J Comput Aided Mol Des (2007) 21:539–548
DOI 10.1007/s10822-007-9137-8

Effective virtual screening protocol for CYP2C9 ligands using a screening site constructed from flurbiprofen and S-warfarin pockets

Tímea Polgár · Dóra K. Menyhárd · György M. Keserű

Received: 8 May 2007 / Accepted: 26 September 2007 / Published online: 25 October 2007

Abstract An effective virtual screening protocol was developed against an extended active site of CYP2C9, which was derived from X-ray structures complexed with flubiprofen and S-warfarin. Virtual screening has been effectively supported by our structure-based pharmacophore model. Importance of hot residues identified by mutation data and structural analysis was first estimated in an enrichment study. Key role of Arg108 and Phe114 in ligand binding was also underlined. Our screening protocol successfully identified 76% of known CYP2C9 ligands in the top 1% of the ranked database resulting 76-fold enrichment relative to random situation. Relevance of the protocol was further confirmed in selectivity studies, when 89% of CYP2C9 ligands were retrieved from a mixture of CYP2C9 and CYP2C8 ligands, while only 22% of CYP2C8 ligands were found applying the structure-based pharmacophore constraints. Moderate discrimination of CYP2C9 ligands from CYP2C18 and CYP2C19 ligands could also be achieved extending the application domain of our virtual screening protocol for the entire CYP2C family. Our findings further demonstrate the existence of an active site comprising of at least two binding pockets and strengthens the need of involvement of protein flexibility in virtual screening.

Keywords CYP2C9 · Docking · Protein-based pharmacophore · Selectivity · Virtual screening

T. Polgár · G. M. Keserű
Gedeon Richter plc, P. O. Box 27, 1475 Budapest, Hungary

D. K. Menyhárd · G. M. Keserű (✉)
Budapest University of Technology and Economics,
Szt. Gellért tér 4, 1111 Budapest, Hungary
e-mail: gy.keseru@richter.hu

Introduction

Cytochrome P450 proteins (CYPs) are membrane-bound, haem containing enzymes that metabolize physiologically important compounds of microorganisms, plants and animals. Mammalian CYPs recognise and transform diverse xenobiotics such as drug molecules, environmental compounds and pollutants [1–3]. Within these, the CYP2C subfamily, which comprises CYP2C8, CYP2C9, CYP2C18, and CYP2C19, is responsible for the metabolism of about 20% of clinically used drugs [3, 4]. The CYP2C enzymes are all genetically polymorphic and share significant sequence identity (of approximately 70%) but are differences in their localization and substrate profile (Fig. 1). Of them, CYP2C9 ranks among the most important drug-metabolizing enzymes [5], taking part in the first pass metabolism of drugs, limiting their oral bioavailability. Among CYP2C9 substrates are anti-inflammatory agents such as diclofenac [6, 7], ibuprofen [8], naproxen [9], and piroxicam [10], anticoagulant compounds like S-warfarin [11, 12], as well as progesterone [13]. Furthermore, CYP2C9 participates in the metabolism of the endogenous substances, such as eicosanoids, arachidonic acid and thereby seems to affect the regulation of vascular homeostasis [11, 12].

Crystal structure of CYP2C9 has been determined in the ligand-free state [14, 15] and in complex with two different substrates, that of S-warfarin [14, 15] and flurbiprofen [16]. CYP2C9 adopts a topology common to all CYPs, where a conserved inner haem-binding core is formed by helices E, I, J, K and L, while the distal haem pocket and the ligand entrance region is subject to variation [17]. Clodfeter et al. has shown in a solvent mapping study that the large binding channels of the ligand-free state lack the ability to anchor small molecules. They concluded that small but

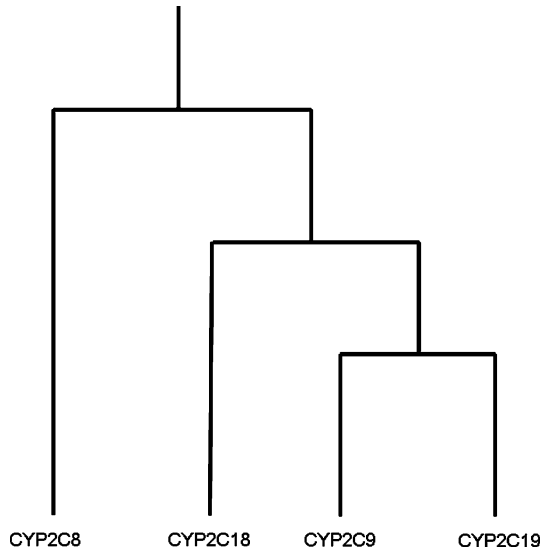

Fig. 1 Phylogenetic tree of human CYP2C family members [37]

essential rearrangements are required for the structure to be able to host ligand molecules [18]. Two examples of this are demonstrated by the complex crystal structures. Reorientation of a single Phe residue enables the enzyme to bind warfarin, although quite far from the haem iron. On the other hand, more extensive, but still quite localized changes take place when recognising flurbiprofen, pertaining to the B–C and F–G loop regions resulting in not only more spacious, but a more polar haem pocket.

The fact that small conformational changes cause significant variation in substrate binding and metabolic efficacy of CYP2C9 is demonstrated by its numerous phenotypes. The human CYP2C9 gene is highly polymorphic. At least 24 CYP2C9 alleles, predominantly of single amino acid change, have been identified to date and most of them show reduced activity [19–26].

CYP2C8—an important member of the CYP2C family (Fig. 1)—shows genetic variations and these lead to marked difference in activity. Although CYP2C8 and CYP2C9 share 78% sequence identity they exhibit relatively minor overlap of their substrate and inhibitor profile. While substrates of CYP2C9 are medium sized acidic molecules with one or two H-bond acceptors, substrates of CYP2C8 can be characterised as large, elongated, acidic or neutral molecules [27]. Crystal structure of CYP2C8 is also available. The active site of CYP2C8 is considerably more spacious than that of CYP2C9 due to an opening at residues Ser114 and Ile476 [28]. Phe residues occupy the corresponding positions of CYP2C9, and of the other CYPs only CYP3A4 shows similarity in this respect, in fact CYP2C8 shares many more substrates with CYP3A4 (of 25% sequence identity) including repaglinide [29], paclitaxel [30], morphine [31], carbamazepine [32], verapamil [33], zopiclone [34], and cerivastatin [35] than with CYP2C9. CYP2C8, in a way, shows a further example of conformational variability in the CYP2C9 type ligand-binding region, resulting in yet another type of selectivity, which makes the comparison of these structures especially intriguing.

