Journal of Medicinal Chemistry

Article

Cite This: J. Med. Chem. 2019, 62, 9824−9836

pubs.acs.org/jmc

The Blood−Brain Barrier (BBB) Score

Mayuri Gupta,[†] Hyeok Jun Lee,[†] Christopher J. Barden,[†] and Donald F. Weaver*[,†,‡,§]

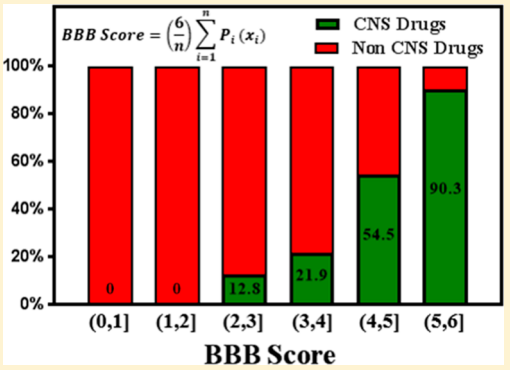
[†]Krembil Research Institute, University Health Network, 60 Leonard Avenue, Toronto M5T 2S8, Canada

[‡]Department of Medicine, University of Toronto, Toronto M5G 2C4, Canada

[§]Department of Chemistry, University of Toronto, Toronto M55 3H6, Canada

Supporting Information

ABSTRACT: The blood−brain barrier (BBB) protects the brain from the toxic side effects of drugs and exogenous molecules. However, it is crucial that medications developed for neurological disorders cross into the brain in therapeutic concentrations. Understanding the BBB interaction with drug molecules based on physicochemical property space can guide effective and efficient drug design. An algorithm, designated "BBB Score", composed of stepwise and polynomial piecewise functions, is herein proposed for predicting BBB penetration based on five physicochemical descriptors: number of aromatic rings, heavy atoms, MWHBN (a descriptor comprising molecular weight, hydrogen bond donor, and hydrogen bond acceptors), topological polar surface area, and pKa. On the basis of statistical analyses of our results, the BBB Score outperformed (AUC = 0.86) currently employed MPO approaches (MPO, AUC = 0.61; MPO_V2, AUC = 0.67). Initial evaluation of physicochemical property space using the BBB Score is a valuable addition to currently available drug design algorithms.



1. INTRODUCTION

Efficient drug development strategies result in cost and time savings when developing a therapeutic agent, whether for central nervous system (CNS) or non-CNS disease indications. While developing non-CNS drugs, one of the important design factors is if it crosses the blood−brain barrier (BBB)[1,2] or not.[3] Significant brain exposure of a CNS drug enables the treatment of neurological disorders but can also lead to variety of untoward side effects for non-CNS drugs. To mitigate this design problem at the design stage before costly in vivo studies,[4] various mathematical algorithms based on analyses of property space of compound physicochemical properties of drugs have been proposed.

Lipinski's benchmark rule of five (RO5)[5] defines desirable drug candidate physicochemical property space as MW < 500 Da, log P < 5, HBD < 5, and HBA < 10. The conventional RO5 has gained popularity as a "good rule of thumb" but is not specific for CNS drugs. Various groups have endeavored to map the physicochemical space of CNS drugs using diversified approaches. A data set of 125 CNS and non-CNS drugs, analyzed by Van der Waterbeemd et al.,[6] suggested the following desirability ranges for CNS penetration: MW < 450, log D [1,4], and PSA < 90 Å.[2] Kelder et al.[7] proposed a polar molecular surface area (PSA) range of <70 Å^2 for most CNS drugs. Norinder et al.[8] suggested that a molecule with O + N < 5 or cLogP − (O + N) > 0 has a higher probability of crossing the BBB. From a study[9] of 329 oral drugs marketed during 1983−2002, CNS drug physicochemical properties were found to be significantly different from non-CNS drugs, with mean values of 310 (MW), 4.32 (O + N), 2.12 (HBA), 4.7 (RB). The recommended physicochemical property ranges provided by Hitchcock et al.[10] for enhanced BBB penetration were PSA < 90 Å^2, HBD < 3, cLogP [2,5], cLogD (pH 7.4) [2,5], and MW < 500.

Recent studies have attempted to improve the RO5 for CNS drug design. In a study of 119 CNS drugs and 108 CNS clinical candidates, Wager et al.[11,12] developed a concise algorithm called the "multiparameter optimization" (MPO) to predict desirable property space for various physicochemical properties of drug molecules. The median values for CNS drugs was as follows: MW = 305.3 Da, cLogP = 2.8, cLogD = 1.7, MW = 305.3, PSA = 44.8 Å^2, HBD = 1, and pKa = 8.4. They re-examined the MPO score in 2016 and found that different approaches for calculating log D and pKa values changes the MPO score; also, based on the method used for the calculation of physicochemical properties, MPO scores can vary significantly (as calculation of few MPO descriptors, i.e., log P, log D, HBD, and pKa can have different values based on software and algorithm selection).[13] The MPO score applies upper limits to every property used, but does not apply lower limits (e.g., clogP, clogD, MW and pKa), which increases the risk of populating the predicted MPO molecule space with very small molecules, which may be of less interest from a medicinal chemistry point of view. The MPO algorithm is based on CNS drugs (119) and CNS candidates (108) and does not include

Received: July 27, 2019
Published: October 11, 2019

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

non-CNS drugs, which may result in lack in efficiency in capturing the physiochemical nature of the blood−brain barrier. Gunaydin has recently published a probabilistic approach to MPO (pMPO) employing five physicochemical descriptors (TPSA, HBD, MW, cLogD, and basic pKa) for a data set of CNS and non-CNS drugs.[14] Ghose et al.[15] has proposed some CNS drug design property profile guidelines in their study of 317 CNS and 626 non-CNS oral drugs: TPSA < 76 Å^2 (25−60 Å^2), 740−970 Å^3 volume, N [1,2], linear chains outside of rings <7 [2,4], HBD < 3, (0,1), and SAS (460−580) Å^2, where the given ranges within the parentheses are preferred. They also proposed the technically extended multiparameter optimization (TEMPO)[16] scheme based on eight physicochemical properties with their preferred ranges "PL−PU" and qualifying ranges "QL-QU". Rankovic et al.[17] have updated the MPO score to MPO_V2, by mapping physicochemical properties of a diverse corporate data set of brain penetrant and peripherally restricted compounds. They discarded log D descriptors from MPO, and added double weight to the HBD contribution (MPO_V2: $\sum$T0 (cLogP, MW, TPSA, pKa, 2 HBD). However, this article has now been retracted.

In this work, we endeavor to understand how the individual physiochemical property spaces of CNS and non-CNS drugs influence each other. We also aim to devise a simple predictive computational model, termed BBB Score, to estimate whether a molecule is preferentially CNS vs non-CNS active. This latter aim is important in that most existing approaches focus on targeting only CNS drug penetration across the BBB.

2. RESULTS AND DISCUSSION

We studied the physicochemical property space of 22 molecular descriptors (Table S1, Supporting Information) and attempted to trial various combinations of these descriptors to produce a model that can identify the best possible correlation between BBB penetration and the physicochemical properties of a given molecule. (Detailed methodology of model development is given in Supporting Information and in the Methods section below.) The final BBB Score model comprises five descriptors: number of aromatic rings (Aro_R), number of heavy atoms (HA), MWHBN, i.e., $(\mathrm{HBN}/\sqrt{\mathrm{MW}})$ (a descriptor composed of molecular weight (MW), number of hydrogen bond donor atoms (HBD), and number of hydrogen bond acceptor atoms (HBA), where HBN = HBD + HBA), topological polar surface area (TPSA), and pKa at physiological pH. (Detailed information on pKa selection criteria for molecules having more than one ionizable groups available at physiological pH is given in Supporting Information Table S1).

Stepwise functions and polynomial piecewise functions were used to define property space of these descriptors. The graphs shown in Figure 1 represent the types of functions that $P(x)$ can adopt, with five descriptors chosen to model the BBB Score. Stepwise functions were used for descriptors that had discrete values and did not have continuous range of values (e.g., Aro_R). For other descriptors, (i.e., HA, MWHBN, TPSA, and pKa), a polynomial piecewise function was used to model $P(x)$. Note that the polynomial expressing $P(x)$ does not necessarily have to pass through the x axis, as $P(x)$ minima can be above 0, and $P(x)$ can output a value of 0 for any descriptor value outside the range of $[x_1, x_2]$.

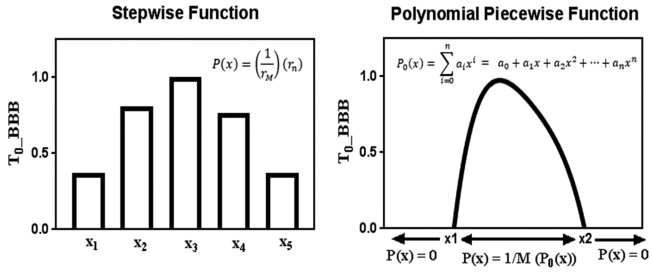


Figure 1. Graphical representation of stepwise function and polynomial piecewise functions utilized in the present work.

