Tuning properties for blood-brain barrier permeation: a statistics-based analysis

ACS Chemical Neuroscience

Article

Tuning Properties for Blood–Brain Barrier (BBB) Permeation: A Statistics-Based Analysis

Maria Dichiara, Benedetto Amata, Rita Turnaturi, Agostino Marrazzo, and Emanuele Amata

ACS Chem. Neurosci., **Just Accepted Manuscript** • DOI: 10.1021/acschemneuro.9b00541 • Publication Date (Web): 03 Dec 2019

Tuning Properties for Blood–Brain Barrier (BBB) Permeation: A Statistics-Based Analysis

Maria Dichiara, Benedetto Amata, Rita Turnaturi, Agostino Marrazzo, Emanuele Amata*

Department of Drug Sciences, Medicinal Chemistry Section, Università degli Studi di Catania, Viale Andrea Doria 6, 95125 Catania, Italy

Abstract

In the effort to define a set of rules useful in tuning the properties for a successful BBB permeation, we statistically analyzed a set of 338 compounds and correlated their experimental *in vivo* logBB with a series of computed descriptors. Contingency tables were constructed, observed and expected distributions calculated, and chi-square distributions evaluated. This allowed to point out a significant dependence of certain physicochemical properties in influencing the BBB permeation. Over 15 computed descriptors, nine resulted to be particularly important showing highly significant chi-square distribution: polar surface area (66.79; p=1.08 x 10^{-13}), nitrogen and oxygen count (51.17; p=2.06 x 10^{-10}), logP (47.38; p=1.27 x 10^{-9}), nitrogen count (38.29; p=9.77 x 10^{-8}), logD (36.80; p=36.80), oxygen count (35.83; p=3.13 x 10^{-7}), ionization state (33.02, p=3.19 x 10^{-7}), hydrogen bond acceptors (30.80; p=3.36 x 10^{-6}), and hydrogen bond donors (29.29; p=6.81 x 10^{-6}). Other parameters describing the mass and size of the molecules (molecular weight: 11.18; p=2.46 x 10^{-2}) resulted not significant since the population within the observed and expected distribution was similar. Depending on the combination of the significative descriptors, we set a three cases probabilistic scenario (BBB^+, BBB^-, BBB^+/BBB^-) that would prospectively be used to tune properties for BBB permeation.

Keywords: BBB permeability, logBB, physicochemical properties, contingency table, chi-square statistic, CNS drug discovery.

Introduction

According to the American Neurological Association (ANA), hundreds of million Americans suffer from at least one neurological disease, with headache and epilepsy at the top of the list, followed by sleep disorders, strokes and dementia. Many of these disorders have an incidence that increases significantly over the age of 65 years with a global annual cost beyond 800 billion dollars.[1] Furthermore, the brain is a common site of metastases in 10−30% of adults with cancer.[2]

The brain is the most complex organ in the human body equipped with a sophisticated protection system needed for maintaining the homeostasis and physiological environment. A key player in this finely regulated equilibrium is the blood-brain barrier (BBB), a restricted gateway allowing specific nutrients and hormones to selectively permeate. This protection system hinders the development of CNS active molecules and the standard pharmacokinetic parameters such as oral bioavailability and plasma concentration are not sufficient to evaluate drug exposure and time course in the brain.[3] A fine interplay between passive membrane permeability and active transport processes is also determinant for brain entry.[4] Certain drugs freely diffuse through BBB, nevertheless the presence of some efflux transport systems on the brain microvasculature, such as the P-glycoprotein (P-gp) and the multidrug resistance-associated protein family (MRP), leads to their immediate extrusion from brain.[5, 6] Furthermore, the lack of a completely satisfying *in vitro* model of BBB strongly prevents drug discovery and development. Experiments with brain capillary or immortalized endothelial cells are impaired from sufficiently restrictive tight junctions, whereas permeability screening assays using non-cerebral cell lines, such as the intestine Caco-2 and kidney MDCK cell lines may result qualitatively poor when compared with BBB endothelial cells.[7] Non-cell-based methods are also available and these include parallel artificial permeability assay (PAMPA),[8] and separation methods like high-performance liquid chromatography, and capillary electrophoresis.[9, 10] Nevertheless, PAMPA lacks in reproducing active influx and/or efflux transport mechanisms and some metabolic transformations, while separation methods, relying on the adjusted relationship between their output results and BBB permeation, may well be considered as predictive methods.[11]

An index of BBB permeability is represented by the logBB ($\log K_{p,brain}$), the logarithm of the ratio of the concentration of a drug in the brain and in the blood measured at the steady state:

$$logBB = log \frac{C_{brain}}{C_{blood}}$$

Lately, there has been a growing acceptance that just the unbound portion of the total drug ($K_{p,uu,brain}$), is free to diffuse across biological barriers and tissues, thus governing, more significantly than logBB, the overall BBB permeation and CNS drug activity.[12] However, at the state of the art the output results of both methods are employed and beneficial for understanding the magnitude of BBB permeation and to some extent the CNS active dose.[13-20]

Over the years, a number of researchers has developed methods to define the properties for a successful BBB permeation, using different starting data sets that inevitably had led to different conclusions.[21] Selected studies have considered population distribution, or variation in minimum, maximum and mean of certain parameters of CNS and non-CNS compounds. Other studies are based on more complex algorithms or quantitative structure-property relationship analysis.[22] In 2012, Ghose *et al.* have reported the plain distribution of a number of physicochemical properties and chemical structural profiles of 626 non-CNS and 317 CNS oral marketed drugs.[23] Using a restricted set of six physicochemical parameters, Wager *et al.* realized an algorithm to optimize the druglike properties of CNS candidates.[24, 25] Based on a structure-property analysis derived from multiple ADMET assays, Gleeson has identified molecular weight and logP as important parameters in determining CNS penetration.[26] In a study conducted by Mahar Doan *et al.*, the *in vitro* permeability and P-gp mediated efflux have been evaluated for 93 structurally diverse marketed CNS and non-CNS drugs.[27] Similarly, Desai *et al.* developed an in silico tool based on a P-gp substrate assay on over 2000 compounds.[28] All these are excellent examples of studies conducted with the aim to better understand and classify the properties required for a drug to better accumulate in the CNS, always providing new understanding and hints. In the present work, we were interested in defining a simple set of rules that would prospectively be used for the design or identification of compounds able or not to cross the BBB. In our analysis, a number of marketed drugs and organic compounds of general interest having experimental *in vivo* BBB permeation data, expressed as logBB, value were included. For each compound a series of computed descriptors has been calculated in order to observe which of them significatively influences the BBB permeation. For this purpose, 2×5 contingency tables were constructed for each descriptor, considering five descriptor domains or range of values in respect to two populations of compounds, BBB permeable (BBB^+) and BBB not-permeable (BBB^-). Observed and expected distributions were calculated, and percent deviations measured. Conditional

probabilities, fold enrichments, and a chi-square contingency table test with Yates correction were performed to evaluate the statistical significance of the chi-squared distribution.[29] The output of these contingency tables together with the chi-square statistic provided evidences of a significant association or dependence of certain physicochemical properties evaluated, while excluded other as the molecular weight (MW). In other words, BBB^+ and BBB^- compounds show variation in the population distribution for some of the descriptors and there are significative differences between expected and observed populations.

Results and Discussion

For the construction of the dataset we combined different scholarly journal published datasets. The literature was scanned and articles selected weather or not they contained information over chemical compounds ability in crossing the BBB expressed as logBB. The first collection consisted of 415 compounds from Li,[30] 380 compounds from Muehlbacher,[31] 57 compounds from Liu,[32] and 207 compounds from Abraham[33] for a total of 1059 compounds. However, several compounds were reported multiple times, even within the same article, with similar or identical logBB values. Duplicates were removed and a match between reported SMILES, nomenclature and depicted structures was performed and erroneous information corrected. Argon, krypton, neon, thoron, and xenon, have been removed from the dataset together with nitrous oxide, carbon disulfide, nitrogen, water, methane, and sulfur hexafluoride.[31] Vinblastine, Cyclosporin A, Cefotetan, and Bromocriptine were removed as well since their descriptors were widely outside range. Fifteen isomers have also been identified and one of the couple removed. Geometrical isomers, and tautomers (*e.g.* clonidine, didanosine) were also checked and one removed from the set. For the latest cases in the event that compounds had different logBB an average has been considered. For compounds whose logBB was taken multiple times, the one maintained in the dataset was the most recently published.

After this first round, we ended up with 559 unique compounds. From this set, we wanted to select only those compounds for which experimental *in vivo* logBB value has been reported.

For the determination of *in vivo* experimental logBB, mice are administered with a subcutaneous dose of drug and, after the reach of equilibrium, the drug concentration is measured in the brain and in the blood using a chromatographic method. In some cases, radiolabeled chemical compounds are administered and the radioactivity in the brain or blood is measured with a liquid scintillation counter. Therefore, from this dataset,

we did not consider those molecules for which logBB values have not been given, determined with *in vitro* procedures or for which the predicted value was reported.[30, 33, 34]

These efforts led to a dataset of 342 compounds. From this set we also wanted to exclude known substrates of P-gp, one of the major transport systems acting at the BBB. Indeed, P-gp substrates, generally lipophilic xenobiotics, can cross cell membranes but are immediately pumped back into the cerebral blood by P-gp. Therefore, searching for substrates of P-gp within our dataset, we decided to exclude known substrates of P-gp obtaining a final dataset of 328 compounds.[35-43]

These 328 compounds present experimental logBB values, calculated with the chromatographic or scintillation methods, ranging from –2.15 to +1.64 (logBB mean +0.04), with 182 logBB values positive or equal to zero and 146 negative (Table S1). For each compound a series of descriptors was computed with JChem for excel provided by ChemAxon (17.4.300.1589). A total of 15 descriptors has been computed, including logP, logD[pH7.4] (logD), polar surface area (PSA), atom count (AC), heavy atom count (HAC), molecular weight (MW), rotatable bond count (RBC), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), ring count (RC), ionization state (IS), nitrogen count (NC), oxygen count (OC), nitrogen and oxygen count (NOC), and molar refractivity (MR). The dataset was split into BBB permeable (BBB^+, logBB >0.0, BB ratio >1.0) and BBB non-permeable (BBB^-, logBB <0.0, BB ratio <1.0). This let to separate compounds based on their higher accumulation inside or outside the CNS. Compounds tagged as BBB^+ were 174, compounds with BB ratio equal to 1, thus excluded, were 8, while compounds with BBB^- features were 146 for a total of 328 compounds. We decided to bin descriptors in five groups which gave a large enough sample to provide a statistically valid analysis as reported in Table 1. The IS was split in four groups that correspond to apolar neutral, polar neutral, acid and basic compounds. Two zwitterions have been found in the set and these have been included in the polar neutral group since as unique group was too small. Maximum, minimum and mean of the values for the selected descriptors are reported in Table S2.

