Chemometric Analysis to probe Blood Brain Barrier Permeability of New Chemical Entities (NCEs)

*Inam Tahir and Ishrat Jabeen**
School of Interdisciplinary Engineering & Sciences (SINES), National University of Science and Technology, Sector H-12, Islamabad 44000, Pakistan

*****Correspondence: ishrat.jabeen@rcms.nust.edu.pk (I.J.);
Tel.: +92-51-90855732

ABSTRACT
In present world one of greatest challenges faced by pharmaceutical industry is to maximize efficacy of various neuroactive agents and minimize the risk of neurotoxicity of drugs designed for peripheral body systems. Bioavailability of a neuroactive active agent depends upon its transport across blood brain barrier and thus, is a factor of vital importance in determining drug efficacy. Over the past few decades, intensive research efforts have been made to elucidate blood brain barrier permeability of compounds. However, these experiments are very costly and time demanding endeavors. In this study various chemometeric models, including Principal Component Analysis (PCA) and Partial Least Square Analysis (PLS) using GRID Independent Descriptors have been developed to predict BBB permeability of a diverse dataset of 218 compounds. Our model elucidates two hydrogen bond donor groups at a mutual distance of 6.00 to 6.40 Å and a hydrophobic group at a distance of 10-10.4Å from one of the hydrogen bond donor groups may have a positive impact on blood brain barrier permeability of already marketed neuroactive agents.

Key words: Blood Brain Barrier, GRIND, Computational Drug Design, Chemometeric Analysis, PCA, PLS

INTRODUCTION

Blood brain barrier is a selective barrier which consists of capillary forming endothelial cells and perivascular neurons[1]. These endothelial cells are joined to each other through transcellular proteins such as claudins, occludinand JAMs (Junctional adhesion molecules) forming tight junctions (TJs).Unlike a normal capillary these tight junctions restrict the passive movement of the molecules across blood brain barrier, providing protection to the brain from pathogens and toxins[2]. Numerous transporters,such as P-glycoprotein (P-gp) are also present on luminal and abluminal membranes of the cerebral endothelial cells that are responsible for controlling transcellular traffic between brain and blood. [3].This highly selective nature of BBB restricts free movement of the molecules and thus, providing protection brain. Therefore, a lot of potentially active drugs are also not allowed to enter brain through normal circulatory system due to tight junctions and high levels of efflux transporters[4]. This protective mechanism of blood brain barrier has been a major obstacle in designing drugs for various psychological and neurological diseases for a couple of decades[5]. Therefore, it is very important to know the BBB permeability of drugs-like entities achieve maximum efficacy of neuro active agents and to avoid psychotropic effects of non CNS drugs. A gold standard parameter used to measure the BBB permeability experimentally is log BB. Log BB is described represents the proportion of drug concentration in the brain compared to drug concentration in the plasma. Log of this ratio of concentration between brain and blood is signified as log BB and is given by Equation mentioned below.

$$\textbf{Log BB} = \textbf{Log} \left(C_{BRAIN}/C_{BLOOD}\right)$$

Higher the value of log BB for a compound, higher is its BBB permeability[6].Various traditional methods have been used previously to calculate Log BB experimentally. These include*in vivo*techniques which involve the dissection of laboratory rats and*invitro* techniques such as PAMPA and IAM.PAMPA has been able to show good prediction ability[7-11]. PAMPA was established in 1998 by Kansy et al to forecast passively permeation via the GI tract. [12]however, Di, *et al* adapted it for BBB studies[13, 14].However, these methods are time consuming and very expensive. Therefore, computational models have been developed in past few decades to predict BBB permeability.[15-20].These and many other studies illustrated the importance of lipophilicity, topological indices, , hydrogen bonding potential and molecular volume parametersin blood brain barrier permeability [21-24], [25], [26], [27],[28][29].

However, these classical 3D QSAR approaches were alignment dependent, time consuming and produced user biased results depending upon the alignment used[6, 30-34]. More over superimposition of a structurally diverse dataset is nearly impossible. To counter this problem, alignment free approaches based on autocorrelation functions have been

suggested by Broto, Gasteiger et al and Clementi et al[35].Broto used classic autocorrelation transform to obtain autocorrelation vectors from 2D and 3D structures. Gasteiger used special autocorrelation function on molecular surface properties. Clementi also used autocorrelation vectors but only for planer compounds[35].All these approaches were quite effective in providing solution to alignment problems faced by classical QSAR models yet necessary transformations produce difficulty in interpretation of resultant models in original descriptors.

In present study, a 3D-QSAR model has been developed usingmolecular interaction field (MIF) based GRID independent (GRIND)descriptors that are independent of superimposition of the data [35]. Instead of absolute 3D coordinates as in classical 3D QSAR, GRIND measures the distances between relevant groups[35].Computation of the mutual distance between important 3D structural features such as Hydrogen bonding, hydrophobic features and shape based features has been identifiedas significant outcome of this study.

Methodology

Dataset Collection

A structurally diverse dataset of 552 compounds along with their experimentally determined log BB values has been collected from different publication, [44], [45], [15]. Utilizing software, 3D structures of each compound in the data set were created. Molecular Operating Environment version 2018-01[35] followed by computation of the partial charges and energy minimization using MMFF94 force field [46]. A complete data curation protocol is provided in figure 1. Briefly, at first step data was cleaned by eliminating redundancies and shards, next to the application of drug like filters as defined by Lipinski *et al*,[47, 48]. This results in a final dataset of 218 compounds as training set for building the GRIND model as shown in SM table 1. Interestingly, more than 53% of the training data include drugs already available in the market such as Bupropion, Sertraline,Maprotiline and Imipramine to deal with a variety of mental & psychological abnormalities[35]. Additionally, a separate dataset of 44 compounds as shown in SM table 2 has been taken from some different publication sources as test set [15, 47-49]

Computation of Physicochemical Parameters

So as to estimate impact of physical & chemical attributes on blood brain barrier permeability, 2D physicochemical descriptors including log P (o/w), molecular weight and topological polar surface area have been computed using software MOE version 2018-01 as provided in SM table 1.