Table 1 lists all descriptors and their corresponding function used to develop the final BBB Score. To normalize the BBB Score results against the MPO results, which are presented as scores out of 6, the sum of contribution of five BBB Score descriptors is multiplied by 6/5 to adjust the final BBB Score output. CNS and non-CNS drugs correspond to a BBB Score range of [4,6] and [0,4), respectively. A stepwise function with corresponding discrete values for number of aromatic rings in the molecule is also shown; $P_{\mathrm{Aro\_R}}(x)$ outputs a perfect score of 1 for molecules having two aromatic rings. A cubic polynomial piecewise function is fitted to represent HA and MWHBN descriptors. A cubic polynomial for HA is defined in the range of heavy atoms between 5 and 45. A quadratic polynomial is employed for pKa. For TPSA, a linear equation is found to represent $P_{\mathrm{TPSA}}(x)$.

Figure 2 illustrates the relative distribution of values for each descriptor in the CNS and non-CNS sets and is used as a guidance tool for selecting various functions used to formulate the BBB Score. A stepwise function is used for Aro_R because the desired range for aromatic rings physicochemical property space is limited compared to other descriptors and is not continuous. The stacked bar graph of HA requires a cubic piecewise function to be fitted to correctly describe its property space within CNS and non-CNS drugs. An attempt to fit a higher order (>3) polynomial covering this range would have resulted in overfitting of the data, which was avoided to retain robustness of the resulting descriptor model equation. A cubic piecewise function was fitted to the bell-shaped stacked bar graphical representation of the MWHBN descriptor. (Graphical representations and fitted equations for the corresponding MW, HBA, and HBD contributor descriptors are given in the Supporting Information, along with graphical representations and fitted equations for the rest of the descriptors analyzed in our search for descriptors for an optimized BBB Score). CNS and non-CNS drug property space for TPSA followed a linear trend, peaking for CNS drugs in TPSA range (10, 20], which declined linearly until TPSA (120, 130]. Passive diffusion of small neutral molecules across the BBB is enabled and a significant percentage of weakly basic/acidic molecules exist in neutral form at physiological pH (7.4), which is reflected in the stacked bar graph of pKa distribution as it peaks at pKa ranges of (7,8] and (8,9] for CNS drugs. Relative percentage of CNS drugs decreases as we approach more acidic or basic pKa ranges. To consider the entire pKa property space for CNS and non-CNS drugs, we fit a quadratic equation encompassing the complete range of pKa values.

Delving further into such statistics for BBB Score descriptors (see Table 2), CNS data set's P10 and median values for Aro_R are higher as compared to the non-CNS data set, which reflects that most CNS drugs contain one or more aromatic

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

Table 1. BBB Score Descriptors, Functions, Weighting, and Descriptor Range

descriptor	function	weight	descriptor range
Aro_R	stepwise	1	$P_{\mathrm{ARO\_R}}(x) = \begin{cases} 0.336376, & x = 0 \\ 0.816016, & x = 1 \\ 1, & x = 2 \\ 0.691115, & x = 3 \\ 0.199399, & x = 4 \\ 0, & x > 4 \end{cases}$
HA	polynomial (cubic) piecewise function	1	$P_{\mathrm{HA}}(x) = \begin{cases} \left(\frac{1}{0.624231}\right)(0.0000443x^3 - 0.004556x^2 + 0.12775x - 0.463), & 5 < x \le 45 \\ 0, & x \le 5 \text{ or } x > 45 \end{cases}$
MWHBN	polynomial (cubic) piecewise function	1.5	$P_{\mathrm{MWHBN}}(x) = \begin{cases} \left(\frac{1}{0.72258}\right)(26.733x^3 - 31.495x^2 + 9.5202x - 0.1358), & 0.05 \le x \le 0.45 \\ 0, & x \le 0.05 \text{ or } x > 0.45 \end{cases}$
TPSA	polynomial (linear) piecewise function	2	$P_{\mathrm{TPSA}}(x) = \begin{cases} \left(\frac{1}{0.9598}\right)(-0.0067x + 0.9598), & 0 < x \le 120 \\ 0, & x = 0 \text{ or } x > 120 \end{cases}$
pKa	polynomial (quadratic) piecewise function	0.5	$P_{\mathrm{pKa}}(x) = \begin{cases} \left(\frac{1}{0.597488}\right)(0.00045068x^4 - 0.016331x^3 + 0.18618x^2 - 0.71043x + 0.8579), & 3 < x \le 11 \\ 0, & x \le 3 \text{ or } x > 11 \end{cases}$

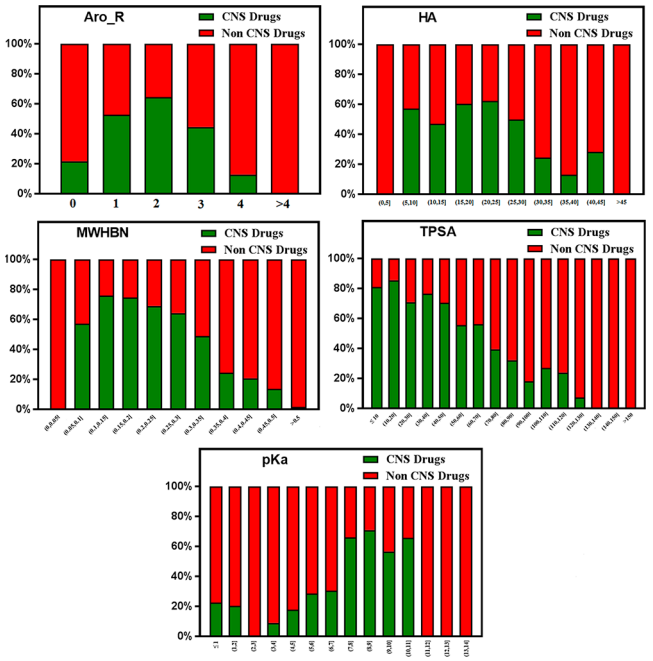


Figure 2. Stacked bar graphs for BBB descriptors (Aro_R, HA, MWHBN, TPSA, and pKa) for CNS and non-CNS databases. CNS and non-CNS data sets have been standardized for production of these graphs.

rings, which adds to lipophilicity and facilitates BBB penetration. However, by P75 and P90, CNS and non-CNS data sets appear to have similar property spaces regarding aromatic rings; in other words, distribution of CNS drugs is higher in property space with aromatic rings lower than 2, which can be interpreted as drugs with Aro_R less than 2 are more likely to be CNS penetrant. Lower value of standard error of the mean (SEM) suggests a good statistical distribution of data set, which is also statistically significant (two tailed Student's t test, $p = 0.054$). In comparison to Aro_R, the upper distribution of HA differentiates CNS and non-CNS data sets more efficiently. Non-CNS drugs are generally larger molecules compared to CNS drugs and thus have higher P75 and P90. Also, mean HA for the non-CNS data set (25.90) is higher than for the CNS drug data set (21.84), which is also statistically significant ($p = 3.19 \times 10^{-14}$).

Statistical analysis of CNS vs non-CNS data sets for MWHBN includes the statistical contribution from three descriptors: MW, HBD, and HBA. All values for the statistical descriptors for the non-CNS data set are higher than for CNS, and it is statistically significant ($p = 6.69 \times 10^{-55}$). The most pronounced differences between CNS and non-CNS data set statistics are seen for the pKa descriptor. P10 (6.63) and P25 (7.48) for CNS data set vs P10 (3.25) and P25 (4.32) for non-CNS data set reflects that CNS drug space is mostly populated with weak acids/bases enabling passive brain penetration, whereas highly acidic or protonated molecules belong to the non-CNS data set. The median and means are also considerably different. As we go to the upper distribution, we note that P75 and P90 are very similar. The observed SEM of 0.10 for the CNS data set is low, as compared to pKa ranges of 6.63 (P10) to 9.77 (P90). Similarly, compared to the pKa range for the non-CNS data set (3.25 (P10) to 9.76 (P90)), the SEM (0.13) is low. Non-CNS drugs have higher mean and median TPSA as compared to CNS drugs, which can be attributed to the preference of small molecules with less polar surface area for BBB penetration. The TPSA for CNS drugs

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

Table 2. Statistics for Descriptors Used for CNS and Non-CNS Databases for Modeling BBB Score

	Aro_R		HA		MWHBN		pKa		TPSA	
	CNS	non-CNS	CNS	non-CNS	CNS	non-CNS	CNS	non-CNS	CNS	non-CNS
P10	1	0	14.9	14	0.11	0.13	6.63	3.25	12.03	34.14
P25	1	1	18	18	0.15	0.24	7.48	4.32	23.47	55.82
median	2	1	22	25	0.21	0.36	8.40	6.99	41.75	87.13
P75	2	2	26	32	0.28	0.48	9.30	8.92	61.36	119.15
P90	3	3	29	39	0.34	0.66	9.77	9.76	81.06	168.88
N	270	720	270	720	270	720	217	450	270	720
mean	1.69	1.56	21.84	25.90	0.22	0.38	8.23	6.71	44.07	94.62
SEM	0.05	0.05	0.37	0.37	0.01	0.01	0.10	0.13	1.63	2.07
p-value	0.054		3.19×10^{-14}		6.69×10^{-55}		3.57×10^{-19}		2.75×10^{-69}	

ranges from 12.03 (P10) to 81.06 (P90), and for non-CNS drugs it ranges from 34.14 (P10) to 168.88 (P90), with statistically significant ($p = 2.75 \times 10^{-69}$) SEM of 1.63 and 2.07 for CNS and non-CNS drugs data set, respectively. Graphical representations of the stepwise (Aro_R) and piecewise polynomial functions (HA, MWHBN, TPSA, and pKa), used for the BBB Score descriptors are given in Figure 3.

