MOLECULAR MODELLING OF THE MOUSE CYTOCHROME P450 CYP2F2 BASED ON THE CYP102 CRYSTAL STRUCTURE TEMPLATE AND SELECTIVE CYP2F2 SUBSTRATE INTERACTIONS

David F.V. Lewis[1]*, Paul T. Bailey[2] and Lawrence K. Low[2]

[1]*School of Biomedical and Life Sciences, University of Surrey, Guildford, Surrey, UK and*
[2]*ExxonMobil Biomedical Sciences, Inc., Annandale, NJ, USA*

SUMMARY

The results of homology modelling of the mouse cytochrome P450, CYP2F2, are reported, based on the CYP102 crystallographic template. It is found that selective CYP2F2 substrates are able to fit the putative active site of the enzyme via aromatic π-π stacking and, in some cases, hydrogen-bonded interactions. Two alkylnaphthalenes were investigated via the model to evaluate whether they are likely to act as CYP2F2 substrates and, of these, 2-isopropyl-naphthalene was found to fit the putative active site, whereas 2-(2-hexadecyl)naphthalene was unable to do this, due to its bulky side-chain. Consequently, it is possible to utilize homology modelling to evaluate the likelihood of enzyme-substrate selectivity for CYP2F2 and predict routes of metabolism mediated by this enzyme. This procedure can therefore be used to investigate the potential for activation of xenobiotics via this enzyme, especially those related to known CYP2F substrates, such as naphthalene.

KEY WORDS

cytochrome P450, xenobiotic metabolism and activation, mouse CYP2F2, modelling, CYP102 structural template

* Author for correspondence:
David F.V. Lewis
School of Biomedical and Life Sciences
University of Surrey
Guildford, Surrey GU2 7XH, UK
e-mail: d.lewis@surrey.ac.uk

INTRODUCTION

Enzymes of the cytochrome P450 (CYP) superfamily are associated with the mono-oxygenation of a large range and variety of organic compounds, both endogenous and exogenous /1,2/. Included in the latter category are pharmaceutical agents and environmental pollutants, many of which are destined for human exposure. Molecular modelling of the enzyme-substrate interactions for P450-mediated pathways can assist in the rationalization of substrate selectivity and site of metabolism, although this process involves initial construction of the enzymes themselves. One of the most appropriate crystal structure templates for homology modelling of mammalian microsomal P450s is represented by the CYP102 haemoprotein domain, which is available as the substrate-bound enzyme for the production of P450 models. There is interest in the role of CYP2F subfamily enzymes /3-5/ in the activation and metabolism of xenobiotics, some of which are likely to show toxicity (e.g. naphthalene) following metabolic activation via the mediation of this enzyme /6-11/. Consequently, it is anticipated that homology modelling of the mouse orthologue, CYP2F2, may yield important structural information regarding substrate preferences for this isoform, and also assist in evaluating routes of metabolism mediated by CYP2F enzymes in general. Although the mouse orthologue was modelled in this study, it is likely that the active sites of CYP2F enzymes in other mammalian species, including *Homo sapiens,* will be similar, due to the fact that there is a relatively high degree of homology between their protein sequences. Furthermore, it would appear that both mouse and human CYP2F enzymes possess similar substrate selectivities.

METHODS

Using the CYP102 haemoprotein domain crystal structure (pdb code: lfag) as a template the mouse CYP2F isoform, CYP2F2, was generated via amino acid sequence homology. Figure 1 provides a multiple sequence alignment of CYP102 against 20 CYP2 family enzymes to indicate similarities and differences between various sequences in this family, and this was employed in homology modelling of CYP2F2 via a procedure involving amino acid residue replacement, deletion and insertion as required, followed by energy

```
                membrane   anchor   region
                1          10       20          30          40
cyp2a1  ...MLDTGLL LVVILASLSV MLLVSLWQQ. KIRGRLPPGP TPLPFIGNYL
cyp2a4  ...MLTSGLL LVAAVAFLSV LVLMSVWKQR KLSGKLPPGP TPLPFVGNFL
cyp2a5  ...MLTSGLL LVAAVAFLSV LVLMSVWKQR KLSGKLPPGP TPLPFIGNFL
cyp2a6  ...MLASGML LVALLVCLTV MVLMSVWQQR KSKGKLPPGP TPLPFIGNYL
cyp2b1  .......MEPT ILLLLALLVG FLLLLVRGHP KSRGNFPPGP RPLPLLGNLL
cyp2b4  .......MEFS LLLLLAFLAG LLLLLFRGHP KAHGRLPPGP SPLPVLGNLL
cyp2b6  .......MELS VLLFLALLTG LLLLLVQRHP NTHDRLPPGP RPLPLLGNLL
cyp2c2  ........MDL VVVLGLCLSC LLLPSLWKQS HGGGKLPPGP TPFPILGNVL
cyp2c3  ........MDL LIILGICLSC VVLLSLWKKT HGKGKLPPGP TPLPVVGNLL
cyp2c4  ........MDP VAGLVLGLCC LLLLSLWKQN SGRGKLPPGP TPFPIIGNIL
cyp2c5  ........MDP VVVLVLGLCC LLLLSIWKQN SGRGKLPPGP TPFPIIGNIL
cyp2c8  ........MEP FVVLVLCLSF MLLFSLWRQS CRRRKLPPGP TPLPIIGNML
cyp2c9  ........MDS LVVLVLCLSC LLLLSLWRQS SGRGKLPPGP TPLPVIGNIL
cyp2c19 ........MDP FVVLVLCLSC LLLLSIWRQS SGRGKLPPGP TPLPVIGNIL
cyp2d1  MELLNGTGLW SMAIFTVIFI LLVDLMHRRH RWTSRYPPGP VPWPVLGNLL
cyp2d6  ...MGLEALV PLAVIVAIFL LLVDLMHRRQ RWAARYPPGP LPLPGLGNLL
cyp2e1r ....MAVLGI TVALLGWMVI LLFISVWKQI HSSWNLPPGP FPLPIIGNLL
cyp2e1h ....MSALGV TVALLVWAAF LLLVSMWRQV HSSWNLPPGP FPLPIIGNLF
cyp2f1  .......MDSI STAILLLLLA LVCLLLTLSS RDKGKLPPGP RPLSILGNLL
cyp2f2  .......MDGV STAILLLLLA VISLSLTFSS RDKGQLPPGP KPLPILGNLL

                < Ahelix >      |1-1|  | 1-2 |   <   B helix   >
                50          60          70          80          90
cyp2a1  QLNTKDVYSS ITQLSERYGP VFTIHLGPRR VVVLYGYDAV KEALVDQAEE
cyp2a4  QLNTEQMYNS LMKISQRYGP VFTIYLGSRR IVVLCGQEAV KEALVDQAEE
cyp2a5  QLNTEQMYNS LMKISQRYGP VFTIYLGPRR IVVLCGQEAV KEALVDQAEE
cyp2a6  QLNTEQMYNS LMKISERYGP VFTIHLGPRR VVVLCGHDAV REALVDQAEE
cyp2b1  QLDRGGLLNS FMQLREKYGD VFTVHLGPRP VVMLCGTDTI KEALVGQAED
cyp2b4  QMDRKGLLRS FLRLREKYGD VFTVYLGSRP VVVLCGTDAI REALVDQAEA
cyp2b6  QMDRRGLLKS FLRFREKYGD VFTVHLGPRP VVMLCGVEAI REALVDKAEA
cyp2c2  QLDFKDLSKS LTNLSKVYGP VFTVYLGMKP TVVVHGYEAV KEALVDLGHE
cyp2c3  QLETKDINKS LSMLAKEYGS IFTLYFGMKP AVVLYGYEGV IEALIYRGEE
cyp2c4  QIDVKDISKS LTKFSERYGP VFTVYLGMKP TVVLHGYKAV KEALVDLGEE
cyp2c5  QIDAKDISKS LTKFSECYGP VFTVYLGMKP TVVLHGYEAV KEALVDLGEE
cyp2c8  QIDVKDICKS FTNFSKVYGP VFTVYFGMNP IVVFHGYEAV KEALIDNGEE
cyp2c9  QIGIKDISKS LTNLSKVYGP VFTLYFGLKP IVVLHGYEAV KEALIDLGEE
cyp2c19 QIDIKDVSKS LTNLSKIYGP VFTLYFGLER MVVLHGYEVV KEALIDLGEE
cyp2d1  QVDLSNMPYS LYKLQHRYGD VFSLQKGWKP MVIVNRLKAV QEVLVTHGED
cyp2d6  HVDFQNTPYC FDQLRRRFGD VFSLQLAWTP VVVLNGLAAV REALVTHGED
cyp2e1r QLDLKDIPKS FGRLAERFGP VFTVYLGSRR VVVLHGYKAV REMLLNHKNE
cyp2e1h QLELKNIPKS FTRLAQRFGP VFTLYVGSQR MVVMHGYKAV KEALLDYKDE
cyp2f1  LLCSQDMLTS LTKLSKEYGS MYTVHLGPRR VVVLSGYQAV KEALVDQGEE
cyp2f2  QLRSQDLLTS LTKLSKEYGS VFTVYLGSRP VIVLSGYQTV KEALVDKGEE
```

