GENETIC POLYMORPHISMS IN THE CYP2A6: IMPLICATIONS FOR INTER-INDIVIDUAL DIFFERENCES IN NICOTINE METABOLISM

by

NURFADHLINA BINTI MUSA

2006

Thesis submitted in fulfillment of the requirements for the "Degree" of: Master of Science

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisors Assoc Prof Foong Kin, Assoc Prof Tan Soo Choon and Dr Teh Lay Kek who have given me this opportunity and guidance to undertake and complete my research that I now submit as a desertation. I express a special thank you to AP Foong Kin for her patience in seeing me through this project.

I would like to express my deepest gratitude to my Head Programme of Pharmacogenetic Research Group, Prof Rusli Ismail. During my project, I have been amazed millions of times by his knowledge, never-ending optimism, and encouragement. His unfailingly enthusiastic attitude and abundance of time for me and my problems have helped me grow as a young scientist. I thank Prof Rusli for our many late evening discussions, which always gave me the stimulation to continue my work with new energy and ideas!! I would also to thank Prof Rusli for sharing important moments with me; we have had many interesting and illuminating discussions touching all aspects of life.

I am also grateful to the Department of Pharmacology, School of Medical Sciences and later the Pharmacogenetic Research Group at the Institute for Research in Molecular Medicine (INFORMM) for the facilities for me to complete this research.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Plates

List of Abbreviations

°C	Celsius centigrade
µl	Microlitre
µmol	Micromole
bp	Base pairs
CV	Coefficient variation
CYP	Cythochrome P450
DME	Drug metabolizing enzyme
DNA	Deoxynucleic acid
dNTP	Deoxynucleoside triphosphate
EM	Extensive metabolizer
ex	exon
F	Forward
GC	Gas chromatography
h	hour
H_2O	Water
HCl	Hydrochloric acid
HPLC	High performance liquid chromatography
K'	Retention factor (capacity factor)
kb	kilobasepair

KCl	Potassium chloride
LOQ	Limit of quantification
$MgCl_2$	Magnesium chloride
min	minute
mmol	millimole
MR	Metabolic ratio
mt	Mutant type
NaOH	Sodium hydroxide
O	Oxygen
OD	Optical density
OH	Hydroxyl
PCR	Polymerase Chain Reaction
PM	Poor metabolizer
pmol	piccomole
R	Reverse
rpm	rounds per minute
SD	Standard deviation
SDS	Sodium dodecyl sulphate
secs	seconds
T_m	Melting temperature
UV	Ultraviolet
wt	Wild type

List of Appendices

ABSTRAK

POLIMORFISME GENETIK DALAM CYP2A6: IMPLIKASI UNTUK PERBEZAAN METABOLISME NIKOTINA ANTARA INDIVIDU

Terdapat pelbagai bentuk *CYP2A6*. Ia memetabolismekan nikotina, kotinina dan sesetengah pro-karsinogen serta drug. Variasi dalam alel *CYP2A6* boleh mengurangkan risiko untuk merokok dan frekuensi bagi 'PM' adalah lebih rendah dikalangan penduduk Barat (Caucasion) berbanding penduduk Asia. Objektif kajian kami adalah untuk membangunkan kaedah genetik molekul dan analitikal bagi mengkaji kepelbagaian *CYP2A6* di Malaysia. Subjek terdiri daripada dewasa yang sihat berketurunan Melayu, Cina dan India yang telah bersetuju menyertai kajian ini. DNA yang telah diesktrak daripada darah menggunakan kaedah 'salting-out' seterusnya digunakan untuk mengesan genotip *CYP2A6* melalui kaedah PCR. Seramai 616 subjek telah menyertai kajian ini. Frekuensi bagi *CYP2A6*1A, CYP2A6*1B, CYP2A6*4, CYP2A6*5, CYP2A6*7, CYP2A6*8, CYP2A6*10* dan *CYP2A6*1x2* dikalangan etnik Melayu masing-masingnya adalah 27.04%, 46.67%, 7.04%, 0.93%, 4.26%, 5.00%, 4.26% dan 0.37%, etnik Cina, 34.30%, 44.48%, 4.94%, 1.16%, 6.98%, 1.45%, 1.74% dan 1.45% dan etnik India 52.01%, 39.37%, 0.29%, 1.15%, 1.44%, 0.86% and 0.86%. *CYP2A6*3* ditemui dikalangan etnik India sahaja. Kami juga menggunakan kaedah HPLC untuk mengukur kepekatan nikotina dan kotinina di dalam plasma dikalangan 40 sukarelawan yang telah mengunyah gam nikotina. Plot log probit nisbah metabolik telah diperolehi.

Kewujudan beberapa sub populasi mencadangkan wujudnya kepelbagaian pada lokus *CYP2A6*. Sebagai kesimpulan, kami telah berjaya membangunkan kaedah untuk mengkaji kepelbagaian *CYP2A6* dikalangan populasi di Malaysia. Kaedah PCR alel-spesifik yang kami bangunkan bukan sahaja spesifik dan sensitif, malah lebih mudah berbanding kaedah di dalam literatur. Kaedah ini telah digunapakai ke atas tiga etnik terbesar di Malaysia dan kami dapati kepelbagaian dalam polimorfisme *CYP2A6* dengan *CYP2A6*3* hanya ditemui dikalangan etnik India. Kaedah HPLC juga spesifik dan sensitif. Ianya dibuktikan dengan membuat ujian pengesahan melalui kontrol yang telah ditetapkan dan juga projek rintis dengan 40 subjek yang telah diketahui polimorfisme *CYP2A6*nya. Perbezaan genotip di antara etnik meramalkan penduduk Malaysia mungkin berbeza dari segi kecenderungan mengalami sesetengah penyakit dan juga kesan sampingan sesetengah drug. Walau bagaimanapun, kajian lebih lanjut perlu dilakukan dengan menggunakan saiz sampel yang lebih besar.