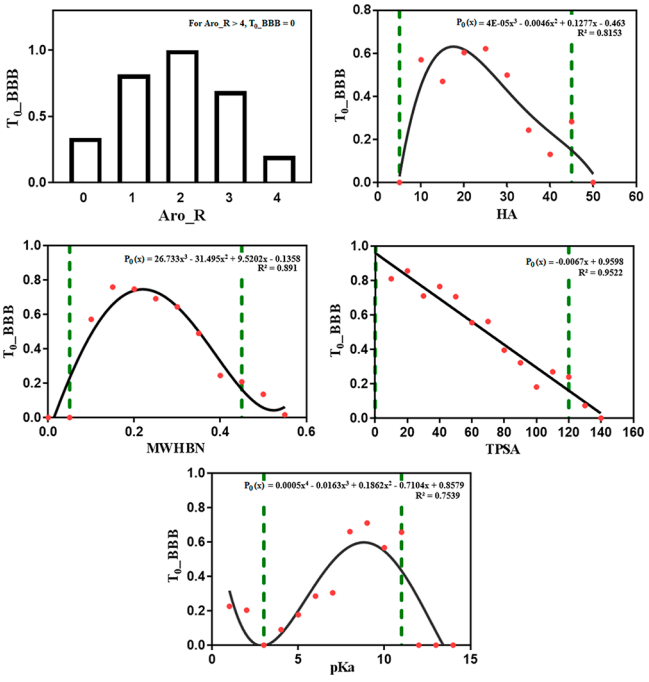


Figure 3. Stepwise and piecewise graphical representation of five descriptors used in BBB Score model, along with corresponding Stepwise function and fitted equations for Aro_R, HA, MWHBN, TPSA, and pKa, respectively.

Figure 4 represents 100% stacked bar graphs for low to high MPO, MPO_V2, and BBB Scores for standardized CNS and non-CNS drug data sets. Ideally, any molecule with an MPO, MPO_V2, and BBB Score in the range of [4,6] should be placed in the CNS data set; current scoring schemes can be interpreted as probability of a molecule attaining a score between 4 and 6 to be a CNS drug. MPO and MPO_V2 predicts 60.5% and 66.1% CNS drugs, respectively, in the range (5,6], which is very low compared to the 90.3% predicted by the BBB Score. For MPO and MPO_V2, CNS drugs in range (3,4] have poor predictability (nearly 50%) in differentiating between CNS and non-CNS drugs. The BBB Score predicts that 21.9% of CNS drugs score in the (3,4] range, which is an improvement over the MPO and MPO_V2 scores.

The graphs in Figure 5 represent sensitivity and specificity for each of the scoring systems. A significant problem with MPO and MPO_V2 is that they exhibit low specificity, i.e., high representation of molecules with scores above 4 for non-CNS. With the BBB Score, we were able to increase both sensitivity and specificity.

Two CNS drugs and two non-CNS drugs were randomly chosen to compare MPO, MPO_V2, TEMPO, and BBB Score predictions and to provide sample calculations of the BBB Score (Table 3). Various multiparameter optimization scores for penfluridol and cinnarizine were calculated. The BBB Score was successful in correctly characterizing penfluridol and cinnarizine as CNS drugs with a BBB score above 4 for both; MPO and MPO_V2 produced a score of less than 4. The two non-CNS drugs, tazobactam and ibandronate, were selected since tazobactam is an acidic drug whereas ibandronate contains a basic group. The BBB Scores for both of these non-CNS drugs are quite low, 1.82 and 1.8, respectively, as compared to other scoring systems. The MPO and MPO_V2 scores failed to correctly characterize tazobactam as a non-CNS drug, predicting a score of above 4. The TEMPO scoring system predicts CNS having ≤5 and non-CNS having >5 scores using a relatively complex algorithm. TEMPO predicts non-CNS drugs tazobactam and ibandronate correctly.

Table 4 illustrates BBB Score output as provided in the Supporting Information for the CNS drug cinnarizine. The provided calculator allows the user to input values for various physicochemical descriptors (aromatic rings, heavy atoms, molecular weight, hydrogen bond donor, hydrogen bond acceptor, TPSA, and pKa).

Table 5 outlines a performance evaluation for the MPO, MPO_V2, and BBB Scores for CNS and non-CNS data sets, with the same number of CNS and non-CNS drugs being analyzed for these scoring systems; the different statistical parameters are explained in the Methods section. The positive predictive value (PPV)[18] entails the percentage of drugs predicted to be CNS drugs that truly are CNS drugs; as the PPV value approaches 1, an increasing percentage of molecules is predicted correctly. The PPV value for the BBB Score is considerably higher than for the MPO and MPO_V2 Scores. Similarly, the negative predictive value (NPV)[19] describes the percentage of drugs predicted to be non-CNS drugs that truly are non-CNS drugs. The NPV value for the BBB Score is higher than for the MPO and MPO_V2 Scores. NPV value for these scores is higher as compared to the PPV, which is an artifact of the large non-CNS data set compared to the CNS

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

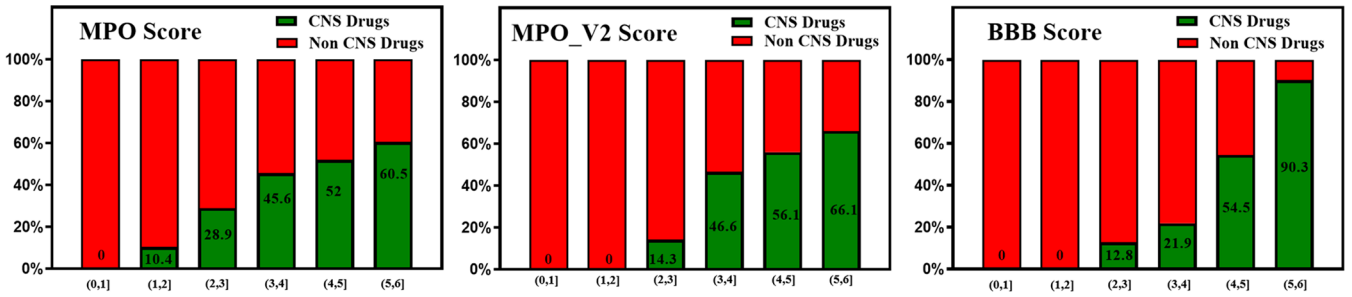


Figure 4. Percentage of CNS drugs vs non-CNS drugs was plotted on 100% stacked bar graph from low to high MPO, MPO_V2, and BBB Scores based on criteria that all molecules with MPO, MPO_V2, and BBB Scores within [0,4) are classified as non-CNS and all molecules within [4,6] are classified as CNS drugs. Number of molecules for CNS data set is 270, and non-CNS data set is 720. (SMILES strings are provided as CSV file in Supporting Information).

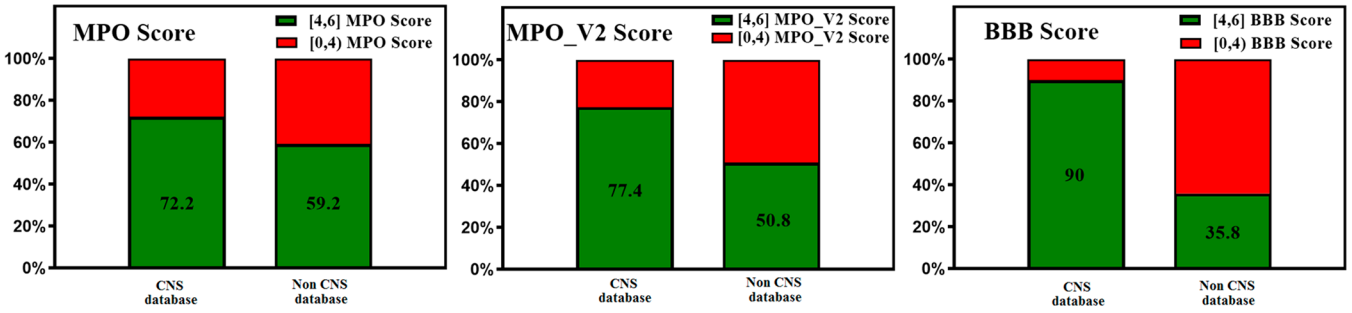
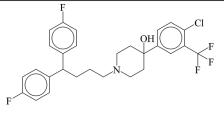
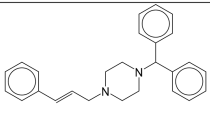
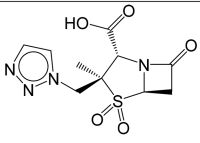
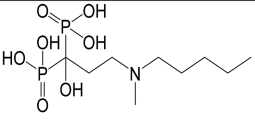


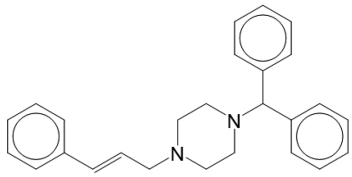
Figure 5. Percentage of CNS drugs and non-CNS drugs correctly identified (for CNS, MPO, MPO_V2, BBB Score [4,6]; for non-CNS, MPO, MPO_V2, BBB Scores [0,4)) in their respective CNS and non-CNS database is plotted on 100% stacked bar graph for MPO, MPO_V2, and BBB Scores. Number of molecules for CNS data set is 270 and non-CNS data set is 720. (SMILES strings are provided as CSV file in Supporting Information).

