RESEARCH ARTICLE

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

Pedro Rosado[1], Gonçalo C. Justino[1], M. Matilde Marques[1]

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 genes 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-napthol 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 and, through LC/HRMS analysis of the obtained metabolites, we observed the occurrence of SULT1B1-mediated sulfonylation and CYP2C8-mediated oxidations, which indicate 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** DILI; CYP450s; SULTs; Bacterial model; Drug metabolism; Nevirapine

Drug-induced liver injury is the most frequent cause of hepatic dysfunction and drug failures during clinical trials. Once these events happen frequently, approved drugs are removed from the market[1–3]. The main organ involved in biotransformation of xenobiotic compounds is the liver comprising mainly phase I and phase II metabolic pathways. Cytochrome P450 (CYP) and sulfotransferase (SULT) enzymes are the main involved in phase I and phase II reactions, respectively. Moreover, these enzymes are also related to the metabolic bioactivation of drugs due to the formation of reactive metabolites which can lead to hepatotoxicity[1,2,4–6]. Regarding CYP450s, the most abundant enzyme expressed in the human liver is CYP3A4 being responsible for oxidation of one-half of the drugs in the human organism. Despite CYP2D6 accounts for only 2% of the CYPs found in the liver, it has been reported to be involved in the metabolism of 25% of commonly prescribed pharmaceuticals, namely, antiarrhythmics, antidepressants, antipsychotics and some anti-cancer drugs[4,7,8]. Additionally, CYP2C8 enzyme is also important once it is mainly responsible for the metabolism of antidiabetics and some cancer treatment drugs, for instance, paclitaxel[8]. So far, SULT1A1 is the most extensively studied sulfonation enzyme, which metabolizes phenols, alcohols and amines[4]. Considering SULT1B1, it appears to be related with the thyroid hormone metabolism and it is present in the liver, small intestine, colon and blood leukocytes. Importantly, this enzyme can be responsible for bioactivating procarcionogens to reactive electrophiles[9]. Microbial models can be applied to assess and predict drug biotransformation and toxicity, overcoming ethical problems related with the use of *in vivo* studies in pre-clinical stages. Moreover these models present several advantages comparing with *in vitro* mammalian models, for instance simpler methods, higher amount of metabolites formed and easier scale-up[10–13].Previous studies in *Escherichia coli* already shown that is possible to express functional human CYPs at high levels. Since *E. coli* is intensively studied, it seems ideal for further drug metabolism research[10,14]. 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.

## Materials and Methods

**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[15,16].

**Strains, plasmids and growth conditions** In this work, *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 proper media with 25% (v/v) glycerol. Prior to use, frozen cells were thawed on ice and cultured in appropriate fresh media at 37ºC overnight (o/n), approximately 16 hours, with orbital agitation. Except where mentioned, all *E. coli* cultures were performed in

[1] Centro de Química Estrutural (CQE), Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal

solid or liquid Lysogeny (LB) growth medium (Liofilchem, Italy).

The plasmids used in this work were pET28a(+) (Novagen, Nottingham, UK), pCR4-TOPO CYP3A4, pCR4-TOPO CYP2D6 and pCMV-SPORT 6 SULT1A1, pET28a_SULT1B1 and pCW_CYP2C8.The pCR4-TOPO CYP3A4, pCR4-TOPO CYP2D6 and pCMV-SPORT 6 SULT1A1 plasmids were provided by 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 (Sigma-Aldrich, St. Louis, MO), ampicillin (100 $\mu g.mL^{-1}$) and kanamycin (30 or 50 $\mu g.mL^{-1}$) were used when needed as recommended. 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, the medium was supplemented with 30 or 50 $\mu g.mL^{-1}$ kanamycin.

**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 (Ilustra™ plasmidPrep Mini Spin Kit (GE Healthcare, UK)). Afterwards, plasmid DNA (pDNA) was extracted, purified and stored at -20ºC for further use. Finally, the DNA concentration and purity was assessed 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[17].

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

**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 min 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). 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.

**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 manufacturers' 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.

**Polymerase Chain Reaction** PCR was set up with One-Fusion DNA Polymerase (GeneON, Germany) using optimal reaction conditions with a cycling program adjusted to the chosen primers (**Table 1**). PCR products were analysed by agarose gel electrophoresis and purified using the GF-1 AmbiClean Kit (PCR & Gel) (Vivantis Technologies, Malaysia).