CYP2C18 and CYP2C19 also belong to the CYP2C family [36, 37] and share 85% and 92% sequence identity with CYP2C9, respectively. Presently no X-ray structures are available for CYP2C18 and CYP2C19. However, several homology models were built for both the CYP2C18 and CYP2C19 [38–40] that attempt to demonstrate the subtle differences between them and can be used as structure-based guides for mutagenesis studies or screening of potential pharmaceuticals or toxins [41, 42]. CYP2C19 is responsible for the oxidative metabolism of many drugs including S-mephenytoin and omeprazole [39] while up to now only a few ligands of CYP2C18 are known [38, 40, 41].

Several compound properties related to CYP enzymes, such as the rate and site of metabolism, inhibition, induction and the selectivity for various isoforms should be considered during lead optimisation [43]. The design of predictive computational methods is therefore of raising interest for pharmaceutical industry. Structure-based virtual screening is one of these methods that can be used to identify inhibitors and substrates of CYP enzymes during library design [44, 45], investigate of isoenzyme selectivity [46, 47], understand structure-metabolism relationship [48, 49] or predict of heteroactivators [71]. Due to their pharmaceutical importance, the need for fast and effective screening techniques for CYP2C ligands is evident.

To best of our knowledge our group described the first structure-based virtual screening study against P450s [50]. Here we report our screening experiments on CYP2C9. A site of metabolism tool has been described for CYP2C9 substrates [51] and regions responsible for selectivity discussed [52] using a homology model of the enzyme. However, the crystal structures obtained since have shown the structure liable to greater variability upon ligand coordination than expected. The aim of the present calculations was to derive an effective virtual screening protocol to identify ligands of CYP2C9 selectively using the wealth of structural information available. Our best virtual screening protocol was when an extended binding site (screening site) was applied by combining S-warfarin and flubiprofen binding sites. The formulated protocol was tested in a selectivity study, CYP2C9 ligands were discriminated from the mixture of CYP2C9, CYP2C18, CYP2C19 and CYP2C8 ligands. The selectivity study further supported the efficiency of our virtual screening protocol as CYP2C9 ligands were separated from not only non-binders but even from other CYP2C ligands. However, the lack of experimental structural information prevented developing effective virtual screening protocols for CYP2C

isoenzymes other than CYP2C9. Since our preliminary docking calculations were ineffective using the apo structure of CYP2C8 we suggest that essential rearrangements are required for ligand-free CYP2C8 structure to be able to bind ligands. Based on high sequence homology and the large number of ligands shared by different CYP2C isoforms we hypothesize similar rearrangements for CYP2C18 and CY2C19 isoenzymes. These results also underlined the importance of protein flexibility in docking calculations.

Methods

Preparation of the screening database

Our screening library involved a subset of World Drug Index (WDI) as inactive molecules that were designated to reduce artificial enrichment. The WDI was filtered to eliminate compounds having molecular weight lower than 200 and >800, logP larger than 7, rotatable bonds more than 15 [53]. Inorganic and organometallic compounds were also filtered. Out of the ~40000 remaining molecules, 5369 were chosen based on a diversity selection using 2D UNITY fingerprints [54]; Tanimoto index was lower than 0.68 within the set. The WDI set did not contain known CYP2C ligands. The CYP2C9 ligand set was compiled manually from public sources [6–13, 29–35, 38–40]. Our main selection criterion was to cover as many different compound classes as possible to establish structural diversity among them. The final input library comprised the 5369 inactive and 42 CYP2C9 ligands (inhibitors and substrates) [55, 56] (content of ligands ~0.5 % such as in a typical real screening situation).

In total 108 CYP2C9 and 86 CYP2C8 ligands (substrates and inhibitors) were collected from Prous Integrity Drugs & Biologics database [55] and public sources [56]. 14 CYP2C18 [38, 40, 41, 56] and 76 CYP2C19 [38–40, 56] ligands (substrates and inhibitors) were compiled from public sources. These were used for CYP2C9/CYP2C selectivity tests. All CYP2C ligand sets are available from the authors upon request.

Preparation of protein coordinates and definition of active sites for docking with FlexX-Pharm

Crystal structures of CYP2C9 (PDB code: 1R90 (complexed with flubiprofen) [16], 1OG5 complexed with S-warfarin) and 1OG2 (apostructure) [14]) were used to screen a benchmark library of known ligands. 'A' chains were used for virtual screening in each case. Actives sites were defined as all the residues within the 6.5 Å of the bound ligand. Active site definition of the ligand-free structures was based on alignment with the ligand-bound ones and selection of corresponding residues. Haem was always included in the active site.

Definition and generation of the extended active site (screening site)

The extended active site was created from X-ray structure 1R90—as it does not incorporate several internal substitutions as well as differences in the N- and C-terminal sequences such as the constructs used for 1OG5 and 1OG2—with two subpockets, which were defined by S-warfarin and flubiprofen.

1R90 and 1OG5 3D structures were aligned by sequence homology, then S-warfarin from 1OG5 was merged into 1R90. The extended active site was defined as all the residues within the 6.5 Å of S-warfarin and flubiprofen. This extended active site, including Arg97, Gly98, Ile99, Phe100, Pro101, Leu102, Ala103, Glu104, Val113, Phe114, Ser115, Leu208, Ile213, Gln214, Asn217, Leu362, Thr364, Ser365, Leu366, Pro367, Leu388, Phe476 and Ala477 from the S-warfarin and Arg97, Glu104, Arg108, Val113, Phe114, Lys200, Leu201, Asn204, Ile205, Leu208, Ser209, Leu233, Val237, Met240, Val292, Asp293, Leu294, Phe295, Gly296, Ala297, Gly298, Glu300, Thr304, Leu366, Asn474, Phe476, Ala477, Ser478 and Val479 from the flubiprofen binding pockets, was modified such that the orientation of residues Phe100, Phe114, Pro367 and Phe476, major hydrophobic interaction partners of S-warfarin, were set according to their conformation in 1OG5 complex (Fig. 3).

