Impact of Incorporating the 2C5 Crystal Structure into Comparative Models of Cytochrome P450 2D6

Stewart B. Kirton,[1] **Carol A. Kemp,**[2] **Nicholas P. Tomkinson,**[3] **Steven St.-Gallay,**[3] **and Michael J. Sutcliffe**[1,2*]

[1]*Department of Chemistry, University of Leicester, Leicester, United Kingdom*
[2]*Department of Biochemistry, University of Leicester, Leicester, United Kingdom*
[3]*Department of Physical and Metabolic Sciences, AstraZeneca R&D Charnwood, Loughborough, United Kingdom*

ABSTRACT Cytochrome P450 2D6 (CYP2D6) metabolizes approximately one third of the drugs in current clinical use. To gain insight into its structure and function, we have produced four different sets of comparative models of 2D6: one based on the structures of P450s from four different microorganisms (P450 terp, P450 eryF, P450 cam, and P450 BM3), another on the only mammalian P450 (2C5) structure available, and the other two based on alternative amino acid sequence alignments of 2D6 with all five of these structures. Principal component analysis suggests that inclusion of the 2C5 crystal structure has a profound effect on the modeling process, altering the general topology of the active site, and that the models produced differ significantly from all of the templates. The four models of 2D6 were also used in conjunction with molecular docking to produce complexes with the substrates codeine and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP); this identified Glu 216 [in the F-helix; substrate recognition site (SRS) 2] as a key determinant in the binding of the basic moiety of the substrate. Our studies suggest that both Asp 301 and Glu 216 are required for metabolism of basic substrates. Furthermore, they suggest that Asp 301 (I-helix, SRS-4), a residue thought from mutagenesis studies to bind directly to the basic moiety of substrates, may play a key role in positioning the B′-C loop (SRS-1) and that the loss of activity on mutating Asp 301 may therefore be the result of an indirect effect (movement of the B′-C loop) on replacing this residue. Proteins 2002;49:216–231.
© 2002 Wiley-Liss, Inc.

Key words: molecular modeling; model validation; substrate specificity; binding determinants; principal component analysis

INTRODUCTION

The superfamily of enzymes known as the cytochromes P450 (P450s) comprises several families that are ubiquitous in animals, plants, and microorganisms. These families can be initially subdivided into two classes, based on the nature of their redox partners.[1] Class I P450s, found predominantly in the mitochondrial membranes of eukaryotes and the majority of bacteria, require a flavin adenine mononucleotide (FAD)-containing reductase and an iron-sulphur-containing protein, such as a ferredoxin, before they can become catalytically active. In contrast, class II P450s, usually found in animals, are bound to the endoplasmic reticulum of cells. They depend on interaction with a reductase containing both a FAD and a flavin mononucleotide (FMN) before they are activated. The function of P450s is in general the same across all the families—the activation of molecular dioxygen (O_2) and subsequent insertion of a single atom of molecular oxygen into an organic molecule[2] (a small number of P450s are not O_2 reductases, e.g., allene oxide synthase and P450nor; nevertheless their structures are similar).

Cytochrome P450 2D6 belongs to family 2 of the P450s—this is concerned with the detoxification, usually in animals, of pharmaceuticals, phytoalexins, and a wide-range of other exogenous and endogenous substances.[3,4] It is believed that at least 30% of the drugs in current clinical use are metabolized preferentially by the hepatic 2D6 isoform.[5] 2D6 also gives rise to the defect in man known as the debrisoquine/sparteine polymorphism.[6,7] Inheritance of the "poor metabolizer" phenotype has been linked with an increased susceptibility to ailments such as Parkinson's disease[7] and certain cancers.[8] Poor metabolizers may not be able to bioactivate a parent drug to its therapeutically active metabolites,[9] and more than 20 compounds have been identified that show impaired oxidation for affected individuals. As medicines they are diverse and include β-adrenergic blocking agents, tricyclic antidepressants, antiarrhythmics, and analgesics,[10–13] but a characteristic common to the vast majority of 2D6 substrates is the presence of at least one basic nitrogen atom. Many models of the active site of 2D6 have postulated the involvement of a carboxylate group in the protein forming a salt bridge with this basic nitrogen[9,14–16]—this has been proposed to

be Asp 301, a residue in the I-helix [substrate recognition site (SRS) 4], both by modeling[16] and by mutagenesis.[17]

A number of structural models of 2D6 have been produced previously using a variety of methods (e.g., Refs. 18–20). The work we present here uses comparative modeling to extend our previous work[19,21.22] and additionally uses principal component analysis (PCA) and molecular docking to explore the proposed salt bridge between the basic nitrogen in the substrate and a carboxylate group in the active site—investigating a number of alternative structural models of 2D6.

METHODS

Comparative Modeling

Four different sets of models of 2D6 have been produced: one based on the structures of P450s from four different microorganisms (P450 terp, P450 eryF, P450 cam, and P450 BM3) used to produce models previously in the absence of a mammalian template, another on the only mammalian P450 (rabbit 2C5) structure available alone, and the other two based on alternative amino acid sequence alignments of 2D6 with this 2C5 structure.

Four-bacterial template approach

Proteins of known three-dimensional (3D) structure homologous to 2D6 (i.e., suitable structural templates) were identified by scanning the amino acid sequence of human 2D6 against the sequences of those structures in the Protein Data Bank[23] using PSI-BLAST.[24] At that time, several years ago, this resulted in the P450s cam (PDB accession code 2cpp[25]), terp (1cpt[26]), BM3 (2hpd[1]), and eryF (1oxa)[27] being identified as suitable templates. These templates were used to generate comparative models by aligning them structurally to produce a profile using Malign3D within Modeller,[28] which was then aligned against a second profile produced by aligning the amino acid sequences of the 2D subfamily using CLUSTALX.[29]

Obtaining a subfamily alignment serves two purposes. First, it highlights any regions across the subfamilies that are highly conserved. Second, appending a subfamily alignment to a profile as opposed to a appending a single amino acid sequence reduces the likelihood of conserved residues in the subfamily being incorrectly aligned with differing, yet still conserved residues in the templates. The resultant alignment was then checked to ensure that (i) all of the secondary structural elements have a minimum number of insertions/deletions in them, (ii) any residues crucial to catalytic activity, such as the proximal cysteinate in the fifth coordination site of the haem,[30] were conserved, and (iii) any residues that have been shown to play an important role in the binding of substrates, e.g., Asp 301 in 2D6,[9,19,21] were conserved. Correlation between the secondary structural elements observed in a template [obtained from the Database of Secondary Structure in Proteins[31] (DSSP)] and the secondary structure predicted for the 2D6 (using PSIPRED[32]) was used to modify the sequence alignment. Once an acceptable alignment had been produced (denoted alignment 1; Fig. 1), an ensemble of 15 models of 2D6 were generated using Modeller.[28]

Single template approach

The sequence of 2D6 was aligned against the sequence of most homologous template—P450 2C5 (1dt6[33]). To achieve this a sequence alignment of the 2D subfamily was aligned against a sequence alignment of the subfamily of the template (2C) using CLUSTALX. This alignment was analysed in an identical manner to that for alignment 1, and the resultant alignment (denoted alignment 2; Fig. 2) used to generate fifteen models using Modeller.

Multiple template approach using FSSP

An initial alignment profile of the sequences of the five structural templates [1dt6 (2C5), 2hpd (BM3), 2cpp (cam), 1oxa (eryF), 1cpt (terp)] was generated using the FSSP (Fold classification and Secondary Structure alignment of Proteins) database.[34] The amino acid sequence of 2D6 was aligned against this by aligning the template-derived profile against a second profile comprising the sequence alignment of the 2D subfamily using CLUSTALX. This alignment was then checked in a manner identical to that used for alignment 1, and the resulting alignment (denoted alignment3; Fig. 3) used to generate fifteen models using Modeller.

Multiple template approach using MALIGN3D

An initial alignment profile of the sequences of the five structural templates (1dt6, 2hpd, 2cpp, 1oxa, 1cpt) was generated using the MALIGN3D routine within Modeller to structurally align the templates. The amino acid sequence of 2D6 was aligned against this by aligning the template-derived profile against a second profile comprising the sequence alignment of the 2D subfamily using the profile alignment option within CLUSTALX. This alignment was then checked in a manner identical to that used for alignment 1, and the resulting alignment (denoted alignment 4; Fig. 4) used to generate 15 models using Modeller.

Model Validation

The lowest energy models (as determined by Modeller) from each ensemble of 15 models (denoted model 1, model 2, model 3, and model 4 for the lowest energy model derived from alignment 1, alignment 2, alignment 3, and alignment 4, respectively) were assessed using a number of validation methods. Two types of validation were used: stereochemical quality (PROCHECK[35] for backbone and side-chain conformation), and side-chain environment (Errat[36] and Verify 3D[37]). For both the stereochemical and environmental checks, the values returned for the templates were used as a baseline against which the models were compared. Also, it has been shown that the RMSD between the mainchain atoms in (i) the template with the greatest homology to the target and (ii) the model is another useful method of validation.[38] Backbone RMSDs were therefore used as an additional validation check; these were calculated with reference to P450 2C5, the most homologous template.

