Computational modelling of the binding of arachidonic acid to the human monooxygenase CYP2J2

G. Proietti[1,3]*, K.K. Abelak[2,3]*, D. Bishop-Bailey[2], A. Macchiarulo[1] & I. Nobeli[3$]

1. Department of Pharmaceutical Sciences, Università di Perugia, via del Liceo 1, 06123 Perugia, Italy.

2. Comparative Biomedical Sciences, Royal Veterinary College, Royal College Street, London NW1 0TU, UK.

3. Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck, Malet Street, London WC1E 7HX, UK.

* These two authors have contributed equally to this work

$ Corresponding author email: i.nobeli@bbk.ac.uk

telephone: +44 (0) 203 073 8014

fax: +44 (0)20 7 631 6803

Abbreviations

CYP:	Cytochrome P450
CYP2J2:	Human cytochrome P450 2J2
AA:	Arachidonic Acid
EET:	Epoxyeicosatrienoic Acid
MD:	Molecular Dynamics
IFD:	Induced Fit Docking

Abstract

An experimentally determined structure for human CYP2J2, a member of the cytochrome P450 family with significant and diverse roles across a number of tissues, does not exist yet. Our understanding of how CYP2J2 accommodates its cognate substrates and how it might be inhibited by other ligands relies thus on our ability to computationally predict such interactions using modelling techniques. In this study we present a computational investigation of the binding of arachidonic acid (AA) to CYP2J2 using homology modelling, induced fit docking (IFD) and molecular dynamics (MD) simulations. Our study reveals a catalytically competent binding mode for AA that is distinct from a recently published study that followed a different computational pipeline. Our proposed binding mode for AA is supported by crystal structures of complexes of related enzymes to inhibitors and evolutionary conservation of a residue whose role appears essential for placing AA in the right site for catalysis.

Keywords:

CYP2J2; Arachidonic Acid; Homology Model; Induced Fit Docking; Molecular Dynamics

Introduction

Cytochrome P450s (CYPs) form a large and diverse family of monooxygenase enzymes involved in the metabolism of both exogenous and endogenous substrates[1]. CYP2J2, the only member of the CYP2J subfamily in humans, is the primary source of arachidonic acid (AA)-derived epoxygenase products 5,6-, 8,9-, 11,12- and 14,15- epoxyeicosatrienoic acids (EETs) in human cardiac tissue[2]. In addition to the heart, the CYP2J2 mRNA is present in the liver, kidney and skeletal muscle, and to a lesser extent, in the small intestine, pancreas, lung and brain[3]. Moreover, CYP2J2 expression can be induced in endothelial cells[4] and monocytes[5] by bacterial lipopolysaccharide. CYP2J2 is also expressed in a wide variety of tumours and tumour cell lines, where it appears to promote cell survival[6]. CYP2J2 and the EETs it produces have been shown to regulate the inflammatory response, vascular tone, cellular proliferation, angiogenesis, and metabolism[3]. In addition to AA, CYP2J2 can also metabolise the structurally related linoleic, docosahexaenoic and eicosapentaenoic acids to epoxy-products with biological activity[7, 8].

In addition to the metabolism of fatty acids, similar to many CYPs, CYP2J2 can accept a wide range of xenobiotic substrates for detoxification[9-11]. Using screens of marketed therapeutic agents, albendazole, amiodarone, cyclosporine A, danazol, mesoridazine, nabumetone, tamoxifen, thioridazine, telmisartan and flunarizine were all shown to be potential xenobiotic substrates for CYP2J2. Moreover, terfenadine[12, 13] and ebastine[13] have been used as the parent compounds to generate structure-based CYP2J2 inhibitors with K(i) values as low as 160 nM[12, 13].

The roles of CYP2J2 in the physiology and patho-physiology of cardiovascular and tumour biology along with drug metabolism, have led to a quest for greater understanding of its structure. There is currently no crystal structure of CYP2J2. In the absence of an experimentally determined crystal structure, homology models represent an invaluable tool to progress our understanding of a protein's structure-function relationship[14]. Indeed, homology models of proteins that elude crystallisation efforts (or have simply not yet been characterised) have contributed significantly to our understanding of their substrate specificity[15] and mode of action[16], and facilitated structure-based ligand/inhibitor design to target these proteins[17, 18].

Homology models of CYP2J2 have been built in previous studies (Supplementary Table 1) using a variety of related structures as templates. Lafite *et al*.[19] used structures of CYPs 2A6, 2B4, 2C5, 2C8 and 2D6 with resolutions ranging from 2.05 to 3.00 Å. The templates were chosen to provide the greatest diversity in the family as well as in the occupancy of the active site (from no ligand to large ligands present). Li *et al*.[20] used a structure of CYP2C9 bound to warfarin as a template (resolution: 2.55Å). However, the template used carried significant mutations in the F-G region and a thorough investigation of changes to substrate binding was not completed prior to crystallisation[21], raising doubts about its ability to represent the wild type protein. Lee *et al*.[9] created a model based on structures for CYPs 2B4, 2C8, and 2A6 at an average resolution of 2.1Å and their analyses focused on accessibility of a variety of ligands to the active site. Cong *et al*.[22] created a model based on the 2.8Å resolution structure of CYP2R1 and concentrated on mutation-induced changes to AA binding based on known polymorphisms. Finally, more recently, Xia *et al*.[23] created a model based on CYP2A6, CYP2E1,

CYP17A1, CYP2R1 and CYP2C8 as templates (resolutions ranging from 2.2Å to 2.8Å). The model was used to dock AA, and polymorphisms known to reduce the metabolism of AA were explained based on the binding mode observed in the docking study.

In the study presented here we have investigated the binding of AA to CYP2J2 using a modelling approach that is also based on the principles of homology modelling and docking simulations. However, our strategy of selecting promising binding poses from induced fit docking and using them as starting points for 50ns-long molecular dynamics (MD) simulations differs significantly from existing studies and our results are also distinct. The only two modelling studies that have examined the binding of AA to CYP2J2 (Xia *et al*.[23] and Cong *et al*.[22]) were partially carried out by the same authors and followed a very similar protocol in the building of their homology model. Cong *et al*. carried out in addition a 10ns MD simulation, but the main conclusions of the two studies are analogous. Both have resulted in a protein-ligand complex, where AA is tethered at the far end of the binding site channel by hydrogen bonds to Leu378 and Gly486. This model of binding relies on a hydrogen bond to the main chain of CYP2J2. We viewed this model with scepticism, as we believe that the need for specificity for AA would be better served by one or more residues anchoring it in place through side-chain interactions. To investigate further, we have built our own homology model and carried out both docking and MD simulations of the binding of AA to CYP2J2.

We present here an alternative binding mode for AA to the one that has been suggested in the literature. This binding mode is supported by existing crystal structure complexes of proteins of the same family bound to inhibitors of diverse structures. Differences in our model and previously suggested ones are reflected in the

distinct list of residues that we put forward as being important for recognition, and, hence, being most interesting for testing with mutagenesis experiments. In the absence of relevant experimental data, it is difficult to ascertain which model, if any, is right. Hence, we believe that until relevant experimental data become available, it is worth keeping in mind that binding modes distinct from the ones published so far are not only possible, but also plausible.

Methods

All computational modelling tasks, including protein preparation, docking and MD simulations were carried out using tools available in the Schrödinger molecular modelling software suite (www.schrodinger.com).