Pharmacophore constraints were given for residues Arg108, Phe100, Phe114, Phe476 and Asn204 that were used as optional interaction constraints.

FlexX-Pharm

Standard parameters were used as implemented in SYBYL 7.3 [54] package. Formal charges, place particle, treat rings flexibly, protonate acids and amines deprotonated options were always checked. Thirty poses were saved in mol2 files for further analysis. ScreenScore was used during ligand building [57]. All stored poses were rescored using the CScore module of SYBYL 7.3 comprising the following functions: Dock, Chem, FlexX, PMF, and Gold scoring function. Pharmacophore constraints assigned by comparative analysis of ligand-bound X-ray structures and mutation experiments available in the literature [14, 16, 60] were given as optional interaction constraints (see details in Results).

Enrichment factor calculation

The effectiveness of different screening protocols has been investigated in enrichment studies aimed at identifying known actives from the screening library. Enrichment factors were used to assess the quality of the rankings in each docking run [58]:

$$EF_{(\%)} = \frac{\left(N_{active(\%)}/N_{(\%)}\right)}{\left(N_{active}/N_{all}\right)}$$

where $EF_{(\%)}$ is given at the % percentage of the ranked database, $N_{active(\%)}$ is the number of active compounds in a selected subset of the ranked database, $N_{(\%)}$ is the number of compounds in the subset, N_{active} and N_{all} are the number of active molecules and the number of compounds in the screening database. Enrichment factors were not scaled, i.e. absolute values depend on the ratio of active and inactive molecules and should be compared to the maximum (ideal) achievable EFs. EFs were calculated at 1% of the ranked database as a fixed threshold that enables the comparison of enrichment factors when studying the effectiveness of different screening protocols.

Using five different scoring functions we first extracted the best poses from the 30 saved ones generated for all compounds in the screening database. Then compounds with the lowest scores were ranked by the same five scoring functions, thereby the pose extraction and ranking was separated resulting in 25 enrichment factors for each docking run. Scoring function used for pose extraction and ranking are represented successively.

Results

Following structural analysis, virtual screens were carried out against active site models based on the ligand-free, the S-warfarin bound and the flurbiprofen bound CYP2C9 X-ray structures. Impact of conformation on the efficiency of virtual screening was investigated. Pharmacophore constraints were set up based on results of mutation experiments available in the literature and our own structural analysis [14, 16, 60]. The ability of our best virtual screening protocol to discriminate between CYP2C9 and other CYP2C isoenzyme ligands was also estimated.

Formulation of CYP2C9 models

Alignment of C_α atoms of the ligand-free structure and the S-warfarin bound complex of CYP2C9 resulted in a weighted root mean square distance of 0.36 Å corresponding to the lack of major structural rearrangement upon complexation (Fig. 2a). Weighted root mean square distance calculated by fitting C_α atoms of the flurbiprofen complex to that of the S-warfarin complex increased to 1.94 Å, as a result of a number of residues (those of the B–C, E–F and F–G loops, the F helix and the residues 467–478) adopting significantly different conformation (Fig. 2b). Sequence alignment revealed that there were several internal substitutions as well as differences in the N- and C-terminal amino acid sequences of 1OG5 and 1OG2 compared to 1R90 [16]. Noteworthy that B–C loops in 1OG5 and 1OG2 are more typical of conformations observed in structures of CYP2C5, CYP2C8 or CYP2B4 [16]. This structural alteration can be a consequence of the several internal substitutions in the sequence of 1OG5 and 1OG2 but it is likely to reflect the conformational flexibility of the protein. To test the effect of active site conformation on docking results, all three were used in independent enrichment studies.

In the S-warfarin complex Phe100, Phe114, Pro367, Phe476 are in hydrophobic contact with the ligand and the backbone amide group of Phe100 forms an H-bond with it. In the flurbiprofen complex Arg108 and Asn204 flip toward the enzyme interior and form H-bonds with the

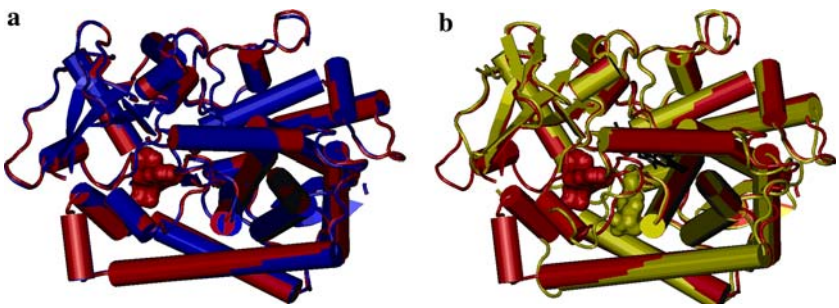

Fig. 2 (**a**) The ligand-free structure of CYP2C9 (PDB code: 1OG2) (shown in blue) and the S-warfarin-bound complex structure (PDB code: 1OG5) (shown in red) aligned by amino acid sequence homology. Van der Waals surface of S-warfarin is also shown (red). (**b**) The crystal structure of the complex of CYP2C9 with S-warfarin (PDB code: 1OG5) (shown in red) and with flurbiprofen (PDB code: 1R90) (shown in yellow) aligned by amino acid sequence homology. Van der Waals surface of S-warfarin (red) and flubiprofen (yellow) is also shown

ligand, while the importance of hydrophobic contacts, in this specific case, is less pronounced. Accordingly, in case of screening with the active site models formed according to the ligand-free state and the S-warfarin complex, interaction pharmacophore constraints included interactions with Phe100, Phe114, and Phe476. In the calculations using the active site of the flurbiprofen-complex the pharmacophore constraint set included interactions with Phe114, Arg108 and Asn204. Besides using the three aforementioned crystal structures, a model with extended binding site (screening site) was also created (Fig. 3). In the calculations using screening site of the pharmacophore constraint set included interactions with Phe100, Phe114, Phe476, Arg108 and Asn204 as well.

