SIGNIFICANCE OF POLYMORPHISMS IN CYP2A6 GENE

HARRIET GULLSTÉN

Department of Pharmacology and Toxicology, University of Oulu

OULU 2000
HARRIET GULLSTÉN

SIGNIFICANCE OF POLYMORPHISMS IN CYP2A6 GENE

Academic Dissertation to be presented with the assent of the Faculty of Medicine, University of Oulu, for public discussion in the Auditorium of the Department of Pharmacology and Toxicology, on January 26th, 2001, at 10 a.m.

OUŁUN YLIOPISTO, OULU 2000
Gullstén, Harriet, Significance of polymorphisms in CYP2A6 gene
Department of Pharmacology and Toxicology, University of Oulu, P.O.Box 5000, FIN-90014 University of Oulu, Finland
2000
Oulu, Finland
(Manuscript received: 18 December 2000)

Abstract
Cytochrome P450 2A6 (CYP2A6) is involved in the 7-hydroxylation of coumarin, C-oxidation of nicotine, and the metabolism of tobacco specific nitrosamines. Initially in 1995 Fernandez-Salguero et al. reported a genotyping method for three alleles: CYP2A6*1 (wild-type), CYP2A6*2 (variant 1), and CYP2A6*3 (variant 2). Later studies presented in this thesis indicated that the original genotyping method produces erroneous results for the CYP2A6*3 allele due to unspecific PCR conditions and previously unknown CYP2A6*1B allele. Furthermore, the CYP2A6*2 allele genotyping caused erroneous genotypes (CYP2A6*2/*2 was misclassified as CYP2A6*1/*2).

In this work, new PCR based genotyping methods were developed for CYP2A6*2 and for several new alleles (CYP2A6*1B, CYP2A6*4A/*4D and CYP2A6*5). In population-based studies, the deletion alleles (pooled as CYP2A6*4) turned out to be more prevalent among Asians (15.1%) than Caucasians (0.5%). The frequencies of the other inactive alleles varied within 0-3% in both populations. Asians totally lacked the CYP2A6*2 allele, whereas Caucasians lacked the CYP2A6*5 allele. The frequencies of two wild-type alleles, CYP2A6*1A and CYP2A6*1B alleles were 66.5% and 30.0% in Caucasians, and 43.2% and 40.6% in Asians, respectively.

Correlation studies between the phenotype, as tested by the administration of coumarin, and the genotype demonstrated that individuals with the CYP2A6*2/*2 genotype were totally defective, while CYP2A6*1/*2 subjects exhibited intermediate and CYP2A6*1/*1 subjects full capability of producing 7-hydroxycoumarin. Upon phenotyping with nicotine, individuals with the CYP2A6*1/*2 or CYP2A6*1/*4 genotype were shown to have a lower enzyme activity (one fourth of the normal activity), compared to those with the CYP2A6*1/*1 genotype.

Defective CYP2A6 activity has been hypothesised to reduce the risk of environmentally (especially tobacco smoke) induced diseases either by decreasing production of genotoxic metabolites or by preventing addiction to tobacco smoking. However, in our case-control studies on Spanish patients with liver cirrhosis (n = 83) and liver cancer (n = 90) and their controls (n = 237) no significant association between the CYP2A6 genotypes and disease proneness was found. The odds ratio (OR) for developing liver cancer was was 1.4 (95% confidence interval [CI] 0.5-3.7) for genotypes containing at least one CYP2A6*2 allele. For liver cancer the respective OR was 1.3 (95% CI 0.4-4.5). Similarly, no statistically association between CYP2A6 alleles and the risk of lung cancer was observed in our Finnish study population consisting of 177 cases and 1089 controls; the OR for combined CYP2A6 variant allele containing genotypes (CYP2A6*1/*2 and CYP2A6*1/*4) was 1.19 (95% CI 0.56-2.45). Our studies therefore do not indicate any major modifying role for the CYP2A6 genotypes in individual susceptibility to environmentally induced diseases.

Keywords: coumarin, nicotine, liver cirrhosis, liver and lung cancer.
Gullstén, Harriet, Significance of polymorphisms in CYP2A6 gene
Department of Pharmacology and Toxicology, University of Oulu, FIN-90014 University of Oulu, Finland
Oulu, Finland
(Manuscript received: 18 December 2000)

Abstract

Cytochrome P450 2A6 (CYP2A6) is involved in the 7-hydroxylation of coumarin, C-oxidation of nicotine, and the metabolism of tobacco specific nitrosamines. Initially in 1995 Fernandez-Salguero et al. reported a genotyping method for three alleles: CYP2A6*1 (wild-type), CYP2A6*2 (variant 1), and CYP2A6*3 (variant 2). Later studies presented in this thesis indicated that the original genotyping method produces erroneous results for the CYP2A6*3 allele due to unspecific PCR conditions and previously unknown CYP2A6*1B allele. Furthermore, the CYP2A6*2 allele genotyping caused erroneous genotypes (CYP2A6*2/*2 was misclassified as CYP2A6*1/*2).

In this work, new PCR based genotyping methods were developed for CYP2A6*2 and for several new alleles (CYP2A6*1B, CYP2A6*4A/*4D and CYP2A6*5). In population-based studies, the deletion alleles (pooled as CYP2A6*4) turned out to be more prevalent among Asians (15.1 %) than Caucasians (0.5 %). The frequencies of the other inactive alleles varied within 0–3 % in both populations. Asians totally lacked the CYP2A6*2 allele, whereas Caucasians lacked the CYP2A6*5 allele. The frequencies of two wild-type alleles, CYP2A6*1A and CYP2A6*1B alleles were 66.5 % and 30.0 % in Caucasians, and 43.2 % and 40.6 % in Asians, respectively.

Correlation studies between the phenotype, as tested by the administration of coumarin, and the genotype demonstrated that individuals with the CYP2A6*2/*2 genotype were totally defective, while CYP2A6*1/*2 subjects exhibited intermediate and CYP2A6*1/*1 subjects full capability of producing 7-hydroxycoumarin. Upon phenotyping with nicotine, individuals with the CYP2A6*1/*2 or CYP2A6*1/*4 genotype were shown to have a lower enzyme activity (one fourth of the normal activity), compared to those with the CYP2A6*1/*1 genotype.

Defective CYP2A6 activity has been hypothesised to reduce the risk of environmentally (especially tobacco smoke) induced diseases either by decreasing production of genotoxic metabolites or by preventing addiction to tobacco smoking. However, in our case-control studies on Spanish patients with liver cirrhosis (n = 83) and liver cancer (n = 90) and their controls (n = 237) no significant association between the CYP2A6 genotypes and disease proneness was found. The odds ratio (OR) for developing liver cancer was was 1.4 (95 % confidence interval [CI] 0.5–3.7) for genotypes containing at least one CYP2A6*2 allele. For liver cancer the respective OR was 1.3 (95 % CI 0.4–4.5). Similarly, no statistically association between CYP2A6 alleles and the risk of lung cancer was observed in our Finnish study population consisting of 177 cases and 1089 controls; the OR for combined CYP2A6 variant allele containing genotypes (CYP2A6*1/*2 and CYP2A6*1/*4) was 1.19 (95 % CI 0.56–2.45). Our studies therefore do not indicate any major modifying role for the CYP2A6 genotypes in individual susceptibility to environmentally induced diseases.

Keywords: coumarin, nicotine, liver cirrhosis, liver and lung cancer
Acknowledgements

The present work was carried out during the years 1995–2000 at the Department of Pharmacology and Toxicology, University of Oulu (1995–1998), and at the Department of Industrial Hygiene and Toxicology, Finnish Institute of Occupational Health, Helsinki (1998–2000). I would like to express my deepest gratitude to my three supervisors Professor Hannu Raunio, Professor Olavi Pelkonen and Docent Ari Hirvonen.

Professor Hannu Raunio was the first person I got acquainted when searching for a research group in which to finalise my graduation paper and to start to do my thesis on cancer. 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 just for me and my problems have helped me grow as a young scientist and a member of the scientific community. I thank Hannu for our many late evening discussions, e-mails and long-distance calls, which always gave me the stimulation to continue my work with new energy and ideas!! Thanks also for lifting up the palm in FST’s banquette…

Professor Olavi Pelkonen, Head of the Department of Pharmacology and Toxicology, has been an enjoyable character who has affected the atmosphere at the department and especially contributed to my project with his never-ending enthusiasm and optimism. His bohemian attitude, encouragement, and constantly expeditious response in everything have been essential for my work, not forgetting the lovely moments outside of scientific work, such as the time in Liseberg, Göteborg or having the special performance at Micke’s dissertation party in Stockholm!

Docent Ari Hirvonen, my third supervisor since June 1999, has provided me with more than perfect facilities, resources and knowledge to continue my work in Helsinki. He supervised also my fifth study included this thesis.

I wish to acknowledge my official referees, Professor Rachel Tyndale and Professor Harri Vainio, for an excellent and careful review of the manuscript of my thesis. I also want to express my thanks to Sirkka-Liisa Leinonen for revising the language.

Department of Pharmacology and Toxicology, University of Oulu:

Several individuals of our research group have had a very significant role during my project, and I would therefore like to express my gratitude to them: Arja Rautio, Markku Pasanen, Pirkko Viitala, Päivi Taavitsainen, Jukka Hakola, Janne Hukkanen, Päivi Tyni,
Ritva Tauriainen, Pauliina Paakki, Marko Uutela, and other younger colleagues at ‘Club Med’. Especially, I would like to thank Päivi Taavitsainen and Pirkko Viitala for sharing important moments with me; we have had many interesting and illuminating discussions touching all aspects of life. I also wish to thank Professor Kirsi Vähäkangas and Professor Heikki Ruskoaho for their role at the Department. My warmest thanks are due to the whole personnel for the CHEERFUL atmosphere at the Department of Pharmacology and Toxicology, especially to Raija Hanni!

Department of Industrial Hygiene and Toxicology, Finnish Institute of Occupational Health, Helsinki:

First, I wish to acknowledge Ari Hirvonen, Kirsti Husgafvel-Pursiainen, Antti Tossavainen, Kai Savolainen and Hannu Norppa for providing perfect facilities for my research in the Institute. I would also like to thank Raija Marttala and Leila Kuusela at the office, and the atmosphere at coffee rooms in the Institute. Concerning my problems with the working practices in the new lab, I wish to give my warmest thank to Tuula Suititala, and for help with the computers, I wish to thank Antti Zitting and Reijo Kuronen. I have shared many nice moments during this project with the researchers in the institute, for which I am especially grateful to Pia Sillanpää, Kristiina Järvinen and Maria Hirvonen. I am especially indebted to Trina Ollikainen for sincere support and fruitful discussions of the thesis, but also in general, during this process!

Collaborators:

I wish to express my sincere gratitude for the very fruitful collaboration with Professor Magnus Ingelman-Sundberg, Mikael Oscarson and Roman A. McLellan. I thank for the opportunity to visit Magnus’s lab in 1998 and the possibility to get to know ‘the atmosphere’ in his lab. I especially wish to thank Micke for INESTIMABLE help at the methodological level during the present work!!

I use this opportunity also to thank all DNA suppliers.

Friends:

I wish to express my deepest gratitude and appreciation to my dearest and most dependable friend and colleague Anna Maria Mattila. The time spent together during the studies and work, the fruitful conversations and e-mails, and her support and encouragement in EVERYTHING have been one of the most important things in my life. I am also very grateful to the following friends of mine, who have, in their own special ways, supported me and brought joy into my life: Petra Könnönen, Minna Salo, Maarit Keckman, Johanna Vartiainen, and Marko Nivalainen. I especially wish to thank following colleagues and friends: Katarina Klintrup, Niina Korpela, and Anne Ruotsalainen.

Above all, I want to dedicate this thesis to my family for their lifetime of love and support: my mother Mirjam, father Reijo, sister Riikka, and grandmother Aino. I also wish to thank some of my relatives and other important people in my life. At the last but not least, I wish to thank Jukka for bringing brightness and craziness into my life outside the lab. His continuous support and love during the last steps of my thesis have been indispensable!!!

This study was supported by the University of Oulu, Finnish Academy of Science, a postgraduate position in the Graduate School of Clinical Drug Research (coordinator Professor Pertti Neuvonen), Finnish Institute of Occupational Health, Finnish Pharmacological Society, Cancer Society of Northern Finland, Finnish Society of


Harriet Gullstén, former Kähkönen
Abbreviations

AFB1  aflatoxin B1  
AhR  Aryl hydrocarbonhydroxylase-receptor  
Arnt  Ah-receptor nuclear translocator  
CYP  cytochrome P450  
EH  epoxide hydrolase  
EM  extensive metaboliser  
FMO  flavin mono-oxygenase  
GST  glutathione S-transferase  
GSH  glutathione (reduced)  
IM  intermediate metaboliser  
kB  kilobase  
MAO  monoamine oxidase  
MR  metabolic ratio  
NDEA  N-nitrosodiethylamine  
NDMA  N-nitrosodimethylamine  
NNAL  4-(methyleneimino)-1-(3-pyridyl)butanol  
NNK  4-(methylnitrosamino)-1-(3-pyridyl)butanone  
NNN  N’-nitrosonornicotine  
NAT  N-acetyltransferase  
PAH  polycyclic aromatic hydrocarbon  
PM  poor metaboliser  
RFLP  restriction fragment length polymorphism  
SCE  sister chromatid exchange  
SNP  single nucleotide polymorphism  
SM-12502  3,5-dimethyl-2-(3-pyridyl)thiazolidin-4-one hydrochloride  
SSCP  single-strand conformation polymorphism  
UM  ultrarapid metaboliser  
UGT  uridinediphosphate-glucuronosyl transferase  
XME  xenobiotic metabolising enzyme
List of tables

1. Frequencies of defective phenotypes of some XMEs.
2. Enzyme activities and expression of $CYP2A$ subfamily.
4. The frequencies and interethnic variation of $CYP2A6$ alleles.
5. The association between $CYP2A6$ and lung cancer.
6. The association between $CYP2A6$ and tobacco smoking.
7. PCR conditions for the original genotyping methods.
8. The primers for PCR in the original genotyping method.
10. SSCP conditions for $CYP2A6$ exon 3.
11. The specificity of primers in the new genotyping methods and the original methods.
12. Published and modified PCR conditions for $CYP2A6^*4$ genotyping.
13. Phenotype-genotype correlation of $CYP2A6$ in the Finnish population.
14. $CYP2A6$ allele frequencies in Caucasian and Asian populations.
15. The $CYP2A6$ genotypes and coumarin and nicotine phenotypes in Finnish individuals.
16. The distribution of $CYP2A6$ genotypes in Spanish cases and controls.
17. The allele frequencies of $CYP2A6$ in Finnish lung cancer and control subjects.
18. The crude and adjusted ORs, and 95% CIs for $CYP2A6$ genotypes.
19. All published genotyping methods for $CYP2A6$ alleles.
21. The differences of CYP2A6 in different ethnic groups in diseases and smoking behavior.
22. The case-control studies done with the new $CYP2A6$ genotyping methods in Caucasian population.
List of figures

1. A schematic pedigree of the development of some XMEs.
2. The CYP2A, CYP2B and CYP2F gene cluster.
3. Routes of coumarin metabolism.
4. Principal metabolic routes of nicotine.
5. The structures of CYP2A6 alleles.
6. Association between metabolism and toxicity of chemical compounds.
9. PCR protocols for the original and new genotyping methods.
10. Schematic structures of CYP2A6 alleles.
List of original articles

This thesis is based on the following articles, which are referred to in the text by Roman numerals:


Contents

Abstract
Acknowledgements
Abbreviations
List of tables
List of figures
List of original articles

1 Introduction .............................................................. 19
2 Review of the literature .............................................. 21
  2.1 Xenobiotic metabolising enzymes (XMEs) ....................... 21
    2.1.1 Endogenous and exogenous substrates ..................... 23
    2.1.2 Phase I enzymes ........................................... 23
    2.1.3 Phase II enzymes ......................................... 24
    2.1.4 Regulation of XMEs ..................................... 24
  2.2 Pharmacogenetics of XMEs ....................................... 25
    2.2.1 Pharmacokinetics and metabolism of drugs ................ 25
    2.2.2 Variation in drug metabolism ............................. 26
    2.2.3 Variation of XMEs in drug metabolism .................... 27
  2.3 Polymorphic CYPs .............................................. 27
  2.4 CYP2A subfamily .............................................. 29
    2.4.1 Expression of CYP2A members ............................. 30
    2.4.2 Substrates ............................................... 31
      2.4.2.1 Coumarin ........................................... 32
      2.4.2.2 Nicotine ........................................... 33
    2.4.3 Genetic regulation and polymorphisms of CYP2A6 ......... 35
  2.5 Chemical-induced diseases and genetic factors ................ 36
    2.5.1 Role of genetic factors in chemical toxicity and carcinogenesis .... 37
    2.5.2 Mechanisms of chemical carcinogenesis .................... 37
    2.5.3 Liver cancer and cirrhosis ................................ 38
    2.5.4 Lung cancer ............................................ 38
    2.5.5 Tobacco smoking and genetic factors ...................... 39
3 Aims of the study .................................................. 41
4 Materials and methods ............................................ 42
  4.1 Study populations ............................................. 42
1 Introduction

Chemical compounds foreign to living organisms are actively metabolised by enzymatic reactions. Collectively, these enzymes are called xenobiotic metabolising enzymes (XMEs). An ancestor of one of the XME families, cytochrome P450 (CYP), probably arose 1.5 billion years ago in primitive organisms. Hundreds of millions of years later, animals moved from water to ground and simultaneously began to use plants as their diet. Therefore, to protect themselves, plants developed phytotoxins, toxins that repel animals, which, in turn, adapted to these toxins by generating a system to metabolise them. During evolution, the ancient CYP genes branched into many families and subfamilies. This is thought to be due to the great variety of toxins present in the diet and ambient surroundings (Gonzalez & Nebert 1990). Nowadays, specialised XME families are found in animals, plants and bacteria. Under the pressure of natural selection and due to the accumulating mutations, deletions, duplications and other changes, CYP families exhibit considerable interspecies and interindividual variation. For example, about 40 % of the human CYP genes have been found to be polymorphic (Ingelman-Sundberg et al. 1999).

