Classification of membrane permeability of small drug candidates:
A methodological investigation

Poster B101

Berith F. Jensen[1], Per B. Brockhoff[1] & Hanne H.F. Refsgaard[2]

[1] (bfj@kvl.dk) Department of Natural Science, The Royal Veterinary- and Agricultural University, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark
[2] Department of Drug Metabolism, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark

Introduction

In the discovery of small molecules (New Chemical Entries, NCEs) drug candidates it is of paramount importance that the compound has the ability to cross several barriers in getting from the site of administration to the site of action.

Most pharmaceutical companies have established high throughput *in vitro* assays to classify compounds according to permeability properties already in the discovery phase.

In Refsgaard et *al*. [1] such data was applied for the classification of compounds into **THREE CLASSES** with respect to apperent membrane penetration by passive diffusion (Papp) in a cell based assay:

1) **Papp < 4 * 10-6 cm/s**
2) **Papp 4 - 20 * 10-6 cm/s**
3) **Papp > 20 * 10-6 cm/s.**

The calibration data consisted of 712 NCEs and the method **k-NEAREST NEIGHBOUR** (k-NN) was applied using nine molecular descriptors.

THIS POSTER is an extention of Refgaard et *al*.
Where we have:
- **1040** small drug candidates
- the same N**ine** M**olecular** D**escriptors**
- divided data in T**hree** C**lasses**
- investigated S**ix** C**lassification** T**echniques**

Results

Table 2: Result of test set validation of six different classification techniques. CV = Cross validation. Test = Test set validation.

	False Positives		True Positives		Not Classified	
	CV	Test	CV	Test	CV	Test
LDA	13%	14%	70%	68%	35%	42%
QDA	13%	10%	61%	62%	21%	20%
k-NN	**6%**	**0%**	**62%**	**67%**	**17%**	**15%**
DANN	7%	1%	63%	68%	16%	19%
SIMCA	26%	32%	86%	82%	0%	0%
ClassTree	**15%**	**11%**	**83%**	**79%**	**0%**	**0%**

Table 3: Result of test set validation of k-NN Model

		K-NN Model			
		Non-Permeable	Partly Permeable	Highly Permeable	Not Classified
Assay	Non-Permeable	**66**	14	3	16
	Partly Permeable	**0**	29	11	7
	Highly Permeable	**0**	11	43	8
	Sum	66	54	57	**31**

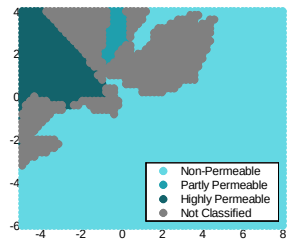

Figure 1: Plot of k-NN model (a PCA based projection)

Summary of Results

The k-NN and ClassTree models had the best classification results (Table 2). The **k-NN Model Was Chosen** as we regard it important to minimize the loss of potential drug candidates.

Validated by cross and test set validation the k-NN model was found to be **Stable** (Table 2). The model did not describe the partly permeable compounds well (Table 3 & Figure 1).

Comparing our results to the results obtained by Refsgaard et al. [1], we found that the percentage of false positives was the same in both studies, while the number of true positives found by Refgaard et al. was higher (75%).

However, in Refsgaard et al. the cross and test validation gave different results (1% (CV) & 19% (test set)) with respect to the percentage of non-classified compounds indicating that the model was not as stable as the k-NN model obtained in the present study.

Conclusion

- **k-NN** was a suitable classification method for permeability data
- Validated by cross and test set validation the k-NN model was found to be **Stable**
- The model classified 83-85% of the compounds, found 62-67% true positives and made **0-6% False Positive Predictions**

Material & Methods

Table 1: Training and test set divided into three classes with respect to membrane penetration as described above.

	Training Set	Test Set
1: Non-Permeable	399	99
2: Partly Permeable	187	47
3: Highly Permeable	246	62
Sum	832	208

Classification Techniques [2]

- Linear Discriminant Analysis **(LDA)**
- Quadratic Discriminant Analysis **(QDA)**
- k-Nearest Neighbor **(k-NN)**
- Discriminant Adaptive Nearest-Neigbor **(DANN)**
- Soft Independent Modeling of Class Analogy **(SIMCA)**
- Classification trees **(ClassTree)**

Restrictions and Goals

•The classification methods provide a set of posterior probabilities of class memberships when predicting new candidates. It was decided that the class posterior **(Pr) should be >0.60** for classification.

•The amount of **False Positives** (compounds that were falsely classified as non-permeable) should be less than 10 percent.

•The percent of **True Positives** (compounds that were correctly classified as non-permeable) should be maximized

•The percentage of **Not Classified** compounds should be less than 25%.

Molecular Descriptors

- ClogP
- mlogP
- Number of flex bonds
- Number of hydrogen bond acceptors
- Number of hydrogen bond donors
- Molecular surface area
- Polar surface area
- Molecular volume
- Molecular weight

References

[1] Refsgaard HHF, Brockhoff PB, Jensen BF, Guldbrandt M, Christensen MS. *In silico* prediction of membrane permeability from calculated molecular parameters (Submitted).

[2] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd ed. New York (NY): Springer-Verlag; 2001.