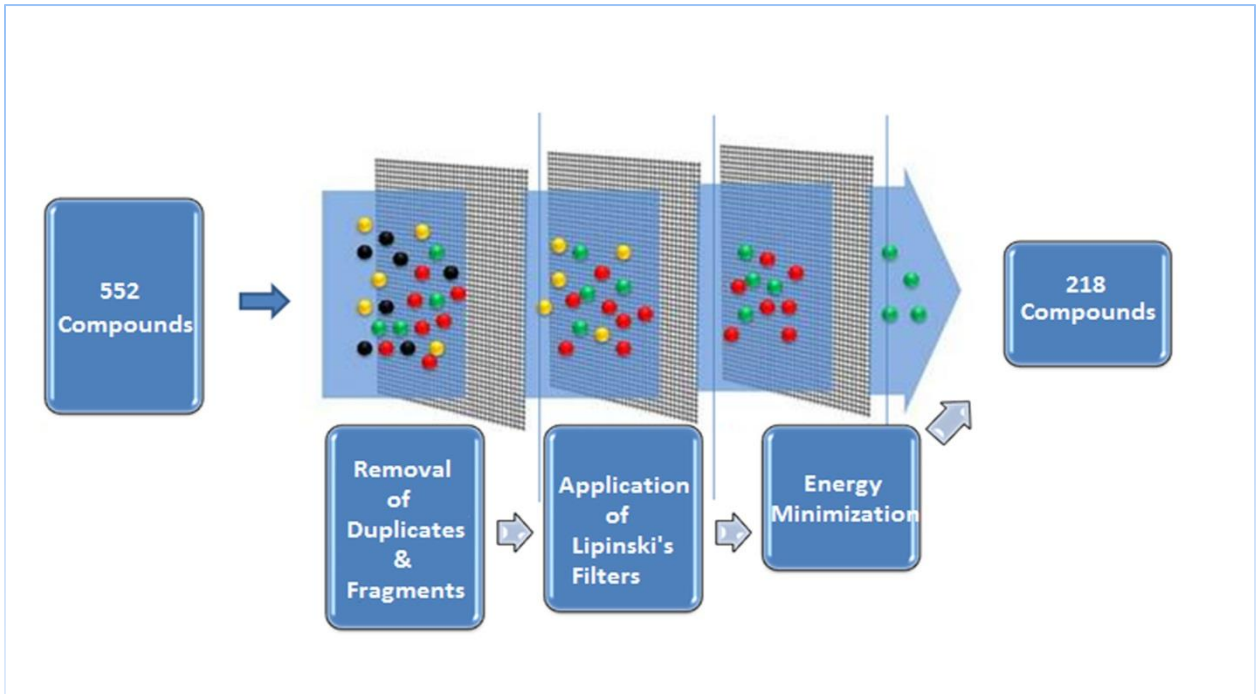

Figure1: Data pre-processing protocol for the refinement of structural and biological data (log BB) data of the Training and test set

3D QSAR Modeling

Applying CORINA CLASSIC, standard 3D variations of the training and test sets were produced. [50]. These extended 3D conformations along with log BB measurements of dataset molecules were imported to software package pentacle v1.05 to compute alignment-free molecular descriptors (GRID-Independent molecular descriptors)[35].

The molecular interaction fields (MIFs) provide information that the GRIND technique seeks to extract [35] and condense it into new categories of variables whose magnitudes are unaffected by the molecule under study's geographical location. Default program values were used for the computation of MIF using four probes (DRY representing hydrophobic interaction, O (Carbonyl Oxygen) representing hydrogen bond acceptor group, N1 (Amide Nitrogen) representing hydrogen bond donor groups and TIP representing the shape descriptor). Using an optimization technique (AMANDA) that employs the mutual node-node distances between the selected nodes and the intensity of the field at a node as a score function, the most pertinent regions were recovered from the MIF.[51]. The Lennard-Jones energy (Elj), hydrogen bond (Ehb), and electrostatic (Eal) interactions are added to determine the interaction energy (Exyz) at each site.

$$E_{TOTAL} = \sum E_{Lj} + \sum E_{es} + \sum E_{hb}$$

For the discretization of MIF, the following default values for the probe cutoff were used: DRY= -0.5, O= -2.6, N1= -4.2, and TIP= -0.74. Nodes whose energy value fell below this cutoff were eliminated. The pre-filtered nodes were then encoded into GRIND using the consistently large auto and cross-correlation (CLACC) approach. Correlogram plots, which show the products of node-node energies reported against distance separating the nodes, were used to directly display the values received from the analysis.

Through the use of GRIND variables, Principle Component Analysis (PCA) and Partial Least Square (PLS) Analysis have been carried out to comprehend the structural variance of the data and its link with logBB values. Briefly, Leave One Out (LOO) cross validation procedure[52] was utilized to correlate the observed verses predicted log BB values.

Results & Discussion

Estimation of Physicochemical Parameters

Principle Component Analysis (PCA)

Principle Component Analysis (PCA) of the GRIND variables revealed40 % structural variance of the training data with the help of first two principle components. Briefly, 1st principle component define the data on the basis of explicit distance ranges of a hydrogen bond donor reference feature from other 3D structural features including, a hydrophobic group, a molecular steric hotspot, another hydrogen bond donor and from a hydrogen bond acceptor group within the respective extended conformation of a molecule as depicted by DRY-O, O-TIP O-O and O-N1correlograms respectively in figure 2 A. However, 2ndprinciple component differentiate the training data by delineating the distance of a hydrogen bond acceptor reference feature from same 3D structural features as characterized by 1st PC. These are illustrated by DRY-N1, N1-TIP N1-N1and O-N1correlograms respectively as shown in figure 2 B. A summary, of the respective distance variables of different 3D structural features as defined by PC1 and PC2 are provided in table 2.

