PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A2

(11) International Publication Number: **WO 97/49832**

(43) International Publication Date: 31 December 1997 (31.12.97)

(21) International Application Number: PCT/CA97/00488

(22) International Filing Date: 23 June 1997 (23.06.97)

(30) Priority Data:

08/667,546	21 June 1996 (21.06.96)	US
08/724,466	1 October 1996 (01.10.96)	US

(60) Parent Applications or Grants

(63) Related by Continuation

US	08/667,546 (CIP)
Filed on	21 June 1996 (21.06.96)
US	08/724,466 (CIP)
Filed on	1 October 1996 (01.10.96)

(71) Applicant *(for all designated States except US)*: QUEEN'S UNIVERSITY AT KINGSTON [CA/CA]; Kingston, Ontario K7L 3N6 (CA).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: PETKOVICH, P., Martin [CA/CA]; 875 Everitt Avenue, Kingston, Ontario K7M 4R1 (CA).

(74) Agents: HUNT, John, C. et al.; Blake, Cassels & Graydon, P.O. Box 25, Commerce Court West, Toronto, Ontario M5L 1A9 (CA).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD OF IDENTIFYING CYTOCHROME P450

(57) Abstract

A method for identifying a nucleotide sequence encoding a cytochrome P450, the cytochrome being inducible or suppressible in cells of a given type by an agent. The method includes: exposing a first group of the cells to the agent so as to induce or suppress expression of the cytochrome P450; isolating first mRNA from the cells; isolating second mRNA from a second group of the cells which have not been exposed to the agent so as to induce or suppress expression of the cytochrome P450; amplifying the first and second mRNA, respectively, in the presence of an oligo(dT) based first nucleic acid primer sufficient to prime synthesis from a poly(A) tail and a second nucleic acid primer substantially complementary to a nucleic acid sequence encoding a conserved region of a known cytochrome P450; displaying amplified products of the first mRNA and amplified products of the second mRNA to detect differences therebetween; and identifying said nucleotide sequence encoding said inducible or suppressible cytochrome P450.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHOD OF IDENTIFYING CYTOCHROME P450

Cross-References to Related Applications

This application is a continuation-in-part application of U.S. patent application Serial No. 08/724,466, filed October 1, 1996, which is a continuation-in-part application of U.S. patent application Serial No. 08/667,546, filed June 21, 1996, the specifications of which applications are incorporated herein by reference.

Field of the Invention

This invention relates to identification of enzymes known as cytochromes that are differentially expressed in a cell in response to variation in the amount of a given factor.

Background of the Invention

Cytochrome P450 genes encode enzymes which are involved in a wide range of biochemical transformations throughout nature. These heme-containing proteins, which in general are bound to membranes of the endoplasmic reticulum, serve in the oxidative metabolism of both endogenous and exogenous compounds, such as steroids, and xenobiotics, such as polycyclic hydrocarbons [Nebert, 1987; Gonzalez, 1992]. Most of the reactions catalyzed by cytochrome P450s involve the addition of one atom of molecular oxygen to a given substrate. The consequence of this activity can be either bioactivation or catabolism. The ability to control the activities of certain cytochrome P450s is desirable therapeutically, in the control of insect populations, and in the control of fungal and bacterial growth and function.

For example, the fungal cytochrome P450 sterol 14α-demethylase is a key enzyme in ergosterol biosynthesis, and a number of highly effective inhibitors of this enzyme have been developed either as fungicides in agriculture or as antimycotic drugs in medical applications [Koller, 1992]. In addition, inhibitors of specific cytochrome P450s, for example, fluconazole, have been useful adjuncts in the treatment of prostate cancer.

More than 200 cytochrome P450 genes have been described and classified into 36 families; within a single family, cytochrome P450 proteins have greater than 40% amino acid identity and within a subfamily, this is usually greater than 55% [Nelson, 1993]. All cytochrome P450s are heme binding proteins, associate with specific electron transfer components (e.g. ferrodoxin, adrenodoxin), and bind molecular oxygen. It is thus not surprising that several regions of known P450 sequences have been found to be conserved across species.

Summary of the Invention

According to the present invention, it is possible to use nucleic acid amplification technology such as PCR (polymerase chain reaction) [Innis, 1990] to identify cytochromes specifically involved in certain biological reactions. In one particular aspect, the invention includes identifying a cytochrome P450 which is expressed in response to high levels of a particular agent. For example, cytochrome P450RAI has been isolated and characterized

[White, 1996] by means of "differential display" of mRNA [Liang, 1992] existing in the presence and absence of retinoic acid, a substrate of P450RAI. This cytochrome specifically metabolizes the active vitamin A derivative, retinoic acid (RA) and is highly induced by RA in certain cell lines and tissues. This is one example where the substrate of a particular cytochrome P450 can regulate the expression of that cytochrome. Another cytochrome P450 shown to be regulated by a substrate is the 1,25-dihydroxyvitamin D3 metabolizing enzyme CYP24 which is involved in vitamin D catabolism. Notably the substrate 1,25-dihydroxyvitamin D3 regulates the CYP24 promoter through a vitamin D response element, resulting in the accelerated catabolism of 1,25-dihydroxyvitamin D3 [Makin, 1989]. The invention described herein provides a means to identify all such cytochromes P450 fom any species--human, animal, plant, parasite, bacteria, fungi- where such a negative feedback loop is operative, ie where the substrate also regulates the expression of the cytochrome P450.

In another particular aspect, the present invention includes identifying a cytochrome P450 which is normally produced by a cell, in which such production is suppressed by exposure of the cell to a particular agent. In either aspect, the agent may or may not be a substrate of the cytochrome being identified. An example of this type of suppression is seen with the negative regulation of the vitamin D 1-alpha hydroxylase expression in kidney by 1,25-dihydroxyvitamin D3 [Henry, 1979].

Inducers aside from RA include certain steroids and xenobiotics, for example.

Specific inhibitors of cytochrome P450RAI find utility, for example, in increasing the effectiveness of RA in the treatment of cancer, such as acute promyelocytic leukemia, breast and prostate cancer, and the treatment of skin disorders such as actinic keratosis, icthyosis, acne and psoriasis. Inhibitors of other P450RAIs, however induced, are useful in reducing the metabolism their substrates. Other cytochrome P450s will also be useful therapeutically, for example, for the design of specific inhibitors of the cytochrome as well as design of compounds that have certain desireable properties of the substrate but which is not subject to metabolism by the cytochrome.

Brief Description of the Drawings

A detailed description of the invention follows, in which:

Figures 1A to 1H show primary structures of exemplary cytochrome P450s (SEQ ID NOs:1 to 11, respectively) taken from the literature, in which boxed portions indicate predicted similar secondary structural portions, conserved amino acid sequences of the primary structures being aligned; and

Figures 2A and 2B illustrate the autoregulatory feedback loop involving RA induction of a cytochrome P450 and metabolism of RA by the induced cytochrome P450.

Figure 3 is a schematic representation of the steps used to isolate retinoid-regulated genes using differential display of mRNA. Schematic representation of the steps involved in the isolation of retinoid-regulated genes using the differential display technique. The cloned products isolated in step 6 can then be used for sequencing, Northern blotting or screening of cDNA libraries. P1, P2 and P3 correspond to fragments from RA induced mRNAs. P4 represents a PCR product from an mRNA which is down-regulated.

Figure 4(a) shows a polyacrylamide gel of PCR amplified mRNA in duplicate obtained using retinoic acid-treated fish (lanes 1 and 2) and dimethyl sulfoxide (DMSO-) treated control fish (lanes 3 and 4). The arrow indicates a PCR amplified band present in the RA-treated samples and not observed in the controls.

Figure 4(b) shows the nucleotide sequence (SEQ ID NO:33) of the 337 base pair PCR product isolated from the band (arrow) of Figure 4(a). The arrows indicate the nucleotide sequences where the upstream and downstream priming sites for differential display PCR amplification were located in the 3'- untranslated portion of zP450RAI.

Figures 4C(i) and 4C(ii) show an amino acid sequence (SEQ ID NO:34) corresponding to cDNA (492 amino acid open reading frame). The boxed residues indicate the heme-binding motif characteristic of cytochrome P450s.

Figure 4(d) shows amino acid sequence comparisons between zP450RAI and several other cytochrome P450s (SEQ ID Nos:40,41,42,43,44) in the area of the conserved heme-binding motif found in the superfamily. The cysteine, designated 0 in the figure, which has been shown to be directly involved in heme-binding [Gotoh, 1989] is surrounded by several highly conserved amino acids.

Figure 5(a) shows Northern blot analysis of mRNAs obtained from regenerate tissue of RA-treated fish in lane 5, and controls (DMSO-treated fish) in lane 4, using a zP450RAI cDNA probe. Comparison to an RNA ladder (lane 1) shows the major zP450RAI transcript to be in the 1.4-2.4 kb range.

Figure 5(b) shows localization of zP450RAI transcripts in regenerating caudal fin tissue 72 hours post-amputation by whole mount *in situ* hybrization. (i) zP450RAI transcripts were found to be undetectable in DMSO-treated regenerates. The original plane of amputation is indicated by the white line with arrowhead; m (soft mesenchyme) and r (bony rays) are labelled. (ii) In a sample obtained from an RA-treated fish, zP450RAI transcripts, indicated by the black arrowhead, were found to be localized to a band of cells extending across the distal tip of the regenerate. Lower levels of expression of zP450RAI were also evident in non-regenerate tissue at the proximal base of the isolated fin, as indicated by the black line with arrowhead. The plane of amputation is indicated by the white line with arrowhead as in Figure 5(b)(i). (iii) A histological section taken through the plane is indicated by the line. (iv) A histological section of RA-treated fins post-hybridization is shown. Localized expression of zP450RAI was detected in a subset of epithelial cells (black arrowhead) which lie at the distal

tip of the regenerate. Basement membrane separating the dense blastemae and the wound epithelium is indicated by the grey arrowhead.

Figure 6 shows elution profiles of lipid soluble extracts obtained from treated media of pSG5-zP450RAI transfected COS-1 cells and pSG5 transfected control cells. Figures 6(a) and 6(b) are plots of cpm (% of total cpm) vs fraction number for cells incubated with 575 pM [11,12-^{3}H]RA for 4 hours and 24 hours, respectively, pSG5-zP450RAI COS-1 cells (—) and control cells (---). Metabolism of [11,12-^{3}H]RA to total aqueous soluble metabolites was measured using aliquots of the aqueous soluble extract subjected to β-scintillation counting. See insets of Figures 6(a) and (b). Figures 6(c) and 6(d) are plots of absorbance vs retention time for cells incubated with 1μM RA for 4 and 24 hours, respectively. Peaks observed in

zP450RAI transfected cell are shaded black. The region of the chromatogram from 4 to 6 min has been expanded (see insets of Figures 6(c) and (d)). In cells transfected with zP450RAI cDNA, the generation of peaks corresponding to 4-oxo RA and 4-OH RA was observed.

Figure 7 shows results obtained with human cell lines probed with a α-[^{32}P]-dATP labeled probe having the sequence identified as SEQ ID NO:37: HEK293; EL-E; HL-60; MCF10A; LC-T; SK-LC6; and MCF7. (+) indicates pretreatment with 10^{-6}M RA and (-) indicates no RA pretreatment. The blot was also probed with hGAPDH to control for RNA loading of the gel, shown in the bottom panel.

Figure 8 is similar to Figure 7 for the cell lines U937 and HepG2.

Figure 9(a) shows elution profiles of lipid soluble extracts obtained from treated media of pSG5-hP450RAI transfected COS-1 cells and pSG5 transfected control cells. Plots of cpm vs fraction number for cells incubated with [11,12-^{3}H]RA for 24 hours of pSG5-hP450RAI COS-1 cells (—) and control cells (—) are shown. Figure 9(b) shows measurement of aliquots of the aqueous soluble extract subjected to β-scintillation counting taken to determine metabolism of [11,12-^{3}H]RA to total aqueous soluble metabolites. Figure 9(c) shows plots of absorbance vs retention time for hP450RAI transfected cell (—) and control cells (—) cells incubated with 1μM RA for 24 hours. The inset is the region around 10 minutes, expanded for clarity.

Figure 10(a) shows 4-oxo-RA production of pSG5-hP450RAI transfected COS-1 cells and pSG5 transfected control cells.

Figure 10(b) shows 4-OH-RA production of pSG5-hP450RAI transfected COS-1 cells and pSG5 transfected control cells.

Figure 10(c) shows formation of aqueous soluble metabolites of pSG5-hP450RAI transfected COS-1 cells and pSG5 transfected control cells.

Figure 10(d) shows unmetabolized RA of pSG5-hP450RAI transfected COS-1 cells and pSG5 transfected control cells.

Figure 11 is similar to Figure 7 for the NT2 cell line.

Figure 12 is similar to Figure 7 for a normal NB4 cell line (first two lanes) and three individually derived retinoic acid resistant NB4 derivative cell lines.

Figure 13(a) shows elution profiles of lipid soluble extracts obtained from media of MCF10A cells exposed to RA and unexposed MCF10A control cells. Plots of cpm vs fraction number for cells incubated with [11,12-^{3}H]RA for 24 hours of RA-induced MCF10A cells (—) and control (—) are shown.

Figure 13(b) shows elution profiles of lipid soluble extracts obtained from treated media of MCF7 cells exposed to RA and unexposed MCF7 control cells. Plots of cpm vs fraction number for cells incubated with [11,12-^{3}H]RA for 24 hours of RA-induced MCF7 cells (—) and control (—) are shown.

Figure 13(c) shows the total aqueous soluble metabolites measured using aliquots of the aqueous soluble extracts of the cell lines described in Figures 13(a) and (b) subjected to β-scintillation counting. The first two bars are for unexposed MCF7 cells and

MCF7 cells exposed to RA, respectively. The third and fourth bars are for unexposed MCF10A cells and MCF10A cells exposed to RA, respectively.

Figure 14(a) shows elution profiles of lipid soluble extracts obtained from microsomal preparations after incubation with radiolabelled RA for ninety minutes, as described in Example 7. Plots of dpm vs fraction number for HeLa P microsomes (■,▲) and HeLa RAI microsomes (▼,◆).

Figure 14(b) shows fractions 5 to 15 of Figure 14(a) on a larger scale.

Figure 15 shows relative luciferase activity induced in cells containing a luciferase vector into which was cloned a portion of the putative promoter for mP450RAI. Luciferase activity was measured in cell extract supernatants from cells transfected with 3 concentrations of expression vectors comprising cDNAs encoding RXRα and RARγ (100 ng, 500 ng, and 1μg) along with a luciferase-based reporter gene including either the sense or antisense promoter sequence, or no promoter sequence, grown in presence and absence of RA.

Figure 16 shows inhibition of P450RAI mRNA in MCF7 cells by 4-hydroxy-phenylretinamide (4-HPR). Cells were treated for twelve hours with the indicated concentrations of *all-trans* retinoic acid (*at*RA) and 4-HPR. Total RNA was extracted using TRIzol, and, following electrophoresis, Northern blotting was performed as described. The nitrocellulose was probed with radiolabelled P450RAI then GAPDH.

Figure 17 shows expression of cytochrome P450RAI in MCF7 cells by northern blot analysis, over time, after administration of all-*trans* retinoic acid.

Figure 18 shows expression of cytochrome P450RAI in MCF7 cells by northern blot analysis, over time, after administration of all-*trans* retinoic acid alone, and in the presence of all-*trans* retinoic acid and ketoconazole.

Figure 19 shows expression of cytochrome P450RAI in MCF7 cells by northern blot analysis, over time, after administration of all-*trans* retinoic acid and after administration of Am580.

Detailed Description of the Invention

The present invention is illustrated below through the use of examples. Figure 1 shows amino acid sequences of eleven cytochrome P450s (SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, respectively), selected as representative of this group of proteins. The cytochrome P450 superfamily is a group of over 100 proteins, subdivided into microsomal and mitochondrial isoforms, that are responsible for the metabolism (e.g. hydroxylation) of endogenous and exogenous (xenobiotic) compounds [Guengerich, 1991]. Their structure is well conserved across the superfamily, with domains for heme binding, ferrodoxin binding, O_2 binding, and substrate binding. These proteins are membrane associated and are thus not easily studied by X-ray crystallographic means. For several mammalian steroidal cytochrome P450 isoforms (e.g. aromatase, cholesterol side-chain cleavage enzyme, 17-hydroxylase , rat 2B1, human 2D6), molecular modeling studies (e.g. Graham-Lorence, 1995; Vijayakumar,

1992) have begun based on information derived from crystal structures of soluble prokaryotic cytochromes P450 (CAM, BM-3, TERF, EryF). Thus, for the mammalian cytochromes P450, this work is in its infancy. Nevertheless, the approach appears highly promising. Such models,

despite being crude first approximations, allow for identification of putative active site residues suitable for site-directed mutagenesis studies and/or for rational drug design. Refinements of the model derived from mutant proteins then follow.

Examples of known cytochrome P450s in addition to those shown in Figure 1 are given in the literature [Nelson, 1995; Nelson Homepage; Directory of P-450 Containing Systems]. The important point of Figure 1, as far as the present invention is concerned, is that these functionally related cytochrome P450 proteins have amino acid sequence portions which are conserved from protein to protein, although the degree to which portions are conserved from protein to protein varies. It will be noted that in Figure 1, conserved sequences are not necessarily shown to be aligned, and this is particularly true of the "meander" sections of the amino acid sequences of the illustrated proteins. In any case, knowledge of these conserved regions permits amplification of mRNA coding for a new cytochrome P450 having at least one similarly conserved portion. That is, reverse transcription is used to amplify partial DNA sequences from subsets of mRNA.

The expression of a cytochrome P450 produced by a particular cell type often varies, depending upon the conditions to which the cells are exposed. It is this type of cytochrome P450 that is the subject of this invention. It has been reported that increases in expression due to changes in growth environment primarily reflect increases in the rates of transcription of these genes [Denison, 1995; Whitlock, 1993; Quattrochi, 1994; Savas, 1994; Sutter, 1994]. Thus, according to the present invention, mRNA is obtained from two groups of cells, the two groups being grown under different conditions from each other such that different levels of a cytochrome P450 are expressed in the two groups. Preferably, the difference in the expression is quite substantial, say at least a 3 to 5 fold difference; most preferably, the cytochrome will be expressed in one group of cells and minimally or not expressed at all in the other. Such differences in expression, being a reflection of the amount of mRNA encoding the cytochrome present in the cell means that mRNA from the two groups can be amplified and differentially displayed, as described in detail below. Such a differentially displayed nucleic acid can in turn be used as a probe of a DNA library containing the cytochrome coding sequence. The coding sequence can then be isolated and used according to conventional techniques to produce the protein itself.

It is often the case that expression of a cytochrome P450 is "inducible" by exposure of a cell to a particular agent. Figure 2 shows one such example where the "inducer" induces the transcription of a cytochrome P450, the function of which is to metabolize the "inducer", that is, the inducer is also the preferred substrate of the P450 enzyme. This is the case with the cytochrome P450RAI which is inducible by treatment of certain cells, such as F9 embryonyl carcinoma cells, and MCF7, the human breast epithelial adenocarcinoma derived cell line, with retinoic acid, and which functions to metabolize retinoic acid. This forms the auto-regulatory negative feedback loop shown in Figure 2. The net result of this negative feedback loop is to "normalize" or bring back to normal, the levels of RA in the surrounding fluid or medium. The "set point" of such a negative feedback loop is defined by the sensitivity of the

regulatory element(s) (e.g., promoter) of the cytochrome P450 gene to a given concentration of inducer.

It is also possible that a cytochrome P450 is normally produced by a cell and that such production is suppressed by exposure of the cell to a particular agent. For example, 1α-hydroxylase is expressed at high levels in the kidneys of animals suffering a deficiency of 1,25-dihydroxy vitamin D3. Both such instances are intended to be encompassed by the present invention.

Having described the invention generally, particular aspects of the invention will be described in greater detail.

Primer Selection

In certain embodiments, a first oligonucleotide primer is selected for binding to the poly(A) tail of an mRNA of a cytochrome P450 and a second oligonucleotide primer is selected for binding to a conserved region of an mRNA of a cytochrome P450. In certain embodiments, first and second primers are selected for binding to two spaced apart conserved regions of an mRNA.

At least one of the primers is based on conserved amino acid sequences selected, for example, from Figure 1. Table One lists conserved amino acid sequences (not all of which are contained in Figure 1) which can be used to derive primers.

Table One: Examples of conserved amino acid motifs in cytochrome P450 family members which can be used to design differential display primers.

Cytochrome	Organism	Motifs		Reference
CYP9B1	Drosophila	SESLRK	PERF	Dunkov, B. C. (1996)
CYP9C1	Drosophila	SESLRK	PERF	Dunkov, B. C. (1996)
CYP4E2	Drosophila	KEAQRL	PERF	Dunkov, B. C. (1996)
CYP4E4	Drosophila	KESLRL	PERH	Dunkov, B. C. (1992)
CYP4D1	Drosophila	KETLRM	PERF	Dunkov, B. C. (1992)
CYP4C3	Drosophila	KDSLRL	PDNF	Dunkov, B. C. (1992)
CYP4G1	Drosophila	LETLRL	PDNF	Dunkov, B. C. (1992)
CYP4P1	Drosophila	KETLRL	PERF	Dunkov, B. C. (1992)
CYP4C2	mosquito	KEGLRL	PDHF	Scott, J. A. (1994)
CYP4J1	mosquito	KESLRL	PDRF	Scott, J. A. (1994)
CYP4H1	mosquito	KETLRM	PERF	Scott, J. A. (1994)
CYP4D5	mosquito	KETLRL	PERF	Scott, J. A. (1994)
CYP4K1	mosquito	KESLRL	PTRF	Scott, J. A. (1994)
CYP4L1	tobacco hornworm	KESMRL	PERF	Snyder, M. J. (1995)
CYP4M1	tobacco hornworm	KESLRL	PDRF	Snyder, M. J. (1995)
bovine SCC	bovine	KETLRL	PDKF	Morohashi, K. (1984)
CYP11	human	KETLRL	PDRY	Chung, B. C. (1986)
CYP27	human	KETLRL	PESF	Cali, J. J. (1991)
CYP24	human	KESMRL	PERW	Ken K. S. (1993)
CYP26	human	KETLRL	PDRF	White, J. A. (1996)

Primers having the following sequences, for example, may thus be made based on the indicated codons corresponding to each amino acid:

KETLRL	KESMRL	Degenerate Sequence
K - AA(G/A)	K - AA(G/A)	K - AA(G/A)
E - GA(G/A)	E - GA(G/A)	E - GA(G/A)
T - AC(G/A/T/C)	S - AG(C/T)	T/S - A(G/C)(G/A/C/T)
L - AA(G/A)	M - ATG	L/M - A(G/A)(G/A/T)
R - CG(G/A/T/C)	R - CG(G/A/T/C)	R - CG(G/A/T/C)
L - AA(G/A)	L - AA(G/A)	L - AA(G/A)

Degenerate primers to correspond to any 3-amino acid subsequence:

K E (T/S)				
K E (T/S) (L/M)	E (T/S) (L/M)			
K E (T/S) (L/M) R	E (T/S) (L/M) R	(T/S) (L/M) R		
K E (T/S) (L/M) R L	E (T/S) (L/M) R L	(T/S) (L/M) R L	(L/M) R L	

Thus, for example, a set of degenerate primers for E (T/S) (L/M) R, GA(G/A)A(G/C)(G/A/C/T)A(G/A)(G/A/T)CG(G/A/T/C) (SEQ ID NO:12) can be used.

The number of oligonucleotide primers in the degenerate set necessary to be made can be reduced: (1) primers may be selected to have 45-55% GC content; (2) primers may be selected to have no more than 2 identical bases in a row; (3) primers may be designed to have no more than 2 bases of uninterrupted self complementarity within a given primer; and (4) A, G or C at the 3' end is preferable. Additionally, the nucleotide analog inosine can be included at any position where any of G/A/T/C is possible, because of inosine's apparent ability to bind to G, A, T or C [Shen, 1993].

For the design of other primers to be used in conjunction with those described above, other regions of the cytochrome P450 can also be used (ie., where conservation of amino acid sequence is observed for stretches of 2 or more amino acids). Thus, primers can be derived from other similarly conserved sequences found in cytochrome P450s, for example, in the L helix adjacent the heme binding region as shown in Table Two.

Table Two: Conserved regions of selected cytochrome P450s taken from the L helix region

Cytochrome P450	Amino Acid Sequence
Bovine SCC	GRR
hum CYP11	GRR
hum CYP27	GRR
hum CYP24	GRR

In this instance, there are 64 possible primers due to the degeneracy of the genetic code:

amino acid	G	R	R
codons	5'-GG(G/A/C/U)	CG(G/A/C/U)	CG(G/A/C/U)-3'

Thus, for example, to identify a retinoic acid inducible cytochrome P450,

differential display is performed on two samples of tissue (or cells), one sample treated with retinoic acid and one sample not treated with retinoic acid. Rather than using a random "second primer" as described by White *et al.* [White, 1996] in the isolation of P450RAI, a primer selected as described above is used as the "second primer".

In another example of the invention, cDNA encoding vitamin D3 extra-renal 1α-hydroxylase is isolated. In this instance, the inducer, cytokine γ-interferon (IFN-γ), is not a substrate of the cytochrome, a different situation from that of the retinoic acid induced cytochromes. According to this aspect of the invention, macrophages are cultured with or without the cytokine inducer. Differential display is performed using either the KET or GRR second primer sets shown above with an oligo(dT) first primer.

PCR conditions for a selected Primer To design a set of oligonucleotide primers for the peptide KETL, for example, the degenerate sequence is: AA(G/A)GA(G/A)AC(G/A/C/T)AA(G/A) (SEQ ID NO:13). Using the four parameters described above, this becomes AAGGAGC(G/A/T/C)AAG. Once a primer sequence has been selected, the melting temperature of a DNA duplex comprising this sequence is determined, where T_m in degrees Celsius is 4x(# of GC pairs) + 2x(# of AT pairs). For this example, this is between 4x(6) + 2x(6) = 36°C and 4x(5) + 2x(7)= 34°C. Once the T_m is determined, the PCR conditions are selected accordingly. PCR includes a denaturing step which, optimally, separates all the template DNA into single strands, for example a 1 minute incubation at 94°C. This is followed by an annealing step, which is close to the T_m of the oligonucleotide primer and for this example is about 40°C for 1 minute. This is followed by an extension step at 72°C to allow for the polymerase (preferably *Taq* DNA polymerase) to extend the PCR product. These steps are repeated a number of times, for example 35 times.

Example 1

By way of illustration of particular aspects of the present invention, isolation of cytochrome P450RAI from zebrafish caudal fins is described. It will be understood that alternative primers selected as described above and in Example 2 could well have been used.

Figure 3 outlines the steps used to isolate retinoid-regulated genes using differential display of mRNA. The cloned products isolated in step 6 of Figure 3 were used for sequencing and screening of *Danio rerio* (*D. rerio*) cDNA libraries. P1, P2 and P3 correspond to fragments from RA induced mRNAs. P4 is a PCR product from a down-regulated mRNA. Details of procedures followed in determination of gene sequences described herein follow.

***Danio rerio* stocks**

D. rerio were kept at 28.5°C in 40 L tanks with 25-30 fish per tank on a 14 hour light-10 hour dark cycle. Tap water was conditioned by the addition of 10 ml of Water Conditioner (Sera Aqutan) and 10 ml of 250 g/L Aquarium Salt (Nutra Fin) per 20 L. 2-3 L of water was changed daily. Amputation of fins was carried out following anaesthetization of the

fish in a solution of 0.2% ethyl-m-aminobenzoate methanesulfonic acid (ICN) in conditioned water. Retinoic acid treatment was performed by adding all-*trans* RA, to a final concentration of 10^{-6} M, directly into the tank water two days following amputation. Both control- and RA-treated fish were kept in the dark during the experiments.

Differential Display of mRNAs

Differential mRNA display was performed essentially as described by Liang and Pardee (1992) with appropriate modifications as described herein. Regenerating tissues were collected 3 days post-amputation (24 hours post-RA addition) and quick frozen in liquid nitrogen. Poly $(A)^+$ RNA was isolated using the Micro Fast-Track kit. Duplicate independent reverse transcription reactions were performed on the isolated poly$(A)^+$ RNA from both the treated and untreated samples for each specific 3' poly-T primer used (5'-T_{12}VN-3'). The symbol "V" represents A or C or G and not Tor U. Several combinations of the 3' poly-T primers given in the first column of Table Three and the upstream primers given in the second column were utilized for PCR amplification. For each reaction 0.1 µg poly$(A)^+$ RNA was reverse transcribed in a 20 µl reaction volume containing 300U Superscript Reverse Transcriptase (Gibco/BRL), 1X Buffer, 20 µM each dGTP, dATP, dCTP and dTTP, 10 µM dithiothreitol (DTT) and 5 pmol of 5'-T_{12}VN-3' primer. The reactions were mixed and incubated at 35°C for 60 minutes, followed by 5 minutes at 95°C. PCR amplification was performed in a Perkin Elmer Cetus PCR machine as follows: 1 µl cDNA synthesis reaction, 5U *Taq* DNA polymerase (Gibco/BRL), 1X PCR Buffer, 2 µM each dGTP, dATP, dCTP and dTTP, 10 µCi α-[^{35}S]dATP (*redi*vue, Amersham) 1.2 mM $MgCl_2$, 0.5 µM upstream primer and 0.5 µM of the corresponding 5'-T_{12}VN-3' primer. PCR conditions were as follows: 1 cycle, 94°C for 5 minutes; 40 cycles, 94°C for 30 seconds, 42°C for 1 minute, 72°C for 30 seconds; followed by a final extension of 5 minutes at 72°C. 4 µl of the PCR reactions were loaded onto a 6% non-denaturing polyacrylamide gel and electrophoresed at 60 watts, 45°C. The gel was dried and exposed for 12 to 24 hours on Kodak XAR film at room temperature.

Table Three	Sequences of the downstream Poly (T) oligonucleotides for the differential display procedure.
3'- Poly(T) primers:	5'-degenerate primers:
5'-TTT TTT TTT TTT GG-3' 5'-TTT TTT TTT TTT GA-3' 5'-TTT TTT TTT TTT GT-3' 5'-TTT TTT TTT TTT GC-3' 5'-TTT TTT TTT TTT AG-3' 5'-TTT TTT TTT TTT AA-3' 5'-TTT TTT TTT TTT AT-3' 5'-TTT TTT TTT TTT AC-3' 5'-TTT TTT TTT TTT CG-3' 5'-TTT TTT TTT TTT CA-3' 5'-TTT TTT TTT TTT CT-3' 5'-TTT TTT TTT TTT CC-3'	5'-AAG CGA CCG A-3' 5'-TGT TCG CCA G-3' 5'-TGC CAG TGG A-3' 5'-GGC TGC AAA C-3' 5'-CCT AGC GTT G-3'

In Table Three, the sequences in the first column are identified as SEQ ID NOs: 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25, respectively. The sequences in the second column are identified as SEQ ID NOs:26, 27, 28, 29 and 30, respectively.

Gel Purification and Reamplification

Bands demonstrating reproducible differential amplifications (see Figure 4a) were found for the upstream-downstream primer combination of 5'-TGCCAGTGGA-3'- poly-T primer, 5'-TTT TTT TTT TTT AG-3' (SEQ ID NOs: 28 and 18, respectively). These bands were excised from the gel by overlaying the X-ray film and cutting out the corresponding piece of dried gel and filter paper. The PCR product corresponding to a fragment of the protein described herein was isolated from the band in Figure 4(a). Samples were placed in 100 µl of nuclease free water, incubated for 10 minutes at room temperature, then boiled for 15 minutes. The supernatant was recovered following a 15 minute centrifugation at 12,000 x g.

In order to facilitate cloning of the PCR products, several changes were made to the reactions. Primers which included EagI restriction endonuclease sites were used in the reamplification. Based on results obtained in the differential display analysis, the upstream 5'-TGCCAGTGGA-3' primer was replaced by 5'-GTAGCGGCCGCTGCCAGTGGA-3' (SEQ ID NO:31) and the downstream poly-T primer, 5'-TTT TTT TTT TTT AG-3', was replaced by 5'-GTAGCGGCCGCT_{12}-3' (SEQ ID NO:32). In addition, the reaction volume was increased to 40 µl, isotope was omitted and 20 as opposed to 40 cycles were performed. 5 µl aliquots of the PCR reactions were removed and the products were visualized by electrophoresis in a 1% agarose gel followed by ethidium bromide staining and UV illumination.

Cloning PCR Products

The reamplified products were purified by phenol/chloroform extraction followed by ethanol precipitation. The resulting DNA pellet was resuspended in 17 µl of sterile

water and digested at 37°C for 1 hour by the inclusion of 10U EagI (New England Biolabs), and 1X NEB 3 buffer. EagI restriction endonuclease was heat inactivated by incubation at 65°C for 20 minutes. pBluescript SK$^+$ vector was prepared by digestion with EagI, followed by dephosphorylation using calf intestinal alkaline phosphatase (CAP, Promega). Restriction digests were purified using the GeneClean II Kit (Bio 101) following electrophoresis in a 1% agarose gel. In a total ligation volume of 10 µl, 2 µl of digested PCR product, 1 µl digested SK$^+$, 1U T4 DNA ligase (Gibco/BRL) and 1X buffer were incubated at 16°C overnight. *E. coli* bacterial strain JM109 was transformed with 1 µl of the ligation product using a BioRad Gene Pulser, then plated on LB+ampicillin plates and incubated overnight at 37°C.

Colony Selection

Individual colonies were transferred in duplicate to fresh LB plates and grown overnight at 37°C. Colonies were tranferred to nitrocellulose membrane and denatured in a solution of 1.5M NaCl, 0.5M NaOH for 5 minutes, neutralized in 1.5M NaCl, 0.5M Tris-HCl, pH 8.0 for 5 minutes, followed by two 5 minute washes in 2X SSC. Membranes were then UV cross-linked (Stratalinker UV Crosslinker, Stratagene). Prehybridization and hybridization were performed using Quickhyb (Stratagene) following the manufacturer's directions. Each colony lift was probed with the corresponding PCR product isolated during the gel reamplification and purification step. α-[^{32}P]-dATP labelled probes were generated using the Prime-It Kit II (Stratagene). Subsequent to hybridization, filters were washed twice for 20 minutes in 2X SSC, 0.1% SDS solution at room temperature and exposed to Kodak X-omat autoradiography film overnight at -70°C. Positive colonies were selected from the duplicate plates, grown overnight in LB+ampicillin (100 µg/ml) and plasmid DNA isolated using the Qiaprep Spin Plasmid Kit (Qiagen).

Clones were sequenced using the T7 Sequencing Kit (Pharmacia Biotech). Sequence comparisons were generated using the GeneWorks software package (Intelligenetics).

Screening of a *D. rerio* cDNA Library

A random primed *D. rerio* 6-18 hour embryo cDNA library constructed in Uni-ZAP II (Stratagene) was produced. 4.5 x 10^5 independent pfu were screened using the random primed, α-[^{32}P]-dATP labelled 337 bp PCR fragment isolated by mRNA differential display as a probe. Filters were prehybridized for 1-4 hours at 42°C in 50% formamide, 5X SSPE, 1X Denhardt's solution, 0.2 mg/ml denatured salmon sperm DNA. Hybridization was performed at 42°C by adding denatured probe to the prehybridization solution. Filters were washed two times for 20 minutes in 2X SSC, 0.05% SDS at room temperature and exposed to Kodak XAR film overnight at -70°C. Positive plaques were picked into 500 µl SM buffer and subjected to additional rounds of rescreening until purified. Positive plaques were exposed to the *in vivo* excision protocol following the manufacturer's directions (Stratagene). pBluescript containing colonies were plated onto LB+amp plates and grown overnight at 37°C. Sequence data were

generated using the T7 Sequencing Kit (Pharmacia) and analysed using the GeneWorks software package (Intelligenetics).