Principal Component Analysis

PCA was used to characterize the active sites of each of the lowest energy models. Models 1–4 and the 2C5 crystal

```
                                         A'                              A
1oxa  ----A----TV--PD---L-E--SDSFH-VD----W--------------YSTYAELRE     25
2cpp  --NLAPLPPHVPE----HLVF---D-F---DMY-N--PSN-LSAGV-----QEAWAVLQ-     46
1cpt  --MDA-RA-TIP--------EHIARTVILP-Q--G---YAD-D--E---VIYPAFKWLR-     35
2hpd  ----TIKEMPQPK-TFGELK---NLPLL--NTDK--------------PVQALMKIAD      34
2D6   RQRWAARYPPGP------LPLPGLGNLLHVDFQN--------------TPYCFDQLRR      63

            β1-1              β1-2       B      β1-5
1oxa  T---------APVTPVRFL--G--QDAWLVTGYDEAKAALSDL-RLSSDPKKKYPGVEVE    71
2cpp  ----E--SNVP--DLVWT-RCN--GGHWIATRGQLIREAYEDYRHFSSEC---------    85
1cpt  ------D-EQPLAMAHI--EGY-DPMWIATKHADVMQIGKQPGLFSNA--E------GS    76
2hpd  EL-------GEIFKFEA----PGRVTRYLSSQRLIKEACDE---SRFDKN---------    70
2D6   R--------FGDVFSLQLA----WTPVVVLNGLAAVREALVTHGEDTADR---------   101
         B'                                                 C
1oxa  -FPAYLGF--PEDVRNYFA--------TNMG-TSDPP--T-HTRLRKLVS------QEF-   109
2cpp  PFI-------PREAGEAY--------DFIPT-SMDPP--E-QRQFRALAN-------QVV   119
1cpt  EIL-YD-QNNEAFMRSISG-GCPHVI-DSLT-SMDPP--T-HTAYRGLTL------NWF   121
2hpd  LS-------QALKFVRDF-----AG--DGLFTSWTH-E-KNWKKAHNILLP--SFSQQAM  112
2D6   P---------PVPITQILG---FGPRSQGVFLARYGP--AWREQRRFSVS-------TLR  140
        C'        D                β3-1  E*            E
1oxa  TV-RRVEAMRPRVEQITAELLDEVGD--S-GVVDIVDRFAHPLPIKVICELLGV---DE-  161
2cpp  GM-PVVDKLENRIQELACSLIESLRP--Q-GQCNFTEDYAEPFPIRIFMLLAGL---PE-  171
1cpt  QPAS-IRKLEENIRRIAQASVQRLLDF-D-GECDFMTDCALYYPLHVVMTALGVPEDDE-  177
2hpd  KGY------HAMMVDIAVQLVQKWERLNADEHIEVPE-DMTRLTLDTIGLCGFNYR--FN  163
2D6   NLGLGKKSLEQWVTEEAACLCAAFAN--HSGRPFRPNGLLDKAVSNVIASLTCGRRFEYD  198
                    F                               G
1oxa  -------AARGAF-GRWSSE-ILVMD--PER----------AEQRGQAAREVVNFILDLV  200
2cpp  ----EDIPHL-KYLTDQ-MTRPD-G-SMTF-----------AEAKEALYDYLIPIIEQ--  210
1cpt  ----PLMLKLTQDFFGV--------------------EAARRFHETIATFYDYFNGFIVD  229
2hpd  SFYRDQPHPPITSMVRAL-DEA--MNKLQR--ANPDDP-AYDENKRQFQEDIKVMNDLVD  217
2D6   DPRFLRLLDLAQEGLKE--ESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLT  256
                             H    β5-1       β5-2
1oxa  ERRRTE-P----------G--D-DLLSALISVQDDDD-G---RLSADELTSIALVLLLAG  242
2cpp  ---RRQKP----------G--T-DAISIVANGQV--NGR---PITSDEAKRMCGLLLVGG  249
1cpt  RRS---------------CP-KDDVMSLLANSKL---DGN--YIDDKYINAYYVAIATAG  268
2hpd  KIIADRKAS--------GEQSDDLLTHMLNGKDPETGE---PLDDENIRYQIITFLIAG  265
2D6   EHRMTWDP---------AQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAG  306
           I              J                  J'         K
1oxa  FEASVSLIGIGTYLLLTHPDQLALVRAD-----------------PSALPNAVEEILRY  284
2cpp  LDTVVNPLSFSMEFLAKSPEHRQELIER-----------------PERIPAACEELLRR  291
1cpt  HDTTSSSSGGAIIGLSRNPEQLALAKSD-----------------PALIPRLVDEAVRW  310
2hpd  HETTSGLLSFALYFLVKNPHVLQKAAEEAARVL-VDPVPSYKQVKQLKYVGMVLNEALRL  324
2D6   MVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRF  366
       β1-4        β2-1     β2-2      β1-3  K'                  K"
1oxa  IAPPE-TTTRFAAEEVEIG-GVAIPQYSTVLVANGAANRDPSQFP-DPHRFDVTR--DTR  339
2cpp  FSLV--ADGRILTSDYEFH-GVQLKKGDQILLPQMLSGLDERENA-CPMHVDFSRQ-KV-  345
1cpt  TAPVK-SFMRTALADTEVR-GQNIKRGDRIMLSYPSANRDEEVFS-NPDEFDITRF-PN-  365
2hpd  WPTAP-AFSLYAKEDTVLGGEYPLEKGDELMVLIPQLHRDKTIWGDDVEEFRPERFEN-P  382
2D6   GDIVPLGMTHMTSRDIEVQ-GFRIPKGTTLITNLSSVLKDEAVWE-KPFRFHPEHFLDAQ  424
                                         L               β3-1        β4-1
1oxa  ---G----HLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSIG--IDADDVVWRRSL  390
2cpp  ---S----HTTFGHGSHLCLGQHLARREIIVTLKEWLTRIPDFSIAPGA---QIQHK-SG  394
1cpt  ---R----HLGFGWGAHMCLGQHLAKLEMKIFFEELLPKLKSVELS--G---PPRLVATN  413
2hpd  -SAIPQHAFKPFGNGQRACIGQQFALHEATLVLGMMLKHFDFEDHT--NYE-LDIKET-   436
2D6   GHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPT--GQPRPSHHGVFA  482
       β6-2  β4-2
1oxa  LLRGIDHLPVRLDG-----------                               404
2cpp  IVS-GVQALPLVWD-PATTKAV----                               414
1cpt  FVGGPKNVPIRFT----KA------                                428
2hpd  LTLKPEGFVVKA--K--SKKIPLGG                                457
2D6   FLVSPSPYELCAVPR----------                                497
```

Fig. 1. Alignment of the four bacterial P450 structural templates (alignment 1; P450s eryF, cam, terp, and BM3, respectively) used to generate comparative models of P450 2D6. Grey shading, α-helices; black shading, β-strands.

structure were superimposed based on their haem moieties. A GRID[39] box (18 × 18 × 18 Å^3; GRID spacing 0.66 Å) large enough to encompass the superposed active sites of all of the 3D structures was generated, and the active site of each model was characterized using 10 GRID probes [methyl (C3), carbonyl (O), carboxylate oxygen (O::), NH-amide like (N1), sp3 amine cation (N3+), phenolic OH (OH), alkyl OH (O1), water (OH2), trimethylammonium

```
                                                                      A'
2C5  -------------------------------------------PPGPTPFPIIGNILQIDA  47
2D6  ------MGLEA---LVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDF  51


         A                  β1-1      β1-2       B
2C5  KDISKSLTKFSECYGPVFTVYLGMKPTVVLHGYEAVKEALVDLGEEF-AGR-GSVPILEK 104
2D6  QNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTA-DRP-PVPITQI 109


                  C                           D
2C5  VSKG---LGIAFS-NAKTWKEMRRFSLMTLRNFGMGKRSIEDRIQEEARCLVEELRKTNA 160
2D6  LGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFANHSG 169


                  E                           F
2C5  SPCDPTFILGCAPCNVICSVIFHNRFDYKDEEFLKLMESLHENVELLGTPWLQVYNNFPA 220
2D6  RPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVP- 228


                  G                          H
2C5  LLDYFPGIHKTLLKNADYIKNFIMEKVKEHQKLLDVNN-PRDFIDCFLIKMEQENN---L 276
2D6  VLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPES 288


                 I                  J
2C5  EFTLESLVIAVSDLFGAGTETTSTTLRYSLLLLLKHPEVAARVQEEIERVIGRHRSPCMQ 336
2D6  SFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMG 348


     J'        K                      β1-4   β2-1   β2-2    β1-3 K'
2C5  DRSRMPYTDAVIHEIQRFIDLLP-TNLPHAVTRDVRFRNYFIPKGTDIITSLTSVLHDEK 395
2D6  DQAHMPYTTAVIHEVQRFGDIVP-LGMTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEA 407


                 K"                               L
2C5  KAFPNPKVFDPGHFLDESGNFKKS-DYFMPFSAGKRMCVGEGLARMELFLFLTSILQNFK 453
2D6  AVWEKPFRFHPEHFLDAQGHFVK-PEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFS 465


     β3-1                        β3-2
2C5  LQSLVEPKDLDITAVVNGFVSVPPSYQLCFIPIHH                          488
2D6  FS-VPTGQPRPSHHGVFAFLVSPSPYELCAVPR--                          497
```

Fig. 2.   Single-template alignment (alignment 2) with P450 2C5 used to generate comparative models of P450 2D6. Grey shading, α-helices; black shading, β-strands.

cation (NM3), and chlorine (Cl)] This generated 10 values for each grid point for each of the five structures. GOLPE[40] was used to analyze these data—loadings and scores plots were generated, giving insight into the macroscopic differences in the topology and chemical identity between the active sites of the different structures. Scores plots were produced with one data point per structure by combining the data for all the probes for each grid point.

Docking

To evaluate the predictive value of the 2D6 models, docking studies were carried out on two of the known substrates of 2D6: codeine[41] and MPTP.[22] Docking calculations were performed using the program GOLD.[42] The distinct advantage of using these substrates is that their interactions with 2D6 have been studied by NMR and the products formed are known[19,22]— hence, these data were used to investigate the predictive value of the models.

To investigate the quality of the dockings, the active site of the model was examined using GRID,[39] producing contour maps that highlight the favorable interactions of various chemical probes within the active site (the parameters for the GRID box and the probes used are given in the previous section). The contour maps generated by GRID

were analysed visually using Sybyl 6.5 (Tripos Inc., St. Louis, MO).

