Cell, Vol. 47, 381–389, November 7, 1986, Copyright © 1986 by Cell Press

Internal Duplication and Homology with Bacterial Transport Proteins in the *mdr*1 (P-Glycoprotein) Gene from Multidrug-Resistant Human Cells

Chang-jie Chen,* Janice E. Chin,* Kazumitsu Ueda,† Douglas P. Clark,†‡ Ira Pastan,† Michael M. Gottesman,† and Igor B. Roninson*
*Center for Genetics
University of Illinois College of Medicine at Chicago
Chicago, Illinois 60612
†Laboratory of Molecular Biology
National Cancer Institute
National Institutes of Health
Bethesda, Maryland 20892
‡Howard Hughes Medical Institute
National Institutes of Health
Bethesda, Maryland 20892

Summary

Resistance of tumor cells to multiple cytotoxic drugs is a major impediment to cancer chemotherapy. Multidrug resistance in human cells is determined by the *mdr*1 gene, encoding a high molecular weight membrane glycoprotein (P-glycoprotein). Complete primary structure of human P-glycoprotein has been determined from the cDNA sequence. The protein, 1280 amino acids long, consists of two homologous parts of approximately equal length. Each half of the protein includes a hydrophobic region with six predicted transmembrane segments and a hydrophilic region. The hydrophilic regions share homology with peripheral membrane components of bacterial active transport systems and include potential nucleotide-binding sites. These results are consistent with a function for P-glycoprotein as an energy-dependent efflux pump responsible for decreased drug accumulation in multidrug-resistant cells.

Introduction

Multidrug-resistant mammalian cells are characterized by cross-resistance to many lipophilic cytotoxic compounds, including various plant alkaloids and anti-tumor antibiotics. The widespread occurrence of multidrug resistance in tumor cells represents a major impediment to successful cancer chemotherapy, since it involves resistance to some of the most commonly used antineoplastic agents. Analysis of different multidrug-resistant cell lines has indicated that this phenomenon is due to decreased drug accumulation in the resistant cells (reviewed by Riordan and Ling, 1985). Studies of drug transport in different multidrug-resistant cell lines have suggested that decreased drug accumulation in the resistant cells results at least in part from an increased rate of drug efflux occurring by an energy-dependent mechanism, since drug accumulation is enhanced by metabolic inhibitors (Dano, 1973; Skovsgaard, 1978; Inaba et al., 1979; Fojo et al., 1985; Willingham et al., 1986). These studies have led to the concept of an efflux pump responsible for the removal of various lipophilic compounds from multidrug-resistant cells (Dano, 1973). Other mechanisms for multidrug resistance have been proposed, including decreased drug influx (Skovsgaard, 1978; Ling et al., 1983) and altered drug binding in the resistant cells (Beck et al., 1983).

The most common biochemical characteristic of multidrug-resistant cells is the increased expression of a membrane glycoprotein with a molecular weight of approximately 170 kd, termed P-glycoprotein (Juliano and Ling, 1976; Kartner et al., 1983). In some studies, high molecular weight glycoproteins appear as a heterogeneous group in multidrug-resistant cells (Peterson et al., 1983), but it is unknown whether this heterogeneity reflects different proteins or variations in the oligosaccharides attached to the same P-glycoprotein. The size of P-glycoprotein in the absence of N-glycosylation has been estimated as approximately 140 kd (Ling et al., 1983). Using cDNA clones for the hamster P-glycoprotein, several investigators have shown that the P-glycoprotein gene is amplified in multidrug-resistant cell lines, and that gene amplification is accompanied by increased expression of 4.5–5.0 kb P-glycoprotein mRNA. Differential amplification of DNA sequences hybridizing to P-glycoprotein clones has suggested that P-glycoproteins may be encoded by a multigene family (Riordan et al., 1985; van der Bliek et al., 1986; Scotto et al., 1986). The P-glycoprotein gene has been mapped to the human chromosome 7 (Trent et al., 1985). It has been speculated that P-glycoprotein may be directly involved in drug transport in multidrug-resistant cells, either as an efflux pump or by changing permeability of the lipid bilayer (Riordan and Ling, 1985). Studies using photoaffinity-labeled analogues of vinblastine have provided direct evidence for drug binding by P-glycoprotein (Cornwell et al., 1986; Safa et al., 1986).

We have previously used the in-gel DNA renaturation procedure (Roninson, 1983) to clone a Chinese hamster gene, designated *mdr*, which spans approximately 80 kb of DNA and is amplified in two independently derived multidrug-resistant Chinese hamster cell lines (Roninson et al., 1984; Gros et al., 1986a). Using the hamster *mdr* gene as a probe, we have cloned segments of two cross-hybridizing human genes that were amplified in multidrug-resistant derivatives of KB epidermoid carcinoma cells (Roninson et al., 1986). One of these genes, designated *mdr*1, is amplified and/or overexpressed in human cell lines of different origins, selected with different cytotoxic drugs (Shen et al., 1986a), as well as in certain normal and tumor tissues (A. Fojo, K. Ueda, D. J. Slamon, D. G. Poplack, H. R. Keiser, M. M. Gottesman, and I. Pastan, submitted). The human *mdr*1 gene is transferred and amplified in multidrug-resistant transfectants of mouse NIH 3T3 cells, transfected with DNA from multidrug-resistant human cells (Shen et al., 1986c). Recently Gros et al. (1986b) have demonstrated that the expression of a full-length mouse *mdr* cDNA clone confers a complete multidrug-resistant phenotype onto sensitive cells. Several properties of the *mdr*1 gene suggest that it is a

member of the P-glycoprotein gene family. These properties include the size of *mdr*1 mRNA (4.5 kb; Roninson et al., 1986; Shen et al., 1986a), its localization on chromosome 7 (Fojo et al., 1986), and amplification of *mdr*1 in the cell lines that have amplified the P-glycoprotein gene or overproduced P-glycoprotein (Roninson et al., 1984; Riordan et al., 1985; Shen et al., 1986a, 1986b). Identification of *mdr*1 as a human P-glycoprotein gene has been recently confirmed by cross-hybridization between P-glycoprotein and *mdr*1 cDNA clones (K. Ueda, M. M. Gottesman, I. Pastan, I. B. Roninson, V. Ling, and J. R. Riordan, unpublished).

In the present communication, we report the complete cDNA sequence of the human *mdr*1 gene. The protein encoded by the *mdr*1 gene consists of two approximately equal parts that are homologous to each other. Both parts contain a hydrophobic and a hydrophilic domain. Each hydrophobic domain includes six potential transmembrane segments, whereas the hydrophilic domains share sequence homology with peripheral membrane components of bacterial periplasmic transport systems and include potential nucleotide-binding sites. The predicted membrane orientation of the protein and homology with bacterial active transport proteins are consistent with the function of P-glycoprotein as an efflux pump responsible for decreased drug accumulation in multidrug-resistant cells.

