Bulgarian Chemical Communications, Volume 56, Issue 3 (pp. 247-254) 2024 DOI: 10.34049/bcc.56.3.GS-M

Carvacrol and 4-IP-2-MeO-1-MB derivatives: DFT computations and drug-likeness studies

G. Serdaroğlu*

Sivas Cumhuriyet University, Faculty of Education, Department of Mathematics and Science Education, 58140, Sivas, Turkey

Received May 26, 2024; Accepted: August 17, 2024

This work deals with the detailed computational investigation of carvacrol and 4-IP-2-MeO-1-MB (4-isopropyl-2-methoxy-1-methylbenzene) derivatives which include an -OCH_3 functionality instead of an -OH group. Molecular orbital studies were performed at B3LYP/6-311G** level of the theory, in the gas phase. FMO analyses disclosed that CM compound would gain higher charge transfer capability in the presence of halogen. Also, the lipophilicity, water-solubility, drug-likeness, and ADMT properties were predicted to enlighten the possible bioavailability, physicochemical, and pharmacokinetic characteristics, as well as the adverse effects on both health and environment.

Keywords: Carvacrol, DFT, FMO & MEP, drug likeness

INTRODUCTION

Carvacrol molecule, as a member of the monoterpenoid phenols, has a characteristic pungent odor and is a component of different volatile oils such as oregano, thyme, Lippia, Nigella Sativa, etc. [1-3]. It has been known in traditional medicine for a long time [4] because of its many bioactivity characteristics such as antimicrobial, antioxidant, anticancer, anti-HIV, etc. [5, 6]. Mouwakeh and co-workers have explored the microbial activity and capability of resistance modifiers of the compounds that are the main components of *Nigella sativa* volatile oil; carvacrol and p-cymene can be used as resistance modifiers in MRSA strains [3]. Recently, Anjos and co-workers [7] have prepared biodegradable films including carvacrol and thymol to evaluate their potency on tick control; the inclusion of thymol and carvacrol in the biofilm has increased the mortality rates of larvae and engorged females of the tick *R. microplus* [7]. Kazemi and colleagues have explored the potencies of carvacrol (CAR) and p-cymene in preventing synaptic plasticity impairment; the results imply that combined treatment with carvacrol and p-cymene to prohibit the destructive effects of Aβ on hippocampal LTP couldn't be successful, in spite of their useful effects against Alzheimer's disease (AD) [8]. In previous works, monoterpenes such as carvone, terpineol, limonene [9], and structurally related pyrimidine and cumene [10] derivatives have been reported: the possible bioavailability, physicochemical, and chemical reactivity properties have been investigated by using computational tools. The main motivation of this work is to enlighten the key electronic, physicochemical, and biomedical properties of carvacrol and structurally related compounds. First, all compounds were optimized at DFT/B3LYP/6-311G ** level and then confirmed by having no negative frequency. Then, physicochemical properties were elucidated in light of the computed lipophilicity and water-solubility scores, which are important in terms of designing/modifying future drug agents. Furthermore, the pharmacokinetic and possible toxicity tendencies of the data set were predicted and evaluated. The obtained results of these simple molecular systems are hoped to provide information on exploring the proper precursors which can be used in further drug-design works.

COMPUTATIONAL DETAILS

All DFT/B3LYP/6-311G** level [11, 12] computations were employed by G09W [13] in the gas phase, and the optimized geometries, FMO plots were visualized by GaussView 6.0.16 [14]. The statistical thermodynamic principles [15, 16] were used for elucidation of the thermochemical quantities of the data set.

The *I* "ionization energy" and *A* "electron affinity" were predicted by HOMO and LUMO energies according to Koopmans' theorem [17]. The *I* and *A* values were used to calculate the other reactivity parameters, which are χ "electronic chemical potential", η "global hardness", ω "electrophilicity", ΔN_{max} "maximum charge transfer capability index" [18, 19], ω^- "electrodonating power", ω^+ "electroaccepting power" [20], and $\Delta E_{back\text{-}donat.}$ "back-donation energy" [21].

