981) Rev.Biochem.Toxicol. 3, 1-32
3. Bock, K.W., Josting, D., Lillienblum, W. and Pfeil, H. (1979) Eur.J.Biochem. 98, 19-26
4. Burchell, B. (1980) FEBS Lett. 111, 131-135
5. Falany, C.N. and Tephly, T.R. (1983) Arch.Biochem.Biophys. 227, 248-258
6. Scragg, I., Celier, C. and Burchell, B. (1985) FEBS Lett. 183, 37-42
7. Burchell, B., Jackson, M.R., McCarthy, L.R. and Barr, G.C. (1985) in 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
8. Burchell, B., Kennedy, S.M.E., Jackson, M.R. and McCarthy, L.R. (1984) Biochem.Soc.Trans. 12, 50-53
9. Burchell, B., Jackson, M.R., McCarthy, L.R. and Barr, G.C. (1985) in Advances in Glucuronide Conjugation, Matern, S., Bock, K.W. and Gerok, W. Eds., p. 117-124. MTP Press, Lancaster, U.K.
10. Jackson, M.R., McCarthy, L.R., Corser, R.B., Barr, G.C. and Burchell, B. (1985) Gene 34, 147-153
11. Burchell, B. (1977) Biochem.J. 161, 543-549
12. Heirweigh, K.P.H., Van der Vijver, M. and Ferery, J. (1972) Biochem.J. 129, 605-608
13. Rao, G.S., Hauester, G., Rao, M.L. and Brener, H. (1976) Anal.Biochem.74, 35-40
14. Wisnes, A. (1969) Biochim.Biophys.Acta. 191, 279-291
15. Lowry, O.H., Rosebrough, N.J., Farr, A.J. and Randall, R.J. (1951) 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 35, 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 9, 287-305
20. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl.Acids Res. 9, 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
23. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5194-5199
24. Lehrach, M.D., Diamonds, J.M., Wozney, H. and Boedtker (1977) Biochemistry 16, 4743-4751
25. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J.Mol.Biol. 113, 237-251
26. Mackenzie, P.I., Gonzalez, F.J. and Owens, S.I. (1984) J.Biol.Chem. 259, 12153-12160
27. Kozak, M. (1983) Microbiol.Rev. 47, 1
28. Leach, B.S., Collawn, J.F. and Fish, W.W. (1980) Biochem. 19, 5734-5741
29. Gonzalez, F.J., Nebert, D.W., Hardwick, J.P. and Kasper, C.B. (1985) J.Biol.Chem. 260, 7435-7441
30. Yabusaki, Y., Shimizu, M., Murakami, H., Nakamura, K., Oeda, K. and Ohkawa, H. (1984) Nucleic Acids Res. 12, 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 80, 3958-3962
33. Leighton, J.K., De Brunner-Yossbrinck, B.A. and Kemper, B. (1984) Biochemistry 23, 204-210
34. Mackenzie, P.I., Gonzalez, F.I. and Owens, S.I. (1984) Arch.Biochem.Biohys. 230, 676-680
35. Lillienblum, W., Walli, A.K. and Bock, K.W. (1982) Biochem.Pharmacol. 31, 907-913
36. Matsui, M., Nagai, F. and Aoyagi, S. (1979) Biochem.J. 179, 483-487
37. Matsui, M. and Watenbe, H.K. (1982) Biochem.J. 204, 441-447
38. Green, M.D., Falany, C.N., Kirkpatrick, R.B. and Tephly, T.R. (1985) Biochem.J. 230, 403-409