IPS Instituto Politécnico de Setúbal
Escola Superior de Tecnologia do Barreiro

JOSIMAR PIRES DA CRUZ

DEVELOPMENT OF A BACTERIAL BIOMIMETIC SYSTEM FOR LIVER DRUG METABOLISM

Relatório de Dissertação do Mestrado em Engenharia Biológica e Química

ORIENTADOR
Professora Doutora Marta Justino

SUPERVISOR
Doutor Gonçalo Justino

Julho de 2021

JOSIMAR PIRES DA CRUZ

DEVELOPMENT OF A BACTERIAL BIOMIMETIC SYSTEM FOR LIVER DRUG METABOLISM

JÚRI

Presidente: Professora Doutora Maria de Lurdes de Figueiredo Gameiro, ESTBarreiro/IPS

Supervisor: Doutor José Gonçalo Deira Duarte de Campos Justino, Centro de Química Estrutural, Instituto Superior Técnico, Universidade de Lisboa

Vogal: Doutora Sofia de Oliveira Dias Duarte, Instituto de Bioengenharia e Biociências, Instituto Superior Técnico, Universidade de Lisboa

Julho de 2021

ACKNOWLEDGEMENTS

I would like to acknowledge Instituto Superior Técnico – Universidade de Lisboa for the opportunity to do my master thesis in a fantastic place with amazing people. I am also very grateful to my professors and supervisors Doctor Gonçalo Justino and Doctor Marta Justino for accepting me to develop this thesis, for teaching me the basic techniques about this field, for helping me to solve a lot of problems in the laboratory, and how to develop a critical sense. I would like to say thank you for all the motivation, understanding, personal and professional support in this stage of my life. It was an amazing experience where I learned a lot with two fantastic professors.

I am very grateful to all the co-workers that supported me in the laboratory during this time, especially Pedro Rosado for all his support in some techniques and for helping me solve a lot of problems. I also thank Sofia Raposo for her help in the maintenance of some culture and preparing some material to develop the work.

I appreciate all the advice related to this thesis and I am very honoured to work with all of you, for making this stage special in my life and especially for making IST a wonderful place to work!

Com toda a sinceridade este ano foi o ano das experiências, dos desafios, de aprendizagem e acima de tudo de muitas lágrimas. Sendo assim tenho o dever de agradecer todas as pessoas que me impediram de desistir e acima de tudo que me apresentaram soluções e que me fizeram acreditar que vale a pena lutar mesmo quando muitas das vezes estamos sem expectativas de vencer. Primeiramente gostaria de agradecer todo o apoio, compreensão e incentivo dos meus orientadores Prof. Gonçalo Justino e Prof. Marta Justino. Agradeço profundamente por serem os principais responsáveis por parte das conquistas que alcancei e por serem pessoas tão maravilhosas, difíceis de adjetivar.

Quero agradecer aos colegas dos almoços pelos momentos de conversa e de muitos sorrisos sem fim, pelos desabafos, reclamações, brincadeiras, enfim por todos os momentos partilhados. Sou muito grato à toda a minha família pela compreensão, pelo apoio prestado, especialmente ao meu pai, a minha mãe e as minhas irmãs, por estarem sempre presentes quando precisei.

Agradeço profundamente à minha noiva Aida Chantre por ter sido maravilhosamente compreensiva durante esta fase, por ter sido sempre o porto seguro na minha vida, pelo apoio prestado e acima de tudo pelo presente maravilhoso que me deu durante esta fase. Muito obrigado pelo teu esforço e por apresentar soluções as quais foram determinantes para enfrentar e ultrapassar os desafios enfrentados. Ao meu pequenino e maravilhoso Luciano da Cruz pela motivação, por mudar a forma como via o mundo, pelos sorrisos maravilhosos e acima de tudo por despertar em mim um amor incapaz de ser definido.

A todas as pessoas que estiveram presentes nesta fase da minha vida e que contribuíram de uma forma ou de outra para que o término desta tese fosse possível, um muito obrigado. O meu mais sincero obrigado a todos vós. Desejo que todos continuam com esta presença na minha vida e que acima de tudo todos possam ser muito felizes.

INDEX

FIGURE INDEX

TABLE INDEX

RESUMO

A avaliação da segurança de fármacos, quer durante a fase clínica quer durante as fases iniciais do desenvolvimento de novos fármacos, é um processo fundamental para o despiste de potenciais efeitos tóxicos.

Os fármacos pertencem ao grupo dos xenobióticos - compostos estranhos ao organismo que serão extensivamente metabolizados a fim de diminuir sua toxicidade para o organismo humano. O metabolismo de fármacos envolve vários sistemas enzimáticos, e dois dos mais importantes são os citocromos P450 (CYP), uma família de oxidorredutases hémicas que catalisam vários tipos de reações, e as sulfotransferases (SULT), uma família de enzimas capazes de catalisar a sulfonação de grupos hidroxilo.

Uma das abordagens para prever o metabolismo de fármacos é o desenvolvimento de reatores enzimáticos que permitem replicar o metabolismo *in vivo*. Com esta abordagem em mente, várias proteínas da família dos CYPs e das SULTs humanas foram subclonadas num vetor de expressão estável para permitir a sobreprodução em *Escherichia coli* competente para expressão. O objetivo foi obter as proteínas purificadas para serem utilizadas numa linha de investigação que visa co-imobilizar essas proteínas e desenvolver um reator enzimático *in vitro*.

A isoforma SULT1A1 e as isoformas CYP 2D6 e 3A4 foram clonadas no vetor pET28a (+) por amplificação por PCR dos cDNAs presentes em vetores disponíveis comercialmente usando *primers* desenvolvidos para introduzir locais de corte para enzimas de restrição. A isoforma CYP2C8 foi subclonada a partir do vetor disponível por excisão com enzimas de restrição adequadas e transferida para o vetor pET28a (+). A isoforma SULT1B1 já estava disponível no vetor desejado. Todas as proteínas foram expressas como proteínas de fusão apresentando uma cauda His_6 N-terminal para facilitar a purificação e, a jusante, a imobilização das proteínas.

Todas as proteínas, com exceção do CYP 3A4, foram sobre-expressas com sucesso em *E. coli*, com bandas de absorvância nos cromatogramas de purificação indicando a existência de proteínas marcadas com histidina. No entanto, em alguns casos, a quantidade de proteína produzida era muito baixa. Não foi possível obter um plasmídeo estável com o CYP 3A4, provavelmente porque a baixa especificidade de substrato deste enzima, associada ao comportamento *leaky* característico do promotor T7 presente no vetor pET28a (+), leva a um ambiente tóxico nas células.

PALAVRAS-CHAVE: Proteínas recombinantes; clonagem de cDNA; expressão de proteínas; purificação de proteínas; biologia molecular.

ABSTRACT

Testing drugs to assess drug safety is a key process not only during drug monitoring but also during the initial stages of drugs design. Clinical drugs belong to the group of xenobiotics – foreign compounds that will be extensively metabolized in order to decrease their toxicity to the human organism. Drug metabolism involves various enzymatic systems, and two of the more important ones are cytochromes P450 (CYP), a family of heme oxidoreductases that catalyse various types of reactions, and sulfotransferases (SULT), a family of enzymes able to catalyse the sulfonylation of hydroxyl groups in drugs.

One of the approaches to predict drug metabolism is the development of *in vitro* enzymatic reactors that allow mimicking the *in vivo* metabolism. With this in mind, a set of human CYP and SULT enzymes were sub-cloned into a stable expression vector to allow over-production in expression-competent *Escherichia coli* cells. The goal was to obtain the purified proteins to be used in a research line that aims at co-immobilizing these proteins and develop an *in vitro* enzymatic reactor.

SULT isoform 1A1 and CYP isoforms 2D6 and 3A4 were cloned into pET28a(+) vector by PCR amplification of the cDNAs from commercially available vectors using primers designed to introduce restriction sites for enzymatic digestion. CYP isoform 2C8 was subcloned from the available vector by excision with adequate restriction enzymes and transferred into the pET28a(+) vector. SULT isoform 1B1 was already available in the desired vector. All proteins were expressed as fusion proteins bearing an N-terminal His_6-tag in order to assist in both purification and, downstream, protein immobilization.

All proteins, except for CYP isoform 3A4, were successfully overexpressed in *E. coli*, with absorbance bands in the purification chromatograms indicating the existence of His-tagged proteins. However, in some cases, the amount of produced protein was too low and could not be further used. A stable plasmid with the CYP3A4 could not be obtained, probably because the low substrate specificity of this enzyme, coupled with the characteristic leaky behaviour of the T7 promoter present in the pET28a(+) vector, leads to a toxic environment in the cells.

KEYWORDS: Recombinant proteins; cDNA cloning; protein expression; protein purification; molecular biology.

SYMBOLS AND ABBREVIATIONS

BL21(DE3)	*Escherichia coli* strain widely used for protein expression of T7 RNA Polymerase-based systems.
C41(DE3) **C43(DE3)**	*Escherichia coli* strains derived from the BL21(DE3) strain that contains uncharacterized mutations preventing cell death associated with the expression of toxic recombinant proteins.
cDNA	Coding DNA
CYP	Cytochrome P450
CYP2C8	Unspecific monooxygenase of the cytochrome P450 gene family 2, subfamily C, gene 8. Broad range oxygenase, epoxygenase and N/O-dealkylase. E.C. 1.14.14.1. Systematic name: substrate, NADPH-hemoprotein reductase: oxygen oxidoreductase (RH-hydroxylating or -epoxidizing) (Schomburg, Schomburg, and Chang, 2006).
CYP2D6	Unspecific monooxygenase of the cytochrome P450 gene family 2, subfamily D, gene 6. Broad range oxygenase and epoxygenase. E.C. 1.14.14.1. Systematic name: substrate, NADPH-hemoprotein reductase: oxygen oxidoreductase (RH-hydroxylating or -epoxidizing) (Schomburg, Schomburg, and Chang, 2006) .
CYP3A4	Unspecific monooxygenase of the cytochrome P450 gene family 3, subfamily A, gene 4. Broad range oxygenase and epoxygenase. E.C. 1.14.14.1. Systematic name: substrate, NADPH-hemoprotein reductase: oxygen oxidoreductase (RH-hydroxylating or -epoxidizing) (Schomburg, Schomburg, and Chang, 2006) .
DNA	Deoxyribonucleic acid
E.C.	Enzyme Commission
HindIII	Type II site-specific deoxyribonuclease restriction enzyme purified from *H aemophilus influenzae* strain Rd that cleaves the DNA palindromic sequence 5'-A↓AGCTT-3' (Roy and Smith, 1973a, 1973b). E.C. 3.1.21.4. No systematic name attributed (Schomburg, Schomburg, and Chang, 2013a).
LB	Lysogeny broth culture media, also known as Luria, Lennox and Luria-Bertani broth.
LBAI	LB supplemented with glucose and lactose in order to promote protein expression auto-induction under control of the T7 promoter without the use of other inducers.
NdeI	Type II site-specific deoxyribonuclease restriction enzyme purified from *Neisseria denitrificans* that cleaves the DNA palindromic sequence 5'-CA↓TATG-3' (Watson *et al.*, 1982). E.C. 3.1.21.4. No systematic name attributed (Schomburg, Schomburg, and Chang, 2013a).

NheI	Type II site-specific deoxyribonuclease restriction enzyme purified from *Neisseria mucosa* variant *heidelbergensis* that cleaves the DNA palindromic sequence 5'-G↓CTAGC-3' (Roberts, 1992). E.C. 3.1.21.4. No systematic name attributed (Schomburg, Schomburg, and Chang, 2013a).
One-Fusion™ DNA polymerase	DNA polymerase type II (DNA-directed DNA polymerase) engineered for high processivity and high fidelity. Bioron (Germany) trademark. E.C. 2.7.7.7. Systematic name: deoxynucleoside-triphosphate:DNA deoxynucleotidyltransferase (DNA-directed) (Schomburg, Schomburg, and Chang, 2013a).
PAPS	3-phosphoadenosine-5'-phosphosufate; sulfonyl donor for SULTs
PCR	Polymerase Chain Reaction, a molecular biology method used to produce copies of a template DNA segment.
pDNA	Plasmid DNA
SlaI	Type II site-specific deoxyribonuclease restriction enzyme purified from *Streptomyces lavendulae* that cleaves the DNA palindromic sequence 5'-C↓TCGAG-3'; an isoschizomer of XhoI (Takahashi *et al.*, 1979). E.C. 3.1.21.4. No systematic name attributed (Schomburg, Schomburg, and Chang, 2013a).
SULT	Sulfotransferase
SULT1A1	Broad range aryl sulfotransferase isoform 1A1. E.C. 2.8.2.1. Systematic name: 3'-phosphoadenylyl-sulfate:phenol sulfotransferase (Schomburg, Schomburg, and Chang, 2013a).
SULT1B1	Broad range aryl sulfotransferase isoform 1B1. E.C. 2.8.2.1. Systematic name: 3'-phosphoadenylyl-sulfate:phenol sulfotransferase (Schomburg, Schomburg, and Chang, 2013a).
T4 DNA ligase	DNA ligase isolated from bacteriophage T4 (*Escherichia virus T4*). DNA ligases form two covalent phosphodiester bonds between the 3' hydroxyl end of one nucleotide ("acceptor") and the 5' phosphate end of another ("donor"). E.C. 6.5.1.1; systematic name: poly(deoxyribonucleotide)-3'-hydroxyl:5'-phospho-poly(deoxyribonucleotide) ligase (ATP) (Schomburg, Schomburg, and Chang, 2013b).
XhoI	Type II site-specific deoxyribonuclease restriction enzyme purified from *Xanthomonas holcicola* that cleaves the DNA palindromic sequence 5'-C↓TCGAG-3'; the prototype for SlaI (Gingeras *et al.*, 1978). E.C. 3.1.21.4. No systematic name attributed (Schomburg, Schomburg, and Chang, 2013b).

1. INTRODUCTION

1.1 Genetic Engineering

Since the discovery of DNA, genetic engineering has been an important tool to improve the function of various metabolic and functional processes within an organism of interest. The progress of genetic engineering is dependent on the availability of techniques and methods that expand the complexity of the experiments that may be performed. Over the past 50 years, it has been demonstrated that genetic engineering has grown in a spectacular way, in many laboratories around the world. Nowadays it is a routine practice to isolate a specific DNA fragment from the genome of an organism, determine its base sequence, and assess its function. This technology is used in many applications, as genome mapping and sequencing and biotechnology industry (Nicholl, 2008).

To improve the function of various metabolic process within an organism of interest, it is possible to use genetic engineering technology. The principles of this technology can be used to express non-native genes in a host cell, or an organism of interest, leading to the expression of previously unavailable protein products.

Key among these was the discovery of two types of enzymes that made DNA cloning possible: cloning, in this sense, refers to the isolation and amplification of defined pieces of DNA. One type of these enzymes are the restrictions enzymes, which cut the DNA from any organism at specific sequences of a few nucleotides, generating a reproducible set of fragments. Another type of enzymes are the DNA ligases, enzymes that can covalently join DNA fragments at their termini that have been created by restriction enzymes (Brandenberg *et al.*, 2011).

Recombinant DNA can be introduced in an appropriate cell (host cell), where, from this transformed cell, it is possible to have a descendant that can be called clone, where it is possible to find the same recombinant DNA molecule. When a clone of cells carrying the desired segment of DNA is isolated, an unlimited amount of DNA sequence can be prepared. In case of DNA fragment that contains protein-coding genes, the recombinant DNA molecule introduced in a host cell can direct the expression of these genes, producing proteins within the host cell (Brandenberg *et al.*, 2011).

During the past 50 years, it has become clear that this process has been valuable for the development of basic science research, but nowadays it is possible to see that there are a multitude of factors that must be considered to properly express exogenous genetic constructs. The ability that the organism has to replicate, repair and express their native genetic constructs with a high level of efficiency is one of those factors to be considered. (Barrera-Saldaña, 2012)

Other factors that have been shown to be important when the interest is the expression of a eukaryotic gene in a prokaryotic organism, or vice versa, for example, is the presence or absence of exons, the functionality of polycistronic expression systems, and differences in ribosomal interaction with the gene sequence. (Barrera-Saldaña, 2012)

Moreover, several studies have focused on understanding the biochemical mechanisms responsible for genetic replication and expression, and shown that the expression of non-native constructs can be affected by differences in the codon usage bias of the surrogate versus the native host, as well as discrepancies in the overall GC content of each organism that affect the efficiency of the gene expression and long-term maintenance of the construct, and the mechanism employed by the host to recognize and remove foreign DNA (Barrera-Saldaña, 2012).