Table 3. BBB Score and BBB Composite Score for Selected CNS and Non-CNS Drugs

BBB Score and BBB Composite Score for Selected Drugs											
CNS Drugs						Non-CNS Drugs					
Penfluridol			Cinnarizine			Tazobactam			Ibandronate		
											
Property	Value	T0	Property	Value	T0	Property	Value	T0	Property	Value	T0
Aro_R	3	0.69	Aro_R	3	0.69	Aro_R	1	0.82	Aro_R	0	0.34
HA	36	0.48	HA	28	0.82	HA	20	1.00	HA	19	1.00
MW	523.97		MW	368.52		MW	300.29		MW	319.23	
HBA	2		HBA	2		HBA	7		HBA	8	
HBD	1		HBD	0		HBD	1		HBD	5	
MWHBN	0.13	0.87	MWHBN	0.10	0.75	MWHBN	0.46	0	MWHBN	0.73	0
TPSA	23.47	0.84	TPSA	6.48	0.95	TPSA	122.46	0	TPSA	138.53	0
pKa	8.96	1.00	pKa	8.1	0.97	pKa	2.86	0	pKa	9.92	0.93
MPO Score		1.53	**MPO Score**		2.89	**MPO Score**		4.83	**MPO Score**		3.04
MPO_V2 Score		2.36	**MPO_V2 Score**		3.88	**MPO_V2 Score**		4.67	**MPO_V2 Score**		2.04
TEMPO Score		8.45	**TEMPO Score**		7.74	**TEMPO Score**		6.61	**TEMPO Score**		12.69
BBB Score		4.65	**BBB Score**		5.04	**BBB Score**		1.82	**BBB Score**		1.8

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

Table 4. Interactive BBB Score Calculator[a] (where BBB Score = ∑T0[Aro_R, HA, (1.5) MWHBN, 2 (TPSA), (0.5) pKa])



Cinnarizine
(CNS Drug)

BBB Score Calculator		
Property	**Value**	**T_0**
Number of Aromatic Rings (Aro_R)	3	0.69
Number of Heavy Atoms (HA)	28	0.82
Molecular Weight (MW)	368.52	
Number of Hydrogen Bond Acceptor (HBA)	2	
Number of Hydrogen Bond Donor (HBD)	0	
MWHBN [MWHBN = (MW^(-0.5)*HBN), where HBN=HBA+HBD]	0.10	0.75
Topological Polar Surface Area(TPSA)	6.48	0.95
pKa	8.1	0.97
BBB SCORE		5.04

[a]Active BBB calculator is available in Supporting Information.

Table 5. Performance Evaluation of MPO, MPO_V2, and BBB Scores for CNS and Non-CNS Training Datasets

	MPO	MPO_V2	BBB
N_CNS drugs	270	270	270
N_non-CNS drugs	720	720	720
TP	195	209	243
TN	294	354	462
FP	426	366	258
FN	75	61	27
PPV	0.31	0.36	0.49
NPV	0.8	0.85	0.94
sensitivity	0.72	0.77	0.9
specificity	0.41	0.49	0.64
Kappa	0.09	0.2	0.42
MCC	0.12	0.23	0.48
AUC	0.61	0.67	0.86

data set. Sensitivity describes the portion of the CNS data set correctly being predicted as CNS drugs. Clearly, the BBB Score has a higher sensitivity (0.9) than the MPO (0.72) and MPO_V2 (0.77) Scores. Specificity presents the portion of the non-CNS drugs correctly being predicted as non-CNS. Specificity is significantly higher for the BBB Score compared to the MPO and MPO_V2 Scores. Specificity for the BBB Score is approximately 1.5 times that of the MPO score.

To further analyze MPO, MPO_V2, and BBB Score statistics, Kappa and Matthews correlation coefficients (MCC)[20] were calculated. The Kappa coefficient describes the probability of accurately segregating molecules adjusted by the probability of random prediction;[20] i.e., the Kappa value accounts for the possibility of a CNS molecule being classified as CNS (and non-CNS being classified as non-CNS) by chance. A Kappa value of ≥0.4 is a general consensus for a scoring system to be considered "substantial".[20] The BBB Score has a Kappa value of 0.42. The MCC reflects accuracy of a scoring system for classifying CNS and non-CNS drugs; a value closer to unity is better. The MCC measure of accuracy is reliable as it takes into account different class sizes.[20] The MCC value for the BBB Score is 0.48, which is better as compared to MPO (0.12) and MPO_V2 (0.23).

Figure 6 presents receiver operating characteristics (ROC) curves[21] for MPO, MPO_V2, and BBB Scores. ROC graphs are helpful for evaluating the performance of classifiers in a two-dimensional depiction. ROC curves can be interpreted as area under the ROC curves (AUC),[22,23] which is a quantitative

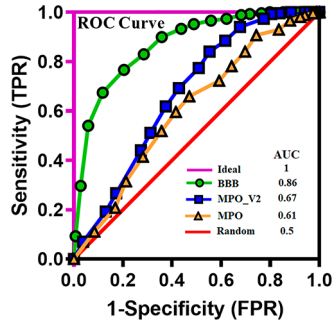


Figure 6. ROC curves for MPO, MPO_V2, and BBB Score.

single scalar value representing area of a unit square under ROC curve and ranges between 0 to 1, where an ideal classifier has an AUC of 1, and a classifier with almost no classification ability and random prediction has an AUC of 0.5; i.e., an AUC value of 0.7 would mean that a randomly selected CNS drug would have higher score in comparison to a randomly selected non-CNS drug 70% of the time,[24] and thus, the higher the AUC value, the closer the classification system is to perfect prediction. For plotting ROC curves, various cutoffs for the different scoring systems were considered. The true positive rate[25] (TPR) and false positive rate[25] (FPR) for each cutoff are calculated and plotted to produce Figure 6. AUC is calculated using the trapezoid method.[26] The trapezoids were drawn for each curve using two parallel bases of trapezoid parallel to the *y* axis, and height of trapezoid corresponds to difference in FPR value on the *x* axis. Total AUC is calculated by summation of area of all smaller trapezoids using adjacent points on the *x*-axis (FPR).

From Figure 6, we note that the MPO and MPO_V2 ROC curves are closer to the red curve which corresponds to random prediction. The ROC curve for the BBB Score is closer to ideal prediction.

Figure 7 shows predicted distributions of MPO, MPO_V2, and BBB Scores for CNS and non-CNS data sets, based on the assumption that the scores follow a normal (i.e., Gaussian) distribution.[23] Ideally, distributions for CNS and non-CNS agents would not overlap and would give no false predictions; the greater the overlap of the two curves (colored red and green), the worse the predictions. A dashed black vertical line signifies a threshold value of 4 for each score, above which are CNS predictions and below are non-CNS predictions. For the BBB Score, the overlap region containing false predictions (FP

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

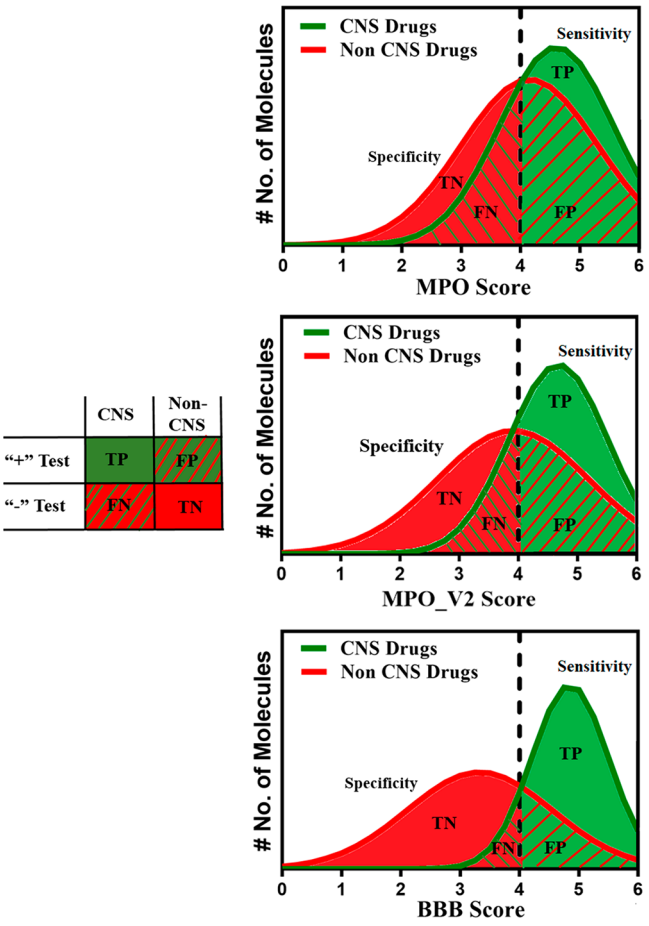


Figure 7. Normal distributions of MPO, MPO_V2, and BBB Scores for CNS and non-CNS data sets, where red curves and green curves correspond to CNS and non-CNS data sets, respectively. It is assumed that the MPO, MPO_V2, and BBB Scores for both CNS and non-CNS data sets have normal (i.e., Gaussian) distribution. The red and green colored areas represent FP and FN, respectively. A dashed black vertical line at score = 4 represents the threshold for each score for being characterized as CNS or non-CNS drug.

and FN) is considerably smaller as compared to the MPO and MPO_V2 score curves. These curves correlate with the performance evaluation statistics presented in Table 5 for MPO, MPO_V2 and BBB scores in which the BBB score has a higher percentage of correctly predicted CNS drugs (TP and TN) and a lower percentage of incorrect predictions (FP and FN) as compared to other scores. Figure 7 also visually accentuates the observation of high sensitivity and specificity for the BBB score.

