

Cloning of a human liver microsomal UDP-glucuronosyltransferase cDNA

Michael R. JACKSON, Lee R. McCARTHY,* David HARDING, Stuart WILSON, Michael W. H. COUGHTRIE and Brian BURCHELL

Department of Biochemistry, University of Dundee, Dundee DD1 4HN, Scotland, U.K.

A cDNA clone (HLUG 25) encoding the complete sequence of a human liver UDP-glucuronosyltransferase was isolated from a λ gt11 human liver cDNA library. The library was screened by hybridization to a partial-length human UDP-glucuronosyltransferase cDNA (pHUDPGT1) identified from a human liver pEX cDNA expression library by using anti-UDP-glucuronosyltransferase antibodies. The authenticity of the cDNA clone was confirmed by hybrid-select translation and extensive sequence homology to rat liver UDP-glucuronosyltransferase cDNAs. The sequence of HLUG 25 cDNA was determined to be 2104 base-pairs long, including a poly(A) tail, and contains a long open reading frame. The possible site of translation initiation of this sequence is discussed with reference to a rat UDP-glucuronosyltransferase cDNA clone (RLUG 38).

INTRODUCTION

Liver microsomal UDP-glucuronosyltransferase (UDPGT, EC 2.4.1.17) is of major importance in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds [1]. Overwhelming evidence now indicates that in rat liver these glucuronidation reactions are catalysed by a family of isoenzymes, which exhibit different substrate specificities and slightly different molecular masses on examination by SDS/polyacrylamide-gel electrophoresis [2,3]. Limited peptide-mapping experiments indicated that the separated rat liver enzymes are different polypeptides [4], and more recently cloning of the rat liver UDPGTs has revealed at least four classes of highly homologous cDNAs with some differences in base and deduced amino acid sequences [5].

Little is known about the ability of human liver to glucuronidate substrates other than bilirubin [6]. The existence of isoenzymes that would specifically glucuronidate bilirubin, bile acids or various drugs such as chloramphenicol and morphine has been suggested from kinetic analyses of results obtained *in vitro* [7–9], although different UDPGT isoenzymes have not yet been identified. Recently our immuno-blot analyses have detected the existence of at least five UDPGTs in human liver microsomal fraction (M. W. H. Coughtrie & B. Burchell, unpublished work, and see Fig. 1).

Inherited defects of bilirubin UDPGT have been suggested as the major factor causing Crigler–Najjar syndrome [10] and Gilbert's syndrome [11]. Deficiency of UDPGT in the newborn infant results in high risk of chloramphenicol toxicity during this period.

The examination of the multiplicity, genetic deficiency, regulation and development of human liver UDPGTs requires the use of molecular probes to the structural genes and mRNAs, which will facilitate the study of these problems *in vitro*.

In the present paper we report the molecular cloning

and sequencing of a human UDPGT cDNA and, by comparison with a rat UDPGT cDNA sequence, demonstrate the authenticity of the clone and indicate the possible translation-initiation site of this sequence.

MATERIALS AND METHODS

Human liver samples

Samples of human liver from a number of donors were kindly provided by R. Wolf, I.C.R.F. Unit, University of Edinburgh, Edinburgh, Scotland, U.K.

Immunological studies

Antibodies against rat testosterone/phenol UDPGT were raised in sheep [12,13]. The specificity of interaction was determined by immuno-blot analysis [14] of human microsomal proteins with detection by the use of immunoperoxidase with 4-chloro-1-naphthol as substrate [15] and by inhibition of UDPGT activity assayed towards the substrates 1-naphthol [16], testosterone [17] and bilirubin [18].

Screening of human liver cDNA libraries and analysis of human cDNA

A human liver cDNA library in the bacterial expression vector pEX2 was kindly provided by K. Stanley and P. Luzio [19]. In all 50000 recombinants were screened by the procedure of Stanley [20] with affinity-purified sheep anti-(rat UDPGT) IgG labelled with 125 I [21]. Positive colonies were taken through successive rounds of screening until all colonies gave a positive signal.

Fusion proteins were analysed by immuno-blotting [21] of bacterial lysate [20,21] with the same 125 I-labelled antibody preparation. Plasmid DNA was prepared from positive clones, and the cDNA insert mapped by using a variety of restriction enzymes [21]. Appropriate restriction

Abbreviations used: UDPGT, UDP-glucuronosyltransferase; poly(A)⁺ RNA, polyadenylated RNA.

* Present address: Department of Obstetrics and Gynaecology, Bristol Maternity Hospital, Bristol BS2 8EG, U.K.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00317.

fragments were subcloned into M13 mp18 and mp19 for sequencing by the dideoxy method [22,23].

Recombinant plasmid DNA was used to hybrid-select human UDPGT mRNA [24] from human liver poly(A)⁺ RNA prepared by using guanidinium isothiocyanate [25] and affinity chromatography on oligo(dT)-cellulose [26]. Hybrid-selected RNA was analysed by translation *in vitro* in a rabbit reticulocyte lysate and selective immunoprecipitation from the translation products with anti-UDPGT IgG [21].

Total RNA prepared from rat and human liver was analysed by Northern blotting after electrophoresis on denaturing formaldehyde/agarose gels [27]. Hybridization conditions were as previously described [5] but with the inclusion of a single-stranded ³²P-labelled cDNA probe transcribed from an M13 cDNA recombinant containing the complete cDNA insert of pHUDPGT1 in the appropriate orientation [28].

A second human liver cDNA library in the vector λgt11 [29] kindly provided by Savio Woo was screened by plaque hybridization [30] with a nick-translated [31] ³²P-labelled 500 bp coding portion of the human cDNA insert of pHUDPGT1. In all 250 000 recombinants were screened and positive plaques were purified by successive rounds of screening. The cDNA inserts in the recombinants were sized by Southern blotting of *Eco*RI digests of bacteriophage mini-preparations as described by Kwok *et al.* [29]. cDNA inserts were isolated from large-scale preparations of recombinant bacteriophage [27] and subcloned into plasmid pUC8 and M13 mp18 and mp19 for restriction-enzyme mapping and sequencing [21,22,27]. DNA sequence was analysed by using the Staden programmes [32].

Isolation of rat liver UDPGT cDNAs from a rat liver λgt11 library was by plaque hybridization as previously described [21] with a nick-translated [31] ³²P-labelled fragment of androsterone UDPGT cDNA [5]. The sequencing of rat UDPGT cDNA was carried out as previously described [5].

