Development of a bacterial model for the production of functional human drug metabolizing enzymes

Pedro da Conceição Rosado

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisors: Professor Maria Matilde Soares Duarte Marques

Doctor José Gonçalo Deira Duarte de Campos Justino

Examination Committee

Chairperson: Prof. Dr. Leonilde de Fátima Morais Moreira

Supervisor: Doctor José Gonçalo Deira Duarte de Campos Justino

Members of the Commitee: Prof. Marília Clemente Velez Mateus

October 2018

Acknowledgments

First of all, I would like to acknowledge my supervisor, Professor Matilde Marques for providing me this great opportunity to join the group of Design, Synthesis, and Toxicology of Bioactive Molecules. I am so grateful for the chance to develop such an incredible work in your lab. Thank so much for believing in my intuition and for all the support, motivation and guidance that you showed me during this year.

To my other supervisor, Dr. Gonçalo Justino, I would like to thank you for everything! It was an amazing experience to share this important phase with you…I think we both share the same feeling. I am fortunate to have been supervised by you, always helping me during this stressful year. It was a privilege to work with you and create this incredible friendship! Thank you for your pessimism that made me work harder and thank you for all the motivation, support and of course, for all the patience you had! In any time, you were always available to me and, definitively, you made me a better professional and certainly a better person. Thank you!

To Pedro Pinheiro, who helped in several steps of my work, it was like having a third supervisor! Your availability, patience and advices were crucial to finish this work. You made this experience better!

I want also to thank Marta Justino for all the help given during my Master Thesis! Your availability and patience were crucial to accomplish this work.

To my lab colleagues, a thanks isn't enough. Thank you for all the help, work discussions and, even more for all the conversations we had! You were a part of this year!

I would like to thank to all my friends for the support! To Débora for all the conversations and friendship that remained since our childhood. Thank you for believing in me and being a very special person in my life!

Thanks to Carolina who always supported me over the years. Thanks for all your love!

Thank you to my godparents for being an example to follow and for encouraging me to be better and ambitious.

Thanks also to my godchildren for giving me huge lessons and for making my university experience fantastic!

Thanks to Beatriz Batista for being a fantastic Erasmus colleague and for continuing to support me unconditionally.

I would also like to thank all my colleagues that FCT gave me, in particular, Pipa, Marisa, Madeira, Diogo and Pires.

I would like to thank Rita Valente, Madalena, Catarina Esteves and Cíntia for being the craziest friends and for all the support in every step. Thank you for all your positive thoughts!

I want to thank Mariana São Pedro for all the conversations and all the help during these two years!

I would also like to thank Miguel Gaspar and Joana Aldeias for being a fantastic company, for making every moment a fun, for being good friends and always so worried with me. Mostly, thank you for being fashion!

Especially, I would like to thank Pedro for being a pillar in my life in these two years. Thank you for helping me to overcome all obstacles and to put a smile every day in my face. Thank you for the patience and pure friendship along the years! Thank you so much for the unconditional support and for believing in me!

Another special thanks goes to Catarina who was the best person that university gave me. I want to thank you for all the patience, help, smiles, sadness, for everything that we have been through, you are very special!

Finally, I would like to thank my entire family for believing in me and for always encouraging me to do better! I want to express my gratitude, which is the size of the universe, to my parents, my sister and my grandfather who have made all this possible and made me the person I am today. I am so grateful for everything you have given me in life! Love you!

Abstract

Drug-induced liver injury is the main cause of drug failure during clinical trials and can lead to market withdrawal of already approved pharmaceuticals. Metabolic bioactivation of drugs by phase I and phase II metabolizing enzymes, mainly cytochromes P450 (CYP450s) and sulfotransferases (SULTs), respectively, are directly related with the formation of reactive metabolites. In this work we aimed to develop a bacterial model for the production of functional human drug metabolizing enzymes, namely, CYP450s and SULTs, to further predict the production of toxic metabolites by these enzymes and contribute to providing safer and more efficient drugs to patients.

Metabolic enzymes cDNAs were subcloned in *E. coli*-pET expression systems. CYP2C8 and SULT1B1 enzymes were selected for further overexpression and purification. Both proteins were purified by Immobilized Metal Affinity Chromatography and also by size exclusion chromatography. SULT1B1 kinetic parameters were assessed using 2-naphthol sulfonylation assay and using different substrates, as resorcinol, phenol and quercetin. Furthermore, to get some insights in both enzymes metabolic activity, SULT1B1 was incubated with resorcinol and CYP2C8 was incubated with nevirapine and tamoxifen. Through LC/MS analysis of the obtained metabolites we observed the occurrence of SULT1B1-mediated sulfonylation of resorcinol and formation of nevirapine metabolites (hydroxy-nevirapine) and tamoxifen metabolites (*N*-desmethyl-tamoxifen, *N,N*-didesmethyl-tamoxifen and hydroxy-tamoxifen) by CYP2C8-mediated oxidation, indicating that both expressed enzymes maintain their described metabolic activity and that our bacterial model is functional. Overall, results suggest that this bacterial model is a promising way of testing already approved or new drugs to avoid further risks in patients' life.

Keywords Drug-induced liver injury; Cytochrome P450; Sulfotransferase; Bacterial model; Drug metabolism; Nevirapine.

Resumo

A hepatoxicidade é a principal causa da rejeição de fármacos durante ensaios clínicos e pode levar à remoção do mercado de fármacos já aprovados. A bioativação de fármacos pela ação de enzimas de fase I e II, nomeadamente citocromos P450 (CYP450s) e sulfotransferases (SULTs), respetivamente, está diretamente relacionada com a formação de metabolitos tóxicos. Neste trabalho, pretendeu-se desenvolver um modelo bacteriano para a produção de CYP450s e SULTs humanas para posteriormente prever a produção de metabolitos tóxicos por ação destas enzimas e assim proporcionar fármacos mais seguros aos pacientes. Os genes que codificam para enzimas metabólicas foram subclonados em sistemas de expressão *E. coli*-pET. As enzimas CYP2C8 e SULT1B1 foram selecionadas para sobreexpressão e purificação. Ambas as proteínas foram purificadas por cromatografias de afinidade por iões metálicos imobilizados e de exclusão molecular. Os parâmetros cinéticos da enzima SULT1B1 foram caracterizados usando ensaios de sulfonação do 2-naftol e diversos substratos, particularmente, resorcina, fenol e quercetina.

Posteriormente, para obter resultados sobre a atividade metabólica das duas enzimas a SULT1B1 foi incubada com resorcinol e o CYP2C8 foi incubado com nevirapine e tamoxifeno. Analisando os resultados obtidos por LC/MS, foram observados metabolitos sulfonados do resorcinol pela SULT1B1 e metabolitos da nevirapina (hidroxi-nevirapina) e do tamoxifeno (*N*-desmetil-tamoxifeno, *N,N*-didesmetil-tamoxifeno e hidroxi-tamoxifeno) por oxidação pelo CYP2C8 indicando assim que ambas as proteínas mantém a atividade metabólica descrita. No geral, os resultados sugerem que o modelo bacteriano é um método promissor para testar fármacos novos ou já aprovados de modo a evitar riscos no quotidiano dos pacientes.

Palavras-chave Hepatoxicidade; Citocromo P450; Sulfotransferase; Modelo Bacteriano; Metabolismo de fármacos; Nevirapina.

Table of contents

Figure Index

Table Index

Abbreviation List

ABC	ATP binding cassettes
COMTs	Catechol *O*-methyltransferases
CYP	Cytochrome P450 (NADPH:oxygen oxidoreductase)
DNA	Deoxyribonucleic acid
DILI	Drug-induced liver injury
EDTA	Ethylenediamine tetraacetic acid
ESI	Electrospray ionization
GSH	Glutathione
GST	Glutathione *S*-transferases
HIV	Human immunodeficiency virus
IMAC	Immobilized metal affinity chromatography
iPSCs	Induced pluripotent stem cells
MS	Mass spectrometry
m/z	mass-to-charge ratio
NADPH	Dihydronicotinamide-adenine dinucleotide phosphate
$NADP^+$	Nicotinamide-adenine dinucleotide phosphate
NATs	*N*-acetyltransferases
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NAC	N-acetyl-L-cysteine
NVP	Nevirapine
o/n	Over-night
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	Rotations per minute
RT	Room temperature
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

SLC	Solute carrier
SULTs	Sulfotransferases
TAE	Tris-Acetate-EDTA
TPMTs	Thiopurine *S*-methyltransferases
UDP	Uridine diphosphate
UGTs	UDP-glucuronosyltransferases

1. Introduction

Xenobiotics, such as drugs, are compounds that are foreign to the human organism. Xenobiotics are eliminated from the organism through a series of conserved metabolic pathways. The liver is the major organ responsible for drug biotransformation, which comprises phase I and phase II reactions.

Phase I drug-metabolizing enzymes are represented mainly by the cytochrome P450 (CYP) superfamily. CYPs are heme-proteins that use molecular oxygen as the donor to catalyse oxidation of a vast range of substrates. The metabolites resulting from phase I pathways are further enzymatically modified during phase II by several enzymes, such as UDP-glucuronosyltransferases (UGTs), sulfotransferases (SULTs), *N*-acetyltransferases (NATs), glutathione *S*-transferases (GSTs), thiopurine *S*-methyltransferases (TPMTs) and catechol *O*-methyltransferases (COMTs). Phase II conjugation is focused on the introduction of large polar groups on the substrates in order to increase their water solubility and ease their urinary excretion.

Sometimes, the normal detoxification of drugs leads to the formation of reactive intermediates, a process called metabolic bioactivation. These toxic metabolites can lead to various toxic side-effects, in particular liver injury or even liver failure.

Drug-induced liver injury (DILI) represents a major clinical problem. Pharmaceutical industries and regulatory authorities agreed that there is the need to improve methodologies to accurately assess the hepatotoxic potential of compounds in early stages of drug development.

To overcome this problem, some models were created to assess the risk of hepatotoxicity by drugs, including mammalian and microbial models. Both models present advantages and limitations and more research is needed to find the best *in vitro* method.

Thus, easy and efficient models to mimic human metabolism in order to assess drug bioactivation are required. The ultimate goal of this work is to obtain a large enzymatic model of the pathways of drug metabolism using different isoforms of various enzymes simultaneously and recover the drug metabolites, validating it with well–studied model substrates and paving the way to a simple and effective drug prediction system.

1.1. Human Drug Metabolism in the Liver: an overview

Xenobiotics are typically small lipophilic molecules that are foreign to the human organism and so do not have a native role in the organism physiology. These compounds enter the body by external sources, namely in the form of drugs, food constituents and additives, natural toxins and pollutants. In addition, one of the main characteristics of xenobiotics is that they are usually not excreted by the organism in their native form and so they need to be modified through metabolic pathways [1–3].

The term *metabolism* refers to a set of chemical transformations catalysed by enzymes. In particular, drug metabolism can involve the biotransformation of lipid-soluble xenobiotics and endogenous compounds by enzymes. These primary enzymatic transformations have the purpose to ease the elimination of endogenous and/or exogenous molecules from the human body by converting lipophilic chemicals to more hydrophilic products[1,4].

There are, however, exceptions as is the case of prodrugs. In this case, the enzymes involved in drug metabolism catalyse a reaction in order to convert the prodrug into a pharmacologically active substance [1].

The major drug-metabolizing pathways are phase I, phase II and phase III[1]. Before drug elimination, most drugs are actively metabolized. Phase I metabolism encompasses a variety of reactions such as dehydrogenation, oxidation, hydrolysis (esterases), reduction and monooxygenation; many of these reactions correspond to the introduction of oxygen atoms in the molecule. Phase I metabolites are then subject to derivatization of functional groups by glucuronidation, sulfonylation, acetylation, GSH-conjugation and methylation[1,4,5]. Phase I and II reactions are the principal via of drug metabolism in the liver, the main organ involved in biotransformation of exogenous compounds[1,3].

Phase III is responsible for the excretion of ionized and large molecules by drug transporters. This pathway includes two main protein superfamilies: 1) ATP-binding cassette (ABC) and 2) solute carrier (SLC) transporters. In the liver, the sodium-taurocholate co-transporting polypeptide, organic cation transporter 1, organic anion transporter 2, and organic anion-transporting polypeptides are the main uptake transporters, while the hepatic efflux transporters are mainly multidrug resistance protein 1, bile salt export pump and multi-drug resistance-associated protein 2 [1,6].

1.1.1. Phase I metabolism by cytochrome P450

CYPs are the most common enzymes involved in the phase I metabolism. During drug metabolism, CYPs catalyse the formation of more water soluble products by the introduction of polar groups, leading into an easily and efficient excretion by the liver and/or kidney [1,5,7].

Beyond CYP450s, other important phase I enzymes responsible for drug clearance include flavin-containing monooxygenases, monoamine oxidases, molybdenum hydroxylases, alcohol dehydrogenases, aldehyde dehydrogenases, aldo-keto reductase, NADPH:quinone reductases, and various hydrolytic enzymes[1].

Biotransformation products are usually less active than the parent drug or may be even inactive (detoxification). Nonetheless, some metabolites can have an enhanced activity or toxic effects (bioactivation)[4,5].

Usually, drug bioactivation is related to CYP-mediated oxidations. The most abundant cytochrome isoforms in the human liver is CYP3A4. Other CYP isoforms, such as CYP 2C9, 2C19, 1A2, 2B6, 2D6 or 2E1, are also involved in the bioactivation of xenobiotics[5].

Cytochrome P450 enzymes are heme-containing proteins located mostly in the endoplasmic reticulum of liver and small intestine cells, the main organs involved in the metabolism and elimination of drugs[1]. They have several endogenous functions, being involved in steroid, fatty acid, eicosanoid and vitamins A and D metabolism, and are a key player in the metabolism of xenobiotics[7,8]. The human CYP genes are the most extensive annotated mammalian genes containing more than 115 gene and pseudogene members. These enzymes are present in several tissues and organs in the human body[1,7].

The most abundant CYP enzyme expressed in the human liver is CYP3A4, followed by CYP2E1 and CYP2C9, constituting approximately 22.1%, 15.3% and 14.6%, respectively, of total CYP protein content (**Figure 1**) [1].

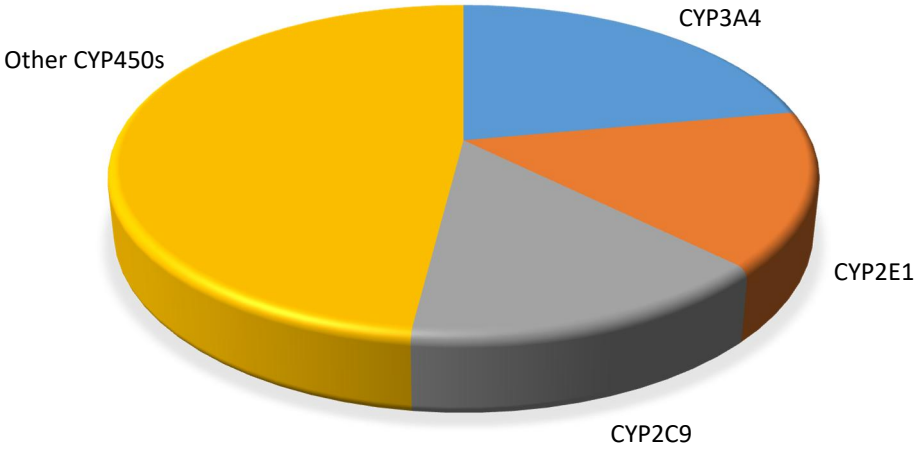

Figure 1 - Representative chart showing the expression of different CYP enzymes in the human liver. In detail, CYP3A4 (22,1%), CYP2E1 (15,3%), CYP2C9 (14,6%) and other CYPs. (Adapted from 1).

Although CYPs are involved in numerous reactions, as oxidation, sulphoxidation, aromatic hydroxylation, aliphatic hydroxylation, *N*- and *O*-dealkylation and deamination, CYPs catalyse primarily the addition of one or more oxygen atoms to the parent drugs (**Equation 1**)[1,7]..

$$NADPH + H^+ + O_2 + RH \xrightarrow{CYP} NADP^+ + H_2O + ROH \qquad \textbf{(Equation 1)}$$

where NADPH, RH and ROH are nicotinamide adenine dinucleotide (a cofactor that acts as electron donor), an oxidizable substrate and an oxidized metabolite, respectively.

Phase I drug metabolism can also lead to the reduction of compounds. This type of reaction is really useful for metabolizing aromatic nitro, nitroso, azo and *N*-oxide compounds. For this reaction to occur a secondary enzymatic system, which can be a NADH cytochrome-reductase system, is required[1].

CYP expression and activity are controlled by several factors. For instance, increased mRNA expression results in higher protein synthesis, leading to an increased total enzyme activity. Increased drug clearance can be achieved by induction of CYP enzymes (**Table 1**) leading to decrease risk of hepatotoxicity. In contrast, there are some endogenous or exogenous compounds that can inhibit CYP

enzymes (**Table 1**), leading to toxicity[1]. **Table 1** also lists the key substrates for each of the isoforms used in this work.

Table 1 – **List of well-known inhibitors, inducers and substrates for different CYPs in phase I metabolism pathway.** (Adapted from references 1, 9 and 10).

Enzymes	Inhibitors	Inducers	Substrates
CYP2C8	Gemfibrozil Fluvoxamine Ketoconazole Trimethoprim	Rifampin	Amodiaquine Buprenorphine Paclitaxel Enzalutamide[9].
CYP2D6	Quinidine	-	Tamoxifen Imipramine Codeine Risperidone[9].
CYP3A4	Ketoconazole Clarithromycin itraconzaole Saquinavir Fluconazole Grapefruit Juice tipranavir/Ritonavir Gemstone Troleoandomycin	Phenytoin Rifampin Efavirenz Etravirine Nafcillin Prednisone, Dexamethasone Phenobarbital	Ciclosporin Tamoxifen Paclitaxel Nevirapine[9].

1.1.2. Phase II metabolism

Phase I drug metabolites are enzymatically conjugated with hydrophilic endogenous cofactors with the help of transferase enzymes (phase II enzymes), resulting in more water-soluble metabolites that are easily excreted. During this drug metabolism phase a diversity of enzymes can catalyse different reactions[1,5], in particular SULTs, that catalyse the sulfonylation of hydroxyl groups.

1.1.2.1. Sulfotransferases

Up to the present date, 13 SULTs have been identified in humans and they are organized in 4 different families (SULT1, SULT2, SULT4 and SULT6)[1,10]. Regarding SULT1 family, it is divided in 4 subfamilies (1A1, 1A2, 1A3, 1A4; 1B1; 1C1 and 1C2; and 1E1) (**Table 2**)[1,10].

Table 2 – SULT1 enzymes and their function[1].

SULT1 Enzymes	Function
SULT1A1	Metabolism of phenols, alcohols, and amines
SULT1A2	Metabolism of amines (aromatic amines are the primary substrate)
SULT1A3	
SULT1B1	Metabolism of thyroid hormones and some small phenolic compounds
SULT1C1	Metabolism of iodothyronines
SULT1C2	Metabolism of 4-nitrophenols
SULT1E1	Metabolism of estrogens

Sulfotransferases catalyse reactions that transfer a sulphonyl moiety from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to oxygen in target compounds. Despite the fact that these enzymes are not as highly expressed in the body as UGTs, they are essential for metabolism of several endogenous compounds, such as steroids, serotonin eicosanoids and retinol. Moreover, these enzymes are also able to metabolize exogenous compounds[1,10,11]. SULTs are expressed throughout the entire human body, but are typically present in higher amounts in the liver, gut, breast, lung, adrenal glands and brain[1]. These enzymes can appear in two different cellular locations and, therefore, exhibit different degrees of importance in phase II drug metabolism: 1) In the cytosol and with metabolic importance, and 2) in the membrane and with less importance[1].

Sulfonylation is an important reaction in the metabolism of several xenobiotics, drugs and endogenous compounds. This enzymatic reaction increases drugs' water solubility and decreases their biological activity by adding a sulfonate moiety to the compound. However, SULT-mediated sulfonylation can also lead to the bioactivation of procarcinogens to reactive electrophiles[10].

1.2. Metabolic bioactivation and hepatotoxicity

Nowadays, it is known that drug metabolism can produce toxic metabolites that are chemically diverse, for instance, reactive electrophiles, free radicals and reactive oxygen species (ROS)[3,5]. The electrophilic intermediates formed are correlated with toxicity produced by drugs or chemicals. They can react with bionucleophilic sites in cells, potentially leading to changes in the structure and activity of the targeted macromolecules. Moreover, the electrophilic metabolites can interact with DNA inducing damage by modifications in purine and pyrimidine bases[5].

Cell damage can be caused either by the formation of radical species produced during drug metabolism, or by the production of ROS due to the interaction of reactive metabolites with oxygen. Free radicals lead to the oxidation of macromolecules, disturbing biological systems through the oxidative damage of cellular components. Accumulated oxidative damage can lead to mutagenic and/or carcinogenic processes[5].

A wide range of xenobiotics are possibly hepatotoxic. This hepatotoxicity can result from the drug itself, from a reactive metabolite produced by bioactivation of the drug, or by the reactive species formed during the process of drug detoxification. Drug-induced liver injury (DILI) represents a major clinical problem and it is a major cause for drugs being withdrawn from the market. It is estimated that 1 in 10000 to 1 in 100000 treated patients can suffer from drug-induced hepatotoxicity. However some drugs have higher risks of liver damage, for instance chlorpromazine, which has an incidence of 1 in 100 treated patients[3,12,13].

The relationship between bioactivation and hepatotoxicity is not quite understood. Taking into consideration the incidence of hepatotoxic events and an increased use of drugs, there is an unmet need to determine and find potentially damaging metabolites that are produced after drug uptake prior to patient administration. Furthermore, to discover more effective and better tolerated drugs, and to anticipate possible drug effects, mammalian and microbial models of liver metabolism can be used. The application of these models to study this problem is a promising strategy instead of *in vivo* animal models[5,14].

1.3. Mammalian and microbial models to study the risk of hepatotoxicity by drugs

A variety of experimental models have been applied to access the role of human liver enzymes in drug metabolism, such as, insect, bacterial, yeast and mammalian models[15].

1.3.1. Mammalian models

Regarding mammalian models, cell cultures are able to provide a connection between biochemical *in vitro* and *in vivo* studies. The use of this technology is growing since the pressure to diminish, refine and substitute the use of animal models is increasing. Moreover, these models allow us to determine drug safety and understand the mechanism behind its action [16–18].

The development of biotechnology permitted the isolation and culturing of many cell lines that can be applied in cell-based models to study drug metabolism. In order to test compounds toxicity *in vitro*, three basic types of cells are used, such as transformed cell lines, primary cells and pluripotent cells[16,17].

1.3.1.1. Primary Hepatocytes

The use of primary human hepatocytes was attempted since they are the most *in vitro* representative of the *in vivo* cell type. Concerning the other cell types, primary hepatocytes have some advantages such as their easy isolation with high viability and yields, and they are also the major cell type inside the liver. However, this type of cells presented some limitations namely low availability of fresh human liver samples, donor variability, complicated isolation procedures, short lifespan and the high cost[3,5,16].

1.3.1.2. Modified cell lines

As an alternative to overcome the limitations of primary human hepatocytes, the human liver cancer cell line (HepG2) was used to predict the toxicity *in vivo* through *in vitro* techniques. The usage of

transformed cell lines presents some advantages, for instance, unlimited supply and no genetic variation, which is useful for reproducibility[5,16]. Despite HepG2 cells displaying morphological features similar to liver parenchymal cells and maintaining many functions of *in vivo* hepatocytes, results cannot be extrapolated to humans because they do not behave as native hepatocytes in drug biotransformation. Moreover, the expression of CYP450 enzymes, compared to primary hepatocytes, is very small[5,16,18].

The restrictions of limited lifespan and low levels of enzyme expression in primary hepatocytes and in HepG2 cells, respectively, was surpassed by using hepatic stem cell line (HepaRG). These types of cells present higher levels of CYP expression and maintain hepatic functions of primary hepatocytes. Furthermore, they exhibit normal levels of liver-specific genes and do not show the inter-donor variability which was the main limitation in the primary hepatocytes[3,13,16].