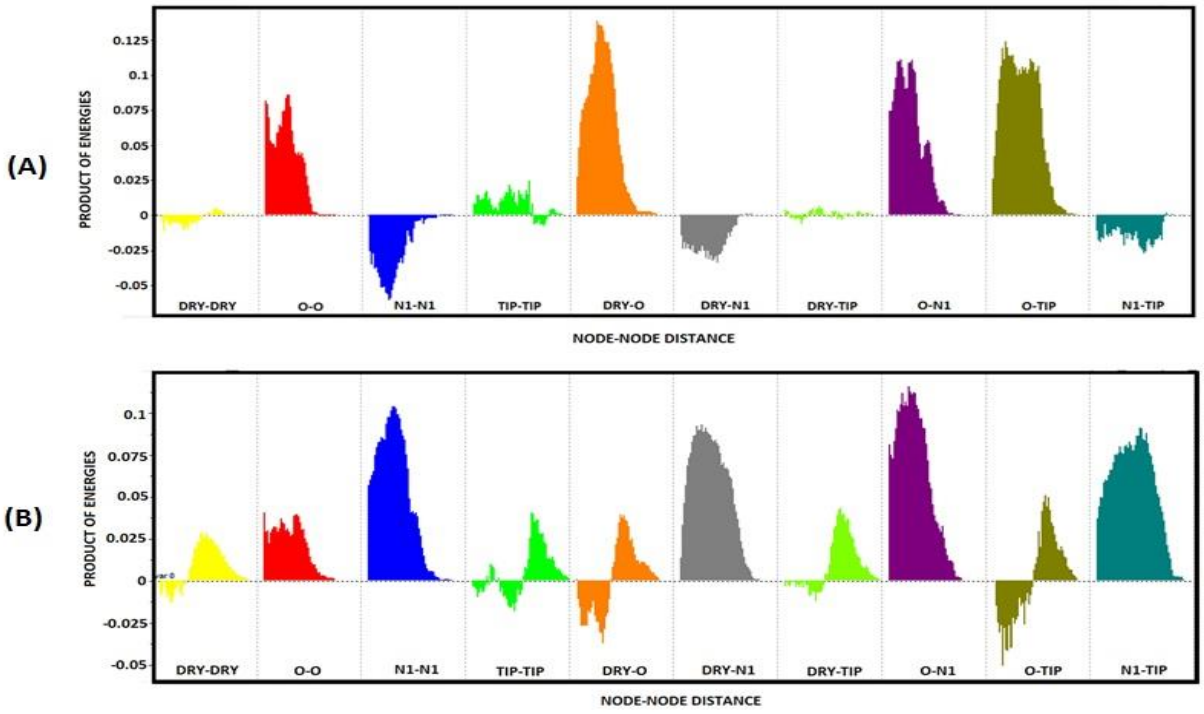

Figure2: (A) Represents GRIND variables of the 1st PC where, O-O, DRY-O, O-N1 and O-TIP correlograms are characterized by the more significant variables defining the data. (B) Represents GRIND variables of the 2nd PC where, DRY-N1, N1-TIP N1-N1 and O-N1correlograms are depicted as the important variables defining the data.

Table1: Representing distances in Angstrom (A) between important 3D structural features of the data depicted by first two principal components

PCs	Probes	Distances b/w GRIND Variables	Comments
PC1	O-O	6.80-7.20	Absent in cluster A & some compounds of cluster B
	DRY-O	6.40-6.80	Absent in only Cluster A
	O-N1	4.40-4.80	Absent in Cluster A & Cluster B
	O-TIP	4.80-5.20	Absent in only Cluster A
PC2	N1-N1	8.00-8.40	Absent in Cluster B & some compounds of Cluster A
	DRY-N1	6.80-7.20	Absent in only Cluster B
	O-N1	6.40-6.80	Absent in Cluster A & Cluster B
	N1-TIP	12.00-12.40	Absent in only Cluster B

A plot between 1st and 2nd principle component in figure 3shows that a H bond donor group is absent within respective chemical scaffolds of compound in cluster A and therefore, the feature distances ranges defined by 1st PC on the basis ofDRY-O, O-TIP O-O and O-N1 correlogram in figure 2A are absent in this cluster of the data set. Similarly, ahydrogen bond acceptor feature is absent in all compounds encircled as cluster B in figure 3 and therefore, a reference point for mapping the distances of characterized 3D structural features defined by DRY-N1, N1-TIP N1-N1and O-N1correlogram is absent in cluster B.Nevertheless, rest of the data exhibit one to two hydrogen bond acceptor as well as donor group within respective chemical scaffold thus, are defined by the presence of all O-O, DRY-O, O-TIP, DRY-N1, N1-TIP,

N1-N1and O-N1 variables depicted by first two PCs as shown in figure 2.

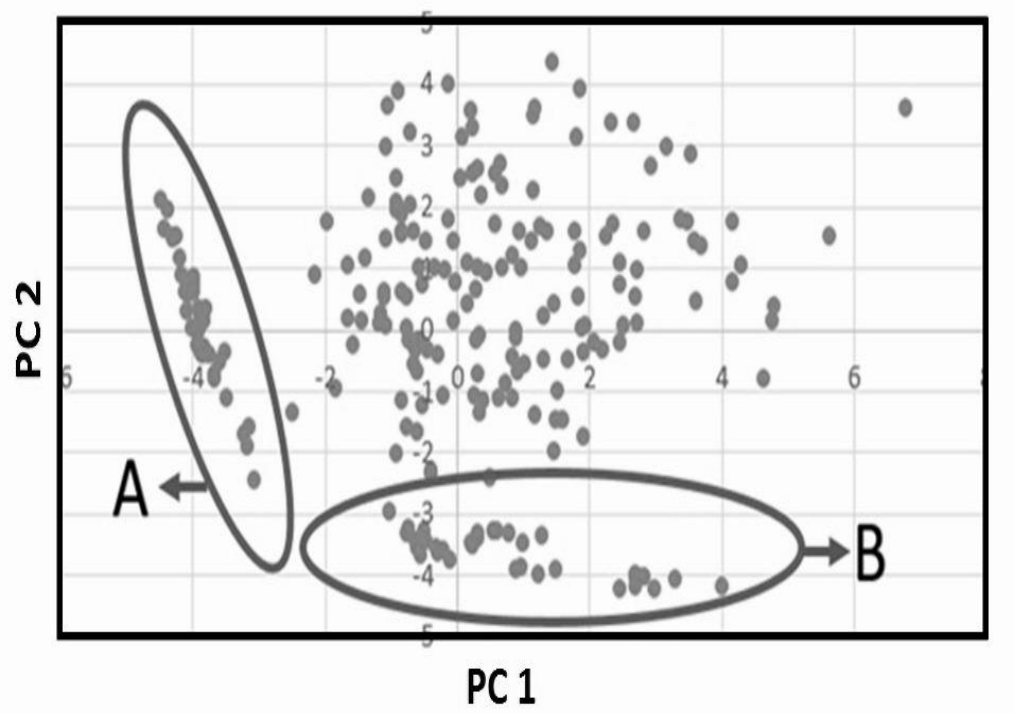

Figure 3: Graphical representation of Principal Component Analysis. Clusters are produced by plot between PC 1 and PC 2.

Furthermore, the impact of presence or absence of acceptor and donor groups for hydrogen bonds at specific distance from rest of the pharmacophoric features on blood brain barrier has been elucidated by Partial Least Square (PLS) Analysis.

Partial Least Square (PLS) Analysis

The full set of GRIND variables were subjected to partial least square (PLS) analysis utilizing Leave One Out (LOO) cross validation, which produced a model with unsatisfactory statistical values of q2= 0.46, R2= 0.59, and SDEP= 0.52. This is might be due to the presence of some inconsistent set of variables as defined by pastor *et al*[35]. Therefore, a variable selection algorithm known as FFD (Fractional factorial design) was applied to remove inconsistent variable [36]. An improvement in the statistical parameters of different GRIND models after subsequent 1st and 2nd FFD run has been observed as shown in table 2. A final GRIND model was obtained after the 2ndFFD cycle with q^2 = 0.50, r^2= 0.63 and standard error of prediction (SDEP) of 0.49 as shown in table 2.

Table 2 Blood brain barrier permeability model statistics after subsequent 1stand 2nd FFD cycles.

FFD Cycle	Variables #	Q^2	R^2	SDEP
0	Complete	0.46	0.59	0.52
1st	440	0.48	0.61	0.51
2nd	408	0.50	0.63	0.50

Figure 4 displays the plot of actual log BB values against predicted values obtained from multiple linear regression analysis utilizing leave one out (LOO) cross validation. As can be seen in SM table 1, nearly all compounds in the training set as well as the test set (44 compounds) are accurately predicted with an error of less than 1.5 log units between the actual and predicted logBB.

