The EMBO Journal vol.6 no. 1 pp.3325–3331, 1987

The human *mdr3* gene encodes a novel P-glycoprotein homologue and gives rise to alternatively spliced mRNAs in liver

A.M.Van der Bliek, F.Baas[1], T.Ten Houte de Lange, P.M.Kooiman, T.Van der Velde-Koerts and P.Borst

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[1]Present address: Department of Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

Communicated by P.Borst

We have found cDNAs corresponding to a novel human P-glycoprotein gene in liver cDNA banks. The sequence of the 3′ part of this cDNA reveals a domainal organization of the derived protein similar to that of the known P-glycoproteins and an 80% amino acid homology with the product of the human *mdr1* gene (Chen *et al.*, 1986). The new gene lies within 500 kb from *mdr1* as determined by pulsed field gradient gel electrophoresis and is designated *mdr3*, as it appears to correspond to the third of the three P-glycoprotein genes mapped in the hamster multidrug resistance domain. *mdr3* yields a transcript of 4100 nucleotides, 400 nucleotides less than the *mdr1* transcript; the difference is accounted for by the shorter 3′-untranslated region of the *mdr3* mRNA. Our cDNAs provide evidence for alternative splicing of *mdr3* pre-mRNAs. One alternative is an insert of seven amino acids between the two major blocks of the nucleotide binding site and another is a deletion of 43 or 47 amino acids covering the putative transmembrane segment 5a. We speculate that these alternatives superimposed on differential expression of P-glycoprotein homologues could provide an explanation for the large variation in cross-resistance patterns observed in cell lines selected for multidrug resistance with different cytostatic drugs.

Key words: P-glycoprotein/alternative splicing/multidrug resistance/drug transport/gene conversion

Introduction

Most multidrug resistant (MDR) cells overproduce a cell membrane protein called P-glycoprotein (Juliano and Ling, 1976; Kartner *et al.*, 1983) and this overproduction can cause MDR, as demonstrated by transfection experiments (Robertson *et al.*, 1984; Gros *et al.*, 1986a; Shen *et al.*, 1986; Deuchars *et al.*, 1987; Sugimoto and Tsuruo, 1987; Ueda *et al.*, 1987). MDR is primarily due to increased drug extrusion from the resistant cell (Dano, 1973; Willingham *et al.*, 1986). The homology of hamster (Gerlach *et al.*, 1986), human (Chen *et al.*, 1986) and mouse (Gros *et al.*, 1986b) P-glycoproteins with bacterial permeases (Ferro-Luzzi Ames, 1986), notably the export system for haemolysin, supports the idea that P-glycoproteins are drug pumps. These proteins possess two homologous halves, each with six hydrophobic segments adjacent to a consensus sequence for nucleotide binding. The series of hydrophobic segments could form a membrane channel analogous to that of other transport systems (Henderson and Maiden, 1987); the nucleotide binding site could be involved in energization of drug export, as the enhanced efflux observed in MDR cells is impaired when cellular ATP is decreased (see Kessel, 1986). P-glycoprotein appears to bind some of the drugs that are actively extruded from MDR cells (Cornwell *et al.*, 1986, 1987; Safa *et al.*, 1986). Together these data support a model in which the overproduced P-glycoproteins directly mediate the energy dependent efflux of cytotoxic drugs.

The large variation of cross-resistance patterns found in cell lines selected with different drugs (Peterson *et al.*, 1983; Riordan and Ling, 1985) suggests that multiple factors determine the precise MDR phenotype, even though overproduction of a single type of P-glycoprotein is sufficient for resistance. Since the cross-resistance patterns associated with particular drug treatments may influence the success of combination chemotherapy, our goal has been to identify these co-determinants. We have isolated cDNAs corresponding to six gene classes that are linked and co-amplified in multidrug resistant Chinese hamster cell lines (Van der Bliek *et al.*, 1986a; De Bruijn *et al.*, 1986; Jongsma *et al.*, 1987; and unpublished results). The proteins encoded by two of the gene classes have been identified: gene class 4 encodes sorcin (V19), a small cytosolic calcium-binding protein with an as yet unknown function (Van der Bliek *et al.*, 1986b) and gene class 2 encodes P-glycoproteins. Amplification always involves the class 2 genes, whereas amplification of the other genes is more variable and not clearly related to the selecting drug or cross-resistance pattern (De Bruijn *et al.*, 1986; Stahl *et al.*, 1987). Hence, the other genes may only co-amplify because of their fortuitous location next to the class 2 (P-glycoprotein) genes and not contribute to resistance at all. Variation in cross-resistance might therefore be due to variation in the nature of the P-glycoprotein expressed.

Circumstantial evidence for the presence of more than one P-glycoprotein gene in the hamster genome has come from the complexity of the genomic DNA fragments hybridizing with P-glycoprotein cDNAs (Riordan *et al.*, 1985); the fact that only some of these fragments are transferred with the MDR genotype in transfection experiments (Deuchars *et al.*, 1987); and the differential amplification of DNA fragments in different MDR cell lines (Van der Bliek *et al.*, 1986a; De Bruijn *et al.*, 1986; Scotto *et al.*, 1986). In man, the presence of at least two P-glycoprotein genes has also been deduced from hybridization experiments (Roninson *et al.*, 1986) and two different, but very homologous, types of P-glycoprotein cDNAs have been obtained from a mouse clone bank (Gros *et al.*, 1986a). Some mammalian cells may therefore be able to produce more than one type of P-glycoprotein. Whether different P-glycoproteins differ in the spectrum of drugs transported remains to be seen.

During an analysis of cDNAs from human liver, we noted that our cDNAs fell into two classes, one corresponding to the *mdr1* gene described by Chen *et al.* (1986), the other to a novel P-glycoprotein gene, substantially different from all P-glycoprotein genes described thus far and designated *mdr3*. We present here a preliminary characterization of this gene; we show that it is part of the human P-glycoprotein cluster; and we present the sequence of three variant cDNAs that must have arisen by alternative splicing of the *mdr3* pre-mRNAs.