Fig. 1: A multiple sequence alignment of CYP102 and various CYP2 family proteins. The substrate recognition sites (SRS) are shaded and site-directed mutagenesis information is provided by emboldening the relevant residues.

```
                    SRS1                          <  C helix >
            100             110                   120         130
cyp2a1  FSGRGEQATY NTLFKG...Y GVAFSS.GER AKQLRRLSIA TLRDFGVGKR
cyp2a4  FSGRGEQATF DWLFKG...Y GIAFSS.GER AKQLRSFSIA TLRDFGVGKR
cyp2a5  FSGRGEQATF DWLFKG...Y GVVFSS.GER AKQLRRFSIA TLRDFGVGKR
cyp2a6  FSGRGEQATF DWVFKG...Y GVVFSN.GER AKQLRRFSIA TLRDFGVGKR
cyp2b1  FSGRGTIAVI EPIFPE...Y GVIFAN.GER WKALRRFSLA TMRDFGMGKR
cyp2b4  FSGRGKIAVV DPIFQG...Y GVIFAN.GER WRALRRFSLA TMRDFGMGKR
cyp2b6  FSGRGKIAMV DPFFRG...Y GVIFAN.GNR WKVLRRFSVT TMRDFGMGKR
cyp2c2  LSGRSRFLVT AKLNKG...F GVIFSN.GKR WTETRRFSLM TLRNFGMGKR
cyp2c3  FSGRGIFPVF DRVTKG...L GIVFSS.GEK WKETRRFSLT VLRNLGMGKK
cyp2c4  FAGRGHFPIA EKVNKG...L GIVFTN.ANT WKEMRRFSLM TLRNFGMGKR
cyp2c5  FAGTGSVPIL EKVSKG...L GIAFSN.AKT WKEMRRFSLM TLRNFGMGKR
cyp2c8  FSGRGNSPIS QRITKG...L GIISSN.GKR WKEIRRFSLT TLRNFGMGKR
cyp2c9  FSGRGIFPLA ERANRG...F GIVFSN.GKK WKEIRRFSLM TLRNFGMGKR
cyp2c19 FSGRGHFPLA ERANRG...F GIVFSN.GKR WKEIRRFSLM TLRNFGMGKR
cyp2d1  TADRPPVPIF KCLGVKPRSQ GVILASYGPE WREQRRFSVS TLRTFGMGKK
cyp2d6  TADRPPVPIT QILGFGPRSQ GVFLARYGPA WREQRRFSVS TLRNLGLGKK
cyp2e1r FSGRGEIPAF .REFKD...K GIIFNN.GPT WKDTRRFSLT TLRDYGMGKQ
cyp2e1h FSGRGDLPAF .HAERD...R GIIFNN.GPT WKDIRRFSLT TLRNYGMGKQ
cyp2f1  FSGRGDYPAF FNFTKG...N GIAFSS.GDR WKVLRQFSIQ ILRNFGMGKR
cyp2f2  FSGRGAYPVF FNFTRG...N GIAFSD.GER WKILRRFSVQ ILRNFGMGKR
```

```
          <     D helix    >          < D'> < Ehelix >
          140         150         160         170         180
cyp2a1  GVEERILEEA GYLIKMLQGT CGAPIDPTIY LSKTVSNVIS SIVFGERFDY
cyp2a4  GIEERIQEEA GFLIDSFRKT NGAFIDPTFY LSRTVSNVIS SIVFGDRFDY
cyp2a5  GIEERIQEEA GFLIDSFRKT NGAFIDPTFY LSRTVSNVIS SIVFGDRFDY
cyp2a6  GIEERIQEEA GFLIDAHRGT GGANIDPTFF LSRTVSNVIS SIVFGDRFDY
cyp2b1  SVEERIQEEA QCLVEELRKS QGAPLDPTFL FQCITANIIC SIVFGERFDY
cyp2b4  SVEERIQEEA RCLVEELRKS KGALLDNTLL FHSITSNIIC SIVFGKRFDY
cyp2b6  SVEERIQEEA QCLIEELRKS KGALMDPTFL FQSITANIIC SIVFGKRFHY
cyp2c2  SIEERVQEEA HCLVEELRKT NASPCDPTFI LGAAPCNVIC SVIFQNRFDY
cyp2c3  TIEERIQEEA LCLIQALRKT NASPCDPTFL LFCVPCNVIC SVIFQNRFDY
cyp2c4  SIEDRVQEEA RCLVEELRKT NALPCDPTFI LGCAPCNVIC SVILHNRFDY
cyp2c5  SIEDRIQEEA RCLVEELRKT NASPCDPTFI LGCAPCNVIC SVIFHNRFDY
cyp2c8  SIEDRVQEEA HCLVEELRKT KASPCDPTFI LGCAPCNVIC SVVFQKRFDY
cyp2c9  SIEDRVQEEA RCLVEELRKT KASPCDPTFI LGCAPCNVIC SIIFHKRFDY
cyp2c19 SIEDRVQEEA RCLVEELRKT KASPCDPTFI LGCAPCNVIC SIIFQKRFDY
cyp2d1  SLEEWVTKEA GHLCDAFTAQ AGQSINPKAM LNKALCNVIA SLIFARRFEY
cyp2d6  SLEQWVTEEA ACLCAAFANH SGRPFRPNGL LDKAVSNVIA SLTCGRRFEY
cyp2e1r GNEDRIQKEA HFLLEELRKT QGQPFDPTFV IGCTPFNVIA KILFNDRFDY
cyp2e1h GNESRIQREA HFLLEALRKT QGQPFDPTFL IGCAPCNVIA DILFRKHFDY
cyp2f1  SIEERILEEG SFLLADVRKT EGEPFDPTFV LSRSVSNIIC SVLFGSRFDY
cyp2f2  SIEERILEEG SFLLEVLRKM EGKPFDPVFI LSRSVSNIIC SVVFGSRFDY
```