Table 1. Descriptor ranges

Descriptor Range				
logP<0	0≤logP<1.5	1.5≤logP<3.0	3.0≤logP<4.5	logP≥4.5
logD<0	0≤logD<1.5	1.5≤logD<3.0	3.0≤logD<4.5	logD≥4.5
0<PSA<25	25≤PSA<50	50≤PSA<75	75≤PSA<100	PSA≥100
AC<15	15≤AC<25	25≤AC<35	35≤AC<45	AC≥45

HAC<7	7≤HAC<14	14≤HAC<21	21≤HAC<28	HAC≥28
MW<100	100≤MW<200	200≤MW<300	300≤MW<400	MW≥400
RBC<1	1≤RBC≤2	3≤RBC≤4	5≤RBC≤6	RBC≥7
HBA=0	HBA=1	HBA=2	HBA=3	HBA≥4
HBD=0	HBD=1	HBD=2	HBD=3	HBD≥4
RC=0	RC=1	RC=2	RC=3	RC≥4
IS=Neutral Apolar	IS=Neutral Polar	IS=Acid	IS=Basic	
NC=0	NC=1	NC=2	NC=3	NC≥4
OC=0	OC=1	OC=2	OC=3	OC≥4
NOC<1	NOC=2	NOC=3	NOC=4	NOC≥5
MR<30	30≤MR<60	60≤MR<90	90≤MR<120	MR≥120

At this stage, for each descriptor a 2 × 5 contingency table with four degree of freedom was constructed, observed and expected distribution in the ranges were calculated and percent deviations measured. Conditional probabilities and fold enrichments were thus calculated, and a chi-square contingency table test with Yates correction was performed to evaluate the statistical significance of the chi-squared distribution. The enrichment of a contingency table was considered significant when its chi-square test *p* value was <0.01.

Descriptors scored by enrichment and their respective *p* values are reported in Table 2. Over 15 descriptors, six resulted to be not significant that means their distribution within the five descriptor ranges and the two BBB groups (BBB^+ and BBB^-) does not show significant variation. In other words, compounds having BBB^+ and BBB^- features share similar distribution for these six descriptors and there are not significative differences between expected and observed populations.

Table 2. Descriptors scored by enrichment and statistical significance

Descriptor	Chi-Squared Test	*p* value	*p*
PSA	66.79	1.08×10^{-13}	<0.01
NOC	51.17	2.06×10^{-10}	<0.01
logP	47.38	1.27×10^{-9}	<0.01
NC	38.29	9.77×10^{-8}	<0.01
logD	36.80	1.98×10^{-7}	<0.01
OC	35.83	3.13×10^{-7}	<0.01
IS	33.02	3.19×10^{-7}	<0.01
HBA	30.80	3.36×10^{-6}	<0.01
HBD	29.29	6.81×10^{-6}	<0.01
MW	11.18	2.46×10^{-2}	-
RC	6.61	1.58×10^{-1}	-
MR	6.01	1.99×10^{-1}	-
HAC	5.68	2.24×10^{-1}	-
AC	5.48	2.42×10^{-1}	-

RBC	4.05	4.00×10^{-1} -

AC (Chi-Squared Test 5.48; p=2.42 x 10^{-1}) and HAC (Chi-Squared Test 5.68, p=2.24 x 10^{-1}) are descriptors whose chi-square distribution between the BBB^+ and BBB^- groups has not given significative variation. Within the five ranges evaluated, the observed distribution is similar to the expected distribution and there are no quadrants with enriched distribution (Figure S1–2). Other not significative descriptors are RC (Chi-Squared Test 6.61; p=1.58 x 10^{-1}), RBC (Chi-Squared Test 4.05; p=4.00 x 10^{-1}) and MR (Chi-Squared Test 6.01; p=1.99 x 10^{-1}) where a random distribution and a poor enrichment within the quadrants have been observed (Figure S3–5). The whole interpretation is that these descriptors do not significantly affect the BBB permeability by themselves, and it is not necessary to evaluate their influence in defining whether or not a compound permeate the BBB.

MW (Chi-Squared Test 11.18; p=2.46 x 10^{-2}) has been found to be a not essential parameter by itself for having a better BBB permeation (Figure 1). Indeed, MW falls in the not-significative group of descriptors, that means the population within the ten quadrants does not provide sufficient variation for correlating MW with a BBB^+/BBB^- variation. MW provides a low variation of its fold enrichments that resulted to be not significative according to chi-squared test. Thus, MW by itself is not a descriptor that could help in designing or defining molecules that better accumulate in the CNS. Overall, the four descriptors here used for describing the mass and size of the molecules that means AC, HAC, MW, and MR, work similarly having a not-significative distribution as measured in the four contingency tables and in the performed chi-square tests. This means that there are no enrichments within the ten considered quadrants and that observed and expected population distribution is similar. For all the non-significative descriptors other considered ranges have given similar results (data not shown).

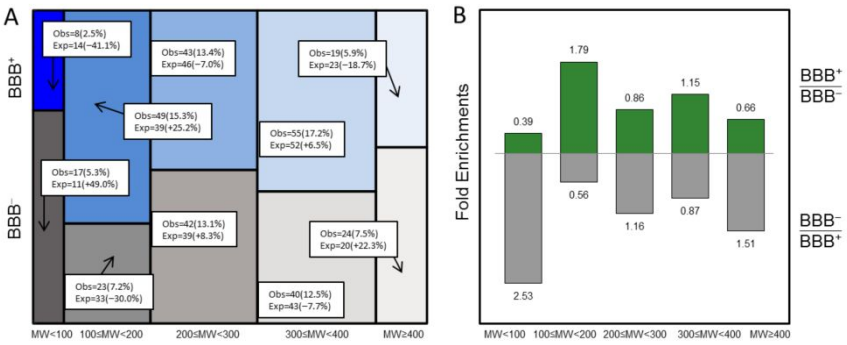


Figure 1. Observed distribution represented as mosaic plot of compounds BBB^+ (blue) and BBB^- (gray) for the MW ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis (Panel A). Fold enrichments for MW (Panel B).

As regard logP (Chi-Squared Test 47.38, p=1.27 x 10^{-9}) and logD (Chi-Squared Test 36.80, p=1.98 x 10^{-7}), significative differences are found with respect to fold enrichment of certain quadrants. Results show that a logP<1.5 is detrimental for reaching the CSN. Indeed, the population 0≤logP<1.5 is composed of 71 compounds and is represented in BBB^+ by 21 observed versus 39 expected compounds (–45.6% deviation) and in BBB^- by 50 observed versus 32 expected compounds (+54.4% deviation). Having a molecule with logP<0, it points out towards a clear inability of it to properly accumulate inside the CNS since the BBB^-/BBB^+ enrichment is 11.92 times [BBB^+ 2 observed versus 12 expected (–83.3% deviation); BBB^- 20 observed versus 10 expected (+99.3% deviation)]. This trend is inverted for logP values above 1.5, where it has been found BBB^+/BBB^- fold enrichment of 1.33, 2.15 and 1.99, respectively for the ranges 1.5≤logP<3, 3≤logP<4.5, and logP≥4.5 (Figure 2). It can be stated that a molecule with logP<1.5 will probably not cross the BBB; while for logP≥1.5, there are more chances that the designed molecule will better accumulate within the CNS.

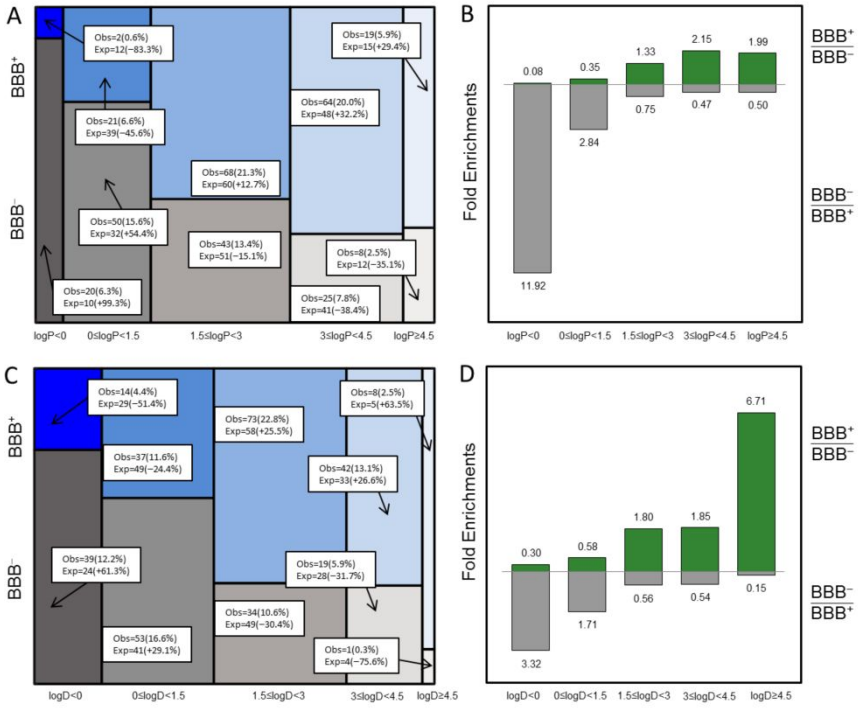


Figure 2. Observed distribution represented as mosaic plot of compounds BBB^+ (blue) and BBB^- (gray) for logP (panel A) and logD (panel C) ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for logP (panel B) and logD (Panel D).

