Supporting Information

for

To Pass or Not To Pass: Predicting the Blood-Brain Barrier Permeability with the 3D-RISM-KH Molecular Solvation Theory

Dipankar Roy,[1] Vijaya Kumar Hinge,[1] Andriy Kovalenko[1,2,]*

[1] Department of Mechanical Engineering, University of Alberta, 10-203 Donadeo Innovation Centre for Engineering, 9211-116 Street NW, Edmonton, Alberta T6G 1H9, Canada

[2] Nanotechnology Research Centre, 11421 Saskatchewan Drive, Edmonton, Alberta T6G 2M9, Canada

Descriptors used for *Model A*:

Descriptor	Details	*Descriptor*	Details
TPSA	Topological Polar Surface Area	*KierFlex*	Molecular flexibility index
lip_acc	Lipinski H-bond acceptor count	*mr*	Molar refractivity
HOMO	Highest occupied molecular orbital	*weinerPath*	weiner path number
Dipole_Moment	Dipole Moment	*pmv2*	Partial molar volume in cyclohexane
lip_don	Lipinski H-bond donor count	*weinerPol*	weiner polarity number
logS	aqueous solubility descriptor	*chi0*	connectivity descriptor
x4	Excess chemical potential in n-octanol	*a_heavy*	Total number of heavy atoms
LUMO	Lowest unoccupied molecular orbital	*a_count*	Total number of atoms
a_hyd	Number of hydrogen atoms	*vdw_vol*	van der Wall's volume
balabanJ	Balaban's J Index	*pmv3*	Partial molar volume in n-hexadecane
pmv4	Partial molar volume in n-octanol	*vdw_area*	van der Wall's area
ASA	Accessible surface area	*a_aro*	Total number of aromatic atoms
zagreb	Zagreb index	*pmv1*	Partial molar volume in chloroform
x2	Excess chemical potential in cyclohexane	*x3*	Excess chemical potential in n-hexadecane
Weight	Molecular weight	*pmv5*	Partial molar volume in water
bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens)	*Apol*	Sum of the atomic polarizabilities (including implicit hydrogens)
rings	ring count	*McGowan_Volume*	Characteristic McGowan volume
x5	Excess chemical potential in n-hexadecane	*PetitjeanNumber*	Petitjean Number
x1	Excess chemical potential in chloroform		

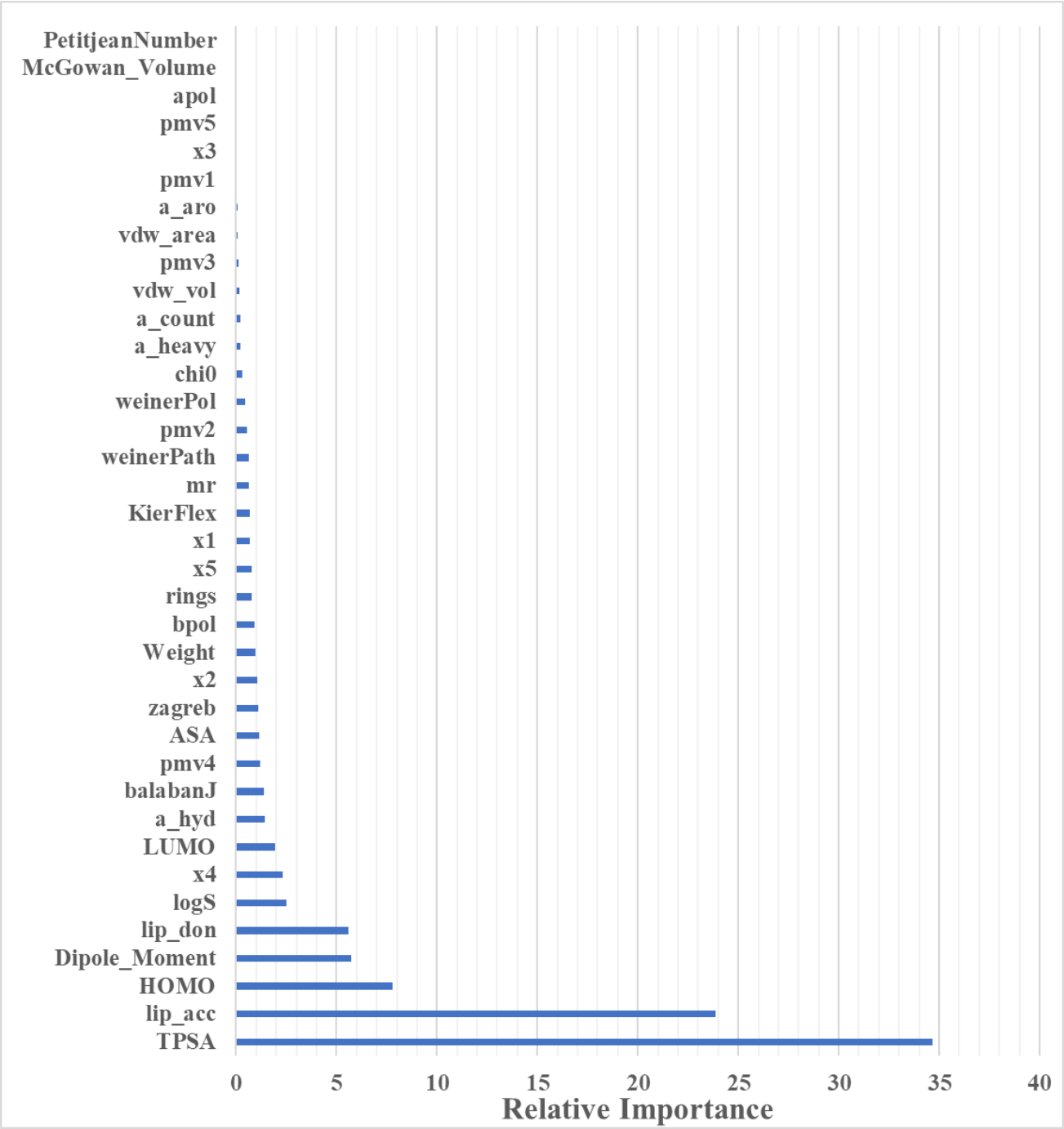

Figure S1: Relative importance (influence) of different variables in *Model A* obtained from GBM analysis.