The use of these cells appears a promising alternative for drug metabolism for *in vitro* studies. However, since it is a recent cell line, that was first described in 2002 by Gripon *et al.*, more knowledge about it is needed in order to correlate the results obtained on these cells with human liver tissues[16].

1.3.1.3. Pluripotent stem cell-derived hepatocytes

The use of stem cells-derived hepatocytes is a promising strategy to overcome limitations in hepatocyte preparation. These cells caught attention since they can provide the origin of differentiated cell lineages, and can proliferate indefinitely. The interest in the use of these cells over primary cells is due to the maintenance in culture for a longer period of time and the fact that they can be grown in a large scale[5,16].

Previous studies have shown that generating hepatocyte-like cells (HLCs) from embryonic stem cells can be successful by mimicking the development pathways of the liver during embryogenesis. It is also possible to obtain HLCs from differentiation of induced pluripotent stem cells (iPSCs) being able to mimic many hepatic functional properties[5,16].

However, stem cell technologies have some weaknesses, namely in their limitation in human stem cells availability to deliver hepatocytes[5,16].

1.3.2. Microbial models

In alternative to mammalian models, microorganisms can be applied in order to reproduce the human metabolism. Nevertheless, microbial models and especially bacterial models, are less researched and, therefore, can be a new and promising approach to create an *in vitro* model able to reproduce and/or anticipate the potential hepatotoxicity of any drug with more efficacy and safety.

The advances in molecular biology allowed stable expression of catalytically active CYPs in a variety of expression systems[19].

Previous studies performed in the *Cunninghamella* genus have shown that it is possible to perform individual expression of cytochromes P450 in a heterologous system (fungi). With those studies, it was

also possible to conclude that, using the same model, the prediction and identification of drug metabolites was also successful[15].

Other possible models to study drug metabolism include bacterial models using *Escherichia coli*. Since *E. coli* is intensively studied, it provides an ideal means of engineering bacteria for biodegradation of xenobiotic compounds[15,20]. Previous studies have shown high levels of functional expression of human CYPs in *E.coli*, which are suitable for further drug metabolism research[8,20].

1.4. Comparison between mammalian and microbial models

Comparing mammalian and microbial models, it is possible to enumerate some advantages and disadvantages of both systems related with drug metabolism.

Concerning mammalian models to mimic drug metabolism the main advantage is related with the possibility to mimic *in vitro* the *in vivo* hepatocyte functions, meaning that the results obtained are reliable and safe. The limitations of these methods depend on the type of cells applied as previously mentioned[3].

Regarding microbial systems as models for drug metabolism, the advantages comparing to mammalian models are: 1) simple, easy, low cost associated with preparation; 2) screening for a large number of proteins is a simple process; 3) high amount of metabolites formed, allowing easier detection, isolation and structural characterization; 4) maintenance of stock cultures of microorganisms is cheaper and easier comparing with cell or tissue cultures or laboratory animals, and 5) higher reproducibility[15,21].

Moreover, the prediction and identification of toxic metabolites can be done earlier, reducing the number of tests in animals[21].

The major drawback in the microorganism utilization is that drug metabolites are not always produced at the same proportion as in humans, which would be useful for further studies about drug-drug interactions and drug disposition[21]. Bacteria are also not able to post-process nascent proteins in the same way eukaryotic cells do, in particular subcellular targeting of the proteins[22]. Besides this, it is important to be aware that in a few cases no metabolism (or a very slow one) can occur due to high drug hydrophobicity meaning that the drug is indefinitely trapped in the cell membrane (lipidic compartment)[21].

1.5. Nevirapine as a model for hepatotoxicity related drugs

Nevirapine (NVP), shown in **Figure 2**, is a non-nucleoside reverse transcriptase inhibitor (NNRTI) used for the treatment of human immunodeficiency virus 1 (HIV-1) infection in humans, as part of combined antiretroviral therapy. This drug is the most prescribed NNRTI worldwide due to its high efficacy, low price and availability as a generic prescription. However, NVP is linked to skin rash and hepatotoxicity. The mechanisms undergoing NVP inducing toxicity are not fully understood, so further studies are needed. [23–27].

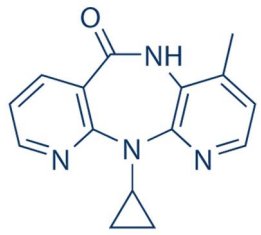

Figure 2 – Chemical structure of nevirapine.

Briefly, NVP is a dipyridodiazepinone that blocks the RNA- and DNA-dependent DNA polymerase activities. Moreover, this activity is controlled through the direct binding to an allosteric site of the reverse transcriptase.[24]

Previous studies demonstrated that NVP biotransformation in humans involves the production of several metabolites by phase I pathways (2-,3-,8-, and 12-OH-NVP, and 4-carboxy-NVP – **Figure 3**)[23,26].

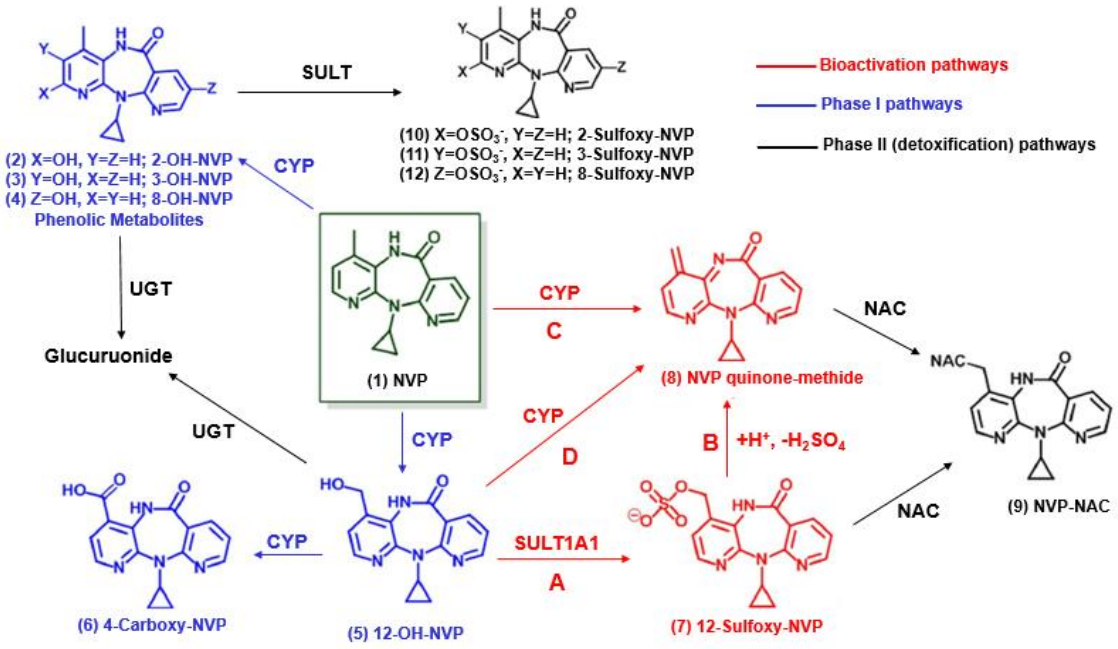

Figure 3 – Nevirapine (1) and its phase I (2-6) and phase II metabolites (7,10-12). Possible bioactivation pathways are represented in red (A, B, C and D), leading to the production of 12-sulfoxy-NVP (7) and NVP quinone methide (8). These reactive metabolites are capable of reacting with bionucleophiles, producing covalent adducts (e.g., NVP-NAC,9)[26]. NAC: *N*-acetyl-L-cysteine; UGT: uridine-5-diphospho-glucuronosyltransferase; CYP: cytochrome P450; SULT: sulfotransferase.

Previous studies confirmed that the formation of the major phase I NVP metabolite (12-OH-NVP) is catalysed by the CYP3A4 enzyme, but CYP2D6 and CYP2C9 can also be involved. Moreover, both 2-OH-NVP and 3-OH-NVP metabolites are formed by participation of CYP3A4 and CYP2B6, respectively.

The intermediate metabolites obtained from phase I metabolism are transformed by enzymes involved in phase II drug metabolism, mainly by UGTs (**Figure 3**). However, previous studies proposed that several bioactivation pathways can occur, for instance SULT1A1-mediated phase II sulfonylation of 12-OH-NVP and formation of NVP quinone methide from NVP or 12-OH-NVP and 12-sulfoxy-NVP, as shown in **Figure 3**.

1.6. Mass Spectrometry

Mass spectrometry (MS) is an analytical technique that involves the formation of ions and their separation and detection according to their mass-to-charge ratio (*m/z)* [28]. A mass spectrum is an intensity *vs. m/z* plot that represents ion distribution in a sample.

Ion generation in mass spectrometry can be accomplished in different ways. Electrospray ionization (ESI) is one of the most frequent ionization approaches to study biologically relevant molecules because it is a soft ionization technique, adequate for samples with low volatility. In ESI, a solution containing the analyte is sprayed into the vaccum chamber of the mass spectrometer ion source and the high electrical potential imparts charge to the mixture leading to evaporation of the solvent and affords charged species of the analyte. The analyte may actually acquire multiple charges through ESI, and the result is a spectrum with *m/z* peaks for each analyte. Once the ions are formed, they are separated and detected in the mass analyser component. Each of the analyte peaks of interest (e.g., metabolites) can then be selected, for ion scanning, which produces a product ion mass spectrum that is characteristic of the specific metabolite and can be compared to expected isotopic patterns for structural confirmation[28,29].

2. Materials and Methods

2.1. Chemicals

Chemicals used in this work were obtained from the following sources: quercetin was from Merck (Darmstadt, Germany). Acetylsalicyclic acid was from Fluka Chemicals (Madrid, Spain). β-NADPH Tetrasodium Salt was purchased from PanReac (Barcelona, Spain). Ibuprofen was from Acofarma (Tarrasa, Barcelona, Spain). All other chemicals were from Sigma-Aldrich (St. Louis, MO). *N*-desmethyltamoxifen, *N,N*-didesmethyl-tamoxifen and *rac*-8,14–dihydroxy-efavirenz were previously synthesized in our lab according to published protocols[30,31].

2.2. Strains, plasmids and growth conditions

Escherichia coli (*E. coli*) strain DH5α was used for plasmid maintenance and general molecular cloning procedures. In order to overexpress recombinant proteins using pET expression system, *E. coli* BL21 (DE3) (NZYTech) was used. *E. coli* strains were preserved at -80ºC in Lysogeny (LB) medium with 25% (v/v) glycerol. Prior to use, frozen cells were thawed on ice and cultured in LB fresh medium at 37ºC overnight (o/n), approximately 16 hours, with orbital agitation. Except where mentioned, all *E. coli* cultures were performed in solid or liquid LB growth medium (Liofilchem, Italy).

Bacterial strains and plasmids used in this study are listed in **Table 3**. The pCR4-TOPO CYP3A4, pCR4-TOPO CYP2D6 and pCMV-SPORT 6 SULT1A1 plasmids were provided by the Mammalian Gene Collection (MGC), PlasmidID clone ID HsCD00341290, PlasmidID clone ID HsCD00335548 and PlasmidID clone ID HsCD00346008, respectively. Storage and distribution of these plasmids was provided by the PlasmID Repository at Harvard Medical School and funded in part by NCI Cancer Center Support Grant # NIH 5 P30 CA06516. pET28a_SULT1B1 was a gift from Cheryl Arrowsmith (Addgene plasmid # 25496). pCW_CYP2C8 was a gift from Joyce Goldstein (Addgene plasmid # 69604). Antibiotics ampicillin (100 $\mu g.mL^{-1}$) and kanamycin (30 or 50 $\mu g.mL^{-1}$) were used when needed as recommended.

The vector map for pET28a(+) is presented in the Annexes (Figure A1). The vector maps for the plasmids acquired with the cDNAs of interest are also given in the Annexes (pET28a-SULT1B1 in Figure A2, pCW-CYP2C8 (data not shown), pCMV-SPORT6-SULT1A1 in Figure A3, pCR4-TOPO-CYP3A4 in Figure A4, pCR4-TOPO-CYP2D6 in Figure A5). The maps for the prepared vectors are also given in Annex – pET28a-CYP2C8 (data not shown), pET28a-CYP2D6 (Figure A6), pET28a-CYP3A4 (Figure A7) and pET28a-SULT1A1 (Figure A8).

Table 3 – **Strains and plasmids used in this study.** CYP – Cytochrome P450; SULT – Sulfotransferase.

Strain	Description	Source
BL21(DE3)	-	NZYTech
DH5α	-	NZYTech
pCCCP	Control plasmid Ampicillin resistance	NZYTech
pCW_CYP2C8	pCW expressing CYP2C8 protein. Ampicillin resistance	Joyce Goldstein *et. al., unpublished*
pCR4-TOPO CYP2D6	pCR4 expressing CYP2D6 protein. Ampicillin resistance	Harvard Medical School
pCR4-TOPO CYP3A4	pCR4 expressing CYP3A4 protein. Ampicillin resistance	Harvard Medical School
pCMV-SPORT 6 SULT1A1	pCMV expressing SULT1A1 protein. Ampicillin resistance	Harvard Medical School
pET28a_SULT1B1	pET28a expressing SULT1B1 protein Kanamycin resistance	Cheryl Arrowsmith *et. al., unpublished*
pET28a(+)	-	Novagen
pET28a_CYP2C8	pET28a expressing CYP2C8 protein Kanamycin resistance	This work
pET28a_CYP2D6	pET28a expressing CYP2D6 protein Kanamycin resistance	This work
pET28a_CYP3A4	pET28a expressing CYP3A4 protein Kanamycin resistance	This work
pET28a_SULT1A1	pET28a expressing SULT1A1 protein Kanamycin resistance	This work

Expression of recombinant proteins (CYP2C8, CYP2D6, CYP3A4, SULT1A1 and SULT1B1) was performed using the pET expression system (Novagen, Nottingham, UK). All pET28a(+) plasmids contained the coding sequence of the respective protein fused in frame with a N-terminal six histidine tail in order to allow subsequent purification by Immobilized Metal Affinity Chromatography (IMAC). These plasmids also contain a selection marker for kanamycin resistance and are inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG). The pCR4-TOPO CYP3A4, pCR4-TOPO CYP2D6, pCW_CYP2C8 and pCMV-SPORT 6 SULT1A1 have a selection marker for ampicillin resistance for *E. coli*. The temperature of bacterial growth, and the time of cell harvesting after IPTG induction that maximize the production of soluble proteins in *E. coli* cells was optimized for SULT1B1 expression.

Bacterial cells were cultured in rich solid or liquid LB growth medium, supplemented with 100 $\mu g.mL^{-1}$ ampicillin or, for cells transformed with pET28a plasmids where the medium was supplemented with 30 or 50 $\mu g.mL^{-1}$ kanamycin.

2.3. Cloning Methods

2.3.1 Plasmid DNA extraction

E. coli DH5α cells containing the plasmids (pET28a, pCW_CYP2C8, pET28a_SULT1B1, pCR4-TOPO CYP2D6, pCR4-TOPO CYP3A4 and pCMV-SPORT 6 SULT1A1) were submitted to plasmid DNA extraction according to the manufacturer's guidelines (Illustra™ plasmidPrep Mini Spin Kit (GE Healthcare, UK)). First, *E. coli* DH5α cells were inoculated in LB medium supplemented with kanamycin (50 $\mu g.mL^{-1}$) for pET28a_SULT1B1 and pET28a empty vector carrying an N-terminal His-tag, and ampicillin (100 $\mu g.mL^{-1}$) for the remaining plasmids, and cultured at 37ºC for 16 hours, and cells were harvested by centrifugation at 10 000 x *g*, during 10 minutes at 4ºC. Afterwards, plasmid DNA (pDNA) was extracted, purified and stored at -20ºC for further use. Finally, the DNA concentration and purity was spectrophotometrically by measuring absorbance at 260 nm assuming an absorbance of 1 at 260 nm corresponds to a double-standard DNA concentration of 50µg/mL[32].

2.3.2 Plasmid DNA digestion by restriction enzymes

Plasmidic DNA digestions were performed using HindIII, NdeI, NheI and XhoI enzymes (Bioron, Germany) with a final concentration of 1 µl. μg^{-1} of pDNA and following the manufacturer's protocol. Plasmidic DNA digestions were performed under optimal conditions: 37ºC for 30 minutes with slow orbital agitation, followed by one cycle at 65ºC for 20 minutes to inactivate the enzymes.

Plasmids pCR4-TOPO CYP2D6 and pCR4-TOPO CYP3A4 were digested by NheI/ HindIII, and NdeI/XhoI(SlaI) restriction, respectively. The pCMV-SPORT 6 SULT1A1 and pCW_CYP2C8 plasmids were cut using NdeI and HindIII restriction enzymes. Digestion of pET28a expression vector was performed by NheI/HindII, NdeI/XhoI(SlaI) or NdeI/HindIII, depending on the cDNA inserted.

2.3.3 Agarose Gel electrophoresis

To verify whether the pDNA digestion was properly done, pDNA fragments obtained were analysed by agarose gel electrophoresis. The gels contained 1% agarose standard (Carl Roth, Germany) dissolved in 1x Tris-Acetate-EDTA (TAE) buffer. The samples were prepared by mixing 10 µL of digested pDNA with 2 µL of RUNSAFE loading buffer with incorporated DNA staining dye (Cleaver, Scientific, UK) and 5 µL of non-digested DNA with 1 µL of RUNSAFE. The DNA Ladder (1 kbp DNA Leiter, Carl Roth, Germany) was also mixed with RUNSAFE in a proportion of 5 µL to 1 µL and ran in parallel to the samples. Electrophoresis was carried out in 1 x TAE buffer at 80V for 60 minutes and with no light exposure. Visualization of pDNA bands was performed by exposure to UV-light using PhotoDoc-It™ Imaging System Benchtop 2 UV™ Transilluminator (UVP Inc., CA, USA).

2.3.3. DNA extraction from agarose gel

The pDNA fragments of interest were cut out from the agarose gel with a clean scalpel. Extraction of the pDNA from the gel already cut was performed using the GF-1 AmbiClean Kit (PCR & Gel) (Vivantis Technologies, Malaysia) following the manufacturer's protocol.

2.3.4. Ligation of pDNA fragments

Ligation of pDNA fragments was performed using T4 DNA Ligase 10x (Bioron, Germany) with the correspondent buffer. Ligation was performed at optimal reaction conditions according with the manufacturer's protocol, with an insert: vector ratio of 5:1 and a total reaction volume of 20 µL. The ligation mixture was incubated o/n at 15ºC.

2.3.5. Polymerase Chain Reaction

Polymerase chain reaction (PCR) was performed in order to amplify pDNA fragments of CYP2D6, CYP3A4 and SULT1A1 using adequate primers. PCR was set up with One-Fusion DNA Polymerase (GeneON, Germany) using optimal reaction conditions with a cycling program adjusted to the chosen primers. PCR products were analysed by agarose gel electrophoresis and purified using the GF-1 AmbiClean Kit (PCR & Gel) (Vivantis Technologies, Malaysia).

All the primers used in this study are listed in **Table 4**.

Table 4 – **List of the primers used in this work.** CYP – cytochrome P450; SULT – sulfotransferase.

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

2.4. Preparation of *E. coli* BL21(DE3) competent cells

E. coli BL21(DE3) cells were growth o/n in LB medium without antibiotics. Cells were harvested by centrifugation at 16 000 x *g* during 5 minutes at 4 ºC. The supernatant was discarded and the cells were resuspended in 750 µl of ice cold $CaCl_2$ 0.1 M, followed by an incubation step in ice for 1h. Then, cells were reharvested and resuspended in 50 µl of fresh ice cold $CaCl_2$ 0.1 M and kept on ice at 4ºC for 16 h before transfomation[33].

2.5. Transformation of competent cells

E. coli BL21(DE3) cells were thawed on ice and transformed with 0.2 µL of pCCCP (control plasmid), 5 µL of pET28a_CYP2C8, pET28a_CYP2D6, pET28a_CYP3A4, pET28a_SULT1A1 or 5 µL of pET28a_SULT1B1, using the $CaCl_2$ classic technique[34]. After plasmid addition, samples were incubated on ice for 30 minutes and were submitted to a heat shock of 42ºC for 40 seconds. The mixture was cooled on ice for 2 minutes and then 0.9 mL of SOC medium were added in order to stabilize the

membrane and promote cell growth. Samples were incubated for 1 hour at 37ºC, with an orbital agitation of 250 rotation per minute (rpm). After incubation, 50 µL of cell suspension containing pCCCP and 150 µL of there other plasmids inocculated into 1 mL of liquid LB medium and were centrifuged at 2000 x *g* for 1 minute. The supernatant was removed and the cells were ressuspended in the remaining medium. Cells containing the pCCCP plasmid were spread in LB/ampicillin plates and the cells transformed with the remaining plasmids in LB+ kanamycin plates. Plates were incubated at 37ºC o/n and the obtained transformants were selected for ampicillin or kanamycin resistance.

2.6. Subcloning of CYP2C8, CYP2D6, CYP3A4 and SULT1A1

The DNA sequences of interest from the pCR4-TOPO CYP2D6, pCR4-TOPO CYP3A4, pCMV-SPORT 6 SULT1A1 and pCW_CYP2C8 plasmids were moved to a pET28a expression vector. The pCR4-TOPO CYP2D6 and pCR4-TOPO CYP3A4 plasmids were cut and cloned into pET28a by NheI/ HindIII, and NdeI/XhoI(SlaI) restriction enzymes, respectively. The pCMV-SPORT 6 SULT1A1 and pCW_CYP2C8 plasmids were restricted and the corresponding DNA was cloned into pET28a(+) using NdeI and HindIII restriction enzymes. The DNA fragments obtained from the digests were separated by agarose gel electrophoresis and the digested vectors and the DNA fragments of interest were extracted from the agarose gel.

2.7. Protein Overexpression of CY2C8 and SULT1B1 recombinant proteins in *Escherichia coli*

Overexpression of proteins CYP2C8 and SULT1B1 was carried out using *E. coli* BL21(DE3) as host cells and, moreover were expressed as N-terminal 6xHis-tag fused proteins.

A single colony of the transformants obtained was inoculated in 5 mL of LB medium in a 50 mL flask, supplemented with kanamycin (100 µg.mL^{-1}) and was incubated overnight. Afterwards, 25 µL of these cells were reinoculated in 5 mL of induction medium with the same conditions mentioned above, using the auto-induction expression method with the NZY Auto-Induction Kit and following the respective protocol. The protein expression induction was also performed by adding IPTG at a final concentration of 1mM to the medium. Using the auto-induction medium (LB auto-induction medium), which consists of LB supplemented with glucose (0.5 g/L) and α-lactose (2.0 g/L), the cell suspensions were incubated at 37ºC o/n for 24 hours with agitation (250 rpm). For IPTG induction, the cells were incubated at 37ºC for *ca.* 4h, until they reached an optical density (OD) at 600 nm of approximately 0.5[35–37].

2.8. Cell Lysis

Cells were harvested from liquid cultures by centrifugation (5 000 x *g* for 10 minutes at 4ºC). Supernatants were discarded and the pellets weighted. Cells were resuspended using 5 mL of NZY Bacterial Cell Lysis Buffer per gram of cell paste, and 2 µL of lysozyme (Sigma-Aldrich, MO, USA) at 50 mg.mL^{-1} and 2 µL of DNAse I (Sigma-Aldrich, MO, USA) at 2 mg.mL^{-1} were added per mL of NZY

Bacterial Cell Lysis Buffer. The cell suspension was incubated at room temperature (RT) for 15 minutes and the insoluble cell debris was removed by centrifugation step (15 000 x *g* for 15 minutes at 4ºC).

2.9. Purification by Immobilized Metal Affinity Chromatography

Since the proteins expressed in this project included an N-terminal 6 x His tag, the purification process chosen was IMAC, using a nickel coated column (HisTrapTM FF Crude – GE Healthcare, UK). The column was first washed with 5 column volumes of water and afterwards equilibrated with 5 column volumes of binding buffer (20 mM sodium phosphate, 0.5 M NaCl, 40 mM imidazole pH 7.4). Cell lysate was loaded on the column and it was washed with 6 column volumes of binding buffer. Protein elution was performed by increasing the imidazole concentration from 40 mM to 500 mM and washing the column with 6 column volumes of elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole pH 7.4). Absorbance at 280 and 320 nm was recorded for the eluted fractions (5 mL each).

