

Article

Not peer-reviewed version

Phenylalanine Residues in the Active Site of CYP2E1 Participate in Determining the Binding Orientation and Metabolism-Dependent Genotoxicity of Aromatic Compounds

[Kegui Hu](#)[#], Hongwei Tu[#], [Jiayi Xie](#), Zongying Yang, Zihuan Li, Yijing Chen, [Yungang Liu](#)^{*}

Posted Date: 29 April 2023

doi: 10.20944/preprints202304.1222.v1

Keywords: aromatic compounds; CYP2E1; phenylalanine; molecular simulation; random forest



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Phenylalanine Residues in the Active Site of CYP2E1 Participate in Determining the Binding Orientation and Metabolism-Dependent Genotoxicity of Aromatic Compounds

Keqi Hu ^{1,2,#}, Hongwei Tu ^{3,#}, Jiayi Xie ^{2,#}, Zongying Yang ², Zihuan Li ^{1,2}, Yijing Chen ² and Yungang Liu ^{2,*}

¹ Department of Science and Education, Guangdong Second Provincial General Hospital, 466 Xingang Middle Road, Guangzhou 510317, China.

² Department of Toxicology, School of Public Health, Southern Medical University (Guangdong Provincial Key Laboratory of Tropical Disease Research), 1023 S. Shatai Road, Guangzhou 510515, China.

³ Guangdong Provincial Center for Disease Control and Prevention, Qunxian Road, Panyu District, Guangzhou 511430, China.

These authors contributed equally to this work.

* Correspondence: Yungang Liu, yungliu@126.com

Abstract: The composition of amino acids forming the active site of a CYP enzyme is impactful in its substrate selectivity. For CYP2E1, the role of PHE residues in the formation of effective binding orientations for its aromatic substrates remains unclear. In this study, molecular docking and molecular dynamics analysis were performed to reflect the interactions between PHEs in the active site of human CYP2E1 and various aromatic compounds known as its substrates. The results indicated that the orientation of 1-methylpyrene (1-MP) in the active site was highly determined by the presence of PHEs, PHE478 contributing to the binding free energy most significantly. Moreover, by building a random forest model the relationship between each of 19 molecular descriptors of polychlorinated biphenyl (PCB) compounds (from molecular docking, quantum mechanics, and physicochemical properties) and their human CYP2E1-dependent mutagenicity as established mostly in our lab, was investigated. The presence of PHEs did not appear to significantly modify the electronic or structural feature of each bound ligand (PCB), instead, the flexibility of the conformation of PHEs contributed substantially to the effective binding energy and orientation. It is supposed that PHE residues adjust their own conformation to permit a suitably shaped cavity for holding the ligand and forming its orientation as favorable for a biochemical reaction. This study has provided some insights into the role of PHEs in guiding the interactive adaptation of the active site of human CYP2E1 for the binding and metabolism of aromatic substrates.

Keywords: aromatic compounds; CYP2E1; phenylalanine; molecular simulation; random forest

1. Introduction

Biotransformation enzyme-dependent metabolism is one of the most important processes experienced by xenobiotics present in the human body. Among various families of biotransformation enzymes, cytochrome P450s (CYPs) as a superfamily of monooxygenases play a major role in the metabolism of various exogenous and endogenous compounds (1). The active site of each CYP enzyme is formed by the residues of involving amino acids (AA) and a Fe ion-containing heme group, which provides a hydrophobic cavity and determines the conformation (orientation) of substrate binding (2-4). Studies have shown that increased flexibility of the conformation of the AAs in the active site of CYP102A1 is favorable for reaching an orientation of substrate binding, particularly for a short distance from ligand to the Fe iron in the heme as valid for electron transfer (5). In human CYP3A4, several AAs in the active site (including ARG 106, ARG 372, GLU 374, PHE 108, and PHE 213) are involved in directing the effective binding of most substrates (6). Additionally, threonine

(THR) 303 in the active site of human CYP2E1 is significant for stable binding of 4-methylpyrazole *via* forming hydrogen bonds (4). Subsequent to directed mutations in the active site of human CYP2B4 (F296A, T302A, I363A, and V367L involved), the affinities of various substrates to the enzyme, the rates of product formation, and the potencies of known inhibitors for the mutated CYP2B4 were all tremendously changed, which evidences the importance of those AAs in the catalyzing activity and substrate specificity of the enzyme (7).

The aromatic AAs (e.g., PHEs) in the active site of CYP2E1 have been proven to be of great importance for stabilizing a hydrophobic cavity, sustaining the substrate binding, guiding its orientation, and determining the site of metabolism (SOM) (3, 4). In our previous studies, some phenyl group-containing chemicals appeared to be activated by human CYP2E1 for cytogenetic toxicity (8-13). For example, 1-methylpyrene (1-MP), which is a rodent carcinogen abundantly present in cigarette smoke (14), vehicle emissions (15), smoked or barbecued foods (16), and spills of crude oil (17), has been identified as a substrate for human CYP2E1, with its benzylic hydroxide as the major metabolite (18). However, CYP2E1 is known to oxidize only small-sized chemicals, such as ethanol and trichloroethylene, since its active site is rather small (4). As 1-MP has four conjugated benzyl groups, seemingly too large to be hold in the active site of CYP2E1, we were interested in the use of this compound in the docking study for understanding how 1-MP could bind into the small active site as effective for a biochemical reaction, especially for a role of the conformational flexibility of PHEs in the active site.