Validation of MPO, MPO_V2, and BBB Score Is Done Using External Test Set of CNS and Non-CNS Data Set. A statistical analysis of a validation set of CNS and non-CNS drugs for the MPO, MPO_V2, and BBB Scores is given in Table 6. To be consistent, we kept the size of the CNS and non-CNS data sets equal. The BBB Score outperforms the MPO and MPO_V2 results.

The better predictive ability of BBB Score, in both training and validation sets, compared to MPO is mainly based upon few fundamental considerations of BBB Score formulation: (1) including non-CNS drugs into the formulation of BBB Score algorithm; (2) comparatively large and diverse data sets of CNS and non-CNS drugs compared to MPO Score; (3) careful curation of data sets, ensuring removal of all brain penetrant drugs with CNS indications and side effects from the non-CNS data set and, similarly, removal of all actively BBB transported drugs from the CNS data set; (4) a robust and powerful algorithm validated by different statistical methods.

Table 6. Statistical Analysis of a Validation Set of CNS and Non-CNS Drugs for the MPO, MPO_V2, and BBB Scores

	MPO	MPO_V2	BBB
N (CNS)	50	50	50
N (non-CNS)	50	50	50
TP	38	38	40
TN	19	21	36
FP	31	29	14
FN	12	12	10
PPV	0.55	0.56	0.74
NPV	0.61	0.63	0.78
sensitivity	0.76	0.76	0.8
specificity	0.38	0.42	0.72
Kappa	0.15	0.18	0.52
MCC	0.14	0.19	0.52
AUC	0.64	0.68	0.78

We performed a binary QSAR for our original data set of 270 CNS and 720 non-CNS drugs. The binary threshold was set at 0.5 for the model, as CNS drugs were represented by 1 and non-CNS drugs by 0. Thus, if a binary QSAR results higher or less than 0.5 for a molecule, it was termed as CNS and non-CNS drug, respectively. Energy optimization and QSAR modeling were performed employing the Molecular Operating Environment (MOE).[27] All available descriptors (more than 330 2D and 3D related to physicochemical, electronic, topologic, and geometric properties) in MOE were calculated and removed following a leave one out methodology to retain the most structure information dense descriptors in the final model. The final binary QSAR model comprised 10 descriptors: vsurf_Wp1 (0.081), ASA_H (0.069), vsurf_W5 (0.063), vsurf_D3 (0.060), pmi2 (0.058), SlogP_VSA0 (0.057), SMR_VSA4 (0.048), vol (0.044), PEOE_VSA-1 (0.039), SlogP_VSA3 (0.025), with their relative contribution in parentheses. Statistical analyses of these binary QSAR results are given in Table 7. By comparing the training set results of the binary QSAR with the statistical analysis results given for the BBB Score in Table 5, it can be seen that Kappa,

Table 7. Statistics for Binary QSAR for CNS and Non-CNS Drug Database

	MOE descriptor		BBB Score descriptor	
	training set	validation set	training set	validation set
# CNS drugs	270	50	270	50
# non-CNS drugs	720	50	720	50
TP	161	20	154	13
TN	622	43	676	46
FP	98	7	44	4
FN	109	30	116	37
PPV	0.62	0.74	0.78	0.76
NPV	0.85	0.59	0.85	0.55
sensitivity	0.6	0.4	0.57	0.26
specificity	0.86	0.86	0.94	0.92
Kappa	0.47	0.29	0.56	0.18
MCC	0.47	0.26	0.57	0.24
AUC	0.84	0.72	0.87	0.71

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

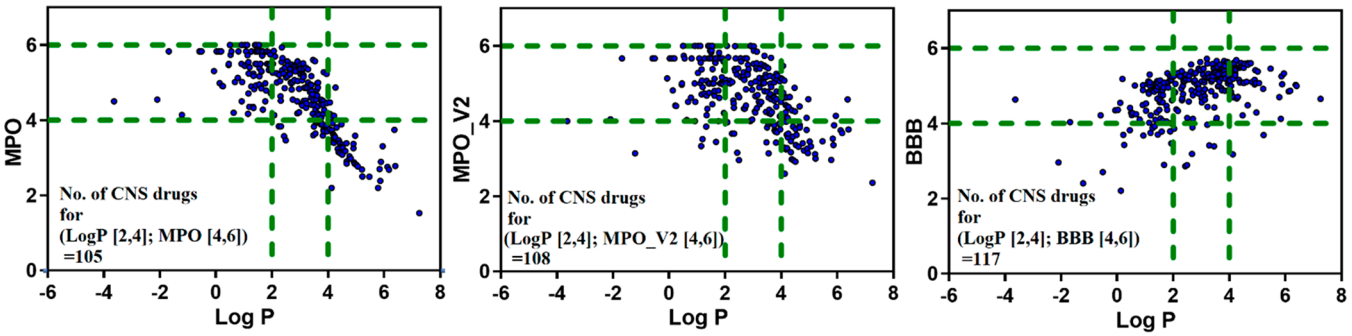


Figure 8. The log *P* vs MPO, log *P* vs MPO_V2, log *P* vs BBB Score property space for CNS drugs.

MCC, and AUC of the binary QSAR and BBB are quite similar. BBB Score sensitivity and NPV have significant improvement over binary QSAR results. However, specificity is better for the binary QSAR as compared with the BBB Score, but it should also be noted that the BBB Score depends upon calculation of five physicochemical descriptors that can be determined using various available software packages. The fact that accuracy for statistical results given by the BBB Score is similar or even better using fewer statistical parameters compared to a QSAR using ten sophisticated MOE descriptors further demonstrates the predictive power of the BBB Score model while retaining simplicity of calculation.

To understand an overlap of physicochemical properties information retained by BBB Score descriptors when used for developing a binary QSAR model vs sophisticated MOE descriptors, a binary QSAR model employing five BBB score descriptors (Aro_R, HA, MWHBN, TPSA, pKa) was developed. The descriptor importance is as follows: TPSA (0.235), pKa (0.107), Aro_R (0.085), HA (0.064), MWHBN (0.047). A comparison of statistical results of binary QSAR model employing MOE descriptors and BBB Score descriptors is given in Table 7. The binary QSAR model employing five BBB Score descriptors has similar or slightly better accuracy as compared to the one which employed 10 MOE descriptors for same CNS and non-CNS training data set. This further confirms that there is no significant information loss while using relatively simple BBB score descriptors as compared to complicated descriptors.

Significantly improved sensitivity, Kappa, MCC, AUC results for BBB Score validation set (Table 6) are observed when compared against binary QSAR models (using MOE descriptors and BBB score descriptors) validation set statistics given in Table 7. BBB Score algorithm validation set Kappa (0.52) and MCC (0.52) have more than twice the value of binary QSAR (employing BBB Score descriptors) validation set Kappa (0.18) and MCC (0.24), which ensures improved accuracy and predictive power of BBB Score algorithm as compared to binary QSAR models developed in this work.

The property spaces of log *P* vs MPO, log *P* vs MPO_V2, and log *P* vs BBB Score are plotted in Figure 8, where the area within log *P* [2,4] and MPO, MPO_V2, BBB [4,6] ranges is of particular interest (i.e., desired property space) for an optimal CNS drug. 105, 108, and 117 CNS drugs occupy this desired property space on the graphs for the MPO, MPO_V2 and BBB Scores. log *P* is not a descriptor of the BBB Score, yet it could predict a similar number of CNS drugs in desired property space as the MPO and MPO_V2 Scores. However, the importance of property space of molecules comprising log *P* [2,4] and MPO, MPO_V2, or BBB [4,6] in Figure 8 emphasizes the need for a correct classification system. Inaccurate predictions will unnecessarily populate the desired area with false positives, and true positives can be incorrectly filtered out. With the proven efficiency of the BBB Score, the population of molecules predicted by the BBB score in desired property space is likely to have better physiochemical properties for enabling brain penetration. While calculating BBB Scores, it is rare for descriptors of a molecule to hit maxima and attain a score of 1; BBB Scores are thus mainly clustered around a maximum value of 5.6.

