

Revisiting Aldehyde Oxidase Mediated Metabolism in Drug-like Molecules: An Improved Computational Model

Jihui Zhao, Rongrong Cui, Lihao Wang, Yingjia Chen, Zunyun Fu, Xiaoyu Ding, Chen Cui, Tianbiao Yang, Xutong Li, Yuan Xu, Kaixian Chen, Xiaomin Luo, Hualiang Jiang, and Mingyue Zheng

J. Med. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.jmedchem.9b01895 • Publication Date (Web): 19 Mar 2020

Downloaded from pubs.acs.org on March 20, 2020

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

1
2
3
4
5
6
7 Revisiting Aldehyde Oxidase Mediated Metabolism
8
9
10
11 in Drug-like Molecules: An Improved
12
13
14
15 Computational Model
16
17
18
19

20 *Jihui Zhao,^{a,b,‡} Rongrong Cui,^{a,c,‡} Lihao Wang,^d Yingjia Chen,^a Zunyun Fu,^{a,c} Xiaoyu Ding,^a*
21
22 *Chen Cui,^a Tianbiao Yang,^a Xutong Li,^a Yuan Xu,^e Kaixian Chen,^{a,c,f} Xiaomin Luo,^a Hualiang*
23
24 *Jiang,^{a,c,f,*} Mingyue Zheng^{a,b,*}*

25
26
27
28 ^aState Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese
29 Academy of Sciences, 555 Zu Chong Zhi Road, Shanghai, 201203, China

30
31
32
33 ^bUniversity of Chinese Academy of Sciences, No.19(A) Yuquan Road, Shijingshan
34 District, Beijing, 100049, China

35
36
37 ^cCollege of Pharmacy, Nanjing University of Chinese Medicine, 138 Xianlin Avenue, Qixia
38 District, Nanjing, 210023, China

39
40
41 ^dGillings School of Public Health, University of North Carolina at Chapel Hill, 135 Dauer
42 Dr, Chapel Hill, NC 27599, the United States

43
44
45 ^eShanghai EnnovaBio Pharmaceuticals Co.,Ltd, Room 404, Building 2, Lane 720, Cailun Road,
46 Pudong New Area, Shanghai

47
48
49 ^fShanghai Institute for Advanced Immunochemical Studies, and School of Life Science
50
51
52 and Technology, ShanghaiTech University, Shanghai 200031, China
53
54
55
56
57
58
59
60

ABSTRACT

Aldehyde oxidase (AOX) is a drug metabolizing molybdo-flavoenzyme that has gained increasing attention because of contribution to the biotransformation in phase I metabolism of xenobiotics. Unfortunately, the intra- and inter-species variations in AOX activity and lack of reliable and predictive animal models make evaluation of AOX-catalyzed metabolism prone to be misleading. In this study, we developed an improved computational model integrating both atom-level and molecule-level features to predict whether a drug-like molecule is a potential human AOX (hAOX) substrate and to identify the corresponding sites of metabolism. Additionally, we combined the proposed computational strategy and in vitro experiments for evaluating the metabolic property of a series of epigenetic-related drug candidates still in early stage of development. In summary, this study provides an improved strategy to evaluate the liability of molecules toward hAOX and offers useful information for accelerating the drug design and optimization stage.

INTRODUCTION

Aldehyde oxidase (AOX), a drug metabolizing molybdo-flavoenzyme distributing mainly in mammal cytoplasm, has broad substrate specificity and capability to catalyze the metabolism of various endogenous and exogenous transformation, including oxidation of aldehydes and nitrogen-containing heterocycles,¹ hydrolysis of amides,² reduction of N-oxides and S-oxides.³⁻⁴ AOX widely distributes with profound species differences in vivo and plays an important role in phase I metabolism of xenobiotics. It has been reported that AOX contributes to clinical failures by metabolizing molecules with N-containing heterocyclic aromatic ring, in which the aromatic N is introduced to stabilize the metabolism by CYP-450s and consequently increase the susceptibility to hAOX.⁵ Ever-growing small molecule candidates have failed due to hAOX-mediated oxidative metabolism.⁶ For example, INC280 (Capmatinib), PF-04217903, SGX-523 and others have been terminated during early phase of clinical trials due to AOX-catalyzed biotransformation.^{3, 6-7} However, the different experimental medium or poor in vitro-in vivo extrapolation may significantly influence the estimation of hAOX-mediated metabolism, not to mention the uncertainty caused by species differences. Under these circumstances, computational modeling can be considered as a complementary strategy to predict hAOX metabolic profiles with homologous experimental data.

Recently, Manevski et al. discussed thoroughly about the biological functions, catalytic mechanism, experimental assessment and the rational drug design approaches to modulate AOX-mediated metabolism in drug design and discovery.³ The structure of hAOX contains three domains: (1) The flavin adenine dinucleotide (FAD) domain, (2) two [2Fe-2S] redox domain and (3) the molybdenum cofactor (MoCo) and substrate binding domain.⁸ The three domains are

essential for hAOX-mediated biotransformation of substrate.⁹ Several studies have been reported about the mechanism of hAOX-catalyzed oxidation of azaheterocycle molecules,¹⁰⁻¹² which can be briefly summarized as follows: (1) The nucleophilic attack at the electron deficient C of the substrate is initialized by the hydroxyl of the MoCo; (2) The H attached to the electron deficient C transfers to the S of MoCo simultaneously, and then the substrate binds to the MoCo by forming the C-O-Mo bond;¹⁰ (3) The electron deficient C is nucleophilically attacked by one water molecule from solvent, and the substrate is released as an oxidation product.¹³ The proposed mechanism of AOX-mediated oxidation was included in the Figure 1.

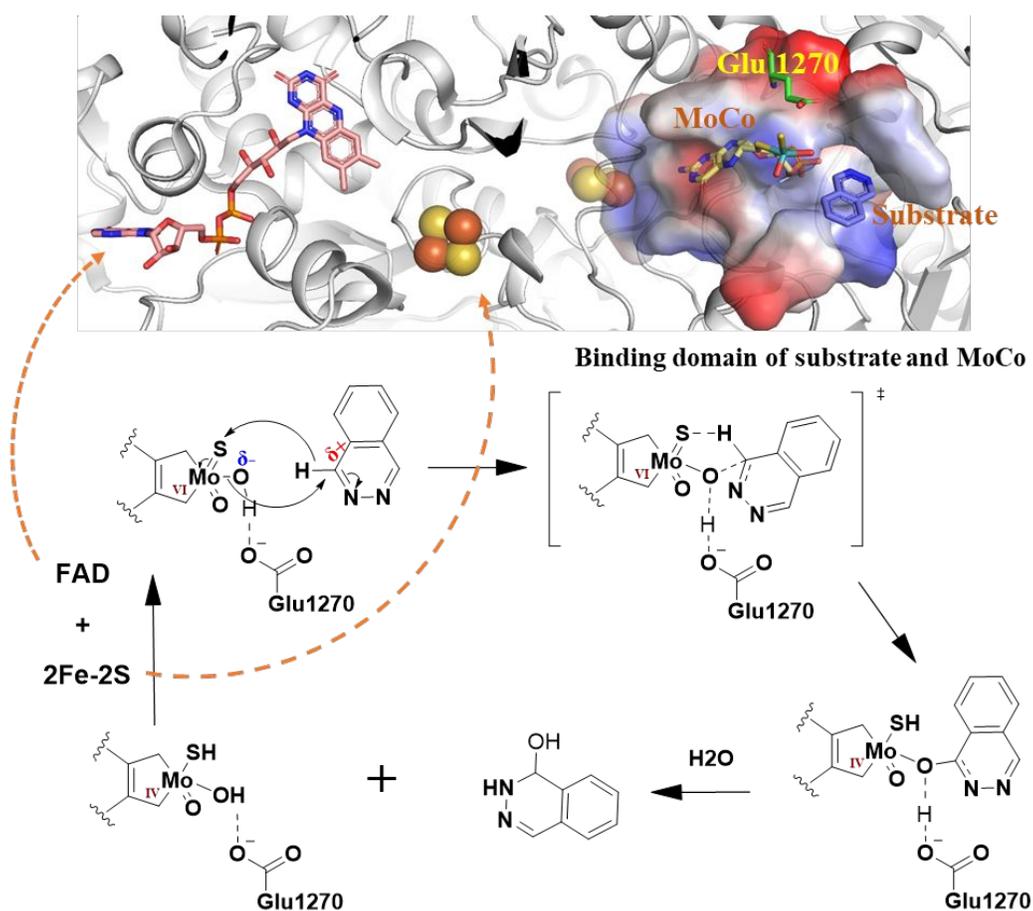


Figure 1. The scheme showing the proposed mechanism of AOX-mediated oxidation. PDB structure used for illustration is 4UHX,⁹ and the substrate is phthalazine.

1
2
3 Inspired by the catalytic mechanism of the hAOX mediated metabolism, several in silico models
4 have been reported to predict whether a molecule can be metabolized by hAOX and where its
5 potential sites of metabolism (SOMs) locate in the structure. Torres et al. used the energy of the
6 tetrahedral intermediate formation to predict the regioselectivity in hAOX-catalyzed
7 metabolisms.¹⁰ Dalvie et al. employed a similar strategy to predict the structure-metabolism
8 relationship (SMR) for zoniporide analogues.¹⁴ They proposed that the energies of formation of
9 tetrahedral intermediates could establish a possible relationship with the occurrence of a
10 nucleophilic attack, as well as the electrostatic potential (ESP) charge of the carbon adjacent to the
11 nitrogen atom and the energy of lowest unoccupied molecular orbital (E_{LUMO}). Besides, Jones et
12 al. used the energy of the tetrahedral intermediate and steric hindrance as two descriptors to predict
13 the intrinsic clearance for eight drugs metabolized by hAOX.¹⁵ Based on similar descriptors, we
14 constructed a decision tree model to predict the susceptibility of molecules toward hAOX.⁶ More
15 recently, Lepri et al. experimentally evaluated a large amount of small molecules containing
16 hAOX-substrate-like motifs and summarized the general SMR rules for hAOX mediated
17 metabolism.¹⁶ Their work proposed that the electronic and exposure effects were two important
18 features and mutually influenced, in which the electronic effects were calculated using density
19 functional theory and exposure effects were assessed by molecular docking analysis to consider
20 the ligand exposure in hAOX catalytic site. Montefiori et al. explored the regioselectivity of a
21 series of 6-Substituted 4-Quinazolinones using the concerted mechanism as a reactive model.¹²
22 Their results indicated that the stability of the hydroxylated tetrahedral intermediate could be used
23 to determine whether a site in the AOX substrate was a likely SOM. In their study, the ESP charge
24 of the electron deficient C showed good correlation with the stability of the hydroxylated
25 tetrahedral intermediate, which could be used to separate the SOMs and nonSOMs.¹² Cruciani et

1
2
3 al. developed a probability function for hAOX-metabolized oxidation with 88% prediction
4 accuracy based on a large set of homogeneous experimental data, in which both exposure and
5 electrophilic effects were taken into account.¹⁷ Furthermore, Montefiori et al. collected a dataset
6 of 78 substrates with known SOMs information and constructed models to predict the AOX-
7 mediated SOMs, in which the importance of the chemical shift and ESP charge of the deficient C
8 of SOMs were highlighted.¹⁸
9
10
11
12
13
14
15
16
17

18 In our previous work, the reaction energy barrier (ΔE_{inter}) and steric hindrance (**Steric**) were
19 adopted to predict the potential SOMs of the molecules toward hAOX-mediated metabolism. As
20 the results illuminated, the DT_{AOX} model exhibited good predictive performance and gave clear
21 explanation of the relationship between the exposure, reaction energy barrier effects and the
22 hAOX-mediated metabolism.⁶ Consequently, the recently published hAOX-mediated oxidative
23 metabolism data was used to evaluate the previous published DT_{AOX} model including
24 substrate/nonsubstrate data.¹⁶⁻¹⁷ The sample distribution was calculated, and more than half of
25 these recently published data falling outside of the applicability domain (AD) defined by DT_{AOX} .⁶
26
27 As a result, the model yielded a lower recall rate (alternately named as the sensitivity score, SE)
28 of 0.74, a specificity score (SP) of 0.64. Clearly, DT_{AOX} needs to be improved to predict whether
29 a molecule was a true substrate, and to identify the potential metabolic site toward hAOX-mediated
30 oxidative metabolism. Checking the incorrect predictions, we further noticed that DT_{AOX} also
31 failed to recognize the potential SOM with similar scaffolds. For example, the tendency toward
32 hAOX of position 2 in quinoline analogues were indistinguishable to the model, because these
33 sites showed similar reaction energy barrier and steric hindrance, as illustrated in Figure 2. Among
34 these quinoline analogues, only MOL038 was a hAOX substrate, and the other three were all
35 mistaken by DT_{AOX} for substrates. Apparently, more features are needed to capture the substituent
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

differences in these similar molecules. Similar observations were also reported by Cruciani et al,¹⁷ although the 5-nitro-phthalazine, 5-amino-phthalazine and phthalazine were analogues with only different substituents at same position of phthalazine, these compounds exhibited different relationship with the empirical SMR concluded by Lepri et al.¹⁶⁻¹⁷ In addition, DT_{AOX} was also evaluated toward the newly published data with known hAOX-catalyzed oxidative sites within the applicability domain.^{7,19} As summarized in Table S1, the low sensitivity score suggested that many SOMs were mispredicted to be negative by the model.

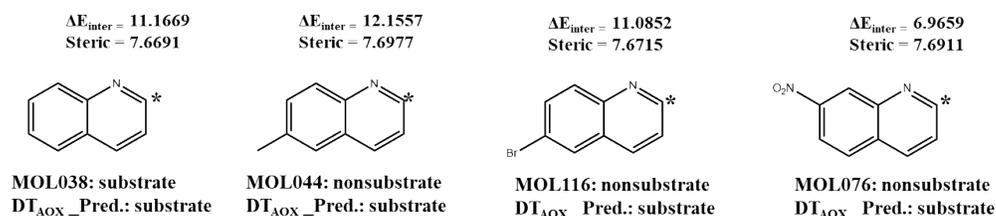


Figure 2. Site of metabolism prediction results of four quinoline analogues by DT_{AOX} model. The position 2 in these quinoline analogues are marked as black asterisk. Pred.: prediction results.

Above retrospective analyses revealed that our previous computational model DT_{AOX} trained using small data set has narrow applicability domain and insufficient generalization performance. In this context, more data and more global and local features need to be considered to improve the model. Several works have been published to predict the metabolizing enzyme mediated metabolic selectivity by constructing machine learning models with impressive predictive performance. For example, Peng et al. predicted UGT-catalyzed metabolism by constructing SVM classification model and identified 84% of experimentally observed sites of metabolisms for an external test.²⁰

1
2
3 A multitask deep autoencoder neural network model was developed by Li et al. to predict the
4 inhibitors of five major isozymes of CYP450 and outperformed traditional machine learning
5 methods.²¹ The recently published homogeneously hAOX-mediated experimental data and
6 structurally diverse molecules have laid a solid foundation for constructing a more accurate hAOX-
7 mediated drug metabolism model.
8
9
10
11
12
13

14
15 There are two objectives of this study, (1) to accurately predict whether the molecule is hAOX
16 substrate, and (2) to predict where the SOMs are located, i.e., to identify the specific site(s) within
17 substrates that is metabolized by hAOX. First, the structurally diverse, experimental homogeneous
18 dataset were collected from the latest published literature. Second, the molecular and atomic
19 properties including topological, physicochemical and electronic descriptors were calculated, and
20 the descriptors were combined to construct the substrate/nonsubstrate classification model. Third,
21 the selected key descriptors were combined to optimize the SOMs classification model. Fourth,
22 external test sets were used to verify the generalization ability of the substrate/nonsubstrate and
23 SOMs classification model. In the end, the epigenetics-related inhibitors containing hAOX-
24 substrate-like motifs were predicted using the combined model and in vitro experiments were
25 carried out for further validation.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 **RESULTS AND DISCUSSION**

42 **hAOX-catalyzed Oxidation: A View from Molecular Level to Atomic Level**

43
44
45
46
47
48
49
50 Usually, some physical-chemical properties such as logP and logD are the important factors to
51 discriminate whether a molecule is a cytochrome P450 substrate or not.²² However, for hAOX
52 substrates and nonsubstrates, their physical-chemical properties are of less discriminatory power.¹⁷
53
54
55
56
57
58
59
60

1
2
3 As mentioned above, a variety of computational models have been developed to predict the hAOX-
4 catalyzed metabolism. Based on homogenously hAOX-mediated experimental data with
5 structurally diverse skeletons recently published, we herein studied whether there were individual
6 topological and electronic features could determine the regioselectivity of hAOX-mediated
7 metabolism.
8
9
10
11
12
13
14