The molecular size of documented substrates for CYP2E1 seems to vary largely. Benzene and its hydroxylated metabolites (phenol and hydroquinone) are simplest aromatic substrates for human CYP2E1, thereby they are activated for mutagenic effects (8). However, the PHEs in the active site appear to have little effect on these small ligands during the metabolism process (19). It is likely that the PHEs in the active site are primarily important for directing the binding of larger (such as multiple phenyl groups-containing) ligands. Polychlorinated biphenyls (PCBs, contain two phenyl rings connected by a single bond), which are a large group of persistent organic pollutants with various health effects, including neuro/behavioral abnormalities (20), endocrine disruptions (21), overweight (22), and cancers (23), are substrates for different CYP2 enzymes giving rise to differently structured products. For example, PCB 95 and PCB 136 may be metabolized by human CYP2B6 to their meta-hydroxylated metabolites, while para-hydroxylated metabolites of PCBs were the major products formed by human CYP2E1 (24). Interestingly, the amino acid sequences of the active sites of both enzymes are quite similar to each other (3), except that PHE 478 in CYP2E1 is different from the ortholog site in CYP2B6 (VAL477), and it is likely that this discrepancy is involved in forming varying orientations of PCBs in the different active sites.

Collectively, the importance of PHEs in the active site of a CYP enzyme for substrates containing varying number of phenyl groups may vary in some way. To gain deeper insights into the role of PHEs in directing the binding and effective orientation of aromatic substrates in the active site of human CYP2E1, computational simulation methods such as molecular docking, molecular simulation, quantum chemistry approaches, and random forest model were employed in this study. Meanwhile, based on the established documents of human CYP2E1-activated mutagenicity of aromatic compounds in mammalian cells, a preliminary exploration of the screening value of the ligand-enzyme molecular docking analysis in the prediction of human CYP2E1-activated promutagens as candidate compounds to be tested in subsequent genotoxicity assays, was also conducted.

2. Materials and Methods

Preparation of Protein and Ligand Structures. The primary 3D structures of human CYP2E1 were obtained from RCSB Protein Data Bank (3LC4, 3KOH, 3GPH, 3E4E, 3T3Z, and 3E6I for CYP2E1). Ligands of this 3D model were removed before protein energy minimization. Charmm36 force field was used for all protein residues (25), and the TIP3P model was applied for water molecules (26). Energy minimization was performed by Gromacs (Version 2018.4) (27) using the steep algorithm with 0.01 nm for step size and was stopped when the maximal force was less than 1000.0 kJ/mol/nm. The

3D structures of ligands were built by Avogadro (28) and preliminarily optimized by using GFN2-XTB method (29).

Molecular Docking. Molecular docking was performed by using Autodock version 4.2 suits (30). Before docking, water molecules were removed from the protein structures. Gasteiger charges were added to the protein and ligands. The grid box was set to cover the active zone of proteins. The center of grid was on element Fe in the heme and the grid point values of X-, Y-, and Z-axes were set to 60, 60, and 60 with a spacing of 0.375 Å (1 Å is equal to 0.1 nm). Lamarckian genetic algorithm (LGA) (31) was performed for ligand-protein docking calculations. The default parameters of LGA were used and the number of LGA runs was set to 50. The 50 conformations were clustered by using a root-mean-square deviations (RMSD)-tolerance of 2.0 Å, then the typical conformation of each cluster was analyzed. The best conformation of protein-ligand complex was selected on the basis of mean binding energy score, the number in each cluster, and the distance between a ligand and the heme.

Molecular Equilibration, Dynamic and Free Energy Calculation. The best conformation of each ligand-protein complex as obtained from molecular docking was further used for molecular dynamics analysis. Protein and ligands were described by Charmm36 force field and CgenFF (the official CHARMM general force field server), respectively (25, 32). The canonical molecular dynamics with periodic boundary condition was performed by using GROMACS (Version 2018.4), with parameters being set according to previous descriptions (19). Briefly, short-range nonbonded interactions were cut off at 1.2 nm, with long-range electrostatics calculated using the particle mesh Ewald (PME) algorithm. The bonds to H were constrained with LINCS algorithm. After energy minimization, isothermic-isopycnic ensemble (NVT) for 100 ps was performed, of which temperature was maintained at 310K. Following NVT equilibration, isothermic-isobaric ensemble (NPT) for 100 ps was performed and the pressure was maintained at 1.0 bar. For both NVT and NPT equilibration, the position of ligands was restrained, and modified Berendsen thermostats were set for protein-ligand complex and other components. Finally, the production MD simulations were conducted for 50 ns without any restraints. Modified Berendsen thermostat and Parrinello-Rahman were applied to maintain temperature and pressure, respectively. The time step of integration was 2 fs and the energy and coordinates of the system were recorded every 10 ps.

Molecular Mechanics/Poisson-Boltzmann Surface Area (MMPBSA) method was used to estimate the free energies of protein-ligand complex (33). In this study, MMPBSA calculation was performed by using gmx_mmpbsa (34). Five hundred snapshots of each complex were extracted from the last 10 ns of molecular simulations.

Tunnel Analysis. As ligands need to pass tunnels to reach the active site of any enzyme, Caver 3.0 software was applied to identify the potential tunnels leading to the active site of human CYP2E1 (35). Five hundred snapshots were sampled from the molecular trajectories of 1-MP-CYP2E1 complex during the last 10 ns. The heme group was set as the starting point for tunnel searching. The probe radius and the clustering threshold was 1.0 and 4.5 Å, respectively. The other parameters were established using default settings. After calculations, tunnels were visualized by VMD (version 1.9) (36) and the parameters of the tunnel bottleneck were collected for further investigation.

GFN2-xTB- and DFT-based simulations. For a semi-empirical description of the each PCB binding to the active site of CYP2E1 (or its F478V mutant), the GFN2-xTB method was applied to perform geometry optimization and calculate the fractional occupation density analysis (FOD) (37). The analytical linearized Poisson-Boltzmann (ALPB) model was employed to provide the implicit solvation in chloroform for the total system.

