Supporting information

A context-based Matched Molecular Pair Analysis identifies structural transformations that reduce CYP1A2 Inhibition.

Janvi A. Raut, and Vaibhav A. Dixit*

Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER Guwahati), Department of Pharmaceuticals, Ministry of Chemicals & Fertilizers, Govt. of India, Sila Katamur (Halugurisuk), P.O.: Changsari, Dist: Kamrup, 781101, Guwahati, Assam, India.

*Corresponding author Email: vaibhavadixit@gmail.com, vaibhav@niperguwahati.in

Contents

Detailed Methodology of MMPA

Data Curation

To find simple and frequently occurring transformations in CYP1A2 inhibitors, we collected data from the ChEMBL database and performed classical (i.e., context-independent) MMPA. Thereafter, the MMPs for these transformations were grouped based on chemotype, and the results of classical MMPA were compared with a context-based approach. The dataset of 6029 compounds (IC_{50}/K_i CYP1A2 Inhibitor) was curated from the target ID ChEMBL3356. The available data was filtered based on the following criteria: (i) Target Organism: *Homo sapiens*, (ii) Activity: IC_{50} and K_i data for a single protein, and (iii) confidence score of 8 or 9. The final prepared dataset of 2454 molecules was used to perform MMPA. The workflow was designed and implemented in the KNIME Analytics Platform (version 4.7.5, see Figure S1).[1]

Dataset preparation and molecule preprocessing

As shown in (Figure S1), a workflow was built using different nodes in the KNIME Tool. Curated input data (SMILES pattern) was fed to the workflow using the "CSV reader" node. The dataset was prepared by removing the missing values (64 SMILES and 1842 activity values) by "Row filter," which gave 4122 molecules. The duplicates (282 SMILES) were removed by the "Duplicate row filter." Finally, only one assay type (A) and a single protein format were selected, giving 2454 compounds for performing MMPA. Molecules in the dataset were pre-processed by desalting and charge standardization, then stripping the stereochemistry and double bond geometry information. RDKit was used to get canonical SMILES.

Fragmentation and Generation of Matched Molecular Pairs

The fragment indexing (F+I) method was used to fragment molecules.[2] Subsequently, fragments were generated using the 'MMP Molecule Fragment' node and analyzed to identify Matched Molecular Pairs. The activity difference for each transformation was calculated with the 'Math Formula' node. Further, all the pairs were grouped based on the type of transformations involved using the 'GroupBy' node. After grouping the pairs with similar types of transformation, the output table was assessed for the transformation with the maximum number of occurrences and its average activity differences. The most occurring transformations were selected ($n \geq 10$ for classical and $n \geq 5$ for context-based MMPA) and written into a new CSV file.

Classical MMP Analysis

The transformations obtained from the above methodology were filtered based on their number of occurrences. Descriptive statistical analysis was performed for each transformation, and outliers were removed. The key count for each transformation was used for further study. The required minimum average activity difference (min ΔAA) was calculated for all the transformations[3] and compared with the calculated/obtained activity difference (ΔA = change in pIC_{50}/pK_i) from the classical MMPA. The transformations that passed the min ΔAA criteria were considered statistically significant. Some transformations were found to be insignificant and thus are not included (see results and discussion section in the main text).