15 As shown in Figure 3, the hAOX substrates have broad substrate specificity with widely
16 different molecular size and geometry. Taking the distribution of molecular volume as an example,
17 we may find from Figure 3A that the interquartile range of hAOX substrates is from 230 to 345.
18 This value is wider than that of a few major isoforms of CYP450s with substrate specificity, such
19 as CYP1A2, CYP2D6 and CYP2C19.²³ Besides, the molecular weights showed similar
20 distribution among substrates and nonsubstrates, as well as the molecular volume and molecular
21 surface area (Figure 3A). Consequently, it is difficult to estimate whether a molecule is a hAOX
22 substrate according to the molecular weight or molecular surface area. Since the hAOX-catalyzed
23 oxidation of azaheterocycles involves the nucleophilic attack initialized by the MoCo to the
24 molecule, the lower molecular electrophilicity might decrease the liability of a molecule toward
25 the hAOX metabolism. In other words, the substrate might have higher electrophilicity than
26 nonsubstrate, which might promote the nucleophilic attack by MoCo. In particular, the interaction
27 of the highest occupied molecular orbital (HOMO) of the nucleophile with the lowest unoccupied
28 molecular orbital (LUMO) of the electrophile is essential for bond formation.²⁴ For hAOX-
29 catalyzed reaction, the reactant with lower E_{LUMO} could be more electrophilic and thus more
30 susceptible to hAOX. Dalvie et al. have verified the correlation between E_{LUMO} and hAOX-
31 mediated susceptibility of zoniporide analogues.¹⁴ In addition to E_{LUMO} , some other properties like
32 electronegativity,²⁵ hardness²⁶ and ESP properties have also been used to explore the nucleophilic
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 attack reaction.²⁴ Accordingly, we calculated these electronic descriptors of reactivity or properties
4 that determined the electrophilicity of the molecules, including: E_{LUMO} , electronegativity, hardness
5 and the maximum and minimum ESP energies. For the electronic properties like E_{LUMO} ,
6 electronegativity and hardness, there were no significant difference between substrates and
7 nonsubstrates (Figure 3B, D). As for the ESP properties, several studies demonstrated that the
8 partial charges such as MK charge or ESP partial charge of the metabolic sites, namely the
9 electron-deficient carbon adjacent to the nitrogen, were more positive than nonSOMs.^{12, 16, 18}
10 Because the ESP partial charge of the atom was the integral of electrostatic potential surface on
11 the atom, and for the nitrogen-containing heterocycle rings, the more positive charge of the SOMs
12 might implicate the higher value of the ESP energy. Thus, we speculated that the substrate might
13 have the higher maximum values of the ESP energies (Max_EIPot) on the electrostatic potential
14 maps or the nonsubstrates might have the lower minimum values of the ESP energies (MIN_EIPot)
15 on the electrostatic potential maps. However, as shown in Figure 3C, D, for these properties there
16 was no significant difference between the substrates and nonsubstrates either.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

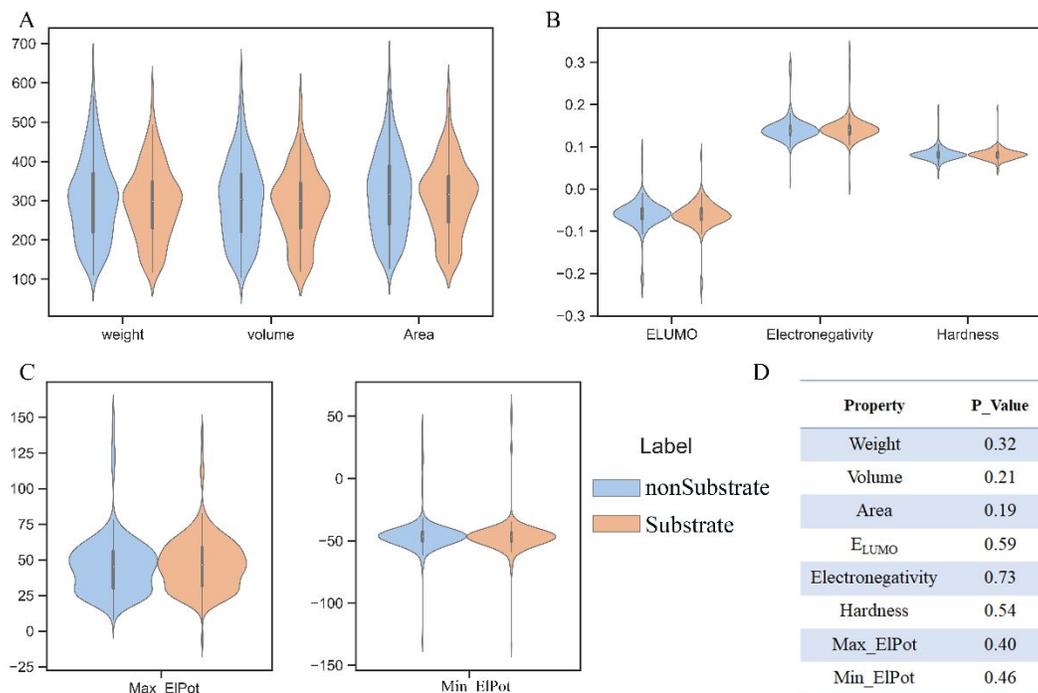


Figure 3. The distribution of topological and electronic descriptors between substrates and nonsubstrates. A: the distribution of molecular weight (unit: amu), molecular volume (unit: \AA^3) and molecular surface area (unit: \AA^2). B: the distribution of the energy of lowest unoccupied molecular orbital (E_{LUMO}) (unit: eV), the molecular electronegativity (unit: eV) and molecular hardness (unit: eV). C: the distribution of the maximum and minimum values of the ESP energies (kJ/mol) as mapped onto the electrostatic potential maps, denoted as Max_EIPot and Min_EIPot, respectively. Blue and orange colors represent the substrates and nonsubstrates, respectively. D: the p_value of corresponding descriptors in Figure 3A, 3B and 3C.

For these topological or electronic features, we found that there was no significant distribution between hAOX substrates and nonsubstrates, highlighting the following two points: 1. The hAOX possessed a broad range of substrate specificity and regioselectivity for oxidation of nitrogen-containing heterocycles. 2. For some electronic features like the E_{LUMO} , the correlation might only

1
2
3 exist in a specific scaffold, but it might diminish when given a large set of structurally diverse
4 molecules. Similar situations also appeared in the Dalvie et al.'s and Torres et al.' work.^{3, 10, 14}
5
6 Therefore, these results indicated that we need to further explore the correlation of other features
7
8 between published SOMs and nonSOMs.
9
10

11
12
13 Next, we compared the distribution of the local property related to the SOMs and nonSOMs.
14
15 Overall, the descriptors explored three types of effects, including: the reactive energy barrier
16 (ΔE_{inter}), hindrance effects (**Steric**, exposed area of reactive atoms), and electronic effects (ESP
17 charge and chemical shift of reactive atoms). Unlike the properties of molecular level, the
18 distribution differences of atomic descriptors were significant in most instances (Figure 4, Table
19 S2). First of all, the ΔE_{inter} representing the reaction energy barrier had the highest significance
20 with p_value $\sim 10^{-36}$ (Figure 4, Table S2). Chemical shift, ESP charge and exposed area properties
21 on the reactive carbon were more discriminating than those on the hydrogens attached to the
22 carbon. In contrast, the distribution of chemical shift and ESP charge of the electron-rich nitrogen
23 showed less significant difference, except for the exposed area (Figure 4, Table S2). The statistical
24 data of the distribution of these calculated descriptors are listed in Table S2.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

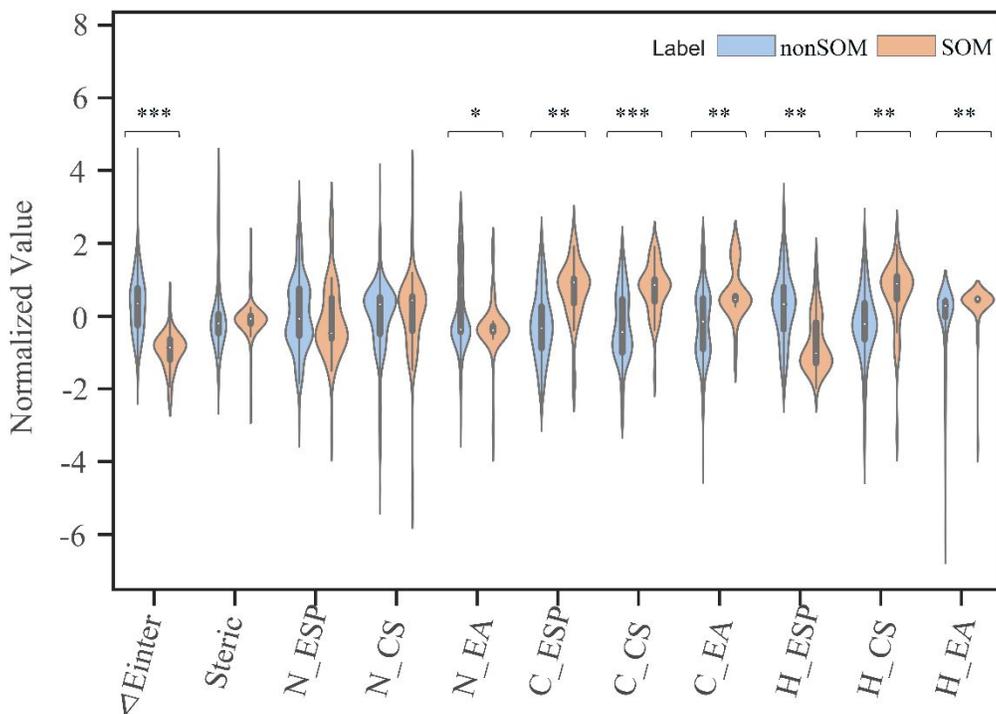


Figure 4. The distribution of some energy-related, hindrance effect related and electronic descriptors between SOMs and nonSOMs. The notation of these descriptors is listed following the order from left to right: ΔE_{inter} : the energy changes between prototype and the tetrahedral intermediate; Steric: steric hindrance; ESP: electrostatic potential charge; CS: chemical shift; EA: exposed area. The prefix N_, C_ or H_ represent the electron-rich nitrogen, electron-deficient carbon and hydrogen attached to the electron-deficient carbon atoms, respectively. P_value of the descriptors between SOMs and nonSOMs was calculated using T-test statistical hypothesis. The corresponding significance of each descriptor is depicted as black asterisk within the underlying range: (*) $10^{-5} < p < 0.05$; (**) $10^{-15} < p < 10^{-5}$; (***) $p < 10^{-15}$. Blue and orange colors represent the SOMs and nonSOMs, respectively. Note that all these descriptors were normalized using StandardScaler implemented in Scikit-learn v0.21.3.

1
2
3 The reaction energy barrier that determines whether a reaction occurs or not has been adopted
4 to discriminate the hAOX-mediated metabolism by many researches.^{6-7, 10} From the above
5
6 analysis, we might find that the relative energy between the substrate and the intermediate (ΔE_{inter})
7
8 indeed showed significant differences on the two classes. In addition, the chemical shift and ESP
9
10 charge of the electron-deficient carbon of the SOMs also showed different distributions.
11
12 Montefiori et al. have proposed that the ESP charge and chemical shift of the reactive carbon had
13
14 certain correlation with the energies of the transition states and were good discriminators to predict
15
16 the SOMs.^{12, 18} Here, we also explored the relationship of ESP charge and chemical shift of the
17
18 reactive carbon with the ΔE_{inter} but the results were not in concert with the Montefiori et al.'s
19
20 (Figure S2). Additionally, we also discovered that the properties of the hydrogen adjacent to the
21
22 aromatic carbon showed significant differences (Figure 4, Table S2). There have been reports
23
24 demonstrated that hydride displacement for a substrate might be a rate-limiting step.²⁷ Attention
25
26 has been drawn to investigate the thermodynamics of the hydride extraction step but the exposed
27
28 effect of this hydrogen has not yet been explored.¹⁷
29
30
31
32
33
34
35

36 Overall, from molecular level to atomic level, these statistical results highlighted that the hAOX
37
38 had quite wide substrate selectivity and diversity, and the common electrophilic descriptors like
39
40 E_{LUMO} , electronegativity and others showed less significant relationship with the tendency toward
41
42 hAOX-mediated metabolism. As for the discrimination between SOMs and nonSOMs, the local
43
44 features such as the reaction energy between substrate and tetrahedral intermediate, properties of
45
46 aromatic carbon and adjacent hydrogen were important. Based on these analyses, we next proposed
47
48 a new scheme to predict whether the molecule was a hAOX substrate and its regioselectivity of
49
50 metabolism.
51
52
53
54

55 Construction of the Substrate/Nonsubstrate Classification Model

56
57
58
59
60

As mentioned above, the decision-tree-based models needed to be refined to incorporate more knowledge of AOX substrates and nonsubstrates. Here, a substrate/nonsubstrate classification model was first constructed. The maximum, minimum and average values of local features of all potential SOMs within each molecule were calculated (described in Experimental Section). Then, all descriptors were used to construct the substrate/nonsubstrate classification model using training dataset of Cruciani et al.¹⁷ During the dataset training stage, for every hyperparameter combination, 5-folds cross-validation was performed and the F1-score was used as model performance evaluation metric. The final selected hyperparameters were: layer size: [400,200], learning rate: 0.001, batch size: 32, dropout rate: 0.3. Table 1 summarized the model performance of the deep neural network (DNN) classification model. For comparison, the methods recently proposed by Montefiori et al.¹⁸ were benchmarked here with the same test and external datasets. We may find that DNN model demonstrated more accurate prediction results on these tests.

Table 1. The performance of DNN substrate/nonsubstrate classification model.

Performance Score	Train	Test				External test			
	DNN	DNN	NMR Shielding*	ESP charge*	Chemical Shift*	DNN	NMR Shielding	ESP charge	Chemical Shift
SE	0.88	0.80	0.79	0.86	0.79	0.85	0.64	0.64	0.64
SP	0.95	1.00	0.38	0.56	0.38	0.94	0.58	0.58	0.63
Acc	0.92	0.93	0.57	0.70	0.57	0.90	0.60	0.60	0.63
F1	0.90	0.89	0.63	0.73	0.63	0.88	0.54	0.54	0.56

* Descriptors for predicting AOX-mediated SOM proposed by Montefiori et al.¹⁸

As stated, it was difficult for DT_{AOX} model to predict the metabolic tendency accurately within similar analogues, and to recognize the nonsubstrate accurately. Thus, we analyzed the classification ability of DNN model toward a series of similar molecules in test dataset. As shown

in Figure 5, the MOL486, MOL487 and MOL488 were mistaken as substrates by DT_{AOX} model, which were stable toward hAOX-mediated metabolism. In DT_{AOX} model, the position 2 in purine ring within MOL486 and in imidazole ring within MOL487 and MOL488 were predicted to be metabolic sites. In contrast, the DNN substrate/nonsubstrate classification model could accurately recognize the metabolic susceptibility of these molecules. (For the molecules in Figure 2, the DNN substrate/nonsubstrate classification model also yielded sound prediction results, as shown in Figure S3 of SI.)

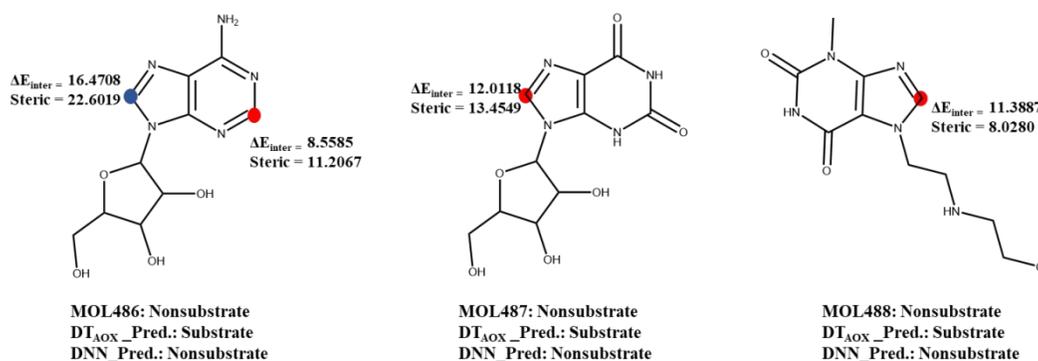


Figure 5. Examples of prediction ability of DNN substrate/nonsubstrate classification model on test datasets. The positive and negative sites predicted by DT_{AOX} are colored in red and blue circles, respectively. Pred.: prediction results.

We also investigated the model performance on external dataset containing different molecules with diverse skeletons to evaluate the general applicability and robustness of this model. Furthermore, the DNN classification model was verified toward the external dataset in which the molecules were collected from different resources with multiple experimental conditions. As summarized in Table 1, the classification performance on external dataset was slightly poor than test dataset, with the exception of the higher SE score. These results on both test dataset and external dataset revealed the generalization ability of the DNN classification model.

