A systematic analysis of intermolecular interactions of cytochrome P450s with their ligands

Yaraslau U. Dzichenka[1], Natalia E. Boboriko[2], Sergey A. Usanov[2], Suzana Jovanović-Šanta[3], Biljana Šmit[4*]

[1] National Academy of Sciences of Belarus, Institute of Bioorganic Chemistry, Minsk, Belarus, 220084, Kuprevich str. 5/2; e-mail: dichenko@iboch.by, usanov@iboch.by
[2] Belarusian State University, Faculty of Chemistry, Department of Inorganic Chemistry, Minsk, Belarus, 220030, Nezavisimosti ave. 4; e-mail: natchem@tut.by
[3] University of Novi Sad, Faculty of Sciences, Novi Sad, Serbia, 21000, Trg Dositeja Obradovića 3; e-mail: suzana.jovanovic-santa@dh.uns.ac.rs
[4] University of Kragujevac, Institute for Information Technology, Kragujevac, 34000, Serbia, Jovana Cvijića bb; e-mail: biljana.smit@uni.kg.ac.rs

* *Corresponding author*

DOI: 10.46793/ICCBIKG25.608D

Abstract: Statistical analysis of 6,859 intermolecular contacts in experimentally determined spatial structures of cytochromes P450 (CYPs) with ligands, stored in the Protein Data Bank (PDB), was performed. The most frequent types of contacts are hydrophobic interactions, hydrogen bonds, water bridges, and π-stacking interactions. Domain-specific analysis of the interaction types allowed identifying of interaction patterns, which are specific for bacterial and eukaryotic proteins. Among these patterns amino acids (AA) involved in ligand binding are preferred. The high fraction of hydrophobic contacts reflects the fact that many of the tested ligands are lead compounds and that the active site of most CYPs is more hydrophobic than hydrophilic. The results highlight cytochromes P450 specific ligand-binding peculiarities, which are of great importance for molecular structure embedding in deep learning neural networks and for the development of efficient scoring functions to analyze molecular docking results.

Keywords: cytochrome P450, intermolecular interaction, Protein Data Bank, ligand.

1. Introduction

Cytochromes P450 (CYPs) are a diverse family of enzymes involved in various critical physiological processes [1]. CYPs are found in organisms across all domains of life, with the exception of certain Archaea and Eubacteria [2], including Escherichia coli [3]. In prokaryotes, there are more than 600 CYPs that participate in the hydroxylation of fatty acids, linear alkanes, aromatic hydrocarbons, and in xenobiotic metabolism [4]. In plants, these enzymes (comprising over 270 families) take part in the biosynthesis of plant hormones, polymeric compounds, in interorganismal communication (e.g., with bacteria), hormonal signaling, herbicide metabolism, and stress tolerance [5]. Fungal

cytochromes P450 (with more than 800 families) are involved in ergosterol biosynthesis, gibberellin production, xenobiotic detoxification (including antifungal medications), hydroxylation of saturated and unsaturated fatty acids, and denitrification processes [6]. In animals, P450s contribute to the biosynthesis of steroid hormones and xenobiotic detoxification. These enzymes are among of the most extensively studied proteins, particularly human CYPs, due to their medical significance: many of them serve as molecular targets for drugs and are associated with severe diseases [7]. Statistical data on protein-ligand interactions are crucial for designing novel, highly specific bio-regulators. First, these data implicitly contain information about all possible interaction types, providing a foundation for computationally efficient and relatively accurate in silico modeling of ligand binding. Second, this analysis enhances the accuracy of scoring functions in computational models used for virtual screening of new bioregulators.