Whole Mount *in situ* hybridization

RA- and DMSO-treated regenerates were isolated 72 hours post-amputation (24 hours post RA/DMSO addition), washed in PBS and prepared for whole mount *in situ* hybridization. *In situ* hybridizations were undertaken as previously described [White, 1994].

Northern Blot Analysis

Fish were allowed to regenerate their caudal fins for 72 hours. At 48 hours 10^{-6}M all-*trans* RA in DMSO vehicle or DMSO alone was added directly to the tank water. mRNA was prepared using the Micro Fast-Track mRNA isolation kit (Invitrogen, CA) according to the manufacturer's directions. 3.0-5.0 µg poly A$^+$ RNA was electrophoresed, blotted and probed using a previously described method [White, 1994] with the full length zP450RAI cDNA obtained as described below. Ethidium bromide stained agarose gel showed that equivalent amounts of mRNA were used in the blotting experiments. See lanes 2 and 3 of Figure 5(a).

HPLC Analysis

Media from transfected cells incubated with 575 pM [11,12-^{3}H]RA (Figures 6(a) and 6(b)) or 1 µM RA (Figures 6(c) and 6(d)) for either 4 hrs (Figures 6(a) and 6(c)) or 24 hrs (Figures 6(b) and 6(d)) were acidified with 0.1% acetic acid. Lipid soluble metabolites were separated from aqueous soluble metabolites using a total lipid extraction of the medium [Bligh, 1957]. Metabolism of [11,12-^{3}H]RA to total aqueous soluble metabolites was measured using aliquots of the aqueous soluble extract subjected to ß-scintillation counting (See the insets of Figures 6(a) and 6(b)). Lipid soluble extracts were evaporated to dryness under a stream of nitrogen and resuspended in 93.5/5/1/0.5 hexane/isopropanol/methanol/acetic acid (H/I/M/AA). Metabolites were separated by HPLC using a Zorbax-SIL (3µ, 8 x 0.62 cm) column eluted with a solvent system of 93.5/5/1/0.5 H/I/M/AA at a flow rate of 1 ml/min.

Characterization of a novel cytochrome P450

Transcripts present in fin tissue regenerating in the presence or absence of RA were compared using the differential display PCR technique developed by Liang and Pardee [Liang, 1992] (Fig 4(a)). One of the differential display products which exhibited a dependence on the presence of RA for its expression, indicated by the arrow in Figure 4(a), was isolated and sequenced. The sequence is identified as SEQ ID NO:33 and is also shown in Figure 4(b). The amino acid sequence corresponding to the cDNA, termed here, "zP450RAI", is shown in Figure 4(c) and identified as SEQ ID NO:34. BLAST search analyses revealed sequence homology between zP450RAI and multiple members of the cytochrome P450 superfamily. Alignments between zP450RAI cDNA deduced amino acid sequence and those of other cytochrome P450s indicated that zP450RAI exhibited less than 30% overall amino acid identity

with members of previously defined subfamilies [Nelson, 1993]. zP450RAI contains many of the structural motifs which are common to cytochrome P450 family members, including the heme-binding domain located in the C-terminal portion of the protein. See Figure 4(d). The P450RAI family has been designated "CYP26".

Cell specific induction of zP450RAI by all-*trans* RA

Northern blot analysis of mRNAs expressed in regenerate tissue isolated from control (dimethyl sulfoxide-treated) and RA-treated fish was performed with a full-length zP450RAI cDNA probe. zP450RAI transcripts were not detectable in regenerate tissue from control fish (Figure 5(a), lane 4) but were very noticeably present in tissues isolated from fish exposed to RA for 24 hours (Figure 5(a), lane 5).

Whole mount *in situ* hybridization was used to determine the cellular localization of zP450RAI expression in regenerating fin tissue. Figure 5(b) shows regenerating fins from control and RA-treated fish. zP450RAI transcripts are not detectable in control fin tissue (Figure 5(b)(i)). In regenerating tissue from RA-treated fish, zP450RAI transcripts were found to be abundant in a layer of epithelial cells extending across the distal edge of the wound epithelium as indicated by the black arrowhead in Figure 5(b)(ii). Some low level staining was also observed in inter-ray tissue as indicated by the black line with arrowhead in Figure 5(b)(ii). A histological section of an RA-treated fin, taken along the line shown in Figure 5(b)(iii), is shown in Figure 5(b)(iv). The section indicates that cells expressing zP450RAI are located deep within the epithelial layer at the distal tip of the blastemal mesenchyme. Whole mount *in situ* hybridization thus illustrates the usefulness of nucleic acid probes of the invention for the localization of cytochrome P450RAI mRNA in whole tissue.

Metabolism of all-*trans* RA by zP450RAI transfected cells

Retinoic acid as a substrate of zP450RAI was studied. The full-length zebrafish zP450RAI cDNA was cloned into the eukaryotic expression vector pSG5 [Green, 1988]. COS-1 cells were transiently transfected with either pSG5 or pSG5-zP450RAI and then incubated with either picomolar concentrations of [11,12-^{3}H]all-*trans*-RA or micromolar concentrations of non-radioactive all-*trans*-RA. COS-1 cells are an African green monkey kidney "fibroblast-like" cell line. zP450RAI expression in COS-1 cells promoted the rapid conversion of RA into both lipid- and aqueous-soluble metabolites. See Figures 6(a) and 6(b). Fractions of total lipid extracts of transfected cells were initially separated by normal-phase HPLC on Zorbax-SIL. Comparison between extracts from pSG5 and pSG5-zP450RAI-transfected cells indicated that zP450RAI significantly increased RA metabolism. Incubation of zP450RAI-transfected cells with 575 pM [11,12-^{3}H]all-*trans*-RA for either 4 or 24 hours resulted in accumulation of RA metabolites, one of which co-migrated on a column with synthetic standards 4-OH-RA and 18-OH-RA, and a second slightly less polar metabolite which co-migrated with 4-oxo-RA standard (Figures 6(a) and 6(b)). Rechromatography of RA metabolites using other HPLC systems confirmed the identity of these two metabolites as 4-

OH-RA and 4-oxo-RA (Table Four). It is possible that the aqueous-soluble radioactivity represents glucuronides of RA metabolites or glucuronides of RA itself. Rapid glucuronidation of 4- and 18-hydroxy-RA in mammalian cell extracts has been reported by others [Wouters, 1992; Takatsuka, 1996].

Table Four Chromatographic properties of RA metabolites.

Metabolite	Retention Time (min)		
	Z-Sil[a]	Z-CN[b]	Z-ODS[c]
RA (std)	2.57	4.47	19.92
4-oxo-RA (std)	4.79	11.33	11.73
4-OH-RA (std)	5.17	9.65	12.65
18-OH-RA (std)	5.06	9.53	14.03
Peak 1 (RA)	2.57	4.48	19.73
Peak 2 (4-oxo-RA)	4.87	11.38	11.57
Peak 3 (4-OH-RA)	5.16	9.68	12.68

[a]HPLC conditions: Zorbax-SIL column eluted with 93.5/5/1/0.5 H/I/M/A.A. (1 ml/min)
[b]HPLC conditions: Zorbax-CN column eluted with 93.5/5/1/0.5 H/I/M/A.A (1 ml/min)
[c]HPLC conditions: Zorbax-ODS column eluted with a 20 min linear gradient with solvent containing 10 mM ammonium acetate which ranged from 55.45 to 5.95 H_2O /MeOH (2 ml/min).

A similar pattern of zP450RAI-dependent metabolism was also observed using a much higher RA concentration (1 μM). zP450RAI-transfected COS-1 cells incubated for 4 or 24 hours with 1 μM RA generated two closely-running peaks which were discernible in a 350nm HPLC trace shown in Figures 6(c) and 6(d), but which were essentially undetectable in control pSG5-transfected cells (See the insets of Figures 6(c) and 6(d)). These peaks co-migrated with those of 4-oxo-RA and 4-OH-RA standards, respectively. Diode array spectrophotometric detection of the zP450RAI-generated peaks showed that the spectral properties of the two metabolite peaks matched the standard retinoids [In hexane-based solvents: 4-OH-RA, λ_{max}=350nm; 4-oxo-RA, λ_{max}=355nm; in methanol-based solvents: 4-OH-RA, λ_{max}=340nm; 4-oxo-RA, λ_{max}=360nm].

Example 2

Treatment of cells The extra-renal 1-α hydroxylase (1α-OHase) is isolated using differential display using human pulmonary macrophages obtained from lavage. These cells are placed in culture [Adams, 1985] and treated with γ-interferon [Reichel, 1987a] or lipopolysaccharide [Reichel, 1987b] to induce the extra-renal 1α-OHase activity.

Differential Display of mRNAs Differential mRNA display is performed essentially as described by Liang *et al.* [Liang, 1992] with modifications described as follows. Cells from cultures treated as described above are harvested. Poly (A)+ RNA is isolated from the samples using the Micro-Fast Track kit (Invitrogen, CA, USA). Reverse transcription reactions are performed in duplicate with isolated poly (A)+RNA from both the treated (i.e., plus interferon) and control samples for each specific 3' oligo(dT) primer used (described as primer

1). For each reaction 0.1 µg poly (A)+RNA is reverse-transcribed in a 20 µl reaction volume containing 300U Superscript Reverse Transcriptase (Gibco/BRL), 1X Buffer, 20 µM each dGTP, dATP, dCTP and dTTP, 10 µM dithiothreitol and 5 pmol of 5'-T_{12}VN-3' primer. The reactions are mixed and incubated at 35°C for 60 minutes followed by 5 minutes at 95° C. PCR amplification is performed in a Perkin Elmer Cetus PCR machine as follows: 1 µl cDNA synthesis reaction, 5U *Taq* DNA polymerase (Gibco/BRL), 1X PCR Buffer, 2 µM each dGTP, dATP, dCTP and dTTP, 10 µCi α-[^{35}S]dATP (*redi*vue, Amersham) 1.2 mM $MgC1_2$, 0.5 µM upstream primer (5'-AA(G/A)GA(G/A)AC(G/A/C/T)AA(G/A)-3') and 0.5 µM of the corresponding 5'-T_{12}NN-3' primer. PCR conditions will vary depending on the primer set used. Typical conditions are as follows: 1 cycle, 94° C for 5 minutes; 40 cycles, 94° C for 30 seconds, 42° C for 1 minute, 72° C for 30 seconds; followed by a final extension of 5 minutes at 72° C. Aliquots (4 µl) of the PCR reactions are loaded onto a 6% non-denaturing polyacrylamide gel and electrophoresed at 60 watts, 45° C. The gel is dried and exposed to Kodak XAR film for 12 to 24 hours at room temperature.

Gel Purification and Reamplification PCR products corresponding to differentially regulated cytochrome P450 genes are isolated from the gel by overlaying the X-ray film and excising the corresponding piece of dried gel and filter paper. Samples are placed in 100 µl of nuclease free water, incubated 10 minutes at room temperature, then boiled for 15 minutes. The supernatant is recovered following a 15 minute centrifugation at 12,000 x g.

A reamplification round of PCR is next performed. In order to facilitate cloning of the PCR products, several alterations are made to the PCR reactions. Primers which include EagI restriction endonuclease sites (CGGCCG) are used in a reamplification step. For example, the upstream 5'-AA(G/A) GA(A/G) AC(G/A/C/U)-3' primer is replaced by 5'-GTAGCGGCCGCAA(G/A) GA(A/G) AC(G/A/U)-3' (SEQ ID NO:35) and the downstream poly-T primer is replaced by 5'-GTAGCGGCCGC$(T)_{12}$-3' (SEQ ID NO:36). (In alternative embodiments of the invention, these primers could correspond to EcoR1 sites (GAATTC) or BgII sites (AGATCT) and so on, in which cases cloning into the corresponding EcoR1 or BgII or other restriction site would be facilitated.) In addition, the reaction volume is increased to 40 µl, isotope is omitted and only 20 cycles are performed. 5 µl aliquots of the PCR reactions are removed and the products are visualized by electrophoresis in a 1% agarose gel followed by ethidium bromide staining and UV illumination.

Cloning PCR Products The reamplified products are purified by phenol/chloroform extraction followed by ethanol precipitation. The resulting DNA pellet is resuspended in 17 µl of sterile water and digested at 37° C for 1 hour with 10U EagI (New England Biolabs) in 1XNEB 3 buffer. EagI is heat inactivated by incubation at 65°C for 20 minutes. pBluescript SK+ vector is prepared by digestion with EagI, followed by dephosphorylation using calf intestinal alkaline phosphatase (CIP, Promega). Restriction digest products are purified using the GeneClean II Kit (Bio 101) following electrophoresis in a 1% agarose gel. In a total ligation volume of 10 µl, 2

µl of digested PCR product, 1 µl digested SK+, 1U T4 DNA ligase (Gibco/BRL) and 1X ligation buffer are incubated at 16° C overnight. *E coli* bacterial strain JM109 is transformed with 1 µl of the ligation products, using the BioRad Gene Pulser, then plated on LB+ampicillin plates and incubated overnight at 37° C.

Colony Selection Individual colonies are transferred in duplicate to fresh LB plates and grown overnight at 37° C. Colonies are transferred to nitrocellulose membranes and denatured in a solution of 1.5 M NaCl, 0.5 M NaOH for 5 minutes, neutralized in 1.5 M NaCl, 0.5 M Tris-HCl, pH8.0 for 5 minutes, followed by two 5 minute washes in 2X SSC. Membranes are then UV cross-linked (Stratalinker UV Crosslinker, Stratagene). Prehybridization and hybridization are performed using Quickhyb (Stratagene) following the manufacturer's directions. Each colony lift is probed with a PCR product isolated during the gel purification and reamplification step. α-[^{32}P]-dATP labeled probes are generated using the Prime-It Kit II (Stratagene). After hybridization, filters are washed twice for 20 minutes in 2X SSC, 0.1% SDS solution at room temperature and exposed to Kodak X-ornat autoradiography film overnight at -70° C. Colonies hybridizing with the PCR probe are selected from the duplicate plates, grown overnight in LB-ampicillin (100 µg/ml) and plasmid DNA is isolated using the Qiaprep Spin Plasmid Kit (Qiagen).

cDNA inserts are sequenced using the T7 Sequencing Kit (Pharmacia Biotech). Sequence comparisons are generated using the GeneWorks software package (Intelligenetics).

Northern Blot Analysis To confirm the inducible nature of the cytochrome p450 encoded by the identified PCR product, northern blot analysis is performed using mRNA prepared using the Micro Fast-Track mRNA isolation kit (Invitrogen, CA) according to the manufacturer's directions. 3.0-5.0 µg poly (A)+RNA is electrophoresed, blotted and probed with the PCR product as previously described [White, 1994].

Screening of a cDNA Library To obtain the full length cDNA corresponding to the isolated PCR product an appropriate genomic or cDNA library, and preferably a cDNA library, will be screened. If a suitable cDNA library is not available, it can be constructed as described previously [White, 1994] from tissue or cells used for differential display. Typically, 4.5 x 10^5 independent plaques are screened using the random-primed, α-[^{32}P]-dATP labeled PCR fragment. Filters are prehybridized for 1-4 hours at 42°C in 50% formamide, 5X SSPE, 1X Denhardt's solution, 0.2 mg/ml denatured salmon sperm DNA. Hybridization is performed overnight at 42°C by adding denatured probe to the prehybridization solution. Filters are washed two times for 20 minutes in 2X SSC, 0.05% SDS at room temperature and exposed to Kodak XAR film overnight at -70°C. Positive plaques are picked into 500 µg SM buffer and rescreened until purified. pBluescript-containing colonies, generated using the *in vivo* excision protocol, are plated onto LB+amp plates and grown overnight at 37°C. Sequence data are

generated using the T7 Sequencing Kit (Pharmacia) and analyzed using the GeneWorks software package (Intelligenetics).

Example 3

In the murine embryonal carcinoma cell lines (F9 and P19), a retinol metabolizing activity, inducible by retinoic acid, has recently been detected [Achkar, 1996]. We have found that the retinoic acid inducible retinoic acid metabolizing cytochrome P450, P450RAI, does not metabolize retinol, suggesting that there is an as yet unidentified cytochrome P450 responsible for the metabolism of retinol in these cells. To isolate the cDNA for this cytochrome, F9 embryonal carcinoma cells are cultured in the presence or absence of 10^{-6}M retinoic acid for the purpose of performing differential display. The RNA is isolated from these cells after a 24 hour incubation period using the methods described in Example 2. Similarly, the differential display is used with primer sets described above. For example, one could use an oligo(dT) primer as a first primer and the second primer could be based on the amino acid sequence KETLRL (5'-AAGGAGACCCTTCGAC-3' (SEQ ID NO:51); as an example of one primer of a degenerate mix). Primers based on other conserved nucleotide sequences could also be used. In this way, eventually all known possibilities of conserved sequences at a given site can be tested. For example, the next group of primers could be based on the nucleotide sequence encoding KESMRL, KESLRL, or KETLRM, and so on. Once candidate PCR products have been selected on the basis of inducibility and size (predicted by the choice of primers) DNA sequencing is performed on individual clones so that unique cytochrome P450- encoding PCR products are selected. Those selected are used as probes for screening a cDNA library prepared from F9 or P19 cells treated with retinoic acid (Stratagene) to obtain cDNAs encompassing the open reading frame of the corresponding probe. Preferably, a cDNA corresponding to the complete open reading frame is obtained. Alternatively, full length cDNAs can be obtained using rapid amplification of cDNA ends (RACE) [Innis, 1990, which reference is incorporated herein by reference](Kit from Clonetech) using the RNA isolated from the treated F9 cells described above for the PCR reactions.

The full length cDNAs are expressed as proteins in mammalian cells such as HeLa or COS 1 cells to determine the function of these cytochrome P450s. These experiments thus confirm the metabolic activity of the cytochrome proteins, allowing the identification of those having retinol metabolizing activity, as previously described [White, 1996].

Example 4

Cloning of Human P450RAI

This example illustrates use of a nucleic acid sequence identified in one species in the identification of an ortholog from another relevant species. Here, the amino acid sequence corresponding to the DNA of zebrafish P450RAI (zP450RAI) (SEQ ID NO:34) was used to search an express sequence tag (EST) database. It is, of course, possible to use a nucleic acid molecule based on the cytochrome P450 identified from a first species as a probe

of a cDNA library of a second species in order to identify an ortholog. A commercially available EST clone (SEQ ID NO:37) having a high degree of homology with a C-terminal portion of the zP450RAI (from Glu 292 to Phe 410 of SEQ ID NO:34 was purchased (Research Genetics, Huntsville, AL). The clone is reportedly from a human infant brain cDNA library (Bento Soares and M. Fatima Bonaldo) and is apparently otherwise unpublished. The purchased clone was sequenced using the T7 sequencing kit (Pharmacia) and sequence data was generated using the Geneworks Software Package (Intelligenetics).

A cDNA library generated from an NT2 cell line treated with retinoic acid is commercially available (Stratagene, cat#939231) and this product was used for further studies. The cDNA library was probed with a nucleic acid having a sequence identified as SEQ ID NO:37. Eleven positively hybridizing clones were isolated and purified according to the manufacturer's directions. Sequence data for these clones were generated using the T7 Sequencing Kit (Pharmacia) and analyzed using the Geneworks package (Intelligenetics). The human DNA sequence is identified as SEQ ID NO:38 and the corresponding polypeptide as SEQ ID NO:39.

Example 5

Transient Tranfection Analysis

Once a nucleic acid encoding a cytochrome P450 has been isolated, it can be expressed in cultured cells. In this particular example, the nucleic acid encoding human P450RAI was used in the transfection of COS-1 cells.

COS-1 cells were subcultured 20 hours prior to transfection which was carried out according to the standard DEAE-dextran method [Sambrook, 1989; Maniatis, 1982]. Cells were transfected with pE-AR (adrenodoxin expression vector, 1 µg/P100 plate) and pE-ADX (adrenodoxin reductase expression vector, 1 µg/P100 plate) together with 3 µg per plate of either pSG5 (control) or hP450RAI-pSG5 (experimental). [11,12-^{3}H]all-*trans* retinoic acid (60,000 cpm per plate) was added 24 hours after transfection. Analyses were carried out as described in Example 3 and results obtained are shown in Figures 9 and 10(a) to 10(d). As indicated in the Figures, hP450RAI expression in COS-1 cells promoted conversion of RA into 4-OH-RA and 4-oxo-RA. Total amounts of 4-oxo-RA and 4-OH-RA produced in the transfected cells in comparison to amounts produced in the control cells are shown in Figures 10(a) and (b), respectively. Overall, greater amounts of aqueous soluble metabolites were produced in the transfected cells (Figure 10(c)) and greater amounts of unmetabolized RA were found in control cells (Figure 10(d)).

The clone sequence (SEQ ID NO:37) was prepared as a ^{32}[P]-dATP labeled probe to study the inducibility of hP450RAI by RA in several cell lines: HEK293; EL-E; HL-60; MCF10A; LC-T; SK-LC6; MCF7; U937; HepG2; NT2 (See Figures 7, 8 and 11). As can be seen, a variety of expression patterns were observed.

The ^{32}[P]-dATP labeled probe was also used to study hP450RAI mRNA expression in a human acute promyelocytic leukemia cell line. Experiments were carried out

using the NB4 cell line, isolated from a human acute promyelocytic leukemia patient, and three retinoic acid resistant cell lines were independently derived from NB4. Results are shown in Figure 12. As can be seen, the normal cells expressed hP450RAI mRNA after treatment with 10^{-6}M RA, while such expression appeared to be absent for the other cell lines both in the absence and presence of RA.

Analysis of metabolites of MCF10A and MCF7 cell lines exposed to RA was carried out, MCF10A cells having displayed no expression of mRNA and latter having displayed a large dependence of mRNA expression on exposure to RA. The results are shown in Figures 13(a) to 13(c). Consistent with the results shown in Figure 7, the results shown in Figure 13(a) indicate there was little difference in the lipid soluble activity profiles of the MCF10A cell line exposed to RA and the control. The last two bars of Figure 13(c) indicate that total aqueous soluble metabolites were about the same for both the induced and control MCF10A cells. As indicated in Figure 13(b), the MCF7 cell line exposed to RA had an elution profile which indicated significantly greater concentrations of 4-OH-RA and 4-oxo-RA than the same cell line not exposed to RA. Figure 13(c) indicates that the amount of total aqueous soluble metabolites of the MCF7 cells exposed to RA was much greater than that for the control cells.

Example 6

Generation of a Stable Cell Line using P450

This example illustrates generation of a stable cell line expressing a cytochrome P450. In this case, the human P450RAI was expressed in HeLa cells.

For expression in HeLa cells, the human cytochrome P450RAI cDNA (SEQ ID NO:38) was inserted into the XhoI-NotI sites of the multiple cloning site of the Epstein-Barr virus- based vector pCEBV7 [Wilson, 1995]. Stable transfection was performed via the calcium phosphate method [Sambrook, 1989]. Prior to the day of transfection, HeLa cells were seeded at 3.0 X 10^6 cells per 100 mm plate. Approximately 12 µg of DNA were transfected per plate and triplicate plates were employed for the transfection. Selection using hygromycin B began three days after the transfection and continued for approximately three weeks until the development of foci on the plates. The concentration of hygromycin B (100 µg/ml) was chosen for selection of cells with high expression of the construct. A killing curve was determined prior to selection which showed that 50 µg/ml of hygromycin was sufficient to kill 50% of the cells in 4 days. Confirmation of the selected HeLa cells expressing the sense construct was determined by Northern blot analysis and probing with full length hP450RAI cDNA (data not shown).

Microsomes can be prepared in the case of microsomal P450s such as human P450RAI, for example. Here, microsomes were prepared from HeLa cells transfected with the pCEBV7 alone (HeLa P) or from the HeLa cells expressing the P450RAI (HeLa RAI) and exposed to radiolabelled RA for ninety minutes. The results are shown in Figures 14(a) and 14(b), in which it can be seen that when microsomes prepared from these cells are incubated with RA, only those from HeLa RAI showed any conversion of retinoic acid to 4-hydroxy-retinoic acid or 4-oxo-retinoic acid .

Example 7

Regulation of P450 transcription - Cloning of P450 Promoters

Given a nucleotide sequence encoding a protein, it is generally possible to identify sequences which regulate expression of the sequence, such as a promoter. Here, two examples are given, which illustrate the isolation of sequences having promoter activity in the expression of human P450RAI and mouse P450RAI.

Cloning Human P450RAI promoter. The full length hP450RAI cDNA was used as a probe to identify PAC (P1 artificial chromosome) clones which contain the hP450RAI gene. cDNA from hP450RAI was sent to Canadian Genome Analysis and Technology Program at the Hospital for Sick Children in Toronto, Ontario, Canada for screening of PAC libraries. 5 PAC clones were obtained from this screening, which were verified to contain the hP450RAI gene by restriction enzyme digestion and Southern blotting using the full length hP450RAI cDNA as a probe. One of these clones, 245C7, was found to hybridize to an N-terminal probe from hP450RAI. The probe used was an approximately 130 bp NotI fragment generated from the hP450RAI cDNA. Digestion and Southern blotting of clone 245C7 identified an approximately 3.5 Kb BamHI fragment which hybridized with the NotI fragment. This fragment was subcloned into the plasmid SK+ and sequence data generated at the Core Facility for Protein/DNA Chemistry at Queen's University, Kingston, Canada. Comparison of the sequence data generated with the hP450RAI cDNA identified this 3.5 Kb clone as containing the potential initial methionine and approximately 675 bp upstream (5').

Cloning of mouse P450RAI promoter. A clone of mouse P450RAI genomic DNA approximately 17 kb long was isolated from an SV129 λDASH library and subcloned into SK. DNA prepared from this plasmid was digested with various restriction endonucleases, electrophoresed on an agarose gel, and Southern blotted onto nitrocellulose. The resulting blot was hybridized with a ^{32}P-labelled 230 base pair probe from the N-terminal region of a mouse P450RAI cDNA clone. A SacI fragment approximately 520 base pairs in length was found to hybridize strongly to the probe. This fragment was subcloned into SK cleaved with SacI. Sequence analysis revealed the presence of a DR5-type RARE in the proximal promoter. Flanking this RARE were two BssHII sites 193 base pairs apart. This 193 base pair BssHII fragment was subcloned into the MluI site of pGL3B. Diagnostic restriction digests with XhoI were done to isolate clones with the BssHII fragments in both forward and reverse orientations. Within the 193 base pair fragment is also found the TATA box, 45 base pairs downstream of the RARE. Also included in the 193 base pair BssHII fragment is 83 base pairs of DNA upstream of the DR5 RARE.

A promoter sequence for the mouse P450RAI was also determined. The human, mouse and fish are identifed as SEQ ID NOs:48, 49 and 50, respectively.

Example 8

Transfection with P450 Promoter

Given a nucletotide sequence having promoter activity, a person skilled in the

art is able to transfect cells with the promoter and thereby induce expression of a structural gene operatively linked with the sequence so as to regulate expression of the gene. The gene itself might naturally be regulated by the regulatory sequence, or another gene(s) or nucleic acid sequence might be linked to the promoter so as to have its expression under its control.

In this case, a 195 basepair fragment of genomic DNA containing a portion of the putative promoter for mP450RAI was cloned into the luciferase vector, pGL3B (Promega). For analyses of promoter activity, HepG2 cells were transfected in 6-well dishes with 2 µg BssHI-pGL3B sense or antisense constructs, described in Example 10, using 5 µl lipofectAMINE reagent (Gibco, BRL).

On the first day, 48 hours prior to transfection, cells were replated into 6-well plates, with 2.5-3 mls Minimal Essential Medium (MEM) (Gibco, BRL)+ 10% Fetal Calf Serum (Gibco, BRL). On the third day, cells are generally about 50% confluent (HepG2). Before beginning transfection, the medium was replaced with fresh medium and the cells were allowed to grow while preparing DNA/lipofectamine mix. In individual wells of a 48 well plate were mixed 1-5 µg DNA in 100 µl optiMEM (Gibco, BRL) and 5 µl lipofectAMINE reagent in 100 µl optiMEM, by addition of the DNA/optiMEM to the lipofectAMINE/optiMEM with gentle mixing. This was left to sit 15 min to 1 hour at room temperature. To each well were added 800 µl optiMEM to obtain a final volume of 1 ml. The cells were washed 2 times in 1X PBS and once in optiMEM. The 1 ml DNA/lipofectamine/optiMEM mixture was added to the cells and incubated for 20 hours at 37°C.

The effects of retinoic acid on promoter activity were analyzed by cotransfecting with varying amounts (100 ng to 1 µg) of expression vectors including cDNA sequences encoding zebrafish retinoic acid receptor gamma (RAR-γ) and zebrafish retinoid x receptor alpha (RXR-α). Comparisons were made between cells transfected with the control pGL3B vector, those with the sense construct and those with the antisense construct by incubating one set of the transfected cells with DMSO and the other set with 10^{-6} M RA in DMSO.

On the fourth day, the medium was removed and replaced with fresh medium (+ 10% FBS). The RA or vehicle was added as required to the wells and mixed gently by swirling the plate. For a 6 well plate 3 µl of 10^{-3} M RA in DMSO were added. Control wells received 3 µl DMSO. The plates were covered in foil and incubated for 24 hours at 37°C.

On the fifth day, cells were harvested by removing the medium and washing twice in 1X PBS. Then, on ice, 100 µl lysis buffer (1% Triton X-100, 25 mM glycylglycine, pH 7.8, 15 mM $MgSO_4$, 4 mM EGTA, 1mM DTT (added fresh just before use)) were added, and the cells were scraped off the bottom of the dish and transferred to a 1.5 ml microcentrifuge tube, spun for 5 minutes at 12000 x g, and the supernatant recovered.

To assay for luciferase activity, 20 µl supernatant from lysed, pelleted cells were transferred to a fresh tube. 80 µl luciferase assay buffer were added and a reading in millivolts in a luminometer was taken immediately. The luciferase assay buffer was 20 mM Tricine, 1.07 mM $(MgCO_3)_4Mg(OH)_2{}^*5H_2O$, 2.67 mM $MgSO_4$, 0.1 mM EDTA, 33.3 mM DTT, 0.27

mM Coenzyme A, 0.47 mM Luciferin, 0.53 mM ATP. The results are shown in Figure 15, in which it can be seen that enhanced luciferase activity was observed in the presence of RA, RXRα and RARγ, for both orientations of the promoter sequence, although enhancement appears to be greater for the sense construct.

Example 9

Inhibition of P450 Induction

Given a nucleotide sequence having promoter activity, it is possible to study the influence of an agent on that promoter activity. For example, it has recently been reported that 4-hydroxyphenylretinamide (4-HPR) inhibits RA-induced RA catabolism by NB4 cells [Taimi, 1997]. It was suggested that 4-HPR may inhibit the cytochrome P450 enzyme(s) responsible for RA oxidation to competitively inhibit RA catabolism. However, any such enzymes were not identified, no explanation of the nature of 4-HPR inhibition of RA catabolism was provided, and no evidence of 4-HPR metabolism was observed.

Experiments to determine the effect of 4-HPR on the induction of P450RAI were thus carried out. Figure 16 illustrates the ability of the synthetic retinoid, 4-HPR to inhibit the induction of P450RAI by RA. MCF7 cells were grown in culture in minimal essential medium (MEM) (Gibco) supplemented with 10 % fetal calf serum, insulin (.01 mg/mL), MEM non-essential amino acids (as directed by the manufacturer - Gibco), sodium pyruvate (500 nM), L-glutamine (2 mM) gentamycin (10 µg/mL), penicillin (5 µg/mL), streptomycin (5 µg/mL), and fungizone (200 ng/mL). MCF7 cells cultured in parallel were treated for 12 hours with: 10 µM 4HPR; 1 µM 4HPR; 1 µM RA; 1 µM RA and 10 µM 4-HPR; or 1 µM RA and 1 µM 4-HPR. At the end of the 12 hour treatment, total RNA was extracted from the cells using TRIzol reagent (as outlined by the manufacturer - Gibco). P450RAI message in the total RNA preparations was analyzed by northern blot hybridization. The blot was reprobed with a probe corresponding to the GAPDH cDNA to control for equivalent loading of RNA in each lane of the blot. The results indicate that 4-HPR treatment alone does not induce the P450RAI message. As expected, RA treatment of MCF7 cells results in a marked induction of P450RAI message following 12 hours of incubation. However, when cells are treated with RA in the presence of 10 µM 4-HPR, there is a noticeable suppression of P450RAI induction.

Example 10

P450 expression in the presence of retinoic acid and ketoconazole

This example illustrates the effect a P450 inhibitor can have on retinoic acid metabolism and hence, expression of a retinoic acid responsive gene, e.g., P450RAI gene. Figure 17 shows a time course of expression of cytochrome p450 RAI following treatment of MCF7 breast epithelium adenocarcinoma cells with 1 µM all-*trans* retinoic acid. In this northern blot analysis, total RNA was extracted at the indicated time points and transferred to nitrocellulose following electrophoresis in a 1% agarose, 0.66 M formaldehyde gel. The nitrocellulose was then probed with radioactively labelled human cytochrome p450RAI cDNA

or GAPDH cDNA to control for equivalence of mRNA loading. This shows that after 3 hours of incubation with retinoic acid, the MCF7 cells express high levels of P450RAI message and following 12 hours of exposure, the message declines sharply, possibly indicating that the metabolic activity of induced P450RAI protein is reducing the concentration of active retinoic acid in the surrounding medium. This strongly suggests that the induction of P450RAI in MCF7 cells forms an autoregulatory negative feedback loop.

Figure 18 shows a time course of P450RAI mRNA expression in MCF7 cells similar to that described in Figure 17, except that the effect of the broad spectrum cytochrome p450 inhibitor ketoconazole on P450RAI expression was examined. In the absence of ketoconazole, lanes 1 to 5 in Figure 18 a time course similar to that shown in Figure 17 is shown. In cells which were exposed to 1 μM ketoconazole (replenished every 12 hours following initial treatment), cytochrome P450RAI message was detectable at high levels at 24 and 48 hour time points indicating that the breakdown of retinoic acid can be inhibited by a cytochrome P450 inhibitor and that P450RAI metabolism may be responsible for the sharp drop in P450RAI message in the absence of ketoconazole. This thus one approach to identifying P450RAI inhibitors.

Example 11

P450RAI expression in the presence of Am580

Figure 19 shows a time course of P450RAI mRNA expression in MCF7 cells similar to that described in Figure 17, except that a comparison was made between the retinobenzoic acid derivative Am580 and all-*trans* retinoic acid. Retinoic acid shows a typical time course of induction of P450RAI message, lanes 1 to 4, Figure 19. Am580 induces P450RAI message to levels comparable to those observed following treatment with retinoic acid. Notably, whereas P450RAI message has declined sharply between 24 and 48 hours, for retinoic acid treated cells, the levels of P450RAI in Am580 treated cells remains high at this time point. This indicates that the synthetic retinoid Am580 is resilient to metabolism in these cells and illustrates the utility of identifying such compounds for therapeutic use. For example, the resistance to retinoic acid treatment observed in acute promyelocytic leukemia due to increased retinoic acid metabolism [Warrell, 1994] might be circumvented by treatment with a metabolism-resistant retinoid. P450RAI protein may be a useful agent for screening for these types of compounds.

Discussion

The foregoing examples illustrate various aspects of the invention, particularly, methods for identifying a nucleotide sequence encoding a cytochrome P450 in which the cytochrome is inducible or suppressible in cells of a given type by an agent. Also described are methods for identifying nucleotide sequences having promoter activity in conjunction with a nucleotide sequence encoding a cytochrome P450.

Given a coding sequence, a person skilled in the art is capable of preparing

the protein or polypeptide having the corresponding amino acid sequence. It is then possible to screen agents which interact with the protein and identify those that act as inhibitors, for example.