1
2
3 For several models introduced before, the focus was placed on determining which aromatic
4 carbon could be metabolized by hAOX. In these studies, thermokinetic, electronic and steric
5 hindrance effects were three main factors that were taken into consideration. The developed
6 models yielded good prediction results toward molecules with similar scaffolds, but showed less
7 generalization ability to molecules with different scaffolds.^{6, 14} One possible reason might be that
8 the metabolism of different scaffolds was influenced by different factors, thus the consensus and
9 limited descriptors were unable to differentiate the metabolic property of structurally diverse
10 molecules. Here, both global and local descriptors were combined to construct the
11 substrate/nonsubstrate classification model, which yielded much improved predictive ability to
12 structurally diverse molecules. Based on this model, we next explored where the SOMs were
13 located and then optimized our previously reported SOM prediction model DT_{AOX}.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 **Optimization of SOMs Classification Model for Predicting SOMs toward hAOX**

30
31
32 We have evaluated the DT_{AOX} toward recently published data with known SOMs/nonSOMs
33 information and the DT_{AOX} showed insufficient generalization ability to these compounds (Table
34 S1). After analyzing the predicted results of every branch of the decision tree, we observed that
35 the ΔE_{inter} could differentiate the true SOMs and false SOMs accurately, while the branch derived
36 from the **Steric** was unsatisfactory. This result was also reflected by the statistical analysis
37 described in previous section, where the ΔE_{inter} was the most discriminating descriptor among all
38 the local properties between SOMs and nonSOMs (Figure 4, Table S2), but Steric was less
39 significant. Here, the decision tree based method was applied to optimize the previous DT_{AOX}
40 model (methods described in the Experimental Section). In decision tree models, a node's impurity
41 is a function that measures the homogeneity of classes in data samples reaching the node. The
42 smaller the degree of impurity, the more skewed the class distribution. By minimizing the gini
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

impurity of every branch,²⁸ we observed that combining the local exposed area of the reactive atom H adjacent to the reactive center C (**H_EA**) with **Steric** could significantly improve the performance derived from the middle range of ΔE_{inter} . We therefore optimized the classification tree model, which will be referred to as DTN_{AOX} hereafter (Figure 6A).

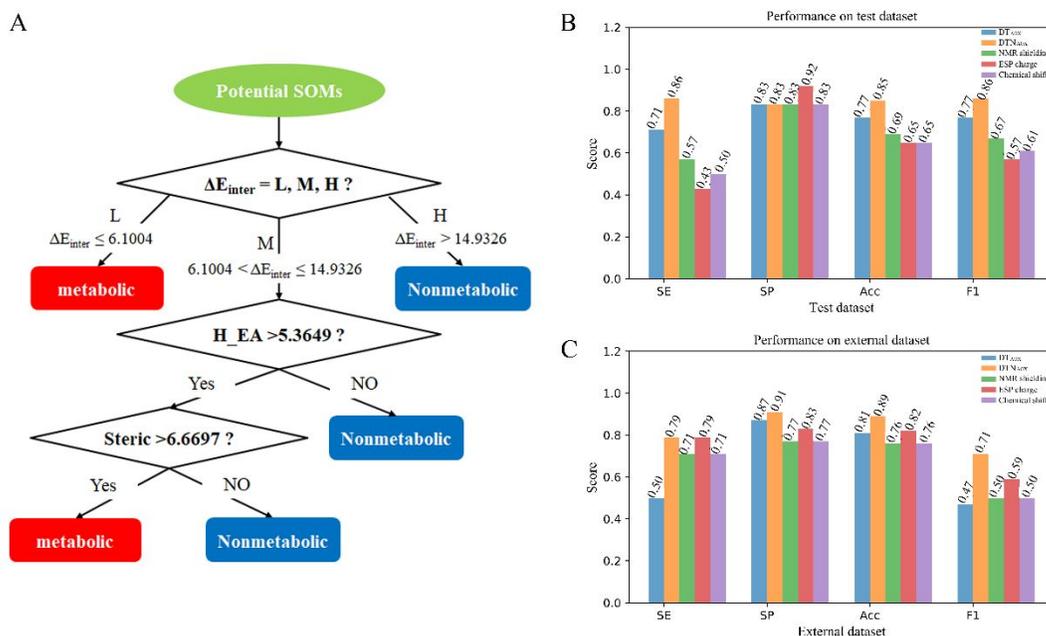


Figure 6. Decision tree model DTN_{AOX} (A) for predicting hAOX-mediated SOMs, and its performance on (B) the test dataset and (C) the external dataset, compared with our previous DT_{AOX} model⁶ and three descriptors proposed by Montefiori et al.¹⁸

Taking **H_EA** into consideration, we noticed that the sensitivity score was improved significantly toward the test dataset (Figure 6B). For further validation, the DTN_{AOX} was evaluated toward an external dataset extracted from several different literatures with diverse scaffolds and different experimental conditions. For this external dataset, the DTN_{AOX} model demonstrated decent performance (Figure 6C), and significantly exceeded DT_{AOX}, especially in terms of SE

1
2
3 score. Prediction results of the methods proposed by Montefiori et al.¹⁸ were also benchmarked
4
5 here for reference.
6
7

8
9 As shown, the SE was improved significantly with the descriptor **H_EA** introduced. Though
10 the significance of the properties of the aromatic C had been demonstrated before (Figure 4, Table
11 S2), after further analysis of these descriptors, we found that the ΔE_{inter} , **H_EA** and **Steric** yielded
12
13 the best performance toward test dataset. First of all, the thermokinetic effect was of high
14
15 the best performance toward test dataset. First of all, the thermokinetic effect was of high
16
17 importance for the metabolic reaction. According to the proposed reaction mechanism
18
19 aforementioned, the H atom attached to the potential metabolic site C of the compound needed to
20
21 be seized by MoCo, thus the exposed area of H also had a great influence on the occurrence of the
22
23 reaction. Actually, the descriptor **Steric** describes the global steric hindrance, which indicated
24
25 whether the MoCo could access to the reaction site, and the **H_EA** represents the local steric
26
27 influence when two reactive atoms access to each other for bond breakage and formation. With
28
29 reaction barrier, global and local hindrance effects considered, we might find that our optimized
30
31 model have not only good generalization ability, but also good interpretation ability for the
32
33 mechanism of AOX-mediated reaction.
34
35
36
37
38

39 **Fusing Model for Prediction of hAOX Substrate and Site of Metabolism**

40

41
42 In the actual drug-discovery-and-optimization program, the researchers concern whether the
43
44 molecule is metabolized then recognition and optimization of the SOMs come next. Thus, the
45
46 DNN substrate/nonsubstrate and DTN_{AOX} classification model were combined here to predict
47
48 whether the molecule was a hAOX substrate and the corresponding SOM or SOMs for practicable
49
50 purpose. Overall, fusing model could efficiently and accurately identify the molecular liability
51
52 toward hAOX-mediated biotransformation and predict the SOMs, providing more comprehensive
53
54
55
56
57
58
59
60

information of the hAOX-catalyzed metabolism. The proposed fusing model flowchart is shown in Figure 7.

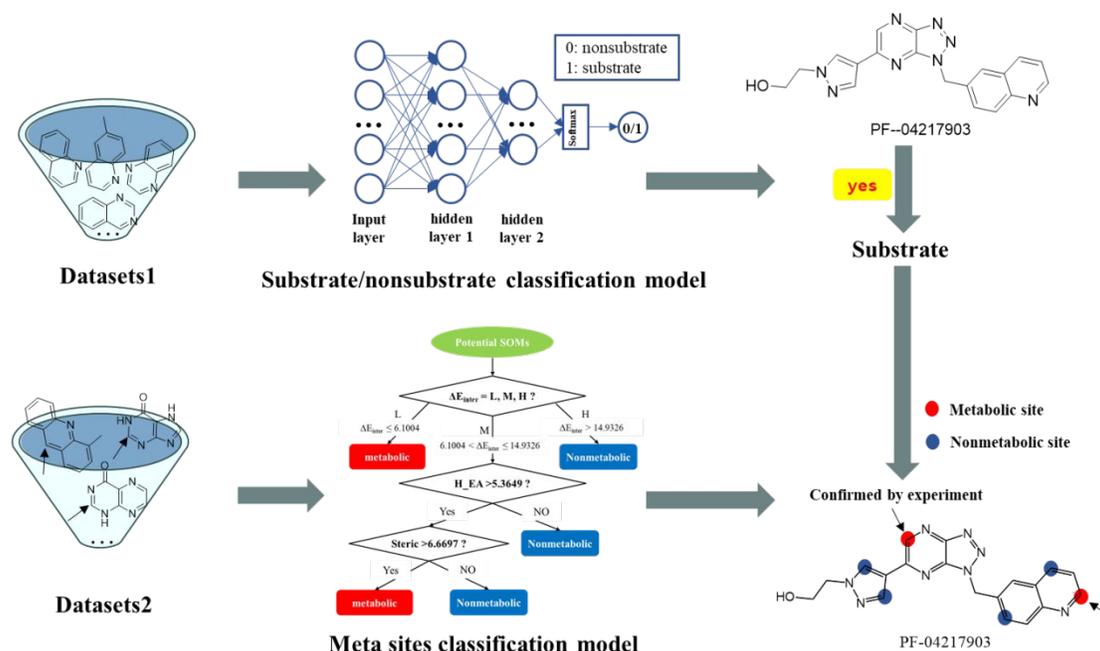


Figure 7. The flowchart of fusing model. The datasets1 and datasets2 contain the data with substrate/nonsubstrate and SOMs/nonSOMs labels, respectively. The red and blue circles indicate the predicted metabolic and nonmetabolic sites, respectively. The arrows indicate the metabolic sites identified in human liver cytosols.

Screen AOX Substrates in Epigenetics-related Inhibitors Using Fusing Model

The compounds regulating epigenetic modification are currently in the early stage of clinical development. Most of their inhibitors are similar to kinase inhibitors that contain nitrogen-containing aromatic heterocyclic fragments, which may be potential hAOX substrates. In order to further verify the practicability of our model, virtual screening of hAOX substrates was conducted for compounds related to epigenetic regulation that have not been reported on hAOX metabolism.

1
2
3 102 epigenetics-related inhibitors containing AOX-substrate-like motif were obtained from
4 Selleck (https://www.selleck.cn/pharmacological_epigenetics.html). All potential AOX metabolic
5 sites of these epigenetics-related inhibitors were labeled by matching predefined SMARTS rules,
6 then the fusing model was used to predict whether the molecule was a hAOX substrate and the
7 corresponding SOMs. To validate the in silico prediction results, 15 compounds were incubated
8 with human liver cytosol separately, and ultraperformance liquid chromatography
9 (UPLC)/quadrupole time-of-flight mass spectrometry (Q-TOF MS) was used to analyze the
10 incubation mixtures to assess their potential to turnover by hAOX (the 15 compounds' information
11 is provided in SI, Table S6). A compound metabolized would display an additional peak in the
12 total ion chromatogram upon mass spectrometric analysis of the incubation mixture, meaning there
13 would be addition of an oxygen atom and occurrence of oxidative metabolism, if the molecular
14 ion (MH^+) corresponding to the peak of the metabolite was 16 daltons greater than the molecular
15 ion of parent compound. The relationship of the observed oxidative metabolite with the parent
16 was confirmed from their respective mass spectral fragment ions. Furthermore, the contribution of
17 hAOX to the oxidative metabolism of a given compound incubated with human liver cytosol was
18 confirmed by chemical inhibition studies. The effects of various molybdenum hydroxylase
19 inhibitors on the formation of oxidative metabolite were examined. The chemical inhibitors
20 raloxifene and hydralazine were used to selectively inhibit cytosolic AOX, whereas allopurinol
21 was used to selectively inhibit cytosolic xanthine oxidase (XO).²⁹⁻³¹ If metabolism by hAOX was
22 involved, a substantial decrease of the liquid chromatography–mass spectrometry (LC-MS) peak
23 area would be observed for an oxidative metabolite after adding raloxifene or hydralazine to the
24 incubation system. In this way, the different inhibitory potential of AOX and XO selective
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

inhibitors to the formation of the oxidative metabolite would suggest which enzyme was mainly responsible for the oxidation metabolism.

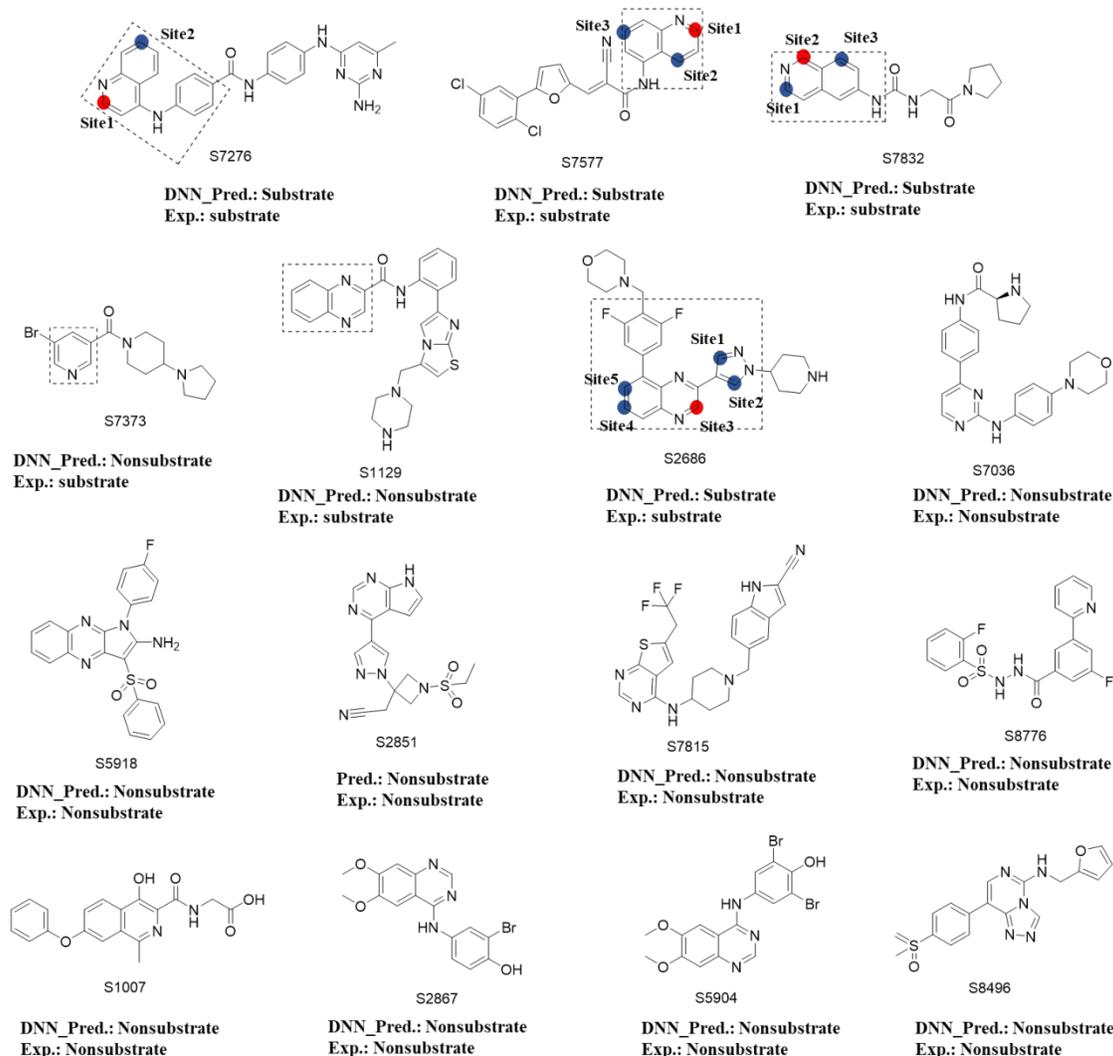


Figure 8. Predicted hAOX-catalyzed SOMs for 15 epigenetics-related inhibitors using the fusing model. Red and blue circles indicate the metabolic and nonmetabolic sites, respectively, predicted by DTN_{AOX}. The proposed metabolic region is marked using dotted box. Pred.: prediction results. Exp.: experimental results.

1
2
3 On the basis of the fusing model, compounds S7276, S7577, S7832 and S2686 were predicted
4 as substrates by DNN substrate/nonsubstrate classification model and the rest of 11 compounds
5 were predicted as nonsubstrates for hAOX (Figure 8). Then, the corresponding SOMs of 4
6 predicted substrates were predicted by using DTN_{AOX} SOMs classification model highlighted in
7 Figure 8 and the descriptor values are listed in Table S3. To validate our in vitro experimental
8 method, known hAOX substrates SGX523 and JNJ-38877605 and Amodiaquine were evaluated
9 with the abovementioned human liver cytosol assay.^{17, 32-33} All these positive controls displayed
10 significant oxidative metabolite formation and high inhibition ratio of the selective AOX inhibitor,
11 which verified the reliability of the assay (data and spectrum are provided in SI, Figure S4, S5 and
12 Table S7). For the 15 epigenetics-related inhibitors, only compounds S7276, S7577, S7832,
13 S7373, S1129 and S2686 had mono-oxidized metabolites detected by LC-MS (Ion extraction
14 chromatographs of parent and their corresponding mono-oxidized metabolites are shown in Figure
15 9). The metabolic region of these 6 compounds was identified based on the fragment ion
16 information from Q-TOF MS and marked using dotted box (Figure 10). No further attempt was
17 taken to confirm the oxidative metabolites structures. Then, to identify the molybdenum
18 hydroxylase (AOX or XO) mainly contributed to the oxidation metabolism of these 6 compounds,
19 chemical inhibitors were added to the incubation system. AOX selective inhibitors, raloxifene and
20 hydralazine inhibited over 93.9% of the oxidation metabolite formation of these compounds, while
21 XO selective inhibitor allopurinol exhibited limited inhibition (less than 12.0%) toward them (data
22 for inhibition study is listed in Table 2 and the LC-MS peak area data is provided in SI, Table S8
23 and S9). In summary, these data demonstrated that AOX not XO was the enzyme responsible for
24 the formation of the oxidative metabolite. And accordingly, S7276, S7577, S7832, S7373, S1129
25 and S2686 were confirmed to be hAOX substrates.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The molecules experimented included miscellaneous scaffolds like quinoline-, isoquinoline-,
4 quinazoline-, quinoxaline- and pyridine-substituted azaheterocycle ring compounds as well as
5 bicyclic skeletons. Unlike the molecules experimented by Lepri et al., electron-donating group
6 (EDG) and electron-withdrawing group (EWG) appeared more than once for most types of
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

performance on the experimental validation.