Katakunci: CYP2A6, polimorfisme, PCR, genotip, HPLC, nikotina, kotinina, Melayu, Cina, India, penduduk Malaysia.

ABSTRACT

GENETIC POLYMORPHISMS IN THE CYP2A6: IMPLICATIONS FOR INTER-INDIVIDUAL DIFFERENCES IN NICOTINE METABOLISM

CYP2A6 is polymorphic. It metabolizes nicotine, cotinine, several pro-carcinogens and drugs. Variant *CYP2A6* alleles reduced risks for smoking and Caucasian have lower frequencies for the PM phenotype than Asians. The objective of our study was to develop analytical and molecular genetic methods for studying *CYP2A6* polymorphism in Malaysia. Subjects were healthy adult Malays, Chinese and Indians who gave informed consents. DNA was extracted from blood using salting out procedures and subjected to PCR-genotyping for CYP2A6. Six-hundred and sixteen subjects were enrolled. The frequencies for *CYP2A6*1A, CYP2A6*1B, CYP2A6*4, CYP2A6*5, CYP2A6*7, CYP2A6*8, CYP2A6*10* and *CYP2A6*1x2* among Malays were 27.04%, 46.67%, 7.04%, 0.93%, 4.26%, 5.00%, 4.26% and 0.37% respectively; among Chinese, 34.30%, 44.48%, 4.94%, 1.16%, 6.98%, 1.45%, 1.74% and 1.45% respectively and among Indians were 52.01%, 39.37%, 0.29%, 1.15%, 1.44%, 0.86% and 0.86% respectively. *CYP2A6*3* was also detected among Indians. We also used an HPLC to determine nicotine and cotinine from plasma of 40 volunteers administered nicotine. A probit plot of log MR for nicotine was obtained. Sub populations exist suggesting existence of a polymorphism at the *CYP2A6* locus. We conclude that we successfully developed

methods for the population study of *CYP2A6* polymorphism in Malaysia. Our allele-specific PCR was not only specific and sensitive but was also less tedious compared to published methods. Applied to the three major ethnic groups in Malaysia we noted heterogeneity in the polymorphism and the presence of *CYP2A6*3* among Indians. Our HPLC methods were sufficiently specific and sensitive as proven by validating against seeded controls and using in a pilot project with 40 subjects in whom *CYP2A6* polymorphism was evident. The ethnic differences we showed in our Malaysian population may imply different ethnic susceptibility to certain diseases and adverse effects of drugs. Further studies are however required using larger sample sizes.

Keywords: CYP2A6, genetic polymorphism, PCR, genotypes, HPLC, nicotine, cotinine, Malays, Chinese, Indians, Malaysian

Chapter One

Introduction and Review of the Literature

1.1 Introduction

No two individuals are the same. Great variability occurs. In as much as individuals differ in hair colour, their skin colour or their heights, they also differ in the way they react to medications and environmental toxins. Thus, when two individuals are given the same dose of the same drug, the first may obtain the best of effects while the other may develop toxicity and indeed may suffer death. Similarly when different individuals are exposed to the same environmental toxins some may suffer no harm while others may develop environmentally-induced diseases.

Factors that contribute to the variability of pharmacologic response are many and varied. Well known are underlying disease of the individuals; drugs and food that are concurrently taken may cause interactions; his age, nutritional status, renal and liver function, and his concomitant other illnesses and exposures.

Increasingly recognized as a factor of paramount importance in pharmacologic variability is the effect of inherited differences. It is well known that disease development is influenced by familial factors. For instance there is an aggregation of diseases like hypertension, diabetes mellitus, bronchial

asthma and even alcohol addiction in certain families. More recently, interest on similar inherited difference contributing to pharmacologic variability is gaining recognition.

Among the first clinical observations of such inherited differences were documented in the 1950s. It was shown that prolonged muscle relaxation after suxamethonium was associated with an inherited deficiency of plasma cholinesterase (1999). It was similarly shown that the haemolysis occurring after the administration of the anti-malarial drug primaquine and the ingestion of fava beans was due to an inherited deficiency of glucose-6-phosphate dehydrogenase (G6PD)(Carson *et al.*, 1956), and peripheral neuropathy with isoniazid occurred with inherited differences in acetylation of this medication (Evans and Relling, 1999). It thus became increasingly evident that genetic variability of certain enzymes significantly contributed to pharmacologic variability and adverse drug reactions. "Pharmacogenetics" was born.

"Pharmacogenetics," is a study which initially focused on the genetic polymorphisms, drug-metabolizing enzymes (DME's) and how this is translated into inherited differences in drug effects. Its definition is now refined as "the study of the hereditary basis for differences in a population's response to a drug" reflecting the multitudes of genetic variations that can impact on drug effects at population levels. In general, two types of pharmacogenetic conditions can be differentiated: first, genetic conditions transmitted as single factors altering the way drugs act on the body (altered drug action); and

second, genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism) (Eichelbaum and Evert, 1996).

The term "pharmacogenetics" owed its origin to Vogel and Motulsky who coined the term 'pharmacogenetics' to describe genetically determined variations in drug response (Eichelbaum and Evert, 1996). These pharmacogenetic conditions can occur either as rare defects or a polymorphism. "Genetic polymorphism" defines a monogenic trait that exists in the population in at least two phenotypes (and presumably at least two genotypes), neither of which is rare (i.e. the rarest phenotype still occurs at the frequency of at least 1%).