RESULTS AND DISCUSSION

Cross-reactivity of anti-(rat UDPGT) antibodies with human UDPGTs

Antibodies raised in sheep against purified rat liver phenol/testosterone UDPGT have previously been shown to inhibit specifically rat liver microsomal UDPGT activities towards the substrates 1-naphthol, bilirubin, testosterone and androsterone [12,13; M. W. H. Coughtrie, B. Burchell, I. M. Shepherd & J. R. Bend, unpublished work]. The antibody also specifically identifies, by immuno-blot analysis [13], the UDPGT isoenzymes that catalyse these glucuronidation reactions (see Fig. 1).

This antibody preparation was tested for its ability to interact with human liver microsomal UDPGTs. UDPGT activities towards 1-naphthol, bilirubin and testosterone were specifically inhibited by up to 85% by incubation with anti-(rat liver UDPGT) antibodies. Further, the antibody recognizes a similar spectrum of UDPGT isoenzymes (52–56 kDa) to those observed on immuno-blot analysis of rat hepatic microsomal fraction (Fig. 1). This result illustrates for the first time the existence of separable UDPGT isoenzymes in human liver, a feature well documented for rat liver [3,13].

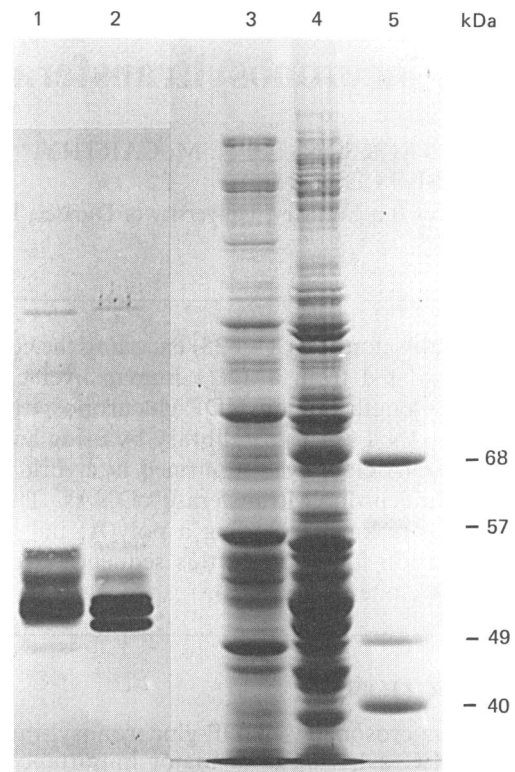


Fig. 1. Immuno-blot analysis of human and rat liver microsomal proteins with anti-(rat UDPGT) IgG

Microsomal proteins from the liver of an adult human [20 µg (lane 1) and 100 µg (lane 3)] or Wistar rat [10 µg (lane 2) and 100 µg (lane 4)] were analysed by electrophoresis on an SDS/7.5% polyacrylamide gel. The proteins in lanes 1 and 2 were transferred to nitrocellulose, and the blots were incubated with anti-(rat UDPGT) IgG. Chromogenic detection of specifically bound IgG involved the immunoperoxidase system with 4-chloro-1-naphthol as the substrate. The other half of the gel (lanes 2 and 4) was stained with Coomassie Brilliant Blue and aligned to the blot with the aid of standard proteins (lane 5).

A comparative analysis of the UDPGT isoenzyme complement present in hepatic microsomal fractions from a Crigler-Najjar child and post-mortem samples from other infants of similar age indicates that the putative bilirubin UDPGT isoenzyme (M_r 53 kDa) is absent from the hepatic microsomal fraction of the Crigler-Najjar child (M. W. H. Coughtrie & B. Burchell, unpublished work). The use of this genetic deficiency further suggests a specific interaction of our anti-(rat UDPGT) antibody with human UDPGT isoenzymes.

Isolation of cDNAs coding for human UDPGTs

Screening of 50 000 recombinants from the pEX2 human cDNA expression library with an affinity-purified iodinated anti-UDPGT IgG identified 11 immunopositive clones. Plasmid DNA prepared from these clones was digested with *Pst*I, and the restriction fragments were analysed by agarose-gel electrophoresis. This procedure showed that all the clones contained the same-sized cDNA insert of approx. 1.1 kb. Further digestion of the DNA with *Sac*I or *Bam*HI showed the cDNA to have the same restriction maps. All of the

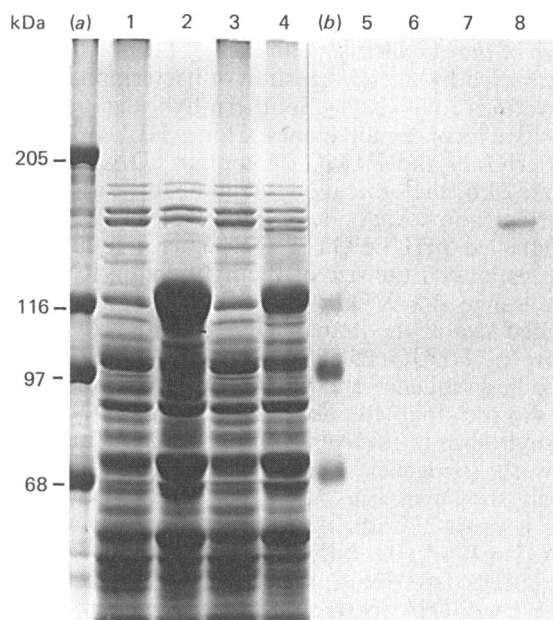


Fig. 2. Characterization of *cro* β -galactosidase-UDPGT fusion proteins

Cultures of *Escherichia coli* transformed with pEX2 (lanes 1 and 2) or pHUDPGT1 (lanes 3 and 4) were grown at 30 °C, and exponential-phase cells were either heat-shocked at 42 °C (lanes 2 and 4) or maintained at 30 °C (lanes 1 and 3) for a further 2 h. Bacterial lysates were prepared from each culture and analysed by SDS/polyacrylamide-gel electrophoresis. (a) shows a Coomassie Brilliant Blue-stained gel; (b) shows an autoradiograph of an identical gel immunoblotted with ^{125}I -labelled affinity-purified anti-UDPGT IgG. Molecular masses (kDa) of protein standards are indicated on the left of the Figure.

recombinant clones produced an identical-sized fusion protein of approx. 145 kDa, some 28 kDa larger than the pEX-encoded *cro* β -galactosidase protein (Fig. 2). We concluded that a single recombinant had apparently been isolated many times as a consequence of its overamplification in the library. Thus only one clone (pHUDPGT1) was further analysed.