Furthermore, the scatter plots (a) and correlation plots (b) between log *P* and number of aromatic rings (Aro_R) descriptors for 270 CNS and 720 non-CNS drugs show a strong correlation with R^2 of 0.97 and 0.95 for CNS and non-CNS data sets, respectively, as observed in the literature[28] (see Figure 9). However, it should be noted that standard deviation

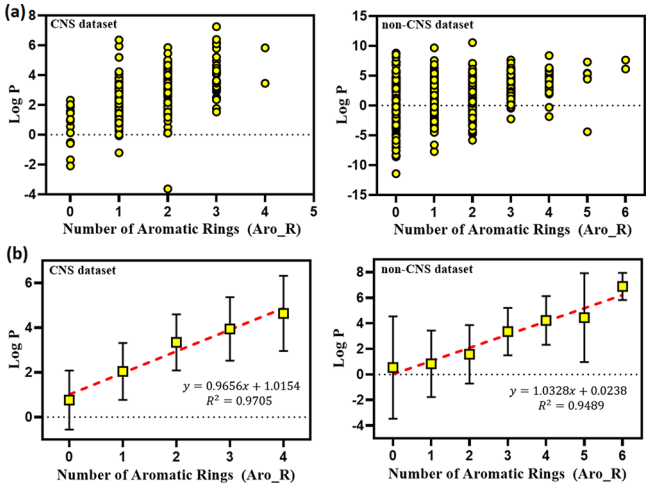


Figure 9. The scatter plots (a) and correlation plots (b) between log *P* and number of aromatic rings (Aro_R) descriptors for 270 CNS and 720 non-CNS drugs.

within log *P* values for discrete values of Aro_R is high. Addition of aromatic rings in a molecule results in increase in log *P*; however, addition of polar heterocyclic rings will increase topological polar surface area (TPSA) of the molecule with compromised log *P* value. For brain permeability, two aromatic rings or three aromatic rings (potentially involving a heterocyclic ring) would be preferable, with optimal balance of log *P* and TPSA. A correlation matrix among the selected BBB Score descriptors is given in Supporting Information (Table S7 and S8).

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

3. CONCLUSIONS

The strength of the BBB Score proposed here is its ability to characterize not only CNS drugs but also non-CNS drugs efficiently, which may prove to be of benefit for improving the medicinal chemist's ability to identify compounds that meet the target product profile. The ease of calculation of the BBB Score, employing BBB Score calculator and its efficient BBB penetration predictive capability as compared with available BBB penetration prediction algorithms, can aid in mapping the property space of drug data sets.

4. METHODS

4.1. Data Set Selection. To analyze the effect of various physicochemical properties on BBB penetration, a diversified data set of CNS and non-CNS drugs was assembled. The CNS and non-CNS drug data set collection was done in two steps.

The first step involved initial collection of CNS and non-CNS drugs from different sources. Supporting Information given by Leeson et al.[9] provided an exhaustive list of pre-1983 (864) and 1983−2002 (329) orally administered drugs. A FDA-approved drug library containing 2580 compounds along with their development status, indication, target, CAS number, pathway, solubility, SMILES, and physiochemical properties data was obtained from https://www.selleckchem.com/screening/fda-approved-drug-library.html. A CNS-penetrant compound library containing 307 compounds was also retrieved from the same Web site https://www.selleckchem.com/screening/cns-penetrant-compound-library.html. This library contains information about drug, pathway, target, SMILES, and selected physicochemical properties. DrugBank (http://www.drugbank.ca/), maintained by the University of Alberta (Edmonton, AB), was used to gather further information about drug structures and biological, biochemical, and pharmaceutical properties. To further expand this database, we included a compilation of FDA approved and CNS drugs derived from Drug Bank and covered by five insurance plans, found at the following link http://www.cureffi.org/2013/10/04/list-of-fda-approved-drugs-and-cns-drugs-with-smiles/. CNS and non-CNS molecules were also taken from various literature sources including Rankovic et al.[17] and Wager et al.[12] We have retained all drugs (including prodrugs) regardless of their route of administration (oral or intravenous infusion (IV)), as all drugs are in the bloodstream before crossing the BBB. (A list of commonly IV administered drugs and prodrugs from our data set, along with their BBB Score predictions, is given in Supporting Information Table S5 and Table S6, respectively.)

The second step involved CNS and non-CNS data set curation, which involved the following steps. First, data sets obtained from various sources for CNS and non-CNS drugs were merged together and duplicates were removed. For initial classification of CNS and non-CNS drugs, "drug category", "indication", "mechanism of action", and "pharmacodynamics" fields in DrugBank were used. Wikipedia (http://en.wikipedia.org/wiki/Main_Page), SciFinder (http://www.cas.org/products/scifinder), Drugs@FDA (http://www.accessdata.fda.gov/scripts/cder/drugsatfda), Pharmacogenomics Knowledge Base (http://www.pharmgkb.org) were referenced for ensuring consistency, when insufficient data were available on Drugbank. As many non-CNS drugs cross the BBB appreciably, the non-CNS data set was manually checked to remove molecules shown to have CNS effects or side effects at normal therapeutic doses. The CNS data set was also screened manually for eliminating molecules that are reported to follow active transport (employing various influx transporters) across the BBB. For example, actively transported compounds such as L-DOPA, glucose, gabapentin, and vitamins were removed. We sought to retain CNS drugs demonstrating only passive diffusion across the BBB in the CNS data set and non-CNS drugs having no observed or documented CNS side effects.

The final CNS and non-CNS training sets contained 270 and 720 molecules, respectively. The test set contained 50 molecules each for both the CNS and non-CNS data sets. Supporting Information contains a comprehensive table for CNS drugs having SMILES, drug name, BBB score, its indication, and mechanism of action. The non-CNS and CNS drug data sets along with SMILES strings, physiological properties calculated using MOE and Chem Axon softwares, MPO and BBB scores are also provided in the Supporting Information as a CSV file.

For the calculation of MPO, MPO_V2, and BBB Scores, 270 CNS drugs and 720 non-CNS drugs were used as a training set. In some cases, pKa values were not available from ChemAxon (CA) or MOE, as for example when drugs were present as mixtures of two or more compounds. These drugs were excluded while fitting polynomial function for pKa distribution of CNS and non-CNS data sets.

A total of 22 physicochemical descriptors were studied to probe the full chemical space of CNS and non-CNS drugs. The MPO score, given by Wager et al., gives a score from 0 to 6 for CNS-like drugs. However, the BBB Score has predictive powers for CNS and non-CNS drugs as it is also trained against a non-CNS data set. Molecular weight (MW), topological polar surface area (TPSA), log *P*, log *D*, pKa, hydrogen bond donor (HBD), carbon/heteroatom ratio (C:Hetero), first kappa shaped index (Kier1), second kappa shaped index (Kier2), third kappa shaped index (Kier3), atom connectivity index (chi1), carbon connectivity index (chi1_C), hydrogen bond acceptor (HBA), hydrogen bond number (HBN = HBD + HBA), number of rotational bonds (RB), square of log *P* ($(\log P)^2$), ($\log P/\sqrt{\mathrm{MW}}$), MWHBN ($\mathrm{HBN}/\sqrt{\mathrm{MW}}$), number of aromatic rings (Aro_R), number of heavy atoms (HA), van der Waals volume (vdw_V), area of van der Waals surface (vdw_A) were analyzed descriptors. The descriptors ($\log P/\sqrt{\mathrm{MW}}$), MWHBN ($\mathrm{HBN}/\sqrt{\mathrm{MW}}$) were chosen based upon simple linear regression literature studies, based upon brain-uptake index (BUI).[29−31] These descriptors were calculated using Molecular Operating Environment (MOE) and ChemAxon (CA) software. The physiochemical property space for each descriptor is mapped for the CNS and non-CNS drugs data sets, and an equation (cutoffs, linear or polynomial) was fitted as a boundary for CNS and non-CNS drug space. For fitting the polynomial equation for pKa, 217 CNS and 450 non-CNS drugs were used. To fit polynomial equations for the remaining descriptors, 270 CNS and 720 non-CNS drugs were used. For testing the validation of the BBB Score, both CNS and non-CNS data sets were composed of 50 drugs.

4.2. Calculation of Physiochemical Properties for CNS and Non-CNS Data Set Molecules. Table S1 in the Supporting Information lists all 22 descriptors used in the current work to map the physiochemical property space of CNS and non-CNS drugs. A brief description of the descriptors used are also provided in the table. MPO and

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

BBB Scores can vary for a molecule if different software packages are used to calculate these descriptors. However, this error is mitigated in the BBB Score as it does not have log P and log D descriptors, which vary significantly based on the algorithm applied for their calculation. The BBB Score algorithm has three descriptors (HBD, HBA, pKa) vs MPO which has four descriptors (HBD, pKa, log P, log D), which have varying values depending upon their calculation method. We have employed ChemAxon (Marvin Sketch) for calculation of pKa, HBD, and HBA, which is a freely available software with reliable accuracy. This is not a requirement for using the BBB Score, but we recommend the users consistently utilize the same software package for calculating particular descriptors when comparing brain penetrance of different molecules. Property space for pKa vs TPSA, TPSA vs MW and graphs of MPO Score for CNS and non-CNS data sets, calculated using MOE vs CA descriptors, are also given in the Supporting Information.