LogD provides similar observations (Figure 2). In particular, the three quadrants BBB^+/BBB^- 1.5≤logD<3, BBB^+/BBB^- 3≤logD<4.5, and BBB^+/BBB^- logD≥4.5 show fold enrichments of 1.80, 1.85 and 6.71. LogD<1.5 points out towards a clear reduced ability of molecules to cross the BBB since the BBB^-/BBB^+ fold enrichment is 1.71 times with 53 observed compounds in the BBB^- (expected 41, +29.1% deviation) versus 37 in the BBB^+ group (expected 49, –24.4% deviation). Similarly to logP, for this descriptor there is a linear improvement of the fold enrichment moving from low to high logD values with positive BBB^+/BBB^- enrichment for logD≥1.5. NC (Chi-Squared Test 38.29, p=9.77 x 10^{-8}) and OC (Chi-Squared Test 35.83, p=3.13 x 10^{-7}) show important effects on the ability of a molecule to cross the BBB having both descriptors a highly significant distribution within the tables (Figure 3). According to chi-square test, NC should be kept between zero and one since at these values the number of BBB^- compounds is under-represented while compounds BBB^+ are over-represented. As can be seen in Figure 3, for NC=0 a BBB^+/BBB^- fold enrichment of 1.50 is found, with 50

observed compounds in the BBB^+ versus the 42 expected (+17.9% deviation). While in the BBB^- group we observe 28 compounds with 36 expected (–21.3% deviation).

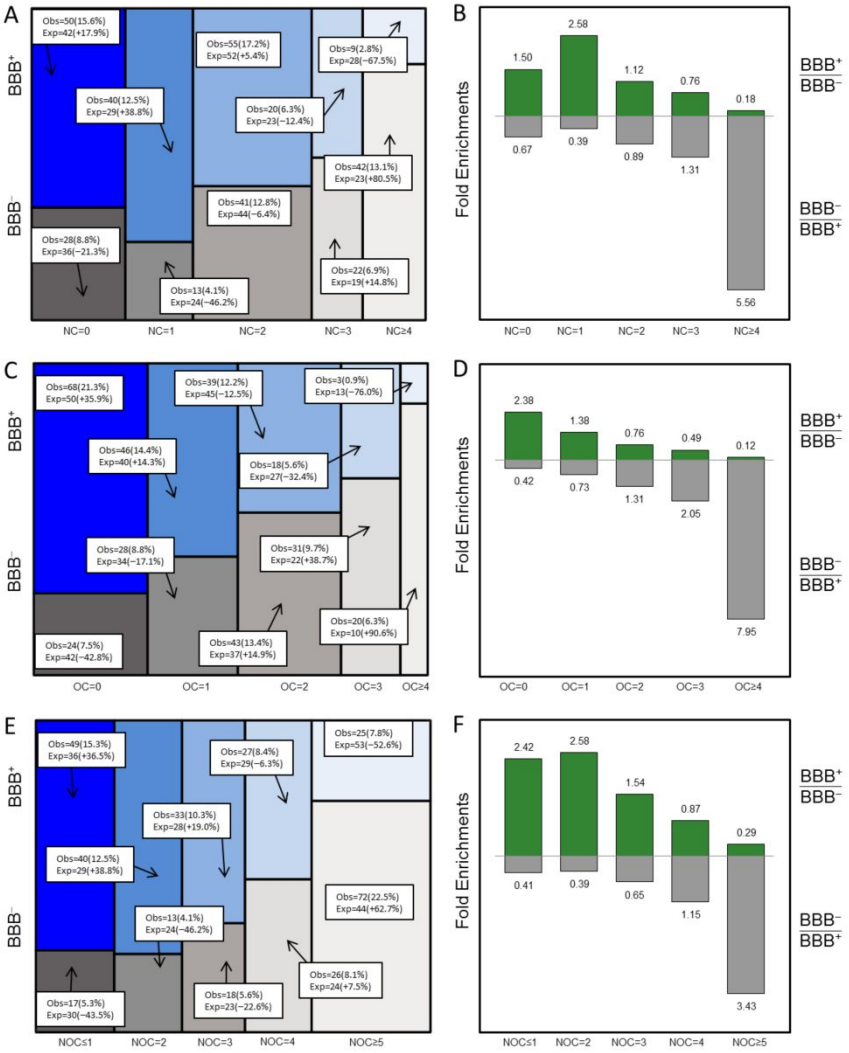


Figure 3. Observed distribution represented as mosaic plot of compounds BBB^+ (blue) and BBB^- (gray) for NC (panel A), OC (panel C), and NOC (panel E) ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for NC (panel B), OC (panel D), and NOC (panel F).

For NC=1 a BBB^+/BBB^- fold enrichment of 2.58 is found. We observe 40 compounds in the BBB^+ set with 29 compounds expected (+38.8% deviation), and 13 compounds in the BBB^- group with 24 compounds expected (–46.2% deviation). That means there is a higher probability that a molecule with NC=0–1 is in the BBB^+ group. Compounds with NC=2 show a similar distribution between BBB^+ and BBB^- thus NC=2 does

not discriminate. For these quadrants, we found a similar observed versus expected distribution [BBB^+ 55 observed compounds versus 52 expected compounds (+5.4% deviation); BBB^- 41 observed compounds versus 44 expected compounds (–6.4% deviation)], and a BBB^+/BBB^- fold enrichment of 1.13 times. At the same time molecules with a higher number of NC (NC=3) have a higher probability of being in the BBB^- group having a BBB^-/BBB^+ fold enrichment of 1.31 times. For NC≥4 (BBB^-/BBB^+ fold enrichment = 5.56) it was observed a significative deviation of –67.5% for the BBB^+ group with 28 expected compounds and 9 observed compounds. Conversely the BBB^- group is over-represented with a +80.5% deviation having observed 42 compounds in this quadrant against the expected 23 compounds. It can be summarized that a NC≥3 would result in a compound with a higher probability of being BBB^-. If we reduce NC to 2, the probability of being BBB^+ or BBB^- is similar, while if we set a value below 2 (NC=0–1) there is a higher probability that the molecule would better permeate the BBB.

More restrictions are found for the OC having OC=0 $BBB^+/BBB^- = 2.38$ and OC=1 $BBB^+/BBB^- = 1.38$, while OC=2 $BBB^+/BBB^- = 0.76$, OC=3 BBB^+/BBB^- 0.49, and for OC≥4 $BBB^+/BBB^- = 0.12$ (Figure 3). OC=0, we observe 68 compounds in BBB^+ versus 50 expected (+35.9% deviation) and 24 compounds in BBB^- versus 42 expected (–42.8% deviation). OC=1, the population is composed by 46 compounds in BBB^+ versus 40 expected (+14.3% deviation) and 28 compounds in BBB^- versus 34 expected (–17.1% deviation). OC=2, there are 39 compounds in BBB^+ versus 45 expected (–12.5% deviation) and 43 compounds in BBB^- versus 37 expected (+14.9% deviation). For this last group while the plain distribution is similar (39 versus 37 observed compounds), according to contingency table and chi-square distribution, the probability points out towards compounds mostly accumulated in the BBB^- with fold enrichment BBB^-/BBB^+ of 1.31 times. If the OC is higher or equal to 4, we observe 3 compounds in BBB^+ versus 13 expected (–76.0% deviation) and 20 compounds versus 10 expected (+90.6% deviation) in BBB^-.

Based on these results we were interested in evaluating the optimal nitrogen and oxygen count (NOC). To accomplish this, we summed NC and OC, built a contingency table and performed chi-squared test. The NOC has been grouped in NOC≤1, NOC=2, NOC=3, NOC=4 and NOC≥5 which gave a good distribution within the groups (Figure 3). For NOC it has been found an enrichment of 51.17, and a $p=2.06 \times 10^{-10}$. This parameter provides a wide enrichment and an high level of probability meaning that, as already anticipated by the NC

and OC analysis, this is a descriptor particularly important for the accumulation of a drug in the CNS with high fold enrichments [NOC≤1 BBB^+/BBB^- = 2.42, NOC=2 BBB^+/BBB^- = 2.58, NOC=3 BBB^+/BBB^- = 1.54, while NOC=4 BBB^+/BBB^- = 0.87 and for NOC≥5 BBB^+/BBB^- = 0.29]. The NOC≤1 population was composed by 49 compounds in BBB^+ versus 36 expected (+36.5% deviation) and 17 compounds versus 30 expected (–43.5% deviation) in BBB^-. A similar trend is found for NOC=2 and NOC=3. In the NOC=4 range it is observed a similar probability of being BBB^+ and BBB^-. Finally, NOC≥5, we observe 25 compounds versus 53 expected (–52.6% deviation) in BBB^+ and 72 compounds versus 44 expected (+62.7% deviation) in BBB^-. NC and OC should be kept as low as possible in designing new molecules that have to cross the BBB; it is preferred to keep NC≤2 and OC≤1. At the same time, NOC should be kept ≤3 in order to have an improved accumulation of the compound in the CNS, although for NOC=4 it is observed a similar probability of being BBB^+ and BBB^-.

HBA (Chi-Squared Test 30.80, p=3.36 x 10^{-6}) and HBD (Chi-Squared Test 29.29, p=6.81 x 10^{-6}) have shown to be fundamental for the ability of a molecule to cross the BBB (Figure 4). Fold enrichments clearly show optimal values for BBB^+ at HBA≤2 [HBA=0 (BBB^+/BBB^- = 3.24); HBA=1 (BBB^+/BBB^- = 2.04); HBA=2 (BBB^+/BBB^- = 1.31), while no significative enrichment is observed for HBA=3 (BBB^+/BBB^- = 0.94). For HBA≥4 the discrimination between BBB^+ and BBB^- is marked, indeed it is observed a BBB^+/BBB^- fold enrichment of 0.40, meaning that molecules in this range have a high probability to do not cross the BBB.