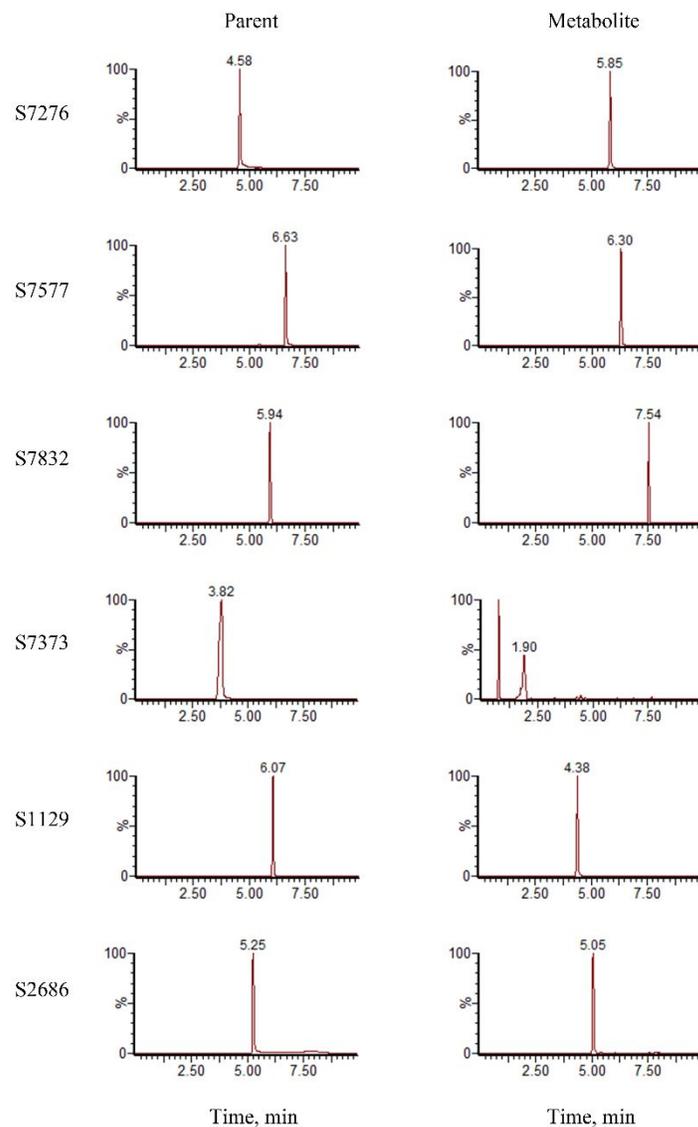


Figure 9. Ion extraction chromatographs of the parent compounds (left) and their corresponding mono-oxidized metabolites (right) formed in human liver cytosol.

For compounds containing quinoline- or isoquinoline- fragments such as S7276, S7577, S7832 and S1007, the model gave sound prediction results and the corresponding SOMs of substrates were in accordance with the proposed metabolic region. As we have discussed, for quinoline-containing substrates, the ΔE_{inter} alone could not determine the susceptibility of potential SOMs

1
2
3 toward oxidation and the **Steric** always gave higher false positive rate. Like S7276, the same
4 instances were also discovered in our experimented molecules that ΔE_{inter} alone was not decisive.
5
6 Like site1 in S7276, after combing **H_EA** with **Steric**, the DTN_{AOX} model could predict the true
7
8 SOM accurately while **Steric** alone gave wrong prediction. Lepri et al. have concluded that the
9
10 EDG at position 4, 6 and 7 of quinolines or isoquinolines might be substrate, and the EWG vice
11
12 versa.¹⁶ The same instance was also discovered in our experimental molecules S7276 and S7832
13
14 with EDG substituents, which were all susceptible to hAOX. Besides, regarding to the metabolism
15
16 of S7577, the EDG at position 8 also increased the susceptibility toward hAOX, which was not
17
18 discovered in existing studies.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

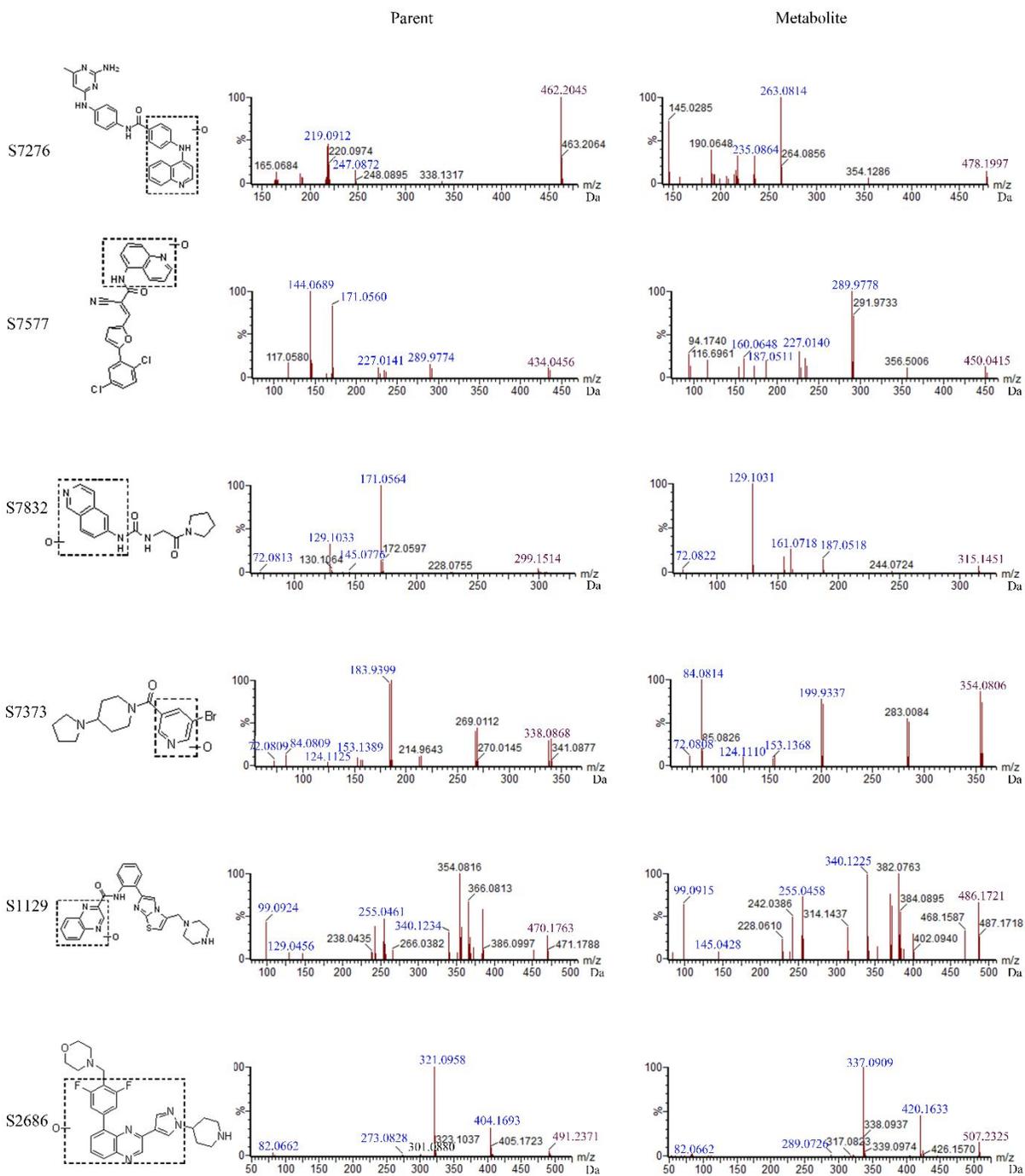


Figure 10. Positive product ion spectra of the parent compounds (left) and their corresponding mono-oxidized metabolites (right) formed in human liver cytosol. The m/z of parent ion was

1
2
3 colored in purple, and the m/z of relevant product ions which were applied to identify the
4
5 oxidized regions was colored in blue.
6
7

8
9 Regarding compounds containing quinazoline-, quinoxaline- and bicyclic-skeletons substituted,
10
11 the metabolism was more complicated. For quinazoline-substituted molecules like S2867 and
12
13 S5904, there were EDG groups at position 4 and two EWG groups at position 6 and 7, and this
14
15 kind of substitution was beyond the SMR scope summarized by Lepri et al.¹⁶ The same situation
16
17 was discovered in quinoxaline- and bicyclic-skeletons substituted compounds that were not mono-
18
19 substituted, namely with only one EDG- or EWG- substituted. Except for S1129, the fusing model
20
21 still gave concerted predictive results for these heterogeneously substituted scaffolds. Lepri et al.
22
23 proposed that pyridine-containing compounds are always stable to hAOX-mediated
24
25 biotransformation.¹⁶ While the in vitro experimental results demonstrated that some pyridine
26
27 substituted molecules are also susceptible toward hAOX. For example, S7373 with pyridine
28
29 skeleton was metabolized by hAOX while the S8776 was stable, and both molecules have EWG
30
31 substituted. For the last two kinds of skeletons, molecules were all correctly predicted except
32
33 S7373.
34
35
36
37
38

39
40 **Table 2.** Inhibition study of AOX/XO selective inhibitor on the formation of oxidative
41
42 metabolite in human liver cytosol.^a
43
44

Compound Name	Raloxifene	Hydralazine	Allopurinol
	(AOX inhibitor)	(AOX inhibitor)	(XO inhibitor)
	Inhibition Ratio (%)	Inhibition Ratio (%)	Inhibition Ratio (%)

S7276	98.3	99.7	7.33
S7577	93.9	97.3	4.97
S7832	100	100	2.45
S7373	100	100	7.50
S1129	99.7	99.7	9.25
S2686	99.6	99.4	12.0

$$^a \text{Inhibition Ratio} = 1 - \frac{\text{peak area of oxidative metabolite in human liver cytosol with inhibitor}}{\text{peak area of oxidative metabolite in human liver cytosol without inhibitor}}$$

Overall, these results highlighted the practicality of our model in recognizing the regioselectivity of molecules and metabolic SOM toward hAOX. With a clear SMARTS-pattern definition of the potential SOM, the fusing model can be used to prioritize compounds on the basis of their susceptibility toward hAOX.

CONCLUSION

As issues caused by hAOX-catalyzed metabolism of new drug molecules continue to emerge, it is critical that medicinal scientists have reliable and efficient methodologies to identify whether a molecule is a hAOX substrate, and where its metabolic site is. Unfortunately, the hAOX has quite wide substrate selectivity and diversity, which poses challenges for the predictive models. In this study, a combined model was developed based on structurally diverse molecules with homogeneously experimental data, and it yielded improved performance on predicting the metabolic propensity and explaining the metabolic regioselectivity of hAOX mediated reaction. Moreover, the improved computational model was used for screening structurally diverse drug

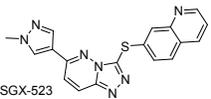
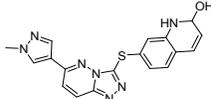
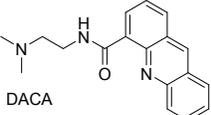
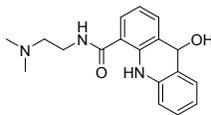
1
2
3 candidates containing potential hAOX metabolic sites, the results were then experimentally
4
5 validated with in vitro assays. Overall, this study provides practical scheme to optimize hAOX
6
7 metabolic profile of chemicals, which is of the essence for the hit-to-lead optimization in drug
8
9 development.
10
11

12 13 EXPERIMENTAL SECTION

14 15 16 17 Potential SOMs and Tetrahedral Intermediate Defining

18
19
20 According to the catalytic mechanism and previously reported hAOX metabolites, we optimized
21
22 the SMiles ARbitrary Target Specification (SMARTS) strings of the two types of potential SOMs
23
24 (Table 3).⁶ The code for extracting potential SOM defined by SMARTS was implemented using
25
26 Python 3.6 with RDKit 2018.09.1 version.
27
28
29

30
31 **Table 3** Two types of potential sites of metabolism by hAOX.
32

33 34 35 36 37 38	Typ	Descriptions	SMARTS	Representative substrates/sites	Representative intermediates
39 40 41 42 43	A	The carbon in the aromatic ring adjacent to the aromatic nitrogen with exactly one hydrogen	<chem>[\$([cR;H]:[nX2R])]</chem>	 SGX-523	
44 45 46 47 48 49	B	The carbon in the aromatic ring conjugated addition with γ -position nitrogen with exactly one hydrogen	<chem>[\$([#6D2R;H][*][*][#7X2R])]</chem>	 DACA	

50 51 52 53 Dataset Preparation and Construction

Existing metabolic datasets may show variance, in the quality of chemical/biological samples, measurement methods, experimental conditions, etc. To address this problem: (1) We only used experimentally determined metabolism data, and in silico data were not considered; (2) We have carefully chosen the data validated in human liver cytosol as our main dataset to reduce the inconsistency of experimental conditions; (3) We further excluded the data of the same molecule collected from different sources but with inconsistent metabolic profiles. These procedures help us increase the reliability of dataset on merging data from different sources. The hAOX-oxidated experimental data from the work published by Lepri et al. and Cruciani et al. were collected.¹⁶⁻¹⁷ The homogenous data was used for model construction and the number of molecules used to construct model are listed in Table 4. For substrate/nonsubstrate classification model, the training dataset was extracted from Cruciani et al.'s work and the test dataset was from Lepri et al.'s work, and the inconsistent data and molecule failed for descriptor calculation were excluded (For detailed molecules that were excluded were listed in SI). For SOM model, our previous training dataset was enriched with recently published data.^{7, 19} The same test dataset in ref 6 was used here for model comparison.

Furthermore, we assessed the predictive ability of the classification model against an external datasets collected from the literatures published recently.^{7, 19, 34-35} Noting that there were no overlapping molecules among the training dataset, test dataset and external dataset for substrate/nonsubstrate classification model construction, which was the same as the construction of SOMs/nonSOMs classification model. The detailed description of the dataset used for substrate/nonsubstrate model construction is provided in SI. And the structures with SOM information collected from newly published work is provided in Table S4. The structures of nonsubstrates collected from newly published work is provided in Table S5.

Table 4. Datasets used in the model construction and validation.

Dataset distribution		Substrate/Nonsubstrate classification model (molecules per dataset)	SOMs classification model (sites per dataset)
Train dataset	Positive	202	45
	Negative	299	128
Test dataset	Positive	10	14
	Negative	20	12
External dataset	Positive	13	14
	Negative	17	70

Descriptors Calculation

A total of 48 descriptors were calculated here. The energy changes between hAOX substrate and the tetrahedral intermediate (ΔE_{inter}) and steric hindrance (**Steric**) were calculated based on the methods described by Xu et al.⁶ The other descriptors were computed using Spartan'18 v. 1.2.0, which could be categorized as topological, physicochemical and electronic features. Table 5 provides a brief description of these descriptors.

Table 5. Condensed list of descriptors

Atom-level descriptors

ExposedArea	the ExposedArea of the reaction related atoms, describing the steric exposure of the reaction related atom
Bond order	The bond order of the C-H and C-N
Atom charge	The Electrostatic potential charge of the reaction related atoms
Chemical shift	The chemical shift of the reaction related atoms
Isotropic	The average three diagonal elements of the magnetic shielding tensor

Molecule-level descriptors

electronegativity	The electronegativity of the molecule, defined as $-(HOMO + LUMO)/2$
hardness	The hardness of the molecule, defined as $-(HOMO - LUMO)/2$
QSAR_ACC.Area	The accessible area of the molecular surface
QSAR_MIN_Elpot	The min local values of the electrostatic potential
Dipole	The dipole moment of the molecule
Solvation	The solvated energies of the molecule

Model Construction and Hyperparameter Optimization

Deep neural networks (DNN) have been successfully applied in addressing drug metabolism related prediction tasks,^{21, 36} in which the algorithm can be considered as a function that maps the input vector, which is the representation of a compound by its predefined features, to an output vector that represents the prediction task. For substrate/nonsubstrate classification, the deep neural network algorithm was used to predict whether the molecule was a hAOX substrate. The maximum, minimum and average value of descriptors of all potential SOMs within each molecule were calculated, which were fed into the model as molecular level descriptors. The descriptor mapping and model construction procedures are illustrated in Figure 11.