Hybrid-select translation of human UDPGT mRNA with pHUDPGT cDNA

To confirm the identity of the isolated human UDPGT cDNA, the cDNA insert of pHUDPGT 1 was used to hybridize to poly(A)⁺ mRNA prepared from the human liver. The selected mRNA was then translated *in vitro* by using a rabbit reticulocyte lysate and a [^{35}S]methionine label. The translation products were shown by SDS/polyacrylamide-gel electrophoresis to consist of polypeptides of molecular mass 51 and 52 kDa (Fig. 3, lane 5), similar in size to rat UDPGTs that had been translated from mRNA *in vitro* (Fig. 3, lane 2). The 51 kDa protein was immunoprecipitated with the anti-UDPGT antibody (Fig. 3, lane 6). The 52 kDa protein may also be immunoprecipitated, but was not resolved from the lower-molecular-mass form in this analysis owing to distortion of its molecular mass by the presence of large amounts of IgG heavy chain, which has a similar molecular mass.

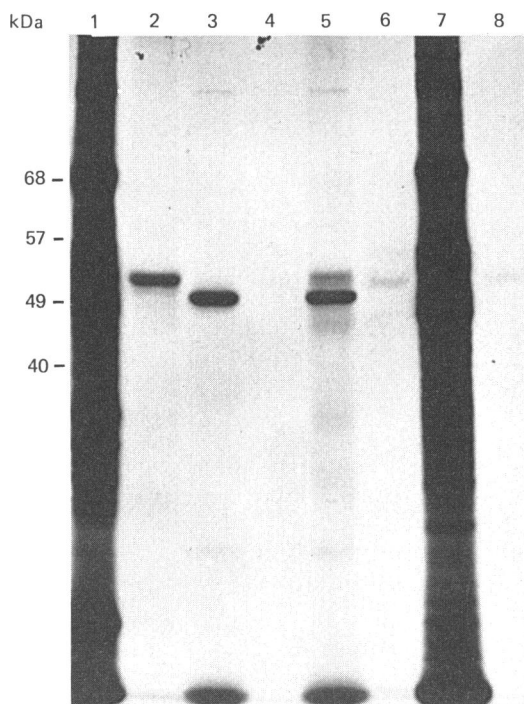


Fig. 3. Selection of translatable human UDPGT mRNA by hybridization to pHUDPGT1 DNA

Total human liver poly(A)⁺ RNA was hybridized to plasmid DNA bound to nitrocellulose filters. Specifically hybridizing mRNA was eluted and translated *in vitro* in a reticulocyte-lysate system including [^{35}S]methionine. Translation products of hybrid-selected and poly(A)⁺ RNA were incubated with anti-UDPGT IgG, and the immunoprecipitation and the total translation products were subjected to SDS/polyacrylamide-gel electrophoresis. The gel was subjected to fluorography for 4 days. Translation products directed by rat liver poly(A)⁺ RNA, by human poly(A)⁺ RNA and by mRNA hybrid selected by pEX2 and pHUDPGT1 are shown in lanes 1, 7, 3 and 5 respectively. The immunoprecipitates from these translation products with anti-(rat UDPGT) IgG are shown in the adjacent lanes 2, 8, 4 and 6. The highly radioactive band in lanes 3 and 5 is an artifact of the reticulocyte-lysate system. Molecular masses of protein standards (in kDa) are indicated on the left.

Characterization of human UDPGT cDNA

The nucleotide sequence of the cDNA insert pHUDPGT 1 was determined by using the dideoxy chain-termination method and the strategy depicted in Fig. 5. This cDNA sequence showed strong homology to rat liver UDPGT cDNAs [5] over the 5'-terminal 850 bp; however, the sequence was incomplete and did not extend far enough 5' to encode a complete amino acid sequence for UDPGT. A second cDNA library was thus screened to obtain a clone containing a complete cDNA. The 5'-terminal *Pst*I-*Sac*I restriction fragment (see Fig. 5) of the cDNA insert of pHUDPGT 1 was used as a hybridization probe to screen a human liver cDNA library in the vector λ gt11. Approximately 200000 recombinants from the library were screened, and 14 positive clones were identified and plaque-purified. None of these positive clones produced a fusion protein

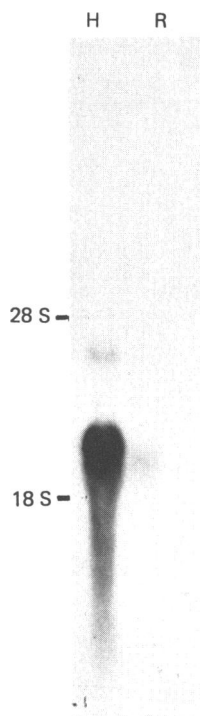


Fig. 4. Northern-blot analysis of human (H) and rat (R) hepatic UDPGT mRNAs

A 10 μ g portion of human liver total RNA was electrophoresed on a 1% formaldehyde/agarose gel, transferred to Zeta probe membrane, and hybridized to 32 P-labelled DNA probe prepared from the cDNA insert of pHUDPGT1 (as described in the Materials and methods section). The blot was washed to a final stringency of $2 \times$ SSC ($1 \times$ SSC is 0.15 M-NaCl/0.015 M-sodium citrate buffer, pH 7.0) at 55 $^{\circ}$ C and then autoradiographed for 24 h. Size markers were rat rRNA and are indicated on the left.

detectable with anti-UDPGT antibody. Determination of the size of the cDNA inserts in each of the positive clones was achieved by *Eco*RI digestion of bacteriophage DNA preparations followed by Southern hybridization of the 14 positive recombinant clones. Clones HLUG 2, HLUG 16, HLUG 24 and HLUG 25 contain cDNA inserts of approx. 2 kb; the other clones contained shorter inserts.