RESULTS AND DISCUSSION

Comparative Modeling

Our original models of 2D6[19] were based on the three bacterial P450 structures available at the time: P450 cam (2cpp[25]), P450 terp (1cpt[26]), and P450 BM3 (2hpd[1]). These were superseded by our models, which also incorporated P450 eryF (1oxa[27])[21,22]; these were superseded in turn by models based on (i) a modified structural alignment of the templates and (ii) aligning a predetermined alignment of the 2D subfamily against this as two profiles (Fig. 1). We present these latter models here and compare them with models produced by incorporating into the modeling a mammalian P450, rabbit 2C5 (1dt6[33]). It should be noted that, subsequent to producing our four bacterial-template based models, higher resolution structures of P450 cam (1phc[43]), BM3 (1bu7[44]), and eryF (1eup[45]) have become available. To make a more meaningful comparison with our four-template based models, the five-template based models incorporating 2C5 were generated using the original (2cpp, 1cpt, 2hpd, 1oxa) structures for the bacterial enzymes. An additional structure has since become avail-

```
                           A'                         A             β1-1              β1-2
1oxa   -------------------------ATVPDLESDSFHVDWYSTYAELR-ETAPVTPVRFL-----GQDAWL   41
2cpp   NLAPLPPHVPEHLVFD--------FDMYNPSNL--SAGVQEAWAVLQESNVPDLVWTRC------NGGHWI   64
1cpt   --------MDARATIPEH-----IARTVILPQGYADDEVIYPAFKWLR-DEQPLAMAHIE----GYDPMWI   53
2hpd   --------TIKEMPQPKTFGELKNLPL-LNTDKP-----VQALMKIAD-ELGEIFKFEAP-----GRVTRY   51
2C5    -----------PGPTPFPIIG-NILQI-D-AKDI-----SKSLTKFSE-CYGPVFTVYLG-----MKPTVV   75
2D6    --RQRWAARYPPGPLPLPGLG-NLLHV-DFQNTP------YCFDQLRR-RFGDVFSLQLA-----WTPVVV   80

            B      β1-5                                       B'
1oxa   VTGYDEAKAALSDL--RLSSDPKKKYPGVEVE-FPAYLGF--PEDVRNYFA-------TNMG-TSDPP-T-   97
2cpp   ATRGQLIREAYEDY-RHFSSEC----------PFI------PREAGEAY-------DFIPT-SMDPP-E-  107
1cpt   ATKHADVMQIGKQP-GLFSNA--E------GSEIL-YD-QNNEAFMRSISGGCPHVI-DSLT-SMDPP-T-  109
2hpd   LSSQRLIKEACDE--SRFDK----------LS------QALKFVRDF-----AG--DGLFTSWTH-EKN   95
2C5    LHGYEAVKEALVDLGEEFAGRG-----------------SVPILEKVS-------KGLGIAFSNAKT-  118
2D6    LNGLAAVREALVTHGEDTADRP-----------------PVPITQILG-FGPRSQ----GVFLARYGPA  127

            C        C'          D            β3-1   E*        E
1oxa   HTRLRKLVS-----QEFTV-RRVEAMRPRVEQITAELLDEVGD--S-GVVDIVDRFAHPLPIKICELLGV-  159
2cpp   QRQFRALAN-----QVVGM-PVVDKLENRIQELACSLIESLRP--QGQCNFTEDYAEPFPIRIFMLLAGL-  169
1cpt   HTAYRGLTL-----NWFQPAS-IRKLEENIRRIAQASVQRLLDF-DGECDFMTDCALYYPLHVVMTALGVP  173
2hpd   WKKAHNILL-----PSFSQQAMKGY-HAMMVDIAVQLVQKWERLNADEHIEVPEDMTRLTLDTIGLCGFNY  160
2C5    WKEMRRFSLMT--LRNFGMGK--RSIEDRIQEEARCLVEELRKT-NASPCDPTFILGCAPCNVICSVIFHN  184
2D6    WREQRRFSVST--LRNLGLGK--KSLEQWVTEEAACLCAAFANH-SGRPFRPNGLLDKAVSNVIASLTCGR  193

            F                                            G
1oxa   --DE--------AARGAF-GRWSSE-ILVMD--PER----------AEQRGQAAREVVNFILDLVERRRTE  206
2cpp   --PE-----EDIPHL-K---YLTDQ-MTRPD-G-SMTF-----------AEAKEALYDYLIPIIEQ--RR  211
1cpt   EDDE-----PLMLKLTQ---DFFGV----------------------EAARRFHETIATFYDYFNGFTVD  229
2hpd   R--FNSFYRDQPH----PFITSMVRALDEA-MNKLQR-ANPDDP-AYDENKRQFQEDIKVMNDLVDKIIAD  222
2C5    RFDYKDEEFLKLMESLHENVELLGTPWLQVYNNFPALLD------YFPGIHKTLLKNADYIKNFIMEKVKE  248
2D6    RFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLH-------IPALAGKVLRFQKAFLTQLDELLTE  257

            H     β5-1   β5-2                   I                  J
1oxa   ---P-G--D---LLSALISVQDDDD-G---RLSADELTSIALVLLLAGFEASVSLIGIGTYLLLTHPDQL  263
2cpp   QK-P-G--T---DAISIVANGQV--NGR--PITSDEAKRMCGLLLVGGLDTVVNFLSFSMEFLAKSPEHR  270
1cpt   RR-S-CP-KD--DVMSLLANSKL---DGN--YIDDKIYNAYYVAIATAGHDTTSSSSGGAIIGLSRNPEQL  290
2hpd   RK-ASGEQSD--DLLTHMLNGKDPETGE---PLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKNPHVL  287
2C5    HQ-KLLDVNNPRDFIDCFLIKME-QENNL--EFTLESLVIAVSDLFGAGTETTSTTLRYSLLLLLKHPEVA  317
2D6    HRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQ  328

            J'          K                β1-4  β2-1   β2-2   β1-3
1oxa   ALVRAD------------------PSALPNAVEEILRYIAPPE--TTRFAAEEVEIG-GVAIPQYSTVLV  313
2cpp   QELIER------------------PERIPAACEELLRRFSLV---ADGRILTSDYEFH-GVQLKKKGDQII  319
1cpt   ALAKSD------------------PALIPRLVDEAVRWTAPVK--SFMRTALADTEVR-GQNIKKRGDRIM  340
2hpd   QKAAEEAARVL-VDPVPSYKQVKQLKYVMVLNEALRLWPTAP--AFSLYAKEDTVIGGEYPLEEKGDELM  355
2C5    ARVQEEIERVIGRHRSPCMQDRSRMPYTDAVIHEIQRFIDLLP-TNLPHAVTRDVRFR-NYFIPKGTDIIT  386
2D6    RRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVP-LGMTHRTSRDIEVQ-G-FRIPKGTTLIT  397

            K'           K"                          L       β3-1
1oxa   ANGAANRDPSQFP-DPHRFDVTR--DTR---G----HLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPAL  374
2cpp   LPQMLSGLDERENA-CPMHVDFSRQ-KV----S---HTTFGHGSHLCLGQHLARREIIVTLKEWLTRIPDF  381
1cpt   LSYPSANRDEEVFS-NPDEFDITRF-PN----R---HLSFGQGSHLAKLAKLEMKIFFEELLPKLKSV  402
2hpd   VLIPQLHRDKTIWGDDVEEFRPERFEN-P-SAIPQHAFKPFGNGQRACIGQQFALHEATLVLGMMLKHFDF  424
2C5    S-LTSVLHDEKAFP-NPKVFDPGHFLDESGNFKKSDYFMPFSAGKRMCVGEGLARMELFLFLTSILQNFKL  455
2D6    N-LSSVLKDEAVWE-KPFRFHPEHFLDAQGHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSF  466

            β4-1       β6-2     β4-2      β3-2
1oxa   SLG--IDADDVVWRRSLLLRGIDHLPVRLDG-----------  404
2cpp   SIAPGA---QIQHK-SGIVSGVQALPLVWD-PATTKAV----  414
1cpt   ELS--G---PPRLVATNFVGGPKNVPIRFT----KA-----  428
2hpd   ED-HT--NYE-LDIKET-LTLKPEGFVVKA--K-SKKIPLGG  457
2C5    QSLVE-----PKDLDITAVVNGFVSVPPSYQLCFIPIHH---  488
2D6    SVPT------GQPRPSHHGVFAFLVSPSPYELCAVPR-----  497
```

Fig. 3. The FSSP-determined alignment (alignment 3) used to generate comparative models of P450 2D6. Grey shading, α-helices; black shading, β-strands.

able: P450 nor (1rom[46]; 19.3% sequence ID with 2D6). This structure was not used.

The results from the initial PSI-BLAST search using the amino acid sequence of 2D6 identified four suitable template structures for comparative modeling (Table I). All the targets identified were P450s, from a variety of microorganisms—the class I enzymes P450 terp (PDB accession code 1cpt[26]) P450 eryF (1oxa[27]), P450 cam (2cpp[25]) as well as the class II enzyme P450 BM3 (2hpd[1]). The class II mammalian (rabbit) P450 2C5 structure (1dt6[33]) was incorporated into alignments 2–4.

Model Validation

The results of the model validation are summarized in Table II; the corresponding values from the templates— the baseline against which these values were compared— are summarized in Table III.

```
                                          A'                          A
1oxa  --------A----TV--PD---L-E----SDSFH-VD------------------WYSTYAELRET-  26
2cpp  ------NLAPLPPHVPE--HLVF-----D-F---DMY-N-PSN-LSAGV----QEAWAVLQ----ESN  49
1cpt  ------MDA-RA-TIP--------EH--IARTVILP-Q--G---YAD-D-----EVIYPA FKWLR-D  36
2hpd  --------TIKEMPQPK-TFGELK-----NLPLL--NTDK-----------PVQALMKI ADEL---  36
2C5   ---------PGPTPFP--------IIGNILQID---------------AKDISKSLTKFS-EC  59
2D6   RQRWAARYPPGPLPLP-----------GLGNLLHVD-------------FQNTPYCFDQLRRR-  64

        β1-1            β1-2      B        β1-5                      B'
1oxa  -APVTPVRFL--G--QDAWLVTGYDEAKAALSDL--RLSSDPKKKYPGVEVE-FPAYLGF--PEDVRN  84
2cpp  VP--DLVWT-RCN--GGHWIATRGQLIREAYEDY-RHFSSEC----------PFI-------PREAGE  94
1cpt  EQPLAMAHI--EGY-DPMWIATKHADVMQIGKQP-GLFSNA--E------GSEIL-YD-QNNEAFMRS  90
2hpd  -GEIFKFEA----PGRVTRYLSSQRLIKEACDE----SRFDKN---------LS------QALKFVRD  80
2C5   YGPVFTVYLG----MKPTVVLHGYEAVKEALVDLGEEFAGRG------------------SVPILEK 104
2D6   FGDVFSLQLA----WTPVVVLNGLAAVREALVTHGEDTADRP-----------PVPITQILG-FGPRSQ 117

        =               C                C'           D
1oxa  YFA-------TNMG-TSDPP-T-HTRLRKLVS----QEFTV-RRVEAMRPRVEQITAELLDEVGD--S 135
2cpp  AY-------DFIPT-SMDPP-E-QRQFRALAN----QVVGM-PVVDKLENRIQELACSLIESLRP--Q 145
1cpt  ISGGCPHVI-DSLT-SMDPP-T-HTAYRGLTL----NWFQPAS-IRKLEENIRRIAQASVQRLLDF-D 148
2hpd  F-----AG--DGLFTSWTH-EKNWKKAHNILL----PSFSQQAMKGY-HAMMVDIAVQLVQKWERLNA 135
2C5   VS-------KGLGIAFSNAKT-WKEMRRFSLMT-LRNFGMGK--RSIEDRIQEEARCLVEELRKT-N 159
2D6   -------------GVFLARYGPAWREQRRFSVST-LRNLGLGK--KSLEQWVTEEAACLCAAFANH-S 168

        β3-1 E*        E                          F
1oxa  -GVVDIVDRFAHPLPIKVICELLGV---DE-------AARGAF-GRWSSE-ILVMD--PER------- 181
2cpp  -GQCNFTEDYAEPFPIRIFMLLAGL---PE----EDIPHL-K---YLTDQ-MTRPD-G-SMTF----- 193
1cpt  -GECDFMTDCALYYPLHVVMTALGVPEDDE----PLMLKLTQ---DFFGV---------------- 190
2hpd  DBHIEVPE-DMTRLTLDTIGLCGFNYR--FNSFYRDQPH----PFITSMVRALDEA-MNKLQR-ANPD 194
2C5   ASPCDPTF-ILGCAPCNVICSVIFHNRFDYKDEEFLKLMESLHENVELLGTPWLQVYNNFPALLD--- 223
2D6   GRPFRPNG-LLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEKLKEESGFLREVLNAVPVLLH--- 232

        G                    H     β5-1  β5-2
1oxa  ----AEQRGQAAREVVNFILDLVERRTE---P-G---D---DLLSALISVQDDDD-G---RLSADELT 232
2cpp  --------AEAKEALYDYLIPIIEQ--RRQK-P-G---T---DAISIVANGQV--NGR---PITSDEAK 239
1cpt  ------EAARRFHETIATFYDYFNGFTVDRR-S-CP-KD--DVMSLLANSKL---DGN--YIDDKYIN 258
2hpd  DP-AYDENKRQFQEDIKVMNDLVDKIIADRK-ASGEQSD--DLLTHMLNGKDPETGE---PLDDENIR 255
2C5   ---YFPGIHKTLLKNADYIKNFIMEKVKEHQ-KLLDVNNPRDFIDCFLIKME-QENNL--EFTLESLV 285
2D6   ----IPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQPPRDLTEAFLAEMEKAKGNPESSFNDENLR 296

        I                    J                   J'          K
1oxa  SIALVLLLAGFEASVSLIGIGTYLLLTHPDQLALVRAD----------------PSALPNAVEEIL 282
2cpp  RMCGLLLVGGLDTVVNFLSFSMEFLAKSPEHRQELIER----------------PERIPAACEELL 289
1cpt  AYYVAIATAGHDTTSSSSGGAIIGLSRNPEQLALAKSD----------------PALIPRLVDEAV 308
2hpd  YQIITFLIAGHETTSGLLSFALYFLVKNPHVLQKAAEEAARVL-VDPVPSYKQVKQLKYVGMVLNEAL 322
2C5   IAVSDLFGAGTETTSTTLRYSLLLLLKHPEVIEKRVIGRHRSPCMQDRSRMPYTDAVIHEIQ 353
2D6   IVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPEMGDDQAHMPYTTAVIHEVQ 364

        β1-4    β2-1     β2-2     β1-3K'            K''
1oxa  RYIAPPE--TTTRFAAEEVEIG-GVAIPQYSTVLVANGAANRDPSQFP-DPHRFDVTR--DTR---G- 340
2cpp  RRFSLV---ADGRILTSDYEFH-GVQLKKGDQILLPQMLSGLDERENA-CPMHVDFSRQ-KV----S- 346
1cpt  RWTAPVK--SFMRTALADTEVR-GQNIKRGDRIMLSYPSANRDEEVFS-NPDEFDITRF-PN----R- 366
2hpd  RLWPTAP--AFSLYAKEDTVIGGEYPLEKGDELMVIIPQLHRDKTIWGDDVEEFRPERFEN-P-SAIP 386
2C5   RFIDLLP-TNLPHAVTRDVRFR-NYFIPKGTDIITSLTSVLHDEKAFP-NPKVFDPGHFLDESGNFKK 418
2D6   RFGDIVP-LGMTHMTSRDIEVQ-GFRIPKGTTLITNLSSVLKDEAVWE-KPFRFHPEHFLDAQGHFVK 429

        L                       β3-1       β4-1    β6-2   β4-2
1oxa  ---HLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSLG--IDADDVVWRRSLLLRGIDHLPVRLD 403
2cpp  ---HTTFGHGSHLCLGQHLARREIIVTLKEWLTRIPDFSIAPGA---QIQHK-SGIVSGVQALPLVWD 407
1cpt  ---HLGFGWGAHMCLGQHLAKLEMKIFFEELLPKLKSVELS--G---PPRLVATNFVGGPKNVPIRFT 426
2hpd  QHAFKPFGNGQRACIGQQFALHEATLVLGMMLKHFDFEDHT--NYE-LDIKET-LTLKPEGFVVKA- 448
2C5   SDYFMPFSAGKRMCVGEGLARMELFLFLTSILQNFKLQSLVE-----PKDLDITAVVNGFVSVPPSYQ 480
2D6   PEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPT------GQPRPSHHGVFAFLVSPSPYE 491

        β3-2
1oxa  G-----------  404
2cpp  -PATTKAV----  414
1cpt  ----KA------  428
2hpd  -K--SKKIPLGG  457
2C5   ICFIPIHH----  488
2D6   ICAVPR------  497
```