* To whom all correspondence should be sent:
E-mail: *goncagul.serdaroglu@gmail.com*

I= -E_{HOMO} and A= -E_{LUMO}

$$\chi = -(\frac{I+A}{2})$$

$$\eta = \frac{I-A}{2}$$

$$\omega = \frac{\mu^2}{2\eta}$$

$$\Delta N_{max} = (I + A)/2(I - A)$$

$$\omega^{+} \approx (I + 3A)^2/(16(I - A))$$

$$\omega^{-} \approx (3I + A)^2/(16(I - A))$$

$$\Delta\varepsilon_{back-donation} = -(\eta/4)$$

The lipophilicity [22-26] and water-solubility [27, 28] features of the data set were determined by using SwissADME [29] tools. Also, drug-likeness [25, 30-33], bioavailability [34], and ADMT [35] parameters of the data set were determined.

RESULTS AND DISCUSSION

Physicochemistry

The optimized structures and the data set's thermodynamic quantities/physical values are given in Fig. 1 and Table 1, respectively. As expected, the μ and α values were calculated for the carvacrol molecule at 1.254 D and 115.190 au, respectively. The ΔE, ΔH, and ΔG (au) values of carvacrol were calculated at -464.639712, -464.627152, and -464.676987 au, respectively. Instead of the -OH group, the -OCH_3 substitution on the substituted methylbenzene (CM) made these quantities increase, that is, they were calculated for CM at -503.921202 (ΔE), -503.907295 (ΔH), and -503.960320 au (ΔG), respectively. Table 1 shows that the thermodynamic values slightly change according to the position of the halogen atom on the core CM structure. For instance, the ΔE, ΔH, and ΔG (au) values of the CM1 were slightly greater predicted -603.194716, -603.179844, and -603.235346, au, respectively, than those of the CM2. Among the carvacrol derivatives, the highest α values were determined for the bromine-substituted derivatives CM5 and CM6 at 146.746 and 146.254 au, respectively.

Furthermore, Table 2 shows the other two key physicochemical parameters lipophilicity and water-solubility in drug-design research. As is well known, these parameters have shown some different orders depending on the used approach, wherein 5 approaches have been used. Looking at the results of lipophilicity, CM4 (4.94) would show the most lipophilic features according to the XLOGP3, the other methods implied that the bromine-substituted derivatives CM5 and CM6 would be more lipophilic among the compounds. On the other hand, the Avg. LogPo/w order, which was CV (2.82) < CM (3.21) < CM1= CM2 (3.44) < CM3 (3.67) < CM4 (3.85) < CM6 (3.75) < CM5 (3.76) implied that carvacrol would be less lipophilic and CM5 could be more lipophilic than the other derivatives. As expected from the lipophilic features of the derivatives, the halogenated compounds would have lower water-solubility than the CV and CM compounds. According to the SILICOS-IT approach, the CV and CM molecules would be more water-soluble, while all halogenated derivatives would have moderate solubility in water. On the other hand, CM4-CM6 derivatives could exhibit a medium-level water-solubility, whereas the remaining compounds would be soluble in water. All methods implied that CM4 would present lower solubility than the others.