Results and Discussion

*mdr*1 cDNA Sequence

Construction of the cDNA library from multidrug-resistant KB-C2.5 cells and isolation of *mdr*1 cDNA clones will be described in detail elsewhere (K. Ueda, D. P. Clark, C.-J. Chen, I. B. Roninson, M. M. Gottesman, and I. Pastan, submitted). Briefly, a cDNA library in the λgt11 phage vector (Young and Davis, 1983), prepared from multidrug-resistant KB-C2.5 cells (Akiyama et al., 1985; Shen et al., 1986b), was screened with a probe (pMDR1) containing a 0.8 kb PvuII fragment of the *mdr*1 genomic clone pHDR4.4 (Roninson et al., 1986). cDNA clones containing the longest inserts were designated λHDR69A, λHDR28, λHDR10, and λHDR5 (Figure 1). Restriction enzyme mapping has indicated considerable sequence heterogeneity among different clones hybridizing with pMDR1, even though all the clones contained the same 267 bp PvuII fragment, which hybridized to the genomic probe. Partial DNA sequencing has shown that clones λHDR10 and λHDR5, as well as several other clones (not shown), contain identical overlapping sequences, and therefore most likely represent the major species of *mdr*1 mRNA. The clones λHDR69A and λHDR28, on the other hand, appear to represent rare variants, which in the case of λHDR69A have resulted from aberrant RNA splicing (J. E. Chin, and I. B. Roninson, unpublished). The rest of the *mdr*1 cDNA sequence has been isolated by screening the library with a 638 bp terminal fragment of λHDR5 (indicated with a striped bar in Figure 1). The resulting clones λHDR103, λHDR104, and λHDR105 showed no divergence from each other or from clone λHDR5 either in their restriction

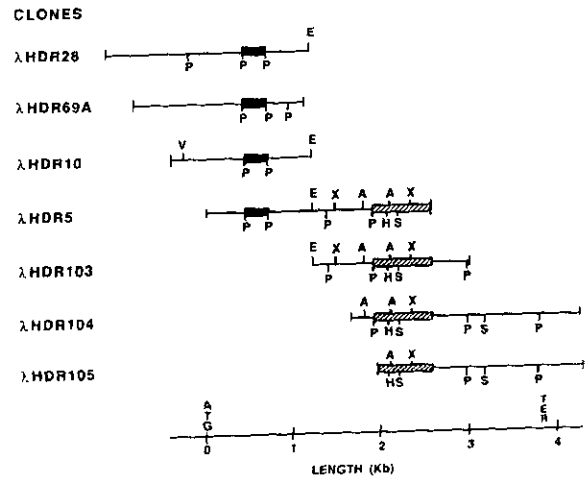

Figure 1. Restriction Maps of *mdr*1 cDNA Clones

Restriction sites: A, AccI; E, EcoRI; H, HindIII; P, PvuII; S, StuI; V, AvaI; X, XmnI. The solid bar indicates the region hybridizing to the genomic pMDR1 probe (Roninson et al., 1986); the striped bar corresponds to the fragment of clone λHDR5 that was used as a probe to isolate the clones containing 3′ terminal sequences. Positions of the translation initiation (ATG) and translation termination (TER) sites are indicated.

maps or in partial DNA sequences, except for a variant 12 bp sequence at the 5′ end of clone λHDR105, which could result from a cloning artifact. Clones λHDR10, λHDR5, and λHDR104 were sequenced in their entirety on both strands, providing a nearly full-length cDNA sequence for *mdr*1. It should be noted that since this sequence was determined from three different overlapping clones and not from a single cDNA molecule, there is a formal possibility that the 5′ and 3′ parts of this sequence have not necessarily been derived from the same mRNA.

The 4669 bp sequence of *mdr*1 cDNA is shown in Figure 2. The sequence contains a long open reading frame between positions −177 and 3840, where position 1 has been assigned to the first ATG codon in the cDNA sequence. This codon is preceded by a purine nucleotide at position −3, found at almost all functional initiation codons in eukaryotic mRNA (Kozak, 1984). The size of the protein encoded by the open reading frame is 1280 amino acids, with a calculated molecular weight of 141,475 daltons, which is in good agreement with the estimated size of the polypeptide moiety of P-glycoprotein (Ling et al., 1983).