Generation of a 3D model of CYP2J2

The sequence of CYP2J2 was extracted from the Uniprot (www.uniprot.org) database entry with accession code P51589. This sequence was used to search the RCSB Protein Data Bank[24] (www.rcsb.org) using the 'Search by Sequences' facility. The selection of the template for the homology model was based on the following considerations: a) the percentage sequence identity between CYP2J2 and putative templates, b) the resolution of the crystal structures returned by the search, c) the presence and type of ligand bound in the template, and d) the completeness of the template protein chain (in terms of missing or mutated residues). Only proteins with sequence identity to the query higher than 35% were considered. Rabbit CYP2B4 (PDB ID: 1suo[25]) was selected as the template for homology modelling based on a combination of relatively high sequence identity to CYP2J2 (39%), completeness of sequence, and high resolution of the corresponding crystal structure (1.9Å). This

structure (PDB ID: 1suo) has not been used in the past for the construction of a homology model of CYP2J2, despite the fact that it has comparable sequence identity and higher resolution among most structures used so far to build models. The sequence alignment between rabbit CYP2B4 and human CYP2J2 (Supplementary Figure 1) justifies the use of the first as a template for modelling the second. The ligand bound to CYP2B4 in the crystal structure is 4-(4-chlorophenyl)imidazole, a compact ligand by comparison to AA. A model based on this structure would have a relatively small active site that would close in further during protein preparation for docking, as the protein side chains would tend to move closer to one another to maximize interactions. To allow the active site to remain open, we carried out protein preparation in the presence of a larger ligand, vitamin D3 (all three ligands mentioned in this study are depicted in Supplementary Figure 2). This ligand was positioned in the CYP2J2 active site by superimposing our model with the crystal structure of CYP2R1 (PDB ID: 3c6g[26]), a human homologue crystallised with vitamin D3. Once the model CYP2J2 was optimized for docking, vitamin D3 was removed to leave a large, open active site ready for docking AA.

The template structure (PDB ID: 1suo) was prepared using the Protein Preparation Wizard tool[27]. Water molecules were deleted whilst the original ligand (CPZ) and heme molecules were retained. Prime[28] was used to generate the wild type homology model. The model was refined using the Protein Preparation Wizard and minimised using Prime Refinement. Analysis of the generated model with Verify3D[29] identified amino acids that were poorly modelled (Verify3D score was less than 0.2). These amino acids were then used as starting points for refinement using the Prime "refinement with proximity" approach. This protocol involves two cycles of minimisation: the first cycle involves amino acids proximal to the selected

ones (starting at 1Å) and the second involves a total minimisation of the whole structure. The proximity cut-off was increased by 2Å in each cycle, resulting in a total of 51 cycles for this protein. The minimised structures after each cycle were analysed with Verify3D and the best structure (in this case the result of the final cycle) was chosen.

The final model was assessed using the following structure assessment programs: Verify3D, ERRAT[30], PROCHECK[31], QMEAN[32]. Verify3D, ERRAT and PROCHECK were run on the server of the Molecular Biology Institute at the University of California, Los Angeles (http://services.mbi.ucla.edu). QMEAN score values were obtained from the QMEAN server at http://swissmodel.expasy.org/qmean.

Generation of a complex between CYP2J2 and AA

The 2D structure of AA was taken from the PubChem Compound[33] database and it was processed using Schrodinger's LigPrep[34] tool. Ionisation states were assigned by Epik[35] at physiological pH (7.0). Parameters were assigned using the OPLS2005 force field[36].

AA was docked into the active site of the CYP2J2 model using the Induced Fit Docking Protocol (IFD[37]). Briefly, the first stage of this protocol scales down the van der Waals radii of both protein and ligand atoms by a factor of 0.5, and ligands are then docked into the fixed receptor using the Glide SP docking protocol. Next, Prime is used to predict the optimal orientation of the side chains of residues lining the binding site. Finally, the ligand is re-docked into the optimised binding site and poses are scored with Glide XP. A cubic box of $10Å^3$ was used for docking, centred on the original co-crystallised ligand in the template structure used for homology

modelling. Twenty top-scoring poses were kept following IFD. The poses were viewed and five poses were manually selected as diverse starting points for MD simulations.

Molecular dynamics simulation of AA binding to CYP2J2

MD simulations were carried out using the Schrödinger interface of the Desmond program. Five MD runs were started from the five poses selected following induced fit docking. The protein-ligand complex was solvated in a cube of TIP3P waters, extended 15Å away from any protein atom. The resulting system was then neutralised with the addition of sodium and chlorine ions at a concentration of 0.15M. Periodic boundary conditions were applied to avoid finite-size effects. Atomic partial charges of ligands were maintained as obtained from OPLS2005 force field. MD simulations were performed using Desmond v.2.4 and the OPLS2005 force field. The simulation protocol included starting relaxation steps and a final production phase of 50 ns at 300K.

In particular, the starting relaxation steps comprised an initial minimisation of the system over a maximum 2000 steps, with a convergence criterion of 50 kcal/mol/Å, and the presence of harmonic restraints on the solute atoms (force constant = 50.0 kcal/mol/$Å^2$); a second minimisation without restraints; a third stage of 12ps at 10K with harmonic restrains on the solute heavy atoms (force constant = 50.0 kcal/mol/$Å^2$), using NVT ensemble and Berendsen thermostat; a fourth 12 ps at 10K, retaining the harmonic restraints, and using NPT ensemble and Berendsen thermostat and barostat; a fifth heating phase of 24 ps at 300 K, retaining the

harmonic restraints, and using NPT ensemble and Berendsen thermostat and barostat; a final 24 ps at 300K without harmonic restraints, using the NPT Berendsen thermostat and barostat. After the relaxation steps, a 50ns MD simulation was carried out at 300K using canonical NPT Berendsen ensemble. During MD simulations, we used a time step of 2 fs while constraining the bond lengths of hydrogen atoms with the M-SHAKE algorithm[38]. The atomic coordinates were saved every 5 ps. Following simulations, the Schrödinger Simulations Interaction Diagram (SID) was employed for analysis. SID is a post-MD analysis tool available in the Schrödinger environment for the easy visualisation and analysis of protein-ligand interactions derived from MD simulations.

Preparation of figures

Figures 1, 3, 4c, 5 have been prepared using the software Chimera[39]. Where a superposition of two or more structures was necessary, Chimera's Match Maker was used in the global alignment mode.

Figure 2 was prepared using output from the programs PROCHECK, QMEAN and ERRAT.

Figures 4a&b, 6, 7, were prepared using the Schrödinger molecular modelling software.

All additional formatting of figures and addition of labels was done using the free image editor Gimp (www.gimp.org).

Results

Generation and validation of a homology model for human CYP2J2 from rabbit CYP2B4

A homology model of human CYP2J2 was built using the rabbit CYP2B4 (PDB ID: 1suo) as template. The model retained the overall fold of CYPs with 12 major helices and four beta sheets (Figure 1). The structural quality of the model was confirmed by a range of computational tools (Figure 2). The Ramachandran plot for the model (Figure 2a) showed good stereochemistry for over 90% of the residues (97.8% of residues are in "most favoured" and "additional allowed" regions, and only two of the 463 residues modelled were in disallowed regions). Verify3D reported 93.3% of residues as having an average 3D-1D score >=0.2. The calculated QMEAN score for this model was 0.77, and the associated Z-score was -0.03. The QMEAN score is an estimate of the "degree of nativeness" of the structural features in the model, and, as shown in Figures 2b&c, this model displayed "native-like" features, when compared with a high quality reference dataset of structures. ERRAT, which uses statistics of non-bonded interactions in the model and compares them to those of high quality crystal structures, produced an overall quality factor of 77.1%, meaning that 77.1% of all residues had calculated error values below the 95% rejection limit (Figure 2d). All these quality metrics are comparable with those reported in publications of existing models of CYP2J2. In addition, our model was compared to those listed in ModBase[40] as well as others built automatically by the Protein Model Portal[41] (Supplementary Table 2). The results show our model to be of comparable, or superior quality (not unexpectedly, given the automated nature of these model builders). Furthermore, the models listed in Supplementary Table 2 lack the heme

moiety whilst our method preserves the co-factor in the final model. Co-factor exclusion or incorrect placement is a major drawback of using automated modelling servers, currently only circumvented by manual inclusion.