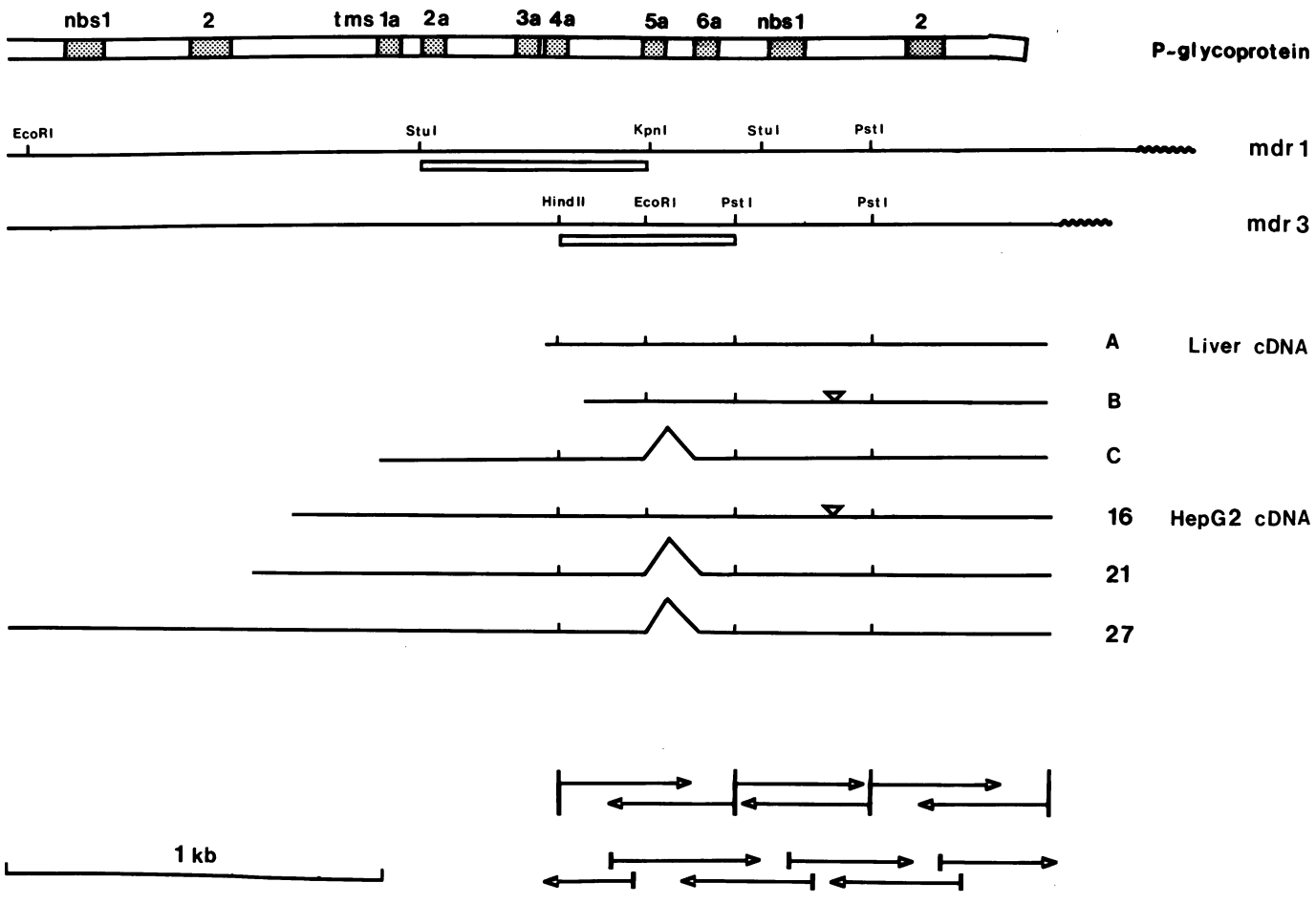

Fig. 1. Restriction maps of the P-glycoprotein-encoding cDNAs. The protein's C-terminal half is shown above with the positions of putative transmembrane segments (tms) and nucleotide binding sites (nbs). The relevant *mdr1* and *mdr3* restriction sites are shown in the physical map with the gene-specific probes indicated by rectangular boxes. Several individual *mdr3* cDNAs are aligned beneath the compiled map, illustrating insertions and deletions, which correspond to splice variants. The sequencing strategy with the positions of the oligonucleotide primers is indicated below the cDNA maps.

Results

P-glycoprotein cDNAs from human liver

We have isolated P-glycoprotein-related cDNAs from human liver cDNA libraries. Restriction site maps (Figure 1) revealed that three independent cDNAs were distinct from human *mdr1* (Chen *et al.*, 1986), although they hybridized strongly to our class 2 cDNA cp22, which corresponds to hamster P-glycoprotein (Van der Bliek *et al.*, 1986a). There were also map differences among the three human cDNAs, variants, A, B and C (Figure 1), which were investigated by sequence analysis. For this purpose appropriate fragments were subcloned in pEMBL vectors and sequenced with the dideoxy method. The sequence of the opposite strand was then completed by priming with custom-made oligonucleotides. The strategy is shown in Figure 1 and the sequence of cDNA variants A, B and C is compiled in Figure 2. Variant B differs from A by a 21-bp insertion and C has a 129-bp deletion as indicated. No other differences were found in the 1250 bp spanned by these three variants. Cloning artefacts can be excluded because we have also isolated these variants from a HepG2 cDNA library, i.e. one additional cDNA with a restriction map corresponding to variant A, two to variant B and two to variant C, besides 15 *mdr1* cDNAs. Because of length differences (only three HepG2 cDNAs are drawn in Figure 1) we conclude that these were all independently synthesized cDNAs. However, sequence analysis of the deletions in HepG2 cDNAs 21 and 27 showed these to be 12 bp longer at the 3′-end than in the liver cDNA variant C. The open reading frame is not interrupted by either the insertion in variant B or the deletions in C and in cDNAs 21 and 27 (Figure 2).