Fig. 4. MALIGN3D-determined alignment (alignment 4) used to generate comparative models of P450 2D6. Grey shading, α-helices; black shading, β-strands.

The percentage of residues with conformations in the most favored regions of the Ramachandran plot[35] suggests that model 2 has the greatest number of residues in the most favored region (87%), followed by model 3 (83%), model 4 (80%), and finally model 1 (71%). This can be put into context by examining the scores achieved by the 3D structural templates (Table III). It is apparent that all but model 1 perform significantly better than the 2C5 crystal structure [71%; note that this is a low resolution structure (3.0 Å)] and compare well with the values obtained from

TABLE I. Sequences of P450 Structures Identified by a PSI-BLAST Search Using the Sequence of P450 2D6

Cytochrome P450	PSI BLAST score[a]	PSI BLAST sequence identity[b]	% Sequence ID to 2D6 from PSI BLAST	% Sequence ID to 2D6 from Alignment 1[c]	% Sequence ID to 2D6 from Alignment 2[d]	% Sequence ID to 2D6 from Alignment 3[e]	% Sequence ID to 2D6 from Alignment 4[f]	Number of amino acid residues
2C5 (1dt6)	614	197/479	41	—	41	41	42	473
Cam (2cpp)	283	71/375	18	14	—	13	12	414
BM3 (2hpd)	257	85/461	18	15	—	16	15	471
terp (1cpt)	210	54/264	20	15	—	14	14	412
eryF (1oxa)	175	64/402	15	14	—	15	15	403

[a]The PSI-BLAST score for an alignment is calculated by summing the scores for each aligned position and the scores for gaps.[24]
[b](Number of identical residues)/(length of sequence fragment identified by PSI-BLAST).
[c]See Figure 1.
[d]See Figure 2.
[e]See Figure 3.
[f]See Figure 4.

TABLE II. Results of Validation Studies Performed on the Lowest Energy Models Produced from the Three Different Sequence Alignments

Model	Ramachandran plot[a] (%)	Verify 3D[b] (total score)	Errat (%)	Mainchain RMSD[c] (Å)
Four-template (bacterial) (Model 1[d])	70	139	68	N/A
Single template (model 2[d])	87	173	75	0.8
Multiple template–FSSP (Model 3[d])	83	191	79	2.3
Multiple template–MALIGN3D (Model 4[d])	80	182	74	2.9

[a]Percentage of residues with ϕ, ψ conformation in the most favoured regions of the Ramachandran plot.[35]
[b]The total Verify 3D score summed over all of the residues. For a protein the same size as P450 2D6 a score of 207 is expected, and a score of <97 would indicate an incorrect structure.
[c]The RMSD between aligned mainchain atoms in the model, and those in the most homologous template P450 2C5 (1dt6).
[d]Model 1, Model 2, Model 3, and Model 4 refer to the lowest energy model derived from Alignment 1 (Fig. 1), Alignment 2 (Fig. 2), Alignment 3 (Fig. 3), and Alignment 4 (Fig. 4), respectively.

the other template structures. This observation can likely be attributed to the restraints within modeller producing models with good stereochemistry.

For a protein the size of 2D6, a Verify 3D score of $\geq$215 would indicate a valid structure, whereas a score <95 would indicate that the model is incorrect. None of the models achieved the expected score (Table II), with model 3 performing best. Again, this can be put into context by examining the crystal structures. Both P450 2C5 (182) and P450 terp (188) fail to achieve their expected scores (209 and 204, respectively). Each of the alignments used to generate the models uses either P450 terp in conjunction with other nonmammalian templates, 2C5 alone, or 2C5 plus P450 terp in conjunction with the other templates; therefore, it is perhaps not surprising that the 2D6 models have Verify 3D scores lower than expected for a structure of this size. All of the templates have Errat scores >90%. The 2D6 models have Errat scores below this, with model 3 once again returning the best result (79%; Table II).

The RMSD between aligned mainchain atoms in the model and those in the most homologous template (2C5) were measured (Table II). This suggests that model 2 (0.8 Å) does not deviate (sample conformational space) as much as would be expected, given the percentage sequence identity, from the crystal structure of 2C5—this is likely the result of using a single structural template (Chothia

and Lesk[47] indicated that a value of >1 Å Cα RMSD is expected for a template with 41% sequence identity). Model 3 and model 4 perform better (RMSDs of 2.3 and 2.9Å, respectively), illustrating that more of conformational space has been explored during the modeling process.

Model 3 has the highest Verify 3D and Errat scores and the second highest Ramachandran score (Table II), indicating that it is the "best," in terms of mainchain stereochemistry and amino acid environment, of all of the models.

Principal Component Analysis

PCA was used to identify the differences between the topology and chemistry of the active sites of the models and the most homologous crystal structure, 2C5. There are two pieces of important data generated when performing PCA: the loadings plot and the scores plot. The scores plot represents the grid data in a reduced dimensionality with the most significant differences between the models showing up in the lower components. The loadings plot describes the components in terms of the original variables (grid points).

The loadings plot is complex, and for this set of data we see the appearance of five distinct lines [Fig. 5(a)].

The different approaches to modeling will produce models of different shapes, and it is not possible to superim-

TABLE III. Results of Validation Studies Performed on the 3D Structural Templates Used to Generate Comparative models of 2D6

Cytochrome P450	Ramachandran plot (%)	Verify 3D (total score)	Errat (%)
2C5 (1dt6)	71	182	94
cam (2cpp)	90	222	97
BM3 (2hpd)	91	221	95
terp (1cpt)	89	188	95
eryF (1oxa)	91	184	96

pose all of the residues in the active site of one structure perfectly onto those of another structure. The active sites of all of the structures are distinctly different shapes, an observation supported by examining the MOLCAD (Tripos Inc.) surfaces (Fig. 6) generated for the active site of each of the models. The major differences in shape are due to the relative positions of inherently flexible regions of the protein, in particular the B′-C loop and the F-G loop, and, to a lesser extent, secondary structural elements that vary in position across the P450 family, in particular the F helix. Therefore, some of the GRID points will be placed in locations that, because of structural differences in the models, are partially or wholly within a protein structure. As a result, the energy of interaction calculated at this point would be phenomenally high; parameterisation within GRID returns the maximum allowed value (5 kcal mol^{-1}).

The straight lines are due to correlation between the two principal components as a result of the dominant effect of those grid points that are within the volume of one or more (but not all) of the proteins. The impact of this is that these points become essentially binary variables. The angle of the straight lines with respect to the two components is a result of the score of the protein on the components. The analysis is therefore dominated by the shape differences, and longer-range chemical differences will be swamped.

This data was eliminated by removing computationally those GRID points across the structures with the maximum (5 kcal mol^{-1}) value for all 10 probes in any single structure. As the loadings plot shows [Fig. 6(b)], this eliminated the straight lines observed previously.