**Table 1 - List of the primers used in this work.** CYP – cytochrome P450; SULT – sulfotransferase

| | Amplification Sequence | Length (nt) |
|---|---|---|
| **CYP2D6_Fw_NheI** | 5' GTGAGGCAGCTAGC GGGCTAGAAGCACTG 3' | 29 |
| **CYP2D6_Rev_Hind** | 5' CCCGCCAAGCTTTTCC CAGTCACGACG 3' | 27 |
| **SULT1A1_Fw_NdeI** | 5' GGCCAGGCATATGCAG CTGATCCAGGACACCTC 3' | 33 |
| **SULT1A1_Rev_Hin** | 5' CGTGAGAAGCTTGGTC AGGTTTGATTCGCACAC 3' | 33 |
| **CYP3A4_Fw_NdeI** | 5' ACAGTACATATGGCT CTCATCCCAGACTTGG 3' | 31 |
| **CYP3A4_Rev_XhoI** | 5' ACTCTCGAGAGGGCGA ATTGAATTTAGCGGC 3' | 31 |

**Preparation of *E. coli* BL21(DE3) competent cells** *E. coli* BL21(DE3) competent cells were prepared as described by *Dagert et. al*[18]. Briefly, cells were growth o/n in LB medium without antibiotics and harvested by centrifugation at 16 000 x *g* during 5 min 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[18].

**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[19]. 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.

**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.

**Protein Overexpression of CYP2C8 and SUT1B1 recombinant proteins in *E. coli*** A single colony of the transformants obtained was inoculated in 5 mL of LB medium in a 50 mL flask, supplemented with kanamycin (100 $\mu g.mL^{-1}$) 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 NZY auto-induction medium, 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 OD at 600 nm of approximately 0.5[20–22].

**Cell Lysis** Cells were harvested from liquid cultures by centrifugation (5 000 x *g* for 10 min at 4ºC). Supernatants were discarded and the pellets weighted. Cells were lysed by using NZY Bacterial Cell Lysis Buffer (NZYtech) and following manufacturer's protocol.

**Purification by Immobilized Metal Affinity Chromatography** Protein purification was performed by Immobilized Metal Affinity Chromatography (HisTrap™ FF Crude – GE Healthcare, UK). 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.

**Desalting** PD-10 Desalting columns containing Sephadex G-25 resin (GE Healthcare, UK) were used 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.

**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. A standard curve was made in the same conditions of the experiments using bovine serum albumin (BSA); concentration of experimental samples was determined using this standard curve.

**SDS-PAGE** 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 min 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 min and then 150V for 60 min. The gel was stained with EzBlue™ Staining Reagent (Sigma-Aldrich, MO, USA) at RT o/n, with mild agitation and destained with water.

**Sulfotransferase 1B1 activity measurement** The activity of SULT 1B1 was assessed using the 2-napthol sulfonation assay, as described by Frame *et al*[23]. Other experiments were performed by changing the 2-napthol 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. SULT1B1 activity was determined by a spectrophotometer following the absorbance at 405 nm in different times of the experiment. Activities were measured with UV-1800 spectrophotometer (Shimadzu, Japan).

**SULT1B1-mediated sulfonylation studies by LC/MS** The standard assay mixture contained 5 µM PAPS solution, 2.5 mM 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 min in Amicon Vivaspin filters with a membrane cutoff of 10 kDa MWCO (GE Healthcare, UK). The eluted fraction was analysed by liquid chromatography-tandem high resolution mass spectrometry (LC/HRMS). 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.

**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 min in Amicon Vivaspin filters with a membrane cutoff of 10 kDa MWCO (GE Healthcare, UK). The permeated fraction was analysed by liquid chromatography-tandem high resolution mass spectrometry (LC/HRMS). Several chemicals were tested as possible CYP2C8 substrates, namely, tamoxifen, N-desmethyltamoxifen, nevirapine, arachidonic acid and novocaine.

**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

spectrophotometer 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. 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 ab acquisition rate of 5 Hz using a dynamic method with a fixed cycle time of 3 s. Dynamic exclusion duration was 0.4 min.

**Data Processing** Acquired data was processed by DataAnalysis 4.1 software (Bruker Daltonics, Bremen, Germany). 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.

## Results and discussion

**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 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. 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 (data not shown) 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 vectors 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 inserts 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.

**Expression and purification of CYP2C8 and SULT1B1 proteins** 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 1** presents typical results for SULT1B1 and CYP2C8 expression.