Finally, the active site volume of our model was calculated to be 234Å^3 using the SiteMap[42] method. This value is similar to the volumes found by Xia *et al*.[23] (330Å^3) and Cong *et al*.[22] (320Å^3) but notably smaller than those reported by Lafite *et al*.[19] (945Å^3) and Lee *et al*.[9] (1420Å^3). There are several potential reasons for this discrepancy. The models may differ significantly in their conformation, but we cannot ascertain this, as we have no access to these models and there are no coordinates deposited in a public repository. However, other more technical reasons may also be responsible for the differences observed. Programs differ widely in their definition of a "cavity" and the corresponding calculation of its volume. For example, the inclusion or exclusion of a solvent accessible channel leading to the heme whilst calculating the volume of the binding site could justify the differences between the reported values in the literature. In addition, the inclusion of heme itself could alter the results significantly due to the large volume occupied by this cofactor. Indeed, calculations of the active site volume using *fpocket*[43] or the cavities analysis from the PDBsum[44] server (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) give much larger estimates for our model (1339Å^3 and 1977Å^3 respectively), but it is clear from Supplementary Figure 3 that this is due to merging of cavities that are forming a channel towards the heme. Moreover, we show in Supplementary Figure 3 that a comparison of the active site volumes of different models may be less relevant, given that the movement of side

chains during the MD run allows the sampling of conformations with considerable variation in the active site volume, at least as perceived by software.

Induced fit docking suggests that AA is tethered to the CYP2J2 active site by a hydrogen bond to Arg117

Following preparation of both the small molecule ligand and the protein, AA was docked into the active site of the CYP2J2 model using a flexible receptor protocol (Induced Fit Docking or IFD[37], as described in the Methods). The twenty resulting poses of AA following the IFD protocol are depicted (superimposed) in Figure 3. In contrast to previously identified poses involving a hydrogen bond to the main chain atoms of Leu378, all our IFD poses showed the AA carboxylic group facing away from Leu378, and interacting with Arg117 instead. Hydrogen bonds to both side and main chain atoms of Arg117 were observed for the top ranking poses (Figure 3b) but there was small variation in the positioning of the carboxylic acid among the lower-ranking of the 20 IFD poses, leading to slightly different interaction profiles with the protein (see examples of protein-ligand interaction diagrams in Supplementary Figure 4). A hydrogen bond to the Arg117 side-chain was observed in 19 cases (the 20th only just failing the cut-off for the minimum acceptor angle), and in 11 of the 20 poses the Arg117 backbone was also involved in a hydrogen bond with AA. Hence, Arg117 plays a major role in stabilizing the observed AA poses. Moreover, Met116 appeared to be playing a supportive role in the recognition of AA, as in just under half the poses (8 of 20), the Met116 backbone was also hydrogen-bonded to the AA carboxylate tail. In addition to the residues mentioned above, a channel to the active site was lined by a network of mostly hydrophobic residues, creating a favourable environment for the

long non-polar AA chain. The following ten residues were in contact with the substrate in all 20 poses from IFD: Thr114, Phe121, Ile127, Asp307, Phe310, Ala311, Glu314, Thr315, Ile376 and Ile487, while there were several others that appeared in at least half the poses (e.g. Pro115, Ile120, Val218, Glu222, Trp251).

Molecular dynamics simulations of AA bound to CYP2J2 confirm the importance of Arg117 as a tether for the substrate

Five MD simulations were run (two for 30 ns and three for 50 ns) starting with five manually selected and diverse poses from the IFD runs. Here we concentrate on the analysis of one of the 50 ns runs (referred to here as run 2), which resulted in the most promising (from a catalysis point of view) and most stable (especially during the last nanoseconds of the simulation) position for the binding of AA to CYP2J2. We occasionally refer to the remaining MD runs in this section, where we believe it is instructive to do so.

The conformation of the protein did not change drastically during the simulation (Figure 4a). Although an increase in the average RMSD values for the protein C-alpha atoms was observed around the 35th ns of the simulation, the RMSD values stayed within the 3Å range, and so were within what might be expected from thermal fluctuations of a small globular protein. Amino acids with the largest fluctuations during the simulation are highlighted in the Root Mean Square Fluctuation (RMSF) plot (Figure 4b) and map to highly flexible loops on the protein structure (Figure 4c), confirming that the increase in the average RMSD was not due to major restructuring of parts of the protein, but to expected fluctuations of flexible loops on the surface. By contrast, amino acids involved in the binding of the ligand

appeared stable during the simulation and exhibited, as expected, the lowest fluctuations (Figure 4b). AA displayed a similar stability throughout the simulation, indicative of a narrow binding site that allows only restrained conformational movements (Figure 4a).

In the final frame of the 50ns simulation, AA was positioned in a catalytically competent orientation within the small hydrophobic tunnel on top of the heme molecule (Figure 5a). Comparison with the final frames from the remaining 4 simulations (Supplementary Figure 5a) highlight a number of ligand-binding poses that potentially represent AA on the route to approaching the heme. Ligand-only RMSD values between the AA atoms of these four last frame snapshots and the AA atoms of the last frame of run 2 vary from 3.4 to 9.3Å. In run 2 the ω9 double bond positioned above the iron at a distance of approximately 3.5Å in a plane parallel to the heme (Figure 5b&c). This distance fluctuated between 3 and 4.5Å throughout most of the simulation (Figure 6a), a range consistent with distances previously described for a heme-catalysed epoxidation reaction[45]. Literature data on the regioselectivity of this enzyme indicates epoxidation rates of 37% for ω6 vs 18% for ω9[3]. Although this suggests a small preference for the ω6 bond, our simulation was not sensitive enough to reproduce this result. However, we did observe ω6 approaching the iron atom at around 30 ns, as is evident from the distance plot in Figure 6b. ω9 showed the opposite profile, with its distance to iron increasing significantly in the same timeframe (Figure 6a). Finally, we also monitored the ω6 and ω9 distances from iron for two of the remaining four simulations (runs 3 and 7; the other two stay too far from the heme for this distance to be meaningful). During both these simulations there are time slots when ω9 is at a reasonable distance (between 3 and 4Å) from the iron

atom but for most of the time the distances are longer than 4.5Å and so unlikely to represent poses representative of a catalytic event (Supplementary Figure 5b).

The interaction of protein residues with AA during the 50ns of the simulation are summarised in Figure 7. A number of hydrophobic residues that line the binding site stabilise the ligand in place (Phe310, Ile127, Ala311, Ile375, Ile376 and others). As the simulation progresses and the complex equilibrates, these interactions become stronger (at about 25 ns) and the ligand appears to stabilise further (Figure 4a). Similarly to what was seen in the induced-fit docking poses, Arg117, Met116 and Thr114 are involved in hydrogen bonds to the small molecule substrate, anchoring it in place. The most persistent hydrogen bonds between AA and CYP2J2 were formed by Arg117 (53.8% and 44.4% of the simulation time the main-chain and side-chain nitrogens respectively were involved in a hydrogen bond to the ligand). Thr114 was involved in an equally persistent hydrogen bond through its side-chain (43.4% of the time), whereas the hydrogen bond to the Met116 main chain nitrogen was observed less than half of that time (17.7%) in this simulation. In the other two MD runs where the ligand approached the heme (runs 3 and 7), hydrogen bonds were again observed predominantly to Arg117, with one simulation (run 3) showing persistent hydrogen bonds to the main chain (76.2%) whereas the other (run7) showed interactions with the side-chain nitrogens (50.7% and 19.9%). Hydrogen bonds to the main-chain of Met116 were only persistent in run 3 (46.2%) and almost absent in run 7 (4.6%), whereas interactions with the Thr114 side chain were present in both runs for significant periods of time (run3: 25.3%; run7: 29.9%).