Genetic polymorphisms have been described for a wide variety of drug and xenobiotic metabolizing enzymes. Many of these variations were first identified by the occurrence of adverse reactions after normal doses of drugs in patients or volunteers. They are indeed common occurrences. Mutations in the genes for numerous drug-metabolizing enzymes cause enzyme variants with higher or lower activity or lead to the partial or total absence of an enzyme protein.

At the molecular level, inherited differences in drug-metabolizing capacity are generally monogenic traits. Their influence on the pharmacokinetics and pharmacologic effects of medications is determined by their importance for the activation or inactivation of drug substrates or their progenies. In general, the effects can either be profound toxicity for active medications that have a narrow

therapeutic index and are inactivated by a polymorphic enzyme (for example, mercaptopurine, azathioprine, thioguanine, and fluorouracil) (Krynetski and Evans, 1998) or reduced efficacy of medications that require activation by an enzyme exhibiting genetic polymorphism (such as codeine) (Desmeules *et al.*, 1991).

In the last half a century, interest in pharmacogenetics, especially the association between decreased drug clearance and decreased activity of a drug-metabolizing enzyme, the inherited nature of the deficiency, and its frequency and clinical importance rapidly gained ground. Researchers compete to evaluate these phenomena. Prior to advances in molecular genetic techniques, phenotyping methods involving administration of probe drugs and measurement of metabolites in body fluids or determination of enzyme activity in families and populations were used extensively. It involves the use of *in vivo* probe drugs that are metabolized by single polymorphically expressed CYPs to one or more metabolites. To characterize CYP2D6 for instance, debrisoquine which is metabolised to 4-hydroxydebrisoquine, is used. To characterize CYP2A6, nicotine is used. It produces cotinine. To characterize CYP2C8 amiadoquine is used. The administered drug and its metabolites are usually measured in plasma, saliva or urine permitting measurement of parent-to-metabolite ratios in plasma (single or multiple time points) or urine (4 to 12 hour collection) (Spina *et al.*, 1997, Gonzalez and Idle, 1994). In high throughput studies, multiple probes are administered as "cocktails", so that the activity of different CYPs can be measured simultaneously.

Phenotyping methods were and remain popular. For examples, there are abundant examples in the literature on the use of phenotyping in population studies. Phenotyping for CYP2D6 and CYP2C19 are the best characterized. For CYP2C19, phenotyping is usually with mephenytoin because the 4'-hydroxylation of (S)-mephenytoin is catalyzed by CYP2C19 exclusively. The (S)-mephenytoin/(R)-mephenytoin ratio ("S/R ratio") in urine, recovery of 4'-hydroxy mephenytoin in urine (8 hr), or a "hydroxylation index" (molar dose of (S)-mephenytoin divided by the molar amount of 4'-hydroxy mephenytoin recovered in urine) are used as measures of CYP2C19 activity and phenotype (Streetman *et al.*, 2000). Such studies revealed that the incidence of PMs in various Caucasian populations of European decent varied from 0.9% to 7.7% (mean of 2.9%) compared to 13-23% (Ibeanu *et al.*, 1998) in Oriental populations. "Metabolic ratios" (MRs) for proguanil (proguanil/ cycloguanil ratio in urine) and omeprazole (5-hydroxy omeprazole/omeprazole ratio in plasma) have also been used as indices of CYP2C19 phenotype (Streetman *et al.*, 2000).

The 2-hydroxylation of desipramine, 4-hydroxylation of debrisoquine, O-demethylation of dextromethorphan, oxidation of sparteine, and hydroxylation of metoprolol are all catalyzed by CYP2D6 and these reactions have been used to phenotype subjects for CYP2D6 polymorphism. The desipramine concentration and metoprolol/α-hydroxy metoprolol ratio methods require plasma, while urine and saliva (dextromethorphan) samples are used in debrisoquine/4-hydroxy debrisoquine, sparteine/ (2, 3- and 5, 6-dehydrosparteine methods) (Streetman *et al.*, 2000). With these methods (Streetman *et al.*, 2000, Bertilsson, 1995), it

has been shown that the incidence of PM subjects in Caucasian and Oriental populations averages 7% and 1%, respectively.

Interest in the molecular mechanism for the polymorphisms soon emerged. With the advent of molecular genetics, during the past several years, several of these polymorphisms have been studied at the protein and gene level. Many laboratories including ours subsequently developed simple DNA tests to predict the phenotypes. Thus, in its humble beginning genetics depended on family studies but more recently, when molecular genetics began to be used, it allowed the investigation of DNA sequences that cause the difference. This change of the science of genetics has also much affected pharmacogenetics. In the not too distant pasts, differences in drug response between ethnic populations were often called cross-cultural differences, especially by psychiatrists, who could not believe a genetic cause of the observed differences.

Historically, initial work to elucidate the molecular genetic basis for inherited traits in pharmacology began in the late 1980s, with the cloning and characterization of a human gene encoding the drug-metabolizing enzyme debrisoquine hydroxylase (CYP2D6), a prototype of genetically polymorphic drug metabolising enzyme (Gonzalez *et al.*, 1988). Genes are considered functionally "polymorphic" when stable allelic variants exist in populations, one or more of which alters the activity of the encoded protein in relation to the wild-type sequence. In many cases, the genetic polymorphism is associated with reduced activity of the encoded protein, but there are also examples where the

allelic variant encodes proteins with enhanced activity. Since the cloning and characterization of CYP2D6, human genes involved in many such pharmacogenetic traits have been isolated, their molecular mechanisms have been elucidated, and their clinical importance has been more clearly defined.