Likewise, given a nucleic acid sequence having promoter activity, it is possible to screen agents for their effects on such promoter activity.

Antisense nucleic acids or oligonucleotides (RNA or preferably DNA) that inhibit cellular P450 production induced (or suppressed) by a given agent can be used to inhibit metabolism of a substrate of the P450 by the P450 [Monia, 1996]. Antisense oligonucleotides, typically 15 to 20 bases long, bind to the sense mRNA or pre mRNA region coding for the protein of interest, which can inhibit translation of the bound mRNA to protein. The cDNA sequence encoding hP450RAI, for example, can thus be used to design a series of oligonucleotides which together span a large portion, or even the entire cDNA sequence. These oligonucleotides can be tested to determine which provides the greatest inhibitory effect on the expression of the protein [Stewart, 1996]. This can be done by exposing cells to the various oligonucleotides and measuring subsequent changes in hP450RAI activity or by using antibodies to screen for inhibition of P450RAI synthesis. The most suitable mRNA target sites include 5'- and 3'-untranslated regions as well as the initiation codon. Other regions might be found to be more or less effective. Alternatively, an antisense nucleic acid or oligonucleotide may bind to P450RAI DNA coding or regulatory sequences.

Rather than reducing substrate metabolism by reducing P450 gene expression (e.g., lowering RA metabolism by inhibiting P450RAI gene expression) at the nucleic acid level, activity of the P450 protein may be directly inhibited by binding to an agent, such as, for example, a suitable small molecule.

The present invention thus includes a method of screening drugs for their effect on activity of a protein inducible by an agent, particularly, retinoic acid. The method includes exposing the protein to a prospective inhibitor drug and determining the effect on protein activity. The measured activity might be hydroxylation of a retinoid, particularly all-*trans* retinoic acid, or hydroxylation of a retinoic acid, particularly all-*trans* retinoic acid, at the 4 position of the β-ionone ring thereof. For screening drugs for use in humans, hP450RAI itself is particularly useful for testing the effectiveness of such drugs. Prospective drugs could also be tested for inhibition of the activity of other P450 cytochromes, which are desired not to be inhibited. In this way, drugs which selectively inhibit hP450RAI over other P450s could be identified.

Another system for screening for potential inhibitors of a P450 protein includes a stably transfected cell line having incorporated therein DNA of a reporter gene (e.g., β-galactosidase, firefly luciferase, or the like) and of the P450, in which expression of both genes is inducible by exposure of the cells to an agent, such as RA. Expression of the reporter gene provides a measure of the induction of the expression system and therefore provides an indication of the amount of agent present. Where the agent is substrate, such as RA for example, exposure of the cells to RA leads to RA metabolism and, with time, such metabolism

leads to a decrease in the degree of induction which is indicated by the reporter protein. Exposure of the cells to RA in the presence of an agent that inhibits P450RAI metabolism of RA results in decreased RA metabolism, whereas exposure of the cells to RA in the presence of an agent that does not inhibit P450RAI metabolism of RA has no effect on RA metabolism. A comparison of expression of the reporter gene in the presence of RA alone and in the presence of both RA and a potential inhibitory drug thus gives a measure of the effectiveness of the drug in inhibiting metabolism of RA by the P450RAI protein.

One system for screening for potential inhibitiors of a P450 protein includes a cell line in which the endogenous P450 gene is not present or not functional or not expressed. In this cell line, a cytochrome P450 expression vector and an inducible reporter gene are incorporated such that exposure of the cell line to an induction agent, (e.g., RA) results in metabolism of the agent by the expressed P450 protein and a degree of induction of the reporter gene based on remaining active agent. The addition of an inhibitor of P450RAI will decrease the rate of metabolism/degradation of the agent and therefore increase the activation/induction of thereporter gene sensitive to the agent.

Given the high degree of conservation of the promoter regions of the mouse, human and zebrafish P450RAI promoter regions, it is likely that RA regulates P450RAI expression through a transcriptional mechanism involving the RARE conserved in the promoters of all three species. This is supported by studies which show the rapid and RA-dependent expression of P450RAI in a number of cell lines. Since P450RAI message is induced so strongly a reporter gene may be a useful indicator of RA activity *in vivo* as well as *in vitro*. Thus, the P450RAI promoter linked to a reporter gene provides a tool for screening retinoids or other compounds which have the ability to block or inhibit P450RAI induction. For example, the P450RAI reporter gene could be stably or transiently introduced into a cell line such that when the cells are exposed to a certain level of retinoid or other agent, the concentration will be reflected in reporter gene activity. Such transfection assays can be carried out in a manner similar to those described by Petkovich *et al.*, for example [Petkovich, 1987; Ohno, 1994].

The invention thus provides a system for screening potential inhibitors of RA catabolism by a P450RAI protein. The system includes a transfected cell line having incorporated therein DNA of a reporter gene, for example the luciferase gene exemplified above, in which expression of the reporter gene is inducible by exposure of the cells to RA. In this system, the P450RAI gene is omitted, that is the reporter gene is under the control of the native promoter for the P450RAI gene. Expression of the reporter gene provides a measure of the induction of the expression system and therefore provides an indication of the amount of mRNA produced in response to exposure of the cells to RA. Exposure of the cells to RA in the presence of an agent that inhibits induction of the expression system indicates that the agent is a potential inhibitor of RA catabolism, i.e., provides a measure of the effectiveness of the agent as a drug in inhibiting the expression of P450RAI gene and thus metabolism of RA.

It will of course be understood, without the intention of being limited thereby,

that a variety of substitutions of amino acids is possible while preserving the structure responsible for retinoid metabolizing acitivity of the protein of the present invention. Conservative substitutions are described in the patent literature, as for example, in United States Patent No. 5,264,558. It is thus expected, for example, that interchange among non-polar aliphatic neutral amino acids, glycine, alanine, proline, valine and isoleucine, would be possible. Likewise, substitutions among the polar aliphatic neutral amino acids, serine, threonine, methionine, asparagine and glutamine could possibly be made. Substitutions among the charged acidic amino acids, aspartic acid and glutamic acid, could probably be made, as could substitutions among the charged basic amino acids, lysine and arginine. Substitutions among the aromatic amino acids, including phenylalanine, histidine, tryptophan and tyrosine would also likely be possible. These sorts of substitutions and interchanges are well known to those skilled in the art. Other substitutions might well be possible. Of course, it would also be expected that the greater the percentage of homology, i.e., sequence similarity, of a variant protein with a naturally occuring protein, the greater the retention of metabolic activity. Of course, as protein variants having the activity of P450 as described herein are intended to be within the scope of this invention, so are nucleic acids encoding such variants.

A further advantage may be obtained through chimeric forms of the protein, as known in the art. A DNA sequence encoding the entire protein, or a portion of the protein, could thus be linked, for example, with a sequence coding for the C-terminal portion of *E. coli* ß-galactosidase to produce a fusion protein. An expression system for human respiratory syncytial virus glycoproteins F and G is described in United States Patent No. 5,288,630 issued February 22, 1994 and references cited therein, for example.

A recombinant expression vector of the invention can be a plasmid, as described above. The recombinant expression vector of the invention further can be a virus, or portion thereof, which allows for expression of a nucleic acid introduced into the viral nucleic acid. For example, replication defective retroviruses, adenoviruses and adeno-associated viruses can be used.

The invention provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression, by transcription of the DNA molecule, of an RNA molecule which is antisense to the nucleotide sequence of SEQ ID NO:45, SEQ ID NO:38, SEQ ID NO:46, for example. Regulatory sequences operatively linked to the antisense nucleic acid can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance a viral promoter and/or enhancer, or other regulatory sequences can be chosen which direct tissue or cell type specific expression of antisense RNA.

The recombinant expression vectors of the invention can be used to make a transformant host cell including the recombinant expression vector. The term "transformant host cell" is intended to include prokaryotic and eukaryotic cells which have been transformed or transfected with a recombinant expression vector of the invention. The terms "transformed

with", "transfected with", "transformation" and "transfection" are intended to encompass introduction of nucleic acid (e.g. a vector) into a cell by one of many possible techniques known in the art. Prokaryotic cells can be transformed with nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. Nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride coprecipitation, DEAE-dextran-mediated transfection, lipofection, electroporation or microinjection. Suitable methods for transforming and transfecting host cells are known [Sambrook, 1989].

The number of host cells transformed with a recombinant expression vector of the invention by techniques such as those described above will depend upon the type of recombinant expression vector used and the type of transformation technique used. Plasmid vectors introduced into mammalian cells are integrated into host cell DNA at only a low frequency. In order to identify these integrants, a gene that contains a selectable marker (e.g. resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to certain drugs, such as G418 and hygromycin. Selectable markers can be introduced on a separate plasmid from the nucleic acid of interest or, preferably, are introduced on the same plasmid. Host cells transformed with one or more recombinant expression vectors containing a nucleic acid of the invention and a gene for a selectable marker can be identified by selecting for cells using the selectable marker. For example, if the selectable marker encodes a gene conferring neomycin resistance (such as pRc/CMV), transformant cells can be selected with G418. Cells that have incorporated the selectable marker gene will survive, while the other cells die.

Certain nucleic acids of the invention encode proteins which "have biological activity of a cytochrome P450", cytochrome P450s being a family of NAPDH-dependent heme-containing enzymes involved in the metabolism of endogenous compounds such as steroids and fatty acids and numerous foreign compounds such as drugs and chemical carcinogens. Such activity can be tested for.

The invention provides purified proteins having biological P450 activity. The terms "isolated" and "purified" each refer to a protein substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. In certain preferred embodiments, the protein having biological activity of P450RAI comprises an amino acid sequence identified as SEQ ID NO:34 or SEQ ID NO:39 or SEQ ID NO:47. Alternatively, preferred proteins encoded by a nucleic acid comprising the nucleotide sequence identified as SEQ ID NO:45 or SEQ ID NO:38 or SEQ ID NO:46, as defined above, are encompassed by the invention. Furthermore, proteins having biological activity of P450RAI that are encoded by nucleic acids which hybridize under stringent conditions, as discussed above, to a nucleic acid comprising a nucleotide sequence identified as SEQ ID NO:45 or SEQ ID NO:38 or SEQ ID NO:46 are encompassed by the invention. P450s, and particularly, P450RAIs of the invention can be obtained by expression in a suitable host cell using techniques known in the art. Suitable host cells include prokaryotic or

eukaryotic organisms or cell lines, for example, yeast, *E. coli*, insect cells and COS 1 cells. The recombinant expression vectors of the invention, described above, can be used to express a protein having P450 activity in a host cell in order to isolate the protein. The invention provides a method of preparing an purified protein of the invention comprising introducing into a host cell a recombinant nucleic acid encoding the protein, allowing the protein to be expressed in the host cell and isolating and purifying the protein. Preferably, the recombinant nucleic acid is a recombinant expression vector. Proteins can be isolated from a host cell expressing the protein and purified according to standard procedures of the art, including ammonium sulfate precipitation, column chromatography (e.g. ion exchange, gel filtration, affinity chromatography, etc.), electrophoresis, and ultimately, crystallization [see generally, "Enzyme Purification and Related Techniques", *Methods in Enzymology*, 22, 233-577 (1971)].

Alternatively, the protein or parts thereof can be prepared by chemical synthesis using techniques well known in the chemistry of proteins such as solid phase synthesis [Merrifield, 1964], or synthesis in homogeneous solution [Houbenwcyl, 1987].

The protein of the invention, or portions thereof, can be used to prepare antibodies specific for the proteins. Antibodies can be prepared which bind to a distinct epitope in an unconserved region of a particular protein. An unconserved region of the protein is one which does not have substantial sequence homology to other proteins, for example other members of the P450 family of cytochromes. Conventional methods can be used to prepare the antibodies. For example, by using a peptide of a P450RAI protein, polyclonal antisera or monoclonal antibodies can be made using standard methods. Techniques for conferring immunogenicity on a peptide include conjugation to carriers or other techniques well known in the art. For example, the peptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassay can be used to assess the levels of antibodies. Following immunization, antisera can be obtained and, if desired, polyclonal antibodies isolated from the sera.

To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized animal and fused with myeloma cells by standard somatic cell fusion procedures, thus immortalizing these cells and yielding hybridoma cells. Such techniques are well known in the art. For example, the hybridoma technique originally developed by Kohler and Milstein [Kohler, 1975] as well as other techniques such as the human B-cell hybridoma technique [Kozbor, 1983], the EBV-hybridoma technique to produce human monoclonal antibodies [Cole, 1985], and screening of combinatorial antibody libraries [Huse, 1989]. Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with the peptide, and monoclonal antibodies isolated.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with a protein having the biological activity of P450, or a peptide fragment thereof. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies.

For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments.

Another method of generating specific antibodies, or antibody fragments, reactive against protein having the biological activity of a P450, or a peptide fragment thereof, is to screen expression libraries encoding immunoglobulin genes, or portions thereof, expressed in bacteria, with peptides produced from the nucleic acid molecules of the present invention. For example, complete Fab fragments, VH regions and FV regions can be expressed in bacteria using phage expression libraries. See for example Ward *et al.*, Huse *et al.*, and McCafferty *et al.* [Ward, 1989; Huse, 1989; McCafferty, 1990]. Screening such libraries with, for example, a P450 peptide can identify immunoglobulin fragments reactive with P450. Alternatively, the SCID-hu mouse developed by Genpharm can be used to produce antibodies, or fragments thereof.

The polyclonal, monoclonal or chimeric monoclonal antibodies can be used to detect the proteins of the invention, portions thereof or closely related isoforms in various biological materials, for example they can be used in an ELISA, radioimmunoassay or histochemical tests. Thus, the antibodies can be used to quantify the amount of a P450 protein of the invention, portions thereof or closely related isoforms in a sample in order to determine the role of P450 proteins in particular cellular events or pathological states. Using methods described hereinbefore, polyclonal, monoclonal antibodies, or chimeric monoclonal antibodies can be raised to nonconserved regions of a P450 and used to distinguish a particular P450 from other proteins.

The polyclonal or monoclonal antibodies can be coupled to a detectable substance or reporter system. The term "coupled" is used to mean that the detectable substance is physically linked to the antibody. Suitable detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, and acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride and phycoerythrin; an example of a luminescent material includes luminol; and examples of suitable radioactive material include ^{125}I; ^{131}I, ^{35}S and ^{3}H. In a preferred embodiment, the reporter system allows quantitation of the amount of protein (antigen) present.

Such an antibody-linked reporter system could be used in a method for determining whether a fluid or tissue sample of a subject contains a deficient amount or an excessive amount of the protein. Given a normal threshold concentration of such a protein for a given type of subject, test kits could thus be developed.

The present invention allows the skilled artisan to prepare bispecific antibodies and tetrameric antibody complexes. Bispecific antibodies can be prepared by forming hybrid hybridomas [Staerz, 1986a &b].

The present invention includes at least three types of compounds and methods for screening such compounds: those that inhibit enzymatic activity of a P450, thereby inhibiting metabolism of a substrate; those with desirable P450 substrate activities that evade metabolism by P450, for example Am580, shown above; and those compounds that repress induction of P450 gene expression, for example 4-HPR, as shown above.

Compositions of the invention are administered to subjects in a biologically compatible form suitable for pharmaceutical administration in vivo. By "biologically compatible from suitable for administration *in vivo*" is meant a form of the composition to be administered in which any toxic effects are outweighed by the therapeutic effects of the composition. The term "subject" is intended to include living organisms in which a desired therapeutic response can be elicited, e.g. mammals. Examples of subjects include human, dogs, cats, mice, rats and transgenic species thereof. Administration of a therapeutically active amount of the therapeutic compositions of the present invention is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, a therapeutically active amount of a compound that inhibits catabolism of a P450 agent such as RA by a P450RAI protein may vary according to factors such as the disease state, age, sex, and weight of the individual, as well as target tissue and mode of delivery. Dosage regimes may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

Compounds of the present invention, such as those that are found to inhibit metabolism of RA by P450RAI enzymes and that are useful as anticancer agents and in the treatment, amelioration, or prevention of skin disorders for which retinoic acid is useful, for example, may be used topically. In this regard they may be included in compositions for therapy in animals, including humans, for premalignant epithelial cell lesions, as a prophylaxis against tumor promotion in epithelial cells and treatment for dermatoses such as ichthyoses, follicular disorders, benign epithelial disorders, and other proliferative skin diseases, such as acne, psoriasis, eczema, atopic dermatitis, nonspecific dermatitis and the like.

Topical compositions are usually formulated with a pharmaceutically acceptable carrier in liquid, semi-solid or solid form. A pharmaceutically acceptable carrier is a material that is nontoxic and generally inert and does not affect the functionality of the active ingredients adversely. Such materials are well known and include those materials sometimes referred to as diluents or vehicles (excipients) in the pharmaceutical formulation art. The carrier may be organic or inorganic in nature. Examples of pharmaceutically acceptable carriers are water, gelatin, lactose, starch, mineral oil, cocoa butter, dextrose, sucrose, sorbitol, mannitol, gum, acacia, alginates, cellulose, talc, magnesium stearate, polyoxyethylene sorbitan monolaurate, and other commonly used pharmaceutical carriers. In addition to an active ingredient and carrier, the formulation may contain minor amounts of additives such as flavoring agents, coloring agents, thickening or gelling agents, emulsifiers, wetting agents, buffers, stabilizers, and preservatives such as antioxidants.

Certain compositions may be administered enterally. For oral administration, suitable forms are, for example, tablets, pills, syrups, suspensions, emulsions, solutions, powders and granules.

As anti-tumor agents or as part of an anti-tumor formulation, for example, compounds of the present invention can be used in a similar manner to retinoids used for treating various tumours, such as all-*trans* retinoic acid. The dose to be administered, whether a single dose, multiple does or daily dose, will of course vary with the particular compound employed because of the varying potency of the active ingredient, the chosen route of administration, the size of the recipient, the type of tumor, and the nature of the patient's condition. The dosage to be administered is not subject to definite bounds, but it will usually be an effective amount, or the equivalent on a molar basis of the pharmacologically active free form produced from a dosage formulation upon the metabolic release of the active drug to achieve its desired pharmacological and physiological effects. An oncologist skilled in the art of cancer treatment will be able to ascertain without undue experimentation, appropriate protocols for the effective administration of the compounds of this present invention.

Nucleic acids which encode proteins having biological activity of a P450 protein can be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (e.g., a mouse) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, a human P450, for example, P450RAI cDNA, comprising the nucleotide sequence shown in SEQ ID NO:38, or an appropriate variant or subsequence thereof, can be used to generate transgenic animals that contain cells which express human P450 protein. Methods for generating transgenic animals, particularly animals such as mice, have become conventional in the art are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. In a preferred embodiment, plasmids containing recombinant molecules of the invention are microinjected into mouse embryos. In particular, the plasmids are microinjected into the male pronuclei of fertilized one-cell mouse eggs; the injected eggs are transferred to pseudo-pregnant foster females; and, the eggs in the foster females are allowed to develop to term. [Hogan, 1986]. Alternatively, an embryonal stem cell line can be transfected with an expression vector comprising nucleic acid encoding a protein having P450 activity, and cells containing the nucleic acid can be used to form aggregation chimeras with embryos from a suitable recipient mouse strain. The chimeric embryos can then be implanted into a suitable pseudopregnant female mouse of the appropriate strain and the embryo brought to term. Progeny harboring the transfected DNA in their germ cells can be used to breed uniformly transgenic mice.

Typically, particular cells would be targeted for P450 transgene incorporation by use of tissue specific enhancers operatively linked to the P450 encoding gene. For example, promoters and/or enhancers which direct expression of a gene to which they are

operatively linked preferentially in cardiac muscle cells can be used to create a transgenic animal which expresses a P450 protein preferentially in cardiac muscle tissue. Examples of suitable promoters and enhancers include those which regulate the expression of the genes for cardiac myosin and cardiac actin. Transgenic animals that include a copy of an P450 transgene introduced into the germ line of the animal at an embryonic stage can also be used to examine the effect of increased P450 expression in various tissues.

The pattern and extent of expression of a recombinant molecule of the invention in a transgenic mouse is facilitated by fusing a reporter gene to the recombinant molecule such that both genes are co-transcribed to form a polycistronic mRNA. The reporter gene can be introduced into the recombinant molecule using conventional methods such as those described in Sambrook *et al.*, [Sambrook,1989]. Efficient expression of both cistrons of the polycistronic mRNA encoding the protein of the invention and the reporter protein can be achieved by inclusion of a known internal translational initiation sequence such as that present in poliovirus mRNA. The reporter gene should be under the control of the regulatory sequence of the recombinant molecule of the invention and the pattern and extent of expression of the gene encoding a protein of the invention can accordingly be determined by assaying for the phenotype of the reporter gene. Preferably the reporter gene codes for a phenotype not displayed by the host cell and the phenotype can be assayed quantitatively. Examples of suitable reporter genes include lacZ (β-galactosidase), neo (neomycin phosphotransferase), CAT (chloramphenicol acetyltransferase) dhfr (dihydrofolate reductase), aphIV (hygromycin phosphotransferase), lux (luciferase), uidA (β-glucuronidase). Preferably, the reporter gene is lacZ which codes for β-galactosidase. β-galactosidase can be assayed using the lactose analogue X-gal (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside) which is broken down by β-galactosidase to a product that is blue in color [Old].

Although experimental animals used in the preferred embodiment disclosed are mice, the invention should not be limited thereto. It can be desirable to use other species such as, for example, rats, hamsters, rabbits and sheep.

The transgenic animals of the invention can be used to investigate the molecular basis of RA metabolism. The transgenic animals of the invention can also be used to test substances for the ability to prevent, slow or enhance RA metabolism. A transgenic animal can be treated with the substance in parallel with an untreated control transgenic animal.

Cells from the transgenic animals of the invention can be cultured using standard tissue culture techniques. In particular, cells carrying the recombinant molecule of the invention can be cultured and used to test substances for the ability to prevent, slow or enhance RA metabolism.

Additionally, the non-human homologues of genes encoding proteins having P450 activity can be used to construct a "knock out" animal which has a defective or altered P450 gene. For example, with established techniques, a portion of murine genomic P450RAI DNA (e.g., an exon), can be deleted or replaced with another gene, such as a gene encoding a

selectable marker which can be used to monitor integration. The altered P450RAI DNA can then be transfected into an embryonal stem cell line. The altered P450RAI DNA will homologously recombine with the endogenous P450RAI gene in certain cells and clones containing the altered gene can be selected. Cells containing the altered gene are injected into a blastocyst of an animal, such as a mouse, to form aggregation chimeras as described for transgenic animals. Chimeric embryos are implanted as described above. Transmission of the altered gene into the germline of a resultant animal can be confirmed using standard techniques and the animal can be used to breed animals having an altered P450RAI gene in every cell [Lemoine, 1996]. Accordingly, a knockout animal can be made which cannot express a functional P450RAI protein. Such a knockout animal can be used, for example, to test the effectiveness of an agent in the absence of a P450RAI protein.

The antisense nucleic acids and oligonucleotides of the invention are useful for inhibiting expression of nucleic acids (e.g. mRNAs) encoding proteins having P450RAI activity. Since proteins having P450RAI activity are associated with metabolism of agents which can act on the cell, e.g., RA, decreasing expression of such proteins can increase sensitivity of the cell to such agents. Antisense nucleic acids can be introduced into a drug resistant cell in culture to inhibit P450RAI expression. One or more antisense nucleic acids, such as oligonucleotides, can be added to cells in culture media, typically, for example, at 200 µg/ml.

The antisense nucleic acids of the invention, or oligonucleotides thereof, can thus be used in gene therapy to correct or prevent retinoic acid or other retinoid resistance in a subject. For example, antisense sequences can be used to render retinoic acid or other retinoid resistant malignant cells sensitive to chemotherapeutic agents. Administration of antisense nucleic acids to a subject may be most effective when the antisense nucleic acid is contained in a recombinant expression vector which allows for continuous production of antisense RNA. Recombinant molecules comprising an antisense nucleic acid or oligonucleotide thereof, can be directly introduced into tissues, including lung tissue *in vivo*, using delivery vehicles such as liposomes, retroviral vectors, adenoviral vectors and DNA virus vectors. A delivery vehicle can be chosen which can be targeted to a cell of interest in the subject (e.g. a retinoid resistant tumor cell). Antisense nucleic acids can also be introduced into isolated cells, such as those of the haematopoietic system, *ex vivo* using viral vectors or physical techniques such as microinjection and electroporation or chemical methods such as coprecipitation and incorporation of DNA into liposomes, and such cells can be returned to the donor. Recombinant molecules can be delivered in the form of an aerosol or by lavage.

Accordingly, the invention provides a method for inhibiting retinoic acid or other retinoid resistance of a resistant cell by introducing into the resistant cell a nucleic acid which is antisense to a nucleic acid which encodes the protein identified as SEQ ID NO:34, SEQ ID NO:47, or particularly, the in case of human cells SEQ ID NO:39.

The nucleic acids of the invention can further be used to design ribozymes which are capable of cleaving a single-stranded nucleic acid encoding a protein having P450 activity, such as an mRNA. A catalytic RNA (ribozyme) having ribonuclease activity can be

designed which has specificity for a P450-encoding mRNA based upon the sequence of a nucleic acid of the invention. For example, a derivative of a *Tetrahymena* L-19IVS RNA can be constructed in which the base sequence of the active site is complementary to the base sequence to be cleaved in a P450-encoding mRNA. [Cech a and b]. Alternatively, a nucleic acid of the invention could be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules [Bartel, 1993].

The isolated nucleic acids and antisense nucleic acids of the invention can be used to construct recombinant expression vectors as described previously. These recombinant expression vectors are then useful for making transformant host cells containing the recombinant expression vectors, for expressing protein encoded by the nucleic acids of the invention, and for isolating proteins of the invention as described previously. The isolated nucleic acids and antisense nucleic acids of the invention can also be used to construct transgenic and knockout animals as described previously.

The isolated proteins of the invention are useful for making antibodies reactive against proteins having P450 activity, and particularly, P450RAI activity, as described previously. Alternatively, the antibodies of the invention can be used to isolate a protein of the invention by standard immunoaffinity techniques. Furthermore, the antibodies of the invention, including bispecific antibodies are useful for diagnostic purposes.

Molecules which bind to a protein comprising an amino acid sequence shown in SEQ ID NO:39, for example, can also be used in a method for killing a cell which expresses the protein, wherein the cell takes up the molecule. Preferably, the cell is a tumor cell. Destruction of such cells can be accomplished by labeling the molecule with a substance having toxic or therapeutic activity. The term "substance having toxic or therapeutic activity" as used herein is intended to include molecules whose action can destroy a cell, such as a radioactive isotope, a toxin (e.g. diphtheria toxin or ricin), or a chemotherapeutic drug, as well as cells whose action can destroy a cell, such as a cytotoxic cell. The molecule binding to the cytochrome can be directly coupled to a substance having a toxic or therapeutic activity or may be indirectly linked to the substance. In one example, the toxicity of the molecule taken up by the cell is activated by P450RAI protein.

The invention also provides a diagnostic kit for identifying tumor cells comprising a molecule which binds to a protein comprising an amino acid sequence shown in SEQ ID NO:39, for example, or other human protein, for incubation with a sample of tumor cells; means for detecting the molecule bound to the protein, unreacted protein or unbound molecule; means for determining the amount of protein in the sample; and means for comparing the amount of protein in the sample with a standard. Preferably, the molecule is a monoclonal antibody. In some embodiments of the invention, the detectability of the molecule which binds to P450 is activated by said binding (e.g., change in fluorescence spectrum, loss of radioisotopic label). The diagnostic kit can also contain an instruction manual for use of the kit.

The invention further provides a diagnostic kit for identifying tumor cells comprising a nucleotide probe complementary to the sequence, or an oligonucleotide fragment

thereof, shown in SEQ ID NO:45, for example, for hybridization with mRNA from a sample of tumor cells; means for detecting the nucleotide probe bound to mRNA in the sample with a standard. The diagnostic kit can also contain an instruction manual for use of the kit.

Those skilled in the art will know, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

REFERENCES

Particulars of references cited above are given below. All of the listed references are incorporated herein by reference.

Achkar, C., *et al.* (1996). Proc. Natl. Acad. Sci. USA, 4879-4884.

Adams *et al.*, (1985). J. Clin Endocrinol Metab **60**, 960-966.

Bligh, E. G. and Dyer, W. J. (1957). A rapid method of total lipid extraction and purification. Canadian Journal of Biochemistry 37, 911-917.

Boylan, J. F., Lufkin, T., Achkar, C. C., Taneha, R., Chambon, P. and Gudas, L. J. (1995). Targeted Disruption of Retinoic Acid Receptor a (RARa) and RARg Results in Receptor-Specific Alterations in Retinoic Acid-Mediated Differentiation and Retinoic Acid Metabolism. Mol. Cell Biol. **15**, 843-851.

Cali, J.J. and Russell, D.W. (1991). Characterization of human sterol 27-hydroxylase. A mitochondrial cytochromeP-450 that catalyzes multiple oxidation reaction in bile acid biosynthesis. Journal of Biological Chemistry **266**, 7774-7778.

Chung, B.C., Matteson, K.J., Voutilainen, R., Mohandas T.K. and Miller, W.L. (1986). Human cholesterol side-chain cleavage enzyme. Proc. Natl. Acad. Sci., USA. **83**, 8962-8966.

Denison, M.S. and Whitlock, Jr., J.P. (1995). Xenobiotic-inducible transcription of cytochrome P450 genes. Journal of Biological Chemistry **270**, 18175-17178.

Duell, E. A., Astrom, A., Griffiths, C. E., Chambon, P. and Voorhees, J. J. (1992). Human skin levels of retinoic acid and cytochrome p-450-derived 4-hydroxyretinoic acid after topical application of retinoic acid *in vivo* compared to concentrations required to stimulate retinoic acid receptor-mediated transcription *in vitro*. Journal of Clinical Investigation **90**, 1269-74.

Frolik, C. A., Roberts, A. B., Tavela, T. E., Roller, P. P., Newton, D. L. and Sporn, M. B. (1979). Isolation and identification of 4-hydroxy- and 4-oxoretinoic acid. *In vitro* metabolites of all-*trans* retinoic acid in hamster trachea and liver. Biochemistry **18**, 2092-7.

Gonzalez, F. (1992). Trends Pharm. Sci. **13**, 346-352.

Graham-Lorence S., Amarne B., White R.E., Peterson J.A., Simpson, E.R. (1995). A three-dimensional model of the aromatase cytochrome p450. Protein Sci **4**,:1065-1080.

Green, S., Issemann, I. and Sheer, E. (1988). A versatile *in vivo* and *in vitro* eukaryotic expression vector for protein engineering. Nucleic Acids Research **16**, 369-370.

Guengerich, F.P. (1991). Reactions and significance of cytochrome p-450 enzymes. J. Biol. Chem. **266**, 10019-10022.

Henry,H.L. (1979). J.Biol. Chem. **254**, 2722-2729.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (1990). PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., San Diego, CA.

Ken, K.S., Prahl, J.M. and Deluca, H.F. (1993). Isolation and expression of human 1,25-dihydroxyvitamin D3 24-hydroxylase cDNA. Proc. Natl. Acad. Sci. USA., **90**, 4543-4547.

Koller, W. (1992). Antifungal agents with target sites in sterol functions and synthesis. In: Target Sites for Fungicide Action (Koller, Editor) CRC press, Boca Raton, Florida. 119-206.

Liang, P. and Pardee, A. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science **257**, 961-971.

Makin,G., Lohnes, D., Byford, V., Ray, R., and Jones, G. (1989) Biochem. J. **262**, 173-180.

Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogawa, K., Hirose, T., Inayama, S. and Omura, T. (1984). Molecular cloning and nucleotide sequence of cDNA of mRNA of mitochonrial cytochrome P-450 (SCC) of bovine adrenal cortex. Proc. Natl. Acad. Sci. USA., **81**, 4647-4651.

Nebert, D.W. and Gonzalez, F. J. (1987). Ann. Rev. Biochem. **56**, 945-993.

Nelson, D. R., Kamataki, T., Waxman, D. J., Guengerich, F. P., Estabrook, R. W., Feyereisen, R., Gonzalez, F. J., Coon, M. J., Gunsalus, I. C., Gotoh, O., Okuda, K. and Nebert, D. W. (1993). The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA & Cell Biology **12**, 1-51.

Nelson, D.R. (1995). Cytochrome P450 nomenclature and alignment of selected sequences", in Cytochrome P450 Structure, Mechanism, and Biochemistry, ed. Paul R. Ortiz de Montellano, Plenum Press, NY & London, 2nd Ed. pp. 575-606.

Quattrochi, L.C., Vu, T. and Tukey, R.H. (1994). Journal of Biological Chemistry **269**, 2949-2954.

Reichel *et al.*, (1987a). Journal of Biological Chemistry **262**, 10931-10937.

Reichel *et al.*, (1987b). J. Clin. Endocrino. Metabol. **64**, 1-9.

Roberts, A. B., Nichols, M. D., Newton, D. L. and Sporn, M. B. (1979a). *In vitro* metabolism of retinoic acid in hamster intestine and liver. Journal of Biological Chemistry **254**, 6296-302.

Roberts, A. B., Frolik, C. A., Nichols, M. D. and Sporn, M. B. (1979b). Retinoid-dependent induction of the *in vivo* and *in vitro* metabolism of retinoic acid in tissues of the vitamin A-deficient hamster. Journal of Biological Chemistry **254**, 6303-9.

Savas, U., Bhettscharyyn, K.K., Christou, M., Alexander, D.L., and Jefenate, C.R. (1994). Journal of Biological Chemistry **269**, 14095-14911.

Shen, Z., *et al.* (1993). Proc. Natl. Acad. Sci. USA., **90**, 11483-11487.

Sutter, T.R., Tang, Y.M., Hayes, C.L., Wo, Y.-Y., Jabs, E.W., Li, X., Yin, H., Cody, C.W., and Greenlee, W.F. (1994). Journal of Biological Chemistry **269**, 13092-13099.

Takatsuka, J., Takahashi, N. and De Luca, L. M. (1996). Retinoic Acid Metabolism and Inhibition of Cell Proliferation: An Unexpected Liaison. Cancer Research **56**, 675-678.

Vijayakumar S., Salerno J.C. (1992). Molecular modelling of the 3-D structure of cytochrome P-450scc. Biochim Biophys. Acta, **1160**, 281-286.

White, J. A., Boffa, M. B., Jones, B. and Petkovich, M. (1994). A zebrafish retinoic acid receptor expressed in the regenerating caudal fin. Development **120**, 1861-72.

White J., *et al.* (1996). Journal of Biological Chemistry **271**, 29922-29927.

Whitlock, Jr., J.P. (1993). Chem. Res. Toxol. **6**, 754-763.