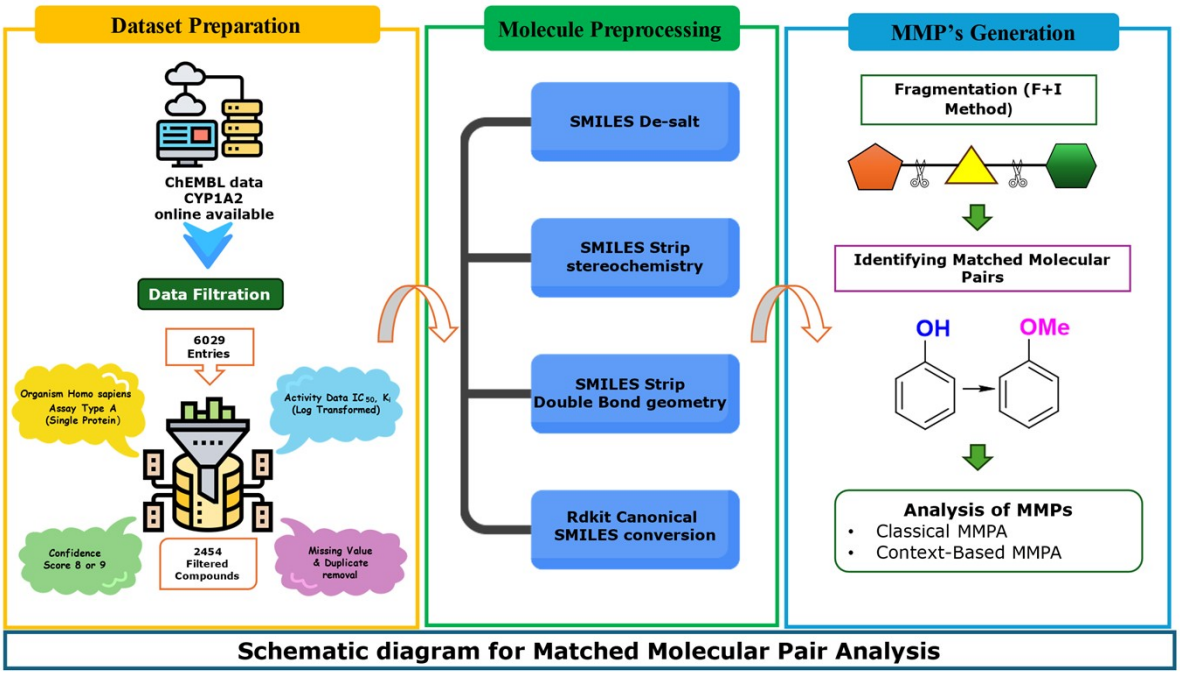

*Figure S1: Conceptual workflow for Matched Molecular Pair Analysis implemented in KNIME. This includes data preparation, molecule preprocessing, fragmentation of molecules, identification and generation of matched molecular pairs, and configuration of results followed by analysis of MMPs (see **Figure S2**) for the detailed KNIME workflow.*

Chemotype-Specific MMP Analysis

The chemotype-specific MMPA was performed after choosing the five most frequently occurring transformations (H to Me, H to F, H to OMe, H to OH, and F to Cl). To avoid the repetition of pairs, directional filtering is performed to select only the left-to-right change (e.g., H to Me) for all transformations. This method divides the large dataset into groups of similar compounds. MMPs with a Tanimoto coefficient ≥ 0.7 were clustered. The overall mean of all pairs in each group was calculated, and the groups

with statistically significant differences in mean activity were chosen for further analysis (see Table 4 main text).[3] Representative scaffolds were obtained for all groups using scaffold analysis in Data Warrior.[4] The same procedure was followed for all five (H to Me, H to F, H to OMe, H to OH, and F to Cl) major transformations (see Figure S2). The canonical SMILES of the right fragment of each pair were considered as a reference for clustering (note that the left fragment gives the same clustering).