A scores plot produces a more easily interpretable picture of the differences between individual models, by taking multidimensional information and reducing it into two dimensions. The relative distance between individual data points along each of the principal components is a measure of the variation between those points; in this case [Fig. 5(c)] the distance along the principal components is an indication of the extent of the differences between the active sites of models 1–4 and the 2C5 crystal structure. There are three distinct groups: group 1: model 1; group 2: model 2 and the 2C5 crystal structure; and group 3: models 3 and 4. The relatively large distance between group 1 and groups 2 and 3 along the first principal component shows that inclusion of the 2C5 structure in the modeling process causes significant differences in the active site of 2D6, not all of which can be attributed merely to a change in gross topology. Hence, model 1 is distinctly different from any of
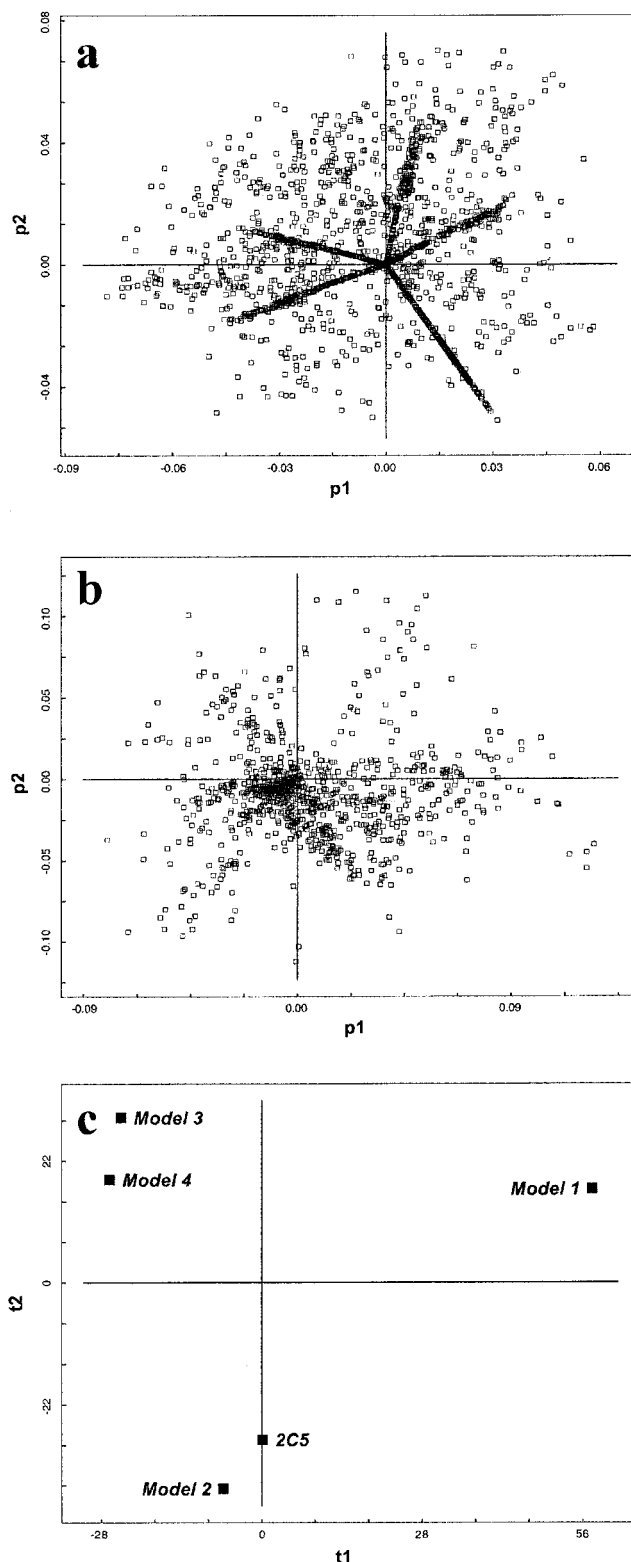
Fig. 5. (a) GOLPE-generated loadings plot for the superimposed active sites of models 1–4 and the 2C5 crystal structure, analyzed using GRID with 10 chemical probes. (b) GOLPE-generated loadings plot corresponding to panel a but from which those GRID points for which a value of 5 kcal mol^{-1} was returned in any structure for all 10 probes investigated have been deleted for all structures. (c) GOLPE-generated concatenated scores plot to highlight the differences between the active sites between the four models and the 2C5 crystal structure. PC1 accounts for 36% of the variance, PC2 accounts for 15% of the variance
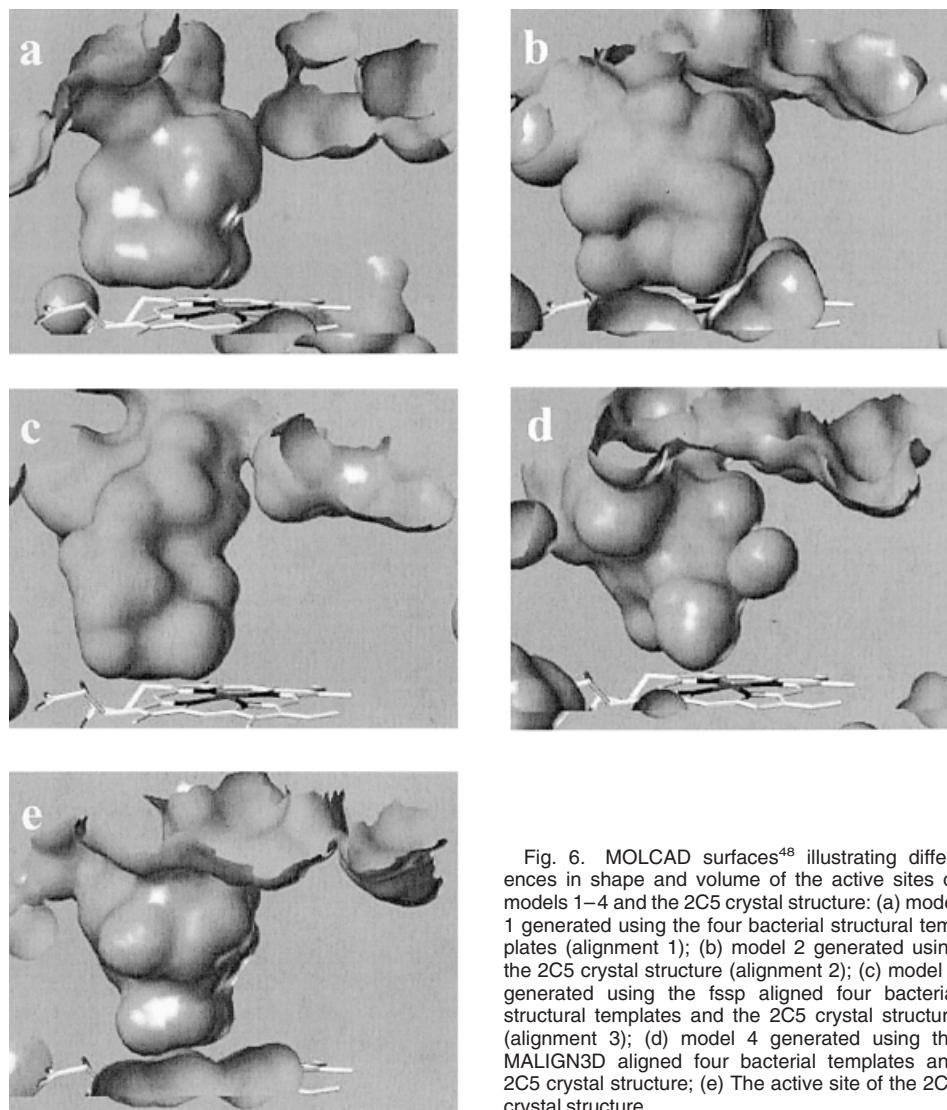
Fig. 6. MOLCAD surfaces[48] illustrating differences in shape and volume of the active sites of models 1–4 and the 2C5 crystal structure: (a) model 1 generated using the four bacterial structural templates (alignment 1); (b) model 2 generated using the 2C5 crystal structure (alignment 2); (c) model 3 generated using the fssp aligned four bacterial structural templates and the 2C5 crystal structure (alignment 3); (d) model 4 generated using the MALIGN3D aligned four bacterial templates and 2C5 crystal structure; (e) The active site of the 2C5 crystal structure.

the other structures analysed; this is not unexpected since it is the only model that does not include information from the 2C5 crystal structure. Group 2 illustrates that model 2 and the 2C5 crystal structure are similar, as indicated by their proximity along both PC1 and PC2 in the scores plot [Fig. 5(c)], but are significantly different from groups 1 and 3, as indicated by the separation of these groups along PC2. One reason for this may be the nature of the modeling process. Model 2 was constructed using information solely from the 2C5 crystal structure and is therefore more likely to share more characteristics with 2C5 than models 3 and 4, which were built using information from 2C5 plus the bacterial templates. In turn, models 3 and 4 would be expected to be more similar to one another than to the other structures and again this is apparent from the scores plot [Fig. 5(c)]. Discarding the bacterial P450-based model (model 1), because of the major differences between it and the remainder of the structures, allows the more subtle differences between the different 2C5-based models to be investigated.

The initial loadings plot for the analysis of models 2–4 and 2C5 shows distinct lines radiating from the origin [Fig. 7(a)], one for each of the structures. This illustrates, as when model 1 was included in the analysis [Fig. 5(a)], that the shape of the active site for each structure is different and to gain further information it is necessary to eliminate the data responsible for the appearance of these lines [Fig. 7(b)]. The scores plot generated shows two distinct groups differentiated along PC1: (i) models 3 and 4, (ii) model 2, and the 2C5 crystal structure. Again, the models produced from the multiple template alignments, group (i), are similar, as indicated by their comparable PC1 values [Fig. 7(c)]. They are different to both model 2 and the 2C5 crystal structure. In contrast to the previous scores plot [Fig. 5(c)] model 3 and model 4 have significantly different values along PC2, which may be due to the differences between the amino acid sequence alignments used to generate model 3 and model 4 becoming more apparent with the removal of the data for the disparate model 1.
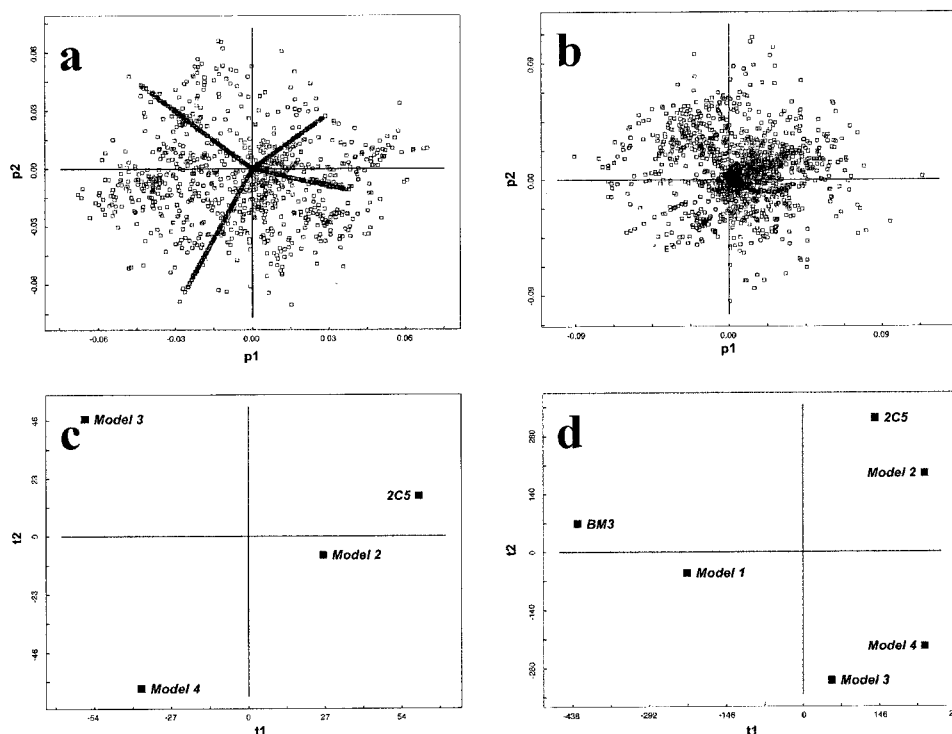
Fig. 7. (a) GOLPE-generated loadings plot from the active sites of the crystal structure of 2C5 and those models (models 2, 3 and 4) that incorporate information from 2C5. (b) GOLPE-generated loadings plot corresponding to panel a but from which those GRID points for which a value of 5 kcal mol^{-1} was returned in any structure for all 10 probes investigated have been deleted for all structures. (c) GOLPE-generated scores plot illustrating the differences in the active sites between models 2, 3, and 4 and the 2C5 crystal structure. PC1 accounts for 36% of the variance, PC2 accounts for 13% of the variance. (d) GOLPE-generated scores plot to illustrate the differences between the most homologous bacterial template (BM3), the 2C5 crystal structure and models 1–4. PC1 accounts for 38% of the variance, PC2 accounts for 11% of the variance.

One cause for concern raised by the scores plot was the difference between the 2C5 crystal structure and models 3 and 4. If models 3 and 4 were observed to be either as different, or more different, from the 2C5 crystal structure than the (more distant) bacterial templates, it would imply that including the 2C5 template in the modeling process has no added value, as it simply gives an answer which is different to the previous one and not necessarily a more accurate representation of 2D6. Hence, the crystal structure of the most homologous bacterial template BM3 (2hpd) was incorporated into the PCA. As can be seen from the resulting scores plot [Fig. 7(d)], models 2, 3, and 4 are closer to, and therefore more similar to 2C5 than BM3, whereas model 1 has greater similarity with the bacterial crystal structure. This implies that the incorporation of 2C5 into the modeling process alters the shape and/or chemistry of the active site substantially in comparison to the bacterial templates.

To ascertain the differences at the atomic level, which give rise to the differences observed in the PCA for the active sites of models 1–4 and the 2C5 crystal structure, a single GRID probe was used to characterize the structures, and the resulting loadings plot mapped back onto the protein structures as a GRID contour map. The three probes investigated (OH2, C3, and N3+) indicated that the major differences (i.e., those areas with the greatest variance for a probe as determined by the loadings plot)

arose from gross changes in relative positioning and shape of inherently flexible regions of the protein such as the F-G loop, the B'-C region, and the β-strands towards the C-terminus. However, the N3+ probe also highlighted chemical differences between the 2C5 crystal structure and the models, in particular the substitution of the acidic Glu 297 in the I helix of 2C5 for the neutral Val 308 in the I helix of the 2D6 models (Fig. 8).

Docking Studies

The validation studies discussed above indicated that model 3 was the "best" of the P450 2D6 models generated. Hence, the subsequent analyses of the active site of the 2C5-inclusive model were carried out using model 3.