Wouters, W., van, D. J., Dillen, A., Coene, M. C., Cools, W. and De, C. R. (1992). Effects of liarazole, a new antitumoral compound, on retinoic acid-induced inhibition of cell growth and on retinoic acid metabolism in MCF-7 human breast cancer cells. Cancer Research 52, 2841-6.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS:
(A) NAME: QUEEN'S UNIVERSITY AT KINGSTON
(B) STREET:
(C) CITY: Kingston
(D) PROVINCE: Ontario
(E) COUNTRY: CA
(F) POSTAL CODE (ZIP) : K7L 1J7

(A) NAME: PETKOVICH, P. Martin
(B) STREET: 875 Everitt Avenue
(C) CITY: Kingston
(D) PROVINCE: Ontario
(E) COUNTRY: CA
(F) POSTAL CODE (ZIP) : K7M 4R1

(ii) TITLE OF INVENTION: METHOD OF IDENTIFYING CYTOCHROMES

(iii) NUMBER OF SEQUENCES: 50

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
(B) COMPUTER: COMPAQ, IBM PC compatible
(C) OPERATING SYSTEM: MS-DOS 5.1
(D) SOFTWARE: WORD PERFECT

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/CA 97/00488
(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/667,546; US 08/724,466
(B) FILING DATE: 21-JUN-1996; 01-OCT-1996

(2) INFORMATION FOR SEQ ID NO:1
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1

```
His His Asp Asn Leu Met Ile Ser Leu Leu Ser Leu Phe Phe Ala Gly
 1               5                  10                  15

Thr Asp Thr Ser Ser Thr Thr Leu Arg Tyr Gly Phe Leu Leu Met Leu
            20                  25                  30

Lys Tyr Pro His Val Ala Asp Lys Val Gln Lys Asp Ile Asp Gln Val
        35                  40                  45

Ile Gly Ser His Arg Leu Pro Thr Leu Asp Asp Arg Ser Lys Met Pro
    50                  55                  60

Tyr Thr Asp Ala Val Ile His Asp Ile Gln Arg Phe Ser Asp Leu Val
65                  70                  75                  80

Pro Ile Gly Val Pro His Arg Val Thr Lys Asp Thr Met Phe Arg Gly
                85                  90                  95
```

```
Tyr Leu Leu Pro Lys Asn Thr Asp Val Tyr Pro Ile Leu Ser Ser Ala
            100                 105                 110

Leu His Asp Pro Gln Tyr Phe Asp His Pro Asp Ser Phe Asn Pro Asp
        115                 120                 125

His Phe Leu Asp Ala Asn Gly Ala Leu Lys Lys Ser Asp Ala Phe Met
    130                 135                 140

Pro Phe Ser Thr Gly Lys Arg Ile Cys Leu Gly Asp Gly Ile Ala Arg
145                 150                 155                 160

Asn Asp Leu Phe Leu Phe Phe Thr Thr Ile Leu Gln Asn Phe Ser Val
                165                 170                 175

Ser Ser His Leu Ala Pro Lys Asp Ile Asp Leu Thr Pro Lys Asp Ser
            180                 185                 190

Gly Ile Gly Lys Ile Pro Pro Thr Tyr Gln Ile Cys Phe Ser Ala Arg
        195                 200                 205
```

(2) INFORMATION FOR SEQ ID NO:2

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 211 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2

```
Arg Asp Asn Val Asn Gln Cys Ile Leu Asp Met Leu Ile Ala Ala Pro
 1               5                  10                  15

Asp Thr Met Ser Val Ser Leu Phe Phe Met Leu Phe Leu Ile Ala Lys
            20                  25                  30

His Pro Asn Val Asp Asp Ala Ile Ile Lys Asp Ile Gln Thr Val Ile
        35                  40                  45

Gly Asp Arg Asp Ile Lys Ile Asp Asp Ile Gln Lys Leu Lys Val Met
    50                  55                  60

Asp Asn Phe Ile Tyr Asp Ser Met Arg Tyr Gln Pro Val Val Asp Leu
65                  70                  75                  80

Val Met Arg Lys Ala Leu Asp Asp Asp Val Ile Asp Gly Tyr Pro Val
                85                  90                  95

Lys Lys Gly Thr Asn Ile Ile Leu Asn Ile Gly Arg Met His Arg Leu
            100                 105                 110

Asp Phe Phe Pro Lys Pro Asn Asp Phe Thr Leu Asp Asn Phe Ala Lys
        115                 120                 125

Asn Val Pro Tyr Arg Tyr Phe Gln Pro Phe Gly Phe Gly Pro Arg Gly
    130                 135                 140

Cys Ala Gly Lys Tyr Ile Ala Met Val Met Met Lys Ala Ile Leu Val
145                 150                 155                 160

Thr Leu Leu Arg Arg Phe His Val Lys Thr Leu Gln Gly Gln Cys Val
                165                 170                 175
```

```
Asp Ser Ile Gln Lys Ile His Asp Leu Ser Leu His Pro Asp Asp Thr
            180                 185                 190

Lys Asn Met Leu Asp Met Ile Phe Thr Pro Arg Asn Ser Asp Arg Cys
        195                 200                 205

Leu Asp His
    210
```

(2) INFORMATION FOR SEQ ID NO:3

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 209 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3

```
Leu Asp Asp Val Lys Ala Asn Ile Thr Asp Met Leu Ala Gly Gly Val
 1               5                  10                  15

Asn Thr Thr Ser Met Thr Leu Gln Trp His Leu Tyr Asp Met Ala Arg
            20                  25                  30

Ser Leu Asn Val Gln Asp Met Leu Arg Asp Asp Val Leu Asn Ala Arg
        35                  40                  45

Arg Gln Ala Asp Gly Asp Ile Ser Lys Met Leu Gln Met Val Pro Leu
    50                  55                  60

Leu Lys Ala Ser Ile Lys Asp Thr Leu Arg Leu His Pro Ile Ser Val
65                  70                  75                  80

Thr Leu Gln Arg Tyr Pro Asp Ser Asp Leu Val Leu Gln Asp Tyr Leu
                85                  90                  95

Ile Pro Ala Lys Thr Leu Val Gln Val Ala Ile Tyr Ala Met Gly Arg
            100                 105                 110

Asp Pro Ala Phe Phe Ser Ser Pro Asp Lys Phe Asp Pro Thr Arg Trp
        115                 120                 125

Leu Ser Lys Asp Lys Asp Leu Ile His Phe Arg Asn Leu Gly Phe Gly
    130                 135                 140

Trp Gly Val Arg Gln Cys Val Gly Arg Arg Ile Ala Asp Leu Asp Met
145                 150                 155                 160

Thr Leu Phe Leu Ile His Ile Leu Asp Asn Phe Lys Val Asp Met Gln
                165                 170                 175

His Ile Gly Asp Val Asp Thr Ile Phe Asn Leu Ile Leu Thr Pro Asp
            180                 185                 190

Lys Pro Ile Phe Leu Val Phe Arg Pro Phe Asn Gln Asp Pro Pro Gln
        195                 200                 205

Ala
```

(2) INFORMATION FOR SEQ ID NO:4
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 204 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4

```
Ser Leu Asp Ala Ile Lys Ala Asn Ser Met Asp Leu Thr Ala Gly Ser
 1               5                  10                  15

Val Asp Thr Thr Ala Phe Pro Leu Leu Met Thr Leu Phe Asp Leu Ala
            20                  25                  30

Arg Asn Pro Asp Val Gln Gln Ile Leu Arg Gln Asp Ser Leu Ala Ala
        35                  40                  45

Ala Ala Ser Ile Ser Asp His Pro Gln Lys Ala Thr Thr Asp Leu Pro
    50                  55                  60

Leu Leu Arg Ala Ala Leu Lys Asp Thr Leu Arg Leu Tyr Pro Val Gly
65                  70                  75                  80

Leu Phe Leu Asp Arg Val Val Ser Ser Asp Leu Val Leu Gln Asn Tyr
                85                  90                  95

His Ile Pro Ala Gly Thr Leu Val Gln Val Phe Leu Tyr Ser Leu Gly
            100                 105                 110

Arg Asn Ala Ala Leu Phe Pro Arg Pro Asp Arg Tyr Asn Pro Gln Arg
        115                 120                 125

Trp Leu Asp Ile Arg Gly Ser Gly Arg Asn Phe His His Val Pro Phe
    130                 135                 140

Gly Phe Gly Met Arg Gln Cys Leu Gly Arg Arg Leu Ala Asp Val Asp
145                 150                 155                 160

Met Leu Leu Leu Leu His His Val Leu Lys His Phe Leu Val Asp Thr
                165                 170                 175

Leu Thr Gln Asp Asp Ile Lys Met Val Tyr Ser Phe Ile Leu Arg Pro
            180                 185                 190

Gly Thr Ser Pro Leu Leu Thr Phe Arg Ala Ile Asn
        195                 200
```

(2) INFORMATION FOR SEQ ID NO:5
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 211 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5

```
Ser Pro Arg Asp Ala Met Gly Ser Leu Pro Asp Leu Leu Met Ala Gly
 1               5                  10                  15

Val Asp Thr Thr Ser Asn Thr Leu Thr Trp Ala Leu Tyr His Leu Ser
            20                  25                  30
```

```
Lys Asp Pro Asp Ile Gln Asp Ala Leu His Asp Asp Val Val Gly Val
        35                  40                  45

Val Pro Ala Gly Gln Val Pro Gln His Lys Asp Phe Ala His Met Pro
    50                  55                  60

Leu Leu Lys Ala Val Leu Lys Asp Thr Leu Arg Leu Tyr Pro Val Val
65                  70                  75                  80

Pro Thr Asn Ser Arg Ile Ile Asp Lys Asp Ile Asp Val Asp Gly Phe
                85                  90                  95

Leu Phe Pro Lys Asn Thr Gln Phe Val Phe Cys His Tyr Val Val Ser
            100                 105                 110

Arg Asp Pro Thr Ala Phe Ser Asp Pro Asp Ser Phe Gln Pro His Arg
        115                 120                 125

Trp Leu Arg Asn Ser Gln Pro Ala Thr Pro Arg Ile Gln His Pro Phe
    130                 135                 140

Gly Ser Val Pro Phe Gly Tyr Gly Val Arg Ala Cys Leu Gly Arg Arg
145                 150                 155                 160

Ile Ala Asp Leu Asp Met Gln Leu Leu Leu Ala Arg Leu Ile Gln Lys
                165                 170                 175

Tyr Lys Val Val Leu Ala Pro Asp Thr Gly Asp Leu Lys Ser Val Ala
            180                 185                 190

Arg Ile Val Leu Val Pro Asn Lys Lys Val Gly Leu Gln Phe Leu Gln
        195                 200                 205

Arg Gln Cys
    210
```

(2) INFORMATION FOR SEQ ID NO:6

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 203 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6

```
Ser Lys Lys Asp Leu Tyr Ala Ala Val Thr Asp Leu Gln Leu Ala Ala
 1               5                  10                  15

Val Asp Thr Thr Ala Asn Ser Leu Met Trp Ile Leu Tyr Asn Leu Ser
            20                  25                  30

Arg Asn Pro Gln Val Gln Gln Lys Leu Leu Lys Asp Ile Gln Ser Val
        35                  40                  45

Leu Pro Asp Asn Gln Arg Pro Arg Asp Asp Asp Leu Arg Asn Met Pro
    50                  55                  60

Tyr Leu Lys Ala Cys Leu Lys Asp Ser Met Arg Leu Thr Pro Gly Val
65                  70                  75                  80

Pro Phe Thr Thr Arg Thr Leu Asp Lys Ala Thr Val Leu Gly Asp Tyr
                85                  90                  95
```

```
Ala Leu Pro Lys Gly Thr Val Leu Met Leu Asn Thr Gln Val Leu Gly
            100                 105                 110

Ser Ser Asp Asp Asn Phe Asp Asp Ser Ser Gln Phe Arg Pro Asp Arg
        115                 120                 125

Trp Leu Gln Asp Lys Asp Lys Ile Asn Pro Phe Ala His Leu Pro Phe
    130                 135                 140

Gly Val Gly Lys Arg Met Cys Ile Gly Arg Arg Leu Ala Asp Leu Gln
145                 150                 155                 160

Leu His Leu Ala Leu Cys Trp Ile Val Arg Lys Tyr Asp Ile Gln Ala
                165                 170                 175

Thr Asp Asn Asp Pro Val Asp Met Leu His Ser Gly Thr Leu Val Pro
            180                 185                 190

Ser Arg Asp Leu Pro Ile Ala Phe Cys Gln Arg
        195                 200
```

(2) INFORMATION FOR SEQ ID NO:7

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 212 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7

```
Asp Met Gln Ala Leu Lys Gln Ser Ser Thr Asp Leu Leu Phe Gly Gly
 1               5                  10                  15

His Asp Thr Thr Ala Ser Ala Ala Thr Ser Leu Ile Thr Tyr Leu Gly
            20                  25                  30

Leu Tyr Pro His Val Leu Gln Lys Val Arg Asp Asp Leu Lys Ser Lys
        35                  40                  45

Gly Leu Leu Cys Lys Ser Asn Gln Asp Asn Lys Leu Asp Met Asp Ile
    50                  55                  60

Leu Asp Gln Leu Lys Tyr Ile Gly Cys Val Ile Lys Asp Thr Leu Arg
65                  70                  75                  80

Leu Asn Pro Pro Val Pro Gly Gly Phe Arg Val Ala Leu Lys Thr Phe
                85                  90                  95

Asp Leu Asn Gly Tyr Gln Ile Pro Lys Gly Trp Asn Val Ile Tyr Ser
            100                 105                 110

Ile Cys Asp Thr His Asp Val Ala Asp Ile Phe Thr Asn Lys Asp Asp
        115                 120                 125

Phe Asn Pro Asp Arg Phe Ser Ala Pro His Pro Asp Asp Ala Ser Arg
    130                 135                 140

Phe Ser Phe Ile Pro Phe Gly Gly Gly Leu Arg Ser Cys Val Gly Lys
145                 150                 155                 160

Asp Phe Ala Lys Ile Leu Leu Lys Ile Phe Thr Val Asp Leu Ala Arg
                165                 170                 175
```

```
His Cys Asp Trp Gln Leu Leu Asn Gly Pro Pro Thr Met Lys Thr Ser
            180                 185                 190

Pro Thr Val Tyr Pro Val Asp Asn Leu Pro Ala Arg Phe Thr His Phe
        195                 200                 205

His Gly Asp Ile
    210
```

(2) INFORMATION FOR SEQ ID NO:8

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 181 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8

```
Thr Ser Asp Asp Ala Lys Arg Met Cys Gly Leu Leu Leu Val Gly Gly
 1               5                  10                  15

Leu Asp Thr Val Val Asn Phe Leu Ser Phe Ser Met Asp Phe Leu Ala
            20                  25                  30

Lys Ser Pro Asp His Arg Gln Asp Leu Ile Asp Arg Pro Asp Arg Ile
        35                  40                  45

Pro Ala Ala Cys Asp Asp Leu Leu Arg Arg Phe Ser Leu Val Ala Asp
    50                  55                  60

Gly Arg Ile Leu Thr Ser Asp Tyr Asp Phe His Gly Val Gln Leu Lys
65                  70                  75                  80

Lys Gly Asp Gln Ile Leu Leu Pro Gln Met Leu Ser Gly Leu Asp Asp
                85                  90                  95

Arg Asp Asn Ala Cys Pro Met His Val Asp Phe Ser Arg Gln Lys Val
            100                 105                 110

Ser His Thr Thr Phe Gly His Gly Ser His Leu Cys Leu Gly Gln His
        115                 120                 125

Leu Ala Arg Arg Asp Ile Ile Val Thr Leu Lys Asp Trp Leu Thr Arg
    130                 135                 140

Ile Pro Asp Phe Ser Ile Ala Pro Gly Ala Gln Ile Gln His Lys Ser
145                 150                 155                 160

Gly Ile Val Ser Gly Val Gln Ala Leu Pro Leu Val Trp Asp Pro Ala
                165                 170                 175

Thr Thr Lys Ala Val
            180
```

(2) INFORMATION FOR SEQ ID NO:9

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9

```
Asp Asp Asp Asn Ile Arg Tyr Gln Ile Ile Thr Phe Leu Ile Ala Gly
 1               5                  10                  15

His Asp Thr Thr Ser Gly Leu Leu Ser Phe Ala Leu Tyr Phe Leu Val
            20                  25                  30

Lys Asn Pro His Val Leu Gln Lys Ala Ala Asp Asp Ala Ala Arg Val
        35                  40                  45

Leu Val Asp Pro Val Pro Ser Tyr Lys Gln Val Lys Gln Leu Lys Tyr
    50                  55                  60

Val Gly Met Val Leu Asn Asp Ala Leu Arg Leu Trp Pro Thr Ala Pro
65                  70                  75                  80

Ala Phe Ser Leu Tyr Ala Lys Asp Asp Thr Val Leu Gly Gly Asp Tyr
                85                  90                  95

Pro Leu Asp Lys Gly Asp Asp Leu Met Val Leu Ile Pro Gln Leu His
            100                 105                 110

Arg Asp Lys Thr Ile Trp Gly Asp Asp Val Asp Asp Phe Arg Pro Asp
        115                 120                 125

Arg Phe Asp Asn Pro Ser Ala Ile Pro Gln His Ala Phe Lys Pro Phe
    130                 135                 140

Gly Asn Gly Gln Arg Ala Cys Ile Gly Gln Gln Phe Ala Leu His Asp
145                 150                 155                 160

Ala Thr Leu Val Leu Gly Met Met Leu Lys His Phe Asp Phe Asp Asp
                165                 170                 175

His Thr Asn Tyr Asp Leu Asp Ile Lys Asp Thr Leu Thr Leu Lys Pro
            180                 185                 190

Asp Gly Phe Val Val Lys Ala Lys Ser Lys Lys Ile Pro Leu Gly Gly
        195                 200                 205

Ile Pro Ser Pro Ser Thr Asp Gln Ser Ala Lys Lys Val Arg
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 176 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10

```
Asp Asp Lys Tyr Ile Asn Ala Tyr Tyr Val Ala Ile Ala Thr Ala Gly
 1               5                  10                  15

His Asp Thr Thr Ser Ser Ser Ser Gly Gly Ala Ile Ile Gly Leu Ser
            20                  25                  30

Arg Asn Pro Asp Gln Leu Ala Leu Ala Lys Ser Asp Pro Ala Leu Ile
        35                  40                  45
```

```
Pro Arg Leu Val Asp Asp Ala Val Arg Trp Thr Ala Pro Val Lys Ser
    50                  55                  60

Phe Met Arg Thr Ala Leu Ala Asp Thr Asp Val Arg Gly Gln Asn Ile
65                  70                  75                  80

Lys Arg Gly Asp Arg Ile Met Leu Ser Tyr Pro Ser Ala Asn Arg Asp
                85                  90                  95

Asp Asp Val Phe Ser Asn Pro Asp Asp Phe Asp Ile Thr Arg Phe Pro
            100                 105                 110

Asn Arg His Leu Gly Phe Gly Trp Gly Ala His Met Cys Leu Gly Gln
        115                 120                 125

His Leu Ala Lys Leu Asp Met Lys Ile Phe Phe Asp Asp Leu Leu Pro
    130                 135                 140

Lys Leu Lys Ser Val Asp Leu Ser Gly Pro Pro Arg Leu Val Ala Thr
145                 150                 155                 160

Asn Phe Val Gly Gly Pro Lys Asn Val Pro Ile Arg Phe Thr Lys Ala
                165                 170                 175
```

(2) INFORMATION FOR SEQ ID NO:11

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 180 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11

```
Ser Ala Asp Asp Leu Thr Ser Ile Ala Leu Val Leu Leu Leu Ala Gly
 1               5                  10                  15

Phe Asp Ala Ser Val Ser Leu Ile Gly Ile Gly Thr Tyr Leu Leu Leu
            20                  25                  30

Thr His Pro Asp Gln Asp Gln Leu Ala Leu Val Arg Arg Asp Pro Ser
        35                  40                  45

Ala Leu Pro Asn Ala Val Asp Asp Ile Leu Arg Tyr Ile Ala Pro Pro
    50                  55                  60

Asp Thr Thr Thr Arg Phe Ala Ala Asp Asp Val Asp Ile Arg Gly Val
65                  70                  75                  80

Ala Ile Pro Gln Tyr Ser Thr Val Leu Val Ala Asn Gly Ala Ala Asn
                85                  90                  95

Arg Asp Pro Lys Gln Phe Pro Asp Pro His Arg Phe Asp Val Thr Arg
            100                 105                 110

Asp Thr Arg Gly His Leu Ser Phe Gly Gln Gly Ile His Phe Cys Met
        115                 120                 125

Gly Arg Pro Leu Ala Lys Leu Asp Gly Asp Val Ala Leu Arg Ala Leu
    130                 135                 140
```

Phe Gly Arg Phe Pro Ala Leu Ser Leu Gly Ile Asp Ala Asp Asp Val
145 150 155 160

Val Trp Arg Arg Ser Leu Leu Leu Arg Gly Ile Asp His Leu Pro Val
165 170 175

Arg Leu Asp Gly
180

(2) INFORMATION FOR SEQ ID NO:12
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12

GARASNARDC GN 12

(2) INFORMATION FOR SEQ ID NO:13
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13

AARGARACNA AR 12

(2) INFORMATION FOR SEQ ID NO:14
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14

TTTTTTTTTT TTGG 14

(2) INFORMATION FOR SEQ ID NO:15
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15

TTTTTTTTTT TTGA 14

(2) INFORMATION FOR SEQ ID NO:16
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16

TTTTTTTTTT TTGT 14

(2) INFORMATION FOR SEQ ID NO:17
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17

TTTTTTTTTT TTGC 14

(2) INFORMATION FOR SEQ ID NO:18
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18

TTTTTTTTTT TTAG 14

(2) INFORMATION FOR SEQ ID NO:19
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19

TTTTTTTTTT TTAA 14

(2) INFORMATION FOR SEQ ID NO:20
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20

TTTTTTTTTT TTAT 14

(2) INFORMATION FOR SEQ ID NO:21
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21

TTTTTTTTTT TTAC 14

(2) INFORMATION FOR SEQ ID NO:22
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22

TTTTTTTTTT TTCG 14

(2) INFORMATION FOR SEQ ID NO:23
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23

TTTTTTTTTT TTCA 14

(2) INFORMATION FOR SEQ ID NO:24
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24

TTTTTTTTTT TTCT 14

(2) INFORMATION FOR SEQ ID NO:25
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25

TTTTTTTTTT TTCC 14

(2) INFORMATION FOR SEQ ID NO:26
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26

AAGCGACCGA 10

(2) INFORMATION FOR SEQ ID NO:27
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27

TGTTCGCCAG 10

(2) INFORMATION FOR SEQ ID NO:28
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28

TGCCAGTGGA 10

(2) INFORMATION FOR SEQ ID NO:29
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29

GGCTGCAAAC 10

(2) INFORMATION FOR SEQ ID NO:30
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30

CCTAGCGTTG 10

(2) INFORMATION FOR SEQ ID NO:31
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31

GTAGCGGCCG CTGCCAGTGG A 21

(2) INFORMATION FOR SEQ ID NO:32
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32

GTAGCGGCCG CT 12

(2) INFORMATION FOR SEQ ID NO:33
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33

TGCCAGTGGA CAATCTCCCT ACCAAATTCA CTAGTTATGT CCAGAAATTA GCCTAAACCG 60

GAGCCTTTGT ACATATGTTT TTATTTTAGA TGAACTGTGA TGTATTGGAT ATTTTCTAAT 120

TTGTTTATAT AAAGCAGATG TGTATATAAG TCTATGCGAA GAAGCGAAAA CGAGGGCACT 180

ACTTTCTCAT GGATCACTGT AATGCTACAG AGTGTCTGTG ATGTATATTT ATAATGTAGT 240

TGTGTCATAT AGCTTTTGTA CTGTATGCAA CTTATTTAAC TCGCTCTTTA TCTCATGGGT 300

TTTATTTAAT AAAACATGTT CTTACAAAAA AAAAAAA 337

(2) INFORMATION FOR SEQ ID NO:34
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 492 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34

Met Gly Leu Tyr Thr Leu Met Val Thr Phe Leu Cys Thr Ile Val Leu
1 5 10 15

Pro Val Leu Leu Phe Leu Ala Ala Val Lys Leu Trp Glu Met Leu Met
20 25 30

Ile Arg Arg Val Asp Pro Asn Cys Arg Ser Pro Leu Pro Pro Gly Thr
35 40 45

```
Met Gly Leu Pro Phe Ile Gly Glu Thr Leu Gln Leu Ile Leu Gln Arg
    50                  55                  60

Arg Lys Phe Leu Arg Met Lys Arg Gln Lys Tyr Gly Cys Ile Tyr Lys
65                  70                  75                  80

Thr His Leu Phe Gly Asn Pro Thr Val Arg Val Met Gly Ala Asp Asn
                85                  90                  95

Val Arg Gln Ile Leu Leu Gly Glu His Lys Leu Val Ser Val Gln Trp
            100                 105                 110

Pro Ala Ser Val Arg Thr Ile Leu Gly Ser Asp Thr Leu Ser Asn Val
        115                 120                 125

His Gly Val Gln His Lys Asn Lys Lys Lys Ala Ile Met Arg Ala Phe
    130                 135                 140

Ser Arg Asp Ala Leu Glu His Tyr Ile Pro Val Ile Gln Gln Glu Val
145                 150                 155                 160

Lys Ser Ala Ile Gln Glu Trp Leu Gln Lys Asp Ser Cys Val Leu Val
                165                 170                 175

Tyr Pro Glu Met Lys Lys Leu Met Phe Arg Ile Ala Met Arg Ile Leu
            180                 185                 190

Leu Gly Phe Glu Pro Glu Gln Ile Lys Thr Asp Glu Gln Glu Leu Val
        195                 200                 205

Glu Ala Phe Glu Glu Met Ile Lys Asn Leu Phe Ser Leu Pro Ile Asp
    210                 215                 220

Val Pro Phe Ser Gly Leu Tyr Arg Gly Leu Arg Ala Arg Asn Phe Ile
225                 230                 235                 240

His Ser Lys Ile Glu Glu Asn Ile Arg Lys Lys Ile Gln Asp Asp Asp
                245                 250                 255

Asn Glu Asn Glu Gln Lys Tyr Lys Asp Ala Leu Gln Leu Leu Ile Glu
            260                 265                 270

Asn Ser Arg Arg Ser Asp Glu Pro Phe Ser Leu Gln Ala Met Lys Glu
        275                 280                 285

Ala Ala Thr Glu Leu Leu Phe Gly Gly His Glu Thr Thr Ala Ser Thr
    290                 295                 300

Ala Thr Ser Leu Val Met Phe Leu Gly Leu Asn Thr Glu Val Val Gln
305                 310                 315                 320

Lys Val Arg Glu Glu Val Gln Glu Lys Val Glu Met Gly Met Tyr Thr
                325                 330                 335

Pro Gly Lys Gly Leu Ser Met Glu Leu Leu Asp Gln Leu Lys Tyr Thr
            340                 345                 350

Gly Cys Val Ile Lys Glu Thr Leu Arg Ile Asn Pro Pro Val Pro Gly
        355                 360                 365

Gly Phe Arg Val Ala Leu Lys Thr Phe Glu Leu Asn Gly Tyr Gln Ile
    370                 375                 380
```

```
Pro Lys Gly Trp Asn Val Ile Tyr Ser Ile Cys Asp Thr His Asp Val
385                 390                 395                 400

Ala Asp Val Phe Pro Asn Lys Glu Glu Phe Gln Pro Glu Arg Phe Met
                405                 410                 415

Ser Lys Gly Leu Glu Asp Gly Ser Arg Phe Asn Tyr Ile Pro Phe Gly
            420                 425                 430

Gly Gly Ser Arg Met Cys Val Gly Lys Glu Phe Ala Lys Val Leu Leu
        435                 440                 445

Lys Ile Phe Leu Val Glu Leu Thr Gln His Cys Asn Trp Ile Leu Ser
    450                 455                 460

Asn Gly Pro Pro Thr Met Lys Thr Gly Pro Thr Ile Tyr Pro Val Asp
465                 470                 475                 480

Asn Leu Pro Thr Lys Phe Thr Ser Tyr Val Arg Asn
                485                 490
```

(2) INFORMATION FOR SEQ ID NO:35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35

GTAGCGGCCG CAARGARACD 20

(2) INFORMATION FOR SEQ ID NO:36

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36

GTAGCGGCCG CTTTTTTTTT TTT 23

(2) INFORMATION FOR SEQ ID NO:37

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 351 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37

GAACTCCTCT TTGGAGGACA CGAAACCACG GCCAGTGCAG CCACATCTCT GATCACTTAC 60

CTGGGGCTCT ACCCACATGT TCTCCAGAAA GTGCGAGAAG AGCTGAAGAG TAAGGGTTTA 120

CTTTGCAAGA GCAATCAAGA CAACAAGTTG GACATGGAAA TTTTGGAACA ACTTAAATAC 180

ATCGGGTGTG TTATTAAGGA GACCCTTCGA CTGAATCCCC CAGTTCCAGG AGGGTTTCGG 240

GTTGCTCTGA AGACTTTTGA ATTAAATGGA TACCAGATTC CCAAGGGCTG GAATGTTATC 300

TACAGTATCT GTGATACTCA TGATGTGGCA GAGATCTTCA CCAACAAGGA A 351

(2) INFORMATION FOR SEQ ID NO:38
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1494 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38

```
ATG GGG CTC CCG GCG CTG CTG GCC AGT GCG CTC TGC ACC TTC GTG CTG          48
Met Gly Leu Pro Ala Leu Leu Ala Ser Ala Leu Cys Thr Phe Val Leu
 1               5                  10                  15

CCG CTG CTG CTC TTC CTG GCT GCG ATC AAG CTC TGG GAC CTG TAC TGC          96
Pro Leu Leu Leu Phe Leu Ala Ala Ile Lys Leu Trp Asp Leu Tyr Cys
            20                  25                  30

GTG AGC GGC CGC GAC CGC AGT TGT GCC CTC CCA TTG CCC CCC GGG ACT         144
Val Ser Gly Arg Asp Arg Ser Cys Ala Leu Pro Leu Pro Pro Gly Thr
        35                  40                  45

ATG GGC TTC CCC TTC TTT GGG GAA ACC TTG CAG ATG GTA CTG CAG CGG         192
Met Gly Phe Pro Phe Phe Gly Glu Thr Leu Gln Met Val Leu Gln Arg
    50                  55                  60

AGG AAG TTC CTG CAG ATG AAG CGC AGG AAA TAC GGC TTC ATC TAC AAG         240
Arg Lys Phe Leu Gln Met Lys Arg Arg Lys Tyr Gly Phe Ile Tyr Lys
65                  70                  75                  80

ACG CAT CTG TTC GGG CGG CCC ACC GTA CGG GTG ATG GGC GCG GAC AAT         288
Thr His Leu Phe Gly Arg Pro Thr Val Arg Val Met Gly Ala Asp Asn
                85                  90                  95

GTG CGG CGC ATC TTG CTC GGA GAC GAC CGG CTG GTG TCG GTC CAC TGG         336
Val Arg Arg Ile Leu Leu Gly Asp Asp Arg Leu Val Ser Val His Trp
            100                 105                 110

CCA GCG TCG GTG CGC ACC ATT CTG GGA TCT GGC TGC CTC TCT AAC CTG         384
Pro Ala Ser Val Arg Thr Ile Leu Gly Ser Gly Cys Leu Ser Asn Leu
        115                 120                 125

CAC GAC TCC TCG CAC AAG CAG CGC AAG AAG GTG ATT ATG CGG GCC TTC         432
His Asp Ser Ser His Lys Gln Arg Lys Lys Val Ile Met Arg Ala Phe
    130                 135                 140

AGC CGC GAG GCA CTC GAA TGC TAC GTG CCG GTG ATC ACC GAG GAA GTG         480
Ser Arg Glu Ala Leu Glu Cys Tyr Val Pro Val Ile Thr Glu Glu Val
145                 150                 155                 160

GGC AGC AGC CTG GAG CAG TGG CTG AGC TGC GGC GAG CGC GGC CTC CTG         528
Gly Ser Ser Leu Glu Gln Trp Leu Ser Cys Gly Glu Arg Gly Leu Leu
                165                 170                 175

GTC TAC CCC GAG GTG AAG CGC CTC ATG TTC CGA ATC GCC ATG CGC ATC         576
Val Tyr Pro Glu Val Lys Arg Leu Met Phe Arg Ile Ala Met Arg Ile
            180                 185                 190
```

```
CTA CTG GGC TGC GAA CCC CAA CTG GCG GGC GAC GGG GAC TCC GAG CAG          624
Leu Leu Gly Cys Glu Pro Gln Leu Ala Gly Asp Gly Asp Ser Glu Gln
        195                 200                 205

CAG CTT GTG GAG GCC TTC GAG GAA ATG ACC CGC AAT CTC TTC TCG CTG          672
Gln Leu Val Glu Ala Phe Glu Glu Met Thr Arg Asn Leu Phe Ser Leu
    210                 215                 220

CCC ATC GAC GTG CCC TTC AGC GGG CTG TAC CGG GGC ATG AAG GCG CGG          720
Pro Ile Asp Val Pro Phe Ser Gly Leu Tyr Arg Gly Met Lys Ala Arg
225                 230                 235                 240

AAC CTC ATT CAC GCG CGC ATC GAG CAG AAC ATT CGC GCC AAG ATC TGC          768
Asn Leu Ile His Ala Arg Ile Glu Gln Asn Ile Arg Ala Lys Ile Cys
                245                 250                 255

GGG CTG CGG GCA TCC GAG GCG GGC CAG GGC TGC AAA GAC GCG CTG CAG          816
Gly Leu Arg Ala Ser Glu Ala Gly Gln Gly Cys Lys Asp Ala Leu Gln
            260                 265                 270

CTG TTG ATC GAG CAC TCG TGG GAG AGG GGA GAG CGG CTG GAC ATG CAG          864
Leu Leu Ile Glu His Ser Trp Glu Arg Gly Glu Arg Leu Asp Met Gln
        275                 280                 285

GCA CTA AAG CAA TCT TCA ACC GAA CTC CTC TTT GGA GGA CAC GAA ACC          912
Ala Leu Lys Gln Ser Ser Thr Glu Leu Leu Phe Gly Gly His Glu Thr
    290                 295                 300

ACG GCC AGT GCA GCC ACA TCT CTG ATC ACT TAC CTG GGG CTC TAC CCA          960
Thr Ala Ser Ala Ala Thr Ser Leu Ile Thr Tyr Leu Gly Leu Tyr Pro
305                 310                 315                 320

CAT GTT CTC CAG AAA GTG CGA GAA GAG CTG AAG AGT AAG GGT TTA CTT         1008
His Val Leu Gln Lys Val Arg Glu Glu Leu Lys Ser Lys Gly Leu Leu
                325                 330                 335

TGC AAG AGC AAT CAA GAC AAC AAG TTG GAC ATG GAA ATT TTG GAA CAA         1056
Cys Lys Ser Asn Gln Asp Asn Lys Leu Asp Met Glu Ile Leu Glu Gln
            340                 345                 350

CTT AAA TAC ATC GGG TGT GTT ATT AAG GAG ACC CTT CGA CTG AAT CCC         1104
Leu Lys Tyr Ile Gly Cys Val Ile Lys Glu Thr Leu Arg Leu Asn Pro
        355                 360                 365

CCA GTT CCA GGA GGG TTT CGG GTT GCT CTG AAG ACT TTT GAA TTA AAT         1152
Pro Val Pro Gly Gly Phe Arg Val Ala Leu Lys Thr Phe Glu Leu Asn
    370                 375                 380

GGA TAC CAG ATT CCC AAG GGC TGG AAT GTT ATC TAC AGT ATC TGT GAT         1200
Gly Tyr Gln Ile Pro Lys Gly Trp Asn Val Ile Tyr Ser Ile Cys Asp
385                 390                 395                 400

ACT CAT GAT GTG GCA GAG ATC TTC ACC AAC AAG GAA GAA TTT AAT CCT         1248
Thr His Asp Val Ala Glu Ile Phe Thr Asn Lys Glu Glu Phe Asn Pro
                405                 410                 415

GAC CGA TTC AGT GCT CCT CAC CCA GAG GAT GCA TCC AGG TTC AGC TTC         1296
Asp Arg Phe Ser Ala Pro His Pro Glu Asp Ala Ser Arg Phe Ser Phe
            420                 425                 430

ATT CCA TTT GGA GGA GGC CTT AGG AGC TGT GTA GGC AAA GAA TTT GCA         1344
Ile Pro Phe Gly Gly Gly Leu Arg Ser Cys Val Gly Lys Glu Phe Ala
        435                 440                 445
```