Northern-blot analysis of human liver RNA, with radiolabelled pHUDPGT 1 cDNA as a hybridization probe, estimated the size of the hybridizing RNA to be in the range 2.3–2.7 kb (Fig. 4), consistent with the predicted size of the human UDPGT mRNA based on the size of HLUG 25 cDNA if additional 5'-terminal non-coding sequence and poly(A) tail is included. Fig. 4 also showed that the isolated human cDNA could cross-hybridize to rat liver UDPGT mRNA (but only at low wash stringencies). It is thus likely that in the analysis shown in Fig. 4 the human UDPGT cDNA probe is cross-hybridizing to several related classes of UDPGT mRNA; weak hybridization to a 3.8 kb human RNA species (see Fig. 4) is particularly interesting, as a similar-sized RNA species is also recognized in Northern blots of rat liver RNA with a class 2 rat UDPGT cDNA hybridization probe [5].

A partial restriction map of HLUG 25 cDNA is shown in Fig. 5. The restriction sites indicated were used to aid directional subcloning of DNA fragments into M13 mp18 and mp19 for subsequent sequencing by the dideoxy method. The correct arrangement of the three *Eco*RI restriction fragments was determined by sequencing through the two internal *Eco*RI sites. Partial sequencing and restriction mapping of nine of the other human cDNAs showed that they all coded for the same mRNA and that they differed in size owing to variation in the length of their 5'-terminal sequence, with all clones containing a poly(A) tail at their 3'-terminus.

The complete nucleotide sequence of HLUG 25 cDNA is shown in Fig. 6, without the *Eco*RI linker

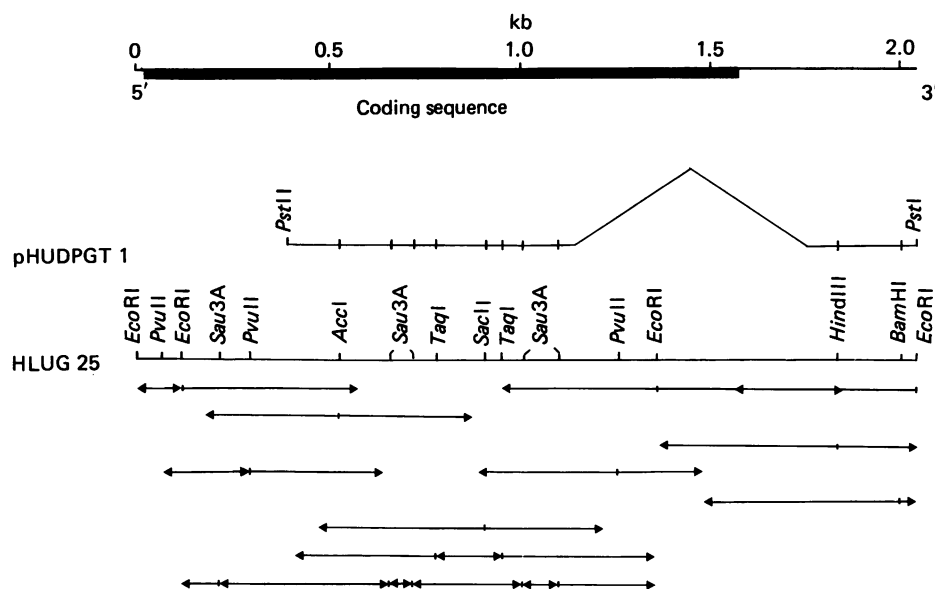


Fig. 5. Partial-restriction map and nucleotide sequencing strategy of the cDNA insert in pHUDPGT1 and λ HLUG 25

Extent of sequencing is shown by horizontal arrows, where the direction of each arrow indicates the strand that was sequenced from the indicated restriction site. The break in the horizontal line of pHUDPGT1 indicates the deletion in this cDNA clone. The *Eco*RI linker sites are shown flanking the HLUG 25 cDNA, and the *Pst*I sites shown in pHUDPGT 1 cDNA are derived from G-C tailing.

sequences. Every position in the sequence was determined by sequencing both strands of the cDNA. The sequence is 2104 bp long and contains a long open reading frame that is terminated by the stop codon UAA; this differs from the stop codon used in the four rat UDPGT mRNAs so far determined [5], where it has been shown to be UAG. There is a putative polyadenylation sequence AATAAA 12 bases upstream from the poly(A) tail; the length of this tail is 11 residues in HLUG 25; however, sequence data from the other human cDNA clones show that the poly(A) tail length was up to 40 bases in one clone. A comparison of this cDNA sequence with that of the cDNA insert in pHUDPGT 1 showed both clones to be coding for the same UDPGT. However, the cDNA insert of pHUDPGT was smaller, extending from position 406 to the poly(A) tail in the HLUG 25 sequence but with a large deletion of 614 bases between bases 1179 and 1793. It is interesting to speculate that the palindromic-sequence nucleotide position 1789–1799 (HLUG 25) of GAAATATTC might be involved in this cloning artifact (perhaps hybridizing to the sequence ATCTATAAGG at nucleotide position 1175–1185).

Comparison of human and rat liver UDPGT cDNAs

The human cDNA sequence showed strong homology to the DNA sequence of rat liver UDPGT cDNAs [5]. Unfortunately we do not have complete cDNAs for all the classes of rat UDPGT mRNA, and this makes meaningful comparisons of the human sequence with individual rat cDNAs difficult. However, examination of the available sequence data indicates that none of the rat cDNA sequences that we have obtained is strikingly more homologous than another to the human sequence reported here.

The human cDNA sequence was compared with the sequence of RLUG38 cDNA (a recently isolated member of class 3 UDPGT cDNA [5]). This comparison (Fig. 6) shows the strong homology of the sequences throughout the coding region (73%); two single-codon gaps in the human sequence are required to maximize homology (see Fig. 6). Homology between the 5'- and 3'-terminal non-coding sequences of HLUG 25 and RLUG38 is not apparent.

Identification of the translation-initiation codon of UDPGT cDNAs

In our previously reported sequence analysis of rat UDPGT cDNAs [5] we considered the ATG at nucleotide position 166–168 underlined in RLUG 38 cDNA (Fig. 6) to be the translation-initiation codon for another rat liver UDPGT cDNA (RLUG 23 androsterone UDPGT cDNA [5]). This assignment was based on the homology of the initiation-codon flanking sequence to the consensus start sequence [33] and on the fact that initiation of translation at this codon predicted a protein whose molecular mass was over 57 kDa. This predicted UDPGT was thus calculated to be some 5 kDa larger than the reported molecular mass of the purified isoform [34]. UDPGT proteins synthesized *in vitro* by using a reticulocyte lysate were apparently of a similar size to the proteins isolated from hepatic microsomal fraction [21], and therefore the difference between the observed and predicted molecular masses was believed to be due to anomalous behaviour of UDPGT on SDS/polyacrylamide-gel electrophoresis.