The purity of collected protein fractions was assessed by SDS-PAGE analysis.

2.10. Desalting

An additional step of purification was performed in order to remove small contaminants, essentially salts by size exclusion chromatography. PD-10 Desalting columns containing Sephadex G-25 resin (GE Healthcare, UK) were used to rapidly clean up the proteins (> 5000 molecular weight) from CYP2C8 and SULT1B1 protein samples. The experiment was performed following the manufacturer's instructions for gravity protocol. The elution of CYP2C8 and SULT1B1 proteins was performed using 1x PBS pH 7.4 buffer and 1x PBS, 2.5 mM $MgCl_2$ pH 7.4 buffer, respectively. The purity of collected protein fractions was assessed by SDS-PAGE analysis. The purest fractions were joined and concentrated using Amicon Vivaspin Filters with a membrane cutoff of 10000 kDa (Sartorius, Göttingen, Germany) according to manufacturer's instructions.

2.11. Protein Quantification

The concentration of solubilized protein obtained was determined by a simple and accurate procedure performing Bio-Rad Protein Assay (Bio-Rad Laboratories, USA), based on the Bradford method and following the standard procedure according to manufacturer's instruction. Upon mixing the reagent with the samples (5 mL of reagent and 100 µL of sample), reaction mixtures were incubated at room temperature for 5 minutes and absorbance at 595 nm was determined. A standard curve was made in the same conditions of the experiments using bovine serum albumin (BSA) at concentrations of 0.2, 0.4, 0.6, 0.8 and 1.0 $mg.mL^{-1}$. The calibration equation obtained is:

$$\text{Absorbance at } 595 \text{ nm} = 1.1441 \times [\text{BSA}](\text{mg.mL}^{-1}) + 0.5652 \quad \textbf{(Equation 2)}$$

Concentration of experimental samples was determined using this **Equation 2**.

2.12. 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[38].

The lysate pellets were resuspended in 200 µL of dH_2O followed by vortex. 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. Enzyme Activity Assays

2.13.1. Sulfotransferase 1B1 activity measurement

The activity of SULT1B1 was assessed using the 2-naphthol sulfonation assay, as described by Frame *et al*[39]. Other experiments were performed by changing the 2-naphthol for other chemicals in order to characterize the activity of SULT1B1 in the presence of those chemicals. SULT1B1 enzyme was used at 11.2 $mg.mL^{-1}$ and 0.224 $mg.mL^{-1}$ for cuvette and microplate assays, respectively. Substrates were used with a final concentration ranging from 0 to 50 µM. The reactions occurred at 37ºC with slow agitation for 4 hours. Both microplate and cuvette assays were performed. Firstly, SULT1B1 activity was assessed using the test tube scale and 96-well plates, by measuring the absorbance at 405 nm in a spectrophotometer at different times of incubation. In the microplate experiments, all the compounds were pipetted to flat bottom 96-well plates and the SULT1B1 activity was determined by a spectrophotometer following the absorbance at 405 nm in different times of the experiment. The enzyme activities in test tube scale and measurements in spectrophotometric cuvette were determined by the same type of protocol, except all volumes were four times larger than in microplate assay and hence total reaction volume was 1000 µL instead of 250 µL. Activities were measured with a UV-1800 spectrophotometer (Shimadzu, Japan).

2.14. SULT1B1-mediated sulfonylation studies by LC/MS

The standard assay mixture contained 5 µM PAPS solution, 2.5 mM potassium 4-nitrophenyl sulfate (KNPS) solution, 100 µM substrate, and purified SULT1B1 protein (final concentration in the assay 2.8 $mg.mL^{-1}$) in 50 mM ammonium bicarbonate pH 6.5, 5 Mm $MgCl_2$, in a final volume of 500 µL. Mixtures were incubated at 37ºC for 5 hours with slow orbital agitation and, centrifuged at 10000 x g, 4ºC for 5 minutes in Amicon Vivaspin filters with a membrane cutoff of 10 kDa (GE Healthcare, UK). The permeated fraction was analysed by liquid chromatography-tandem high resolution mass spectrometry (LC/MS). Several chemicals were tested as possible SULT1B1 substrates, namely, resorcinol,

quercetin, 1,5 – dihydroxyanthraquinone, 1,8 – dihydroxy-anthraquinone, phenol, acetaminophen, α-hydroxytamoxifen, *N*-desmethyltamoxifen, rac 8,14-dihydroxy-efavirenz, cinchonine and antipyrine.

2.15. CYP2C8-mediated oxidation metabolism studies by LC/MS

The standard assay mixture contained 30 μM NADPH, 30 μM substrate, and purified CYP2C8 protein (final concentration in the assay 0.312 $mg.mL^{-1}$) in 50 mM Tris buffer pH 8.0 in a final volume of 1 mL. The mixtures were incubated at 37ºC for 4-5-hours with slow orbital agitation and next, centrifuged at 10000 x g, 4ºC for 5 minutes in Amicon Vivaspin filters with a membrane cutoff of 10 kDa (GE Healthcare, UK). The permeated fraction was analysed by liquid chromatography-tandem high resolution mass spectrometry (LC/MS). Several chemicals were tested as possible CYP2C8 substrates, namely, tamoxifen, *N*-desmethyltamoxifen, nevirapine, arachidonic acid and novocaine.

2.16. Liquid chromatography-tandem high resolution mass spectrometry (LC/MS) analysis

Samples were analysed by liquid chromatography (Ultimate 3000 RSLCnano system, ThermoFisher Scientific, San Jose, CA, USA) interfaced with a Bruker Impact II quadrupole time-of-flight mass spectrometer equipped with an electrospray source (Bruker Daltoniks, Bremen, Germany). Chromatographic separation was carried out on a HypersilGold C18 column (2.1 mm x 150 mm, 1.9 μm particle size; ThermoFisher Scientific, San Jose, CA, USA). The mobile phase consisted of water containing 0.1% formic acid (A) and the acetonitrile containing 0.1% formic acid (B), at a flow rate of 200 μL/min. The elution conditions were as follows: 5% B for 2.4 min; 5-25% B for 2.1 min; 25-70% B for 4.1 min; 70-100% B for 6 min. The column and the autosampler were maintained at 37ºC and 4ºC, respectively.

The high resolution mass spectra were acquired in both positive and negative ion modes. The mass spectrometric parameters were set as follows: end plate offset: 500 V; capillary voltage: 4.5 and -2.5 kV (positive and negative mode, respectively); nebulizer: 40 psi; dry gas: 8L/min; heater temperature: 200ºC. Internal calibration was achieved with a sodium formate solution introduced to the ion source *via* a 20 μL loop at the beginning of each analysis using a six-port valve. Calibration was then performed using a high-precision calibration mode (HPC). Acquisition was performed in full scan mode in the *m/z* 50 – 1000 range with an acquisition rate of 5 Hz using a dynamic method with a fixed cycle time of 3 s. Dynamic exclusion duration was 0.4 min.

2.16.1. Data Processing

Acquired data was processed by DataAnalysis 4.1 software (Bruker Daltoniks, Bremen, Germany). This particular analysis allows searching for compounds having the same number of atoms (by their molecular mass) with a relevant isotopic pattern. All spectra corresponding to either substrates or hypothetical metabolites were then manually checked. Other intense ions in the full mass spectrometry (MS) spectra were also analysed. Ions with a mass-to-charge ratio (*m/z)* deviation values lower than 5

ppm were considered acceptable for positive identification. Isotope cluster analysis was also performed in order to validate the identity of the species of interest.

3. Results and Discussion

3.1. Cloning procedures

Plasmid pET28a(+) (Novagen) was used as the expression vector. Standard genetic engineering techniques were used to generate the recombinant vectors containing the human CYP2C8, CYP2D6, CYP3A4 and SULT1A1 cDNAs fused to an N-terminal 6xHis-Tag, present in the pET28a plasmid backbone.

The CYP2C8 cDNA was excised NdeI/HindIII from a recombinant vector derived from pCW ori+ (created by Joyce Goldstein, unpublished, and deposited in the Addgene repository) and was subcloned into an NdeI/HindIII treated pET28(+)vector. To this end, both plasmids were sequentially digested with the restriction enzymes under optimal conditions as previously mentioned in the Materials and Methods section. The DNA fragments obtained from each digest were separated by agarose gel electrophoresis (data not shown), and the bands of interest (of approximately 1.5 kbp for the insert, and 5.5 kbp for the linearized vector) were cut an purified from the agarose fragment using the agarose gel purification kit as previously mentioned in the Materials and Methods section.

The other cDNAs were PCR-amplified from plasmids belonging to human cDNA libraries from the PlasmID Repository of the DF/HCC DNA Resource Core from Harvard Medical School. To this end, pairs of primers of ~30 nt were designed containing specific restriction sites. The forward primers were designed inserting an NdeI (for CYP3A4 and SULT1A1) or NheI (for CYP2D6) restriction sites positioned to allow the in-frame N-terminal tag fusion. The reverse primers were designed to maintain the STOP codon and inserting afterwards a restriction site for HindIII (for CYP2D6 and SULT1A1) or for XhoI (for CYP3A4). PCR reactions were done using Phusion DNA polymerase, according the manufacturer's instructions, and with annealing temperatures that were optimized for each PCR product. PCR reactions were analyzed by agarose gel electrophoresis (data not shown), DNA fragments were of approximately 1.5, 1.5 and 1.0 kbp length, corresponding to the expected size of CYP2D6, CYP3A4 and SULT1A1 gene coding sequences, respectively. The PCR products were purified with a PCR-purification kit and were then incubated under optimal conditions, with restriction enzymes, as previously mentioned in the Materials and Methods section. Following similar procedures as those described for the preparation of the NdeI/HindIII treated pET28(+)vector, the pET28(a)+ vector was linearized by sequential double digestion with NdeI/XhoI or NheI/HindIII and the DNA fragments were extracted and purified from an agarose gel.

The DNA fragments (inserts and vectors) obtained from each digest were quantified by their absorption at 260 nm (**Table 5**) and ligated with T4 DNA ligase, at molar ratios of 5 or more times insert to vector. *E. coli* DH5α cells were transformed with the ligation reactions and selected in the presence of kanamycin, the vector selective marker. In general, all ligations attempted resulted in the presence of *E. coli* transformants, but naturally not all were positive clones. Confirmation of the insert integration was done by restriction gel electrophoresis after purification of the recombinant plasmids or by colony PCR using T7 promoter and T7 terminator primers that anneal in pET28 backbone (data not shown).

Recombinant vectors were successfully prepared for CYP2C8, while for SULT1A1, CYP2D6 and CYP3A4 it was not possible until now to isolate positive transformants.

Table 5 - DNA fragments (inserts and vectors) concentration obtained from each digest.

DNA digests	[DNA] ng/µL
pET28 (NdeI / HindIII)	10.7
pET28 (NheI / HindIII)	37.7
pET28 (NdeI / XhoI)	16.6
SULT1A1 (NdeI / HindIII)	36.5
CYP2C8 (NdeI / HindIII)	9.00
CYP2D6 (NheI / HindIII)	61.3
CYP3A4 (NdeI / XhoI)	20.6

3.2. Expression and purification of CYP2C8 and SULT1B1 proteins

The overexpression of the recombinant CYP2C8 and SULT1B1 proteins was carried out using *E. coli* BL21 (DE3) as host cells and using the pET expression system. Induction of plasmid expression was carried out by addition of IPTG to the bacterial cell culture and by auto induction medium.

The genes to be expressed from the recombinant plasmid have a fused N-terminal 6x His-tag, allowing the desired proteins purification by IMAC; in order to remove salts from samples, an additional step of desalting was performed by size-exclusion chromatography.

To assess protein induction with IPTG for both CYP2C8 and SULT1B1 proteins, electrophoresis under denaturing conditions (SDS-PAGE) was performed. Protein expression after induction was analysed; **Figure 4** presents typical results for SULT1B1 and CYP2C8 expression.

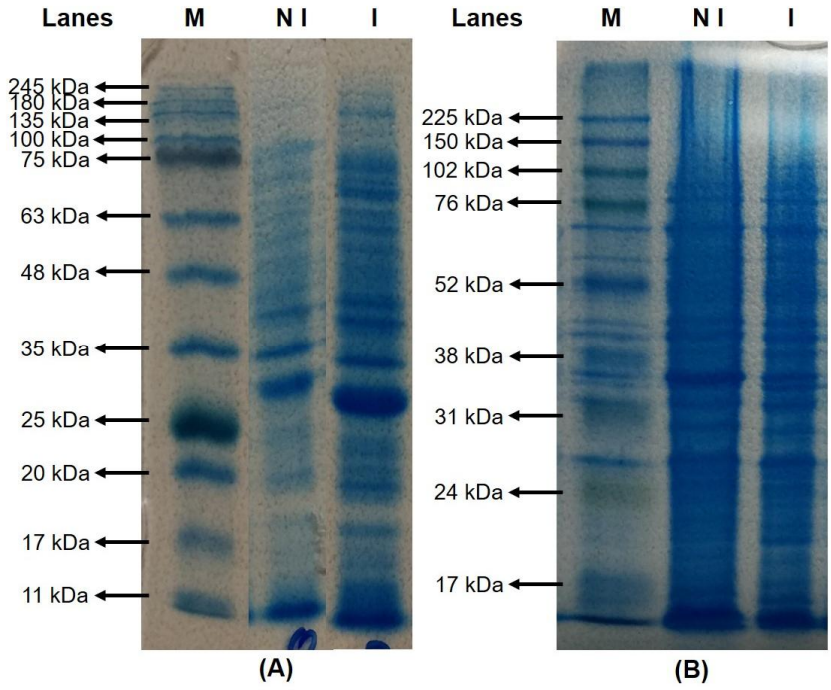

Figure 4 - SDS-PAGE analysis of target protein expression induction in *E. coli* BL21(DE3) cells. (A) *E. coli* BL21(DE3) cells containing SULT1B1 plasmid were analysed prior to (NI, not induced) and after induction (I, induced) with 1 mM IPTG. (M) Molecular weight marker Protein Marker II (Nzytech); (B) *E. coli* BL21(DE3) cells containing CYP2C8 plasmid were analysed prior to (NI) and after induction (I) with 1 mM IPTG. (M) Molecular weight marker Amersham™ ECL™ Rainbow Marker – Full Range RPN800E (GE Healthcare).

The addition of IPTG to the bacterial cell culture should increase the expression of the target proteins since pET vector has a T7 promoter system. The coding sequence for T7 RNA polymerase is present in the chromosome under control of the inducible *lacUV5* promotors in hosts such as BL21 (DE3) used in this project. In these systems, the gene of interest is cloned behind a T7 promoter recognized specifically by the phage T7 RNA polymerase. When the *lacUV5* promoter is not induced, insignificant amounts of T7 RNA polymerase or target protein should be present and in general the cells should grow. Contrariwise, when an inducer, for instance, IPTG is added to the bacterial culture, the synthesis of T7 RNA polymerase increases, leading to transcription of the DNA controlled by the T7 promoter[36,40,41]. In this work, induction of SULT1B1 protein expression by adding IPTG to the bacterial culture medium led into an increase in the expression of the target protein shown by a significantly more intense band between the 25 and 35 kDa comparing with the same band in the *E. coli* cells analysed before induction (**Figure 4A**); human SULT1B1 has a molecular mass of 34.9 kDa, in agreement with the observed band[42]. In addition to a higher intensity in the band corresponding to the protein of interest, an increase in intensity was observed in all bands corresponding to an increase in the expression of all the proteins in the *E. coli* cells.

Optimization of SULT1B1 expression was attempted by changing culture medium, time of incubation and antibiotic concentration; whole-cell samples of cultures in the different conditions were collected and analysed by SDS-PAGE (**Figure 5**).

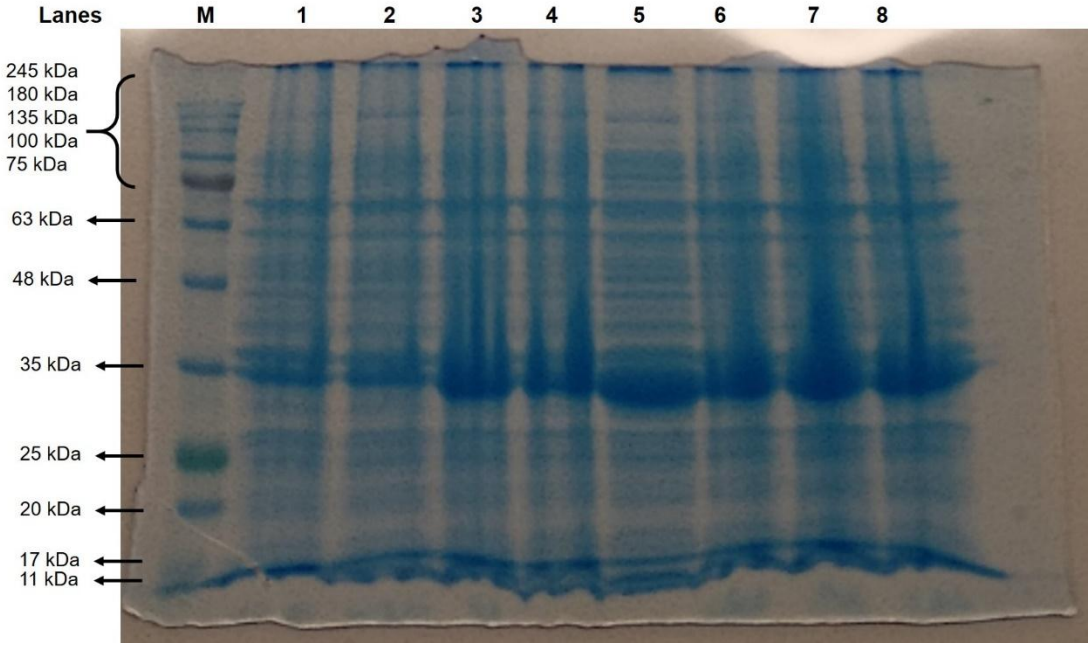

Figure 5 - SDS-PAGE analysis of target SULT1B1 protein expression in *E. coli* BL21(DE3) cells. (M) Molecular weight marker Protein Marker II (Nzytech). (1) LB medium supplemented with kanamycin 30 μg.mL^{-1} with 4h of growth, (2) LB medium supplemented with kanamycin 50 μg.mL^{-1} with 4h of growth, (3) LB medium supplemented with kanamycin 30 μg.mL^{-1} with 16h of growth, (4) LB medium supplemented with kanamycin 50 μg.mL^{-1} with 16h of growth, (5) LB medium supplemented with kanamycin 30 μg.mL^{-1} with 24h of growth, (6) LB medium supplemented with kanamycin 50 μg.mL^{-1} with 24h of growth; (7) LBAI medium supplemented with kanamycin 30 μg.mL^{-1} with 24h of growth; (8) LBAI medium supplemented with kanamycin 50 μg.mL^{-1} with 24h of growth.

By analysing **Figure 5**, it is possible to observe that in all conditions there was an overexpression of the target protein, visible by an intense band around 34.9 kDa, corresponding to the molecular weight of SULT1B1. Moreover, all lanes present other bands which correspond to expression of all host proteins. It is also possible to observe that protein expression is higher when the cells grow for 24h (**Figure 5**, lanes 5 and 6) comparing with the other incubation times (**Figure 5**, lanes 1 to 4). The concentration of kanamycin also influences protein expression, being visible that a concentration of 30 μg.mL^{-1} (**Figure 5**, lanes 1, 3 and 5) is preferable comparing with 50 μg.mL^{-1} (**Figure 5**, lanes 2, 4 and 6), since the band around 34.9 kDa is more intense in lower concentrations of antibiotic. Moreover, the LB auto-induction (LBAI) medium (**Figure 5**, lanes 7 and 8) does not appear to be advantageous in relation with the LB medium since the bands are less intense. The conditions that present higher expression levels are evident in lane 5 (**Figure 5**), since the band around 34.9 kDa is wider and more intense and so the optimal conditions for protein expression are LB medium supplemented with kanamycin at a final concentration of 30 μg.mL^{-1} and with 24h of cell growth. These results do not exclude the possibility to investigate other conditions in order to achieve a higher amount of produced protein.

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 addition of the inducer (**Figure 4B**). 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 an harmful effect that heterologous proteins exert in the cells[36,40,43].

Since CYPs are able to 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[36,40]. 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[36,40,41].

To test this hypothesis, cultures of *E. coli* transformed with the pET-28-CYP2C8 plasmid were also performed in LB auto-induction medium, 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[35]. 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.

Results obtained for CYP2C8 during expression in *E. coli* can also be explained by the production of the target protein in inclusion bodies. Several factors can influence inclusion bodies formation, such as, pH, osmolarity, redox potential, cofactors and folding mechanisms because the expression of the recombinant protein is performed in different conditions comparing to the original source. These factors can lead to protein instability and aggregation. The high-translational rates required for overexpression can lead to saturation in folding mechanisms, and can be overcome by the co-expression molecular chaperons, by the supplementation of the culture medium with osmolytes (proline, glycine-betaine, and trehalose) which are chemical chaperons, and by using other cofactors (for example, iron-sulfur and magnesium) to achieve the correct final conformation and stabilization of the target protein [40,44,45].

Another approach to avoid inclusion bodies formation it is based in slowing down production rate giving the necessary time for the transcribed recombinant proteins to fold properly. Low temperatures also can be used in order to decrease aggregation since hydrophobic interactions are temperature dependent. Actually, as previously mentioned, CYP2C8 protein expression was performed without any of these supplementations and at 37ºC, which agrees with the possibility of having the target protein in inclusion bodies. Cultures performed at room temperature (*ca.* 25 °C) for 36 hours did not exhibit any inclusion bodies, but no significant expression was observed (data not shown).

Furthermore, the overexpression of heterologous proteins in this study was performed in BL21(DE3) strains which can be an issue for plasmid stability leading to lower expression levels of the target protein. To reduce the plasmid instability for toxic proteins, different expression strains are being tested, such as either C41(DE3) or C43(DE3) strains, which are mutant host strains from BL21(DE3). These mutant strains were selected because they grew to high saturation densities and continue to produce proteins at an elevated level without toxic effects[43].

As already mentioned, both proteins were purified by IMAC. A representative chromatogram for SULT1B1 (**Figure 6A**) and CYP2C8 (**Figure 6B**) purification was done by measuring the absorbance of the eluted fractions at 280 and 320 nm and the concentration of imidazole. To evaluate protein elution during this purification process and purity of the eluted fractions for both CYP2C8 and SULT1B1 proteins with different imidazole concentrations, a SDS-PAGE bisacrylamide gel followed by protein staining was performed. A representative gel of the 5 mL eluted fractions of SULT1B1 and CYP2C8 are shown in **Figure 6C** and **Figure 6D**, respectively.

Regarding SULT1B1 purification by IMAC, from the analysis of the chromatogram (**Figure 6A**) it is possible to observe that the absorbance at 280 and 320 nm abruptly increases in the beginning of the experiment where low concentrations of imidazole (40 mM) were used in order to minimize binding of host cell proteins. Since proteins have a characteristic maximal absorption at 280 nm, the obtained peak corresponds to the elution of all host proteins that do not bind the nickel-based column. Since histidine is the amino acid that presents the strongest interaction with immobilized ion matrices, the histidine-tagged protein SULT1B1 is retained in the column[46].

When the absorbance at 280 nm reaches low levels, around 30 to 35 mL of elution volume, the buffer is changed to one with a higher concentration of imidazole (500 mM). Imidazole is a histidine analogue and competitively interacts with the stationary phase leading to SULT1B1 protein elution as shown in **Figure 6A** by a peak in absorbance at 280 nm around 35 to 50 mL of elution volume[46]. The absorbance after His-tagged protein elution decreases and stabilizes, meaning that all proteins have already been eluted from the column. The absorbance at 320 nm was also followed to check other contaminants besides proteins and most contaminants were removed at the beginning of the elution around 5 to 20 mL of elution volume since in the rest of the elution process the values were almost insignificant.