```
AAA ATT CTT CTC AAA ATA TTT ACA GTG GAG CTG GCC AGG CAT TGT GAC        1392
Lys Ile Leu Leu Lys Ile Phe Thr Val Glu Leu Ala Arg His Cys Asp
    450                 455                 460

TGG CAG CTT CTA AAT GGA CCT CCT ACA ATG AAA ACC AGT CCC ACC GTG        1440
Trp Gln Leu Leu Asn Gly Pro Pro Thr Met Lys Thr Ser Pro Thr Val
465                 470                 475                 480

TAT CCT GTG GAC AAT CTC CCT GCA AGA TTC ACC CAT TTC CAT GGG GAA        1488
Tyr Pro Val Asp Asn Leu Pro Ala Arg Phe Thr His Phe His Gly Glu
                485                 490                 495

ATC TGA                                                                1494
Ile
```

(2) INFORMATION FOR SEQ ID NO:39

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 497 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39

```
Met Gly Leu Pro Ala Leu Leu Ala Ser Ala Leu Cys Thr Phe Val Leu
 1               5                  10                  15

Pro Leu Leu Leu Phe Leu Ala Ala Ile Lys Leu Trp Asp Leu Tyr Cys
            20                  25                  30

Val Ser Gly Arg Asp Arg Ser Cys Ala Leu Pro Leu Pro Pro Gly Thr
        35                  40                  45

Met Gly Phe Pro Phe Phe Gly Glu Thr Leu Gln Met Val Leu Gln Arg
    50                  55                  60

Arg Lys Phe Leu Gln Met Lys Arg Arg Lys Tyr Gly Phe Ile Tyr Lys
65                  70                  75                  80

Thr His Leu Phe Gly Arg Pro Thr Val Arg Val Met Gly Ala Asp Asn
                85                  90                  95

Val Arg Arg Ile Leu Leu Gly Asp Asp Arg Leu Val Ser Val His Trp
            100                 105                 110

Pro Ala Ser Val Arg Thr Ile Leu Gly Ser Gly Cys Leu Ser Asn Leu
        115                 120                 125

His Asp Ser Ser His Lys Gln Arg Lys Lys Val Ile Met Arg Ala Phe
    130                 135                 140

Ser Arg Glu Ala Leu Glu Cys Tyr Val Pro Val Ile Thr Glu Glu Val
145                 150                 155                 160

Gly Ser Ser Leu Glu Gln Trp Leu Ser Cys Gly Glu Arg Gly Leu Leu
                165                 170                 175

Val Tyr Pro Glu Val Lys Arg Leu Met Phe Arg Ile Ala Met Arg Ile
            180                 185                 190

Leu Leu Gly Cys Glu Pro Gln Leu Ala Gly Asp Gly Asp Ser Glu Gln
        195                 200                 205
```

```
Gln Leu Val Glu Ala Phe Glu Glu Met Thr Arg Asn Leu Phe Ser Leu
    210                 215                 220

Pro Ile Asp Val Pro Phe Ser Gly Leu Tyr Arg Gly Met Lys Ala Arg
225                 230                 235                 240

Asn Leu Ile His Ala Arg Ile Glu Gln Asn Ile Arg Ala Lys Ile Cys
                245                 250                 255

Gly Leu Arg Ala Ser Glu Ala Gly Gln Gly Cys Lys Asp Ala Leu Gln
            260                 265                 270

Leu Leu Ile Glu His Ser Trp Glu Arg Gly Glu Arg Leu Asp Met Gln
        275                 280                 285

Ala Leu Lys Gln Ser Ser Thr Glu Leu Leu Phe Gly Gly His Glu Thr
    290                 295                 300

Thr Ala Ser Ala Ala Thr Ser Leu Ile Thr Tyr Leu Gly Leu Tyr Pro
305                 310                 315                 320

His Val Leu Gln Lys Val Arg Glu Glu Leu Lys Ser Lys Gly Leu Leu
                325                 330                 335

Cys Lys Ser Asn Gln Asp Asn Lys Leu Asp Met Glu Ile Leu Glu Gln
            340                 345                 350

Leu Lys Tyr Ile Gly Cys Val Ile Lys Glu Thr Leu Arg Leu Asn Pro
        355                 360                 365

Pro Val Pro Gly Gly Phe Arg Val Ala Leu Lys Thr Phe Glu Leu Asn
    370                 375                 380

Gly Tyr Gln Ile Pro Lys Gly Trp Asn Val Ile Tyr Ser Ile Cys Asp
385                 390                 395                 400

Thr His Asp Val Ala Glu Ile Phe Thr Asn Lys Glu Glu Phe Asn Pro
                405                 410                 415

Asp Arg Phe Ser Ala Pro His Pro Glu Asp Ala Ser Arg Phe Ser Phe
            420                 425                 430

Ile Pro Phe Gly Gly Gly Leu Arg Ser Cys Val Gly Lys Glu Phe Ala
        435                 440                 445

Lys Ile Leu Leu Lys Ile Phe Thr Val Glu Leu Ala Arg His Cys Asp
    450                 455                 460

Trp Gln Leu Leu Asn Gly Pro Pro Thr Met Lys Thr Ser Pro Thr Val
465                 470                 475                 480

Tyr Pro Val Asp Asn Leu Pro Ala Arg Phe Thr His Phe His Gly Glu
                485                 490                 495

Ile
```

(2) INFORMATION FOR SEQ ID NO:40
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40

Pro Phe Gly Gly Gly Pro Arg Leu Cys Pro Gly Tyr Glu Leu Ala Arg
1 5 10 15

Val Ala Leu Ser
20

(2) INFORMATION FOR SEQ ID NO:41
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41

Pro Phe Ser Gly Gly Ala Arg Asn Cys Ile Gly Lys Gln Phe Ala Met
1 5 10 15

Ser Glu Met Lys
20

(2) INFORMATION FOR SEQ ID NO:42
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42

Pro Phe Ser Gly Gly Ala Arg Asn Cys Ile Gly Lys Gln Phe Ala Met
1 5 10 15

Asn Glu Leu Lys
20

(2) INFORMATION FOR SEQ ID NO:43
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43

Pro Phe Gly Thr Gly Pro Arg Asn Cys Ile Gly Met Arg Phe Ala Ile
1 5 10 15

Met Asn Met Lys
20

(2) INFORMATION FOR SEQ ID NO:44
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44

```
Pro Phe Ser Gly Gly Ser Arg Asn Cys Ile Gly Lys Gln Phe Ala Met
 1               5                  10                  15

Asn Glu Leu Lys
            20
```

(2) INFORMATION FOR SEQ ID NO:45
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1850 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45

```
TGTCGCCGTT GCTGTCGGTT GCTGTCGGAC GCTGTCTCCT CTCCAGAAGC TTGTTTTTCG        60

TTTTGGCGAT CAGTTGCGCG CTTCAAC ATG GGG CTG TAC ACC CTT ATG GTC ACC       114
                              Met Gly Leu Tyr Thr Leu Met Val Thr
                              1               5

TTT CTC TGC ACC ATC GTG CTA CCC GTT TTA CTC TTT CTC GCC GCG GTG         162
Phe Leu Cys Thr Ile Val Leu Pro Val Leu Leu Phe Leu Ala Ala Val
10                  15                  20                  25

AAG TTG TGG GAG ATG TTA ATG ATC CGA CGA GTC GAT CCG AAC TGC AGA         210
Lys Leu Trp Glu Met Leu Met Ile Arg Arg Val Asp Pro Asn Cys Arg
                30                  35                  40

AGT CCT CTA CCG CCA GGT ACC ATG GGC TTG CCG TTC ATT GGA GAA ACG         258
Ser Pro Leu Pro Pro Gly Thr Met Gly Leu Pro Phe Ile Gly Glu Thr
            45                  50                  55

CTC CAG CTG ATC CTC CAG AGA AGG AAG TTT CTG CGC ATG AAA CGG CAG         306
Leu Gln Leu Ile Leu Gln Arg Arg Lys Phe Leu Arg Met Lys Arg Gln
        60                  65                  70

AAA TAC GGG TGC ATC TAC AAG ACG CAC CTC TTC GGG AAC CCG ACT GTC         354
Lys Tyr Gly Cys Ile Tyr Lys Thr His Leu Phe Gly Asn Pro Thr Val
    75                  80                  85

AGG GTG ATG GGA GCT GAT AAT GTG AGG CAG ATT CTG CTG GGC GAA CAC         402
Arg Val Met Gly Ala Asp Asn Val Arg Gln Ile Leu Leu Gly Glu His
90                  95                  100                 105

AAG CTG GTG TCT GTT CAG TGG CCA GCA TCA GTG AGA ACC ATC CTG GGC         450
Lys Leu Val Ser Val Gln Trp Pro Ala Ser Val Arg Thr Ile Leu Gly
                110                 115                 120

TCT GAC ACC CTC TCC AAT GTC CAT GGA GTT CAA CAC AAA AAC AAG AAA         498
Ser Asp Thr Leu Ser Asn Val His Gly Val Gln His Lys Asn Lys Lys
            125                 130                 135
```

```
AAG GCC ATT ATG AGG GCG TTC TCT CGA GAT GCT CTG GAG CAC TAC ATT          546
Lys Ala Ile Met Arg Ala Phe Ser Arg Asp Ala Leu Glu His Tyr Ile
        140                 145                 150

CCC GTG ATC CAG CAG GAG GTG AAG AGC GCC ATA CAG GAA TGG CTG CAA          594
Pro Val Ile Gln Gln Glu Val Lys Ser Ala Ile Gln Glu Trp Leu Gln
    155                 160                 165

AAA GAC TCC TGC GTG CTG GTT TAT CCA GAA ATG AAG AAA CTC ATG TTT          642
Lys Asp Ser Cys Val Leu Val Tyr Pro Glu Met Lys Lys Leu Met Phe
170                 175                 180                 185

CGG ATA GCT ATG AGA ATC CTG CTT GGT TTT GAA CCA GAG CAA ATA AAG          690
Arg Ile Ala Met Arg Ile Leu Leu Gly Phe Glu Pro Glu Gln Ile Lys
                190                 195                 200

ACG GAC GAG CAA GAA CTG GTG GAA GCT TTT GAG GAA ATG ATC AAA AAC          738
Thr Asp Glu Gln Glu Leu Val Glu Ala Phe Glu Glu Met Ile Lys Asn
            205                 210                 215

TTG TTC TCC TTG CCA ATC GAC GTT CCT TTC AGT GGT CTG TAC AGG GGT          786
Leu Phe Ser Leu Pro Ile Asp Val Pro Phe Ser Gly Leu Tyr Arg Gly
        220                 225                 230

TTG AGG GCA CGC AAT TTC ATT CAC TCC AAA ATT GAG GAA AAC ATC AGG          834
Leu Arg Ala Arg Asn Phe Ile His Ser Lys Ile Glu Glu Asn Ile Arg
    235                 240                 245

AAG AAA ATT CAA GAT GAC GAC AAT GAA AAC GAA CAG AAA TAC AAA GAC          882
Lys Lys Ile Gln Asp Asp Asp Asn Glu Asn Glu gln Lys Tyr Lys Asp
250                 255                 260                 265

GCC CTT CAG CTG TTG ATC GAG AAC AGC AGA AGA AGT GAC GAA CCT TTT          930
Ala Leu Gln Leu Leu Ile Glu Asn Ser Arg Arg Ser Asp Glu Pro Phe
                270                 275                 280

AGT TTG CAG GCG ATG AAA GAA GCA GCT ACA GAG CTT CTA TTT GGA GGT          978
Ser Leu Gln Ala Met Lys Glu Ala Ala Thr Glu Leu Leu Phe Gly Gly
            285                 290                 295

CAT GAA ACC ACC GCC AGC ACT GCA ACC TCA CTT GTC ATG TTT CTG GGT         1026
His Glu Thr Thr Ala Ser Thr Ala Thr Ser Leu Val Met Phe Leu Gly
        300                 305                 310

CTG AAC ACA GAA GTG GTG CAG AAG GTC AGA GAG GAG GTT CAG GAG AAG         1074
Leu Asn Thr Glu Val Val Gln Lys Val Arg Glu Glu Val Gln Glu Lys
    315                 320                 325

GTT GAA ATG GGC ATG TAT ACA CCT GGA AAG GGC TTG AGT ATG GAG CTG         1122
Val Glu Met Gly Met Tyr Thr Pro Gly Lys Gly Leu Ser Met Glu Leu
330                 335                 340                 345

TTG GAC CAG CTG AAG TAC ACT GGA TGT GTG ATT AAA GAG ACT CTT AGA         1170
Leu Asp Gln Leu Lys Tyr Thr Gly Cys Val Ile Lys Glu Thr Leu Arg
                350                 355                 360

ATC AAC CCT CCT GTT CCC GGA GGA TTC AGA GTC GCA CTC AAA ACC TTT         1218
Ile Asn Pro Pro Val Pro Gly Gly Phe Arg Val Ala Leu Lys Thr Phe
            365                 370                 375

GAA TTG AAT GGT TAC CAA ATT CCT AAA GGA TGG AAC GTC ATT TAC AGC         1266
Glu Leu Asn Gly Tyr Gln Ile Pro Lys Gly Trp Asn Val Ile Tyr Ser
        380                 385                 390
```

```
ATC TGT GAC ACG CAC GAT GTG GCC GAC GTC TTT CCA AAC AAA GAG GAG        1314
Ile Cys Asp Thr His Asp Val Ala Asp Val Phe Pro Asn Lys Glu Glu
    395                 400                 405

TTC CAG CCG GAG AGA TTC ATG AGC AAA GGT CTG GAG GAC GGG TCC AGG        1362
Phe Gln Pro Glu Arg Phe Met Ser Lys Gly Leu Glu Asp Gly Ser Arg
410                 415                 420                 425

TTT AAC TAC ATC CCC TTC GGA GGA GGA TCC AGG ATG TGT GTG GGC AAA        1410
Phe Asn Tyr Ile Pro Phe Gly Gly Gly Ser Arg Met Cys Val Gly Lys
                430                 435                 440

GAG TTC GCC AAA GTG TTA CTC AAG ATC TTT TTA GTT GAG TTA ACG CAG        1458
Glu Phe Ala Lys Val Leu Leu Lys Ile Phe Leu Val Glu Leu Thr Gln
            445                 450                 455

CAT TGC AAT TGG ATT CTC TCA AAC GGA CCC CCG ACA ATG AAA ACA GGC        1506
His Cys Asn Trp Ile Leu Ser Asn Gly Pro Pro Thr Met Lys Thr Gly
        460                 465                 470

CCG ACT ATT TAC CCA GTG GAC AAT CTC CCT ACC AAA TTC ACT AGT TAT        1554
Pro Thr Ile Tyr Pro Val Asp Asn Leu Pro Thr Lys Phe Thr Ser Tyr
    475                 480                 485

GTC AGA AAT TAGCCTAACC GGAGCTTTGT ACATATGTTT TTATTTTAGA                1603
Val Arg Asn
490

TGAACTGTGA TGTATTGGAT ATTTTCTATT TTGTTTATAT AAAGCAGATG TGTATATAAG      1663

TCTATGCGAG GAAGCGAAAA CGAGGGCACT ACTTTCTCAT GGATCACTGT AATGCTACAG      1723

AGTGTCTGTG ATGTATATTT ATAATGTAGT TGTGTTATAT AGCTTTTGTA CTGTATGCAA      1783

CTTATTTAAC TCGCTCTTTA TCTCATGGGT TTTATTTAAT AAAACATGTT CTTACAAAAA      1843

AAAAAAA                                                               1850
```

(2) INFORMATION FOR SEQ ID NO:46

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1725 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46

```
GCACGAGGGA GGCTGAAGCG TGCC ATG GGG CTC CCG GCG CTG CTG GCC AGT          51
                           Met Gly Leu Pro Ala Leu Leu Ala Ser
                            1               5

GCG CTC TGC ACC TTC GTG CTG CCG CTG CTG CTC TTC CTG GCG GCG CTC         99
Ala Leu Cys Thr Phe Val Leu Pro Leu Leu Leu Phe Leu Ala Ala Leu
10                  15                  20                  25

AAG CTC TGG GAC CTG TAC TGT GTG AGC AGC CGC GAT CGC AGC TGC GCC        147
Lys Leu Trp Asp Leu Tyr Cys Val Ser Ser Arg Asp Arg Ser Cys Ala
                30                  35                  40
```

```
CTC CCC TTG CCC CCC GGT ACC ATG GGC TTC CCA TTC TTT GGG GAA ACA          195
Leu Pro Leu Pro Pro Gly Thr Met Gly Phe Pro Phe Phe Gly Glu Thr
            45                  50                  55

TTG CAG ATG GTG CTT CAG CGG AGG AAG TTT CTG CAG ATG AAG CGC AGG          243
Leu Gln Met Val Leu Gln Arg Arg Lys Phe Leu Gln Met Lys Arg Arg
        60                  65                  70

AAA TAC GGC TTC ATC TAC AAG ACG CAT CTG TTT GGG CGG CCC ACG GTG          291
Lys Tyr Gly Phe Ile Tyr Lys Thr His Leu Phe Gly Arg Pro Thr Val
    75                  80                  85

CGG GTG ATG GGC GCG GAT AAT GTG CGG CGC ATC TTG CTG GGA GAG CAC          339
Arg Val Met Gly Ala Asp Asn Val Arg Arg Ile Leu Leu Gly Glu His
90                  95                  100                 105

CGG TTG GTG TCG GTG CAC TGG CCC GCG TCG GTG CGC ACC ATC CTG GGC          387
Arg Leu Val Ser Val His Trp Pro Ala Ser Val Arg Thr Ile Leu Gly
                110                 115                 120

GCT GGC TGC CTC TCC AAC CTG CAC GAT TCC TCG CAC AAG CAG CGA AAG          435
Ala Gly Cys Leu Ser Asn Leu His Asp Ser Ser His Lys Gln Arg Lys
            125                 130                 135

AAG GTG ATT ATG CAG GCC TTC AGC CGC GAG GCA CTC CAG TGC TAC GTG          483
Lys Val Ile Met Gln Ala Phe Ser Arg Glu Ala Leu Gln Cys Tyr Val
        140                 145                 150

CTC GTG ATC GCT GAG GAA GTC AGC AGT TGT CTG GAG CAG TGG CTA AGC          531
Leu Val Ile Ala Glu Glu Val Ser Ser Cys Leu Glu Gln Trp Leu Ser
    155                 160                 165

TGC GGC GAG CGC GGC CTC CTG GTC TAC CCC GAG GTG AAG CGC CTC ATG          579
Cys Gly Glu Arg Gly Leu Leu Val Tyr Pro Glu Val Lys Arg Leu Met
170                 175                 180                 185

TTC CGC ATC GCC ATG CGC ATC CTG CTG GGC TGC GAG CCG GGT CCA GCG          627
Phe Arg Ile Ala Met Arg Ile Leu Leu Gly Cys Glu Pro Gly Pro Ala
                190                 195                 200

GGC GGC GGG GAG GAC GAG CAA CAG CTC GTG GAG GCT TTC GAG GAG ATG          675
Gly Gly Gly Glu Asp Glu Gln Gln Leu Val Glu Ala Phe Glu Glu Met
            205                 210                 215

ACC CGC AAT CTC TTC TCT CTT CCC ATT GAC GTG CCC TTT AGC GGC CTG          723
Thr Arg Asn Leu Phe Ser Leu Pro Ile Asp Val Pro Phe Ser Gly Leu
        220                 225                 230

TAC CGG GGC GTG AAG GCG CGG AAC CTT ATA CAC GCG CGC ATC GAG GAG          771
Tyr Arg Gly Val Lys Ala Arg Asn Leu Ile His Ala Arg Ile Glu Glu
    235                 240                 245

AAC ATT CGC GCC AAG ATC CGC CGG CTT CAG GCT ACA GAG CCG GAT GGG          819
Asn Ile Arg Ala Lys Ile Arg Arg Leu Gln Ala Thr Glu Pro Asp Gly
250                 255                 260                 265

GGT TGC AAG GAC GCG CTG CAG CTC CTG ATT GAG CAC TCG TGG GAG AGG          867
Gly Cys Lys Asp Ala Leu Gln Leu Leu Ile Glu His Ser Trp Glu Arg
                270                 275                 280

GGA GAG AGG CTG GAT ATG CAG GCA CTA AAA CAA TCG TCA ACA GAG CTC          915
Gly Glu Arg Leu Asp Met Gln Ala Leu Lys Gln Ser Ser Thr Glu Leu
            285                 290                 295
```

```
CTC TTT GGT GGT CAT GAA ACT ACA GCC AGT GCT GCG ACA TCA CTG ATC          963
Leu Phe Gly Gly His Glu Thr Thr Ala Ser Ala Ala Thr Ser Leu Ile
        300                 305                 310

ACT TAC CTA GGA CTC TAC CCA CAT GTC CTC CAG AAA GTT CGA GAA GAG         1011
Thr Tyr Leu Gly Leu Tyr Pro His Val Leu Gln Lys Val Arg Glu Glu
    315                 320                 325

ATA AAG AGC AAG GGC TTA CTT TGC AAG AGC AAT CAA GAC AAC AAG TTA         1059
Ile Lys Ser Lys Gly Leu Leu Cys Lys Ser Asn Gln Asp Asn Lys Leu
330                 335                 340                 345

GAC ATG GAA ACT TTG GAA CAG CTT AAA TAC ATT GGG TGT GTC ATT AAG         1107
Asp Met Glu Thr Leu Glu Gln Leu Lys Tyr Ile Gly Cys Val Ile Lys
                350                 355                 360

GAG ACC CTG CGA TTG AAT CCT CCG GTT CCA GGA GGG TTT CGG GTT GCT         1155
Glu Thr Leu Arg Leu Asn Pro Pro Val Pro Gly Gly Phe Arg Val Ala
            365                 370                 375

CTG AAG ACT TTT GAG CTG AAT GGA TAC CAG ATC CCC AAG GGC TGG AAT         1203
Leu Lys Thr Phe Glu Leu Asn Gly Tyr Gln Ile Pro Lys Gly Trp Asn
        380                 385                 390

GTT ATT TAC AGT ATC TGT GAC ACC CAC GAT GTG GCA GAT ATC TTC ACT         1251
Val Ile Tyr Ser Ile Cys Asp Thr His Asp Val Ala Asp Ile Phe Thr
    395                 400                 405

AAC AAG GAG GAA TTT AAT CCC GAC CGC TTT ATA GTG CCT CAT CCA GAG         1299
Asn Lys Glu Glu Phe Asn Pro Asp Arg Phe Ile Val Pro His Pro Glu
410                 415                 420                 425

GAT GCT TCC CGG TTC AGC TTC ATT CCA TTT GGA GGA GGC CTT CGG AGC         1347
Asp Ala Ser Arg Phe Ser Phe Ile Pro Phe Gly Gly Gly Leu Arg Ser
                430                 435                 440

TGT GTA GGC AAA GAG TTT GCA AAA ATT CTT CTT AAG ATA TTT ACA GTG         1395
Cys Val Gly Lys Glu Phe Ala Lys Ile Leu Leu Lys Ile Phe Thr Val
            445                 450                 455

GAG CTG GCT AGG CAC TGT GAT TGG CAG CTT CTA AAT GGA CCT CCT ACA         1443
Glu Leu Ala Arg His Cys Asp Trp Gln Leu Leu Asn Gly Pro Pro Thr
        460                 465                 470

ATG AAG ACA AGC CCC ACT GTG TAC CCT GTG GAC AAT CTC CCT GCA AGA         1491
Met Lys Thr Ser Pro Thr Val Tyr Pro Val Asp Asn Leu Pro Ala Arg
    475                 480                 485

TTC ACC TAC TTC CAG GGA GAT ATC TGATAGCTAT TTCAATTCTT                   1535
Phe Thr Tyr Phe Gln Gly Asp Ile
490                 495

GGACTTATTT GAAGTGTATA TTGGTTTTTT TTAAAAATAG TGTCATGTTG ACTTTATTTA      1595

ATTTCTAAAT GTATAGTATG ATATTTATGT GTCTCTACTA CAGTCCCGTG GTCTTTAAAT      1655

ATTAAAATAA TGAATTTGTA TGATTTCCCA ATAAAGTAAA ATTAAAAAGT GAAAAAAAAA      1715

AAAAAAAAAA                                                             1725
```

(2) INFORMATION FOR SEQ ID NO:47
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 497 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47

```
Met Gly Leu Pro Ala Leu Leu Ala Ser Ala Leu Cys Thr Phe Val Leu
 1               5                  10                  15

Pro Leu Leu Leu Phe Leu Ala Ala Leu Lys Leu Trp Asp Leu Tyr Cys
            20                  25                  30

Val Ser Ser Arg Asp Arg Ser Cys Ala Leu Pro Leu Pro Pro Gly Thr
        35                  40                  45

Met Gly Phe Pro Phe Phe Gly Glu Thr Leu Gln Met Val Leu Gln Arg
    50                  55                  60

Arg Lys Phe Leu Gln Met Lys Arg Arg Lys Tyr Gly Phe Ile Tyr Lys
65                  70                  75                  80

Thr His Leu Phe Gly Arg Pro Thr Val Arg Val Met Gly Ala Asp Asn
                85                  90                  95

Val Arg Arg Ile Leu Leu Gly Glu His Arg Leu Val Ser Val His Trp
            100                 105                 110

Pro Ala Ser Val Arg Thr Ile Leu Gly Ala Gly Cys Leu Ser Asn Leu
        115                 120                 125

His Asp Ser Ser His Lys Gln Arg Lys Lys Val Ile Met Gln Ala Phe
    130                 135                 140

Ser Arg Glu Ala Leu Gln Cys Tyr Val Leu Val Ile Ala Glu Glu Val
145                 150                 155                 160

Ser Ser Cys Leu Glu Gln Trp Leu Ser Cys Gly Glu Arg Gly Leu Leu
                165                 170                 175

Val Tyr Pro Glu Val Lys Arg Leu Met Phe Arg Ile Ala Met Arg Ile
            180                 185                 190

Leu Leu Gly Cys Glu Pro Gly Pro Ala Gly Gly Gly Glu Asp Glu Gln
        195                 200                 205

Gln Leu Val Glu Ala Phe Glu Glu Met Thr Arg Asn Leu Phe Ser Leu
    210                 215                 220

Pro Ile Asp Val Pro Phe Ser Gly Leu Tyr Arg Gly Val Lys Ala Arg
225                 230                 235                 240

Asn Leu Ile His Ala Arg Ile Glu Glu Asn Ile Arg Ala Lys Ile Arg
                245                 250                 255

Arg Leu Gln Ala Thr Glu Pro Asp Gly Gly Cys Lys Asp Ala Leu Gln
            260                 265                 270

Leu Leu Ile Glu His Ser Trp Glu Arg Gly Glu Arg Leu Asp Met Gln
        275                 280                 285
```

```
Ala Leu Lys Gln Ser Ser Thr Glu Leu Leu Phe Gly Gly His Glu Thr
    290                 295                 300

Thr Ala Ser Ala Ala Thr Ser Leu Ile Thr Tyr Leu Gly Leu Tyr Pro
305                 310                 315                 320

His Val Leu Gln Lys Val Arg Glu Glu Ile Lys Ser Lys Gly Leu Leu
                325                 330                 335

Cys Lys Ser Asn Gln Asp Asn Lys Leu Asp Met Glu Thr Leu Glu Gln
            340                 345                 350

Leu Lys Tyr Ile Gly Cys Val Ile Lys Glu Thr Leu Arg Leu Asn Pro
        355                 360                 365

Pro Val Pro Gly Gly Phe Arg Val Ala Leu Lys Thr Phe Glu Leu Asn
    370                 375                 380

Gly Tyr Gln Ile Pro Lys Gly Trp Asn Val Ile Tyr Ser Ile Cys Asp
385                 390                 395                 400

Thr His Asp Val Ala Asp Ile Phe Thr Asn Lys Glu Glu Phe Asn Pro
                405                 410                 415

Asp Arg Phe Ile Val Pro His Pro Glu Asp Ala Ser Arg Phe Ser Phe
            420                 425                 430

Ile Pro Phe Gly Gly Gly Leu Arg Ser Cys Val Gly Lys Glu Phe Ala
        435                 440                 445

Lys Ile Leu Leu Lys Ile Phe Thr Val Glu Leu Ala Arg His Cys Asp
    450                 455                 460

Trp Gln Leu Leu Asn Gly Pro Pro Thr Met Lys Thr Ser Pro Thr Val
465                 470                 475                 480

Tyr Pro Val Asp Asn Leu Pro Ala Arg Phe Thr Tyr Phe Gln Gly Asp
                485                 490                 495

Ile
```

(2) INFORMATION FOR SEQ ID NO:48

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 273 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48

