Determining Blood-Brain-Barrier Penetration of Small Molecule Drugs Using Machine-Learning Modelling Algorithms

T Bogdanowicz, M Gupta, C Barden, D Weaver

Krembil Research Institute, UHN, 60 Leonard Ave, TO, Canada

UNIVERSITY OF TORONTO LESLIE DAN FACULTY OF PHARMACY

Krembil UHN Relentless.

Context:

Determining BBB penetration of drugs is a critical step in drug development. The following are key parameters that computational chemists refer to when investigating small molecule CNS druggability:

- Kp = total brain:blood concentration ratio
- Kpuu = unbound brain:blood concentration ratio
- Kpuu,cell = unbound intracellular:extracellular concentration ratio
- BBB transporter effects:
  - Efflux: Pgp, BCRP, MRP
  - Influx: OCT1, OCT2

Various BBB models exist to guide CNS druggability of small molecules:

- 1997: Pfizer's *Lipinski Rule of Five*
- 2010: Pfizer's *MPO*
- 2016: Teva Pharmaceuticals' *TEMPO*

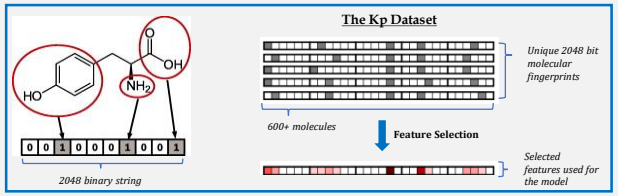
Objectives:

- Support drug design of small molecules via predicting key BBB penetration parameters.
- Build robust machine-learning models & package them into a user-friendly calculator, available for public access.
- Create a novel BBB predictive model to guide CNS druggability:
  - The Brain Exposure Efficiency Score

Methods:

For each dataset:

- Binary 2048 bit molecular fingerprints were computed for each molecule.
  - Each bit represents a structural feature that a molecule may have.
- Feature selection was conducted via recursive gradient-boosting regressors to identify the most critical bits of the fingerprint.
  - A data-dependent amount of bits are used for each model.
  - E.g. the Kp model needs to use 200/2048 bits while a Kpuu model only needs 100/2048.



- Each dataset was split into a training set (75%) & test set (25%).
- The coefficient of determination (R^2) is used to assess the accuracy of the models at predicting the training & test sets' target parameter.

Results:

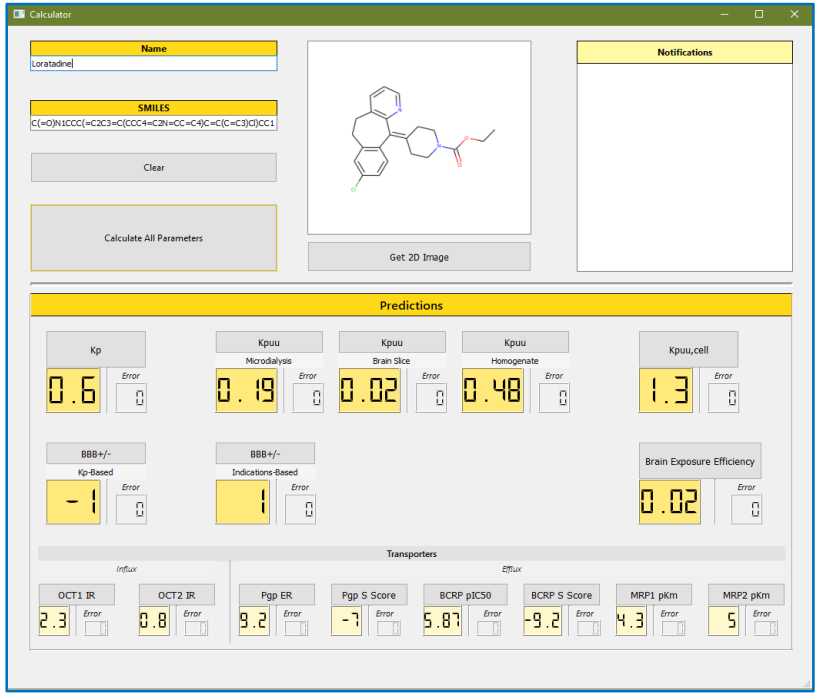
- Deep-learning artificial neural networks, boosted decision trees, gradient boosting regressors, Bayesian gaussian processes, gradient boosted classifiers, and multilayer perceptron models were constructed and used to model each parameter.
- The *Gradient Boosting Regressor (GBR)* & the *Bayesian Gaussian Process (BGP)* were identified to be the most accurate.