Molecular methods have many advantages. Genotyping methodologies can be easier to use than biochemical measurements in a clinical setting. They are less invasive. Genotyping requires only a single sample, whereas biochemical measurements often require various body fluids and/or tissue biopsies taken at multiple time points. Genotyping yield the same results across different laboratories, thus making it ideally suited to a diagnostic application (Kashuba *et al.*, 1998); in contrast, phenotyping may yield more variable data between laboratories and at different time points for the same patient. Intra-individual variation does not exist in genotype data; on any given day a patient's genotype will be exactly the same. Furthermore, often a very small amount of material, such as blood from a single finger prick or a buccal swab, is all that is needed to complete genotyping assays. Genotyping results can also often be obtained more quickly than phenotyping results, and future technologies may offer rapid "bedside" assessment.

Genotyping is more stable. It is not influenced by drug-drug interactions, or food-drug interactions. There are no problems with compliance or occurrence of adverse drug reactions among the test subjects, simplifying ethical considerations. It is also relatively easy to perform with samples of genomic DNA from either tissue, leukocytes, buccal swabs, fingernails and hair.

Characterization is through the use of different assay methods like PCR-restriction fragment length polymorphism (RFLP), allele specific PCR, fluorescent dye-based genotyping, mass spectrometry, and gene chip technology (Shi *et al.*, 2001, Cronin *et al.*, 2001, Bertilsson, 1995, Linder and Vales, 1999). It is also possible to obtain genotype data for multiple CYPs, or different allelic variant forms of the same CYP, using a single sample of DNA (Tamminga *et al.*, 2001). PCR-based genotyping is used extensively to investigate the genetic mechanism of the genetic polymorphisms in different populations, races and ethnic groups such as African-Americans, Caucasian-Americans, Caucasian-Europeans (Northern and Southern), Japanese, Chinese, African and Canadian Native Indian and even Malaysians, like in this study.

The developed technologies have also yielded very interesting observations of interethnic differences in pharmacogenetics. Phenotyping studies have previously clearly revealed interethnic and population differences in pharmacogenetics. At the molecular level, the differences have been more clearly defined. Differences in the types and frequencies of alleles are for instance observed with two of the allelic forms of CYP2D6 (CYP2D6*4 and CYP2D6*10). In Chinese, CYP2D6*4 is rare or occurs at low frequencies but is relatively more common in the West. On the other hand, CYP2D6*10 is a predominantly Oriental allele (Gaedigk, 2000). Similarly, clinically relevant interethnic differences in CYP2C9 genotype have been observed, where the allele frequency of CYP2C9*2 (8 to 14%) and CYP2C9*3 (6 to 16%) varies in different Caucasian populations (Garcia-Martin *et al.*, 2001, Aynacioglu *et al.*,

1999, Yasar *et al.*, 1999, Nasu *et al.*, 1997). These alleles occur very rarely or at very low frequencies in Japanese and other Orientals (Nasu *et al.*, 1997).

The recent completions of the Human Genome Project have further changed the face of genetics, creating genomics and pharmacogenomics, where high technology is blended with pharmacogenetics. Whereas pharmacogenetics was based on biochemistry, the new pharmacogenomics takes advantage of high throughput DNA sequencing, gene mapping, and bioinformatics. The result will be a quantum leap in the ability to discover genes, whether they are for physical attributes, disease susceptibility, or the response to drugs. By understanding which genetic factors are responsible for whether a patient will benefit from a drug or toxin or be at risk for a particular side effect, it is possible to develop tests to predict these responses before exposure to the drug or toxin.

1.2 Drug Metabolizing Enzymes (DMEs)

In the above section, drug metabolising enzymes, chemical entities that have great influence on pharmacologic variability were extensively referred to. Generally, chemical compounds foreign to living organisms are actively metabolized by enzymatic reactions. Collectively, these enzymes are called xenobiotic or drug metabolizing enzymes (DMEs). Their evolution is most definitely not related to the human use of drugs as we know it today. Their evolution was probably related to the 'wars' life had to fight when it moved from

the sea to the land. And they are most important in causing variability in human exposures to exogenous substances.

An ancestor of one of the DME families, cytochrome P450, probably arose some 1.5 billion years ago in primitive organisms. Hundred millions of years later, animals moved from water to land and began to use plants as diet. To protect themselves, plants developed phytotoxins, toxins that repel animals. Animals in turn, adapted to these toxins by generating a system to metabolise them and rendered them less harmful. The prehistoric P450 genes evolved into many families and subfamilies, thought to be a response to the great variety of toxins present in the diet and ambient surroundings (Gonzalez and Nebert, 1990). With further move into land, and under the pressure of natural selection, due to the accumulating mutations, deletions, duplications and other changes, P450 families exhibit considerable interspecies and inter-individual variation that later became 'characteristics' to the species and their localities and now understood to be the root to geographical difference in pharmacology of drugs and toxins.

The principal function of DMEs is to metabolise the large number of endogenous and exogenous compounds. In human, they are importantly involved in the biotransformation of exogenous agents, such as drugs and solvents and they also are involved in the activation of pro-carcinogens to carcinogens (Pelkonen and Raunio, 1997). Although their physiologic role is not fully elucidated, they also participate in the synthesis and degradation of several steroid hormones and other small molecular endogenous compounds.

Metabolism usually results in the inactivation of the parent compound (metabolic activation) (Nebert, 1997a) but in some cases, as with pro-carcinogens, active products are formed.