CYP enzymes metabolise lipid-soluble endogenous and exogenous compounds into metabolites that are usually inactive (detoxification), but active products (metabolic activation) are formed in some cases. A majority of pharmaceutical agents (drugs) are also metabolised by CYP enzymes. This has opened up a completely new area of research, pharmacogenetics, which focuses on the occurrence, function and consequences of polymorphic XMEs as well as genetically based interindividual variations in pharmacokinetics and pharmacodynamics of drugs (Boddy & Ratain 1997, Kalow et al. 1999a, Kalow et al. 1999b). For example, the most thoroughly studied polymorphic CYP enzyme, CYP2D6 (originally measured by debrisoquine hydroxylation assay), exhibits a number of phenotypes based on polymorphic alleles. Based on these phenotypes individuals may be categorised as poor (PM), extensive (EM), and ultra-rapid metabolisers (UM). These variations are important for drug therapy, because CYP2D6 is involved in the metabolism of about 25 % of all pharmaceuticals (Meyer et al. 1990, Kalow 1997, Wong et al. 2000).

Because CYPs are important in the detoxification and metabolic activation of numerous foreign chemicals, it is reasonable to hypothesise that the polymorphisms of CYP enzymes are associated with various toxicities, such as chemical carcinogenesis.
(e.g., lung cancer) (Pelkonen & Raunio 1995). Lung cancer is known to be caused by tobacco smoke, which contains carcinogens activated by CYPs (e.g., PAHs and NNK) (Gonzalez 1997, Hecht 1998).

This study focused on CYP2A6, a CYP form that was recently shown to be an efficient metaboliser of key agents in tobacco smoke, especially nicotine. The studies presented here are based on earlier results concerning significant interindividual and interethnic differences in the metabolism of CYP2A6 substrates in *in vitro* and *in vivo* (Pelkonen *et al.* 1985, Rautio *et al.* 1992). These findings led to further research on the genetic background of CYP2A6 polymorphism (Hecht 1998) and its possible associations with selected chemically induced diseases.
2 Review of the literature

2.1 Xenobiotic metabolising enzymes (XMEs)

XMEs have been found in all organisms, including very primitive microorganisms and eubacteriums. About 400 million years ago, ancient XME families began to diversify at the time when organisms moved from water to the ground. At this time, animals began to use plants as diet, and plants in turn developed animal repelling toxins (Nebert & Gonzalez 1987, Nebert 1997a). The XME families in animals gradually diversified to adapt to these toxins. Subsequently, more specific XME genes emerged through selection pressure in different species. As a result of this continuous selection pressure and diversification, almost every compound is metabolised by one or more XMEs in higher organisms, such as humans (Gonzalez & Nebert 1990, Negishi et al. 1996). In humans, more than 20 XME families have been found, which is more than in any other species (Nelson et al. 1993, Nebert & Dieter 2000). The specialised XME families developed through various mutations, and nowadays interindividual and interethnic differences have also been observed (Nebert 1997a, Ingelman-Sundberg et al. 1999) (http://www.incyte.com/web site). During the last 400 million years, the CYP2 family has been duplicated several times (Nebert & McKinnon 1994, Nebert 1997a). The development of XMEs, especially the CYP2 family (Nelson et al. 1996a), is presented in Fig. 1.

Nowadays, the principal function of XMEs is to metabolise a large number of endogenous and exogenous compounds. These compounds vary from highly hydrophilic to highly lipophilic. XMEs are involved in the biotransformation of exogenous agents, such as solvents, procarcinogens and drugs (Guengerich & Shimada 1991, Pelkonen & Raunio 1997, Lewis 2000). 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, but some of the metabolites generated are more toxic than the parent compound (metabolic activation) (Nebert 1997a).

XMEs are categorised into two groups, i.e., phase I (functionalisation reactions) and phase II enzymes (conjugation reactions) (Nebert & Gonzalez 1987, Nebert et al. 1996, Nelson et al. 1996b).
Fig. 1. A schematic pedigree of the development of some XMEs, especially the CYP2 family (Gonzalez & Nebert 1990, Nebert & McKinnon 1994, Nelson et al. 1996a).
2.1.1 Endogenous and exogenous substrates

The substrates of XMEs include both endogenous and exogenous compounds. Endogenous compounds, such as steroids, participate in many different functions ranging from fertilisation to insulin production. Also other endogenous compounds existing in the body, including prostaglandins and bile acids, are metabolised by XMEs (Gonzalez & Nebert 1990, Gonzalez 1992, Wolff & Strecker 1992, Nelson et al. 1996a, Nelson et al. 1996b). Exogenous compounds are inhaled, ingested or diffused through the skin. Examples of exogenous agents are various toxins, several carcinogens and mutagens, such as radon in the indoor environment, and occupational agents, such as asbestos and polycyclic aromatic hydrocarbons (PAHs) in the ambient environment. The most important source of exogenous compounds is diet. Preparation of food modifies the compounds present in the diet. For instance, boiled meat contains less carcinogenic compounds than fried meat (Vineis 1997). Recreational substances, such as tobacco and alcoholic beverages, are known to contain large amounts of toxic compounds. For example, tobacco smoke contains PAHs, such as benzo(a)pyrene, nitrosamines and other toxicants, including nicotine, cresol, carbon monoxide, and pyridine (Guengerich & Shimada 1991, Wexler 1998).

Drugs, being exogenous compounds, are inherently active or metabolically activated by XMEs (Ingelman-Sundberg et al. 1999). In the case of polymorphic XMEs, individuals with altered metabolic capacity such as those exhibiting PM and UM CYP2D6 phenotypes, may be an altered metabolism of drugs, which may cause serious adverse effects in the patients (Guengerich 1993, Kalow et al. 1999a, Kalow et al. 1999b).

2.1.2 Phase I enzymes

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 metabolize different substrates (Guengerich 1992a, Guengerich 1992b). CYP enzymes monooxidise or reduce and epoxide hydrolases (EHs) hydrolyse their substrates. Hydroxylases, flavin-containing monoxygenases (FMOs) and monoaminoxidases (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, Guengerich 1995a).

CYPs are a well characterized group of phase I enzymes. In endoplasmic reticulum (ER), a CYP enzyme together with appropriate cofactors, such as NADPH, metabolises chemicals by breaking up oxygen molecules (O2). The chemical to be metabolised receives a hydroxyl group (-OH), while the other oxygen atom is inserted into a water molecule (H2O). This reaction is called monoxygenation (Degtyarenko & Archakov 1993). The active centre of the CYP 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, Guengerich 1994, Guengerich 1995b).
In humans, CYP enzymes are divided into four main families (http://drnelson.utmem.edu/cytochromeP450.html/): CYP1, CYP2, CYP3 and CYP4A (Guengerich 1995b, Nelson et al. 1996b). CYP polymorphisms are hypothesised to have emerged and being propagated by the natural selection exercised by toxic compounds in the diet or the environment during evolution (Gonzalez 1992, Nelson et al. 1996a, Hong & Yang 1997, Pelkonen et al. 1998). This topic will be elaborated in more detail later (chapter 2.3).

2.1.3 Phase II enzymes

Phase II enzymes usually detoxify, but sometimes also activate, endogenous or exogenous compounds. The substrate is converted into a 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 follows: 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 instance, 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, Saarikoski et al. 1998, Weinberg & Sandler 1999, Bennett et al. 1999). GSTs metabolise aflatoxin B1, PAHs, and some other compounds present in tobacco smoke. Synergistic effects of the combined GSTM1-1 and GSTT1-1 null genotypes may confer a highly increased risk to lung cancer, but also the combination of GSTM1, GSTT1 and GSTP1 alleles has been found to confer a synergistic risk to lung and breast cancer (Nebert 1997b, Wormhoudt et al. 1999).

In general, NATs acetylate a wide variety of arylamines and tobacco-derived aromatic amines. There are two polymorphic enzymes, NAT1 and NAT2 (Payton & Sim 1998, Wormhoudt et al. 1999) (http://louisville.edu/medschool/pharmacology/NAT.html/). Controversial results of NAT1 and NAT 2 role in lung cancer have been observed (Bouchardy et al. 1998, Hein et al. 2000a, Hein et al. 2000b).

2.1.4 Regulation of XMEs

Regulation of xenobiotic metabolism is highly complex. The large number of factors affecting the expression and activity of xenobiotic metabolising enzymes are divided into three major groups, i.e., genetic factors, non-genetic host factors (such as diseases, age, stress, obesity, physical exercise, etc.) and environmental factors (environmental pollutants, occupational chemicals, drugs, etc.) (Vesell 1980, Pelkonen 1992, Pelkonen et al. 1998). It is clear that in most instances the activity of an XME is determined by the interplay between these factors. The extent to which any one factor affects the activity depends on the specific XME, the specific individual and the specific factor. For example, the activity of the mouse CYP2A5 (analogous to human CYP2A6) is induced
by a large number of compounds (such as phenobarbital, pyrazole, cobalt) and biological factors (liver infestations, tumours) and inhibited by numerous substances (such as methoxsalen and many furanocoumarins), and the activity and inducibility depend on mouse strain. CYP2A5 has been found to be also regulated by cAMP and protein kinase A-mediated systems (Raunio et al. 1988a, Salonpää et al. 1994, Posti et al. 1999, Viitala et al. 2000). This complex regulation points to an endogenous role of the CYP2A5 enzyme. As another example, the regulation of CYP1A2 is controlled by the AhR. The cytosolic receptor, AhR, binds to a xenobiotic (for example TCDD) and regulates CYP1A2 activity by binding to the xenobiotic responsive element (XRE) of the receptor. Both CYP1A2 and AhR have been found to be polymorphic (Hankinson 1995, FitzGerald et al. 1996, Micka et al. 1997, Nebert et al. 2000).

### 2.2 Pharmacogenetics of XMEs

In the early 20th century, the variation of an inherited deficient gene, an inability to taste phenylthiourea, was primarily discovered (Snyder 1932). The next observation concerning variation in drug responses was made in Italy during World War II. Especially black soldiers were found to suffer from adverse effects caused by the antimalarial drug primaquine, due to a deficiency in glucose-6-phosphate dehydrogenase enzyme (Carson et al. 1956). The concept of pharmacogenetics was later defined by Friedrich Vogel as the study of the role of genetics in drug response (Vogel 1959, Nebert 1997b). Later on, various definitions have been given to pharmacogenetics, including the following: the effect of polymorphic hereditary factors on the metabolism of drugs (Park et al. 1996, Hasler 1999).

#### 2.2.1 Pharmacokinetics and metabolism of drugs

To enter their target sites, drugs (as well as other xenobiotics) have to be absorbed from the gut (ingestion), through skin and the airways (pollutants, smoking, environmental chemicals). After absorption, xenobiotics are distributed into systemic circulation (Park et al. 1996), bind to blood elements (e.g., albumin), are metabolised by XMEs mostly in liver cells (but also in intestine and other extrahepatic tissues) (Scherer et al. 1988), and are finally excreted through kidneys into urine or through liver into bile (Park et al. 1995, Nebert 1997b, Pelkonen & Ruskoaho 1998, Kalow et al. 1999). At all these different steps, transport proteins and pumps, carriers and enzymes participate in the process, and it is possible that all these macromolecules demonstrate pharmacogenetic variability (Pelkonen & Ruskoaho 1998). This variability has been most convincingly shown for XMEs.
2.2.2 Variation in drug metabolism

An ever-increasing number of pharmacogenetic alterations in XMEs have been detected since the discovery of polymorphic acetylation (Nebert 1994, Nebert 2000). Probably the most seminal finding was the discovery of defective debrisoquine 4'-hydroxylation, later shown to be due to polymorphisms of CYP2D6, which is responsible for the metabolism of at least three dozen drugs (Daly et al. 1996). CYP2D6 polymorphism has been found to have practical consequences, such as adverse effects or therapeutic resistance, in individuals with defective or multiduplicated genetic backgrounds (PM and UM) (Meyer & Zanger 1997, Kalow et al. 1999b). Generally, UM, EM and IM are functional phenotypes, and EM individuals metabolise normally, whereas IM has approximately half of the activity of EM (Guengerich 1994, Nebert et al. 1999). However, as is practically always the rule in pharmacogenetic conditions, the correspondence between the genotype and the phenotype is rarely perfect, and there is a lot of phenotypic variability within each genotype category (Pelkonen et al. 1999, Nebert 1999a, Kalow et al. 1999b).

Another noticeable finding in pharmacogenetics is interethnic variability. As an example, the frequencies of variant (defective) phenotypes in several populations of some representative XMEs, CYP2C19, CYP2D6, GSTM1, GSTT1 and NAT, are presented in Table 1.

Table 1. Frequencies of defective phenotypes of CYP2C19, CYP2D6, GSTM1, GSTT1 and NAT in several populations.

<table>
<thead>
<tr>
<th>Population</th>
<th>CYP2C19 PM</th>
<th>CYP2D6 PM</th>
<th>GSTM1 null</th>
<th>GSTT1 null</th>
<th>NAT Slow</th>
</tr>
</thead>
<tbody>
<tr>
<td>European</td>
<td>3.5</td>
<td>7.4</td>
<td>48.2</td>
<td>10.0</td>
<td>59.0</td>
</tr>
<tr>
<td>White Americans</td>
<td>2.6</td>
<td>7.7</td>
<td>52.0</td>
<td>14.7</td>
<td>55.0</td>
</tr>
<tr>
<td>Black Americans</td>
<td>–</td>
<td>1.9</td>
<td>–</td>
<td>24.1</td>
<td>–</td>
</tr>
<tr>
<td>African-Americans</td>
<td>–</td>
<td>–</td>
<td>27.0</td>
<td>21.8</td>
<td>41.0</td>
</tr>
<tr>
<td>Japanese</td>
<td>22.5</td>
<td>0.5</td>
<td>48.6</td>
<td>–</td>
<td>8.0</td>
</tr>
<tr>
<td>Chinese</td>
<td>17.4</td>
<td>0.95</td>
<td>–</td>
<td>61.2</td>
<td>20.0</td>
</tr>
<tr>
<td>Korean</td>
<td>12.6</td>
<td>–</td>
<td>–</td>
<td>60.2</td>
<td>–</td>
</tr>
</tbody>
</table>

Data adapted from (Wormhoudt et al. 1999).

Depending on the correspondence between the genotype and the phenotype, determination of the genotype can predict the phenotype, if the main defective alleles are known. High frequencies of variant or deleted alleles, resulting in significant interindividual differences in the activities of XMEs, cause huge variability in the metabolism of substrates and are valuable information in the planning of drug therapy (Gonzalez & Idle J.R. 1994, Nebert & Carvan, III 1997). (Pelkonen et al. 1999) Furthermore, because of the importance of metabolic factors in the effects and kinetics of drugs and other xenobiotics, certain phenotypes have been proposed to be risk factors for some diseases (Guengerich & Shimada 1991, Connor et al. 1993, Guengerich 1994, Daly et al. 1996, Nebert 1999b).
Many adverse effects and therapeutic failures are due to deviant metabolism by CYPs and other XMEs, which decrease the therapeutic efficiency of drugs (Kalow et al. 1999a, Nebert 1999a). Adverse effects associated with the use of a particular drug metabolised by a polymorphic enzyme can be avoided if the phenotype (or the genotype) of the individual is known. For drug industry, specific information about the XMEs participating in the metabolism of a particular compound at a very early stage is useful for the development of drug candidates (Guengerich 1994, Nebert et al. 1999), because later decisions will be increasingly time- and money-consuming (Linder et al. 1997, Vermes et al. 1997).