To confirm that the cDNA clones were in fact derived from *mdr*1 and not from another member of the *mdr* gene family, the *mdr*1 genomic clone pHDR4.4 has been sequenced in the region that hybridized to the cDNA clones. Figure 3 shows that the genomic and cDNA clones contain two identical exon sequences, corresponding to residues 339–530 and 531–702 of the cDNA. The 808 bp PvuII fragment, corresponding to the pMDR1 clone used in earlier studies (Shen et al., 1986a) as a probe for *mdr*1 mRNA expression, includes 267 bp of exon sequences and a 541 bp intron.

```
                                                  -424   CCTACTCTATTCAGATATTCTCCAGATTCCTAAAGATTAGAGATCATTTCTCATTCTCCTAGGA   -361
GTACTCACTTCAGGAAGCAACCAGATAAAAGAGAGGTGCAACGGAAGCCAGAACATTCCTCCTGGAAATTCAACCTGTTTCGCAGTTTCTCGAGGAATCAGCATTCAGTCAATCCGGGCC   -241
                                                                  ***
GGGAGCAGTCATCTGTGGTGAGGCTGATTGGCTGGGCAGGAACAGCGCCGGGGCGTGGGCTGAGCACAGCGCTTCGCTCTCTTTGCCACAGGAAGCCTGAGCTCATTCGAGTAGCGGCTC   -121
TTCCAAGCTCAAAGAAGCAGAGGCCGCTGTTCGTTTCCTTTAGGTCTTTCCACTAAAGTCGGAGTATCTTCTTCCAAGATTTCACGTCTTGGTGGCCGTTCCAAGGAGCGCCAGGTCGGG   -1
1
ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAAGAAGAAGAACTTTTTTAAACTGAACAATAAAAGTGAAAAAGATAAGAAGGAAAAGAAACCAACTGTCAGTGTATTTTCAATGTTT   120
CGCTATTCAAATTGGCTTGACAAGTTGTATATGGTGGTGGGAACTTTGGCTGCCATCATCCATGGGGCTGGACTTCCTCTCATGATGCTGGTGTTTGGAGAAATGACAGATATCTTTGCA   240
AATGCAGGAAATTTAGAAGATCTGATGTCAAACATCACTAATAGAAGTGATATCAATGATACAGGGTTCTTCATGAATCTGGAGGAAGACATGACCAGGTATGCCTATTATTACAGTGGA   360
ATTGGTGCTGGGGTGCTGGTTGCTGCTTACATTCAGGTTTCATTTTGGTGCCTGGCAGCTGGAAGACAAATACACAAAATTAGAAAACAGTTTTTTCATGCTATAATGCGACAGGAGATA   480
GGCTGGTTTGATGTGCACGATGTTGGGGAGCTTAACACCCGACTTACAGATGATGTCTCTAAGATTAATGAAGTTATTGGTGACAAAATTGGAATGTTCTTTCAGTCAATGGCAACATTT   600
TTCACTGGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCTTGTGATTTTGGCCATCAGTCCTGTTCTTGGACTGTCAGCTGCTGTCTGGGCAAAGATACTATCTTCATTTACT   720
GATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTCTTGGCAGCAATTAGAACTGTGATTGCATTTGGAGGACAAAAGAAAGAACTTGAAAGGTACAACAAAAAT   840
TTAGAAGAAGCTAAAAGAATTGGGATAAAGAAAGCTATTACAGCCAATATTTCTATAGGTGCTGCTTTCCTGCTGATCTATGCATCTTATGCTCTGGCCTTCTGGTATGGGACCACCTTG   960
GTCCTCTCAGGGGAATATTCTATTGGACAAGTACTCACTGTATTCTTTTCTGTATTAATTGGGGCTTTTAGTGTTGGACAGGCATCTCCAAGCATTGAAGCATTTGCAAATGCAAGAGGA   1080
GCAGCTTATGAAATCTTCAAGATAATTGATAATAAGCCAAGTATTGACAGCTATTCGAAGAGTGGGCACAAACCAGATAATATTAAGGGAAATTTGGAATTCAGAAATGTTCACTTCAGT   1200
TACCCATCTCGAAAAGAAGTTAAGATCTTGAAGGGCCTGAACCTGAAGGTGCAGAGTGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGAGCACAACAGTCCAGCTGATG   1320
CAGAGGCTCTATGACCCCACAGAGGGGATGGTCAGTGTTGATGGACAGGATATTAGGACCATAAATGTAAGGTTTCTACGGGAAATCATTGGTGTGGTGAGTCAGGAACCTGTATTGTTT   1440
GCCACCACGATAGCTGAAAACATTCGCTATGGCCGTGAAAATGTCACCATGGATGAGATTGAGAAAGCTGTCAAGGAAGCCAATGCCTATGACTTTATCATGAAACTGCCTCATAAATTT   1560
GACACCCTGGTTGGACAGAGAGGGGCCCAGTTGAGTGGTGGGCAGAAGCAGAGGATCGCCATTGCACGTGCCCTGGTTCGCAACCCCAAGATCCTCCTGCTGGATGAGGCCACGTCAGCC   1680
TTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCTGGATAAGGCCAGAAAAGGTCGGACCACCATTGTGATAGCTCATCGTTTGTCTACAGTTCGTAATGCTGACGTCATCGCTGGT   1800
TTCGATGATGGAGTCATTGTGGAGAAAGGAAATCATGATGAACTCATGAAAGAGAAAGGCATTTACTTCAAACTTGTCACAATGCAGACAGCAGGAAATGAAGTTGAATTAGAAAATGCA   1920
GCTGATGAATCCAAAAGTGAAATTGATGCCTTGGAAATGTCTTCAAATGATTCAAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAGCCCAAGACAGA   2040
AAGCTTAGTACCAAAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCTTTTGGAGGATTATGAAGCTAAATTTAACTGAATGGCCTTATTTTGTTGTTGGTGTATTTTGTGCCATTATA   2160
AATGGAGGCCTGCAACCAGCATTTGCAATAATATTTTCAAAGATTATAGGGGTTTTTACAAGAATTGATGATCCTGAAACAAAACGACAGAATAGTAACTTGTTTTCACTATTGTTTCTA   2280
GCCCTTGGAATTATTTCTTTTATTACATTTTTCCTTCAGGGTTTCACATTTGGCAAAGCTGGAGAGATCCTCACCAAGCGGCTCCGATACATGGTTTTCCGATCCATGCTCAGACAGGAT   2400
GTGAGTTGGTTTGATGACCCTAAAAACACCACTGGAGCATTGACTACCAGGCTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATTACCCAGAATATA   2520
GCAAATCTTGGGACAGGAATAATTATATCCTTCATCTATGGTTGGCAACTAACACTGTTACTCTTAGCAATTGTACCCATCATTGCAATAGCAGGAGTTGTTGAAATGAAAATGTTGTCT   2640
GGACAAGCACTGAAAGATAAGAAAGAACTAGAAGGTGCTGGGAAGATCGCTACTGAAGCAATAGAAAACTTCCGAACCGTTGTTTCTTTGACTCAGGAGCAGAAGTTTGAACATATGTAT   2760
GCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTTGGAATTACATTTTCCTTCACCCAGGCAATGATGTATTTTTCCTATGCTGGATGTTTCCGGTTTGGA   2880
GCCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTTAGTATTTTCAGCTGTTGTCTTTGGTGCCATGGCCGTGGGGCAAGTCAGTTCATTTGCTCCTGACTATGCCAAA   3000
GCCAAAATATCAGCAGCCCACATCATCATGATCATTGAAAAAACCCCTTTGATTGACAGCTACAGCACGGAAGGCCTAATGCCGAACACATTGGAAGGAAATGTCACATTTGGTGAAGTT   3120
GTATTCAACTATCCCACCCGACCGGACATCCCAGTGCTTCAGGGACTGAGCCTGGAGGTGAAGAAGGGCCAGACGCTGGCTCTGGTGGGCAGCAGTGGCTGTGGGAAGAGCACAGTGGTC   3240
CAGCTCCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAAGTGCTGCTTGATGGCAAAGAAATAAAGCGACTGAATGTTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCCCAGGAGCCC   3360
ATCCTGTTTGACTGCAGCATTGCTGAGAACATTGCCTATGGAGACAACAGCCGGGTGGTGTCACAGGAAGAGATCGTGAGGGCAGCAAAGGAGGCCAACATACATGCCTTCATCGAGTCA   3480
CTGCCTAATAAATATAGCACTAAAGTAGGAGACAAAGGAACTCAGCTCTCTGGTGGCCAGAAACAACGCATTGCCATAGCTCGTGCCCTTGTTAGACAGCCTCATATTTTGCTTTTGGAT   3600
GAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAAGGTTGTCCAAGAAGCCCTGGACAAAGCCAGAGAAGGCCGCACCTGCATTGTGATTGCTCACCGCCTGTCCACCATCCAGAATGCA   3720
GACTTAATAGTGGTGTTCAGAATGGCAGAGTCAAGGAGCATGGCACGCATCAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACAAAGCGCCAG   3840
***
TGAACTCTGACTGTATGAGATGTTAAATACTTTTTAATATTTGTTTAGATATGACATTTATTCAAAGTTAAAAGCAAACACTTACAGAATTATGAAGAGGTATCTGTTTAACATTTCCTC   3960
AGTCAAGTTCAGAGTCTTCAGAGACTTCGTAATTAAAGGAACAGAGTGAGAGACATCATCAAGTGGAGAGAAATCATAGTTTAAACTGCATTATAAATTTTATAACAGAATTAAAGTAGA   4080
TTTTAAAAGATAAAATGTGTAATTTTGTTTATATTTCCCATTTGGACTGTAACTGACTGCCTTGCTAAAAGATTATAGAAGTAGCAAAAAGTATTGAAATGTTTGCATAAAGTGTCTAT   4200
AATAAACTAAACTTTCATGTGAAAAAAAAAAAAAAAAAAAAAAAA   4255
```

Figure 2. *mdr*1 cDNA Sequence

Asterisks indicate the termination codons delineating the longest open reading frame. The sequence is numbered from the first ATG codon. The complete cDNA sequence has been determined from clones λHDR10 (from −424 to 1176), λHDR5 (from 25 to 2555) and λHDR104 (from 1346 to 4255), as described in Experimental Procedures.

```
                Intron | exon                                                                                  PvuII
pHDR4.4   TTCTTTTTCAGGTATGCCTATTATTACAGTGGAATTGGTGCTGGGGTGCTGGTTGCTGCTTACATTCAGGTTTCATTTTGGTGCCTGGCAGCTGGAAGACAAATACACAAAATTAGAAAA
cDNA      GACATGACCAGGTATGCCTATTATTACAGTGGAATTGGTGCTGGGGTGCTGGTTGCTGCTTACATTCAGGTTTCATTTTGGTGCCTGGCAGCTGGAAGACAAATACACAAAATTAGAAAA
            330       340       350       360       370       380       390       400       410       420       430       440

                                                                                          exon | Intron
pHDR4.4   CAGTTTTTTCATGCTATAATGCGACAGGAGATAGGCTGGTTTGATGTGCACGATGTTGGGGAGCTTAACACCCGACTTACAGAGTAAGTATTTAGTTTTATGTTGAACTTGGGTGTCGTT
cDNA      CAGTTTTTTCATGCTATAATGCGACAGGAGATAGGCTGGTTTGATGTGCACGATGTTGGGGAGCTTAACACCCGACTTACAGA----------------------------------
            450       460       470       480       490       500       510       520       530

pHDR4.4   CTTATCCTTAGTAAAATGAAATAGATGTCATCACATCTGTTAGGAGGTGTTAATGTATCATTCAAAGGTACTTATGAGACAAAATTCCTTCTAAGCAGCAACAATGTCGTGTGCATCCTT
cDNA      ------------------------------------------------------------------------------------------------------------------------

pHDR4.4   TTGTTCCCAGTGCCTTGACAGGGTATGGGGGGACCTGCATGACTAGCATTAAATGAAGGACTGGGCTTTCCAGAATGAAGAAATCCTCTGAGAATGTGCAGTAGAGCAAAACAAGATACT
cDNA      ------------------------------------------------------------------------------------------------------------------------

pHDR4.4   TTCTGAGGAAATTTCTGAGCAATTTGAAATTCCTAGGTTGAATACTTCTTGTGTACACGATGTCCATTTCCTGGGGCCATGTGGCTATGGATTTTTGTTGTTAATGACAAATATCCTAGT
cDNA      ------------------------------------------------------------------------------------------------------------------------

pHDR4.4   AGAAACTTCTACCCTGCTAAATAAAACAAAGCATAGGCACAAAATACTCTAGCCATAAACTACCCTACACTCAAAACAGGCTTCACGAGAAAAGTTGATGTTTACAATTCTGACAATTAT
cDNA      ------------------------------------------------------------------------------------------------------------------------

                        Intron | exon
pHDR4.4   TTCTAACACTATCTGTTCTTTCAGTGATGTCTCTAAGATTAATGAAGTTATTGGTGACAAAATTGGAATGTTCTTTCAGTCAATGGCAACATTTTTCACTGGGTTTATAGTAGGATTTAC
cDNA      ----------------------TGATGTCTCTAAGATTAATGAAGTTATTGGTGACAAAATTGGAATGTTCTTTCAGTCAATGGCAACATTTTTCACTGGGTTTATAGTAGGATTTAC
                                540       550       560       570       580       590       600       610       620

                                                PvuII
                                                                                  exon | Intron
pHDR4.4   ACGTGGTTGGAAGCTAACCCTTGTGATTTTGGCCATCAGTCCTGTTCTTGGACTGTCAGCTGCTGTCTGGGCAAAGGTAGGTGAAGCCTGTGAATCCAGATTTTGAACTGCACCTTCTCC
cDNA      ACGTGGTTGGAAGCTAACCCTTGTGATTTTGGCCATCAGTCCTGTTCTTGGACTGTCAGCTGCTGTCTGGGCAAAGATACTATCTTCATTTACTGATAAAGAACTCTTAGCGTATGCAAA
            630       640       650       660       670       680       690       700       710       720       730       740
```