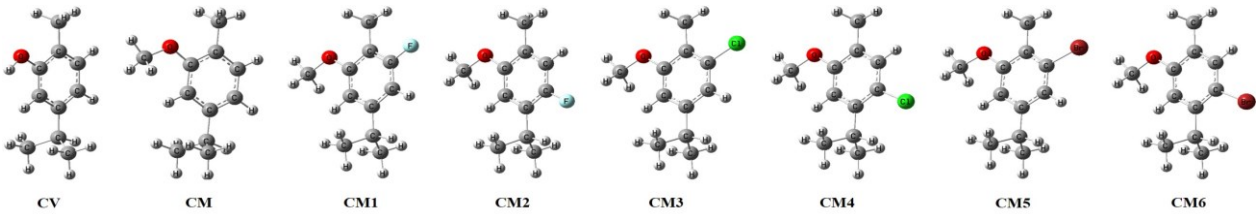


Fig. 1. Optimized chemical structures of the data set

Table 1. Thermochemical and physical values of the data set

Compound	ΔE (au)	ΔH (au)	ΔG (au)	$E_{therm.}$	Cv	S	μ	α
CV	-464.639712	-464.627152	-464.676987	142.606	44.249	104.886	1.254	115.190
CM	-503.921202	-503.907295	-503.960320	161.181	48.726	111.600	1.319	128.045
CM1	-603.194716	-603.179844	-603.235346	156.700	51.744	116.814	2.864	127.588
CM2	-603.193057	-603.178217	-603.233994	156.693	51.715	117.393	2.370	127.536
CM3	-963.551686	-963.536369	-963.593622	156.026	52.660	120.499	3.423	140.193
CM4	-963.550795	-963.535546	-963.593304	156.079	52.633	121.562	3.008	139.798
CM5	-3077.472454	-3077.456881	-3077.515348	155.796	53.086	123.055	3.345	146.746
CM6	-3077.471787	-3077.456280	-3077.514697	155.841	53.077	122.950	2.937	146.254

Table 2. Lipophilicity and water-solubility

	CV	CM	CM1	CM2	CM3	CM4	CM5	CM6
Lipophilicity								
iLOGP	2.24	2.76	2.87	2.87	3.03	3.00	3.12	3.08
XLOGP3	3.49	3.82	3.48	3.48	4.01	4.94	4.08	4.08
WLOGP	2.82	3.13	3.69	3.69	3.78	3.78	3.89	3.89
MLOGP	2.76	3.05	3.46	3.46	3.60	3.60	3.74	3.74
SILICOS-IT	2.79	3.29	3.72	3.72	3.94	3.94	3.97	3.97
Avg. LogPo/w	2.82	3.21	3.44	3.44	3.67	3.85	3.76	3.75
Water-solubility								
Log S (ESOL)	-3.31	-3.50	-3.37	-3.37	-3.81	-4.39	-4.13	-4.13
Solubility (mg/mL)×10^{-2}	7.40	5.16	7.74	7.74	3.09	0.803	1.81	1.81
Class	S	S	S	S	S	MS	MS	MS
Log S (Ali)	-3.60	-3.71	-3.36	-3.36	-3.91	-4.87	-3.98	-3.98
Solubility (mg/mL)×10^{-2}	3.79	3.21	8.02	8.02	2.47	0.267	2.55	2.55
Class	S	S	S	S	S	MS	S	S
Log S (SILICOS-IT)	-3.01	-3.73	-4.01	-4.01	-4.36	-4.36	-4.59	-4.59
Solubility (mg/mL)×10^{-2}	14.6	3.09	1.76	1.76	0.875	0.875	0.619	0.619
Class	S	S	MS	MS	MS	MS	MS	MS

Drug-likeness and ADMT study

According to the Lipinski, Veber and Egan rules, all compounds could be prospective structures in terms of drug-likeness (see Table S1). On the other hand, the Muegge method implied that none of the compounds would have been a proper structure in view of the drug-likeness; heteroatom numbers for all compounds are smaller than 2. In addition to heteroatom numbers, except for CM5 and CM6 compounds, the other violation is the molecular weight < 200. Except for CV, all compounds seem to obey to Ghose rules (Table S1).

From Table S2 (suppl. data), Caco-2 permeability scores of all compounds were calculated in the range of -4.400 and -4.775, which were higher than that of the optimal value (-5.15 Log unit). All compounds could fail by looking at the MDCK permeability and Pgp-inhibitor. Fortunately, the PAMPA, Pgp-substrate, and HIA scores of the data set seem to satisfy expectations. Except for CM3, the VD (volume distribution) of all compounds was predicted in the optimal range of 0.04-20L/kg. CV, CM, CM2, and CM4 would have the capability of BBB (Blood-Brain Barrier) penetration, whereas the others would also be capable of BBB penetration, but not as much as these compounds. The CYP2C9 and CYP2D6 substrate potencies of all compounds would be satisfying. Furthermore, the hERG blockers, genotoxicity, RPMI-8226, A549, and Hek293 cytotoxicity scores implied that all compounds could be safe in terms of avoiding possible medicinal toxicity (Table S3). The calculated scores of the Tox21 pathway implied that none of the compounds could show any toxic effect, except for the CV compound. Namely, the calculated SR-MMP score of CV implied that this compound could affect the mitochondrial membrane potential.