To verify protein elution and purity during SULT1B1 purification process, eluted fractions were analysed by SDS-PAGE, as presented in **Figure 6C**. Analysing **Figure 6C** it is possible to observe that in lane 1, which corresponds to the cell lysate injection (5 mL of elution volume in **Figure 6A**) a smear is present since all the host proteins that do not present affinity to the nickel column were eluted. The lanes 2 and 3 correspond to the elution volumes of 10-15 and 25-30 mL, respectively. In this case, proteins that do not present affinity to the nickel column continue to be eluted and there is any presence of an intense band corresponding to SULT1B1 protein which has a molecular weight around 34.9 kDa. The samples analysed in lanes 4, 5 and 6 (**Figure 6C)** correspond to fractions 35-40, 40-45 and 50-55 mL of elution volume obtained from purification process. A band around 34.9 kDa which is in agreement with the theoretical molecular weight of SULT1B1 protein, is present in lanes 4, 5 and 6 (**Figure 6C)**, but the intensity is lower in lane 6 and in the remaining the same band presents higher intensity. It is important

to note that in lane 4 (**Figure 6C)** although there is an intense band at the molecular weight of SULT1B1 protein, other bands at other molecular weights are shown, which indicates the presence of other contaminating proteins. The lanes 5 and 6 (**Figure 6C)** already present less protein contaminants since there are almost no other bands and so purification of SULT1B1 protein was successful.

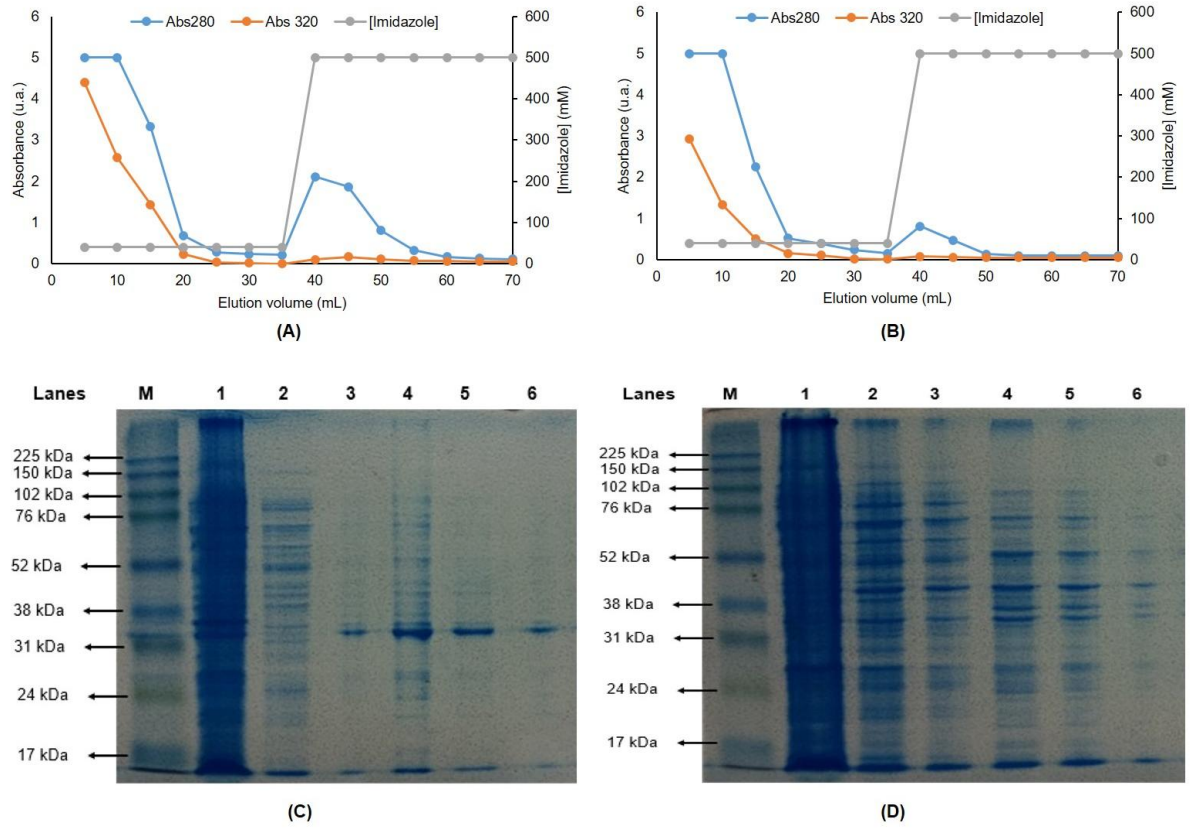

Figure 6 - Analysis of the elution profile of the recombinant his-tagged SULT1B1 and CYP2C8 proteins, purified by Ni^{2+} - based IMAC. A) Chromatogram of SULT1B1 protein purification by IMAC. The blue, orange and grey lines represent the absorbance at 280 nm, absorbance at 320 nm and the concentration of imidazole. **B)** Chromatogram of CYP2C8 protein purification by IMAC. The blue, orange and grey lines represent the absorbance at 280 nm, absorbance at 320 nm and the concentration of imidazole. **C)** 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. 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. **D)** 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. 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.

CYP2C8 protein was also purified by IMAC and the elution chromatogram obtained is represented in **Figure 6B**. The pattern of protein elution is similar to SULT1B1 elution (**Figure 6A**) which begins with an absorption peak at 280 nm at low concentrations of imidazole (40 mM) corresponding to host cell proteins elution that do not present affinity to the column matrix. Since CYP2C8 protein presents a His-tag tail it binds the nickel-loaded column matrix. When the absorbance at 280 nm reaches a plateau which is around 25 to 35 mL of elution volume, the imidazole concentration is raised to 500 mM. The

change of the buffer led to elution of His-tagged proteins around 35-50 mL of elution volume, since the free imidazole competes for the interaction with immobilized ion matrices. A plateau state is once again achieved after CYP2C8 elution since all the proteins were already eluted from the column. The absorbance at 320 nm was also followed to check other contaminants besides proteins and most contaminants were removed at the beginning of the elution, around 5 to 20 mL of elution volume since in the rest of the elution process the values measured were almost insignificant.

Once again, a SDS-PAGE was performed in order to evaluate protein elution and purity for CYP2C8 purification process (**Figure 6D**). It is possible to observe by analysing **Figure 6D**, lane 1, a smear formed by the presence of several intense bands which correspond to the elution of all proteins that do not present affinity to the nickel column. In lane 2 and 3 (**Figure 6D**), which correspond to the elution volumes of 10-15 and 25-30 mL respectively, several bands are present meaning that the host proteins continue to be eluted in these fractions. Importantly, no intense bands around 52 kDa in these lanes were detected, meaning that the CY2PC8 protein is not eluted yet, being immobilized in the column matrix. In order to elute the target protein, the elution buffer was changed as already mentioned. The fractions 35-40, 40-45 and 50-55 mL of elution volume which correspond to the elution buffer with higher imidazole concentration were analysed in lanes 4, 5 and 6 (**Figure 6D**), respectively. In these lanes, a band around 52 kDa, which is the same molecular weight as CYP2C8 protein, can be observed. In lane 6 (**Figure 6D**), there are less contaminating proteins since there are fewer bands, although the 52 kDa band appears with lower intensity comparing with lanes 4 and 5. In lanes 4 and 5 the band around 52 kDa has higher intensity but there are several other bands corresponding to contaminating proteins. In summary, the eluted protein was not pure, as multiple bands on SDS bisacrylamide gel can be observed (**Figure 6**), which can mean that we should wash the column with more volumes of binding buffer in order to let the contaminant proteins elute before eluting the protein of interest. Also, some native proteins present in *E.coli* strain, such as BL21(DE3), present affinity for metal chelating resins commonly used in IMAC. The binding of these proteins is determined by several factors, for example, accessibility of surface histidine residues to the metal ions present in chelating resins, local conformations and cooperation between neighbour amino acid side groups[47,48]. If the contaminants are in fact interacting with the column, the solution may be eluting the proteins in a stepwise gradient with the main goal of separating successfully the target protein from the contaminating proteins.

Comparing both chromatograms (**Figure 6A** and **Figure 6C**) it is possible to observe that the purification process was quite similar for both proteins. The major difference was in the values of absorbance at 280 nm obtained for each protein, since SULT1B1 protein reached a maximum value of 2.113 units of absorbance while CYP2C8 protein only achieved 0.812, meaning that SULT1B1 protein is present in higher concentration. Beyond the low absorbance at 280 nm obtained for CYP2C8 protein, the presence of other contaminating proteins can contribute to the value of absorbance measured, meaning that the target protein may be present at very low concentrations.

An additional step of desalting was performed in order to increase proteins purification, and representative gels of the 3.5 mL of eluted fractions of SULT1B1 and CYP2C8 are shown in **Figure 7A** and **Figure 7B**, respectively.

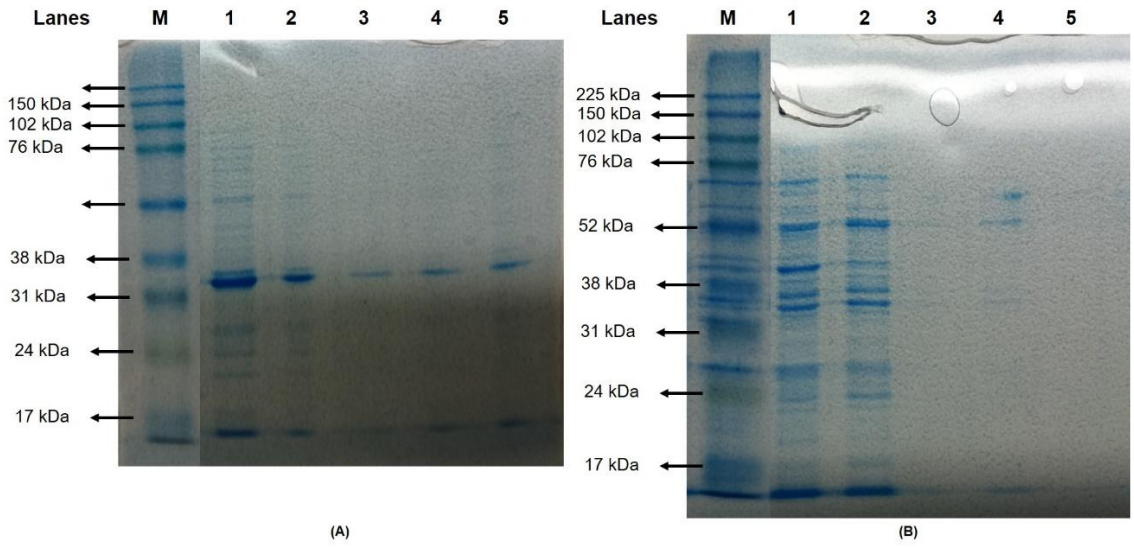

Figure 7 - Analysis of the elution profile of the recombinant SULT1B1 and CYP2C8 proteins, purified by desalting process. A) SDS-PAGE analysis of the elution profile for the SULT1B1 protein by desalting. Gel was stained with EzBlueTM Staining Reagent (Sigma-Aldrich, St. Louis, MO). M – Molecular weight marker Amersham^{TM} ECL^{TM} Rainbow Marker – Full Range RPN800E (GE Healthcare). Lane 1 – Desalting fraction 2 (6 mL). Lane 2 – Desalting fraction 1 which corresponds to sample injection (2.5 mL). Lane 3 – Desalting fraction 3 (9.5 mL). Lane 4 – Desalting fraction 4 (13 mL). Lane 5 – Desalting fraction 5 (25 mL); **B)** SDS-PAGE analysis of the elution profile for the CYP2C8 protein by desalting. Gel was stained with EzBlueTM Staining Reagent (Sigma-Aldrich, St. Louis, MO). M – Molecular weight marker Amersham^{TM} ECL^{TM} Rainbow Marker – Full Range RPN800E (GE Healthcare). Lane 1 – Desalting fraction 2 (6 mL). Lane 2 – Desalting fraction 1 which corresponds to sample injection (2.5 mL). Lane 3 – Desalting fraction 3 (9.5 mL). Lane 4 – Desalting fraction 4 (13 mL). Lane 5 – Desalting fraction 5 (25 mL).

The desalting column consisted of a packed well-known size exclusion medium Sephadex G-25 superfine resin. The medium is based on cross-linked dextran beads which allow excellent resolution and high flow rates. The main goal of this step in the purification of SULT1B1 and CYP2C8 was to remove imidazole from both samples. Imidazole is an aromatic molecule that can adsorb to the column due to anion-π interactions with the matrix OH groups. Since both proteins have high molecular weight they are not retained in the matrix being the first ones to be eluted while the imidazole, because of the interaction with the matrix will migrate later[49].

The SULT1B1 sample was applied to the desalting column in a volume of 2.5 ml represented in the **Figure 7A**, lane 2; eluted fractions were recovered at 6 mL, 9.5 mL, 13 mL and 25 mL, represented in lanes 1, 3, 4 and 5, respectively (**Figure 7A**). Since SULT1B1 protein is not retained in the column, it is immediately eluted at the beginning of the process as shown in **Figure 7A**, lanes 1 and 2, where there are intense bands around 34.9 kDa which correspond to the molecular weight of the target protein. In later fractions some protein was also collected, but in less quantity, as can be seen by the presence of less intense bands at the same molecular weight (lanes 3, 4 and 5 in **Figure 7A**).

Concerning CYP2C8 protein purification, the sample was injected in a volume of 2.5 ml represented in the **Figure 7B**, lane 2 and since the protein is not retained in the column it is eluted immediately in the mobile phase which is in accordance with the gel analysis (**Figure 7B**, lane 2), where there is an intense band around 52 kDa corresponding to the molecular weight of the target protein. Lanes 1, 3, 4 and 5 in **Figure 7B** correspond to 6 mL, 9.5 mL, 13 mL and 25 mL of mobile phase volume passed through the column. Since the protein sample was already contaminated with other proteins that have a molecular

weight above the exclusion size of this matrix, those proteins were also eluted at the beginning of the purification step with the target protein (52 kDa), seen by the presence of an intense band at 52 kDa and other several bands with different molecular weights in lanes 1 and 2 (**Figure 7B**). Lanes 3, 4 and 5 in **Figure 7B** do not present bands, meaning that proteins were not retained in the column matrix.

In summary, the attempt to overexpress both recombinant proteins was achieved for SULT1B1, although CYP2C8 overexpression seemed not be so successful. An additional technique that should be performed in order to confirm the expression of the desired protein is a western blot using an anti-histidine six-tag specific antibody[50]. This technique appears as an additional tool to guarantee that the right protein was produced. The purification of SULT1B1 was quite successful while in CYP2C8 other contaminating proteins were present in the final sample. Once more, it would be necessary to optimize purification and expression processes in order to obtain the highest amount of protein production and the best possible protein purity.

3.3. Enzyme Activity Assays

The activity of SULT1B1 was assessed using the 2-naphthol sulfonation assay, a model substrate of the enzyme (**Figure 8**). This method also uses *p*-nitrophenol sulfate as sulfate donor to regenerate the active form of the cofactor and lift the 3'-phosphoandenosine-5'-phosphate-inihibition of the enzyme[39,51]. In this assay, SULT1B1 enzyme catalyses the synthesis of 2-naphthyl sulfate from 2-naphthol and 5'-phosphoadenosine 3'- phosphosulfate (PAPS). The addition of *p*-nitrophenyl sulfate to the experiment leads to an effective PAPS-regenerating system. The produced 3' – phosphoandenosine 5'-phosphate (PAP) serves as a cofactor to remove the sulfate group from *p*-nitrophenyl sulfate and regenerate PAPS. This reaction produces *p*-nitrophenol that is quantified colorimetrically at 405 nm, giving an indirect measure of sulfotransferase activity. By knowing that these enzymes catalyse the transfer of a sulfo moiety from PAPS to the acceptor substrate, it is possible to change the 2-naphthol for other substrates in order to study the activity of SULT1B1 enzyme in other conditions. The molar extinction coefficient used was 18200 $M^{-1}.cm^{-1}$ [39,52].

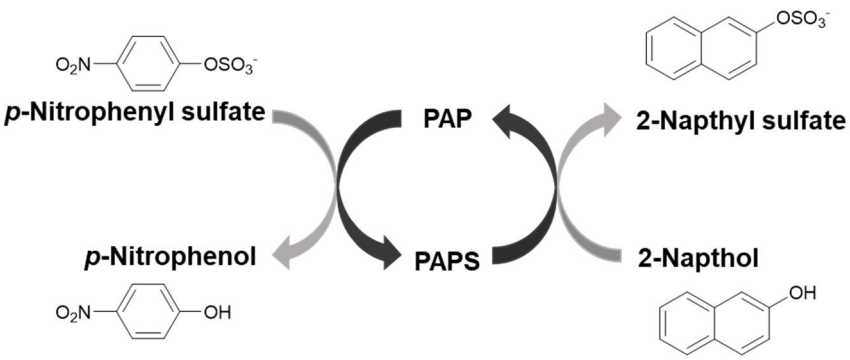

Figure 8 – Schematic representation of 2-naphthol sulfonation assay.

Firstly, it was important to know which are the specific substrates for SULT1B1 enzyme and also to determine kinetic constants. The enzyme was incubated with several possible substrates with different

concentrations and the absorbance at 405 nm was measured over time. The concentration of *p*-nitrophenol was then calculated and it was possible to determine the rate of reaction (amount of product made per unit time). A plot of reaction rate *vs.* substrate concentration allows determining the kinetic parameters of the enzyme (**Figure 9**), in order to know the substrates that present the highest affinity for the enzyme[39,53].

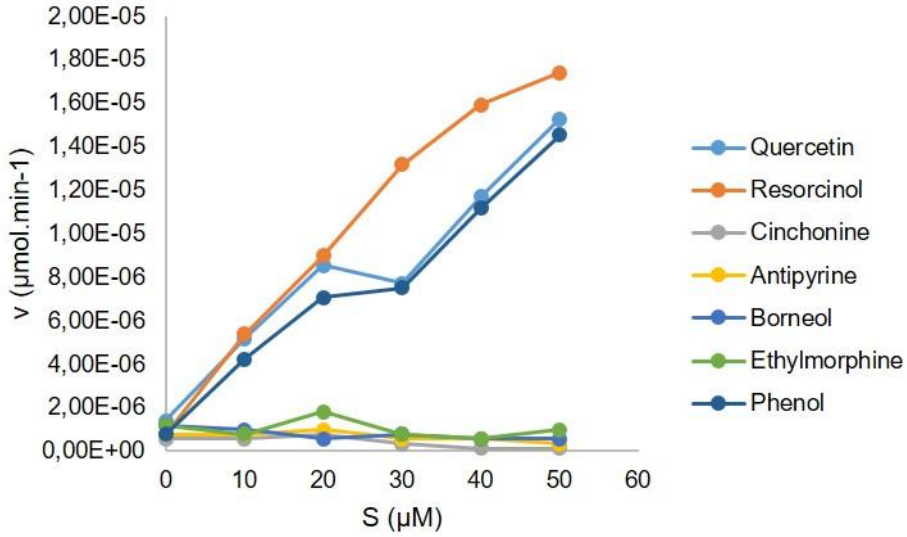

Figure 9 – Plot of initial velocity, v, against substrate concentration, S, for a SULT1B1 reaction in the presence of different substrates. SULT1B1 enzyme was used at 11.2 mg.mL^{-1} and the substrates with a final concentration ranging from 0 to 50 µM. The reactions occurred at 37ºC with slow agitation for 4 hours.

By analysing **Figure 9**, it is possible to observe that cinchonine, antipyrine, borneol and ethylmporhine are not substrates of the SULT1B1 enzyme, since the concentration of *p*-nitrophenol obtained was almost insignificant.

Antipyrine was used in this experiment as a negative control, since it does not have any hydroxyl-groups, as shown in **Figure 10A**. SULT1B1 is an enzyme that catalyses the sulfonylation of various hydroxyl-containing compounds and no sulfonylation was expected to occur in the presence of this substrate.

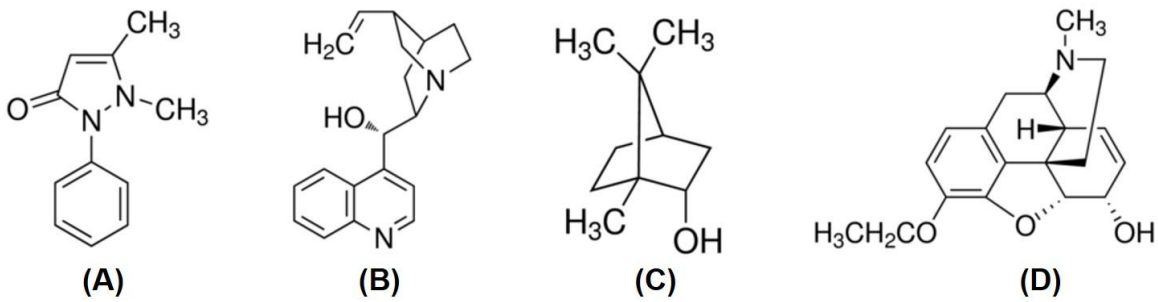

Figure 10 – Chemical structures of antipyrine (A), cinchonine (B), borneol (C) and ethylmorphine (D). These chemicals were used in this work.

In contrast, although the substrates cinchonine (**Figure 10B**), borneol (**Figure 10C**) and ethylmorphine (**Figure 10D**) have hydroxyl groups, they did not suffer sulfonation. Hydroxyl groups in these substrates are alkylic, in opposition to the aromatic OH groups in phenol, resorcinol and quercetin

(**Figure 11**), for which catalysis occurred. This implies that the produced SULT1B1 is a phenol-acting sulfotransferase, in agreement with published data[54].

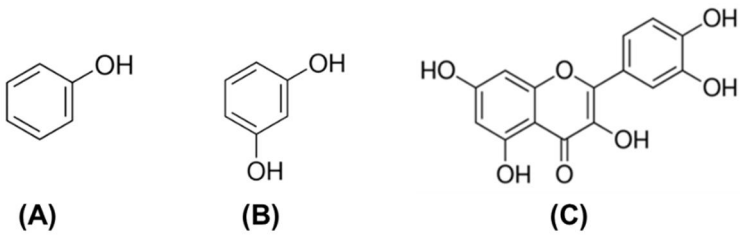

Figure 11 - Chemical structures of phenol (A), resorcinol (B) and quercetin (C). These chemicals were used in this work.

Resorcinol, quercetin and phenol present a hyperbolic rate profile which indicates that the kinetics follows a Michaelis-Menten equation (**Equation 3**)[55]. The values analysed at 30 µM of substrate concentration for quercetin and phenol were outliers, possibly to experimental errors.

$$v = \frac{V_{max} \cdot [S]}{[S] + K_M} \quad \textbf{(Equation 3)}$$

In order to determine the kinetic parameters for SULT1B1 in the presence of these substrates, a non-linear fit was performed using Prism 6.0 (GraphPad, USA) in order to determine K_M and V_{max}, as represented in **Figure 12;** the values obtained are present in ***Table 6***. The turnover number, k_{cat}, and the catalytic efficiency, k_{cat}/K_M, were computed from the maximal velocity using $V_{max} = [E_{free}]k_{cat} \approx [E_{total}]k_{cat}$[56].

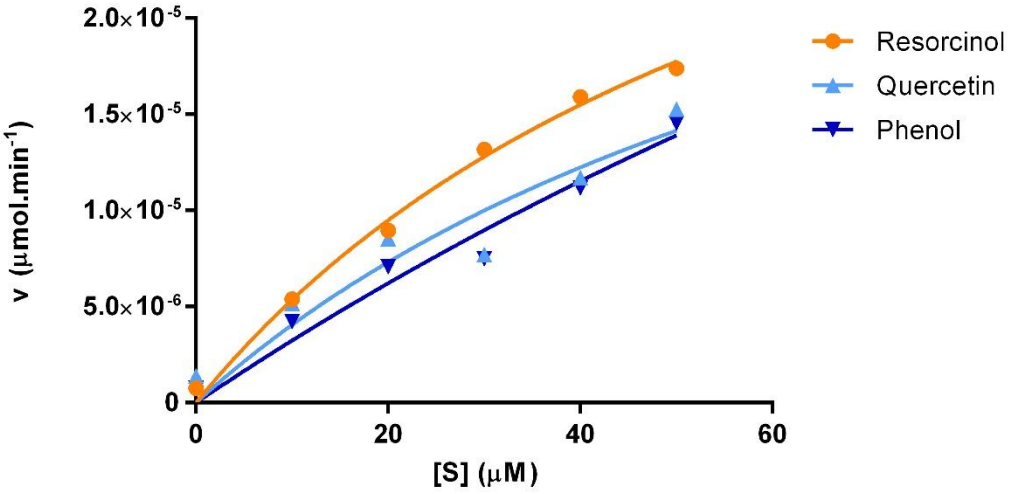

Figure 12 – Fitted curves for resorcinol, quercetin and phenol in order to determine SULT1B1 kinetic parameters. SULT1B1 enzyme was used at 11.2 mg.mL^{-1} and the substrates with a final concentration ranging from 0 to 50 µM. The reactions occurred at 37ºC with slow agitation for 4 hours.