Codeine orients itself within the active site of 2D6 in an orientation consistent with the known role of 2D6 as an *O*-demethylase of codeine,[49–51] and consistent with NMR data[19]—with the methoxy group situated close to the iron atom of the haem. It has been postulated that a carboxylate group in 2D6 forms a salt bridge with the basic nitrogen[9,14–16] of the substrate; this has been proposed both by modeling[16] and by mutagenesis[17] to be Asp 301. However, this was not observed in any of the highly ranked dockings. Instead, the basic nitrogen was observed to interact with a second acidic residue in the active site, Glu 216 (in the F-helix, SRS-2)—this was true for all of the highly ranked dockings from GOLD. Analysis of the dis-

tances from the haem iron to the hydrogen atoms of the substrate showed that, for these dockings, the Fe—H distances were closely similar to those obtained from NMR
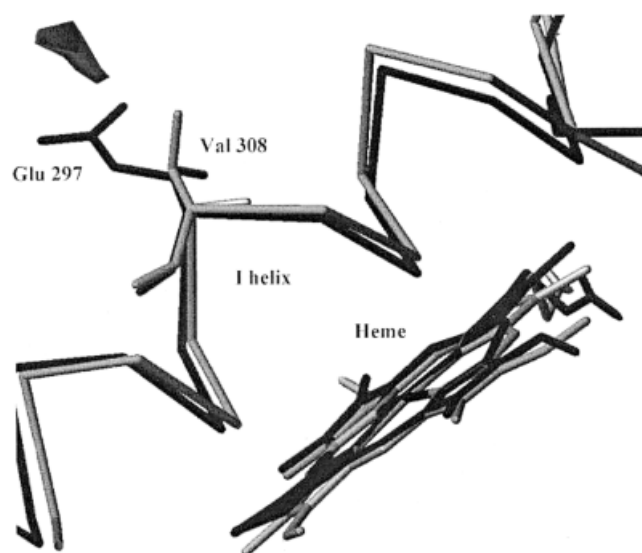

Fig. 8.   The GOLPE generated GRID-like loadings contours for the trimethylammonium probe, comparing the active sites of model 3 and the 2C5 crystal structure. This illustrates that the major difference between 2C5 and model 3 (i.e., 2D6) results from the chemical difference between (acidic) Glu 297 in the I-helix of 2C5 (black) and the corresponding (hydrophobic) Val 308 in 2D6 (model 3; grey).

studies of the codeine complex (Table IV).[19] These results thus implicate Glu 216, rather than Asp 301, as the major binding determinant of substrates in the active site of P450 2D6; nevertheless, they required further substantiation. To this end, docking investigations were carried out into another substrate of 2D6, 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), which has also been studied by NMR.[22]

Two metabolites of MPTP have been observed when it is metabolized by 2D6, the N-demethylated and *para*-hydroxylated product, which are thought to arise because of the different orientations this small molecule is able to adopt in the active site of 2D6. Once again, it is widely believed that the acidic residue Asp 301 is responsible for sequestering the basic nitrogen of the substrate—but again the GOLD generated dockings suggested that this was not the case and that in all of the highest ranked dockings the basic nitrogen of the substrate was positioned much closer to Glu 216 than Asp 301.
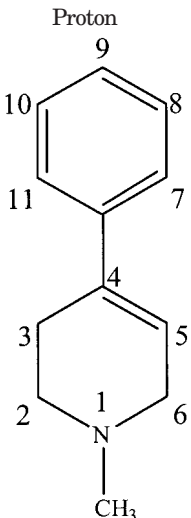
In the first instance GOLD predicted that only hydroxylation of the benzene ring would occur—none of the 20 dockings produced displayed an orientation that would facilitate N-demethylation. However, only 2 of the 20 dockings oriented MPTP in such a manner as to facilitate *para*-hydroxylation, with the remainder implying that *meta*-hydroxylation occurred. This is not consistent with the experimentally observed products and may be due to the relatively small size of the substrate molecule, allow-

TABLE IV. Interatomic Fe-H distances for codeine docked into the active site using GOLD compared with experimentally determined NMR distances[‡]

Proton	Experimental distance (Å)[19]	Average distance in models[a] (Å)	Difference between experimental and average distance in models (Å)
C1H	7.5 ± 0.2	6.3 ± 0.3	1.2
C2H	5.0 ± 0.1	4.1 ± 0.3	0.9
–OCH$_3$	3.1 ± 0.1	2.9 ± 0.1	0.2
C5H	9.1 ± 0.1	8.6 ± 0.2	0.5
C6H	10.0 ± 0.2	8.4 ± 0.1	1.6
C7H	9.3 ± 0.3	8.5 ± 0.1	0.8
C8H	10.2 ± 0.2	9.7 ± 0.1	0.5
C9H	11.2 ± 0.2	10.4 ± 0.2	0.8
C10H$_2$	9.8 ± 0.1	8.8 ± 0.2	1.0
C15H$_A$	9.3 ± 0.1	8.5 ± 0.1	0.8
C15H$_B$	10.8 ± 0.2	9.6 ± 0.1	1.2
NCH$_3$	12.1 ± 0.2	11.8 ± 0.3	0.3

[‡]From Modi et al.[19]
[a]The mean distance and standard deviation derived from 15 dockings of codeine into the 2D6 active site.

TABLE V. Interatomic Fe–H Distances for MPTP Docked into the Active Site Using GOLD Compared with Experimentally Determined NMR Distances

Proton

9
10   8
11   7
4
3   5
2   1   6
N
CH₃

	Experimental distance (Å)[a]	Average distance dockings (Å)	Average distance restrained dockings (Å)	Difference (unrestrained/restrained) between experimental and docking values (Å)
NCH$_3$	12.4 ± 0.6	12.4 ± 0.3	12.7 ± 0.1	0.0/0.3
C2H^A	11.1 ± 0.5	12.0 ± 0.5	10.8 ± 0.3	0.9/0.3
C2H^B	10.9 ± 0.7	11.0 ± 0.3	10.3 ± 0.3	0.1/0.6
C3H^A	9.1 ± 0.4	10.3 ± 0.5	8.8 ± 0.1	1.2/0.3
C3H^B	8.9 ± 0.2	10.0 ± 0.7	8.3 ± 0.3	1.1/0.6
C5H	8.3 ± 0.6	7.9 ± 0.2	8.6 ± 0.2	0.4/0.3
C6H^A	10.5 ± 0.5	10.2 ± 0.3	10.7 ± 0.2	0.3/0.2
C6H^B	10.6 ± 0.5	10.2 ± 0.2	10.8 ± 0.2	0.4/0.2
C7H, C11H	6.9 ± 0.4	6.6 ± 0.2	6.6 ± 0.5	0.3/0.3
C8H, C10H	4.5 ± 0.2	4.8 ± 0.28	4.1 ± 0.36	0.3/0.4
C9H	3.0 ± 0.1	5.2 ± 1.0	2.9 ± 0.08	2.2/0.1

[a]From Modi et al.[22]

ing it to move comparatively freely in the active site. Comparison of the interatomic iron-proton distances from the dockings with those determined by NMR shows that the only atom that falls outside of the expected range is the *para*-hydrogen (C9; Table V). To further investigate this, a docking experiment was carried out which restrained the C9 atom of MPTP to within 3.0 Å of the Fe atom of the haem, thus ensuring that the orientation of the substrate in all of the subsequent dockings would be poised for *para*-hydroxylation.

The results for all of the restrained dockings place the substrate molecule in an orientation where *para*-hydroxylation can occur, and the interatomic Fe—H distances for the dockings compare favorably to the experimental results (Table V). Additionally, the basic nitrogen atom only associates with Glu 216 (and not Asp 301), lending further credence to the importance of this residue as a major determinant of substrate binding.

Validation of Docking Studies

The contour maps generated by GRID for the trimethyl-ammonium cation (NM3) probe overlap the position occupied by the tertiary nitrogen atom in the substrates [Fig. 9(a)] at a relatively low contour level. This is an indication that the dockings, which place the basic nitrogen in the vicinity of Glu 216, are representative of the modeled active site. This also implies a key role for Glu 216 in binding of a substrate containing a basic nitrogen atom.

Mutagenesis experiments have shown that Asp 301 is critical to the activity of the P450 2D6.[17] It was believed that this was due to a direct interaction between the basic nitrogen atoms of the substrate molecules and the acidic residue, Asp 301, for those substrates containing a basic nitrogen atom 5–7 Å from the site of oxidation. Codeine, and MPTP undergoing *para*-hydroxylation fall, into this category. The role of Glu 216 as a binding determinant has been alluded to in previous studies as a way of explaining the metabolism of larger substrates, i.e., those substrates with have basic nitrogen atoms at a distance approximately ≥10 Å from the site of oxidation.[4,18,52,53] Our results suggest that Glu 216 additionally plays an important role in binding of 5–7 Å substrates. However, the proposed crucial role of Asp 301 cannot be overlooked. To address this, models were built using alignment 3 but additionally incorporating the mutation of acidic Asp 301 to a hydrophobic alanine (D301A). Dockings of codeine and subsequent GRID analysis of the mutant were then carried out.

In the first instance, the dockings did not place codeine in a position that would result in O-demethylation. Consequently, a distance restraint was incorporated anchoring the methoxy group to the haem iron. The docking was
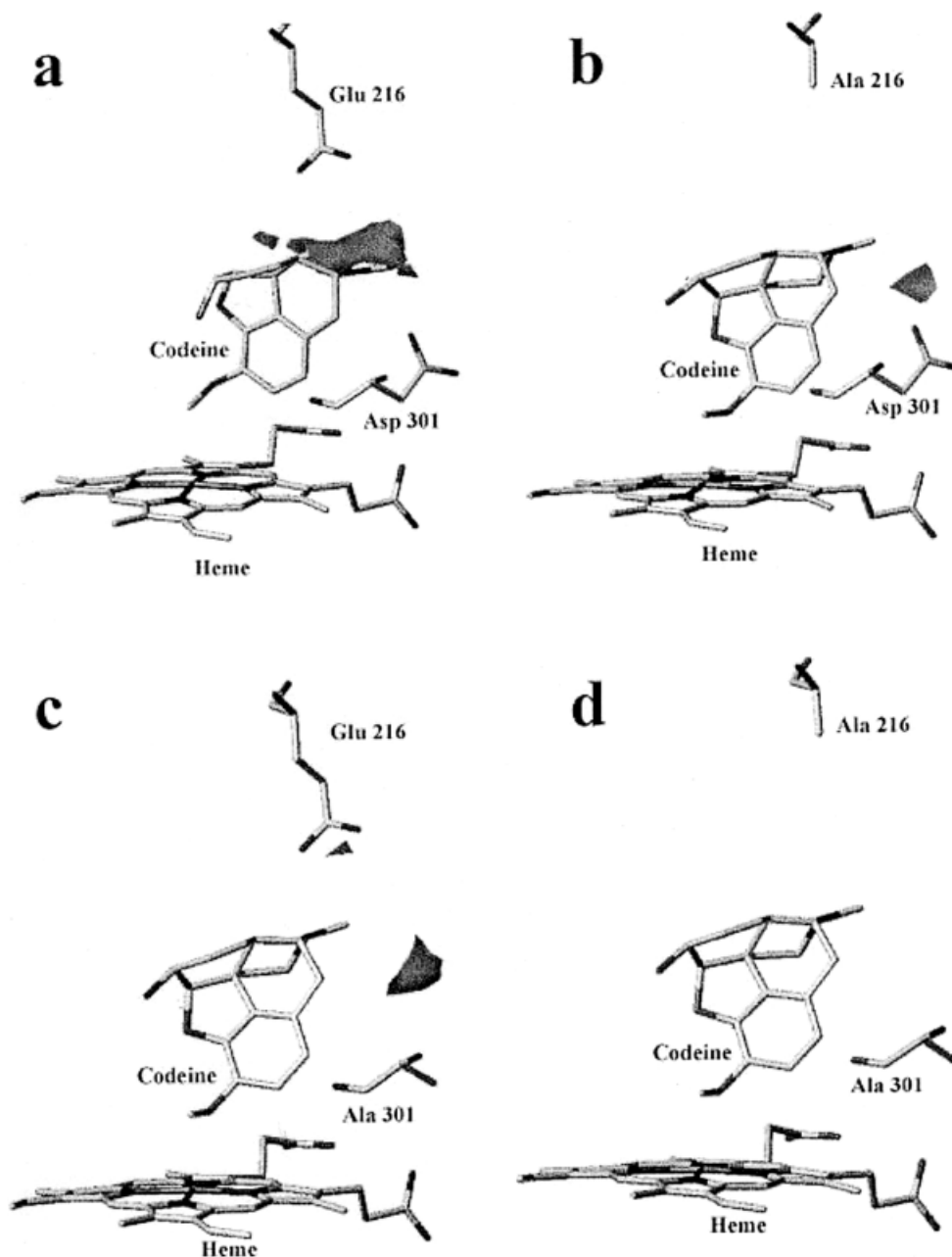
Fig. 9.   The active site of 2D6, illustrating the effects on the binding of basic nitrogen-containing ligands of mutating Glu 216 and/or Asp 301: (a) model 3, (b) the E216A mutant, (c) the D301A mutant, and (d) the E216A/D301A double mutant. In all cases, the highest ranked GOLD docking of codeine is overlaid on the contour map of the GRID trimethylammonium cation (NM3) probe contoured at -8 kcal mol^{-1}.