Descriptors used for *Model B*:

Descriptor	Details	*Descriptor*	Details
TPSA	Topological Polar Surface Area	*x4*	Excess chemical potential in n-octanol
lip_acc	Lipinski H-bond acceptor count	*pmv4*	Partial molar volume in n-octanol
HOMO	Highest occupied molecular orbital	*x2*	Excess chemical potential in cyclohexane
Dipole_Moment	Dipole Moment	*Weight*	Molecular weight
lip_don	Lipinski H-bond donor count	*chi0*	connectivity descriptor
logS	aqueous solubility descriptor		

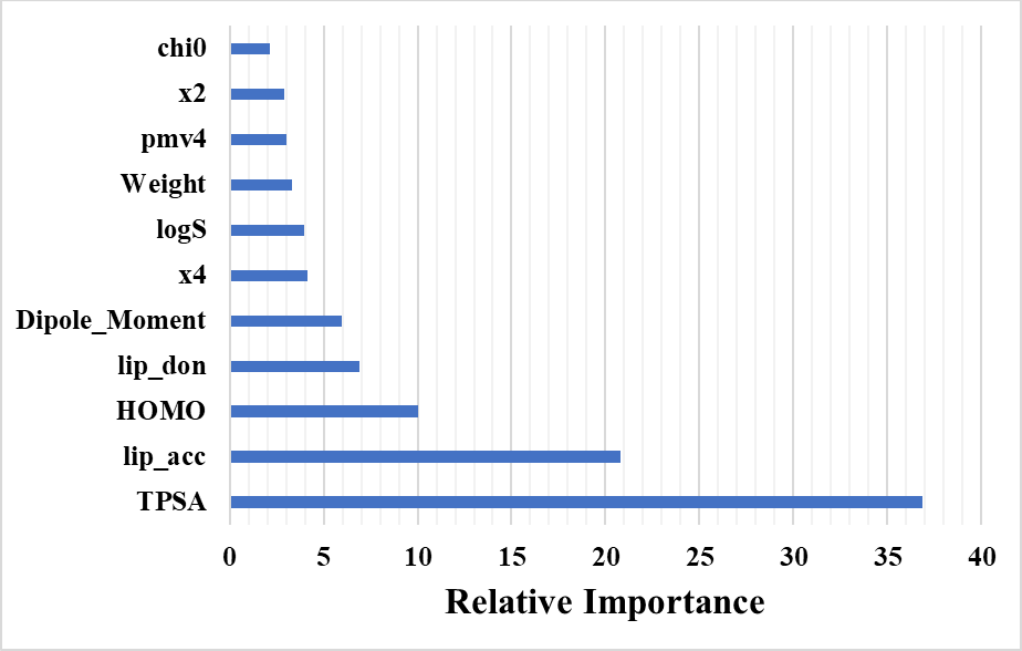

Figure S2: Relative importance (influence) of different variables in *Model B* obtained from GBM analysis.

Descriptors used for *Model C*:

Descriptor	Details	*Descriptor*	Details
TPSA	Topological Polar Surface Area	*KierFlex*	Molecular flexibility index
lip_acc	Lipinski H-bond acceptor count	*x2*	Excess chemical potential in cyclohexane
HOMO	Highest occupied molecular orbital	*balabanJ*	Balaban's J Index
Dipole_Moment	Dipole Moment	*x4*	Excess chemical potential in n-octanol
lip_don	Lipinski H-bond donor count	*LUMO*	Lowest unoccupied molecular orbital
GapHL	HOMO-LUMO energy gap (in ev)		