1.1.1 Vectors

A vector is a DNA molecule that can carry genetic material into a cell from a suitable host organism, where it can replicate. The use of vectors allows the introduction of foreign DNA into the host in order to make use of the host protein production systems. Vectors should be easily introducible in the host organism where they must replicate, producing copies of itself, with the foreign DNA introduced in it. It must also be easy to isolate the vector from the host cell. Most of those vectors used in molecular biology are based on bacterial plasmids (Green and Sambrook, 2012; Lodish *et al.*, 2000).

They need to have the following characteristics:

- They need to possess an origin of replication (ori), which allows the vectors to replicate independently of the host genome;
- They need to have a site where they can be cleaved by a restriction enzyme, where a foreign DNA can be inserted;
- They need to have a convenient marker (for example, an antibiotic resistance gene) for identifying the host cell contains the vector with the insert DNA of interest.

Also, vectors designed for expression require a promotor that can be engaged by the host expression machinery, ideally with a high and controllable activity; a ribosome binding site (RBS) compatible with the host; a terminator region, in the case of prokaryotic hosts; and, optionally, have encoded tags to ease the post-translational processing, such as stabilization, purification or simply detection (Green and Sambrook, 2012; Barrera-Saldaña, 2012).

1.2 Sulfotransferases

In 1876, Eugen Baumann showed that sulfonate conjugation is an important pathway in the biotransformation. This reaction is responsible for the activation and inactivation of numerous xenobiotics, including therapeutic drugs, environmental toxicants and pro-toxicants. Nowadays, it is possible to affirm that this reaction is catalysed by sulfotransferases (SULTs) (Nimmagadda, Cherala,

and Ghatta, 2006). These enzymes catalyse the transfer of a sulfonyl group from a donor molecule to an acceptor molecule, one of the most important reactions in the metabolism of drugs (Gamage *et al.*, 2006).

In detail, a sulfuryl group from the donor 3-phosphoadenosine-5'-phosphosufate (PAPS) is transferred to an acceptor molecule; the products of the sulfonylation reaction are adenosine 3',5'-diphosphate (PAP) and a sulfuric acid ester. The reactions of sulfonation take place in the cytosol or associated with membranes of the Golgi apparatus of the cell (Nimmagadda, Cherala, and Ghatta, 2006).

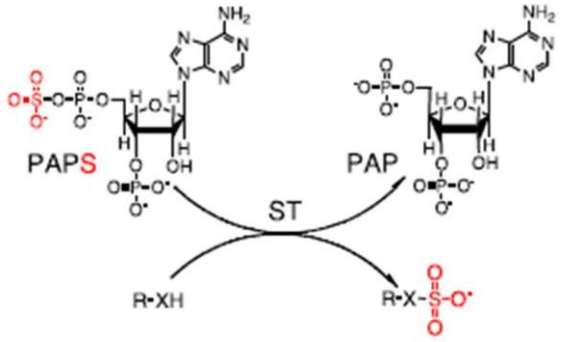

Figure 1 - Schematic representation of a sulfonylation reaction.
ST-catalysed transfer of a sulfonyl group (red) from PAPS donor to R-XH acceptor (R - sugar or peptide or small cytosolic molecules. When X - O the ST is an O-ST and when X-NH/NR the ST is an N-ST. Adapted from Paul *et al.*, 2012

Two different groups of sulfotransferases are known, the membrane-bound SULTs, located in the Golgi apparatus of the cell, responsible for sulfonation of proteins and complex carbohydrates, and cytosolic sulfotransferases, which are responsible for the metabolism of xenobiotics and small endogenous substrates, such as steroids and bile acids (Gamage *et al.*, 2006).

The cytosolic SULTs are divided into two different groups of sulfotransferases, the aryl-sulfotransferases (SULT1 family) and the hydroxysteroid sulfotransferases (SULT2 family). The SULT1 family includes five subfamilies - the 1A, 1B, 1C, 1D and 1E, each with different substrate specificity. When this reaction takes place with xenobiotics and small endogenous substrates, it is considered a detoxification pathway, leading to more water-soluble products and thereby aiding their excretion via the kidney and bile (Paul *et al.*, 2012; Gamage *et al.*, 2006) .

These enzymes exhibit quite broad, overlapping specificities, however, individual enzymes often demonstrate strict regio-specificity toward a substrate. The catalytic mechanism and function of those enzymes were demonstrated after understanding the structural basis of such specificity. It has been demonstrated using crystal structures that those enzymes are globular with a single α/β domain that forms characteristic five-stranded parallel β sheet surrounded on either side with α-helices (Gamage *et al.*, 2006; Paul *et al.*, 2012).

The first identified sulfotransferase was the 1A1 isoform, which is involved in the normal metabolism of a large number of endogenous aromatic substrates, such as estrogenic hormones and

melatonin, and is also the SULT isoform with the largest substrate specificity, making it a very important drug-metabolizing enzyme (Albin *et al.*, 1993; Rohn-Glowacki and Falany, 2014; Aust *et al.*, 2005; Nowell and Falany, 2006).

Another member of the cytosolic SULTs, the SULT1B1 subfamily is responsible for the sulfonation of 3,5,3–triiodothyronine, an important step in thyroid hormone metabolism. In 1995, Sakakibara *et al.* (1995) showed that this enzyme had activity towards various thyroid hormone substrate, pNP and dopamine. Since then this enzyme was isolated and characterised by Fujita *et al.* (1997), showing the major thyroid hormone sulfotransferase, having a slightly higher affinity for the tri-iodothyronine. This enzyme was cloned from a rat liver cDNA library and was described as the dopa/tyrosine sulfotransferase activity towards tyrosine and dopa (Sakakibara *et al.*, 1995; Fujita *et al.*, 1997).

1.3 Cytochromes P450

Cytochrome P450 (CYPs) are a superfamily of heme enzymes found from bacteria to human, that can catalyse the oxidative biotransformation of drugs, involved in the phase I metabolism. This biotransformation can generate some products less active than the parent drug, or in some cases even inactive products. Attention has been focused on human CYPs, because they are involved in the biosynthesis of low-molecular-weight compounds, such as steroids, prostaglandins, thromboxanes, fatty acid derivatives and derivatives of retinoic acid (Anzenbacher and Anzenbacherová, 2001).

In humans, those enzymes are mainly found embedded in the phospholipid bilayer of the endoplasmic reticulum (ER) in the cytosolic surface. CYPs enzymes require a redox partner, cytochrome P450 reductase (CPR) to perform their function. The transference of electrons from NADPH to the heme cofactor of the CYPs is the main role of CPR. They can be found in the human liver and play an important role in phase I drug metabolism. Those metabolites produced by one enzyme can be the substrate for a different isoform of P450, which can be complex because a different oxidative reaction can take place on different attack position (Di Nardo and Gilardi, 2012).

1.3.1 CYP3A4

The 3A4 isoform of CYP 450 is the most abundant one in the human liver, accounting for about 22.1 % of CYP protein content alone. It is one of the CYP isoforms with the broadest substrate range, being involved in the metabolism of a huge number of exogenous drugs, and playing an important role both in drug resistance and inactivation and the onset of adverse drug reactions (Ganesh *et al.*, 2017; Almazroo, Miah, and Venkataramanan, 2017).

1.3.2 CYP 2D6

CYP2D6 is the most popular cytochrome P450 among health professionals. This popularity is because of its genetic polymorphism that causes the presence of three main phenotypes of oxidative metabolism of drug substrate of this enzyme (Anzenbacher and Anzenbacherová, 2001).

The impact of decreased activity of CYP2D6 on drug treatment is extremely important, because elevated levels of tricyclic antidepressants may lead to cardiotoxic effects. Substrates of these enzymes must possess a basic atom of nitrogen that will interact with a carboxyl residue in the active site (Anzenbacher and Anzenbacherová, 2001).

1.3.3 CYP2C8

Thanks to the study of Totah and Rettie (2005), it was recognised the importance of CYP2C8 in drug metabolisms. Those type of enzymes are about 6% of the total liver content. Besides hepatocytes, CYP2C8 was detected also in salivary ducts, intestine, kidney and adrenal cortical cells. The crystal structure of this enzyme was determined to 2.7 Å resolution. A large active site cavity was observed, similar in size to that of CYP3A4, but the shapes of the cavities are considerably different, which explains why CYP2C8 and CYP3A4 share many substrates, but they produce different metabolites (Totah and Rettie, 2005).

In vitro drug interaction studies related to drug metabolism led to the recognition of this enzyme as a major player in the metabolism of several drugs, for example, cerivastatin, repaglinide, rosiglitazone. The importance of CYP2C8 understood thanks to the in vitro studies, and drug interaction studies related to the metabolism of drugs such as cerivastatin (Totah and Rettie, 2005).

2. EXPERIMENTAL

2.1 *Escherichia coli* strains

Escherichia coli (*E. coli*) strain DH5α was used for plasmid maintenance and general molecular cloning procedures. This is a cloning strain that transforms with high efficiency. Like many cloning strains, DH5α has several features that make it useful for recombinant DNA methods. The endA1 mutation inactivates an intracellular endonuclease that degrades plasmid DNA in many miniprep methods (Taylor, Walker, and McInnes, 1993). The hsdR17 mutation eliminates the restriction endonuclease of the EcoKI restriction-modification system, so DNA lacking the EcoKI methylation will not be degraded (Chan *et al.*, 2013). The recA1 mutation eliminates homologous recombination. This makes the strain somewhat sickly but reduces deletion formation and plasmid multimerization (Bryant, 1988).

For protein expression, the *E. coli* BL21(DE3) strain and derived C41(DE3) and C43(DE3) strains were used. The BL21(DE3) strain expresses T7 RNA polymerase, which is encoded by the lambda *DE3* prophage present within the chromosome. The T7 RNA polymerase is expressed from the *lacUV5* promoter, which is less sensitive to catabolite repression than the wt *lac* promoter. These *DE3* strains may exhibit uninduced target protein expression due to the leaky promoter. Any target gene cloned downstream from the T7 promoter is transcribed by the T7 polymerase.

E. coli B strains are deficient in Lon and OmpT proteases, which serve to degrade misfolded proteins and to prevent some cell cycle-specific proteins from accumulating. Cells deficient in both these proteases are much more amenable to the production of proteins from cloned genes. B strains also lack flagella component genes, the DNA cytosine methylase *dcm* and have an additional type II secretion system, used to translocate some proteins across the outer membrane of the cell (Jeong *et al.*, 2009; Studier *et al.*, 2009; Patrick *et al.*, 2010).

OverExpress strains C41(DE3) and C43(DE3) (Lucigen, USA) contain genetic mutations phenotypically selected for conferring toxicity tolerance. The strain C41(DE3) was derived from BL21(DE3) and has at least one uncharacterized mutation that prevents cell death associated with the expression of many toxic recombinant proteins. The strain C43(DE3) was derived from C41(DE3) by selecting a mutant strain for resistance to a different toxic protein. It can express a different set of toxic proteins than C41(DE3) (Studier *et al.*, 1990; Dumon-Seignovert, Cariot, and Vuillard, 2004).

2.2 cDNA source and general cloning strategy

The cDNAs coding the proteins to be overexpressed were obtained as plasmids. pET28a(+)-SULT1B1 (a gift from Cheryl Arrowsmith; Addgene plasmid #25496; http://n2t.net/addgene:25496; RRID:Addgene_25496) and pCW-CYP2C8 (a gift from Joyce Goldstein; Addgene plasmid #69604; http://n2t.net/addgene:69604; RRID:Addgene_69604) were obtained through Addgene, USA (plasmids #25496 and #69604, respectively). pCMV-SPORT6-SULT1A1 (clone HsCD00335548), pCR4-TOPO-CYP3A4 (clone HsCD00341290) and pCR4-TOPO-CYP2D6 (clone HsCD00346008), from the Mammalian Gene Collection (Strausberg *et al.*, 1999; Gerhard *et al.*, 2004), were obtained from the PlasmID Repository at the Dana-Farber/Harvard Cancer Center DNA Resource Core at Harvard Medical School; storage and distribution from these plasmids is funded in part by NCI Cancer Center Support Grant # NIH 5 P30 CA06516.

The pET-28a(+) vector (Novagen, USA) that was used for routine protein expression carries an N-terminal His-Tag/thrombin/T7-tag configuration in addition to an optional C-terminal His-Tag sequence. The T7 expression region is reversed on the circular map due to the sequence numbering from the pBR322 convention. This vector also features a T7 promoter, a multiple cloning site (MCS) and the *lacI* operon, allowing for protein expression mediated by the T7 RNA polymerase under control of the *lacI* repressor. The vector map for pET28a(+) is presented in the Annex (Figure A1). The vector maps for the plasmids acquired with the cDNAs of interest are also given in the Annex – pET28a(+)-SULT1B1 in Figure A2, pCW-CYP2C8 in Figure A3, pCMV-SPORT6-SULT1A1 in Figure A4, pCR4-TOPO-CYP3A4 in Figure A5, and pCR4-TOPO-CYP2D6 in Figure A6. The maps for the prepared vectors are also given in Annex – pET28a(+)-CYP2C8 (Figure A7), pET28a(+)-CYP2D6 (Figure A8), pET28a(+)-CYP3A4 (Figure A9) and pET28a(+)-SULT1A1 (Figure A10).

The pET28a(+)-SULT1B1 vector was used as provided for sulfotransferase 1B1 expression. pCW-CYP2C8 was digested with NdeI and HindIII restriction enzymes to excise the CYP2C8 cDNA, in which the initial methionine was encoded by the ATG sequence of the NdeI cut site, and then inserted into the pET28a(+) vector, thus inserting the N-terminal His-tag fusion in frame with the CYP2C8 cDNA. The remaining plasmids were used for PCR amplification of the cDNA regions using primers that also introduced new restriction sites, allowing the digested PCR products to be cloned into the pET28a(+) plasmid, creating an N-terminal fused His-tag construct.

2.2.1 Restriction enzymes

NdeI, an endonuclease isolated from *Neisseria denitrificans*, was obtained from NEB, USA, or NZYTech, Portugal. HindIII, from *Haemophilus influenzae Rd*, NheI, from *Neisseria mucosa heidelbergensis (ATCC 25999)* and SlaI, an isoschizomer of XhoI, from *Streptomyces lavendulae*, were from Bioron GmbH (Germany). The recognition sequences are given in Table 1.

Table 1 – Recognition sequences of the restriction enzymes used in this work.

Restriction enzyme	Recognition sequence
HindIII	5'-A↓AGCTT-3' 3'-TTCGA↑A-5'
NdeI	5´-CA↓TATG-3´ 3´-GTAT↑AC-5´
NheI	5'-G↓CTAGC-3' 3'-CGATC↑G-5'
XhoI (SlaI)	5'-C↓TCGAG-3' 3'-GAGCT↑C-5'

2.3 pET28a(+) digestion

The pET28a(+) was chosen as the vector to express the recombinant proteins and was digested with the appropriate enzyme pairs. For CYP2C8 subcloning and SULT1A1 cloning, NdeI and HindIII were used; for CYP2D6 cloning, NheI and HindIII were used, and CYP3A4 was tentatively cloned using the NdeI and XhoI(SlaI) restriction enzymes.

2.3.1 NdeI/HindIII pET28a(+) digestion (for CYP2C8 subcloning and SULT1A1 cloning)

The pET28a(+) plasmid was digested with NdeI and HindIII restriction enzymes, according to the standard manufacturer instructions. Briefly, 4.51 µg of pET28a(+) (200 µL) were digested with 5 µL NdeI restriction enzyme (NZYTech, Portugal) in a reaction with 25 µL of Buffer (10x) and 20 µL of deionized water, in a final volume of 250 µL. Plasmid digestion proceeded at 37 °C, 150 rpm, for 1 hour, followed by enzyme inactivation at 65 °C for 20 minutes.