Fig. 1 *cont.*

```
          < F helix SRS2 >     <  F' >          <       SRS3
         190        200        210        220        230
cyp2a1   EDTEFLSLLQ MMGQMNRFAA SPTGQLYDMF HSVMKYLPGP QQQIIKVTQK
cyp2a4   EDKEFLSLLR MMLGSLQFTA TSMGQVYEMF SSVMKHLPGP QQQAFKELQG
cyp2a5   EDKEFLSLLR MMLGSFQFTA TSMGQLYEMF SSVMKHLPGP QQQAFKELQG
cyp2a6   KDKEFLSLLR MMLGIFQFTS TSTGQLYEMF SSVMKHLPGP QQQAFQLLQG
cyp2b1   TDRQFLRLLE LFYRTFSLLS SFSSQVFEFF SGFLKYFPGA HRQISKNLQE
cyp2b4   KDPVFLRLLD LFFQSFSLIS SFSSQVFELF PGFLKHFPGT HRQIYRNLQE
cyp2b6   QDQEFLKMLN LFYQTFSLIS SVFGQLFELF SGFLKYFPGA HRQVYKNLQE
cyp2c2   TDQDFLSLMG KFNENFKILN SPWVQFCNCF PILFDYFPGS HRKAVKNIFY
cyp2c3   DDEKFKTLIK YFHENFELLG TPWIQLYNIF PILGHYLPGS HRQLFKNIDG
cyp2c4   KDEEFLKLME RLNENIRILS SPWLQVYNNF PALLDYFPGI HKTLLKNADY
cyp2c5   KDEEFLKLME SLNENVRILS SPWLQVYNNF PALLDYFPGI HKTLLKNADY
cyp2c8   KDQNFLTLMK RFNENFRILN SPWIQVCNNF PLLIDCFPGT HNKVLKNVAL
cyp2c9   KDQQFLNLME KLNENIKILS SPWIQICNNF SPIIDYFPGT HNKLLKNVAF
cyp2c19  KDQQFLNLME KLNENIRIVS TPWIQICNNF PTIIDYFPGT HNKLLKNLAF
cyp2d1   EDPYLIRMVK LVEESLTEVS GFIPEVLNTF PALLR.IPGL ADKVFQGQKT
cyp2d6   DDPRFLRLLD LAQEGLKEES GFLREVLNAV PVLLH.IPAL AGKVLRFQKA
cyp2e1r  KDKQALRLMS LFNENFYLLS TPWLQVYNNF SNYLQYMPGS HRKVIKNVSE
cyp2e1h  NDEKFLRLMY LFNENFHLLS TPWLQLYNNF PSFLHYLPGS HRKVIKNVAE
cyp2f1   DDERLLTIIR LINDNFQIMS SPWGELYDIF PSLLDWVPGP HQRIFQNFKC
cyp2f2   DDERLLTIIH FINDNFKIMS SPWGEMYNIF PSVLDWIPGP HKRLFRNFGG

            G helix   >          < H  >                <
         240        250        260        270        280
cyp2a1   LEDFMIEKVR QNHSTLDPN. SPRNFIDSFL IRMQEE.KNG NSEFHMKNLV
cyp2a4   LEDFITKKVE HNQRTLDPN. SPRDFIDSFL IRMLEEKKNP NTEFYMKNLV
cyp2a5   LEDFITKKVE HNQRTLDPN. SPRDFIDSFL IRMLEEKKNP NTEFYMKNLV
cyp2a6   LEDFIAKKVE HNQRTLDPN. SPRDFIDSFL IRMQEEEKNP NTEFYLKNLV
cyp2b1   ILDYIGHIVE KHRATLDPS. APRDFIDTYL LRMEKEKSNH HTEFHHENLM
cyp2b4   INTFIGQSVE KHRATLDPS. NPRDFIDVYL LRMEKDKSDP SSEFHHQNLI
cyp2b6   INAYIGHSVE KHRETLDPS. APKDLIDTYL LHMEKEKSNA HSEFSHQNLN
cyp2c2   VKNYITEQIK EHQKSLDIN. NPRDFIDCFL IKMEQEKCNQ QSEFTIENLL
cyp2c3   QIKFILEKVQ EHQESLDSN. NPRDFVDHFL IKMEKEKHKK QSEFTMDNLI
cyp2c4   TKNFIMEKVK EHQKLLDVN. NPRDFIDCFL IKMEKE...N NLEFTLGSLV
cyp2c5   IKNFIMEKVK EHEKLLDVN. NPRDFIDCFL IKMEQE...N NLEFTLESLV
cyp2c8   TRSYIREKVK EHQASLDVN. NPRDFIDCFL IKMEQEKDNQ KSEFNIENLV
cyp2c9   MKSYILEKVK EHQESMDMN. NPQDFIDCFL MKMEKEKHNQ PSEFTIESLE
cyp2c19  MESDILEKVK EHQESMDIN. NPRDFIDCFL IKMEKEKQNQ QNEFTIENLV
cyp2d1   FMALLDNLLA ENRTTWDPAQ PPRNLTDAFL AEVEKAKGNP ESSFNDENLR
cyp2d6   FLTQLDELLT EHRMTWDPAQ PPRDLTEAFL AEMEKAKGNP ESSFNDENLR
cyp2e1r  IKEYTLARVK EHHKSLDPS. CPRDFIDSLL IEMEKDKHST EPLYTLENIA
cyp2e1h  VKEYVSERVK EHHQSLDPN. CPRDLTDCLL VEMEKEKHSA ERLYTMDGIT
cyp2f1   LRDLIAHSVH DHQASLDPR. SPRDFIQCFL TKMAEEKEDP LSHFHMDTLL
cyp2f2   MKDLIARSVR EHQDSLDPN. SPRDFIDCFL TKMAQEKQDF LSHFNMDTLL
```