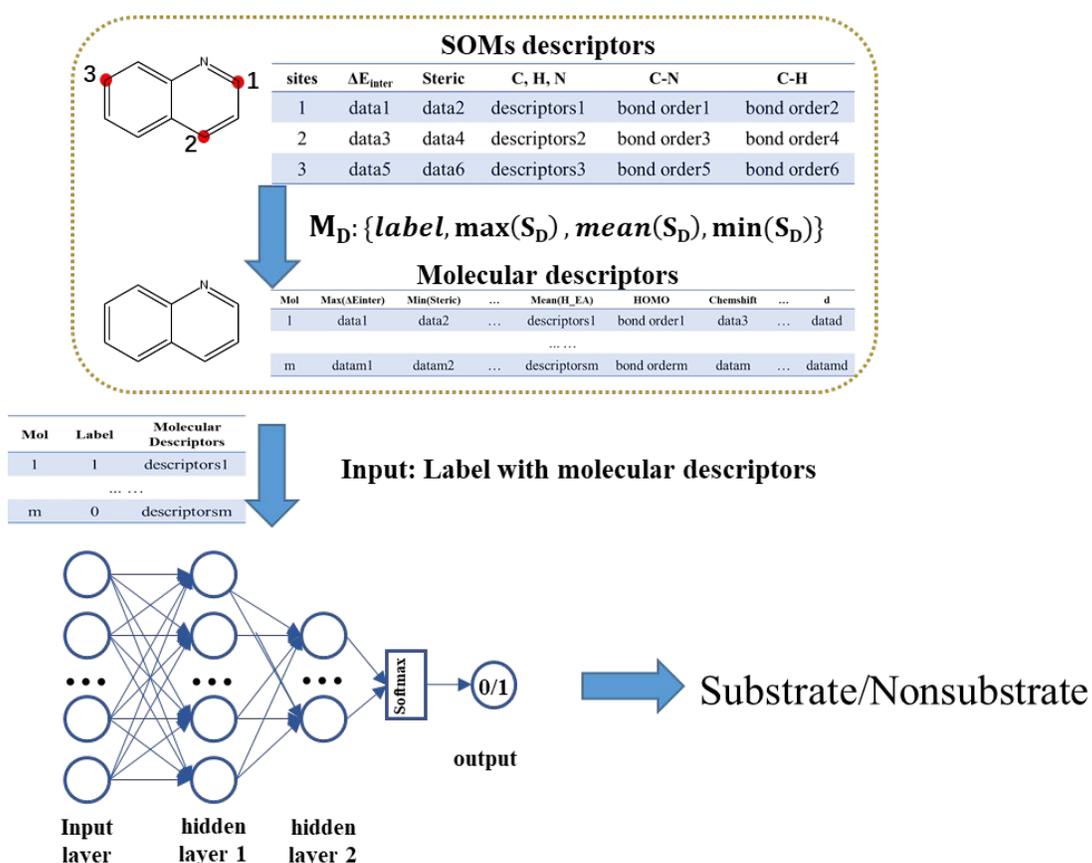


Figure 11. The descriptor mapping and model construction flowchart. The quinoline and DNN model structure were used to elucidate the flow chart.

To optimize the performance of DNN model, the hyperparameter optimization was carried out via a grid search algorithm.³⁷ The searched hyperparameter included network architecture (layer size), learning rate, batch size and dropout rate. The detailed hyperparameter setting is listed in Table 6.

Table 6. The hyperparameter setting

Hyperparameters	Setting
Layer size	([100,200,300,400,500,600,700,800,900,1000]) ([100,50], [200,100], [300,150], [400,200], [500,250], [600,300], [700,350], [800,400], [900,450], [1000,500]) ([400,200,100],[600,300,150],[800,400,200])
Learning rate	(0.00001, 0.00005, 0.0001, 0.0005, 0.001, 0.005, 0.01)
Batch size	(8, 16, 32, 64)
Dropout rate	(0.1, 0.2, 0.3, 0.4, 0.5)

For SOMs classification model, the decision-tree-based classification model was constructed with the Scikit-learn v0.21.3 using the Classification and Regression Trees (CART) algorithm.³⁸ CART is a nonparametric approach that can deal with both categorical and numeric dependent variables, and it constructs binary trees using the feature and threshold that yield the largest information gain at each node.³⁹ During the model training phase, 5-folds cross validation was implemented on the training dataset with F1-score as validation criterion, and the obtained models were evaluated by the following metrics:

$$\text{Sensitivity(SE)} = \frac{TP}{(TP + FN)}$$

$$\text{Specificity(SP)} = \frac{TN}{(TN + FP)}$$

$$\text{Accuracy(Acc)} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$F1 = \frac{2 * TP}{(2 * TP + FN + FP)}$$

TP: the number of true positive predicted by the model. TN: the number of true negative predicted by the model. FP: the number of false positive predicted by the model. FN: the number of false negative predicted by the model.

Human Liver Cytosol Incubations of Epigenetics related Inhibitors

Human liver cytosol for assays (mixed sex; pool of 50 donors; catalog no. H0610.C, lot no. 1610027; 30 males and 20 females) was purchased from Sekisui XenoTech. The incubations were carried out in 1.5 mL micro-centrifuge tubes. Briefly, the incubation mixture was composed of human liver cytosol (2 mg/mL final protein concentration), 2 mM MgCl₂, 10 μM test compound and 100 mM potassium phosphate buffer, pH 7.4. The total volume of each incubation was 200 μL and the final concentration of dimethyl sulfoxide (DMSO) used in the assay was 0.1% (v/v). Reactions were commenced with the addition of test compound after pre-warming at 37°C for 3 min in a low-speed shaking thermomixer and terminated by adding an equal volume of ice-cold acetonitrile after 3 h of incubation. Each sample was immediately mixed, and subsequently centrifuged at 12000 rpm for 10 min at room temperature. The supernatant was transferred into a 96-well plate for analysis by UPLC/Q-TOF MS. The control samples were prepared with no test compound or with inactivated enzymes. All samples for the experiment were performed in triplicate.

UPLC/Q-TOF MS Analysis

1
2
3 The UPLC/Q-TOF MS system consisted of a Waters Acquity UPLC system coupled to a Waters
4 Synapt G2 Q-TOF MS. Chromatography was performed on a Waters Acquity UPLC HSS T3
5 column (2.1*100 mm, 1.8 μm ; maintained at 45°C) using a mobile phase that consisted of solvent
6 A (2.5 mM aqueous ammonium acetate containing 0.1% formic acid) and solvent B (methanol or
7 acetonitrile/methanol, 9:1, v/v). The mobile phase was delivered at 0.4 mL/min in a gradient
8 manner, e.g., 0–1 min, at 10% solvent B; 1–6.5 min, from 10% to 55% solvent B; 6.5–7.5 min,
9 from 55% to 95% solvent B; 7.5–9.0 min, at 95% solvent B; and 9.0–10.0 min, at 10% solvent B.
10 A 7 μL aliquot of the sample in the 96-well plate was injected into the system. The Synapt G2 Q-
11 TOF MS was operated in resolution mode with electrospray ionization. The capillary and cone
12 voltages were set at 3 kV and 30 V. The source and desolvation temperatures were set at 120 and
13 400 °C, respectively. Nitrogen was applied as the cone gas (50 L/h) and desolvation gas (800 L/h).
14 All data were acquired in positive ionization mode. Mass was externally calibrated with a 0.5 mM
15 sodium formate solution at 25 $\mu\text{L}/\text{min}$ over a range of m/z 50-1000 and corrected during
16 acquisition using an external reference (LockSpray™) consisting of a 40 ng/mL solution of leucine
17 enkephalin infused at a flow rate of 10 $\mu\text{L}/\text{min}$ via a LockSpray interface, generating a reference
18 ion for positive ion mode ($m/z = 556.2771$) to ensure accuracy during the analysis. Metabolites
19 were characterized by MS full scan and product ion scan that both ranged from m/z 50 to 1000.
20 Product ion scan was obtained either from the MS^E approach or the dedicated MS/MS experiment.
21 For MS^E approach, accurate mass data were acquired by two separate scan functions programmed
22 with independent collision energies (e.g., trap collision energy in function 1 was 2 V and an energy
23 ramp of 15-25 V was used in function 2). With MS/MS experiment, for example, the trap collision
24 energy was ramped from 10 to 16 V, whereas the transfer collision energy was ramped from 14 to
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 22 V. All the instrument control and data acquisition were achieved using Waters MassLynx
4 software (version 4.1).
5
6

7 8 **Molybdenum Hydroxylase Inhibition Study** 9

10
11 Aldehyde oxidase selective inhibitors raloxifene, hydralazine and xanthine oxidase selective
12 inhibitor allopurinol were used in the study to further investigate the involvement of human
13 aldehyde oxidase in the oxidative metabolism of the test compounds. The incubation sample
14 contained human liver cytosol (2 mg/mL final protein concentration), 2 mM MgCl₂, 10 μM test
15 compound, 100 μM inhibitor and 100 mM potassium phosphate buffer, pH 7.4. The test
16 compounds and inhibitors were dissolved in DMSO respectively and the final concentration of
17 DMSO in each incubation (200 μL total volume) was 1.1% (v/v). The mixtures were pre-incubated
18 for 3 min at 37°C, and each reaction was initiated by adding test compound. After incubation at
19 37°C for 3 h, reactions were terminated by adding 200 μL ice-cold acetonitrile. The mixtures were
20 centrifuged and 7 μL of the supernatant was injected into the UPLC/Q-TOF MS. Controls without
21 inhibitors were also prepared. Experiments were performed in triplicate.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 **ASSOCIATED CONTENT** 39

40 41 **Supporting Information.** 42 43 44 45

46 Additional figures and tables illustrating the sample distribution according to the
47 applicability domain defined by DT_{AOX} (Figure S1); The relationship of ΔE_{inter} versus the
48 ESP charge (A) and the chemical shift (B) of the reactive carbon (Figure S2); The
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 prediction results of molecules in Figure 2 by DNN substrate/nonsubstrate classification
4
5
6
7 model (Figure S3); Ion extraction chromatographs of the positive control (Figure S4);
8
9
10 Positive product ion spectra of the positive control (Figure S5); The predictive
11
12
13 performance of DT_{AOX} on newly published data (Table S1); Distributions of descriptors
14
15
16 between SOMs and nonSOMs (Table S2); Descriptor values of 4 substrates predicted by
17
18
19 fusing model (Table S3); The substrates with SOM information collected from newly
20
21
22 published work (Table S4); The nonsubstrates' structure collected from newly published
23
24
25 work (Table S5); The structures used for experimental validation (Table S6); Inhibition
26
27
28 study of oxidative metabolites for positive control incubated in human liver cytosol with or
29
30
31 without human AOX chemical inhibitors (Table S7); LC-MS peak areas of oxidative
32
33
34 metabolites for compounds incubated in human liver cytosol with or without human AOX
35
36
37 chemical inhibitors (Table S8); LC-MS peak areas of oxidative metabolites for compounds
38
39
40 incubated in human liver cytosol with or without human XO chemical inhibitor (Table S9)
41
42
43
44
45
46
47
48
49 (PDF) with Molecular formula strings (csv) used in model construction.
50
51
52

53 AUTHOR INFORMATION

54
55
56
57
58
59
60

Corresponding Author

*Mingyue Zheng, Phone: 86-21-50806600-1308. E-mail: myzheng@simm.ac.cn.

Hualiang Jiang, Phone: 86-21-50806600-1303. E-mail: hljiang@simm.ac.cn.

Author Contributions

‡J.Z. and R.C. contributed equally to this work.

Notes

The authors declare no competing financial interest.

The DNN model is available at github (<https://github.com/AOXpredict/>).

ACKNOWLEDGMENT

We gratefully acknowledge financial support from the National Natural Science Foundation of China (81773634 to M.Z), National Science & Technology Major Project “Key New Drug Creation and Manufacturing Program”, China (Number: 2018ZX09711002 to H.J.), and “Personalized Medicines—Molecular Signature-based Drug Discovery and Development”, Strategic Priority Research Program of the Chinese Academy of Sciences (XDA12050201 to M.Z.).

ABBREVIATIONS USED

1
2
3 AOX, Aldehyde oxidase; hAOX, human AOX; SOM, site of metabolism; FAD, flavin
4
5
6
7 adenine dinucleotide; MoCo, molybdenum cofactor; SMR, structure-metabolism
8
9
10 relationship; ESP, electrostatic potential; LUMO, lowest unoccupied molecular orbital.
11
12
13
14 E_{LUMO} , energy of LUMO; AD, applicability domain; SE, sensitivity score; SP, specificity
15
16
17 score; Acc, accuracy; HOMO, highest occupied molecular orbital; DNN, deep neural network;
18
19
20 DT, decision tree; UPLC, ultraperformance liquid chromatography; Q-TOF MS, quadrupole time-
21
22 of-flight mass spectrometry; LC-MS, liquid chromatography–mass spectrometry; XO, xanthine
23
24 oxidase; EDG, electron-donating group; EWG, electron-withdrawing group; SMARTS, SMiles
25
26
27 ARbitrary Target Specification; DMSO, dimethyl sulfoxide.
28
29
30

31 REFERENCE

- 32
33
34
35 1. Pryde, D. C.; Tran, T. D.; Jones, P.; Duckworth, J.; Howard, M.; Gardner, I.;
36 Hyland, R.; Webster, R.; Wenham, T.; Bagal, S.; Omoto, K.; Schneider, R. P.; Lin, J.,
37 Medicinal Chemistry Approaches to Avoid Aldehyde Oxidase Metabolism. *Bioorganic &*
38 *Medicinal Chemistry Letters* **2012**, *22* (8), 2856-2860.
39
40
41 2. Sodhi, J. K.; Wong, S.; Kirkpatrick, D. S.; Liu, L.; Khojasteh, S. C.; Hop, C. E.;
42 Barr, J. T.; Jones, J. P.; Halladay, J. S., A Novel Reaction Mediated by Human
43 Aldehyde Oxidase: Amide Hydrolysis of GDC-0834. *Drug Metabolism & Disposition*
44 **2015**, *43* (6), 908-915.
45
46
47 3. Manevski, N.; King, L.; Pitt, W. R.; Lecomte, F.; Toselli, F., Metabolism by
48 Aldehyde Oxidase: Drug Design and Complementary Approaches to Challenges in
49 Drug Discovery. *Journal of Medicinal Chemistry* **2019**, *62* (24), 10955-10994.
50
51
52
53
54
55
56
57
58
59
60

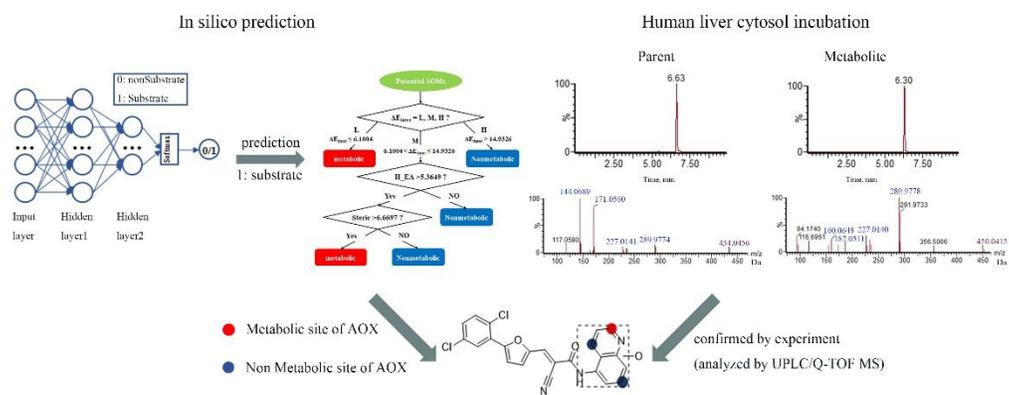
- 1
2
3
4 4. Paragas, E. M.; Humphreys, S. C.; Min, J.; Joswig-Jones, C. A.; Jones, J. P.,
5 The Two Faces of Aldehyde Oxidase: Oxidative and Reductive Transformations of 5-
6 Nitroquinoline. *Biochemical Pharmacology* **2017**, *145*, 210-217.
- 7
8
9 5. Cerny, M. A., Prevalence of Non-Cytochrome P450-Mediated Metabolism in
10 Food and Drug Administration-Approved Oral and Intravenous Drugs: 2006-2015. *Drug*
11 *Metabolism & Disposition* **2016**, *44* (8), 1246-1252.
- 12
13
14 6. Xu, Y.; Li, L.; Wang, Y.; Xing, J.; Zhou, L.; Zhong, D.; Luo, X.; Jiang, H.; Chen,
15 K.; Zheng, M.; Deng, P.; Chen, X., Aldehyde Oxidase Mediated Metabolism in Drug-like
16 Molecules: A Combined Computational and Experimental Study. *Journal of Medicinal*
17 *Chemistry* **2017**, *60* (7), 2973-2982.
- 18
19
20
21 7. Zhang, J. W.; Xiao, W.; Gao, Z. T.; Yu, Z. T.; Zhang, J. Y. J., Metabolism of c-
22 Met Kinase Inhibitors Containing Quinoline by Aldehyde Oxidase, Electron Donating,
23 and Steric Hindrance Effect. *Drug Metabolism & Disposition* **2018**, *46* (12), 1847-1855.
- 24
25
26 8. Coelho, C.; Mahro, M.; Trincao, J.; Carvalho, A. T.; Ramos, M. J.; Terao, M.;
27 Garattini, E.; Leimkuhler, S.; Romao, M. J., The First Mammalian Aldehyde Oxidase
28 Crystal Structure: Insights into Substrate Specificity. *The Journal of Biological Chemistry*
29 **2012**, *287* (48), 40690-40702.
- 30
31
32
33 9. Coelho, C.; Foti, A.; Hartmann, T.; Santos-Silva, T.; Leimkuhler, S.; Romao, M.
34 J., Structural Insights Into Xenobiotic and Inhibitor Binding to Human Aldehyde Oxidase.
35 *Nature chemical biology* **2015**, *11* (10), 779-783.
- 36
37
38
39 10. Torres, R. A.; Korzekwa, K. R.; McMasters, D. R.; Fandozzi, C. M.; Jones, J. P.,
40 Use of Density Functional Calculations to Predict the Regioselectivity of Drugs and
41 Molecules Metabolized by Aldehyde Oxidase. *Journal of Medicinal Chemistry* **2007**, *50*
42 (19), 4642-4647.
- 43
44
45 11. Alfaro, J. F.; Jones, J. P., Studies On the Mechanism of Aldehyde Oxidase
46 and Xanthine Oxidase. *The Journal of Organic Chemistry* **2008**, *73* (23), 9469-9472.
- 47
48
49 12. Montefiori, M.; Jorgensen, F. S.; Olsen, L., Aldehyde Oxidase: Reaction
50 Mechanism and Prediction of Site of Metabolism. *ACS Omega* **2017**, *2* (8), 4237-4244.
- 51
52
53 13. Pryde, D. C.; Dalvie, D.; Hu, Q.; Jones, P.; Obach, R. S.; Tran, T. D., Aldehyde
54 Oxidase: An Enzyme of Emerging Importance in Drug Discovery. *Journal of Medicinal*
55 *Chemistry* **2010**, *53* (24), 8441-8460.
- 56
57
58
59
60