repeated; all dockings placing the methoxy group of the substrate in a position amenable to O-demethylation. The GRID analysis showed two interesting things. First, the most negative interaction energies for the NM3 probe in the D301A mutant and the original structure are comparable despite the removal of a negative charge in the D301A active site. Second, the relative propensity for the positively charged NM3 moiety in the D301A mutant is significantly reduced when compared with the original model, signifying that there are less favorable interactions between the active site of the D301A mutant and the NM3

probe. This is exemplified by the fact that there is no longer a correlation between the GRID contour maps and the dockings of codeine [Fig, 9(c)], which is apparent in the docking studies carried out using model 3 [Fig. 9(a)]. This suggests that Asp 301 plays a crucial electrostatic role in the binding of basic substrates by increasing the net negative charge within the active site.

Further studies show that an E216A mutation also leads to a reduction in the volume of the NM3 contour map comparable to that of the D301A mutant [Fig. 9(b)], and a structure that contains both E216A and D301A mutations

illustrates virtually no favorable interaction sites for the NM3 probe [Fig. 9(d)]. Hence, it appears that both Glu 216 and Asp301 play important roles in the binding of substrates.

Because the docking studies indicate that a salt bridge is formed with Glu 216 and Asp 301 does not directly sequester the basic nitrogen atom in substrates, it must affect substrate binding in a more subtle manner. We have already seen that one such way would be electrostatic i.e., increasing the net negative charge in the active site. Another would be if Asp 301 played a structural role, a hypothesis supported by recent work carried out by Guengerich and coworkers,[54] in particular the formation of a hydrogen bond with a residue in the flexible B′-C loop region of the protein. Analysis of all 15 of the models of 2D6 produced by alignment 3 showed that one of the carboxylate oxygen atoms of Asp 301 and the amide groups of Val 119 and Phe 120 is positioned appropriately to form a hydrogen bond; thus, there is a likely interaction between Asp 301 (I-helix) and the B′-C loop. Analysis of the most homologous structural template lends credence to this argument—the 2C5 crystal structure contains a possible hydrogen bond between Asp 290 (equivalent to Asp 301 in 2D6) and the amide of Ile 112 (equivalent to Val 119 in 2D6) of the B′-C loop. The temperature factors of the 2C5 crystal structure add additional support to this hypothesis, as the backbone of the inherently flexible B′-C loop becomes more ordered around Ile 112. Hence, one of the carboxyl oxygen atoms of Asp 301 could anchor (to some extent) the inherently flexible B′-C loop. Therefore, Asp301 could be subject to steric restraints, which would limit its ability to bind substrates directly.

Further support for the suggested role of Glu 216 in binding the basic moiety of substrate comes from analysis of the alignment of the amino acid sequences of 431 P450s.[55] This showed that only 2.8% (12/431) of all of the P450s analyzed have an acidic residue in the position equivalent to Glu 216 in 2D6, compared with 26% (112/431) that have an acidic residue in the position equivalent to Asp 301. There are only four sequences that contain residues equivalent to both Glu 216 and Asp 301: 2D6 (human), 2D14 (bovine), 2D4 (rat), and 2J1 (rabbit). These three 2D enzymes metabolize basic substrates,[6,56,57] with benzphetamine, another basic compound, the preferred substrate for the 2J1 enzyme.[58] This is important because it shows that Asp 301 is present in a large number of other P450s that do not metabolize basic substrates. For example, members of the 2C family that are known to hydroxylate steroids such as testosterone and progesterone[59,60]; 2E1, which is responsible for the metabolism of a variety of substances including ethanol[61]; 2H1 and 2H2, which metabolize steroids[62]; various mammalian orthologues of 1A1 and 1A2[63] and 21A1,[64] which hydroxylate steroids and fatty acids also fall into this category, along with the rat, bovine, and human orthologues of 1B1, which play an important part in steroid regulation for the development of the eye,[65] 17A1 which hydroxylates pregnenol and progesterone for a variety of vertebrates,[66] and P450-like subfamilies found in plants such as 71A,[67] which, along with members of the 71C family,[68] act as

trans-cinnamic acid hydroxylases. Some other members of the 2D family are also found to contain an Asp 301 equivalent, but no Glu 216 equivalent, i.e., 2D1, 2D2, 2D3, 2D5, 2D9, 2D10, 2D11, and 2D15. Of these only 2D2 contains an acidic equivalent (Asp) to Glu216,[69] and all are able to metabolize basic substrates such as debrisoquine. Despite this the preferred substrates of the remainder of the 2D family (i.e., excluding 2D4, 2D6, and 2D14) are fatty acids and steroids (see e.g., Matsunaga et al.[70]), and the majority of basic substrates are metabolized preferentially by 2D6. Finally, the N-hydroxylases CYP79[70] and CYP83[71] are also P450s with an equivalent of Asp301, but not Glu 216.

If both a Glu 216 equivalent and an Asp301 equivalent are present in a P450, basic substrates appear to be preferred; in contrast, Glu 216 alone is not sufficient for this preference. For those systems that contain a Glu 216, but no Asp 301, equivalent the story is very different. Members of the CYP73A family that have a Glu 216, but no Asp 301, equivalent, i.e., 73A1,[73] 73A2,[74] 73A3,[75] 73A4,[76] and 73A9[77], are known to preferentially metabolize cinnamic acid, whereas 4F2[78] and 7B[79] metabolize neurosteroids. This suggests that both Asp 301 and Glu 216 are important for the metabolism of basic substrates, and this could be due to the requirement for Asp 301 to position the B′-C loop in an appropriate position and Glu 216 to interact with the basic nitrogen atom of the substrate, before metabolism can occur.

CONCLUSIONS

In the absence of crystal structures, comparative modeling has given valuable insight into the interaction between basic substrates and human P450 2D6. The four sets of models of 2D6 generated illustrated that, in the case of 2D6, use of multiple structural templates (including rabbit 2C5) produce better quality models than either a single (most homologous; 2C5) template or four bacterial templates. Principal component analysis illustrates that the incorporation of information from the mammalian 2C5 crystal structure has a profound effect on the modeling process, altering the general topology and chemical characteristics of the active site, and that the models produced are significantly different from all of the templates. Docking studies predict an acidic residue, Glu 216 [in the F-helix (SRS-2) of 2D6], as a major determinant in the binding of basic substrates and suggest that Asp 301, thought previously to be a direct binding determinant, plays a key role in positioning the B′-C loop (SRS-1) via hydrogen bonds between its side-chain carboxylate group and mainchain amides in the B′-C loop. Investigation into the relationship between Asp 301- and Glu 216-containing systems suggests that both residues are necessary for metabolism of a basic substrate—if either residue is missing then the specificity of the enzyme is altered. Hence, Glu 216 does not replace Asp 301 in the binding of basic substrates; both appear necessary for the metabolism of basic substrates.

