XenoSite: Supporting Information

Jed Zaretzki, Matthew Matlock, and S. Joshua Swamidass*

Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63130

Table 1: Percentage of each substrate set with an experimentally observed SOM predicted in the top two rank-positions by the given method[a]

isozyme	1A2	2A6	2B6	2C8	2C9	2C19	2D6	2E1	3A4	HLM	Average
number of substrates	271	105	151	142	226	218	270	145	475	680	
XenoSite (MOL FP SQC)	87.1	85.7	79.5	86.7	85.9	86.7	88.5	83.5	87.6	88.7	86.0
XenoSite (FP SQC)	86.0	82.9	82.1	88.7	85.8	87.2	86.7	81.4	87.0	87.7	85.6
XenoSite (FP MOL)	83.0	84.8	80.1	85.2	85.8	86.2	88.2	82.8	84.2	88.5	84.9
XenoSite (MOL SQC)	86.0	82.9	79.5	87.3	83.2	87.6	86.3	80.7	86.5	87.5	84.8
XenoSite (FP)	84.1	81.9	80.8	84.5	86.3	87.2	86.7	82.1	84.8	88.1	84.7
XenoSite (FP QC)	86.4	80.0	80.8	85.2	86.3	89.0	86.7	79.3	85.9		84.4
XenoSite (SQC)	85.6	80.0	83.4	82.4	84.5	85.3	87.4	81.4	84.2	89.0	84.3
XenoSite (MOL)	83.0	81.0	78.8	88.0	86.3	84.9	85.2	81.4	85.5	88.5	84.3
XenoSite (TOP SCR)	84.9	81.0	78.8	84.5	85.4	84.4	86.3	80.7	84.8	89.4	84.0
XenoSite (MOL FP QC)	85.2	81.0	80.8	85.9	85.4	87.6	85.9	78.6	85.9		84.0
RS-Predictor (TOP QC SCR)	83.0	81.0	82.1	83.8	84.5	86.2	85.9	82.8	82.3	84.1	83.6
RS-Predictor (TOP SCR)	82.3	85.7	76.8	83.8	84.1	86.2	83.7	80.7	82.1	86.0	83.1
XenoSite (MOL QC)	85.2	78.1	76.2	83.1	84.5	88.1	84.8	80.0	83.2		82.6
XenoSite (QC)	84.9	76.2	78.8	83.1	84.1	87.6	87.0	77.9	83.0		82.5
Fingerprint (ATOM)	84.6	79.5	82.4	77.3	70.8	81.3	72.3	82.5	83.5	83.4	79.8
Fingerprint (ATOM·MOL)	82.1	75.3	79.9	76.7	69.9	81.1	67.8	81.6	81.5	81.8	77.8
StarDrop					78.0		75.3		74.1		75.8
SMARTCyp	80.0	86.0	77.0	83.0	84.0	86.0	83.0	82.0	78.0		73.9
Schrödinger					72.1		68.1		76.4		72.2
random model	26.0	31.9	24.8	22.6	22.2	20.2	21.1	36.5	21.0	26.3	25.3

[a]All NeuralNet models were trained using SOMs encoded with TOP and SCR descriptors, as well as the descriptors listed in parentheses.

*To whom correspondence should be addressed

Table 2: Original RS-Predictor definitions of topological (TOP) and quantum chemical (QC) SOM descriptors.[2]