The clinical consequence of the conversion therefore varies. Whether a genetic polymorphism has relevance for drug therapy depends on the characteristics of the drug in question. The quantitative role of a DME in the overall kinetics of a drug and the agent's therapeutic range will determine how much the dose has to be adjusted in poor metabolisers or ultrarapid metabolisers. The example of the CYP2D6 polymorphism provides clinical evidence for these ideas. Most patients (about 90%) require 75-150 mg/day of nortriptylline to reach a "therapeutic" plasma steady-state concentration of 200-600 nmol/L, but poor metabolisers need only 10-20 mg/day to reach the same concentrations. Ultrarapid metabolisers, on the other hand, may require 300-500 mg/day or even more to reach the same plasma concentration. Obviously, if the genotype or phenotype of the patient is not known, poor metabolisers will be overdosed and be at high risk of drug toxicity, whereas ultrarapid metabolisers will be under-dosed (Bertilsson *et al.*, 1997, Dalen *et al.*, 1998). On the other hand, for drugs like codeine and enalapril, metabolism produces the active compounds required for efficacy and individuals lacking in the particular enzyme will not enjoy the benefits of the prescribed drugs. Of a further interest, almost all exogenous carcinogens require activation by metabolic enzymes, and detoxification enzymes frequently exist to deactivate carcinogens or their intermediate metabolites. Inherited polymorphisms in these enzymes may alter their rate of activation or detoxification, thus increasing or

decreasing the carcinogenic potential of the environmental exposures they act on.

Most drugs initially possess lipophilic characteristics that promote the passage of the drug through cell membranes to its site of action. Lipophilic drugs must be bio transformed into more hydrophilic metabolites to facilitate elimination and excretion. The cytochrome P450 system (CYP) is responsible for metabolizing many endogenous and exogenous substances, including 40% to 50% of all medications (Ingelman-Sundberg *et al.*, 1999), into more hydrophilic substances. Overall, approximately 10 P450s are responsible for the metabolism of a large number of pharmacologic agents in humans.

DMEs are categorised into two groups, i.e., phase I (functionalisation reactions) and phase II enzymes (conjugation reactions) (Nebert and Gonzalez, 1987). Phase I enzymes mediate the first step in the metabolism of xenobiotics into more hydrophilic forms. These enzyme families are categorised by their ability to metabolise different substrates (Guengerich, 1992).

P450 enzymes are phase I enzymes that catalyse the mono-oxidation or reduction and epoxidation of their substrates. Hydroxylases, flavin-containing mono-oxygenases (FMOs) and mono-aminoxidases (MAOs) are other phase I enzymes (Nebert, 1994). These enzymes add small molecules (e.g., OH^- or O-groups) to their substrates. The metabolism of xenobiotics occurs mainly in liver (Daly *et al.*, 1993).

Phase II enzymes usually detoxify, but sometimes also activate, endogenous or exogenous compounds. The substrate is converted into more hydrophilic form by conjugating with, for instance, glutathione. Very lipophilic compounds are usually first metabolised by phase I enzymes and then by phase II enzymes, whereas less lipophilic compounds can be metabolised directly by phase II enzymes. Phase II enzymes are categorised as glucuronyl transferases (UGTs), sulfotransferases (SULTs), glutathione-S-transferases (GSTs) and N-acetyltransferases (NATs). Many phase II enzymes have been shown to exhibit genetic polymorphisms (Mackenzie *et al.*, 1997, Nebert, 1997b, Pelkonen *et al.*, 1998, Hirvonen, 1999). For example, the GSTM1 enzyme has been found to be polymorphic, and a total gene deletion (*GSTM1-1* null) has been suggested to be a risk factor in lung cancer caused by tobacco smoke (London *et al.*, 1995, Weinberg and Sandler, 1999, Bennett *et al.*, 1999). GSTs metabolise aflatoxin B1, PAHs, and some other compounds present in tobacco smoke.

1.3 What is Cytochrome P450?

Cytochrome P450 is a cellular chromophore that was first named in 1961, because the pigment (P) has a 450 nm spectral peak when reduced and bound to carbon monoxide (Gonzalez, 1989). P450s comprise well characterized group of phase I enzymes. In endoplasmic reticulum (ER), a P450 enzyme together with appropriate cofactors, such as NADPH, metabolises chemicals by breaking up oxygen molecules. The chemical to be metabolised receives a hydroxyl group, while the other oxygen atom is inserted into a water molecule. This reaction is called mono-oxygenation (Degtyarenko and Archakov, 1993).

The active centre of the P450 enzyme contains heme, and the substrate is in a proper orientation with respect to the heme iron and oxygen molecule for the reaction to take place (Daly *et al.*, 1993).

P450 is involved in the metabolism of drugs, environmental pollutants, dietary chemicals and endogenous compounds (Nelson *et al.*, 1996). Since first described decades ago, tens of thousands scientific papers have been published on the P450. This is a clear reflection of the importance that the system plays in the metabolism of a wide range of both endogenous compounds and drugs. Initially the system was thought to be a single enzyme, but it soon became clear in the 1970's that there exists multiple forms of this haemoprotein, and they are products of discrete genetic loci. The organisation of the P450 superfamily into families and subfamilies has been reviewed on multiple occasions (Gonzalez, 1992, Nelson *et al.*, 1993, Archakov and Degtyarenko, 1993), but it has always been clear that understanding the role of individual forms of P450 in the metabolism of drugs would be a gigantic task, given the number and overlapping specificity of the enzymes involved.

