

NOTE: The following document describes the methods used to produce the data uploaded in this repository (Zenodo, doi: 10.5281/zenodo.3465884) and all the ones that supplement it. All parts of this document are extracted (with minor editing) from Kavin Abelak's PhD thesis.*

* Abelak, K. K. (2018). Molecular mechanisms for eicosanoid metabolism by fatty acid metabolising cytochrome P450 enzymes. PhD thesis. The Royal Veterinary College, University of London.

1.1 Homology Modelling

The protein sequence of human CYP2J2 (UID: P51589) was obtained from UNIPROT (<http://www.uniprot.org/>). The sequence was submitted to BLAST-P (<http://blast.ncbi.nlm.nih.gov/>; McWilliam *et al.*, 2013) for a search against the Protein Data Bank (PDB) structure database (Berman, 2000). Sequences with more than 40% sequence identity were selected and representative PDB structures were filtered to discard those with low resolution and mutations affecting folding. Of the 21 total matches with >40% sequence identity, 14 were discarded due to lack of ligand bound (1PO5, 2PG6, 2PG5, 2PG7, 1DT6, 1OG2, 2F9Q) and/or resolution worse than 2.5Å (2Q6N, 3TK3, 4H1N, 2PG6, 2PG7, 2PG5, 3C6G, 1PQ2, 1DT6, 3E4E, 1OG2, 4GQS, 3QM4, 2F9Q) and three had mutations affecting the protein fold (4MGJ, 3IDB and 1R9O). The remaining templates (1SUO, 2P85, 3EBS and 1Z10) were chosen for further steps.

1.1.1 Modelling wild type CYP2J2

The sequence of CYP2J2 was trimmed of the N-terminal transmembrane domain residues 1-43 prior to modelling in order to avoid spurious alignments due to the lack of this region in crystal-structure-derived sequences. The four selected PDB structures 1SUO, 2P85, 3EBS and 1Z10 were run through a MODELLER (<http://salilab.org/modeller/>; (Eswar *et al.*, 2006)) script personalised to automate PDB download, alignment and model creation (MOD_salign.py). The intermediate alignment file produced by MODELLER was re-submitted to ClustalW2 for a better multiple sequence alignment (MSA) using standard parameters and "NBRF/PIR" as output format. MODELLER makes use of the combined spatial restraints provided by the 3D structures and applies it locally to the target sequence based on the MSA. Thus, the best quality structural data is used at each position in the sequence. The MSA file was further modified to force the inclusion of the haem HETATM record from 1SUO during modelling.

1.1.2 Modelling CYP2J2 Mutants

The final model of CYP2J2 was loaded into UCSF Chimera (Pettersen *et al.*, 2004) and the *swapaa* function was run to change residues 111, 117, 382 and 446 from arginine to alanine to remove the charge on the side chain. Alanine was chosen over the other hydrophobic amino acids for small size and lack of branching, while glycine was disregarded due to the extra flexibility afforded by the lack of any side chain.

1.2 Docking

1.2.1 Wild type CYP2J2

Docking was carried out using AutoDOCK VINA (Trott and Olson, 2009). Coordinates for the protein were carried forward from MD minimisation using mol2 format while the structure of AA was obtained from the Zinc Dock database (Irwin & Shoichet 2005; substance ref: 12504416). As the pKa of AA is 4.82 (<https://www.drugbank.ca/drugs/DB04557>) the unprotonated form is most likely to occur at physiological pH (7.4), and thus this was the form of AA used herein. A grid box of dimensions 20x20x20Å³ was centred approximately 4Å above the plane of the haem iron and the VINA algorithm was run using random seed and exhaustiveness = 8. All AA C-C bonds except double bonds were considered rotatable while a rigid docking procedure was applied to the protein, meaning all CYP2J2 atoms were fixed.

The structures of unprotonated DHA and EPA were derived using Automated Topology Builder (Malde *et al.*, 2011). The pKa of DHA and EPA were 4.89 (<https://www.drugbank.ca/drugs/DB03756>) and 4.82 (<https://www.drugbank.ca/drugs/DB00159>) respectively, making them likely to be unprotonated at physiological pH. Additionally, a flexible docking approach was carried out. Residues within 12Å radius of the centre of the grid box but excluding those below the plane of the haem were allowed to move whilst keeping the protein backbone fixed. The docking for each ligand was run five times with exhaustiveness = 50 for flexible docking, and exhaustiveness = 2,000 for rigid docking.

1.2.2 CYP2J2 Mutants

The previously docked poses from section 1.2.1 were maintained while the appropriate residue(s) were mutated as per section 1.1.2 above.