The phase I enzymes CYP2C9, CYP2C19, CYP2D6 and CYP3A4 are significant in the metabolism of clinically used drugs (Puga et al. 1997). In clinical practice, i.e., drug treatment of individuals, CYP phenotypes are relevant knowledge, e.g., whether an individual is PM or UM with respect to CYP2D6 activity (Bertilsson et al. 1992, Meyer & Zanger 1997, Linder et al. 1997, Wong et al. 2000). A register of all individual is polymorphic genotypes would be helpful for correct administration of drugs, for choosing the right dosage, and for avoiding possible adverse effects (Mahgoub et al. 1977, Bertilsson et al. 1992, Idle & Smith 1995, Ingelman-Sundberg et al. 1999, Smith 2000).

2.3 Polymorphic CYPs

According to the current definition, if the frequency of a variant allele is higher than one percent, it is called a genetic polymorphism (Pelkonen et al. 1999). Polymorphisms in genes have developed during the evolution through various events, such as point mutations (missense, nonsense or frameshift), gene conversions, deletions, and insertions (Nebert 1997a). Numerous SNPs are also found throughout the genome (Cargill et al. 1999, Nebert 2000).

Polymorphisms were first observed at the phenotype level. Variations in enzyme activities or in the rate of elimination were observed in, for example, liver microsomes in vitro or in animals and humans in vivo. In humans, certain polymorphisms of XMEs were found by observing the impaired metabolism of certain drugs or the failure of drug treatment (Linder et al. 1997). The enzymes presented below are polymorphic at the phenotype and genotype levels (Linder et al. 1997, Wormhoudt et al. 1999, Nebert 2000). Many XME genes are presented at the web site (http://www.gene.ucl.ac.uk/nomenclature/).

Many carcinogens are metabolised by CYP1A1 and CYP1A2 enzymes. These substrates are mainly planar aromatic hydrocarbons, for example, benzo[a]pyrene, dimethylbenz[a]anthracene and 6-nitrochrysene. Four polymorphic alleles of the CYP1A1 gene have been found: CYP1A1*2A, CYP1A1*2B, CYP1A1*3, and CYP1A1*4. The CYP1A1*2A and CYP1A1*2B alleles are in linkage disequilibrium. The frequencies of these variant alleles in different populations are quite variable: 2.7–5.1 % in Caucasians, 10–15 % in Japanese, 13.6–25.5 % in Africans, and 2.7–22 % in African-Americans. Some variant CYP1A1 genotypes have been found to increase the risk of lung cancer,
especially when accompanied by the synergistic effect of mutated p53, Ki-ras or GSTM1 (Wormhoudt et al. 1999, Hirvonen 1999). Polycyclic aromatic hydrocarbons (PAHs) are also major substrates of CYP1A2. Caffeine is another important substrate, and the demethylation of caffeine is used as a marker for CYP1A2 activity in vivo (Tang & Kalow 1996). CYP1A2 is inducible by cigarette smoking (Zevin & Benowitz 1999) and omeprazole, and it is regulated at least partly by the AhR-Arnt system (Nebert et al. 1991, Swanson & Bradfield 1993). Although phenotyping studies have demonstrated the presence of interindividual variation, the possible genetic basis has not been elucidated yet (Omiecinski et al. 1999).

The CYP2A6 enzyme will be introduced later (chapter 2.4).

The CYP2C subfamily contains three polymorphic enzymes: CYP2C8, CYP2C9 and CYP2C19. CYP2C9 is involved in the metabolism of S-warfarin, phenytoin, tolbutamide and tienilic acid, but some non-steroid anti-inflammatory drugs (e.g., diclofenac, piroxicam, tenoxicam, ibuprofen and acetylsalicylic acid) are also substrates for this enzyme. Interindividual variability in the metabolism of both tolbutamide and phenytoin have been observed. Two variant alleles (CYP2C9*2 and CYP2C9*3) have been reported. The frequency of the defective genotypes is 6–19.2 % in Caucasians, 0–2.6 % in Orientals and 0.5–1 % in African-Americans (Wormhoudt et al. 1999, Hirvonen 1999).

CYP2C19, also known as mephenytoin 4-hydroxylase, metabolises several drugs (S-mephenytoin, omeprazole, diazepam, imipramine, propranolol, progynon, mephobarbital and hexobarbital) (Meyer & Zanger 1997). Two major variant alleles (CYP2C19*2, CYP2C19*3, CYP2A6*4) and also clear-cut phenotypes (PM and EM) of CYP2C19 have been observed (Garcia-Barcelo et al. 1999, Itoh et al. 2000). In addition to interindividual variation, interethnic variation also exists. The frequency of the CYP2C19 PM phenotype is 2.5–6 % in Caucasians, 18–23 % in Japanese and 15–17 % in Chinese (Meyer & Zanger 1997, Ingelman-Sundberg et al. 1999, Wormhoudt et al. 1999). CYP2C19 PM with HCV-seropositive was found to associate with a high risk for developing hepatocellular carcinoma (HCC) (Chau et al. 2000).

About 25 years ago the impaired metabolism of debrisoquine and sparteine was discovered (Mahgoub et al. 1977, Eichelbaum et al. 1979, Cholerton et al. 1992a). After administration of debrisoquine, idiosyncratic drug reactions were observed in some individuals (Nebert 1997a). Several years later, polymorphic alleles of the CYP2D6 gene were demonstrated to underlie this phenomenon (Gonzalez et al. 1988). The CYP2D6 enzyme metabolises at least 40 drugs (Cholerton et al. 1992a). Debrisoquine, sparteine and dextromethorphan have been used as probe drugs for CYP2D6 (Meyer & Zanger 1997). Because of the importance of this enzyme in the elimination of a number of important drugs, routine genotyping or phenotyping of patients beginning long-term drug treatment with certain CYP2D6 substrate drugs has been recommended (Pelkonen & Saarni 1980, Wormhoudt et al. 1999). Currently, the number of known variant CYP2D6 alleles is close to 50, but less than 10 of these are frequent enough in the population to be designated as polymorphic. According to the nomenclature of CYP2D6 (Daly et al. 1996), the most common alleles associated with the PM phenotype are CYP2D6*3, CYP2D6*4 and CYP2D6*5 (Daly et al. 1996, Wormhoudt et al. 1999). An interesting feature of CYP2D6 is that the gene is sometimes duplicated or multiduplicated, resulting in an ultrarapid metaboliser phenotype (UM) (Daly et al. 1991, Kroemer & Eichelbaum 1995). Interethnic variation in the frequencies of these polymorphic alleles is
considerable. For example, the frequency of the PM phenotypes is 5–10 % in Caucasians and 1–2 % in Orientals (Meyer & Zanger 1997, Nebert 1997b, Ingelman-Sundberg et al. 1999). Although very few carcinogens or toxicants have been found to be substrates of CYP2D6 (Chen et al. 1996), special interest has been focused on its association with lung cancer since the observation that PM individuals are underrepresented among lung cancer patients (Ayesh et al. 1984, Bertilsson et al. 1992), suggesting that the lack of CYP2D6 protects against contracting lung cancer. Later, extensive studies have yielded contradictory results (Boustead et al. 1997, Rostami-Hodjegan et al. 1998). Recently, the relationship between CYP2D6 polymorphisms and tobacco smoking habits have also been studied: the defective phenotype may decrease probability of smoking, whereas individuals with UM phenotype may be more prone to develop addiction to smoking (Boustead et al. 1997, Christensen et al. 1997, Lerman et al. 1998, Saarikoski et al. 2000). Individuals with CYP2D6 PM phenotypes have also been suggested to be protected against developing addiction to opiates, such as codeine (Tyndale et al. 1997, Romach et al. 2000).

2.4 CYP2A subfamily

The genes in CYP2A, CYP2B and CYP2F subfamilies are located in chromosome 19p13.2 in a region spanning 350 kb (Fernandez-Salguero & Gonzalez 1995, Hoffman et al. 1995). These three subfamilies are thought to descend from the same ancestral gene. They may have differentiated during evolution to specialise in metabolism of certain preferred substrates. Beginning about 400 million years ago, the CYP2 family has developed into more than 50 genes, probably due to environmental pressure (i.e., natural selection) (Nelson et al. 1993).

The human CYP2A subfamily consists of three genes and two pseudogenes: CYP2A6, CYP2A7, CYP2A13, CYP2A7P(T) and CYP2A7P(C) (Fernandez-Salguero et al. 1995, Honkakoski & Negishi 1997). The CYP2A7 gene is located next to the CYP2A6 gene. The CYP2A6 and CYP2A7 genes have a 96 % similarity in the nucleotide sequence and a 94 % identity at the amino acid level (Miles et al. 1989, Yamano et al. 1990).

The structure of CYP2A genes consists of 9 exons and 8 introns, and the size of the whole gene is about 6 kb. The sequence of the pseudogenes (the CYP2A7P(T) and CYP2A7P(C)) breaks up in exon 5, and they therefore do not encode any protein (Fernandez-Salguero et al. 1995). The CYP2A6 gene encodes a functional protein (Raunio et al. 1988b, Miles et al. 1990, Fentem & Fry 1993, Raunio et al. 1999a), but the CYP2A7 gene (and the aberrantly spliced CYP2A7AS) produces an inactive protein (Yamano et al. 1990, Ding et al. 1995). The CYP2A13 protein has been found to produce an active protein (Su et al. 2000). The structure of the CYP2A-CYP2B-CYP2F gene cluster is presented in Fig. 2.
2.4.1 Expression of CYP2A members

The CYP2A6 gene is expressed at a high level in liver and at lower levels in nasal mucosa and respiratory tract. CYP2A6 mRNA is also found in some cell types in lungs (Fernandez-Salguero et al. 1995, Koskela et al. 1999). CYP2A7 mRNA is expressed in liver at approximately the same levels as CYP2A6 (Ding et al. 1995). CYP2A13 and two pseudogenes (CYP2A7(C) and CYP2A7(T)) are expressed in liver. CYP2A13 is also expressed in nasal mucosa at a relatively high level (Yamano et al. 1990, Miles et al. 1990, Pearce et al. 1992) and there are some indications of expression in the respiratory tract (Koskela et al. 1999, Raunio et al. 1999b). Recent studies suggest that CYP2A13 mRNA is present in lungs at a higher level than CYP2A6 mRNA (Hong et al. 2000, Su et al. 2000).

The expression of CYP2A enzymes in various tissues is presented in Table 2.

Table 2. Enzyme activities and expression of CYP2A subfamily.

<table>
<thead>
<tr>
<th>Alleles/genes</th>
<th>Expression tissue (mRNA)</th>
<th>Expression (protein)</th>
<th>Enzyme activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2A6*1</td>
<td>liver (+++) nasal mucosa (+)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>–</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CYP2A6*3</td>
<td>–</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>CYP2A7</td>
<td>liver (+++)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CYP2A7P(C)</td>
<td>liver (+++)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CYP2A7P(T)</td>
<td>liver (+++)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CYP2A13</td>
<td>nasal mucosa (+++)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>respiratory tract (+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>liver (+)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data adapted from (Koskela et al. 1999, Su et al. 2000).
CYP2A6 enzyme activities have been measured in vitro in liver microsomes, and an in vivo test is also available (Rautio et al. 1992, Cholerton et al. 1992b, Satarug et al. 1996, Rautio et al. 1998). Considerable interindividual differences have been observed in these studies (Iscan et al. 1994), indicating a polymorphic pattern (Pelkonen & Raunio 1995, Honkakoski & Negishi 1997).

2.4.2 Substrates

The CYP2A6 enzyme metabolises substrates that can be categorised into toxic (procarcinogens, promutagens and other toxins) and pharmaceutical agents (drugs). The substrates are usually detoxicated, but some are also activated. Most substrates of CYP2A6 have been found in in vitro studies of liver microsomes (Pelkonen & Raunio 1995). Such drugs as coumarin, halothane, losigamone, letrozole and SM-12502 are substrates of CYP2A6 (Pearce et al. 1992, Nunoya et al. 1996, Torchin et al. 1996, Wirz et al. 1996, Spracklin & Kharasch 1998, Kharasch et al. 2000a). The metabolism of the procarcinogen 1,3-butadiene and the strong liver toxin aflatoxin B1 (AFB1) is also mediated by CYP2A6 (Aoyama et al. 1990, Duescher & Elfarra 1994, Elfarra et al. 1996, Krause & Elfarra 1997, Sadeque et al. 1997). AFB1 is thought to be a major cause of the initiation of liver cancer and cirrhosis.

Some well-known toxins and procarcinogens are known to exist in tobacco smoke: nicotine and nicotine-derived nitrosamines, such as NNK, NNAL, NDEA and NDMA (Hoffmann & Hecht 1985, Hoffmann et al. 1985a, Crespi et al. 1990, Yamazaki et al. 1992, Camus et al. 1993, Tiano et al. 1994, Smith et al. 1995, Hoffmann et al. 1996, Patten et al. 1997). NNK, a causative agent for esophageal cancer, requires metabolic activation to exert its carcinogenic potential (Hecht 1996a). CYP2A6 mediates the 5'-hydroxylation and CYP3A4 catalyses the 2'-hydroxylation of NNK (Staretz et al. 1997). The potent procarcinogens NNK and NNN are thought to initiate lung cancer (Hoffmann et al. 1985b, Hecht & Hoffmann 1988, Gonzalez & Gelboin 1994, Hecht 1996b). Molecular modelling has been used to characterise the activity and substrate selectivity of CYP2A6, but the specific mechanisms are not yet known (Lewis et al. 1999). CYP2A6 has also been found to metabolise gasoline additives (methyl tert-butyl ether (MTBE), ethyl tert-butyl ether (ETME) and tert-amyl methyl ether (TAME)) in liver microsomes (Hong et al. 1999). Methoxsalen has been found to be an efficient inhibitor of CYP2A6 (Sellers et al. 2000, Kharasch et al. 2000b).

The most relevant pharmaceuticals and toxic substrates of CYP2A6 are summarised in Table 3. Some other compounds are less specific substrates, or the kinetics of metabolism have not been worked out in detail (Rautio et al. 1992, Nunoya et al. 1996). For instance, the oxidation of halothane is catalysed by CYP2E1 and CYP2A6 as the low and high Km enzymes (no role for CYP3A4). On the contrary, CYP2A6 and CYP3A4 mediate the reduction of halothane as the low and high Km enzymes (no role for CYP2E1) (Spracklin et al. 1996). Thus, the enzyme specificity of halothane metabolism is dependent on many factors, especially the substrate concentration and redox conditions at the site of metabolism (Spracklin et al. 1996).
Table 3. Some pharmaceutical and toxic substrates of CYP2A6.

<table>
<thead>
<tr>
<th>Pharmaceutical agents</th>
<th>Reaction/end-point</th>
<th>Toxic agents</th>
<th>Reaction/end-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coumarin</td>
<td>7-hydroxylation</td>
<td>Nicotine</td>
<td>N-1’-oxidation</td>
</tr>
<tr>
<td>Methoxyflurane</td>
<td>Dehalogenation</td>
<td>Cotinine</td>
<td>3’-hydroxylation</td>
</tr>
<tr>
<td>Halothane</td>
<td>Reduction</td>
<td>NNK</td>
<td>Mutagenicity</td>
</tr>
<tr>
<td>SM-12502</td>
<td>S-oxidation</td>
<td>NNN</td>
<td>Mutagenicity</td>
</tr>
<tr>
<td>Losigamone</td>
<td>Oxidation</td>
<td>NDEA</td>
<td>Mutagenicity</td>
</tr>
<tr>
<td>Valproic acid</td>
<td>Oxidation</td>
<td>AFB1</td>
<td>Mutagenicity</td>
</tr>
<tr>
<td>Letrozole</td>
<td>Oxidation</td>
<td>MOCA</td>
<td>N-oxidation</td>
</tr>
<tr>
<td>Disulfiram</td>
<td>Sulfoxidation</td>
<td>1,3-butadiene</td>
<td>Monoxide formation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quinoline</td>
<td>1-oxidation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCBN</td>
<td>Protein adduct formation</td>
</tr>
</tbody>
</table>

SM-12502, 3,5-dimethyl-2-(3-pyridyl)thiazolidin-4-one hydrochloride; NNK, 4-(methyl(nitrosamino)-1-(3-pyridyl)-1-butane; NNN, N’-nitrosonornicotine; NDEA, N-nitrosodiethylamine; AFB1, aflatoxin B1; MOCA, 4,4’-methylene-bis(2-chloroaniline); DCBN, 2,6-dichlorobenzonitrile.