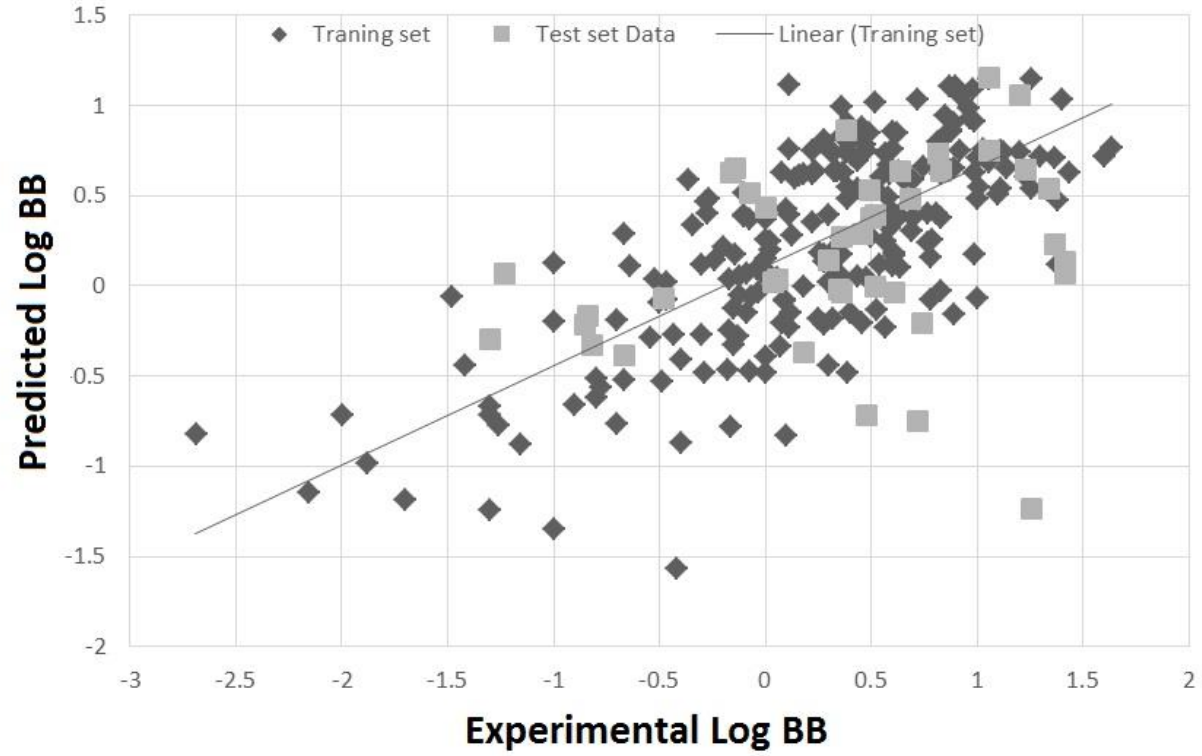

Figure 4: Plot of experimental versus predicted log BB values obtained from multiple linear regression model.

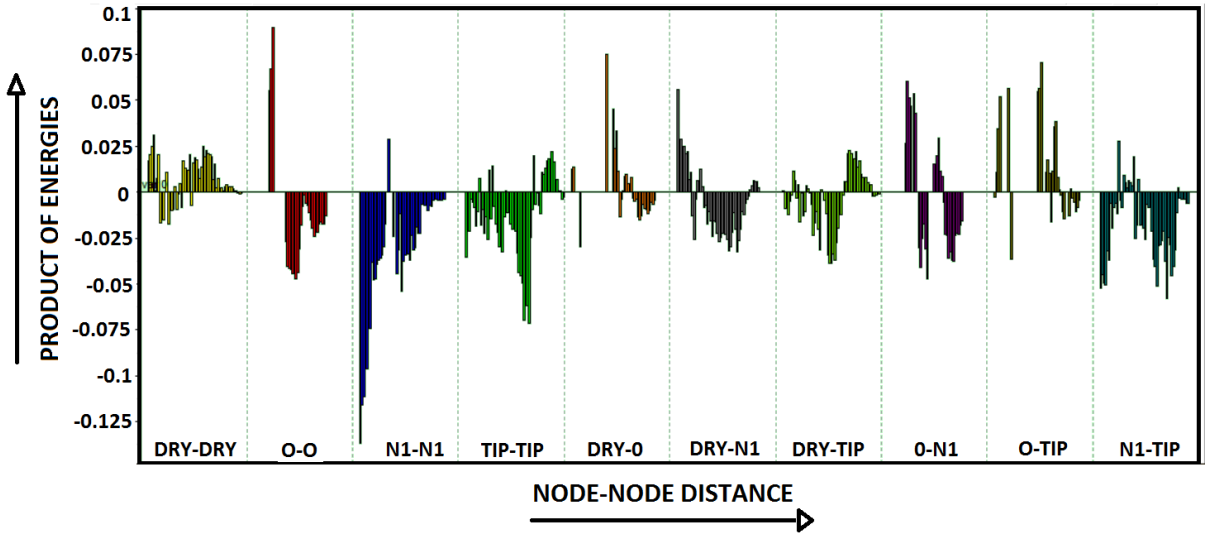

Figure 5: PLS coefficient correlograms plot representing the influence of 3D structural features on Log BB.

Figure 5 represent a PLS coefficient correlograms plot where positive and negative auto and cross correlogram peaks represent the 3D structural features having direct and inverse correlation with blood brain barrier permeability. Highly negative variable values of N1-N1, TIP-TIP, and N1-TIP correlogramsin figure 5 represent the 3D features that are present in non-permeable compounds. However, highly positive O-O, DRY-O, O-N1 and O-TIP correlogramsvariables depict the 3D structural features at certain mutual distance having positive impact on blood brain barrier permeability.

Figure 5 represent a PLS coefficient correlograms plot where positive and negative auto and cross correlogram peaks represent the 3D structural features having direct and inverse correlation with blood brain barrier permeability. Highly negative variable values of N1-N1, TIP-TIP, and N1-TIP correlogramsin figure 5 represent the 3D features that are present in non-permeable compounds. However, highly positive O-O, DRY-O, O-N1 and O-TIP correlograms variables depict the 3D structural features at certain mutual distance having positive impact on blood brain barrier permeability.

Briefly, the most positive O-O correlogram depicts the presence of two hydrogen bond donor groups (Don1 and Don2) at a mutual distance of 6.00 - 6.40 Å in highly permeable compounds having logBB values range 0.32 to 1.40 as shown in figure 6. Within our training data, two hydrogen bond donor groups at a mutual distance of 6.00 - 6.40 Å have been observed in extended 3D conformations of various antipsychotic agents including Nebivolol, Bupropion, Bromperidol, Zanapezil, Tramadol, Oxazepam, Lubeluzole, Biperiden and Donepezil exhibiting log BB of 0.48 to 1.44. However, a decrease in blood brain barrier permeability (log BB) has been observed as the distance between two hydrogen bond donor group increases. For instance, the two features, are present at a longer distance range of 11.6-12.0Å in least permeable compounds (log BB: 0.30 to -2.69) including 9-hydroxy risperidone, SKF89124 (7-hydroxy ropinirole) and SKF 93319 as shown in figure 6 and thus, are absent CNS active agents in our training data. Interestingly, O-N1 correlogram variables complement the O-O correlogram and represent a hydrogen bond donor group (Don1) at a distance of 6.0 to 6.4 Å from a hydrogen bond acceptor group in highly permeable compounds having a log BB range of 0.39 to 1.64 Å.