```
CGCACCCCAG GAGGCGCGCT CGGAGGGAAG CCGCCACCGC CGCCGCCTCT GCCTCGGCGC   60

GGAACAAACG GTTAAAGATT TTGGGCCASC GCCTCCGCGG GGGGAGGAGC CAGGGGCCCC  120

AATCCCGCAA TTAAAGATGA ACTTTGGGTG AACTAATTGT CTGACCAAGG TAACGTGGGC  180

AGCAACCTGG GCCGCCTATA AAGCGGCAGC GCCGTGGGGT TTGAAGCGCT GGCGGCGGCG  240

GCAGGTGGCG CGGGAGGTCG CGGCGCGCCA TGG                                273
```

(2) INFORMATION FOR SEQ ID NO:49
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 274 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49

CGCACCCCCA GGAGGCGCGC TCAGAGGGAA GCCGCCAGTG CGCCGCCTCT GCCTCGGCGC 60

GGAACAAACG GTTAAAGATT TTTTTGGGCA GCGCCTCGAG GGGGGAGGAG CCAGGGGCCC 120

GATCCGCAAT TAAAGATGAA CTTTGGGTGA ACTAATTTGT CTGACCAAGG TAACGTGGGC 180

AGTAACCTGG GCGGCCTTAT AAAGAGGGCG CGCGGCGGGG TTCGGAGCTA GGGAGGCGGC 240

GGCAGGTGGC GCGGGAGGCT GAAGCGTGCC ATGG 274

(2) INFORMATION FOR SEQ ID NO:50
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 319 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50

TCGGGGGAAT TAACACCTTT TCAAAGTGAA ATCTCAGGAT TGTCTGCCTT CTACAGGAGG 60

TGGTATTAAA ATGCGCCTAT AACAAATGGT TGAGAGTTTG GAGCCGCTTC TGCCCTGTGG 120

GCGGGGCGAG ATGACACCAC AATTAAAGAT GAACTTTGGG TGAACTAATT TATCTGAGGA 180

AGTTAACAGG AGGAGACCTG CGCGCAATGG ATATATAAGG GCGCGCAGGC GAGGACGCCC 240

TCAGTTTGTG CGTAAAGACG CGTCTCCTCT CCAGAAGCTT GTTTTTCGTT TTGGCGATCA 300

GTTGCGCGCT TCAACATGG 319

WHAT IS CLAIMED IS:

1. A method for identifying a nucleotide sequence encoding a cytochrome P450, the cytochrome being inducible or suppressible in cells of a given type by an agent, the method comprising:

exposing a first group of the cells to the agent so as to induce or suppress expression of the cytochrome P450;

isolating first mRNA from the cells;

isolating second mRNA from a second group of the cells which have not been exposed to the agent so as to induce or suppress expression of the cytochrome P450;

amplifying the first and second mRNA, respectively, in the presence of an oligo(dT) based first nucleic acid primer sufficient to prime synthesis from a poly(A) tail and a second nucleic acid primer substantially complementary to a nucleic acid sequence encoding a conserved region of a known cytochrome P450;

displaying amplified products of the first mRNA and amplified products of the second mRNA to detect differences therebetween; and

identifying said nucleotide sequence encoding said inducible or supressible cytochrome P450.

2. The method of claim 1 wherein identifying said nucleotide sequence includes:

preparing a probe from a said amplified product of a said first or second mRNA which displays an enhanced presence relative to a corresponding said product of the other mRNA; and

screening a DNA library with the probe to identify a clone including hybridizing DNA encoding at least a portion of the cytochrome P450 to be identified.

3. The method of claim 1 wherein the second nucleic acid primer is substantially complementary to a nucleic acid sequence encoding a heme binding site of a cytochrome P450.

4. The method of claim 2, further comprising:

preparing a protein encoded by the identified nucleotide sequence; and

testing the protein for cytochrome P450 activity, wherein cytochrome P450 activity indicates that said clone contains a nucleotide sequence encoding the cytochrome P450.

5. The method of claim 1 wherein the first nucleic acid primer includes the sequence 5'-N-oligo (dT)-3'.

6. The method of claim 5 wherein the first nucleic acid primer includes the sequence 5'-NN-oligo(dT)-3'.

7. The method of claim 1 wherein the 5'-end of the first primer consists of NN-oligo(dT).

8. The method of claim 1 wherein the first nucleic acid primer consists of 5'-NN-oligo(dT)-3'.

9. The method of claim 1 wherein the second nucleic acid primer has between about 6 and about 30 nucleotides, or between about 6 and about 20 nucleotides, or between about 6 and about 15 nucleotides, or between about 8 and about 12 nucleotides, or between about 9 and

about 11 nucleotides, or 10 nucleotides.

10. The method of claim 1 wherein amplifying the first mRNA includes subjecting the first mRNA to PCR.

11. The method of claim 10 wherein the first mRNA is reverse transcribed prior to PCR.

12. The method of claim 4, further comprising the step of purifying said protein.

13. A purified protein obtained according to claim 12.

14. The method of claim 2, further comprising isolating the nucleotide sequence encoding the at least a portion of the cytochrome P450.

15. An isolated nucleotide sequence encoding a cytochrome P450, obtained according to the method of claim 14.

16. The method of claim 1 wherein the conserved portion is between 30 and 2000 bases from the 3'-end of the nucleotide sequence encoding the known cytochrome P450.

17. The method of claim 1 wherein the second nucleic primer contains a sequence which encodes an amino acid sequence selected from the group consisting of KETLRM, ETLRM, TLRM, LRM, PERF, ERF, PDHF, DHF, PDRF, DRF, PTRF, TRF, PDKF, DKF, PERW and ERW.

18. The method of claim 1 wherein the agent is a substrate of the cytochrome P450 to be identified.

19. The method of claim 18 wherein the agent is a substrate of the known cytochrome P450.

20. The method of claim 19 wherein the agent is retinoic acid.

21. The method of claim 1 wherein the agent is not a substrate of the cytochrome P450 to be identified.

22. The method of claim 21 wherein the agent is not a substrate of the known cytochrome P450.

23. The method of claim 22 wherein the agent is an interferon, particuarly IFN-γ.

24. A method for identifying a nucleotide sequence encoding a cytochrome P450, the cytochrome being inducible or suppressible in cells of a given type by an agent, the method comprising:

exposing a first group of the cells to the agent so as to induce or suppress expression of the cytochrome P450;

isolating first mRNA from the cells;

isolating second mRNA from a second group of the cells which have not been exposed to the agent so as to induce or suppress expression of the cytochrome P450;

amplifying the first and second mRNA, respectively, in the presence of a first nucleic acid primer substantially complementary to a first nucleic acid sequence encoding a first conserved region of a first known cytochrome P450 and a second nucleic acid primer substantially complementary to a second nucleic acid sequence encoding a second conserved region of a second known cytochrome P450;

displaying amplified products of the first mRNA and amplified products of the second mRNA to detect differences therebetween; and

identifying said nucleotide sequence encoding said inducible or suppressible cytochrome P450.

25. The method of claim 24 wherein identifying said nucleotide sequence includes:

preparing a probe from a said amplified product of a said first or second mRNA which displays an enhanced presence relative to a corresponding said product of the other mRNA; and

screening a DNA library with the probe to identify a clone including hybridizing DNA encoding at least a portion of the cytochrome P450 to be identified.

26. The method of claim 24 wherein the first and second known cytochrome P450s are the same cytochrome P450.

27. The method of claim 26 wherein the first conserved region is spaced from the second conserved region by between about 30 and about 3000 bases.

28. The method of claim 27 wherein the first nucleic acid primer has between about 6 and about 30 nucleotides, or between about 6 and about 20 nucleotdies, or between about 6 and about 15 nucleotides, or between about 8 and about 12 nucleotides, or between about 9 and about 11 nucleotides, or 10 nucleotides.

29. The method of claim 28 wherein the second nucleic acid primer contains between about 6 and about 30 nucleotides, or between about 6 and about 20 nucleotides, or between about 6 and about 15 nucleotides, or between about 8 and about 12 nucleotides, or between about 9 and about 11 nucleotides, or 10 nucleic acids.

30. The method of claim 24 wherein amplifying the first mRNA includes subjecting the first mRNA to PCR.

31. The method of claim 30 wherein the first mRNA is reverse transcribed prior to PCR.

32. The method of claim 24 wherein one of the first and second nucleic acid primers is substantially complementary to a nucleic acid sequence encoding a heme binding site of a cytochrome P450.

33. The method of claim 24, further comprising:

preparing a protein encoded by a nucleic acid sequence of the identified nucleotide sequence; and

testing the protein for cytochrome P450 activity, wherein cytochrome P450 activity indicates that said clone contains a nucleotide sequence encoding the cytochrome P450.

34. The method of claim 33, further comprising the step of purifying said protein.

35. A purified protein obtained according to claim 35.

36. The method of claim 24, further comprising isolating the nucleotide sequence encoding the at least a portion of the cytochrome P450.

37. An isolated nucleotide sequence encoding a cytochrome P450, obtained according to the method of claim 36.

38. The method of claim 24 wherein the agent is a substrate of the cytochrome P450 to be

identified.

39. The method of claim 38 wherein the agent is a substrate of the known cytochrome P450.

40. The method of claim 39 wherein the agent is retinoic acid.

41. The method of claim 24 wherein the agent is not a substrate of the cytochrome P450 to be identified.

42. The method of claim 41 wherein the agent is not a substrate of the known cytochrome P450.

43. The method of claim 42 wherein the agent is an interferon, particuarly IFN-γ.

44. A method of screening a drug for its effect on activity of a protein prepared according to claim 43, comprising exposing a said protein to a said drug and determining the effect on the activity of the protein.

45. The method of claim 44 wherein a substrate of the protein is a retinoic acid and determining the effect on the activity of the protein includes comparing enzymatic activity of the protein on the retinoic acid in the presence of the drug.

46. The method of claim 45 wherein the retinoic acid is all *trans*-retinoic acid.

47. A method of reducing catabolism of retinoic acid in a mammal in need thereof, comprising administering to the mammal an effective amount of a drug screened according to the method of claim 45, wherein the activity of said protein on the retinoic acid is reduced in the presence of the drug.

48. A method for identifying a first nucleotide sequence having native promoter activity in conjunction with a second nucleotide sequence wherein the second nucleotide sequence is identified according to claim 1, comprising:

providing a genomic library constructed to contain the first and second nucleotide sequences;

providing a probe capable of hybridizing to the region of the second nucleotide sequence encoding the N-terminus;

screening the library with the probe to identify hybridizing DNA including the region of the second nucleotide sequence encoding the N-terminus; and

sequencing hybridizing DNA obtained in the screening step to identify the first nucleotide sequence.

49. A method of screening a drug for its effect on the activity of a nucleotide sequence having promoter activity identified according to claim 48, comprising:

providing an expression system containing the nucleotide sequence operatively linked to a reporter gene;

exposing the system to the drug in the presence of an agent which normally induces or suppresses expression of a coding sequence under control of the nucleotide sequence so as to determine the effect of the drug on expression of the reporter gene.

50. The method of claim 49 wherein the agent is retinoic acid.

51. A method of screening a drug for its effect on expression of a nucleotide sequence encoding a cytochrome P450, wherein the sequence has been identified according to claim 1

and is incorporated into an expression system so as to be under control of a nucleotide sequence having promoter activity identified according to claim 50, comprising,

exposing the system to the drug in the presence of an agent which normally induces or suppresses expression of a coding sequence under control of the nucleotide sequence having promoter activity so as to determine the effect of the drug on expression of the nucleotide sequence encoding the cytochrome P450.

52. The method of claim 51, wherein the agent is retinoic acid and retinoic acid normally induces expression of said coding sequence.

53. The method of claim 52 wherein the nucleotide sequence encoding the cytochrome P450 is the coding sequence.

54. A method of screening a drug for its effect on the metabolism of a retinoid by a cytochrome P450 encoded by a nucleotide sequence identified according to claim 1 and incorporated into an expression system so as to be under control of a nucleotide sequence having promoter activity identified according to claim 48, wherein the retinoid normally induces expression of a gene under the control of the nucleotide sequence having promoter activity, comprising:

exposing the system to the drug in the presence of the retinoid so as to determine the effect of the drug on metabolism of the retinoid.

55. A method of screening a drug for its effect on the metabolism of a retinoid by a cytochrome P450 encoded by a nucleotide sequence identified according to claim 1 and incorporated into an expression system so as to be under control of the nucleotide sequence having native promoter activity identified according to claim 50, wherein the retinoid normally induces expression of the nucleotide sequence identified according to claim 1, comprising:

exposing the system to the drug in the presence of the retinoid so as to determine the effect of the drug on metabolism of the retinoid.

56. The method of claim 5 wherein the retinoid is retinoic acid.

57. The method of claim 55 wherein the retinoid is retinoic acid.

58. The method of claim 56 or claim 57 wherein the retinoic acid all *trans*-retinoic acid.

59. A drug screened according to any of claims 44, 45, 46, 49, 50, 52, 53, 56 or 57.

60. A drug screened according to claim 51.

61. A drug screened according to claim 54.

62. A drug screened according to claim 55.

63. A drug screened according to claim 58.

64. A transfected cell line capable of expressing a nucleotide sequence identified according to any of claims 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 38, 39, 40, 41, 42 or 43.

65. A cell line of claim 66 wherein the cell line is stably transfected.

66. A method of claim 1 or 24 wherein the agent induces expression of the cytochrome P450.

67. The method of claim 45 wherein determining the effect on the activity of the protein includes comparing enzymatic activity of the protein on the retinoic acid in the presence and absence of the drug.

1/35

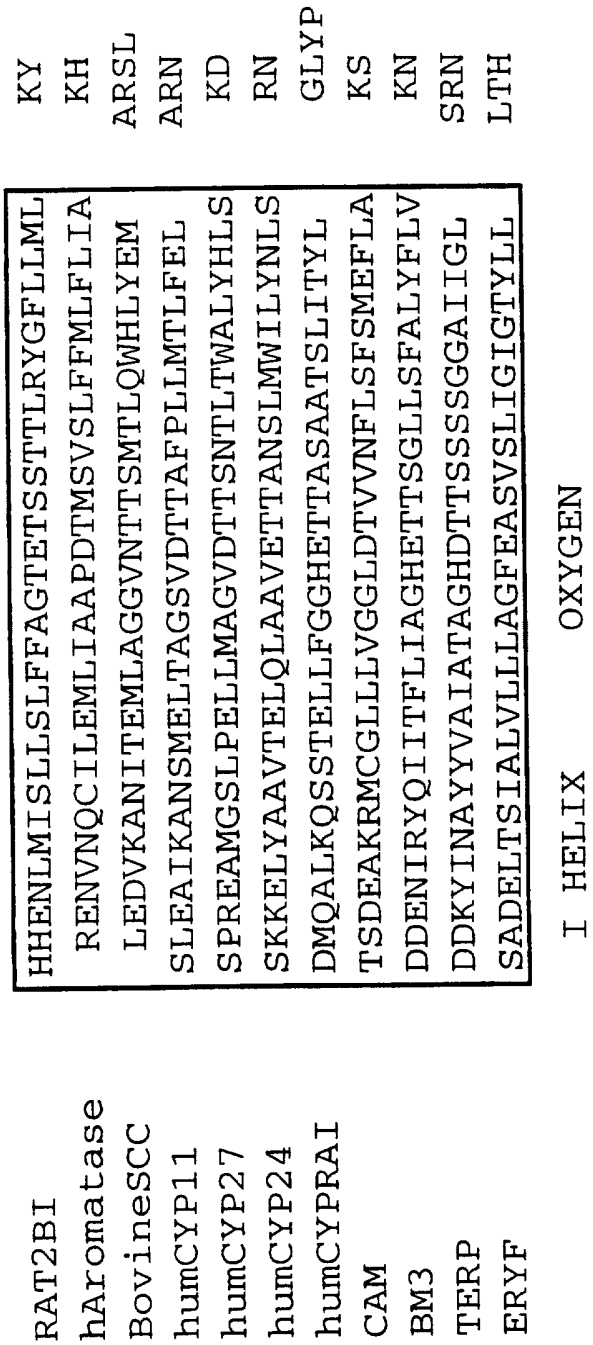

RAT2BI
hAromatase
BovineSCC
humCYP11
humCYP27
humCYP24
humCYPRAI
CAM
BM3
TERP
ERYF
HHENLMISLLSLFFAGTETSSTTLRYGFLLML
RENVNQCILEMLIAAPDTMSVSLFFMLFLIA
LEDVKANITEMLAGGVNTTSMTLQWHLYEM
SLEAIKANSMELTAGSVDTTAFPLLMTLFEL
SPREAMGSLPELLMAGVDTTSNTLTWALYHLS
SKKELYAAVTELQLAAVETTANSLMWILYNLS
DMQALKQSSTELLFGGHETTASAATSLITYL
TSDEAKRMCGLLLVGGLDTVVNFLSFSMEFLA
DDENIRYQIITFLIAGHETTSGLLSFALYFLV
DDKYINAYYVAIATAGHDTTSSSSGGAIIGL
SADELTSIALVLLLAGFEASVSLIGIGTYLL
I HELIX
OXYGEN
KY
KH
ARSL
ARN
KD
RN
GLYP
KS
KN
SRN
LTH

FIG. 1A

RAT2BI	PHVAEKVQKEIDQVI	GSHRLPT
hAromatase	PNVEEAIIKEIQTVI	GERDI
BovineSCC	NVQEMLREEVLNARR	QAEGDI
humCYP11	PDVQQILRQESLAAA	ASISE
humCYP27	PEIQEALHEEVVGVV	PAGQVPQ
humCYP24	PQVQQKLLKEIQSVL	PENQRPR
humCYPRAI	HVLQKVREELKSKGLL	CKSNQDNKLD
CAM	PEHRQELIER	PE
BM3	PHVLQKAAEEAARVLV	DPVPS
TERP	PEQLALAKS	DP
ERYF	PDQDQLALVRR	DPS

HELICAL REGION

FIG. 1B

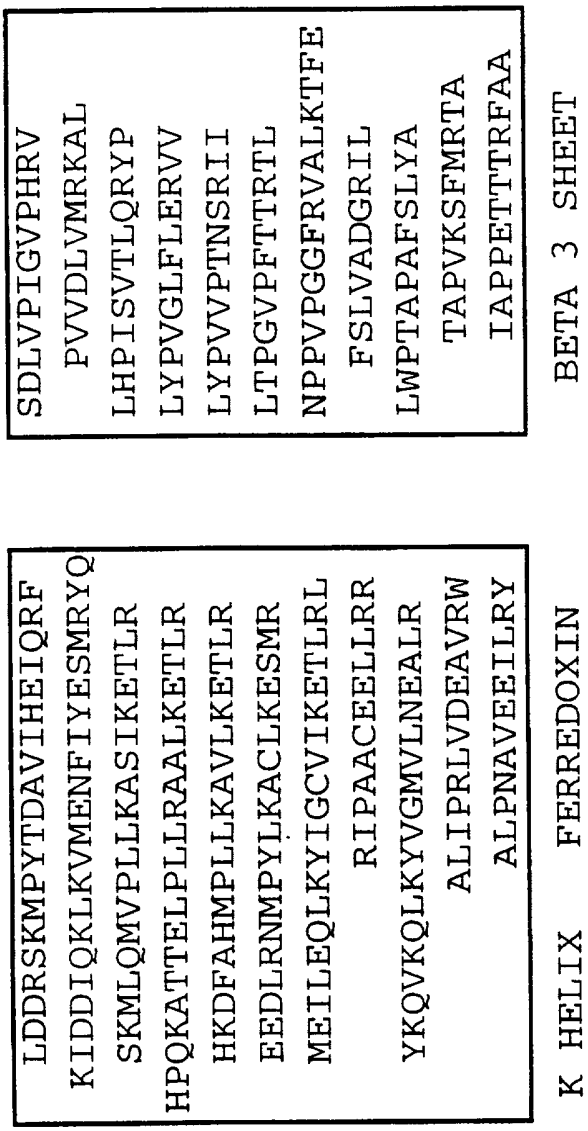

SDLVPIGVPHRV
PVVDLVMRKAL
LHPISVTLQRYP
LYPVGLFLERVV
LYPVVPTNSRII
LTPGVPFTTRTL
NPPVPGGFRVALKTFE
FSLVADGRIL
LWPTAPAFSLYA
TAPVKSFMRTA
IAPPETTTRFAA
BETA 3 SHEET
LDDRSKMPYTDAVIHEIQRF
KIDDIQKLKVMENFIYESMRYQ
SKMLQMVPLLKASIKETLR
HPQKATTELPLLRAALKETLR
HKDFAHMPLLKAVLKETLR
EEDLRNMPYLKACLKESMR
MEILEQLKYIGCVIKETLRL
RIPAACEELLRR
YKQVKQLKYVGMVLNEALR
ALIPRLVDEAVRW
ALPNAVEEILRY
K HELIX FERREDOXIN
RAT2BI
hAromatase
BovineSCC
humCYP11
humCYP27
humCYP24
humCYPRAI
CAM
BM3
TERP
ERYF

FIG. 1C

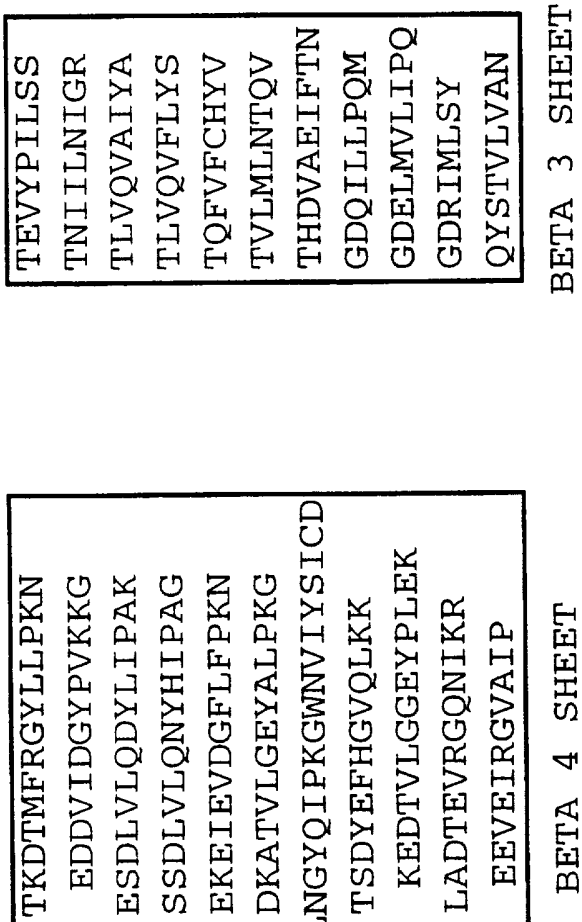
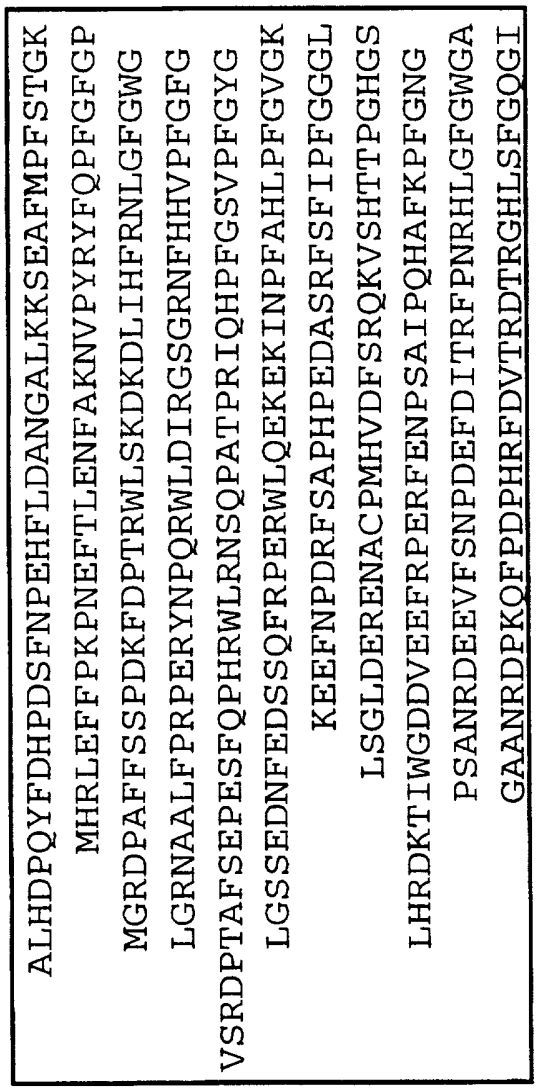

TEVYPILSS
TNIILNIGR
TLVQVAIYA
TLVQVFLYS
TQFVFCHYV
TVLMLNTQV
THDVAEIFTN
GDQILLPQM
GDELMVLIPQ
GDRIMLSY
QYSTVLVAN
BETA 3 SHEET
TKDTMFRGYLLPKN
EDDVIDGYPVKKG
ESDLVLQDYLIPAK
SSDLVLQNYHIPAG
EKEIEVDGFLFPKN
DKATVLGEYALPKG
LNGYQIPKGWNVIYSICD
TSDYEFHGVQLKK
KEDTVLGGEYPLEK
LADTEVRGQNIKR
EEVEIRGVAIP
BETA 4 SHEET
RAT2BI
hAromatase
BovineSCC
humCYP11
humCYP27
humCYP24
humCYPRAI
CAM
BM3
TERP
ERYF

FIG. 1D



RAT2BI
hAromatase
BovineSCC
humCYP11
humCYP27
humCYP24
humCYPRAI
CAM
BM3
TERP
ERYF

FIG. 1E

GEGIARNELFLFFTTILQNF
GKYIAMVMMKAILVTLLRR
GRRIAELEMTLFLIHILENF
GRRLAEVEMLLLLHHVLKH
GRRIAELEMQLLLARLIQK
GRRLAELQLHLALCWIVRKY
GKEFAKILLKIFTVELARH
GQHLARREIIVTLKEWLTRI
GQQFALHEATLVLGMMLKH
GQHLAKLEMKIFFEELLPK
GRPLAKLEGEVALRALFGRF
L HELIX
RICL
RGCA
VRQCV
MRQCL
VRACL
RMCI
RSCV
HLCL
QRACI
HMCL
HFCM
THIOLOATE
RAT2BI
hAromatase
BovineSCC
humCYP11
humCYP27
humCYP24
humCYPRAI
CAM
BM3
TERP
ERYF

FIG. 1F

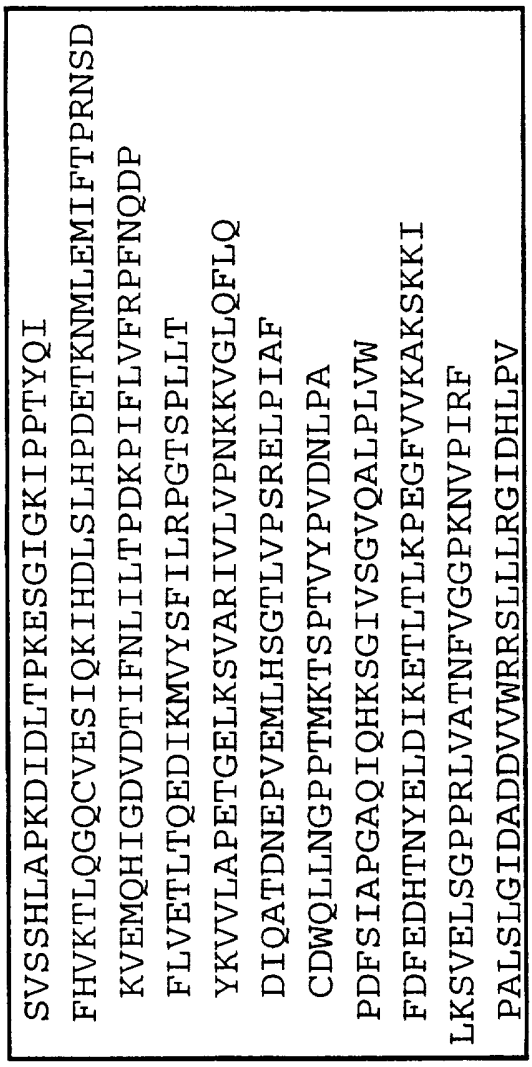

SVSSHLAPKDIDLTPKESGIGKIPPTYQI
FHVKTLQGQCVESIQKIHDLSLHPDETKNMLEMIFTPRNSD
KVEMQHIGDVDTIFNLILTPDKPIFLVFRPFNQDP
FLVETLTQEDIKMVYSFILRPGTSPLLT
YKVVLAPETGELKSVARIVLVPNKKVGLQFLQ
DIQATDNEPVEMLHSGTLVPSRELPIAF
CDWQLLNGPPTMKTSPTVYPVDNLPA
PDFSIAPGAQIQHKSGIVSGVQALPLVW
FDFEDHTNYELDIKETLTLKPEGFVVKAKSKKI
LKSVELSGPPRLVATNFVGGPKNVPIRF
PALSLGIDADDVVWRRSLLLRGIDHLPV
BETA 5 SHEET
RAT2BI
hAromatase
BovineSCC
humCYP11
humCYP27
humCYP24
humCYPRAI
CAM
BM3
TERP
ERYF

FIG. 1G

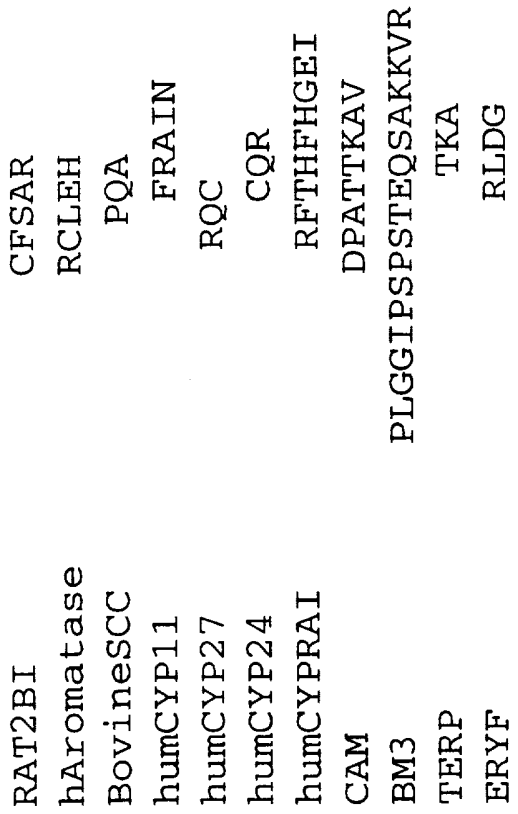

CFSAR
RCLEH
PQA
FRAIN
RQC
CQR
RFTHFHGEI
DPATTKAV
PLGGIPSPSTEQSAKKVR
TKA
RLDG
RAT2BI
hAromatase
BovineSCC
humCYP11
humCYP27
humCYP24
humCYPRAI
CAM
BM3
TERP
ERYF

FIG. 1H

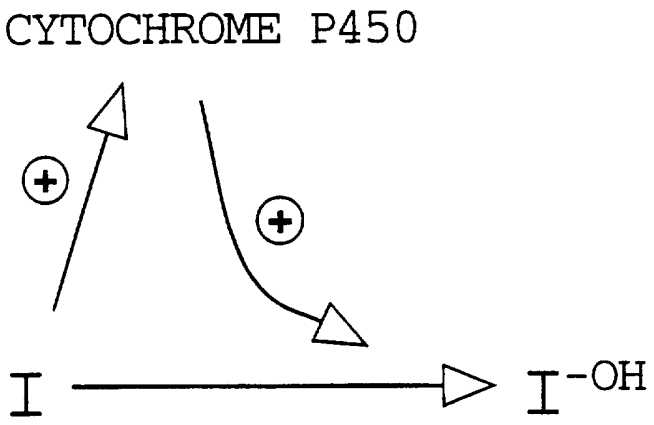
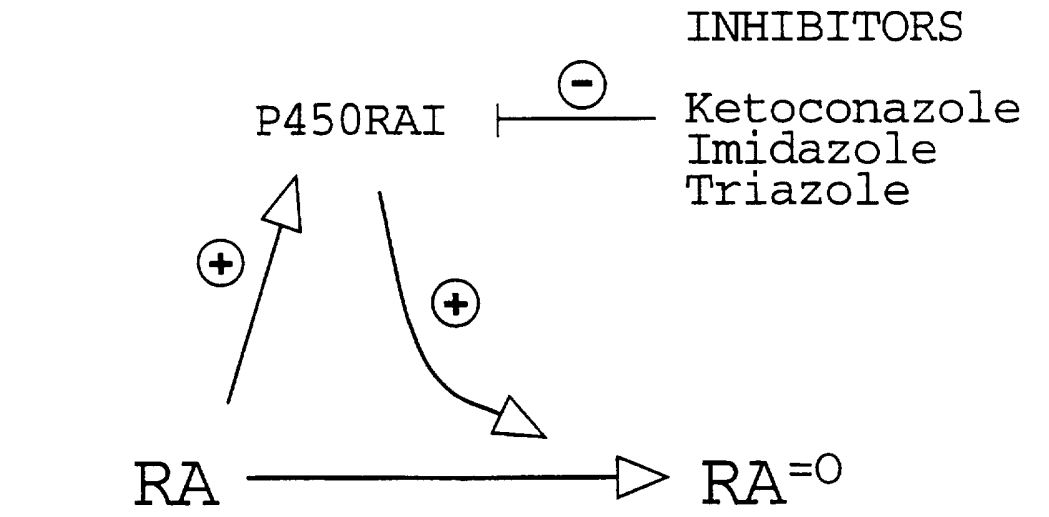


FIG. 2A



retinoic acid 4-oxoretinoic acid

FIG. 2B

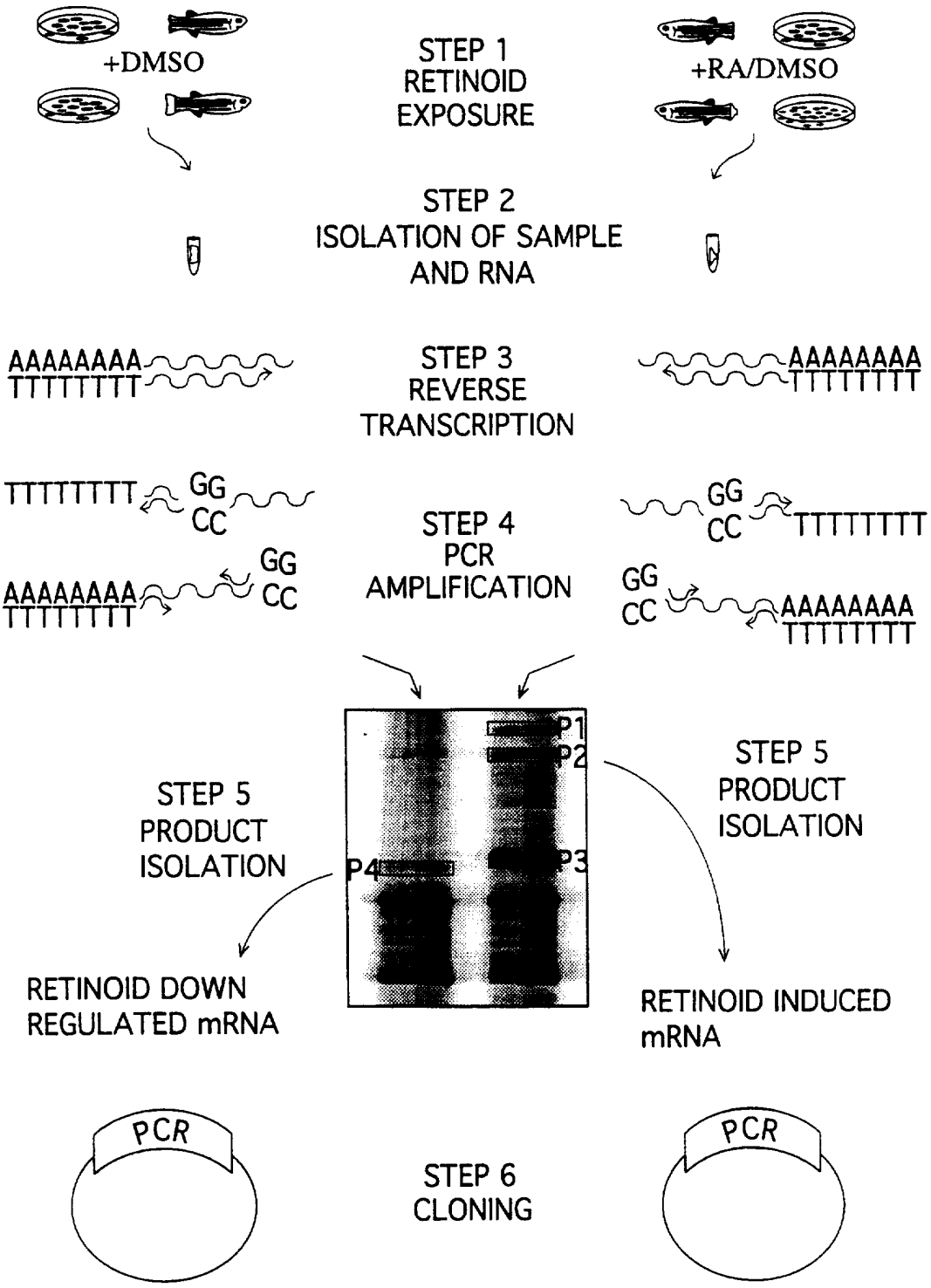

+DMSO
STEP 1
RETINOID
EXPOSURE
+RA/DMSO
STEP 2
ISOLATION OF SAMPLE
AND RNA
AAAAAAAA
TTTTTTTT
STEP 3
REVERSE
TRANSCRIPTION
AAAAAAAA
TTTTTTTT
TTTTTTTT
GG
CC
STEP 4
PCR
AMPLIFICATION
GG
CC
TTTTTTTT
AAAAAAAA
TTTTTTTT
GG
CC
AAAAAAAA
TTTTTTTT
P1
P2
P3
P4
STEP 5
PRODUCT
ISOLATION
STEP 5
PRODUCT
ISOLATION
RETINOID DOWN
REGULATED mRNA
RETINOID INDUCED
mRNA
PCR
PCR
STEP 6
CLONING

FIG. 3

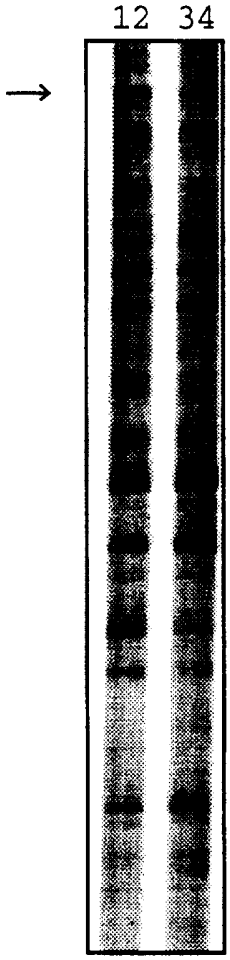


FIG. 4A

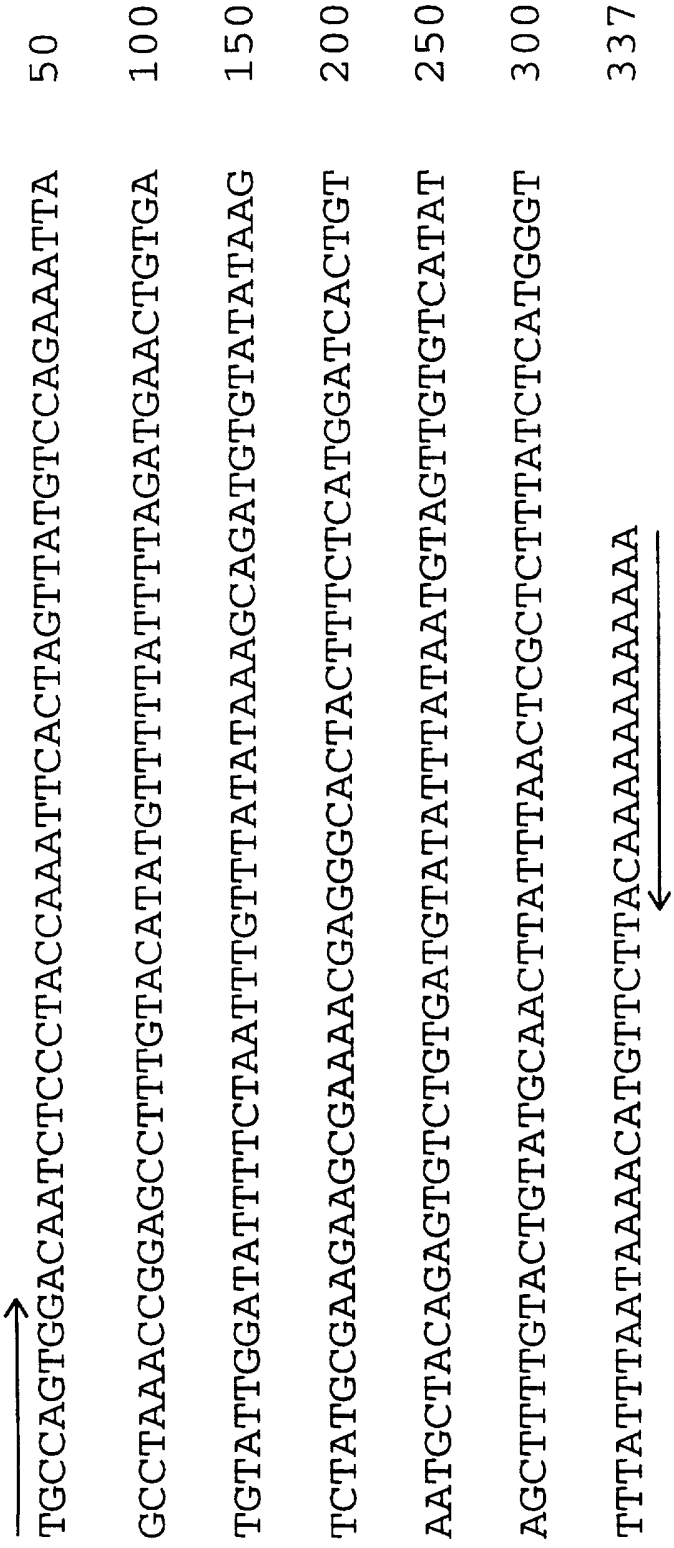

50
100
150
200
250
300
337
TGCCAGTGGACAATCTCCCTACCAAATTCACTAGTTATGTCCAGAAATTA
GCCTAAACCGGAGCCTTTGTACATATGTTTTTATTTTAGATGAACTGTGA
TGTATTGGATATTTTCTAATTTGTTTATATAAAGCAGATGTGTATATAAG
TCTATGCGAAGAAGCGAAAACGAGGGCACTACTTTCTCATGGATCACTGT
AATGCTACAGAGTGTCTGTGATGTATATTTATAATGTAGTTGTGTCATAT
AGCTTTTTGTACTGTATGCAACTTATTTAACTCGCTCTTTATCTCATGGGT
TTTATTTAATAAACATGTTCTTACAAAAAAAAAAAA

FIG. 4B

MetGlyLeuTyrThrLeuMetValThrPhe 10
LeuCysThrIleValLeuProValLeuLeu 20
PheLeuAlaAlaValLysLeuTrpGluMet 30
LeuMetIleArgArgValAspProAsnCys 40
ArgSerProLeuProProGlyThrMetGly 50
LeuProPheIleGlyGluThrLeuGlnLeu 60
IleLeuGlnArgArgLysPheLeuArgMet 70
LysArgGlnLysTyrGlyCysIleTyrLys 80
ThrHisLeuPheGlyAsnProThrValArg 90
ValMetGlyAlaAspAsnValArgGlnIle 100
LeuLeuGlyGluHisLysLeuValSerVal 110
GlnTrpProAlaSerValArgThrIleLeu 120
GlySerAspThrLeuSerAsnValHisGly 130
ValGlnHisLysAsnLysLysLysAlaIle 140
MetArgAlaPheSerArgAspAlaLeuGlu 150
HisTyrIleProValIleGlnGlnGluVal 160
LysSerAlaIleGlnGluTrpLeuGlnLys 170
AspSerCysValLeuValTyrProGluMet 180
LysLysLeuMetPheArgIleAlaMetArg 190
IleLeuLeuGlyPheGluProGluGlnIle 200
LysThrAspGluGlnGluLeuValGluAla 210
PheGluGluMetIleLysAsnLeuPheSer 220
LeuProIleAspValProPheSerGlyLeu 230
TyrArgGlyLeuArgAlaArgAsnPheIle 240
HisSerLysIleGluGluAsnIleArgLys 250
LysIleGlnAspAspAspAsnGluAsnGlu 260
GlnLysTyrLysAspAlaLeuGlnLeuLeu 270

FIG. 4C(i)