4.3. Model Development. *4.3.1. Representation of Descriptors Using Piecewise or Stepwise Function.* Each descriptor (except HBD, HBA, Aro_R) was fitted to a polynomial equation. The range was calculated for CNS drugs and divided into intervals such that at least 10 different intervals compared the ratios of CNS to non-CNS drugs with descriptor values within that interval. These ratios were taken as points for the graph where (x, y) = (highest value of interval, ratio of CNS to (CNS + non-CNS) on the Cartesian plane. The size of the CNS and the size non-CNS sets were standardized when calculating the ratios by dividing the number of drugs within that calculated descriptor range by the total number of drugs (separate for CNS and non-CNS). After plotting, the curve/polynomial of best fit was created with GraphPad Prism Software. The equation was chosen based on [i] the equation having a degree less than 5 (to avoid overfitting), [ii] how well it fits the plotted points and if the curve was "simplistic" when observing qualitatively (making sure the curve was not all over the place and had a relatively simply described trend). A higher degree polynomial was employed in case there was a significant difference between the shape of the curve or its R^2 value was higher, compared against a lower degree polynomial. After, a piecewise function $P(x)$ was defined for each descriptor with the general equation being

$$P(x) = \begin{cases} \left(\frac{1}{M}\right)(P_0(x)), & x_1 < x \le x_2 \\ 0, & x \le x_1 \text{ or } x > x_2 \end{cases} \quad (1)$$

where

$$P_0(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \quad (2)$$

M = local max of $P_0(x)$ between (x_1, x_2). This was so that the polynomial should have a local max of 1. The cutoff points (x_1 and x_2) were deduced by seeing if there was little or no representation of CNS drugs below x_1 and above x_2. The ranges may differ based on different descriptors (e.g., inclusion of x_1 for the polynomial instead of outputting 0).

For descriptors HBD, HBA, and Aro_R (having discrete values), a continuous equation was not used but mimicked more of a stepwise function.

$$P(x) = \begin{cases} \left(\frac{1}{r_M}\right)(r_0), & x = 0 \\ \left(\frac{1}{r_M}\right)(r_1), & x = 1 \\ \vdots \quad\quad \vdots & \\ \left(\frac{1}{r_M}\right)(r_n), & x = n \\ 0, & x > n \end{cases} \quad (3)$$

r_n = ratio of CNS to total (CNS + non-CNS) drugs when the descriptor value is n, r_M = the largest CNS to total (CNS + non-CNS) drugs ratio out of $r_0, r_1, \ldots, r_n$.

Finally, a combination of descriptors was chosen to create a final BBB Score

$$\text{BBB Score} = \left(\frac{6}{n}\right) \sum_{i=1}^{n} P_i(x_i) \quad (4)$$

n = number of descriptors used, P_i = piecewise equation of descriptor where x_i is the calculated value for that descriptor.

4.3.2. Python Script for Evaluating Various Models (Containing 5−10 Descriptors, with Equal Weightage) Performance. The code describing a script for creating five descriptor models with at least an 80% TPR, 65% TNR, and 87.5% representation of CNS drugs between molecules that output a score between 5 and 6 is given in Supporting Information as Python script S1. After the code was run, the final models were compared based on statistical results to determine the three best models. The code was also modified for generating a model containing 6−10 descriptors accordingly. Forty-four text files (22 for both CNS and non-CNS data sets) were prepared before running the code where each file contains a list of $P(x)$ values for each individual descriptor (the piecewise equation outputs a score between 0 and 1 for each descriptor). These lists are ordered according to compound (e.g., the fifth value from each CNS list represents the same molecule). The BBB score models (BBB Score V1, BBB Score V2) with equal weights are given in Supporting Information pp S39−S42.

4.3.3. Python Script for Adding Different Weights to Descriptors of the Five Descriptor BBB Model. The code describing a script that evaluates different weightings for a five descriptor model that we have chosen (Aro_R, HA, MWHBN, pKa, TPSA) from the three best models selected from step 2.3.2 is given in Supporting Information as Python script S2. It outputs a list of models as described in 2.3.2, and few models were chosen manually to be compared against the one with equal weights. The BBB Score model (BBB Score V3) with varying weights final equation has been given in Supporting Information pp S43−S44. We also endeavored to apply hard cutoffs to descriptor equations but were unable to improve the performance of the model. Additional explanation and performance of this model is given in Supporting Information pp S45−S47. Piecewise and stepwise functions "$P(x)$" used for the 22 descriptors and graphical representation of distribution of CNS and non-CNS data set for corresponding descriptor values are given in Supporting Information.

4.4. Statistical Significance of Results. *4.4.1. Student t Test.* Two tailed Student's t test has been applied to all 22

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

descriptors analyzed to determine statistical significance of each descriptor for CNS and non-CNS database. The results for the five descriptors used to make the BBB Score are given in Table 2.

4.4.2. Receiver Operating Curves (ROC Curves). To analyze the accuracy of various scoring models, ROC curves were plotted. To plot these curves and to analyze performance of different CNS penetration scoring algorithms, true positive (TP), false positive (FP), true negative (TN), false negative (FN), positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity, true predictive rate (TPR), false predictive rate (FPR), Kappa and Matthews correlation coefficients (MCC) were calculated as explained below.

TP represents number of CNS drugs correctly identified as CNS (i.e., MPO, MPO_V2, and BBB Scores in CNS data set is within [4,6]), TN represents number of non-CNS drugs correctly identified as non-CNS (i.e., MPO, MPO_V2, and BBB Score in non-CNS data set is within [0,4)), FP represents number of non-CNS drugs incorrectly identified as CNS (i.e., MPO, MPO_V2, and BBB Scores in non-CNS data set is within [4,6]), and FN represents number of CNS drugs incorrectly identified as non-CNS (i.e., MPO, MPO_V2, and BBB Scores in CNS data set is within [0,4)).

	CNS Drug	Non CNS Drug	
Positive Test (BBB Score > 4)	TP	FP	$PPV = \frac{TP}{TP + FP}$
Negative Test (BBB Score < 4)	FN	TN	$NPV = \frac{TN}{TN + FN}$
	$Sensitivity = \frac{TP}{TP + FN}$	$Specificity = \frac{TN}{FP + TN}$	

Positive predictive value (PPV) and negative predictive value (NPV) are given as

$$PPV = \frac{TP}{TP + FP} \tag{5}$$

$$NPV = \frac{TN}{TN + FN} \tag{6}$$

Sensitivity and specificity are given as

$$\text{sensitivity} = \text{true positive rate (TPR)} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{specificity} = \frac{TN}{FP + TN}, \quad \text{false positive rate (FPR)} = 1 - \text{specificity} = \frac{FP}{FP + TN} \tag{8}$$

Kappa and Matthews correlation coefficients, ranging within −1 to +1 (where 1 means perfect prediction, 0 means random guessing, and −1 mean perfect disagreement), demonstrates the probability of accurately classifying molecules modified by the probability of random agreement. Typically for a classification model to be considered "moderate" or "substantial", Kappa or Matthews correlation coefficients should have a value of 0.4.

$$\text{Kappa} = \frac{\frac{TP + TN}{TP + FP + FN + TN} - \frac{(TP + FP)(TP + FN) + (TN + FN)(TN + FP)}{(TP + FP + TN + FN)^2}}{1 - \frac{(TP + FP)(TP + FN) + (TN + FN)\cdot(TN + FP)}{(TP + FP + TN + FN)^2}} \tag{9}$$

$$\text{MCC} = \frac{TP\cdot TN - FN\cdot FP}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \tag{10}$$

Receiver operating curves (ROCs) and area under the curve (AUC) were calculated using Microsoft Excel. MPO, MPO_V2, and BBB Scores cutoffs were taken from 0 to 6 in 0.25 intervals. For each cutoff, any score above or equal to cutoff indicates the drug is CNS and below the cutoff would indicate the drug is non-CNS. For each cutoff, the TPR and FPR were calculated and they were plotted as a connected line graph. The area under the curve was calculated using the trapezoid method. AUC reflects how well the MPO, MPO_V2, and BBB Score algorithm can predict a CNS drug with a higher value than non-CNS (the higher the score, more likely to be CNS) where AUC = 0.5 means random guessing and 1 is perfect prediction.

For plotting normal distribution curves for MPO, MPO_V2, and BBB Score, as given in Figure 7, a function "NORM.DIST" was employed in Excel by inputting mean and standard deviation for CNS and non-CNS data set. Score range of 0−6 is divided into intervals of 0.25, and the result from normal distribution function for each range is plotted as points on the graph.

4.5. Binary QSAR. Gas phase geometry optimization of the training and test set molecules for the binary QSAR was done using the Chemistry at Harvard Macromolecular Mechanics (CHARMM) force field[32] as implemented in the MOE. Initial input structures (SMILES strings) of the training and test set molecules were obtained from structural databases (as explained in section 4.1 above) or generated with a model building program (builder in MOE). Optimization in MOE uses three sequential algorithms of steepest descent, conjugate gradient, and truncated Newton to identify a minimum on the potential energy surface. The list of molecules used for the training and test sets and the detailed binary QSAR results can be found in the Supporting Information CSV file.

4.6. User Friendly BBB Score Calculator Excel Sheet (BBB Score Calculator Tool). A user-friendly BBB Score calculator is given in the Supporting Information, as shown in Table 4. The BBB Score for a molecule can be calculated by inputting the number of aromatic rings, heavy atoms, molecular weight, hydrogen bond donors, hydrogen bond acceptors, topological polar surface area, and pKa for a molecule. The BBB calculator then outputs the MWHBN value and BBB Scores. An example of the BBB Score calculation is given in the Supporting Information.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.9b01220.