Similarly, the DRY-O correlogram shown in the figure 5 illustrates the presence of hydrophobic group at a mutual distance of 10.0–10.4 Å from a hydrogen bond donor group (Don1) in only highly permeable compounds having log BB range between 0.30 to 1.6 including Sertraline, Desipramine, Tamozifin and Methadone. Thus, it highlights the influence of hydrophobic feature of the molecules on blood brain barrier permeability which is in accordance with previous study conducted by Gulyaeva*et al*, in which relative hydrophobicity and lipophilicity of drugs on log BB was measured by aqueous two-phase partitioning, octanol-buffer partitioning and HPLC[37]. Nevertheless, the most positive variable in O-TIP correlogram depicts the presence of a hydrogen bond donor group (Don1) at a mutual distance of 4.80 to 5.20 Å from molecular steric hotspots in highly permeable compounds exhibiting a log BB range of 0.28 to 1.60 Å.

As a whole, our PLS model revealed the significance of one hydrogen bond donor group, Don1, which is depicted in figure 6 and may serve as an anchor for determining the locations of additional pharmacophoric characteristics. Moreover, PC1 also separated our training data on the basis of presence/absence of a hydrogen bond donor reference feature. Therefore, for the compounds of cluster A, distance between 3D structural features defined by O-O, DRY-O, O-N1 and O-TIP correlogram could not be mapped mostly because of scarcity of a hydrogen bond donor reference point in their respective chemical scaffolds. Moreover, the influence of hydrogen bond donor group on BBB permeability of the drug like compounds have been demonstrated in previous studies by considering descriptors like number of hydrogen bond donors/acceptors present within molecules and polarity [12, 22, 38]. These studies associate higher blood brain barrier permeability with lower counts of hydrogen bond donors and acceptors within a molecule. Moreover, a study conducted by Prashant *et al*, correlates blood brain barrier permeability with the strength of hydrogen bond acceptor or donor groups within a molecule[39]. Additionally, van de Water beemd*et al* demonstrated the importance of shape of the molecule towards blood brain barrier permeability using molecular size, shape and hydrogen bonding descriptors [40]. These previous investigations are in line with our study and thus, strengthen our models. Additionally, our models utilizes > 50% of already marketed drugs including CNS active agents which may provide a better scaffold to correlate and understand the 3D structural features important for BBB permeability. Moreover, our models were able to define and map the distances between these 3D structural features.

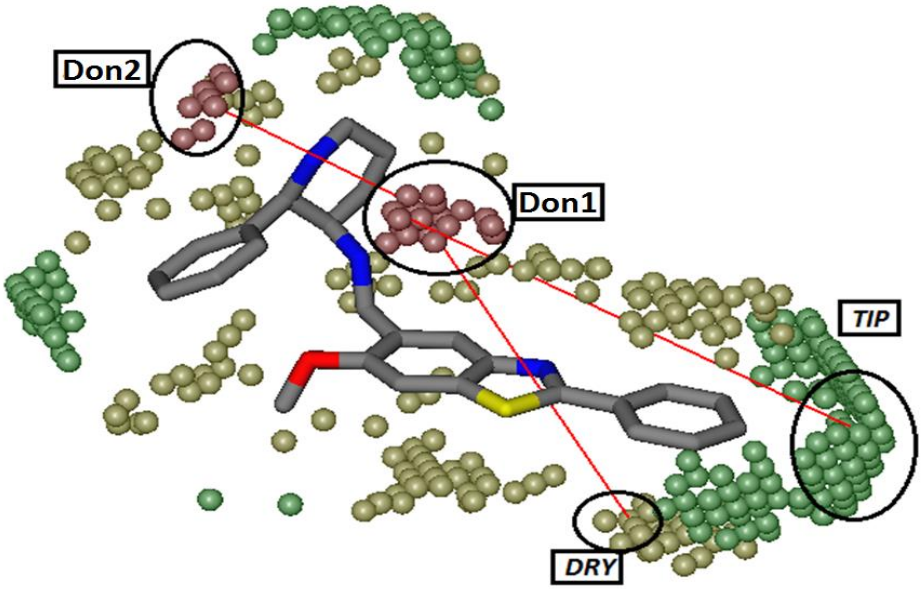

Figure 6: Represents a blood brain barrier permeable compound. In this figure blood red contour shows hydrogen bond donor hot-spots, Moss green contour represents hydrophobic hotspots and sea green contour represents molecular edges.

Table 3: Summary of GRIND variables, their mutual distance and impact on blood brain barrier permeability.

Correlogram	**Distance**	**Influence on log BB**
O-O	**6.00-6.40 Å**	+
DRY-O	**10.00-10.40 Å**	+
O-N1	**6.00-6.40 Å**	+
O-TIP	**4.80-5.20 Å**	+
N1-N1	**1.60-2.00 Å**	-

Likewise, N1-N1 correlogramvariables in figure 5 depictthe presence of two hydrogen bond acceptor groups(Acc1, Acc2) at a mutual distance of 1.60 to 2.0 Å in least permeable compounds (-2.6 to -0.18) of the training data as shown in figure 7. Similarly, the most negative N1-TIP correlogramvariable illustrates a distance of 17.6 to 18 Å between a hydrogen bond acceptor group (Acc1) and a molecular steric hotspot edge in non-permeable compounds of our dataset.

Highly negative TIP-TIP correlogramvariable depicts the presence of two molecular steric boundaries (TIP1 and TIP2) at a mutual distance of 15.6 to16.0 Å in non-permeable (log BB -0.54 to -2.6) compounds of the dataset as shown in figure 7. Thus, it may suggests that as the distance between the molecular edges increase and molecule gets more flexible then the permeability of the molecule decreases which has also been mentioned in a previous study by Pardridge*et al*[41]