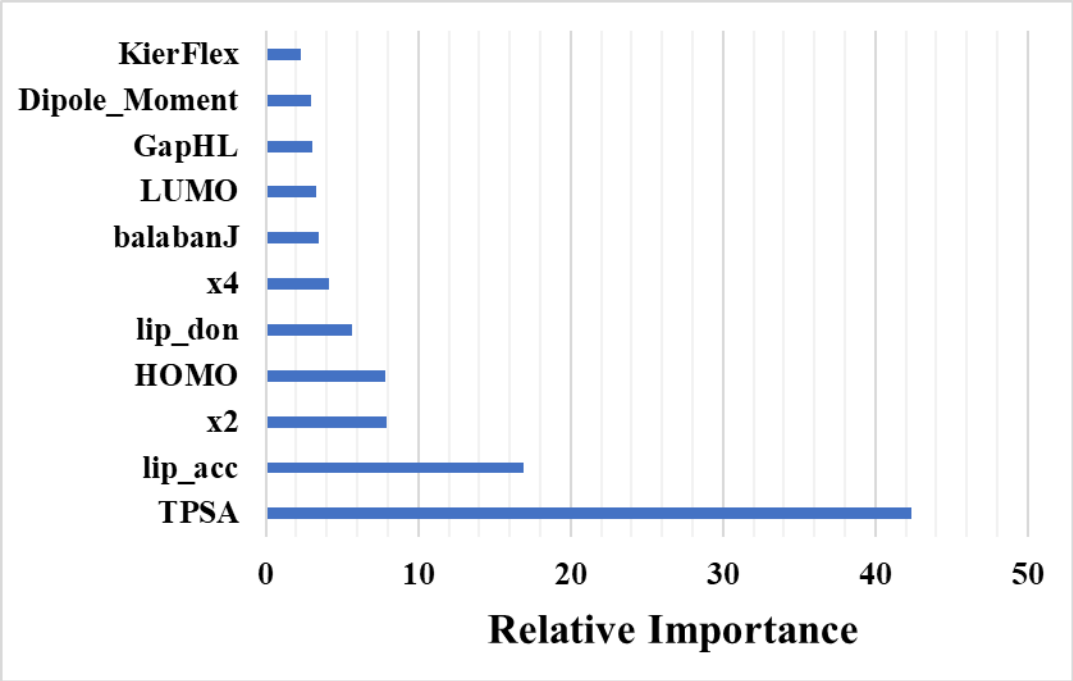

Figure S3: Relative importance (influence) of different variables in *Model C* obtained from GBM analysis.

Descriptors used for *Model D*:

Descriptor	Details	*Descriptor*	Details
TPSA	Topological Polar Surface Area	*x4*	Excess chemical potential in n-octanol
lip_acc	Lipinski H-bond acceptor count	*pmv4*	Partial molar volume in n-octanol
HOMO	Highest occupied molecular orbital	*x2*	Excess chemical potential in cyclohexane
Dipole_Moment	Dipole Moment	*lip_don*	Lipinski H-bond donor count

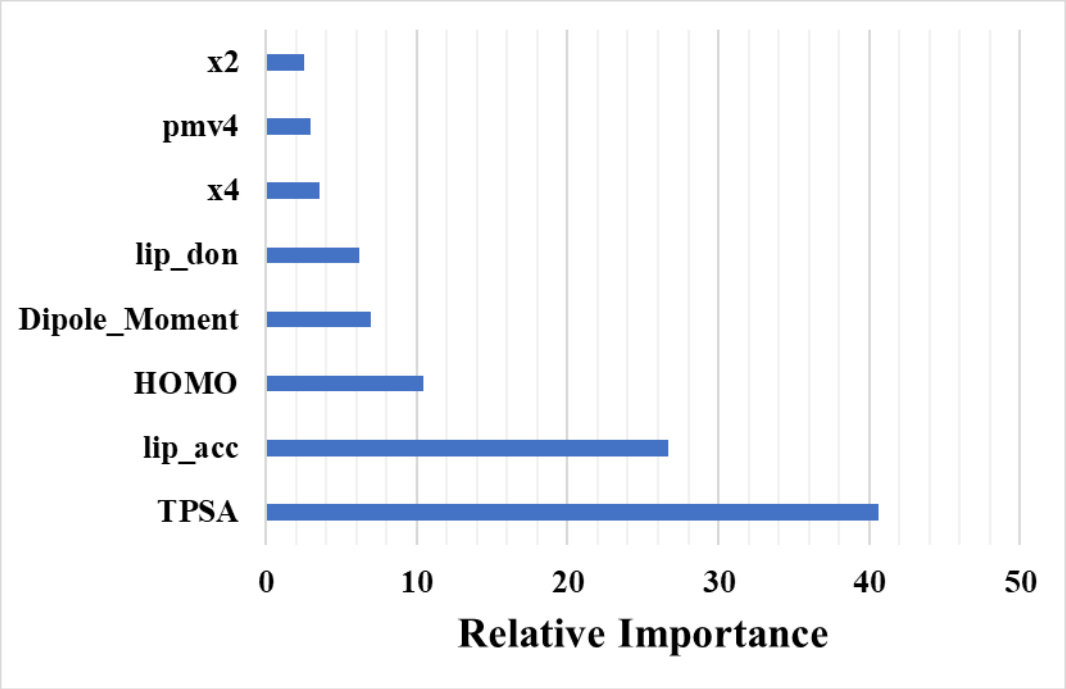

Figure S4: Relative importance (influence) of different variables in *Model D* obtained from GBM analysis.

Descriptors used for *Model E*:

Descriptor	Details
TPSA	Topological Polar Surface Area
lip_acc	Lipinski H-bond acceptor count
HOMO	Highest occupied molecular orbital
Dipole_Moment	Dipole Moment
x4	Excess chemical potential in n-octanol

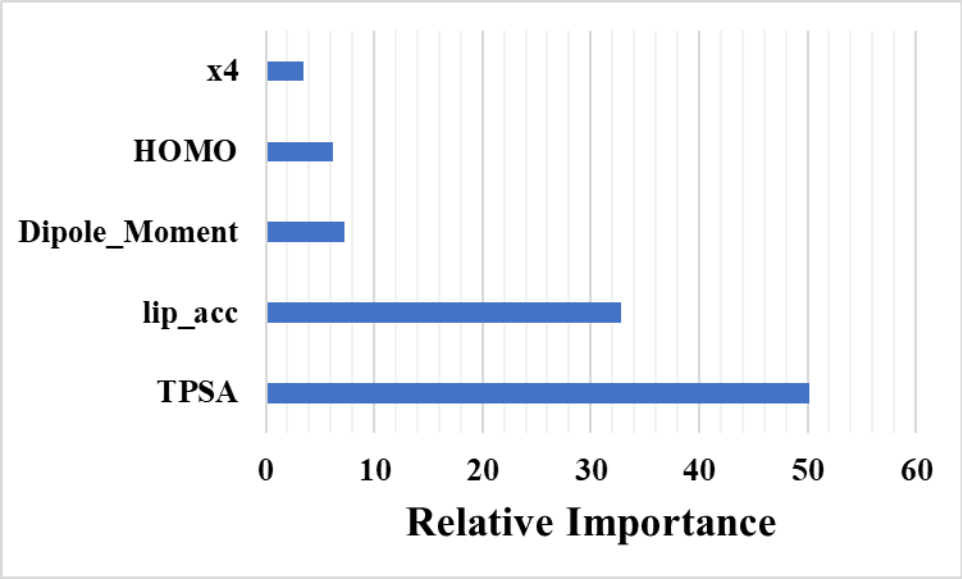

Figure S5: Relative importance (influence) of different variables in *Model E* obtained from GBM analysis.