REFERENCES

1. Ravichandran KG, Boddupalli SS, Hasemann CA, Peterson JA, Deisenhofer J. Crystal structure of hemoprotein domain of P450BM3, a prototype for microsomal P450s. Science 1993;261: 731–736.
2. Schlichting I, Berendzen J, Chu K, Stock AM, Maves SA, Benson DE, Sweet RM, Ringe D, Petsko GA, Sligar SG. The catalytic pathway of cytochrome p450cam at atomic resolution. Science 2000;287:1615–1622.
3. 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.
4. Lewis DF. The CYP2 family: models, mutants and interactions. Xenobiotica 1998;28:617–661.
5. Jones BC, Tyman CA, Smith DA. Identification of the cytochrome P450 isoforms involved in the O-demethylation of 4-nitroanisole in human liver microsomes. Xenobiotica 1997;27:1025–1037.
6. Mahgoub A, Idle JR, Dring LG, Lancaster R, Smith RL. Polymorphic hydroxylation of Debrisoquine in man. Lancet 1977;2:584–586.
7. Eichelbaum M, Spannbrucker N, Dengler HJ. Influence of the defective metabolism of sparteine on its pharmacokinetics. Eur J Clin Pharmacol 1979;16:189–194.
8. Smith G, Stanley LA, Sim E, Strange RC, Wolf CR. Metabolic polymorphisms and cancer susceptibility. Cancer Surv 1995;25:27–65.
9. Koymans L, Vermeulen NP, van Acker SA, te Koppele JM, Heykants JJ, Lavrijsen K, Meuldermans W, Donne-Op den Kelder GM. A predictive model for substrates of cytochrome P450-debrisoquine (2D6). Chem Res Toxicol 1992;5:211–219.
10. Kalow W. Genetics of drug transformation. Clin Biochem 1986;19: 76–82.
11. Guengerich FP. Polymorphism of cytochrome P-450 in humans. Trends Pharmacol Sci 1989;10:107–109.
12. Relling MV. Polymorphic drug metabolism. Clin Pharm. 1989;8: 852–863.
13. Meyer UA. Genetic polymorphisms of drug metabolism. Fundam Clin Pharmacol 1990;4:595–615.
14. Wolff T, Distlerath LM, Worthington MT, Groopman JD, Hammons GJ, Kadlubar FF, Prough RA, Martin MV, Guengerich FP. Substrate specificity of human liver cytochrome P-450 debrisoquine 4-hydroxylase probed using immunochemical inhibition and chemical modeling. Cancer Res 1985;45:2116–2122.
15. Meyer UA, Gut J, Kronbach T, Skoda C, Meier UT, Catin T, Dayer P. The molecular mechanisms of two common polymorphisms of drug oxidation—evidence for functional changes in cytochrome P-450 isozymes catalysing bufuralol and mephenytoin oxidation. Xenobiotica 1986;16:449–464.
16. Islam SA, Wolf CR, Lennard MS, Sternberg MJ. A three-dimensional molecular template for substrates of human cytochrome P450 involved in debrisoquine 4-hydroxylation. Carcinogenesis 1991;12:2211–2219.
17. Mackman R, Tschirret-Guth RA, Smith G, Hayhurst GP, Ellis SW, Lennard MS, Tucker GT, Wolf CR, de Montellano ctive-site topologies of human CYP2D6 and its aspartate-301 → glutamate, asparagine, and glycine mutants. Arch Biochem Biophys 1996;331: 134–140.
18. Lewis DF, Eddershaw PJ, Goldfarb PS, Tarbit MH. Molecular modelling of cytochrome P4502D6 (CYP2D6) based on an alignment with CYP102: structural studies on specific CYP2D6 substrate metabolism. Xenobiotica 1997;27:319–339.
19. Modi S, Paine MJ, Sutcliffe MJ, Lian LY, Primrose WU, Wolf CR, Roberts GC. A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. Biochemistry 1996;35:4540–4550.
20. de Groot MJ, Vermeulen NP, Kramer JD, van Acker FA, Donne-Op den Kelder GM. A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102, and P450 108. Chem Res Toxicol 1996;9:1079–1091.
21. Smith G, Modi S, Pillai I, Lian LY, Sutcliffe MJ, Pritchard MP, Friedberg T, Roberts GC, Wolf CR. Determinants of the substrate specificity of human cytochrome P-450 CYP2D6: design and construction of a mutant with testosterone hydroxylase activity. Biochem J 1998;331:783–792.
22. Modi S, Gilham DE, Sutcliffe MJ, Lian LY, Primrose WU, Wolf CR, Roberts GC. 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine as a substrate of cytochrome P450 2D6: allosteric effects of NADPH-cytochrome P450 reductase. Biochemistry 1997;36:4461–4470.
23. Berman HM, Westbrook J, Feng Z, Gilliland, G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
24. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25: 3389–3402.
25. Poulos TL, Finzel BC, Gunsalus IC, Wagner GC, Kraut J. The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. J Biol Chem 1985;260:16122–16130.
26. Hasemann CA, Ravichandran KG, Peterson JA, Deisenhofer J. Crystal structure and refinement of cytochrome P450terp at 2.3 A resolution. J Mol Biol 1994;236:1169–1185.
27. Cupp-Vickery JR, Poulos TL. Structure of cytochrome P450eryF involved in erythromycin biosynthesis. Nat. Struct. Biol 1995;2: 144–153.
28. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
29. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
30. Lewis DF. Physical methods in the study of the active site geometry of cytochromes P-450. Drug. Metab. Rev. 1986;17:1–66.
31. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
32. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
33. Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE. Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. J Inorg. Biochem 2000;81:183–190.
34. Holm L, Sander C. The FSSP database: fold classification based on structure-structure alignment of proteins. Nucleic Acids Res 1996;24:206–209.
35. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. J Mol Biol 1993;231:1049–1067.
36. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 1993;2:1511–1519.
37. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356:83–85.
38. Venclovas C, Zemla A, Fidelis K, Moult J. Criteria for evaluating protein structures derived from comparative modeling. Proteins 1997;Suppl.:7–13.
39. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 1985;28:849–857.
40. Cruciani G, Watson KA. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. J Med Chem 1994;37:2589–2601.
41. Wrighton SA, Stevens JC. The human hepatic cytochromes P450 involved in drug metabolism. Crit Rev Toxicol 1992;22:1–21.
42. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727-748.
43. Poulos TL, Finzel BC, Howard AJ. Crystal structure of substrate-free Pseudomonas putida cytochrome P-450. Biochemistry 1986;25: 5314–5322.
44. Sevrioukova IF, Li H, Zhang H, Peterson JA, Poulos TL. Structure of a cytochrome P450-redox partner electron-transfer complex. Proc. Natl. Acad. Sci U S A 1999;96:1863–1868.
45. Cupp-Vickery J, Anderson R, Hatziris Z. Crystal structures of ligand complexes of P450eryF exhibiting homotropic cooperativity. Proc. Natl. Acad. Sci U S A 2000;97:3050–3055.

46. Park SY, Shimizu H, Adachi S, Nakagawa A, Tanaka I, Nakahara K, Shoun H, Obayashi E, Nakamura H, Iizuka T, Shiro Y. Crystal structure of nitric oxide reductase from denitrifying fungus Fusarium oxysporum. Nat. Struct. Biol 1997;4:827–832.

47. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;5:823–826.

48. Brickmann J. A new approach to the display of local molecular surface. J Comp Aid Mol Des 1993;7:503.

49. Mortimer O, Persson K, Ladona MG, Spalding D, Zanger UM, Meyer UA, Rane A. Polymorphic formation of morphine from codeine in poor and extensive metabolizers of dextromethorphan: relationship to the presence of immunoidentified cytochrome P-450IID1. Clin Pharmacol Ther 1990;47:27–35.

50. Desmeules J, Gascon MP, Dayer P, Magistris M. Impact of environmental and genetic factors on codeine analgesia. Eur J Clin Pharmacol 1991;41:23–26.

51. Ladona MG, Lindstrom B, Thyr C, Dun-Ren P, Rane A. Differential foetal development of the O- and N-demethylation of codeine and dextromethorphan in man. Br J Clin Pharmacol 1991;32:295–302.

52. Venhorst J, Onderwater RC, Meerman JH, Commandeur JN, Vermeulen NP. Influence of N-substitution of 7-methoxy-4-(aminomethyl)-coumarin on cytochrome P450 metabolism and selectivity. Drug Metab Dispos 2000;28:1524–1532.

53. de Groot MJ, Ackland MJ, Horne VA, Alex AA, Jones BC. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. J Med Chem 1999;42:1515–1524.

54. Hanna IH, Kim M-S, Guengerich FP. Heterologous expression of cytochrome P450 2D6 mutants, electron transfer and catalysis of bufuralol hydroxylation: The role of aspartate 301 in structural integrity. Arch Biochem Biophys 2001:393:255–261

55. http://www.icgeb.trieste.it/~p450srv/

56. Matsunaga E, Umeno M, Gonzalez FJ. The rat P450 IID subfamily: complete sequences of four closely linked genes and evidence that gene conversions maintained sequence homogeneity at the heme-binding region of the cytochrome P450 active site. J Mol Evol 1990;30:155–169.

57. Tsuneoka Y, Matsuo Y, Higuchi R, Ichikawa Y. Characterization of the cytochrome P-450IID subfamily in bovine liver. Nucleotide sequences and microheterogeneity. Eur J Biochem 1992;208:739–746.

58. Kikuta Y, Sogawa K, Haniu M, Kinosaki M, Kusunose E, Nojima Y, Yamamoto S, Ichihara K, Kusunose M, Fujii-Kuriyama Y. A novel species of cytochrome P-450 (P-450ib) specific for the small intestine of rabbits. cDNA cloning and its expression in COS cells. J Biol Chem 1991;266:17821–17825.

59. Oguri K, Yamada H, Yoshimura H. Regiochemistry of cytochrome P450 isozymes. Annu Rev Pharmacol Toxicol 1994;34:251–279.

60. Leemann TD, Transon C, Bonnabry P, Dayer P. A major role for cytochrome P450TB (CYP2C subfamily) in the actions of non-steroidal antiinflammatory drugs. Drugs Exp Clin Res 1993;19:189–195.

61. Song BJ, Gelboin HV, Park SS, Yang CS, Gonzalez FJ. Complementary DNA and protein sequences of ethanol-inducible rat and human cytochrome P-450s. Transcriptional and post-transcriptional regulation of the rat enzyme. J Biol Chem 1986;261:16689–16697.

62. Hobbs AA, Mattschoss LA, May BK, Williams KE, Elliott WH. The cDNA and protein sequence of a phenobarbital-induced chicken cytochrome P-450. J Biol Chem 1986;261:9444–9449.

63. Kawajiri K, Watanabe J, Gotoh O, Tagashira Y, Sogawa K, Fujii-Kuriyama Y. Structure and drug inducibility of the human cytochrome P-450c gene. Eur J Biochem 1986;159:219–225.

64. Chung BC, Matteson KJ, Miller WL. Structure of a bovine gene for P-450c21 (steroid 21-hydroxylase) defines a novel cytochrome P-450 gene family. Proc Natl Acad Sci USA 1986;83:4243–4247.

65. Sutter TR, Tang YM, Hayes CL, Wo YY, Jabs EW, Li X, Yin H, Cody, CW, Greenlee WF. Complete cDNA sequence of a human dioxin-inducible mRNA identifies a new gene subfamily of cytochrome P450 that maps to chromosome 2. J Biol Chem 1994;269: 13092–13099.

66. Chung BC, Picado-Leonard J, Haniu M, Bienkowski M, Hall PF, Shively JE, Miller WL. Cytochrome P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase): cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues. Proc Natl Acad Sci USA 1987;84:407–411.

67. Bozak KR, Yu H, Sirevag R, Christoffersen RE. Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. Proc Natl Acad Sci USA 1990;87:3904–3908.

68. Frey M, Kliem R, Saedler H, Gierl A. Expression of a cytochrome P450 gene family in maize. Mol Gen Genet 1995;246:100–109.

69. Gonzalez FJ, Matsunaga T, Nagata K, Meyer UA, Nebert DW, Pastewka J, Kozak CA, Gillette J, Gelboin HV, Hardwick, JP. Debrisoquine 4-hydroxylase: characterization of a new P450 gene subfamily, regulation, chromosomal mapping, and molecular analysis of the DA rat polymorphism. DNA 1987;6:149–161.

70. Matsunaga E, Zanger UM, Hardwick, JP, Gelboin HV, Meyer UA, Gonzalez FJ. The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4-hydroxylase deficiency in DA rats. Biochemistry 1989;28:7349–7355.

71. Koch BM, Sibbesen O, Halkier BA, Svendsen I, Moller BL. The primary sequence of cytochrome P450tyr, the multifunctional N-hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor (L.) Moench Arch Biochem Biophys 1995;323:177–186.

72. Chapple CC. A cDNA encoding a novel cytochrome P450-dependent monooxygenase from Arabidopsis thaliana. Plant Physiol 1995;108:875–876.

73. Teutsch HG, Hasenfratz MP, Lesot A, Stoltz C, Garnier JM, Jeltsch JM, Durst F, Werck-Reichhart D. Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. Proc Natl Acad Sci USA 1993;90:4102–4106.

74. Mizutani M, Ward E, DiMaio J, Ohta D, Ryals J, Sato R. Molecular cloning and sequencing of a cDNA encoding mung bean cytochrome P450 (P450C4H) possessing cinnamate 4-hydroxylase activity. Biochem Biophys Res Commun 1993;190:875–880.

75. Fahrendorf T, Dixon RA. Stress responses in alfalfa (Medicago sativa L.). XVIII: Molecular cloning and expression of the elicitor-inducible cinnamic acid 4- hydroxylase cytochrome P450. Arch Biochem Biophys 1993;305:509–515.

76. Hotze M, Schroder G, Schroder J. Cinnamate 4-hydroxylase from Catharanthus roseus and a strategy for the functional expression of plant cytochrome P450 proteins as translational fusions with P450 reductase in Escherichia coli. FEBS Lett 1995;374:345–350.

77. Frank, MR, Deyneka JM, Schuler MA. Cloning of wound-induced cytochrome P450 monooxygenases expressed in pea. Plant Physiol 1996;110:1035–1046.

78. Kikuta Y, Miyauchi Y, Kusunose E, Kusunose M. Expression and molecular cloning of human liver leukotriene B4 omega- hydroxylase (CYP4F2) gene. DNA Cell Biol 1999;18:723–730.

79. Stapleton G, Steel M, Richardson M, Mason JO, Rose KA, Morris RG, Lathe R. A novel cytochrome P450 expressed primarily in brain. J Biol Chem 1995;270:29739–29745.