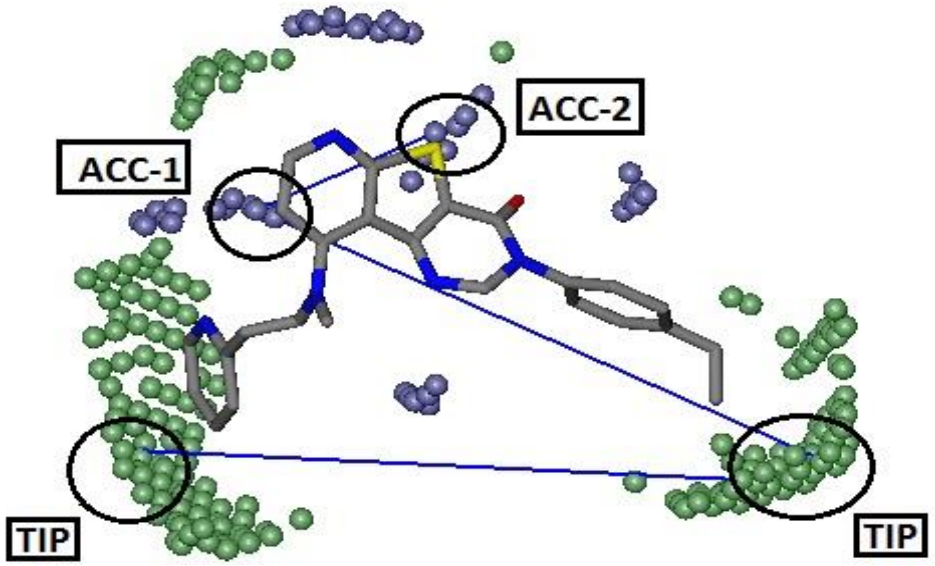

Figure 7: Represents a non-permeable compound in which blue contour represents hydrogen bond acceptor hotspots where as sea green contour refers to molecular edges.

Overall, GRIND model suggests that the permeability of the compounds has been influenced by the distance between four 3D structural features such as hydrogen bond donors, hydrogen bond acceptors, shape and hydrophobic features which is in agreement with previously developed QSAR models in which Lipophilicity and molecular volume have been identified as important properties along with hydrogen bonding potential to cross BBB [22, 24-27, 29, 42]. Several QSAR models correlates physiochemical descriptors such as lipophilicity, water-accessible volume, Molecular weight, potential to ionize, charge, topology polarized area, hydrogen bonds, rotator bonds and numerous others[43]. These descriptors reflects the presence of certain functional groups within the molecules[38]. However, in our study, the effect of 3D structural features has been found to be dependent on the distances present between these features within the extended conformations of the molecules. The presence of two hydrogen bond donor will show positive behavior towards BBB permeability only if they are present within a molecule at 6.00 - 6.40 Å distance from each other. However, hydrogen bond donor groups present at a mutual distance of 11.6-12.0 Å will show a negative effect on blood brain barrier permeability of compounds. Similarly, a molecule will exhibit permeability, if the hydrogen bond acceptor groups such as carbonyl oxygen is present at distance of 6.00 - 6.40 Å from hydrogen bond donor group with a molecule whereas, a negative effect will be observed for a mutual distance of 10.8-11.2 Å. The aromatic moieties present at distance of 10.0-10.4 Å from the donor groups within molecules will contribute positively towards BBB permeability of the molecule while a negative effect on BBB permeability of the molecules will be observed for a mutual distance of 20.0-20.4 Å. These facts signify the important role of relative distance between 3D structural features for BBB permeability of compounds. Moreover, a mutual distance of 4.80-5.20 Å between hydrogen bond donor and molecular edges will act as a positive contributor towards BBB permeability in our model. It is very interesting to note that protonated nitrogen in straight chain has been found as a common hydrogen bond donor in highly permeable compounds as shown in the figure 6.

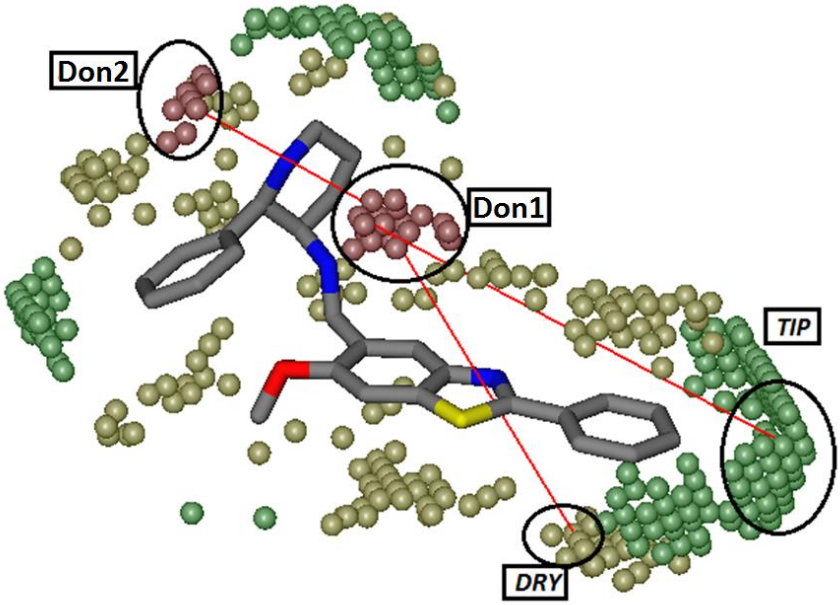

Figure 6: Represents a blood brain barrier permeable compound. In this figure blood red contour shows hydrogen bond donor hot-spots, Moss green contour represents hydrophobic hotspots and sea green contour represents molecular edges.

Table 3: Summary of GRIND variables, their mutual distance and impact on blood brain barrier permeability.

Correlogram	Distance	Influence on log BB
O-O	**6.00-6.40 Å**	+
DRY-O	**10.00-10.40 Å**	+
O-N1	**6.00-6.40 Å**	+
O-TIP	**4.80-5.20 Å**	+
N1-N1	**1.60-2.00 Å**	-

Likewise, N1-N1 correlogramvariables in figure 5 depictthe presence of two hydrogen bond acceptor groups(Acc1, Acc2) at a mutual distance of 1.60 to 2.0 Å in least permeable compounds (-2.6 to -0.18) of the training data as shown in figure 7. Similarly, the most negative N1-TIP correlogramvariable illustrates a distance of 17.6 to 18 Å between a hydrogen bond acceptor group (Acc1) and a molecular steric hotspot edge in non-permeable compounds of our dataset. Highly negative TIP-TIP correlogramvariable depicts the presence of two molecular steric boundaries (TIP1 and TIP2) at a mutual distance of 15.6 to16.0 Å in non-permeable (log BB -0.54 to -2.6) compounds of the dataset as shown in figure 7. Thus, it may suggests that as the distance between the molecular edges increase and molecule gets more flexible then the permeability of the molecule decreases which has also been mentioned in a previous study by Pardridge*et al*[41]

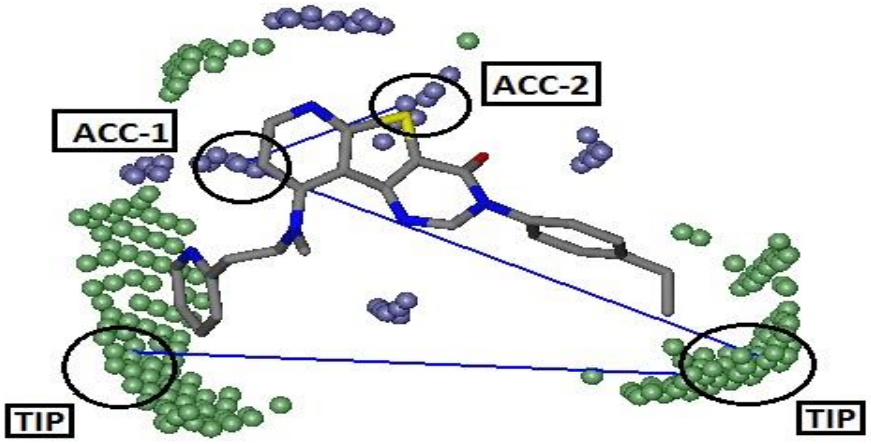

Figure 7: Represents a non-permeable compound in which blue contour represents hydrogen bond acceptor hotspots where as sea green contour refers to molecular edges.