The quantum mechanics calculation was performed by using ORCA 5.0 (38). The structure of each PCBs and its PHE complexe were optimized by using the unrestricted hybrid B3LYP functional (39) in combining with the def2-svp basis set (40). The wave functions and more accurate energies of the N, N+1, and N-1 electronic states were obtained using single-point calculations with PWPB95 functional (41) and the def2-TZVPP basis was set for the optimized structures. Since the B3LYP functional cannot well describe the medium to long range dispersion interactions of the reaction system, D3-BJ correction developed by Grimme was used to correct the dispersion calculation in all

systems (42, 43). In addition, structure optimization and frequency calculations were performed synchronously to confirm that the structure was in the ground state.

Random Forest Model. The dataset used in this study includes 33 PCBs with 19 features (Table S3) to build the binary classification model for predicting the human CYP2E1-dependent mutagenicity of PCBs. These features were collected from molecular docking, GFN2-xtb calculation, DFT calculation and CompTox Chemicals Dashboard (44). Thirty % of the data were randomly sampled from the dataset and set as a test dataset to quantify the prediction performance of the developed Random Forest models (45). The number of trees in the ensemble (*ntree*) was set to 1000 and the terminal node size (*nodesize*) was 5. The number of randomly selected predictor variables at each node (*mtry*) was determined by grid search tuning based on lowest RMSE (root-mean-square error) with a 5-fold cross-validation and 20 repeats (46, 47). To determine the performance of the model the mean values of the parameters (including accuracy, precision, recall rate, and F1 score) were determined.

3. Results

3.1. The significance of conformational flexibility of PHE residues for the affinity of aromatic substrates to human CYP2E1

As previously reported, under a molecular docking condition with the residues of AAs forming the active site of human CYP2E1 set as rigid conformations, the results of the docking of benzene and its oxidized products to the enzyme are in good agreement with the experimentally obtained positive genotoxicity results in mammalian cell lines genetically engineered for stable expression of human CYP2E1 (19). It appears that setting PHEs as flexible residues is generally unnecessary for docking small-sized ligands to human CYP2E1. However, bulky ligands like 1-MP, which have four conjugated phenyl rings, cannot be docked into the active site of human CYP2E1 under rigid docking settings, as shown in the Supplementary Material (Table S1, based on the structure of 3E6I). Therefore, PHE 478, which plays an important role in forming some tunnels in the human CYP2E1 for substrate binding, was set as a flexible residue so that a larger cavity as the active site can be formed. As a result, 1-MP was successfully docked into the active site of human CYP2E1, with a binding energy of -9.99 kcal/mol, where the conformational flexibility of PHE 478 residue (through rotating and moving) provided sufficient space for the entry of 1-MP (Figure 1 B). However, the orientations of the benzyl group in 1-MP, which has been proved to be the main hydroxylation site, did not permit a proper distance (< 6.0 Å) to heme (actually 10.4 Å).

To further understand the contribution of PHEs to effective binding of 1-MP to human CYP2E1, all the PHEs (including PHE 106, PHE 115, PHE 207, PHE 298, and PHE 478) in the active site of human CYP2E1 were set as flexible residues. As shown in Figure 1C, the binding of 1-MP was thus in a favorable orientation for its hydroxylation at the benzyl group. The PHE residues induced a way of 1-MP binding to the active site by rotation and forming π - π bonds to stabilize the ligand binding. Moreover, similar docking results optimized by setting flexible PHEs were observed in other CYP2E1 structures derived from different PDB sources (Table S1).

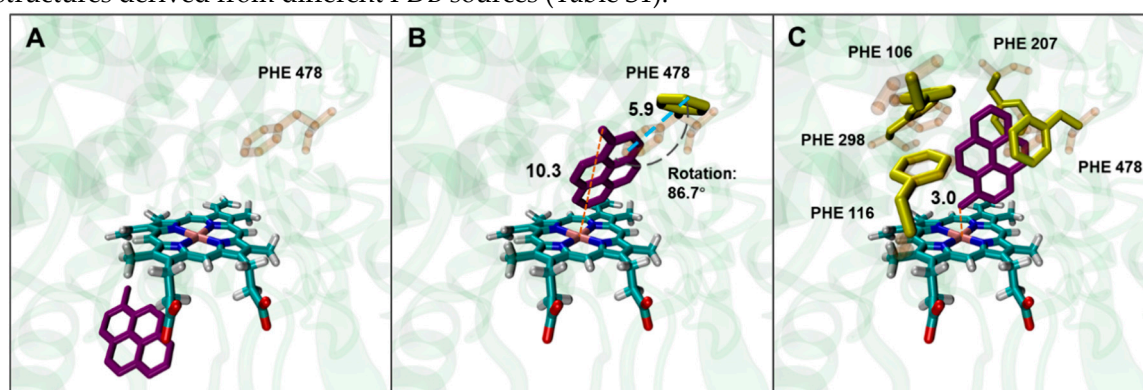


Figure 1. Conformations of 1-MP docked into human CYP2E1 under different settings for PHE residues. A: Docking with rigid settings for all AA residues in the active site; B: PHE 478 being set as a flexible residue in the active site for docking; C: all PHEs in the active site being set as flexible residues for docking. Orange dotted lines represent distances between the C α atom of 1-MP and Fe ion (Å); the blue dotted line represents the range of the movement of PHE 478 (Å); the black dotted line represents the rotation of PHE 478.

3.2. Impact of PHE 298 and PHE 478 residues on the binding of 1-MP to human CYP2E1 and its orientation

Tunnel analysis showed that tunnels 2c and 2b were favorable for 1-MP to approach the active site of human CYP2E1 (Table S2, Figure S1). As PHE 298 and PHE 478 are important controllers of the “gate” of 2c and 2b tunnel, respectively (4, 48), we constructed mutated protein models with alanine residue at both sites (F289A and F478A), then the role of these PHEs in the binding of 1-MP to the enzyme was analyzed. The docking results indicated that 1-MP was able to bind to the active site of each mutant CYP2E1, but with a lower affinity than that of the wild enzyme (Table 1). Further molecular dynamics simulations were performed based on the best conformations obtained from molecular docking. As shown in Figure 2A, the RMSD of each complex was gradually convergent and the system remained in equilibrium status during the last 40 ns. However, the distances between 1-MP (C α atom) and the heme (Fe ion) differed greatly in various complexes. Illustrated in Figure 2B, 1-MP stably bound to the wild-type model with distances around 5 Å, while its binding to F298 mutant was unstable, with long distances (mostly >10 Å) from the heme; and in case of F478A mutant drastic fluctuation in the distance was observed, making stable binding impossible.