```
IleGluAsnSerArgArgSerAspGluPro   280
PheSerLeuGlnAlaMetLysGluAlaAla   290
ThrGluLeuLeuPheGlyGlyHisGluThr   300
ThrAlaSerThrAlaThrSerLeuValMet   310
PheLeuGlyLeuAsnThrGluValValGln   320
LysValArgGluGluValGlnGluLysVal   330
GluMetGlyMetTyrThrProGlyLysGly   340
LeuSerMetGluLeuLeuAspGlnLeuLys   350
TyrThrGlyCysValIleLysGluThrLeu   360
ArgIleAsnProProValProGlyGlyPhe   370
ArgValAlaLeuLysThrPheGluLeuAsn   380
GlyTyrGlnIleProLysGlyTrpAsnVal   390
IleTyrSerIleCysAspThrHisAspVal   400
AlaAspValPheProAsnLysGluGluPhe   410
GlnProGluArgPheMetSerLysGlyLeu   420
GluAspGlySerArgPheAsnTyrIlePro   430
PheGlyGlyGlySerArgMetCysValGly   440
LysGluPheAlaLysValLeuLeuLysIle   450
PheLeuValGluLeuThrGlnHisCysAsn   460
TrpIleLeuSerAsnGlyProProThrMet   470
LysThrGlyProThrIleTyrProValAsp   480
AsnLeuProThrLysPheThrSerTyrVal   490
ArgAsn                           492
```

FIG. 4C(ii)