FMO and MEP analyses

The reactivity values obtained from FMOs' energies were used to predict the possible reactivity tendency and region of the molecular systems; wherein the determined reactivity values of the compounds change in the following orders:

ΔE (L-H): CM (5.821)< CM1 (5.965)< CM2 (5.542)< CM3 (5.826)< CM4 (5.587)< CM5 (5.799)< CM6 (5.565)< CV (5.852)

μ: CM (-2.966)> CM1 (-3.148)> CM2 (-3.118)> CM3 (-3.292)>CM4 (3.217)> CM5 (-3.298)> CM6 (-3.206)> CV (-3.051)

η: CM2 (2.771)< CM6 (2.783)< CM4 (2.793)< CM5 (2.900)< CM (2.911)< CM3 (2.913)< CV (2.926)< CM1 (2.982)

ω: CM (0.056)< CV (0.058)< CM1 (0.061)< CM2 (0.064)< CM3= CM4=CM6 (0.068)< CM5 (0.069)

ω^{+}: CM (0.014)< CV (0.016)< CM1 (0.017)< CM2 (0.020)< CM3 (0.021)< CM4=CM5=CM6 (0.022)

ω^{-}: CM (0.123)< CV (0.128)< CM1 (0.133)< CM2 (0.135)< CM4= CM6 (0.140)< CM3 (0.142)< CM5 (0.143)

ΔN_{max}: CM (1.019)< CV (1.043)< CM1 (1.055)< CM2 (1.125)< CM3 (1.130) < CM5 (1.137)< CM4= CM6 (1.152)

$\Delta E_{back.}$: CM1 (-0.746)< CV (-0.731)< CM=CM3 (-0.728)< CM5 (-0.725)< CM4 (-0.698)< CM6 (-0.696)< CM2 (-0.693)

Accordingly, the CM molecule would prefer the intra-molecular interactions than an action toward an external system, less stable electronically, and *vice*

versa for CV. Also, the CM molecule would be less electrophile and have less charge transfer capability. Furthermore, the halogen substitution would gain the main structure more charge transfer capability, especially the -Cl and -Br substitution. On the other hand, the CM1 could gain stability by back-donation more than the other molecules, whereas the CM2 would benefit from back-donation, but not as much as the other molecules.

Fig. 2 shows that the HOMO mostly extended on the aromatic rings, methyl, and methoxy groups, except for CM3, which HOMO did not seem on the methoxy group. The LUMO of CM1 and CM3 expanded on the oxygen of the methoxy group a little, whereas it did not seem on the halogen of CM2 (-F) and CM4 (-Cl). The HOMO and LUMO representations imply the nucleophilic and electrophilic attack sites, respectively. Herein, the -F and -Cl halogens and -OCH_3 substitutions for CM2 and CM4 molecules could not have any role in electrophilic attacks; the isopropyl unit for all compounds would not relate to the nucleophilic attacks. Furthermore, the MEP presentations provide information on the possible reactive regions of the systems by showing the electron-rich (V<0, red) and electron-poor (V>0, blue) sites as a function of the electrostatic potential on the surface. As expected, the -H atom of the -OH group of the CV molecule was seen as blue, and the aromatic ring was covered by a yellow color. First, the halogen substitution would affect the electron density on the aromatic ring for all derivatives due to the resonance donation to the ring. Thus, the oxygen atoms and aromatic rings of all compounds seem red colored indicating the regions for the electrophiles, and -OCH_3 groups' H is seen with blue color indicating the electron-poor site for the nucleophiles.

CONCLUSION

Herein, comprehensive computational works are presented to the evaluation of the reactivity, physicochemical, and bioavailability features of carvacrol and its derivatives. The lipophilicity scores indicated that the CM4 compound would be more lipophilic, and CV was less lipophilic. In contrast, the CV molecule would have the best water-soluble capability among the compounds, while the -Br substituted compounds CM5 and CM6 could be more soluble in water. FMO analyses implied that the halogen substitution on the CM molecule provides higher electron density on the aromatic ring *via* resonance. According to the Lipinski, Veber and Egan rules, all compounds could be promising agents, while none of them, depending on the Muegge method, would be a proper structure in terms of drug-likeness.