mdr3, A human P-glycoprotein gene

Translation of our cDNA sequences revealed extensive homology to, but also major discrepancies with, the C-terminal part of the recently published P-glycoprotein sequences. An alignment with corresponding segments from the mouse *mdr* (Gros *et al.*, 1986b), human *mdr1* (Chen *et al.*, 1986), and a partial Chinese hamster P-glycoprotein sequence (Gerlach *et al.*, 1986) is shown in Figure 3. With our cDNA variant A, insertions and deletions are not required for optimal alignment with the other P-glycoprotein sequences. The amino acid sequences are virtually identical in large blocks, which represent one of the putative ATPase domains of P-glycoproteins (Chen *et al.*, 1986; Gros *et al.*, 1986b; Gerlach *et al.*, 1986). The often clustered replacements outside these blocks do not affect the hydropathy plots with which the multiple membrane-spanning segments have been predicted (data not shown). The new gene is not a half-gene combined at the RNA level with the 5′ half of the *mdr1* gene by alternative splicing, because we have very recently isolated cDNAs corresponding to a complete gene that differs over its entire length from *mdr1* (C.Schneider and A.M.v.d.B., unpublished).

These results show that we have identified a human P-glycoprotein with a sequence distinct from *mdr1*. We have designated the corresponding gene *mdr3*, since it is the human counterpart

```
LeuThrLeuLeuLeuLeuAlaValValProIleIleAlaValSerGlyIleValGluMetLysLeuLeuAlaGlyAsnAlaLysArgAspLysLysGluLeuGluAlaAlaGlyLysIle
TTAACCCTATTGCTATTAGCAGTTGTTCCAATTATTGCTGTGTCAGGAATTGTTGAAATGAAATTGTTGGCTGGAAATGCCAAAAGAGATAAAAAAGAACTGGAAGCTGCTGGAAAGATT
HindII                                                                                                                  120

AlaThrGluAlaIleGluAsnIleArgThrValValSerLeuThrGlnGluArgLysPheGluSerMetTyrValGluLysLeuTyrGlyProTyrArgAsnSerValGlnLysAlaHis
GCAACAGAGGCAATAGAAAATATTAGGACAGTTGTGTCTTTGACCCAGGAAAGAAAATTTGAATCAATGTATGTTGAAAAATTGTATGGACCTTACAGGAATTCTGTGCAGAAGGCACAC
                                                                                                   EcoRI                240
                                                                  delC
IleTyrGlyIleThrPheSerIleSerGlnAlaPheMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuIleValAsnGlyHisMetArgPheArgAspValIleLeu
ATCTATGGAATTACTTTTAGTATCTCACAAGCATTTATGTATTTTTCCTATGCCGGTTGTTTTCGATTTGGTGCATATCTCATTGTGAATGGACATATGCGCTTCAGAGATGTTATTCTG
                                                                                                                        360

ValPheSerAlaIleValPheGlyAlaValaLALeuGlyHisAlaSerSerPheAlaProAspTyrAlaLysAlaLysLeuSerAlaAlaHisLeuPheMetLeuPheGluArgGlnPro
GTGTTTTCTGCAATTGTATTTGGTGCAGTGGCTCTAGGACATGCCAGTTCATTTGCTCCAGACTATGCTAAAGCTAAGCTGTCTGCAGCCCACTTATTCATGCTGTTTGAAAGACAACCT
                                                                                                                        480

LeuIleAspSerTyrSerGluGluGlyLeuLysProAspLysPheGluGlyAsnIleThrPheAsnGluValValPheAsnTyrProThrArgAlaAsnValProValLeuGlnGlyLeu
CTGATTGACAGCTACAGTGAAGAGGGGCTGAAGCCTGATAAATTTGAAGGAAATATAACATTTAATGAAGTCGTGTTCAACTATCCCACCCGAGCAAACGTGCCAGTGCTTCAGGGGCTG
                                                                                                                        600

SerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlyThrValPheVal
AGCCTGGAGGTGAAGAAAGGCCAGACACTAGCCCTGGTGGGCAGCAGTGGCTGTGGGAAGAGCACGGTGGTCCAGCTCCTGGAGCGGTTCTACGACCCCTTGGCGGGGACAGTGTTTGTG
                                                                                                                        720
   insB
AspPheGlyPheGlnLeuLeuAspGlyGlnGluAlaLysLysLeuAsnValGlnTrpLeuArgAlaGlnLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGlu
GACTTTGGTTTTCAGCTTCTCGATGGTCAAGAAGCAAAGAAACTCAATGTCCAGTGGCTCAGAGCTCAACTCGGAATCGTGTCTCAGGAGCCTATCCTATTTGACTGCAGCATTGCCGAG
                                                                                                                        840

AsnIleAlaTyrGlyAspAsnSerArgValValSerGlnAspGluIleValSerAlaAlaLysAlaAlaAsnIleHisProPheIleGluThrLeuProHisLysTyrGluThrArgVal
AATATTGCCTATGGAGACAACAGCCGGGTTGTATCACAGGATGAAATTGTGAGTGCAGCCAAAGCTGCCAACATACATCCTTTCATCGAGACGTTACCCCACAAATATGAAACAAGAGTG
                                                                                                                        960

GlyAspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuIleArgGlnProGlnIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGlu
GGAGATAAGGGGACTCAGCTCTCAGGAGGTCAAAAACAGAGGATTGCTATTGCCCGAGCCCTCATCAGACAACCTCAAATCCTCCTGTTGGATGAAGCTACATCAGCTCTGGATACTGAA
                                                                                                                       1020

SerGluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGly
AGTGAAAAGGTTGTCCAAGAAGCCCTGGACAAAGCCAGAGAAGGCCGCACCTGCATTGTGATTGCTCACCGCCTGTCCACCATCCAGAATGCAGACTTAATAGTGGTGTTTCAGAATGGG
                                                                                                                       1240

ArgValLysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyThrGlnAsnLeu***
AGAGTCAAGGAGCATGGCACGCATCAGCAGCTGCTGGCACAGAAAGGCATCTATTTTTCAATGGTCAGTGTCCAGGCTGGGACACAGAACTTATGAACTTTTGCTACAGTATATTTTAAA
                                                                                                                       1360

AATAAATTCAAATTATTCTACCCATTTTAAAAAAAAAAAA
```