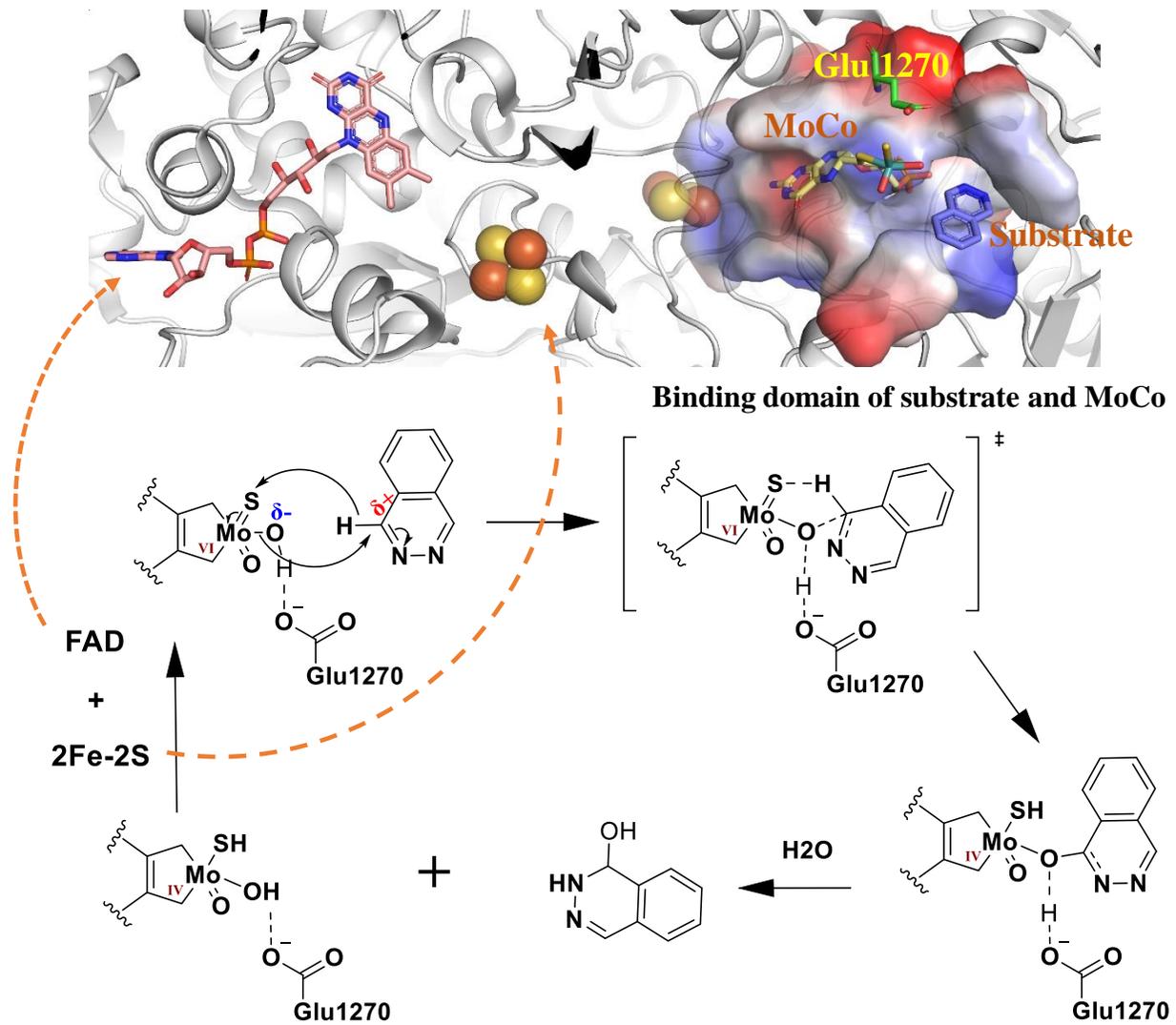
- 1
2
3
4 14. Dalvie, D.; Sun, H.; Xiang, C.; Hu, Q.; Jiang, Y.; Kang, P., Effect of Structural
5 Variation on Aldehyde Oxidase-Catalyzed Oxidation of Zoniporide. *Drug Metabolism &*
6 *Disposition* **2012**, *40* (8), 1575-1587.
7
8
9 15. Jones, J. P.; Korzekwa, K. R., Predicting Intrinsic Clearance for Drugs and Drug
10 Candidates Metabolized by Aldehyde Oxidase. *Molecular Pharmaceutics* **2013**, *10* (4),
11 1262-1268.
12
13
14 16. Lepri, S.; Ceccarelli, M.; Milani, N.; Tortorella, S.; Cucco, A.; Valeri, A.; Goracci,
15 L.; Brink, A.; Cruciani, G., Structure-Metabolism Relationships in Human-AOX:
16 Chemical Insights From a Large Database of Aza-Aromatic and Amide Compounds.
17 *Proceedings of the National Academy of Sciences of the United States of America*
18 **2017**, *114* (16), E3178-E3187.
19
20
21 17. Cruciani, G.; Milani, N.; Benedetti, P.; Lepri, S.; Cesarini, L.; Baroni, M.; Spyraakis,
22 F.; Tortorella, S.; Mosconi, E.; Goracci, L., From Experiments to a Fast Easy-to-Use
23 Computational Methodology to Predict Human Aldehyde Oxidase Selectivity and
24 Metabolic Reactions. *Journal of Medicinal Chemistry* **2018**, *61* (1), 360-371.
25
26
27 18. Montefiori, M.; Lyngholm-Kjaerby, C.; Long, A.; Olsen, L.; Jorgensen, F. S., Fast
28 Methods for Prediction of Aldehyde Oxidase-Mediated Site-of-Metabolism.
29 *Computational and Structural Biotechnology Journal* **2019**, *17*, 345-351.
30
31
32 19. Dick, R. A., Refinement of In Vitro Methods for Identification of Aldehyde Oxidase
33 Substrates Reveals Metabolites of Kinase Inhibitors. *Drug Metabolism & Disposition*
34 **2018**, *46* (6), 846-859.
35
36
37 20. Peng, J.; Lu, J.; Shen, Q.; Zheng, M.; Luo, X.; Zhu, W.; Jiang, H.; Chen, K., In
38 Silico Site of Metabolism Prediction for Human UGT-Catalyzed Reactions.
39 *Bioinformatics* **2014**, *30* (3), 398-405.
40
41
42 21. Li, X.; Xu, Y.; Lai, L.; Pei, J., Prediction of Human Cytochrome P450 Inhibition
43 Using a Multitask Deep Autoencoder Neural Network. *Molecular Pharmaceutics* **2018**,
44 4336-4345.
45
46
47 22. Smith, D. A., Discovery and ADMET: Where Are We Now. *Current Topics in*
48 *Medicinal Chemistry* **2011**, *11* (4), 467-481.
49
50
51 23. Shoombuatong, W.; Prathipati, P.; Prachayasittikul, V.; Schaduangrat, N.; Malik,
52 A. A.; Pratiwi, R.; Wanwimolruk, S.; Wikberg, J. E. S.; Gleeson, M. P.; Spjuth, O.;

- 1
2
3
4 Nantasenamat, C., Towards Predicting the Cytochrome P450 Modulation: From QSAR
5 to Proteochemometric Modeling. *Current Drug Metabolism* **2017**, *18* (6), 540-555.
- 6
7 24. Schwobel, J. A.; Koleva, Y. K.; Enoch, S. J.; Bajot, F.; Hewitt, M.; Madden, J. C.;
8 Roberts, D. W.; Schultz, T. W.; Cronin, M. T., Measurement and Estimation of
9 Electrophilic Reactivity for Predictive Toxicology. *Chemical Reviews* **2011**, *111* (4),
10 2562-2596.
11
12
13
14 25. Schultz, T. W.; Yarbrough, J. W.; Johnson, E. L., Structure-Activity Relationships
15 for Reactivity of Carbonyl-Containing Compounds with Glutathione. *SAR and QSAR in*
16 *Environmental Research* **2005**, *16* (4), 313-322.
17
18
19 26. Roy, R. K.; Krishnamurti, S.; Geerlings, P.; Pal, S., Local Softness and Hardness
20 Based Reactivity Descriptors for Predicting Intra- and Intermolecular Reactivity
21 Sequences: Carbonyl Compounds. *Journal of Physical Chemistry A* **1998**, *102* (21),
22 3746-3755.
23
24
25
26 27. Metz, S.; Thiel, W., A Combined QM/MM Study on the Reductive Half-Reaction
27 of Xanthine Oxidase: Substrate Orientation and Mechanism. *Journal of the American*
28 *Chemical Society* **2009**, *131* (41), 14885-14902.
29
30
31 28. Raileanu, L. E.; Stoffel, K., Theoretical Comparison Between the Gini Index and
32 Information Gain Criteria. *Annals of Mathematics and Artificial Intelligence* **2004**, *41* (1),
33 77-93.
34
35
36
37 29. Obach, R. S., Potent Inhibition of Human Liver Aldehyde Oxidase by Raloxifene.
38 *Drug Metabolism & Disposition* **2004**, *32* (1), 89-97.
39
40
41 30. Strelevitz, T. J.; Orozco, C. C.; Obach, R. S., Hydralazine as a Selective Probe
42 Inactivator of Aldehyde Oxidase in Human Hepatocytes: Estimation of the Contribution
43 of Aldehyde Oxidase to Metabolic Clearance. *Drug Metabolism & Disposition* **2012**, *40*
44 (7), 1441-1448.
45
46
47 31. Massey, V.; Komai, H.; Palmer, G.; Elion, G. B., On the Mechanism of
48 Inactivation of Xanthine Oxidase by Allopurinol and Other Pyrazolo[3,4-d]Pyrimidines.
49 *The Journal of Biological Chemistry* **1970**, *245* (11), 2837-2844.
50
51
52 32. Diamond, S.; Boer, J.; Maduskuie, T. P., Jr.; Falahatpisheh, N.; Li, Y.;
53 Yeleswaram, S., Species-Specific Metabolism of SGX523 by Aldehyde Oxidase and the
54 Toxicological Implications. *Drug Metabolism & Disposition* **2010**, *38* (8), 1277-1285.
55
56
57
58
59
60

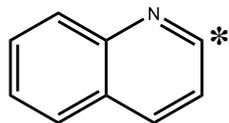
- 1
2
3
4 33. Lolkema, M. P.; Bohets, H. H.; Arkenau, H. T.; Lampo, A.; Barale, E.; de Jonge,
5 M. J. A.; van Doorn, L.; Hellemans, P.; de Bono, J. S.; Eskens, F., The c-Met Tyrosine
6 Kinase Inhibitor JNJ-38877605 Causes Renal Toxicity Through Species-Specific
7 Insoluble Metabolite Formation. *Clinical Cancer Research* **2015**, *21* (10), 2297-2304.
8
9
10 34. Jia, H.; Dai, G.; Weng, J.; Zhang, Z.; Wang, Q.; Zhou, F.; Jiao, L.; Cui, Y.; Ren,
11 Y.; Fan, S.; Zhou, J.; Qing, W.; Gu, Y.; Wang, J.; Sai, Y.; Su, W., Discovery of (S)-1-(1-
12 (Imidazo[1,2-a]pyridin-6-yl)ethyl)-6-(1-methyl-1H-pyrazol-4-yl)-1H-[1,2, 3]triazolo[4,5-
13 b]pyrazine (Volitinib) as a Highly Potent and Selective Mesenchymal-Epithelial
14 Transition Factor (c-Met) Inhibitor in Clinical Development for Treatment of Cancer.
15 *Journal of Medicinal Chemistry* **2014**, *57* (18), 7577-7589.
16
17
18 35. Mao, Z.; Wu, Y.; Li, Q.; Wang, X.; Liu, Y.; Di, X., Aldehyde Oxidase-Dependent
19 Species Difference in Hepatic Metabolism of Fasudil to Hydroxyfasudil. *Xenobiotica*
20 **2018**, *48* (2), 170-177.
21
22
23 36. Hughes, T. B.; Dang, N. L.; Miller, G. P.; Swamidass, S. J., Modeling Reactivity
24 to Biological Macromolecules With a Deep Multitask Network. *ACS Central Science*
25 **2016**, *2* (8), 529-537.
26
27
28 37. Hinton, G. E., In Neural Networks: Tricks of the Trade. *Springer* **2012**, 599-619.
29
30
31 38. Speybroeck, N., Classification and Regression Trees. *International Journal of*
32 *Public Health* **2012**, *57* (1), 243-246.
33
34
35 39. Henrard, S.; Speybroeck, N.; Hermans, C., Classification and Regression Tree
36 Analysis vs. Multivariable Linear and Logistic Regression Methods as Statistical Tools
37 for Studying Haemophilia. *Haemophilia* **2015**, *21* (6), 715-722.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table of Contents Graphic



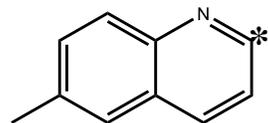


$\Delta E_{\text{inter}} = 11.1669$
Steric = 7.6691



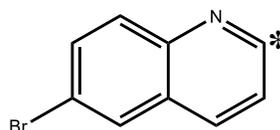
MOL038: substrate
DT_{AOX}_Pred.: substrate

$\Delta E_{\text{inter}} = 12.1557$
Steric = 7.6977



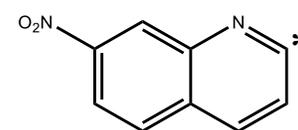
MOL044: nonsubstrate
DT_{AOX}_Pred.: substrate

$\Delta E_{\text{inter}} = 11.0852$
Steric = 7.6715

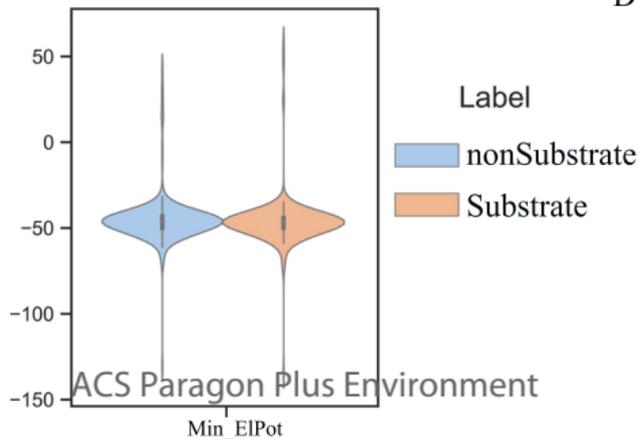
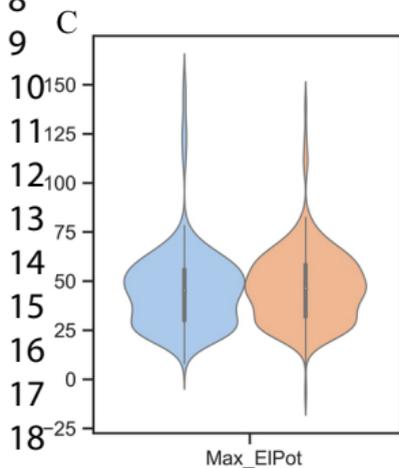
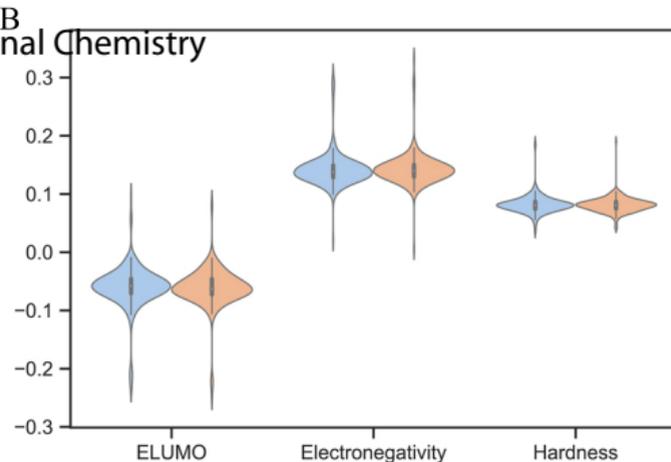
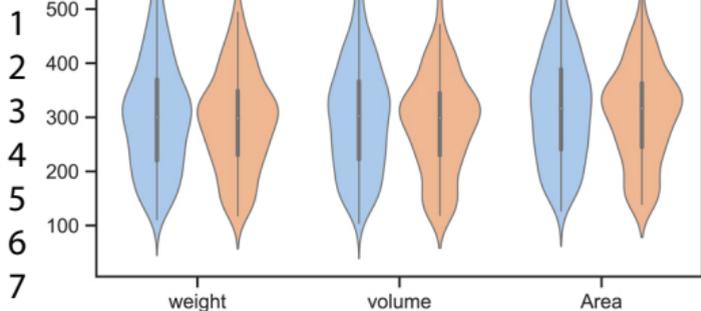