Virtual screens on CYP2C9

Virtual screens were carried out first to assess the dependence of enrichment factors on the active site conformation. Virtual screening against the ligand-free state resulted in a maximal enrichment factor of 3, calculated at 1% of the ranked database ($EF_{(1\%)}$) using FlexX/Chem or FlexX/PMF scores (Fig. 4). The performance of the virtual screen using the active site conformation derived from S-warfarin complex slightly improved (Fig. 3), with $EF_{(1\%)}$ of 5 (FlexX/Gold). Virtual screening against the conformation seen in the flubiprofen complex resulted in further modest improvement—$EF_{(1\%)}$ was found to be 7 when scoring by FlexX/Gold—but still not in acceptable enrichment (Fig. 4).

However, drastic improvements in enrichment and thus a reliable active site model could be achieved by use of our self-constructed screening site, combining the ligand coordinating features of the two complexed states. The best $EF_{(1\%)}$ using this model-system, was 76 (Fig. 4). FlexX scoring function for pose extraction and ranking resulted in the most efficient protocol. Chem score for pose extraction and FlexX scoring function for ranking provided the second best enrichment factor of 67.90% of ligands were ranked in the top 3% of the ranked database.

Next we investigated the influence of Arg108 on enrichment (Fig. 4). As Arg108 adopts significantly different conformations in complex X-ray structures and it was supposed to be crucial not only for enzyme activity but for selectivity [59, 60] we aimed at investigating its role in virtual screening.

When virtual screening with our optimal protocol was repeated such that Arg108 was not constrained as an H-bonding partner, the effectiveness drastically dropped ($EF_{(1\%)}$ was 0).

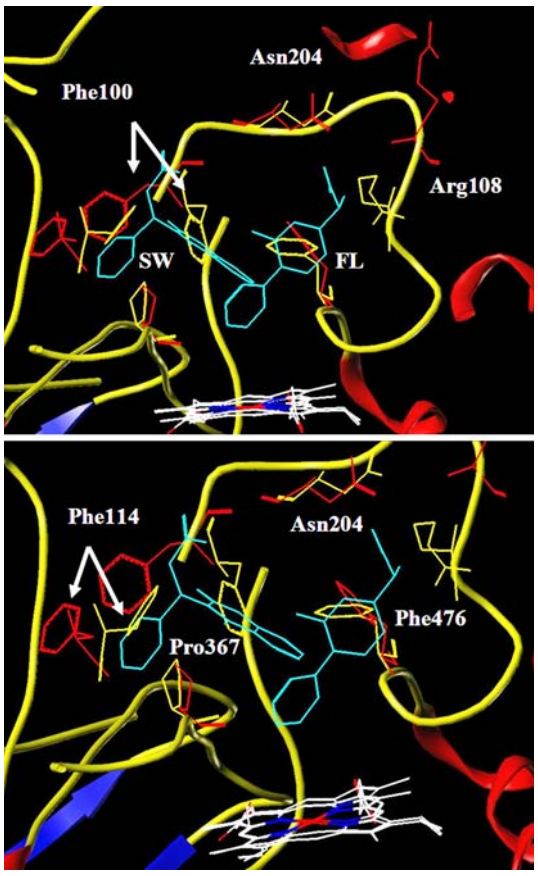

Fig. 3 Ribbon representation of the hybrid active site. Changed residues (Phe100, Phe114, Arg108, Pro367 and Phe476) are signed. Asn204 is also shown. Conformations in 1R90 are coloured yellow, while in 1OG5 are coloured red. Haem, S-warfarin (SW) and flubiprofen (FL) are also shown

Selectivity studies

CYP2C9, CYP2C18, CYP2C19 and CYP2C8 ligands collected from public sources were discriminated by the best

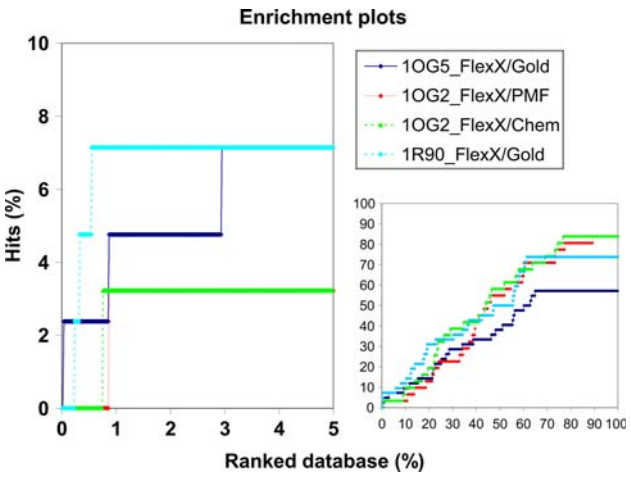

Fig. 4 Enrichment plots calculated for 1R90, 1OG5 and 1OG2 X-ray structures. Scoring functions, used for pose extraction and ranking respectively, are shown in the legend

virtual screening protocol using our CYP2C9 screening site (Fig. 5). Compounds binding more than one CYP2C isoform were omitted from this study. Chemical space of CYP2C sets on WDI compounds was analysed (Fig. 6) that revealed all CYP2C sets equally similar to WDI decoys.

During selectivity studies CYP2C8 ligands were effectively separated from CYP2C9 ligands. 19 (22%) of 86 CYP2C8 ligands and 97 (89%) of 108 CYP2C9 ligands were identified as CYP2C9 ligands (Table 1). Considering that only 22% of CYP2C8 ligands were identified as CYP2C9 ligands our virtual screening protocol resulted in a selectivity ratio of four for CYP2C9 ligands.

Selectivity studies provided somewhat lower efficiency when CYP2C18 or CYP2C19 ligands were docked into the screening site. 42% of CYP2C18 and 39% of the CYP2C19 ligands were classified as CYP2C9 ligands. Thus the selectivity ratio for CYP2C9 was about two in both cases.

Discussion

CYP2C9 is one of the principle members of human CYPs taking part in xenobiotic metabolism. Many CYP2C9 ligands exist as an anion (and/or possess hydrogen bond acceptor sites) and contain hydrophobic groups (including aromatic rings) between the polar hetero-atom and the site of metabolism [61, 62], therefore corresponding interaction partners should be present within the enzyme interior.