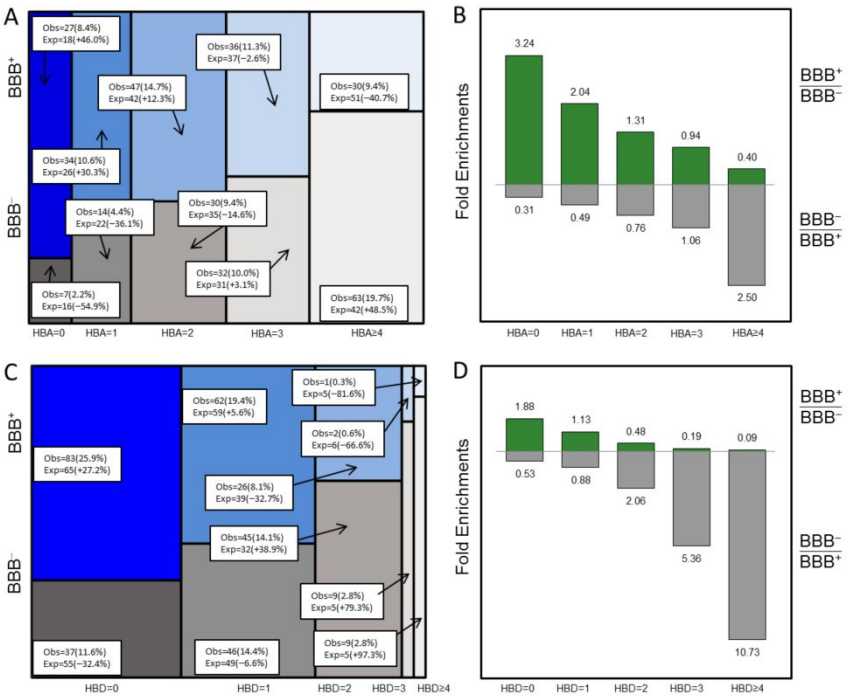


Figure 4. Observed distribution represented as mosaic plot of compounds BBB^+ (blue) and BBB^- (gray) for HBA (panel A) and HBD (panel C) ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for HBA (panel B) and HBD (Panel D).

A different trend is found for HBD (Figure 4). Regarding this parameter in order to have an optimal BBB^+/BBB^- the selection of the number of donor atom is limited. Indeed, just for HBD=0 the molecule properly accumulates inside the CNS having a BBB^+/BBB^- of 1.88. For HBD=1 the distribution is neutral with BBB^+/BBB^- fold enrichment of 1.13, while once the HBD is brought to a value of 2 the probability of a molecule to do not cross the BBB is higher being the ratio BBB^+/BBB^- equal to 0.48. While HBD≥3, clearly points towards molecules that are not able to cross the BBB with an BBB^-/BBB^+ fold enrichment of 5.36 for HBD=3, and 10.73 for HBD≥4. Overall for these two descriptors, while there is some space with HBA being a count of two not detrimental for the ability of a molecule to cross the BBB, the HBD must be kept ≤1 in order to have a molecule that cross the BBB otherwise the probability for it to do not cross is high.

The IS (Chi-Squared Test 33.02; p=3.19 x 10^{-7}) has given significative variation within the eight quadrants considered that means that the IS, is an essential parameter in determining compounds able to cross the BBB

(Figure 5). Compounds have been binned in four groups and of these the neutral apolar and the basic compounds have shown to be over-represented in the BBB^+ group [apolar neutral 42 compounds observed in BBB^+ versus 30 expected (+37.9% deviation) and 14 compounds observed in BBB^- versus 26 expected (–45.2% deviation); basic 88 compounds observed in BBB^+ versus 76 expected (+15.6% deviation) and 52 compounds observed in BBB^- versus 64 expected (–18.6% deviation)]. At the same time, acid and neutral polar compounds that include also two zwitterions, have shown to be more enriched in the BBB^- group with BBB^-/BBB^+ fold enrichment of 1.86 for neutral polar and 15.49 for acid compounds. For example, observed acid compound in the BBB^+ is 1 versus 8 expected (–86.9% deviation) while observed acid compounds in the BBB^- are 13 versus 6 expected (+103.5% deviation). This descriptor provided high variation within the eight subgroups and thus it is recommended to consider neutral polar and acid compounds as having lower probability of crossing the BBB with respect to basic and neutral apolar compounds that show higher probabilities to permeate the BBB.

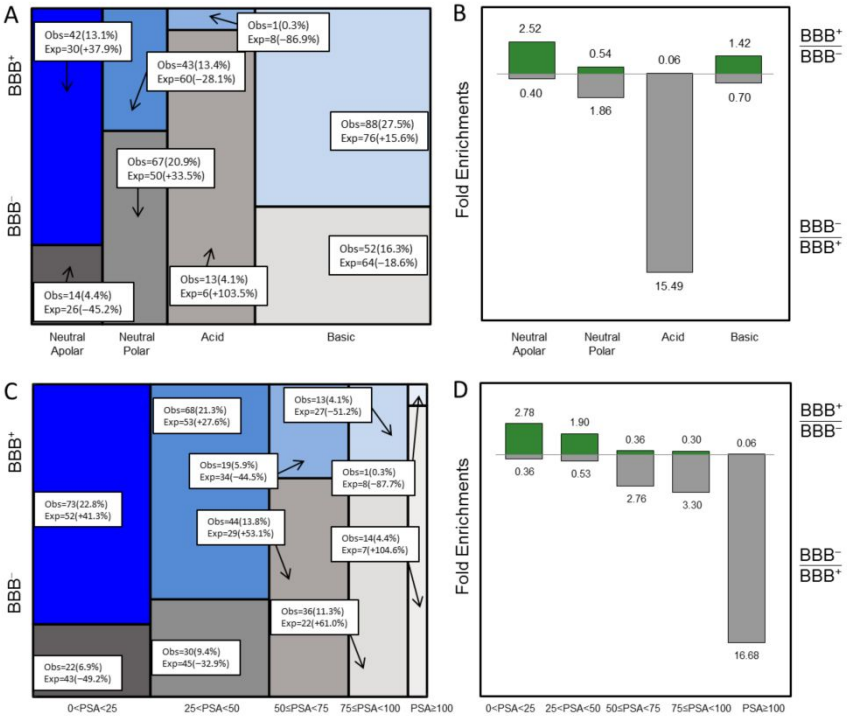


Figure 5. Observed distribution represented as mosaic plot of compounds BBB^+ (blue) and BBB^- (gray) for IS (panel A) and PSA (panel C) ranges. White boxes indicate the number of observed (obs) compounds and its

relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for IS (panel B) and PSA (Panel D).

Finally, PSA (Figure 5) has proved to be the most significative descriptor within the series with enrichment of 66.79 and p=1.08 x 10^{-13}. For the quadrant BBB^+/0<PSA<25 a fold enrichment of 2.78 times respect to BBB^-/0<PSA<30 has been found. For BBB^+ we observed 73 compounds versus 52 expected (+41.3% deviation), while for BBB^- 22 compounds are observed against 43 expected (–49.2% deviation). This trend is maintained in the group 25≤PSA<50 (BBB^+/BBB^- fold enrichment 1.90), with 68 observed compounds versus 53 expected in BBB^+ (+27.6% deviation) and 30 observed compounds versus 45 expected (–32.9% deviation) in BBB^-. While the tendency is inverted in quadrants representing higher value of PSA (50≤PSA<75; 75≤PSA<100; PSA≥100), which are total prerogative of molecules that do not cross the BBB. In particular for 50≤PSA<75 over a total population of 63 compounds, 19 are observed in the BBB^+ versus 34 expected (–44.5% deviation) and 44 observed versus 29 expected in the BBB^- population (+43.1% deviation) and a fold enrichment of 2.73 times (BBB^-/BBB^+). Similar observation can be summarized for the quadrant 75≤PSA<100. The last range, PSA≥100, shows a BBB^-/BBB^+ fold enrichment of 16.68, with a +104.6% deviation for observed BBB^- compounds versus a deviation of –87.7% for BBB^+ compounds. For this descriptor overall a value below 50 is optimal for improving the probability of reaching a desirable accumulation of the drug inside the CNS.

At this stage we were interested in setting cumulative rules that could help prospectively in the design or to define molecules with higher or lower probability to permeate the BBB. A general summary of the descriptor probabilistic ranges is given in Table 3. We consider only those descriptors that have shown a significative chi-square distribution. To this purpose, three different probabilistic scenarios have been identified, BBB^+, BBB^+/BBB^-, and BBB^-. This classification thus contemplates a scenario, BBB^+, where there is a higher probability for the molecule of crossing the BBB (BBB^+/BBB^- fold enrichment≥1.33); BBB^+/BBB^-, where the probability for the two BBB groups is similar (BBB^+/BBB^- fold enrichment<1.33; BBB^-/BBB^+ fold enrichment<1.33), meaning that compounds having descriptors falling in this range have similar probabilities of being BBB^+ or BBB^-; and a BBB^- scenario, where the probability of permeating the BBB is low

(BBB$^-$/BBB$^+$ fold enrichment≥1.33). The not-significative computed descriptors may be considered as BBB$^+$/BBB$^-$, indeed for them the population did not show any statistically significant difference between BBB$^+$ and BBB$^-$. These descriptors could eventually be taken in consideration although their contribution is limited.

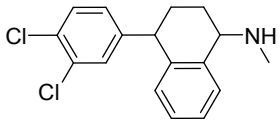
Table 3. Probabilistic scenarios.

Descriptor Range[a]				
logP<0	0≤logP<1.5	1.5≤logP<3.0	3.0≤logP<4.5	logP≥4.5
logD<0	0≤logD<1.5	1.5≤logD<3.0	3.0≤logD<4.5	logD≥4.5
0<PSA<25	25≤PSA<50	50≤PSA<75	75≤PSA<100	PSA≥100
IS=Neutral Apolar	IS=Neutral Polar	IS=Acid	IS=Basic	
HBA=0	HBA=1	HBA=2	HBA=3	HBA≥4
HBD=0	HBD=1	HBD=2	HBD=3	HBD≥4
NC=0	NC=1	NC=2	NC=3	NC≥4
OC=0	OC=1	OC=2	OC=3	OC≥4
NOC<1	NOC=2	NOC=3	NOC=4	NOC≥5

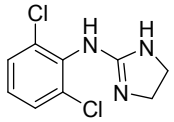
[a]red cells indicate BBB$^-$ scenario;
yellow cells indicate BBB$^+$/BBB$^-$ scenario;
green cells indicate BBB$^+$ scenario.