2. Methodology

Only experimentally determined 3D structures of cytochromes P450 with ligands from the PDB archive were analysed. The structures were extracted from a snapshot of PDB captured on July 7, 2025 according to the next procedure. Firstly, PDB identifiers of all heme-containing structures were extracted from the list of PDB and chemical component identifier correspondences (http://ligand-expo.rcsb.org/dictionaries/cc-to-pdb.tdd). Next step, structures of cytochromes P450 were selected from the list of heme-containing proteins (based on the description of the entity and a brief title that describes the contents of the entry stored in mmcif file). Missing hydrogen atoms in protein structure were added based on known bond lengths and angles using functionality of Hydride (v. 1.2.3) Python library. Openbabel (v. 3.0.1) was used for hydrogen atoms addition to the structures of the ligands. After the structures were filtered according to the following criteria: ligand should be located in the active site of the enzyme and the minimum distance between ligand's atoms and Fe of heme should be ≤ 5.2 Å; ligand was considered as a small molecule (up to 1000 Da) that is not a part of buffer or crystallization agent. Each structure of protein-ligand complex from elementary cell as well as structures with disordered residues or atoms were split into different files and considered as a separate complex. Complexes with the same ligand and protein were clustered on the basis of the RMSD of ligand's atomic coordinates. Structure with better resolution was considered as a representative structure of each cluster for the further analysis of intermolecular interactions. Identification of protein-ligand contacts, interaction types and parameters of interaction was performed using PLIP [8] python module (v. 2.2.0). Pandas (v. 2.2.3) and Seaborn (v. 0.3.12) Python libraries were used for the statistical analysis of the data obtained and visualization of the results. All steps were automated using custom Python (v. 3.10.13.) scripts.

3. Results and Discussion

At the moment, PDB contains 238936 experimentally determined structures. Among them only 5777 heme-containing structures and 1334 are spatial structures of cytochromes P450. After filtration only 867 separate protein-ligand complexes were

picked up with 481 different ligands and more than 190 different enzymes. 56.7% of them are bacterial (the three most presented species are *Rhodopseudomonas palustris*, *Mycobacterium tuberculosis* and *Pseudomonas putida*), 42.3% eukaryotic (the three most presented species are *Homo sapiens*, *Trypanosoma cruzi* and *Saccharomyces cerevisiae*) and 1.0% archaeal proteins. Profiling of protein-ligand interactions allowed to identify 6859 contacts: hydrophobic contacts (72.0%), hydrogen bonds (12.9%), water bridges (9.3%), π-stacking bonds (3.2%), salt bridges (1.9%), halogen contacts (0.5%) and π-cation bonds (0.3%). There are some differences between interaction types for bacterial and eukaryotic enzymes. According to the normalized frequencies of interaction types (each type was normalized on the total number of contacts for given domains followed by the normalization on the frequency of hydrophobic contacts) hydrogen bonds are almost twice as common among bacterial than eukaryotic cytochromes P450. Also, water and salt bridges are overrepresented in bacterial CYPs in comparison with eukaryotic CYPs. On the other hand π-stacking and π-cation interactions occur more often in eukaryotic CYPs than in bacteria. Such high proportion of hydrophobic contact reflects the following facts. Firstly, many of the tested compounds are lead compounds that were optimized for binding, so one of the most efficient and easy strategies for this is to optimize hydrophobic contacts. Secondly, the active site of many tested CYPs is mostly hydrophobic. Analysis of *hydrophobic contacts* showed that the most frequent AA taking part in corresponding interactions are Leu for bacterial and Phe for eukaryotic CYPs. In case of bacterial CYPs ligands also form hydrophobic contacts more often with Val, Thr, Pro, Lys, His and Gln. From the other hand eukaryotic cytochromes P450 showed more often interactions with Glu, Met, Tyr, Arg. As it was found earlier in the frame of analysis of AK frequencies in proteins for different taxa [9] Pro, Gln, Lys and His are more common in eukaryotic proteins as well as Arg and Met in bacteria. Inverse relationship is evidenced between frequency of forming hydrophobic contacts with these AA and their distribution between proteins for the corresponding domain of life, so one can state that the differences obtained are likely connected not with evolutionary processes but with ligand binding properties of the enzymes. In case of *hydrogen bonds* the most frequent AA taking part in the interaction in eukaryotic proteins are Tyr, Thr, Asn, Met, Trp and Ala, Glu, Asp. On the other hand, in bacterial proteins the highest frequencies possess Ser, Gln, Leu, Val, Arg and His. Formation of hydrogen bonds with Lys was detected only in bacterial protein-ligand complexes. There are differences in preferable ligand's atom types that are donors of hydrogen bond. In case of eukaryotic proteins sp^3-hybridized N and sp^3-hybridized O are the most frequent hydrogen bond donors. But in case of bacterial proteins amide N and sp^2-hybridized N with lone pair mainly act as hydrogen bond donors. *Water bridges* in bacterial protein-ligand complexes are mainly formed with atoms belonging to Arg, Ser, Val, Ala, Leu, His, Lys, Met. In contrast for eukaryotic proteins bond frequency is higher for Thr, Tyr, Asn, Ile. *π-Stacking bonds* are mainly formed with Phe residue in both groups of proteins. The most frequent residues for bacterial proteins are Phe and Trp, while Phe and Tyr are for eukaryotic proteins. *Salt bridges* in bacterial CYPs are mainly formed between ligand's atoms and atoms of Arg, Glu, His, while in eukaryotic CYPs key residues are Glu, Asp and Lys.