Data adapted from (Pelkonen et al. 2000).

2.4.2.1 Coumarin

Coumarin is a simple plant alkaloid, and it has been used as a sweetener, fixative, stabiliser and previously also as a food additive. It is also present in certain tobacco products, alcoholic beverages, and various kinds of soaps, detergents and cosmetic preparations (Egan et al. 1990, Pelkonen et al. 1993). Coumarin-containing preparations have been used as natural medicines in various countries, especially in Germany, it is thought to possess anti-carcinogenic properties. For example, the combination of coumarin and cimetidine has been used for the treatment of melanoma and renal cell carcinoma (Cholerton et al. 1992b, Runkel et al. 1996, Lake 1999). Coumarin has also been used in the treatment of diseases of venous and lymphatic vessels (inflammation and oedema, cerebrovascular insufficiencies).

Coumarin is the main probe drug for CYP2A6 (Rautio et al. 1992, Pelkonen et al. 2000). It is fully specific and rapidly metabolised by CYP2A6; on an average, about 70–80% is converted into 7-hydroxycoumarin within two hours. The half-life of coumarin is one hour. CYP2A6 was observed to catalyse the metabolism of coumarin first in in vitro in experiments with liver microsomes (Raunio et al. 1988b, Yun et al. 1991). If the principal metabolic pathway is blocked, coumarin is metabolised mainly into 3-hydroxycoumarin. In rodents this pathway is more common and other metabolites presented in Figure 3 are produced at very minor levels in humans (Nakajima et al. 1996a, Hadidi et al. 1997, Hadidi et al. 1998). Coumarin 7-hydroxylation activities display notably wide interindividual and interethnic variability when measured in human liver microsomes (Pelkonen et al. 1985, Yamano et al. 1990, Pearce et al. 1992).
phenotyping method based on measuring 7-hydroxycoumarin from urine is presented in more detail in chapter 4.3.1. (Rautio et al. 1992, Cholerton et al. 1992b). The various routes of coumarin metabolism are illustrated in Fig. 3.

![Diagram of coumarin metabolism](image)

**Fig. 3. Routes of coumarin metabolism.** Note that only 7-hydroxylation is catalysed by CYP2A6. Enzymes catalysing the other routes have not been identified (Yamano et al. 1990, Yun et al. 1991).

### 2.4.2.2 Nicotine

Tobacco products contain nicotine, which is absorbed via oral mucosa, respiratory tract, gastrointestinal tract, and skin. 80–90 % of absorbed nicotine is metabolised in liver at the first pass (Wexler 1998). Nicotine has well-known pharmacological effects, such as acceleration of heart rate and elevation of blood pressure. The long-term toxic effects of nicotine are partially dependent on its ability to cause addiction in regular use; nicotine is actually comparable to cocaine in its ability to cause addiction (Wexler 1998).

Early studies identified three CYPs, CYP2A6, CYP2B6 and CYP2D6, capable of metabolising nicotine (Flammang et al. 1992). In later *in vitro* studies with human liver microsomes, nicotine was found to be a specific substrate of CYP2A6; this finding was later confirmed *in vivo* (Cashman et al. 1992, Berkman et al. 1995, Nakajima et al. 1996a, Nakajima et al. 1996b). The elimination half-life for nicotine is 2 hours in blood in adults (Benowitz 1996, Zevin et al. 1997). Nicotine is C-oxidised into cotinine, and
cotinine is further metabolised into 3-or 7-hydroxycotinine mainly by the high-affinity metaboliser, CYP2A6 (Scherer et al. 1988, Benowitz et al. 1995, Nakajima et al. 1996a). Some evidence of methoxsalen as an inhibitor of CYP2A6 has been found in the first-pass metabolism by administrating orally nicotine at the same time (Sellers et al. 2000). Some interethnic variation in nicotine C-oxidation activity was also observed (Inoue et al. 2000). The conversion of nicotine into cotinine is a two-step reaction, in which nicotine is first oxidised into nicotine iminium ion and subsequently into cotinine by cytosolic aldehyde oxidase. The CYP2A6-mediated formation of nicotine iminium ion from nicotine is the rate-limiting step in this reaction, and up to 80% of nicotine is metabolised into cotinine (McCracken et al. 1992). Cotinine is further converted (Berkman et al. 1995, Nakajima et al. 1996b, Messina et al. 1997) into several metabolites (Nebert 1997b, Inoue et al. 2000). CYP2A6 appears to be the sole catalyst of nicotine catabolism at low nicotine concentrations (50 μM). At high concentrations (500 μM), other CYP forms, such as CYP2B6, also participate in nicotine oxidation (McCracken et al. 1992, Cashman et al. 1992, Berkman et al. 1995, Nakajima et al. 1996a, Nakajima et al. 1996b, Messina et al. 1997, Yamazaki et al. 1999, Murphy et al. 1999).

The CYP2A6-mediated metabolic route of nicotine is presented in Fig. 4.

![Fig. 4. Principal metabolic routes of nicotine mediated by CYP2A6 (Berkman et al. 1995, Messina et al. 1997).]
2.4.3 Genetic regulation and polymorphisms of CYP2A6

In vitro studies with human liver microsomes and in vivo phenotyping studies have displayed marked interindividual variation in CYP2A6 activity (Rautio et al. 1992, Pelkonen & Raunio 1995). In early phenotyping studies of CYP2A6, coumarin was used as a specific probe drug (Rautio et al. 1992, Cholerton et al. 1992b, Shimada et al. 1996). Subsequently, nicotine was found to be a good probe drug similarly to SM-12502 (Nunoya et al. 1996, Nakajima et al. 1996a, Nakajima et al. 1996b, Nunoya et al. 1999a, Nakajima et al. 2000). These agents have been used as probe drugs in phenotyping CYP2A6 activity, and revealed interindividual and interethnic variations (Hadidi et al. 1998). For instance, SM-12502 has been found in vitro (Nunoya et al. 1996) and in vivo (Nunoya et al. 1998, Nunoya et al. 1999b) as a good substrate for CYP2A6.

At the genomic level, a polymorphism was first observed about 10 years ago, when the wild-type (CYP2A6*1), variant 1 (CYP2A6*2) and variant 2 (CYP2A6*3) alleles were published (Fernandez-Salguero et al. 1995). The CYP2A6*2 allele has one point mutation in exon 3 (L160H). CYP2A6*3 has many mutations in three exons (3, 6 and 8), and this was suggested to be due to crossing-over events with CYP2A7 (Ding et al. 1995). Whole-gene deletion alleles (CYP2A6*4B and CYP2A6*4C) exist in the Japanese population, resulting in defective phenotypes (Nunoya et al. 1998, Nunoya et al. 1999b). The CYP2A6*4C (E-type) allele is suggested to have arisen due to an unequal crossing-over event with CYP2A7 (Nunoya et al. 1999a). The structures of the CYP2A6 alleles are presented in Fig. 5.

![Fig. 5. The structures of CYP2A6 alleles according to Fernandez-Salguero et al. 1995 (Fernandez-Salguero et al. 1995): CYP2A6*1 (wild-type allele), CYP2A6*2 (variant 1) has one point mutation in exon 3 codon 160, and CYP2A6*3 (variant 2) has several mutations in exons 3, 6 and 8.](image)

According to the initial studies, these variant alleles appeared to be more frequent in the Asian and Caucasian populations than in the African-American population (Fernandez-Salguero et al. 1995). The allele frequencies and the interethnic variation of CYP2A6 are presented in Table 4.
Table 4. The allele frequencies and the interethnic variation of CYP2A6.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Caucasian (%; n = 42)</th>
<th>Asian (%; n = 109)</th>
<th>African-American (%; n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2A6*1</td>
<td>80.5</td>
<td>67.5</td>
<td>97.5</td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>16</td>
<td>15.5</td>
<td>0</td>
</tr>
<tr>
<td>CYP2A6*3</td>
<td>3.5</td>
<td>17</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Data adapted from (Fernandez-Salguero et al. 1995).

Recently, correlation with other genes has been suggested to play a role in the interindividual difference in CYP2A6 activity. Nowak et al. reported co-occurrence of mutant CYP2A6 and CYP2C19 alleles in Canadian Native Indians, suggesting an association between these genes (Nowak et al. 1998, Lam et al. 1999).

2.5 Chemical-induced diseases and genetic factors

Toxic chemicals are active as such or are activated into reactive metabolites by XMEs (Pelkonen 1992, Raunio & Pelkonen 1995, Kitagawa et al. 1999). Toxicity affects target tissues by causing damage to cells, especially the nucleus. The damage caused can be reversible or irreversible.

Prevention, early diagnosis and intervention are key determinants in disease management (Nebert et al. 1996). Procarcinogen activators (including many XMEs) are targets of research elucidating risk factors in carcinogenesis initiation (Vermes et al. 1997, Nebert 1997a, Nebert et al. 1999, Pelkonen et al. 1999). The basic association between metabolism of toxic substances and various toxic outcomes is presented in Fig. 6.

Fig. 6. Association between metabolism and toxicity of chemical compounds (Pelkonen & Ruskoaho 1998).
2.5.1 Role of genetic factors in chemical toxicity and carcinogenesis

Individuals are known to differ in their susceptibility to cancer and birth defects. Genetic polymorphisms are often the reason for this phenomenon (Vineis 1997), including mutations in XME genes, oncogenes and tumour suppressor genes (p53, k-ras) (Vainio & Husgafvel-Pursiainen 1996, Spivack et al. 1997). Variability in the genome occurs at the levels of genes (mutation/deletions/insertions/DNA adducts) or chromosomes (SCE, sister chromatid exchanges). Some of these changes may be associated with interindividual differences in susceptibility to toxicity and the tumour initiation (Inada et al. 1995, Pelkonen & Ruskoaho 1998, Klaassen 1998).

CYPs, mEHs, GSTs, NATs and NQO1 are the main polymorphic XME groups found to be associated with certain diseases (Hirvonen 1999). For instance, the NAT2 slow acetylator status in phenotyping and genotyping studies has been found to be a significant risk factor for arylamine-induced bladder cancer (Hirvonen 1999). Likewise, GSTM1 and GSTT1 null genotypes have been found to increase the risk of lung and larynx cancer (Jourenkova et al. 1988, Saarikoski et al. 1998). A third example is the tumour suppressor gene p53, which is mutated by benzo(a)pyrene and may therefore affect lung cancer development (Pelkonen & Ruskoaho 1998, Klaassen 1998).

2.5.2 Mechanisms of chemical carcinogenesis

The process of carcinogenesis is divided into the following steps: metabolic activation of carcinogens by XMEs, DNA adduct formation, unsuccessful DNA damage recognition, unsuccessful DNA repair and failure of the immunological system to recognise tumour formation (Raunio et al. 1995, Niesink et al. 1996, Klaassen 1998, Hirvonen 1999). The initiation of cancer is followed by promotion and progression, leading to tumour formation (Harris 1991). After additional mutations, tumour cells turn invasive (malignant) (Vainio & Husgafvel-Pursiainen 1996). Mutations and other genetic changes may occur at many phases of the growing tumour cell, and many genes and proteins take part in these events (Perera 1996, Nebert 2000).

In chemical carcinogenesis, metabolic activation of procarcinogens by XMEs is often needed to initiate the process (Pelkonen 1992, Raunio & Pelkonen 1995). Subsequent pathways for damage could be the formation of DNA adducts, and mutations in genes encoding cell cycle or DNA repair systems (Anttila 1999). In the target tissue, active carcinogens may cause effects either in cytosol (proteins, mRNA or cytoskeleton) or, after being transported, in the nucleus (DNA). As a consequence, gross chromosomal changes (sister chromatid exchanges, chromosome aberrations) or mutations, deletions and insertions of DNA may occur (Nebert et al. 1999, Hirvonen 1999).
2.5.3 Liver cancer and cirrhosis

Liver cancer and cirrhosis are important diseases due to their high prevalence world-wide. The main causes of these diseases, especially cirrhosis, are alcohol consumption, hepatitis B virus (HBV), and aflatoxin B1 (Harris & Sun 1984, Montesano et al. 1997, Lieber 2000). Liver cancer sometimes stems from liver cirrhosis. The mechanisms of initiation of these diseases are not known completely, but many genetic (such as XMEs) and non-genetic factors (exogenous compounds present in the diet and the environment, such as toxins, carcinogens, and mutagens) are presumably required to be present at the same time (Niesink et al. 1996). The expression of most CYP forms is typically downregulated in some liver diseases, such as cirrhosis, fatty liver and alcoholic steatohepatitis (Wolff & Strecker 1992, Philip et al. 1994, Wilhelmsen 2000).

The role of CYP2A6 in hepatocellular diseases is intriguing, since CYP2A6 was found to be overexpressed in chronic hepatic inflammation and cirrhosis (Raunio et al. 1998). Concerning other CYPs, the functional CYP2D6 phenotype was observed to be a risk factor on liver cancer and cirrhosis formation in a case-control study in the Spanish population (Agúndez et al. 1995).

2.5.4 Lung cancer


Theoretically, metabolism of the procarcinogenic compounds in tobacco smoke by polymorphic XMEs may be associated with such consequences as lung cancer and smoking behaviour. Benzo(a)pyrene is known to cause mutations in the p53 gene and thus cause perturbations in cell cycle control. PAHs are thought to be responsible for the induction of CYP1A1, CYP1A2 and probably also CYP2E1 (Zevin & Benowitz 1999). Inducibility of CYP1A1 in lung cancer has been observed, and GSTM1 and GSTT1 null phenotypes act in a synergistic way in increasing the risk of lung cancer (Saarikoski et al. 1998).

The CYP2A6 enzyme metabolises tobacco specific nitrosamines, such as NNK, and it is therefore theoretically plausible to postulate a role for CYP2A6 in the initiation of lung cancer (Kamataki T. et al. 1999). The results of the case-control studies conducted thus far on the association between CYP2A6 and lung cancer have yielded controversial results. One reason for this is the genotyping methods used in these studies. In an early study on Americans by London et al., a 50 % reduction in the risk of lung cancer was observed with the defective phenotype (PM) (London S.J. et al. 1999). However, the genotyping in this study was done by the original method (Fernandez-Salgueiro et al.
1995, London et al. 1999). Next, a protective effect of the deleted allele (CYP2A6*4C) on lung cancer formation was observed in the Japanese population by Kamataki et al. (Kamataki et al. 1999) and Miyamoto et al. (Miyamoto et al. 1999). In a recent study on the French population by Loriot et al. no correlation between CYP2A6 status and lung cancer was found (Loriot et al. 2000). Similar results were obtained in a Chinese population with regard to lung and esophageal cancer (Tan et al. 2000).

Table 5 is a summary of all the case-control studies published thus far on the association between CYP2A6 and lung cancer.

**Table 5. Association between CYP2A6 and lung cancer.**

<table>
<thead>
<tr>
<th>Case-control Study</th>
<th>Population</th>
<th>Association with lung cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original method</td>
<td>London et al. 1999</td>
<td>American (US, CA)</td>
</tr>
<tr>
<td>New methods</td>
<td>Miyamoto et al. 1999</td>
<td>Japanese</td>
</tr>
<tr>
<td></td>
<td>Kamataki et al. 1999</td>
<td>Japanese</td>
</tr>
<tr>
<td></td>
<td>Loriot et al. 2000</td>
<td>French</td>
</tr>
<tr>
<td></td>
<td>Tan et al. 2000</td>
<td>Chinese</td>
</tr>
</tbody>
</table>


### 2.5.5 Tobacco smoking and genetic factors

Smoking behaviour is governed by the need to maintain desirable levels of nicotine in the body (Benowitz 1996, Benowitz et al. 1998). There is therefore considerable interest in the genetic factors affecting the metabolism of nicotine (Vastag 1998). Other factors have also received attention, such as heavy smokers inhaling less smoke from cigarettes than light smokers (Law et al. 1997) and transdermal nicotine suppressing the intake of tobacco smoke (Benowitz et al. 1998). In a study on smoking habits and its relation to tobacco-induced diseases, some interethnic variation was observed, as higher levels of nicotine per cigarette were inhaled by Blacks as compared with Whites. This was explained by the slower clearance of cotinine and the higher intake of nicotine in Blacks (Perez-Stable et al. 1998). Genetic factors, such as CYP2D6 polymorphism, have been suggested to play a role in modifying smoking behaviour. For example, a case-control study on the Finnish population showed that individuals with the CYP2D6 UM phenotype smoke more cigarettes than PM individuals (Boustead et al. 1997, Saarikoski et al. 2000).