The digested DNA was purified using the Clean-Easy PCR Purification Kit (Canvax, Spain). Briefly, 195 µL of the reaction were diluted with 5 volumes of plasmid binding PB buffer and loaded into the column and centrifuged at 13 000 x *g* for 1 minute. The flowthrough was discarded, and the column was washed with 700 µL of washing PE buffer by centrifugation at 13 000 x *g* for 1 minute. A second centrifugation at 13 000 x *g* for 1 minute was performed to remove all traces of PE buffer. The column was transferred into a collection tube and the plasmid was eluted by adding 30 µL of elution EB buffer onto the membrane, incubated at room temperature for 1 minute, and centrifuged at 13 000 x *g* for 1 minute. The digested and clean plasmid DNA was quantified by measuring its absorbance at 260 nm, assuming that Abs (260 nm) of 1 corresponds to a concentration of dsDNA of 50 ng/µL (Green and Sambrook, 2012).

The NdeI-digested DNA was further digested with HindIII enzyme. Briefly, 2.25 µg of plasmid DNA (100 µL) were digested with 10 U (5 µL) of Hind III restriction enzyme (Bioron, Germany) in a reaction with 15 µL of 10x reaction buffer (10 mM Tris-HCl pH 7.9, 50 mM NaCl, 10 mM MgCl2, 1 mM dithiothreitol), and adding PCR-grade water to a final reaction volume of 150 µL. The digestion was performed at 37 °C, 150 rpm, for 1 hour, followed by enzyme inactivation at 65 °C for 20 min. After inactivation, the digested DNA was cleaned using the Clean-Easy PCR Purification Kit (Canvax, Spain) as described above. This NdeI/HindIII digested pET28a(+) was quantified and stored at -20 °C for further use.

2.3.2 NheI/HindIII pET28a (+) digestion (for CYP2D6 subcloning)

The pET28a(+) plasmid was digested with NheI and HindIII restriction enzymes, according to the manufacturer instructions. Briefly, 6.8 µg of pET28a were double-digested with 150 U (15 µL) of Nhe I and with 15 µL of HindIII restriction enzymes (Bioron, Germany) in a reaction with 31 µL of 10x reaction buffer K (20 mM Tris-acetate pH 7.9, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM dithiothreitol), adding PCR-grade water to a final reaction volume of 310 µL. Plasmid digestion proceeded at 37 °C for 15 minutes, followed by heat inactivation at 65 °C for 20 minutes. After the digestion, the digested DNA was purified as described above, using the Clean-Easy PCR Purification Kit, and quantified by measuring its absorbance at 260 nm. This NheI/HindIII digested pET28a(+) was stored at -20 °C for further use.

2.3.3 NdeI/XhoI pET28a (+) digestion (for CYP3A4 subcloning)

To prepare digested pET28a(+) for CYP3A4 subcloning, the plasmid was first digested with NdeI (NEB, USA), following the manufacturer protocol. Briefly, 10 µg of DNA was digested with 200 units (20 µL) of NdeI with 30 µL of 10x CutSmart buffer (20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, and 100 µg/ml BSA, pH 7.9), adding nuclease-free water to a final reaction volume of 300 µL. Upon 2h30 minutes incubation at 37 °C, the enzyme was heat inactivated at 65 °C for 20 minutes.

The NdeI digested plasmid was purified using the Clean-Easy PCR Purification Kit, quantified by measuring its absorbance at 260 nm, and further digested in two separate tubes with XhoI (SlaI) (Bioron, Germany), following the manufacturer protocols. Briefly, ~4.5 µg of plasmid were digested with 10 U (5 µL) of XhoI (SlaI) with 15 µL of 10x reaction buffer K (10 mM Tris-HCl pH 7.9), 150 mM NaCl, 10 mM MgCl2, 1 mM dithiothreitol, BSA 100 1 mg/mL), adding nuclease-free water to a final reaction volume of 150 µL. Both reaction mixtures were incubated for 2 h at 37 °C and the enzyme was heat inactivated at 65 °C for 20 minutes. Upon purification with the Clean-Easy PCR Purification Kit and quantification, this NdeI/XhoI digested plasmid was stored at -20 °C until further use.

2.3.4 pCW-CYP2C8 digestion

The cDNA for CYP2C8 was excised from the pCW plasmid by digestion of 6.5 µg of DNA using the NdeI and HindIII enzymes, following the same protocol used for pET28a(+) digestion with the same enzymes. All the reaction was loaded on a 1% agarose gel and the band corresponding to the cDNA insert *#1.6 kb(was cut and weighted. The DNA fragment was purified from the agarose slices Clean-Easy Agarose Purification Kit, as previously described, quantified and stored at -20 °C.

2.4 PCR amplification of the cDNAs for SULT1A1, CYP3A4 and CYP2D6

The cDNAs for SULT1A1, CYP3A4 and CYP2D6 were PCR amplified from the acquired plasmids using the primers presented in Table 1. These primers were designed to amplify the region of interest and simultaneously insert the restriction sites required for further digestion, as indicated in the table. Temperatures for the annealing step in the PCR program were computed using the “Tm calculator”

from ThermoFisher Scientific, based on the thermodynamic method of Tm estimation (Allawi and SantaLucia, 1997).

Table 2 – List of the primers used in this work.
CYP – cytochrome P450; SULT – sulfotransferase; nt – nucleotide; Tm – annealing temperature.

Gene	Primer	Restriction site	Sequence (5'□3')	Length (nt)	T_m
SULT1A1	forward	NdeI	GGCCAGGCATATGCAGCTGATCCAGGACACCTC	33	72.0 °C
SULT1A1	reverse	HindIII	CGTGAGAAGCTTGGTCAGGTTTGATTCGCACAC	33	
CYP2D6	forward	NheI	GTGAGGCAGCTAGCGGGCTAGAAGCACTG	29	72.0 °C
CYP2D6	reverse	HindIII	CCCGCCAAGCTTTTCCCAGTCACGACG	27	
CYP3A4	forward	NdeI	ACAGTACATATGGCTCTCATCCCAGACTTGG	31	71.4 °C
CYP3A4	reverse	XhoI	ACTCTCGAGAGGGCGAATTGAATTTAGCGGC	31	

PCR amplification was performed using the One-Fusion high-speed-fidelity Polymerase (Geneon, Germany), according to the manufacturer instructions. Briefly, reactions were performed using the 2.5x One-Fusion buffer (20 µL), 1 µL of 10 mM mix of DNTPs (Bioron, Germany),10 µM of each primer, 50 ng of the template plasmid, and 2 U (1 µL) of One-Fusion polymerase, in a total reaction volume of 50 µL.

PCR amplification was performed on a Prime Thermal Cycler (Techne, UK) with the following program - an initial denaturation at 98 °C for 3 min was followed by 35 amplification cycles, consisting of 3 steps each: denaturation at 98 °C for 10 s, annealing for 20 s at a gradient of temperatures ranging from 58 to 70 °C, and extension at 72 °C for 45 s. A final extension step was performed at 72 °C for 3 min.

The PCR reactions were analysed by agarose gel electrophoresis. The gels contained 1% agarose (Carl Roth, Germany) dissolved in 1x Tris-Acetate-EDTA (TAE) buffer. The samples were prepared by mixing 5 µL of PCR reaction with 2 µL of RUNSAFE loading buffer with incorporated DNA staining dye (Cleaver Scientific, UK). The DNA Ladder (1 kbp DNA Leiter, Carl Roth, Germany) was also mixed with RUNSAFE in a proportion of 5 µL to 2 µL and ran in parallel to the samples. Electrophoresis was carried out in 1 x TAE buffer at 80V for 60 minutes with no light exposure. Visualization of DNA bands was performed by exposure to UV-light using PhotoDoc-It Imaging System Benchtop 2 UV™ Transilluminator (UVP Inc., USA).

PCR products from 3 to 5 reactions were pooled and purified using the Clean-Easy PCR Purification Kit (Canvax, Spain) as described above and quantified by measuring their absorbance at 260 nm.

These purified PCR products were then digested with the appropriate restriction enzymes, following the same procedure as described above for plasmid digestion with each enzyme pair. SULT1A1 PCR products were digested with NdeI and HindIII, CYP2D6 PCR products were digeted with NheI and HindIII, and CYP3A4 PCR products were digested with NdeI and XhoI. Digestion products were purified with the Clean-Easy™ PCR Purification Kit, as previously described, quantified and stored at -20 °C.

2.5 pDNA ligation

To perform the ligation reactions using the digested pET28a(+) plasmid and the fragments containing the cDNAs of interest, the T4 DNA ligase (Bioron, Germany) was used according to the manufacturer protocol. Briefly, 70 ng of linear vector DNA (NdeI/HindIII -pET28) were mixed with 99ng of DNA insert (NdeI/HindIII - CYP2C8), ensuring and insert vector ratio of 5. This mixture was incubated with 2 µL of T4 DNA ligase buffer (10x) and 100U (1µL) T4 DNA ligase, in a total reaction of 20µL. To perform the ligation of pET28a(+)-3A4, 83.25 ng of linear vector DNA (NdeI/XhoI – pET28) were mixed with 246.6 ng of DNA insert (NdeI/XhoI – CYP3A4), ensuring an insert: vector ratio of 9.55. This mixture was incubated with 2 µL of 10x T4 DNA ligase buffer and 100 U (1 µL) of T4 DNA ligase, in a total reaction volume of 20 µL with nuclease-free water. With the same aim, 84.13 ng of linear vector DNA (NdeI/HindIII – pET28) were mixed with 888 ng of DNA insert (NdeI/HindIII -CYP2D6), ensuring an insert: vector ratio of 31.3. This mixture was incubated with 2µL of 10x T4 DNA ligase buffer and 100 U (1µL) of T4 DNA ligase, in a total reaction of 20 µL. All reactions were incubated at 16 °C overnight and heat-inactivated at 65 °C for 15 min. This mixture was stored at -20 ° for further use in cell transformation.

2.6 General pDNA extraction protocol

When necessary, plasmid DNA was extracted and isolated from bacterial cultures using the Illustra Plasmid Miniprep Spin Kit (GE Healthcare, UK). Briefly, 3 mL of overnight grown culture was centrifuged for 5 minutes at 16 000 x g at 4 °C; the supernatant was discarded, and the cells were vortex-resuspended in 175 µL of lysis buffer type 7; alkaline lysis was performed by adding 175 µL solution of lysis buffer type 8, with mix by gentle inversion, and left at room temperature for no more than 5 min. The solution was neutralized by adding 350 µL of lysis buffer type 9, followed by gentle inversion. The reaction mixture was loaded into a spin-column, and centrifuged for 4 min at 4 °C at 16 000 x *g*. The flow-through was discarded, and the column was washed with 400 µL of wash buffer type 1 and 1 min centrifugation at 16 000 x *g* at 4 °C. The column was transferred to a new microtube and the plasmid DNA was eluted by applying 50 µL of elution buffer type 4 on the membrane, incubating for 1 min at room temperature, and centrifuging for 30 seconds at 16 000 x *g* at 4 °C.

2.7 Preparation of *E. coli* competent cells for transformation

Cells competent for transformation were prepared using the classic calcium chloride method. Briefly, *E. coli* cells were inoculated from a solid LB plate into 5 mL of fresh LB medium (LB Broth Miller, Fisher Bioreagents) and incubated in an orbital incubator (Orbital Shaker-Incubator ES-20, Grant - Bio) at 37 °C, 220 rpm for 16 h. After the incubation time, 1.5 mL aliquots of culture were transferred to different microtubes and centrifuged (Gyrozen 1750 R) for 10 min at 16 000 x *g* at 4 °C. The supernatant was discarded, and the pellet was re-suspended in 750 µL of ice-cold 0.1 M $CaCl_2$. The re-suspended cells were kept on ice for 1 hour and were then collected by centrifugation as before. The cells were then re-suspended in 50 µL of ice-cold 0.1 M $CaCl_2$ and kept on ice at 4 °C overnight until being

transformed in the following day. When these cells were not to be used in the next day, the last re-suspension step was performed using 50 µL of ice-cold 0.1 M $CaCl_2$: glycerol 50:50 v/v mixture, and the cells were stored at -80 °C.

2.8 General transformation protocol

To perform the transformation process, *E. coli* competent cells were prepared as described above and thawed on ice. 50 µL of competent cells were mixed with 20 µL of the ligation reactions or with 5 µL of isolated pET28a+CYP2C8, pET28a+CYP2D6, pET28a+CYP3A4, pET28a++SULT1B1 or pET28a+SULT1A1. The bottom of each microtube was flicked gently for 10 seconds to ensure a proper mixing of the plasmid DNA and the cells.

The mixture of competent cells and plasmid DNA was placed on ice for 30 minutes. A heat shock was applied to the microtube at 42 °C for 45 seconds. The tubes were placed back on ice for 2 minutes, and 900 µL of SOC medium (Super Optimal broth with Catabolite repression, from NZYTech, Portugal) were added. After one-hour incubation at 37 °C at 220 rpm, cells were collected by centrifugation for 10 min at 16 000 x *g*, and 850 µL of the supernatant were discarded. The cells were re-suspended in the remaining 100 µL of supernantant, and were plated in LB agar plates, supplemented with kanamycin at a final concentration of 30µg/µl; 25 µL of the cell suspension were used to inoculate one plate, and the remaining 75 µL were used to inoculate another plate. All plates were incubated overnight at 37 °C.

2.9 Transformant selection by functional assay

Upon overnight incubation, each colony on the DH5α transformation plates was individually picked, inoculated into 5 mL of LB medium supplemented with kanamycin at a final concentration of 30µg/µL (LBK30), and streaked on a LB Agar plate; the liquid cultures were grown for 16 h at 37 °C with agitation at 250 rpm.

Using these overnight growths, plasmid was extracted from each and transformed into *E. coli* BL21(DE3) competent cells, as previously described; upon plating and overnight growth at 37 °C, cells were grown in LBK30 media until an OD of *ca*. 0.6 at 600 nm was obtained, and then IPTG was added, to obtain a final 1 mM concentration, to induce protein expression. Cells were harvested 4 hours upon induction and re-suspended in 1 mL of water. 20 µL of these suspensions were mixed with 5 µL of SDS-PAGE loading buffer and prepared for electrophoresis for analysis of the total protein expression pattern (see below for details).

2.10 Routine protein expression

To express protein (CYP-2C8, SULT1B1 and CYP-2D6), *E. coli* (DE3 strains) transformed with different plasmids (pET28a(+)-CYP2C8, pET28a(+)-CYP2D6, pET28a(+)-SULT1A1 and pET28-SULT1B1) were picked from Petri dishes stored at 4 °C. In each Petri dish, different colonies were selected and

they were inoculated in three different 50 mL falcon tubes with 15 mL sterile LB supplemented with kanamycin at a final concentration of 30 μg/mL. Those falcon tubes were incubated at 37 °C, 220 rpm, overnight, and used to inoculate the expression assays.

Cytochrome P450 isoforms expression was performed by growing cells in LB medium, supplemented with kanamycin (final concentration of 30 μg/mL), and expression was induced when the culture reached an OD of 0.6 by adding IPTG to a final concentration of 1 mM. Sulfotransferases 1A1 and 1B1 were expressed by growing cells in LB auto-idnuction medium (LB medium supplemented with D-glucose and α-lactose at final concentrations of 0.50 and 2.00 g/L) with kanamycin 30 μg/mL (Studier *et al.*, 1990).

From each falcon tube of the overnight growth, 1 mL was collected and transferred to different identified flask shots (1000 mL) with 750 mL sterile LB medium with kanamycine (30 μg/mL). To express (SULT1B1), 750 mL of LBAIK30 was inoculated with 1 mL of BL21-pET28-SULT1B1 and the culture grow for 24 hours at 37 °C. After 16 hours of grow was inoculated in 750 mL of LBAIK30 24h a 37hose flasks shot inoculated with different culture (BL21-pET28-SULT1B1, pET28-CYP2C8 and pET28-CYP2D6) were incubated at 37 °C, 220 rpm for 24 hours.

After the incubation time, the culture of BL21(pET28-SULT1B1) was induced with IPTG (1 mM) and after induction incubated again at 37 °C, 220 rpm, for 6 hours. After the expression, the culture was collected in 50 mL falcon tubes and centrifuged at 4000 x *g*, 10 minutes. The supernatant was discarded, and the pellet stored at -20 °C for further cell lysis.