As representative example, sertraline is a compound with good descriptors distribution and with a logBB value of +1.6. Over the nine significative descriptors, sertraline shows eight descriptors in BBB$^+$ and one in BBB$^+$/BBB$^-$. Clonidine, with a logBB value of +0.11, shows six descriptors in BBB$^+$, two in BBB$^+$/BBB$^-$ and one in BBB$^-$. Methotrexate has a BBB$^-$ distribution for the nine descriptors. Indeed, methotrexate with logBB value of −1.51 is a compound with very low ability to cross the BBB (Table 4).

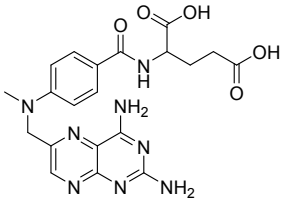
Table 4. Distribution of descriptor values for three representative compounds.



Sertraline[a]			
logP	logD	PSA	HBA
5.15	3.02	16.61	1
HBD	NC	OC	NOC
1	1	0	1
IS			
Basic			
logBB=1.6			



Clonidine			
logP	logD	PSA	HBA
2.49	1.66	38.03	3
HBD	NC	OC	NOC
2	3	0	3
IS			
Basic			
logBB=0.11			



Methotrexate			
logP	logD	PSA	HBA
−0.24	−6.56	216.20	12
HBD	NC	OC	NOC
5	8	5	13
IS			
Acid			
logBB=−1.51			

[a]red cells indicate BBB⁻ scenario;
yellow cells indicate BBB⁺/BBB⁻ scenario;
green cells indicate BBB⁺ scenario.

Conclusion

The purpose of this work was to create a simple set of rules that could prospectively support the development and identification of CNS, non-CNS or mixed drugs. We statistically analyzed a set of 328 compounds and correlated their logBB with common computed descriptors. The work has contemplated the use of contingency tables and chi-square distributions for evaluating if there were significant variations in the properties of the BBB⁺ and BBB⁻ populations. Over the computed descriptors nine resulted to be significant for the ability of a molecule to cross the BBB, while six have shown to not influence the BBB permeation. The work has allowed to define a set of rules that would eventually be used to tune the properties for BBB permeation. We set a three cases probabilistic scenario depending on the combination of multiple descriptors. Based on the data analysis, the suggested ranges should be considered as optimal values to be approached respect to stringent cutoffs that have to be tightly followed. Indeed, the distribution inside and outside the CNS is an equilibrium process resulting from a number of different variables.

Methods

A 2×5 contingency table was constructed, crossing BBB "positive/negative" by the mean of logBB (BBB⁺, logBB >0.0; BBB⁻, logBB <0.0) with a descriptor binned in five ranges (I–V) (Figure 6).

		Descriptor (A^s)*					
		Range^{I} (A^{I})	Range^{II} (A^{II})	Range^{III} (A^{III})	Range^{IV} (A^{IV})	Range^{V} (A^{V})	
logBB (A^r)*	BBB⁺ (A^+) (>cutoff)	BBB⁺|Range^{I} Number of BBB⁺ compounds with descriptor value in Range^{I} (A^{I+})	BBB⁺|Range^{II} Number of BBB⁺ compounds with descriptor value in Range^{II} (A^{II+})	BBB⁺|Range^{III} Number of BBB⁺ compounds with descriptor value in Range^{III} (A^{III+})	BBB⁺|Range^{IV} Number of BBB⁺ compounds with descriptor value in Range^{IV} (A^{IV+})	BBB⁺|Range^{V} Number of BBB⁺ compounds with descriptor value in Range^{V} (A^{V+})	$\sum_s A^{s+}$
	BBB⁻ (A^-) (<cutoff)	BBB⁻|Range^{I} Number of BBB⁻ compounds with descriptor value in Range^{I} (A^{I-})	BBB⁻|Range^{II} Number of BBB⁻ compounds with descriptor value in Range^{II} (A^{II-})	BBB⁻|Range^{III} Number of BBB⁻ compounds with descriptor value in Range^{III} (A^{III-})	BBB⁻|Range^{IV} Number of BBB⁻ compounds with descriptor value in Range^{IV} (A^{IV-})	BBB⁻|Range^{V} Number of BBB⁻ compounds with descriptor value in Range^{V} (A^{V-})	$\sum_s A^{s-}$
		$\sum_r A^{Ir}$	$\sum_r A^{IIr}$	$\sum_r A^{IIIr}$	$\sum_r A^{IVr}$	$\sum_r A^{Vr}$	$\sum_{s,r} A^{sr}$

*where s labels the ranges (I–V) and r the BBB groups (BBB⁺ or BBB⁻)

Figure 6. Example of 2 × 5 contingency table

From these tables expected distribution has been calculated for each quadrant as reported below:

$$Expected\ distribution\ (A^{sr}) = \frac{\sum_l A^{sr} \cdot \sum_s A^{sr}}{\sum_{s,r} A^{sr}}$$

where s labels the ranges (I–V) and r the BBB groups (BBB^+ or BBB^-).

While percent deviations between observed and expected distribution have been calculated with the following formula:

$$Percent\ Deviation\ (A^{sr}) = \frac{Obs\ A^{(sr)} - Exp\ A^{(sr)}}{Exp\ A^{(sr)}} \cdot 100$$

where s labels the ranges (I–V) and r the BBB groups (BBB^+ or BBB^-).

Conditional probabilities have been calculated as reported below:

$$pct^{sr} = \frac{A^{sr}}{\sum_{s,r} A^{sr}}$$

where s labels the ranges (I–V) and r the BBB groups (BBB^+ or BBB^-).

Finally, two fold enrichments have been calculated for each descriptor range as ratio of the two conditional probabilities:

$$\left(\frac{BBB^+}{BBB^-}\right) fold\ enrichment^{s+} = \frac{pct^{s+}}{pct^{s-}}$$

$$\left(\frac{BBB^-}{BBB^+}\right) fold\ enrichment^{s-} = \frac{pct^{s-}}{pct^{s+}}$$

where s labels the ranges (I–V).

p values (chi-squared test) with Yates correction were calculated using R version 3.3.3 (2017-03-06). Mosaic plots have been generated using implemented functions of MATLAB (version R2018b).[44]

Abbreviations

ADMET, absorption, distribution, metabolism, and excretion – toxicity; ANA, American Neurological Association; AC, atom count; BBB, blood brain barrier; BBB^+, BBB permeable; BBB^-, BBB non-permeable; CNS, central nervous system; HAC, heavy atom count; HBA, hydrogen bond acceptors; HBD, hydrogen bond donors; IS, ionization state; MRP, multidrug resistance-protein; MW, molecular weight; NC, nitrogen count; NOC, nitrogen and oxygen count; OC, oxygen count; P-gp, P-glycoprotein; PSA, polar surface area; RBC, rotatable bound count; RC, ring count.

Author information

Corresponding Author

*E-mail: eamata@unict.it

ORCID

Emanuele Amata: 0000-0002-4750-3479

Rita Turnaturi: 0000-0002-5895-7820

Agostino Marrazzo: 0000-0002-8728-8857

Author Contributions

E.A. was responsible for the study design, analysis and interpretation of the data. M.D. and A.M. were responsible for the acquisition of data. E.A. and B.A. were responsible for the graphics design. E.A. and M.D. discussed the results and wrote the manuscript. All authors have participated in the writing refinement and given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest

Acknowledgement

This work was supported by the Italian MIUR, PRIN 2017, Code: 201744BN5T. All the authors gratefully acknowledge Dr. Giovanni Nastasi for his helpful support.

Associated content

Supporting information

Mosaic plots and fold enrichment plots for non significative descriptors AC, HAC, RC, RBC, and MR; dataset in tabulated form: SMILES, ID original article, logBB, reference; descriptors values maximum, minimum and mean.

References

1. Gooch, C. L.; Pracht, E.; Borenstein, A. R. The burden of neurological disease in the United States: A summary report and call to action. *Ann. Neurol.* **2017,** 81, 479-484.
2. Achrol, A. S.; Rennert, R. C.; Anders, C.; Soffietti, R.; Ahluwalia, M. S.; Nayak, L.; Peters, S.; Arvold, N. D.; Harsh, G. R.; Steeg, P. S.; Chang, S. D. Brain metastases. *Nat. Rev. Dis. Primers* **2019,** 5, 5.
3. Di, L.; Rong, H.; Feng, B. Demystifying brain penetration in central nervous system drug discovery. Miniperspective. *J. Med. Chem.* **2013,** 56, 2-12.
4. Hitchcock, S. A.; Pennington, L. D. Structure-brain exposure relationships. *J. Med. Chem.* **2006,** 49, 7559-83.
5. Kusuhara, H.; Sugiyama, Y. Active efflux across the blood-brain barrier: role of the solute carrier family. *NeuroRx* **2005,** 2, 73-85.
6. Hitchcock, S. A. Structural modifications that alter the P-glycoprotein efflux properties of compounds. *J. Med. Chem.* **2012,** 55, 4877-95.
7. Cecchelli, R.; Berezowski, V.; Lundquist, S.; Culot, M.; Renftel, M.; Dehouck, M. P.; Fenart, L. Modelling of the blood-brain barrier in drug discovery and development. *Nat. Rev. Drug Discovery* **2007,** 6, 650-61.
8. Mensch, J.; Melis, A.; Mackie, C.; Verreck, G.; Brewster, M. E.; Augustijns, P. Evaluation of various PAMPA models to identify the most discriminating method for the prediction of BBB permeability. *Eur. J. Pharm. Biopharm.* **2010,** 74, 495-502.
9. Ciura, K.; Dziomba, S. Application of separation methods for in vitro prediction of blood-brain barrier permeability-The state of the art. *J. Pharm. Biomed. Anal.* **2020,** 177, 112891.
10. Espada, A.; Molina-Martin, M. Capillary electrophoresis and small molecule drug discovery: a perfect match? *Drug discovery today* **2012,** 17, 396-404.
11. Naik, P.; Cucullo, L. In vitro blood-brain barrier models: current and perspective technologies. *J. Pharm. Sci.* **2012,** 101, 1337-54.
12. Reichel, A. Integrated Approach to Optimizing CNS Penetration in Drug Discovery: From the Old to the New Paradigm and Assessment of Drug–Transporter Interactions. In *Drug Delivery to the Brain: Physiological Concepts, Methodologies and Approaches*, Hammarlund-Udenaes, M.; de Lange, E. C. M.; Thorne, R. G., Eds. Springer New York: New York, NY, 2014; pp 339-374.
13. Rabal, O.; Sánchez-Arias, J.A.; Cuadrado-Tejedor, M.; de Miguel, I.; Pérez-González, M.; García-Barroso, C.; Ugarte, A.; Estella-Hermoso de Mendoza, A.; Sáez, E.; Espelosin, M.; Ursua, S.; Haizhong, T.; Wei, W.; Musheng, X.; Garcia-Osta, A.; Oyarzabal, J. Design, Synthesis, and Biological Evaluation of First-in-Class Dual Acting Histone Deacetylases (HDACs) and Phosphodiesterase 5 (PDE5) Inhibitors for the Treatment of Alzheimer's Disease. *J. Med. Chem*. **2016**, 59, 8967-9004.