MMPA Workflow

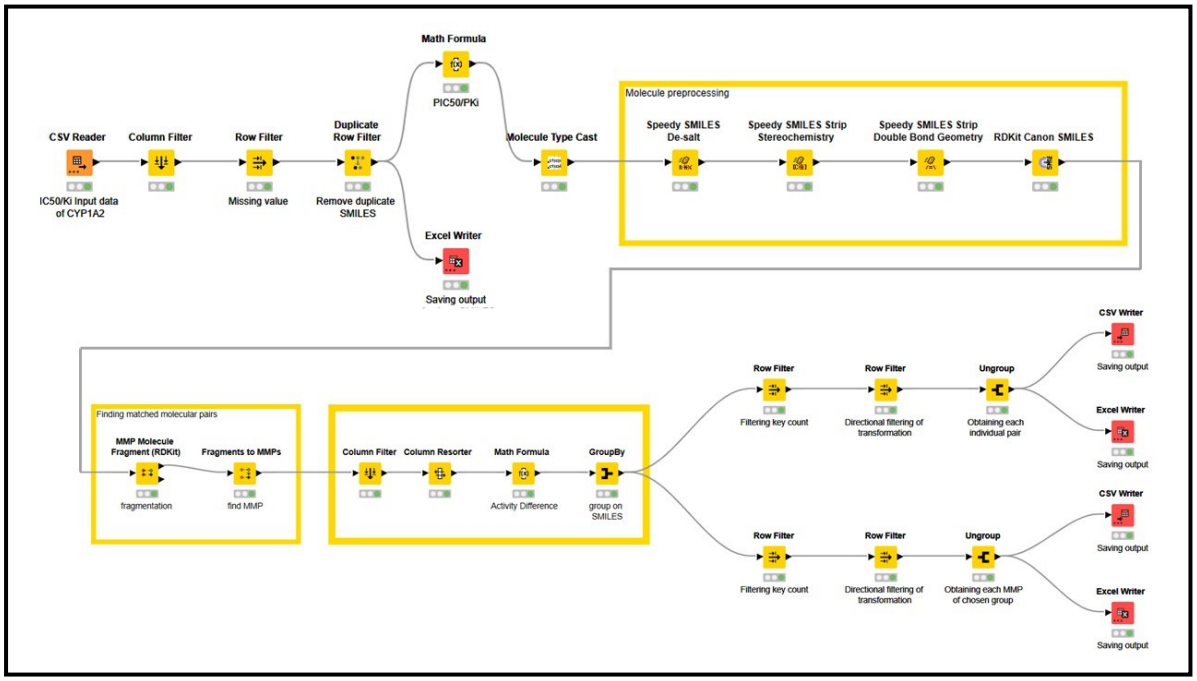

Figure S2: MMPA workflow developed in KNIME Analytics Platform for CYP1A2 inhibitors.

Structure-based analysis of chemotype specific MMPA using molecular docking.

The molecular docking was performed for three matched molecular pairs (structures **1**, **2**, **3**, and **4** in Figure S3) containing H to Me transformation. Biovia Discovery Studio Client 22 was used to perform the docking. The docking score was obtained for the first three poses of each chosen compound in the form of negative CDOCKER interaction energy. Validation was performed by redocking the co-crystalized ligand in the crystal structure of the protein (2HI4).

Protein Preparation:

The CYP1A2 protein structure with PDB ID: 2HI4 was downloaded from RCSB PDB.[5] The protein was prepared for docking by removing the water molecules and adding

polar hydrogens and Kollman charges. The prepared protein structure was saved as PDB.

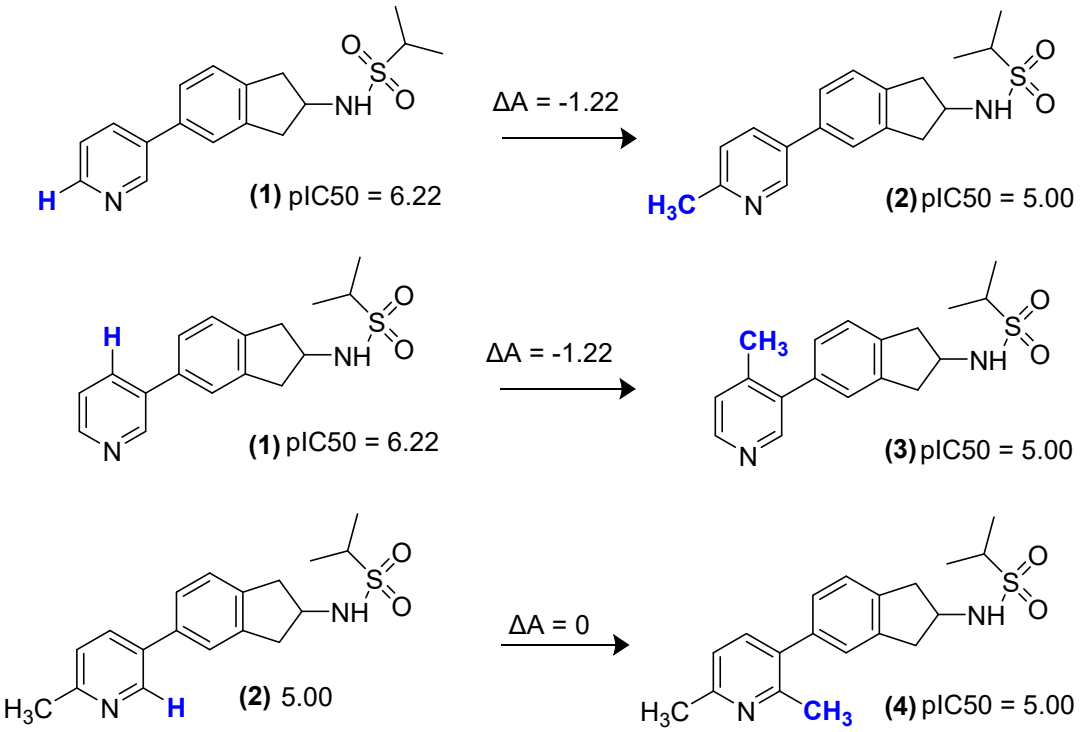

*Figure S3. Structure of molecules from the **A** scaffold with indanylpyridine.*

Ligand Preparation:

The 3D structures for ligands were drawn and minimized using ChemDraw20 and CHEM3D software and saved as SDF files. The SDF file was further used for ligand preparation using Biovia Discovery Studio. This includes ligand minimization and conformer generation.