1.3 Molecular Dynamics

Molecular dynamics simulations were carried out in order to investigate the interactions in the active site between substrate ligands and CYP2J2 with haem bound. A list of the simulations carried out is given below:

- Arachidonic acid in the active site of wild type CYP2J2 (24 simulations of 1 μ s each corresponding to 4 repeat runs for each simulation starting from one of 6 low-energy docked poses).
- Arachidonic acid in the active site of mutant CYP2J2. 4 mutant structures were investigated (Arg111Ala, Arg117Ala, double mutant Arg111Ala and Arg117Ala, quadruple mutant Arg111Ala + Arg117Ala + Arg382Ala + Arg446Ala). For each mutant, 18 simulations in total were carried out: 3 repeats for each simulation starting from the same 6 low-energy docked poses as the wild type, each lasting 500 ns.
- DHA and EPA in the active site of wild type CYP2J2 (12 simulations of 300 ns each corresponding to 3 repeat runs for each simulation starting from one of 4 low-energy docked poses).

1.3.1 Setup

MD simulations were carried out using AMBER14 and AmberTools15 (Salomon-Ferrer *et al.*, 2013) with the ff14SB force-field (Maier *et al.*, 2015) and TIP3P water model (Jorgensen, 1982). Force-field parameters for haem and the cysteine residue coordinated to the haem iron were obtained from QM/MM calculations carried out by Shahrokh *et al.* (2012). The parameters represented the haem state in Step 1 of the haem catalytic cycle where the iron was in a penta-coordinated high-spin ferric form (Meunier *et al.*, 2004; Munro *et al.*, 2007; Shaik *et al.*, 2010). Charges for unprotonated AA were obtained from the Zinc Dock database (Irwin & Shoichet 2005; substance ref: 12504416) while charges for unprotonated DHA and EPA were derived using Automated Topology Builder (Malde *et al.*, 2011) and R.E.D. Server (Bayly *et al.*, 1993; Dupradeau *et al.*, 2010; Vanqualef *et al.*, 2011).

Prior to the creation of topology and coordinate files in LEaP (Link, Edit and Parm; (Amber, 2015)), side chain orientations and protonation states at neutral pH were set using MolProbity's Reduce function (Chen *et al.*, 2010) adapted within AmberTools15. LEaP allows the merging of various inputs files and parameters into a format (.prmtop) readable by AMBER. Periodic boundary conditions (PBCs) were used and the CYP was solvated in a truncated octahedral box

with sides 10Å from protein boundaries. PBCs are used in MD to approximate the simulation into an infinite size system, and thus prevent boundary effects from affecting the species of interest. The overall charge within the box was neutralised by the addition of Na⁺ or Cl⁻ ions as appropriate. Where required, force-field parameters for AA, DHA and EPA were generated from the general Amber force-field (GAFF; Wang et al. 2004) using the Antechamber and parmchk2 modules. In all simulations, the non-bonded interactions cut-off radius was set at 10Å and the Particle-Mesh Ewald (PME) method was used for treatment of long-range electrostatics. PME is a computationally efficient method of summing interaction energies between nuclei.

1.3.2 Minimisation, Heating, Equilibration and Production

The homology model was energy minimised to find the point on the potential energy surface with zero net force. This was done in three sequential 10,000-step runs, each of which included 3,000 steps of steepest descent minimisation followed by 7,000 steps of conjugated gradient minimisation. In the first run, all solute atoms were fixed with high restraints (restraint weight = 100.0 kcalmol⁻¹Å⁻²). In the second run the restraint was reduced (restraint weight = 10.0 kcalmol⁻¹Å⁻²) and applied only to heavy atoms while in the third run only the backbone atoms were fixed with the same low restraint value. The resultant structure was exported as a mol2 file to maintain charges and connections and used in docking experiments.

Energy minimisation of the docked complexes proceeded as above, with an additional unrestrained run comprised of 30,000 steps of steepest gradient minimisation followed by 70,000 steps of conjugated gradient minimisation.

The solute was then restrained (restraint weight = 10.0 kcalmol⁻¹Å⁻²) and the system was gradually heated to 310K over 180ps and equilibrated at 310K over 20ps in an NVT ensemble using Langevin dynamics. A system under NVT regulation has a constant number of particles (N), volume (V) and temperature (T).

After the required kinetic energy had been introduced, all simulations, including further equilibrations, were run in an ensemble where the pressure was kept constant instead of the volume (NPT ensemble). Pressure scaling was applied to equilibrate the system at 1.0bar over 100ps before the restraint was reduced (restraint weight = 2.0 kcalmol⁻¹Å⁻²) and set to backbone atoms only for 40ps. All restraints were removed for the final 100ps of the NPT equilibration. All subsequent simulations were run in 50ns chunks with a 2fs time-step and energy and coordinates being recorded every 20ps.

Outside of minimisations, the SHAKE algorithm was employed to constrain the vibration in covalent bonds involving hydrogen. Representative AMBER input files are uploaded in the repository.