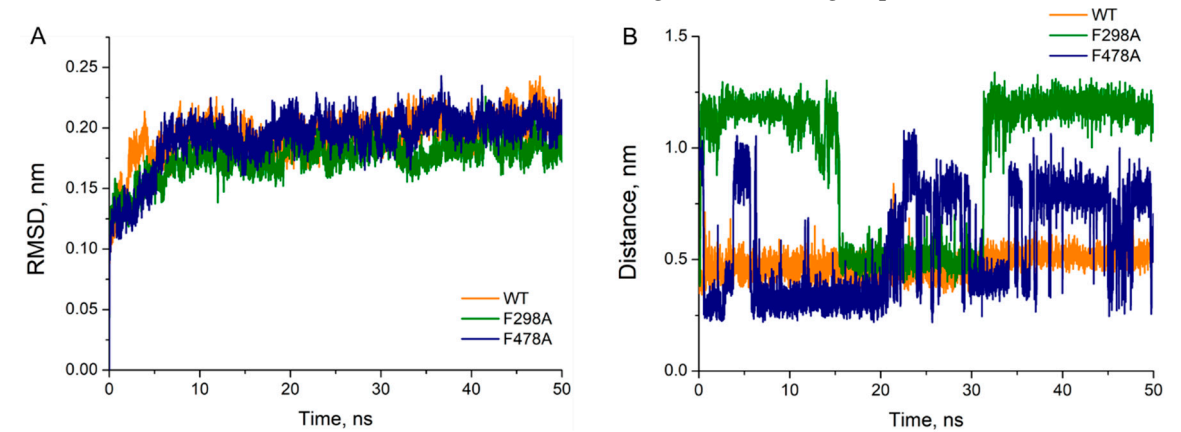


Figure 2. RMSD (A) of wild-type and mutant human CYP2E1 protein bound with 1-MP and the variation of distance between the C α in 1-MP and the heme in a period of 50 ns (B).

The last 10 ns of each complex’s trajectory were extracted for further MMPBSA calculations. As shown in Table 3, 1-MP was tightly bound to the active site of each CYP2E1 model, but the binding energy of each 1-MP-mutant protein complex was less negative than that of 1-MP-wild protein complex. Moreover, the energy contributions from the residues at the active site of the mutant CYP2E1 models differed from that of the wild type model (Figuer 3 and Table S3). Although the energy contributions from PHE 106 and PHE 116 were slightly increased (with that from PHE 207 and PHE 298 unchanged) in both mutants, mutations on PHE 478 caused a steep energy change, which suggests an important role of PHE 478 in the ligand binding and orientation (Figure 3).

Table 1. Free energy of 1-MP for binding into the active site of wild type and mutant human CYP2E1 protein.

CYP2E1 models	Energy for 1-MP binding to the active site of wild-type/mutant human CYP2E1, kJ/mol				
	Coulombic	Lennard-Jones	Polar solvation	Non-polar solvation	Total

WT	-7.55 ± 1.08	-148.50 ± 7.43	31.59 ± 2.63	-16.75 ± 0.32	-141.20 ± 7.65
F298A	-1.92 ± 1.77	-135.69 ± 7.86	34.12 ± 2.28	-17.20 ± 0.20	-120.68 ± 7.56
F478A	-2.12 ± 1.83	-134.83 ± 7.35	44.45 ± 5.21	-17.52 ± 0.37	-110.02 ± 9.35

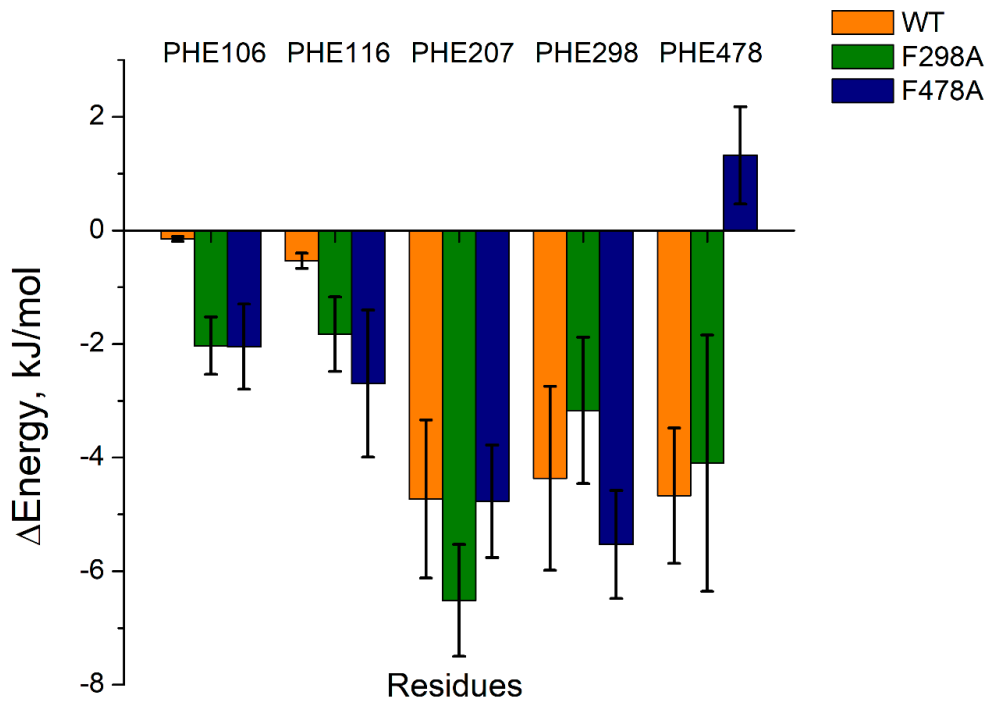


Figure 3. The energy contributions from different PHEs in the active site of 1-MP-wild-type CYP2E1 and that of each 1-MP-mutant CYP2E1 complex.

