Predicting Sites of Metabolism with Artificial Neural Network Ensembles

Marvin Waldman, Robert Fraczkiewicz, Jinhua Zhang, Robert D. Clark, and Walter S. Woltosz

Simulations Plus, Inc., 42505 10th Street West, Lancaster, CA 93534, USA (www.simulations-plus.com)

Introduction

Hepatic first-pass metabolism of many drugs and prodrugs plays a key role in their oral bioavailability. The human cytochrome P450 enzymes are responsible for the metabolism of most drugs. Knowledge of likely sites of metabolic attack in a drug molecule can aid in designing out unwanted metabolic liabilities early on in the drug discovery process, as well as in the design of prodrugs where metabolic transformation is desired. Using datasets constructed from literature compilations and the industry's largest commercially available metabolite database*, we have constructed models based on artificial neural network ensembles that predict likely sites of metabolism for several CYP isoforms, including 2C9, 2D6, and 3A4. The models employ rapidly calculated critical atomic descriptors that describe charge, reactivity, steric accessibility, and other properties of candidate atoms and their local environments. The resulting excellent model performance is shown based on various statistical criteria, as well as specific examples demonstrating scope and limitations.

Typical Oxidations Mediated by P450

- Carbon Hydroxylation

$$[FeO]^{3+} + HC- \rightarrow [Fe]^{3+} + HOC-$$

- Heteroatom Dealkylation

$$[FeO]^{3+} + HC{-}X \rightarrow [Fe]^{3+} + HOC{-}X \rightarrow O{=}C + HX \qquad X{=}O, N, S$$

- Heteroatom Oxidation

$$[FeO]^{3+} + :X- \rightarrow [Fe]^{3+} + [O^{-}]{-}[X^{+}]- \qquad X{=}N, S$$

- Epoxidation (unsaturated systems)

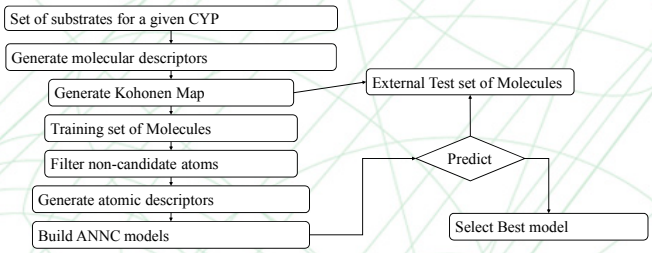
$$[FeO]^{3+} + >C{=}C< \rightarrow [Fe]^{3+} + >C(O)C<$$

Sources of Experimental Data

Extensively curated compilations:

- *Accelrys Metabolite™ Database (formerly Symyx).
- Data published by Sheridan et al. [1].
- Review of literature – both old and new.

Model Building Process



Atomic Descriptors

Descriptors generated by ADMET Predictor™ v5.0:

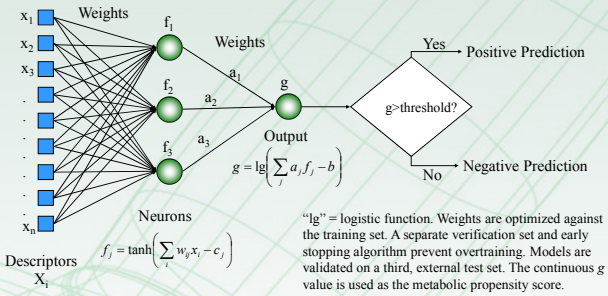
- Partial Atomic Charges and Reactivities
  - EEM-Hückel charge model (accounts for resonance effects)
    - Parameterized from our *ab initio* database of partial atomic σ and π charges
    - EEM (Electronegativity Equalization Method) for σ charges
    - Hückel model for π charges
    - Reactivities
      - EEM σ atomic Fukui indices
      - Hückel π frontier orbital atomic densities
- E-State indices
- Local shape descriptors
  - Sheridan's SPAN
  - Atomic volumes
- Others
  - polarizability, electronegativity, autocorrelation vectors
  - special proprietary