Discussion

In this study we present an investigation into the binding of AA to CYP2J2 based on a homology model of the enzyme structure, induced fit docking of AA to the enzyme and a 50ns MD simulation that starts from realistic poses identified at the IFD stage. Our model structure of CYP2J2 is of high quality, as assessed by several tools (Verify3D, ERRAT, PROCHECK, and QMEAN). The core of the structure, and importantly, the access channel, is similar to the ones reported in recent studies[19],[23],[46] and lined in each case by the same hydrophobic residues (or a subset thereof). The convergence of models to the same core structure with a common access channel increases our confidence in our model. Where our model differs significantly from the pose suggested by Cong *et al*[22] is in the suggested mode of recognition of the substrate AA. The difference appears to originate at least partially by the assignment of a protonated state to the AA carboxylate group. The pKa of AA is 4.82 (http://www.drugbank.ca/drugs/DB04557) making it likely to be mostly unprotonated at physiological pH. In our study, AA is unprotonated and hence able to take advantage of the strong hydrogen bonding capabilities of Arg117. Further evidence supporting the validity of our model is presented below.

A number of crystal structures of complexes of CYP2 with inhibitors support the binding mode for AA suggested by our study. In the structure of the inhibitor troglitazone bound to human CYP2C8 (PDB ID: 2vn0[47]), the inhibitor's position overlaps with the region occupied by the carboxylic acid of AA and interacts with Ser103, the residue corresponding to Arg117 in the sequence alignment of the two enzymes (Supplementary Figure 6a). In the same structure, Thr364, which corresponds to Leu378 in CYP2J2, shows no interactions with the co-crystallised inhibitor. Other ligands in complex with related cytochrome P450s also appear to bind

closer to residues equivalent to Arg117 than to residues equivalent to Leu378 (e.g. abiraterone bound to CYP17A1 (PDB ID: 3ruk[48]), vitamin D3 bound to CYP2R1 (PDB ID: 3c6g[49]), or bifonazole bound to CYP2B4 (PDB ID: 2bdm[50]) – see Supplementary Figure 6b). There are, of course, crystal structures of CYP450s with inhibitors bound in different sites. Notably, the second bifonazole molecule bound to CYP2B4 is positioned closer to the Leu378 side of the pocket but further away from the heme moiety, and so likely represents a less catalytically relevant pose of the ligand. Similarly, warfarin bound to CYP2C9 (PDB ID: 1og5[51]) is positioned closer to Thr364 (the residue equivalent to Leu378) than to the Arg117 equivalent. However, warfarin is a bulky ligand with a very different shape to the flexible AA, and so differences between its binding and that of AA's are not surprising.

Several docking studies further support our model. Lafite *et al.*[19] used their CYP2J2 model for docking the inhibitor terfenadone and its derivatives and found that the keto group of terfenadone is held in place through hydrogen bonding interactions with the guanidine moiety of Arg117. Additionally, the hydrophobic part of terfenadone was accommodated in the narrow tunnel leading to the heme through a series of contacts to the same residues that we have identified as important in holding the hydrophobic part of AA. In the same study, three derivatives of terfenadone presented a similar interaction profile with CYP2J2. Likewise, Li *et al.*[20] used a CYP2J2 model for docking studies of four ligands sharing the same scaffold: ebastine, terfenadine, terfenadone and a terfenadone derivative. Arg117 formed a hydrogen bond to ebastine and, although it did not hydrogen bond to the other three ligands, it was listed as one of the residues contributing significant amounts of energy to the interaction. Similarly, in the Ren *et al*.[46] study (where the Li *et al*. homology model was used), inhibitors telmisartan and flunarizine bind in a pocket that includes

Arg117, and the orientation of telmisartan's carboxylic group after 2 ns of MD simulations suggests that a longer simulation might include poses that hydrogen bond to Arg117. Interestingly, telmisartan binds far from the heme group, yet it is a potent inhibitor of CYP2J2. In our model, blocking access of the substrate to Arg117 leads to the loss of an important hydrogen bond, which should make it easier for a ligand to compete for binding the enzyme.

Finally, a model where AA interacts with Arg117 is supported by evolutionary conservation of this residue. In the absence of a manually curated dataset of CYP2J2 enzymes from other organisms, we obtained an alignment of all members of the same protein family from Ensembl (protein family ID: PTHR24300_SF91). This is a broadly defined family with members that share the same Enzyme Commission number (E.C. 1.14.14.1) but not necessarily the same substrates. Examining the family sequence alignment, it appears at first that Arg117 is conserved in this family (61.8% conservation within a group of 165 proteins) but to a lesser extent than Leu378, which shows 80% conservation. However, on closer inspection, the majority of sequences with a residue other than arginine at this position either have much shorter sequences and are thus unlikely to have the same function, or they originate from Uniprot/TrEMBL entries where functional annotation is either not present or assigned by sequence similarity only. Removing these sequences that are unlikely to be true orthologues, leaves a group of 64 sequences where arginine is prevalent (68% conserved) but glutamine is also common at this position. Restricting the selection further to 15 sequences that both carry the label CYP2J2 in Ensembl and share at least 70% sequence identity at the gene level (both for the target and the query sequences), reveals that Arg117 is conserved in all but one sequence (where it is replaced by glutamine – see Supplementary Figure 7). Although it is hard to speculate about the

significance of these findings, given the overall high sequence identities shared by these protein homologues, there appears to be some evolutionary pressure for conserving a good hydrogen bond donor at this position. Perhaps more interesting are the results of comparing CYP2J2 with its rat homologue CYP2J3, which shares 73.9% sequence identity with the human enzyme (see alignment in Supplementary Figure 8). CYP2J3, like CYP2J2, metabolises AA to EETs[52]. In both CYP2J2 and CYP2J3, Arg117 and the amino acids of the narrow hydrophobic tunnel to the heme, are conserved. In the rat enzyme, Met116 is substituted by a leucine, a very conservative substitution. In contrast, Leu378 and Gly486, the two residues suggested in the previous study[22] as interacting with AA are replaced in CYP2J3 by phenylalanine and serine respectively, despite the very high conservation of the two enzyme sequences. Although the Cong *et al*. model predicts an interaction of AA with the backbone of Leu378 and Gly486, which would lessen the impact of mutations in the sequence, conservation of Arg117 in a homologue known to bind AA is consistent with an important functional role for this residue.

Conclusions

Our homology model of CYP2J2, in combination with IFD calculations and 50 ns MD simulations have led us to suggest a new binding mode for AA, that differs significantly from the recent suggestions in the literature. The main difference in our model is the residue employed to anchor the hydrophilic tail of AA in place (primarily Arg117). The support for our results is two-pronged: first, the similarity of our proposed binding mode to that of the binding of inhibitors CYP2J2 in the literature; and second, the conservation of Arg117 and conservative substitution of Met116 in an

enzyme in rat that performs the same reaction. Until relevant experimental data become available, we believe it is worth keeping in mind this alternative mode of binding, especially when designing mutagenesis experiments.