Figure 3. Sequence Comparison of the Genomic pHDR4.4 and cDNA Clones

The first intron/exon boundary is shown three residues 3′ from the start of sequence homology, on the assumption that the actual boundary occurs at the conventional AG/GT splice site. PvuII sites flanking the previously described pMDR1 subclone (Roninson et al., 1986) are indicated with arrows.

Internal Homology and Transmembrane Domains

Analysis of the primary structure of the *mdr*1 gene product reveals that this protein consists of two approximately equal parts sharing considerable amino acid sequence homology with each other. An alignment of the N-terminal (residues 1–637) and C-terminal (residues 638–1280) halves of the protein is shown in Figure 4. For an optimal alignment, two large gaps (11 and 19 residues) and four small gaps (1–2 residues) had to be introduced. Of the amino acids aligned, 43% (263 out of 613) are identical. Another 215 (35%) represent pairs of functionally similar residues, defined as having a "relative rate of acceptance of point mutations" (Dayhoff et al., 1972) larger than 20. At the DNA level, the homology is 52%, using the same

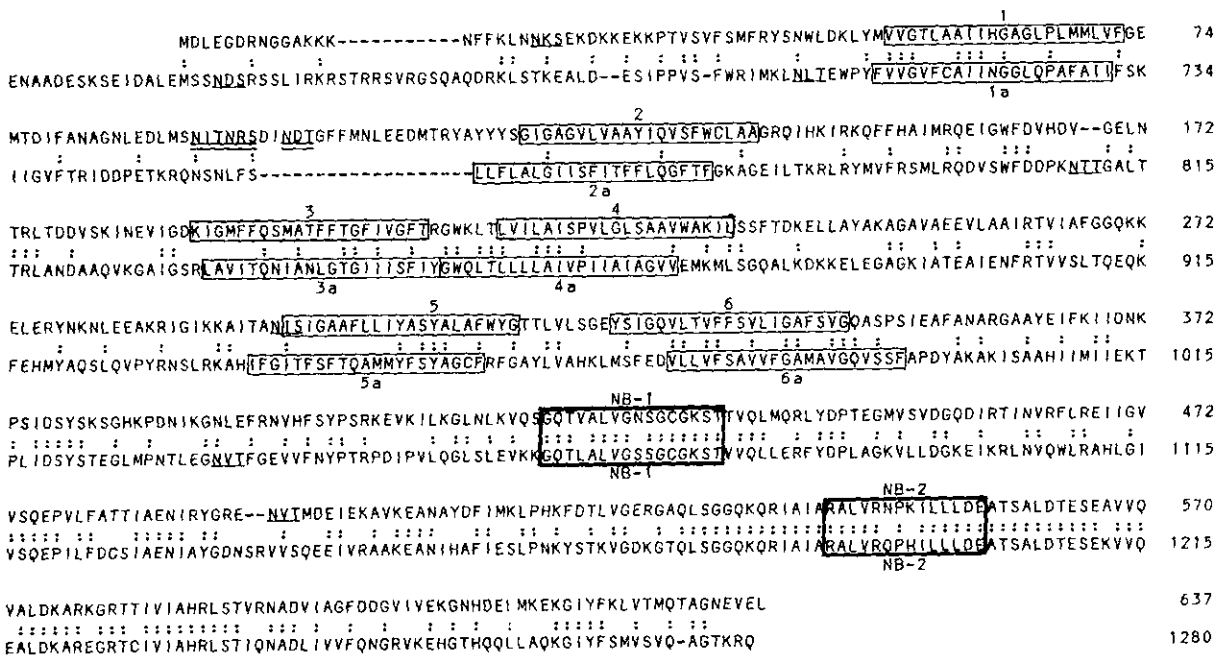

Figure 4. Primary Structure of the mdr1 Gene Product and Alignment of the N-Terminal and C-Terminal Halves

The standard single-letter amino acid code is used. Colons indicate identical residues. Potential N-glycosylation sites (Asn-X-Ser/Thr) are underlined; the sites localized in the predicted extracytoplasmic region (see Figure 7) are underlined twice. Potential transmembrane segments (1–6 and 1a–6a), predicted by the algorithm of Eisenberg et al. (1984), are enclosed in thin boxes. Potential nucleotide-binding sites (NB-1 and NB-2) are enclosed in thick boxes.

alignment as for the protein sequences. The level of homology is considerably higher in the C-terminal than in the N-terminal parts of the aligned sequences. Two pairs of segments have a particularly high homology level. These include residues 529–591 and 1174–1236 (56 identical residues out of 62), as well as residues 419–446 and 1061–1088 (21 out of 27 identical residues). These segments contain the potential nucleotide-binding sites, NB-1 and NB-2, which are described below. The internal homology in the *mdr*1 cDNA sequence suggests that the *mdr*1 gene has likely arisen as a consequence of an internal duplication.

The similarity of the N-terminal and C-terminal halves of the protein is also obvious from their hydropathy plots (Kyte and Doolittle, 1982; Figure 5). Based on the hydropathy plots, each half of the protein can be subdivided into a short hydrophilic region at the N-terminus, a long hydrophobic region and a long relatively hydrophilic region near the C-terminus. These regions are referred to as the N-terminus (residues 1–48), hydrophobic region 1 (residues 49–350), and hydrophilic region 1 (residues 351–637) in the N-terminal half, and as the linker (residues 638–708), hydrophobic region 2 (residues 709–993), and hydrophilic region 2 (residues 994–1280) in the C-terminal half. The protein sequence of *mdr*1 was further analyzed for the presence of potential transmembrane domains using the algorithm of Eisenberg et al. (1984), which can distinguish membrane-spanning regions from hydrophobic cores of globular proteins. This analysis has predicted six 21 amino acid long transmembrane domains within each hydrophobic region of the *mdr*1 protein. The transmembrane segments are designated 1–6 and 1a–6a in Figure 4 and Figure 5. The imprecise alignment of the transmembrane segments between the N-terminal and C-terminal halves (Figure 4) suggests that this assay does not provide exact delineation of the membrane-spanning domains. The assignment of the termini of transmembrane segments in our model should therefore be viewed only as approximate. The average hydropathy of these segments according to the scale of Eisenberg et al. (1984) ranges from 0.55 (segment 3) to 0.81 (segments 2a and 4a), well above 0.42, the minimum value of hydropathy for transmembrane segments. The predicted transmembrane localization of the *mdr*1 gene product is consistent with its identification as the human P-glycoprotein. Analysis of the amphiphilicity of the transmembrane segments by a hydrophobic moment plot (Eisenberg et al., 1984) has indicated that most transmembrane segments of *mdr*1 are more amphiphilic than the corresponding segments in proteins that span the membrane only once (data not shown). However, similar values of amphiphilicity have been observed in other proteins containing multiple transmembrane segments, including some channel-forming proteins (Eisenberg et al., 1984).