Fig. 1 *cont.*

```
          SRS4   I helix            > <   J helix    >
            290        300        310        320        330
cyp2a1   MTTLSLFFAG SETVSSTLRY GFLLLMKHPD VEAKVHEEIE QVIGRNRQPQ
cyp2a4   LTTLNLFFAG TETVSTTLRY GFLLLMKYPD IEAKVHEEID RVIGRNRQPK
cyp2a5   LTTLNLFFAG TETVSTTLRY GFLLLMKHPD IEAKVHEEID RVIGRNRQPK
cyp2a6   MTTLNLFIGG TETVSTTLRY GFLLLMKHPE VEAKVHEEID RVIGKNRQPK
cyp2b1   ISLLSLFFAG TETSSTTLRY GFLLMLKYPH VAEKVQKEID QVIGSHRLPT
cyp2b4   LTVLSLFFAG TETTSTTLRY GFLLMLKYPH VTERVQKEIE QVIGSHRPPA
cyp2b6   LNTLSLFFAG TETTSTTLRY GFLLMLKYPH VAERVYREIE QVIGPHRPPE
cyp2c2   TTVSDVFMAG TETTSTTLRY GLLLLMKHPE VIAKVQEEIE RVIGRHRSPC
cyp2c3   TTIWDVFSAG TDTTSNTLKF ALLLLLKHPE ITAKVQEEIE HVIGRHRSPC
cyp2c4   IAVFDLFGAG TETTSTTLRY SLLLLLKHPE VAARVQEEIE RVIGRHRSPC
cyp2c5   IAVSDLFGAG TETTSTTLRY SLLLLLKHPE VAARVQEEIE RVIGRHRSPC
cyp2c8   CTVADLFVAG TETTSTTLRY GLLLLLKHPE VTAKVQEEID HVIGRHRSPC
cyp2c9   NTAVDLFGAG TETTSTTLRY ALLLLLKHPE VTAKVQEEIE RVIGRNRSPC
cyp2c19  ITAADLLGAG TETTSTTLRY ALLLLLKHPE VTAKVQEEIE RVIGRNRSPC
cyp2d1   MVVVDLFTAG MVTTATTLTW ALLLMILYPD VQRRVQQEID EVIGQVRCPE
cyp2d6   IVVADLFSAG MVTTSTTLAW GLLLMILHPD VQRRVQQEID DVIGQVRRPE
cyp2e1r  VTVADMFFAG TETTSTTLRY GLLILLKHPE IEEKLHEEID RVIGPSRMPS
cyp2e1h  VTVADLFFAG TETTSTTLRY GLLILMKYPE IEEKLHEEID RVIGPSRIPA
cyp2f1   MTTHNLLFGG TKTVSTTLHH AFLALMKYPK VQARVQEEID LVVGRARLPA
cyp2f2   MTTHNLLFGG TETVGTTLRH AFLILMKYPK VQARVQEEID RVVGRSRMPT

         < J'>  <     K helix    > SRS5   |1-4|              |1-3|
            340        350        360        370        380
cyp2a1   YEDHMKMPYT QAVINEIQRF SNLAPLGIPR RIIKNTTFRG FFLPKATDVF
cyp2a4   YEDRMKMPYT EAVIHEIQRF ADLIPMGLAR RVTKDTKFRD FLLPKGTEVF
cyp2a5   YEDRMKMPYT EAVIHEIQRF ADMIPMGLAR RVTKDTKFRD FLLPKGTEVF
cyp2a6   FEDRAKMPYM EAVIHEIQRF GDVIPMSLAR RVKKDTKFRD FFLPKGTEVY
cyp2b1   LDDRSKMPYT DAVIHEIQRF SDLVPIGVPH RVTKDTMFRG YLLPKNTEVY
cyp2b4   LDDRAKMPYT DAVIHEIQRL GDLIPFGVPH TVTKDTQFRG YVIPKNTEVF
cyp2b6   LHDRAKMPYT EAVIYEIQRF SDLLPMGVPH IVTQHTSFRG YIIPKDTEVF
cyp2c2   MQDRSRMPYT DATVHEIQRY INLIPNNVPH TTICNLKFRN YLIPKGTDVL
cyp2c3   SQDRSRMPYT DAVMHEIQRY VDLVPTSLPH AVTQDIEFNG YLIPKGTDII
cyp2c4   MQDRSHMPYT DAVIHEIQRF IDLLPTNLPH AVTRDVKFRN YFIPKGTDII
cyp2c5   MQDRSRMPYT DAVIHEIQRF IDLLPTNLPH AVTRDVRFRN YFIPKGTDII
cyp2c8   MQDRSHMPYT DAVVHEIQRY SDLVPTGVPH AVTTDTKFRN YLIPKGTTIM
cyp2c9   MQDRSHMPYT DAVVHEVQRY IDLLPTSLPH AVTCDIKFRN YLIPKGTTIL
cyp2c19  MQDRGHMPYT DAVVHEVQRY IDLIPTSLPH AVTCDVKFRN YLIPKGTTIL
cyp2d1   MTDQAHMPYT NAVIHEVQRF GDIAPLNLPR FTSCDIEVQD FVIPKGTTLI
cyp2d6   MGDQAHMPYT TAVIHEVQRF GDIVPLGVTH MTSRDIEVQG FRIPKGTTLI
cyp2e1r  VRDRVQMPYM DAVVHEIQRF IDLVPSNLPH EATRDTTFQG YVIPKGTVVI
cyp2e1h  IKDRQEMPYM DAVVHEIQRF ITLVPSNLPH EATRDTIFRG YLIPKGTVVV
cyp2f1   LKDRAAMPYT DAVIHEVQRF ADIIPMNLPH RVTRDTAFRG FLIPKGTDVI
cyp2f2   LEDRTSMPYT DAVIHEVQRF ADVIPMNLPH RVTRDTPFRG FLIPKGTDVI
```