Various models have been developed to explain CYP2C9's preference for anionic substrates. In particular, Lys72, Arg97, Arg105, Arg108, and Arg241 were nominated as key cationic residues in the CYP2C9 active site based on different computational models [63, 64]. In site-directed mutagenesis experiments, substitution of Lys72

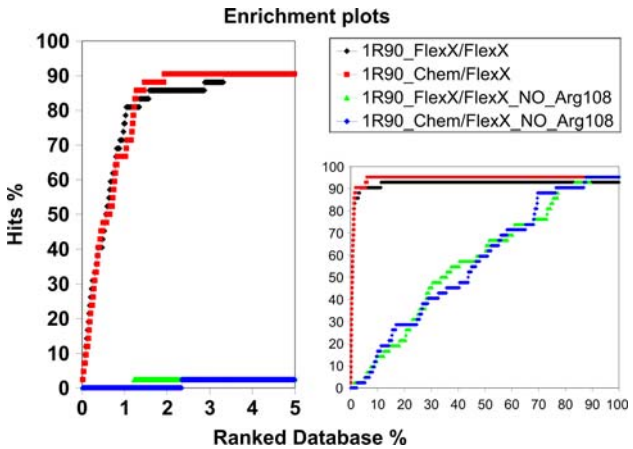

Fig. 5 For ligand retrieval Dock, Gold, Chem and FlexX scoring functions performed the best in pose extraction and FlexX ranked the poses best resulting an $EF_{(1\%)}$ of 76. Enrichment plots omitting interaction constraints for Arg108 (NO_Arg108) are also shown

and Arg105 had little or no effect on interaction with substrates such as diclofenac and ibuprofen [59, 65]. Mutation of both Arg97 and Arg241 resulted in drastically altered catalytic efficiency, however both were shown to play a key role in structure stabilization rather than in selective interaction with the substrates [60, 65]. On the other hand, Arg108, member of the B–C loop, directed away from the active site in homology models and the crystal structure of the ligand-free state as well as in the warfarin complex, was shown not only to be crucial for enzyme activity [59] but to participate in the selective binding of negatively charged substrates too [60]. Furthermore, Dickmann et al. found that the Arg108His mutant formed a nitrogen ligation to the haem which suggests that Arg108 also might sample the active site interior in the ligand-free state also, in contrast to that seen in the crystal structure [14, 15] and homology models.

The most important hydrophobic/aromatic complimentary sites of CYP2C9 ligands, as Phe114 and Phe476, at the entrance of the active site [18], were verified by a number of theoretical and mutational studies [51, 63, 64, 66–68].

Three crystal structures are currently available to rationalise the results of mutation experiments. In 2003, the crystal structures of both the native state of CYP2C9 and the structure of its complex with S-warfarin have been determined [27]. Binding of S-warfarin did not cause significant disturbance of the backbone structure of the enzyme. The positively charged amino acids within the active site (Arg97, Arg124 and Arg433) form numerous intra-molecular H-bonds with nearby protein segments and the haem group, in accordance with their proposed role in structure stabilization, however only a single H-bond was formed between the drug molecule and the host, the backbone amide NH of Phe100 coordinates to the negatively polarised oxygen of S-warfarin. The sidechain of Phe114 is flipped to allow the aromatic part of the drug molecule to reach within substrate recognition site 1 (SRS-1). Based on this, the authors speculate that the obtained arrangement corresponds to either an initial recognition or allosteric site for the ligand. Accordingly, Seifert and coworkers, based on a detailed molecular dynamics analysis, concluded that the crystal structure of the complex of S-warfarin with CYP2C9 describes a catalytically non-productive state where from the ligand has to move toward the haem for direct association [28].

The crystal structure of another CYP2C9 complex that with flurbiprofen has also been determined [29]. The most important difference between the two complex structures is the conformational rearrangement of the B–C loop (residues Gly98-Ile112) of the flurbiprofen complex with respect to both the ligand-free and the S-warfarin bound structure. This rearrangement causes the appearance of a new positively charged amino acid in the active region: Arg108 that

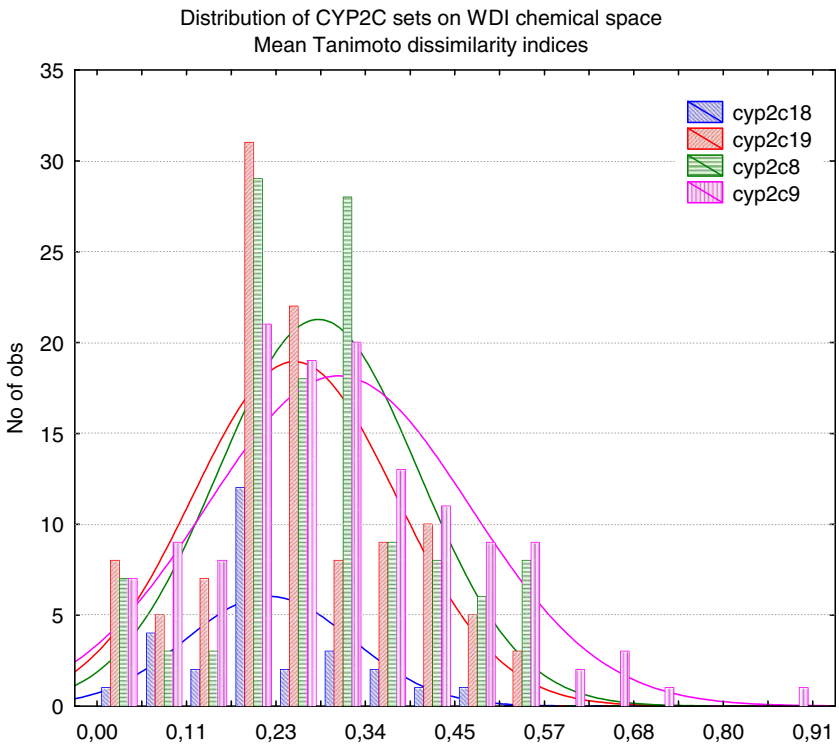

Fig. 6 Analysis of the distribution of the CYP2C sets on the WDI chemical space calculating mean Tanimoto dissimilarity indices

Table 1 Summary of the CYP2C ligand separation in the extended screening site achieved by FlexX-Pharm

	CYP2C9	Non CYP2C9
CYP2C8	19 (22%)	67 (78%)
CYP2C18	6 (42%)	8 (58%)
CYP2C19	30 (39%)	46 (61%)
CYP2C9	97 (89%)	11 (11%)

forms two H-bonds with flurbiprofen. Arg108 possesses a long side chain to be flexible enough to span considerable part of the ligand binding pocket, therefore can be well supposed to coordinate the negative charge centres of ligands of different topology also. Interestingly, the orientation of Phe114 in this structure is more similar to that seen in the ligand-free state of the enzyme than that of the S-warfarin bound—suggesting that the opening of SRS-1 is only required for the anchorage of certain but not all substrates.