Homology to Bacterial Transport Proteins

The National Biomedical Research Foundation (Georgetown University) database was used to search for homology between *mdr*1 and other protein sequences. The sequences of hydrophilic regions 1 and 2 were found to

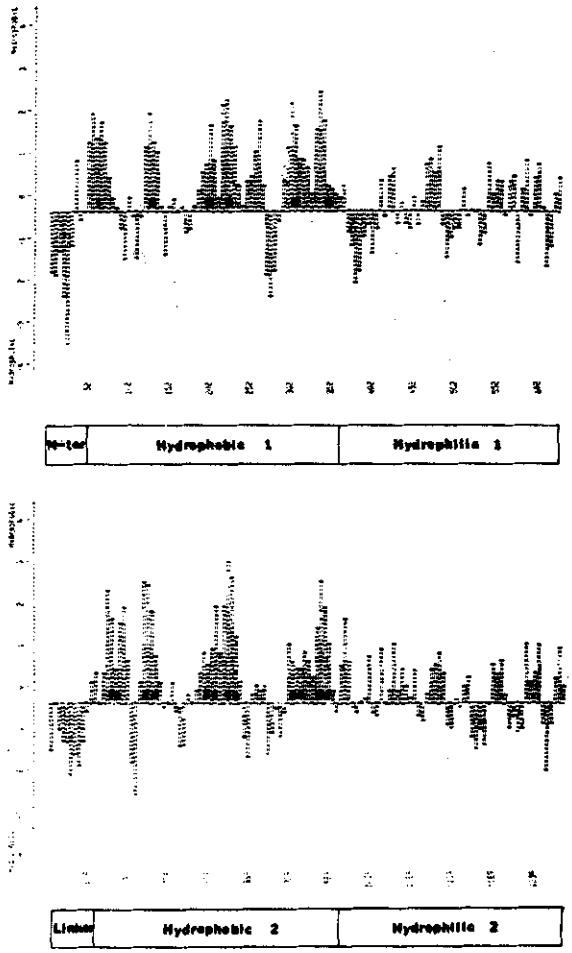

Figure 5. Hydropathy Plots of the N-Terminal and C-Terminal Halves of Human P-Glycoprotein

The plots were derived using the algorithm of Kyte and Doolittle (1982). Each point corresponds to an average of the hydropathic indices of 9 consecutive amino acids. Positions of potential transmembrane segments (1–6; 1a–6a), identified as described by Eisenberg et al. (1984), are indicated. The joined boxes depict the hydrophilic and hydrophobic domains in each half of the protein (see text for detailed description).

share significant homology with bacterial proteins *his*P of Salmonella typhimurium (Higgins et al., 1982) and *mal*K of E. coli (Gilson et al., 1982). These proteins are peripheral membrane components of periplasmic active transport systems for histidine, lysine, arginine, and ornithine (*his*P; Higgins et al., 1982) or maltose and maltodextrins (*mal*K; Gilson et al., 1982). Homology with *mdr*1 has also been found in recently published sequences of two other bacterial proteins with similar transport functions: *opp*D protein of the oligopeptide permease complex of Salmonella (Higgins et al., 1985) and *pst*B protein of the phosphate-specific transport complex of E. coli (Surin et al., 1985). Different bacterial periplasmic transport systems are characterized by a similar organization (Ames and Higgins, 1983). Within these multicomponent systems, *his*P, *mal*K, *opp*D, and *pst*B proteins correspond to relatively hydrophilic peripheral membrane components. These proteins interact with periplasmic substrate-binding proteins (*his*J or LAO in the case of *his*P), as well as with hydrophobic integral membrane proteins (*his*Q and *his*M in the case of *his*P). The interaction with integral membrane proteins is apparently required for membrane association of the peripheral membrane components (Shuman and Silhavy, 1981). *his*P, *mal*K, and *opp*D have been tentatively identified as the energy-coupling components in their respective active transport systems. This identification has been made on the basis of their capacity to bind ATP and from the presence of a consensus nucleotide-binding sequence in these proteins (Hobson et al., 1984; Higgins et al., 1985). As shown below, the same nucleotide-binding sequence is also present in the *pst*B protein.

A pairwise alignment of the *his*P and *mdr*1 proteins (not shown) has indicated that approximately 33% of all residues of the *his*P protein can be matched with identical residues of *mdr*1, and another 28% match with functionally similar residues. Similar levels of homology were observed with *mal*K, *opp*D, and *pst*B proteins. To identify the most highly conserved regions between bacterial transport proteins and *mdr*1, we have prepared a multiple alignment of the amino acid sequences of hydrophilic regions 1 and 2 and all four bacterial proteins (Figure 6). The highest level of homology is observed in the regions designated NB-1 and NB-2. These regions correspond to two parts of the nucleotide-binding fold (Walker et al., 1982), which has been previously detected in the *his*P, *mal*K, and *opp*D proteins (Higgins et al., 1985). In Figure 6, NB-1 and NB-2 regions have been aligned with the corresponding sequences of several known nucleotide-binding enzymes. The sequences of NB-1 and NB-2 are in agreement with the consensus sequence of a nucleotide-binding site, as identified by Walker et al. (1982). Though the highest homology between *mdr*1 and bacterial periplasmic transport proteins is observed within the nucleotide-binding sites, the homology is not limited to these sites. On the other hand, no significant homology outside of NB-1 and NB-2 regions could be detected between *mdr*1 and any other nucleotide-binding proteins, suggesting that the homology of *mdr*1 with periplasmic transport proteins indicates additional functional similarities aside from the potential nucleotide-binding properties.