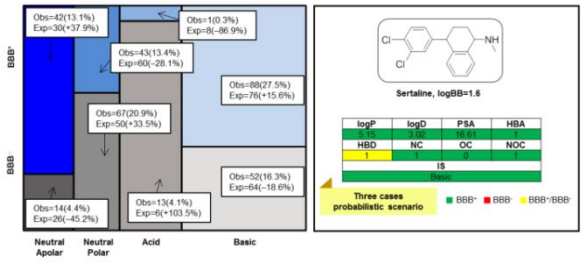
14. Fiorito, J.; Vendome, J.; Saeed, F.; Staniszewski, A.; Zhang, H.; Yan, S.; Deng, S.X.; Arancio, O.; Landry, D.W. Identification of a Novel 1,2,3,4-Tetrahydrobenzo[b][1,6]naphthyridine Analogue as a Potent Phosphodiesterase 5 Inhibitor with Improved Aqueous Solubility for the Treatment of Alzheimer's Disease. *J. Med. Chem*. **2017,** 60, 8858-8875.
15. Frantz, M.C.; Pellissier, L.P.; Pflimlin, E.; Loison, S.; Gandía, J.; Marsol, C.; Durroux, T.; Mouillac, B.; Becker, J.A.J.; Le Merrer, J.; Valencia, C.; Villa, P.; Bonnet, D.; Hibert, M. LIT-001, the First Nonpeptide Oxytocin Receptor Agonist that Improves Social Interaction in a Mouse Model of Autism. *J. Med. Chem*. **2018,** 61, 8670-8692.
16. Sánchez-Arias, J.A.; Rabal, O.; Cuadrado-Tejedor, M.; de Miguel, I.; Pérez-González, M.; Ugarte, A.; Sáez, E.; Espelosin, M.; Ursua, S.; Haizhong, T.; Wei, W.; Musheng, X.; Garcia-Osta, A.; Oyarzabal, J. Impact of Scaffold Exploration on Novel Dual-Acting Histone Deacetylases and Phosphodiesterase 5 Inhibitors for the Treatment of Alzheimer's Disease. *ACS Chem. Neurosci.* **2017,** 8, 638-661
17. Han, J.; Lee, H.J.; Kim, K.Y.; Lee, S.J.C.; Suh, J.M.; Cho, J.; Chae, J.; Lim, M.H. Tuning Structures and Properties for Developing Novel Chemical Tools toward Distinct Pathogenic Elements in Alzheimer's Disease. *ACS Chem. Neurosci.* **2018,** 9, 800-808.
18. Rabal, O.; Sánchez-Arias, J.A.; Cuadrado-Tejedor, M.; de Miguel, I.; Pérez-González, M.; García-Barroso, C.; Ugarte, A.; Estella-Hermoso de Mendoza, A.; Sáez, E.; Espelosin, M.; Ursua, S.; Haizhong, T.; Wei, W.; Musheng, X.; Garcia-Osta, A.; Oyarzabal, J. Discovery of in Vivo Chemical Probes for Treating Alzheimer's Disease: Dual Phosphodiesterase 5 (PDE5) and Class I Histone Deacetylase Selective Inhibitors. *ACS Chem. Neurosci.* **2019,** 10, 1765-1782.
19. Rook, J.M.; Abe, M.; Cho, H.P.; Nance, K.D.; Luscombe, V.B.; Adams, J.J.; Dickerson, J.W.; Remke, D.H.; Garcia-Barrantes, P.M.; Engers, D.W.; Engers, J.L.; Chang, S.; Foster, J.J.; Blobaum; A.L.; Niswender, C.M.; Jones, C.K.; Conn, P.J.; Lindsley, C.W. Diverse Effects on M1 Signaling and Adverse Effect Liability within a Series of M1 Ago-PAMs. *ACS Chem. Neurosci.* **2017,** 8, 866-883.
20. Reed, C.W.; Yohn, S.E.; Washecheck, J.P.; Roenfanz, H.F.; Quitalig, M.C.; Luscombe, V.B.; Jenkins M.T.; Rodriguez, A.L.; Engers, D.W.; Blobaum, A.L.; Conn, P.J.; Niswender, C.M.; Lindsley, C.W. Discovery of an Orally Bioavailable and Central Nervous System (CNS) Penetrant mGlu7 Negative Allosteric Modulator (NAM) in Vivo Tool Compound: N-(2-(1H-1,2,4-triazol-1-yl)-5-(trifluoromethoxy)phenyl)-4-(cyclopropylmethoxy)-3-methoxybenzamide (VU6012962). *J. Med. Chem.* **2019,** 62, 1690-1695.
21. Pajouhesh, H.; Lenz, G. R. Medicinal chemical properties of successful central nervous system drugs. *NeuroRx* **2005,** 2, 541-53.
22. Fan, Y.; Unwalla, R.; Denny, R. A.; Di, L.; Kerns, E. H.; Diller, D. J.; Humblet, C. Insights for predicting blood-brain barrier penetration of CNS targeted molecules using QSPR approaches. *J. Chem. Inf. Model.* **2010,** 50, 1123-33.
23. Ghose, A. K.; Herbertz, T.; Hudkins, R. L.; Dorsey, B. D.; Mallamo, J. P. Knowledge-Based, Central Nervous System (CNS) Lead Selection and Lead Optimization for CNS Drug Discovery. *ACS Chem. Neurosci.* **2012,** 3, 50-68.
24. Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci.* **2010,** 1, 435-49.
25. Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery. *ACS Chem. Neurosci.* **2016,** 7, 767-75.
26. Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **2008,** 51, 817-34.
27. Mahar Doan, K. M.; Humphreys, J. E.; Webster, L. O.; Wring, S. A.; Shampine, L. J.; Serabjit-Singh, C. J.; Adkison, K. K.; Polli, J. W. Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs. *J. Pharmacol. Exp. Ther.* **2002,** 303, 1029-37.
28. Desai, P. V.; Sawada, G. A.; Watson, I. A.; Raub, T. J. Integration of in silico and in vitro tools for scaffold optimization during drug discovery: predicting P-glycoprotein efflux. *Mol. Pharmaceutics* **2013,** 10, 1249-61.
29. Amata, E.; Xi, H.; Colmenarejo, G.; Gonzalez-Diaz, R.; Cordon-Obras, C.; Berlanga, M.; Manzano, P.; Erath, J.; Roncal, N. E.; Lee, P. J.; Leed, S. E.; Rodriguez, A.; Sciotti, R. J.; Navarro, M.; Pollastri, M. P. Identification of "Preferred" Human Kinase Inhibitors for Sleeping Sickness Lead Discovery. Are Some Kinases Better than Others for Inhibitor Repurposing? *ACS Infect. Dis.* **2016,** 2, 180-186.

30. Li, H.; Yap, C. W.; Ung, C. Y.; Xue, Y.; Cao, Z. W.; Chen, Y. Z. 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.* **2005,** 45, 1376-84.
31. Muehlbacher, M.; Spitzer, G. M.; Liedl, K. R.; Kornhuber, J. Qualitative prediction of blood-brain barrier permeability on a large and refined dataset. *J. Comput.-Aided Mol. Des.* **2011,** 25, 1095-106.
32. Liu, R.; Sun, H.; So, S. S. Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 2. Blood-brain barrier penetration. *J. Chem. Inf. Model.* **2001,** 41, 1623-32.
33. Abraham, M. H.; Ibrahim, A.; Zhao, Y.; Acree, W. E., Jr. A data base for partition of volatile organic compounds and drugs from blood/plasma/serum to brain, and an LFER analysis of the data. *J. Pharm. Sci.* **2006,** 95, 2091-100.
34. Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000,** 43, 2204-16.
35. Bendayan, R.; Lee, G.; Bendayan, M. Functional expression and localization of P-glycoprotein at the blood brain barrier. *Microsc. Res. Tech.* **2002,** 57, 365-80.
36. Kim, R. B. Drugs as P-glycoprotein substrates, inhibitors, and inducers. *Drug Metab. Rev.* **2002,** 34, 47-54.
37. Doran, A.; Obach, R. S.; Smith, B. J.; Hosea, N. A.; Becker, S.; Callegari, E.; Chen, C.; Chen, X.; Choo, E.; Cianfrogna, J.; Cox, L. M.; Gibbs, J. P.; Gibbs, M. A.; Hatch, H.; Hop, C. E.; Kasman, I. N.; Laperle, J.; Liu, J.; Liu, X.; Logman, M.; Maclin, D.; Nedza, F. M.; Nelson, F.; Olson, E.; Rahematpura, S.; Raunig, D.; Rogers, S.; Schmidt, K.; Spracklin, D. K.; Szewc, M.; Troutman, M.; Tseng, E.; Tu, M.; Van Deusen, J. W.; Venkatakrishnan, K.; Walens, G.; Wang, E. Q.; Wong, D.; Yasgar, A. S.; Zhang, C. The impact of P-glycoprotein on the disposition of drugs targeted for indications of the central nervous system: evaluation using the MDR1A/1B knockout mouse model. *Drug Metab. Dispos.* **2005,** 33, 165-74.
38. Matsuzaki, J.; Yamamoto, C.; Miyama, T.; Takanaga, H.; Matsuo, H.; Ishizuka, H.; Kawahara, Y.; Kuwano, M.; Naito, M.; Tsuruo, T.; Sawada, Y. Contribution of P-glycoprotein to bunitrolol efflux across blood-brain barrier. *Biopharm. Drug Dispos.* **1999,** 20, 85-90.
39. El Ela, A. A.; Hartter, S.; Schmitt, U.; Hiemke, C.; Spahn-Langguth, H.; Langguth, P. Identification of P-glycoprotein substrates and inhibitors among psychoactive compounds--implications for pharmacokinetics of selected substrates. *J. Pharm. Pharmacol.* **2004,** 56, 967-75.
40. Park, J.; Zhang, Y.; Vykhodtseva, N.; Jolesz, F. A.; McDannold, N. J. The kinetics of blood brain barrier permeability and targeted doxorubicin delivery into brain induced by focused ultrasound. *J. Controlled Release* **2012,** 162, 134-42.
41. Edwards, J. E.; Alcorn, J.; Savolainen, J.; Anderson, B. D.; McNamara, P. J. Role of P-glycoprotein in distribution of nelfinavir across the blood-mammary tissue barrier and blood-brain barrier. *Antimicrob. Agents Chemother.* **2005,** 49, 1626-8.
40. Toornvliet, R.; van Berckel, B. N.; Luurtsema, G.; Lubberink, M.; Geldof, A. A.; Bosch, T. M.; Oerlemans, R.; Lammertsma, A. A.; Franssen, E. J. Effect of age on functional P-glycoprotein in the blood-brain barrier measured by use of (R)-[(11)C]verapamil and positron emission tomography. *Clin. Pharmacol. Ther.* **2006,** 79, 540-8.
42. Kim, R. B.; Wandel, C.; Leake, B.; Cvetkovic, M.; Fromm, M. F.; Dempsey, P. J.; Roden, M. M.; Belas, F.; Chaudhary, A. K.; Roden, D. M.; Wood, A. J.; Wilkinson, G. R. Interrelationship between substrates and inhibitors of human CYP3A and P-glycoprotein. *Pharm. Res.* **1999,** 16, 408-14.
43. Lobell, M.; Molnar, L.; Keseru, G. M. Recent advances in the prediction of blood-brain partitioning from molecular structure. *J. Pharma. Sci.* **2003,** 92, 360-70.
44. MATLAB. Matrix laboratory. http://www.mathworks.com/products/matlab/