Performance of Final Models

P450 Site	Train/Verify Set						Test Set					
Model	No. Molecules	No. Atoms	No. Sites	Top 1	Top 2	Top 3	No. Molecules	No. Atoms	No. Sites	Top 1	Top 2	Top 3
1A2	273	2796	442	76.9%	90.8%	97.4%	37	344	60	73.0%	91.9%	94.6%
2C9	211	2421	308	72.0%	90.0%	95.7%	41	408	58	85.4%	87.8%	92.7%
2C19	202	2238	305	85.1%	97.5%	99.5%	30	267	40	90.0%	96.7%	96.7%
2D6	284	3386	428	80.7%	93.7%	97.2%	44	469	65	88.6%	95.5%	97.7%
3A4	578	8181	909	72.0%	87.7%	94.1%	136	1981	215	72.8%	87.5%	93.4%

"Top *n*" columns show the percentage of molecules with an observed site of metabolism among its top *n* scoring atoms.

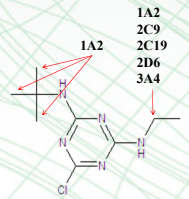
Classification ANN



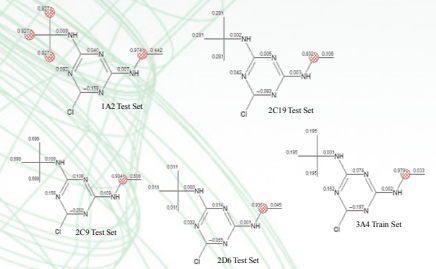
"lg" = logistic function. Weights are optimized against the training set. A separate verification set and early stopping algorithm prevent overtraining. Models are validated on a third, external test set. The continuous *g* value is used as the metabolic propensity score.

Examples

Terbuthylazine [2]

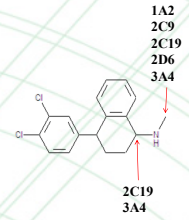


Reported sites [2]

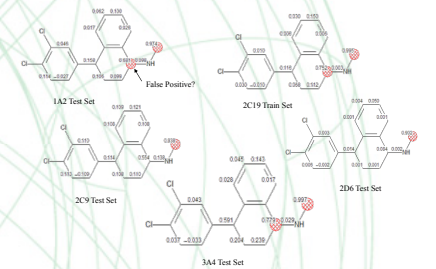


Predicted sites

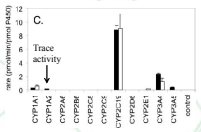
Sertraline [3]



Reported sites [3]

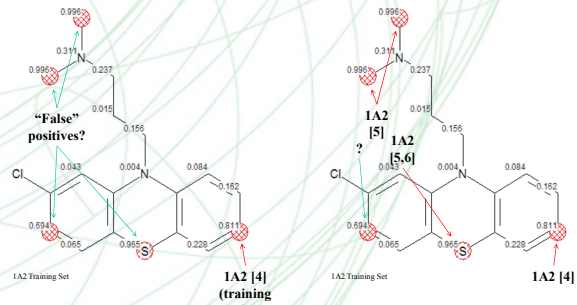


Predicted sites



Raw experimental results from Obach et al. [3] – rates of ketone formation in recombinant CYP assays. Solid bars correspond to sertraline, open bars to desmethylsertraline. CYP 1A2 does indeed attack the ring site, albeit the reaction is relatively slow.

Chlorpromazine [4-6]. CYP 1A2 Model.



"False" positives in 2009 became true positives in 2010!

"If a tree falls in the forest, but no one is there to hear it, does it still make a sound?"

References

1. Sheridan RP, Korzekwa KR, Torres RA, Walker MJ. J Med Chem. *50*: 3173-3184 (2007)
2. Lang DH, Rettie AE, Böcker RH. Chem Res Tox. *10*:1037-1044 (1997)
3. Obach RS, Cox LM, Tremaine LM. Drug Metab Dispos. *33*:262-270 (2005)
4. Yoshii K, Kobayashi K, Tsumuji M, Tani M, Shimada N, Chiba K. Life Sciences *67*:175-184 (2000)
5. Wójcikowski J, Boksa J, Daniel WA. Biochem Pharmacol. *80*:1252-1259 (2010)
6. Cashman JR, Yang Z, Yang L, Wrighton SA. Drug Metab Dispos. *21*:492-501 (1993)

ACKNOWLEDGEMENTS

This work has been funded by the SBIR Grant 5R44CA130388-03 from the National Institute of Health.

simulationsplus, inc.