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[21,24,25]. 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 1A**); human SULT1B1 has a molecular mass of 34.9 kDa, in agreement with the observed band[26]. 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 (data not shown). The optimal conditions for protein expression are LB medium supplemented with kanamycin at a final concentration of 30 $\mu g.mL^{-1}$ and with 24h of cell growth, since the band around 34.9 kDa which is SULT1B1 molecular weight is wider and more intense.

Regarding CYP2C8 expression induction, the expected increase in protein production after incubation of the bacterial cells with IPTG did not occur, since overall the bands intensities are similar prior and after the addition of the inducer (**Figure 1B**). The protein of interest, which has approximately 52 kDa, was not detected among all the bands, presumably because it was expressed at very low

levels due to an harmful effect that heterologous proteins exert in the cells[21,24,27].

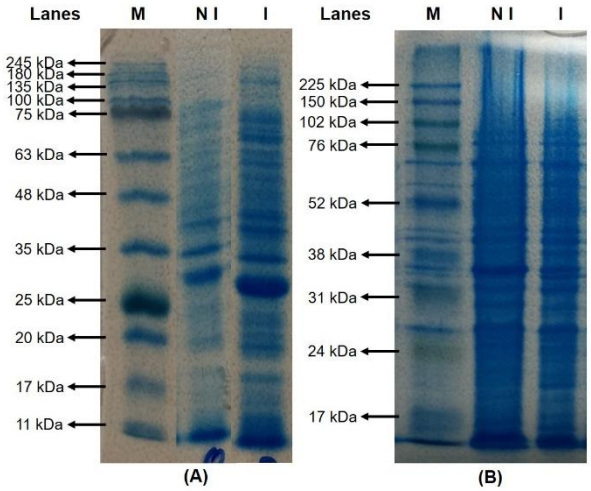


**Figure 1 - SDS-PAGE analysis of target protein expression induction in *E. coli* BL21(DE3) cells.** (A) *E. coli* BL21(DE3) cells containing SULT1B1 plasmid were analysed prior to (NI, not induced) and after induction (I, induced) with 1 mM IPTG. (M) Molecular weight marker Protein Marker II (Nzytech); (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).

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[21,24]. 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[21,24,25].

To test this hypothesis, cultures of *E. coli* transformed with the pET-28-CYP2C8 plasmid were also performed in LB auto-induction media, which consists of LB supplemented with glucose (0.5 g/L) and α-lactose (2.0 g/L), which allows cells to initially grow exclusively on glucose, promoting high cell density[21]. 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 media 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 [24,28,29]. 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[27].

As already mentioned, both proteins were purified by IMAC. A representative chromatogram for SULT1B1 and CYP2C8 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 (data not shown).

In sum, SULT1B1 presented more intense band at 34.9 kDa in comparison with CYP2C8 band (52 kDa), meaning that SULT1B1 protein was expressed in higher amounts. CYP2C8 purification was not so successful comparing with SULT1B1, since several bands were visible in the SDS-PAGE bisacrylamide gel, meaning that other proteins were present in the sample. The alternatives to improve CYP2C8 purification can be increase washing steps 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[30,31]. If the contaminants are in fact interacting with the column, the solution can be eluting the proteins in a stepwise gradient with the main goal of separating successfully the target protein from the contaminating proteins.

An additional step of desalting was performed in order to increase proteins purification, and a representative gel of the 3.5 mL of eluted fractions of SULT1B1 and CYP2C8 are shown in **Figure 2A** and **Figure 2B**, respectively.

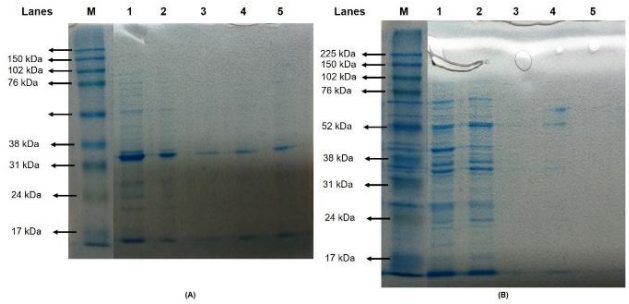


**Figure 2 - Analysis of the elution profile of the recombinant CYP2C8 and SULT1B1 proteins, purified by desalting process. A)** SDS-PAGE analysis of the elution profile for the SULT1B1 protein by desalting. Gel was stained with EzBlueTM Staining Reagent (Sigma-Aldrich, St. Louis, MO). M – Molecular weight marker Amersham™ ECL™ 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™ ECL™ 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 SULT1B1 sample was applied to the desalting column in a volume of 2.5 ml represented in the **Figure 2A**, 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 2A**). Since SULT1B1 protein is not retained in the column, it is immediately eluted at the beginning of the process as shown in **Figure 2A**, 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 2A**).