Table 6. Chemical reactivity parameters

	H (-I)/ eV	L (-A)/ eV	ΔE (L-H)/ eV	μ/ eV	η/ eV	ω/ au	ω^+/ au	ω^-/ au	ΔN_{max}/ eV	$\Delta E_{back.}$/ eV
CV	-5.977	-0.125	5.852	-3.051	2.926	0.058	0.016	0.128	1.043	-0.731
CM	-5.877	-0.055	5.821	-2.966	2.911	0.056	0.014	0.123	1.019	-0.728
CM1	-6.13	-0.165	5.965	-3.148	2.982	0.061	0.017	0.133	1.055	-0.746
CM2	-5.889	-0.347	5.542	-3.118	2.771	0.064	0.020	0.135	1.125	-0.693
CM3	-6.205	-0.379	5.826	-3.292	2.913	0.068	0.021	0.142	1.130	-0.728
CM4	-6.011	-0.424	5.587	-3.217	2.793	0.068	0.022	0.140	1.152	-0.698
CM5	-6.197	-0.398	5.799	-3.298	2.900	0.069	0.022	0.143	1.137	-0.725
CM6	-5.988	-0.423	5.565	-3.206	2.783	0.068	0.022	0.140	1.152	-0.696

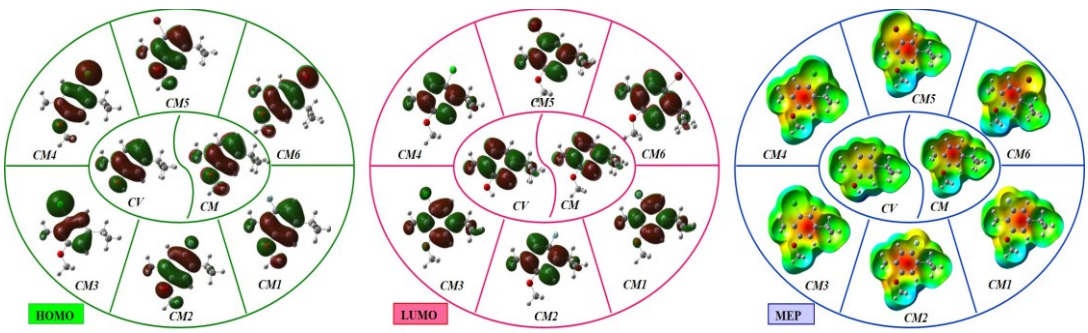


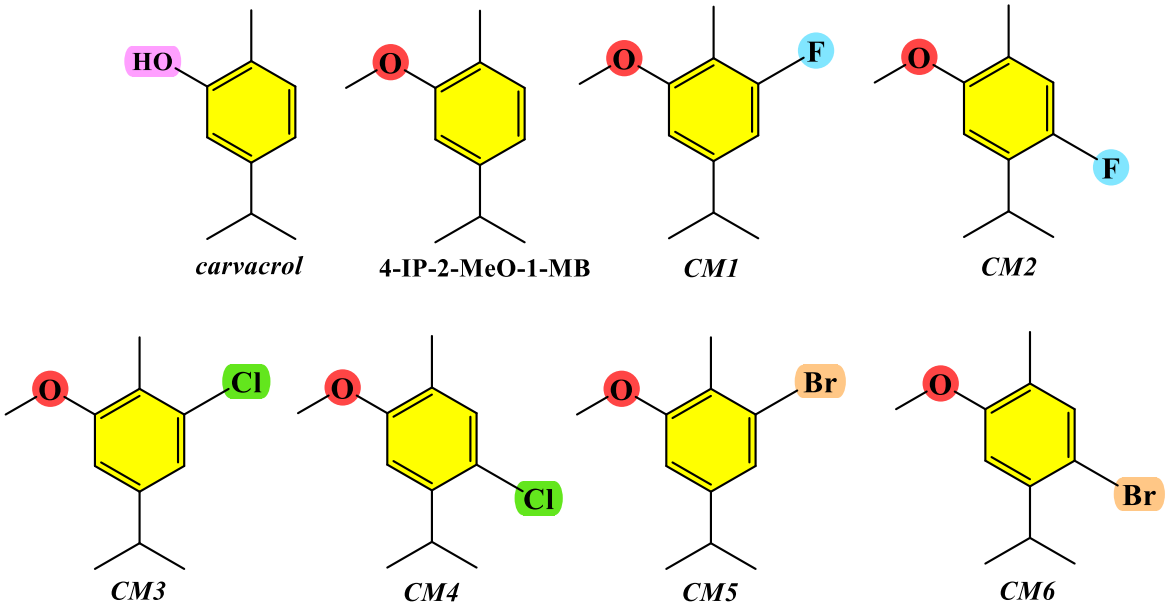
Fig. 2. HOMO and LUMO amplitudes of the data set