MOL116: nonsubstrate
DT_{AOX}_Pred.: substrate

$\Delta E_{\text{inter}} = 6.9659$
Steric = 7.6911

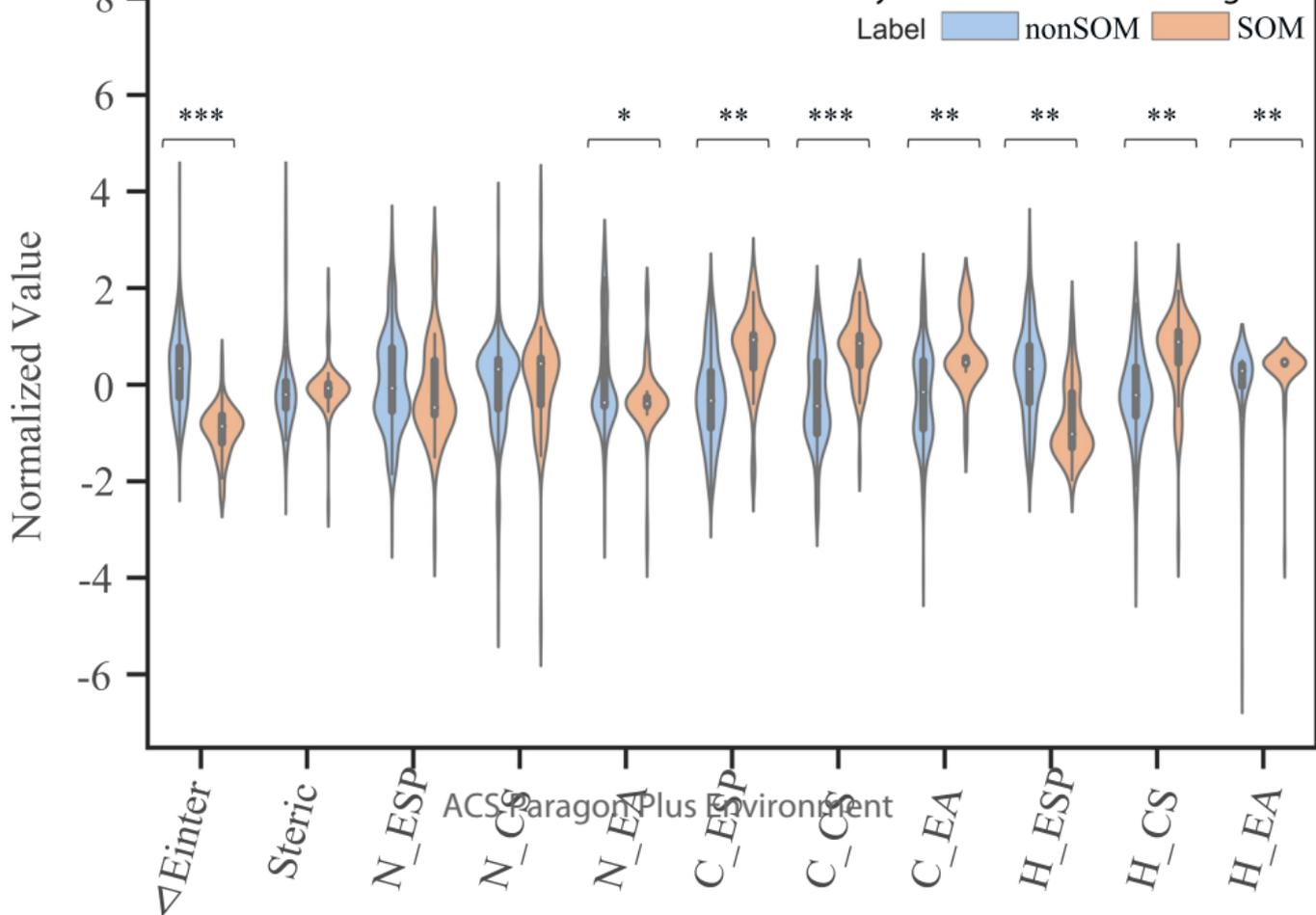


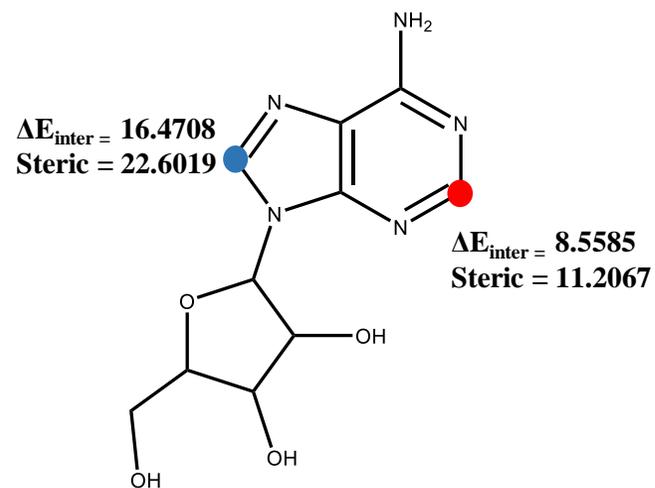
MOL076: nonsubstrate
DT_{AOX}_Pred.: substrate



D

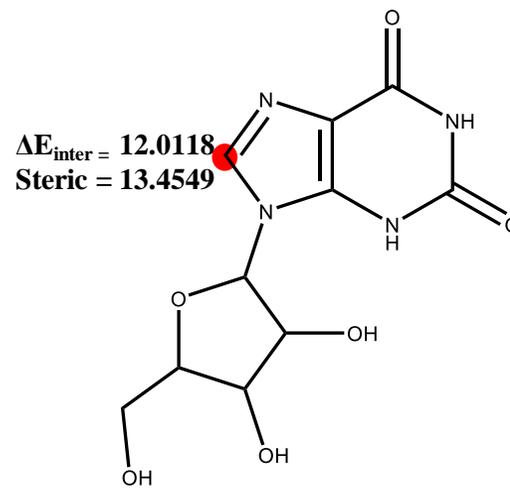
Property	P_Value
Weight	0.32
Volume	0.21
Area	0.19
E_{LUMO}	0.59
Electronegativity	0.73
Hardness	0.54
Max_EIPot	0.40
Min_EIPot	0.46