Fig. 1 ***cont.***

```
          < K'>  <K"">                                          <
            390        400        410        420        430
cyp2a1   PILGSLMTDP KFFPSPKDFD PQNFLDDKGQ LKKNAAFLPF STGKRFCLGD
cyp2a4   PMLGSVLKDP KFFSNPKDFN PKHFLDDKGQ FKKSDAFVPF SIGKRYCFGE
cyp2a5   PMLGSVLKDP KFFSNPKDFN PKHFLDDKGQ FKKNDAFVPF SIGKRYCFGE
cyp2a6   PMLGSVLRDP SFFSNPQDFN PQHFLNEKGQ FKKSDAFVPF SIGKRNCFGE
cyp2b1   PILSSALHDP QYFDHPDSFN PEHFLDANGA LKKSEAFMPF STGKRICLGE
cyp2b4   PVLSSALHDP RYFETPNTFN PGHFLDANGA LKRNEGFMPF SLGKRICLGE
cyp2b6   LILSTALHDP HYFEKPDAFN PDHFLDANGA LKKTEAFIPF SLGKRICLGE
cyp2c2   TSLSSVLHDD KEFFNPDRFD PGHFLDASGN FRKSDYFMPF STGKRVCVGE
cyp2c3   PSLTSVLYDD KEFPNPEKFD PGHFLDESGN FKKSDYFMPF STGKRACVGE
cyp2c4   TSLTSVLHDE KAFPNPKVFD PGHFLDESGN FKKSDYFMPF SAGKRMCVGE
cyp2c5   TSLTSVLHDE KAFPNPKVFD PGHFLDESGN FKKSDYFMPF SAGKRMCVGE
cyp2c8   ALLTSVLHDD KEFPNPNIFD PGHFLDKNGN FKKSDYFMPF SAGKRICAGE
cyp2c9   ISLTSVLHDN KEFPNPEMFD PHHFLDEGGN FKKSKYFMPF SAGKRICVGE
cyp2c19  TSLTSVLHDN KEFPNPEMFD PRHFLDEGGN FKKSNYFMPF SAGKRICVGE
cyp2d1   INLSSVLKDE TVWEKPHRFH PEHFLDAQGN FVKHEAFMPF SAGRRACLGE
cyp2d6   TNLSSVLKDE AVWEKPFRFH PEHFLDAQGH FVKPEAFLPF SAGRRACLGE
cyp2e1r  PTLDSLLYDK QEFPDPEKFK PEHFLNEEGK FKYSDYFKPF SAGKRVCVGE
cyp2e1h  PTLDSVLYDN QEFPDPEKFK PEHFLNENGK FKYSDYFKPF STGKRVCAGE
cyp2f1   TLLNTVHYDP SQFLTPQEFN PEHFLDANQS FKKSPAFMPF SAGRRLCLGE
cyp2f2   TLLNTVHYDS DQFKTPQEFN PEHFLDDNHS FKKSPAFMPF SAGRRLCLGE

           L helix     > |3-:.|              SRS6          |3-2|
            440        450        460        470        480
cyp2a1   GLAKMELFLL LTTILQNFRF KFPMKLEDIN ESPKPLGFTRIIPKYTMSFMPI
cyp2a4   GLARMELFLF LTNIMQNFHF KSTQAPQDID VSPRLVGFVTIPPTYTMSFLSR
cyp2a5   GLARMELFLF LTNIMQNFHF KSTQAPQDID VSPRLVGFATIPPTYTMSFLSR
cyp2a6   GLARMELFLF FTTVMQNFRL KSSQSPKDID VSPKHVGFATIPRNYTMSFLPR
cyp2b1   GIARNELFLF FTTILQNFSV SSHLAPKDID LTPKESGIGKIPPTYQICFSAR
cyp2b4   GIARTELFLF FTTILQNFSI ASPVPPEDID LTPRESGVGNVPPSYQIRFLAR
cyp2b6   GIARAELFLF FTTILQNFSM ASPVAPEDID LTPQECGVGKIPPTYQIRFLPR
cyp2c2   ALARMELFLF LTAILQNFTP KPLVNPNNVD ENPFSSGIVRVPPLYRVSFIPV
cyp2c3   GLARMELFLL LTTILQHFTL KPLVDPKDID PTPVENGFVSVPPSYELCFVPV
cyp2c4   GLARMELFLF LTSILQNFKL QSLVEPKDLD ITAVVNGFVSVPPSYQLCFIPI
cyp2c5   GLARMELFLF LTSILQNFKL QSLVEPKDLD ITAVVNGFVSVPPSYQLCFIPI
cyp2c8   GLARMELFLF LTTILQNFNL KSVDDLKNLN TTAVTKGIVSLPPSYQICFPIV
cyp2c9   ALAGMELFLF LTSILQNFNL KSLVDPKNLD TTPVVNGFASVPPFYQLCFIPV
cyp2c19  GLARMELFLF LTFILQNFNL KSLIDPKDLD TTPVVNGFASVPPFYQLCFIPV
cyp2d1   PLARMELFLF FTCLLQRFSF .SVPVGQPRP STHGFFAFPVAPLPYQLCAVVREQGL
cyp2d6   PLARMELFLF FTSLLQHFSF .SVPTGQPRP SHHGVTAFLVSPSPYELCAVPR
cyp2e1r  GLARMELFLL LSAILQHFNL KPLVDPEDID LRNITVGFGRVPPRYKLCVIPRS
cyp2e1h  GLARMELFLL LCAILQHFNL KPLVDPKDID LSPIHIGFGCIPPRYKLCVIPRS
cyp2f1   LLARMELFLY LTAILQSFSL QPLGAPEDID LTPLSSGLGNLPRPFQLCLRPR
cyp2f2   PLARMELFIY FTSILQNFTL QPLVDPEDID LTPLSSGLGNLPRPFQLCMHIR
```

Fig. 1 *cont.*

minimization of the raw structure using molecular mechanics. After production of a low energy geometry of CYP2F2, the model was probed using selective substrates of the enzyme, including naphthalene, 3-methylindole, 4-nitrophenol, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and ipomeanol. Modelling was assisted by the presence of bound substrate (palmitoleic acid) in the CYP102 crystal structure, and interactive docking procedures (e.g. Dock and FlexiDock) were employed at this stage of the investigation. Following the generation of a structural template using an overlay, these substrate molecules, the candidate compounds 2-isopropylnaphthalene and 2-(2-hexadecyl)naphthalene, were employed as substrate probes of the CYP2F2 model to ascertain their likely selectivity for the enzyme and subsequent route of metabolism on the basis of analogy with known substrates of CYP2F enzymes. All molecular modelling procedures (including homology modelling, energy minimization and substrate docking) were carried out using the Sybyl software package (Tripos Associates, St. Louis, MO) implemented on a Silicon Graphics Indigo^2 IMPACT 10000 graphics workstation operating under UNIX.