After 24 hours of growth, the two cultures (pET28-CYP2C8 and pET28-CYP2D6) were induced with IPTG (1 mM) and supplemented with δ-aminoevulinic acid (ALA, 5-amino-4-oxo-pentanoic acid) (1mM), incubated at 37 °C, 220 rpm, for 24 hours. The cultures were collected in 50 mL falcon tubes and centrifuged at 4000 x *g*, 10 minutes. After the centrifugation, the supernatant was discarded, and the pellet stored at -20 °C until further use.

2.11 Cell Lysis

To assess the induction of protein expression (CYP2C8, SULT1B1 and CYP2D6), all the pellets from liquid cultures were analysed in SDS-PAGE. All the pellet were weight and the weigh registered to know the volume of Bacterial Cell Lysis Buffer (NZYTech, Portugal) to apply to each pellet.

After adding the volume of Bacterial Cell Lysis Buffer, 2 μL of lysozyme 5 mg/mL and RNAse 20 mg/mL were added to each; the tubes were incubated at room temperature with shaking at 70 rpm for 15 minutes. The lysates were centrifuged at 4000 x *g* for 10 min, and stored at 4 °C until use.

2.12 General protein electrophoresis under denaturing conditions (SDS-PAGE)

To assess the induction of protein expression, pellets obtained from cell cultures after induction were analysed by SDS-PAGE followed by protein staining (Shapiro, Viñuela, and Maizel, 1967).

The lysate pellets were resuspended in 200 µL of dH_2O followed by vortexing. Then, 40 µL of cell suspension was mixed with 10 µL of 5x loading buffer (NZYTech, Portugal) and boiled for 5 minutes at 100ºC. The mixture was loaded on a 12% SDS bisacrylamide gel, using a stacking 4% gel. The molecular weight markers (11 to 245 kDa and 17 to 225 kDa) used was Protein Marker II (NZYTech, Portugal) or Amersham™ ECL™ Rainbow™ – Full range RPN800E (GE Healthcare, UK), respectively. The electrophoresis was run at 100V for 30 minutes and then 150V for 60 minutes. The gel was stained with EzBlue™ Staining Reagent (Sigma-Aldrich, MO, USA) at RT o/n, with mild agitation and destained with water.

2.13 Protein purification by affinity chromatography

In order to purify the expressed proteins, a nickel-loaded Sepharose chromatography column (His Trap TM FF Crude – GE Healthcare, UK). was used. Briefly, the cell lysates were diluted 4 fold with binding buffer (0.5 M NaCl in 20 mM sodium phosphate buffer at pH 7.4) and applied onto a column pre-washed with 5 volumes of binding buffer. Each sample (cell lysate) was loaded on the column and it was washed with 6 column volumes of binding buffer. The elution of each protein was performed using a 20 mM sodium phosphate pH 7.4 elution buffer supplemented with 0.5 M NaCl with increasing imidazole concentration, from 5 mM to 500 mM. The column was washed with a further 10 volumes of 500 mM imidazole elution buffer and 10 volumes of distilled water and stored in ethanol 20%(v/v). The protein content in the eluted fractions was estimated by measuring the absorbance at 280 nm of each fraction.

To improve protein stability, and upon a preliminary test purification performed as described to determine the correct imidazole concentration required to elute each column, the specific elution buffers for each protein were supplemented with 5% (v/v) glycerol, 0.025 % (v/v) Tween-20 and 0.1 mM dithiothreitol.

2.14 Protein quantification

To quantify the amount of each protein, a simple and accurate procedure was used, performing the Bio-Rad Protein Assay (Bio-Rad Laboratories, USA) based on Bradford method and following the standard procedure according to manufacturer's instructions. This method measures the binding of Coomassie brilliant blue to the protein (Bradford, 1976). The reagent used in this assay was mixed with the samples (5 mL reagent and 100 µL of sample). After proper mixing, the samples were incubated at room temperature for 5 minutes at room temperature and absorbance at 595 nm was determined. A standard curve was made in the same conditions of the experiments using bovine serum albumin (BSA) standards.

3. RESULTS AND DISCUSSION

3.1 Cloning

To clone human cDNAs fused with N-terminal 6xHis-Tag (CYP2C8, CYP2D6, CYP3A4 and SULT1B1) in expression vector pET28a (+), the fragment cDNAs were digested with different pair of enzymes: SULT1A1 digested with NdeI/HindIII, CYP2D6 digested with NheI/HindIII, CYP2C8 digested with NdeI/HindIII and CYP3A4 digested with NdeI/XhoI. pET28a (+) was digested with the same pairs of enzymes: NdeI/HindIII, NheI/HindIII, NdeI/HindIII and NdeI/XhoI.

All the fragments digested with those enzymes were cleaned using the Clean-Easy™ PCR Purification Kit and DNA fragments quantified; the quantification results are shown in Table 3.

Table 3 - DNA concentration obtained in each digestion.

DNA digests	[DNA]ng/µL
pET28 (NheI/HindIII)	37.65
pET28 (NdeI/HindIII)	10.7
pET28 (NdeI/XhoI)	16.65
SULT1A1 (NdeI/HindIII)	36.5
CYP2D6 (NheI/HindIII)	61.25
CYP2C8 (NdeI/HindIII)	9.00
CYP3A4 (NdeI/XhoI)	20.55

While the CYP2C8 cDNA was obtained directly from the pCW-CYP2C8 vector, by excising it with restriction enzymes, all the other cDNAs were obtained by PCR amplification which required procedure optimization.

To amplify SULT1A1 cDNA two sets of PCR reaction mixtures (with and without 1mM Mg^{2+} supplementation) were prepared. The optimal annealing temperature of these primers was assayed in a range of 60ºC – 70ºC and the extension time was 25s. All reactions were successful, with only one band of approximately 960 bp having been amplified as seen by gel electrophoresis (Figure 2, panel A). Higher yields were obtained with the higher annealing temperatures tested and as there are no major differences between the two groups of reactions, it indicates that for this gene Mg^{2+} supplementation is not essential.

To continue the cloning process, all PCR reactions were pooled together and used to prepare the Nde*I* /Hind*III* digested insert later used in a ligation reaction containing 70ng of the vector and an insert-vector ratio of 28. The transformation into *E. coli* DH5α competent cells was not very successful, as only one colony was obtained. However, after extraction and purification of the plasmid from these cells, it was confirmed by PCR amplification of the insert, to contain the SULT1A1 insert of approximately 960 bp (Figure 2, panel B).

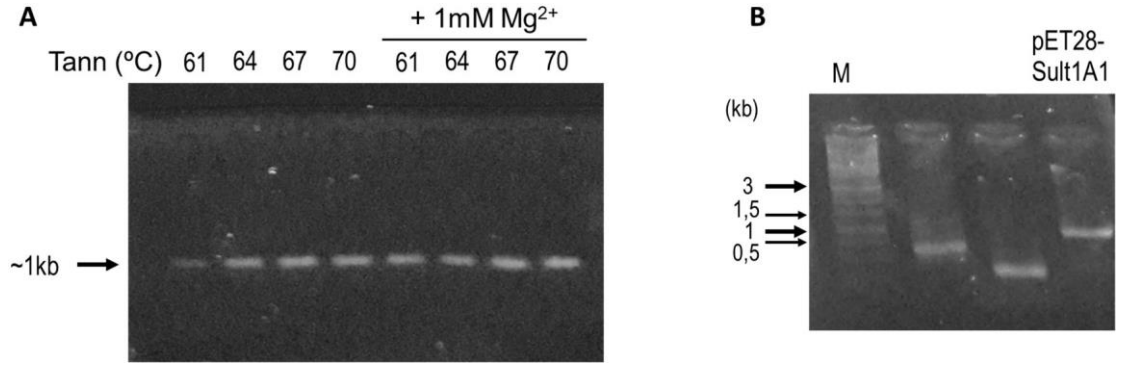

Figure 2 – PCR amplification and cloning of the SULT1A1 cDNA into pET28a(+).
(A) Agarose gel electrophoresis of the PCR reactions for the amplification of the SULT1A1 cDNA from pCMV-Sport6-SULT1A1 were carried out without (lanes 1 – 4) and with 1mM magnesium chloride supplementation (lanes 5 – 8). The PCR program was: 3min 98ºC; 35x [30s 98ºC; 20s Tann (as indicated for each lane); 35s 72ºC]; 3 min 72ºC. (B) Agarose gel electrophoresis of the PCR reaction for the amplification of the SULT1A1 cDNA using as template purified pET28-SULT1A1. Lane M contains the molecular marker 1kb DNA ladder (Carl-Roth). The arrows indicate the position and size of marker bands, with thickness that mimics the intensity of the bands of the DNA ladder. Lane 2 and 3 contain other nonrelated PCR products.

Amplification of the CYP3A4 and CYP2D6 cDNAs required further optimizations when compared to that of SULT1A1. Initially, and for each gene, two sets of PCR reaction mixtures (with and without 1mM Mg^{2+} supplementation) were prepared with 50 ng of the template plasmid DNA in each reaction. Since it was seen for SULT1A1 that higher annealing temperatures (closer to the *in-silico* predictions) gave better yields, the optimal annealing temperature for each pair of primers was assayed in a range of 66ºC – 70ºC. The extension time was set to 30s as these products are larger than SULT1A1.

In the set of reactions done with 1mM Mg^{2+} supplementation, analysed in the gel electrophoresis presented in Figure 3, panel A, amplification of some products of specific size was observed, but none were of the estimated product size (1,6 kb). Moreover, the presence of very large products at more stringent temperatures suggests an excess of template DNA.

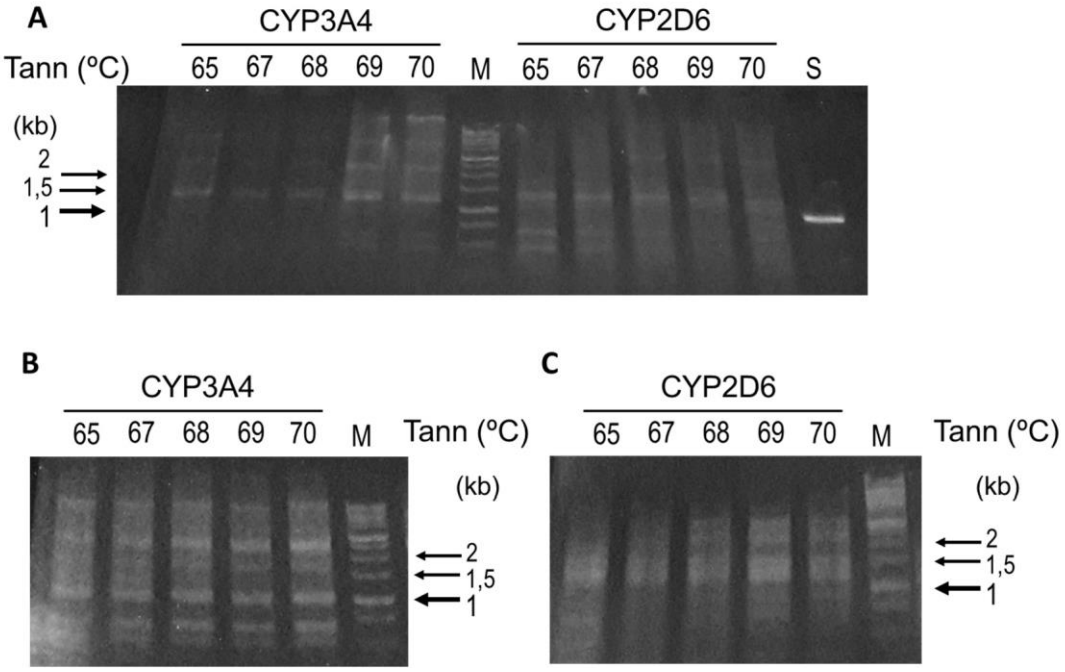

Figure 3 – PCR amplification CYP3A4 and CYP2D6 cDNAs for cloning into pET28a(+).
(A) Agarose gel electrophoresis of the PCR reactions for the amplification of the CYP3A4 (lanes 1 – 4) and CYP2D6 (lanes 6 – 9) cDNAs carried out with 1mM magnesium chloride supplementation, using the indicated annealing temperatures. (B) Agarose gel electrophoresis of the PCR reactions for the amplification of the CYP3A4 cDNA carried out with 1mM magnesium chloride supplementation, using the indicated annealing temperatures. (C) Agarose gel electrophoresis of the PCR reactions for the amplification of the CYP2D6 cDNA carried out with 1mM magnesium chloride supplementation, using the indicated annealing temperatures. In all panels, Lanes M contain the marker 1kb DNA ladder (Carl-Roth). The arrows indicate the position and size of relevant marker bands, with thickness that mimics the intensity of the bands of the DNA ladder. For all reactions the PCR program was: 3min 98ºC; 35x [30s 98ºC; 20s Tann (as indicated for each lane); 40s 72ºC]; 3 min 72ºC. Reactions were done with approximately 50 ng of pCR4-TOPO-CYP3A4 and pCR4-TOPO-CYP3A4 as template, respectively.

Gel electrophoresis analysis of the reactions for CYP3A4, done without the Mg^{2+} supplement, presented in Figure 3, panel B, revealed a large array of non-specific products as previously seen when Magnesium was added. However, for those reactions done at higher temperatures, a product with a size between 1,5 and 2kb is seen, suggesting the amplification of the desired product.

The CYP2D6 PCR reactions without the Mg^{2+} supplement were also analysed by gel electrophoresis, presented in Figure 3, panel C. However for this gene the reactions were unsuccessful and besides multiple non-specific products and smearing, there was no apparent amplification of a product of 1,6 kb with any of the annealing temperatures tested.

For CYP3A4, two new PCR reactions were done without adding magnesium, decreasing the amount of template DNA to 25ng. The amplification program was also altered, increasing the extension time to 45s, as Phusion DNA polymerase amplifies 1kb in 15 - 30s, and using as annealing temperatures 63ºC and 67ºC. The results were analysed by gel electrophoresis (Figure 4, panel A). The reaction with 63ºC as annealing temperature gave only two products, with similar intensities – one consistent with the desired product of ~1,6kb, and a larger one of ~2,5 kb. A higher annealing temperature did not decrease the formation of the non-specific larger product and had in fact less amount of the 1.6kb product and other non-specific products of smaller size.

To produce the insert for CYP3A4 cloning, 8 new reactions were done replicating the previous conditions and using 63ºC as annealing temperature. The reactions were analysed in gel, to confirm a similar result (data not shown) and the band corresponding to the desired CYP3A4 cDNA amplification was cut from a gel a purified with a gel extraction and DNA purification kit. The product was then digested with Nde*I* and Xho*I* and purified at a concentration of 20.6 ng/µL. The ligation reaction was done with insert-vector ratio of ~10 and transformed into *E. coli* DH5α competent cells, but no colonies were obtained. Despite several attempts, that included increasing the volume of the ligation reaction and insert-vector molar ratio up to 16, and that resulted in several colonies appearing in the transformation plates, no positive clone was obtained as confirmed by PCR colony analysis.

For the amplification of the CYP2D6 cDNA, other PCR attempts were made: first, two reactions mimicking the ones for CYP3A4 were done, but no amplification of a product of the right size was seen (data not shown). Then, further PCR reactions reducing the amount of template DNA 5 ng and decreasing significantly the annealing temperature (ranging from 58ºC to 65ºC) were performed but without success (data not shown). Lastly, a set of reactions without added Mg^{2+}, using 25 ng of template plasmid and a PCR program with 40s only of amplification time and annealing temperatures ranging from 56 to 70 ºC was attempted. As can be seen in Figure 4, panel B, one of the reactions (with the annealing temperature of 64ºC) gave a single PCR product with a size consistent with that of the desired product (1.6 kb). The reactions with higher annealing temperatures were inconclusive as a problem occurred in the right side of the gel, affecting the migration. However, with these optimal conditions, six new PCR reactions were done, pooled together, purified and digested with Nhe*I* and Hind*III*. The resulting insert was purified at 61.2ng/µL and a ligation reaction with 60 ng of vector and an insert-vector molar ratio of 32 was prepared and transformed into *E. coli* DH5α competent cells. Of the 40 colonies obtained, twelve randomly chosen were analysed by colony PCR using the primers T7 promoter and T7 terminator that anneal in the pET28a(+) plasmid flanking the multiple cloning site region. Of these, clone #16 was positive showing a PCR product with the correct size (data not shown).