The initiation codon indicated above is not, however, present in the equivalent position in the human sequence, although an ATG 18 bp downstream is available. The problem is further complicated by analysis of the 5'-terminal sequences of HLUG 25 and RLUG 38 cDNA (see Fig. 6). These sequences contain longer 5'-terminal UDPGT cDNA sequences than that of the RLUG 23 UDPGT cDNA originally reported [5]. The open reading frames in both these sequences extend further upstream than either of the above proposed initiation sites. Thus the ATG at nucleotide position 76–78 in RLUG 38 (see Fig. 6) might be used to initiate translation of some of this sequence in the rat clone. Likewise, translation of the human cDNA clone, which is very highly homologous to the rat cDNA over this 5'-terminal region (Fig. 6), might then be initiated at the ATG at nucleotide position 38–40 (HLUG 25) or at a second ATG located two codons downstream, at nucleotide position 44–46 (HLUG 25). The predicted molecular masses of the unmodified rat and human UDPGTs, translated from this earlier initiation codon, are now calculated to be over 60 kDa, whereas human and rat UDPGTs synthesized in reticulocyte lysates have been estimated to have molecular masses of approx. 52 kDa (see Fig. 3). Anomalous migration of UDPGT on SDS/polyacrylamide-gel electrophoresis might explain some of this discrepancy, as mentioned above, and indeed other highly hydrophobic proteins have been previously reported to behave in this manner [35,36].

Inspection of the *N*-terminal sequences of the predicted human and rat proteins translated from the earlier initiation site shows a series of highly hydrophobic residues, which might act as a signal sequence for co-translational insertion of the protein into the endoplasmic-reticulum membrane [37]. Some evidence for post-translational cleavage of UDPGT that had been translated from mRNA *in vitro* has been suggested [38]. Preliminary *N*-terminal sequence analysis of purified rat liver androsterone UDPGT (M. D. Green & T. Tephly, personal communication) indicates that the mature protein begins seven residues upstream of the originally proposed start site [5], supporting the hypothesis of a cleaved signal sequence. Therefore the mature human UDPGT, encoded by the cDNA sequence reported in the present paper, may start at residue 24 (Fig. 6), in good agreement with a consensus cleavage site [39].

The predicted human (HLUG 25) and rat (RLUG 38) UDPGT protein sequences are 64% homologous, and a comparison of the computer-derived hydrophobicity plots [40] of these two sequences (Fig. 7) shows the profiles to be very similar, with the overall pattern of hydrophobic segments conserved. The plots also emphasize a highly hydrophobic segment immediately followed by a very highly charged sequence at the *C*-terminus; this could act as a stop transfer signal sequence [41,42].

The above features suggest UDPGT to be anchored into the endoplasmic-reticulum membrane by the *C*-terminal transmembrane domain, with most of the protein, including the *N*-terminus, being located on the luminal side. The short very highly charged *C*-terminal tail protruding from the membrane into the cytoplasm might help to maintain the membrane orientation of the protein. A putative glycosylation site [43] is found in the predicted human UDPGT sequence at residue 315, consistent with the above model. The proposed topology suggests the active site of the enzyme has a luminal