Funding

GP was funded for this work by an Erasmus Placement Grant (2013/2014) received from the University of Perugia. KKA is funded by a Bloomsbury Colleges PhD studentship.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Bishop-Bailey D, Thomson S, Askari A, et al (2014) Lipid-metabolizing CYPs in the regulation and dysregulation of metabolism. Annu Rev Nutr 34:261–279. doi: 10.1146/annurev-nutr-071813-105747

2. Wu S, Moomaw CR, Tomer KB, et al (1996) Molecular cloning and expression of CYP2J2, a human cytochrome P450 arachidonic acid epoxygenase highly expressed in heart. J Biol Chem 271:3460–3468. doi: 10.1074/jbc.271.7.3460

3. Xu M, Ju W, Hao H, et al (2013) Cytochrome P450 2J2: distribution, function, regulation, genetic polymorphisms and clinical significance. Drug Metab Rev 45:311–352. doi: 10.3109/03602532.2013.806537

4. Askari AA, Thomson S, Edin ML, et al (2014) Basal and inducible anti-inflammatory epoxygenase activity in endothelial cells. Biochem Biophys Res Commun 446:633–637. doi: 10.1016/j.bbrc.2014.03.020

5. Bystrom J, Thomson SJ, Johansson J, et al (2013) Inducible CYP2J2 and its product 11,12-EET promotes bacterial phagocytosis: a role for CYP2J2 deficiency in the pathogenesis of Crohn's disease? PLoS ONE 8:e75107. doi: 10.1371/journal.pone.0075107

6. El-Serafi I, Fares M, Abedi-Valugerdi M, et al (2015) Cytochrome P450 2J2, a new key enzyme in cyclophosphamide bioactivation and a potential biomarker

for hematological malignancies. Pharmacogenomics J 15:405–413. doi: 10.1038/tpj.2014.82

7. Zeldin DC (2001) Epoxygenase pathways of arachidonic acid metabolism. J Biol Chem 276:36059–36062. doi: 10.1074/jbc.R100030200

8. Westphal C, Konkel A, Schunck W-H (2011) CYP-eicosanoids—A new link between omega-3 fatty acids and cardiac disease? Prostaglandins & Other Lipid Mediators 96:99–108. doi: 10.1016/j.prostaglandins.2011.09.001

9. Lee CA, Neul D, Clouser-Roche A, et al (2010) Identification of novel substrates for human cytochrome P450 2J2. Drug Metab Dispos 38:347–356. doi: 10.1124/dmd.109.030270

10. Liu K-H, Kim M-G, Lee D-J, et al (2006) Characterization of ebastine, hydroxyebastine, and carebastine metabolism by human liver microsomes and expressed cytochrome P450 enzymes: major roles for CYP2J2 and CYP3A. Drug Metab Dispos 34:1793–1797. doi: 10.1124/dmd.106.010488

11. Askari A, Thomson SJ, Edin ML, et al (2013) Roles of the epoxygenase CYP2J2 in the endothelium. Prostaglandins & Other Lipid Mediators 107:56–63. doi: 10.1016/j.prostaglandins.2013.02.003

12. Lafite P, Dijols S, Zeldin DC, et al (2007) Selective, competitive and mechanism-based inhibitors of human cytochrome P450 2J2. Arch Biochem Biophys 464:155–168. doi: 10.1016/j.abb.2007.03.028

13. Lee CA, Jones JP, Katayama J, et al (2012) Identifying a selective substrate and inhibitor pair for the evaluation of CYP2J2 activity. Drug Metab Dispos 40:943–951. doi: 10.1124/dmd.111.043505

14. Du H, Brender JR, Zhang J, Zhang Y (2015) Protein structure prediction provides comparable performance to crystallographic structures in docking-based virtual screening. Methods 71:77–84. doi: 10.1016/j.ymeth.2014.08.017

15. Lukk T, Sakai A, Kalyanaraman C, et al (2012) Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. Proc Natl Acad Sci USA 109:4122–4127. doi: 10.1073/pnas.1112081109

16. Ramachandran S, Dokholyan NV (2012) Homology Modeling: Generating Structural Models to Understand Protein Function and Mechanism. In: Computational Modeling of Biological Systems. Springer US, Boston, MA, pp 97–116

17. Carlsson J, Coleman RG, Setola V, et al (2011) Ligand discovery from a dopamine D3 receptor homology model and crystal structure. Nat Chem Biol 7:769–778. doi: 10.1038/nchembio.662

18. Kannan S, Melesina J, Hauser A-T, et al (2014) Discovery of inhibitors of Schistosoma mansoni HDAC8 by combining homology modeling, virtual screening, and in vitro validation. J Chem Inf Model 54:3005–3019. doi:

10.1021/ci5004653

19. Lafite P, André F, Zeldin DC, et al (2007) Unusual regioselectivity and active site topology of human cytochrome P450 2J2. Biochemistry 46:10237–10247. doi: 10.1021/bi700876a

20. Li W, Tang Y, Liu H, et al (2008) Probing ligand binding modes of human cytochrome P450 2J2 by homology modeling, molecular dynamics simulation, and flexible molecular docking. Proteins 71:938–949. doi: 10.1002/prot.21778

21. Williams PA, Cosme J, Ward A, et al (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature 424:464–468. doi: 10.1038/nature01862

22. Cong S, Ma X-T, Li Y-X, Wang J-F (2013) Structural basis for the mutation-induced dysfunction of human CYP2J2: a computational study. J Chem Inf Model 53:1350–1357. doi: 10.1021/ci400003p

23. Xia X-L, Fa B-T, Cong S, et al (2014) Research/review: Insights into the mutation-induced dysfunction of arachidonic acid metabolism from modeling of human CYP2J2. Curr Drug Metab 15:502–513.

24. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242.

25. Scott EE, White MA, He YA, et al (2004) Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. J Biol Chem 279:27294–27301. doi: 10.1074/jbc.M403349200

26. Strushkevich N, Usanov SA, Plotnikov AN, et al (2008) Structural analysis of CYP2R1 in complex with vitamin D3. J Mol Biol 380:95–106. doi: 10.1016/j.jmb.2008.03.065

27. Schrödinger, LLC Schrödinger Suite 2014-1 Protein Preparation Wizard; Epik version 2.7, Schrödinger, LLC, New York, NY, 2013; Impact version 6.2, Schrödinger, LLC, New York, NY, 2014; Prime version 3.5, Schrödinger, LLC, New York, NY, 2014.

28. Schrödinger, LLC Prime, version 3.5, Schrödinger, LLC, New York, NY, 2014.

29. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85. doi: 10.1038/356083a0

30. Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Science 2:1511–1519. doi: 10.1002/pro.5560020916

31. Laskowski RA, MacArthur MW, Moss DS, et al (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291. doi: 10.1107/S0021889892009944

32. Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. Nucleic Acids Res 37:gkp322–W514. doi: 10.1093/nar/gkp322

33. Kim S, Thiessen PA, Bolton EE, et al (2016) PubChem Substance and Compound databases. Nucleic Acids Res 44:D1202–13. doi: 10.1093/nar/gkv951

34. Schrödinger, LLC **Schrödinger Release 2014-1**: LigPrep, version 2.9, Schrödinger, LLC, New York, NY, 2014.

35. Schrödinger, LLC **Schrödinger Release 2014-1**: Epik, version 2.7, Schrödinger, LLC, New York, NY, 2014.

36. Banks JL, Beard HS, Cao Y, et al (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comput Chem 26:1752–1780. doi: 10.1002/jcc.20292

37. Sherman W, Day T, Jacobson MP, et al (2006) Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem 49:534–553. doi: 10.1021/jm050540c

38. Kräutler V, van Gunsteren WF, Hünenberger PH (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J Comput Chem 22:501–508. doi: 10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V

39. Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. doi: 10.1002/jcc.20084

40. Pieper U, Webb BM, Dong GQ, et al (2014) ModBase, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 42:D336–46. doi: 10.1093/nar/gkt1144

41. Haas J, Roth S, Arnold K, et al (2013) The Protein Model Portal—a comprehensive resource for protein structure and model information. Database (Oxford) 2013:bat031–bat031. doi: 10.1093/database/bat031

42. Halgren T (2007) New Method for Fast and Accurate Binding□site Identification and Analysis. Chemical Biology & Drug Design 69:146–148. doi: 10.1111/j.1747-0285.2007.00483.x

43. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. BMC Bioinformatics 10:168. doi: 10.1186/1471-2105-10-168

44. Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res 29:221–222.