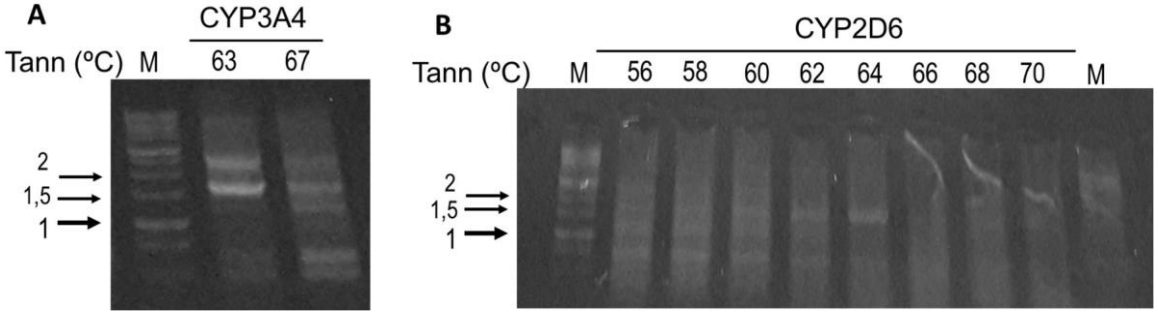

Figure 4 – PCR amplification CYP3A4 and CYP2D6 cDNAs for cloning into pET28a(+)
(A) Agarose gel electrophoresis of the PCR reactions for the amplification of the CYP3A4 cDNA carried out with 25 ng of pCR4-TOPO-CYP3A4 as template DNA and using the PCR program: 3min 98ºC; 35x [30s 98ºC; 20s Tann (as indicated for each lane); 45s 72ºC]; 3 min 72ºC. (B) Agarose gel electrophoresis of the PCR reactions for the amplification of the CYP2D6 cDNA carried out with 25 ng of pCR4-TOPO-CYP2D6 and using the PCR program: 3min 98ºC; 35x [30s 98ºC; 20s Tann (as indicated for each lane); 40s 72ºC]; 3 min 72ºC. In all panels, Lanes M contain the marker 1kb DNA ladder (Carl-Roth). The arrows indicate the position and size of relevant marker bands, with thickness that mimics the intensity of the bands of the DNA ladder.

3.1 Selection of growth media for protein expression

One of the most common growth media used for *E. coli* growth is the LB medium, a nutritionally rich medium. For expression of proteins coded in pET28a(+) plasmid under control of the *lac* operon, the typical protocol is growing the culture until a pre-specified OD is achieved and then inducing the expression of the protein by adding IPTG to the media; IPTG, isopropyl β-D-1-thiogalactopyranoside, is a surrogate for the lactose metabolite allolactose; like allolactose, IPTG binds the lac repressor, a protein that binds the *lac* operon, causing it to release the DNA via an allosteric-induced conformational change. Upon repressor release, transcription of the plasmid DNA can proceed, ultimately leading to protein expression. (Green and Sambrook, 2012; Brandenberg *et al.*, 2011; Nicholl, 2008; Studier *et al.*, 2009; Studier *et al.*, 1990)

An alternative approach is growing the transformed cells in glucose and lactose supplemented growth media. Under these conditions, *E. coli* will use glucose as the primary carbon and energy source. When glucose is completely depleted, the bacterial metabolism will shift to use lactose as a carbon source, leading to the production of allolactose, which will act as a transcription repression releaser. This is called an auto-induction media, that does not require any other intervention other than the initial setup and is also economically more advantageous. (Studier *et al.*, 2009; Studier *et al.*, 1990; Fox and Blommel, 2009)

In order to choose which of the media better suit our goals, transformed cells were grown in both LB media and in LB auto-induction media (LBAI). Protein expression patterns were analysed by lysing the whole cells and running the cellular content in an SDS-PAGE gel, in order to check for protein over-expression. Sulfotransferase 1B1 expression was analysed in LB media at 37 °C and in LBAI media at both 37 °C and at room temperature. The SDS-PAGE gel obtained is presented in Figure 5.

Analysing Figure 5, it is possible to observe that in all the lanes that have expression induction with IPTG (1 mM) leads to the appearance of a broader band, although faint, at around 35 kDa, in agreement with the known molecular weight of SULT1B1, 34.9 kDa (Fujita *et al.*, 1997). A corresponding band is much more intense in lanes 4 and 6, corresponding to the use of auto-induction media for a 24 h growth at 37 °C.

These results indicate that the best conditions for SULT1B1 expression are the use of the auto-induction LB-based media, with a growth duration of 24 h at 37 °C.

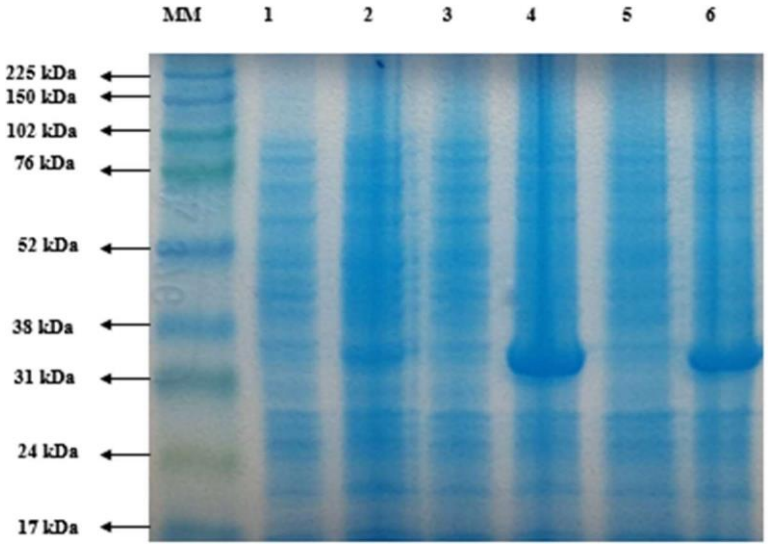

Figure 5 - SDS-PAGE analysis of SULT1B1 protein expression by *E. coli* BL21(DE3).
E.coli BL21(DE3) cells transformed with the pET28a(+)-SULT1B1 plasmid were grown in different conditions. All cultures were grown with media supplemented with kanamycin at a final 30 µg.mL-1 concentration. Lane MM displays the molecular weight marker AmershamTM ECL Rainbow Marker – Full Range RPN800E (GE Healthcare); for the remaining lanes, growth conditions are: Lane 1 - LB medium, 37 °C, no induction (uninduced control growth); Lane 2 – LB medium, 37 °C, expression induction with IPTG addition at 16h and continuing growth at 37 °C for 6 hours; Lane 3 – LB medium, 37 °C, expression induction with IPTG addition at 16h and continuing growth at room temperature; Lane 4 – LBAI medium, 24 h growth at 37 °C; Lane 5 – LBAI medium, 24 h growth at room temperature; Lane 6 - LBAI medium, 24 h growth at 37 °C (duplicate growth).

Regarding CYP2C8 expression induction, the expected increase in protein production after incubation of the bacterial cells with IPTG did not occur, since overall, the band intensities are similar prior and after the addition of the inducer (Figure 6). The protein of interest, which has approximately 52 kDa, was not detected in any of the bands, presumably because it was expressed at very low levels due to a harmful effect that the heterologous proteins exert in the cells (Studier, 2005; Rosano and Ceccarelli, 2014; Dumon-Seignovert, Cariot, and Vuillard, 2004).

Since CYPs can oxidize a large number of substrates, they can present some toxicity in the host cell interfering with normal proliferation of the microorganism leading to high basal levels of protein expression. In this case, since an inducible T7 expression system was used, T7RNA polymerase can be expressed in small basal levels but its high activity can lead to substantial expression of the target protein even in the absence of added inducer. Moreover, protein toxicity to the host cell can difficult the establishment of the target plasmid in the expression host or the expression strain can be unstable or accumulate mutations. An alternative to surpass this problem has the main target of reducing basal expression by placing the lac operator sequence after the start site of a T7 promoter, but by analysing the CYP2C8 plasmid sequence used it is already as suggested (Studier, 2005; Rosano and Ceccarelli, 2014). Another approach to overcome lower expression would be to use a glucose-rich medium, as the presence of glucose stops lactose uptake by inactivating lactose permease in the early stages, blocking induction by lactose and letting the host cells grow and maintain functional plasmid until induction of the toxic protein. Nonetheless excess glucose must be avoided because bacterial cultures can achieve sufficient acidic levels to stop cell growth (Studier, 2005; Rosano and Ceccarelli, 2014; Sørensen and Mortensen, 2005).

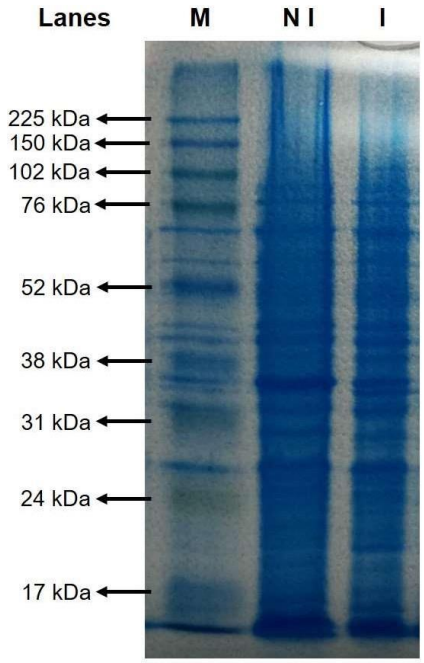

Figure 6 - SDS-PAGE analysis of CYP2C8 protein expression by *E. coli* BL21(DE3).
E. coli BL21(DE3) cells transformed with the pET28a(+)-CYP2C8 plasmid were grown in different conditions. All cultures were grown with media supplemented with kanamycin at a final 30 $\mu g.mL^{-1}$ concentration. Lane MM displays the molecular weight marker Amersham™ ECL Rainbow Marker – Full Range RPN800E (GE Healthcare); for the remaining lanes, growth conditions are: Lane 1 - LB medium, 37 °C, no induction (uninduced control growth); Lane 2 – LB medium, 37 °C, expression induction with IPTG addition at 16h and continuing growth at 37 °C for 6 hours.

To test this hypothesis, cultures of E. coli transformed with the pET28-CYP2C8 plasmid were also performed in LB auto-induction media, which consists of LB supplemented with glucose (0.5 g/L) and α-lactose (2.0 g/L), which allows cells to initially grow exclusively on glucose, promoting high cell density(Studier, 2005). Once glucose is depleted, usually in mid to late log phase, lactose enters the cell where it is converted by ß-galactosidase into allolactose, which in turn serves as the inducer of the IPTG-inducible promoter, resulting in protein expression. Results obtained (data not shown) indicate a very slight overexpression of a protein in the approximate mass range, but not enough to be purified, requiring further optimization of growth conditions.

Despite these results, it was possible to purify cytochrome P450 2C8 from IPTG induced bacterial cultures but not from LBAI grown bacteria, although at very low levels. The SDS-PAGE gels from the cell extracts from these cultures do not show any difference (Figure 6), but further protein isolation and purification lead to the elution of a fraction absorbing at 280 nm.

With these results, further bacterial cultures were grown in LB media using IPTG for induction of CYP450 expression, and in LBAI media for expression of sulfotransferases.

3.2 Protein purification

The recombinant proteins under study in this work were expressed using *E. coli* BL21(DE3) as host, with the required cDNA cloned in pET28a(+) plasmids. Protein expression was induced either using IPTG as inducing agent in LB cultures, in the case of cytochromes P450, or using the glucose/lactose auto-induction system (LBAI media) in the case of sulfotransferases. Also, in the case of cytochromes P450, a hemic protein, the media was supplemented with 1 mM of levulinic acid, a precursor of the haem synthesis pathway (Chung *et al.*, 1995).

Upon bacterial growth, cells were harvested and lysed, and the whole cell lysate was used to isolate the protein using an immobilized metal affinity chromatography (IMAC) system. A Ni-loaded chromatography resin was used in order to preferentially bind proteins expressing a six histidine tag. This chromatographic system was used because of its advantages of having strong, specific binding, mild elution condition and ability to control selectivity by including low concentration of imidazole in chromatography buffers. The final purified protein can be optimized by controlling the ratio of recombinant protein to the column size; lower-affinity contaminants can be out-competed with a relative excess of the histidine-tagged recombinant protein. To analyse the purification procedure, all eluted fractions from each purified protein were quantified measuring the absorbance at 280 nm, and a chromatogram of each purified protein are shown. Each chromatogram relates the variation of absorption at 280 nm of each fraction of eluted volume and the concentration of imidazole in each fraction on eluted volume. This wavelength was used because according to the study of proteins usually show absorption maxima between 275 and 280 nm, which are cause by the absorbance of the two aromatic amino acids tryptophan (Trp) and tyrosine (Tyr) and, to a small extent by the absorbance of cysteine. (; Bolanos-Garcia and Davies, 2006) The chromatograms obtained are presented hereafter.

3.2.1 Sulfotransferase 1B1 purification

The chromatogram corresponding to the purification of sulfotransferase 1B1 is shown in Figure 7; it was obtained using a multi-step imidazole gradient from 0 to 25 mM and finishing with a 500 mM imidazole concentration. After an initial absorbance peak corresponding to non-binding proteins, a sharp intense peak is observed with 500 mM imidazole eluent (195 and 200 mL elution volume), corresponding to elution of a protein that binds strongly to the resin.

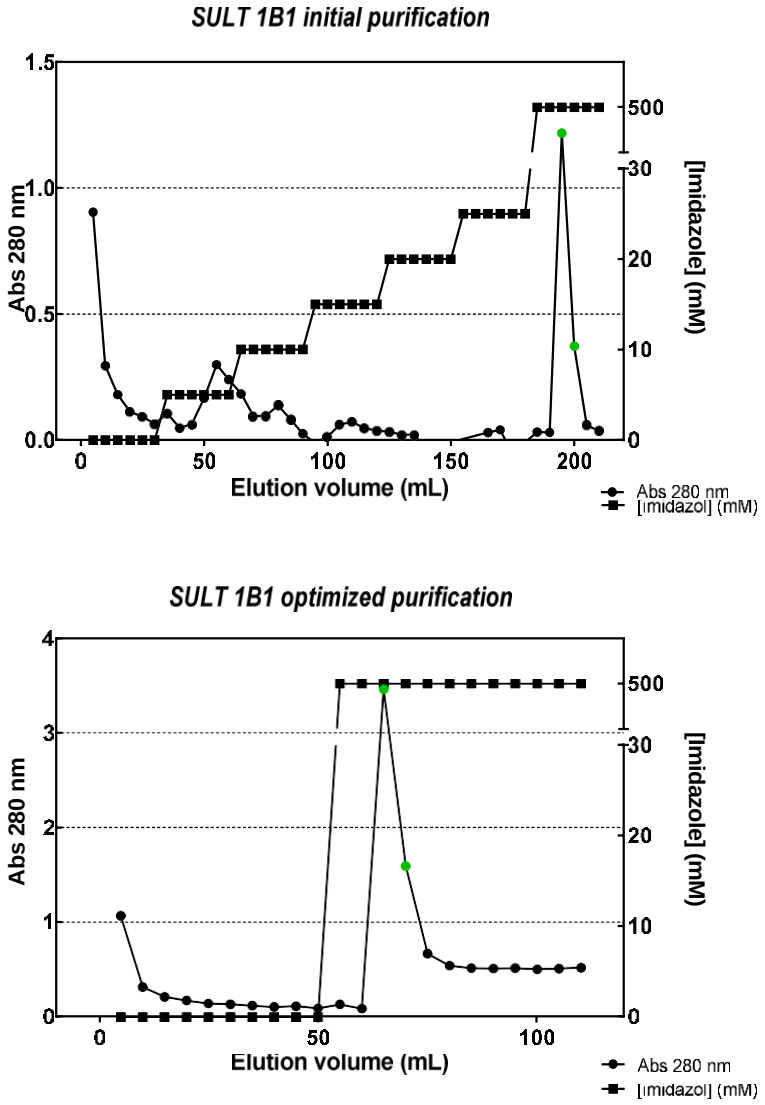

Figure 7 –Purification chromatograms of the expressed His-tagged sulfotransferase1B1.
Purification was performed on an immobilized nickel affinity resin, and absorbance at 280 nm for each fraction was measured off-line. The green dots correspond to the fractions where the recombinant protein was eluted. Upper panel – elution was accomplished via a multi-step imidazole gradient in phosphate buffer supplemented with NaCl; lower panel - elution was accomplished via a three-step imidazole gradient in phosphate buffer supplemented with NaCl, glycerol, Tween-20 and dithiothreitol. Details are given in the experimental section.