Acknowledgement: *All calculations have been carried out at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TR-Grid e-Infrastructure). The author thanks the Scientific Research Projects Department of Sivas Cumhuriyet University (Project No: EĞT-2023-098).*

REFERENCES

1. M. Sharifi-Rad, E. M. Varoni, M. Iriti, M. Martorell, W. N. Setzer, M. Contreras, B. Saleh, A. Soltani-Nejad, S. Rajabi, M. Tajbakhsh, J. Sharifi-Rad, *Phytother. Res.*, **32**(9), 1675 (2018).
2. Ş. Yıldız, S. Turan, *Atatürk Üniversitesi Ziraat Fakültesi Dergisi*, **52** (1), 108 (2021).
3. A. Mouwakeh, A. Kincses, M. Nové, T. Mosolygó, C. Mohácsi-Farkas, G. Kiskó, G. Spengler. *Phytother. Res*., **33** (4), 1010 (2019).
4. N. B. Rathod, P. Kulawik, F. Ozogul, J. M. Regenstein, Y. Ozogul. *Trends Food Sci. Tech.,* **116,** 733 (2021).
5. S. Peter, N. Sotondoshe, B. A. Aderibigbe, *Molecules*, **29** (10), 2277 (2023).
6. C. Zhang, Z. Li, D. Hu, *J. Drug. Deliv. Sci. and Tech.*, **97,** 105779 (2024).
7. O. O. Anjos, M. N. Gomes, C. P. Tavares, D. M. Sousa, C. J. Mendonça, J. Reck, A. P. Maciel, L. M. Costa-Junior, *Vet. Parasitol.*, **327**, 110149 (2024).
8. S. Kazemi, S. Safari, S. Komaki, S. A. Karimi, Z. Golipoor, A. Komaki, *CNS Neurosci. Ther.*, **30** (3), e14459 (2024).
9. G. Serdaroğlu, *JOTCSA*.;11(2), 869 (2024).
10. G. Serdaroğlu, E. Soyutek, Ş. Koçarslan, C. Uludağ, *Results Chem.*, **6,** 101106 **(**2023)
11. A. D. Becke, *J. Chem. Phys*., **98**, 1372 **(**1993)
12. C. Lee, W. Yang, R.G. Parr, *Phys. Rev*., **B37**, 785 (1988**)**.
13. M. J. Frisch et al. Gaussian 09W, Revision D.01, Gaussian, Inc, Wallingford CT, 2013.
14. GaussView 6.0.16, Gaussian, Inc, Wallingford CT, (2016).
15. D. A. McQuarrie, Statistical Thermodynamics, Harper & Row Publishers, New York, **1973**.
16. G. Serdaroglu, S. Durmaz, *Indian J. Chem*., **49**, 861 (2010**).**
17. T. Koopmans, *Physica*, **1** (1), 104 (1934).
18. R. G. Pearson, *Proc. Natl. Acad. Sci. USA,* **83,** 8440 (1986**)**.
19. R. G. Parr, L.V. Szentpaly, S. Liu, *J. Am. Chem. Soc*., **121**, 1922 (1999).
20. J. L. Gazquez, A. Cedillo, A. Vela, *J. Phys. Chem. A*, **111** (10), 1966 (2007).
21. B. Gomez, N. V. Likhanova, M. A. Domínguez-Aguilar, R. Martínez-Palou, A. Vela, J. L. Gazquez, *J. Phys. Chem., B* **110** (18), 8928 (2006).
22. A. Daina, O. Michielin, V. Zoete, *J. Chem. Inf. Model*., **54**(12) 3284 (2014).
23. T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, *J. Chem. Inf. Model*., **47**(6), 2140 (2007).
24. S. A. Wildman, G. M. Crippen, *J. Chem. Inf. Comput. Sci*., **39,** 868 (1999).
25. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliver. Rev*., **46,** 3 (2001).
26. Silicos-IT, [online], Available: https://www.silicos-it.be.
27. S. Delaney, *J. Chem. Inf. Comput. Sci*., **44,** 1000 (2004).
28. J. Ali, P. Camilleri, M. B. Brown, A. J. Hutt, S. B. Kirton, *J. Chem. Inf. Model*., **52,** 2950 (2012).
29. A. Daina, O. Michielin, V. Zoete, *Sci Rep*., **7,** 42717 (2017).
30. A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, *J. Comb. Chem*., **1,** 55 (1999).
31. D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem*., **45,** 2615 (2002).
32. W. J. Egan, K. M. Merz, Jr., J. J. Baldwin, *J. Med. Chem*. **43,** 3867 (2000).
33. I. Muegge, S. L. Heald, D. Brittelli, *J. Med. Chem*., **44**(12), 1841 (2001).
34. Y. C. Martin, *J. Med. Chem*., **48** 3164 (2005).
35. ADMETlab 2.0