RESULTS AND DISCUSSION

Figure 2 indicates how the selective substrate naphthalene is able to fit within the putative active site of CYP2F2 such that 1,2-epoxidation can occur. This figure also shows how the putative active site region of CYP2F2 contains residues which interact with the bound substrate, naphthalene, thus orientating it for 1,2-epoxidation. The substrate molecule is held in position via a number of π-π stacking interactions with aromatic amino acid residues within the haem environment. The spatial arrangement of these aromatic side-chains cooperatively assist in orientating the naphthalene molecule such that the 1,2-bond lies directly about the haem iron for oxygenation to occur at that site. Amino acid residues are numbered according to the alignment with CYP102 as presented in Figure 1, which also contains information on substrate recognition site (SRS) regions. In particular, it is the combination of π-π stacking interactions with two phenylalanine residues, Phe181 and Phe78 in the alignment, which orientates the substrate for 1R,2S-oxidation in accordance with known experimental observations /4,6-8/.

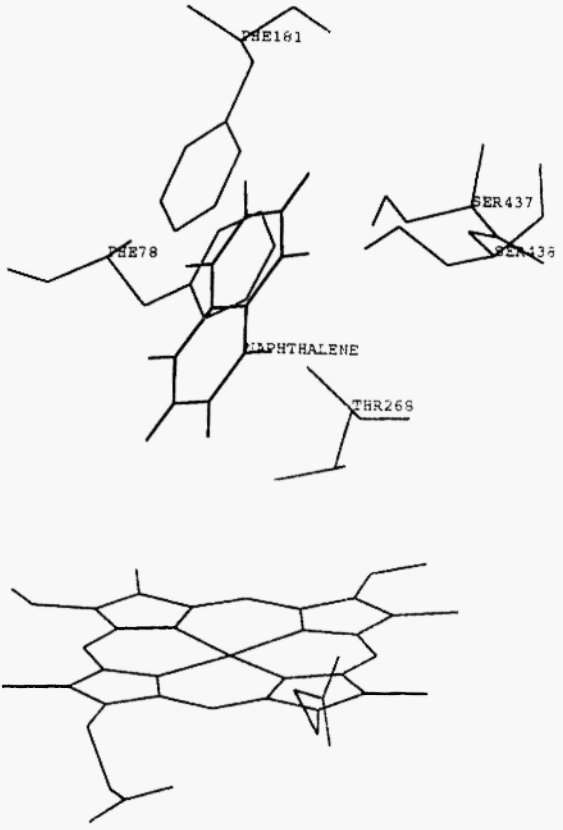

Fig. 2: Naphthalene fitted within the putative active site of CYP2F2. Amino acid residues are numbered according to the alignment with CYP102 shown in Figure 1.

Another selective substrate for CYP2F2 is 3-methylindole, and Figure 3 indicates a likely mode of interaction within the putative active site. In this case, it is a combination of hydrogen bonding to Ser438 (although this only occurs upon entry to the active site) and π-π stacking with Phe181 which orientates the substrate for oxygenation at the 3-methyl position, once again this being in line with the observed findings.

Hydrogen bonding at the same site (i.e. Ser438 in the alignment) also occurs for NNK, 4-nitrophenol and ipomeanol, as shown in

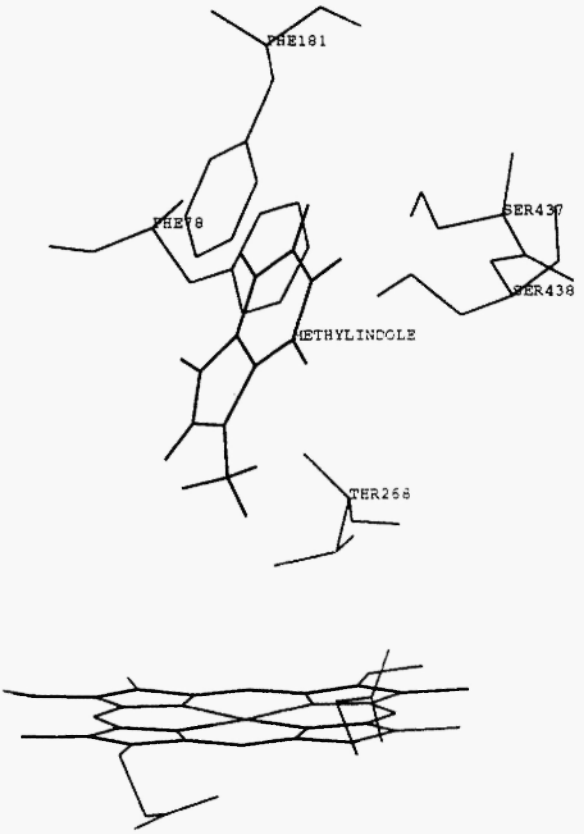

Fig. 3: The active site region of CYP2F2 containing the docked substrate 3-methylindole. Residues are numbered according to the CYP102 alignment (Figure 1).

Figure 4 (where Ser437 is also involved in the latter example) to give the expected metabolites in each case, and this is due to specific active site orientations brought about by amino acid residues in contact with the relevant substrates. Furthermore, the selective inhibitor 5-phenyl-1-pentyne is able to occupy the CYP2F2 site by π-π stacking with Phe181 and Phe78 such that the alkyne group is situated directly above the haem iron (see Fig. 4). This figure includes an overlay of the substrates mentioned above, including naphthalene, 3-methylindole,