R38	+GAAAGCTTACTAGATAAAAGGTCAAGCAGTGACAGAAAG A CAC TG CAG AGG TTTGATTT T A ___ C GG A G T	93
H25	+AGCAGCAACTGAAAAACAAGCATTGCATCAGG ATG TCT ATG AAA TGG ACT Met Ser Met Lys Trp Thr	55
R38	T C A A T C GT C CAA AC G T G TG	183
H25	TCA GCT CTT CTG CTG ATA CAG CTG AGC TGT TAC TTT AGC TCT GGG AGT TGT GGA AAG GTG CTG GTG TGG CCC ACA GAA TTC AGC CAC TGG Ser Ala Leu Leu Leu Ile Gln Leu Ser Cys Tyr Phe Ser Ser Gly Ser Cys Gly Lys Val Leu Val Trp Pro Thr Glu Phe Ser His Trp 10 20 30	145
R38	___ A A A G C A C T C AA C G A TA GT C T G	273
H25	ATG AAT ATA AAG ACA ATC CTG GAT GAA CTT GTC CAG AGA GGT CAT GAG GTG ACT GTA TTG GCA TCT TCA GCT TCC ATT TCT TTC GAT CCC Met Asn Ile Lys Thr Ile Leu Asp Glu Leu Val Gln Arg Gly His Glu Val Thr Val Leu Ala Ser Ser Ala Ser Ile Ser Phe Asp Pro 40 50 (CTG) 60	235
R38	A AA T G C A GA G AC TC AC G C G GA ATGAA A T T T C A TT G A GT T G	366
H25	AAC AGC CCA TCT ACT CTT AAA TTT GAA GTT TAT CCT GTA TCT TTA ACT AAA ACT GAG TTT GAG GAT ATT ATC AAG CAG CTG GTT AAG AGA Asn Ser Pro Ser Thr Leu Lys Phe Glu Val Tyr Pro Val Ser Leu Thr Lys Thr Glu Phe Glu Asp Ile Ile Lys Gln Leu Val Lys Arg 70 80 90	325
R38	ACT TAT G T G A G T G TT T C C T TT C A T G A GAT GA TC T TAC TA CT GT C T	456
H25	TGG GCA GAA CTT CCA AAA GAC ACA TTT TGG TCA TAT TTT TCA CAA GTA CAA GAA ATC ATG TGG ACA TTT AAT GAC ATA CTT AGA AAG TTC Trp Ala Glu Leu Pro Lys Asp Thr Phe Trp Ser Tyr Phe Ser Gln Val Gln Glu Ile Met Trp Thr Phe Asn Asp Ile Leu Arg Lys Phe 100 110 120	415
R38	A C ACC C C G C GCA A C AG C T G T C G C T G G	546
H25	TGT AAG GAT ATA GTT TCA AAT AAG AAA CTT ATG AAG AAA CTA CAG GAG TCA AGA TTT GAT GTT GTT CTT GCA GAT GCT GTT TTC CCC TTT Cys Lys Asp Ile Val Ser Asn Lys Lys Leu Met Lys Lys Leu Gln Glu Ser Arg Phe Asp Val Val Leu Ala Asp Ala Val Phe Pro Phe 130 140 150	505
R38	G A A A G G C C C T T C G T T A AA TCC A	636
H25	GGT GAG CTG CTG GCC GAG TTA CTT AAA ATA CCC TTT GTC TAC AGG CCT CGC TTC TCT CCT GGC TAC GCA ATT GAA AAG CAT AGT GGA GGA Gly Glu Leu Leu Ala Glu Leu Leu Lys Ile Pro Phe Val Tyr Arg Pro Arg Phe Ser Pro Gly Tyr Ala Ile Glu Lys His Ser Gly Gly 160 170 180	595
R38	T A A A A T A A A A T G A G G G C A C A G	726
H25	CTT CTG TTC CCT CCT TCC TAT GTG CCT GTT GTT ATG TCA GAA CTA AGT GAC CAA ATG ACT TTC ATA GAG AGG GTA AAA AAT ATG ATC TAT Leu Leu Phe Pro Pro Ser Tyr Val Pro Val Val Met Ser Glu Leu Ser Asp Gln Met Thr Phe Ile Glu Arg Val Lys Asn Met Ile Tyr 190 200 210	685
R38	AC C T G A T G C C A T G A T T G C A	816
H25	GTG CTT TAT TTT GAA TTT TGG TTC CAA ATA TTT GAC ATG AAG AAG TGG GAT CAG TTC TAC AGT GAA GTT CTA GGA AGA CCC ACT ACG TTA Val Leu Tyr Phe Glu Phe Trp Phe Gln Ile Phe Asp Met Lys Lys Trp Asp Gln Phe Tyr Ser Glu Val Leu Gly Arg Pro Thr Thr Leu 220 230 240	775
R38	G GC A A G C A TC GG C ACA C AT	906
H25	TCT GAG ACA ATG GCA AAA GCT GAC ATA TGG CTT ATT CGA AAC TAC TGG GAT TTT CAA TTT CCT CAC CCA CTC TTA CCA AAT GTT GAG TTC Ser Glu Thr Met Ala Lys Ala Asp Ile Trp Leu Ile Arg Asn Tyr Trp Asp Phe Gln Phe Pro His Pro Leu Leu Pro Asn Val Glu Phe 250 260 270	865
R38	A A G CCT T G C T T G C C G	996
H25	GTT GGA GGA CTC CAC TGC AAA CCT GCC AAA CCC CTA CCG AAG GAA ATG GAA GAG TTT GTC CAG AGC TCT GGA GAA AAT GGT GTT GTG GTG Val Gly Gly Leu His Cys Lys Pro Ala Lys Pro Leu Pro Lys Glu Met Glu Glu Phe Val Gln Ser Ser Gly Glu Asn Gly Val Val Val 280 290 300	955
R38	A C G T A A C GG C T T	1086
H25	TTT TCT CTG GGG TCG ATG GTC AGT AAC ACG TCA GAA GAA AGG GCC AAT GTA ATT GCA TCA GCC CTT GCC AAG ATC CCA CAA AAG GTT CTG Phe Ser Leu Gly Ser Met Val Ser Asn Thr Ser Glu Glu Arg Ala Asn Val Ile Ala Ser Ala Leu Ala Lys Ile Pro Gln Lys Val Leu 310 320 330	1045
R38	A C A CC GCA C C C A A G C C T C C T	1176
H25	TGG AGA TTT GAT GGG AAT AAA CCA GAT ACT TTA GGA CTC AAT ACT CGG CTG TAC AAG TGG ATA CCC CAG AAT GAT CTT CTT GGT CAC CCA Trp Arg Phe Asp Gly Asn Lys Pro Asp Thr Leu Gly Leu Asn Thr Arg Leu Tyr Lys Trp Ile Pro Gln Asn Asp Leu Leu Gly His Pro 340 350 360	1135
R38	A C G T G G C A A G G T A T A	1266
H25	AAA ACC AGA GCT TTT ATA ACT CAT GGT GGA GCC AAT GCC ATC TAT AAG GCA ATC TCT CCT AGA ATC CCT ATG GTG GGC GTT CCA TTG TTT Lys Thr Arg Ala Phe Ile Thr His Gly Gly Ala Asn Gly Ile Tyr Lys Ala Ile Ser Pro Arg Ile Pro Met Val Gly Val Pro Leu Phe 370 380 390	1225
R38	G A A C GT A CA C A T A AGG A AG T G T	1356
H25	GCA GAT CAA CCT GAT AAC ATT GCA CAC ATG AAG GCC AAG GGA GCA GCT GTT AGT TTG GAC TTC CAC ACA ATG TCG AGT ACA GAC TTA CTC Ala Asp Gln Pro Asp Asn Ile Ala His Met Lys Ala Lys Gly Ala Ala Val Ser Leu Asp Phe His Thr Met Ser Ser Thr Asp Leu Leu 400 410 420	1315
R38	A GA A A C A A C A A G TGG G CC C T A	1446
H25	AAT GCA CTG AAG ACA GTA ATT AAT GAT CCT TTA TAT AAA GAG AAT GCT ATG AAA TTA TCA AGA ATT CAT CAT GAT CAA CCA GTG AAG CCC Asn Ala Leu Lys Thr Val Ile Asn Asp Pro Leu Tyr Lys Glu Asn Ala Met Lys Leu Ser Arg Ile His His Asp Gln Pro Val Lys Pro 430 440 450	1405
R38	G C AAG T G G C G A CCA CTT GA T T C A	1536
H25	CTT GAT CGA GCA GTC TTC TGG ATT GAA TTT GTC ATG CGC CAT AAA GGA GCC AAG CAC CTT CGG GTT GCA GCC CAC GAC CTC ACC TGG TTC Leu Asp Arg Ala Val Phe Trp Ile Glu Phe Val Met Arg His Lys Gly Ala Lys His Leu Arg Val Ala Ala His Asp Leu Thr Trp Phe 460 470 480	1495
R38	C T A C T TC GTC A T GC G C T CT GT C T C G T A T AC	1626
H25	CAG TAC CAC TCT TTG GAT GTG ACT GGG TTC CTG CTG GCC TGT GTG GCA ACT GTG ATA TTC ATC ATC ACA AAA TGT CTG TTT TGT GTC TGG Gln Tyr His Ser Leu Asp Val Thr Gly Phe Leu Leu Ala Cys Val Ala Thr Val Ile Phe Ile Ile Thr Lys Cys Leu Phe Cys Val Trp 490 (CTC) 500 510	1585
R38	CGA G AG AG A A AT G AT G G AGCTCAT T CAA C CTACA TGAA T TC GCC ATTC A +	1691
H25	AAG TTT GTT AGA ACA GGA AAG AAG GGG AAA AGA GAT TAA TTACGTCTGAGGCTGGAAGCTGGGAAACCCAATAAATGAACCTCTTTAGTTTATTACAACAAGA Lys Phe Val Arg Thr Gly Lys Lys Gly Lys Arg Asp 520	

```

H25  CGTTGTGATACAAGAGATTCCCTTCTCTTGTGACAAAACATCTTCAAACCTACCTTGTCAGTCAAATTTGTTTTAGTACCTGTTAACCATAGAAAATATTTTCATGTCAAGGAG  1810
H25  GAAAACATTAGGAAAAACAAAATGATATAAAGCCATATGAGGTTATATGAAATGTATTGAGCTTATATTGAAATTTATTGTTCCAATTCACAGGTTACATGAAAAAAATTTACTAA  1929
H25  GCTTAACATACATGTCACACATTGTACATGAAACAAGAACATTAAGAAGTCCGACTGACAGTATCAGTACTGTTTGGCAAATACTCAGCATACTTTGGATCCATTTTCATGCAGGATTGT  2048
H25  GTTGTTTTAACTGTTGTTGAGGAAGCTAATAAATAAATAAATTGTAATAAAAAAAAAA+