Grid Box Generation and Docking:

The coordinates for the grid box were chosen from the ligand in the protein's crystal structure. The binding site was chosen from "the current selection" (co-crystallized ligand) option provided in Discovery Studio software. The docking was carried out using prepared protein and ligands using CDOCKER.

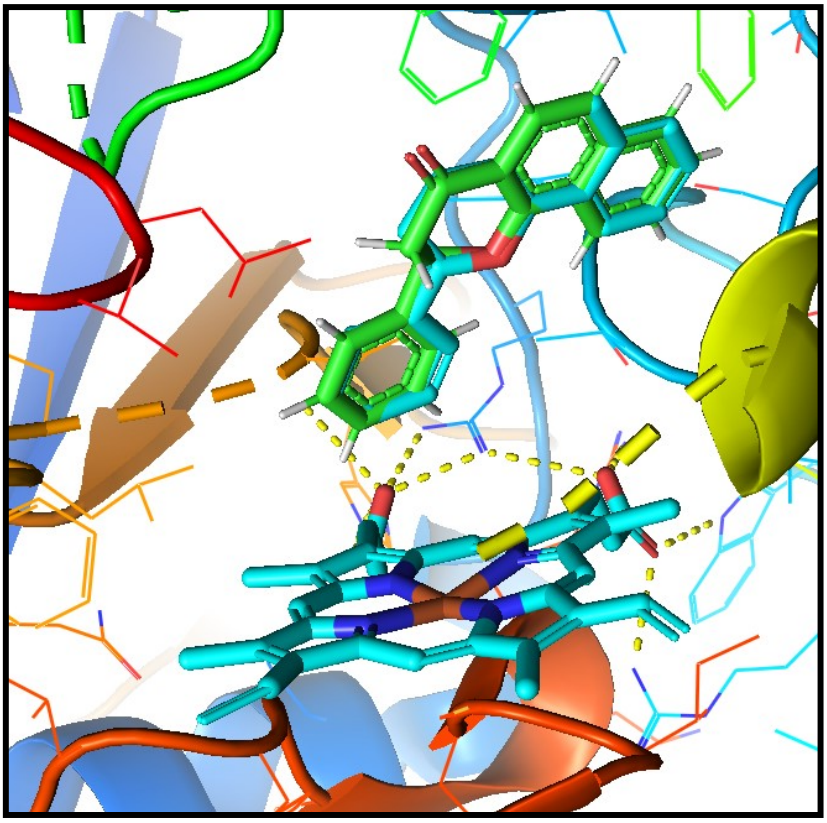

Figure S4. The docked pose for alpha naphthoflavone is shown in cyan color. The green-color overlapping structure is a crystal structure posed inside the CYP1A2 active site (PDB ID: 2HI4).

The docking protocol was validated by comparing the crystal structure pose with the docked pose, as shown in (Figure S4). The crystal structure pose of the ligand overlaps almost perfectly with the docked pose.