Overall, GRIND model suggests that the permeability of the compounds has been influenced by the distance between four 3D structural features such as hydrogen bond donors, hydrogen bond acceptors, shape and hydrophobic features which is in agreement with previously developed QSAR models in which Lipophilicity and molecular volume have been identified as important properties along with hydrogen bonding potential to cross BBB [22, 24-27, 29, 42]. Several QSAR models correlates physiochemical descriptors such as lipophilicity, water-accessible volume, Molecular weight, potential to ionize, charge, topology polarized area, hydrogen bonds, rotator bonds and numerous others[43]. These descriptors reflects the presence of certain functional groups within the molecules[38]. However, in our study, the effect of 3D structural features has been found to be dependent on the distances present between these features within the extended conformations of the molecules. The presence of two hydrogen bond donor will show positive behavior towards BBB permeability only if they are present within a molecule at 6.00 - 6.40 Å distance from each other. However, hydrogen bond donor groups present at a mutual distance of 11.6-12.0 Å will show a negative effect on blood brain barrier permeability of compounds. Similarly, a molecule will exhibit permeability, if the hydrogen bond acceptor groups such as carbonyl oxygen is present at distance of 6.00 - 6.40 Å from hydrogen bond donor group with a molecule whereas, a negative effect will be observed for a mutual distance of 10.8-11.2 Å. The aromatic moieties present at distance of 10.0-10.4 Å from the donor groups within molecules will contribute positively towards BBB permeability of the molecule while a negative effect on BBB permeability of the molecules will be observed for a mutual distance of 20.0-20.4 Å. These facts signify the important role of relative distance between 3D structural features for BBB permeability of compounds. Moreover, a mutual distance of 4.80-5.20 Å between hydrogen bond donor and molecular edges will act as a positive contributor towards BBB permeability in our model. It is very interesting to note that protonated nitrogen in straight chain has been found as a common hydrogen bond donor in highly permeable compounds as shown in the figure 6.

Conclusion:

In the light of the predictions made by 3D QSAR model using GRID independent descriptors (GRIND) it has been demonstrated that the relative distance between important pharmacophoric features such as hydrogen bond donors/acceptors, hydrophobicity and shape based features may affect BBB permeability of a compound.Optimum distance of a hydrogen bond donor pharmacophore from the other features which include hydrophobic group, another donor group and molecular edges, determines the fate of molecule to cross the BBB. According to our model, two hydrogen bond donor groupsare present at a mutual distance of 6.00 to 6.40 Å in CNS active agents having high logBB values. One of the hydrogen bond donor group, at a distance of 10Å-10.4Å from a hydrophobic group and at a distance of 6.00 - 6.40 Åfrom a hydrogen bond acceptor functional group also represent important attributes for blood brain barrier permeability. Similarly the optimum distance between the reference hydrogen bond donor group and molecular edges should be in a range of 4.8Å- 5.2Å for BBB permeability as suggested by our GRIND model. Overall, our GRIND model not only highlights the important 3D structural features of a molecule for

BBB penetration but also maps the relative distance between the functional groups optimum for BBB permeability. Since the descriptors used to build this model, are highly relevant to biological activity, alignment free and easy to obtain therefore, this model can prove to be useful for prediction of blood brain barrier permeability of new chemical entities in early stages of drug development.

References

Abbott, N.J., *Blood–brain barrier structure and function and the challenges for CNS drug delivery.* Journal of inherited metabolic disease, 2013. **36**(3): p. 437-449.

Ballabh, P., A. Braun, and M. Nedergaard, *The blood–brain barrier: an overview: structure, regulation, and clinical implications.* Neurobiology of disease, 2004. **16**(1): p. 1-13.

Gloor, S.M., et al., *Molecular and cellular permeability control at the blood–brain barrier.* Brain research reviews, 2001. **36**(2): p. 258-264.

Francisco, P.-P., E.-C. Manuel, and G.-M. Xerardo, *Review of Bioinformatics and QSAR Studies of β-Secretase Inhibitors.* Current Bioinformatics, 2011. **6**(1): p. 3-15.

Pangalos, M.N., L.E. Schechter, and O. Hurko, *Drug development for CNS disorders: strategies for balancing risk and reducing attrition.* Nature Reviews Drug Discovery, 2007. **6**: p. 521.

Carpenter, Timothy S., et al., *A Method to Predict Blood-Brain Barrier Permeability of Drug-Like Compounds Using Molecular Dynamics Simulations.* Biophysical Journal, 2014. **107**(3): p. 630-641.

Masungi, C., et al., *Parallel artificial membrane permeability assay (PAMPA) combined with a 10-day multiscreen Caco-2 cell culture as a tool for assessing new drug candidates.* Die Pharmazie-An International Journal of Pharmaceutical Sciences, 2008. **63**(3): p. 194-199.

Mensch, J., et al., *Evaluation of various PAMPA models to identify the most discriminating method for the prediction of BBB permeability.* European Journal of Pharmaceutics and Biopharmaceutics, 2010. **74**(3): p. 495-502.

Tsinman, O., et al., *Physicochemical selectivity of the BBB microenvironment governing passive diffusion—matching with a porcine brain lipid extract artificial membrane permeability model.* Pharmaceutical research, 2011. **28**(2): p. 337-363.

10. Campbell, S.D., K.J. Regina, and E.D. Kharasch, *Significance of lipid composition in a blood-brain barrier–mimetic pampa assay.* Journal of biomolecular screening, 2014. **19**(3): p. 437-444.

Könczöl, A.r.d., et al., *Applicability of a blood–brain barrier specific artificial membrane permeability assay at the early stage of natural product-based CNS drug discovery.* Journal of natural products, 2013. **76**(4): p. 655-663.

Kansy, M., F. Senner, and K. Gubernator, *Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes.* J Med Chem, 1998. **41**(7): p. 1007-10.

Di, L., et al., *High throughput artificial membrane permeability assay for blood–brain barrier.* European journal of medicinal chemistry, 2003. **38**(3): p. 223-232.

Yang, C.Y., et al., *Immobilized artificial membranes—screens for drug membrane interactions.* Advanced Drug Delivery Reviews, 1997. **23**(1): p. 229-256.