4. Conclusions

In this study, statistical analysis of the nature and frequency of atomic interactions between inhibitors/substrates and various CYPs was conducted. Specific interactions between ligands and cytochromes were found out, as well as AA residues responsible for different types of the interactions in bacterial and eukaryotic CYPs. The results obtained are of great importance for the creation of protein-specific scoring functions for use in molecular docking. The data obtained is also important for the creation of molecular embedding for the usage in deep learning-based models.

Acknowledgment
This research was partially funded by the Belarusian Republican Foundation for Fundamental Research (Grant No. X25MC-012), Belarus-Serbia bilateral projects (X18SRBG- 002/51-03-003036/2017-09/02 and X20SRBG-004/337-00-00612/2019-09/04) and the Ministry of science, technological development and innovation of the Republic of Serbia (Contract No. 451-03-136/2025-03/200378).

References

[1] J.D. Finnigan, C. Young, D.J. Cook, S.J. Charnock, G.W. Black., *Chapter Nine - Cytochromes P450 (P450s): A review of the class system with a focus on prokaryotic P450s*, Adv. Protein Chem. Struct. Biol., 122 (2020) 289-320.
[2] S.L. Kelly, D.E. Kelly., *Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us?*, Philos. Trans. R. Soc. B, Biol. Sci., 2013 (19) 20120476.
[3] V.B. Urlacher, M. Girhard., *Cytochrome P450 monooxygenases: an update on perspectives for synthetic application*, Trends Biotech., 30 (2012) 26-36.
[4] D.C. Lamb, M.R. Waterman., *Unusual properties of the cytochrome P450 superfamily*, Philos. Trans. R. Soc. B, Biol. Sci., 368 (2013) 20120434.
[5] D. Minerdi, S. Savoi, P. Sabbatini., *Role of Cytochrome P450 Enzyme in Plant Microorganisms' Communication: A Focus on Grapevine*, Int. J. Mol. Sci. 24 (2023) 4695.
[6] B. Črešnar, Š. Petrič., *Cytochrome P450 enzymes in the fungal kingdom*, Biochim. Biophys. Acta - Proteins Proteom., 1814 (2011) 29-35.
[7] F.P. Guengerich, M.R. Waterman, M.Egli., *Recent Structural Insights into Cytochrome P450 Function*, Trends Pharmacol. Sci., 37 (2016) 625-640.
[8] M.F. Adasme, K.L. Linnemann, S.N. Bolz, F. Kaiser, S. Salentin, V.J. Haupt, M. Schroeder., *PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA*, Nucleic Acids Res., 49 (2021) W530–W534.