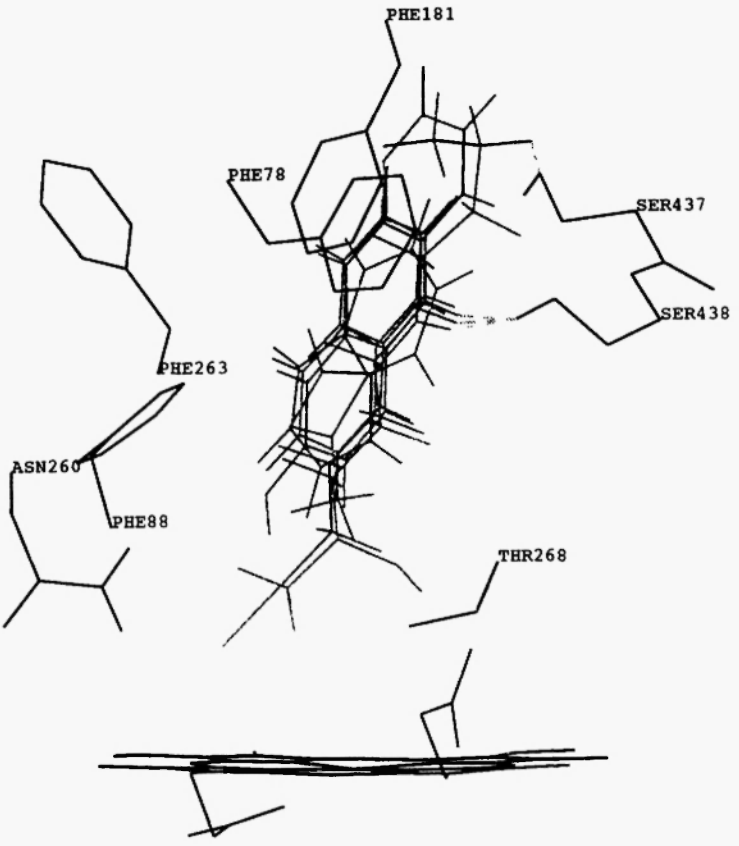

Fig. 4: An overlay of naphthalene, 3-methylindole, 2-isopropylnaphthalene, 5-phenyl-1-pentyne, 4-nitrophenol, NNK and ipomeanol orientated within the putative active site of CYP2F2 for metabolism at the experimentally observed positions. Hydrogen-bonded interactions are displayed as dashed lines, and residues are numbered according to the alignment with CYP102.

NNK, 4-nitrophenol, ipomeanol and 2-isopropyl-naphthalene, together with the inhibitor 5-phenyl-1-pentyne.

However, in the case of 2-isopropylnaphthalene, it is unlikely that ring epoxidation occurs and, instead, oxidation of the isopropyl group is more favourable, as shown in the orientation presented in Figure 5,

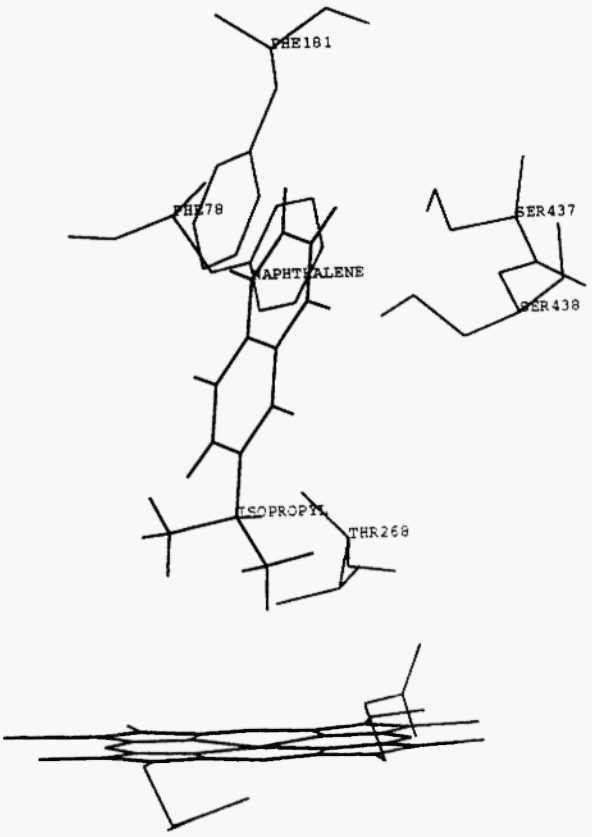

Fig. 5: A potential CYP2F2 substrate, 2-isopropylnaphthalene, fitted within the putative active site of this enzyme.

and this is also consistent with information from metabolism studies in the rat /12-14/ in which side-chain hydroxylated products are observed. For 2-(2-hexadecyl)naphthalene, the situation is less clear but the presence of the bulky alkyl group would appear to preclude interaction with CYP2F2 and, moreover, the log P value is considerably higher than those of other substrates (see Table 1). Consequently, the model presented in this study indicates that this compound is probably not metaboliszed via CYP2F2 involvement although this possibility cannot be entirely ruled out.

TABLE 1: CYP2F substrates [a]

	Compound	log P	pKa	log $D_{7.4}$	M_r	log M_r	K_m (μM)	ΔG_{bind}
1.	Ipomeanol	1.30 [c]	neutral	1.30	168.21	2.2259	N/A	N/A
2.	Toluene	2.73	neutral	2.73	92.14	1.9644	N/A	N/A
3.	Naphthalene	3.37	neutral	3.37	128.17	2.1078	18[d]	-6.7302
4.	3-Methylindole	2.60	neutral	2.60	131.18	2.1179	340[e]	-4.9200
5.	4-Nitrophenol	1.91	7.14	1.91	139.12	2.1434	N/A	N/A
6.	2-Naphthylamine	2.40	4.58 [b]	2.26	143.19	2.1559	N/A	N/A
7.	1-Nitronaphthalene	3.19	neutral	3.19	173.17	2.2385	N/A	N/A
8.	7-Ethoxycoumarin	2.30	neutral	2.30	190.21	2.2792	N/A	N/A
9.	7-Propoxycoumarin	2.86	neutral	2.86	204.24	2.3101	N/A	N/A
10.	7-Ethoxyresorufin	3.61 [c]	9.85 [b]	1.19	241.26	2.3825	N/A	N/A
11.	7-Pentoxyresorufin	5.14 [c]	10.06 [b]	2.50	283.35	2.4523	N/A	N/A
12.	7-Benzyloxyresorufin	4.75 [c]	10.22 [b]	1.98	303.33	2.4819	N/A	N/A

[a] Chemicals are either CYP2F1 (human) or CYP2F2 (mouse) substrates or both.
[b] Basic (the effect of basicity lowers the range of ionization - corrected log P value), i.e. log $D_{7.4}$, indicating a relatively narrow window of substrate lipophilicity).
[c] Calculated value (Pallas Software, CompuDrug Limited, Budapest). [d] Value quoted for CYP2F1 enzyme which is the human orthologue.
[e] Value quoted for CYP2F3 enzyme which is a goat ortholog. N/A = data not available.
$\Delta G_{bind} = RT\ln K_m$ where K_m is the apparent Michaelis constant.
The pK_a values for alkoxyresorufins were calculated using the Pallas Software.
N/K is also a substrate of this enzyme subfamily.
References: /1,2,5,6,8,9,16/.

TABLE 2

Estimation of enzyme-substrate binding energies (kcal.mol^{-1})

Substrate	Naphthalene	3-Methylindole
log P_{expt}	3.37	1.91
ΔG_{part}	-4.780	-2.709
$\Delta G_{\pi-\pi}$	-2.70 (3π-π interactions)	-0.9 (1 π-π interaction)
ΔG_{hbond}	0	-2.0 (1 hydrogen bond)
ΔG_{rotors}	0	+0.6 (1 rotatable bond)
ΔG^{calc}_{bind}	-7.480	-5.009
K_m (μM)	4.7	340 #
ΔG^{expt}_{bind}	-7.557	-4.920

$\log P_{expt}$ = experimental log P value, where P is the octanol/water partition coefficient.