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20



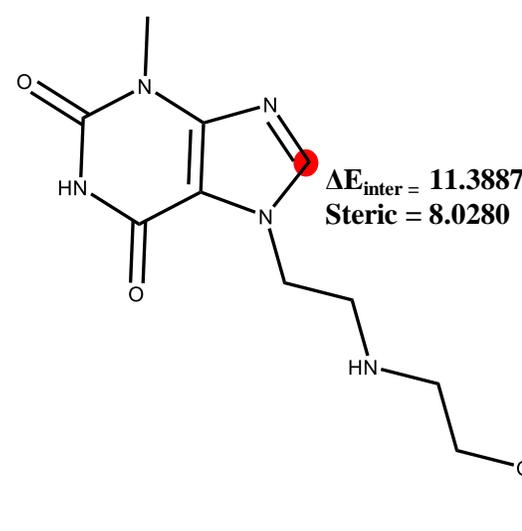
22
23
24
25
26
27

MOL486: Nonsubstrate
DT_{AOX}_Pred.: Substrate
DNN_Pred.: Nonsubstrate



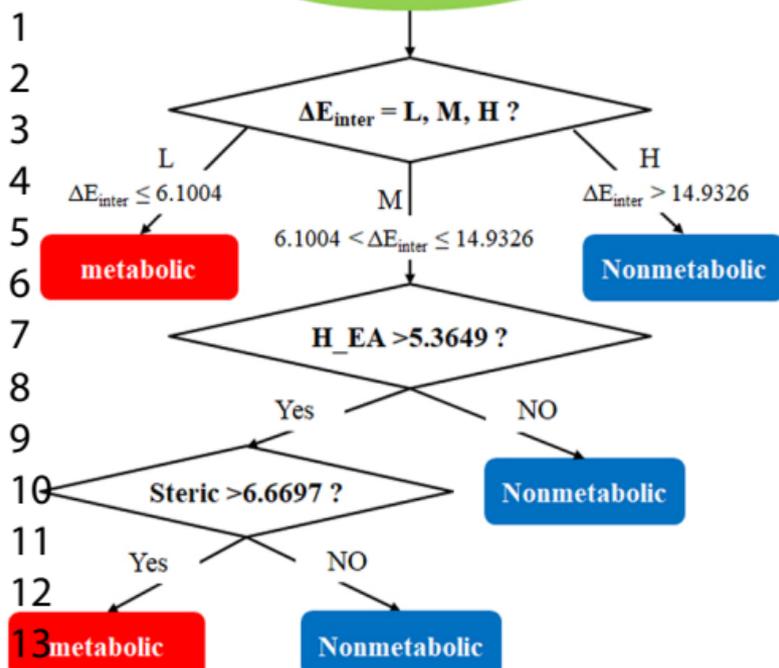
38
39
40
41

MOL487: Nonsubstrate
DT_{AOX}_Pred.: Substrate
DNN_Pred.: Nonsubstrate

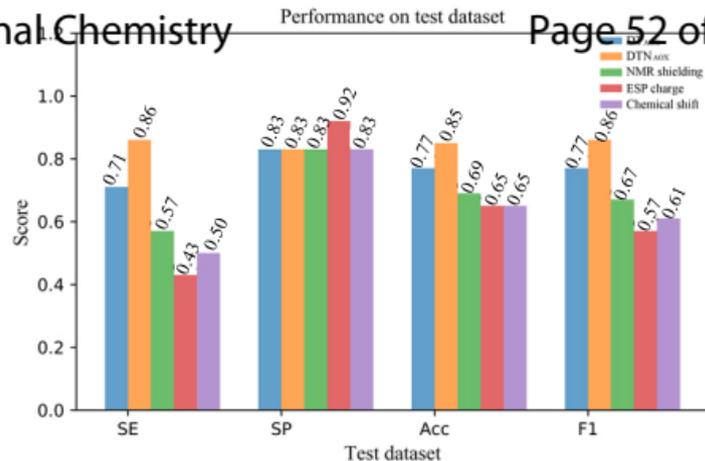


MOL488: Nonsubstrate
DT_{AOX}_Pred.: Substrate
DNN_Pred.: Nonsubstrate

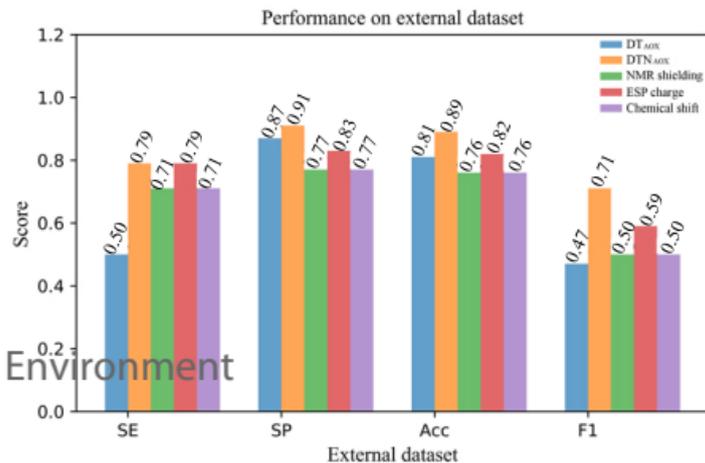
A

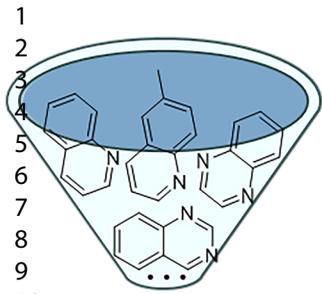


B

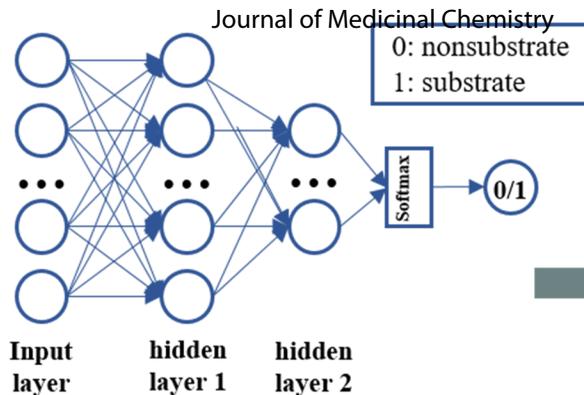


C

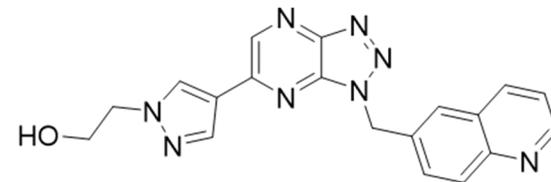




Datasets1



Substrate/nonsubstrate classification model

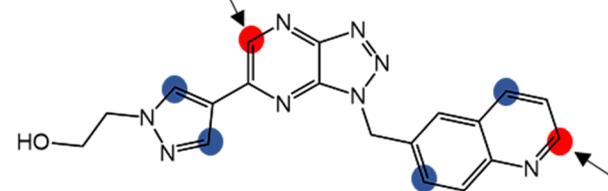


yes

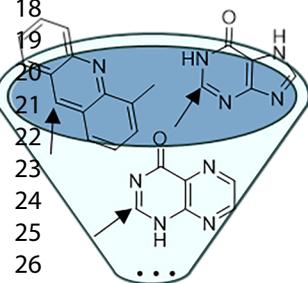
Substrate



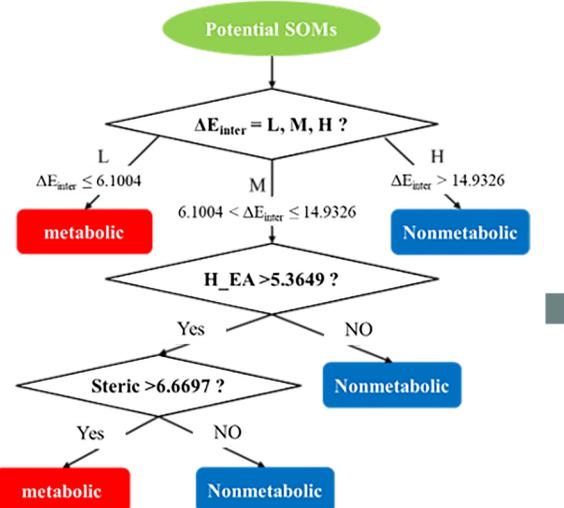
Confirmed by experiment



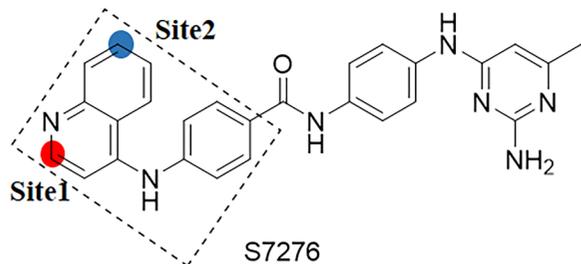
PF-04217903



Datasets2

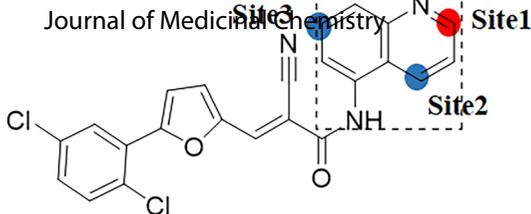


Meta sites classification model



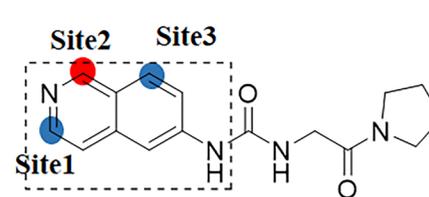
S7276

DNN_Pred.: Substrate
Exp.: substrate



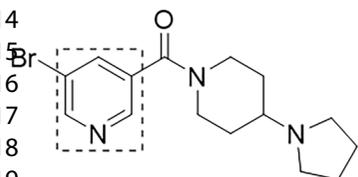
S7577

DNN_Pred.: Substrate
Exp.: substrate



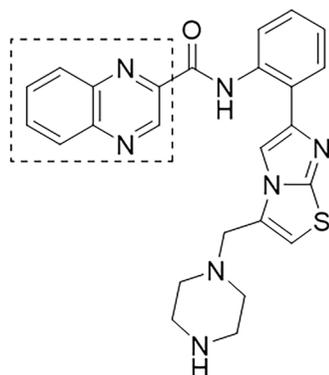
S7832

DNN_Pred.: Substrate
Exp.: substrate



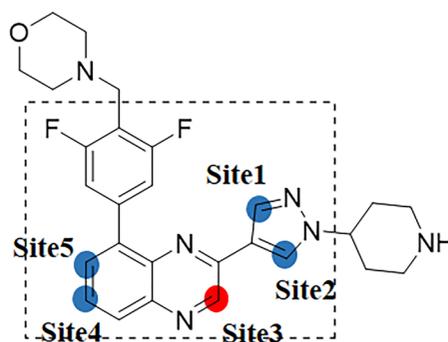
S7373

DNN_Pred.: Nonsubstrate
Exp.: substrate



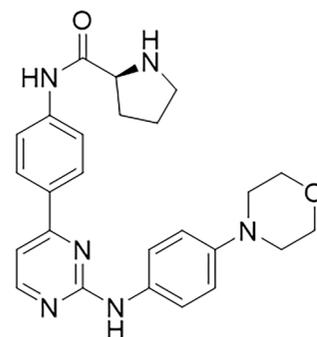
S1129

DNN_Pred.: Nonsubstrate
Exp.: substrate



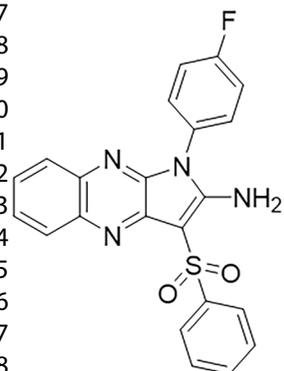
S2686

DNN_Pred.: Substrate
Exp.: substrate



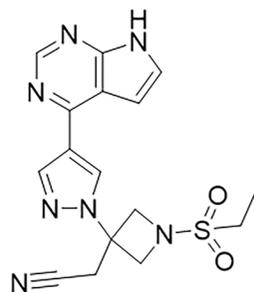
S7036

DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate



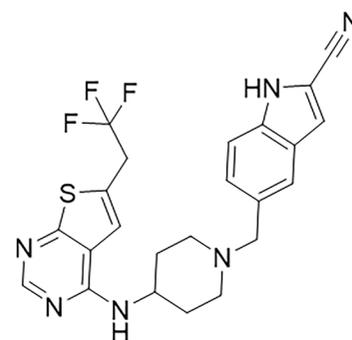
S5918

DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate



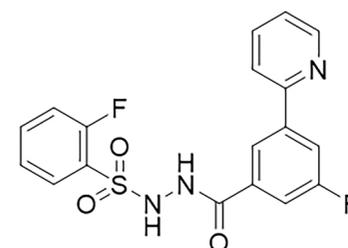
S2851

Pred.: Nonsubstrate
Exp.: Nonsubstrate



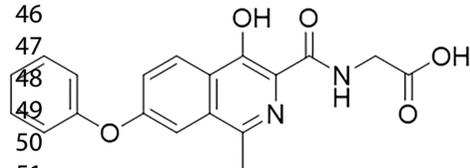
S7815

DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate



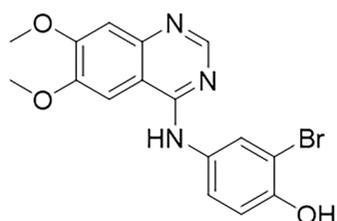
S8776

DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate



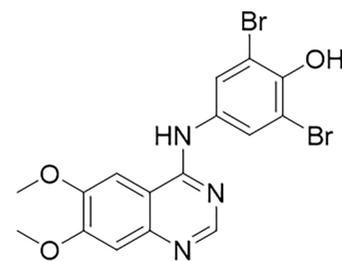
S1007

DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate



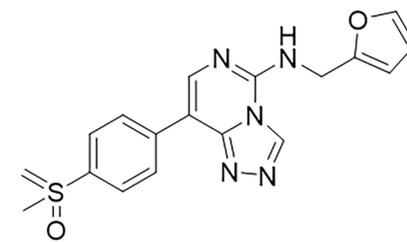
S2867

DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate



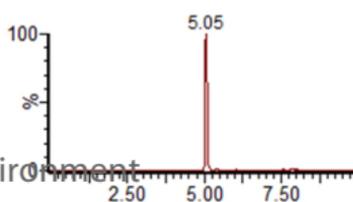
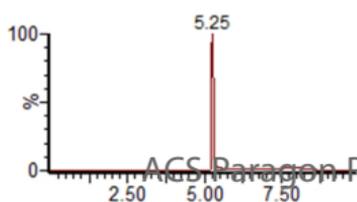
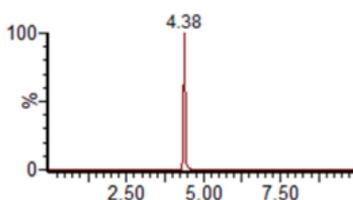
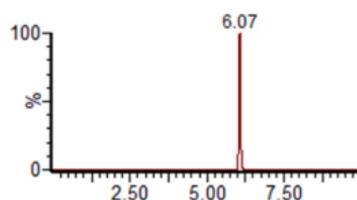
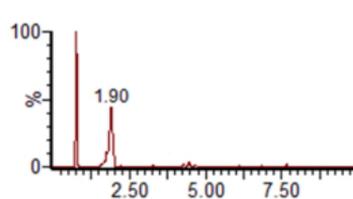
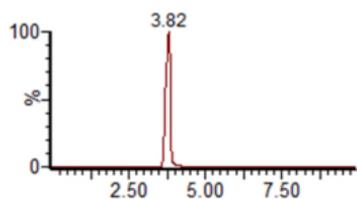
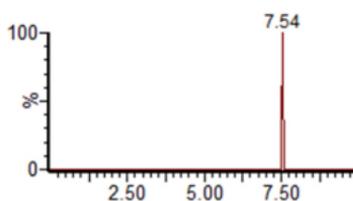
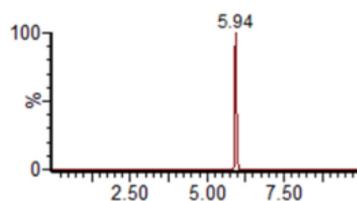
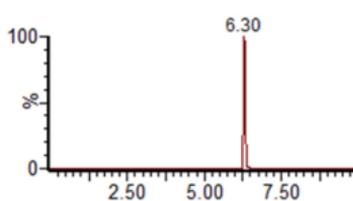
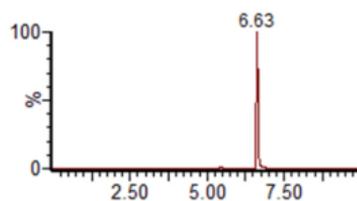
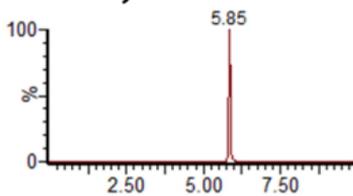
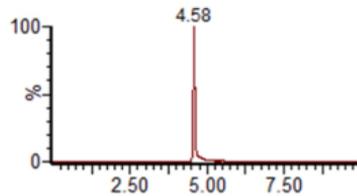
S5904

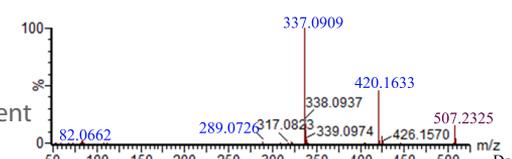
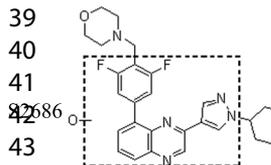
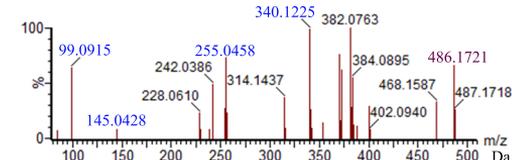
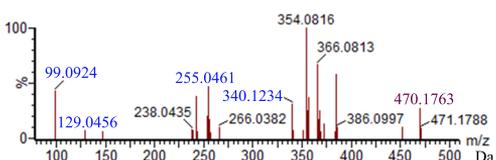
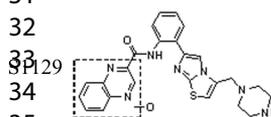
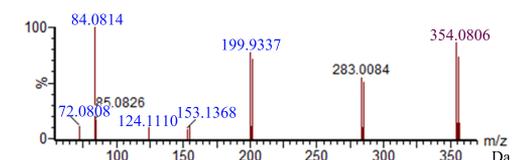
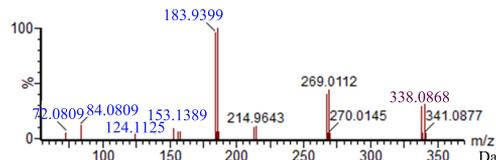
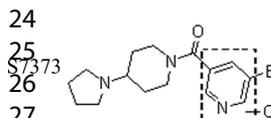
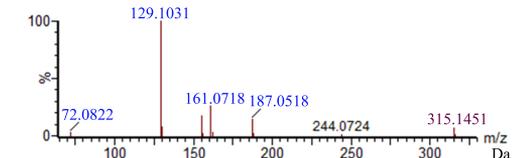
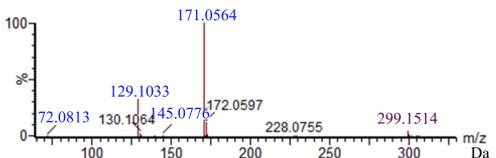
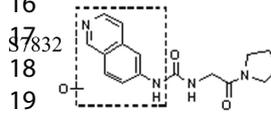
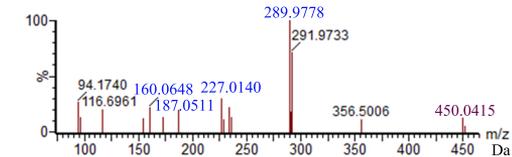
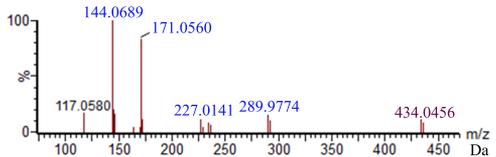
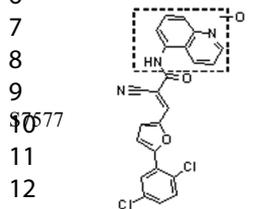
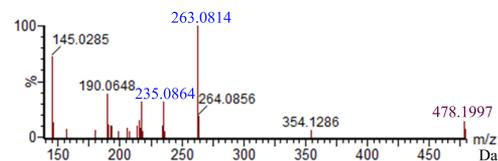
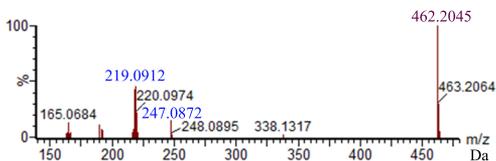
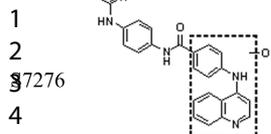
DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate

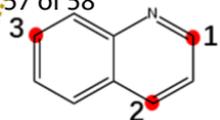


S8496

DNN_Pred.: Nonsubstrate
Exp.: Nonsubstrate





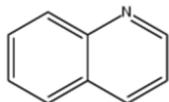


sites	ΔE_{inter}	Steric	C, H, N	C-N	C-H
1	data1	data2	descriptors1	bond order1	bond order2
2	data3	data4	descriptors2	bond order3	bond order4
3	data5	data6	descriptors3	bond order5	bond order6

$$M_D: \{label, \max(S_D), \text{mean}(S_D), \min(S_D)\}$$

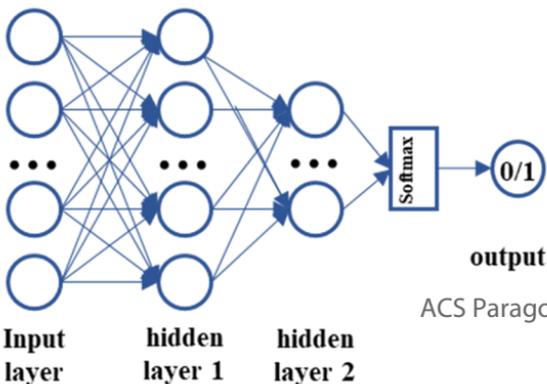
Molecular descriptors

Mol	Max(ΔE_{inter})	Min(Steric)	...	Mean(H_EA)	HOMO	Chemshift	...	d
1	data1	data2	...	descriptors1	bond order1	data3	...	datad
...								
m	datam1	datam2	...	descriptorsm	bond orderm	datam	...	datamd



Label	Molecular Descriptors
1	descriptors1
...	...
0	descriptorsm

Input: Label with molecular descriptors

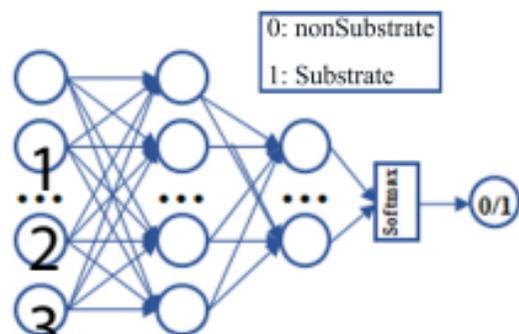


Substrate/Nonsubstrate

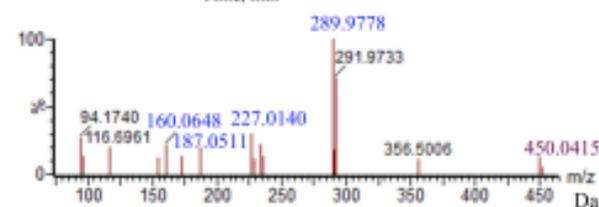
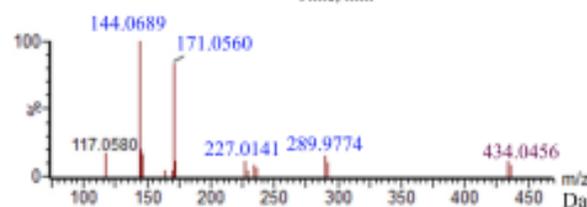
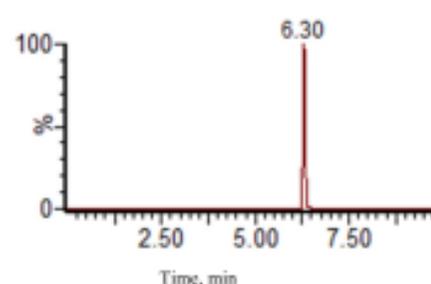
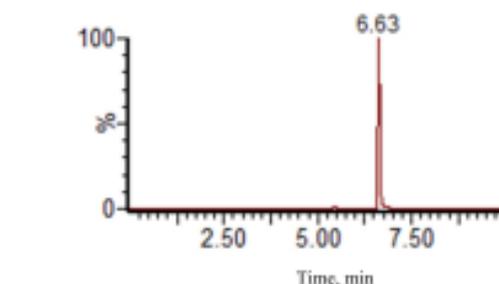
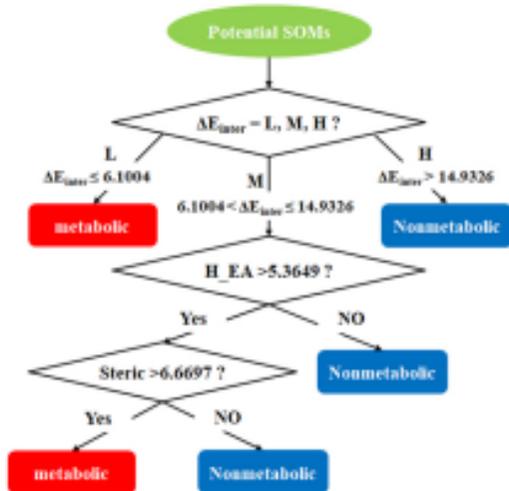
In silico prediction

Parent

Metabolite



prediction
1: substrate

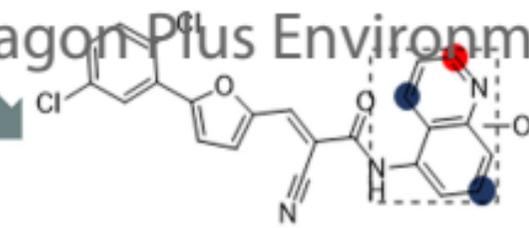


Input layer
Hidden layer1
Hidden layer2

1
2
3
4
5
6
7
8
9

- Metabolic site of AOX
- Non Metabolic site of AOX

ACS Paragon Plus Environment



confirmed by experiment
(analyzed by UPLC/Q-TOF MS)