1.3.3 Analysis

Analysis of trajectories was carried out using the CPPTRAJ program of AmberTools15 (Roe and Cheatham, 2013). Prior to all analyses, excluding the energy, temperature and density investigations, the simulation ensemble was pre-processed. This consisted of stripping waters and ions, removing the box, and fitting all the frames to the first restart file. Representative CPPTRAJ input files are uploaded in the repository.

1.4 References

- Bayly, C. I. C. C. I., Cieplak, P., Cornell, W. D. and Kollman, P. A. (1993) 'A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model', *The Journal of Physical Chemistry . American Chemical Society*, 97, pp. 10269–10280. doi: 10.1021/j100142a004.
- Berman, H. M. (2000) 'The Protein Data Bank', *Nucleic Acids Research . Oxford University Press*, 28(1), pp.235–242. doi: 10.1093/nar/28.1.235.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity : all-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D Biological Crystallography . International Union of Crystallography*, 66(1), pp. 12–21. doi: 10.1107/S0907444909042073.
- Dupradeau, F.-Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W. and Cieplak, P. (2010) 'The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building', *Physical Chemistry Chemical Physics . The Royal Society of Chemistry*, 12(28), p. 7821. doi: 10.1039/c0cp00111b.
- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., Pieper, U. and Sali, A. (2006) 'Comparative Protein Structure Modeling Using Modeller', in *Current Protocols in Bioinformatics . Hoboken, NJ, USA: John Wiley & Sons, Inc.*, p. 5.6.1-5.6.30. doi:10.1002/0471250953.bi0506s15.
- Irwin, J. J. and Shoichet, B. K. (2005) 'ZINC - A free database of commercially available compounds for virtual screening', *Journal of Chemical Information and Modeling*, 45(1), pp. 177–182. doi: 10.1021/ci049714.
- Jorgensen, W. L. (1982) 'Revised TIPS for simulations of liquid water and aqueous solutions', *The Journal of Chemical Physics*, 77(8), p. 4156. doi: 10.1063/1.444325.
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. and Simmerling, C. (2015) 'ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB', *Journal of Chemical Theory and Computation . American Chemical Society*, 11(8), pp. 3696–3713. doi: 10.1021/acs.jctc.5b00255.
- Malde, A. K., Zuo, L., Breeze, M., Stroet, M., Poger, D., Nair, P. C., Oostenbrink, C. and Mark, A. E. (2011) 'An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0', *Journal of Chemical*

Theory and Computation , 7(12), pp. 4026–4037. doi: 10.1021/ct200196m.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. and Lopez, R. (2013) 'Analysis Tool Web Services from the EMBL-EBI', *Nucleic Acids Research* , 41(W1), pp. W597–W600. doi: 10.1093/nar/gkt376.

Meunier, B., de Visser, S. P. and Shaik, S. (2004) 'Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes', *Chemical Reviews* , 104(9), pp. 3947–3980. doi: 10.1021/cr020443g.

Munro, A. W., Girvan, H. M. and McLean, K. J. (2007) 'Variations on a (t)heme—novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily', *Natural Product Reports* , 24(3), pp. 585–609. doi: 10.1039/B604190F.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) 'UCSF Chimera - A visualization system for exploratory research and analysis', *Journal of Computational Chemistry* , 25(13), pp. 1605–1612. doi: 10.1002/jcc.20084.

Roe, D. R. and Cheatham, T. E. (2013) 'PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data', *Journal of Chemical Theory and Computation*, 9(7), pp. 3084–3095. doi: 10.1021/ct400341p.

Salomon-Ferrer, R., Case, D. a and Walker, R. C. (2013) 'An overview of the Amber biomolecular simulation package', *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(2), pp. 198–210. doi: 10.1002/wcms.1121.

Shahrokh, K., Orendt, A., Yost, G. S. and Cheatham, T. E. (2012) 'Quantum mechanically derived AMBER compatible heme parameters for various states of the cytochrome P450 catalytic cycle', *Journal of Computational Chemistry*, 33(2), pp. 119–133. doi: 10.1002/jcc.21922.

Shaik, S., Cohen, S., Wang, Y., Chen, H., Kumar, D. and Thiel, W. (2010) 'P450 Enzymes: Their Structure, Reactivity, and Selectivity—Modeled by QM/MM Calculations', *Chemical Reviews. American Chemical Society*, 110(2), pp. 949–1017. doi: 10.1021/cr900121s.

Trott, O. and Olson, A. J. (2009) 'AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading', *Journal of Computational Chemistry*, 31(16). doi: 10.1002/jcc.21334.

Vanquelef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J. C., Cieplak, P. and Dupradeau, F.-Y. (2011) 'R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments', *Nucleic Acids Research. Oxford University Press*, 39(suppl_2), pp. W511–W517. doi: 10.1093/nar/gkr288.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. and Case, D. A. (2004) 'Development and testing of a general amber force field', *Journal of Computational Chemistry* , 25(9), pp. 1157–1174. doi: 10.1002/jcc.20035.