Descriptors used for building model with CPCM calculated solvation energy:

Descriptor	Details	*Descriptor*	Details
TPSA	Topological Polar Surface Area	*KierFlex*	Molecular flexibility index
lip_acc	Lipinski H-bond acceptor count	*mr*	Molar refractivity
HOMO	Highest occupied molecular orbital	*weinerPath*	weiner path number
Dipole_Moment	Dipole Moment		
lip_don	Lipinski H-bond donor count	*weinerPol*	weiner polarity number
logS	aqueous solubility descriptor	*chi0*	connectivity descriptor
LUMO	Lowest unoccupied molecular orbital	*a_heavy*	Total number of heavy atoms
a_hyd	Number of hydrogen atoms	*CPCM1*	$\Delta\Delta G_{Chloroform}$
balabanJ	Balaban's J Index	*CPCM2*	$\Delta\Delta G_{Cyclohexane}$
ASA	Accessible surface area	*CPCM3*	$\Delta\Delta G_{n\text{-}Hexadecane}$
Zagreb	Zagreb index	*CPCM4*	$\Delta\Delta G_{n\text{-}octanol}$
		CPCM5	$\Delta\Delta G_{Water}$
Weight	Molecular weight	*vdw_area*	van der Wall's area
Bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens)	*vdw_vol*	van der Wall's volume
Rings	ring count	*a_aro*	Total number of aromatic atoms
Apol	Sum of the atomic polarizabilities (including implicit hydrogens)	*a_count*	Total number of atoms
PetitjeanNumber	Petitjean Number	*McGowan_Volume*	Characteristic McGowan volume

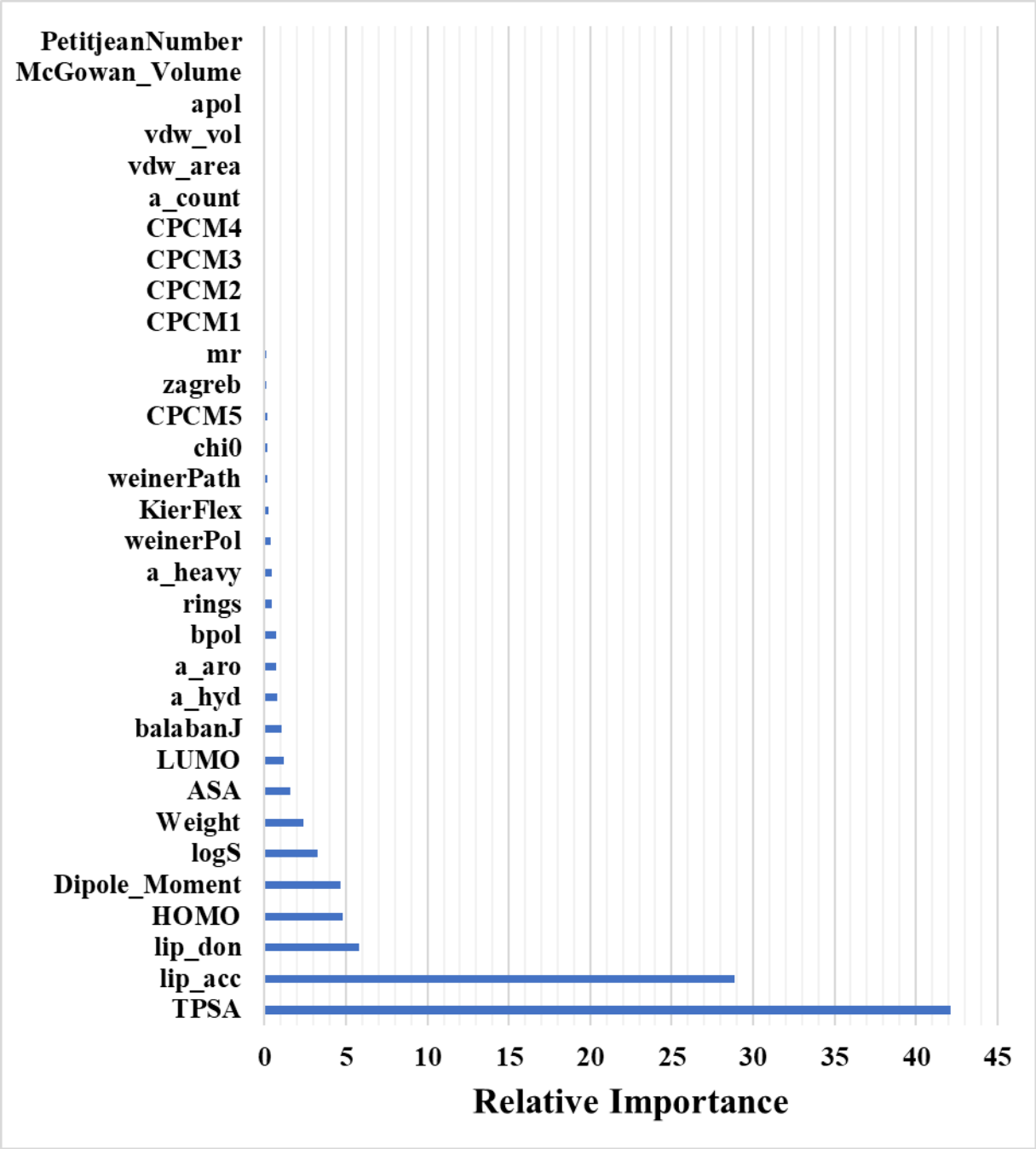

Figure S6: Relative importance (influence) of different variables in of model, built using CPCM calculated solvation energy, obtained from GBM analysis.