After the initial multi-step purification, elution conditions for sulfotransferase were simplified and optimized. Instead of a 5-step gradient, a single 500 mM step was used, also affording a well-defined absorbance peak. On the other hand, and as in the previous purification the protein was found to precipitate easily upon refreezing, the elution buffer was also supplemented with 5% (v/v) glycerol and 0.025 % (v/v) Tween-20, to help to prevent protein aggregation, and 0.1 mM dithiothreitol, to diminish protein oxidation.

3.2.2. Sulfotransferase 1A1 purification

Following the same protocol used for purification of SULT1B1, SULT1A1 was initially purified with an imidazole multi-step gradient, from 5 to 500 mM (Figure 8, upper panel), and after the initial elution of non-binding proteins an absorbance band obtained with 5 and 10 mM imidazole was observed, corresponding to the expressed protein.

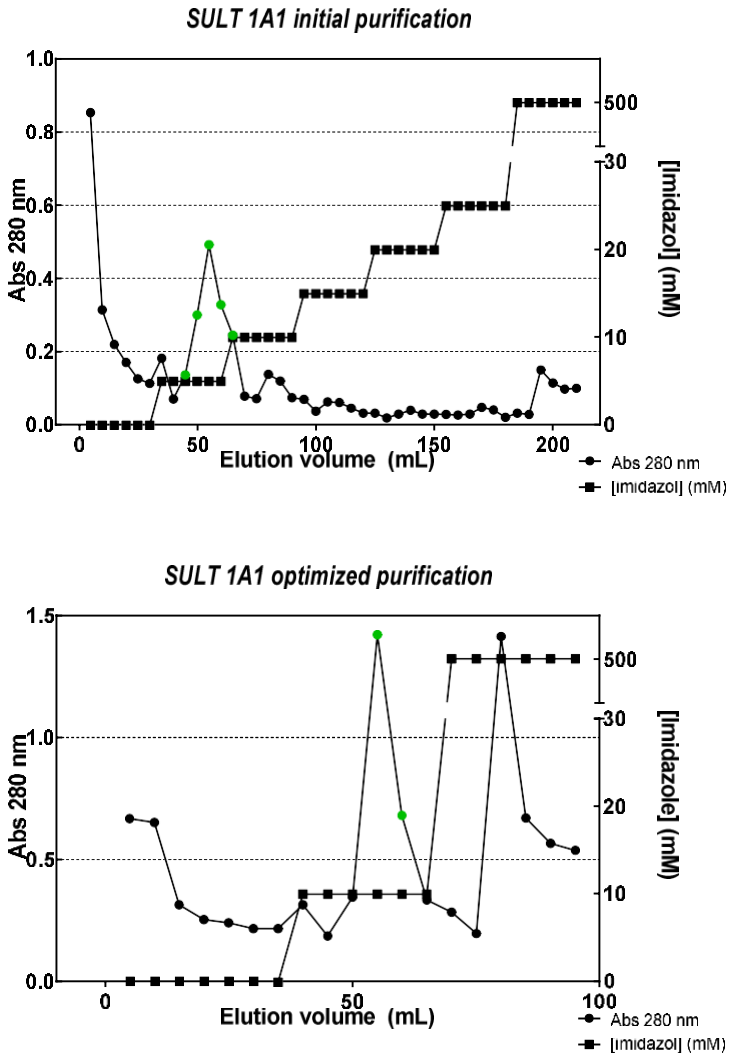

Figure 8 – Purification chromatograms of the expressed His-tagged sulfotransferase 1A1.
Purification was performed on an immobilized nickel affinity resin, and absorbance at 280 nm for each fraction was measured off-line. The green dots correspond to the fractions where the recombinant protein was eluted. Upper panel – elution was accomplished via a multi-step imidazole gradient in phosphate buffer supplemented with NaCl; lower panel - elution was accomplished via a three-step imidazole gradient in phosphate buffer supplemented with NaCl, glycerol, Tween-20 and dithiothreitol. Details are given in the experimental section.

In subsequent purifications, a single elution step with 10 mM was used (Figure 8, lower panel). Although a band corresponding to a protein that interacts with the resin was consistently observed, further work with this protein was not performed. All SULT1A1 batch's grew very slowly and to low OD values, and yielded low amounts of purified protein.

Further work, performed by other group researchers, has led to the obtention of higher protein amounts when using *E. coli* C41(DE3). These cells are a different variant of expression-competent *E. coli* cells, that were selected because they were able to grow proteins that are toxic to the expression host. In fact, SULT1A1 is the most wide-range sulfotransferase in terms of substrate specificity, and the observed low culture growth and low amounts of purified protein are in agreement with these observations.

3.2.3 Cytochrome P4502C8 purification

The chromatograms relative to the purification of the 2C8 isoform of cytochrome P450 are presented in Figure 9. The upper panel shows the multi-step purification of the protein, and no clear absorbance peak was observed, only two small bands at 55 and 75 mL of elution volume (5 and 10 mM imidazole). However, when the elution was performed using only a 25 mM imidazole step, a band, although small, was observed. Like SULT1A1, CYP2C8 is one of the members of the cytochrome P450 family with a largest substrate promiscuity, and cultures performed to express this protein usually grow slower and lead to low protein amounts (Anzenbacher and Anzenbacherová, 2001)

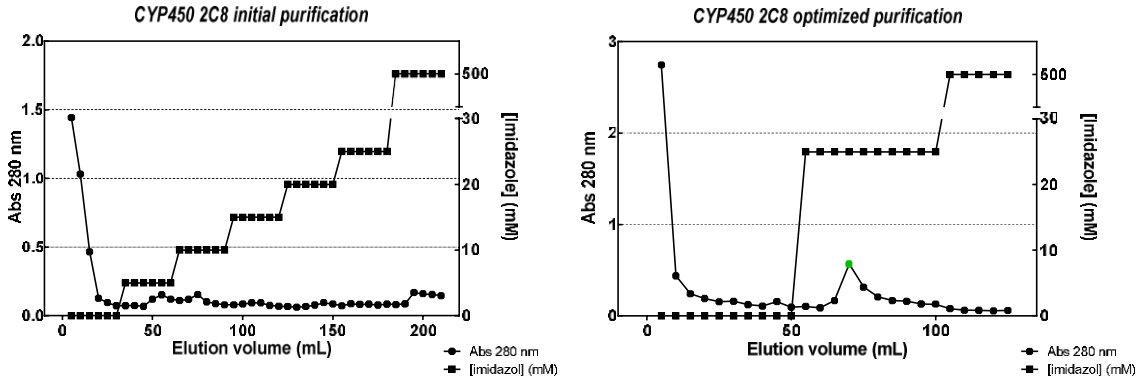

Figure 9 – Purification chromatograms of the expressed His-tagged cytochrome P450 2C8.
Purification was performed on an immobilized nickel affinity resin. Absorbance at 280 nm for each fraction was measured off-line. The green dots correspond to the fractions where the recombinant protein was eluted. Left panel - elution was accomplished via a multi-step imidazole gradient in phosphate buffer supplemented with NaCl. Right panel - elution was accomplished via a three-step imidazole gradient in phosphate buffer supplemented with NaCl, glycerol, Tween-20 and dithiothreitol. Details are given in the experimental section.

3.2.4 Cytochrome P450 2D6 purification

The purification of CYP2D6 with a multi-step gradient yielded an absorbance peak corresponding to a weakly interacting protein that was released from the resin with the 5 mM imidazole eluent. However, when a single 5 mM step was used, no absorbance peak was registered (data not shown). For that reason, a 25 mM imidazole eluent was used, and a well-defined peak was observed at approximately the same elution volume. These results are presented in Figure 10.

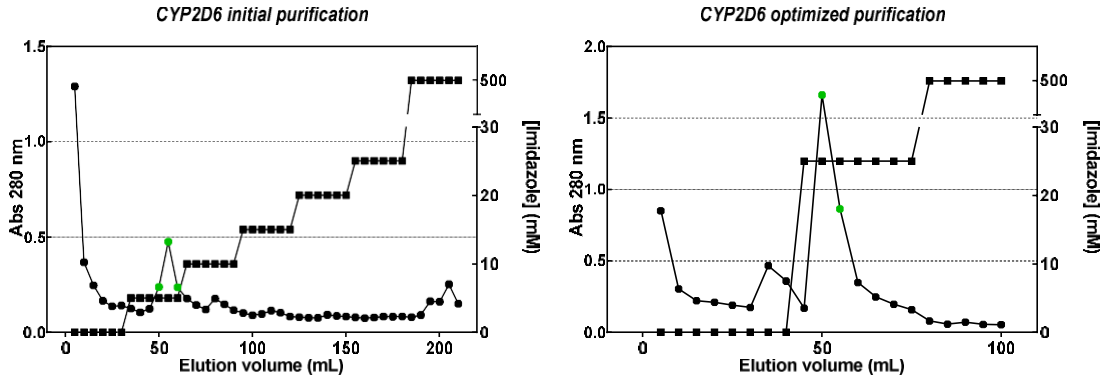

Figure 10 – Purification chromatograms of the expressed His-tagged cytochrome P450 2D6.
Purification was performed on an immobilized nickel affinity resin. Absorbance at 280 nm for each fraction was measured off-line. The green dots correspond to the fractions where the recombinant protein was eluted. Upper panel - elution was accomplished via a multi-step imidazole gradient in phosphate buffer supplemented with NaCl. Lower panel - elution was accomplished via a three-step imidazole gradient in phosphate buffer supplemented with NaCl, glycerol, Tween-20 and dithiothreitol. Details are given in the experimental section.

The UV/visible absorption spectra of the fraction corresponding to the eluted over-expressed protein confirmed that the eluted protein is a cytochrome (Figure 11). Besides the band with an absorption maxima at 280 nm, corresponding to the peptide chain, a second band is observed with a maxima at 412 nm, corresponding to the Soret band of iron-coordinating heme. The nomenclature P450 comes from the characteristic absorption band at 450 nm, when the iron-heme is in a reduced carbxoy (CO complexed) state, considerably distinct from the usual Soret absorption peaks of hemoproteins. This peak at 412 nm is likely to correspond to the low-spin form of the enzyme, in the oxidised (ferric) state, characterized by a band at 417 nm (Anzenbacher and Anzenbacherová, 2001; Schenkman and Jansson, 1998).

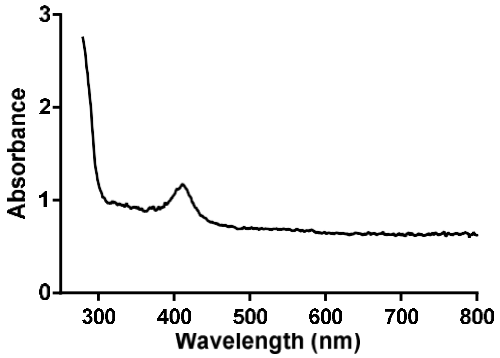

Figure 11 – UV/Visible absorption spectra of the produced CYP 450 2D6 isoform.
The spectra of the fraction corresponding to the eluted over-expressed protein is shown. Maxima are present at 412 nm, corresponding to the heme Soret band, and at 280 nm, corresponding to the peptide chain.

To confirm this, a protein sample was reduced with sodium dithionite, following the established procedure of adding a few milligrams of sodium dithionite on the tip of a spatula to a sample of CYP

2D6 in 20% pyridine in 0.2 M NaOH and recording its visible spectrum against a non-reduced sample in the same conditions(Munro, Girvan, and McLean, 2007). The differential spectra obtained in shown in Figure 12.

Upon reduction, the band at 412 nm disappears, and a transient band at 556 nm appears in the first minute upon dithionite addition (dashed line) but is absent at 3 (pointed line) and 5 minutes (full line) upon addition; this is characteristic of CYP 450 reduction by agents such as dithionite or NADPH (Schenkman and Jansson, 1998; Guengerich *et al.*, 2009).

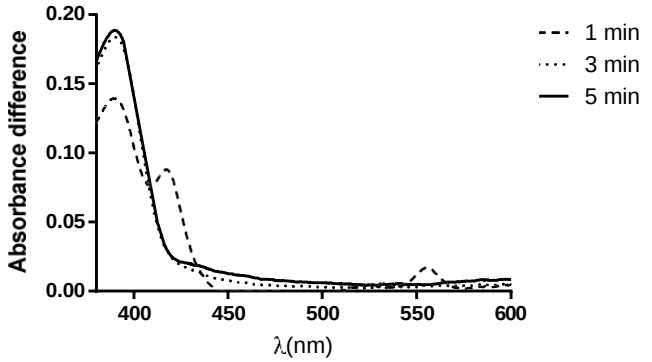

Figure 12 – Differential visible spectrum of cytochrome 2D6.
The differential spectra were recorded with two cuvettes with CYP2D6 in 20 % pyridine in 0.2 M NaOH by reducing the sample with solid sodium dithionite.

~~3.2.5~~ Control of protein purity during purification

In order to monitor the efficiency of the purification process, eluted fractions were analysed by running them in a polyacrylamide gel, under denaturing conditions (SDS-PAGE). Results are shown in Figure 13.

Regarding SULT1B1 purification (left panel), a band around 34 kDa is observed in lanes 4 to 6, in agreement with the molecular weight of SULT1B1 protein, but the intensity is lower in lane 6 and in the remaining the same band presents higher intensity. Although there is an intense band in lane 4 at the molecular weight of SULT1B1 protein, other bands are present, indicating the presence of contaminating proteins. The lanes 5 and 6 already present less protein contaminants since there are almost no other bands and so purification of SULT1B1 protein was successful in these fractions.

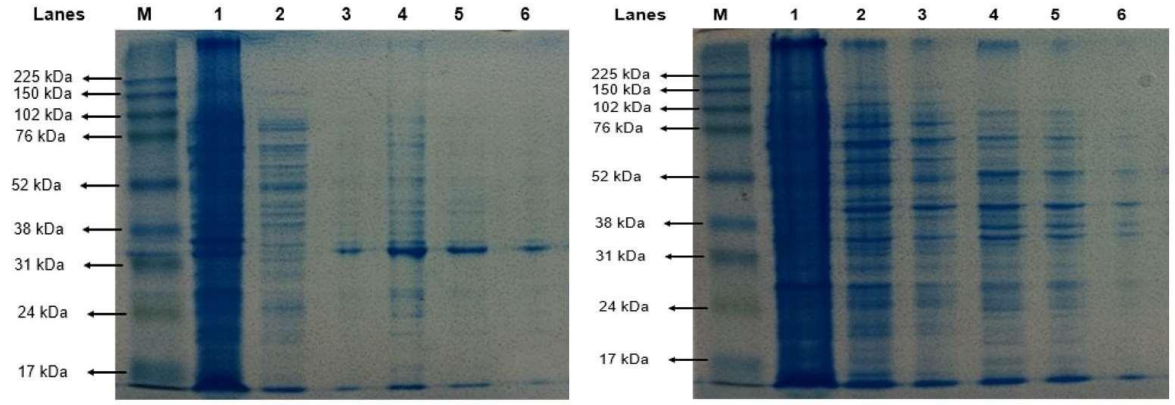

Figure 13 - Analysis of the purity of recombinant His-tagCYP2C8 and SULT1B1 proteins.
Left panel – SDS PAGE analysis of the elution profile for the SULT1B1 protein with an increased gradient of imidazole. Gel was stained with EzBlueTM Staining Reagent (Sigma-Aldrich, St. Louis, MO). Lane M –Molecular weight marker Amersham™ ECL™ Rainbow Marker – Full Range RPN800E (GE Healthcare). Lane 1 – Fraction 0-5 mL of elution volume, corresponding to elution of the non-binding fractions. Lane 2 - Fraction 10-15 mL of elution volume, corresponding to the final washing step without imidazole. Lane 3 – Fraction 25-30 mL of elution volume, corresponding to the first fraction collected with the elution buffer (500 mM imidazole). Lane 4 – Fraction 35-40 mL of elution volume, corresponding to the fraction with highest Abs 280 nm value during the elution step (with 500 mM imidazole). Lane 5 – Fraction 40-45 mL of elution volume, and Lane 6 – Fraction 50-55 mL of elution volume; lanes 5 and 6 correspond to the two final fractions during the elution step (with 500 mM imidazole)
Right panel – SDS-PAGE analysis of the elution profile for the CYP2C8 protein with an increased gradient of imidazole. Gel was stained with EzBlueTM Staining Reagent (Sigma-Aldrich, St. Louis, MO). Lane M –Molecular weight marker Amersham™ ECL™ Rainbow Marker – Full Range RPN800E (GE Healthcare). Lane 1 – Fraction 0-5 mL of elution volume, corresponding to elution of non-binding proteins. Lane 2 - Fraction 10-15 mL of elution volume. Lane 3 – Fraction 25-30 mL of elution volume. Lane 4 – Fraction 35-40 mL of elution volume. Lane 5 – Fraction 40-45 mL of elution volume. Lane 6 – Fraction 50-55 mL of elution volume. Lanes 2 to 4 correspond to different fractions obtained during the elution step, with 25 mM imidazole, and lanes 5 and 6 to fractions collected during the final 500 mM imidazole step.