The role of CYP2A6 as a modifier of smoking habits (Vastag 1998) is currently under intensive research due to its participation in nicotine clearance (Perez-Stable et al. 1998). In a Japanese study, individuals with the CYP2A6*4C/4C genotype were found to produce only 14% of the normal amount of cotinine after nicotine intake (Kitagawa et al. 1999, Nakajima et al. 2000).
Controversial results have been obtained on the association between CYP2A6 status and smoking behaviour. The first case-control study was done in a Canadian population by Pianezza et al., results showed that at least one variant allele of the CYP2A6 genotype reduced the risk to smoking (Pianezza et al. 1998). In the next study on Americans by London et al. (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. Sabol et al. (Sabol & Hamer 1999) and Tan et al. (Tan et al. 2000) found no correlation between the CYP2A6 genotype and cigarette smoking behaviour. Similar results were obtained in French and Finnish populations (Loriot et al. 2000, Tiihonen et al. 2000). The defective genotype was found also to reduce smoking (Rao et al. 2000). Recently, the total inhibition of CYP2A6 function was suggested to reduce smoking, and therefore might have a role in smoking cessation and tobacco exposure reduction (Sellers et al. 2000, Sellers & Tyndale 2000).

Table 6 is a summary of all the case-control studies published thus far on the association between CYP2A6 and tobacco smoking.

Table 6. Association between CYP2A6 and tobacco smoking.

<table>
<thead>
<tr>
<th>Case-control Study</th>
<th>Population</th>
<th>Association with tobacco smoking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Canadian</td>
</tr>
<tr>
<td>Pianezza et al. 1998</td>
<td>American (US, CA)</td>
<td>No</td>
</tr>
<tr>
<td>London et al. 1999</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Sabol et al. 1999</td>
<td>Mainly non-Hispanic</td>
<td>No</td>
</tr>
<tr>
<td>Loriot et al. 2000</td>
<td>French</td>
<td>No</td>
</tr>
<tr>
<td>Tan et al. 2000</td>
<td>Chinese</td>
<td>No</td>
</tr>
<tr>
<td>Rao et al. 2000</td>
<td>Caucasian</td>
<td>Yes</td>
</tr>
<tr>
<td>Tiihonen et al. 2000</td>
<td>Finnish</td>
<td>No</td>
</tr>
</tbody>
</table>

3 Aims of the study

This study was started by applying the published CYP2A6 genotyping method (Fernandez-Salguero et al. 1995) to a genetic epidemiological search for risk factors of environmentally caused human diseases. In the course of these studies, however, some inconsistencies in the original genotyping method were found and, consequently, the project expanded in the direction of eliminating these inconsistencies and finding new CYP2A6 alleles. Thus, relatively early on during this project, the aims were crystallised as follows:

1. **Methods development:**
   a) To identify and characterise CYP2A6 alleles and to develop genotyping methods for them.
   b) To find out CYP2A6 phenotype–genotype correlations.

2. **Population and case-control studies:**
   a) To investigate CYP2A6 allele frequencies in different populations.
   b) To genotype liver cancer and cirrhosis cases and controls with the aim of identifying the risk of certain genotypes in the initiation of these diseases.
   c) To find out the potential association between the CYP2A6 genotype and the risk of lung cancer.

Because of the reasons mentioned above, the Results section of this thesis contains a substantial amount of new, partly unpublished and partly revised information, which will also be reviewed in the Discussion section.
4 Materials and methods

4.1 Study populations

4.1.1 Index cases

Individuals with potentially variant CYP2A6 alleles were found based on their impaired phenotypes or anomalous genotypes. One such case belonged to the Spanish control population mentioned below (III, IV), while some others were found in the course of other genotyping studies on a different Spanish population (II, III), and a few cases were identified in a Finnish population (I). One Thai individual with a potential variant genotype was detected on the basis of a phenotyping study (III). These cases served as index cases for further studies on phenotype-genotype associations and for the development of genotyping assays.

4.1.2 Study populations

Seven Finnish volunteers from the southern part of Finland participating in a phenotype-genotype correlation study were selected from a previous population study done by the National Public Health Institute (Bardy et al. 1993). These individuals were smokers and non-smokers, and 5 of them were female and 2 male, age ranging from 32 to 66 years.

In the first three articles of this thesis (I–III), Caucasian subjects were analysed for their CYP2A6 genotypes. In the first article Finnish (n = 144) volunteers were sampled from the FINRISK’92 study material (Vartiainen et al. 1994). Phenotyping was carried out in volunteers (n = 11) selected from the staff of the Department of Pharmacology and Toxicology, University of Oulu (eight female and three male; age range 30–65 years). Furthermore, three individuals were shown previously to display a discrepancy between genotype and phenotype (LV26, LV29, and LV37) (Fernandez-Salguero et al. 1995). In the second article Finnish (n = 100) volunteers were collected from a larger study (V) (age range 27–67 years) living in the South Western area of Finland.
Unrelated Spanish white individuals with liver cirrhosis, liver cancer and control groups (IV) had been studied previously with respect to CYP2D6 genotypes (Agúndez J.A.G. et al. 1995). The control samples \( n = 237 \) were collected mostly from medical students, (129 female and 109 male, aged 18 –95 years). The liver cancer cases with hepatocellular carcinoma (HCC) \( n = 90 \) (15 female and 75 male, age range 34–82 years) and cirrhosis cases \( n = 83 \) (9 female and 74 male, age range 23–89 years) were collected from hospitals in the Madrid area.

The population of the fifth article of this thesis (V) consisted of 177 lung cancer patients (39 females and 138 males) and 1089 controls (535 females and 554 males) living in the South Western area of Finland. The lung cancer patients came to Helsinki University Central Hospital during a 9-year period (1988–1997) for surgical pneumonectomy or lobectomy due to suspected, operable lung cancer. All consecutive patients from two of the three surgical units of the Department of Thoracic and Cardiovascular Surgery were included. All the study subjects were confirmed to be of Finnish Caucasian origin. The mean age was 61.7 years \( \pm \)SD 14.7 years for the cases and 51.3 years \( \pm \)SD 14.7 years for the controls.

The sources of DNA samples and their specific information are presented in the articles I, II, III, IV and V.

### 4.1.3 DNA samples

DNA preparations used in the genotyping analyses were extracted from blood samples collected into EDTA tubes and stored at -20 °C until diluted and used. In some Finnish cancer cases, DNA was isolated from tissue samples (lungs). Purified genomic DNA samples were diluted for PCR analysis in water to a concentration of 10 ng/μl and stored at 4 °C.

### 4.2 Genotyping methods

The quality of DNA was found to critically affect the performance of the complex amplifications done in this work. The PCR conditions need to be standardised, and the first PCR reactions need to take place physically isolated from the second, to avoid contamination of the second PCR by the product of the first PCR reaction. The amplifications were carried out with the following PCR machines: MJ Research PTC-200 Peltier Thermal Cycler DNA-Engine, Perkin Elmer 2400 and 9600.
4.2.1 Original PCR-RFLP-based method

In the original method (Fernandez-Salguero et al. 1995), a two-step PCR protocol and digestion of the second PCR product with specific restriction enzymes (CYP2A6*2 by XcmI and CYP2A6*3 by DdeI) were used. This method was designed to identify the CYP2A6*1, CYP2A6*2 and CYP2A6*3 alleles. The conditions for amplification in the first and the second PCRs are presented in Table 7.

**Table 7. PCR conditions for the original genotyping method.**

<table>
<thead>
<tr>
<th>PCR I (29 cycles)</th>
<th>T (°C)</th>
<th>t (min.)</th>
<th>PCR II (35 cycles)</th>
<th>T (°C)</th>
<th>t (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturing</td>
<td>93</td>
<td>1:00</td>
<td>Denaturing</td>
<td>94</td>
<td>1:00</td>
</tr>
<tr>
<td>Annealing</td>
<td></td>
<td></td>
<td>Annealing</td>
<td>60</td>
<td>1:00</td>
</tr>
<tr>
<td>Extension</td>
<td>66</td>
<td>6:00</td>
<td>Extension</td>
<td>72</td>
<td>1:00</td>
</tr>
</tbody>
</table>

Data adapted from (15).

The PCR primers of the first and second PCRs are presented in Table 8.

**Table 8. The primers for PCR in the original genotyping method.**

<table>
<thead>
<tr>
<th>Name of primer</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR I F4</td>
<td>CCT CCC TTG CTG GCT GTG TCC CAA GCT AGG</td>
</tr>
<tr>
<td>PCR I R4</td>
<td>CGC CCC TTC CTT TCC GCC ATC CTG CCC CCA G</td>
</tr>
<tr>
<td>PCR II E3F</td>
<td>GCG TGG TAT TCA GCA ACG GG</td>
</tr>
<tr>
<td>PCR II E3R</td>
<td>TCG TCC TGG GTC TTT TCC TTC</td>
</tr>
</tbody>
</table>

Data adapted from (15).

In Fig. 7, a scheme of the original CYP2A6 genotyping method with the two-step PCR (primers in the first PCR F4 and R4; primers in the second PCR E3F and E3R) is presented. The second PCR product is digested by two restriction enzymes (RFLP with XcmI and DdeI) (Fernandez-Salguero et al. 1995).
4.2.2 SSCP method

An SSCP method was developed to replace the original genotyping method, which was found to misclassify some samples. In this method, the original two-step PCR procedure was used together with more specific primers. The PCR products were purified (QIAquick PCR purification Kit, Qiagen) and diluted with water to about 60 ng/µl. 30 ng of the purified PCR product mixed in a stop solution (US Biochemicals) was denatured at 100 °C, loaded on to acrylamide gel and run in an SSCP apparatus (Phastsystem Separation and Control Unit and Phast Gel Homogenous 20, Phastgel native buffer strips, Pharmacia Biotech). A pre-run was done in the following conditions: 400 V, 10 mA, 1 W and 100 aVh (average Volt hours); and the conditions for sample applications were 25 V, 10 mA, 1 W and 2 aVh, as recommended by the manufacturer. SSCP was performed at two different temperatures in the following conditions: 4 °C 600 aVh for three hours and 20 °C 500 aVh for two and a half hours. Allele strands were dissociated under specific conditions according to their conformation characteristics and analysed by non-denaturing polyacrylamide gel electrophoresis (PAGE). Gels were stained with silver staining according to the protocol recommended by the manufacturer (Phastsystem Development Unit and Phastgel Silverkit, Pharmacia Biotech). The gels were preserved with a special Gel Drying Kit (Promega). Optimally, with a well-purified heterozygous sample (one mutant and one wild-type allele), the gel exhibits five bands according to the mobility of all single strands (two wild-type and two mutant complementary strands) and
one band for the double-stranded allele. The resolution of the bands in the PAGE gel depends on how specific the PCR conditions are and how well purified the sample is (Orita et al. 1989, Hayashi 1992, Hayashi & Yandell 1993, Welsh et al. 1997).

To improve specificity, the protocol for amplification was modified by changing the PCR conditions (i.e., the number of cycles) and the pairs of primers. The modifications were based on a previous study, in which mutations of the p53 gene were explored in different exons (Welsh et al. 1997, Castrén et al. 1998). The PCRs were done as both one-step (F8 and R6 primers) and two-step procedures (the first and second PCR primers are presented in Table 9). The pairs of primers were designed to avoid nonspecific amplification caused mainly by CYP2A7 and CYP2A13, but acceptable amplification was not obtained. CYP2A6*3 differs from CYP2A6*1 in the exons 3, 6 and 8, and primers were designed for all these exons. Ultimately, however, this method was found to be too complicated and nonspecific (Gullstén et al., unpublished results). In Table 9, the specific primers of CYP2A6 for the first and second PCRs are shown. The SSCP conditions for CYP2A6 genotyping are shown in Table 10.

Table 9. Primers for CYP2A6 in the one-step and two-step PCR methods. F stands for forward (5’-side region of DNA) and R for reverse (3’-side).

<table>
<thead>
<tr>
<th>Name of primer</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-step PCR</strong></td>
<td></td>
</tr>
<tr>
<td>F8</td>
<td>CTG ATC GAC TAG GCG TGG TA</td>
</tr>
<tr>
<td>R6</td>
<td>CTG CCC CGT GGA GGT TGA</td>
</tr>
<tr>
<td><strong>Two-step PCR</strong></td>
<td></td>
</tr>
<tr>
<td>PCRI</td>
<td></td>
</tr>
<tr>
<td>F4</td>
<td>CCT CCC TTG CTG GCT GTG TCC CAA GCT AGG C</td>
</tr>
<tr>
<td>F5</td>
<td>AGG TAA TTA TGT AAT CAG CCA AAG TCC ATC CC</td>
</tr>
<tr>
<td>F6</td>
<td>CAC AGA TTG AGT CTG GAG GCC CCC TCT CTG</td>
</tr>
<tr>
<td>F7</td>
<td>GAA CAC AGA GCA GAT GTA CA</td>
</tr>
<tr>
<td>R4</td>
<td>CGC CCC TTC CTT TCC GCC ATC CTG CCC CCA G</td>
</tr>
<tr>
<td>R5</td>
<td>ATG TCC TTA GGT GAC TGG GA</td>
</tr>
<tr>
<td>R7</td>
<td>TTT CCG CCA TCC TGC CCC CAG TCT TAG CTG</td>
</tr>
<tr>
<td><strong>PCRII-exon 3</strong></td>
<td></td>
</tr>
<tr>
<td>E3F</td>
<td>GCG TGG TAT TCA GCA ACG GG</td>
</tr>
<tr>
<td>E3R</td>
<td>TCG TCC TGG GTG TTT TCC TTC</td>
</tr>
<tr>
<td>E3R1</td>
<td>AAC GGC GGC GGG TTC TCT CTC GT</td>
</tr>
<tr>
<td>E3R2</td>
<td>TTT CCC CAC CTA GTC CCC AT</td>
</tr>
<tr>
<td>E3R3</td>
<td>GGT CCC CTG CTC ACC GTG CG</td>
</tr>
<tr>
<td><strong>PCRII-exon 6</strong></td>
<td></td>
</tr>
<tr>
<td>E6F</td>
<td>AGT GAG GTT GTC CTA AAG CC</td>
</tr>
<tr>
<td>E6R</td>
<td>GGA CAT TGC ACC AGT CGA A</td>
</tr>
<tr>
<td><strong>PCRII-exon 8</strong></td>
<td></td>
</tr>
<tr>
<td>E8F</td>
<td>TGT GTA CTC TCA ACA ATC CC</td>
</tr>
<tr>
<td>E8F1</td>
<td>TCC CAG GGC ATA GAA GTG TT</td>
</tr>
<tr>
<td>E8R</td>
<td>GCA CCA AAC AGT GGT CTC TT</td>
</tr>
</tbody>
</table>
4.2.3 New allele-specific PCR methods

For the CYP2A6*2, CYP2A6*4A, CYP2A6*4D and CYP2A6*5 alleles, specific two-step PCR genotyping methods were developed. They were found to be reliable and specific enough for genotyping. By designing primers carefully and specifically for CYP2A6, amplification of the CYP2A7, CYP2A13 and CYP2A6*3 genes were avoided (Table 11). The PCR conditions were also optimised (magnesium, primers, buffer and DNA polymerase). The DNA concentration in the first PCR reaction was 10–50 ng. In comparison with the original method, the principal differences were the location of primers used in both PCRs, especially the allele-specific primers in the second PCR instead of restriction enzyme digestion (RFLP). More specific PCR conditions were also used, i.e. not too long annealing or extension times. These protocols are presented in more detail in articles I, II, and III.

Table 11. The specificity of primers in the original method (Fernandez-Salguero et al. 1995) and new genotyping methods.

<table>
<thead>
<tr>
<th>Methods and alleles</th>
<th>Primers and non-amplified member of CYP2A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First PCR</td>
</tr>
<tr>
<td>Fernandez-Salguero et al. 1995</td>
<td></td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>F4 – R4</td>
</tr>
<tr>
<td>CYP2A6*3</td>
<td>5‘-flanking CYP2A7 region</td>
</tr>
<tr>
<td>Oscarson et al. 1998,1999a, 1999b (I–III)</td>
<td></td>
</tr>
<tr>
<td>CYP2A6*1B</td>
<td>2A6ex8F – 2A6R1</td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>2A6ex1 – 2A6ex4R</td>
</tr>
<tr>
<td>CYP2A6*4</td>
<td>2Aex7F – 2A6R1</td>
</tr>
<tr>
<td>CYP2A6*5</td>
<td>2A6ex8F – 2A6R1</td>
</tr>
</tbody>
</table>

From the original references, see publications I, II, and III.

In the protocol some modifications were introduced in the first and second PCRs for the CYP2A6*4 allele, which was done as follows: initial denaturation was carried out at 95°C for 1 minute and final extension at 72°C for 7 minutes. The PCR conditions used for the CYP2A6*4 alleles are presented in Table 12.
Table 12. Published and modified PCR conditions for the first and second PCRs for genotyping CYP2A6*4.