Fig. 2. Nucleotide sequence of *mdr3* cDNAs and deduced amino acid sequence. The sequence presented starts at the *Hin*dII site indicated in Figure 1 and ends at the cloned poly(A) tail. The polyadenylation signal AATAAA is underlined. The deletion in variant C (43 amino acids) and its extended version in the HepG2 cDNAs 21 and 27 (47 amino acids) are boxed, as well as the seven amino acid acids inserted in variant B.

```
                                                                                  tm4a
mdr3                                                                  LTLLLLAVVPIIAVSGIVEMKLLAGNAKRDKKELEAAGKIATEAIENIRTVVSLTQERKFESMYVEKLYGP
mdr1                                                                  LTLLLLAIVPIIAIAGVVEMKMLSGQALKDKKELEGAGKIATEAIENFRTVVSLTQEQKFEHMYAQSLQVP
mouse                                                                 LTLLLVVIIPLIVLGGIIEMKLLSGQALKDKKQLEISGKIATEAIENFRTIVSLTREQKFETMYAQSLQVP

                                   delC
                tm5a                                                  tm6a
mdr3    YRNSVQKAHIYGITFSISQAFMYFSYAGCFRFGAYLIVNGHMRFRDVILVFSAIVFGAVALGHASSFAPDYAKAKLSAAHLFMLFERQPLIDSYSEEGLKPDKFEGNITFNEVVFNYPTR
mdr1    YRNSLRKAHIFGITFSFTQAMMYFSYAGCFRFGAYLVAHKLMSFEDVLLVFSAVVFGAMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPLIDSYSTEGLMPNTLEGNVTFVEVVFNYPTR
mouse   YRNAMKKAHVFGITFSFTQAMMYFSYAACFRFGAYLVAQQLMTFENVMLVFSAVVFGAMAAGNTSSFAPDYAKAKVSASHIIRIIEKTPEIDSYSTEGLKPTLLEGNVKFNGVQFNYPTR

                nbs1                                  insB
mdr3    ANVPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAGTVFVDFGFQLLDGQEAKKLNVQWLRAQLGIVSQEPILFDCSIAENIAYGDNSRVVSQDEIVSAAKAANIHPFIET
mdr1    PDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPLAGKV-------LLDGKEIKRLNVQWLRAHLGIVSQEPILFDCSIAENIAYGDNSRVVSQEEIVRAAKEANIHAFIES
mouse   PNIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPMAGSV-------FLDGKEIKQLNVQWLRAHLGIVSQEPILFDCSIAENIAYGDNSRAVSHEEIVRAAKEANIHQFIDS
hamster PDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPMAGTV-------FLDGKEIKQLNVQWLRAHLGIVSQEPILFDCSIAENIAYGDNSRVVSQDEIERAAKEANIHQFIES

                                nbs2
mdr3    LPHKYETRVGDKGTQLSGGQKQRIAIARALIRQPQILLLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSVQAGTQNL
mdr1    LPNKYSTKVGDKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQLLAQKGIYFSMVSVQAGTKRQ
mouse   LPDKYNTRVGDKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVIENGKVKEHGTHQQLLAQKGIYFSMV--QAGAKRS
hamster LPDKYNTRVGDKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVIQNGKVKEHGTHQQLLAQKGIYFSMV--QAGAKRL
```

Fig. 3. A comparison of the published P-glycoprotein sequences. The C-terminal amino acid sequences of *mdr3*, human *mdr1* (Chen *et al.*, 1986), mouse *mdr* (Gros *et al.*, 1986b) and hamster P-glycoprotein (Gerlach *et al.*, 1986) were aligned with a gap crossing the insert in variant B of *mdr3*. Boxes delineate putative transmembrane segments, the nucleotide binding site homology and the alternative splice products of *mdr3*: an insert in variant B, a deletion in variant C and its extension in the HepG2 cDNAs 21 and 27. Amino acids that deviate from the other sequences are highlighted by an asterisk.

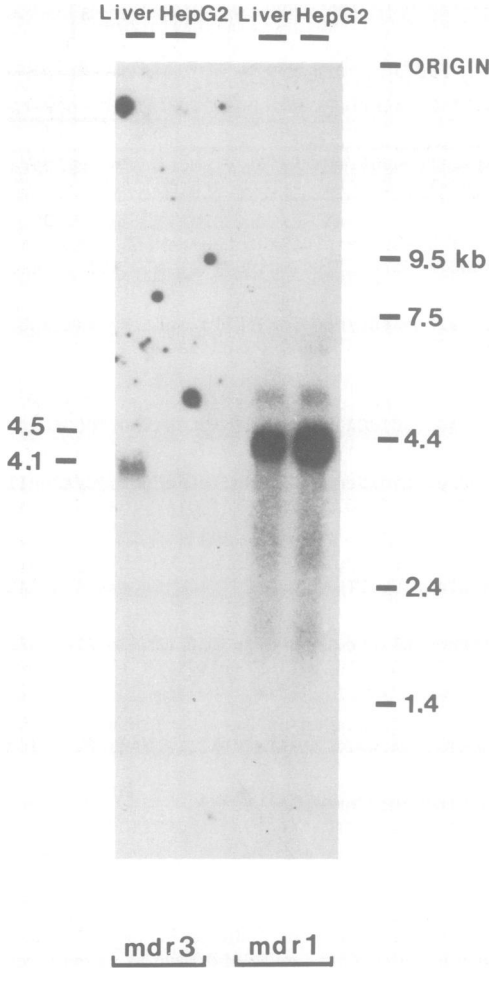

Fig. 4. Hybridization of *mdr3* and *mdr1* RNAs from human liver and HepG2 cells with gene-specific probes. RNA was size-fractionated by agarose gel electrophoresis, transferred to nitrocellulose and hybridized to the gene-specific *mdr1* and *mdr3* probes, boxed in Figure 1. See Materials and methods for details.