```

Fig. 6. Comparison of nucleotide sequences of HLUg 25 cDNA and RLUG 38 cDNA

The nucleotide sequence of HLUg 25 (H25) is shown with the predicted human UDPGT protein sequence, amino acid residues being numbered from the first full codon. The homology of HLUg 25 cDNA (H25) to a rat UDPGT cDNA, RLUG 38 (R38), is shown with the substitutions from the human sequence occurring as indicated. Two additional codons are required in the rat sequence (R38) at nucleotide positions 330–332 and 1630–1632 (shown in parentheses in the correct positions). Potential initiation codons are shown underlined in the 5'-terminal sequences of HLUg 25 and RLUG 38. A putative glycosylation site Asn-Xaa-Ser in the human UDPGT sequence is shown underlined at position 315–317. The beginning and end of the cDNA sequences are marked by + signs.

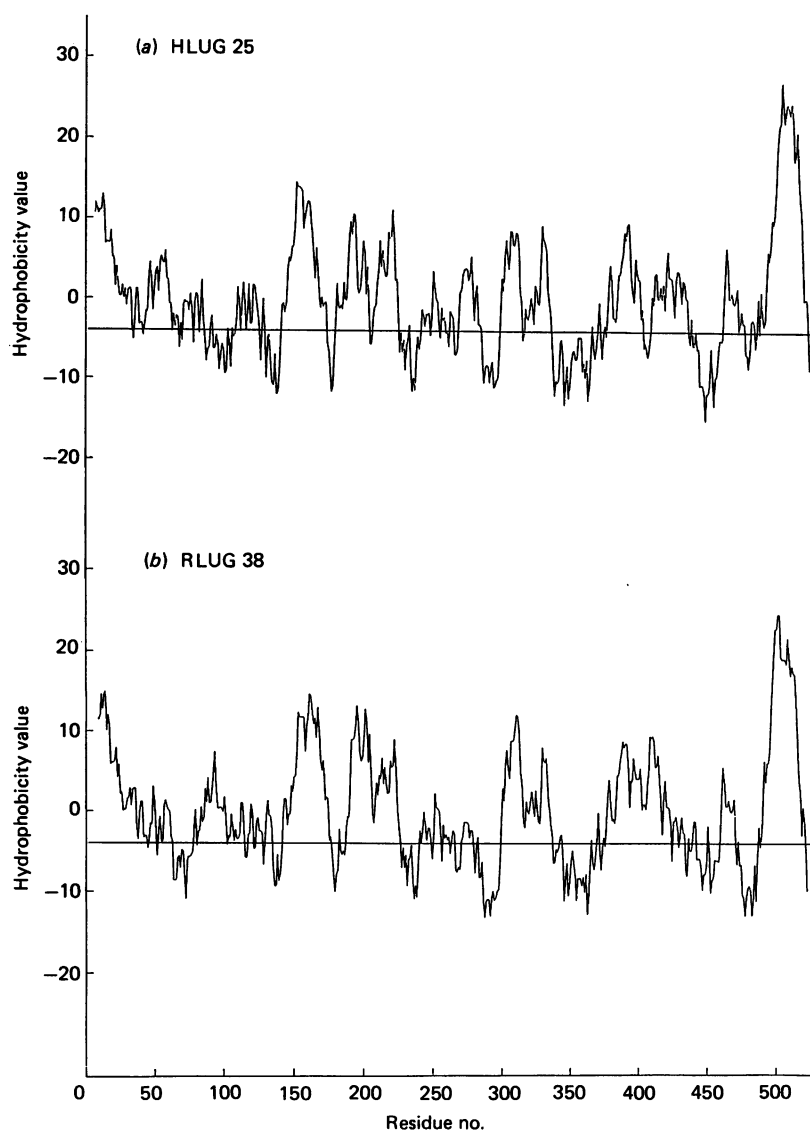


Fig. 7. Hydrophobicity profiles of human (HLUG 25) and rat (RLUG 38) UDPGTs

The UDPGT amino acid sequences deduced from the translation of the open reading frames in HLUg 25 cDNA (a) and RLUG 38 cDNA (b) respectively were analysed by computer for hydrophobicity by using the program of Kyte & Doolittle [40] with a 16-residue window. The horizontal line at -4 indicates the average hydrophobicity of a typical protein.

location, and this would certainly help to explain the trypsin-insensitivity of UDPGT in intact microsomal fraction [44] and influence theories concerning the known latency of the enzyme [1].