<table>
<thead>
<tr>
<th>PCRI (35 cycles)</th>
<th>T (°C)</th>
<th>t (min.)</th>
<th>PCRII (14/17 cycles)</th>
<th>T (°C)</th>
<th>t (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation</td>
<td>95/94</td>
<td>0:15/0:10</td>
<td>Denaturation</td>
<td>95</td>
<td>0:15</td>
</tr>
<tr>
<td>Annealing</td>
<td>60</td>
<td>0:20/0:30</td>
<td>Annealing</td>
<td>52</td>
<td>0:20</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>3:00/2:00</td>
<td>Extension</td>
<td>72</td>
<td>2:00/1:00</td>
</tr>
</tbody>
</table>

4.3 Phenotyping methods

Specific probe drugs, metabolised selectively by the CYP2A6 enzyme, were used in the phenotyping studies. Two different phenotyping methods have been developed previously (Rautio et al. 1992, Cholerton et al. 1992b). In the standard protocol, a constant dose of a probe drug is given, urine or plasma is collected at certain time points, and the amount of the produced metabolite is determined (Rautio et al. 1992). Most phenotyping methods are based on HPLC (high performance liquid chromatography) detection of the parent drug and/or a metabolite in plasma or urine. The most commonly used index for assessing phenotype is the Metabolic Ratio (MR), which usually indicates a relationship between the parent drug and the produced metabolite, such as the ratio between the amounts of debrisoquine and 4-hydroxydebrisoquine excreted in urine during a certain period. On this basis, individuals are usually classified into PMs, IMs, EMs, and UMs according to the function of the enzyme (Daly et al. 1993).

4.3.1 Coumarin test

Coumarin was used as the first in vivo probe drug for CYP2A6. A German “natural drug”, Venalot® (contains 5 mg coumarin and 25 mg rutosides; manufactured by Schaper & Brummer, Ringelheim) is still used in routine testing for the CYP2A6 phenotype. This drug is not used in Finland, but it is still used in Germany. In the phenotyping test, the testee first gives a blank urine sample. Then one Venalot® capsule is taken together with a glass of water. In the standard test, urine is collected in fractions for 8 to 24 h. In a shortened version of the test, a blood sample is taken after one hour and the second cumulative urine sample is taken after two hours, when most of coumarin has been metabolised into 7-hydroxycoumarin.

The plasma half-life of coumarin is one hour and it is normally metabolised into 7-hydroxycoumarin within the first two hours. The amount of 7-hydroxycoumarin is measured by HPLC in urine, and the coumarin index (%) is calculated according to the following equation: Coumarin index (%) = [Amount of 7-hydroxycoumarin excreted in 2 hours versus the total coumarin dose] x 100 % (Rautio et al. 1992).

The following protocol for detecting 7-hydroxycoumarin with HPLC is used. The urine sample is centrifuged, and the supernatant is mixed with 1 M acetate buffer (pH 4.5) containing β-glucuronidase from bovine liver. The mixture is incubated in a metabolic
shaker at 37°C for two hours. After incubation, H₂O and chloroform are added to the mixture, and it is shaken at 37°C for 10 min. The chloroform phase is mixed with 1 M NaCl/0.01 M NaOH and, vortexed, and the alkaline phase is measured immediately by a spectrophotofluorometer at excitation and emission wavelengths of 365 and 454 nm. Standards and samples with known amounts of 7OHC are determined simultaneously (Rautio et al. 1992).

4.3.2 Nicotine test

Nicotine has been shown to be a specific probe drug for CYP2A6. Nicotine is first metabolised into cotinine, which is further metabolised into 3-hydroxycotinine, both reactions being catalysed predominantly by CYP2A6. The half-life of nicotine is short, two hour on an average, but the half-lives of cotinine and 3-hydroxycotinine are much longer, up to 24 hours. Because CYP2A6 metabolises nicotine, it was of interest to assess whether the metabolism of nicotine displays similar interindividual variation as that of coumarin. In liver microsomes in vitro, a 30-fold variation has been found in nicotine metabolism, and this variation correlated well with the variable expression of CYP2A6 (Foley et al. 1993, Nakajima et al. 1996a, Nakajima et al. 1996b, Messina et al. 1997, Yamazaki et al. 1999).

In the in vivo phenotyping test, the concentration of cotinine is measured in blood (plasma) or urine, depending on the method used. In the protocol used in the unpublished study, a blank urine sample is taken first. One piece of chewing gum (Nicorette, 2 mg nicotine; Pharmacia & Upjohn AB) is then given to the testee together with one glass of water. After an hour a blood sample and after two hours another urine sample are taken.

The protocol for measuring the concentrations of cotinine from the plasma sample by gas chromatography mass spectrometry (GC/MS) with a selected ion-monitoring mode is performed as follows. The concentration of cotinine and nicotine were measured. The serum is mixed with 0.5 M NaOH and dichloromethane containing pyribenzamine as an internal standard. After centrifugation of the mixture, the organic phase is taken and evaporated, the residue is dissolved in ethanol and the sample is injected into a GC/MS column. A fused silica capillary column coated with cross-linked methyl silicone gum is used, and helium is used as the carrier gas. The ions monitored here were m/z 98 and 176 for cotinine and m/z 91 for pyribenzamine (Bardy et al. 1993).

4.4 Heterologous expression

Expression systems are tools to find out whether a particular gene produces mRNA and/ or protein and whether the protein produced is functional or defective. In principle, the gene construct to be studied is first transfected into a suitable host cell (bacterium, yeast, etc.) and its expression is then measured at the mRNA, protein or activity level. Such systems are needed especially when the expression level of the gene is low or when there are no specific probes for the gene to study its function, i.e. phenotype, in vivo.
In this study, two expression systems were used to measure the function of the gene in yeast or mammalian COS cells. The results obtained in different host cells may sometimes differ from each other, but are usually identical. The wt CYP2A6 cDNA (Salonpää et al. 1993) was used as a template for subcloning it into the pYe60 yeast expression vector using different combinations of the specific primers excluding any potential artefact (III). *Saccharomyces cerevisiae* strain *W*, a genetically modified yeast cell expressing the yeast P450 reductase gene (Yred), which enhances CYP activities, was transfected with the plasmid consisting of the CYP2A6 construct produced in the system mentioned above (Truan et al. 1993). The expression of CYP2A6 was detected mainly as described previously (Bellamine et al. 1994, Oscarson et al. 1997). The microsomal protein concentration and the total CYP content were determined with the previously described methods (Omura & Sato 1964).

**In vitro** enzyme activities were measured in COS cells. CYP2A6 cDNA is transfected to the cells, after which enzyme activity with the specific substrate, coumarin, is measured. With this method, it has been possible to study whether the alleles under study produce proteins that incorporate heme and are able to metabolise the substrate. Coumarin 7-hydroxylase activity (COH) was measured in reaction mixtures containing Tris-HCl pH 7.4, yeast microsomes and coumarin. The reaction was initiated by NAPDH and terminated by 6 % trichloroacetic acid. The 7-hydroxycoumarin thereby formed was detected by a fluorometric method, as presented in detail in article III (Aitio 1978).

### 4.5 Characterisation and identification of new alleles

Since lack of correlation between genotype and phenotype was observed in several mistakes this led to a need to characterise more thoroughly the genotype/phenotype correlation. New alleles were found in course of these studies. At the phenotype level deviant phenotypes were observed as an abnormal *in vivo* activity (low or totally absent of coumarin 7-hydroxylase activity).

Sequencing and Southern blotting were used to characterise and identify new regions of the genome. All exons and some introns of the new alleles were sequenced and, on the basis of the sequence information, genotyping methods for these alleles were developed. Genomic DNA was amplified using two specific primers. The PCR fragments and cosmids (19296 and 19019) were partially sequenced on both strands. The sequencing method was described in detail in the articles II and III.

With Southern blotting, the *CYP2A* locus was digested by a restriction enzyme (*Eco*RI), and bands were visualised in agarose gels. Densitometric analysis of the Southern blot fragments was performed and revealed three fragments (6.6 kb, 7.5 kb and 7.7 kb) corresponding to the *CYP2A6*, *CYP2A7* and *CYP2A13* genes. The deletion *CYP2A6* genotype was observed in one sample lacking the 7.5 kb fragment completely, which was suggested to occur due to an unequal crossing-over event with 3’-flanking regions of *CYP2A6* and *CYP2A7*. Densitometric quantification suggested that the 7.5 kb fragment corresponded to 0, 1, and 2 copies of *CYP2A6* alleles. This method made it possible to recognise gene deletions, and are described in detail in article II.
4.6 Statistical analyses

The relative risks of liver cancer and cirrhosis associated with the CYP2A6 genotype in the case-control study were estimated using the Statxact –3 for Windows program (Cytel Software Corporation, Cambridge, MA: see Metha & Patel, 1995) (IV).

In the case-control study of Finnish lung cancer patients and controls (V), adjusted odds ratios (ORs) and 95% confidence intervals (CIs) were calculated by unconditional logistic regression. The analyses were performed by adjusting for the possible confounders age, gender, and smoking habits (McCullagh & Nelder 1994).
5 Results

5.1 Development of a new genotyping method for CYP2A6*2 (I)

A new and reliable genotyping method was first developed for detecting the CYP2A6*2 allele (L160H), because of the initial suspicion that the restriction enzyme XcmI digested it incompletely. For example, an individual with the heterozygous genotype (CYP2A6*1/*2) was originally classified as a homozygote (CYP2A6*2/*2), although the coumarin index of this individual was 40 % (the mean for a CYP2A6*1/*1 individual is 65 % and that for a CYP2A6*2/*2 individual 0 %). The development of the new genotyping method is presented in detail in article I.

Individuals from Caucasian populations were genotyped with this new method. The allele frequencies were found to be 1.1–3 % among Caucasians (Finns, Swedes and Spaniards). The allelic distribution as measured by this method conformed to the Hardy-Weinberg equilibrium. In conclusion, it appears that the CYP2A6*2 allele frequencies in Caucasian populations are rather low.

Some Finnish individuals were found to have aberrant phenotypes in previous phenotyping studies. A good correlation between phenotype and genotype was found in individuals homozygous for the CYP2A6*2 allele, since they were completely incapable of 7-hydroxylation coumarin. Both wild-type homozygous (CYP2A6*1/*1) and heterozygous (CYP2A6*1/*2) individuals were active in this respect, although there was some overlap. A correlation analysis between the CYP2A6 phenotype and the genotype is presented in Table 13.

Table 13. Phenotype-genotype correlation of CYP2A6 in the Finnish population.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Coumarin index (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2A6*1/*1 (n = 9)</td>
<td>Range 44–84</td>
</tr>
<tr>
<td></td>
<td>Mean 68</td>
</tr>
<tr>
<td>CYP2A6*1/*2 (n = 3)</td>
<td>Range 40–73</td>
</tr>
<tr>
<td></td>
<td>Mean 54</td>
</tr>
<tr>
<td>CYP2A6*2/*2 (n = 2)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Mean 0</td>
</tr>
</tbody>
</table>
5.2 Genotyping method for CYP2A6*4A (II)

A new whole gene-deletion allele (CYP2A6*4A) was found using the CYP2A6*2 genotyping method in a Spanish population, because the DNA of one individual with a genotype later shown to be CYP2A6*4A/CYP2A6*4A did not initially amplify at all. This sample was analysed by Southern blotting and sequencing and found to be a deletion of the whole CYP2A6 gene. An allele-specific genotyping method was designed for this deletion, and it is presented in detail in article II.

The frequencies of this deletion allele in Caucasian (Finnish and Spanish) and Asian (Chinese) populations were 0.5–1 % and 15.1 %, respectively. The CYP2A6*2 allele frequencies were 3 % in the same Caucasian populations and 0 % in the Chinese population, i.e., not a single CYP2A6*2 allele was found among 100 Chinese subjects. This deletion was hypothesised to have occurred due to unequal crossing-over events during the evolution; only the 3’-flanking region of the CYP2A6 gene was found to be compatible with CYP2A7. A CYP2A6 gene duplication was hypothesised to have occurred during the same crossing-over event, leading to the formation in the other DNA strand a hybrid of CYP2A7/CYP2A6 genes (deletion) and to the other strand one CYP2A7 gene, one CYP2A6 gene, and one hybrid CYP2A7/CYP2A6 gene. Nevertheless, no proof of the existence of duplication was found in our studies. Very recently, however, Rao et al. (Rao et al. 2000) described this duplication (CYP2A61X2).

According to the present results, the frequencies of CYP2A6 alleles exhibit clear interethnic variation. The Chinese population completely lacks the CYP2A6*2 allele, and the CYP2A6*4 deletion is more frequent among Chinese than Caucasians. Altogether, the inactive allele frequencies appear to be low in Caucasian populations.

5.3 Genotyping methods for CYP2A6*1B, CYP2A6*4D, CYP2A6*5 alleles (III)

Potential new variant alleles were identified by phenotyping analysis in Spanish and Thai populations. Upon closer analysis, several previously uncharacterised alleles (CYP2A6*1B, CYP2A6*4D, and CYP2A6*5) were found, and genotyping methods for CYP2A6*1B and CYP2A6*5 were developed. CYP2A6*1B may have been created by gene conversion with the CYP2A7 3’-flanking region. Its coding region is similar to the previously characterised wild-type allele (CYP2A6*1A), and CYP2A6*1B produces a functional protein as does CYP2A6*1A. CYP2A6*4D is a new whole-deletion allele (Fig. 10.). Because these two deletion alleles behave in the same way, they have been pooled as the CYP2A6*4 allele. CYP2A6*5 allele has one base substitution in exon 9 (G479V). The genotyping methods to measure these new alleles are presented in detail in article III.

The CYP2A6*1B allele frequencies were 30.0 % in Caucasians (Spaniards) and 40.6 % in an Asian population (Chinese). The allele frequencies for pooled CYP2A6*4 were 0.5 % in Caucasians (Spaniards) and 15.1 % in Chinese. The CYP2A6*5 allele frequencies in Caucasian and Asian populations were 0 % and 1 %, respectively. The CYP2A6*1B allele was found to be almost equally frequent as CYP2A6*1A. The CYP2A6*5 allele frequency was low.
Ethnic variation in CYP2A6 alleles is summarised in Table 14.

**Table 14. CYP2A6 allele frequencies in Caucasian and Asian populations.**

<table>
<thead>
<tr>
<th>Studies</th>
<th>CYP2A6 allele frequencies (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*1</td>
</tr>
<tr>
<td>Oscarson et al. 1998 (I)</td>
<td></td>
</tr>
<tr>
<td>Finns (n = 144)</td>
<td>98.6</td>
</tr>
<tr>
<td>Spaniards (n = 100)</td>
<td>97.0</td>
</tr>
<tr>
<td>Swedes (n = 90)</td>
<td>98.9</td>
</tr>
<tr>
<td>Oscarson et al. 1999a (II)</td>
<td></td>
</tr>
<tr>
<td>Finns (n = 100)</td>
<td>96.0</td>
</tr>
<tr>
<td>Spaniards (n = 100)</td>
<td>96.5</td>
</tr>
<tr>
<td>Chinese (n = 96)</td>
<td>84.9</td>
</tr>
<tr>
<td>Oscarson et al. 1999b (III)</td>
<td></td>
</tr>
<tr>
<td>Spaniards (n = 100)</td>
<td>(96.5)</td>
</tr>
<tr>
<td>Chinese (n = 96)</td>
<td>(83.8)</td>
</tr>
</tbody>
</table>

From the original references, see publications I, II, and III.

### 5.4 Existence of the CYP2A6*3 allele

Based on the studies presented in article III, the existence of the CYP2A6*3 allele was questioned. It was thought that this allele was probably the consequence of an erroneous classification (Bourian et al. 2000) caused by the carry-over of genomic CYP2A7 when the original genotyping method is used (Fernandez-Salguero et al. 1995). Because CYP2A6*1A and CYP2A6*1B differ only in their 3'-flanking regions and the original reverse primer based on the 3'-sequence preferred, at least in certain cases, the CYP2A6*1B-like sequences of CYP2A6 or CYP2A7 (carry-over), the outcome was that CYP2A6*1B alleles were often misclassified as CYP2A6*3. In the original genotyping method, the restriction enzyme DdeI probably also digested CYP2A6*3 insufficiently. Hence, the CYP2A6*3 alleles did not match the Hardy-Weinberg equilibrium and a CYP2A7 sequence was found in the second PCR product. To explain the origin of the (probably) artefactual CYP2A6*3 allele, it is worth noting that the sequences of CYP2A6 and CYP2A7 have 94 % similarity and the differences in the flanking region may also be crucial with certain primers.