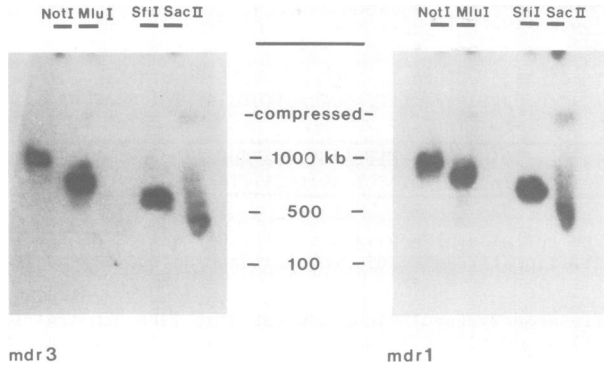

Fig. 5. Hybridization of large DNA fragments size-fractionated by pulsed field gradient gel electrophoresis with gene-specific probes shows linkage of the *mdr1* and *mdr3* genes. DNA was cut with the restriction enzymes indicated, size-fractionated, transferred to a nitrocellulose filter and hybridized with the gene-specific *mdr1* or *mdr3* probe, boxed in Figure 1, in two parallel experiments. The filters were stripped of the probes after autoradiography and a subsequent hybridization in which the probes were switched confirmed identity of the fragments (data not shown). See Materials and methods for details.

Table I. The homology between 3′ parts of mammalian P-glycoprotein cDNAs

Position	human *mdr1*/ mouse *mdr2*	human *mdr1*/ human *mdr3*	mouse *mdr2*/ human *mdr3*
45–285 (tms. 5a)	70	60	46
286–525 (tms. 6a)	71	51	54
526–765 (nbs. 1a)	80	78	80
766–1005 (nbs. 2a)	65	50	41
1006–1245 (C-term.)	75	89	69

For clarity only the third base pair of each triplet was used in the calculation, thus avoiding constraint at the amino acid level and enhancing the divergence. The percentage identity between human *mdr1* (Chen *et al.*, 1986), human *mdr3* and mouse *mdr* (Gros *et al.*, 1986b) is given for 240-bp segments with the sequence numbered as in Figure 2. The transmembrane segments (tms) and nucleotide binding site (nbs) homologous sequences represent hallmarks of these segments. We have assumed that the mouse sequence represents *mdr2* (see Results).

of the third P-glycoprotein gene in the hamster multidrug-resistance domain identified through the sequence of genomic DNA (V.Ling, personal communication). The *mdr3* gene is the homologue of what was formerly called gene class 2a of the hamster *mdr* domain (De Bruijn *et al.*, 1986).

mdr3 transcripts

We have examined the human liver and HepG2 transcripts using gene-specific probes corresponding to *mdr1* and to our newly identified P-glycoprotein gene (*mdr3*). The sequence of cDNA fragments encoding part of the transmembrane segments (boxes in Figure 1) is sufficiently divergent for discrimination between these two genes. The RNA blot analysis in Figure 4 shows that the major transcript in liver corresponding to *mdr3* is distinctly shorter than that of *mdr1* (4100 versus 4500 nucleotides). The length of the 3′-untranslated sequences can account for this difference (Figure 1). Our strain of HepG2 cells had selectively lost detectable *mdr3* expression, even though *mdr3* was obviously expressed at substantial levels in the HepG2 strain from which the cDNA library was made.

Linkage of mdr1 and mdr3

We have tested linkage of *mdr1* and *mdr3* in the human genome by pulsed field gradient electrophoresis of large DNA fragments. With the gene-specific hybridization probes and DNA from the drug-sensitive human cell line COLO 320 DM we can detect linkage of the corresponding genes on *Not*I, *Mlu*I, *Sfi*I and *Sac*II fragments in the range of 500–1000 kb (Figure 5). We have also detected co-amplification of *mdr1* and *mdr3* in human MDR cell lines (unpublished experiments). This suggests a genomic organization similar to that in hamster where at least two, but most likely three, P-glycoprotein genes are physically linked and co-amplified (Van der Bliek *et al.*, 1986a; De Bruijn *et al.*, 1986).

The relationship between P-glycoprotein genes

Table I presents a comparison between the human *mdr3* sequence and the corresponding nucleotide sequences of the human *mdr1* and mouse *mdr* genes. The mouse *mdr* sequence has higher homology to hamster *pgp2*, than to hamster *pgp1*, human *mdr1* or human *mdr3* (V.Ling, personal communication) and therefore represents *mdr2* in this comparison. Both overall and individual segments of the mouse *mdr* sequences have a higher homology to human *mdr1* than to human *mdr3* (Table I). The relationship is even more clear-cut in the sequence of the 200-bp linker between the two halves of the P-glycoproteins, where there is only 41% homology between human *mdr1* and *mdr3*, 27% homology between human *mdr3* and mouse *mdr*, but 71% homology be-

tween human *mdr1* and mouse *mdr* (unpublished results). It is therefore remarkable that the homology between human *mdr1* and *mdr3* sequences in the carboxy-terminal segment is even higher than that between human *mdr1* and mouse *mdr* (Table I). Indeed the sequence is completely identical in stretches as long as 120 bp, while there are 11 substitutions in the same part of the mouse *mdr* sequence. The frequency of third base pair substitutions in the nucleotide binding segments is also much lower than in the transmembrane segments (Table I; to position 525).

These results indicate that the transfer of gene segments between the human *mdr* genes has erased part of the differences between the C-terminal segments of *mdr1* and *mdr3* that have arisen since primates and rodents diverged.

Alternative splicing within mdr3

The complete identity of the *mdr3* cDNAs, except for an insertion in variant B and a deletion in C together with the corresponding HepG2 cDNAs, provides evidence that they are derived from the same gene. As already three different cDNA variants have been obtained from a single liver, the differences cannot be attributed to allelic variations in exons. We also see only a single *mdr3* gene in genomic blots probed with *mdr3*-specific probes (results not shown). The boundaries of the insertion and of the larger deletion coincide with exon/intron boundaries in hamster P-glycoprotein genes (W.Ng and V.Ling, personal communication). The cDNA variations must therefore be the result of alternative splicing events, a possibility first raised by I.Roninson (see Chen *et al.*, 1986).