3.3. Consistency between molecular simulation results and historical data of human CYP2E1-activated mutagenicity of PCBs in genetically engineered mammalian cells

Based on previous experimental studies, a series of PCBs are metabolically activated by human CYP2E1 for mutagenic effects in mammalian cells. Therefore, this study utilized a random forest model that incorporated various factors, such as flexible molecular docking of PHEs, semi-empirical calculations of PCB-CYP2E1 complex, conceptual density functional parameters of PCB-PHE complex, and the physical/chemical property of each PCB (such as molecular weight, chlorine substitution pattern, octanol-water partition coefficient, concentrations of hydroxylated PCB metabolites in air, and their half-lives in fish) to investigate the association between these factors and human CYP2E1-activated PCB mutagenicity, as well as the role of PHEs in the activating process (Table S4).

The results indicated that the random forest model, which was constructed by taking multiple factors into consideration, was able to predict human CYP2E1-dependent PCB mutagenicity with some accuracy, as demonstrated by a coincidence rate of 79%, a precision of 79.3%, a recall of 90%, and an f1 score of 0.838. Furthermore, a detailed analysis of the importance of each factor indicated that some factors, such as the ligand-to-heme distance and docking score obtained by molecular docking, the nucleophilicity index (based on conceptual density functional theory), HOMO-LUMO gap of the ligand, and half-life in fish, were significant determinants for predicting the human CYP2E1-dependent PCB mutagenicity (Figure 4). Among these factors, the flexible molecular docking of PHEs exhibited a dominant influence. Conversely, through quantum chemistry and semi-empirical analysis PHEs were not significantly influential on the intrinsic properties of PCBs, which suggests that PHEs may play an important role in regulating the binding of each PCB to the active site rather than affecting the electronic/chemical properties of ligands.

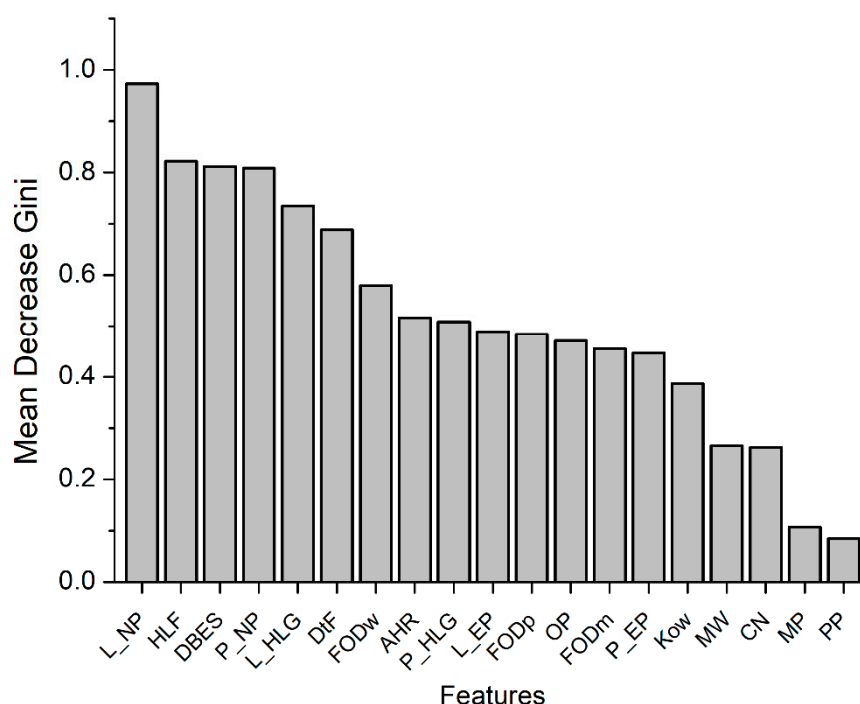


Figure 4. Relative importance of each molecular feature of PCBs in the random forest model. L_NP, the nucleophilicity of ligand; HLF, fish biotransformation half-life; DBES, binding energy score obtained from molecular docking with PHE478 set flexible; P_NP, the nucleophilicity of ligand-PHE complex; L_HLG, the HOMO-LUMO gap of ligand; DtF, the distance from SOM to Fe ion in heme; FODw, the FOD value of ligand binding to the active site of human CYP2E1; AHR, atmospheric hydroxylation rate; P_HLG, the HOMO-LUMO gap of ligand-PHE complex; L_EP, the electrophilicity of ligand; FODp, the FOD value of ligand; OP, the number of ortho-Cl-substitution; FODm, the FOD value of ligand binding to the active site of F478A mutant; P_EP, the electrophilicity of ligand-PHE complex; Kow, LogKow: octanol-water partition coefficient; MW, the molecular weight; CN, the number of Cl-substitution; MP, the number of meta-Cl-substitution; PP, the number of para-Cl-substitution.