For Table of Contents only

Tuning Properties for Blood–Brain Barrier (BBB) Permeation: A Statistics-Based Analysis

Maria Dichiara, Benedetto Amata, Rita Turnaturi, Agostino Marrazzo, Emanuele Amata*



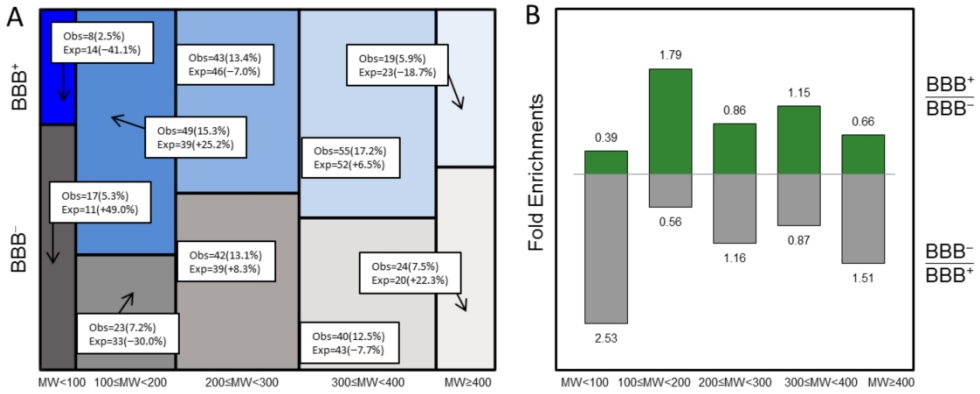


Figure 1. Observed distribution represented as mosaic plot of compounds BBB+ (blue) and BBB– (gray) for the MW ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis (Panel A). Fold enrichments for MW (Panel B).

329x135mm (300 x 300 DPI)

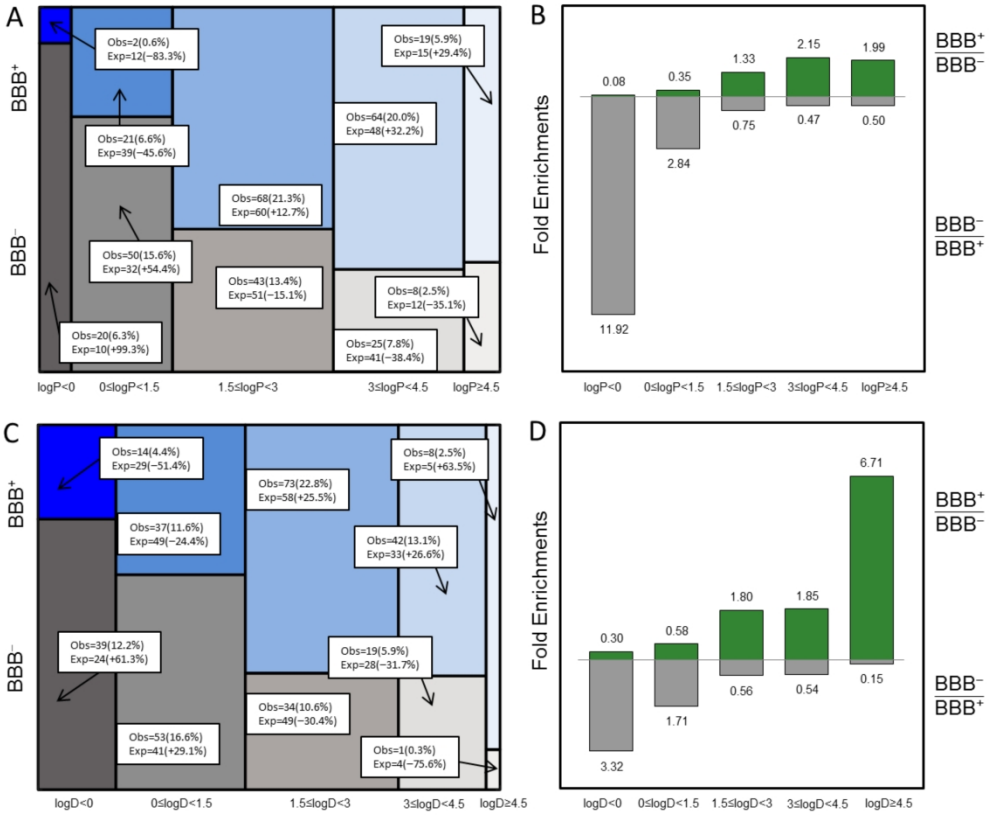


Figure 2. Observed distribution represented as mosaic plot of compounds BBB+ (blue) and BBB– (gray) for logP (panel A) and logD (panel C) ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for logP (panel B) and logD (Panel D).

325x269mm (300 x 300 DPI)

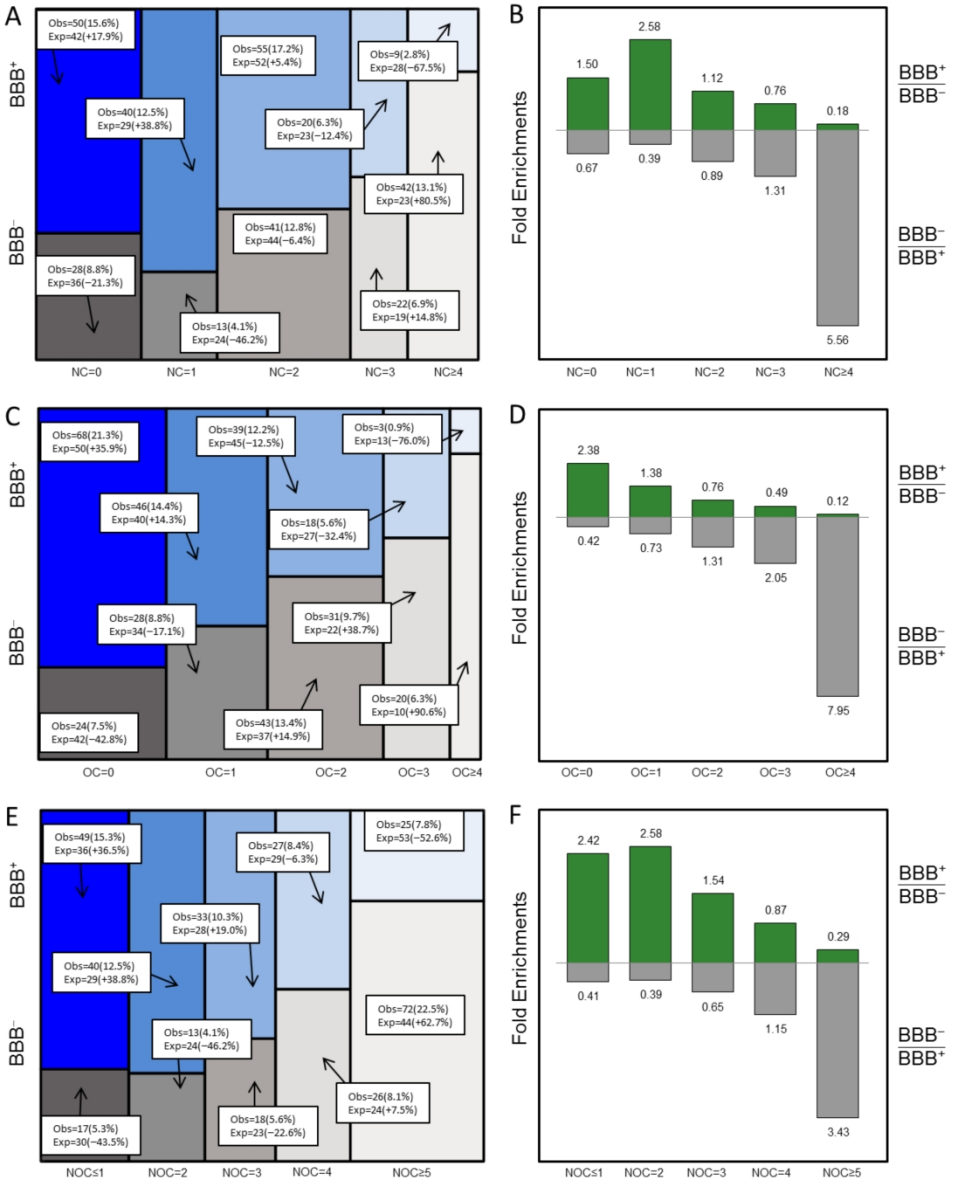


Figure 3. Observed distribution represented as mosaic plot of compounds BBB+ (blue) and BBB– (gray) for NC (panel A), OC (panel C), and NOC (panel E) ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for NC (panel B), OC (panel D), and NOC (panel F).