Concerning CYP2C8 protein purification, the sample was injected in a volume of 2.5 ml represented in the **Figure 2B**, 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 2B**, 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 2B** 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 which 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) as visible by the presence of an intense band at 52 kDa and other several bands with different molecular weights in lanes 1 and 2 (**Figure 2B**). Lanes 3, 4 and 5 in **Figure 2B** 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[32]. 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.

**Enzyme Activity Assays** The activity of SULT1B1 was assessed using the 2-naphthol sulfonation assay, a model substrate of the enzyme. The molar extinction coefficient used was 18200 $M^{-1}.cm^{-1}$ [23,33]. A plot of reaction rate *vs.* substrate concentration allows determining the kinetic parameters of the enzyme, where it was 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.

SULT1B1 is an enzyme that catalyses the sulfonylation of various hydroxyl-groups and no sulfonylation was expected to occur in the presence of antipyrine. In contrast, although the substrates cinchonine, borneol and ethylmorphine have hydroxyl groups, they did not suffer sulfonylation. Hydroxyl groups in these substrates are alkylic, in opposition to the aromatic OH groups in phenol, resorcionol and quercetin, for which catalysis occurred. This implies that the produced SULT1B1 is a phenol-acting sulfotransferase, in agreement with published data[34].

Resorcinol, quercetin and phenol present a hyperbolic rate profile which means that follows a Michaelis-Menten equation[35]. The values analysed at 30 µM of substrate concentration for quercetin and phenol did not follow the hyperbolic curve, which can be related with experimental errors. 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 $V_{max}$ (data not shown). The turnover number, $k_{cat}$, and the catalytic efficiency, $k_{cat}/K_M$, were computed from the maximal velocity using $V_{max} = [E_{free}]k_{cat} \approx [E_{total}]k_{cat}$[36].

It was possible to determine the maximum velocity ($V_{max}$) values in the experiment which were 4.227 x$10^{-5}$ µmol.min$^{-1}$, 3.764 x$10^{-5}$ µmol.min$^{-1}$ and 7.945 x$10^{-5}$ µmol.min$^{-1}$ for resorcinol, quercetin and phenol, respectively (**Table *2***), 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 x$10^{-6}$ $\mu M^{-1}$ $s^{-1}$, 4.241x$10^{-6}$ $\mu M^{-1}$ $s^{-1}$ and 3.152 x$10^{-6}$ $\mu M^{-1}$ $s^{-1}$. 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 2**). 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[35,37,38].

**Table 2 - 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$^{-2}$ |

Resorcinol and quercetin present a similar affinity to SULT1B1 enzyme since both present similar $K_M$ values (**Table *2***), although resorcinol has a lower $K_M$ (69.09 µM) which means that it 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 ontained[39,40]. 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[41]. 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[34].

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 (data not shown) 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. 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[33,42,43].

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

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 (data not shown).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 (data not shown). 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 (data not shown) was analysed and the isotopic pattern for sulfoxy-resorcinol was obtained, matching the predicted isotopic profile for a compound with that chemical formula.Therefore, resorcinol was shown to be metabolised by SULT1B1 enzyme, producing a monosulfoxy-resorcinol.

**CYP2C8-mediated oxidation studies 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.

**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. 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 (data not shown).

These ions were further studied using tandem mass spectrometry, and the resulting MS/MS spectra are shown in **Figure 3**.