4. Discussion

The steric shape of the cavity and the AA composition of the active site are important factors determining the substrate selectivity of an enzyme. In the active sites of the CYP1 enzymes a PHE cluster exists above the heme, thereby they tend to metabolize planar substrates such as polycyclic aromatic hydrocarbons, dioxins, and dioxin-like PCBs (49, 50). The substrate selectivity of CYP2 isoforms varies greatly, among which CYP2A and 2B prefer neutral compounds as their substrates, where the substrates of CYP2B are generally more lipophilic than those of CYP2A. CYP2C and CYP2D, which contain relatively low numbers of PHEs in their active sites, are more likely to catalyze the metabolism of weakly acidic and weakly basic substances, respectively (49, 51). The differences in the number and position of PHE residues may affect the selectivity for, and the affinity of, substrates. Studies have shown that in human CYP3A4, PHE108, PHE213, PHE220, and PHE241 are involved in the allosteric effects in F helix/G helix region, resulting in an extension of helix F and a larger active site; while in CYP3A5, the replacement of PHE at position 108 with LEU leads to changed plasticity of the F/G helix region, producing a substrate-binding pattern different from that in CYP3A4 (52). After mutating the HIS85 of CYP101B1 to PHE, there appeared several changes: significantly increased rate of the binding of hydrophobic substrates to the enzyme, decreased affinity for β -ionone, and shifted site of oxidation (53). For CYP2E1, the high proportion of PHE residues in

its active site may also play an important role in its selectivity of substrates and their binding orientation (4).

For benzene and its oxidized metabolites, which are known mutagens activated by human CYP2E1, the docking results were highly consistent with their mutagenic effects observed *in vitro* (19). Benzene exhibited the strongest affinity toward the enzyme, while the ultimate mutagen 1,4-benzoquinone was difficult to enter the active site of CYP2E1. As indicated in this study, since benzene and its hydroxylated products are relatively small compared to aromatic compounds containing two and more phenyl groups, the PHEs in the active site of CYP2E1 did not show significant regulation on the orientation of the small ligands. On the other hand, under rigid docking conditions, the space of the active site of human CYP2E1 is too small to accommodate larger molecules which contain multiple benzene rings (such as 1-MP), but this barrier (caused by limited cavity size) can be eliminated by increased conformational flexibility of PHEs (4, 54). *In vitro* tests indicated that human CYP2E1 can bind to substrates via ligand recognition and induced fit (55). Furthermore, molecular dynamics simulations showed that human CYP proteins can make conformational change through the activities of the F/G loop, B/C loop, and corresponding α helices, thus the volume of the active site of CYP2E1 can expand from 220 Å³ to 1310 Å³ (when F/G loop is fully open) (55).

Some AAs play an important role in the conformational changes and cavity expansion in CYP enzymes, among which PHE 478 or PHE at other positions (such as PHE 476 in human CYP2C9) are important "gatekeeper" amino acids for regulating the entry of substrate to, and its exit from, the active site of each CYP enzyme, and they have an important impact on the long-range molecular interactions between substrates and CYPs (4, 56). In this study, *via* molecular simulation two PHE mutants of CYP2E1 were built for the analysis of the potential functions of PHE298 and PHE478 in substrate binding and orientation. The results showed that the mutation of either PHE residue had a negative impact on the binding affinity and orientation (for catalyzing a biotransformation reaction) of 1-MP to CYP2E1, and the mutation of PHE478 showed the most significant impact on the binding status. In addition, it is predicted that the rotation of PHE 478 can connect the catalytic site to the secondary binding region and form a larger space to hold bulky substrates (54). Therefore, the conformational changes of PHE 478 in human CYP2E1 might be necessary for the access of large-sized substrates, including 1-MP and PCBs, to the active site. Our docking results showed that after the rotation of PHE 478 to some extent, the cavity of the active site can be enlarged, allowing these bulky molecules to bind to the active site of human CYP2E1. Nevertheless, even under flexible settings for PHE478 the benzylic group of 1-MP is still too far from the heme to permit a biochemical reaction (13). After setting all PHE residues in the active site as flexible, most conformations showed that the benzylic group became closer to the heme, which is consistent with the observed production of benzylic hydroxylation by the enzyme (18). Therefore, our results suggest that the flexibility of PHEs may be critical for catalyzing reactions of aromatic substrates, as the aromatic AA residues in the active site can easily form non-covalent interactions such as π - π interaction with ligands, thus provide a sufficiently large cavity for the binding of large substrates, and further adjust their orientation (54, 57, 58).

It is established that the interactions between the residues of some AAs and a substrate may affect the process of the transfer of proton/electron, thereby enhancing their reactivity of the substrate and/or reducing the barrier from a reaction (59, 60). This study has further explored potential factors involved in substrate-enzyme interaction for the catalysis of biotransformation reactions, by using a model constructed with data calculated in various ways. The results indicated that some factors with the substrates, including nucleophilicity, HOMO-LUMO gap, *in vivo* half-life, and molecular docking results based on flexible setting of PHE conformation, contributed substantially to the prediction of human CYP2E1-activated mutagenicity of PCBs. It looks that PHEs in the active site of human CYP2E1 may participate in locating the site of reaction and the formation of products by enlarging cavity size for holding bulky substrates and regulating their orientations. However, the list of PCBs in this study was not based on random choice, but was guided by ever observed structure-genotoxicity relationships (10-13), thus a bias may have been introduced, leading to relatively high

predictive value of molecular docking for human CYP2E1-activated PCB mutagenicity (12). Therefore, in the future it is necessary to use substrates of random and more diverse structural patterns in the study of PHE-CYP2E1 interaction, to further identify the role of PHEs in regulating the metabolism of aromatic substrates by human CYP2E1 and the prediction of metabolically activated toxicity.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Two major tunnels for ligands entering the active site in human CYP2E1; Table S1: Molecular docking of 1-MP to human CYP2E1 under rigid and flexible settings with regard to the cavity volume, binding energy, and ligand-heme distance; Table S2: Features of high rank tunnels for the binding of 1-MP to the active center of human CYP2E1; Table S3: The energy decomposition of non-PHE AA residues in the active sites of 1-MP-bound wild-type and mutated human CYP2E1; Table S4: The molecular descriptors of PCB congeners as candidate factors influencing their binding to human CYP2E1 and orientation.

Author Contributions: methodology, K.H. and Z.Y.; software, K.H. and Z.Y.; validation, Y.C. and Z.L.; formal analysis, K.H. and J.X.; writing—original draft preparation, K.H. and H.T.; writing—review and editing, Y.L.; funding acquisition, K.H., H.T., and Y.L.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 21577054, 42107310 and 82003487. .