45. Bhattarai S, Niraula NP, Sohng JK, Oh T-J (2012) In-silico and In-vitro based studies of Streptomyces peucetius CYP107N3 for oleic acid epoxidation. BMB

Reports 45:736–741. doi: 10.5483/BMBRep.2012.45.12.080

46. Ren S, Zeng J, Mei Y, et al (2013) Discovery and characterization of novel, potent, and selective cytochrome P450 2J2 inhibitors. Drug Metab Dispos 41:60–71. doi: 10.1124/dmd.112.048264

47. Schoch GA, Yano JK, Sansen S, et al (2008) Determinants of cytochrome P450 2C8 substrate binding: structures of complexes with montelukast, troglitazone, felodipine, and 9-cis-retinoic acid. J Biol Chem 283:17227–17237. doi: 10.1074/jbc.M802180200

48. DeVore NM, Scott EE (2012) Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. Nature 482:116–119. doi: 10.1038/nature10743

49. Strushkevich N, Usanov SA, Plotnikov AN, et al (2008) Structural Analysis of CYP2R1 in Complex with Vitamin D3. J Mol Biol 380:95–106. doi: 10.1016/j.jmb.2008.03.065

50. Zhao Y, White MA, Muralidhara BK, et al (2006) Structure of microsomal cytochrome P450 2B4 complexed with the antifungal drug bifonazole: insight into P450 conformational plasticity and membrane interaction. J Biol Chem 281:5973–5981. doi: 10.1074/jbc.M511464200

51. Williams PA, Cosme J, Ward A, et al (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature 424:464–468. doi: 10.1038/nature01862

52. Wu S, Chen W, Murphy E, et al (1997) Molecular cloning, expression, and functional significance of a cytochrome P450 highly expressed in rat heart myocytes. J Biol Chem 272:12551–12559. doi: 10.1074/jbc.272.19.12551

Figure Captions

Fig. 1 Homology model of human CYP2J2

a & b. Ribbon representation of front (a) and back (b) views of the human CYP2J2 model. The 12 major helices are coloured in the rainbow depiction shown in Figure 1 of Cong *et al*.[22] The rest of the structure is shown in grey.

c. Superposition of the crystal structure of the template structure (rabbit CYP2B4; PDB ID: 1suo; grey) onto the homology model of human CYP2J2 (orange).

Fig. 2 Quality assessment of our CYP2J2 homology model

Assessment of our CYP2J2 homology model by PROCHECK (A), QMEAN (B&C) and ERRAT (D).

a. A Ramachandran plot confirms good stereochemical quality of our model of CYP2J2. Results from PROCHECK: residues in most favoured regions: 81.1%; residues in additional allowed regions: 16.7%; residues in generously allowed regions: 1.7%; residues in disallowed regions: 0.5%.

b. The QMEAN score of our model shown in red against the background of scores of high quality crystal structures of similar size. The calculated Z-score of -0.03 confirms a good degree of "nativeness" of structural features in our model.

c. Density plot of QMEAN scores for all reference models used in the calculation of the Z-score in plot (b). The score for our model is shown by the red line on the plot.

d. ERRAT plots of the error values associated with each residue in the homology model.

Fig. 3 Induced fit docking of arachidonic acid to CYP2J2 suggests a binding mode anchored primarily by a hydrogen bond to Arg117

a. Superposition of 20 poses from IFD of AA to CYP2J2 show AA positioned to hydrogen bond to the Arg117 side-chain, and away from Leu378. All 20 poses are superimposed on the heme cofactor (only one copy of heme is shown), with AA

depicted in dark grey wireframe, and heme and the protein residues Arg117 and Leu378 depicted as sticks.

b. Top-ranking pose from IFD (Glide GScore -11.2 kcal/mol) depicts AA anchored in its place by hydrogen bonds to both the side-chain and main-chain nitrogens of Arg117. AA is depicted in stick mode (only heavy atoms shown), whereas Arg117 is shown in an all-atom representation. Hydrogen bonds are shown as cyan lines.

c. Six of the twenty IFD poses with the most diverse positions for the AA carboxylic acid are depicted in this figure. Only AA and the heme molecule from these six poses are shown.

Fig. 4 Protein and ligand molecules remain relatively stable during 50ns of MD simulation

a. Time-dependent root mean square deviation (RMSD) of the protein C-alpha carbons (blue) and ligand (purple) atoms during the time of the MD simulation. The first frame is used as the reference for calculating RMSD values. The ligand RMSD is measured for the ligand heavy atoms after superimposing the protein-ligand complex on the protein backbone of the reference structure. Thus, the ligand RMSD in this case reflects how stable the ligand is with respect to the protein and its binding site.

b. Residue-specific root mean square fluctuations (RMSF) during the trajectory of the MD simulation. Highest-fluctuating residues are labelled on the plot. The original numbering of the residues can be obtained by adding 39 to the numbers on the x-axis. The labels on the peaks correspond to correct (canonical) labelling of CYP2J2.

c. Superposition of 10 structures extracted every one nanosecond from the period 35-45 ns of the MD simulation. Highly flexible residues (labelled and highlighted in orange) are, as expected, all in loops facing the solvent and away from the heme (sticks coloured by heteroatom) and arachidonic acid (blue sticks) in the middle of the enzyme.

Fig. 5 AA adopts a catalytically competent position above the heme group during the MD simulation

a. A primarily hydrophobic tunnel accommodates AA in a catalytically competent position above heme (only the last frame of the MD simulation is shown here). Protein residues shown are within 3Å of any atom of AA (shown as yellow sticks). The protein residues are coloured by a hydrophobic scale that ranges from red (very hydrophobic) to blue (very hydrophilic).
b. & c. Side (b) and top (c) views of the CYP2J2–AA complex, as seen in the last frame of the 50ns MD simulation. Distances from the heme iron to the carbon atoms of the ω9 double bond are shown in (b).

Fig. 6 The ω9 and ω6 bonds of AA are swapping positions during the MD simulation confirming the possibility of epoxidation at both bonds

The distance between the heme iron and either the ω9 (a) or ω6 (b) bonds of AA are plotted against the time of the simulation. Whilst most of the time we observe ω9 at a distance ready for catalysis, at approximately 30ns the ligand flips and ω6 becomes the bond approaching the heme in a catalytically competent distance.

Fig. 7 Protein-ligand contacts during the 50ns simulation
a. Hydrogen bonds (green) and hydrophobic interactions (purple) between protein and ligand atoms are monitored during the MD simulation and summarised in this plot. The height of the bars indicates the time that a given protein residue spends being involved in a type of interaction with the ligand. Met316, for example, is associated with a bar of approximate height 1, indicating that it is involved in a hydrogen bond across the full trajectory. Heights greater than 1 (such as the one seen for Arg117) are due to multiple contacts of the same type and of the same protein residue with a ligand molecule.

b. Protein-ligand contacts are monitored across the time of the simulation. Contacts by specific protein residues per frame of the trajectory are depicted in the bottom part with orange lines indicating a contact and a darker shade of orange indicating more than one contact per residue. The total number of contacts for all protein residues is monitored in the top panel (blue lines). Arg117, Thr114 and Met116 stand out as the residues with the highest number of contacts across the time of the simulation.