Descriptors used for building model with SMD calculated solvation energy:

Descriptor	Details	*Descriptor*	Details
TPSA	Topological Polar Surface Area	*KierFlex*	Molecular flexibility index
lip_acc	Lipinski H-bond acceptor count	*mr*	Molar refractivity
HOMO	Highest occupied molecular orbital	*weinerPath*	weiner path number
Dipole_Moment	Dipole Moment		
lip_don	Lipinski H-bond donor count	*weinerPol*	weiner polarity number
logS	aqueous solubility descriptor	*chi0*	connectivity descriptor
LUMO	Lowest unoccupied molecular orbital	*a_heavy*	Total number of heavy atoms
a_hyd	Number of hydrogen atoms	*SMD1*	$\Delta\Delta G_{Chloroform}$
balabanJ	Balaban's J Index	*SMD2*	$\Delta\Delta G_{Cyclohexane}$
ASA	Accessible surface area	*SMD3*	$\Delta\Delta G_{n\text{-}Hexadecane}$
Zagreb	Zagreb index	*SMD4*	$\Delta\Delta G_{n\text{-}octanol}$
		SMD5	$\Delta\Delta G_{Water}$
Weight	Molecular weight	*vdw_area*	van der Wall's area
Bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens)	*vdw_vol*	van der Wall's volume
Rings	ring count	*a_aro*	Total number of aromatic atoms
Apol	Sum of the atomic polarizabilities (including implicit hydrogens)	*a_count*	Total number of atoms
PetitjeanNumber	Petitjean Number	*McGowan_Volume*	Characteristic McGowan volume

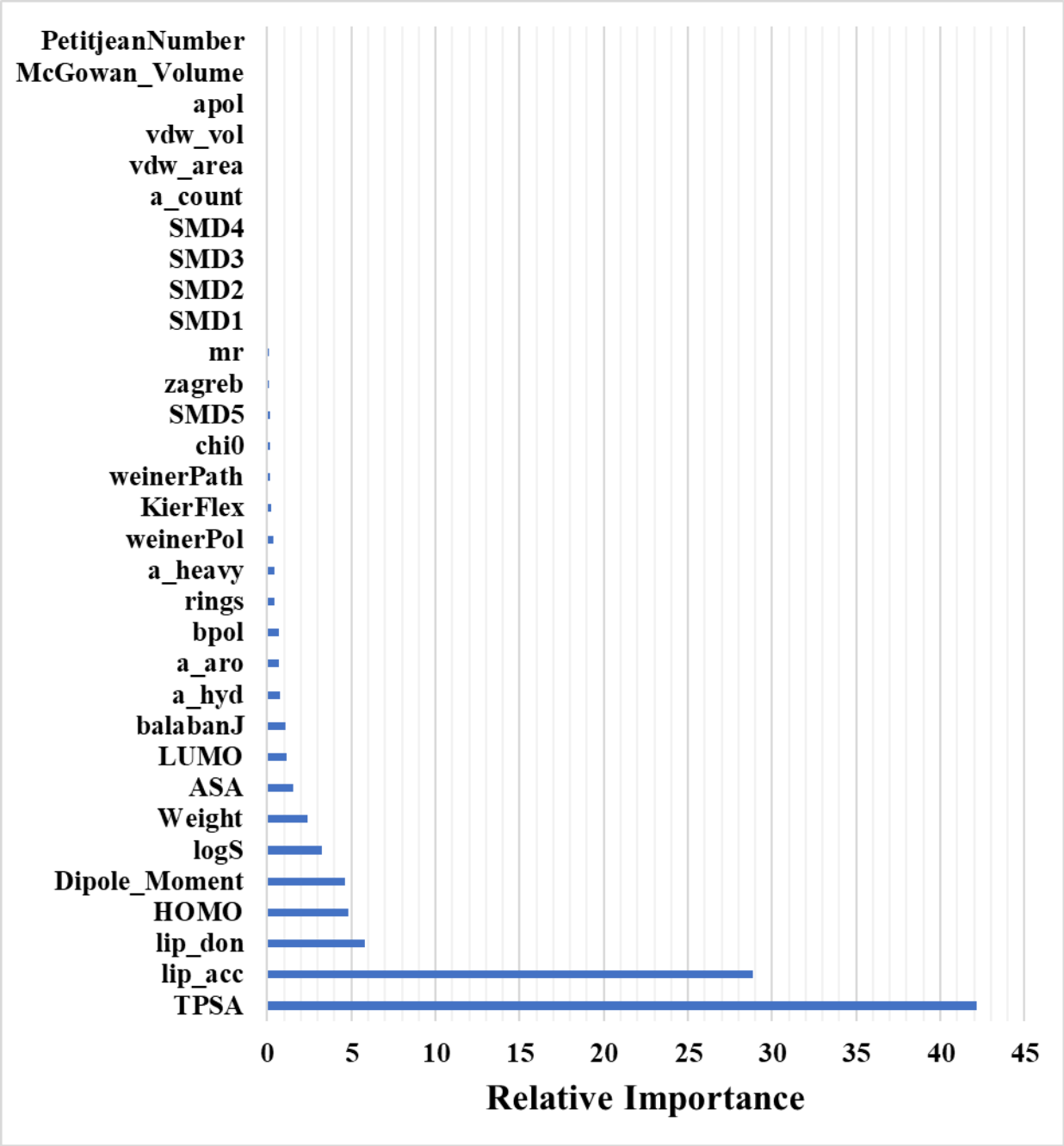

Figure S7: Relative importance (influence) of different variables of model, built using SMD calculated solvation energy, obtained from GBM analysis.