Regarding CYP2C8 purification (right panel in Figure 13), only low-intensity band at 52 kDa (the molecular weight of this protein) were observed, appearing a little more intense on lane 4, corresponding to the elution step. In lanes 5 and 6, there are fewer contaminating proteins since there are fewer bands, although the 52 kDa band appears with lower intensity comparing with lane 4.

4. CONCLUSIONS

The development of a biomimetic enzymatic reactor, in order to predict *in vitro* the most likely drug metabolites, requires the use of large protein amounts, that can only be achieved by overexpressing the human proteins in adequate hosts. Many drugs are associated with toxic side effects of varying degree, and in many cases, these are due to the conversion of the parent drug to a toxic metabolite by cytochrome P450 (CYP) systems or by sulfotransferases.

The goal of this work was to clone and over-express human CYP isoforms 2C8, 2D6 and 3A4, and human sulfotransferase isoforms 1A1 and 1B1, using *E. coli*-pET expression systems. For that, pET28a(+) plasmids with the corresponding cDNA were developed and *E. coli* cells were transformed in order to over-express the protein under control of the T7 promoter by inducing the expression with either IPTG or a glucose/lactose media.

Sulfotransferase 1A1 was successfully cloned and transformed BL21(DE3) cells grown in auto-induction LB media were able to grow and express the protein, although in low levels. Further optimization performed after the experimental work for this thesis was finished has identified C41(DE3) as a better host for expressing this protein, although also at low levels.

Sulfotransferase 1B1, already available in a kanamycin-resistant pET28a(+) vector, was successfully over-expressed in auto-induction LB media.

Cytochrome P450 isoforms 2C8 and 2D6 were also successfully cloned and transformed into BL21(DE3) cells, but further work has shown that these two proteins are better expressed in C43(DE3) cells.

Cytochrome P450 3A4 could not be successfully cloned, despite multiple attempts. This can be due to the very low substrate specificity of this enzyme, which, coupled to the leaky T7 promoter behaviour, may lead to the fast onset of toxic effects on the various host cells, including non-DE3-containting strains. Plating the transformation into glucose supplemented medium may be a good working possibility, that will be attempted.

Overall, the work performed has allowed the production of four human recombinant proteins, that were used in the research group where this work was performed for the initial development of an enzymatic reactor, by immobilizing the proteins on a solid matrix and assessing that their substrate specificity and regioselectivity is maintained.

5. REFERENCES

.

Albin, N., L. Massaad, C. Toussaint, M. C. Mathieu, J. Morizet, O. Parise, A. Gouyette, and G. G. Chabot. 1993. 'Main drug-metabolizing enzyme systems in human breast tumors and peritumoral tissues', *Cancer research*, 53: 3541-46.

Allawi, H. T., and J. SantaLucia, Jr. 1997. 'Thermodynamics and NMR of internal G.T mismatches in DNA', *Biochemistry*, 36: 10581-94.

Almazroo, Omar Abdulhameed, Mohammad Kowser Miah, and Raman Venkataramanan. 2017. 'Drug Metabolism in the Liver', *Clinics in liver disease*, 21: 1-20.

Anzenbacher, P., and E. Anzenbacherová. 2001. 'Cytochromes P450 and metabolism of xenobiotics', *Cellular and molecular life sciences : CMLS*, 58: 737-47.

Aust, Sylvia, Walter Jaeger, Martin Klimpfinger, Klaus Mayer, Günther Baravalle, Cem Ekmekcioglu, and Theresia Thalhammer. 2005. 'Biotransformation of melatonin in human breast cancer cell lines: role of sulfotransferase 1A1', *Journal of pineal research*, 39: 276-82.

Barrera-Saldaña, Hugo. 2012. *Genetic Engineering* (InTech Open).

Bolanos-Garcia, Victor Martin, and Owen Richard Davies. 2006. 'Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography', *Biochimica et biophysica acta*, 1760: 1304-13.

Bradford, M. M. 1976. 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Analytical biochemistry*, 72: 248-54.

Brandenberg, Oliver, Zephaniah Dhlamini, Alessandra Sensi, Kakoli Ghosh, and Andrea Sonnino. 2011. *Introduction to Molecular Biology and Genetic Engineering* (Food and Agriculture Organization of the United Nations).

Bryant, F. R. 1988. 'Construction of a recombinase-deficient mutant recA protein that retains single-stranded DNA-dependent ATPase activity', *The Journal of biological chemistry*, 263: 8716-23.

Chan, Weng-Tat, Chandra S. Verma, David P. Lane, and Samuel Ken-En Gan. 2013. 'A comparison and optimization of methods and factors affecting the transformation of Escherichia coli', *Bioscience reports*, 33: e00086.

Chung, B. C., M. C. Hu, V. M. Guzov, and D. A. Wu. 1995. 'Structure and expression of the CYP21 (P450c21, steroid 21-hydroxylase) gene with respect to its deficiency', *Endocrine research*, 21: 343-52.

Di Nardo, Giovanna, and Gianfranco Gilardi. 2012. 'Optimization of the bacterial cytochrome P450 BM3 system for the production of human drug metabolites', *International journal of molecular sciences*, 13: 15901-24.

Dumon-Seignovert, Laurence, Guillaume Cariot, and Laurent Vuillard. 2004. 'The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3)', *Protein expression and purification*, 37: 203-06.

Fox, Brian G., and Paul G. Blommel. 2009. 'Autoinduction of protein expression', *Current protocols in protein science*, Chapter 5: Unit-5.23.

Fujita, K., K. Nagata, S. Ozawa, H. Sasano, and Y. Yamazoe. 1997. 'Molecular cloning and characterization of rat ST1B1 and human ST1B2 cDNAs, encoding thyroid hormone sulfotransferases', *Journal of biochemistry*, 122: 1052-61.

Gamage, Niranjali, Amanda Barnett, Nadine Hempel, Ronald G. Duggleby, Kelly F. Windmill, Jennifer L. Martin, and Michael E. McManus. 2006. 'Human sulfotransferases and their role in chemical metabolism', *Toxicological sciences : an official journal of the Society of Toxicology*, 90: 5-22.

Ganesh, Swaytha, Omar Abdulhameed Almazroo, Amit Tevar, Abhinav Humar, and Raman Venkataramanan. 2017. 'Drug Metabolism, Drug Interactions, and Drug-Induced Liver Injury in Living Donor Liver Transplant Patients', *Clinics in liver disease*, 21: 181-96.

Gerhard, Daniela S., Lukas Wagner, Elise A. Feingold, Carolyn M. Shenmen, Lynette H. Grouse, Greg Schuler, Steven L. Klein, Susan Old, Rebekah Rasooly, Peter Good, Mark Guyer, Allison M. Peck, Jeffery G. Derge, David Lipman, Francis S. Collins, Wonhee Jang, Steven Sherry, Mike Feolo, Leonie Misquitta, Eduardo Lee, Kirill Rotmistrovsky, Susan F. Greenhut, Carl F. Schaefer, Kenneth Buetow, Tom I. Bonner, David Haussler, Jim Kent, Mark Kiekhaus, Terry Furey, Michael Brent, Christa Prange, Kirsten Schreiber, Nicole Shapiro, Narayan K. Bhat, Ralph F. Hopkins, Florence Hsie, Tom Driscoll, M. Bento Soares, Tom L. Casavant, Todd E. Scheetz, Michael J. Brown-stein, Ted B. Usdin, Shiraki Toshiyuki, Piero Carninci, Yulan Piao, Dawood B. Dudekula, Minoru S. H. Ko, Koichi Kawakami, Yutaka Suzuki, Sumio Sugano, C. E. Gruber, M. R. Smith, Blake Simmons, Troy Moore, Richard Waterman, Stephen L. Johnson, Yijun Ruan, Chia Lin Wei, S. Mathavan, Preethi H. Gunaratne, Jiaqian Wu, Angela M. Garcia, Stephen W. Hulyk, Edwin Fuh, Ye Yuan, Anna Sneed, Carla Kowis, Anne Hodgson, Donna M. Muzny, John McPherson, Richard A. Gibbs, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez, Michelle Whiting, Anup Madari, Alice C. Young, Keith D. Wetherby, Steven J. Granite, Peggy N. Kwong, Charles P. Brinkley, Russell L. Pearson, Gerard G. Bouffard, Robert W. Blakesly, Eric D. Green, Mark C. Dickson, Alex C. Rodriguez, Jane Grimwood, Jeremy Schmutz, Richard M. Myers, Yaron S. N. Butterfield, Malachi Griffith, Obi L. Griffith, Martin I. Krzywinski, Nancy Liao, Ryan Morin, Diana Palmquist, Anca S. Petrescu, Ursula Skalska, Duane E. Smailus, Jeff M. Stott, Angelique Schnerch, Jacqueline E. Schein, Steven J. M. Jones, Robert A. Holt, Agnes Baross, Marco A. Marra, Sandra Clifton, Kathryn A. Makowski, Stephanie Bosak, Joel Malek, and M. G. C. Project Team. 2004. 'The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC)', *Genome research*, 14: 2121-27.

Gingeras, Thomas R., Phyllis A. Myers, Julie A. Olson, Francine A. Hanberg, and Richard J. Roberts. 1978. 'A new specific endonuclease present in Xanthomonas holcicola, Xanthomonas papavericola and Brevibacterium luteum', *Journal of molecular biology*, 118: 113-22.

Green, Michael R., and Joseph Sambrook. 2012. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press).

Guengerich, F. Peter, Martha V. Martin, Christal D. Sohl, and Qian Cheng. 2009. 'Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase', *Nature protocols*, 4: 1245-51.

Jeong, Haeyoung, Valérie Barbe, Choong Hoon Lee, David Vallenet, Dong Su Yu, Sang-Haeng Choi, Arnaud Couloux, Seung-Won Lee, Sung Ho Yoon, Laurence Cattolico, Cheol-Goo Hur, Hong-Seog Park, Béatrice Ségurens, Sun Chang Kim, Tae Kwang Oh, Richard E. Lenski, F. William Studier, Patrick Daegelen, and Jihyun F. Kim. 2009. 'Genome sequences of Escherichia coli B strains REL606 and BL21(DE3)', *Journal of molecular biology*, 394: 644-52.

Lodish, Harvey, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. 2000. *Molecular Cell Biology* (W. H. Freeman).

Munro, Andrew W., Hazel M. Girvan, and Kirsty J. McLean. 2007. 'Variations on a (t)heme--novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily', *Natural product reports*, 24: 585-609.

Nicholl, Desmond S. T. 2008. *An Introduction to Genetic Engineering* (Cambridge University Press: Cambridge).

Nimmagadda, Deepthi, Ganesh Cherala, and Srinivas Ghatta. 2006. 'Cytosolic sulfotransferases', *Indian journal of experimental biology*, 44: 171-82.

Nowell, S., and C. N. Falany. 2006. 'Pharmacogenetics of human cytosolic sulfotransferases', *Oncogene*, 25: 1673-78.

Patrick, Marcella, Miranda D. Gray, Maria Sandkvist, and Tanya L. Johnson. 2010. 'Type II Secretion in Escherichia coli', *EcoSal Plus*, 4: 10.1128/ecosalplus.4.3.4.

Paul, Priscilla, Jiraporn Suwan, Jian Liu, Jonathan S. Dordick, and Robert J. Linhardt. 2012. 'Recent advances in sulfotransferase enzyme activity assays', *Analytical and bioanalytical chemistry*, 403: 1491-500.

Roberts, R. J. 1992. 'Restriction enzymes and their isoschizomers', *Nucleic acids research*, 20 Suppl: 2167-80.

Rohn-Glowacki, Katie Jo, and Charles N. Falany. 2014. 'The potent inhibition of human cytosolic sulfotransferase 1A1 by 17α-ethinylestradiol is due to interactions with isoleucine 89 on loop 1', *Hormone molecular biology and clinical investigation*, 20: 81-90.

Rosano, Germán L., and Eduardo A. Ceccarelli. 2014. 'Recombinant protein expression in Escherichia coli: advances and challenges', *Frontiers in microbiology*, 5: 172-72.

Roy, P. H., and H. O. Smith. 1973a. 'DNA methylases of Hemophilus influenzae Rd. I. Purification and properties', *Journal of molecular biology*, 81: 427-44.

———. 1973b. 'DNA methylases of Hemophilus influenzae Rd. II. Partial recognition site base sequences', *Journal of molecular biology*, 81: 445-59.

Sakakibara, Y., Y. Takami, C. Zwieb, T. Nakayama, M. Suiko, H. Nakajima, and M. C. Liu. 1995. 'Purification, characterization, and molecular cloning of a novel rat liver Dopa/tyrosine sulfotransferase', *The Journal of biological chemistry*, 270: 30470-78.

Schenkman, J. B., and I. Jansson. 1998. 'Spectral analyses of cytochromes P450', *Methods in molecular biology (Clifton, N.J.)*, 107: 25-33.

Schomburg, Dietmar, Ida Schomburg, and Antje Chang. 2006. *Class 1 - Oxidoreductases XI EC 1.14.11 - 1.14.14* (Springer: Germany).

———. 2013a. *Class 2–3.2 Transferases, Hydrolases* (Springer: Germany).

———. 2013b. *Class 3.4–6 Hydrolases, Lyases, Isomerases, Ligases* (Springer: Germany).

Shapiro, A. L., E. Viñuela, and J. V. Maizel, Jr. 1967. 'Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels', *Biochemical and biophysical research communications*, 28: 815-20.

Sørensen, Hans Peter, and Kim Kusk Mortensen. 2005. 'Advanced genetic strategies for recombinant protein expression in Escherichia coli', *Journal of biotechnology*, 115: 113-28.

Strausberg, R. L., E. A. Feingold, R. D. Klausner, and F. S. Collins. 1999. 'The mammalian gene collection', *Science (New York, N.Y.)*, 286: 455-57.

Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. 'Use of T7 RNA polymerase to direct expression of cloned genes', *Methods in enzymology*, 185: 60-89.

Studier, F. William. 2005. 'Protein production by auto-induction in high density shaking cultures', *Protein expression and purification*, 41: 207-34.

Studier, F. William, Patrick Daegelen, Richard E. Lenski, Sergei Maslov, and Jihyun F. Kim. 2009. 'Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes', *Journal of molecular biology*, 394: 653-80.

Takahashi, Hideo, Mariko Shimizu, Hiuga Saito, Yonosuke Ikeda, and Hiroyuki Sugisaki. 1979. 'A new site-specific endonuclease from Streptomyces lavendulae (SlaI)', *Gene*, 5: 9-18.

Taylor, R. G., D. C. Walker, and R. R. McInnes. 1993. 'E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing', *Nucleic acids research*, 21: 1677-78.

Totah, Rheem A., and Allan E. Rettie. 2005. 'Cytochrome P450 2C8: substrates, inhibitors, pharmacogenetics, and clinical relevance', *Clinical pharmacology and therapeutics*, 77: 341-52.

Watson, Robert J., Ira Schildkraut, B.-Q. Qiang, Stan M. Martin, and Louis P. Visentin. 1982. 'NdeI: a restriction endonuclease from Neisseria denitrificans which cleaves DNA at 5′-CATATG-3′ sequences', *FEBS Letters*, 150: 114-16.