ΔG^{calc}_{bind} = $\Delta G_{part} + \Delta G_{\pi-\pi} + \Delta G_{hbond} + \Delta G_{rotors}$. The use of this relation and the average values of the following components has been reported recently /15/.

ΔG_{part} = RTlnP where P is the octanol/water partition coefficient, R is the gas constant and T is the absolute temperature (310°K).

$\Delta G_{\pi-\pi}$ = -0.9 kcal.mol^{-1} for each π-π stacking interaction.

ΔG_{hbond} = -2.0 kcal.mol^{-1} for each hydrogen bond interaction.

ΔG_{rotors} = +0.6 kcal.mol^{-1} for each rotatable bond in the molecule.

Data for CYP2F3, the goat orthologue of CYP2F2.

References for K_m data: /6,8/.

From the published K_m values of naphthalene and 3-methylindole, the binding energy (ΔG_{bind}) can be estimated* using the relationship $\Delta G = RTlnK$ (see Table 2) and it is possible to calculate values for binding affinity from the modelled interactions (see Figs. 2 and 3)

* Assuming that the dissociation constant, KD, is approximately equal to Km, the Michaelis constant for the reaction catalyzed by CYP2F.

based on the average energies for π-π stacking and hydrogen-bonded interactions, together with the partitioning free energy (calculated from experimental log P values) as shown in Table 2. The data presented in Table 2 show that it is possible to obtain good estimates of the experimental binding affinities for these two substrates via a linear combination of the factors contributing to their overall ΔG_{bind}. It is also interesting that this relatively straightforward procedure gives reasonably good results for the calculation of substrate binding affinity, and similar findings have been observed for other CYP2 family substrates /15/.

CONCLUSIONS

The results of homology modelling of the mouse P450, CYP2F2, are consistent with substrate metabolism data in showing that selective substrates are able to fit the putative active site for metabolism at the experimentally observed positions. Orientation for metabolism is achieved via complementary contacts with a relatively small number of key amino acid residues which enter into π-π stacking and/or hydrogen-bonded interactions with typical CYP2F2 substrates, such as naphthalene, ipomeanol, NNK, 3-methylindole and 3-nitrophenol. The selective inhibitor, 5-phenyl-1-pentyne, also fits within the active site such that the alkynic group is likely to become activated and subsequently bind to the haem moiety (i.e. act as a suicide substrate). Alkylnaphthalenes, such as 2-isopropylnaphthalene, may also interact with CYP2F2 but will probably become metabolized at the alkyl group rather than via epoxidation of the aromatic ring system. However, 2-(2-hexadecyl)naphthalene is unlikely to be a CYP2F2 substrate due to the bulky aliphatic side-chain. Consequently, homology modelling of CYP2F2 enables one to evaluate likely substrate selectivity and route of metabolism for this subfamily.

ACKNOWLEDGEMENTS

The financial support of GlaxoSmithKline Research & Development Ltd., Merck, Sharp & Dohme Ltd., the University of Surrey Foundation Fund and ExxonMobil is gratefully acknowledged by one of us (DFVL).

REFERENCES

1. Rendic S, Di Carlo FJ. Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers and inhibitors. Drug Metab Rev 1997; 29: 413-580.
2. Guengerich FP. Human cytochrome P405 enzymes. In: Ortiz de Montellano PR, ed. Cytochrome P450. New York: Plenum Press, 1995; 473-525.
3. Nhamburo PT, Kimura S, McBride OW, Kozak CA, Gelboin HV, Gonzalez FJ. The human CYP2F gene subfamily: identification of a cDNA encoding a new cytochrome P450, cDNA-directed expression and chromosome mapping. Biochemistry 1990; 29: 5491-5499.
4. Ritter JK, Owens IS, Negishi M, Nagata K, Sheen YY, Gillette JR, Sasame HA. Mouse pulmonary cytochrome P450 naphthalene hydroxylase: cDNA cloning, sequence and expression in *Saccharomyces cerevisiae*, Biochemistry 1991; 30: 11430-11437.
5. Gonzalez FJ, Gelboin HV. Role of human cytochromes P450 in the metabolic activation of chemical carcinogens and toxins. Drug Metab Rev 1994; 26: 165-183.
6. Wang H, Lanza DL, Yost GS. Cloning and expression of CYP2F3, a cytochrome P450 that bioactivates the selective pneumotoxins 3-methylindole and naphthalene. Arch Biochem Biophys 1998; 349: 329-340.
7. Lanza DL, Code E, Crespi CL, Gonzalez FJ, Yost GS. Specific dehydrogenation of 3-methylindole and epoxidation of naphthalene by recombinant human CYP2F1 expressed in lymphoblastoid cells. Drug Metab Dispos 1999; 27: 798-803.
8. Schultz MA, Choudary PV, Buckpitt AR. Role of murine cytochrome P450 2F2 in metabolic activation of naphthalene and metabolism of other xenobiotics. J Pharmacol Exp Ther 1999; 290: 281-288.
9. Buckpitt AR, Cruikshank MK. Biochemical function of the respiratory tract: metabolism of xenobiotics. In: Roth RA, ed. Comprehensive Toxicology, Vol 8. Amsterdam: Elsevier, 1997; 159-186.
10. Carlson GP. Effects of inducers and inhibitors on the microsomal metabolism of styrene to styrene oxide in mice. J Toxicol Environ Health 1997; 51: 477-488.
11. Powley MW, Carlson GP. Cytochromes P450 involved with benzene metabolism in hepatic and pulmonary microsomes. J Biochem Mol Toxicol 2000; 14: 303-309.
12. Kojima S, Maruyama K, Babsaki T. Identification of urinary metabolites of 20-isopropylnaphthalene in rats. Drug Metab Dispos 1980; 8: 463-466.
13. Kojima S, Babsaki T, Kiyozumi M, Nakagawa M. Distribution and excretion of monoisopropylnaphthalene in rats. Bull Environ Contam Toxicol 1981; 26: 626-633.
14. Kojima S, Honda T, Babasaki T, Kiyozuim M, Nakagawa M. Identification and determination of urinary and biliary metabolites of 2-isopropylnaphthalene in rats. Eisei Kagaku 1984; 30: 91-95.

15. Lewis DFV. Homology modelling of human CYP2 family enzymes based on the CYP2C5 crystal structure. Xenobiotica 2002; 32: 305-323.
16. Hansch C, Leo A. Substituent Constants for Correlation Analysis in Chemistry and Biology. New York: Wiley, 1978.