In conclusion, we have isolated a human UDPGT cDNA that will undoubtedly be useful to probe the molecular biology of human UDPGTs. At present the biochemistry of human UDPGT is poorly understood, and there have been no reports of any isoenzyme having been purified to homogeneity. While that work is underway, the exact identity of the protein encoded by this cDNA (in terms of substrate specificity) will probably require expression of the enzyme activity in yeast [45] or cultured cells [46,47]. The isolated cDNA is being used to screen further the human cDNA library at lower hybridization stringencies in an attempt to identify other related human cDNAs. In this way we should rapidly obtain an understanding of the molecular heterogeneity of this family of proteins, which play such a key role in the metabolism of drugs and detoxification of foreign compounds.

We are especially grateful to Paul Luzio, Richard Hynes and Savio Woo for supplying their liver cDNA libraries, to Roland Wolf for making available human liver samples and to Mitchell Green and Tom Tephly for communicating preliminary *N*-terminal sequence analysis of purified rat liver androsterone UDPGT before publication. We also thank the Medical Research Council and Wellcome Trust for grants supporting this work. M. W. H. C. held a Science and Engineering Research Council studentship tenable overseas. B. B. is a Wellcome Trust senior lecturer.

REFERENCES

- Dutton, G. J. (1980) *Glucuronidation of Drugs and other Compounds*, CRC Press, Boca Raton
- Burchell, B. (1981) *Rev. Biochem. Toxicol.* **3**, 1–32
- Burchell, B., Jackson, M. R., McCarthy, L. R. & Barr, G. C. (1985) in *Microsomes and Drug Oxidations* (Boobis, A. R., Caldwell, J., De Matteis, F. & Elcombe, C. R., eds.), pp. 212–220, Taylor and Francis, London and Philadelphia
- Roy Chowdhury, J., Roy Chowdhury, N., Falany, C. N. & Tephly, T. R. (1986) *Biochem. J.* **233**, 827–837
- Jackson, M. R. & Burchell, B. (1986) *Nucleic Acids Res.* **14**, 779–795
- Black, M., Billing, B. H. & Heirwegh, K. P. M. (1970) *Clin. Chim. Acta* **29**, 27–35
- Bock, K. W., Brunner, G., Hoensch, H., Huber, E. & Josting, D. (1978) *Eur. J. Clin. Pharmacol.* **14**, 367–373
- Mahu, J., Preaux, A., Mavier, P. & Berthelot, P. (1981) *Enzyme* **26**, 93–102
- Parquet, M., Pessah, M., Sacquet, E., Salvat, C., Raizman, A. & Infante, R. (1985) *FEBS Lett.* **189**, 183–187
- Arias, I. M., Gartner, L. M., Cohen, M., Ezzer, J. & Levi, A. J. (1969) *Am. J. Med.* **47**, 395–409
- Black, M. & Billing, B. H. (1969) *N. Engl. J. Med.* **280**, 1266–1271
- Burchell, B., Kennedy, S. M. E., Jackson, M. R., McCarthy, L. R. & Barr, G. C. (1984) *Biochem. Soc. Trans.* **12**, 50–53
- Scragg, I., Celier, C. & Burchell, B. (1985) *FEBS Lett.* **183**, 37–42
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- Domin, B. A., Serabjit-Singh, C. J. & Philpot, R. M. (1984) *Anal. Biochem.* **136**, 390–396
- Otani, G., Abou-el-Makerem, M. M. & Bock, K. W. (1976) *Biochem. Pharmacol.* **25**, 1293–1297
- Rao, G. S., Haueter, G., Rao, M. L. & Breuer, H. (1976) *Anal. Biochem.* **74**, 35–40
- Heirwegh, K. P. H., Van der Vijver, M. & Fevery, J. (1972) *Biochem. J.* **129**, 605–608
- Stanley, K. K. & Luzio, J. P. (1984) *EMBO J.* **3**, 1429–1434
- Stanley, K. K. (1983) *Nucleic Acids Res.* **11**, 4077–4092
- Jackson, M. R., McCarthy, L. R., Corser, R. B., Barr, G. C. & Burchell, B. (1985) *Gene* **34**, 147–153
- Messing, J., Crea, R. & Seeburgh, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Parnes, J. R., Velan, B., Felsingeld, A., Ramanathan, L., Ferrini, U., Appela, E. & Seidmen, J. C. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2253–2257
- Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408–1412
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **16**, 4743–4751
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Jeffreys, A. L., Wilson, V. & Thein, S. L. (1985) *Nature (London)* **314**, 67–73
- Kwok, S. C. M., Ledley, F. D., DiLella, A. G., Robson, K. J. H. & Woo, S. L. C. (1985) *Biochemistry* **24**, 556–561
- Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251
- Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694
- Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–32
- Falany, C. N. & Tephly, T. (1983) *Arch. Biochem. Biophys.* **227**, 248–258
- Gonzalez, F. J., Nebert, D. W., Hardwick, J. P. & Kasper, C. B. (1985) *J. Biol. Chem.* **260**, 7435–7441
- Yabusaki, Y., Shimuzi, M., Murakami, H., Nakamura, K., Oeda, K. & Ohkawa, H. (1984) *Nucleic Acids Res.* **12**, 2929–2938
- Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–851
- Mackenzie, P. & Owens, I. (1984) *Biochem. Biophys. Res. Commun.* **122**, 1441–1449
- von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Sabatini, D. D., Kreibich, G., Morimoto, T. & Adesnik, M. (1982) *J. Cell Biol.* **92**, 1–22
- Yost, C. S., Hedgpeth, J. & Lingappa, V. R. (1983) *Cell (Cambridge, Mass.)* **34**, 759–766
- Hanover, J. A. & Lennarz, W. J. (1981) *Arch. Biochem. Biophys.* **211**, 1–19
- Wilkinson, J. & Hallinan, T. (1977) *FEBS Lett.* **75**, 138–140
- Oeda, K., Sakaki, T. & Ohkawa, H. (1985) *DNA* **4**, 203–210
- Ledley, F. D., Grenett, H. E., DiLella, A. G., Kwok, S. C. M. & Woo, S. L. C. (1985) *Science* **228**, 77–83
- Shibahara, S., Mullar, R., Taguchi, H. & Yoshida, T. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7865–7869