Table 6 – Kinetic parameters for SULT1B1 in the presence of different substrates.

Michaelis-Menten	Substrates		
	Resorcinol	**Quercetin**	**Phenol**
	Best-fit values		
V_{max} **(μmol.min^{-1})**	4.227x10^{-5} (± 17%)	3.764 x10^{-5} (± 74%)	7.945 x10^{-5} (± 130%)
K_M **(μM)**	69.09 (± 26%)	82.97 (± 108%)	235.6 (± 152%)
k_{cat} **(s^{-1})**	3.951	3.519	7.247
k_{cat}/K_M **(μM^{-1} s^{-1})**	5.719x10^{-2}	4.241x10^{-2}	3.152x10^{-2}

From **Figure 12**, it was possible to determine the maximum velocity (V_{max}) values in the experiment which were 4.227 x10^{-5} μmol.min^{-1}, 3.764 x10^{-5} μmol.min^{-1} and 7.945 x10^{-5} μmol.min^{-1} for resorcinol, quercetin and phenol, respectively (**Table 6**), corresponding to turnover numbers of 3.951 s^{-1}, 3.519 s^{-1}, 7.247 s^{-1} and to a catalytic efficiency of 5.719 x10^{-6} μM^{-1} s^{-1}, 4.241x10^{-6} μM^{-1} s^{-1} and 3.152 x10^{-6} μM^{-1} s^{-1}. Moreover, it was also possible to estimate apparent K_M values of 69.09 μM, 82.97 μM and 235.6 μM for resorcinol, quercetin and phenol, respectively (**Table 6**). The K_M value is a measure for the enzyme binding affinity to the substrate. All substrate concentrations should be further increased to allow complete enzyme saturation, achieving a plateau state, in order to have more reliable results[53,55,57].

Resorcinol and quercetin present a similar affinity to SULT1B1 enzyme since both present similar K_M values (**Table 6**), although resorcinol has a lower K_M (69.09 μM) which means that presents a higher affinity to SULT1B1. Phenol was the substrate that presented the higher K_m (235.6 μM) value meaning that has lowest affinity to the enzyme in comparison with the other substrates analysed.

In the literature, the K_M obtained for SULT1B1 in the presence of quercetin and phenol is 35.1 μM and 40 μM, respectively, while in our study values of 69.09 μM and 235.6 μM, respectively, were obtained[58,59]. The experimental result obtained for SULT1B1 in the presence of quercetin is of the same order of magnitude of the result reported in the literature, while the value for phenol is quite different. However, the standard errors of the results obtained are too large, meaning that further studies are needed in order to optimize the experiment and obtain more reliable results for SULT1B1 activity in the presence of phenol.

The turnover number measures the maximum molecules of substrate that can be converted into product per catalytic site[60]. In the literature, the k_{cat} obtained for phenol was 0.113-0.122 s^{-1} while in our experiment it was obtained 7.247 s^{-1}, meaning that our enzyme appears capable of converting more molecules of phenol per cycle[54].

Further studies were performed in order to obtain more reliable results in a microplate with smaller volumes. Several compounds were tested as possible SULT1B1 enzyme substrates. The results obtained for phenol, α-hydroxy-tamoxifen, *rac*-8,14–dihydroxy-efavirenz, 1,5–dihydroxyanthraquinone and 1,8-dihydroxyanthraquinone did not shown any significant results for *p*-nitrophenol concentration meaning that no reaction occurred between the enzyme and those substrates (data not shown). Once more, resorcinol and quercetin were the substrates that in the presence of SULT1B1 showed production of *p*-nitrophenol (**Figure 13**) meaning that sulfonylation was occurring. It is important to notice that α-

hydroxy-tamoxifen is an allylic alcohol which is in agreement with the evidence of SULT1B1 being a specific phenol sulfotransferase.

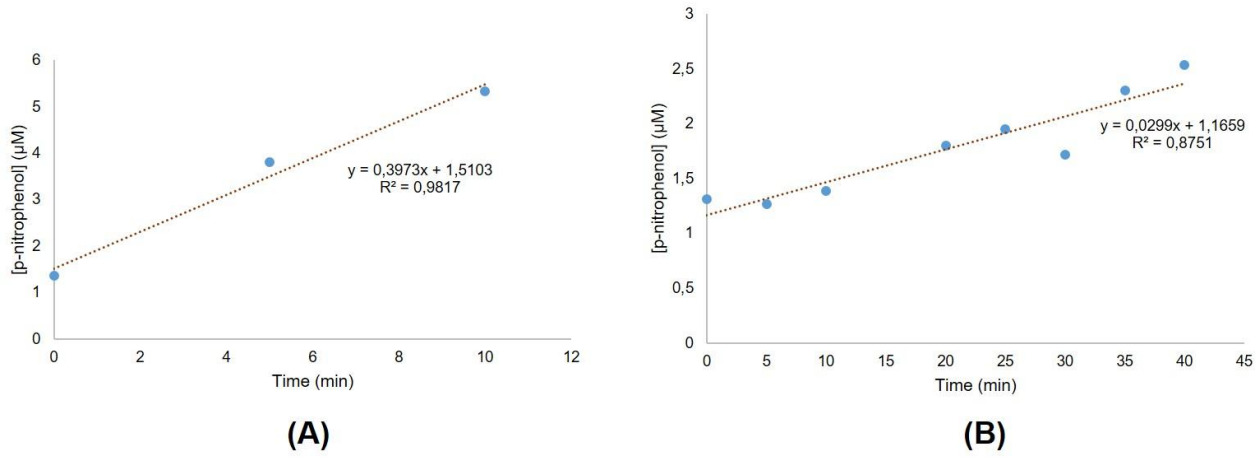

Figure 13 – Plot of *p*-nitrophenol concentration, [*p*-nitrophenol], against time for a SULT1B1 reaction in the presence of quercetin (A) and resorcinol (B). SULT1B1 enzyme was used at 0.224 mg.mL^{-1} and the substrates with a final concentration ranging from 0 to 50 µM. The reactions occurred at 37ºC with slow agitation.

By analysing **Figure 13**, it is possible to observe that the levels of *p*-nitrophenol concentration increase over time for both quercetin (**Figure 13A**) and resorcinol (**Figure 13B**), meaning that *p*-nitrophenyl sulfate is being consumed. Higher concentrations of *p*-nitrophenol were obtained for incubation of the enzyme with quercetin obtaining a maximum value of 6.16 µM (**Figure 13A**) while for resorcinol a maximum value of 2.53 µM was obtained (**Figure 13B**). Taking in consideration these results, sulfonylation is occurring and so the initial velocity of the enzymatic reaction and enzymatic activity can be estimated for both substrates. The initial velocity obtained for the enzyme in the presence of quercetin was 3.97x10^{-5} µmol.min^{-1} and the enzymatic activity was 1.77 x10^{-4} µmol.min^{-1}.mg^{-1}. The values obtained for the same parameters for the same enzyme but in the presence of resorcinol were lower, being obtained an initial velocity of 2.99 x10^{-10} µmol.min^{-1} and an enzymatic activity of 1.33 x10^{-9} µmol.min^{-1}.mg^{-1}.

The substrates that were shown to be metabolised by SULT1B1 enzyme are structurally similar, since they present phenolic OH groups, meaning that this enzyme is an aryl sulfotransferase. Aryl sulfotransferases are characterized by being able to transfer a sulfo group from a donor molecule, usually PAPS, to a phenolic acceptor substrate[51,52,61].

3.4. SULT1B1-mediated sulfonylation studies by LC/MS

In order to validate the usability of recombinant enzymes, expressed in *E. coli*, to predict the metabolic fate of tested compounds, 5 h incubations of each enzyme with relevant substrates and the required cofactors were performed at 37 °C. Reactions were stopped by removing the enzyme by centrifugation in Amicon Vivaspin filters with a membrane cutoff of 10 kDa, and the permeate was analysed by liquid chromatography-tandem high resolution mass spectrometry (LC/MS). Tested substrates for SULT1B1

were resorcinol, quercetin, 1,5–dihydroxyanthraquinone, 1,8–dihydroxy-anthraquinone, phenol, acetaminophen, α-hydroxytamoxifen, *N*-desmethyltamoxifen, *rac*-8,14-dihydroxy-efavirenz, cinchonine and antipyrine.

No drug-derived sulfate-containing ions were found on the full scan MS spectra of samples containing quercetin, 1,5–dihydroxy-anthraquinone (**Figure 14A**), 1,8–dihydroxy-anthraquinone (**Figure 14B)**, phenol (**Figure 11A**), acetaminophen (**Figure 14C**), *rac*-8,14-dihydroxy-efavirenz (**Figure 14D**), α-hydroxytamoxifen (**Figure 14E)**, *N*-desmethyltamoxifen (**Figure 14F**), antipyrine (**Figure 10A**) and cinchonine (**Figure 10B**).

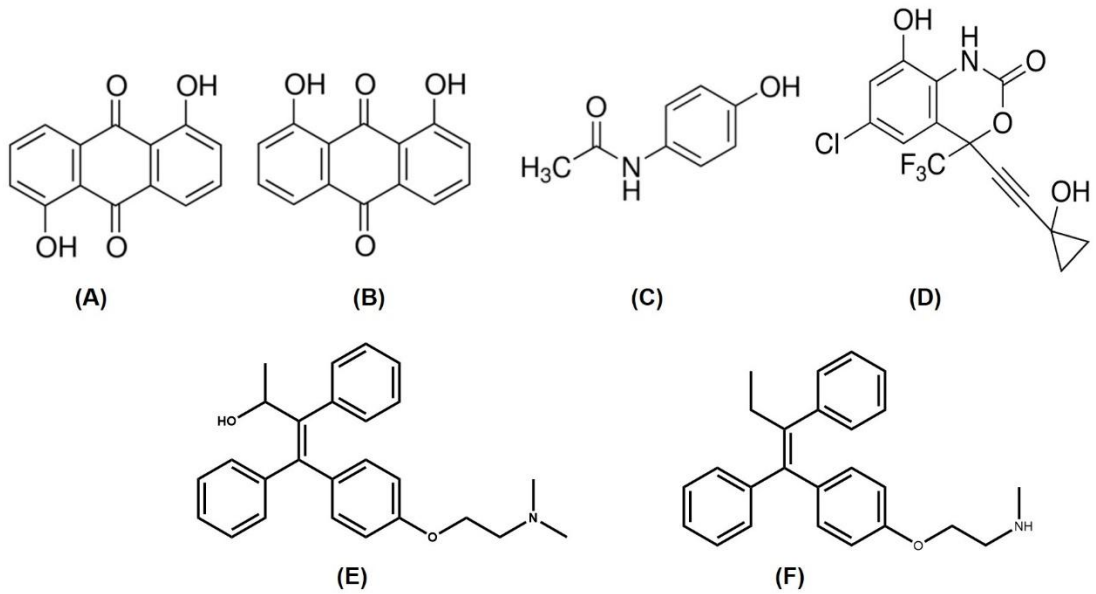

Figure 14 - **Chemical structures of 1,5-dihydroxy-anthraquinone (A), 1,8-dihydroxy-anthraquinone (B), acetaminophen (C), *rac*-8,14-dihydroxy-efavirenz (D), α-hydroxytamoxifen (E) and *N*-desmethyltamoxifen (F).** These chemicals were used in this work.

For resorcinol, a peak in the ion chromatogram for *m/z* = 188.9852 was observed at 10.5 min, corresponding to a significantly more polar compound than resorcinol itself, that was observed in the ion chromatogram for *m/z* = 109.0284 at 14.7 min (**Figure 15**).

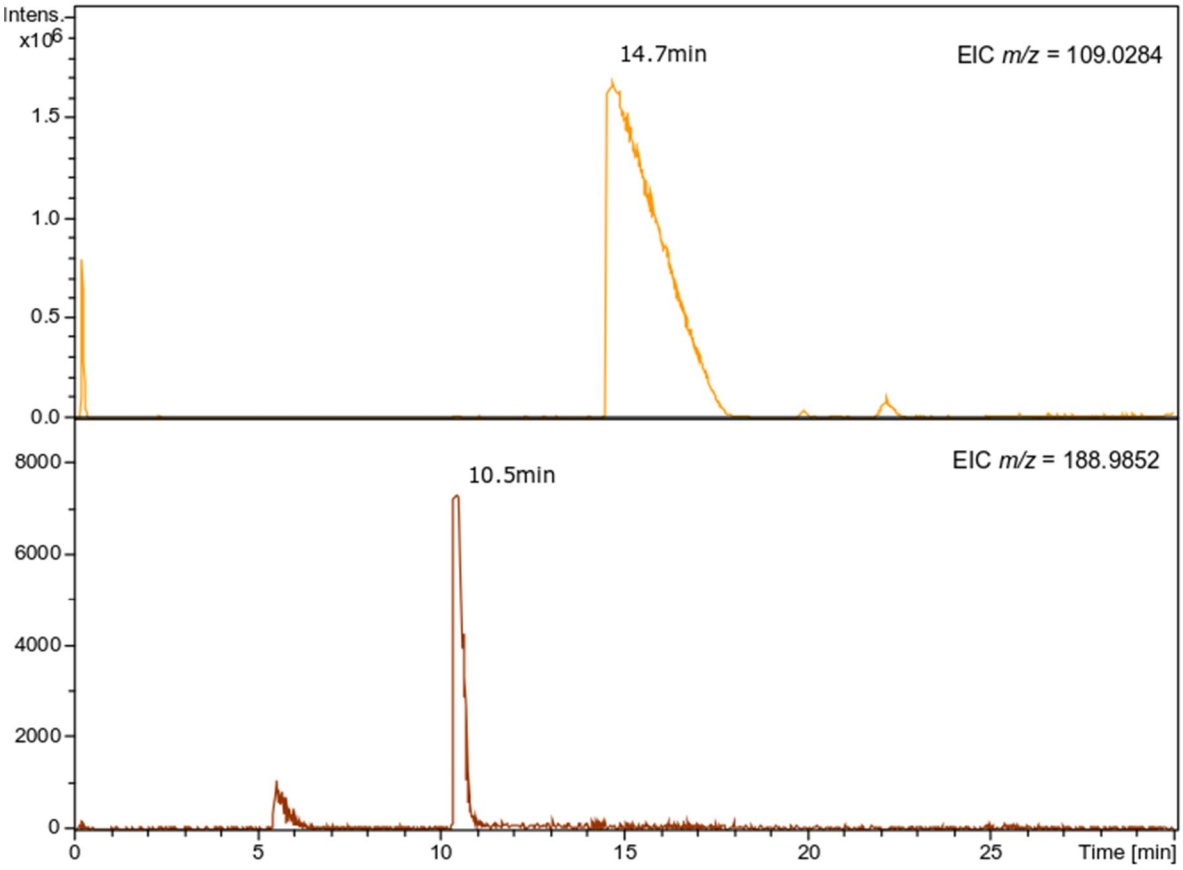

Figure 15 – Extracted ion chromatograms (ESI –) from SULT1B1 incubated with resorcinol. Resorcinol (*m/z*=109.0284) and sulfoxy-resorcinol (*m/z*=188.9852) with retention times of 14.7 and 10.5 minutes, respectively, were identified.

Chromatographic separation was carried out on a HypersilGold C18 stationary reverse phase column, in which the column matrix holds more hydrophobic compounds longer. Taking this in consideration, more polar compounds, such as sulfoxy-resorcinol, compared to resorcinol, elute first, presenting a lower retention time, as shown in **Figure 15**. The top chromatogram in **Figure 15** was obtained by selecting only ions with an *m/z* value of 109.0284, corresponding to the theoretical *m/z* for deprotonated resorcinol, $[M-H]^-$.The lower chromatogram was obtained by selecting only ions with an *m/z* value of 188.9852, corresponding to the calculated mass for deprotonated sulfoxy-resorcinol. The difference between *m/z* values for both peaks shown in **Figure 15** is 79.9568, which is the *m/z* value for the SO_3 group, in accordance with the introduction of this moiety in the parent resorcinol molecule.

SULT1B1 is a metabolising enzyme that facilitates drugs elimination by increasing their hydrophilicity. The sulfoxy-resorcinol is a metabolite of this enzyme and so the sulfoxy-group increases molecule polarity, which facilitates its subsequent elimination *in vivo*.

In order to confirm the presence of sulfoxy-resorcinol in the sample, the full MS^- spectra (**Figure 16**) was analysed and the isotopic pattern for sulfoxy-resorcinol was obtained, as shown in **Figure 16**, matching the predicted isotopic profile for a compound with that chemical formula (**Table 7**).

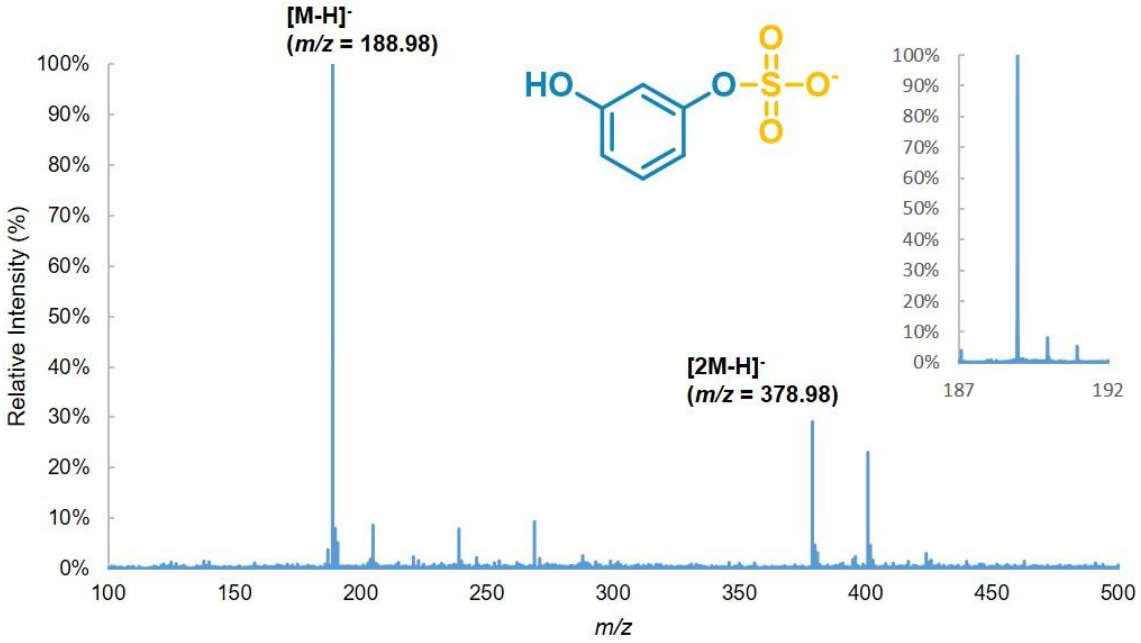

Figure 16 – Full ESI(-) MS scan of resorcinol reaction with SULT1B1 after 5 h of incubation. Inset is the observed isotopic pattern of sulfoxy-resorcinol in the MS spectrum.

The spectrum obtained in **Figure 16** is dominated by ions at *m/z* 188.98 and 378.98, corresponding to deprotonated molecular ions $[M-H]^-$ and $[2M-H]^-$, respectively. The ions were identified based on their *m/z* values and on the observed isotopic profiles, by comparing them with theoretical values (**Table 7**).

Table 7 – Observed isotopic patterns for Res-OSO_3^- ions $[M-H]^-$ and $[2M-H]^-$ and theoretical values, with observed mass deviations (σ, in ppm).

	Theoretical values		Observed values		*σ(m/z)*
	m/z	I %	*m/z*	I %	(ppm)
Resorcinol sulfate [M – H]⁻	188.9852	100	188.9864	100	4.65
	189.9882	7.50	189.9890	7.90	4.19
	190.9830	5.70	190.9839	5.20	5.03
Resorcinol sulfate [2M – H]⁻	378.9788	100	378.9811	29.2	6.03
	379.9818	15.1	379.9842	4.30	6.32
	380.9769	12.1	380.9810	3.20	10.7

Analysing **Figure 16** and **Table 7**, the *m/z* values and the relative intensity for the obtained peaks were similar to the theoretical values. The $[M-H]^-$ ions obtained presented *m/z* deviation values of 5 ppm or lower, which is considered acceptable. Therefore, resorcinol was shown to be metabolised by SULT1B1 enzyme, producing a monosulfoxy-resorcinol.

3.5. CYP2C8-mediated oxidation studies by LC/MS

Several compounds were tested as possible CYP2C8 substrates, namely tamoxifen, nevirapine, arachidonic acid and novocaine. The reaction mixtures were incubated for 5 hours at 37ºC, and were analysed by LC/MS upon protein removal.

Reaction of CYP2C8 with novocaine and arachidonic acid did not lead to the appearance of new peaks in the corresponding chromatograms; furthermore, manual searching for ions at *m/z* values corresponding to possible oxidation products of these compounds was also unsuccessful. In the case of nevirapine and tamoxifen, a number of CYP2C8-mediated oxidation products were found, and are discussed below.

3.5.1. Nevirapine

Electrospray MS analysis in the positive mode of samples obtained by incubating CYP2C8 and nevirapine (in the presence of NADPH) was performed in order to identify the produced metabolites.

Chromatographic separation was carried out on a C18 stationary phase and so, differences in polarity of the obtained metabolites influence the relative retention time as shown in **Figure 17**. Extracted ion chromatograms show the presence of 3 relevant peaks, at 14.6 min, with *m/z* = 267.1241, corresponding to nevirapine; at 13.4 min, with *m/z* = 283.1190, corresponding to hydroxylated nevirapine, more polar that nevirapine and thus with lower column retention; and at 14.6 min (with a small front peak at 14.0 min), tentatively assigned to descyclopropyl-nevirapine.

These ions were further studied using tandem mass spectrometry, and the resulting MS/MS spectra are shown in **Figure 18**.