Computational studies, based on the crystal structures, suggest that not only flurbiprofen but both diclofenac and lornoxicam establish H-bonds with Arg108 within the active site of CYP2C9 [69], and once again underline the significance of Phe114 positioning [18].

In the present work, high-throughput docking calculations were carried out using the crystal structures of the ligand-free state, the S-warfarin complex, the complex with flurbiprofen, and a model that unified structural features of the latter two.

Virtual screening protocols developed for 1R90, 1OG2 and 1OG5 were all ineffective. The reason for the inability of these virtual screening protocols in retrieving active compounds from the test database was assumed to be the improper conformation and/or definition of active sites and/or the inefficacy of pharmacophore constraints. Low ranks of compounds placed in the active site, separately either in S-warfarin or flubiprofen binding pocket, can be explained by incorrect docking solutions—i.e. compounds that might tend to bind in the S-warfarin (flubiprofen) binding pocket were placed in the flubiprofen (S-warfarin) site and vice versa, in these cases proper interactions could not be formed and low ranks were assigned to them. Since we found not only the conformation but also the definition of the active sites responsible for low enrichments a modified active site was created comprising all residues that interact with the ligand both in 1R90 and 1OG5. This unified active site (screening site) can bind S-warfarin and flubiprofen simultaneously. CYP2C9 ligands (inhibitors and substrates) were docked into the screening site with two subpockets, which improved the enrichment significantly. This also indicates that compounds can bind into two binding sites. Conformational rearrangements were needed to interact at the S-warfarin site highlight the importance of protein flexibility in virtual screening too.

To characterise the three X-ray based active sites in short, it could be said, that the ligand-free state represents the open conformation of the B–C loop and contains Phe114 in a conformation blocking the small pocket of

SRS-1, in the S-warfarin complex B–C loop is still in the open state, while Phe114 is switched so that SRS-1 becomes approachable. In the flurbiprofen complex the B–C loop moves toward the active site to participate in ligand coordination (binding conformation), while Phe114 is blocking SRS-1 again and Phe100 and Phe476 block part of the ligand-entrance channel, and in the screening site created by us, the B–C loop is in the binding conformation while Phe100, Phe114 and Phe476 leave both the entrance region and SRS-1 open.

Even though results improved somewhat as moving along the different model systems, effective and reliable screening protocol could only be deduced using our extended screening site—with $EF_{(1\%)}$ up to 76. The results also support that in CYP2C9 coordinating a wide variety of ligands, both Arg108 and the aromatic residue Phe114 play important role. Significance of Arg108 was confirmed in a calculation where—keeping the best obtained setup otherwise—this residue was withdrawn from the pharmacophore constraint set, which alteration practically diminished enrichment. The fact that omitting Arg108 from the constrained interaction list obstructed enrichment, indicate that a significant proportion of ligands must be directed toward the haem-close interior of the active site pocket by our virtual screen, since Arg108 is near-neighbours of the haem. On the other hand Arg108 could play an important role in positioning ligands so molecules could be placed more likely in incorrect orientations. However, presence of Arg108 within the active site in itself was not enough for satisfactory enrichment (see result obtained using the structure of the flurbiprofen complex), creating a more spacious active site by repositioning Phe100, Phe114 and Phe476 was also necessary which is in accordance with the work of Locuson et al. [71]. Phe100 and Phe476 are gatekeepers of the ligand entrance region while Phe114 was shown in numerous studies to interact with ligands.

Application of the extended screening site lead to a better ranking because more interactions could be formed within the active site and possibly favoured interactions could contribute to score values. Repositioning Phe100, Phe114, Pro367 and Phe476 most probably strengthened the effect of the latter one. Applying a pharmacophore model and restricting candidate molecules by interaction constraints decreased the number of possible poses and excluded molecules. Extending active site and restricting the orientation of molecules by pharmacophore constraints simultaneously resulted better ranking of active ligands compared to inactive molecules.

The conformational change caused by the binding of flurbiprofen is most evident in the unstructured B–C loop region (residues 98–112) of CYP2C9. When comparing the crystal structures of CYP2C9 with that of CYP2C8, the primary region of difference is again the same. These three structures, the ligand-free state of CYP2C9, the flurbiprofen complex of it and that of CYP2C8, represent three quite different conformations of this 14-residue sequence, thus providing a singular possibility for the study of their significance.

The active site of CYP2C8 is surrounded by Pro211, Trp212, Ile213, Asn214, Ile476 and a cluster of small, polar residues of Ser100, Ser103 and Ser114, members of the B–C loop. Of these, Ser114 not only provides a more spacious accommodation for substrates than the corresponding Phe residue in CYP2C9, but also seems to interact with them, since its mutation to neutral or aromatic residues abolishes enzyme activity [52, 70–72].

We restrict ourselves to only mention that a virtual screening protocol could not be developed for CYP2C8 ligand-free structure. In a large active site without a correct pharmacophore model and possibly inaccurate side-chain conformations rigid protein docking was not able to rank active ligands to give a high enrichment such in case of CYP2C9. This suggests that because of relevant alterations in corresponding residues of 14-amino acid sequence a pharmacophore model could not be deducted from CYP2C9-ligand complexes.

To further test our virtual screening protocols their abilities to discriminate between CYP2C8, CYP2C18, CYP2C19 and CYP2C9 ligands were also estimated. Noteworthy that score values computed for these ligands covered an overlapping score interval.