Variant B has an insertion of seven amino acids between the segments proposed to form a nucleotide binding site (Figure 2). The insert has three phenylalanines, which is an unusually high incidence for this part of the protein. It is tempting to speculate that the aromatic drugs associated with multidrug resistance could interact with these phenylalanines. This speculation remains to be tested.

The deletion of 43 amino acids in variant C (Figure 2) covers the predicted transmembrane segment 5a and this is extended into segment 6a in the HepG2 cDNAs 21 and 27. A 5′-overhanging arginine triplet is maintained at the splice junction of both deletions. Preliminary sequence data from the upstream area do not reveal other alternative exons, which could compensate for the absent transmembrane segment (unpublished experiments). The odd number of transmembrane segments in variant C might result in a non-functional P-glycoprotein, with one of the flanking ATPases improperly oriented. It seems more likely, however, that a flanking hydrophobic segment is prevented from traversing the membrane in this variant. In line with this explanation, the extended deletion in HepG2 cDNAs 21 and 27 truncates the hydrophobic segment 6a and places an arginine at 20 amino acids from the next charged residue, an aspartate, which suggests that the hydrophobic segment 6a protrudes only partly into the membrane. This may represent a break from the general rule for transport proteins, i.e. two sets of six transmembrane segments (Henderson and Maiden, 1987), but transfection experiments with full length cDNAs of these variants will be needed to determine whether these cDNAs specify functional P-glycoproteins.

Discussion

How cells adapt their spectrum of cross-resistance to the selective drug remains unknown, although the elements that may contribute now seem largely identified. The P-glycoprotein structure provides the cornerstone: an energy-dependent pump analogous to bacterial transport systems (Gros *et al.*, 1986b; Gerlach *et al.*, 1986; Chen *et al.*, 1986). Its overproduction is consistent with (De Bruijn *et al.*, 1986) and also sufficient for (Gros *et al.*, 1986a; Ueda *et al.*, 1987) multidrug resistance. However, a single overproduced protein cannot account for the variable cross-resistance patterns of highly resistant cell lines. Three other elements could be involved: (i) differential expression of two or possibly three related P-glycoprotein genes, each with a different spectrum of drug affinity; (ii) production of multiple P-glycoproteins, differing in drug transport, from a single P-glycoprotein gene by alternative splicing, a possibility first raised by I.Roninson (see Chen *et al.*, 1986); (iii) altered expression of unrelated proteins. It now seems unlikely that the genes co-amplified with P-glycoprotein genes in hamster (Van der Bliek *et al.*, 1986a,b; De Bruijn *et al.*, 1986), mouse (Stahl *et al.*, 1987) and human (unpublished results) MDR cell lines have a major role to play (see De Bruijn *et al.*, 1986). However, mutations in genes not known to be linked to the P-glycoprotein gene cluster could affect the MDR phenotype. An example is the mutant described by Pommier *et al.* (1986) that appears to combine a low level of classical MDR with an altered topoisomerase II. Increased levels of a glutathione transferase have also been reported (Cowan *et al.*, 1986).

Multiplicity of P-glycoprotein genes was thus far mainly inferred from the differential amplification of genomic DNA hybridizing with P-glycoprotein encoding cDNAs in Chinese hamster (De Bruijn *et al.*, 1986; Scotto *et al.*, 1986), in mouse (Stahl *et al.*, 1987) and in human cell lines (Roninson *et al.*, 1986). The sequences of human *mdr1* and mouse *mdr* correspond to hamster *mdr1* and *mdr2*, respectively (V.Ling, personal communication). With our cDNAs we have now shown that an additional P-glycoprotein is encoded by the *mdr3* gene and is expressed in human liver. The *mdr3* gene corresponds to what was formerly called gene class 2a in hamsters (De Bruijn *et al.*, 1986), and its existence, initially deduced from differential amplification, has now been established by our cDNA sequence combined with that of genomic DNA from hamster (W.Ng and V.Ling, personal communication). Roninson *et al.* (1986) have inferred the existence of a second P-glycoprotein in man from hybridization of a DNA probe with two genomic regions. Transcripts of this gene were not detected even in MDR cell lines in which the gene was amplified. It remains unclear whether this gene is the same as our *mdr3* or the homologue of *mdr2*. It is possible, however, that a gene corresponding to the hamster *mdr2* gene does not exist in man. Southern blots hybridized with cDNA probes from the most conserved parts of the hamster P-glycoprotein genes always show more bands with rodent DNAs than with human DNA (Van der Bliek *et al.*, 1986a). Our attempts to detect a counterpart of the hamster *mdr2* gene in human liver cDNA or by differential hybridization to blots of human genomic DNA have failed thus far (unpublished). A genomic map of the *mdr* area should settle the issue.

Whether the P-glycoproteins encoded by *mdr1* and *mdr3* differ in the spectrum of drugs transported remains to be tested. It is of interest, however, that the putative transmembrane segments are very different, whereas the ATP-binding sites are not. This suggests that gene conversion events continue to erase differences between those parts of the genes that are not functionally different. Hence the differences between the transmembrane parts should be actively selected for, possibly because they determine useful differences in transport properties.

From the cDNAs with either an insert (B) or a deletion (C) in frame with the predominant message (A) we conclude that alternative splicing can vary the P-glycoprotein structure, and may be superimposed on the variation resulting from differen-

tially expressed homologous genes. If alternative splicing modulates the cross-resistance pattern, this could require alterations secondary to the P-glycoprotein overproduction in multidrug-resistant cells. For instance, *trans*-acting factors that govern splice site utilization might be affected, but also preferential export from the nucleus and RNA degradation. On the other hand, simply overexpressing the P-glycoprotein genes might influence the levels of alternatively spliced products by changing rate-limiting steps in RNA processing.