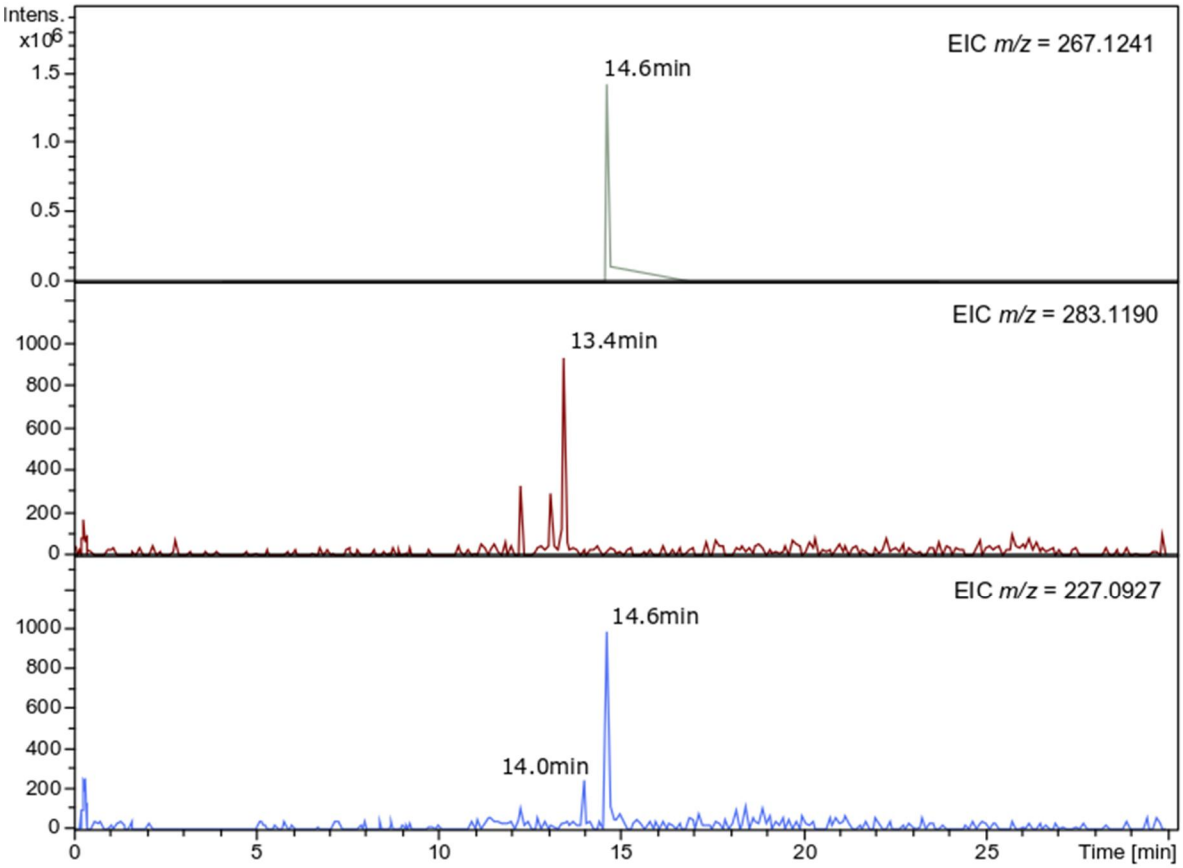

Figure 17 - Extracted ion chromatograms of nevirapine metabolites recovered from samples after incubation with CYP2C8 enzyme. Nevirapine (EIC *m/z*=267.1241), hydroxy-nevirapine (EIC *m/z*=283.1190) and nevirapine without cyclopropyl group (EIC *m/z*=227.0927), with retention times of 14.6, 13.4 and 14.0/14.6 minutes, respectively. These compounds were separated by C18 stationary phase and detected as $[M+H]^+$ ions with ESI(+)-MS.

The MS/MS spectra of protonated nevirapine (**Figure 18A**), at *m/z* = 267.1241, is dominated by the peaks corresponding to the precursor ion, $[M+H]^+$, and to the loss of descyclopropylnevirapine, [M+H-cyclopropyl]$^+$, at *m/z* = 227.0927, in agreement with the existing data on the fragmentation of NVP[62].

Additional smaller peaks were also observed (**Figure 18A**). The proposed fragmentation mechanisms are displayed in **Figure 19**.

The difference between *m/z*= 267.1241 for nevirapine and m/z = 283.1190 shown in **Figure 17** corresponds to an addition of 15.9949, which is the *m/z* change corresponding to the introduction of an oxygen atom, presumably by becoming part of a hydroxyl group. This result is in accordance with a possible CYP2C8-mediated oxidation of nevirapine.

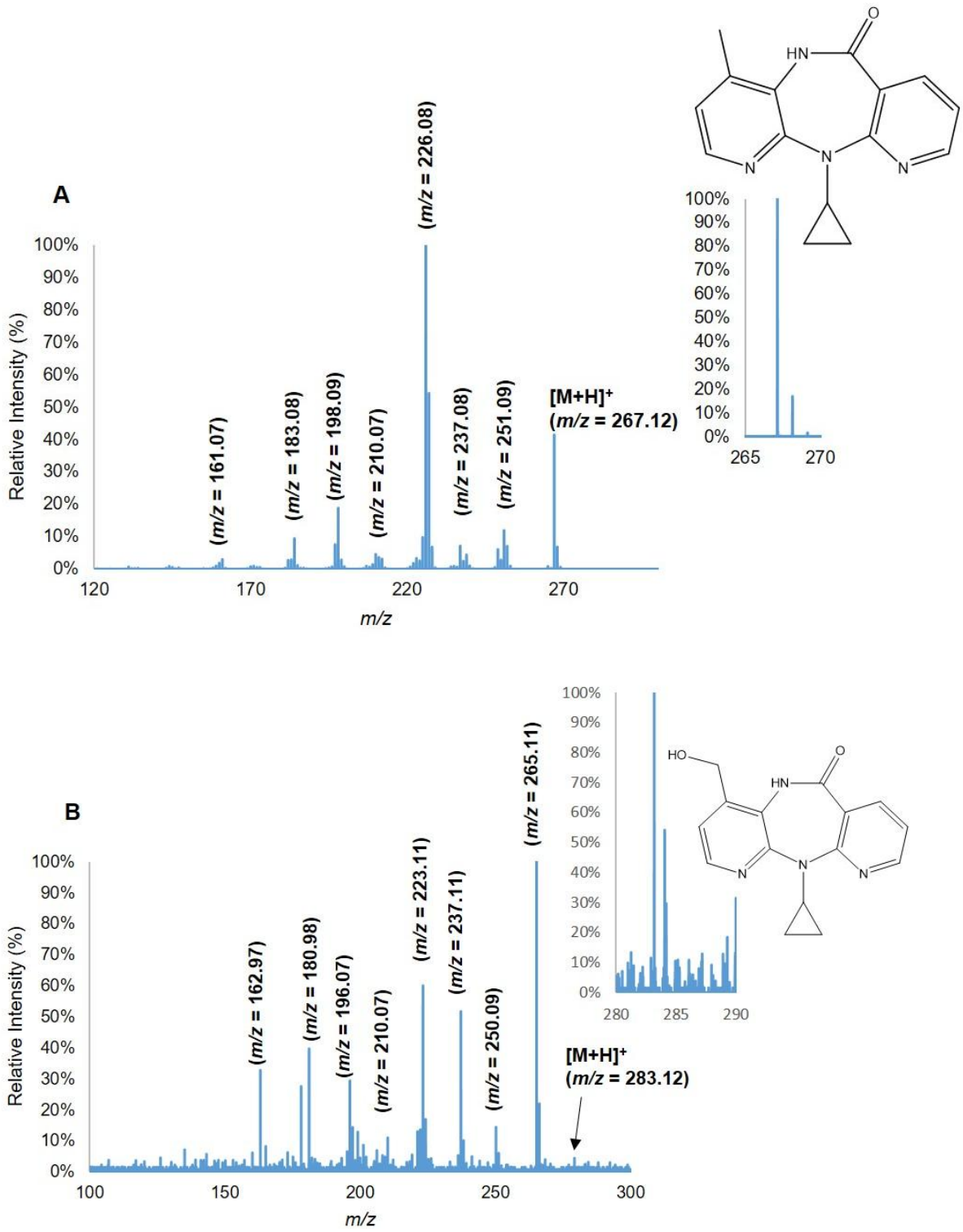

Figure 18 – ESI-MS/MS spectra from protonated nevirapine metabolites [M + H]⁺ ions. (A) Protonated nevirapine (*m/z* 267.12 ion); **(B)** Protonated hydroxylated nevirapine (*m/z* 283.12 ion). Insets are the isotopic patterns for nevirapine (A) and 12-hydroxy-nevirapine (B) in the MS spectra.

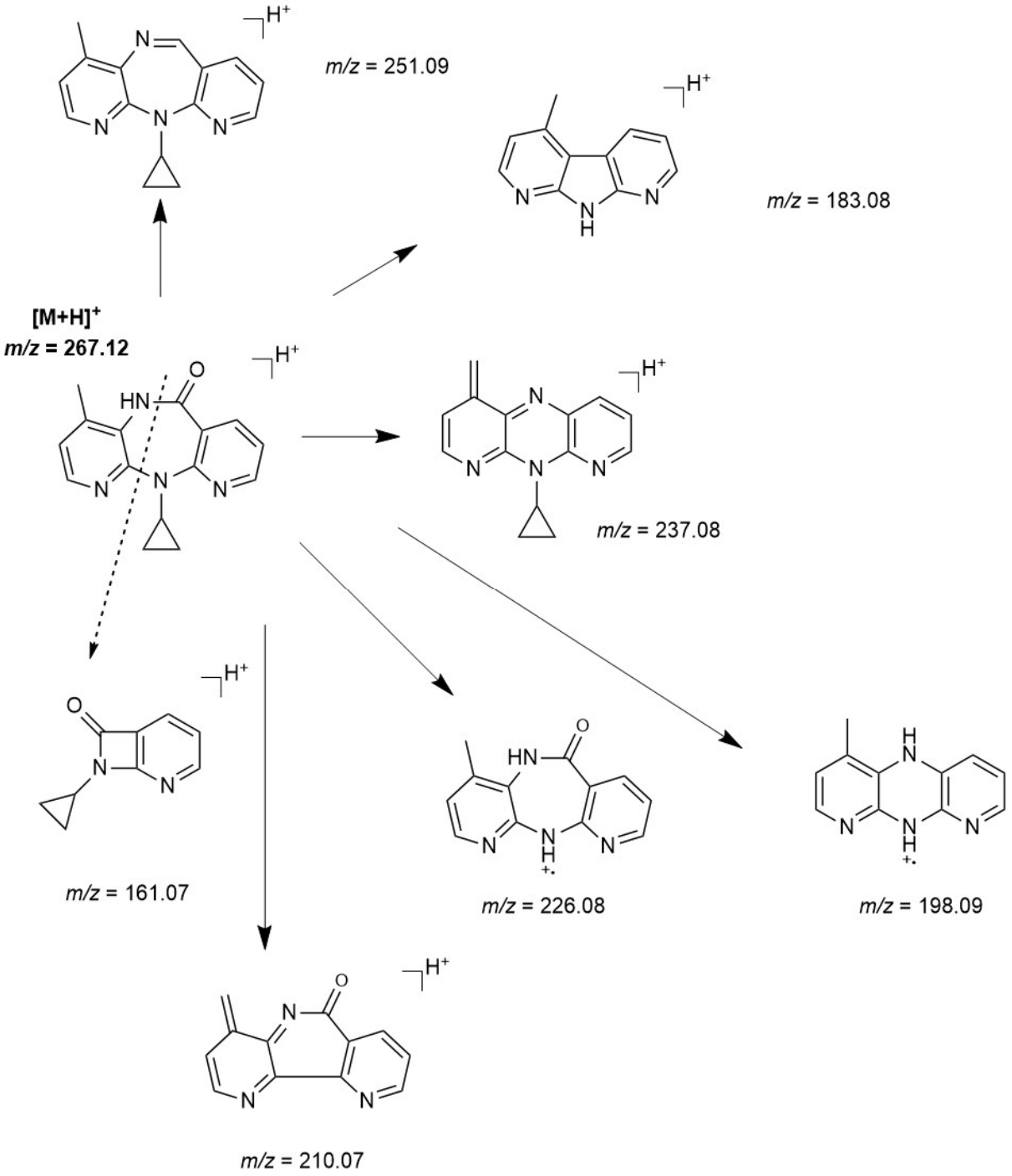

Figure 19 – Proposed mass spectrometry fragmentation pathways for the protonated molecule of nevirapine ([M+H]$^+$). This work.

The MS/MS spectrum obtained for the ion at *m/z* 283.1190 is shown in **Figure 18B**. Analysis of the observed ions and comparison with the MS/MS spectra for the various known hydroxylated metabolites of NVP indicate that this ion at *m/z* 283.1190 corresponds to 12-OH-nevirapine[62]. In particular, the base peak at *m/z* 265.11 (**Figure 18B**) was fully consistent with loss of water to yield a benzylic-type cation that rearranged to a tropylium-like structure (**Figure 20**).

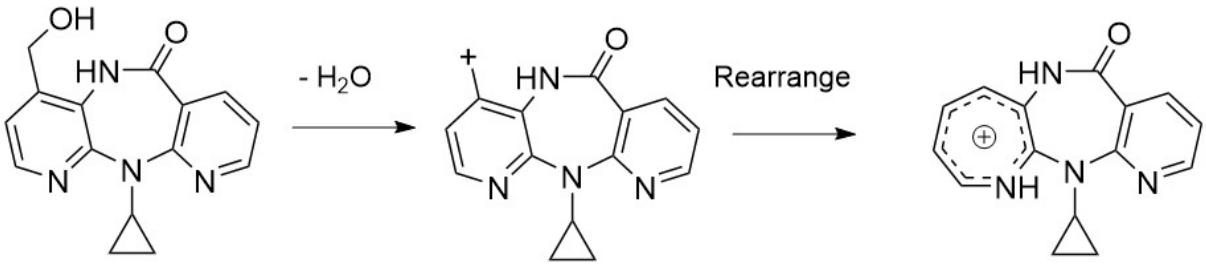

Figure 20 – The putative fragmentation pathway for 12-OHNVP.

The identity of the ions attributed to each peak was also verified by comparing the observed isotopic profiles with the predicted ones, as summarized in **Table 8**. The *m/z* values and the relative intensity for the obtained peaks were similar to the theoretical values. The ions obtained presented *m/z* deviation values much lower than 5 ppm in comparison with the theoretical values, meaning that they are reliable and significant.

Table 8 – Observed isotopic patterns for $[M+H]^+$ ions of nevirapine and hydroxy-nevirapine and theoretical values, with observed mass deviations (σ, in ppm).

	Retention time (min)	Theoretical values		Observed values		Sigma
		m/z	I %	m/z	I %	(ppm)
Nevirapine $[NVP + H]^+$	14.6	267.1240	100	267.1240	100	0.140
		268.1269	17.9	268.1268	16.8	0.400
		269.1296	1.70	269.1298	1.50	0.870
		270.1321	0.100	-	-	-
		271.1346	0.000	-	-	-
Hydroxy-nevirapine $[NVPOH + H]^+$	13.4	283.1190	100	283.1192	100	0.880
		284.1218	17.9	284.1222	12.8	1.32
		285.1243	1.93	-	-	-
		286.1268	0.150	-	-	-
		287.1293	0.0100	-	-	-

When analysing the extracted ion chromatogram at *m/z* 227.0927 (**Figure 17**), the peak at 14.6 min corresponds to a hydrogen atom loss from an ion with *m/z* at 226.0849, compatible with [M+H-cyclopropyl]$^+$, and that, upon fragmentations, yields an MS/MS spectra coherent with *N*-descyclopropylnevirapine; however, as it is present at the same retention time of the protonated NVP, it is most likely a byproduct of ionization at the source.

On the other hand, the peak with retention time 14.0 min corresponds to an ion at *m/z* 227.0927 (relative abundance 14.4%), with only a 0.2 ppm mass deviation from the expected *m/z* value for protonated *N*-descyclopropylnevirapine. However, its low intensity did not allow proper fragmentation. This ion is accompanied by another one at *m/z* 228.0967 (relative abundance 2.9%), corresponding to the ^{13}C isotopomer, with a mass deviation of 5.3 ppm. The observed relative abundances for these ions, 14.4:2.9, or 100:20.1, are not very distant from the expected relative abundances, 100:14.6, further supporting the hypothesis that this ion corresponds to the *N*-dealklyation product of NVP catalysed by CYP2C8. The putative mechanism for such dealkylation is shown in **Figure 21**, and involves the oxidation of the tertiary carbon of the cyclopropyl group, yielding a hemiaminal that undergoes spontaneous collapse with elimination of cyclopropanone.

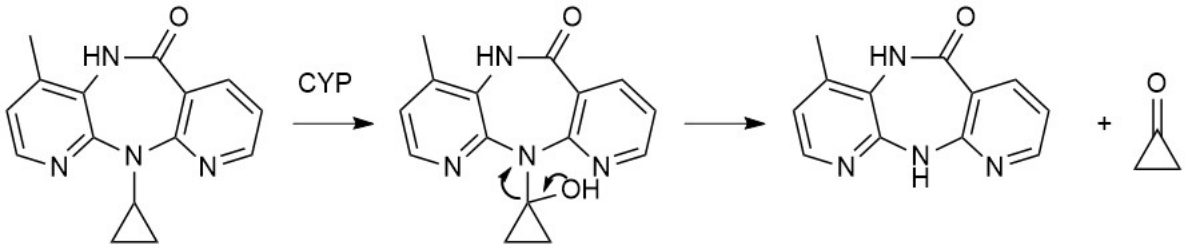

Figure 21 – Proposed elimination of the NVP cyclopropyl group stemming from CYP2C8-mediated hydroxylation at the tertiary carbon.

3.5.2. Tamoxifen

Reaction mixtures of the antiestrogen tamoxifen with CYP2C8 were also analysed by LC/MS. Extracted ion chromatograms for protonated tamoxifen (*m/z* = 372.2311, retention time 17.5 min), protonated *N*-desmethyl-tamoxifen (*m/z* = 358.2155, retention time 17.3 min), protonated *N,N*-didesmethyl-tamoxifen (*m/z* = 344.2000, retention time 17.1 min) and protonated hydroxylated tamoxifen (*m/z* 388.2259, retention time 17.6 min) are shown in **Figure 22**.

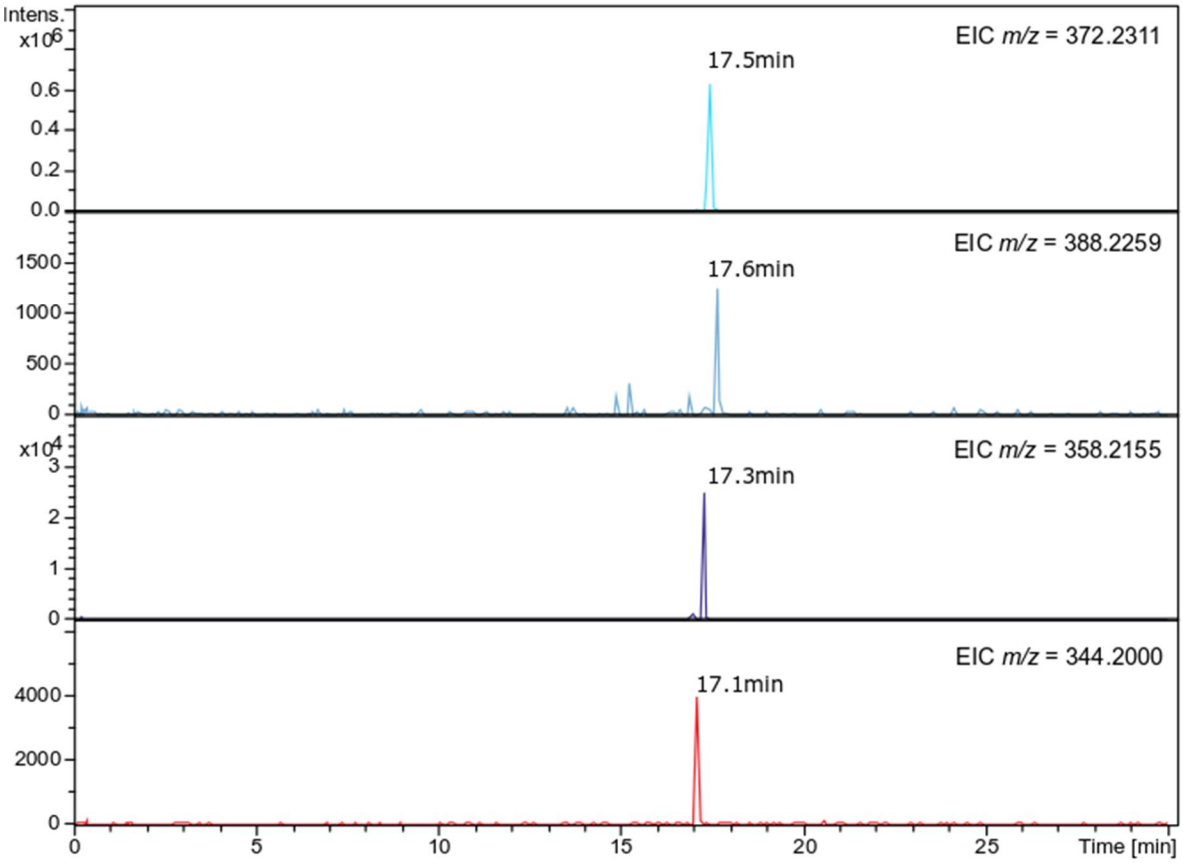

***Figure 22* - Extracted ion chromatograms of tamoxifen metabolites recovered from samples after incubation with CYP2C8 enzyme.** Tamoxifen (EIC *m/z*=372.2311), hydroxyl-tamoxifen (EIC *m/z*=388.2259), *N*-desMetTam (EIC *m/z*=358.2155) and *N,N*-didesMetTam (EIC *m/z*=344.2000) with retention times of 17.5, 17.6, 17.3 and 17.1 minutes, respectively. These compounds were separated by C18 stationary phase and detected as $[M+H]^+$ ions with ESI(+)-MS.

Chromatographic separation was carried out on a C18 stationary phase and so, differences in polarity of obtained metabolites influence the relative retention time as shown in **Figure 22**. The first peak obtained with an EIC *m/z* = 372.2311 (**Figure 22**) is in agreement with the expected *m/z* value for tamoxifen and the retention time for this compound is around 17.5 minutes. The other peak (EIC *m/z* =388.2259) obtained presents a retention time of 17.6 minutes (**Figure 22**), meaning that the compound has similar polarity to that of tamoxifen. The other two peaks obtained have EICs *m/z* = 358.215 and 344.2000 (**Figure 22**) and present retention times of 17.3 and 17.1 minutes, respectively. Both previous peaks present lower retention times than tamoxifen, which makes sense since both compounds suffer demethylation leading into slightly more polar compounds. The retention time for the EIC *m/z* = 344.200 is lower than for 358.215 (**Figure 22**) because the first one goes through a demethylation process while the other suffers a disdemethylation. Nonetheless, the retention times are very close because the properties of the parent compound and the metabolites are primarily governed by the highly hydrophobic triarylethylene structure.

Moreover, the peak obtained for tamoxifen with EIC *m/z* = 372.2311 presents higher intensity values than the other peaks obtained, meaning that tamoxifen is in higher concentration than the metabolites obtained (**Figure 22**). The other metabolite obtained in high intensities is the *N*-desmethyl-tamoxifen (EIC *m/z*= 358.2155), which is significant since this tamoxifen metabolite is the most abundant *in vivo*[63].

All the peaks shown in **Figure 22** were selected for fragmentation and their tandem mass spectra are shown in **Figure 23** and **Figure 24**.

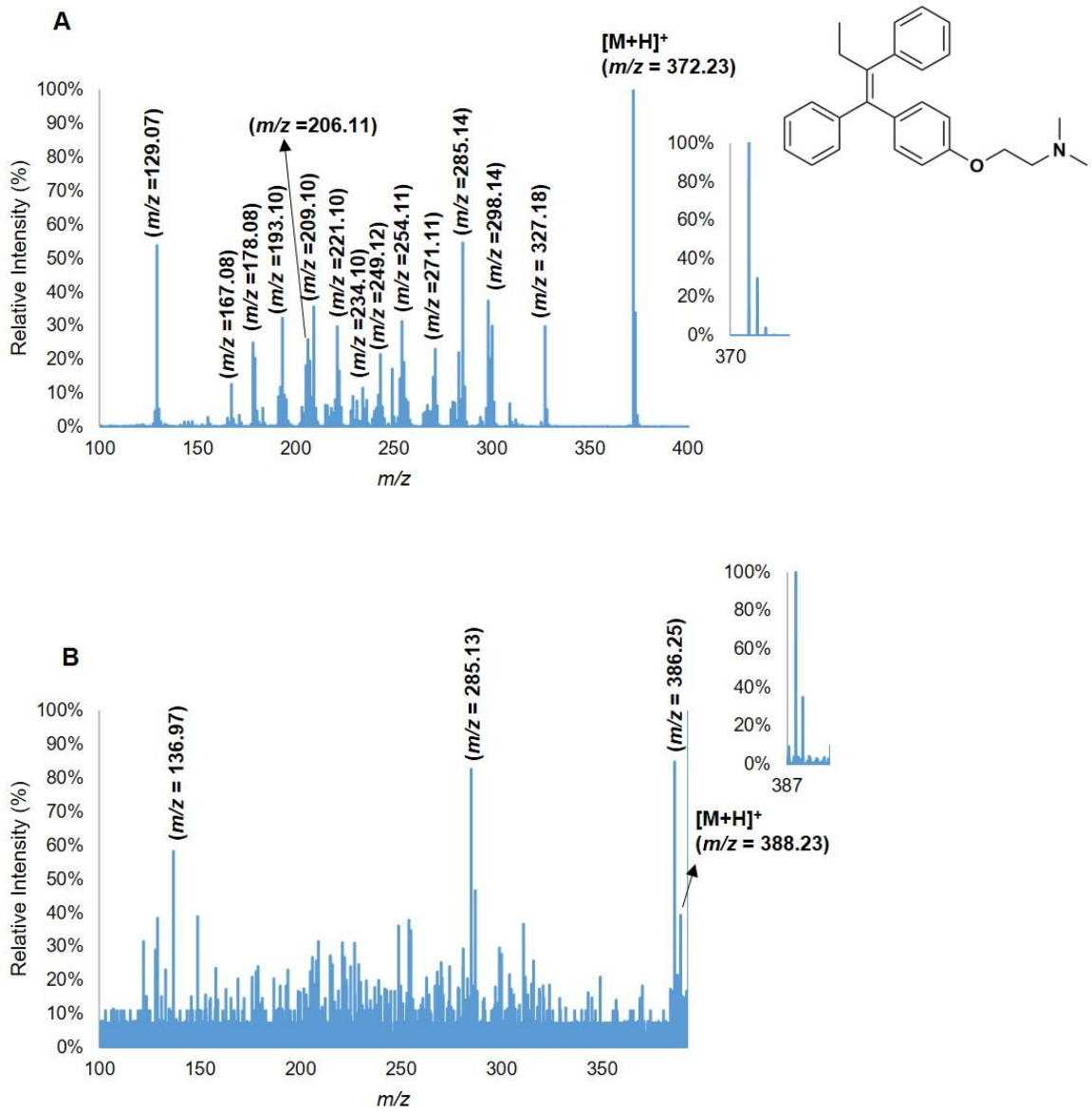

Figure 23 - ESI-MS/MS spectra of protonated tamoxifen (A) and hydroxy-tamoxifen (B) ions. (A) Protonated tamoxifen (*m/z* 372.23 ion); **(B)** Protonated hydroxylated tamoxifen (*m/z* 388.23 ion)**.** Insets are the isotopic patterns for tamoxifen (A) and hydroxy-tamoxifen (B) in MS spectra.