In CYP2C9 structure 89% of the CYP2C9 ligands were selected while only 22% of the CYP2C8 ligands were picked up indicating that the protocol could discriminate between CYP2C8 and CYP2C9 ligands. Our objective during these studies was to separate CYP2C9 ligands from other CYP2C ligands, which could be done successfully for CYP2C8 by applying pharmacophore constraints. That was in accordance with our expectations because virtual screening protocol was very effective in separation active and inactive molecules. However CYP2C18 and CYP2C19 ligands could not be separated from CYP2C9 ligands with similar efficacy. 42% of CYP2C18 and 39% of CYP2C19 ligands were identified as CYP2C9 ligands. Since CYP2C18 and CYP2C19 are closer relatives (Fig. 1, [37]) of CYP2C9 separation of these ligands is more intriguing. Besides our findings relative sequence identities (CYP2C9/CYP2C8: 0.82, CYP2C9/CYP2C18: 0.82, CYP2C9/CYP2C19: 0.93, [68]) further support that CYP2C8 is the most different member among the CYP2C enzymes. Corresponding residues of CYP2C9 Phe100 and Phe114 residues are Phe residues in CYP2C18, CYP2C19 and CYP2C8, while of CYP2C9 Arg108 is Lys in CYP2C8 and CYP2C18, and Arg in CYP2C19. Corresponding residue of CYP2C9 Phe476 is Ile in CYP2C8, while Phe again in CYP2C18 and CYP2C19. These residue's variations and

results of Ridderström et al. [68] support that selectivity could be modelled the best between CYP2C8 and CYP2C9. The lack of structural information for CYP2C isoforms other than CYP2C9 and neglecting protein flexibility in docking calculations might be both responsible for moderate discrimination of CYP2C18 and CYP2C19 ligands. Significant overlap between CYP2C binders observed experimentally further limits the discrimination power of in silico methods.

Conclusion

Results are consistent with that Arg108—identified by Williams et al. in the flurbiprofen complex [16]—and Phe114—described by Wester et al. in the warfarin complex [14] - are crucial interaction sites within the CYP2C9 haem pocket. The protein conformation represented by the crystal structure of the ligand-free state was found ineffective in hosting a wide range of ligands. However, using the structural information provided by the complex structures of CYP2C9 a highly useful representation of its ligand-binding state could be formulated that enabled the construction of selective screens for the enzyme. This in turn should greatly support the identification of CYP2C9 ligands.

Our results further support the existence of a multiple binding pocket that comprises at least the warfarin and flubiprofen binding pocket. This finding is also consistent with the two-site model of CYP2C9 [73, 74]. Applying this extended screening site 76% of the ligands were found in the top 1% of the ranked screening database. Furthermore, nearly 90% of CYP2C9 ligands were retrieved from the mixture of CYP2C9 and CYP2C8 ligands using pharmacophore filters. Lower efficiency of CYP2C18 and CYP2C19 separation suggests that further flexible side-chains should be involved to improve enrichment in virtual screening because of their higher similarity to CYP2C9. Our results also emphasized the importance of protein flexibility in docking calculations as adjustments in CYP2C9 side-chain conformations were needed to achieve this success rate.

In summary identifying a diverse set of molecules that are readily available for in vitro testing can benefit the study of CYP2C selectivity, drug-drug interactions or site of metabolism during drug design.

Acknowledgement The financial support from National Science Foundation (OTKA) under grant number T42933 is gratefully acknowledged. The authors gratefully thank Dr. Zoltán Kovári for commenting US patent applications filed on CYP2C9 crystal structure. DKM is a Bolyai research fellow.

References

1. Denisov IG, Makris TM, Sligar SG, Schlichting I (2005) Chem Rev 105:2253
2. Guengerich FP (2001) Chem Res Toxicol 14:611
3. Nebert DW, Gonzalez FJ (1987) Annu Rev Biochem 56:945
4. Evans WE, Relling MV (2004) Nature 429:464
5. Miners JO, Birkett DJ (1998) Br J Clin Pharmacol 45:525
6. Transon T, Leemann T, Vogt N, Dayer P (1995) Clin Pharmacol Ther 58:412
7. Poli-Scaife S, Attias R, Dansette PM, Mansuy D (1997) Biochemistry 36:12672
8. Hamman MA, Thompson GA, Hall SD (1997) Biochem Pharmacol 54:33
9. Miners JO, Coulter S, Tukey RH, Veronese ME, Birkett DJ (1996) Biochem Pharmacol 51:1003
10. Tracy TS, Marra C, Wrighton SA, Gonzalez FJ, Korzekwa KR (1996) Biochem Pharmacol 52:1305
11. Rettie AE, Korzekwa KR, Kunze KL, Lawrence RF, Craig Eddy A, Aoyama T, Gelboin HV, Gonzalez FJ, Trager WF (1992) Chem Res Toxicol 5:54
12. Thijssen HH, Flinois JP, Beaune PH (2000) Drug Metab Dispos 28:1284
13. Yamazaki H, Shimada T (1997) Arch Biochem Biophys 346:161
14. Williams PA, Cosme J, Ward A, Angove HC, Vinkovic DM, Jhoti H (2003) Nature 424:464
15. Totah RA, Rettie AE (2005) Clin Pharmacol Ther 77:341
16. Wester MR, Yano JK, Schoch GA, Yang C, Griffin KJ, Stout CD, Johnson EF (2004) J Biol Chem 279:35630
17. Johnson EF, Stout CD (2005) Biochem Biophys Res Commun 338:331
18. Clodfelter KH, Waxman DJ, Vajda S (2006) Biochemistry 45:9393
19. Schwarz UI (2003) Eur J Clin Invest Suppl 33:23
20. Yamazaki H, Inoue K, Chiba K, Ozawa N, Kawai T, Suzuki Y, Goldstein JA, Guengerich FP, Shimada T (1998) Biochem Pharmacol 56:243
21. Takanashi K, Tainaka H, Kobayashi K, Yasumori T, Hosakawa M, Chiba K (2000) Pharmacogenetics 10:95
22. Yasar U, Eliasson E, Forslund-Bergengren C, Tybring G, Gadd M, Sjoqvist F, Dahl ML (2001) Eur J Clin Pharmacol 57:729
23. Guo Y, Zhang Y, Wang Y, Chen X, Si D, Zhong D, Fawcett JP, Zhou H (2005) Drug Metab Dispos 33:749
24. Suzuki K, Yanagawa T, Shibasaki T, Kaniwa N, Hasegawa R, Tohkin M (2006) Diabetes Res Clin Pract 72:148
25. Voora CE, Linder MW, Milligan PE, Bukaveckas BL, McLeod HL, Maloney W, Clohisy J, Burnett RS, Grosso L, Gatchel SK, Gage BF (2005) Throm Haemost 93:700
26. Aithal GP, Day CP, Kesteves PJ, Daly AK (1999) Lancet 353:9154
27. Lewis DF, Dickins M (2002) Drug Discov Today 7:918
28. Schoch GA, Yano JK, Wester MR, Griffin KJ, Stout CD, Johnson EF (2004) J Biol Chem 279:9497
29. Bidstrup TB, Bjornsdottir I, Sidelmann UG, Thomsen MS, Hansen KT (2003) British J Clin Pharmacol 56:305
30. Rahman A, Korzekwa KR, Grogan J, Gonzalez FJ, Harris JW (1994) Cancer Res 54:5543
31. Projean D, Morin PE, Tu TM, Ducharme J (2003) Xenobiotica 33:841
32. Ong CE, Coulter S, Birkett DJ, Bhasker CR, Miners JO (2000) J Clin Pharmacol 50:573
33. Tracy TS, Korzekwa KR, Gonzalez FJ, Wainer IW (1999) Br J Clin Pharmacol 47:545