Active BBB Score calculator (XLSX)

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836

List of physiochemical descriptors used to map property space of CNS and non-CNS drug databases; piecewise and stepwise functions "$P(x)$" used for each descriptor and graphical representation of distribution of CNS and non-CNS data set for particular descriptor values; statistics for different descriptors used for CNS and non-CNS databases for modeling BBB Score; Python script for finding best model using descriptors from pool of 22 descriptors with equal weight; Python script for adding different weights to descriptors of the five descriptor BBB model; BBB Score calculation having equal weightage of final BBB Score descriptors (BBB Score V1); BBB Score calculation having equal weightage of descriptors chosen from a pool of 22 descriptors (BBB Score V2); BBB Score calculation having different weightage of final BBB Score descriptors (BBB Score V3); BBB Score using hard cutoffs for $P(x)$ for each descriptor, MPO Score calculation for a set of CNS and non-CNS data set using descriptor's values (LogP, LogD, TPSA, pKa, HBD) from Molecular Operating Environment (MOE) and ChemAxon (Marvin Sketch) softwares; property space for pKa vs TPSA and TPSA vs MW; list of commonly intravenous infusion (IV) administered drugs from CNS and non-CNS data sets, along with their BBB Score predictions; list of prodrugs from CNS and non-CNS data sets, along with their BBB Score predictions; correlation matrix for BBB Score descriptors, results of one-way ANOVA, and Sidak's multiple comparison test; correlation matrix for BBB Score descriptors (weighted to BBB Score algorithm); CNS drug training data set with SMILES strings, BBB Scores, indication, mechanism of action; CNS drug test data set with SMILES strings, BBB Scores, indication, mechanism of action (PDF)

Molecular formula strings for databases of CNS and non-CNS drugs with calculated physiochemical properties (CSV)

List of molecules and detailed results for binary QSAR training and test sets (CSV)

AUTHOR INFORMATION

Corresponding Author

*E-mail: dweaver@uhnres.utoronto.ca.

ORCID

Mayuri Gupta: 0000-0001-8588-1262

Donald F. Weaver: 0000-0003-2042-6220

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge many helpful discussions with Dr. Fan Wu. The authors thank the Krembil Foundation for financial support. D.F.W. thanks the Canada Research Chair program for support as a Tier I CRC in Protein Misfolding Diseases.

ABBREVIATIONS USED

BBB, blood–brain barrier; CNS, central nervous system; MPO, multiparameter optimization; Aro_R, aromatic rings; HA, heavy atom; MW, molecular weight; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HBN, hydrogen bond number (i.e., HBA + HBD); TPSA, topological polar surface area; ROC, receiver operating curve; AUC, area under curve; QSAR, quantitative structure–activity relationship; TEMPO, technically extended multiparameter optimization; TPR, true positive rate; FPR, false positive rate; MOE, molecular operating environment; CA, ChemAxon; MCC, Matthews correlation coefficient

REFERENCES

(1) Abbott, N. J.; Patabendige, A. A. K.; Dolman, D. E. M.; Yusof, S. R.; Begley, D. J. Structure and Function of the Blood–Brain Barrier. *Neurobiol. Dis.* **2010**, *37*, 13–25.

(2) Di, L.; Kerns, E. H. *Blood-Brain Barrier in Drug Discovery: Optimizing Brain Exposure of CNS Drugs and Minimizing Brain Side Effects for Peripheral Drugs*; John Wiley& Sons Inc.: Hoboken, NJ, 2015.

(3) Andersen, P. H.; Moscicki, R.; Sahakian, B.; Quirion, R.; Krishnan, R.; Race, T.; Phillips, A. Securing the Future of Drug Discovery for Central Nervous System Disorders. *Nat. Rev. Drug Discovery* **2014**, *13* (12), 871–872.

(4) Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An Analysis of the Attrition of Drug Candidates from Four Major Pharmaceutical Companies. *Nat. Rev. Drug Discovery* **2015**, *14*, 475–486.

(5) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

(6) Van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Chretien, J. R.; Raevsky, O. A. Estimation of Blood-Brain Barrier Crossing of Drugs Using Molecular Size and Shape, and H-Bonding Descriptors. *J. Drug. Target.* **1998**, *6*, 151–165.

(7) Kelder, J.; Grootenhuis, P. D. J.; Bayada, D. M.; Delbressine, L. P. C.; Ploemen, J.-P. Polar Molecular Surface as a Dominating Determinant for Oral Absorption and Brain Penetration of Drugs. *Pharm. Res.* **1999**, *16*, 1514–1519.

(8) Norinder, U.; Haeberlein, M. Computational Approaches to the Prediction of the Blood–Brain Distribution. *Adv. Drug Delivery Rev.* **2002**, *54*, 291–313.

(9) Leeson, P. D.; Davis, A. M. Time-Related Differences in the Physical Property Profiles of Oral Drugs. *J. Med. Chem.* **2004**, *47*, 6338–6348.

(10) Hitchcock, S. A.; Pennington, L. D. Structure-Brain Exposure Relationships. *J. Med. Chem.* **2006**, *49*, 7559–7583.

(11) Wager, T. T.; Chandrasekaran, R. Y.; Hou, X.; Troutman, M. D.; Verhoest, P. R.; Villalobos, A.; Will, Y. Defining Desirable Central Nervous System Drug Space Through the Alignment of Molecular Properties, in Vitro ADME, and Safety Attributes. *ACS Chem. Neurosci.* **2010**, *1*, 420–434.

(12) 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*, 435–449.

(13) 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*, 767–775.

(14) Gunaydin, H. Probabilistic Approach to Generating MPOs and Its Application as a Scoring Function for CNS Drugs. *ACS Med. Chem. Lett.* **2016**, *7*, 89–93.

(15) 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.

(16) Ghose, A. K.; Ott, G. R.; Hudkins, R. L. Technically Extended MultiParameter Optimization (TEMPO): An Advanced Robust Scoring Scheme To Calculate Central Nervous System Druggability

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824–9836

and Monitor Lead Optimization. *ACS Chem. Neurosci.* **2017**, *8*, 147−154.
(17) Rankovic, Z. CNS Physicochemical Property Space Shaped by a Diverse Set of Molecules with Experimentally Determined Exposure in the Mouse Brain. *J. Med. Chem.* **2017**, *60*, 5943−5954; Retraction notice: *J. Med. Chem.* **2019**, *62* (3), 1699.
(18) Blakely, T.; Salmond, C. Probabilistic Record Linkage and A Method to Calculate the Positive Predictive Value. *Int. J. Epidemiol.* **2002**, *31*, 1246−1252.
(19) Bewick, V.; Cheek, L.; Ball, J. Statistics Review 13: Receiver Operating Characteristic Curves. *Crit. Care.* **2004**, *8*, 508−512.
(20) Dolgikh, E.; Watson, I. A.; Desai, P. V.; Sawada, G. A.; Morton, S.; Jones, T. M.; Raub, T. J. QSAR Model of Unbound Brain-to-Plasma Partition Coefficient, Kp, uu, brain: Incorporating P-glycoprotein Efflux as a Variable. *J. Chem. Inf. Model.* **2016**, *56*, 2225−2233.
(21) Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861−874.
(22) Bradley, A. P. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recogn* **1997**, *30*, 1145−1159.
(23) Hanley, J. A.; McNeil, B. J. The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve. *Radiology* **1982**, *143*, 29−36.
(24) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* **2005**, *48*, 2534−2547.
(25) Davis, J.; Goadrich, M. The Relationship between Precision-Recall and ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA, USA. 2006; ACM: New York, 2006; pp 233−240.
(26) Zhang, D. D.; Zhou, X.-H.; Freeman, D. H.; Freeman, J. L. A Non-Parametric Method for the Comparison of Partial Areas Under ROC Curves and its Application to Large Health Care Data sets. *Stat. Med.* **2002**, *21*, 701−715.
(27) *Molecular Operating Environment (MOE)*, version 2013.08; Chemical Computing Group ULC (1010 Sherbooke St. West, Suite 910, Montreal, QC H3A 2R7, Canada), 2018.
(28) Ritchie, T. J.; Macdonald, S. J. The Impact of Aromatic Ring Count on Compound Developability- Are too Many Aromatic Rings a Liability in Drug Design? *Drug Discovery Today* **2009**, *14*, 1011−1020.
(29) Burns, J.; Weaver, D. F. A Mathematical Model for Prediction of Drug Molecule Diffusion across the Blood-Brain Barrier. *Can. J. Neurol. Sci.* **2004**, *31*, 520−527.
(30) Cornford, E. M.; Braun, L. D.; Oldendorf, W. H.; Hill, M. A. Comparison of Lipid-Mediated Blood-Brain Barrier Penetrability in Neonates and Adults. *Am. J. Physiol.* **1982**, *243* (3), C161−C168.
(31) Bezek, S.; Trnovec, T.; Scasnar, V.; Durisova, M.; Kukan, M.; Kallay, Z.; Laginova, V.; Svoboda, V. Irradiation of the Head by ^{60}Co Opens the Blood-Brain Barrier for Drugs in Rats. *Experientia* **1990**, *46*, 1017−1020.
(32) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

DOI: 10.1021/acs.jmedchem.9b01220
J. Med. Chem. 2019, 62, 9824−9836