Figure 1

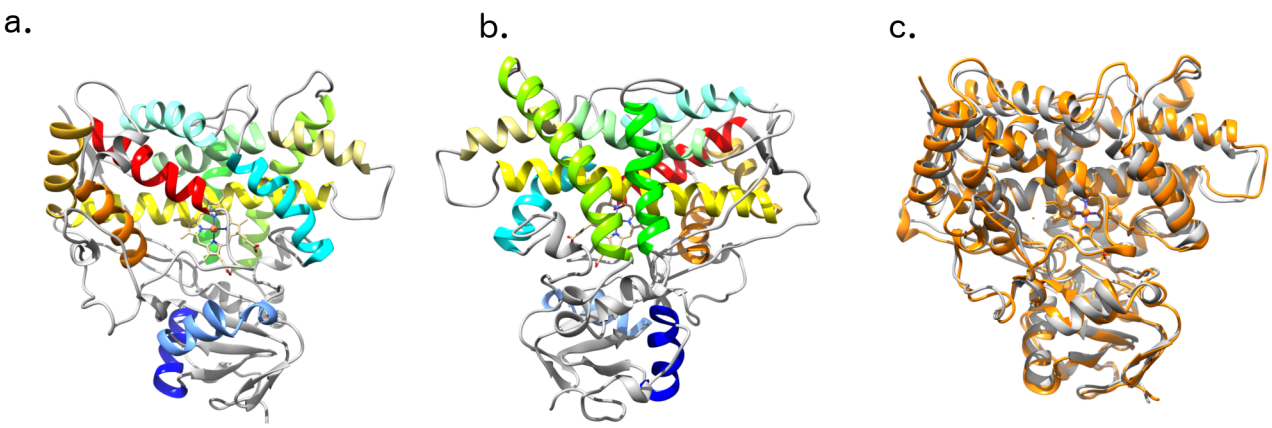


Figure 2

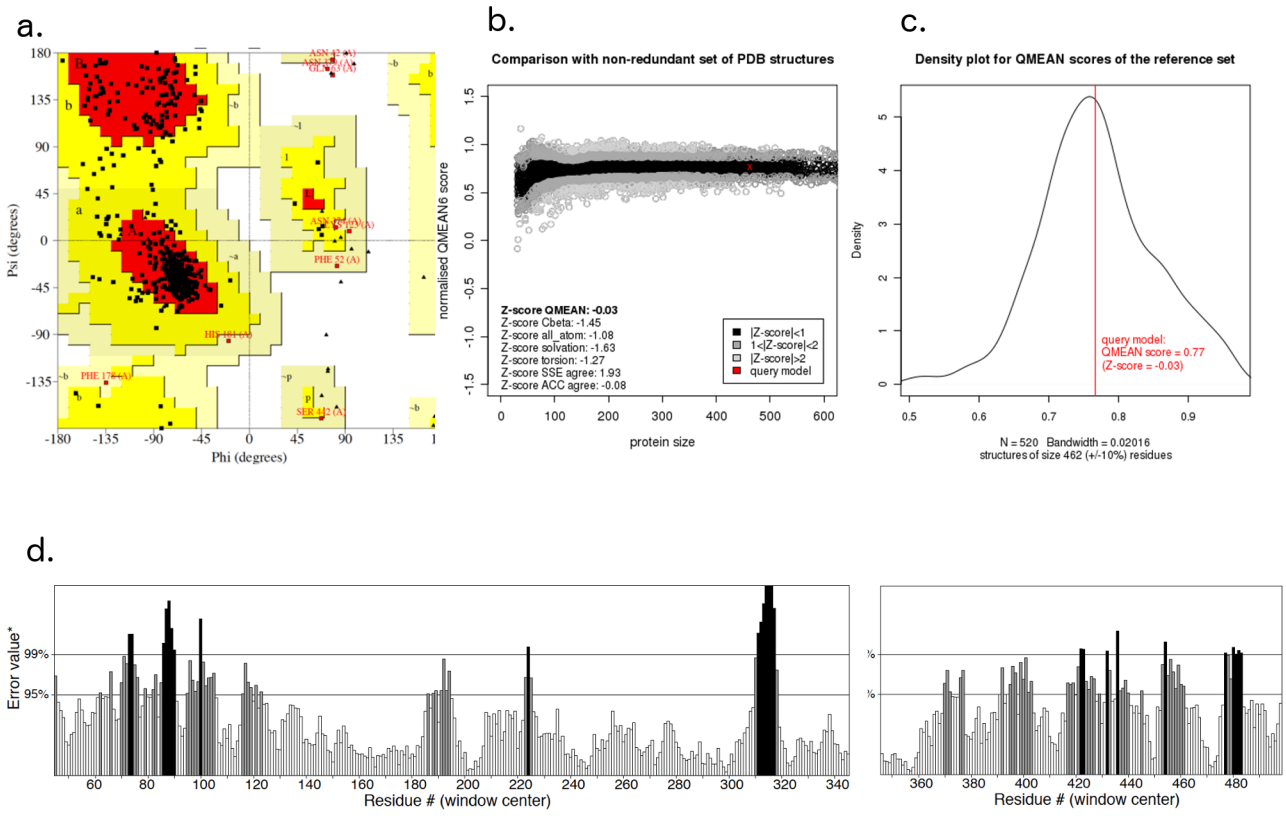

a.
b.
c.
Psi (degrees)
Phi (degrees)
normalised QMEAN6 score
Comparison with non-redundant set of PDB structures
Z-score QMEAN: -0.03
Z-score Cbeta: -1.45
Z-score all_atom: -1.08
Z-score solvation: -1.63
Z-score torsion: -1.27
Z-score SSE agree: 1.93
Z-score ACC agree: -0.08
|Z-score|<1
1<|Z-score|<2
|Z-score|>2
query model
protein size
Density plot for QMEAN scores of the reference set
query model:
QMEAN score = 0.77
(Z-score = -0.03)
Density
N = 520 Bandwidth = 0.02016
structures of size 462 (+/-10%) residues
d.
Error value*
99%
95%
Residue # (window center)
Residue # (window center)

Figure 3

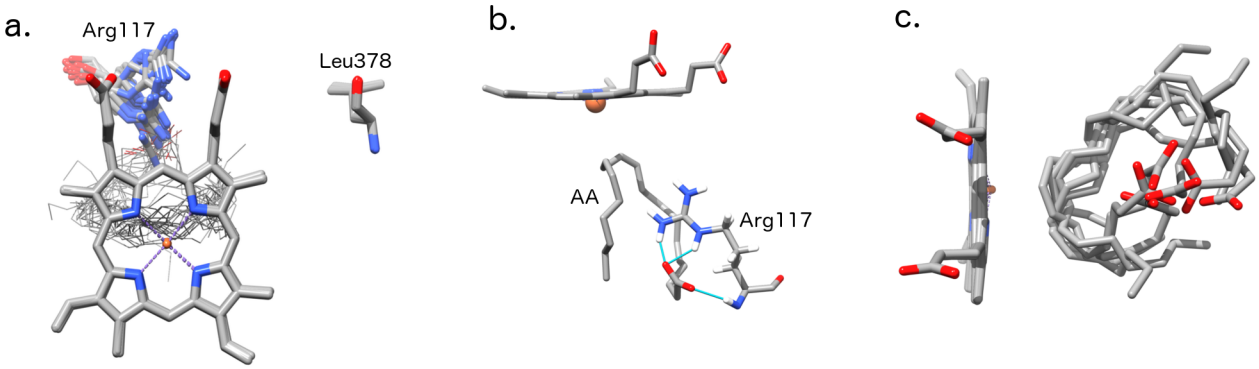
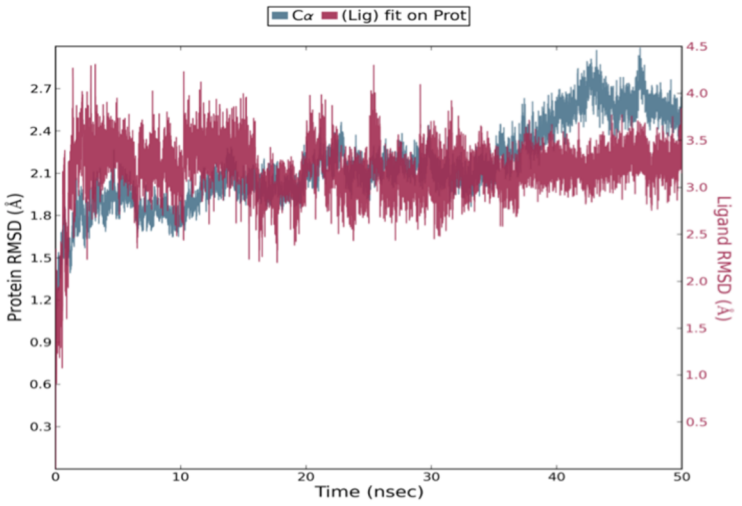
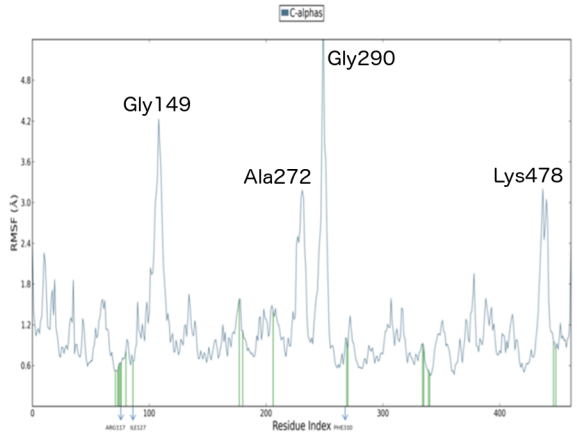


Figure 4

a.