34. Becquemont L, Mouajjah S, Escaffre O, Beaune P, Funck-Brentano C, Jaillon P (1999) Drug Metab Dispos 27:1068
35. Muck W (1998) Drugs 56:15
36. http://www.cypalleles.ki.se
37. Chuang SS, Helvig C, Taimi M, Ramshaw HA, Collop AH, Amad M, White JA, Petkovich M, Jones G, Korczak B (2004) J Biol Chem 279:6305
38. Payne VA, Chang YT, Loew GH (1999) Proteins 37:204
39. Wang JF, Wei DQ, Li L, Zheng SY, Li YX, Chou KC (2007) Biochem Biophys Res Commun 355:513
40. Nguyet-Thanh HD, Dijols S, Marquaes-Soares C, Minoletti C, Dansette PM, Mansuy D (2001) J Med Chem 44:3622
41. Minoletti C, Dijols S Dansette PM, Mansuy D (1999) Biochemistry 38:7828
42. Oda A, Yamaotsu N, Hirono S (2004) Pharm Res 21:2270
43. de Groot MJ (2006) Drug Discov Today 11:601
44. Herz T, Wolf K, Kraus J, Kramer B (2006) Expert Opini Drug Metab Toxicol 2:471
45. Huwe CM (2006) Drug Discov Today 11:763
46. Lill MA, Dobler M, Vedani A (2006) Chem Med Chem 1:73
47. de Graaf C, Oostenbrink C, Keizers PH, van der Wijst T, Jongejan A, Vermeulen NP (2006) J Med Chem 49:2417
48. Ahlstrom MM, Ridderstrom M, Zamora I, Luthman K (2007) J Med Chem, DOI: 10.1021/jm0705096
49. Zhou D, Afzelius L, Grimm SW, Andersson TB, Zauhar RJ, Zamora I (2006) Drug Metab Dispos 34:976
50. Keseru GM (2001) J Comput Aided Mol Des 15:649
51. Zamora I, Afzelius L, Cruciani G (2003) J Med Chem 46:2313
52. Kerdpin O, Elliot DJ, Boye SL, Birkett DJ, Yoovathaworn K, Miners JO (2004) Biochemistry 43:7834
53. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson PJ (2004) J Chem Inf Comput Sci 44:793
54. Tripos Inc., SYBYL 7.3, 1699 South Hanley Road, St. Louis MO 63144-2319
55. Prous Integrity Database, Prous Science, http://www.prous.com. Accessed Feb 2007
56. Rendic S, Guengerich FP (2002) Drug Metab Rev 34
57. Stahl M, Rarey M (2001) J Med Chem 44:1035
58. Polgar T, Keseru GM (2006) In: Swarbrick J (ed) Encyclopedia of pharmaceutical technology, Taylor and Francis LLC, Dekker Encyclopedias, New York, pp 4013–4038
59. Ridderstrom M, Masimirembwa C, Trump-Kallmeyer S, Ahlefelt M, Otter C, Andersson TB (2000) Biochem Biophys Res Commun 270:983
60. Dickmann LJ, Locuson CW, Jones JP, Rettie AE (2004) Mol Pharmacol 65:842
61. Ekins S, De Groot MJ, Jones JP (2001) Drug Metab Dispos 29:936
62. Locuson CW, Wahlstrom JL, Rock DA, Rock DA, Jones JP (2003) Drug Metab Dispos 31:967
63. Payne VA, Chang YT, Loew GH (1999) Proteins 37:176
64. Rao S, Aoyama R, Schrag M, Trager WF, Rettie A, Jones JP (2000) J Med Chem 43:2789
65. Davies C, Witham K, Scott JR, Pearson A, DeVoss JJ, Graham SE, Gillam EMJ (2004) Drug Metabolism and Disposition 32:431
66. Melet A, Assrir N, Jean P, Lopez-Garcia MP, Marques-Soares C, Jaouen M, Dansette PM, Sari MA, Mansuy D (2003) Arch Biochem Biophys 409:80
67. Haining RL, Jones JP, Henne KR, Fisher MB, Koop DR, Trager WF, Rettie AE (1999) Biochemistry 38:3285
68. Ridderstrom M, Zamora I, Fjellstrom O, Andersson TB (2001) J Med Chem 44:4072
69. Zhou YH, Zheng QC, Li ZS, Zhang Y, Sun M, Sun CC, Si D, Cai L, Guo Y, Zhou H (2006) Biochimie 88:1457
70. Melet A, Marques-Soares C, Schoch GA, Macherey AC, Jaouen M, Dansette PM, Sari MA, Johnson EF, Mansuy D (2004) Biochemistry 43:15379
71. Locuson CW, Gannett PM, Ayscue R, Tracy TS (2007) J Med Chem 50:1158
72. Nelson DR, Zeldin DC, Hoffman SMG, Maltais LJ, Wain HM, Nebert DW (2004) Pharmacogenetics 14:1
73. Hutzler JM, Hauer MJ, Tracy TS (2001) Drug Metab Dispos 29:1029
74. Hutzler JM, Wienkers LC, Wahhlstrom JL, Carlson TJ, Tracy TS (2003) Arch Biochem Biophysics 410:16