The P450 enzymes are widely distributed across species. It was not initially known that different species and tissues had similar isoforms. Names were assigned to the enzymes according to their spectral properties, electrophoretic mobility or their substrates. However, with new knowledge of the amino acid sequences for the enzymes, a general nomenclature based on the presence of common amino acid sequence was proposed by Nebert (Nelson *et al.*, 1996). This system groups the enzymes and genes into families and

subfamilies with the prefix 'CYP' to designate P450 enzymes in all species (except Drosophila and mouse gene where 'Cyp' is used). Fourteen families and 26 subfamilies exist in all mammals to date, 20 of the subfamilies have been mapped to the human genome (Benet *et al.*, 1996). Families are defined as having more than 40% homology in their amino acid sequence and are designated by an Arabic numeral, and subfamilies have more than 55% analogy and are designated by a capital letter. An Arabic numeral after the letter denotes the individual enzyme and the gene associated with the enzyme is denoted in italics. For example, the CYP2 family (Nelson *et al.*, 1996, Slaughter and Edwards, 1995), has several subfamilies such as CYP2C, CYP2D and CYP2E. The individual enzyme is denoted by a numeral, as in CYP2D6, and the gene is denoted as *CYP2D6*.

Three of the CYP families – CYP1, CYP2 and CYP3 – are the most important in metabolizing xenobiotics and have been intensively studied by toxicologists and biochemists. The largest and most complex of these is the CYP2 family, which consists of multiple subfamilies; in both the mouse and human genomes, individual genes within its subfamilies have similar sequences and are physically clustered (Nelson *et al.*, 1993).

That P450 is important in therapeutics is undeniable even though efforts are made to design the new generation of therapeutic agents that comprises a number of molecules of higher molecular weight which are not subject to oxidative metabolism. The fact remains that the most existing drugs and indeed many drugs of the future (Bossart and Pearson, 1995, DiMasi, 1995, Drews,

1995, Kleinberg and Wanke, 1995, MacInnes *et al.*, 1994, Smith, 1994) depend on, or will depend on the system to terminate their biological effect, or indeed cause their side effects or adverse reactions (Smith, 1994).

As alluded above, P450 also plays established roles in the metabolic activation of carcinogens (Miller, 1994, Waller, 1994, Shubik, 1995). There is growing evidence that variability in P450 involved, qualitatively or quantitatively, may contribute to interindividual differences in cancer susceptibility (Wolf *et al.*, 1994, Gonzalez and Gelboin, 1994, Caporaso and Goldstein, 1995, Rannug *et al.*, 1995). It is therefore important that the function of the individual forms of P450, together with their specificity and regulation, be understood for many reasons including the design of drugs with a more desirable pharmacokinetic profile and reduced toxicity, the detection of `at risk' groups and the targeting of new therapeutic agents to specific metabolic pathways. Indeed recognition of the importance of these issues led to the establishment of a Human Liver Bank (Boobis *et al.*, 1980).

1.4 Why Study CYP2A6?

CYP2 gene family-CYP2 is the largest P450 family in mammals. In human, CYP2C8, CYP2C9, CYP2C18, and CYP2C19 are the major DME's and they together metabolise more than 50% of prescribed drugs. Other CYP's involved in drug metabolism include CYP2A6, CYP2A13, CYP2B6, CYP2E1, CYP2F1, and CYP2J2.

CYP2A6 was first identified as the human coumarin 7-hydroxylase (Miles *et al.*, 1990, Yamano *et al.*, 1990, Yun *et al.*, 1991) and is now known to metabolise several compounds including nicotine (Messina *et al.*, 1997), coumarin (Miles *et al.*, 1990) and the anticancer tegafur (Ikeda *et al.*, 2000). Apart from that, CYP2A6 also activates pro-toxins and procarcinogens, including many nitrosamines and aflatoxin B1 and several tobacco-specific nitrosamines (Fujita and Kamataki, 2001, Sellers *et al.*, 2003). Some compounds have been found to inhibit CYP2A6 enzyme *in vitro* and they include methoxsalen (8-methoxypsoralen) (Koenings and Trager, 1998, Draper *et al.*, 1997, MaÈenpaÈaÈ *et al.*, 1993) , menthofuran (Khojasteh-Bakht *et al.*, 1998), pilocarpine (Kimonen *et al.*, 1995) and tranylcypromine.

CYP2A6 is predominantly expressed in the liver (Koskela *et al.*, 1999). Expression of CYP2A6 mRNA is highly variable (Koskela *et al.*, 1999) and this variation can be partly explained by the polymorphic nature of the CYP2A6 gene. There is thus a pronounced interindividual and interethnic variability in CYP2A6 levels and activity. To date, at least 20 variants of the CYP2A6 gene and numerous single nucleotide polymorphisms (SNPs) (see http://www.imm.ki.se/ CYPalleles/cyp2a6.htm) have been described.

Table 1.1 CYP2A6 substrates

Drugs	*Assay/end-point*	*Toxic agents*	*Assay/end-point*
Coumarin	7-hydroxylation	Nicotine	*N*-1′-oxidation
Methoxyflurane	Dehalogenation	Cotinine	3′-hydroxylation
Halothane	Reduction	NNK	Mutagenicity
SM-12502	S-oxidation	NDEA	Mutagenicity
Losigamone	Oxidation	AFB_1	Mutagenicity
Valproic acid	Oxidation	MOCA	*N*-oxidation
Letrozole	Oxidation	1,3-butadiene	Monoxide formation
Disulfiram	Sulfoxidation	Quinoline	1-oxidation
		DCBN	Protein adduct formation
		MTBE	*O*-demethylation

Adapted from (Raunio, 2001). SM-12502, 3,5-dimethyl-2-(3-pyridyl)thiazolidin-4-one hydrochloride; NNK, 4-methylnitrosamino-1-(3-pyridyl)-1-butanone; NDEA, N-nitrosodiethylamine; AFB1, aflatoxin B1; MOCA, 4,4k-methylenebis(2-chloroaniline); DCBN, 2,6-dichlorobenzonitrile; MTBE, methyl tert-butyl ether .