```
                -8  -4   0   4   8
P450RAI          PFGGGSRMCVGKEFAKVLLK
ATCYTP450        *****P*L*P*Y*L*R*A*S
RATCYP4A1        **S**A*N*I**Q**MSEM*
RABCYP4A5        **S**A*N*I**Q**MNE**
CYP4503A12       ***T*P*N*I*MR**IMNM*
hCYTFAOH         **S****N*I**Q**MNE**
```

FIG. 4D

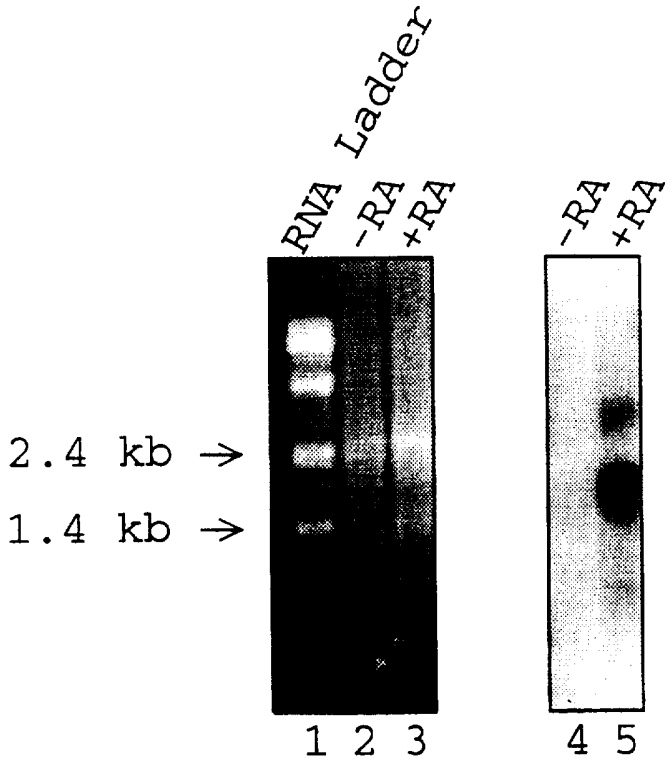


FIG. 5A

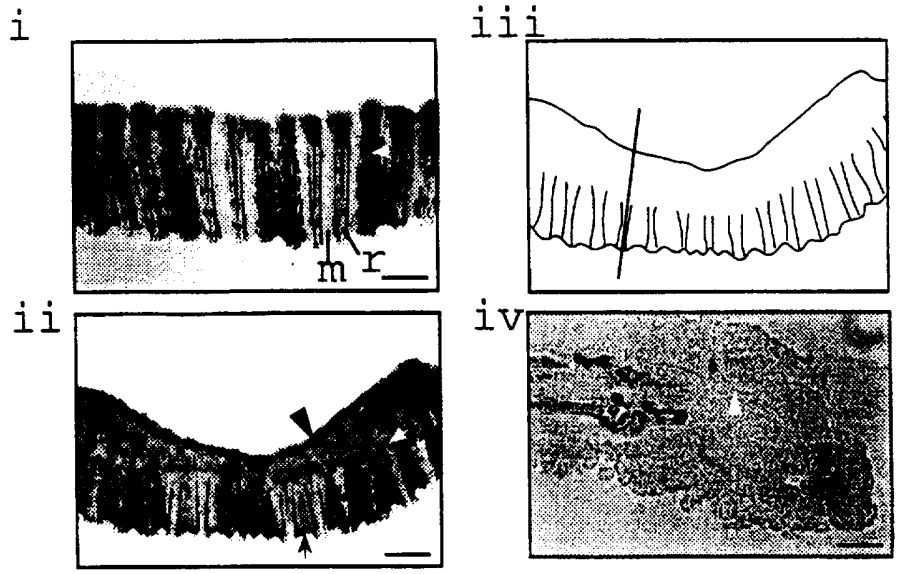
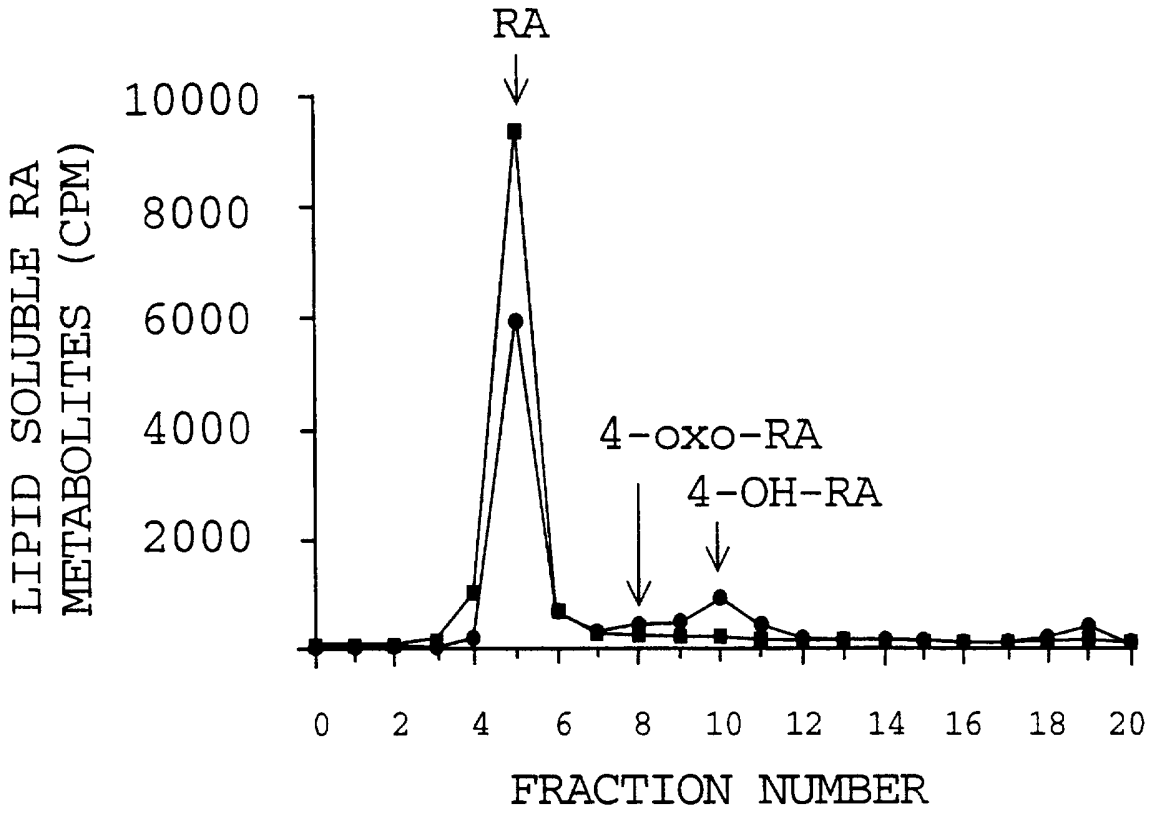


FIG. 5B


RA
10000
8000
6000
4000
2000
LIPID SOLUBLE RA METABOLITES (CPM)
4-oxo-RA
4-OH-RA
0 2 4 6 8 10 12 14 16 18 20
FRACTION NUMBER

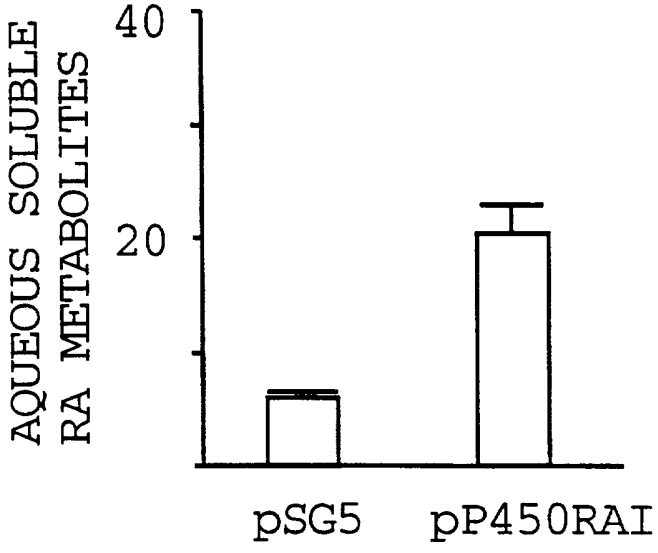
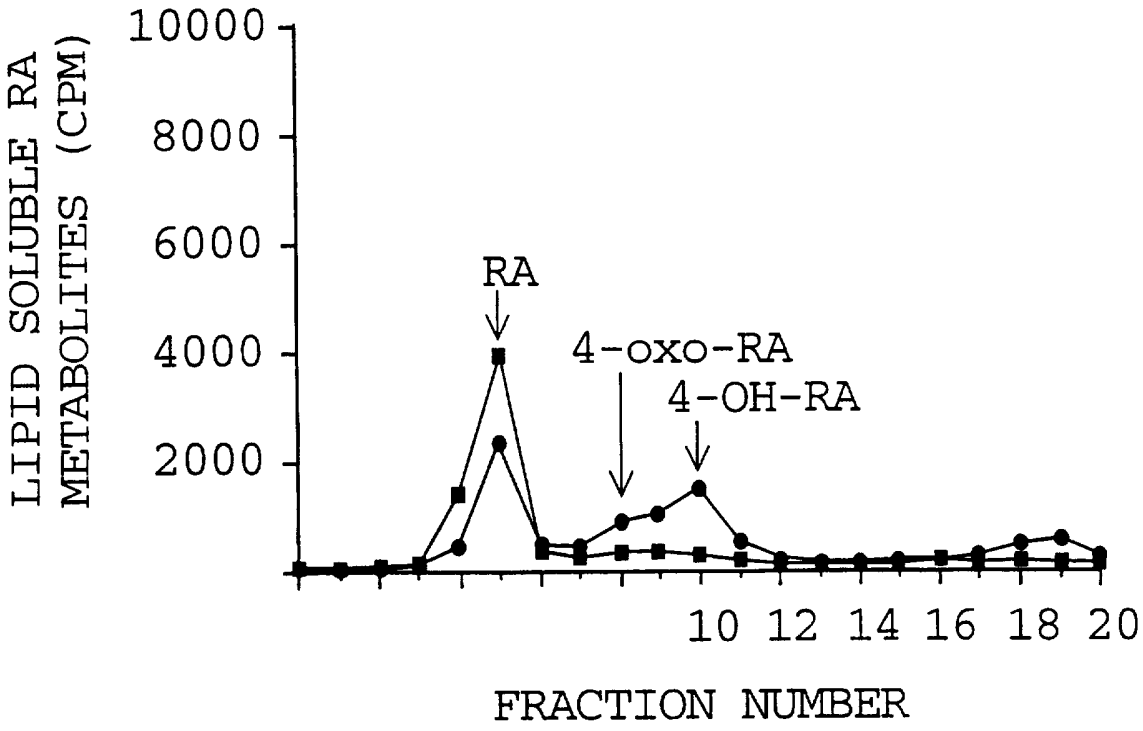

40
20
AQUEOUS SOLUBLE RA METABOLITES
pSG5
pP450RAI

FIG. 6A


LIPID SOLUBLE RA METABOLITES (CPM)
10000
8000
6000
4000
2000
RA
4-oxo-RA
4-OH-RA
10 12 14 16 18 20
FRACTION NUMBER

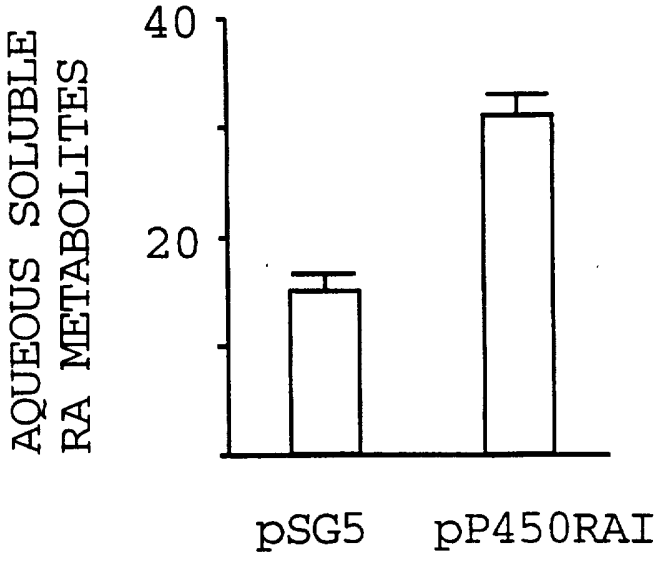
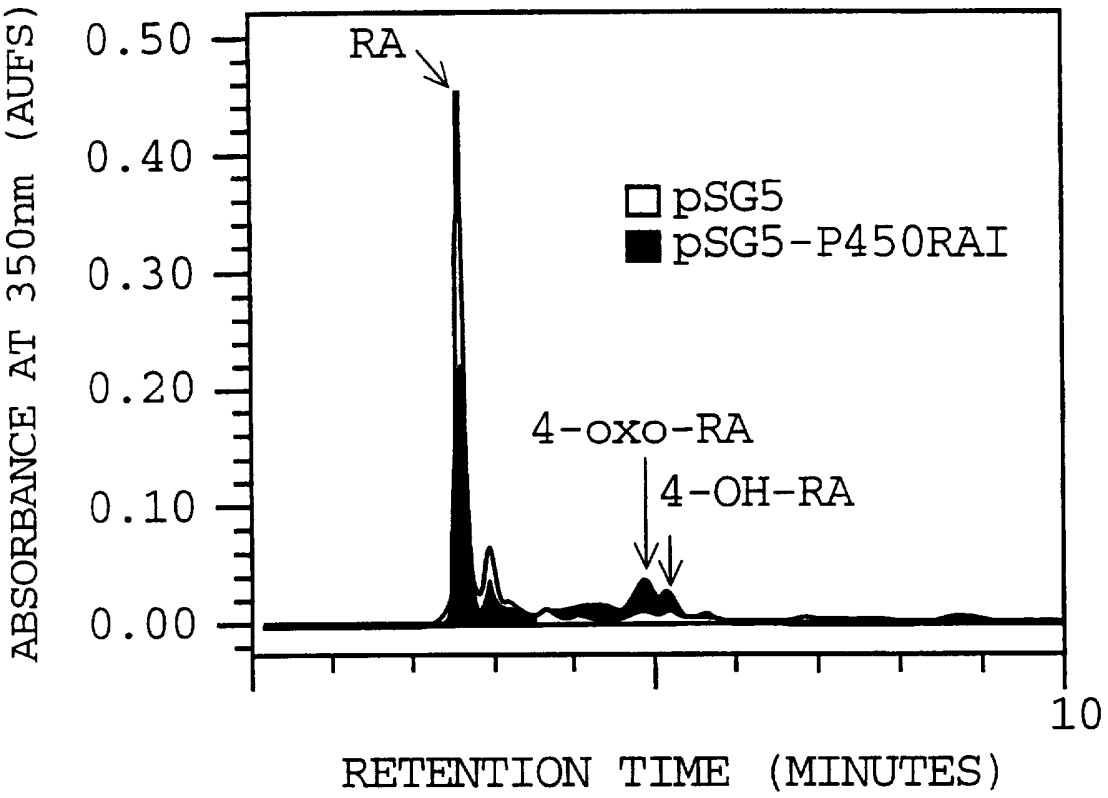

AQUEOUS SOLUBLE RA METABOLITES
40
20
pSG5
pP450RAI

FIG. 6B


ABSORBANCE AT 350nm (AUFS)
0.50
0.40
0.30
0.20
0.10
0.00
RA
pSG5
pSG5-P450RAI
4-oxo-RA
4-OH-RA
10
RETENTION TIME (MINUTES)

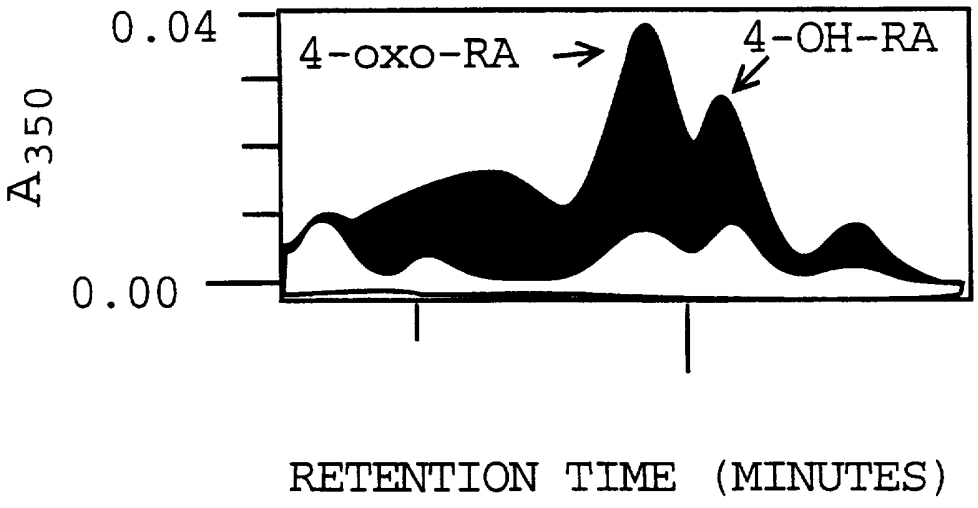
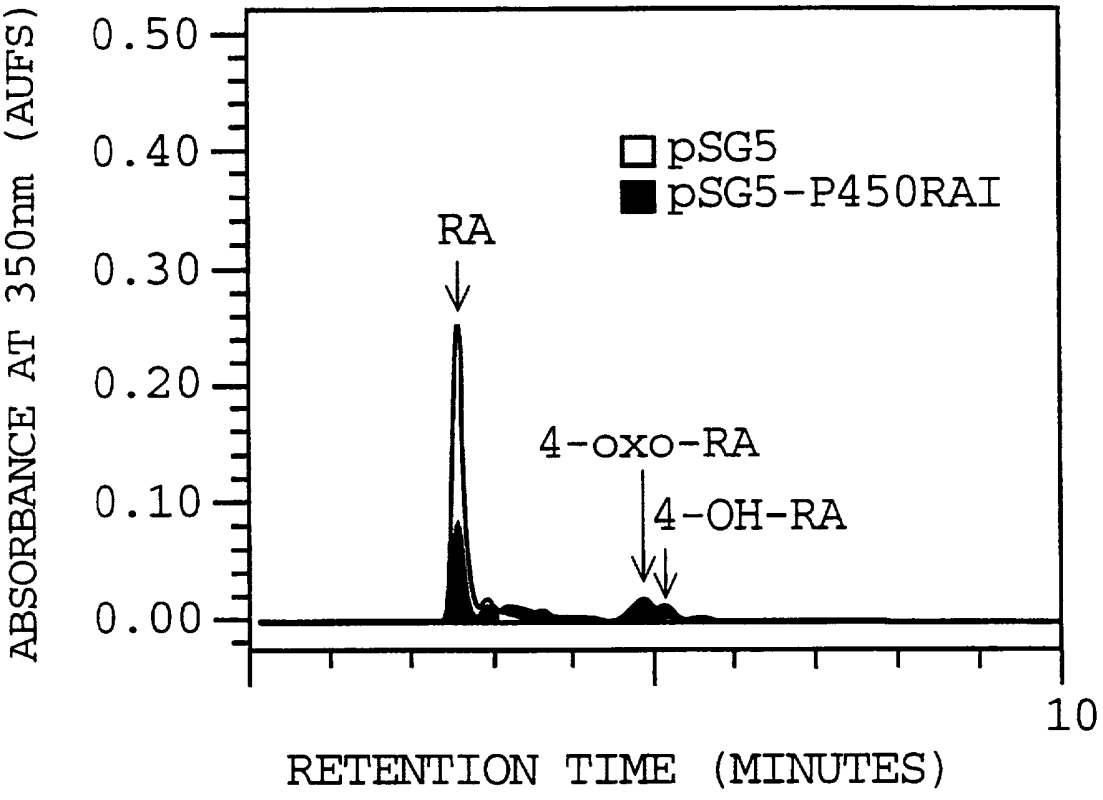

A_{350}
0.04
0.00
4-oxo-RA
4-OH-RA
RETENTION TIME (MINUTES)

FIG. 6C


ABSORBANCE AT 350nm (AUFS)
0.50
0.40
0.30
0.20
0.10
0.00
pSG5
pSG5-P450RAI
RA
4-oxo-RA
4-OH-RA
10
RETENTION TIME (MINUTES)

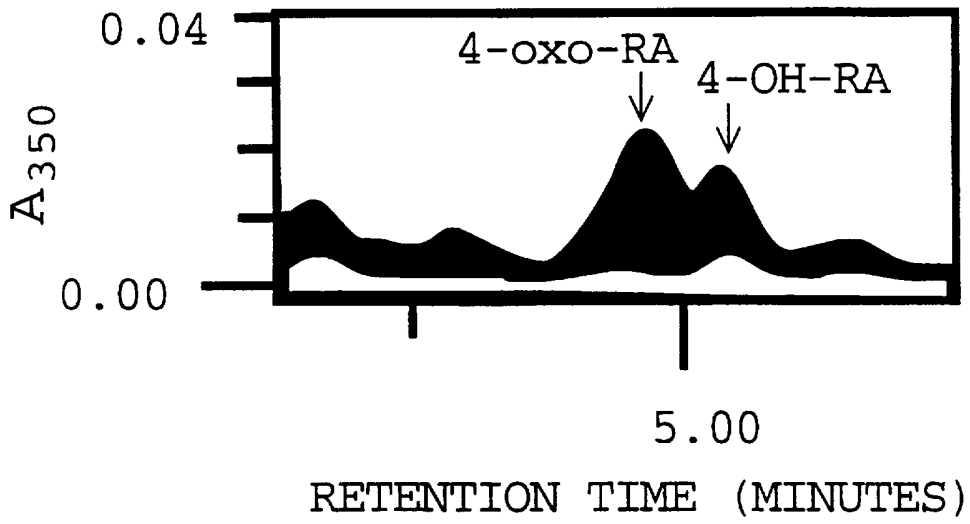

A350
0.04
0.00
4-oxo-RA
4-OH-RA
5.00
RETENTION TIME (MINUTES)

FIG. 6D

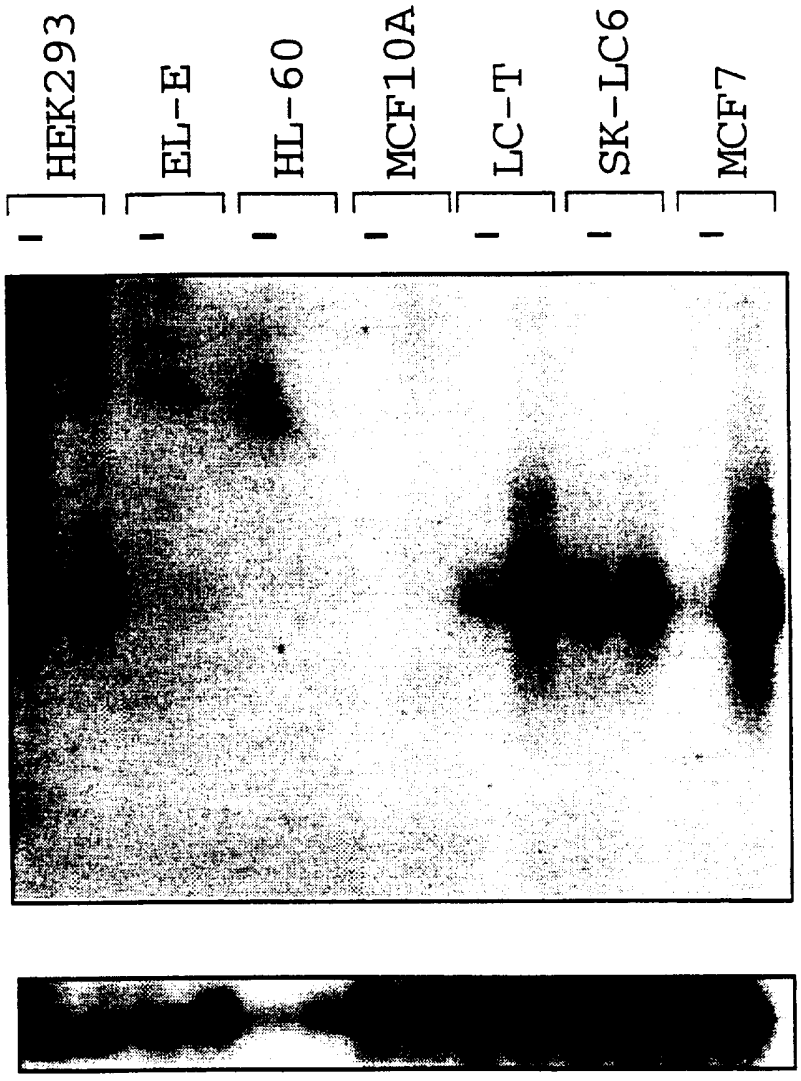


FIG. 7

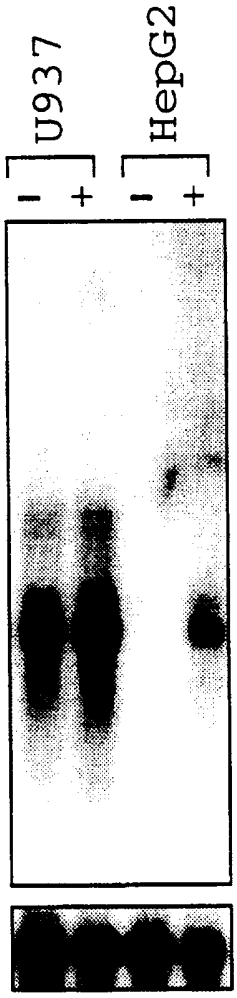


FIG. 8

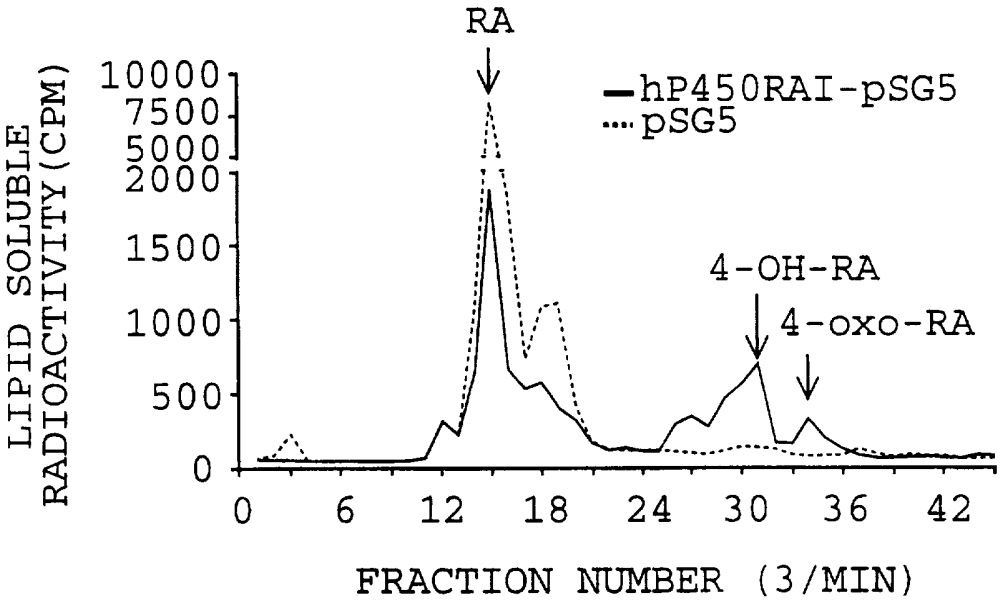

RA
hP450RAI-pSG5
pSG5
4-OH-RA
4-oxo-RA
LIPID SOLUBLE RADIOACTIVITY (CPM)
10000
7500
5000
2000
1500
1000
500
0
0
6
12
18
24
30
36
42
FRACTION NUMBER (3/MIN)

FIG. 9A

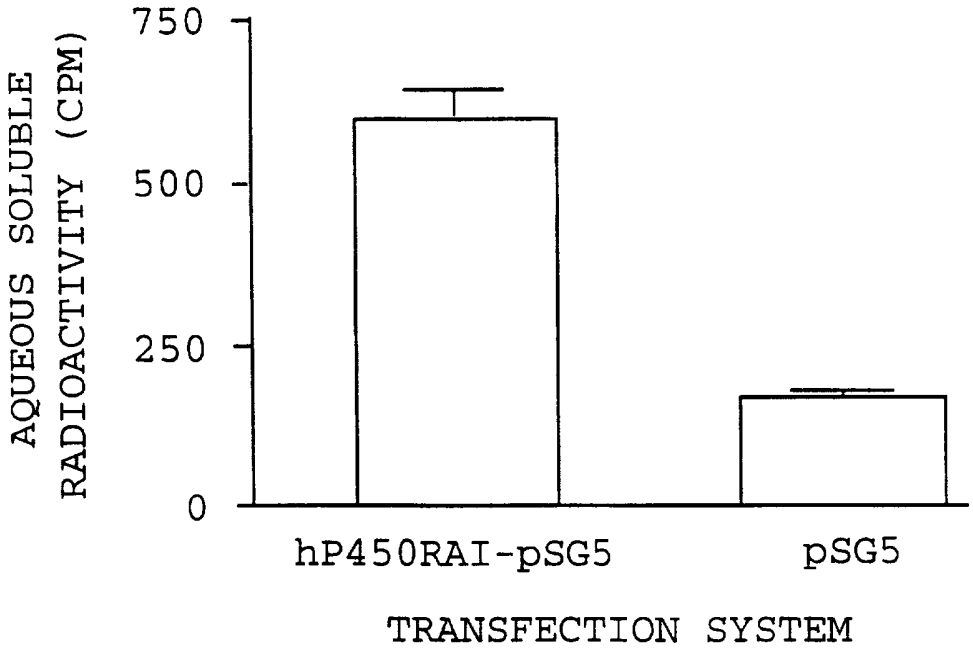

AQUEOUS SOLUBLE RADIOACTIVITY (CPM)
750
500
250
0
hP450RAI-pSG5
pSG5
TRANSFECTION SYSTEM

FIG. 9B

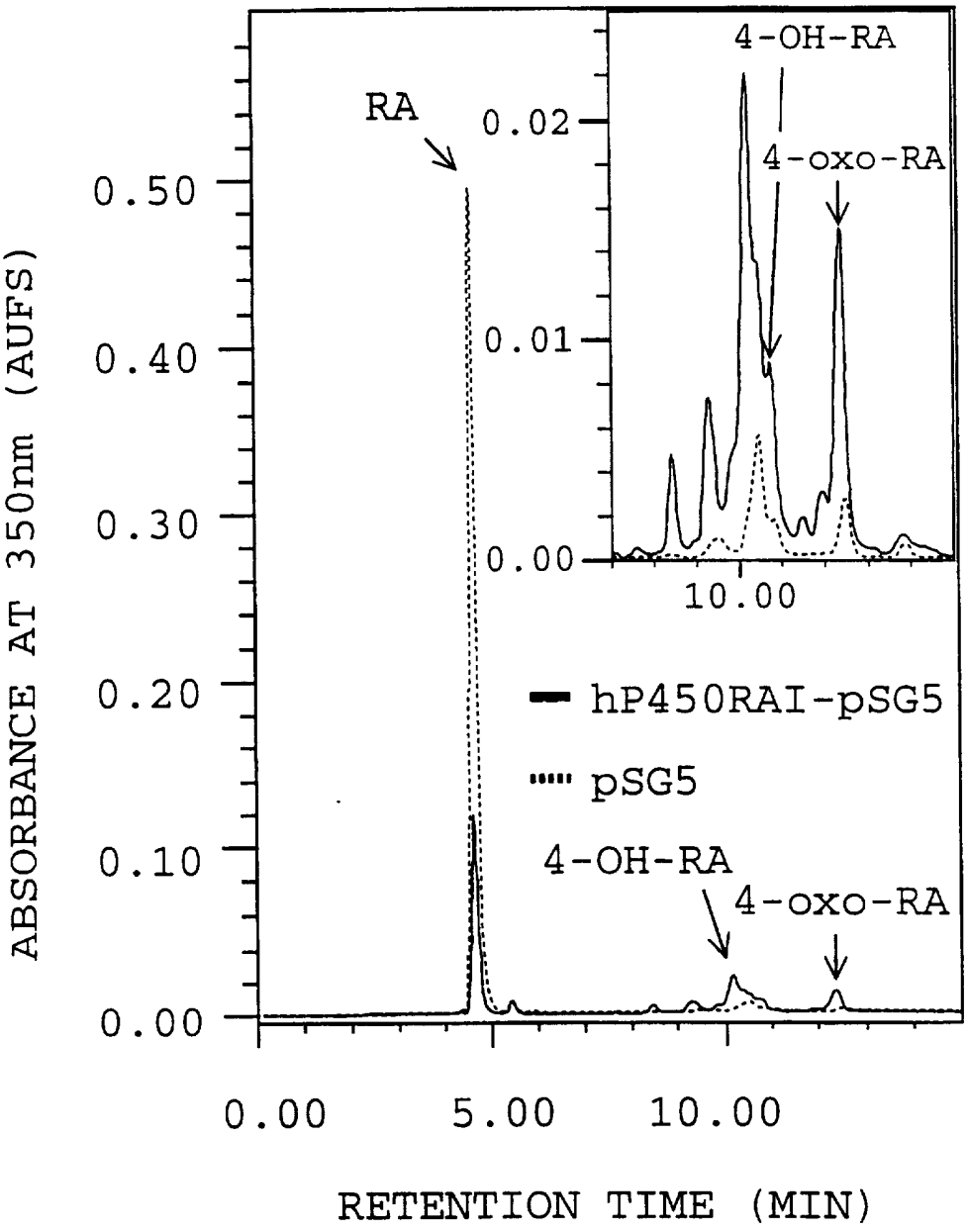

4-OH-RA
4-oxo-RA
RA
0.02
0.01
0.00
10.00
0.50
0.40
0.30
0.20
0.10
0.00
hP450RAI-pSG5
pSG5
4-OH-RA
4-oxo-RA
ABSORBANCE AT 350nm (AUFS)
0.00
5.00
10.00
RETENTION TIME (MIN)

FIG. 9C

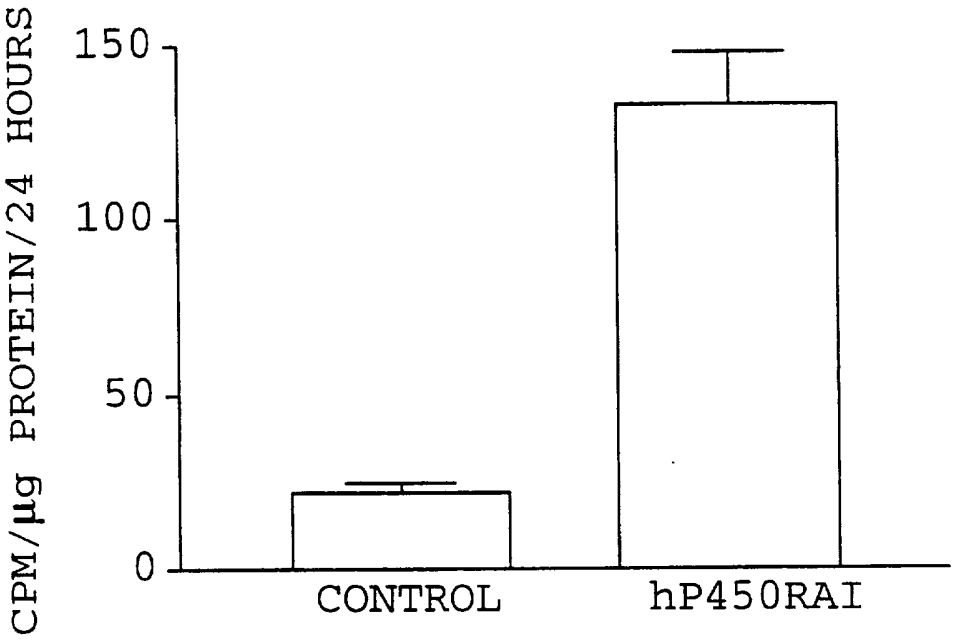


FIG. 10A

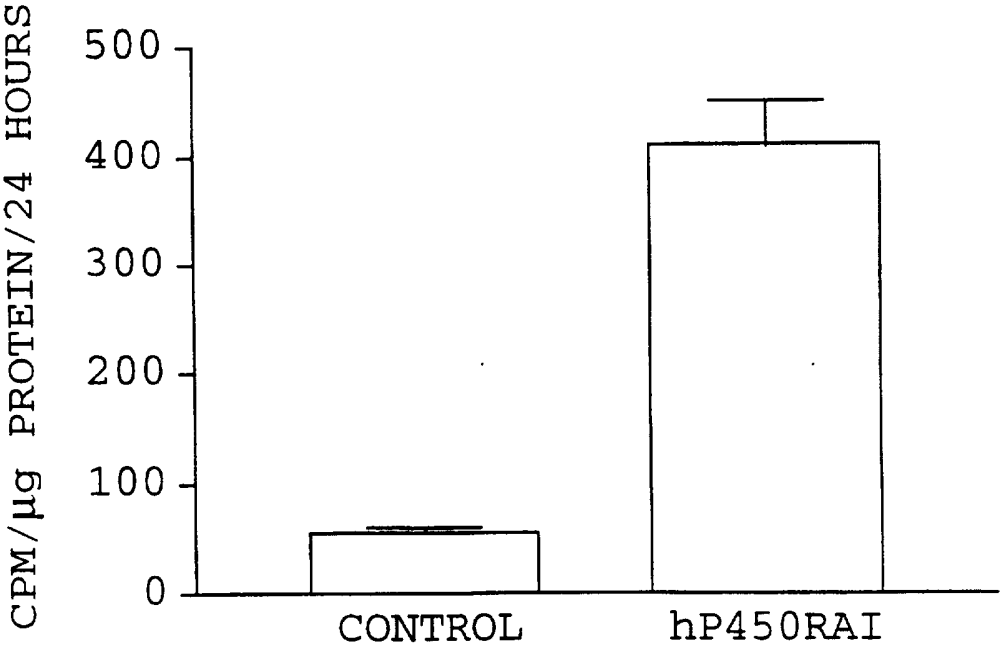


FIG. 10B

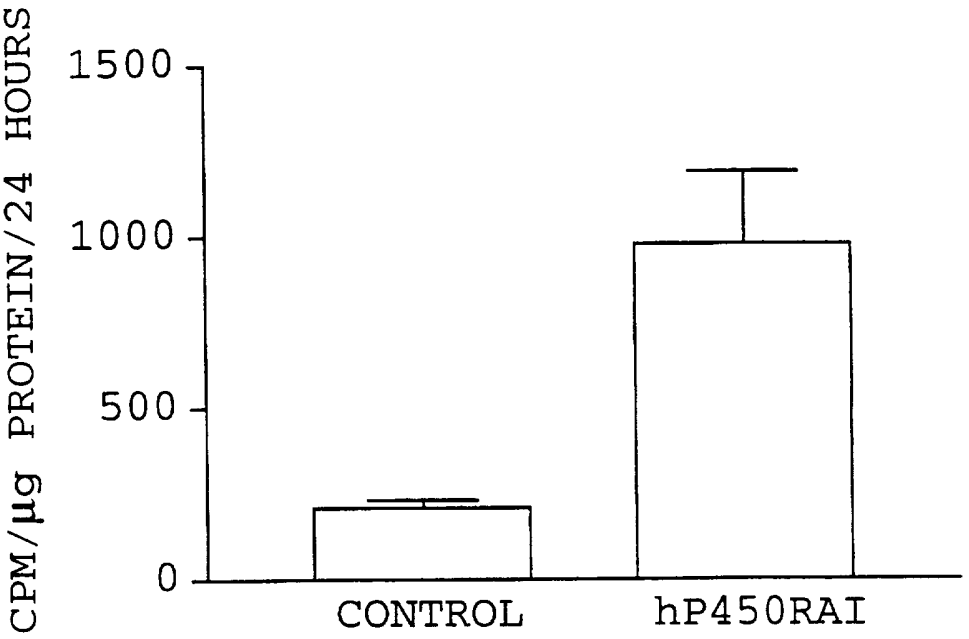


FIG. 10C

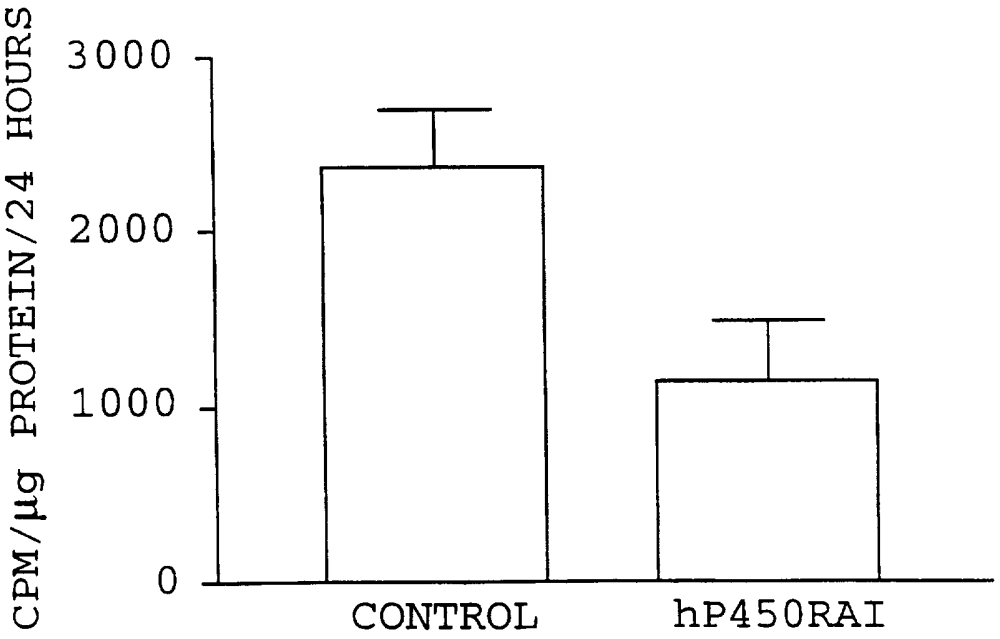


FIG. 10D

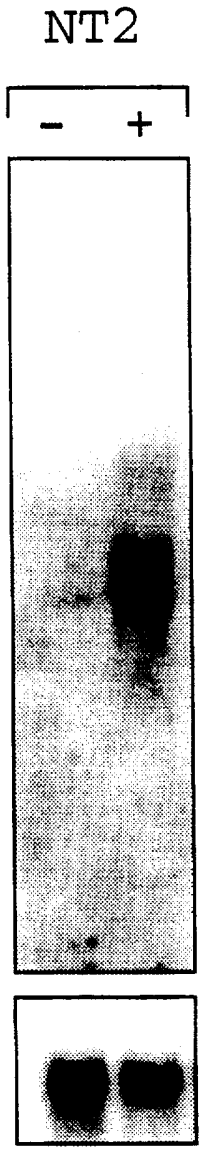


FIG. 11

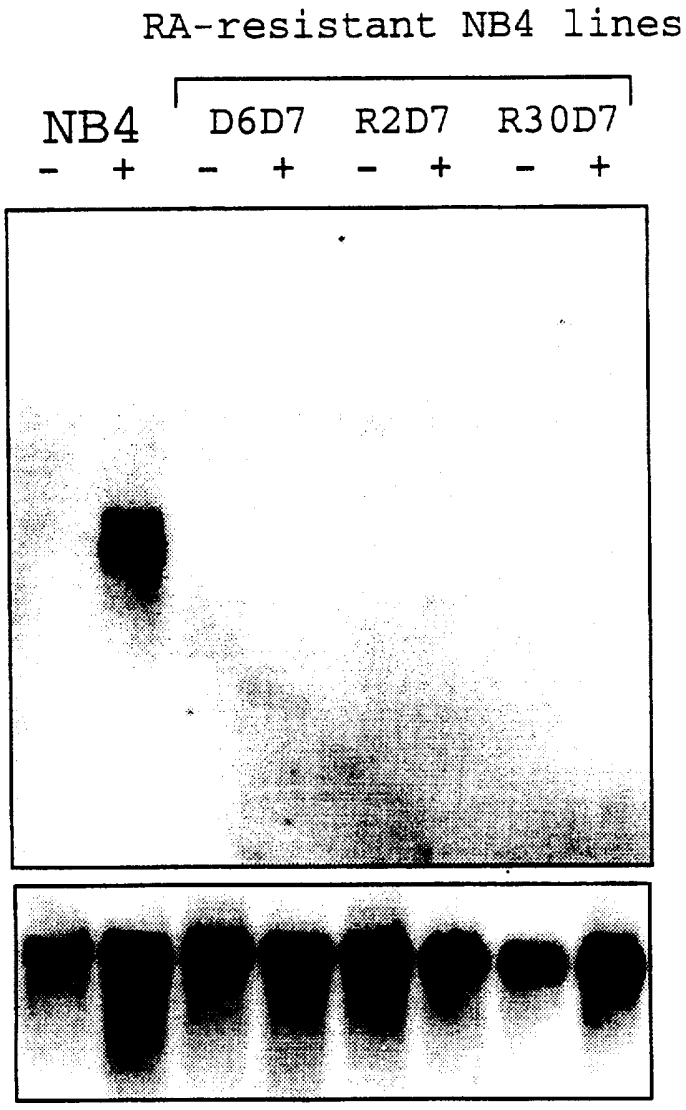


FIG. 12

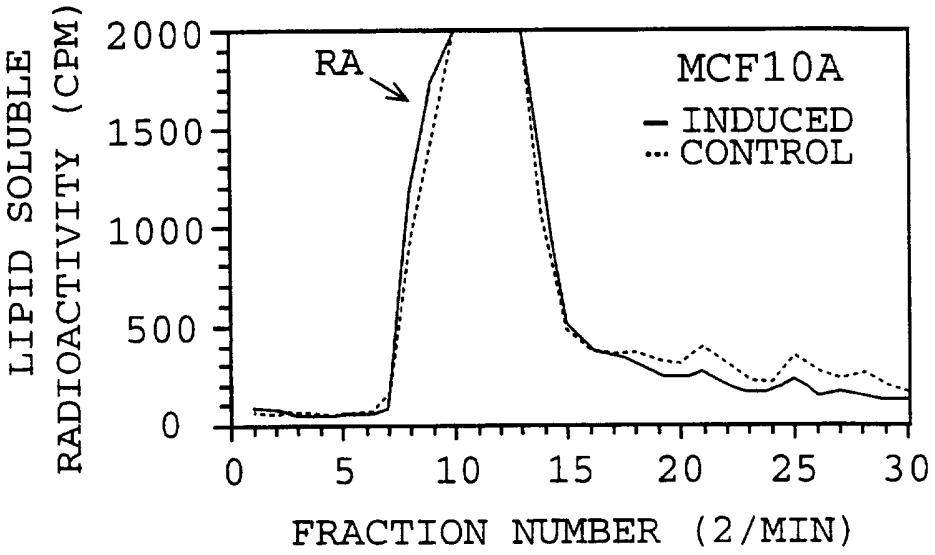


FIG. 13A

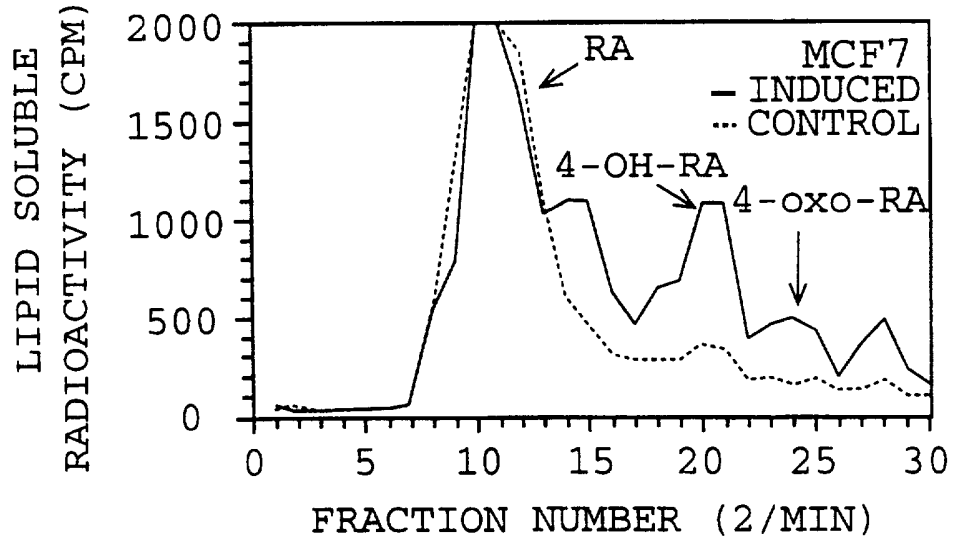


FIG. 13B

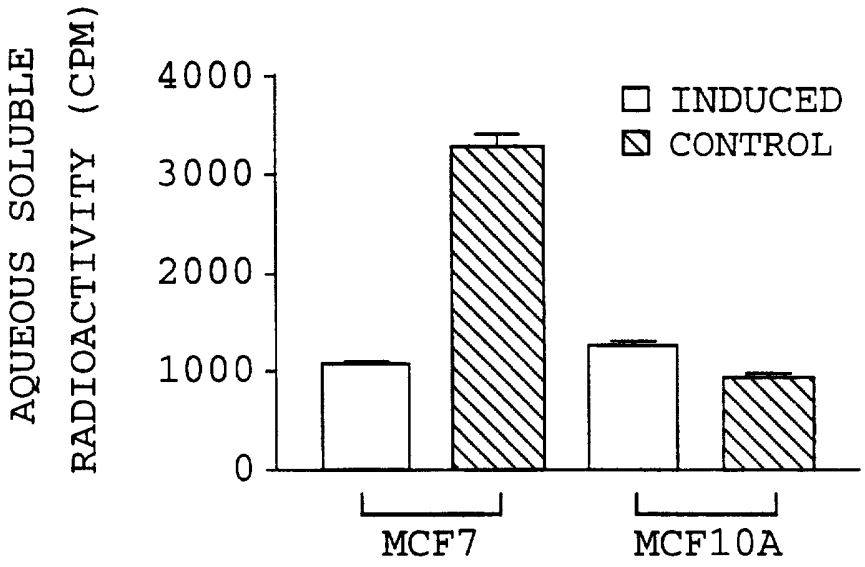

AQUEOUS SOLUBLE RADIOACTIVITY (CPM)
4000
3000
2000
1000
0
INDUCED
CONTROL
MCF7
MCF10A

FIG. 13C

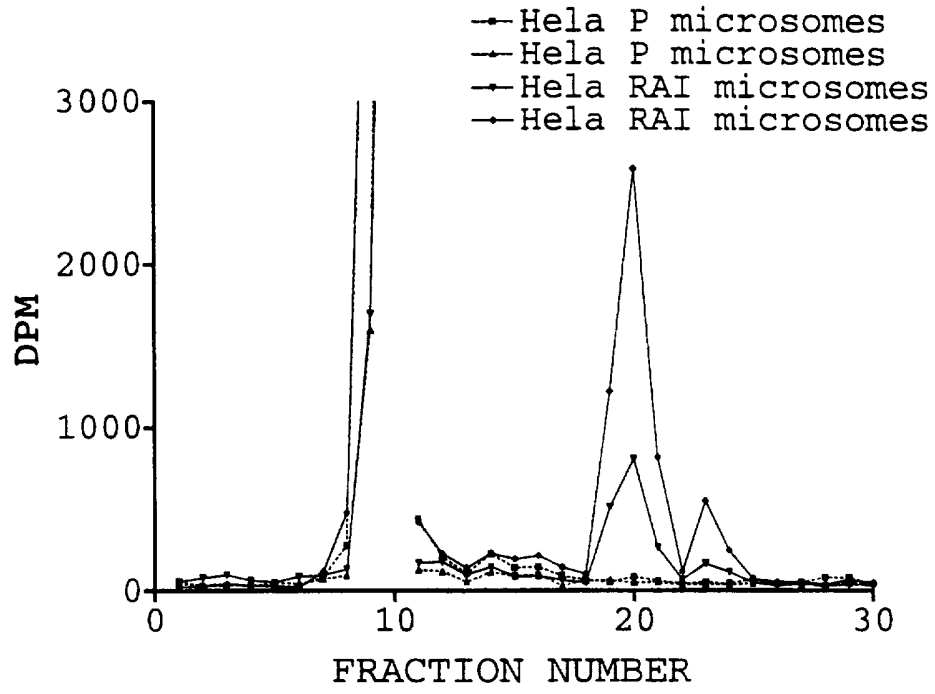


FIG. 14A

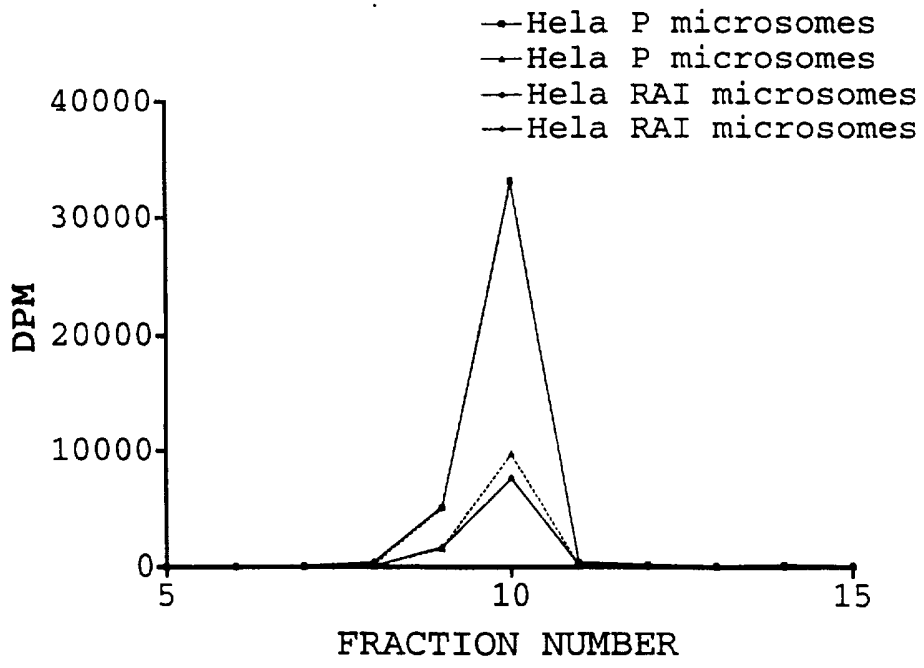


FIG. 14B

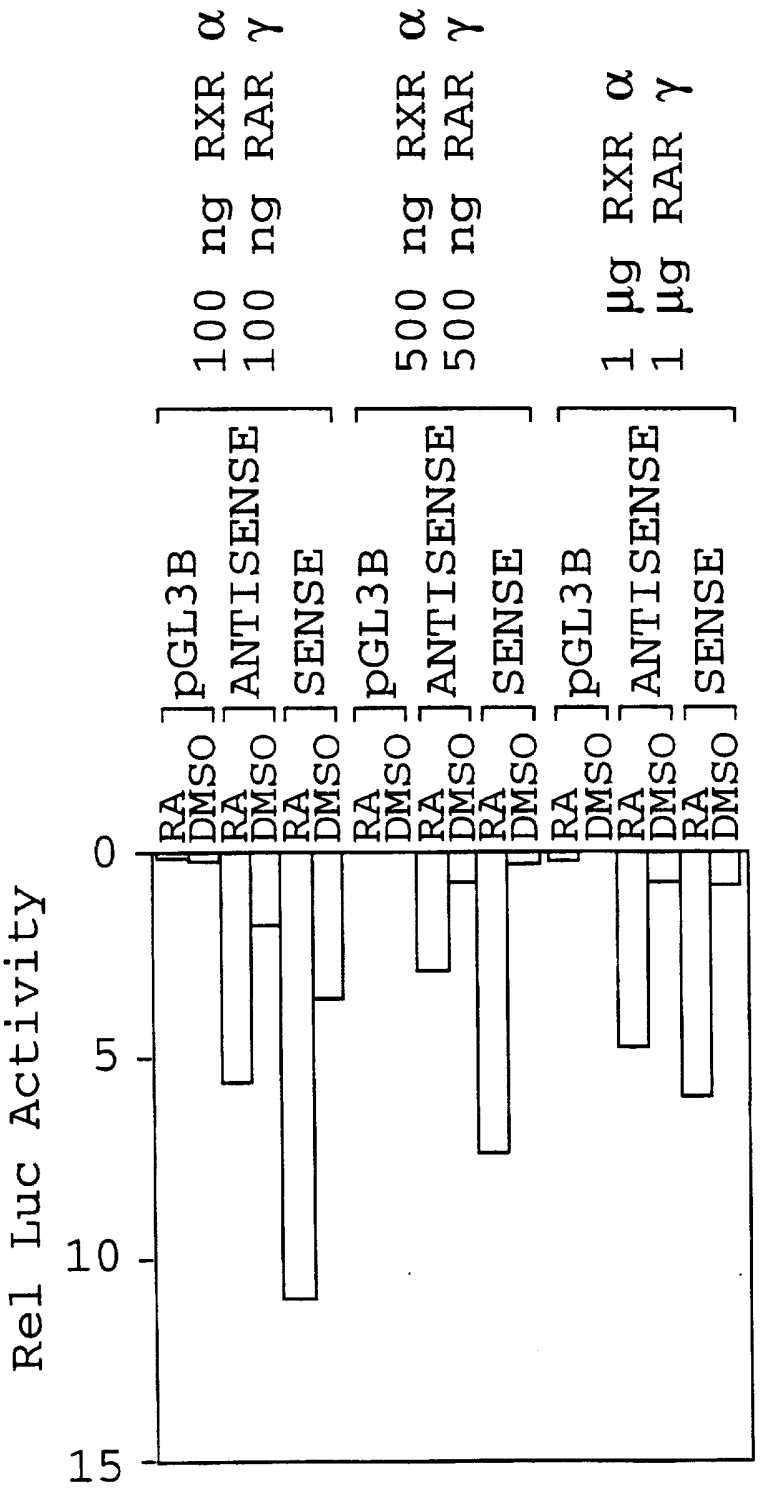

Rel Luc Activity
0
5
10
15
RA
DMSO
pGL3B
ANTISENSE
SENSE
100 ng RXR α
100 ng RAR γ
500 ng RXR α
500 ng RAR γ
1 µg RXR α
1 µg RAR γ

FIG. 15

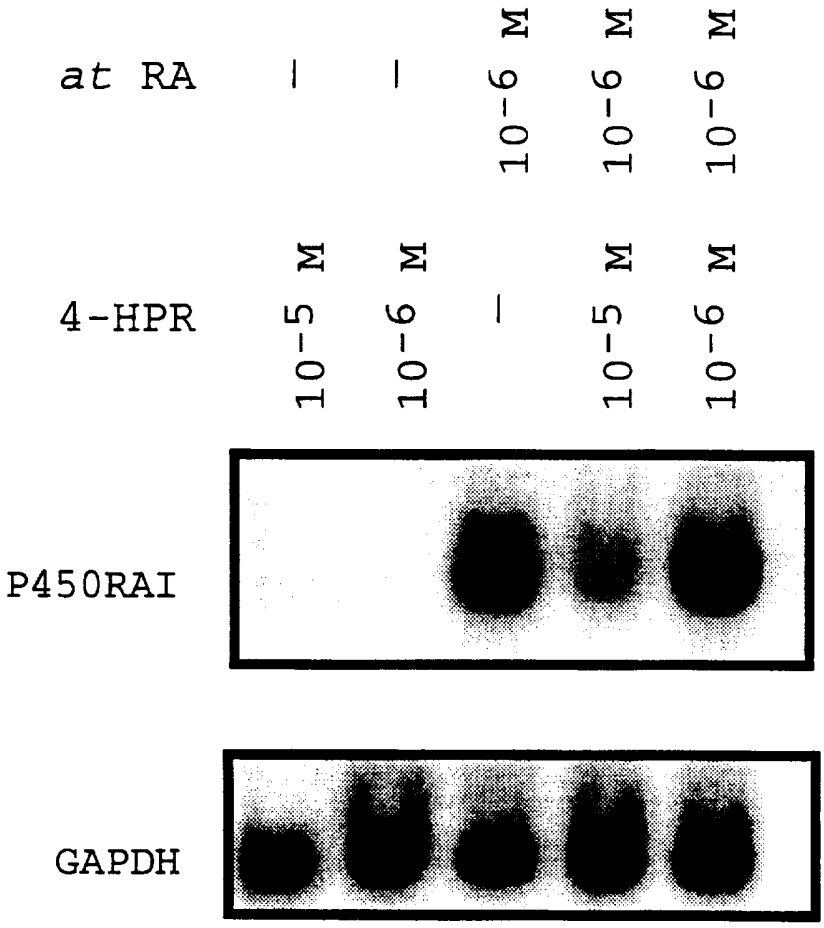


FIG. 16

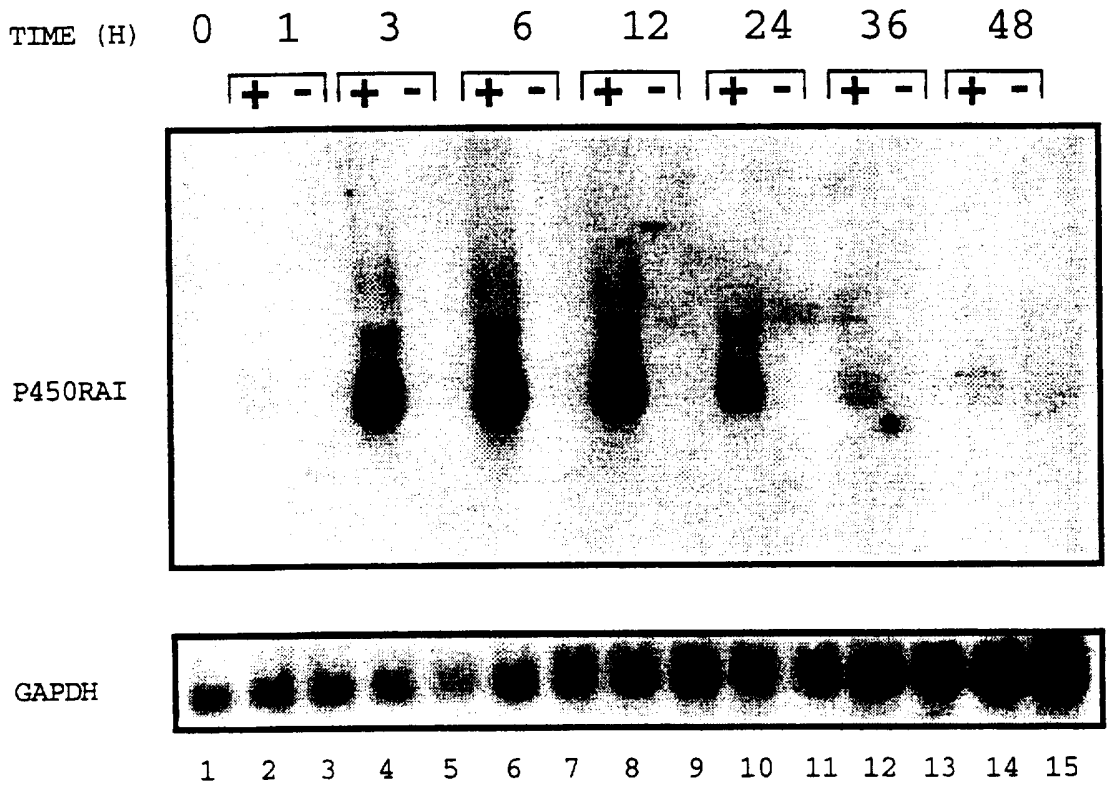

TIME (H)
0 1 3 6 12 24 36 48
+ - + - + - + - + - + - + -
P450RAI
GAPDH
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

FIG. 17

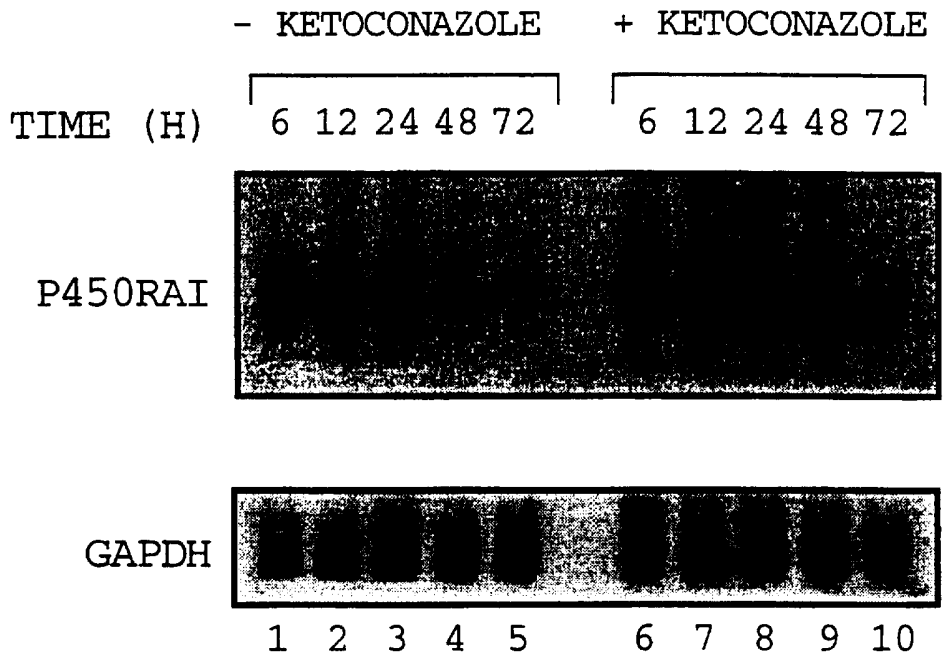

- KETOCONAZOLE
+ KETOCONAZOLE
TIME (H)
6 12 24 48 72
6 12 24 48 72
P450RAI
GAPDH
1 2 3 4 5 6 7 8 9 10

FIG. 18

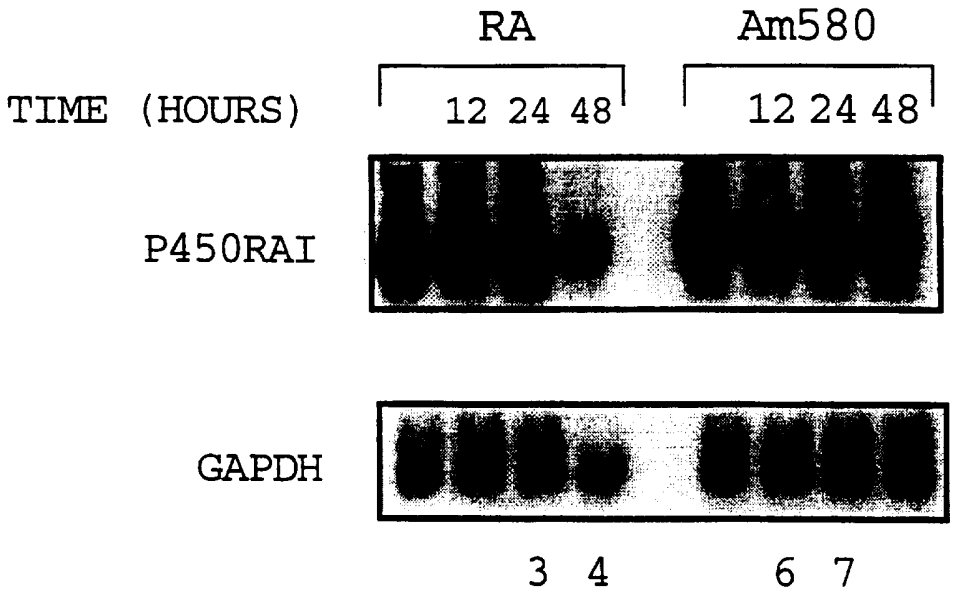

RA
Am580
TIME (HOURS)
12 24 48
12 24 48
P450RAI
GAPDH
3 4 6 7

FIG. 19