Cα (Lig) fit on Prot
Protein RMSD (Å)
Ligand RMSD (Å)
Time (nsec)

b.


C-alphas
Gly290
Gly149
Ala272
Lys478
RMSF (Å)
Residue Index

c.

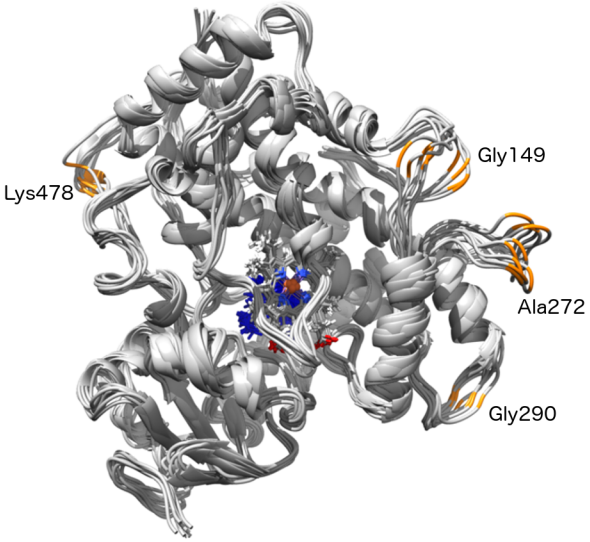

Gly149
Lys478
Ala272
Gly290

Figure 5

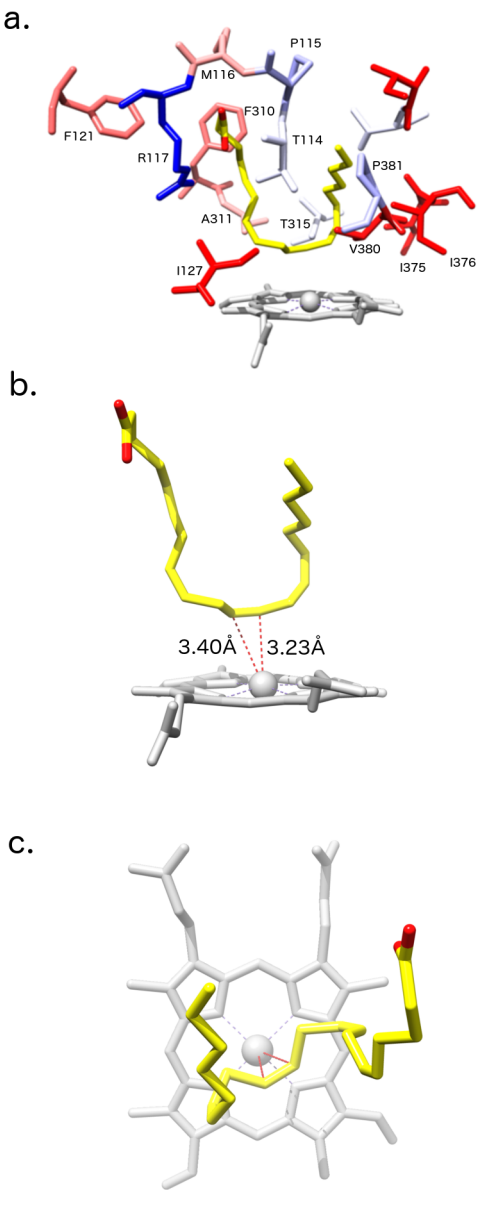
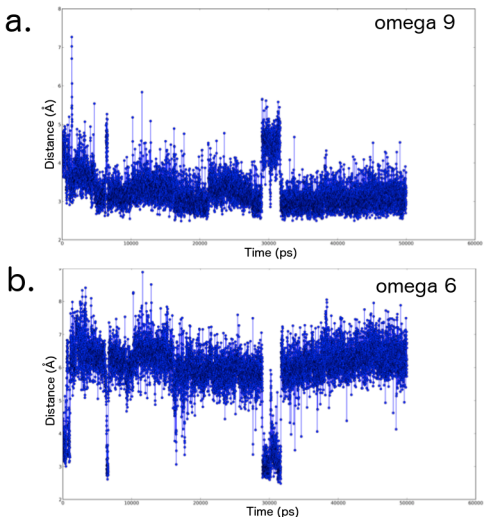
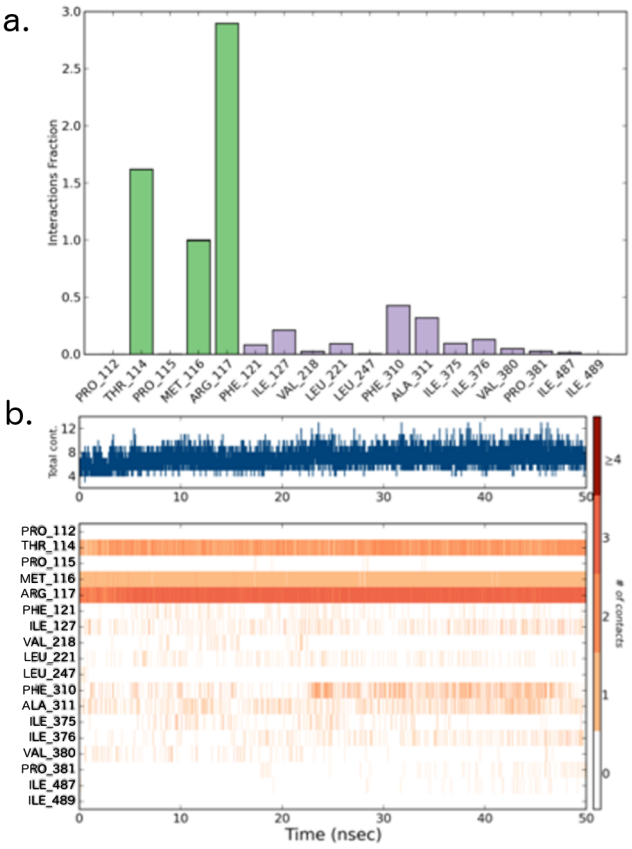

a.
P115
M116
F310
F121
T114
R117
P381
A311
T315
V380
I375
I376
I127
b.
3.40Å
3.23Å
c.

Figure 6




a.
Interactions Fraction
3.0
2.5
2.0
1.5
1.0
0.5
0.0
PRO_112
THR_114
PRO_115
MET_116
ARG_117
PHE_121
ILE_127
VAL_218
LEU_221
LEU_247
PHE_310
ALA_311
ILE_375
ILE_376
VAL_380
PRO_381
ILE_487
ILE_489
b.
Total cont.
12
8
4
0
10
20
30
40
50
≥4
3
2
1
0
# of contacts
Time (nsec)