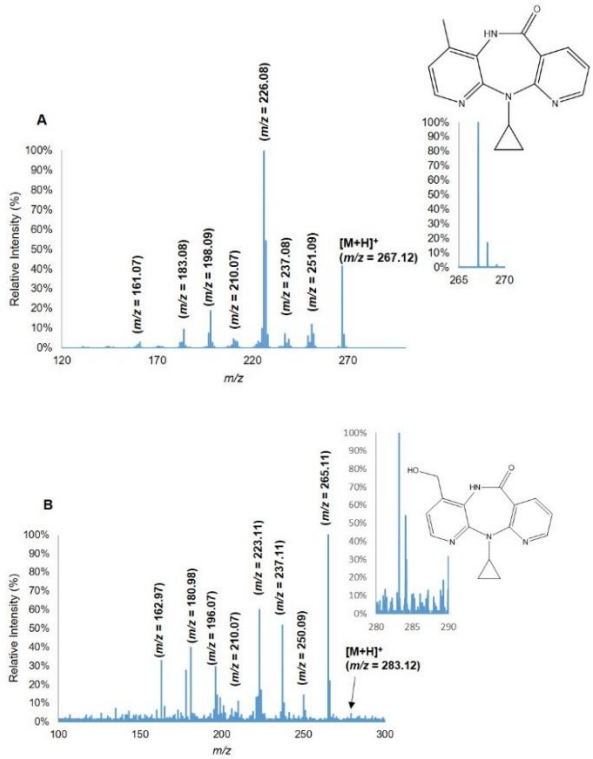


**Figure 3 - ESI-MS/MS spectra from protonated nevirapine metabolites [M + H]$^+$ ions.** (A) Protonated nevirapine (*m/z* 267.12 ion); (B) Protonated hydroxylated nevirapine (*m/z* 283.12 ion). Insets are the isotopic patterns for nevirapine (A) and 12-hydroxy-nevirapine (B) in the MS spectra.

The MS/MS spectra of protonated nevirapine (**Figure 3A**), 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[44].Additional smaller peaks were also observed (**Figure 3 A**).

The MS/MS spectrum obtained for the ion at *m/z* 283.1190 is shown in **Figure 3**. 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[44]. In particular, the base peak at *m/z* 265.11 (**Figure 3B**) was fully consistent with loss of water to yield a benzylic-type cation that rearranged to a tropylium-like structure.

The identity of the ions attributed to each peak was also verified by comparing the observed isotopic profiles with the predicted ones (data not shown). 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.

When analysing the extracted ion chromatogram at *m/z* 227.0927, 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 (data not shown); 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 (data not shown). 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 , and involves the oxidation of the tertiary carbon of the cyclopropyl group, yielding a hemiaminal that undergoes spontaneous collapse with elimination of cyclopropanone.

**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) (data not shown).

Chromatographic separation was carried out on a C18 stationary phase and so, differences in polarity of obtained metabolites influence the relative retention. The first peak obtained with an EIC *m/z* = 372.2311 is in agreement with the expected *m/z* value for tamoxifen and the retention time for this compound is around 17.5 min (data not shown). The other peak (EIC *m/z* =388.2259) obtained presents a retention time of 17.6 min, 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 present retention times of 17.3 and 17.1 min, respectively (data not shown). 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* = 344.2000 is lower than for 358.215 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.

All the peaks previously mentioned were selected for fragmentation and their tandem mass spectra.

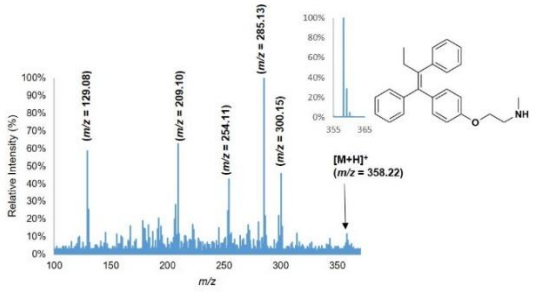


**Figure 4 - ESI-MS/MS spectra of protonated *N*-desmethyl-tamoxifen ions.** Protonated *N*-desMetTamoxifen (m/z 358.22 ion). Inset is the isotopic pattern for desMetTamoxifen.

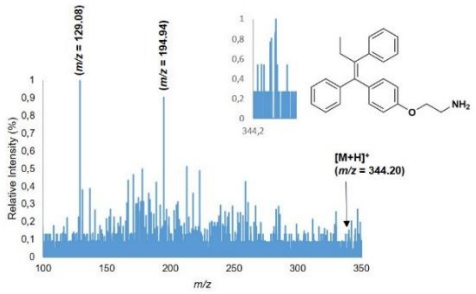


**Figure 5 - ESI-MS/MS spectra of protonated *N,N*-didesmethyl-tamoxifen ions.** Protonated *N,N*-didesMetTamoxifen (*m/z* 344.20 ion)**.** Inset is the isotopic pattern didesMetTamoxifen in MS spectra.

Several peaks were obtained in tamoxifen fragmentation 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*[45,46]. The EIC *m/z*=388.259 peak obtained (data not shown) corresponds to hydroxyl-tamoxifen, but due to the low relative intensity 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 was analysed and by comparison with the theoretical patterns (**Figure 4**) it was discovered that this mass-to-charge value corresponds to the *N*-desmethyl-tamoxifen. In this case, the peaks obtained present high intensities (data not shown) 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[47].