### 5.5 SSCP as a genotyping method for CYP2A6

SSCP has been found to be a good method in distinguishing alleles in p53 mutation studies (Welsh et al. 1997, Castrén et al. 1998). Because of erroneous genotype allocation in the original CYP2A6 method, an attempt was made to develop a more specific SSCP-based method. Samples were amplified according to the original genotyping method with some changes in primers (exon 3 of CYP2A6). DNA samples containing the
following genotypes were used as positive controls to optimise the SSCP conditions: CYP2A6*1/*1, CYP2A6*1/*2, CYP2A6*2/*2 and CYP2A6*3/*3 (afterwards this sample was genotyped as CYP2A6*4D/*5 by the new genotyping methods). These PCR conditions were found not to be specific enough, and the whole protocol (the two-step PCR, the purification of the PCR product and allele separation by SSCP) was too complicated, laborious and expensive for routine use.

It was possible to separate the CYP2A6*1 and CYP2A6*2 alleles and single strands with the PCR-SSCP method (4°C, 600 aVh, F7-R5 primers in the first PCR and E3F-E3R primers in the second PCR). Occasionally, the CYP2A6*5 allele was also distinguished. The CYP2A6*4D deletion allele was not amplified with the specific primers, and the CYP2A6*4/*5 genotype therefore yielded only CYP2A6*5 allele strands in PAGE gels. Fig. 8 is an example of an SSCP gel after the separation of CYP2A6 allele strands (in the region of exon 3 and part of intron 3) at the temperature of 4°C. CYP2A6*1, CYP2A6*2 and CYP2A6*5 alleles appeared to be distinguishable. The bands present single complementary strands.

![Fig. 8. Schematic presentation of the SSCP method (4°C) for CYP2A6 genotyping.](image)

It might have been possible to separate the various alleles correctly in well-optimised SSCP conditions. The disadvantage of this method would have been the time-consuming SSCP analysis and the region studied would have been only a small part of the whole CYP2A6 gene (about 200 bp) (Gullstén et al., unpublished results). The allele-specific PCR methods developed at the same time were found to be much quicker and more reliable, and these genotyping methods were therefore chosen for routine use and developed further for the analysis of new alleles.
5.6 Phenotyping and genotyping of CYP2A6 in subjects with suspected defects in nicotine metabolism

In a previous study conducted by the National Public Health Institute, some smokers were found who seemingly did not produce any cotinine. Some of these individuals were identified and requested to volunteer for a study involving an analysis of nicotine kinetics and CYP2A6 genotyping and phenotyping with coumarin and nicotine. The following results were obtained. Five of the test subjects earlier classified as cotinine non-producers turned out to have low but detectable amounts of cotinine in their plasma after chewing nicotine gum. It is probable that the previously used cotinine detection system was not sensitive enough and therefore gave false negative results. Table 15 shows the data on the genotypes and phenotypes (metabolism of nicotine and coumarin) of these volunteers. Two heterozygous variant genotypes were found, and the metabolism of nicotine to cotinine in these individuals was much slower than in the wild-type homozygous individuals (Gullstén et al., unpublished results). In the phenotyping analyses, the amount of 7-hydroxycoumarin was measured in urine according to Rautio et al. (Rautio et al. 1992) (presented in chapter 4.3.1), and cotinine plasma levels were assayed according to Bardy et al. (Bardy et al. 1993) (presented in chapter 4.3.2).

Table 15. CYP2A6 genotypes and coumarin and nicotine phenotypes of selected Finnish individuals.

<table>
<thead>
<tr>
<th>Individuals (CYP2A6 Genotype)</th>
<th>Coumarin Index (%)</th>
<th>Nicotine (µg/l)</th>
<th>Cotinine (µg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (*1/*1)</td>
<td>69.3</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>2 (*1/*2)</td>
<td>72.8</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>3 (*1/*4)</td>
<td>72.8</td>
<td>1</td>
<td>92</td>
</tr>
<tr>
<td>4 (*1/*4)</td>
<td>72.8</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>5 (*1/*2)</td>
<td>Nd</td>
<td>2</td>
<td>77</td>
</tr>
<tr>
<td>Mean of CYP2A6*1/*1 (n = 5)</td>
<td>69.7</td>
<td>1.8</td>
<td>42</td>
</tr>
<tr>
<td>CYP2A6*1/*2 (n = 1)</td>
<td>73.2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>CYP2A6*1/*4 (n = 1)</td>
<td>36.5</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Nd = Not determined</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.7 Correlation between variant CYP2A6 alleles and liver cancer and cirrhosis (IV)

A case-control study was performed using liver cancer (n = 90) and liver cirrhosis cases (n = 83) and controls (n = 237) in the Spanish population. In this study, the original genotyping method was used (Fernandez-Salguero et al. 1995). The CYP2A6*1, CYP2A6*2 and CYP2A6*3 alleles were analysed, and their frequencies and associations with liver cancer or cirrhosis were determined.
In the original analysis, no association between CYP2A6 alleles and liver cancer was observed. There was a slight, but statistically non-significant correlation with liver cirrhosis, as the variant CYP2A6 homozygote genotype appeared to protect individuals from liver cirrhosis. The results of the original correlation study between CYP2A6 and liver cancer and cirrhosis are shown in Table 16.

Table 16. The distribution of CYP2A6 genotypes in Spanish cases and controls.

<table>
<thead>
<tr>
<th>CYP2A6 Genotype</th>
<th>Controls</th>
<th>Liver Cancer</th>
<th>Cirrhosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>% (OR 95 % CI)</td>
<td>% OR 95 % CI</td>
</tr>
<tr>
<td>Original method</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wt/wt</td>
<td>76.4</td>
<td>74.4 (1.0 referent)</td>
<td>83.1 (1.0 referent)</td>
</tr>
<tr>
<td>wt/m</td>
<td>21.1</td>
<td>23.3 (0.6–2.1)</td>
<td>16.9 (0.3–1.5)</td>
</tr>
<tr>
<td>m/m</td>
<td>2.5</td>
<td>2.2 (0.1–5.2)</td>
<td>–</td>
</tr>
<tr>
<td>Reanalysis of the results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*1/*2</td>
<td>3.4</td>
<td>5.5 (1.7 (0.5–5.3))</td>
<td>4.8 (1.3 (0.4–4.5))</td>
</tr>
<tr>
<td>*1/*3</td>
<td>17.7</td>
<td>17.8 (0.5–4.6)</td>
<td>–</td>
</tr>
<tr>
<td>*2/*2</td>
<td>2.1</td>
<td>1.1 (0.4–4.5)</td>
<td>–</td>
</tr>
<tr>
<td>*2/*3</td>
<td>–</td>
<td>1.1 (0.4–4.5)</td>
<td>–</td>
</tr>
<tr>
<td>*3/*3</td>
<td>0.4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Pooled</td>
<td>5.5</td>
<td>7.8 (1.4 (0.5–3.7))</td>
<td>4.8 (1.3 (0.4–4.5))</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

For original reference, see publication IV.

As described previously, a defect in the original genotyping method resulted in misclassification of genotypes due to the (probably) non-existent CYP2A6*3 allele. As a consequence, known phenotypes of certain individuals did not match their genotypes. There were also some problems with restriction enzymes (XcmI), which did not always work correctly. For instance, the CYP2A6*1/*2 genotype was erroneously classified as a CYP2A6*2/*2 genotype.

In a reanalysis of these results, in which only the CYP2A6*2 allele was taken into consideration (Table 16), a protective association was found between the CYP2A6*2/*2 genotype and liver cancer (OR 0.5, CI 0.1–4.6 %), which was not, however, statistically significant. In the group with CYP2A6*1/*2, no association was found to liver cancer or cirrhosis (OR 1.7, CI % 0.5–5.3; OR 1.3, CI % 0.5–4.5). A similar tendency was also seen in the group of pooled genotypes, containing at least one CYP2A6*2 allele in the genotype (OR 1.4, CI 0.5–3.7 %; OR 1.3, CI 0.4–4.5 %). One CYP2A6*2/*3 genotype was found among the liver cancer cases, and the actual genotype of this individual is not known.

Calculations were also performed concerning the CYP2A6*3 allele. CYP2A6*1/*3 heterozygotes were overrepresented compared with homozygotes, i.e., it did not conform to the Hardy-Weinberg equilibrium. The liver cirrhosis group had a reduced number of CYP2A6*1/*3 genotypes, and the CYP2A6*3/*3 genotype was only found in the control group (frequency 0.4 %). ORs and CIs were not calculated for genotypes consisting of CYP2A6*3 allele(s) because of its artefactual nature (see above).
Previously, defective CYP2D6 genotypes were found to have a similar association to liver cancer as CYP2A6 (Agúndez et al. 1995). However, no definitive conclusions can be drawn due to the small number of samples and the low number of defective alleles present in the Caucasian population. As the genotyping was done with the original method, some erroneous results might also have been obtained for the CYP2A6*2 allele, as detailed in Materials and Methods.

5.8 Correlation between CYP2A6 and lung cancer (V)

In a Finnish population of 177 lung cancer cases and 1089 controls, the CYP2A6*2 and CYP2A6*4 alleles were analysed. The allele frequencies among lung cancer cases and controls were very similar (Table 17), and no association between the alleles and lung cancer risk was found. Due to the large size of the study population the results are anticipated to give a reliable estimate of the role of CYP2A6 in susceptibility to lung cancer in the Finnish population.

Table 17. Allele frequencies of CYP2A6 in Finnish lung cancer and control subjects.

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Controls</th>
<th>Lung cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% (n)</td>
<td>% (n)</td>
</tr>
<tr>
<td>CYP2A6*1</td>
<td>95.2 (2074)</td>
<td>96.3 (341)</td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>2.7 (58)</td>
<td>2.0 (7)</td>
</tr>
<tr>
<td>CYP2A6*4</td>
<td>2.1 (46)</td>
<td>1.7 (6)</td>
</tr>
<tr>
<td>Total</td>
<td>100 (2178)</td>
<td>100 (354)</td>
</tr>
</tbody>
</table>

From original reference, see publication V.

In Table 18, statistical parameters (crude ORs, adjusted ORs and 95 % CI) are presented for at least one inactive allele containing genotypes in controls and lung cancer cases.

Table 18. Crude and adjusted ORs, and confidence intervals for CYP2A6*1/*2, CYP2A6*1/*4 and CYP2A6var genotypes in lung cancer.

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Crude ORs</th>
<th>95% CI</th>
<th>Adjusted ORs</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2A6*1/*1</td>
<td>1.00 (referent)</td>
<td></td>
<td>1.00 (referent)</td>
<td></td>
</tr>
<tr>
<td>CYP2A6*1/*2</td>
<td>0.75 (0.34–1.68)</td>
<td>0.92 (0.36–2.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*1/*4</td>
<td>0.82 (0.34–1.95)</td>
<td>0.79 (0.27–2.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6var*</td>
<td>0.76 (0.42–1.40)</td>
<td>1.19 (0.56–2.45)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Combined genotypes containing at least one inactive allele.

From original reference, see publication V.

Adjusted ORs indicated no association between CYP2A6var genotypes and lung cancer (OR, 1.19; 95 % CI, 0.56–2.45). Adjusting with confounding factors (age, sex and smoking behavior) in the study groups balanced the slight difference between the adjusted and crude ORs.

Adjusted ORs indicated no association between CYP2A6var genotypes and lung cancer (OR, 1.19; 95 % CI, 0.56–2.45). Adjusting with confounding factors (age, sex and smoking behavior) in the study groups balanced the slight difference between the adjusted and crude ORs.
6 Discussion

6.1 Development of genotyping methods and CYP2A6 alleles

The first genotyping method for variant CYP2A6 alleles was published in 1995 by Fernandez-Salguero et al. (Fernandez-Salguero et al. 1995). The chromosomal organisation of the CYP2A subfamily, polymorphic alleles (CYP2A6*1, CYP2A6*2 and CYP2A6*3) and a description of genotyping methods were introduced in this article. A two-step PCR genotyping method using two restriction enzymes (XcmI and DdeI) was developed for the detection of CYP2A6 variant alleles. During the present research project several problems with this genotyping method were encountered. The CYP2A6 genotypes were misclassified due to the insufficient specificity of PCR primers and the occasional amplification of CYP2A7 and CYP2A13 genes. The primers used in the first PCR reaction were located in noncoding regions of the CYP2A6 gene and one primer in the intron region in the second PCR reaction. The restriction enzymes also digested inconsistently: XcmI erroneously digested also CYP2A6*1/*2 instead of only CYP2A6*2/*2, and DdeI digested mainly CYP2A7 (-like) sequences resulting in CYP2A6*3 classification. The third reason for misclassification, especially concerning CYP2A6*3, turned out to be another wild-type allele (CYP2A6*1B), which differs from the first reported wild-type allele (CYP2A6*1A) in the area of the 3'-flanking region. Due to these problems in the original genotyping method, the development of more reliable genotyping methods was considered to be necessary early during the project.

One attempt to overcome the genotyping problems was to develop an SSCP method, which had been used to separate mutations in certain exons of the p53 gene under specific conditions (Welsh et al. 1997). The developed SSCP-PCR method was found to give reliable results depending on the PCR and electrophoresis conditions (see Results). At 4°C temperature it was possible to distinguish CYP2A6*1, CYP2A6*2 and CYP2A6*5 allele strands (Fig. 8). The optimising of specific separation conditions was found, however, to be too complicated, time-consuming and expensive in practice. Another disadvantage of this method is that only short regions (about 200 bp) are amenable to analysis at a time. Thus, with an SSCP method optimised for one region (for example, CYP2A6 exon 3), it is not possible to analyse other mutations in other exons of the gene (Gullstén et al., unpublished results).
Finally, the allele-specific two-step PCR methods with highly specific primers were found to be the most reliable genotyping methods for \textit{CYP2A6}. In addition to designing a more reliable method for the \textit{CYP2A6*2} allele (I), new alleles were identified and characterised, and genotyping methods with special primers were designed for them (II and III). The genotyping methods presented in this thesis (I-III) are very time-consuming especially in large population studies. In the present studies DNA samples were genotyped at least twice to obtain correct genotypes. All developed genotyping methods for \textit{CYP2A6} alleles are shown in Table 19. Schematic protocols of the original and new genotyping methods are presented in Fig. 9.

In conclusion, these studies demonstrated clearly that genotyping methods need to be designed and validated carefully. Every allele should be sequenced and the PCR primers should be designed to known target regions rather than intron/uncoding regions. More \textit{CYP2A6} alleles will be probably found in the future.

During this project, the following alleles were found: two wild-type alleles (\textit{CYP2A6*1A} and \textit{CYP2A6*1B}, pooled as \textit{CYP2A6*1}), two point mutation alleles (\textit{CYP2A6*2} and \textit{CYP2A6*5}), and two gene deletion alleles (\textit{CYP2A6*4A} and \textit{CYP2A6*4D}, pooled as \textit{CYP2A6*4}) (articles I, II and III). The previously reported \textit{CYP2A6*3} allele was not found in these studies. Both \textit{CYP2A6*5} and \textit{CYP2A6*1B} allele were not amplified by the original genotyping method (Fernandez-Salguero et al. 1995) due to the mutated 3'-flanking regions (II). Therefore, \textit{CYP2A7} was amplified instead of \textit{CYP2A6*1B} and \textit{CYP2A6*5} alleles due to amplification of carry-over DNA (too many cyces in the second PCR). A similar conclusion on the existence of the \textit{CYP2A6*3} allele has also been made by two other research groups (Kitagawa et al. 1999, Sabol & Hamer 1999).

The \textit{CYP2A6*1B} allele has some silent mutations in its 3'-flanking region, but appears to produce a functional protein. \textit{CYP2A6*5} has one point mutation in exon 9 (substitution, G479V), and the encoded protein is inactive. \textit{CYP2A6*4A} is a gene deletion allele; only a short part of the 3'-flanking region of the \textit{CYP2A6} gene is present. \textit{CYP2A6*4D} has the whole gene deleted as well, but part of the 3'-region of the \textit{CYP2A7} gene is also deleted. The deletions are suggested to have developed due to unequal crossing-over events, and gene duplications are also thought to have developed at same time (III). For obvious reasons, the deletion alleles are completely inactive. The structures of these \textit{CYP2A6} alleles are presented in Fig. 10.
### Table 19. All published genotyping methods for CYP2A6 alleles.