Binding poses and distance measurement

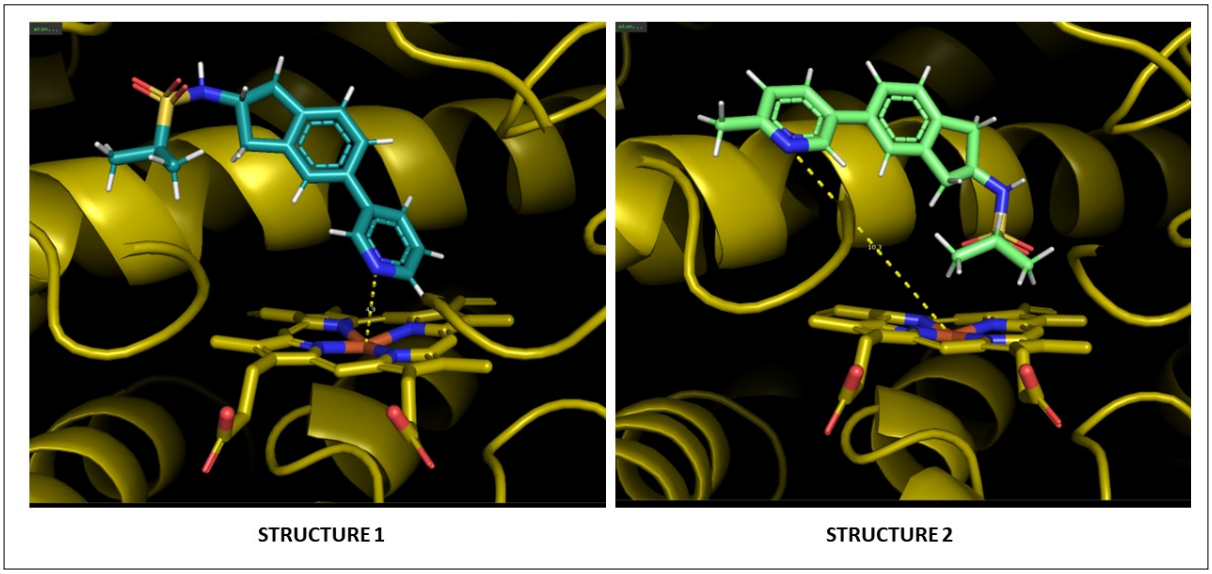

Figure S5: The distance measurement between Heme-Fe and ligand SP2-N for structure 1 (4.9 Å) and 2 (10.3 Å) in Figure S3.

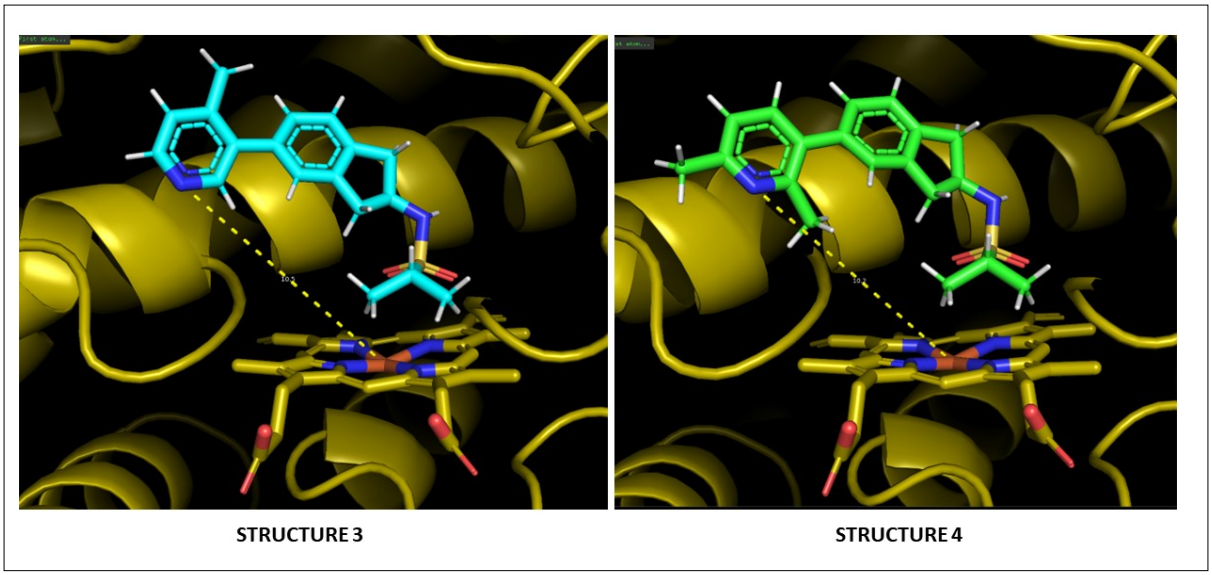

Figure S6: The distance measurement between Heme-Fe and ligand SP2-N for Structure 3 (10.5 Å) and 4 (10.2 Å) in Figure S3.

Summary of the additional context-based MMPA for H to Me transformation

The MMPs for scaffold **B** in Figure 5 (main text) for the H to Me transformation are shown in (Figure S7). All three pairs for this scaffold differ in the position of the added substituent. Structures **9** and **10** show the highest change in activity compared to the other two pairs. In Summary, Structure **10** is expected to show a decrease in CYP1A2 inhibition.