The circumstantial evidence assembled by us and by Ling and co-workers (personal communication) indicates that mammalian cells may be able to produce three different P-glycoproteins, at least one of which can be further diversified by alternative splicing. The overall level of *mdr* mRNA varies from tissue to tissue both in human (Fojo *et al.*, 1987) and in hamster (F.B., unpublished results) tissues and expression can be induced by partial hepatectomy or hepatocarcinogenesis in rats (Thorgeirsson *et al.*, 1987). Differential expression of *mdr1* and *mdr3* is possible, as we have found a HepG2 human liver cell line that has lost expression of *mdr3* but not of *mdr1*. One can see why the inducible production of a large array of P-glycoproteins could be advantageous in the removal of xenobiotics from cells, but not why differential expression would be required in different tissues and why the P-glycoprotein mRNA level should be high in adrenals and low in ovary (Fojo *et al.*, 1987). A comparison of the levels of hydrophobic metabolites in MDR cells and their drug-sensitive parents may provide a clue to the nature of the metabolites normally transported by P-glycoproteins.

Materials and methods

The isolation and mapping of P-glycoprotein encoding cDNAs

P-glycoprotein encoding cDNAs were obtained by screening a human liver cDNA library (kindly provided by Dr R.Cortese, EMBL, Heidelberg) cloned in the *Pvu*II site of a pAT153 derived vector (Twig and Sheratt, 1980). The hybridization probe was the 1.3-kb cDNA cp22 encompassing the C-terminal nucleotide binding site of the hamster *mdr1* gene (Van der Bliek *et al.*, 1986a). The insert fragment was separated on LMP-agarose (Bethesda Research Laboratories, Inc., Gaithersburg, MD) and labelled by nick-translation (Rigby *et al.*, 1977). Hybridization and autoradiography was as described previously (Van der Bliek *et al.*, 1986a) with a 30 min final wash in 1 × SSC at 65°C (1 × SSC = 150 mM NaCl, 15 mM Na-citrate pH 7.0). Plasmid DNA was isolated and characterized by restriction site mapping using standard techniques (Maniatis *et al.*, 1982). A second cDNA library (kindly provided by Dr M.McPhaul, Stanford University Medical Center, Stanford, CA) was derived from HepG2 RNA and cloned in the pcD vector (Okayama and Berg, 1983). This was screened with a human P-glycoprotein probe, the insert fragment of the variant C (Figure 1) isolated from the first library. The restriction site maps were verified by DNA blotting (Southern, 1975) and hybridization with the appropriate fragments under the previously described conditions.

Sequence analysis of P-glycoprotein encoding cDNAs

The sequence of the fragments in Figure 1 was determined with the dideoxy method (Sanger *et al.*, 1977) using *Hin*dII–*Pst*I, *Pst*I–*Pst*I and *Pst*I–*Bam*HI subclones in pEMBL9 (Dente *et al.*, 1983) or in pEP40, a vector kindly provided by P.W. Laird with the 514-bp *Rsa*I fragment of pEMBL9, containing the F1-ori, cloned in the *Pvu*II site of pGEM4 (Promega Biotec, Inc., Madison, WI). The sequence was completed with oligonucleotides priming on clones with *Bam*HI–*Hin*dII inserts of the variants A, B and C. The oligonucleotides were made on an Applied Biosystems DNA synthesizer model 381A. The sequence of the HepG2 cDNAs was determined by double strand sequencing (Promega Biotec, Inc., Madison, WI) with the same oligonucleotides.

Pulsed field gradient gel electrophoresis

The electrophoresis conditions were as described by Johnson and Borst (1986) with a pulse duration of 90 s, run for 22 h at 18°C in a 1% agarose gel. In-gel restriction enzyme digests were as described (Van der Bliek *et al.*, 1986a) using the conditions recommended by the manufacturer (New England Biolabs, Inc., Beverly, MA) with 25 U *Not*I, *Mlu*I or *Sfi*I and 50 U *Sac*II/10 μg DNA in each of two consecutive 5-h incubations. The DNA was isolated by in-gel lysis (Van der Bliek *et al.*, 1986a) from the human drug-sensitive cell line COLO 320 DM cultured under the conditions described (Alitalo *et al.*, 1983). *Trypanosoma evansi* chromosomes (Gibson and Borst, 1986) were run alongside the digests as size markers. These had been calibrated separately with multimers of phage λ DNA. The DNA was blotted and hybridized as described above with, as gene specific hybridization probes, a 600-bp *Stu*I–*Kpn*I fragment of a HepG2 cDNA derived from *mdr1* and a 450-bp *Hin*dII–*Pst*I fragment of the HepG2 cDNA 16 derived from *mdr3*. We found that a final wash in 0.1 × SSC at 65°C was sufficient for complete specificity. Two parallel experiments were conducted, in which the filters were first hybridized to the respective gene-specific probes, autoradiographed, then stripped by extensive washes in 0.1% SDS at 65°C and rehybridized with the remaining probe.

RNA blot analysis

Total human liver RNA was a gift from Dr H.Pannekoek (Central Laboratory of the Blood Transfusion Service, Amsterdam) and HepG2 RNA was isolated by the LiCl–urea method (Auffray and Rougeon, 1980). Poly(A)$^+$ RNA was then isolated with oligo(dT)–cellulose size-fractionated on a 1% agarose–formaldehyde gel (Maniatis *et al.*, 1982), incubated in 20 × SSC and blotted on nitrocellulose (Southern, 1977). Approximately 10 μg poly(A)$^+$ RNA was loaded in each lane as judged by ethidium bromide staining. A commercially available RNA ladder was used as size marker (Bethesda Research Laboratories, Inc., Gaithersburg, MD). The hybridization conditions were as described previously with the gene specific probes adjusted to the same amount of DNA (3×10^8 d.p.m./ μg).