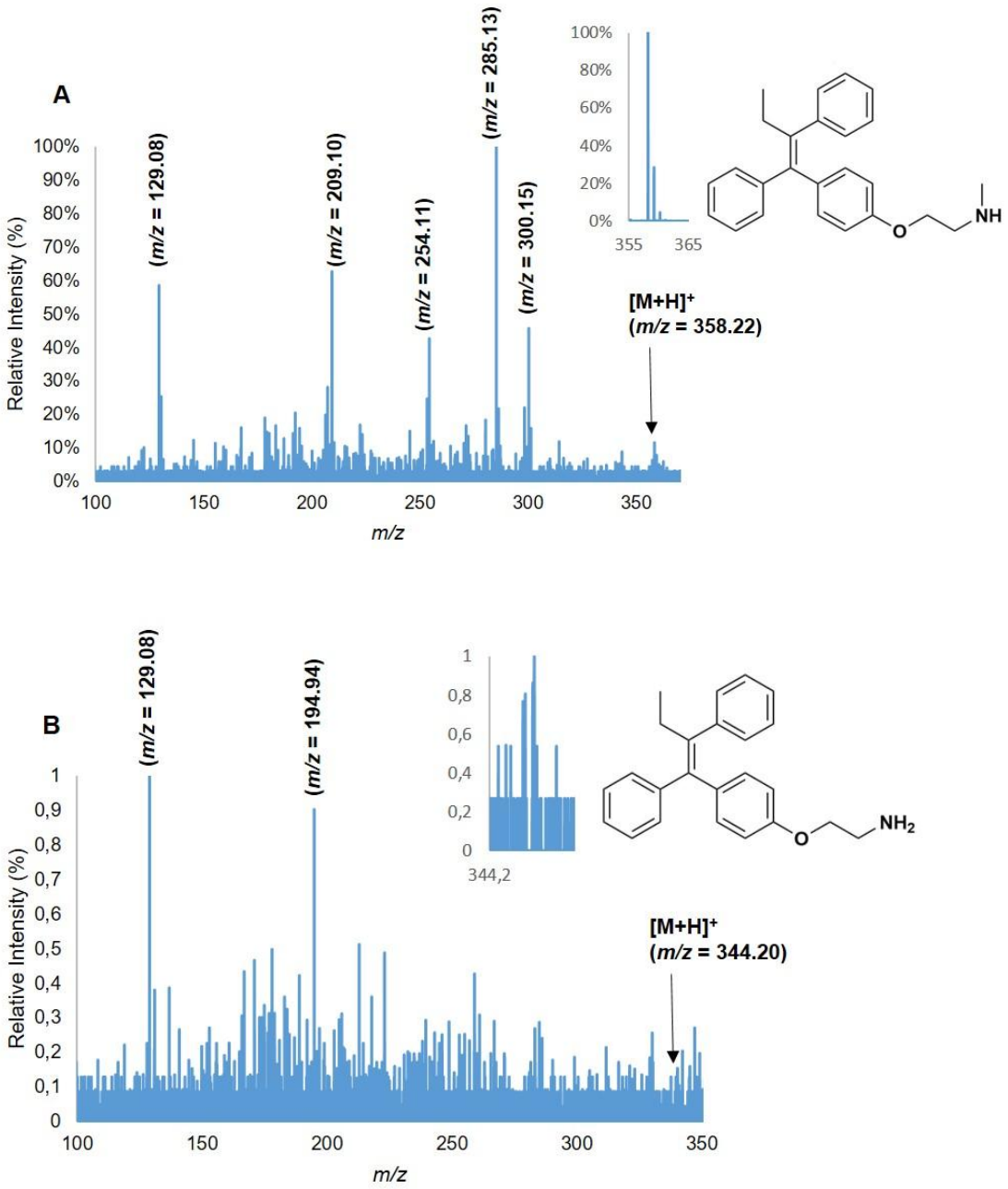

Figure 24 - ESI-MS/MS spectra of protonated *N*-desmethyl-tamoxifen (A) and *N.N'*-didesmethyl-tamoxifen (B) ions. (A) Protonated *N*-desMetTamoxifen (m/z 358.22 ion); **(B)** Protonated *N,N*-didesMetTamoxifen (*m/z* 344.20 ion)**.** Insets are the isotopic pattern for desMetTamoxifen (A) and didesMetTamoxifen (B) in MS spectra.

Several peaks were obtained in tamoxifen fragmentation (**Figure 23A**) and in order to understand the molecular mechanism of their fragmentation, the results obtained were interpreted taking in account the fragmentation pattern proposed by *S.F.Teunissen et. al* and *Gamboa da Costa et. al*[64,65]. The main fragmented ions resulting from the protonated molecules, their isotopic patterns and theoretical values with the associated errors are summarized in **Table 9**.

Table 9 - Main fragment ions with their isotopic patterns, theoretical values and associated errors for tamoxifen and the metabolites obtained.

	Retention time (min)	Peaks	Theoretical values		Observed values		Sigma
			m/z	I %	m/z	I %	(ppm)
Tamoxifen [M + H]$^+$	17.5	1	372.2322	100	372.2331	100	2.44
		2	373.2355	28.9	373.2365	29.8	2.67
		3	374.2387	4.20	374.2398	4.10	2.95
		4	375.2418	0.400	375.2400	0.400	3.24
Hydroxy-tamoxifen [OH-M + H]$^+$	17.6/15.3	1	388.2271	100	388.2280	100	2.30
		2	389.2304	28.9	389.2302	34.5	0.560
		3	390.2335	4.44	-	-	-
		4	391.2364	0.480	-	-	-
N-desMeTam-tamoxifen [M-NdesMe + H]$^+$	17.3	1	358.2165	100	358.2175	100	2.68
		2	359.2198	27.8	359.2207	28.4	2.37
		3	360.2230	3.90	360.2248	4.50	4.92
		4	361.2261	0.400	-	-	-
N-didesMeTam-tamoxifen [M-NdidesMe + H]$^+$	17.1	1	344.2009	100	344.2013	100	1.19
		2	345.2042	26.7	345.2055	29.2	3.79
		3	346.2074	3.60	346.2056	4.55	5.08
		4	347.2104	0.300	-	-	-

Analysing **Figure 23**, **Figure 24** and **Table 9** the *m/z* values and the relative intensities for the obtained peaks were similar to the theoretical values. The ions obtained presented *m/z* deviation values below or very close to 5 ppm, which is considered acceptable.

The main fragmentation patterns proposed for most abundant fragment ions of protonated metabolites under study are present in **Figure 25**.

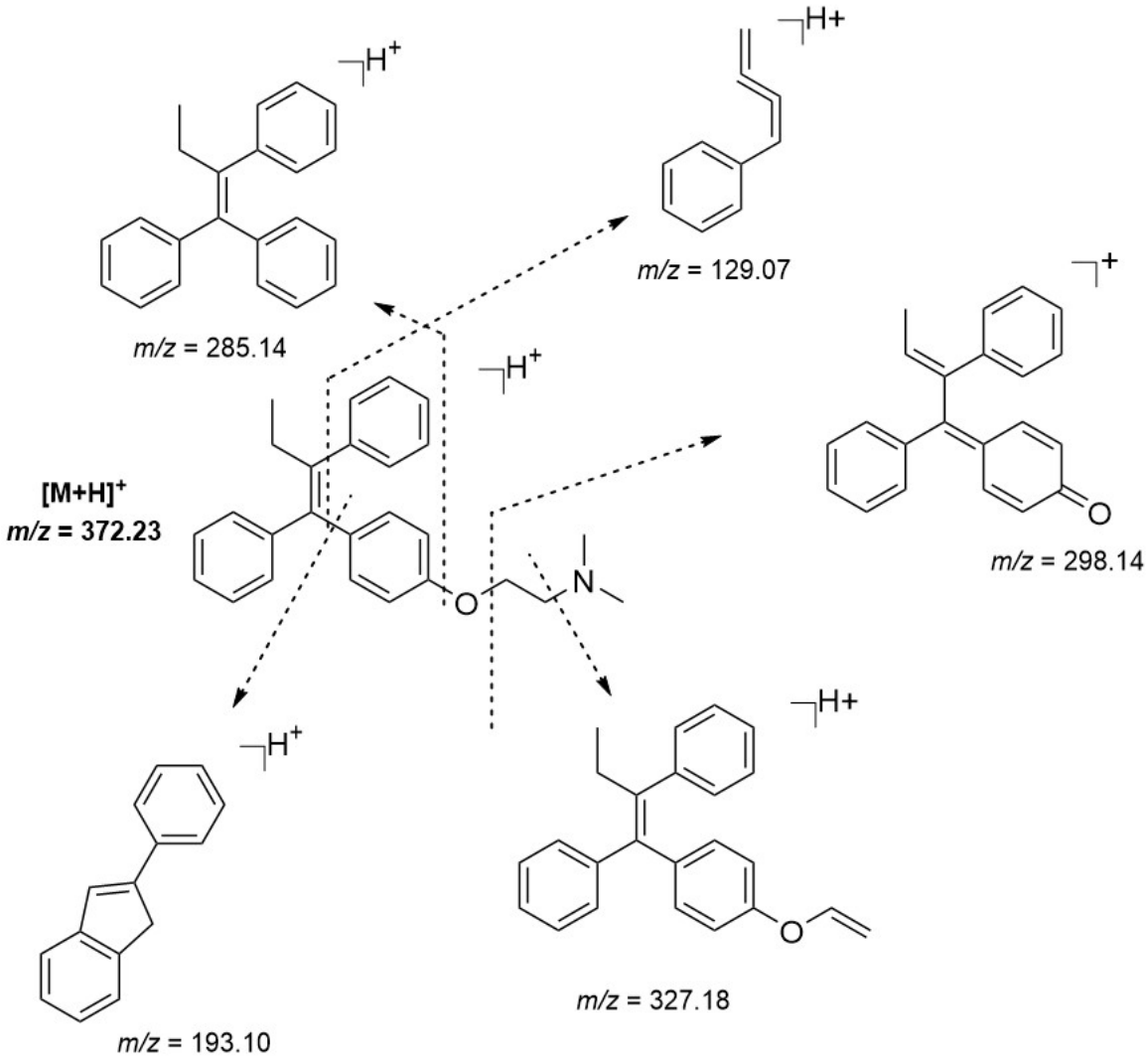

Figure 25 - Proposed mass spectrometry fragmentation pathways for the protonated molecule of tamoxifen (*m/z* 372.23). This work.

The EIC *m/z*=388.259 peak obtained in **Figure 22** corresponds to hydroxyl-tamoxifen, but due to the low relative intensity (**Figure 23B**) of these peaks it wasn't possible to verify if it was 4-hydroxy-tamoxifen or α-hydroxy-tamoxifen.

The other peak EIC *m/z*=358.2155 peak obtained in **Figure 22**, was analysed and by comparison with the theoretical patterns (**Figure 24A**) it was discovered that this mass-to-charge value corresponds to the *N*-desmethyl-tamoxifen. In this case, the peaks obtained present high intensities (**Figure 22**) meaning that this metabolite is produced in high quantities by CYP2C8 protein in comparison with the other metabolites. This is in line with the fact that *N*-desmethyltamoxifen is a major tamoxifen metabolite[63].

Finally, the last peak obtained with EIC *m/z*=344.200 in **Figure 22**, was analysed and by comparison with the theoretical patterns (**Figure 24B**) it was discovered that it corresponds to *N*,*N*-didesmethyl-tamoxifen. The peaks obtained in **Figure 24B** presented low intensities. The production of this metabolite occurs by oxidation of *N*-desmethyl-tamoxifen by CYP2C8, which means that from first

oxidation process of tamoxifen the first product formed is always *N*-desmethyl-tamoxifen, which is in accordance with the higher levels of this metabolite and also with the reported relative abundances of both metabolites when formed *in vivo*[63].

In summary, CYP2C8-mediated oxidation of tamoxifen was observed in this study and the major metabolites formed are *N*-desmethyl-tamoxifen and *N,N*-didesmethyl tamoxifen, presented in **Figure 26**.

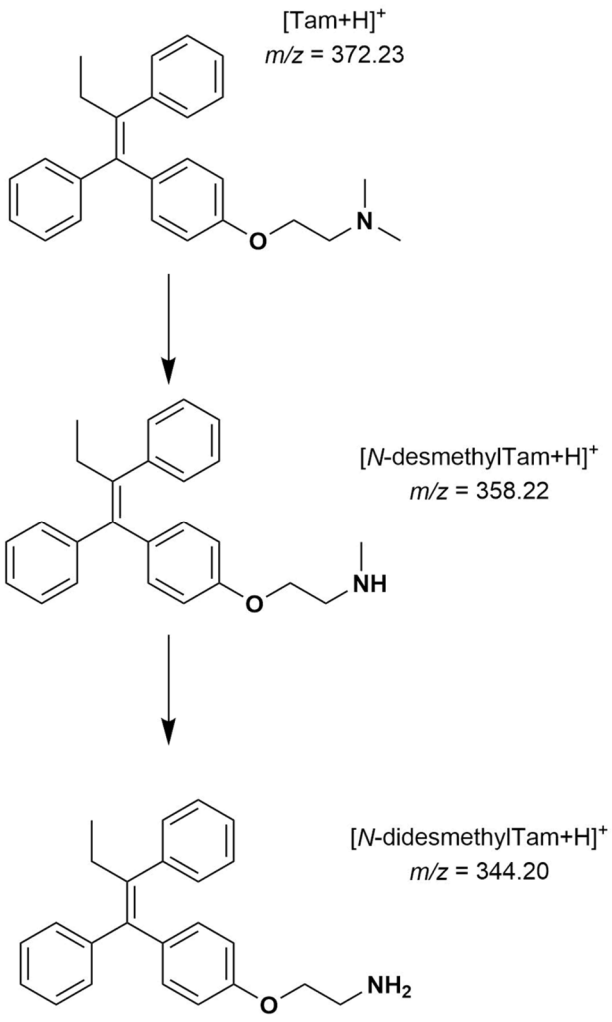

Figure 26 – Tamoxifen metabolites produced from CYP2C8-mediated oxidation.

4. Conclusions

Nowadays, drug-induced liver injury (DILI) is a serious clinical problem that can lead to acute liver failure and therefore, can require liver transplant and even cause death. Both pharmaceutical industries and regulatory authorities are concerned with this problematic issue, thereupon raising safety standards. Importantly, hepatotoxicity is the leading cause of drug withdrawal from the market.

Several drugs are associated with toxic side effects of varying degree, ranging from minor to serious and lethal. Metabolic activation, particularly during phase I and phase II metabolism, may lead to toxic metabolites. Phase I reactions include the formation of oxidation products by cytochrome P450 (CYP) systems while phase II reactions include, amongst others, sulfonylation of either the drug or its phase I metabolites. Both of these types of drug metabolites have the potential to react with bionucleophiles, such as protein and DNA, forming covalent adducts that may lead to the impairment of the cellular function of those molecules.

The relation between bioactivation and hepatotoxicity is not quite understood, reinforcing the need to develop metabolic efficient *in vitro* models capable of assessing drug bioactivation in pre-market phases. Moreover, the development of these *in vitro* models overcomes ethical issues related with dosage and administration of compounds to subjects in *in vivo* studies

The ultimate goal of this work was to obtain a large bacteria-based enzymatic model of the drug metabolism pathways using different isoforms of various enzymes simultaneously and recover the drug metabolites, validating it with well–studied model substrates and paving the way to a simple and effective drug metabolism prediction system.

Thus, recombinant vectors were successfully prepared for CYP2C8, while for SULT1A1, CYP2D6 and CYP3A4 it was not possible until now to isolate positive transformants. CYP2C8 and SULT1B1 enzymes were the only proteins selected for further overexpression and purification. SULT1B1 protein was expressed at high quantities, while CYP2C8 expression has to be optimized to achieve higher scale production. Both proteins were purified by Immobilized Metal Affinity Chromatography and by size exclusion chromatography.

The produced SULT1B1 protein presented the desired activity and its kinetic parameters were characterized by the 2-naphthol sulfonation assay and using different substrates, such as resorcinol, phenol and quercetin.

His-tagged human recombinant sulfotransferase isoform SULT1B1 and CYP isoform 2C8 expressed in *Escherichia coli* were shown to maintain their described metabolic activity using model substrates. SULT1B1 was shown to catalyse the formation of sulfoxy-resorcinol from resorcinol, while CYP2C8 was found to metabolise nevirapine to hydroxyl-nevirapine and tamoxifen to its desmethylated and hydroxylated metabolites.

Overall, our results suggest that our bacterial model is a promising way to obtain active enzymes useful to probe the metabolic pathways of already approved or new drugs to avoid further risks in patients' life.

Further studies will be conducted in order to study the metabolic activity of other CYP and SULT isoforms already subcloned to pET expression vectors in this work.

Moreover, the immobilization of these proteins on a solid matrix will be attempted in order to develop a flow-through reactor in which drugs and co-factors can be loaded and metabolites eluted in two simple sequential steps. This will largely reduce the costs of a drug metabolism study, allowing for the faster identification of possible reactive metabolites.

5. References

1. Almazroo OA, Miah MK, Venkataramanan R. Drug Metabolism in the Liver. *Clin Liver Dis*. 2017;21(1):1-20. doi:10.1016/j.cld.2016.08.001.

2. Testa B, Krämer SD. The Biochemistry of Drug Metabolism – An Introduction. *Chem Biodivers*. 2006;3(10):1053-1101. doi:10.1002/cbdv.200690111.

3. Anthérieu S, Chesné C, Li R, Guguen-Guillouzo C, Guillouzo A. Optimization of the HepaRG cell model for drug metabolism and toxicity studies. *Toxicol in Vitro*. 2012;26(8):1278-1285. doi:10.1016/j.tiv.2012.05.008.

4. Meyer UA. Overview of enzymes of drug metabolism. *J Pharmacokinet Biopharm*. 1996;24(5):449-459. doi:10.1007/BF02353473.

5. Gómez-Lechón MJ, Tolosa L, Donato MT. Metabolic activation and drug-induced liver injury: *In vitro* approaches for the safety risk assessment of new drugs. *J Appl Toxicol*. 2016;36(6):752-768. doi:10.1002/jat.3277.

6. Xu C, Li CY-T, Kong A-NT. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Arch Pharm Res*. 2005;28(3):249-268. doi:10.1007/BF02977789.

7. Furge LL, Guengerich FP. Cytochrome P450 enzymes in drug metabolism and chemical toxicology: An introduction. *Biochem Mol Biol Educ*. 2006;34(2):66-74. doi:10.1002/bmb.2006.49403402066.

8. Belloc C, Baird S, Cosme J, Lecoeur S, Gautier JC, Challine D, de Waziers I, Flinois JP, Beaune PH . Human cytochromes P450 expressed in *Escherichia coli*: production of specific antibodies. *Toxicology*. 1996;106(1-3):207-219. doi:10.1016/0300-483X(95)03178-I.

9. Flockhart D. Drug Interactions: Cytochrome P450 Drug Interaction Table. Indiana University School of Medicine. https://drug-interactions.medicine.iu.edu/Home.aspx. Published 2007. Accessed October 10, 2018.

10. Gamage N, Barnett A, Hempel N, Duggleby RG, Windmill KF, Martin JL, McManus ME. Human sulfotransferases and their role in chemical metabolism. *Toxicol Sci*. 2006;90(1):5-22. doi:10.1093/toxsci/kfj061.

11. Wang J, Falany JL, Falany CN. Expression and characterization of a novel thyroid hormone-sulfating form of cytosolic sulfotransferase from human liver. *Mol Pharmacol*. 1998;53(2):274-282. http://www.ncbi.nlm.nih.gov/pubmed/9463486.

12. Underhill GH, Khetani SR. Bioengineered Liver models for drug testing and cell differentiation Studies. *Cell Mol Gastroenterol Hepatol*. 2017. doi:10.1016/j.jcmgh.2017.11.012.

13. Ott LM, Ramachandran K, Stehno-Bittel L. An automated multiplexed hepatotoxicity and CYP induction assay using HepaRG cells in 2D and 3D. *SLAS Discov*. 2017;22(5):614-625.

doi:10.1177/2472555217701058.doi:10.1177/2472555217701058.

14. Thompson RA, Isin EM, Li Y, Weidolf L, Page K, Wilson I, Swallow S, Middleton B, Stahl S, Foster AJ, Dolgos H, Weaver R, Kenna JG. *In vitro* approach to assess the potential for risk of idiosyncratic adverse reactions caused by candidate drugs. *Chem Res Toxicol*. 2012;25(8):1616-1632. doi:10.1021/tx300091x.

15. Asha S, Vidyavathi M. *Cunninghamella* - A microbial model for drug metabolism studies - A review. *Biotechnol Adv*. 2009;27(1):16-29. doi:10.1016/j.biotechadv.2008.07.005.

16. Tonder JJ Van, Steenkamp V, Gulumian M. Pre-Clinical Assessment of the Potential Intrinsic Hepatotoxicity of Candidate Drugs. In: *New Insights into Toxicity and Drug Testing*. ; 2013:3-28. doi:10.5772/54792.

17. Amelian A, Wasilewska K, Megias D, Winnicka K. Application of standard cell cultures and 3D in vitro tissue models as an effective tool in drug design and development. *Pharmacol Reports*. 2017;69(5):861-870. doi:10.1016/j.pharep.2017.03.014.

18. Jaroch K, Jaroch A, Bojko B. Cell cultures in drug discovery and development: The need of reliable in vitro-in vivo extrapolation for pharmacodynamics and pharmacokinetics assessment. *J Pharm Biomed Anal*. 2018;147:297-312. doi:10.1016/j.jpba.2017.07.023.

19. Ekins S, Ring BJ, Grace J, McRobie-Belle DJ, Wrighton SA. Present and future in vitro approaches for drug metabolism. *J Pharmacol Toxicol Methods*. 2001;44:313-324. doi:S1056-8719(00)00110-6 [pii].

20. Parikh A, Gillam EMJ, Guengerich FP. Drug metabolism by *Escherichia coli* expressing human cytochromes P450. *Nat Biotechnol*. 1997;15(8):784-788. doi:10.1038/nbt0897-784.