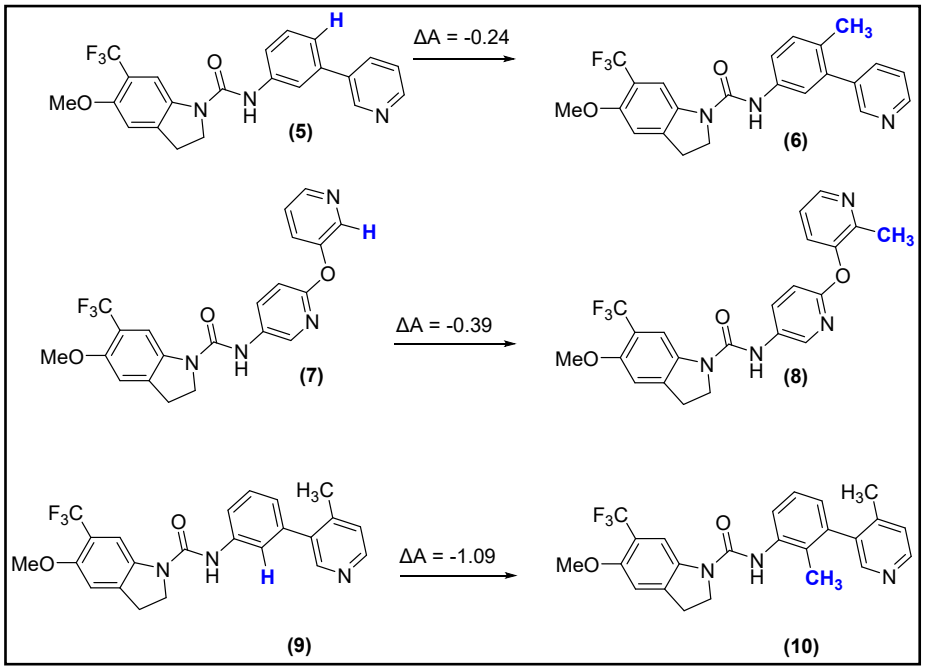

Figure S7: MMPs belonging to the scaffold B given in figure 4 (main text) for transforming H to CH_3.

A similar kind of context-based analysis was performed for the other significant transformations obtained from classical MMPA. All the transformations except one (H to F) gave a set of contextual MMPs (same chemotype) with significant activity changes compared to the global transformation.

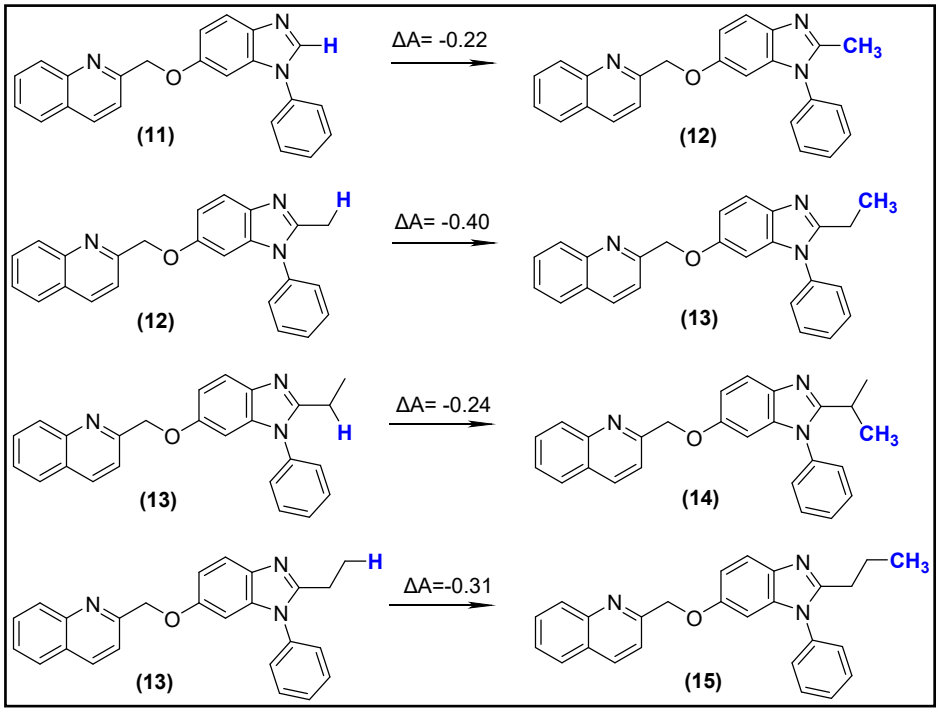

Figure S8: MMPs belonging to the scaffold C given in Figure 4 (main text) for transforming H to CH_3.