Figure 7 shows a hypothetical model for membrane orientation of the human P-glycoprotein. This model is based on the predictions of transmembrane segments 1–6 and 1a–6a by the algorithm of Eisenberg et al. (1984). The protein is oriented so that the potential nucleotide-binding folds are located on the cytoplasmic side. This orientation is in agreement with the results of Kartner et al. (1985), who have used antibodies against the C-terminal region of hamster P-glycoprotein to show that the C-terminus is located inside the cell. The protein sequence of *mdr*1 contains ten potential N-glycosylation sites, Asn-X-Ser/Thr (Kornfeld and Kornfeld, 1985). Seven of these sites are, according to our model, located on the cytoplasmic side of the membrane, and therefore are not expected to be glycosylated. The remaining three sites, corresponding to residues 91–93, 94–96, and 99–102, are

```
mdr1                  351-415   SIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEF--RNVHFSYPSRKEVKI--LKGLNL
mdr1                  994-1058  FAPDYAKAKISAAHIIMIIEKTPLIDSYSTEGLMPNTLEGNVTF--GEVVFNYPTRPDIPV--LQGLSL
pstB                  1-31                                   MSMVETA-----------PSKIQVRNLNFYYGKFHALKNTNL
hisP                  1-27                                              MMSENKLHVIDLHKRYGGHEVLKGVSL
malK                  1-24                                                   MASVQLQNVTKAWGEVVSKDINL
oppD                  1-42                                   MSLSETATQAPQPANVLLEVNDLRVTFATPDGDVTAVNDLNF

                                    NB-1
                                  ------------------
mdr1                  416-465   KVQSGQTVALV-GNSGCGKS-TTVQLMQR-----LY-DPT-EGMVSVDGQDI---RT--INVRF-----
mdr1                  1059-1108 EVKKGQTLALV-GSSGCGKS-TVVQLLER-----FY-DPLA-GKVLLDGKEIK--R---LNVQW-----
pstB                  32-88     DIAKNQVTAFI-GPSGCGKS-T--LLRTFNKMFELYPEQRAEGEILLDGDNII-LTNSQDIAL-------
hisP                  28-87     QARAGDVISII-GSSGSGKS-T--FLRCINFL-----EKPSEGAIIVNGQNIINLVRDKDGQLKVADKNQ
malK                  25-74     DIHEGEFVVFV-GPSGCGKS-T--LLRMIAGL-----ETITSGDLFI-GEK----RMNDTP--PAE---
oppD                  43-98     TLRAGETLGIV-GESGSGKSQS--RLR-LMGL-----LATN-GRIGGSATF-NG-REILNLPERELNT-
Bovine ATPase β       148-169     AKGGKIGLF-GGAGVGKT-V--FIM
E. coli ATPase α      160-180     GRGGRELII-GDRGTGKT-A--LAI
Adenylate kinase      5-26        LKKSKIIFVVGGPGSGKG-T--QCE
K-ras                 1-24        MTEYKLVVV-GASGVGKSALTIQLI
Consensus                           G         G  G GKS T
                                                      T

mdr1                  466-522   ---LREIIGVVSQEPV--LFA-TTIAENIRYG-R--E---NVTMDEIEKAVKEANAYDFIMKLPHKFDT
mdr1                  1109-1167 ---LRAHLGIVSQEPI--LFD-CSIAENIAYGDN----SRVVSQEEIVRAAKEANIHAFIESLPNKYST
pstB                  89-145    ---LRAKVGMVFQKPT--FPP-MSIYDNIAFGVRLFEKLSRADMDERVQWALTLAA------LWNETKD
hisP                  88-145    LRLLRTRLTMVFQHFN--LWSHMTVLENVMEAPIQVLGLSKHDARERALKYL-A-K------VGIDER-
malK                  75-125    -RG----VGMVFQSYA--LYPHLSVAEN-MSFGLKPAGAKKEVINQRVNQ-V-A-E------VLQLAH-
oppD                  99-158    RRA--EQISMIFQDHMTSLNPYMRVGEQLMEVLMLHKGMSKAEAFEESVRMLDAVK------MLEARK-

                                                    NB-2
                                              ---------------
mdr1                  523-583   LVGERGAQLSGGQKQRIAIARALVRNPKILLLDEATSALDTES-------EAVVQVALDKARKGRT-TI
mdr1                  1168-1228 KVGDKGTQLSGGQKQRIAIARALVRQPHILLLDEATSALDTES-------EKVVQEALDKAREGRT-CI
pstB                  146-205   KLHQSGYSLSGGQQQRLCIARGIAIRPEVLLLDEPCSALDPISTG---RIE-E----LITELKQDY-TV
hisP                  146-206   AQGKYPVHLSGGQQQRVSIARALAMEPDVLLFDEPTSALDPELVGEVLRI-MQ----QLAEE-GK--TM
malK                  126-187   LLDRKPKALSGGQRQRVAIGRTLVAEPSVFLLDEPLSNLDAALRVQ-MRIEIS----RLHLRLGR--TM
oppD                  159-219   RMKMYPHEFSGGMRQRVMIAMALLCRPKLLIADEPTTALDVT--VQ-AQI-MT----LNELKREFNTA
Bovine ATPase β       241-267                 VAEYFRDQEGQDVLLFIDNIFRFTQA
E. coli ATPase α      265-290                 MGEYFRD-RGEDALIIYDDLSKQAVA
Adenylate kinase      102-126                 GEEFERK-IGQPTLLLYVDAGPETMT
ATP/ADP translocase   275-300                 SNVL-RGMGGAFVLVLYDEIKKFV
Consensus                                        R   G    hhhhDE
                                                              D

mdr1                  584-637   VI--------AHRLSTVRNADVIAGFDDGVIVEKGNHDELMKEKGIYF--KLVTMQTAG
mdr1                  1229-1280 VI--------AHRLSTIQNADLLVVFQNGRVKEHGTHQQLLAQKGIYF--SMVSVQ-AG--TKRQ
pstB                  206-257   VI-VTHNMQQAARCS-----DHTAFMYLIGELIEFSNTDDL-------F-TKPAKKQTEDYITGRYG
hisP                  207-258   VV-VTHEMGFARHVS-----SHVIFLHQGKIEEEGDPEQV-------F-GNPQSPRLQQFLKGSLK
malK                  188-244   IY-VTHDQVEAMTLA-----DKIVVLDAGRVAQVGKPLAVP------LSGRPFCRRIYRFAKDELLPVK
oppD                  220-275   IIMITHDLGVVAGIC-----DKVLVMYAGRTMEYGKARDV-------F-YQPVHPYSIGLLNAVPRLDS

malK                  245-312   VTATAIDQVQVELPMPNRQQVWLPVESRDVQVGANMSLGIRPEHLLPSDIADVILEGEVQVVEQLGNET
oppD                  276-335   EGAEMLTIPGNPPNLLRLPKGCPFQPRCPHAMEICNNAPPLEAFSPGRLRACFLPVEELL

malK                  313-370   QIHIQIPSIRQNLVYRQNDVVLVEEGATFAIGLPPERCHLFREDGTACRRLHKEPGV
```

Figure 6. Alignment of Amino Acid Sequences of Two Hydrophilic Regions of *mdr*1 and Bacterial Transport Proteins *pst*B, *his*P, *mal*K, and *opp*D

In the regions of consensus nucleotide-binding sites (NB-1 and NB-2), these sequences have also been aligned with several known nucleotide-binding proteins. The sequences and alignment of *his*P, *mal*K, and *opp*D are from Higgins et al. (1985). The sequence of *pst*B is from Surin et al. (1985). The sequences and alignment of nucleotide-binding proteins (except for *K-ras*), as well as the consensus nucleotide-binding sequence are from Walker et al. (1982). The amino acid sequence of viral *K-ras* is from Tsuchida et al. (1982). Amino acids identical in three or more of the proteins, including at least one of *mdr*1 sequences, are boxed. h in the consensus sequence indicates a conserved hydrophobic residue.

all clustered in one predicted extracytoplasmic domain. These three sites are tentatively shown as glycosylated in Figure 7.

Functional Considerations

The predicted transmembrane localization, the presence of potential nucleotide-binding sites, and homology of the hydrophilic regions of the *mdr*1 gene product with bacterial transport proteins are consistent with the hypothesis that P-glycoprotein functions as an efflux pump responsible for the removal of drugs from the cell by an ATP-dependent mechanism. Since genetic evidence has implicated *mdr*1 as the gene responsible for the multidrug-resistant phenotype, analysis of the P-glycoprotein sequence, together with the data on increased drug efflux in multidrug-resistant cells and the sensitivity of the efflux to inhibitors of energy metabolism provides a strong argument that active drug efflux is the main mechanism of multidrug resistance.

While homology of *mdr*1 with *his*P, *mal*K, *opp*D, and

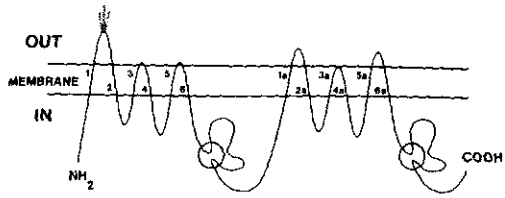

Figure 7. Model for the Transmembrane Orientation of P-Glycoprotein

The transmembrane segments, predicted by the algorithm of Eisenberg et al. (1984), are as follows: 1, residues 52–72; 2, residues 120–140; 3, residues 189–209; 4, residues 216–236; 5, residues 297–317; 6, residues 326–346; 1a, residues 711–731; 2a, residues 757–777; 3a, residues 833–853; 4a, residues 854–874; 5a, residues 937–957; and 6a, residues 974–994. The predicted glycosylation sites are marked by chains. The predicted nucleotide-binding folds are circled.