label	definition	depth range	total number
	Topological Descriptors		148
NA_*d*_*e*	number of atoms depth bonds away of type element	0-4	35
PA_*d*_*e*	percentage of atoms depth bonds away of type element	1-4	28
NA_*d*_*p*	number of atoms depth bonds away of PATTY atom type	0-4	35
PA_*d*_*p*	percentage of atoms depth bonds away of PATTY atom type	1-4	28
span	(maximum path length from current atom) / (maximum path length from all atoms within the molecule)	0	1
$\text{Lip}_{\alpha,\beta}$	lipophilicity according to 5 different (α,β) dependent metrics	0	5
HBonded	number of hydrogen atoms bonded to base atom of metabolophore	0	1
NHBonded	number of non-hydrogen atoms bonded to base atom of metabolophore	0	1
RS	base atom of metabolophore is within a ring of size (0, 3-8)	0	7
AR	base atom of metabolophore is within an aromatic ring	0	1
MR	base atom of metabolophore is within multiple rings	0	1
RB	number of rotatable bonds for base atom of metabolophore (0-3)	0	5
	Quantum Chemical		392
HM	the projection of the given atom onto the 3 dimensional vector defined by the hydrophobic atoms of the molecule	0	1
BL	the average distance between the atom and all atoms it is bound too	0	1
SC	the charge kept by the atoms involved in the bond	0-2	11
AC	the charge not involved in bonding	0-2	11
ED	aromatic orbital electron density	0-2	11
F	Fukui reactivity index	0-2	11
N	nucleophilicity index	0-2	11
E	electrophilicity index	0-2	11
EERE	one-center electron-electron repulsion energy	0-2	11
ENAE	one-center electron-nuclear attraction energy	0-2	11
EE	total one-center electronic energy	0-2	11
area	solvent accessible surface area	0-2	11
$\delta_{\sigma-\sigma}$	$\sigma-\sigma$ component of atom-atom interactions from Mulliken Population analysis	0-1	20
$\delta_{\sigma-\pi}$	$\sigma-\pi$ component of atom-atom interactions from Mulliken Population analysis	0-1	20
$\delta_{\pi-\pi}$	$\pi-\pi$ component of atom-atom interactions from Mulliken Population analysis	0-1	20
$P_{\sigma-\sigma}$	$\sigma-\sigma$ bond order	0-1	20
$P_{\sigma-\pi}$	$\sigma-\pi$ bond order	0-1	20
$P_{\pi-\pi}$	$\pi-\pi$ bond order	0-1	20
P	Bond degree	0-1	20
ERE	Electronic resonance energy	0-1	20
EEE	Electronic exchange energy	0-1	20
ERPE	Electronic repulsion energy	0-1	20
NEAE	Nuclear-electron attraction energy	0-1	20

NNRE	Nuclear-nuclear repulsion energy	0-1	20
C	Coulomb interaction energy	0-1	20
TENE	Total of electronic and nuclear energy	0-1	20

New SMARTCyp Descriptors

After obtaining the preliminary results shown in Table 1 we applied the up-to-date version of SMARTCyp (v2.4) to calculate a new SOM reactivity descriptor to replace the previously encoded SMARTCyp reactivity (SCR) descriptor obtained from an antiquated version of SMARTCyp (v1.5). In addition, a new set of fragment-based atomic descriptors partially developed by the creator of SMARTCyp was also used to encode putative SOMs.[1] These two descriptors classes were incorporated into the optimal XenoSite models that use SOMs encoded with TOP, SCR, SQC, FP, and MOL descriptors. We found that replacing SCR reactivities from version 1.5 of SMARTCyp with those from version 2.4 resulted in models with poorer overall performance. Models trained with both the fragment-based descriptors and the new SCR descriptors also had poorer overall performance than previous optimal models. These results were not expected and seem to indicate that the additional signal represented in these two new classes of descriptors has already been elucidated by neural network optimization of the descriptor sets that have already been implemented.

References

(1) Long, A.; Rydberg, P. Enrichment of True Positives from Structural Alerts Through the Use of Novel Atomic Fragment Based Descriptors. *Molecular Informatics*. **2013**, 32 (1), 81-86.

(2) Zaretzki, J.; Bergeron, C., Rydberg, P., Huang, Tao-wei, Bennett, K., Breneman, C. A New Tool for Predicting Sites of Cytochrome P450-Mediated Metabolism Applied to CYP 3A4. *Journal of Chemical Information and Modeling*. **2011**, 51 (7), 1667-1689.