As shown in (*Table S1*) out of two clusters for H to OMe transformation, one with a key count of 23 shows a statistically significant activity change. Compared to a -0.2 log unit change in activity for the global transformation of H to OMe in classical MMPA, the context-based analysis gave a -0.36 log unit change for structure **11** (Figure S8).

Table S1: Clusters of pairs with different chemical contexts for three different transformations detailing the mean change in activity for each chemical context and its statistical significance in terms of minimum average activity difference as suggested by Kramer et al.[3]

Sr. no	Transformation	Clusters	Key Count	Mean ΔpIC50 /ΔpKi	Std dev	t value	SEM	Min. ΔAA	Stat significance	Benefit in LO
1	**H to F**	1	6	0.20	0.29	4.24	0.12	0.26	NO	NO
		2	19	0.43	1.28	6.37	0.29	0.98	NO	NO
2	**H to OMe**	1	12	0.12	1.22	1.16	0.35	0.22	NO	NO
		2	**23**	**-0.36**	**0.59**	**-14.10**	**0.12**	**-0.91**	**YES**	**YES**
3	**H to OH**	1	**31**	**-0.18**	**0.82**	**-6.91**	**0.15**	**-0.53**	**YES**	**YES**
4	**F to Cl**	1	**15**	**-0.15**	**0.33**	**-7.02**	**0.08**	**-0.31**	**YES**	**YES**

In summary, our analysis shows that a context-based MMPA identifies key transformations required to reduce CYP1A2 liabilities and might be useful for lead optimization. Additionally, the inhibitory potential of compounds with Type II mechanism of inhibition (via interaction with heteroatoms and Heme-Fe) can be predicted or explained retrospectively using a docking-based analysis. However, for compounds that lack a Heme-Fe coordinating atom/groups, the mechanism of inhibition is expected to be Type I and dependent upon other complicated inhibitor-enzyme factors for e.g., thermodynamic (enthalpy and entropy contributions), redox potentia[6,7] biochemical and binding kinetics mechanisms. These are very difficult to explain using simple molecular modeling methods and are expected to require extensive investigation beyond the scope of this manuscript.

References

(1) Sydow, D.; Wichmann, M.; Rodríguez-Guerra, J.; Goldmann, D.; Landrum, G.; Volkamer, A. TeachOpenCADD-KNIME: A Teaching Platform for Computer-Aided Drug Design Using KNIME Workflows. *J Chem Inf Model* **2019**, *59* (10), 4083–4086. https://doi.org/10.1021/acs.jcim.9b00662.

(2) Yang, Z.; Shi, S.; Fu, L.; Lu, A.; Hou, T.; Cao, D. Matched Molecular Pair Analysis in Drug Discovery: Methods and Recent Applications. *J Med Chem* **2023**, *66* (7), 4361–4377. https://doi.org/10.1021/acs.jmedchem.2c01787.

(3) Kramer, C.; Fuchs, J. E.; Whitebread, S.; Gedeck, P.; Liedl, K. R. Matched Molecular Pair Analysis: Significance and the Impact of Experimental Uncertainty. *J Med Chem* **2014**, *57* (9), 3786–3802. https://doi.org/10.1021/jm500317a.

(4) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J Chem Inf Model* **2015**, *55* (2), 460–473. https://doi.org/10.1021/ci500588j.
(5) Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2. *Journal of Biological Chemistry* **2007**, *282* (19), 14348–14355. https://doi.org/10.1074/jbc.M611692200.
(6) Dixit, V. A.; Warwicker, J.; de Visser, S. P. How Do Metal Ions Modulate the Rate-Determining Electron-Transfer Step in Cytochrome P450 Reactions? *Chemistry – A European Journal* **2020**, *26* (66), 15270–15281. https://doi.org/10.1002/chem.202003024.
(7) Dixit, V. A.; Murty, U. S.; Bajaj, P.; Blumberger, J.; de Visser, S. P. Mechanisms of Electron Transfer Rate Modulations in Cytochrome P450 BM3. *J Phys Chem B* **2022**, *126* (47), 9737–9747. https://doi.org/10.1021/acs.jpcb.2c03967.