<table>
<thead>
<tr>
<th>Articles</th>
<th>Region of CYP2A6</th>
<th>First PCR</th>
<th>Normal Second PCR</th>
<th>Allele Specific Second PCR</th>
<th>RFLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fernandez-Salgueiro et al. 1995</td>
<td>5’-Uncoding Region – 3’-Uncoding Region (F4 – R4)</td>
<td>Exon 3 - Intron 3 (E3F – E3R)</td>
<td>-</td>
<td>XcmI</td>
<td></td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numoya et al. 1998</td>
<td>Exon 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*4B</td>
<td>(2A6-A8/2A7-B1 – 2A6-B6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oscarsen et al. 1998</td>
<td>Exon 1 – Exon 4 (ex1 – ex4R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numoya et al. 1999a</td>
<td>Exon 1 (2A6-F1 – 2A6-1AS-2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*4C</td>
<td>Exon 8 (2A-6-B1 – 2A6-B6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*4C</td>
<td>Exon 9 (2A6-9S – 2A6-B1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen et al. 1999</td>
<td>Intron 2 – Intron 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>(CYP2A6F93 – CYP2A7F06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oscarsen et al. 1999a</td>
<td>Exon 7 – 3’-Uncoding Region (2Aex7F – 2A6R1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*4A</td>
<td>Overlapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sabol et al. 1999</td>
<td>Exon 8 – 3’-Uncoding Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>(CYP2A6 – E3R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kitagawa et al. 1999</td>
<td>Intron 2 &amp; Exon 3 – Intron 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*1</td>
<td>Overlapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*2</td>
<td>(KallF – E3R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oscarsen et al. 1999b</td>
<td>Exon 8 – 3’-Uncoding Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*4D</td>
<td>(2A6eX8F – 2A6R1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*5</td>
<td>Same as CYP2A6*4A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rao et al. 2000</td>
<td>Exon 7 – 3’-Uncoding Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6*4A</td>
<td>(2A6eX8F/2A7ex8F – 2A7R1 – 2A6R1) or</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zabetian et al. 2000</td>
<td>3’ end of exon 2 – exon 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6<em>1 (CYP2A6</em>8A)</td>
<td>Exon 5 - Intron 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6<em>2 (as CYP2A6</em>4D)</td>
<td>(E3F – E3R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2A6<em>3 (as CYP2A6</em>8B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 9. PCR protocols for the original and new genotyping methods (location of primers and amplified PCR products): a) original method for \textit{CYP2A6*1}, \textit{CYP2A6*2} and \textit{CYP2A6*3}; b) new allele-specific PCR methods for \textit{CYP2A6*1} and \textit{CYP2A6*2}, c) \textit{CYP2A6*1} and \textit{CYP2A6*4} d) \textit{CYP2A6*1B} and \textit{CYP2A6*5}.
6.2 CYP2A6 phenotype-genotype correlation

The phenotype of an individual can be predicted from the genotype, and tedious phenotyping can therefore be avoided, but only if the phenotype-genotype correlation is known. On the other hand, when searching for new variant alleles that are also functionally significant, it is preferable to perform phenotyping analysis first.
In the substudies of this project, the CYP2A6 phenotyping was done using coumarin and nicotine as probe drugs, both of which are specific for CYP2A6. Some volunteers were phenotyped with these probe drugs (coumarin capsule/nicotine chewing gum), and their ability to metabolise these probe agents was found to correspond to their genotypes (Gullstén et al., unpublished results). The genotype-phenotype correlation for CYP2A6 was found: the \textit{CYP2A6*2} or deleted \textit{CYP2A6} homozygotes were defective, the heterozygotes had lowered activity and the wild-type homozygotes had normal metabolic activity (I). A similar tendency was also found with nicotine, the variant heterozygotes having clearly lowered ability of nicotine metabolism (Gullstén et al., unpublished results). Although the number of samples in our unpublished study (see Results) was low, the results were similar to those obtained in other studies (Kitagawa K et al. 1999).

### 6.3 Interethnic variation in genotype frequencies

Most XMEs have been found to display considerable genetic heterogeneity. Environment and diet are thought to be important modifiers during the evolution, the genome being adapted to the living circumstances through natural selection (see section 2.1.). For instance, the \textit{CYP2D6} gene has been found to have about 50 variant alleles (causing interindividual differences), and the allele frequencies differ greatly between regions and continents (interethnic differences). For example, the frequency of \textit{CYP2D6} multiduplication alleles (leading to the UM phenotype) is 1-2 \% in the Nordic countries, but 29 \% among Ethiopians (Ingelman-Sundberg 1999, Aynacioglu et al. 1999).

\textit{CYP2A6} enzyme activities have been found to be very variable \textit{in vitro} in liver microsome samples and \textit{in vivo} in humans (see section 2.4.3.) and there are marked differences between various ethnic groups. In the present studies, \textit{CYP2A6} alleles were found to display distinct interethnic differences (I, II and III) (see section 5.3.). The most conspicuous of these differences was that the deleted \textit{CYP2A6*4} alleles are more frequent in Asian than Caucasian populations. On the other hand, \textit{CYP2A6*2} is lacking in Asian and \textit{CYP2A6*5} does not exist at all in Caucasian population (I, II and III). However, because the populations studied thus far are not very large, allele frequencies, especially those of rare alleles, may change when larger populations are assessed. The ethnic differences are summarised in Table 21.
Table 21. The differences of CYP2A6 in different ethnic groups in relation to selected diseases and smoking behavior.

<table>
<thead>
<tr>
<th>Diseases/ smoking behavior</th>
<th>Population</th>
<th>Association</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other diseases (than lung cancer)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gullstén et al. 1997</td>
<td>Caucasian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Kamataki et al. 1999</td>
<td>Asian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Tan et al. 2000</td>
<td>Asian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Lung cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London et al. 1999</td>
<td>American</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Miyamoto et al. 1999</td>
<td>Asian</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Kamataki et al. 1999</td>
<td>Asian</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Loriot et al. 2000</td>
<td>Caucasian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Tan et al. 2000</td>
<td>Asian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Gullstén et al. submitted</td>
<td>Caucasian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Smoking behavior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pianezza et al. 1998</td>
<td>American</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>London et al. 1999</td>
<td>American</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Tan et al. 2000</td>
<td>Asian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Tiihonen et al. 2000</td>
<td>Caucasian</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Rao et al. 2000</td>
<td>Caucasian</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Loriot et al. 2000</td>
<td>Caucasian</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>


No association between CYP2A6 and other diseases (liver cancer, cirrhosis, esophageal cancer, small cell carcinoma, colorectal cancer, neuroblastoma, and others) was found either in Caucasian and Asian population (three studies). Controversial results between CYP2A6 and lung cancer have been obtained: no association was found in Caucasians (two studies), whereas an association was found in American (one study) and Asian populations (two studies). Furthermore, controversial results were found also in smoking behavior: no association was found in Caucasians (two out of three studies), Americans (one out of two studies), and Asians (one study). In conclusion, CYP2A6 may have an association in lung cancer in American and Asian population, but not in Caucasian population, this conclusion refers to an ethnic variation between these populations.

6.4 Role of CYP2A6 in liver cancer and cirrhosis

Because the CYP2A6 enzyme metabolises and activates many hepatocarcinogens, it may play a role as a modulator in the initiation of liver cancer and cirrhosis. The presence of functionally significant polymorphisms gives an opportunity to perform molecular epidemiological investigations to study the possible associations between specific genotypes and the risk of cancer and cirrhosis.
In the case-control studies performed here, it was found that the variant homozygote genotype might exert a slight protective effect on the risk of liver cirrhosis in the Spanish population, although the association was not statistically significant (V). No association was found between CYP2A6 polymorphism and liver cancer risk. However, the method of genotyping CYP2A6 used in this study was the original one (Fernandez-Salguero et al. 1995) and, consequently, a reanalysis of the results was performed (see Results). Upon reanalysis, a weak protective association between the CYP2A6*2/*2 genotype and liver cancer was found. However, the number of samples and the incidence of variant alleles were quite low, and no definitive conclusions can be drawn. Studies on larger populations including all the known variant alleles are required before a definite answer concerning the possible association between CYP2A6 polymorphisms and the risk of liver cancer can be obtained. In the same Spanish population, a protective effect of defective CYP2D6 genotypes against liver cancer was observed (Agúndez et al. 1995). Therefore, it would be of interest in the future to study the possible synergistic effects of CYP2A6 and CYP2D6 polymorphisms.

6.5 Role of CYP2A6 in lung cancer

The underlying hypothesis is that individuals with the CYP2A6 PM phenotype might be protected from lung cancer due to their inability to activate procarcinogens present in tobacco smoke. The following case-control studies were performed with new and improved genotyping methods. Miyamoto et al. and Kamataki et al. found an association between the defective CYP2A6 phenotype and lung cancer in the Japanese (Miyamoto et al. 1999, Kamataki et al. 1999). However, our own study of the Finnish population did not show any correlation between the inactive/deleted CYP2A6 genotypes and lung cancer (V), and similar conclusions have also been made in French and Chinese populations (Chen et al. 1999, Tan et al. 2000, Loriot et al. 2000).

In Table 22 case-control studies carried out with similar new CYP2A6 genotyping methods in Caucasian populations are shown.
Table 22. Case-control studies with new CYP2A6 genotyping methods (CYP2A6*2 and CYP2A6*4) in Caucasian populations.

<table>
<thead>
<tr>
<th>Articles</th>
<th>Association with</th>
<th>Smokers (Cigarettes per day, mean ± SD)</th>
<th>Defective alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lung cancer</td>
<td>Smoking behavior</td>
<td></td>
</tr>
<tr>
<td>Tiihonen et al. 2000</td>
<td>-</td>
<td>No</td>
<td>18.4 ± 9.9 17.4 ± 6.2</td>
</tr>
<tr>
<td>Rao et al. 2000</td>
<td>-</td>
<td>Yes</td>
<td>19.5 ± 0.7 13.5 ± 2.3</td>
</tr>
<tr>
<td>Loriot et al. 2000</td>
<td>No</td>
<td>No</td>
<td>16.5 ± 12.7 14.9 ± 8.9</td>
</tr>
<tr>
<td>Gullstén et al. submitted</td>
<td>No</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>No (2/2)</td>
<td>No (2/3)</td>
<td></td>
</tr>
</tbody>
</table>

Data adapted from (Rao et al. 2000, Loriot et al. 2000, Tiihonen et al. 2000) and V.

No association was found between CYP2A6 and lung cancer in Caucasian populations (two out of two studies). The number of samples in these studies were large (Loriot et al. 250 controls and 244 cases: OR, 1.1; 95 % CI, 0.7–1.9; Gullstén et al., 1089 controls and 177 cases: OR, 1.19; 95 % CI, 0.56–2.45) (Loriot et al. 2000) and V. In addition, two out of three studies found no association between CYP2A6 and smoking behavior.

All together, these case-control studies have yielded discrepant results. Some differences in the results can be ascribed to the different genotyping methods used, but interethnic differences in CYP2A6 allele distributions are anticipated to also contribute significantly to the results.

6.6 Future

To date there are a vast number of unanswered questions in this research area. For instance could the explanation for interindividual and interethnic differences in XMEs be in the lifestyle (nutrition, stimulants, working or home environment or air-borne pollutants) or/and related differences in other parts of the genome (such as linkage to another polymorphic gene, polymorphisms in the system regulating expression)? Does there exist an unknown mechanism affecting the expression level or function of XMEs? What consequences does the high frequency of defective phenotypes in some population have? Could certain defective XME phenotypes have an association with some diseases or smoking behaviour? Could there also be effects on other systems, such as reproduction or pregnancy (implantation, prenatal growth, postnatal development or even resistance to bacterial/viral infections), and what consequences could this have? How large an effect do the polymorphic XMEs have on the metabolism of drugs, on finding significant drug candidates for drug industries and on administering the right dose of a drug to an individual with a defective phenotype of XMEs.

How do these questions then apply to research on CYP2A6 polymorphisms? CYP2A6 has been found to have a PM phenotype lacking the activity, and a duplicated allele possible resulting in an UM phenotype with increased activity. One explanation for the associations with lung cancer or other diseases could be a combination of one or two
other polymorphic genes having the same effect and being linked with CYP2A6. Such
linkage has been hypothesised to exist between CYP2A6 and CYP2C19 and also with
CYP2A6 and CYP2B6.

Much effort has recently been put on examining potential association between CYP2A6
tobacco smoking behaviour. The underlying hypothesis is that the concentration of
nicotine in blood would presumably reach higher levels and persists for a longer time in
PM individuals. Thus, PM subjects might reduce smoking or do not start smoking at all.
A study performed using the original genotyping method indicated that individuals with
IM or PM phenotypes smoked fewer cigarettes or did not start to smoke at all (the variant
alleles were pooled in this study) (Pianezza et al. 1998). Recently, in three studies carried
out with the new genotyping methods no association between CYP2A6 status and
smoking behaviour was found (Sabol & Hamer 1999, Loriot et al. 2000, Tiihonen et al.
2000). In one recent study, however, an association between CYP2A6 and smoking
behaviour was found, individuals with defective alleles smoked fewer cigarettes, and
individuals with the duplicated CYP2A6 allele smoked more cigarettes per day than
individuals with normal genotypes (Rao et al. 2000). Sellers et al. suggested that it is
possible to reduce addiction to smoking by inhibiting the function of CYP2A6 enzyme by
methoxsalen (Rao et al. 2000, Sellers & Tyndale 2000). The future studies done will
reveal whether this inhibitor really could have an important role in smoking cessation,
thereby decreasing availability of tobacco smoke containing procarcinogens, and this way
reducing individuals risk to have lung cancer.

Altogether in the general level of research will be fascinating to learn more, if the
frequencies of the defective alleles are also higher in some other populations, such as
Africans and Australians. To accomplish this goal, higher throughput genotyping methods
are needed, possibly based on oligonucleotide chip technology. The data produced by the
Human Genome Project will soon allow mass research of candidate genes for common
diseases. SNPs (single nucleotide polymorphisms) will be located and their effects on the
function of genes, and their probable role in diseases will be solved.
7 Summary and conclusions

It is important to clarify the role of polymorphic XMEs in disease susceptibility, as they are possible risk factors in the initiation of chemical cancer or other diseases. Polymorphisms, if associated with disease risk, help to define possible risk groups, so that appropriate preventive measures can be developed, such as protecting susceptible individuals from exposure to toxic substances or procarcinogens.

The polymorphic XMEs – through causing impaired/enhanced metabolism of drugs – are important parameters also in practical drug therapy. Knowledge of the polymorphic XME status in individual patients might be useful in anticipating dosage and therapeutic efficacy of the planned drug administration. For the pharmaceutical industry, detailed knowledge on CYP2A6 polymorphisms will help in planning and executing clinical studies on drug candidates that are substrates of the CYP2A6 enzyme.

1. Methods development:
   a) Novel CYP2A6 alleles were found, and reliable genotyping methods for them were developed (CYP2A6*1B, CYP2A6*2, CYP2A6*4, CYP2A6*5). CYP2A6*1B is a wild-type allele differing from the originally published CYP2A6*1A allele in that it has mutations in the 3'-flanking region. CYP2A6*2 and CYP2A6*5 have one point mutation each in the exon region (CYP2A6*2 in exon 3 and CYP2A6*5 in exon 9) and CYP2A6*4 alleles are composed of two types of whole-gene deletions (CYP2A6*4A and CYP2A6*4D). All of these alleles were found to produce inactive or unstable enzymes. A hypothesis on the existence of a duplicated allele was presented. The CYP2A6*3 allele is suggested to be an artefact caused by unspecific genotyping methods (carry-over of genomic DNA and previously unknown CYP2A6*1B allele).
   b) Interindividual differences in CYP2A6 enzyme activity proved to exist in phenotype-genotype studies. The phenotype-genotype correlation was found to be rather clear, PM individuals not producing any 7-hydroxycoumarin, and IM individuals producing decreased amounts of 7-hydroxycoumarin (or cotinine in nicotine test).

2. Population and case-control studies:
   a) Interethnic differences in allele frequencies were observed in Asian and Caucasian populations. The frequencies of defective alleles are rather low (0–3 %) in the Caucasian population, and the deletion alleles (CYP2A6*4) are quite frequent (15 %) in Asians.
b) Studies on liver cirrhosis revealed a slight, but non-significant protective effect of the CYP2A6 PM genotype in a Spanish population. A reanalysis of CYP2A6*2 alleles also revealed a slight protective association with liver cancer. Clearly, a study with a larger number of samples is required to obtain a definitive result on this matter.

c) A case-control study pointed out no association between CYP2A6 and lung cancer in the Finnish population. Due to the low frequency of variant CYP2A6 alleles in Caucasians, it is probable that a correlation between inactive CYP2A6 and lung cancer is found only in Asian populations.
8 References


Duescher RJ & Elfarra AA (1994) Human liver microsomes are efficient catalysts of 1,3-butadiene oxidation: evidence for major roles by cytochromes P450 2A6 and 2E1. Arch Biochem Biophys 311: 342-349.


www-pages:
http://www.incyte.com/web site
http://www.gene.ucl.ac.uk/nomenclature/
http://drnelson.utmem.edu/cytochromeP450.html/
http://louisville.edu/medschool/pharmacology/NAT.html/
http://www.imm.ki.se/CYPalleles/