Descriptors used for building model without 3D-RISM-KH calculated solvation descriptors:

Descriptor	Details	*Descriptor*	Details
TPSA	Topological Polar Surface Area	*ASA*	Accessible surface area
lip_acc	Lipinski H-bond acceptor count	*Rings*	Number of rings
HOMO	Highest occupied molecular orbital	*lip_don*	Lipinski H-bond donor count
Dipole_Moment	Dipole Moment	*logS*	aqueous solubility descriptor

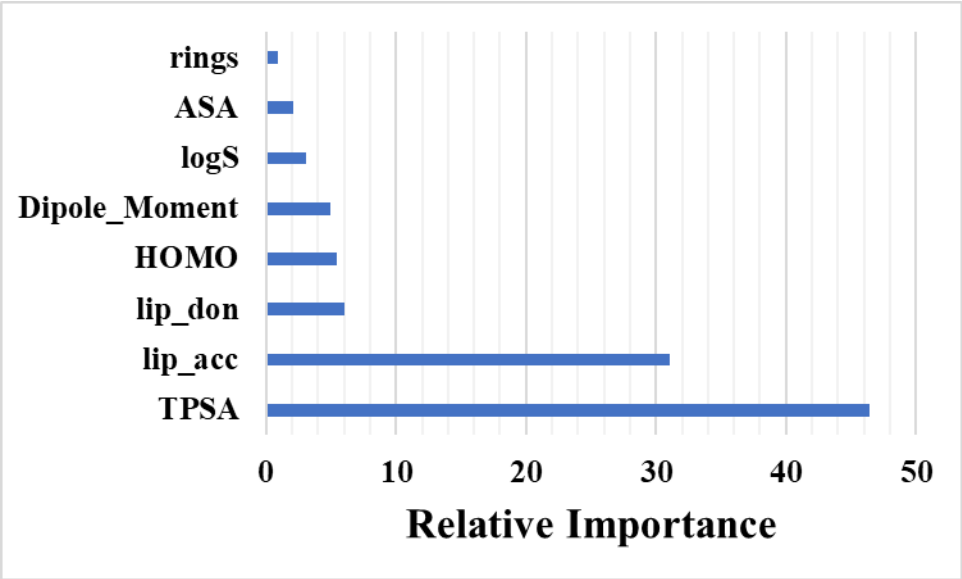

Figure S8: Relative importance (influence) of different variables in model, built without using 3D-RISM-KH calculated solvation descriptors, obtained from GBM analysis.

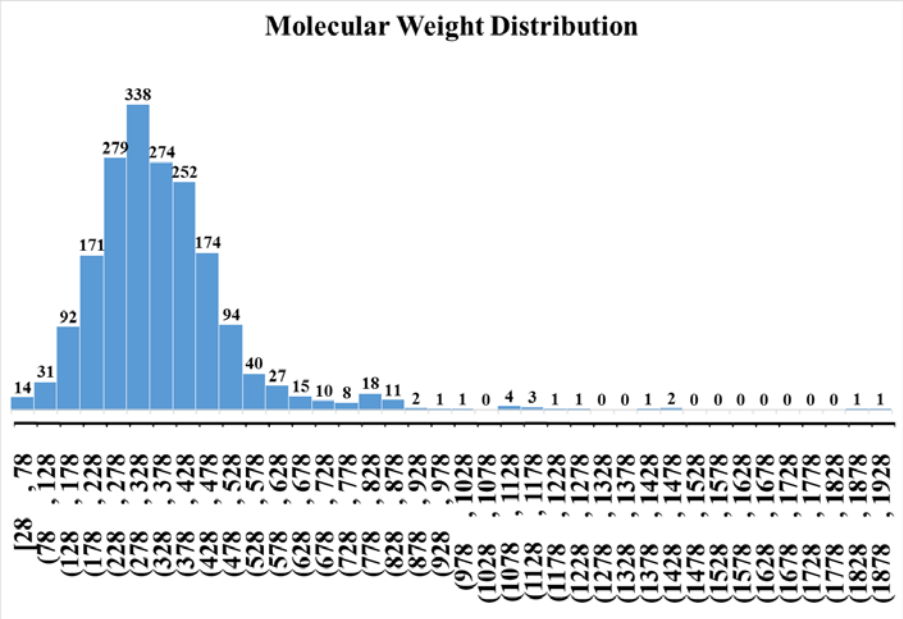

Figure S9: Molecular weight distribution of the entire dataset (bin width 50).