Finally, the last peak obtained with EIC *m/z*=344.200 was analysed and by comparison with the theoretical patterns (**Figure 5**) it was discovered that it corresponds to *N,N*-didesmethyl-tamoxifen. The peaks obtained in **Figure 5** presented low intensities. The production of this metabolite occurs from *N*-desmethyl-tamoxifen, which 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*[47]. 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.

**Final Remarks** The goal of this work was to obtain a large bacteria-based enzymatic model of the drug metabolism pathways using different enzyme isoforms simultaneously and recover the drug metabolites, validating it with well–studied model substrates. 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. 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.

**Abbreviations** CYP – Cytochrome P450 (NADPH:oxygen oxidoreductase); DILI – drug-induced liver injury; DNA – Deoxyribonucleic acid; ESI – Electrospray ionization; IMAC – Immobilized metal affinity chromatography; LB – Lysogeny Broth; o/n – overnight; MS – Mass Spectrometry; *m/z* – mass-to-charge ratio; NVP – Nevirapine; PRC – Polymerase Chain Reaction; SDS-PAGE – Sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SULT – sulfotransferases.

## References


1. 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.
2. Anthérieu S, Chesné C, Li R, Guguen-Guillouzo C, Guillouzo A. Optimization of the HepaRG cell model for drug metabolism and toxicity studies. *Toxicol Vitr*. 2012;26(8):1278-1285. doi:10.1016/j.tiv.2012.05.008.
3. 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.
4. 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.
5. Meyer UA. Overview of enzymes of drug metabolism. *J Pharmacokinet Biopharm*. 1996;24(5):449-459. doi:10.1007/BF02353473.
6. 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.
7. Pan Y, Abd-Rashid BA, Ismail Z, Ismail R, Mak JW, Ong CE. Heterologous Expression of Human Cytochromes P450 2D6 and CYP3A4 in *Escherichia coli* and Their Functional Characterization. *Protein J*. 2011;30(8):581-591. doi:10.1007/s10930-011-9365-6.
8. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug

metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther*. 2013;138(1):103-141. doi:10.1016/j.pharmthera.2012.12.007.

9. Gamage N, Barnett A, Hempel N, et al. Human sulfotransferases and their role in chemical metabolism. *Toxicol Sci*. 2006;90(1):5-22. doi:10.1093/toxsci/kfj061.

10. 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.

11. 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.

12. 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(September 2000):313-324. doi:S1056-8719(00)00110-6 [pii].

13. Faber K, Azerad R. Biotransformations. In: *Advances is Biochemical Engineering*. 1st ed. Berlin: Springer-Verlag Berlin Heidelberg; 1999:175-202.

14. Belloc C, Baird S, Cosme J, et al. 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.

15. 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.

16. Harjivan SG. Chemical modification pof bionucleophiles by NNRTI metabolites. 2017. PhD thesis in Chemistry. Instituto Superior Técnico, Universidade de Lisboa.

17. Sambrook JF, Russell DW. *Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harbor Laboratory Press; 2001.

18. 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.

19. 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. Accessed October 10, 2018.

20. 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.

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

22. 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. Accessed October 10, 2018.

23. Frame LT, Ozawa S, Nowell SA, et al. 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. Accessed October 2, 2018.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. Bartlow P, Uechi GT, Cardamone JJ, et al. 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.

32. Lindner P, Bauer K, Krebber A, et al. 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.

33. 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.

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

35. English BP, Min W, van Oijen AM, et al. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. *Nat Chem Biol*. 2006;2(2):87-94. doi:10.1038/nchembio759.

36. 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.

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

38. 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.

39. 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.

40. 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. Accessed October 11, 2018.

41. Bisswanger H. *Enzyme Kinetics : Principles and Methods*. https://www.wiley.com/en-us/Enzyme+Kinetics%3A+Principles+and+Methods%2C+3rd+Edition-p-9783527342518. Accessed October 13, 2018.

42. 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.

43. 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.

44. 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.

45. Gamboa da Costa G, Marques MM, Beland FA, Freeman JP,Churchwell MI and Doerge DR. Quantification of tamoxifen DNA adducts using on-line sample preparation and HPLC-electrospray ionization tandem mass spectrometry. 2003. doi:10.1021/TX020090G.

46. Teunissen SF, Rosing H, Seoane MD, et al. 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.

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