21. Faber K, Azerad R. Biotransformations. In: *Advances in Biochemical Engineering*. 1st ed. Berlin: Springer-Verlag Berlin Heidelberg; 1999:175-202. doi: 10.1007/3-540-69791-8

22. Berg J, Tymoczko J, Stryer L. Eukaryotic Protein Synthesis Differs from Prokaryotic Protein Synthesis Primarily in Translation Initiation. In: *Biochemistry*. 5th ed. New York: W H Freeman; 2002.

23. Marinho AT, Godinho ALA, Novais DA, Antunes AMM, Marques MM, Ramos T, Dias CG, Monteiro EC, Pereira SA. Development and validation of an HPLC-UV method for quantifying nevirapine and its main phase I metabolites in human blood. *Anal Methods*. 2014;6(5):1575. doi:10.1039/c3ay41911h.

24. Riska P, Lamson M, Macgregor T, Sabo J, Hattox S, Pav J, Keirns J. Disposition and biotransformation of the antiretroviral drug nevirapine in humans. *Drug Metab Dispos*. 1999;27(8):895-901. doi:10.1021/es5024493.

25. Erickson DA, Mather G, Trager WF, Levy RH, Keirns JJ. Characterization of the *in vitro* biotransformation of the HIV-1 reverse transcriptase inhibitor nevirapine by human hepatic

cytochromes P-450. *Drug Metab Dispos*. 1999;27(12):1488-1495.

26. Pinheiro PF, Pereira SA, Harjivan SG, Martins IL, Marinho AT, Cipriano M, Jacob CC, Oliveira NG, Castro MF, Marques MM, Antunes AMM, Miranda JP. Hepatocyte spheroids as a competent *in vitro* system for drug biotransformation studies: nevirapine as a bioactivation case study. *Arch Toxicol*. 2017;91(3):1199-1211. doi:10.1007/s00204-016-1792-x.

27. Sulkowski MS, Thomas DL, Mehta SH, Chaisson RE, Moore RD. Hepatotoxicity associated with nevirapine or efavirenz-containing antiretroviral therapy: Role of hepatitis C and B infections. *Hepatology*. 2002;35(1):182-189. doi:10.1053/jhep.2002.30319.

28. Solomons TW., Fryhle CB. Nuclear Magnetic Ressonance and Mass Spectrometry. In: *Organic Chemistry*. 10th ed. John Wiley & Sons, Ltd; 2009:426-458. doi: 10.1021/ed073pA313.1

29. Korfmacher WA. Principles and applications of LC-MS in new drug discovery. *Drug Discov Today*. 2005;10(20):1357-1367. doi:10.1016/S1359-6446(05)03620-2.

30. Gamboa da Costa G, McDaniel-Hamilton LP, Heflich RH, Marques MM, Beland FA. DNA adduct formation and mutant induction in Sprague-Dawley rats treated with tamoxifen and its derivatives. *Carcinogenesis*. 2001;22(8):1307-1315. http://www.ncbi.nlm.nih.gov/pubmed/11470763.

31. Harjivan, S. G. Chemical modification of bionucleophiles by NNRTI metabolites. 2017. PhD thesis in Chemistry. Instituto Superior Técnico, Universidade de Lisboa.

32. Sambrook JF, Russell DW. *Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harbor Laboratory Press; 2001. doi: 10.1016/0167-7799(91)90068-S

33. Dagert M, Ehrlich SD. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene*. 1979;6(1):23-28. http://www.ncbi.nlm.nih.gov/pubmed/383576.

34. Hanahan D. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol*. 1983;166(4):557-580. http://www.ncbi.nlm.nih.gov/pubmed/6345791.

35. Deutscher J. The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol*. 2008;11(2):87-93. doi:10.1016/j.mib.2008.02.007.

36. Studier FW. Protein production by auto-induction in high-density shaking cultures. 2005. doi:10.1016/j.pep.2005.01.016.

37. Hansen LH, Knudsen S, Sørensen SJ. The effect of the lacY gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens*. *Curr Microbiol*. 1998;36(6):341-347. http://www.ncbi.nlm.nih.gov/pubmed/9608745.

38. Shapiro AL, Viñuela E, Maizel J V. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem Biophys Res Commun*. 1967;28(5):815-820. http://www.ncbi.nlm.nih.gov/pubmed/4861258.

39. Frame LT, Ozawa S, Nowell SA, Chou HC, DeLongchamp RR, Doerge DR, Lang NP, Kadlubar FF. A simple colorimetric assay for phenotyping the major human thermostable phenol sulfotransferase (SULT1A1) using platelet cytosols. *Drug Metab Dispos*. 2000;28(9):1063-1068. http://www.ncbi.nlm.nih.gov/pubmed/10950850.

40. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front Microbiol*. 2014;5(APR):172. doi:10.3389/fmicb.2014.00172.

41. Sørensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol*. 2005;115(2):113-128. doi:10.1016/j.jbiotec.2004.08.004.

42. Fujita K, Nagata K, Ozawa S, Sasano H, Yamazoe Y. Molecular cloning and characterization of rat ST1B1 and human ST1B2 cDNAs, encoding thyroid hormone sulfotransferases. *J Biochem*. 1997;122(5):1052-1061. http://www.ncbi.nlm.nih.gov/pubmed/9443824. Accessed October 12, 2018.

43. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif*. 2004;37(1):203-206. doi:10.1016/j.pep.2004.04.025.

44. Carrió M., Villaverde A. Construction and deconstruction of bacterial inclusion bodies. *J Biotechnol*. 2002;96(1):3-12. doi:10.1016/S0168-1656(02)00032-9

45. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb Cell Fact*. 2015;14:41. doi:10.1186/s12934-015-0222-8.

46. Bornhorst JA, Falke JJ. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol*. 2000;326:245-254. http://www.ncbi.nlm.nih.gov/pubmed/11036646. Accessed October 3, 2018.

47. Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta - Gen Subj*. 2006;1760(9):1304-1313. doi:10.1016/j.bbagen.2006.03.027.

48. Bartlow P, Uechi GT, Cardamone JJ, Sultana T, Fruchtl M, Beitle RR, Ataai MM. Identification of native *Escherichia coli* BL21 (DE3) proteins that bind to immobilized metal affinity chromatography under high imidazole conditions and use of 2D-DIGE to evaluate contamination pools with respect to recombinant protein expression level. *Protein Expr Purif*. 2011;78(2):216-224. doi:10.1016/J.PEP.2011.04.021.

49. Noubhani AM, Dieryck W, Bakalara N, Latxague L, Santarelli X. Evaluation of chromatographic recycling for imidazole used in the chromatographic purification of His-tag recombinant proteins. *J Chromatogr B*. 2003;790(1-2):153-159. doi:10.1016/S1570-0232(03)00088-6.

50. Lindner P, Bauer K, Krebber A, Nieba L, Kremmer E, Krebber C, Honegger A, Klinger B, Mocikat

R, Plückthun A . Specific detection of His-tagged proteins with recombinant anti-His tag scFv-phosphatase or scFv-Phage fusions. *Biotechniques*. 1997;22(1):140-149. doi:10.2144/97221rr01.

51. Riches Z, Stanley EL, Bloomer JC, Coughtrie MWH. Quantitative evaluation of the expression and activity of five major sulfotransferases (SULTs) in human tissues: The SULT "Pie"; *Drug Metab Dispos*. 2009;37(11):2255-2261. doi:10.1124/dmd.109.028399.

52. Tabrett CA, Coughtrie MWH. Phenol sulfotransferase 1A1 activity in human liver: kinetic properties, interindividual variation and re-evaluation of the suitability of 4-nitrophenol as a probe substrate. *Biochem Pharmacol*. 2003;66(11):2089-2097. doi:10.1016/S0006-2952(03)00582-3.

53. Bisswanger H. Enzyme assays. *Perspect Sci*. 2014;1(1-6):41-55. doi:10.1016/J.PISC.2014.02.005.

54. Enzyme Database - BRENDA. https://www.brenda-enzymes.org/. Accessed October 11, 2018.

55. English BP, Min W, van Oijen AM, Lee KT, Luo G, Sun H, Cherayil BJ, Kou SC, Xie XS. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat Chem Biol*. 2006;2(2):87-94. doi:10.1038/nchembio759.

56. Eisenthal R, Danson MJ, Hough DW. Catalytic efficiency and kcat/KM: a useful comparator? *Trends Biotechnol*. 2007;25(6):247-249. doi:10.1016/j.tibtech.2007.03.010.

57. Hutzler JM, Tracy TS. Atypical kinetic profiles in drug metabolism reactions. *Drug Metab Dispos*. 2002;30(4):355-362. http://www.ncbi.nlm.nih.gov/pubmed/11901086. Accessed October 9, 2018.

58. Ung D, Nagar S. Variable Sulfation of Dietary Polyphenols by Recombinant Human Sulfotransferase (SULT) 1A1 Genetic Variants and SULT1E1. *Drug Metab Dispos*. 2007;35(5):740-746. doi:10.1124/dmd.106.013987.

59. Foldes A, Meek JL. Rat brain phenolsulfotransferase: partial purification and some properties. *Biochim Biophys Acta*. 1973;327(2):365-374. http://www.ncbi.nlm.nih.gov/pubmed/4778939.

60. Bisswanger H. *Enzyme Kinetics : Principles and Methods*. 3rd Edition. Weinhem, Germany. Wiley-VCH Vrlag GmbH Co. 2017. doi: 10.1002/9783527622023

61. Malojcic G, Owen RL, Grimshaw JPA, Brozzo MS, Dreher-Teo H, Glockshuber R. A structural and biochemical basis for PAPS-independent sulfuryl transfer by aryl sulfotransferase from uropathogenic *Escherichia coli. Proc Natl Acad Sci*. 2008;105(49):19217-19222. doi:10.1073/pnas.0806997105.

62. Ren C, Fan-Havard P, Schlabritz-Loutsevitch N, Ling Y, Chan KK, Liu Z. A sensitive and specific liquid chromatography/tandem mass spectrometry method for quantification of nevirapine and its five metabolites and their pharmacokinetics in baboons. *Biomed Chromatogr*. 2010;24(7):717-726. doi:10.1002/bmc.1353.

63. Gamboa da Costa G, Hamilton LP, Beland FA, Marques MM. Characterization of the major DNA adduct formed by α-Hydroxy-N-desmethyltamoxifen *in vitro* and *in vivo*. 2000. doi:10.1021/TX990187B.

64. Gamboa da Costa G, Marques MM, Beland FA, Freeman JP, Churchwell MI, Doerge DR. Quantification of tamoxifen DNA adducts using on-line sample preparation and HPLC-electrospray ionization tandem mass spectrometry. *Chem Res Toxicol.* 2003;16(3):357-366. doi:10.1021/TX020090G.

65. Teunissen SF, Rosing H, Seoane MD, Brunsveld L, Schellens JH, Schinkel AH, Beijnen JH. Investigational study of tamoxifen phase I metabolites using chromatographic and spectroscopic analytical techniques. J Pharm Biomed Anal. 2011;55(3):518-526. doi:10.1016/J.JPBA.2011.02.009.

6. Annexes

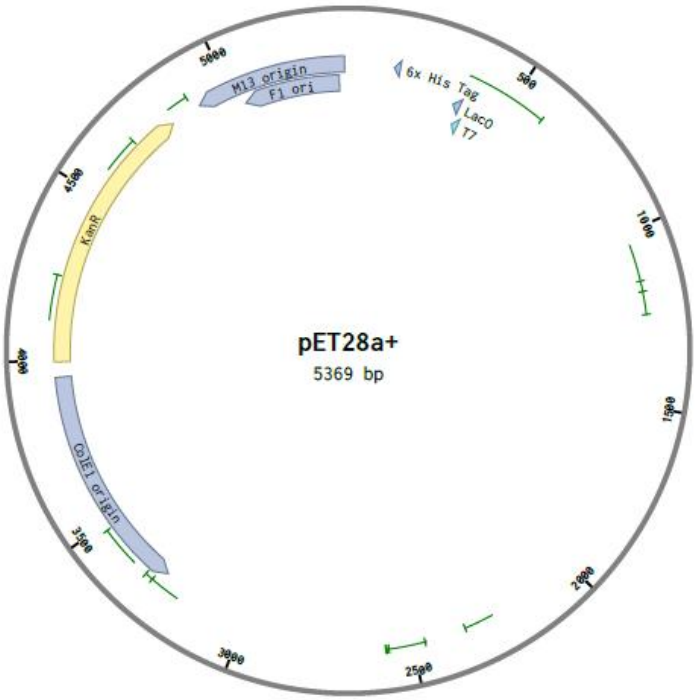

Figure A1 - Vector map for pET28a(+).

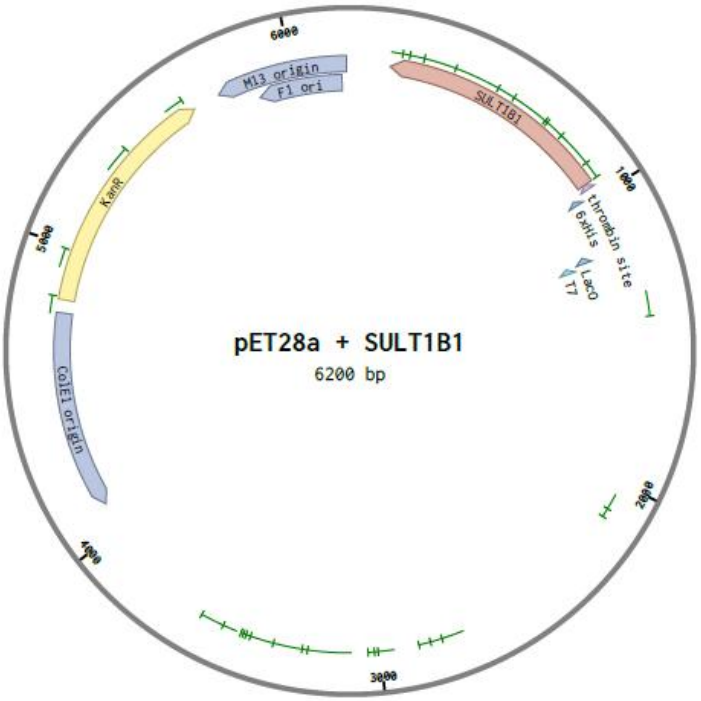

Figure A2 - Vector map for pET28a-SULT1B1.

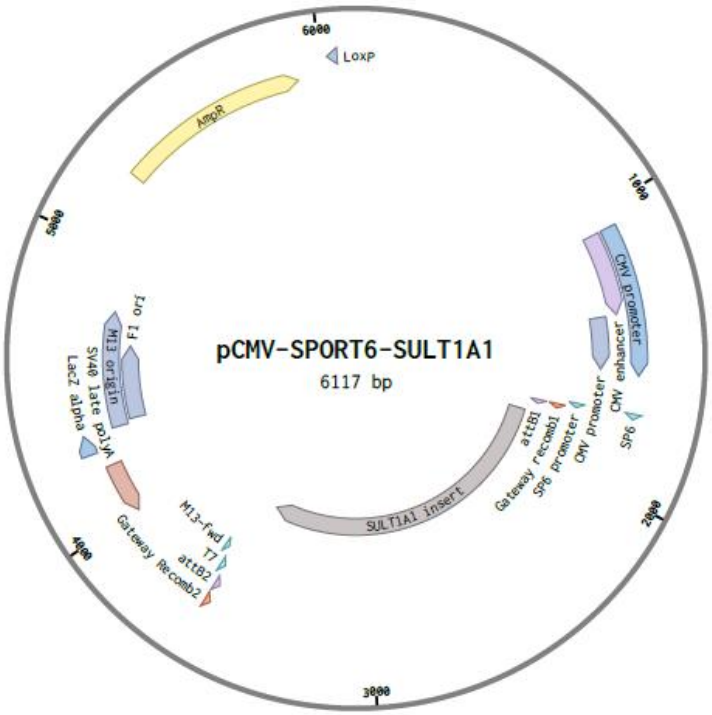

Figure A3 - Vector map for pCMV-SPORT6-SULT1A1.

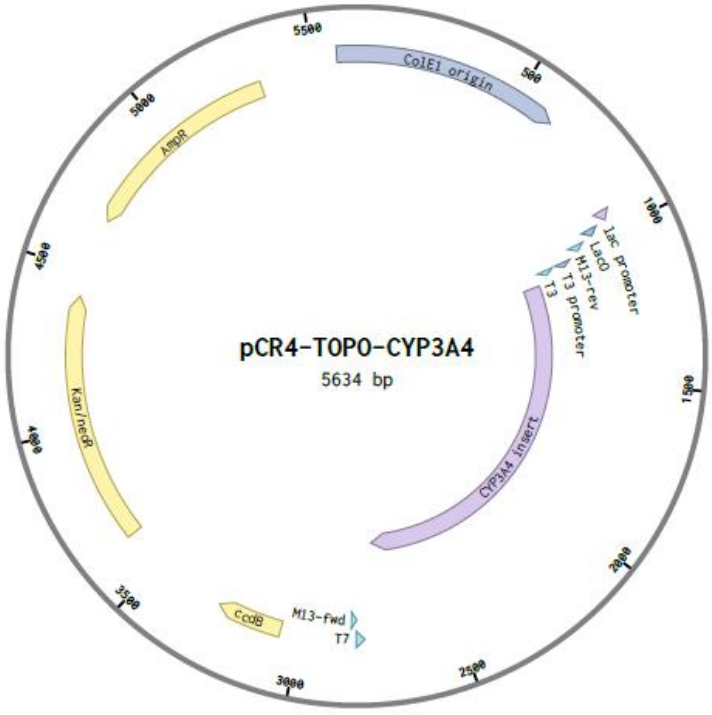

Figure A4 - Vector map for pCR4-TOPO-CYP3A4.

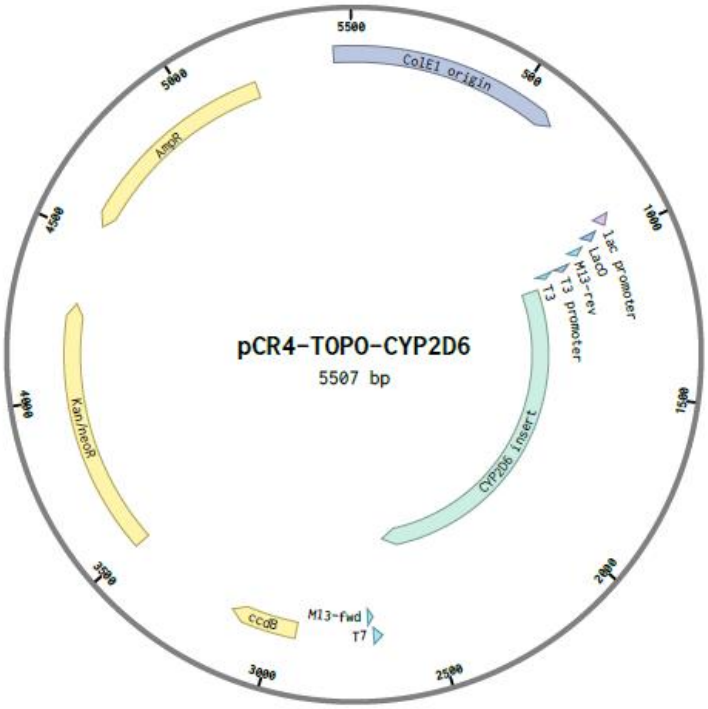

Figure A5 - Vector map for pCR4-TOPO-CYP2D6.

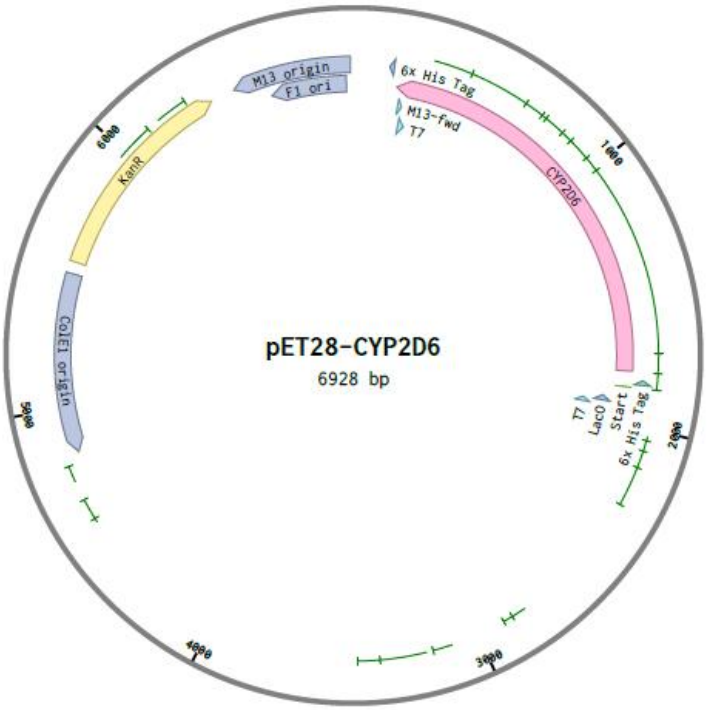

Figure A6 - Vector map for the prepared pET28a-CYP2D6.

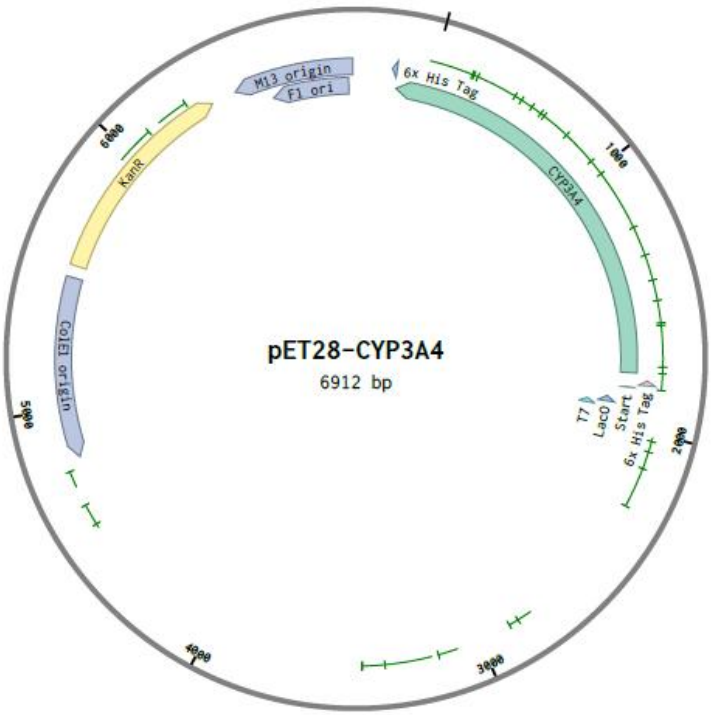

Figure A7 - Vector map for the prepared pET28a-CYP3A4.

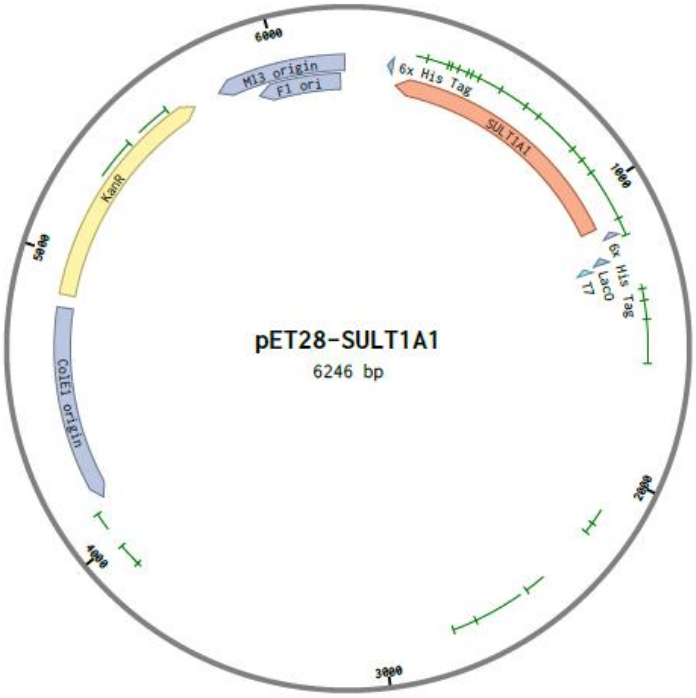

Figure A8 - Vector map for the prepared pET28a-SULT1A1.