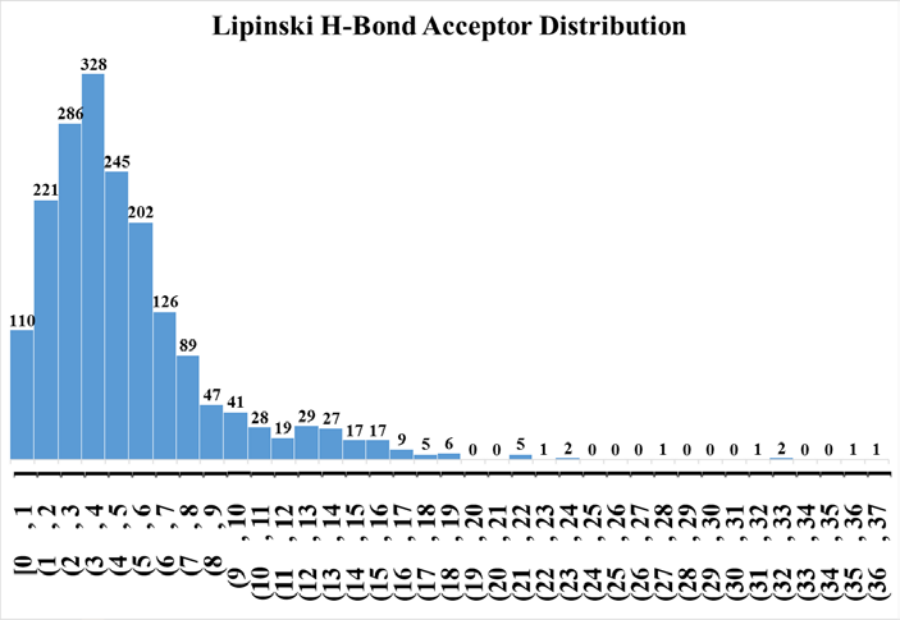

Figure S10: Lipinski acceptor distribution of the entire dataset (bin width 2).

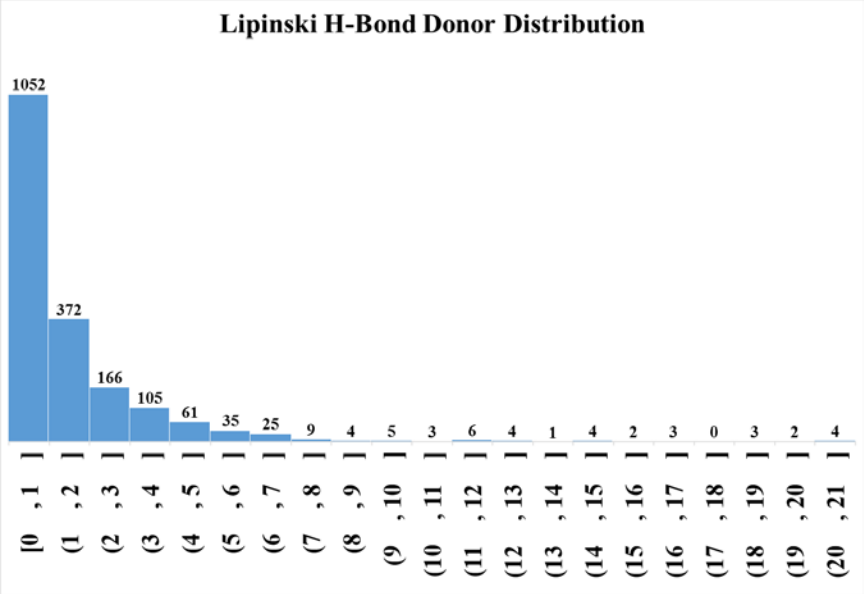

Figure S11: Lipinski donor distribution of the entire dataset (bin width 1).

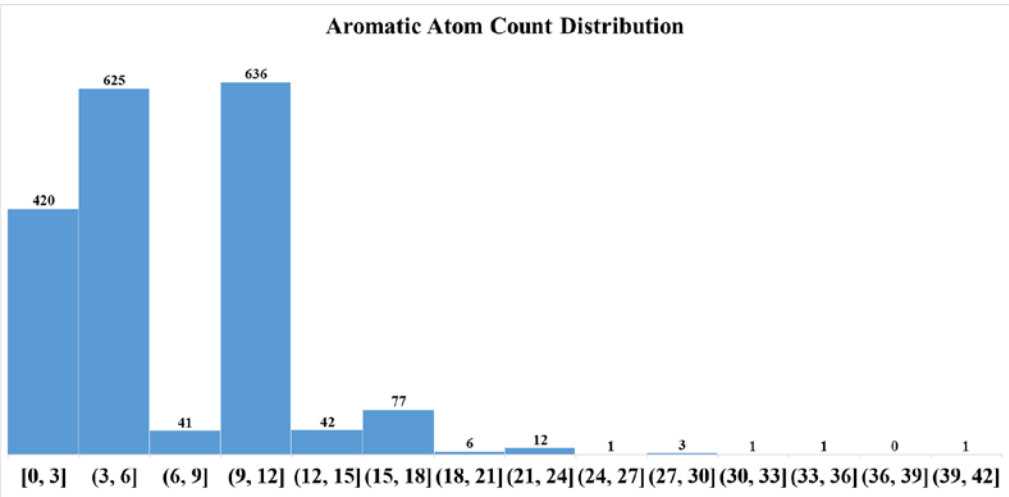

Figure S12: Aromatic atom count distribution of the entire dataset (bin width 3).

