Pharmaceutical Research, Vol. 26, No. 5, May 2009 (© 2009)
DOI: 10.1007/s11095-009-9832-3

Commentary

A Turning Point For Blood–Brain Barrier Modeling

Sean Ekins[1,2,3,5] and Alexander Tropsha[4]

Received January 9, 2009; accepted January 12, 2009; published online January 23, 2009

Recently we published independently in the same issue of this journal two papers dealing with computational modeling of Blood-Brain Barrier (BBB) penetration by organic compounds (1,2). The BBB is a physiological barrier expressing numerous transporters that seperates the brain from the bloodstream and functions to maintain homeostasis while also acting as a barrier to many pharmaceuticals. There have been at least 40 studies published since 1988 which we have summarized (1,2) that derive various quantitative structure activity relationship (QSAR) or classification models for BBB penetration data. The publication of our papers covering the same topic in itself perhaps is not unusual due to the importance of predicting brain penetration of drugs in the pharmaceutical industry. Perhaps more surprising however was that despite the important differences in our modeling approaches, overall we were able to achieve comparable levels of prediction accuracy. This concurrence of results indicates that perhaps we have captured the intrinsic relationship between compound structure and its ability to cross the BBB barrier, indicating that it is possible to achieve highly predictive models. This perspective summarizes these studies briefly and highlights areas of future research.

First, let us look at the two studies in more detail. The study by Zhang *et al.*, generated *k*-nearest neighbors (*k*-NN) and support vector machine (SVM) QSAR quantitative models using between 184 and 346 Dragon, MOE or MolConnZ descriptors with a training set of 144 molecules and a test set of 15 molecules (2). In addition these models were tested with two binary classification sets of 99 and 267 compounds. The study of Kortagere *et al.*, generated a simple regression model with eight descriptors (logP, TPSA, logS, mass, volume, number of rotatable bonds, number of oxygen atoms and number of nitrogen atoms calculated in MOE) for 78 compounds and tested it with 100 compounds (1). In addition, SVM classification models with the eight MOE or shape signatures descriptors were built with training sets of 376 or 351 compounds (in both cases the datasets were composed of more BBB+ than BBB- compounds), leave 20%-out and 10-fold cross validation was performed. All QSAR and classification models in this study were also tested on an external set of 389 drugs in order to predict known BBB+ compounds (1). So in both cases there were significant differences in the approaches used but also areas of overlap. One approach focused on generating continuous property QSAR models while the other produced classification as well as a continuous QSAR models. In both studies the binary test sets were also of a comparable size to the training sets. In contrast to many of the prior BBB models reviewed in each paper, both studies placed a significant emphasis on the validation of the models generated using statistics for testing as well as measures for the applicability domain (2) or chemical space (1) of the training and test sets. The use of the Euclidean distance applicability domain cut-off could be used to improve test set prediction accuracy (2), while the regression and classification studies illustrated a good overlap of the chemical space as described using principal component analysis with the molecular descriptors for the training and test sets (1).

The methodological differences notwithstanding the performance of the best models in both studies appear remarkably congruent. The SVM models with shape signatures descriptors had prediction accuracies of 80–83% for the 10 fold cross validation and could correctly predict 84% of BBB+ drugs (1). The *k*-NN-MOE model performed the best in the second study with an accuracy of 82% for the 99 compound test set and the *k*-NN-Dragon model had an accuracy of 100% with the applicability domain (2). In both cases when the datasets contain more BBB+ compounds than BBB- there is a clear bias towards more accurate prediction of BBB+ (1,2). One dataset we used for SVM modeling was more balanced (3) and the resultant SVM models were also approximately equivalent in the prediction of BBB+ and BBB− [Table III in (1)]. Interestingly the 78 molecule regression model performed better when predicting BBB− compounds in all of the different test sets evaluated.