Acknowledgements

We thank Dr V.Ling (Toronto) for advice and unpublished information; Drs R.Cortese (EMBL, Heidelberg) and M.McPhaul (Stanford University Medical Center, Stanford) for providing us with cDNA libraries; Dr H.Pannekoek (Central Laboratory of the Blood Transfusion Service, Amsterdam) for a gift of human liver RNA; and our colleagues in the laboratory for their critical review of the manuscript. This project was supported in part by grant NKI 84-20 of the Queen Wilhelmina Fund. F.B. was supported by the collaborative project on molecular genetics of the University of Amsterdam and the Netherlands Cancer Institute.

References

Alitalo,K., Schwab,M., Lin,C.C., Varmus,H.E. and Bishop,J.E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1707–1711.
Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.
Chen,C., Chin,J.E., Ueda,K., Clark,D.P., Pastan,I., Gottesman,M.M. and Roninson,I.B. (1986) *Cell*, **47**, 381–389.
Cornwell,M.M., Safa,A.R., Felsted,R.L., Gottesman,M.M. and Pastan,I. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3847–3850.
Cornwell,M.M., Pastan,I. and Gottesman,M.M. (1987) *J. Biol. Chem.*, **262**, 2166–2170.
Cowan,K.H., Batist,G., Tulpule,A., Sinha,B.K. and Myers,C.E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9328–9332.
Dano,K. (1973) *Biochim. Biophys. Acta*, **323**, 466–483.
De Bruijn,M.H.L., Van der Bliek,A.M., Biedler,J.L. and Borst,P. (1986) *Mol. Cell. Biol.*, **6**, 4717–4722.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
Deuchars,K.L., Run-Pan,D., Naik,M., Evernden-Porelle,D., Kartner,N., Van der Bliek,A.M. and Ling,V. (1987) *Mol. Cell. Biol.*, **7**, 718–724.
Ferro-Luzzi Ames,G. (1986) *Cell*, **47**, 323–324.
Fojo,A.T., Ueda,K., Slamon,D.L., Poplack,D.G., Gottesman,M.M. and Pastan,I. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 265–269.
Gerlach,J.H., Endicott,J.A., Juranka,P.F., Henderson,G., Sarangi,F., Deuchars, K.L. and Ling,V. (1986) *Nature*, **324**, 485–489.
Gibson,W.C. and Borst,P. (1986) *Mol. Biochem. Parasitol.*, **18**, 127–140.
Gros,P., Ben-Neriah,Y., Croop,J. and Housman,D.E. (1986a) *Nature*, **323**, 728–731.
Gros,P., Croop,J. and Housman,D. (1986b) *Cell*, **47**, 371–380.
Henderson,P.J.F. and Maiden,M.C.J. (1987) *Trends Genet.*, **3**, 62–64.
Johnson,P.J. and Borst,P. (1986) *Gene*, **43**, 213–220.
Jongsma,A.P.M., Spengler,B.A., Van der Bliek,A.M., Borst,P. and Biedler,J.L. (1987) *Cancer Res.*, **47**, 2875–2878.
Juliano,R.L. and Ling,V. (1976) *Biochim. Biophys. Acta*, **455**, 152–162.
Kartner,N., Riordan,J.R. and Ling,V. (1983) *Science*, **221**, 1285–1288.
Kessel,D. (1986) *Cancer Surveys*, **5**, 100–127.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Okayama,H. and Berg,P. (1983) *Mol. Cell. Biol.*, **3**, 280–289.
Peterson,R.H.F., Meyers,M.B., Spengler,B.A. and Biedler,J.L. (1983) *Cancer Res.*, **43**, 222–228.

Pommier,Y., Schwartz,R.E., Zwelling,L.A., Kerrigan,D., Mattern,M.R., Carcosset,J.Y., Jacquemin-Sablon,A. and Kohn,K.W. (1986) *Cancer Res.*, **46**, 611–616.
Rigby,P.W.J., Dieckman,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.
Riordan,J.R. and Ling,V. (1985) *Pharmacol. Ther.*, **28**, 51–75.
Riordan,J.R., Deuchars,K., Kartner,N., Alon,N., Trent,J. and Ling,V. (1985) *Nature*, **316**, 817–819.
Robertson,S.M., Ling,V. and Stanners,C.P. (1984) *Mol. Cell. Biol.*, **4**, 500–506.
Roninson,I.B., Chin,J.E., Choi,K., Gros,P., Housman,D.E., Fojo,A., Shen, D.-w., Gottesman,M.M. and Pastan,I. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4538–4542.
Safa,A.R., Glover,C.J., Meyers,M.B., Biedler,J.L. and Felsted,R.L. (1986) *J. Biol. Chem.*, **261**, 6137–6140.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Scotto,K.W., Biedler,J.L. and Melera,P.W. (1986) *Science*, **232**, 751–759.
Shen,D.-w., Fojo,A., Roninson,I.B., Chin,J.E., Soffier,R., Pastan,I. and Gottesman,M.M. (1986) *Mol. Cell. Biol.*, **6**, 4039–4044.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Stahl,F., Martinsson,T., Dåhllof,B. and Levan,G. (1987) *Hereditas*, in press.
Sugimoto,Y. and Tsuruo,T. (1987) *Cancer Res.*, **47**, 2620–2625.
Thorgeirsson,S.S., Huber,B.E., Sorrell,S., Fojo,A., Pastan,I. and Gottesman, M.M. (1987) *Science*, **236**, 1120–1122.
Twig,A.J. and Sheratt,D. (1980) *Nature*, **283**, 216–218.
Ueda,K., Cardarelli,C., Gottesman,M.M. and Pastan,I. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3004–3008.
Van der Bliek,A.M., Van der Velde-Koerts,T., Ling,V. and Borst,P. (1986a) *Mol. Cell. Biol.*, **6**, 1671–1678.
Van der Bliek,A.M., Meyers,M.B., Biedler,J.L., Hes,E. and Borst,P. (1986b) *EMBO J.*, **5**, 3201–3208.
Willingham,M.C., Cornwell,M.M., Cardarelli,C.O., Gottesman,M.M. and Pastan,I. (1986) *Cancer Res.*, **46**, 5941–5946.

Received on July 31, 1987