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Montellano PROd. Cytochrome P450 Structure, Mechanism, and Biochemistry. 4 ed: Springer International Publishing; 2015.
2. Williams PA, Cosme J, Vinkovic DM, Ward A, Angove HC, Day PJ, et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science*. 2004;305(5684):683-6.
3. Spatzenegger M, Liu H, Wang Q, Debarber A, Koop DR, Halpert JR. Analysis of differential substrate selectivities of CYP2B6 and CYP2E1 by site-directed mutagenesis and molecular modeling. *J Pharmacol Exp Ther*. 2003;304(1):477-87.
4. Porubsky PR, Meneely KM, Scott EE. Structures of human cytochrome P-450 2E1. Insights into the binding of inhibitors and both small molecular weight and fatty acid substrates. *J Biol Chem*. 2008;283(48):33698-707.
5. Li QS, Ogawa J, Schmid RD, Shimizu S. Indole hydroxylation by bacterial cytochrome P450 BM-3 and modulation of activity by cumene hydroperoxide. *Biosci Biotechnol Biochem*. 2005;69(2):293-300.
6. Hayes C, Ansbro D, Kontoyianni M. Elucidating substrate promiscuity in the human cytochrome 3A4. *J Chem Inf Model*. 2014;54(3):857-69.
7. Muralidhara BK, Sun L, Negi S, Halpert JR. Thermodynamic fidelity of the mammalian cytochrome P450 2B4 active site in binding substrates and inhibitors. *J Mol Biol*. 2008;377(1):232-45.
8. Jiang H, Lai Y, Hu K, Wei Q, Liu Y. Human CYP2E1-dependent and human sulfotransferase 1A1-modulated induction of micronuclei by benzene and its hydroxylated metabolites in Chinese hamster V79-derived cells. *Mutat Res*. 2014;770:37-44.
9. Jiang H, Lai Y, Hu K, Chen D, Liu B, Liu Y. Genotoxicity of 1-methylpyrene and 1-hydroxymethylpyrene in Chinese hamster V79-derived cells expressing both human CYP2E1 and SULT1A1. *Environ Mol Mutagen*. 2015;56(4):404-11.
10. Zhang C, Lai Y, Jin G, Glatt H, Wei Q, Liu Y. Human CYP2E1-dependent mutagenicity of mono- and dichlorobiphenyls in Chinese hamster (V79)-derived cells. *Chemosphere*. 2016;144:1908-15.
11. Liu Y, Hu K, Jia H, Jin G, Glatt H, Jiang H. Potent mutagenicity of some non-planar tri- and tetrachlorinated biphenyls in mammalian cells, human CYP2E1 being a major activating enzyme. *Arch Toxicol*. 2017;91(7):2663-76.
12. Chen Y, Zhu N, Luo Y, Hu K, Liu Y. Featured structure-activity relationships for some tri- and tetrachlorobiphenyls in human CYP2E1-activated mutagenicity - Impact of the extent of ortho-chlorination. *Chemosphere*. 2018;210:467-75.
13. Hu K, Yu H, Li Z, Jin G, Jia H, Song M, et al. Human CYP2E1-activated mutagenicity of dioxin-like PCBs 105 and 118-Experimental data consistent with molecular docking results. *Toxicology*. 2020;437:152438.
14. Bi X, Sheng G, Feng Y, Fu J, Xie J. Gas- and particulate-phase specific tracer and toxic organic compounds in environmental tobacco smoke. *Chemosphere*. 2005;61(10):1512-22.

15. Lian JJ, Ren Y, Chen JM, Wang T, Cheng TT. Distribution and source of alkyl polycyclic aromatic hydrocarbons in dustfall in Shanghai, China: the effect on the coastal area. *Journal of environmental monitoring* : JEM. 2009;11(1):187-92.
16. Larsson BK, Sahlberg GP, Eriksson AT, Busk LA. Polycyclic aromatic hydrocarbons in grilled food. *Journal of agricultural and food chemistry*. 1983;31(4):867-73.
17. Richter-Brockmann S, Achten C. Analysis and toxicity of 59 PAH in petrogenic and pyrogenic environmental samples including dibenzopyrenes, 7H-benzo[c]fluorene, 5-methylchrysene and 1-methylpyrene. *Chemosphere*. 2018;200:495-503.
18. Engst W, Landsiedel R, Hermersdörfer H, Doehmer J, Glatt H. Benzylic hydroxylation of 1-methylpyrene and 1-ethylpyrene by human and rat cytochromes P450 individually expressed in V79 Chinese hamster cells. *Carcinogenesis*. 1999;20(9):1777-85.
19. Hu K, Cai L, Li Z, Glatt H, Shi M, Liu Y. Human CYP2E1-Dependent Mutagenicity of Benzene and Its Hydroxylated Metabolites in V79-derived cells: Suppression and Enhancement by Ethanol Pretreatment. *Environ Mol Mutagen*. 2020.
20. Zhang H, Yolton K, Webster GM, Sjodin A, Calafat AM, Dietrich KN, et al. Prenatal PBDE and PCB exposures and reading, cognition, and externalizing behavior in children. *Environ Health Perspect*. 2017;125(4):746-52.
21. Berg V, Nost TH, Pettersen RD, Hansen S, Veyhe AS, Jorde R, et al. Persistent organic pollutants and the association with maternal and infant thyroid homeostasis: a multipollutant assessment. *Environ Health Perspect*. 2017;125(1):127-33.
22. Valvi D, Mendez MA, Martinez D, Grimalt JO, Torrent M, Sunyer J, et al. Prenatal concentrations of polychlorinated biphenyls, DDE, and DDT and overweight in children: a prospective birth cohort study. *Environ Health Perspect*. 2012;120(3):451-7.
23. Ghosh S, Loffredo CA, Mitra PS, Trnovec T, Palkovicova Murinova L, Sovcikova E, et al. PCB exposure and potential future cancer incidence in Slovak children: an assessment from molecular finger printing by Ingenuity Pathway Analysis (IPA®) derived from experimental and epidemiological investigations. *Environmental science and pollution research international*. 2018;25(17):16493-507.
24. Uwimana E, Ruiz P, Li X, Lehmler HJ. Human CYP2A6, CYP2B6, and CYP2E1 atropselectively metabolize polychlorinated biphenyls to hydroxylated metabolites. *Environ Sci Technol*. 2019;53(4):2114-23.
25. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *Journal of chemical theory and computation*. 2012;8(9):3257-73.
26. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1983;79(2):926-35.
27. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19-25.
28. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*. 2012;4(1):17.
29. Bannwarth C, Ehlert S, Grimme S. GFN2-xTB-An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of chemical theory and computation*. 2019;15(3):1652-71.
30. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*. 2009;30(16):2785-91.
31. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*. 1998;19(14):1639-62.
32. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*. 2010;31(4):671-90.
33. Wang C, Nguyen PH, Pham K, Huynh D, Le TB, Wang H, et al. Calculating protein-ligand binding affinities with MMPBSA: Method and error analysis. *Journal of computational chemistry*. 2016;37(27):2436-46.
34. Li J. The introduction of gmx_mmpbsa script 2019 [Available from: https://jerkwin.github.io/2019/07/31/gmx_mmpbsa%E4%BD%BF%E7%94%A8%E8%AF%B4%E6%98%8E/].
35. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol*. 2012;8(10):e1002708.
36. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-8.
37. Grimme S, Hansen A. A practicable real-space measure and visualization of static electron-correlation effects. *Angewandte Chemie (International ed in English)*. 2015;54(42):12308-13.
38. Neese F, Wennmohs F, Becker U, Riplinger C. The ORCA quantum chemistry program package. *The Journal of chemical physics*. 2020;152(22):224108.

39. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry*. 1994;98(45):11623-7.
40. Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical chemistry chemical physics : PCCP*. 2005;7(18):3297-305.
41. Zhao Y, Truhlar DG. Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *The Journal of Physical Chemistry A*. 2005;109(25):5656-67.
42. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics*. 2010;132(15):154104.
43. Grimme S, Ehrlich S, Goerigk L. Effect of the damping function in dispersion corrected density functional theory. *Journal of computational chemistry*. 2011;32(7):1456-65.
44. Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of cheminformatics*. 2017;9(1):61.
45. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32.
46. Shao J. Linear Model Selection by Cross-validation. *Journal of the American Statistical Association*. 1993;88(422):486-94.
47. Konovalov DA, Llewellyn LE, Vander Heyden Y, Coomans D. Robust Cross-Validation of Linear Regression QSAR Models. *Journal of Chemical Information and Modeling*. 2008;48(10):2081-94.
48. Porubsky PR, Battaile KP, Scott EE. Human cytochrome P450 2E1 structures with fatty acid analogs reveal a previously unobserved binding mode. *J Biol Chem*. 2010;285(29):22282-90.
49. Lewis DF. Human cytochromes P450 associated with the phase 1 metabolism of drugs and other xenobiotics: a compilation of substrates and inhibitors of the CYP1, CYP2 and CYP3 families. *Curr Med Chem*. 2003;10(19):1955-72.
50. Lewis DF. Quantitative structure-activity relationships in substrates, inducers, and inhibitors of cytochrome P4501 (CYP1). *Drug metabolism reviews*. 1997;29(3):589-650.
51. Lewis DF, Jacobs MN, Dickins M. Compound lipophilicity for substrate binding to human P450s in drug metabolism. *Drug Discov Today*. 2004;9(12):530-7.
52. Hsu MH, Johnson EF. Active-site differences between substrate-free and ritonavir-bound cytochrome P450 (CYP) 3A5 reveal plasticity differences between CYP3A5 and CYP3A4. *J Biol Chem*. 2019;294(20):8015-22.
53. Sarkar MR, Lee JHZ, Bell SG. The oxidation of hydrophobic aromatic substrates by using a variant of the P450 monooxygenase CYP101B1. *Chembiochem*. 2017;18(21):2119-28.
54. Liu Y, Liu BY, Hao P, Li X, Li YX, Wang JF. π - π Stacking mediated drug-drug interactions in human CYP2E1. *Proteins*. 2013;81(6):945-54.
55. Hendrychova T, Berka K, Navratilova V, Anzenbacher P, Otyepka M. Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr Drug Metab*. 2012;13(2):177-89.
56. Cojocar V, Balali-Mood K, Sansom MS, Wade RC. Structure and dynamics of the membrane-bound cytochrome P450 2C9. *PLoS Comput Biol*. 2011;7(8):e1002152.
57. Mena-Ulecia K, MacLeod-Carey D. Interactions of 2-phenyl-benzotriazole xenobiotic compounds with human Cytochrome P450-CYP1A1 by means of docking, molecular dynamics simulations and MM-GBSA calculations. *Comput Biol Chem*. 2018;74:253-62.
58. Ang SJ, Mak AM. Nature of halogen bonding involving π -systems, nitroxide radicals and carbenes: a highlight of the importance of charge transfer. 2018;20(41):26463-78.
59. Chowdhury AS, Ali HS, Faponle AS, de Visser SP. How external perturbations affect the chemoselectivity of substrate activation by cytochrome P450 OleT(JE). *Physical chemistry chemical physics : PCCP*. 2020;22(46):27178-90.
60. de Visser SP, Mukherjee G, Ali HS, Sastri CV. Local Charge Distributions, Electric Dipole Moments, and Local Electric Fields Influence Reactivity Patterns and Guide Regioselectivities in α -Ketoglutarate-Dependent Non-heme Iron Dioxygenases. *Accounts of chemical research*. 2022;55(1):65-74.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.