1.5 Cigarette Smoking and CYP2A6

Cigarette smoking is governed by the need to maintain desirable levels of nicotine in the body (Benowitz *et al.*, 1998, Benowitz, 1996). Approximately 80% of nicotine is C-oxidized to the inactive metabolite cotinine (Benowitz and Jacob, 1994); approximately 90% of this pathway is mediated by CYP2A6 (Nakajima *et al.*, 1996). *In vitro* studies using human liver microsomes and recombinant P450s have shown that CYP2A6 is the most important P450 responsible for the C-oxidation of nicotine (Nakajima *et al.*, 1996, Messina *et al.*, 1997).

CYP2A6 has a high-affinity for both nicotine and its oxidized metabolite cotinine. These have been revealed in studies with heterologously expressed CYP enzymes and human liver microsomes *in vitro* (Figure 1). The reaction is a two-step reaction in which nicotine is first oxidized to nicotine iminium ion and subsequently to cotinine by cytosolic aldehyde oxidase. The reaction rate is limited by the CYP2A6-mediated formation of nicotine iminium ion from nicotine but up to 80% of nicotine is metabolized to cotinine. Cotinine is turn, oxidized to several metabolites, also by CYP enzymes, including possibly CYP2A6 (Murphy *et al.*, 1999). At low physiologic concentration, CYP2A6 appears to be the only enzyme involved in the metabolism of nicotine (50 m M) but at high nicotine concentrations (500 mM), other CYP isoforms, such as CYP2B6, may also play a part (Yamazaki *et al.*, 1999).

Of further importance, individuals homozygous for a CYP2A6 gene deletion displayed only 15% of urinary cotinine levels compared with individuals carrying at least one active CYP2A6 gene after smoking the same number of cigarettes (Kitagawa *et al.*, 1999). Because of the substantial involvement of CYP2A6 in nicotine elimination, it has been proposed that the *CYP2A6* polymorphism is a major determinant of an individual's smoking behaviour. Recently, the total inhibition of CYP2A6 function was also suggested to reduce smoking, and therefore may have a role in smoking cessation and tobacco exposure reduction (Sellers and Tyndale, 2000a, Sellers *et al.*, 2000b).

The first case-control study to determine the role of CYP2A6 in maintaining smoking was done in a Canadian population by Pianezza *et al* (1998). Results showed that at least one variant allele of the *CYP2A6* reduced the risk to smoking (Pianezza *et al.*, 1998). The defective genotype was similarly also found to reduce smoking (Rao *et al.*, 2000). Contradictory results have however also been obtained. In a study on Americans by London *et al.* (1999) no evidence was found for an association between reduced CYP2A6 activity and the number of cigarettes consumed per day in smokers. Similarly, Sabol and Hamer (1999) and Tan *et al.* (2000) found no correlation between the *CYP2A6* genotype and cigarette smoking behaviour and further similar results were obtained in French and Finnish populations (Tiihonen *et al.*, 2000, Loriot *et al.*, 2001).

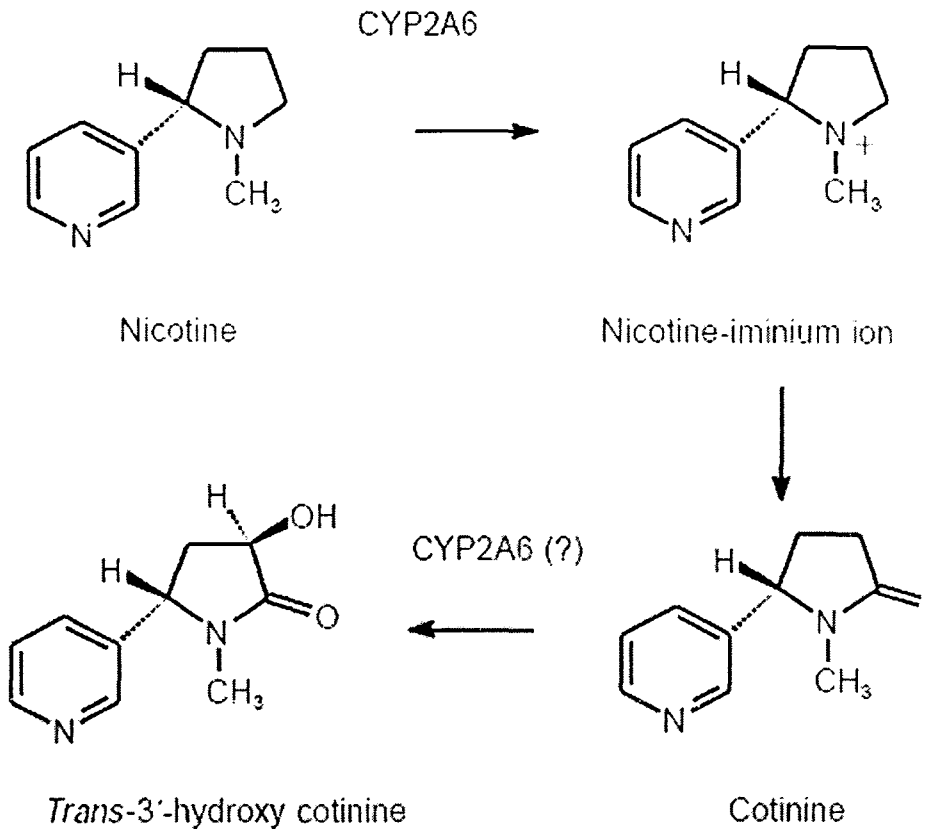

Figure 1.1 Participation of CYP2A6 in the oxidative metabolism of nicotine and cotinine