6. ANNEXES

6.1 Annex A1 - vector map of pET28a(+).

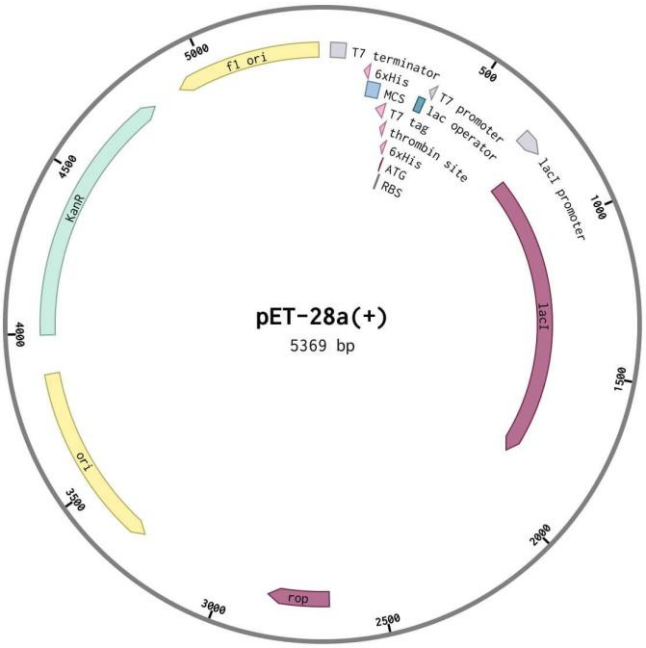

Figure A1 – Vector map for pET28a (+).
Plasmid map source: Novagen/MerckMillipore.

6.2 Annex A2 - vector map of pET28a(+)-SULT1B1.

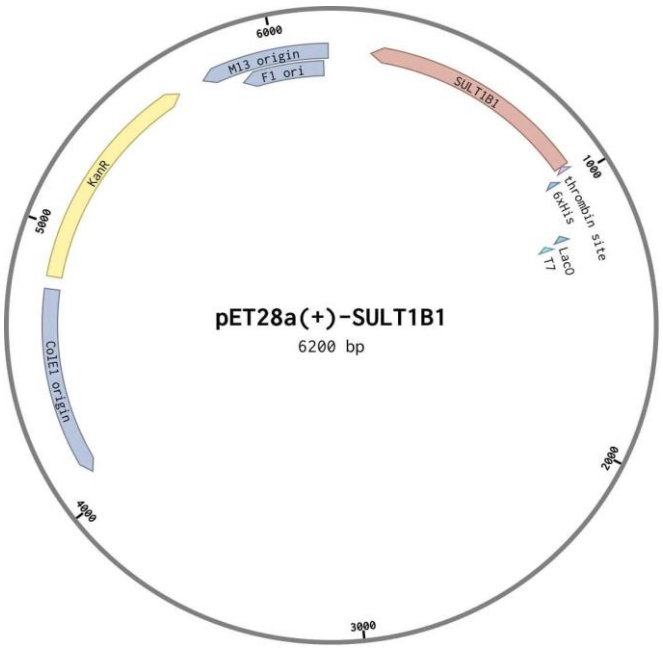

Figure A2 - Vector map for pET28a(+)-SULT1B1.
Built from the pET28a(+) map with the cloning information provided by Addgene (plasmid # 25496; http://n2t.net/addgene:25496)

6.3 Annex A3 - vector map of pCW-CYP2C8.

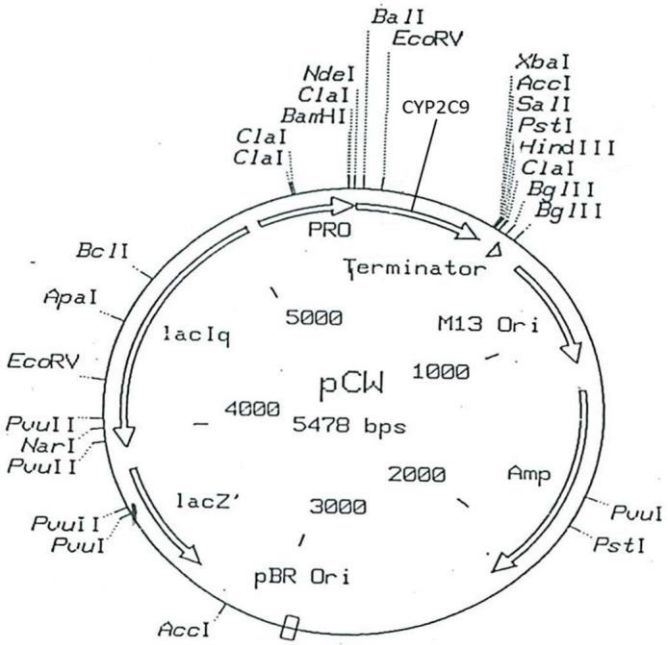

Figure A3 - Vector map for pCW-Ori+-CYP2C9.
Supplied by Addgene (CYP2C9 in vector pCW ori+, Addgene plasmid #69554; http://n2t.net/addgene:69554; RRID:Addgene_69554))

6.4 Annex A4 - vector map of pCMV-SPORT6-SULT1A1.

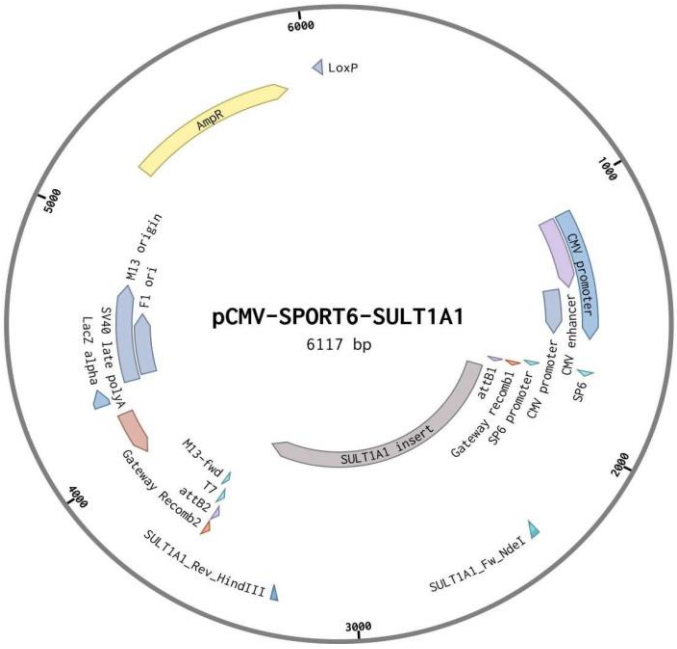

Figure A4 - Vector map for pCMV-SPORT6-SULT1A1.
Supplied by PlasmID database, clone HsCD00335548.

6.5 Annex A5 - vector map of pCR4-TOPO-CYP3A4.

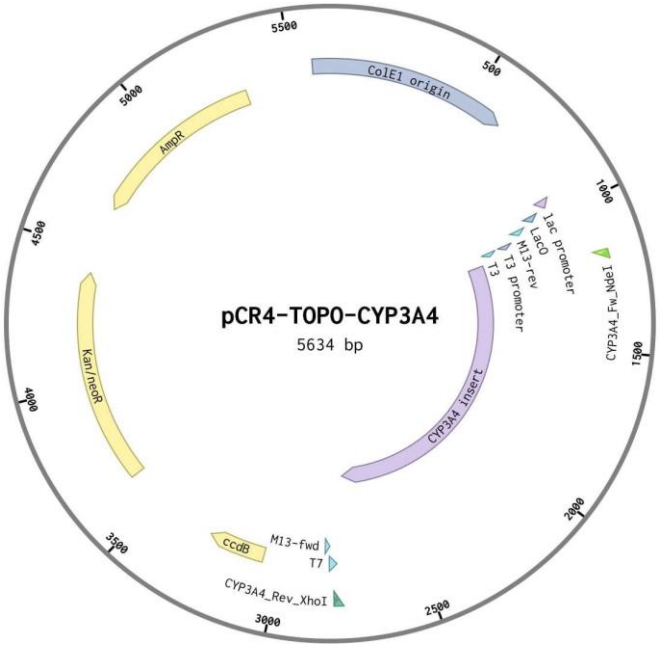

Figure A5 - Vector map for pCR4-TOPO-CYP3A4.
Supplied by PlasmID database, clone HsCD00341290.

6.6 Annex A6 - vector map of pCR4-TOPO-CYP2D6.

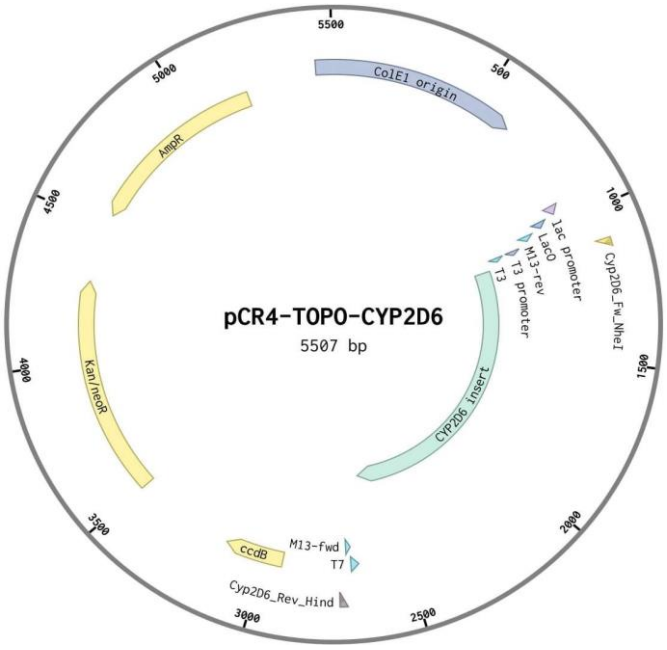

Figure A6 - Vector map for pCR4-TOPO-CYP3A4.
Supplied by PlasmID database, clone HsCD00346008.

6.7 Annex A7 - vector map of pET28a(+)-CYP2C8.

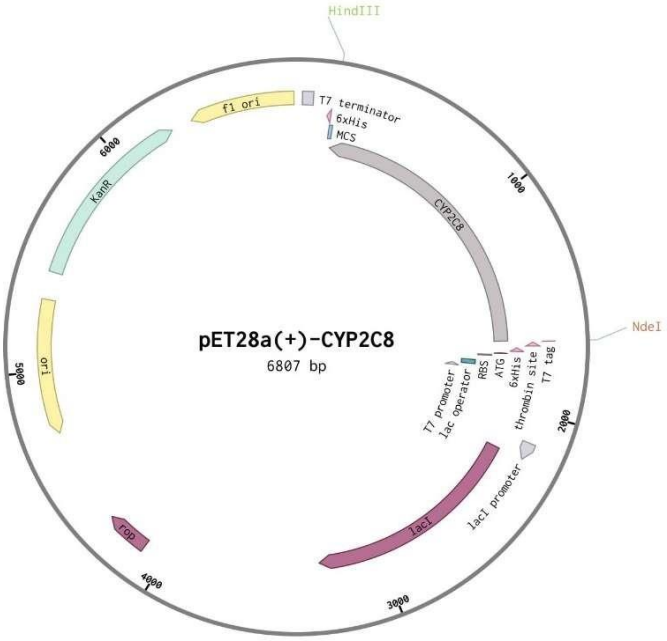

Figure A7 - Vector map for pET28a(+)-CYP2C8.
This vector was produced in this work, as described in the Experimental section.

6.8 Annex A8 - vector map of pET28a(+)-CYP2D6.

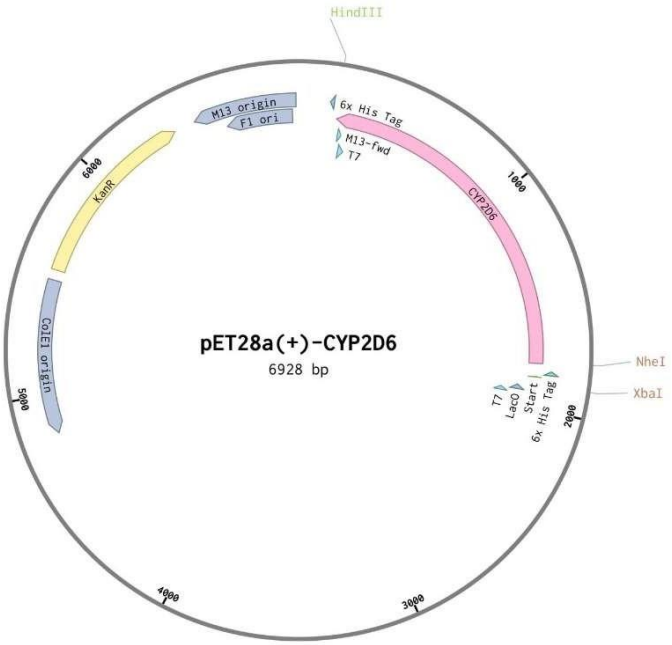

Figure A8 - Vector map for pET28a(+)-CYP2D6.
This vector was produced in this work, as described in the Experimental section.

6.9 Annex A9 - vector map of pET28a(+)-CYP3A4.

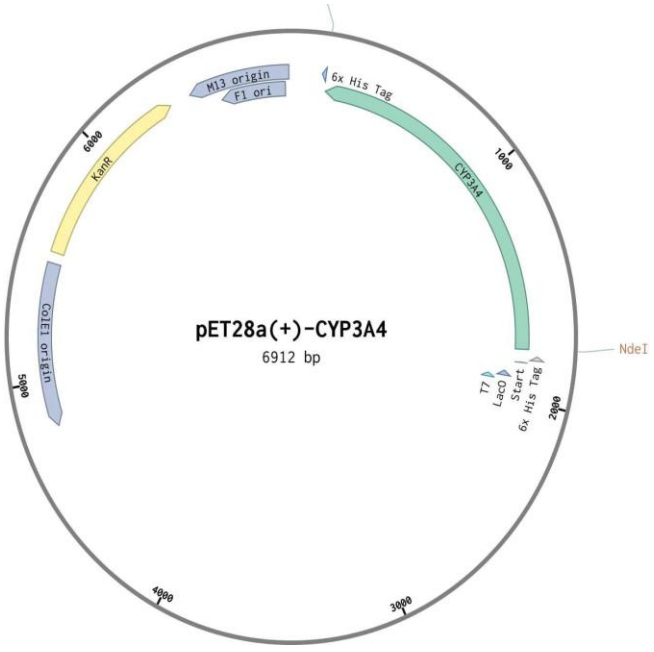

Figure A9 - Vector map for pET28a(+)-CYP3A4.
This vector was produced in this work, as described in the Experimental section.

6.10 Annex A10 - vector map of pET28a(+)-SULT1A1.

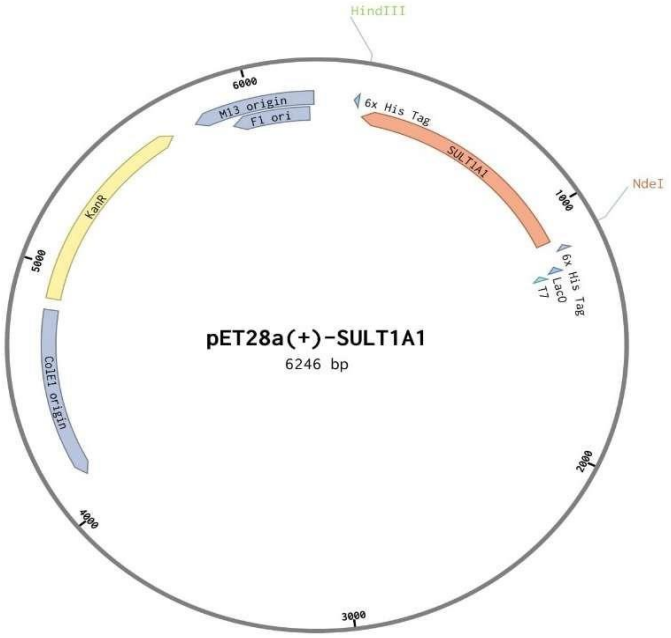

Figure A10 - Vector map for pET28a(+)-SULT1A1.
This vector was produced in this work, as described in the Experimental section.