*pst*B indicates either a common evolutionary origin for the P-glycoprotein and bacterial periplasmic transport systems or functional similarities in their transport mechanisms, there are clear structural and functional differences between these systems. Structurally, *his*P, *mal*K, *opp*D, and *pst*B are separate proteins that form a quarternary complex with the binding protein and hydrophobic integral membrane components, whereas the hydrophilic regions of P-glycoprotein are parts of a larger protein, which includes its own hydrophobic membrane-spanning domains. No significant homology has been observed between the hydrophobic regions of P-glycoprotein and the integral membrane components of bacterial periplasmic transport systems (data not shown). It should be noted, however, that the sequences of integral membrane proteins are not highly conserved even among different bacterial transport systems (Ames and Higgins, 1983). The most obvious functional difference between P-glycoprotein and the bacterial complexes concerns the direction of the transport, since P-glycoprotein is presumed to pump its substrates out of the cell, whereas the bacterial transport systems function to deliver the substrate from the periplasmic space into the cell. Another important functional difference concerns the limited substrate specificity of the bacterial transport systems, as opposed to a very broad range of lipophilic compounds transported by P-glycoproteins. In the bacterial systems, the specificity is achieved primarily through the recognition of the substrate by specialized periplasmic binding proteins, which then transmit the substrate to the membrane components. The substrate specificity in the histidine and maltose transport systems, however, can also be affected by mutations in the integral membrane components (Payne et al., 1985; Shuman, 1982). Furthermore, mutations in the integral membrane components of the maltose transport system may result in direct binding and translocation of maltose by the membrane complex, even in the absence of the periplasmic binding protein (Shuman, 1982; Treptow and Shuman, 1985). A similar observation has been reported for the methylgalactoside transport system of E. coli (Robbins and Rotman, 1975). In the P-glycoprotein system, biochemical evidence indicates that the P-glycoprotein itself is capable of substrate binding (Cornwell et al., 1986), but the existence of additional drug-binding cytoplasmic protein(s) cannot be excluded by the available data. Future studies will indicate whether heterogeneity of P-glycoproteins, observed in different cell lines, contributes to the broad substrate specificity of this transport system.

To understand the significance of the internal duplication in the P-glycoprotein, it will be important to determine whether all twelve predicted transmembrane segments interact to form a single transmembrane channel or whether each hydrophobic region forms a separate channel composed of six segments. At present, we cannot distinguish between these possibilities. Each half of the protein contains a potential nucleotide-binding site. If the function of these sites involves hydrolysis of ATP or a related nucleotide, P-glycoprotein would represent the first known example of a single protein with two ATPase sites. It is possible, however, that nucleotide binding at either one or both sites does not lead to hydrolysis of ATP but rather plays a regulatory role. It would also be important to determine whether the function of the potential nucleotide-binding sites is differentially affected by phosphorylation of P-glycoprotein (Carlsen et al., 1977). With regard to the drug-binding properties of the P-glycoprotein, it is tempting to speculate that internal duplication in the P-glycoprotein structure gives rise to two different drug-binding sites, thereby increasing the range of substrates for the P-glycoprotein system. There is presently no information regarding the location of drug-binding sites in P-glycoprotein.

The availability of *mdr1* cDNA clones and the complete amino acid sequence of the human P-glycoprotein make it possible now to initiate detailed biochemical and genetic analysis of the mechanisms by which normal and tumor cells protect themselves against potentially cytotoxic lipophilic compounds. The predictions of the transmembrane orientation and structural organization of the human P-glycoprotein also provide a convenient working model for designing new chemical and immunological approaches to the problem of drug resistance in cancer chemotherapy.

Experimental Procedures

DNA Sequencing

Prior to sequencing, cDNA inserts from λgt11 phage clones were recloned into the EcoRI sites of plasmid vectors pUC18 or pGEM4 (Promega Biotec). Clone λHDR5, containing an internal EcoRI site, was subcloned as two separate fragments, designated pHDR5A (3′ end) and pHDR5B (5′ end). The genomic clone pHDR4.4 (Roninson et al., 1986) was mapped with several restriction enzymes, and individual 200–1000 bp fragments were subcloned into pUC18. DNA was sequenced using modifications of the enzymatic chain-termination technique (Sanger et al., 1977). A part of the cDNA sequence was determined by subcloning the inserts into an M13 vector (Messing, 1983) and generating a series of overlapping deletion subclones (Henikoff, 1984). Most of the sequence was obtained by using supercoiled plasmid DNA as a template (Zagursky et al., 1985) and synthesizing a series of specific oligonucleotide primers with a DNA synthesizer (Applied Biosystems). All the sequences were determined on both strands, with a minimum of two gel readings per sequence.

Sequence Analysis

Homology search and initial alignments of amino acid and nucleotide sequences were done by the program of D. J. Lipman and W. R. Pearson based on the algorithm of Wilbur and Lipman (1983). The hydropa-

thy plots (Kyte and Doolittle, 1982) were obtained using sequence analysis software from International Biotechnologies, Inc. The hydropathy and hydrophobic moment analysis was done by the algorithm of Eisenberg et al. (1984).

Acknowledgments

We are grateful to Dr. David Eisenberg (UCLA) for his kind gift of a computer program for hydropathy analysis and helpful discussions. We also thank Rena Soffir for excellent technical assistance. This work was supported by the National Cancer Institute grant CA40333 to I. B. R. D. P. C. is an HHMI/NIH Research Scholar.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 29, 1986.

References

Akiyama, S.-I., Fojo, A., Hanover, J. A., Pastan, I., and Gottesman, M. M. (1985). Isolation and genetic characterization of human KB cell lines resistant to multiple drugs. Somat. Cell Mol. Genet. *11*, 117–126.

Ames, G. F.-L., and Higgins, C. F. (1983). The organization, mechanism of action and evolution of periplasmic transport systems. Trends Biochem. Sci. *8*, 97–100.

Beck, W. T., Cirtain, M. C., and Lefko, J. L. (1983). Energy-dependent reduced drug binding as a mechanism of Vinca alkaloid resistance in human leukemic lymphoblasts. Molec. Pharmac. *24*, 485–492.

Carlsen, S. A., Till, J. E., and Ling, V. (1977). Modulation of drug permeability in Chinese hamster ovary cells. Possible role for phosphorylation at surface glycoproteins. Biochim. Biophys. Acta *467*, 238–250.

Cornwell, M. M., Safa, A. R., Felsted, R. L., Gottesman, M. M., and Pastan, I. (1986). Membrane vesicles from multidrug-resistant cancer cells contain a specific 150- to 170-kDa protein detected by photoaffinity labeling. Proc. Natl. Acad. Sci. USA *83*, 3847–3850.

Dano, K. (1973). Active outward transport of daunomycin in resistant Ehrlich ascites tumor cells. Biochim. Biophys. Acta *323*, 466–483.

Dayhoff, M. O., Eck, R. V., and Park, C. M. (1972). A model of evolutionary change in proteins. In Atlas of Protein Sequence and Structure 1972, Vol. 5. M. O. Dayhoff, ed. (Washington, D.C.: National Biomedical Research Foundation), pp. 89–99.

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface protein sequences with a hydrophobic moment plot. J. Mol. Biol. *179*, 125–142.

Fojo, A., Akiyama, S. I., Gottesman, M. M., and Pastan, I. (1985). Reduced drug accumulation in multiply drug-resistant human KB carcinoma cell lines. Cancer Res. *45*, 3002–3007.

Fojo, A., Lebo, R., Shimizu, N., Chin, J. E., Roninson, I. B., Merlino, G. T., Gottesman, M. M., and Pastan, I. (1986). Localization of multidrug resistance-associated DNA sequences to human chromosome 7. Somat. Cell Mol. Genet. *12*, 415–420.