1.6 *CYP2A6* Polymorphism and its Analysis by Polymerase Chain Reaction

As alluded, *CYP2A6* is genetically polymorphic. Phenotyping results revealed a pronounced inter-individual variability in the CYP2A6 activity (Rautio *et al.*, 1992, Cholerton *et al.*, 1992, Iscan *et al.*, 1994). In Caucasian populations, the poor metaboliser (PM) phenotype for CYP2A6 appears to occur at very low frequencies and very rarely are cases reported where individuals completely lacked CYP2A6 activity (Hadidi *et al.*, 1997, Oscarson *et al.*, 1998). On the other hand, the frequency of PMs for CYP2A6 among Japanese seems to be higher and in vitro studies using microsomes from Japanese subjects demonstrated that eight out of 30 livers had a very low or no CYP2A6 immunoreactivity and activity (Shimada *et al.*, 1996).

The subfamilies *CYP2A, CYP2B, CYP2F* and *CYP2S* comprise a gene cluster on chromosome 19 (Kwon *et al.*, 2001, Hoffman *et al.*, 1995, Sheng *et al.*, 2001, Rylander *et al.*, 2001) with *CYP2A* comprising of *CYP2A6*, *CYP2A7*, *CYP2A13*, and more than 3 pseudogenes (Hoffman *et al.*, 1995). *CYP2A7* and *CYP2A6* have 94% amino acid homology but *CYP2A7* is functionally inactive (Yamano *et al.*, 1990). *CYP2A13* is highly homologous and although it is also able to metabolize coumarin, nicotine and tobacco-smoke nitrosamine, its role is limited in this (Su *et al.*, 2000).

Several variant alleles of the CYP2A6 gene have been characterized, beginning about 10 years ago when the sequences of the wild-type and one

variant allele were reported. The wild-type alleles *CYP2A6*1A* and *CYP2A6*1B* (containing 58 bp of *CYP2A7* in the 3'UTR) is a result of a conversion in the 3' flanking region of the gene but they are similar in enzymatic function (Oscarson *et al.*, 1999a, Miyamoto *et al.*, 1999). *CYP2A6*2* encodes a protein with a L160H substitution which does not incorporate heme and is inactive *in vitro* (Inoue *et al.*, 2000) and in vivo (Benowitz *et al.*, 1995, Hadidi *et al.*, 1997). *CYP2A6*3*, originally thought to be a true variant allele, has later been shown to be lacking in different populations (Oscarson *et al.*, 1999b). One reason for the original misclassification turned out to be due to the very common *CYP2A6*1B* allele exhibiting a gene conversion between 3' flanking regions of the CYP2A6 and CYP2A7 genes (Oscarson *et al.*, 1999a). Consequently, *CYP2A7* sequences are also amplified when the original genotyping method was used (Salguero *et al.*, 1995), causing misclassification. Nevertheless, *CYP2A6*3* may indeed be a true but an inactive and rare allele created by multiple sequence conversions between *CYP2A6* and *CYP2A7* (Salguero *et al.*, 1995). The occurrence of multiple gene and pseudogene duplications, high sequence homology, and the gene conversion events in *CYP2A6* loci, have led to the originally erroneous genotyping methods (Salguero *et al.*, 1995). This lead to misidentification of *CYP2A7* as *CYP2A6*2*, and *CYP2A6*1B* as *CYP2A6*3* alleles. Several other *CYP2A6* alleles have been created by the unequal crossover events. Thus, the *CYP2A6*4A* deletion allele is generated by an unequal crossover between the 3' flanking regions of the CYP2A6 and CYP2A7 genes. Similarly, In the *CYP2A6*4D* allele, the crossover region occurs in intron 8 or exon 9 (Oscarson *et al.*, 1999b). Two point mutations (*CYP2A6*2* and

*CYP2A6*5*) abolish the catalytic activity of the encoded enzyme, similar with the gene deletions (Yamazaki *et al.*, 1999, Oscarson *et al.*, 1999a).

The frequency of the inactive alleles is low in European populations and very few poor metabolisers for the probe drug coumarin have been described in these population. In contrast, a relatively high allele frequency (15-20%) of the CYP2A6 gene deletion has been found in Asians, resulting in a generally reduced activity in these populations.

Basic to the study of the genetic polymorphism of DMEs is the availability of methods to identify the variations. Therefore to study the genetic polymorphism of *CYP2A6* specific detection of *CYP2A6* allelic variants would be useful to study human populations for the presence of alleles that predict variable responses.

A variety of genotyping techniques have been developed to identify polymorphisms. Commonly used methods include gel electrophoresis-based techniques, such as polymerase chain reaction (PCR) coupled with restriction fragment length polymorphism (RFLP) analysis, multiplex PCR, and allele-specific amplification. Fluorescent dye-based high-throughput genotyping procedures have gained increased popularity, including the oligonucleotide ligation assay (OLA), heterozygote sequencing, TaqMan (Perkin Elmer, Foster City, CA) allelic discrimination, and high-density chip array technology. Novel polymorphisms also can be identified by single-strand conformation polymorphism (SSCP) and direct heterozygote sequencing. Mass spectrometry