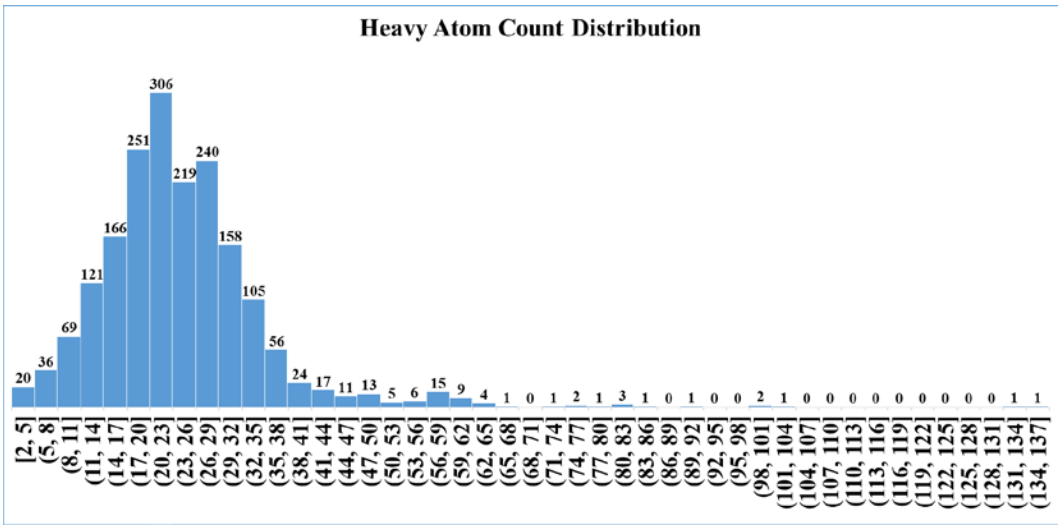

Figure S13: Heavy atom count distribution of the entire dataset (bin width 3).

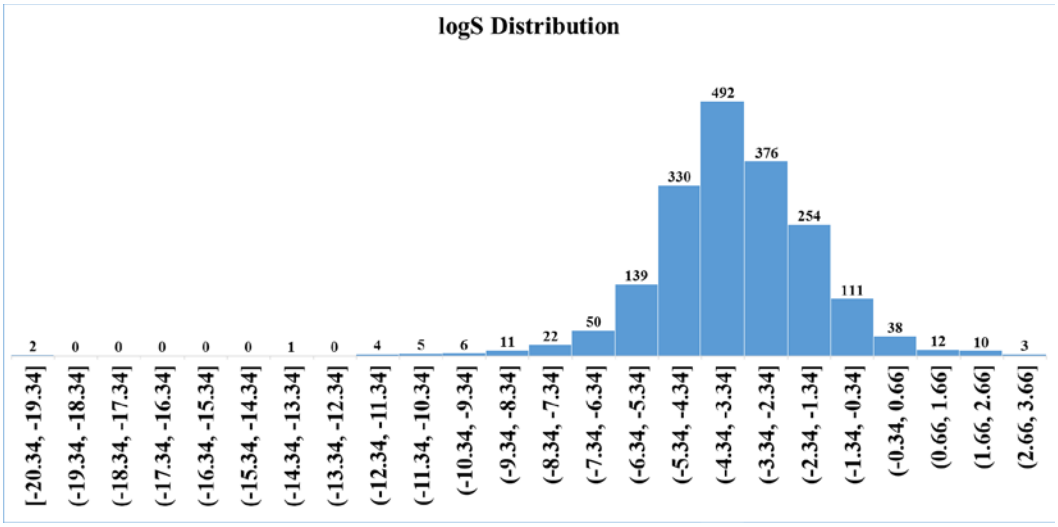

Figure S14: logS distribution of the entire dataset (bin width 1 log unit).

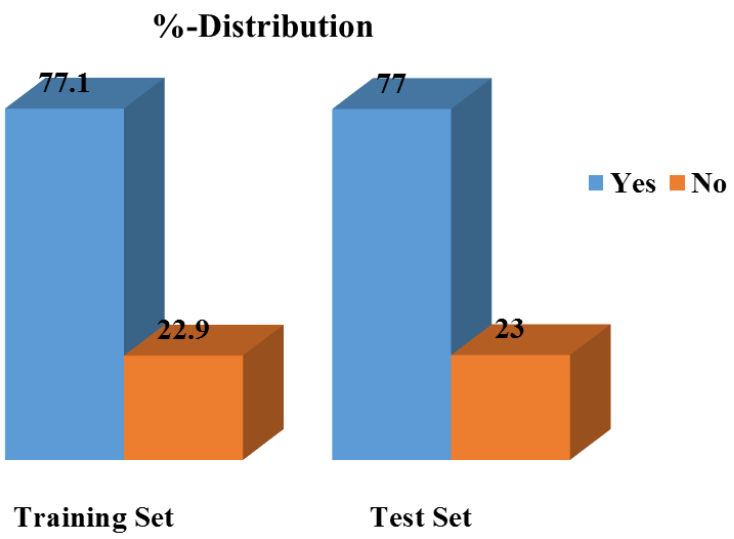

Figure S15: %-Distribution of blood-brain permeability in the training and test sets.

Table S1: Performance of logP(octanol-water) based prediction schemes with 7-variables in classification models using the SVM technique[a]

logP Prediction Method	Accuracy	Precision	Sensitivity	Specificity	F1-Score
ALogP	0.95	0.96	0.98	0.85	0.96
CrippenLogP	0.96	0.96	0.98	0.87	0.97
ManholandLogP	0.96	0.96	0.98	0.88	0.97
XLogP	0.96	0.96	0.99	0.88	0.98

[a]For a measure of percentage of accuracy indices, please multiply individual values by 100%. Please see the footnote in Table 1 for the definitions of performance indices used. Descriptors used for these calculations are HOMO, Dipole moment, molecular connectivity index, Lipinski H-bond donor, Lipinski H-bond acceptor, TPSA, and logP.

Full Citation of Reference 36:

Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; ; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. et al. Gaussian16, revision B.01; Gaussian Inc.: Wallingford, CT, 2016.

Complete Citation for Reference 39:

Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K., et al.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER; University of California: San Francisco, 2018.