Parameter	Kp	Kpuu			Kpuu,cell
		Microdialysis	*Brain Slice*	*Homogenate*	
# of mol	600+	50+	60+	80+	50+
Model	GBR	BGP	GBR	GBR	GBR
Training R^2	0.97	0.99	1.0	1.0	1.0
Test R^2	**0.81**	**0.92**	**0.90**	**0.87**	**0.95**

- Two BBB+/- models were made:
  - Indications-based: collecting 1900+ drugs and grouping them via CNS indication (+) or nonCNS indication (-)
  - Kp-based: in which Kp >0.1 is (+) and ≤ 0.1 is (-)

Parameter	BBB +/-	
	Indications-based	Kp-based
# of mol	1900+	600+
Model	MLP	GBC
Training R^2	0.99	0.96
Test R^2	**0.99**	**0.97**

Parameter	# of mol	Model	Training R^2	Test R^2
Pgp ER	160+	GBR	1.0	**0.85**
Pgp S		GBR	1.0	**0.96**
BCRP pIC50	160+	GBR	0.99	**0.90**
BCRP S	30+	GBR	1.0	**0.99**
MRP1 pKm	20+	GBR	1.0	**0.99**
MRP2 pKm	20+	GBR	1.0	**0.98**
OCT1 IR	190+	GBR	1.0	**0.92**
OCT2 IR	120+	GBR	0.92	**0.84**



Brain Exposure Efficiency:

- The Brain Exposure Efficiency score (BEE) was created to be a simplified measurement of CNS druggability.

For a given drug,

$$BEE = \frac{[unbound\ concentration\ in\ brain\ ISF]}{dose}$$

→ Dose normalized, free concentration in brain

- 41 drugs were selected, covering a diverse molecular space.
- Their experimental BEE values were calculated.
- Machine-learning predictive regressors were built.
- The most accurate BEE model is the Bayesian Gaussian Process that predicts BEE values of an 11 molecule test set with an R^2 of 0.97

Parameter	Brain Exposure Efficiency
# of mol	40+
Model	BGP
Training R^2	1.0
Test R^2	**0.97**

The Development of a Blood-Brain-Barrier Calculator:

To increase practical usability for drug design teams & academic research groups, all predictive models have been integrated into an executable program, available for use on any computer.

Input a molecule's SMILES code to compute any available parameter and generate its 2D image.

Conclusion & Future Direction:

- Key BBB penetration parameters have been accurately modelled & packaged into a practical calculator.
- The BEE model is a new CNS druggability tool, which predicts brain concentration of small molecules.
- Future work lies on improving the models, expanding datasets, modelling new BBB parameters (new transporters), increasing usability of the calculator (autofilling SMILES code if molecule name is given).

References:

1. Dolgikh E, *et al*. QSAR model of unbound brain-to-plasma partition coefficient, Kp,uu,brain: incorporating P-glycoprotein efflux as a variable. 2016.
2. Friden M, *et al*. Structure-brain exposure relationships in rat and human using a novel data set of unbound drug concentrations in brain interstitial and cerebrospinal fluids. 2009.
3. Pedregosa F, *et al*; Scikit-learn: Machine Learning in Python. 2011.
4. RDKit: Open-source cheminformatics; http://www.rdkit.org