Gilson, E., Nikaido, H., and Hofnung, M. (1982). Sequence of the *malK* gene in E. coli K12. Nucl. Acids Res. *10*, 7449–7458.

Gros, P., Croop, J., Roninson, I. B., Varshavsky, A., and Housman, D. E. (1986a). Isolation and characterization of DNA sequences amplified in multidrug-resistant hamster cells. Proc. Natl. Acad. Sci. USA *83*, 337–341.

Gros, P., Ben Neriah, Y., Croop, J. M., and Housman, D. E. (1986b). Isolation and expression of a cDNA (*mdr*) that confers multidrug resistance. Nature, in press.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene *28*, 351–357.

Higgins, C. F., Haag, P. D., Nikaido, K., Ardeshir, F., Garcia, G., and Ames, G. F.-L. (1982). Complete nucleotide sequence and identification of membrane components of the histidine transport operon of S. typhimurium. Nature *298*, 723–727.

Higgins, C. F., Hiles, I. D., Whalley, K., and Jamieson, D. J. (1985). Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. EMBO J. *4*, 1033–1040.

Hobson, A. C., Weatherwax, R., and Ames, G. F.-L. (1984). ATP-binding sites in the membrane components of histidine permease, a periplasmic transport system. Proc. Natl. Acad. Sci. USA *81*, 7333–7337.

Inaba, M., Kobayashi, H., Sakurai, Y., and Johnson, R. K. (1979). Active efflux of daunorubicin and adriamycin in sensitive and resistant sublines of P388 leukemia. Cancer Res. *39*, 2200–2203.

Juliano, R. L., and Ling, V. (1976). A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim. Biophys. Acta *455*, 152–162.

Kartner, N., Riordan, J. R., and Ling, V. (1983). Cell surface P-glycoprotein associated with multidrug resistance in mammalian cell lines. Science *221*, 1285–1288.

Kartner, N., Evernden-Porelle, D., Bradley, G., and Ling, V. (1985). Detection of P-glycoprotein in multidrug-resistant cell lines by monoclonal antibodies. Nature *316*, 820–823.

Kornfeld, K., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Ann. Rev. Biochem. *54*, 631–664.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acids Res. *12*, 857–872.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.

Ling, V., Kartner, N., Sudo, T., Siminovitch, L., and Riordan, J. R. (1983). The multidrug resistance phenotype in Chinese hamster ovary cells. Cancer Treat. Rep. *67*, 869–874.

Messing, J. (1983). New M13 vectors for cloning. Meth. Enzymol. *101*, 20–36.

Payne, G. M., Spudich, E. N., and Ames, G. F.-L. (1985). A mutational hot-spot in the *hisM* gene of the histidine transport operon in Salmonella typhimurium is due to deletion of repeated sequences and results in an altered specificity of transport. Mol. Gen. Genet. *200*, 493–496.

Peterson, R. H. F., Meyers, M. B., Spengler, B. A., and Biedler, J. L. (1983). Alteration of plasma membrane glycopeptides and gangliosides of Chinese hamster cells accompanying development of resistance to daunorubicin and vincristine. Cancer Res. *43*, 222–228.

Riordan, J. R., and Ling, V. (1985). Genetic and biochemical characterization of multidrug resistance. Pharmac. Ther. *28*, 51–75.

Riordan, J. R., Deuchars, K., Kartner, N., Alon, N., Trent, J., and Ling, V. (1985). Amplification of P-glycoprotein genes in multidrug-resistant mammalian cell lines. Nature *316*, 817–819.

Robbins, A. R., and Rotman, B. (1975). Evidence for binding protein-independent substrate translocation by the methylgalactoside transport system of Escherichia coli K12. Proc. Natl. Acad. Sci. USA *72*, 423–427.

Roninson, I. B. (1983). Detection and mapping of homologous, repeated and amplified DNA sequences by DNA renaturation in agarose gels. Nucl. Acids Res. *11*, 5413–5431.

Roninson, I. B., Abelson, H. T., Housman, D. E., Howell, N., and Varshavsky, A. (1984). Amplification of specific DNA sequences correlates with multidrug resistance in Chinese hamster cells. Nature *309*, 626–628.

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). Isolation of human *mdr* DNA sequences amplified in multidrug-resistant KB carcinoma cells. 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). Vinblastine photoaffinity labeling of a high molecular weight surface membrane glycoprotein specific for multidrug-resistant cells. J. Biol. Chem. *261*, 6137–6140.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Scotto, K. W., Biedler, J. L., and Melera, P. W. (1986). Amplification and expression of genes associated with multidrug resistance in mammalian cells. Science *232*, 751–755.

Shen, D.-w., Fojo, A., Chin, J. E., Roninson, I. B., Richert, N., Pastan, I., and Gottesman, M. M. (1986a). Human multidrug-resistant cell lines: increased *mdr1* expression can precede gene amplification. Science *232*, 643–645.

Shen, D.-w., Cardarelli, C., Hwang, J., Cornwell, M., Richert, N., Ishii, S., Pastan, I., and Gottesman, M. (1986b). Multiple drug-resistant human KB carcinoma cells independently selected for high-level resistance to colchicine, adriamycin, or vinblastine show changes in expression of specific proteins. J. Biol. Chem. *261*, 7762–7770.

Shen, D.-w., Fojo, A., Roninson, I. B., Chin, J. E., Soffir, R., Pastan, I., and Gottesman, M. M. (1986c). Multidrug-resistance in DNA mediated transformants is linked to transfer of the human *mdr1* gene. Mol. Cell. Biol., in press.

Shuman, H. A. (1982). Active transport of maltose in Escherichia coli K12. J. Biol. Chem. *257*, 5455–5461.

Shuman, H. A., and Silhavy, T. J. (1981). Identification of the *malK* gene product. J. Biol. Chem. *256*, 560–562.

Skovsgaard, T. (1978). Mechanisms of resistance to daunorubicin in Ehrlich ascites tumor cells. Cancer Res. *38*, 1785–1791.

Surin, B. P., Rosenberg, H., and Cox, G. B. (1985). Phosphate-specific transport system of Escherichia coli: nucleotide sequence and gene-polypeptide relationships. J. Bacteriol. *161*, 189–198.

Trent, J., Bell, D., Willard, H., and Ling, V. (1985). Chromosomal localization in normal human cells and CHO cells of a sequence derived from P-glycoprotein (PGY1). Hum. Gene Mapping *8*, 761–762.

Treptow, N. A., and Shuman, H. A. (1985). Genetic evidence for substrate and periplasmic-binding protein recognition by the MalF and MalG proteins, cytoplasmic membrane components of the Escherichia coli maltose transport system. J. Bacteriol. *163*, 654–660.

Tsuchida, M., Ohtsubo, E., and Ryder, T. (1982). Nucleotide sequence of the oncogene encoding the p21 transforming protein of Kirsten murine sarcoma virus. Science *217*, 937–939.

Van der Bliek, A. M., Van der Velde-Koerts, T., Ling, V., and Borst, P. (1986). Overexpression and amplification of five genes in a multidrug-resistant Chinese hamster ovary cell line. Mol. Cell. Biol. *6*, 1671–1678.

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide-binding fold. EMBO J. *8*, 945–951.

Wilbur, W. J., and Lipman, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA *80*, 726–730.

Willingham, M. C., Cornwell, M. M., Cardarelli, C. O., Gottesman, M. M., and Pastan, I. (1986). Single cell analysis of daunomycin uptake and efflux in multidrug-resistant and sensitive KB cells: effects of verapamil and other drugs. Cancer Res., in press.

Young, R. A., and Davis, R. W. (1983). Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA *80*, 1194–1198.

Zagursky, R. J., Baumeister, K., Lomax, N., and Berman, M. L. (1985). Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. Gene Anal. Techn. *2*, 89–94.