Liu, X., et al., *Development of a computational approach to predict blood-brain barrier permeability.* Drug metabolism and disposition, 2004. **32**(1): p. 132-139.

Bujak, R., et al., *Blood–brain barrier permeability mechanisms in view of quantitative structure–activity relationships (QSAR).* Journal of pharmaceutical and biomedical analysis, 2015. **108**: p. 29-37.

Vilar, S., M. Chakrabarti, and S. Costanzi, *Prediction of passive blood–brain partitioning: straightforward and effective classification models based on in silico derived physicochemical descriptors.* Journal of Molecular Graphics and Modelling, 2010. **28**(8): p. 899-903.

Katritzky, A.R., et al., *Correlation of blood–brain penetration using structural descriptors.* Bioorganic & medicinal chemistry, 2006. **14**(14): p. 4888-4917.

Wu, Z.-Y., et al., *Comparison of prediction models for blood brain barrier permeability and analysis of the molecular descriptors.* Die Pharmazie-An International Journal of Pharmaceutical Sciences, 2012. **67**(7): p. 628-634.

Young, R.C., et al., *Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists.* Journal of medicinal chemistry, 1988. **31**(3): p. 656-671.

Abraham, M.H., H.S. Chadha, and R.C. Mitchell, *Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain.* Journal of pharmaceutical sciences, 1994. **83**(9): p. 1257-1268.

Abraham, M., H. Chadha, and R. Mitchell, *Hydrogen-bonding. Part 36. Determination of blood brain distribution using octanol-water partition coefficients.* Drug design and discovery, 1995. **13**(2): p. 123-131.

Norinder, U., P. Sjöberg, and T. Österberg, *Theoretical calculation and prediction of brain–blood partitioning of organic solutes using MolSurf parametrization and PLS statistics.* Journal of pharmaceutical sciences, 1998. **87**(8): p. 952-959.

Brewster, M.E., et al., *AM1-based model system for estimation of brain/blood concentration ratios.* International journal of quantum chemistry, 1996. **60**(8): p. 1775-1787.
Kelder, J., et al., *Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs.* Pharmaceutical research, 1999. **16**(10): p. 1514-1519.
Clark, D.E., *Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood–brain barrier penetration.* Journal of pharmaceutical sciences, 1999. **88**(8): p. 815-821.
Subramanian, G. and D.B. Kitchen, *Computational models to predict blood–brain barrier permeation and CNS activity.* Journal of computer-aided molecular design, 2003. **17**(10): p. 643-664.
Luco, J.M., *Prediction of the brain− blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modeling.* Journal of chemical information and computer sciences, 1999. **39**(2): p. 396-404.
Feher, M., E. Sourial, and J.M. Schmidt, *A simple model for the prediction of blood–brain partitioning.* International journal of pharmaceutics, 2000. **201**(2): p. 239-247.
Kubinyi, H., *3D QSAR in drug design: volume 1: theory methods and applications*. Vol. 1. 1993: Springer Science & Business Media.
Kubinyi, H., *QSAR and 3D QSAR in drug design Part 1: methodology.* Drug discovery today, 1997. **2**(11): p. 457-467.
Verma, J., V.M. Khedkar, and E.C. Coutinho, *3D-QSAR in drug design-a review.* Current topics in medicinal chemistry, 2010. **10**(1): p. 95-115.
Jagiello, K., et al., *Advantages and limitations of classic and 3D QSAR approaches in nano-QSAR studies based on biological activity of fullerene derivatives.* Journal of Nanoparticle Research, 2016. **18**(9): p. 256.
34. F Morales, J., et al., *Current State and Future Perspectives in QSAR Models to Predict Blood-Brain Barrier Penetration in Central Nervous System Drug R&D.* Mini reviews in medicinal chemistry, 2017. **17**(3): p. 247-257.
35. Pastor, M., et al., *GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors.* Journal of medicinal chemistry, 2000. **43**(17): p. 3233-3243.
Gunst, R.F. and R.L. Mason, *Fractional factorial design.* Wiley Interdisciplinary Reviews: Computational Statistics, 2009. **1**(2): p. 234-244.
Gulyaeva, N., et al., *Relative hydrophobicity and lipophilicity of drugs measured by aqueous two-phase partitioning, octanol-buffer partitioning and HPLC. A simple model for predicting blood–brain distribution.* European journal of medicinal chemistry, 2003. **38**(4): p. 391-396.
Geldenhuys, W.J., et al., *Molecular determinants of blood-brain barrier permeation.* Ther Deliv, 2015. **6**(8): p. 961-71.
Desai, P.V., T.J. Raub, and M.-J. Blanco, *How hydrogen bonds impact P-glycoprotein transport and permeability.* Bioorganic & medicinal chemistry letters, 2012. **22**(21): p. 6540-6548.
van de Waterbeemd, H., et al., *Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors.* Journal of drug targeting, 1998. **6**(2): p. 151-165.
Pardridge, W.M., *The Blood-Brain Barrier: Bottleneck in Brain Drug Development.* NeuroRX, 2005. **2**(1): p. 3-14.
Suenderhauf, C., F. Hammann, and J. Huwyler, *Computational prediction of blood-brain barrier permeability using decision tree induction.* Molecules, 2012. **17**(9): p. 10429-10445.
Gao, Z., et al., *Predict drug permeability to blood-brain-barrier from clinical phenotypes: drug side effects and drug indications.* Bioinformatics, 2017. **33**(6): p. 901-908.
Ooms, F., et al., *A simple model to predict blood–brain barrier permeation from 3D molecular fields.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2002. **1587**(2): p. 118-125.
Garg, P. and J. Verma, *In Silico Prediction of Blood Brain Barrier Permeability: An Artificial Neural Network Model.* Journal of Chemical Information and Modeling, 2006. **46**(1): p. 289-297.
Sadowski, J., M. Wagener, and J. Gasteiger, *Assessing similarity and diversity of combinatorial libraries by spatial autocorrelation functions and neural networks.* Angewandte Chemie International Edition, 1996. **34**(23-24): p. 2674-2677.
Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1.* Advanced drug delivery reviews, 2001. **46**(1-3): p. 3-26.
Lipinski, C.A., *Lead-and drug-like compounds: the rule-of-five revolution.* Drug Discovery Today: Technologies, 2004. **1**(4): p. 337-341.
Zhang, L., et al., *QSAR modeling of the blood–brain barrier permeability for diverse organic compounds.* Pharmaceutical research, 2008. **25**(8): p. 1902.
Sadowski, J., *3D structure generation.* Handbook of Chemoinformatics: From Data to Knowledge in 4 Volumes, 2003: p. 231-261.
Durán, Á., G.C. Martínez, and M. Pastor, *Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in molecular interaction fields.* Journal of chemical information and modeling, 2008. **48**(9): p. 1813-1823.

Cawley, G.C. and N.L. Talbot, *Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers.* Pattern Recognition, 2003. **36**(11): p. 2585-2592.