329x406mm (300 x 300 DPI)

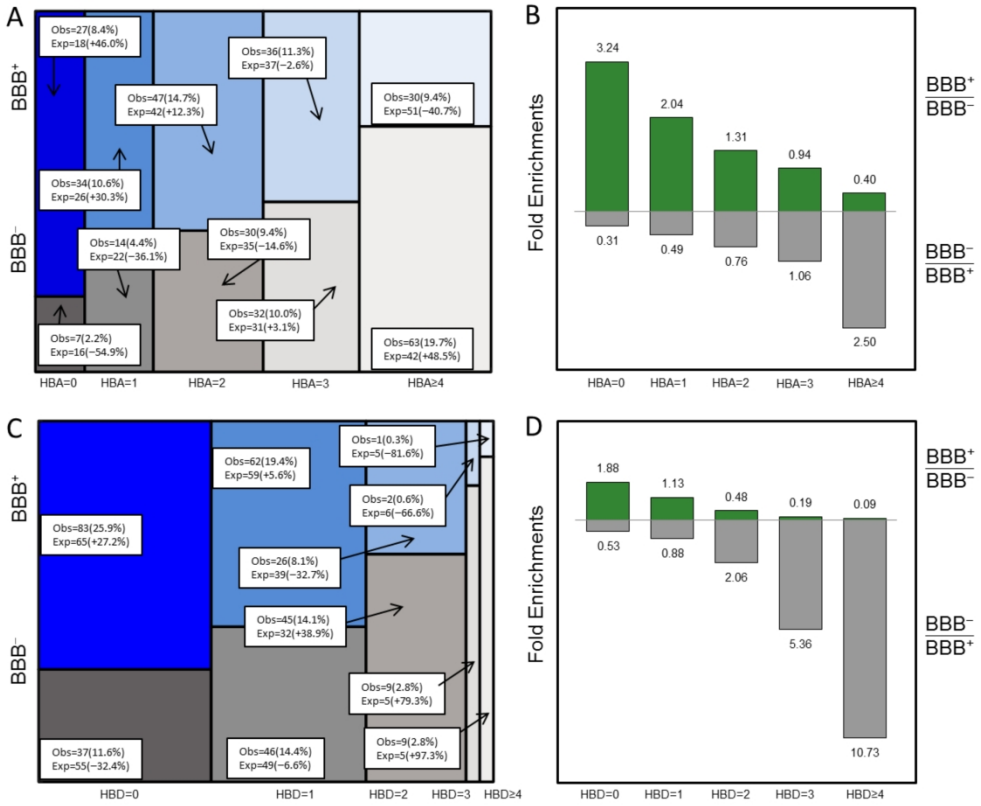
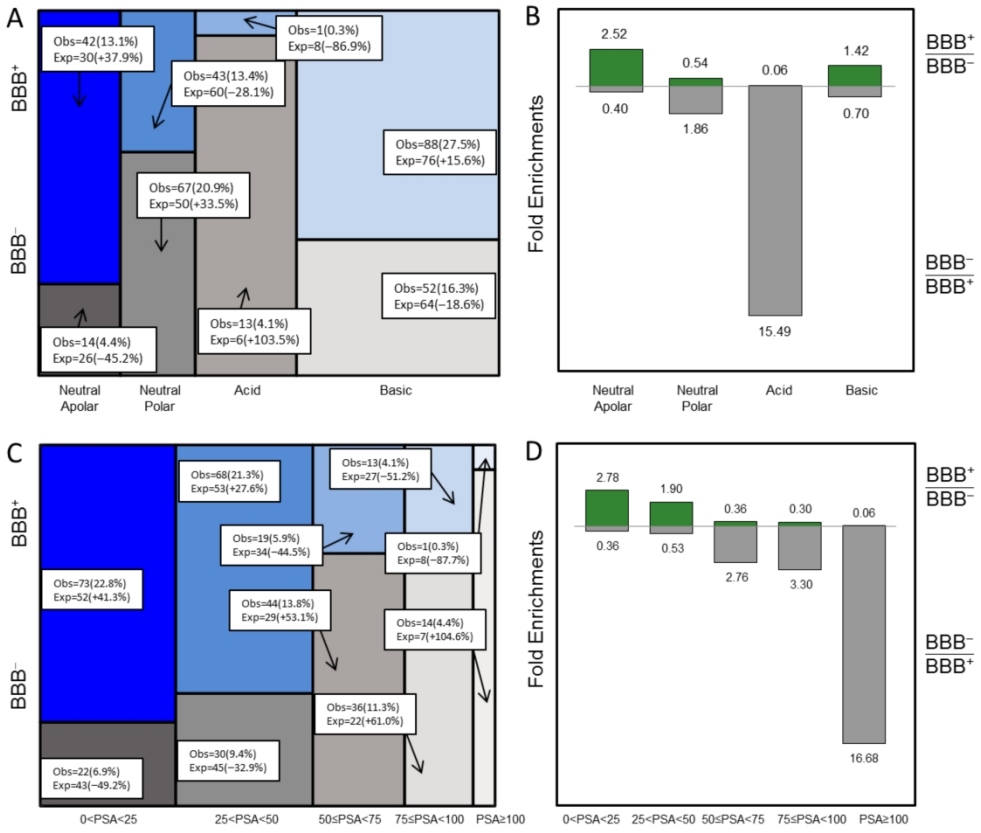


Figure 4. Observed distribution represented as mosaic plot of compounds BBB+ (blue) and BBB– (gray) for HBA (panel A) and HBD (panel C) ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for HBA (panel B) and HBD (Panel D).

329x270mm (300 x 300 DPI)



Caption : Figure 5. Observed distribution represented as mosaic plot of compounds BBB+ (blue) and BBB– (gray) for IS (panel A) and PSA (panel C) ranges. White boxes indicate the number of observed (obs) compounds and its relative percent respect the total set and the expected (exp) number of compounds with the percent deviation in parenthesis. Fold enrichments for IS (panel B) and PSA (Panel D).

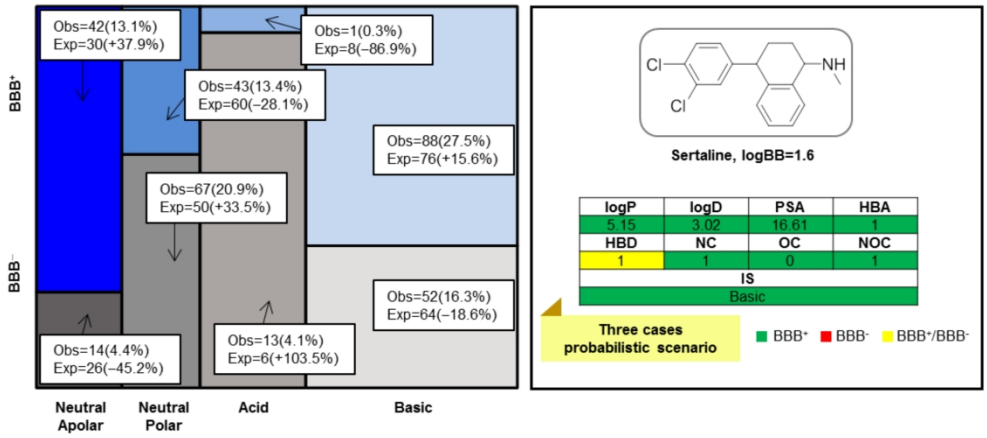
329x278mm (300 x 300 DPI)

		Descriptor (A^s)*					
		Range^{I} (A^{I})	Range^{II} (A^{II})	Range^{III} (A^{III})	Range^{IV} (A^{IV})	Range^{V} (A^{V})	
logBB (A^r)*	BBB$^+$ (A^+) (>cutoff)	**BBB$^+$|Range^{I}** Number of BBB$^+$ compounds with descriptor value in Range^{I} (A^{I+})	**BBB$^+$|Range^{II}** Number of BBB$^+$ compounds with descriptor value in Range^{II} (A^{II+})	**BBB$^+$|Range^{III}** Number of BBB$^+$ compounds with descriptor value in Range^{III} (A^{III+})	**BBB$^+$|Range^{IV}** Number of BBB$^+$ compounds with descriptor value in Range^{IV} (A^{IV+})	**BBB$^+$|Range^{V}** Number of BBB$^+$ compounds with descriptor value in Range^{V} (A^{V+})	$\sum_s A^{s+}$
	BBB$^-$ (A^-) (<cutoff)	**BBB$^-$|Range^{I}** Number of BBB$^-$ compounds with descriptor value in Range^{I} (A^{I-})	**BBB$^-$|Range^{II}** Number of BBB$^-$ compounds with descriptor value in Range^{II} (A^{II-})	**BBB$^-$|Range^{III}** Number of BBB$^-$ compounds with descriptor value in Range^{III} (A^{III-})	**BBB$^-$|Range^{IV}** Number of BBB$^-$ compounds with descriptor value in Range^{IV} (A^{IV-})	**BBB$^-$|Range^{V}** Number of BBB$^-$ compounds with descriptor value in Range^{V} (A^{V-})	$\sum_s A^{s-}$
		$\sum_r A^{Ir}$	$\sum_r A^{IIr}$	$\sum_r A^{IIIr}$	$\sum_r A^{IVr}$	$\sum_r A^{Vr}$	$\sum_{s,r} A^{sr}$

*where s labels the ranges (I–V) and r the BBB groups (BBB$^+$ or BBB$^-$)

Figure 6. Example of 2×5 contingency table

307x157mm (300 x 300 DPI)



295x130mm (300 x 300 DPI)