It would appear that we may have demonstrated in both papers an upper limit on test set classification (in the absence of applicability domain measures) of approximately 80%. Both

[1] Collaborations in Chemistry, 601 Runnymede Avenue, Jenkintown, Pensylvannia 19046, USA.
[2] Department of Pharmaceutical Sciences, University of Maryland, Baltimore, Maryland 21201, USA.
[3] Department of Pharmacology, University of Medicine & Dentistry of New Jersey (UMDNJ)–Robert Wood Johnson Medical School, 675 Hoes lane, Piscataway, New Jersey 08854, USA.
[4] The Laboratory for Molecular Modeling, School of Pharmacy, University of North Carolina at Chapel Hill, CB# 7360, Chapel Hill, North Carolina 27599-7360, USA.
[5] To whom correspondence should be addressed. (e-mail:ekinssean@yahoo.com)

groups also suggested a consensus of the different models in each individual paper. In one case we suggested it represented an average of all the models (1) while in another the prediction accuracy was slightly lower than the best individual model (2). Perhaps we should explore developing a joint consensus prediction model by combining models developed in both studies; as we have convincingly demonstrated recently, the consensus modeling appears highly beneficial as compared to any of the contributing models (4). Overall both studies are clearly complimentary, reflecting at least the state of the art in algorithms, interpretable descriptors that describe shape and surface area (2D and 3D) and the largest BBB datasets currently available from the literature.

So what should we focus on in the future? Certainly we require larger training and test sets that are well balanced in the distribution of logBB values or in the binary categories. In current studies we have had access to at most, nearly 400 compounds with classification data or 159 compounds with quantitative logBB data. These molecules may not all be drugs or drug-like and the representation of chemical space, lead-likeness, drug-likeness etc., will be sparse. It is important therefore that future general BBB models are used to predict drug databases (not in the model) as we attempted in both our studies. At the very least a series of lead compounds should be compared to show the limitations of a global model *versus* a local model for predicting BBB. Also as indicated in one of our studies (2) we need to identify molecules that may be substrates for efflux transporters like P-glycoprotein (P-gp) which could confound our predictions of BBB permeability. This might be achieved by parallel prediction of molecules P-gp substrate liability using a QSAR model, pharmacophore or classification model (in the same way as carried out in these BBB studies) using machine learning methods.

After over 20 years of computational studies and many BBB models (1,2) our independent studies suggest that we may be approaching a turning point in the modeling of this important property. Without the luxury of large datasets (as available for solubility and logP, logD which include many thousands of data points) validation of the latest BBB models with up to several hundred molecules, suggests we have reached a peak in prediction accuracy for known drugs and drug-like chemicals. Although there is certainly still room for methodological improvements, perhaps the next stage is for these BBB models to become more widely disseminated and used in drug discovery to promote experimental validation studies that would in turn enrich experimental datasets available for modeling. This could be facilitated in the same way that other physicochemical and molecular properties are predicted routinely in databases of small molecules such as ChemSpider (www.chemspider.com) and eMolecules (www.emolecules.com). We hope that other researchers will continue to validate these and future models and in turn publish their valuable data in this and other journals.

REFERENCES

1. S. Kortagere, D. S. Chekmarev, W. J. Welsh, and S. Ekins. New predictive models for blood brain barrier permeability of drug-like molecules. *Pharm. Res.* **25**:1836–1845 (2008). doi:10.1007/s11095-008-9584-5.
2. L. Zhang, H. Zhu, T. I. Oprea, A. Golbraikh, and A. Tropsha. QSAR modeling of the blood–brain barrier permeability for diverse organic compounds. *Pharm. Res.* **25**:1902–1914 (2008). doi:10.1007/s11095-008-9609-0.
3. H. Li, C. W. Yap, C. Y. Ung, Y. Xue, Z. W. Cao, and Y. Z. Chen. Effect of selection of molecular descriptors on the prediction of blood-brain barrier penetrating and nonpenetrating agents by statistical learning methods. *J. Chem. Inf. Model.* **45**:1376–1384 (2005). doi:10.1021/ci050135u.
4. H. Zhu, A. Tropsha, D. Fourches, A. Varnek, E. Papa, P. Gramatica, T. Oberg, P. Dao, A. Cherkasov, and I. V. Tetko. Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis. *J. Chem. Inf. Model.* **48**:766–784 (2008). doi:10.1021/ci700443v.