SUPPLEMENTARY DATA
Table of Contents



1-chloro-2-isopropyl-4-methoxy-5-methylbenzene

1-bromo-2-isopropyl-4-methoxy-5-methylbenzene

4-isopropyl-2-methoxy-1-methylbenzene

Fig. S1. The chemical structures of *the data set*

Table S1. Drug-likeness and bioavailability scores

	Lipinski	Ghose	Veber	Egan	Muegge	Bioavail. Score
CV	Yes	No; MW<160	Yes	Yes	No; MW<200, heteroatoms<2	0.55
CM	Yes	Yes	Yes	Yes	No; MW<200, Heteroatoms<2	0.55
CM1	Yes	Yes	Yes	Yes	No; MW<200, Heteroatoms<2	0.55
CM2	Yes	Yes	Yes	Yes	No; MW<200, Heteroatoms<2	0.55
CM3	Yes	Yes	Yes	Yes	No; MW<200, Heteroatoms<2	0.55
CM4	Yes	Yes	Yes	Yes	No; MW<200, Heteroatoms<2	0.55
CM5	Yes	Yes	Yes	Yes	No; Heteroatoms<2	0.55
CM6	Yes	Yes	Yes	Yes	No; Heteroatoms<2	0.55

Table S2. Absorption, Distribution, and Metabolism

	CV	CM	CM1	CM2	CM3	CM4	CM5
Absorption							
Caco-2 Pe.	-4.4	-4.576	-4.68	-4.652	-4.775	-4.737	-4.763
MDCK Pe. ($x10^{-5}$) cm/s	-4.681	-4.627	-4.651	-4.665	-4.681	-4.687	-4.663
PAMPA	0.027	0.011	0.047	0.002	0.032	0.001	0.022
Pgp-inh.	0.931	0.987	0.898	0.993	0.909	0.987	0.982
Pgp-subs.	0.064	0.163	0.216	0.081	0.079	0.046	0.033
HIA	0.012	0.017	0.021	0.002	0.025	0.001	0.100
$F_{20\%}$	0.385	0.227	0.166	0.037	0.128	0.034	0.158
$F_{30\%}$	0.719	0.393	0.307	0.074	0.400	0.134	0.555
$F_{50\%}$	0.933	0.917	0.906	0.259	0.955	0.615	0.922
Distribution							
PPB %	91.894	97.603	97.716	98.178	98.91	98.928	98.139
VD (L/kg)	0.206	0.217	0.05	0.264	-0.077	0.234	0.187
BBB Pen.	0.179	0.135	0.331	0.142	0.338	0.296	0.401
Fu %	7.482	2.041	1.998	1.392	0.762	0.731	1.703
Metabolism							
CYP1A2 inh.	0.952	0.966	0.964	0.930	0.991	0.994	0.995
CYP1A2 subs.	0.948	0.981	0.991	0.922	0.997	0.938	0.792
CYP2C19 inh.	0.799	0.941	0.923	0.98	0.952	0.993	0.974
CYP2C19 subs.	0.889	0.994	0.971	0.985	0.985	0.996	0.185
CYP2C9 inh.	0.859	0.691	0.894	0.959	0.726	0.859	0.967
CYP2C9 subs.	0.004	0.004	0.043	0.004	0.272	0.228	0.003
CYP2D6 inh.	0.006	0.678	0.152	0.348	0.533	0.859	0.234
CYP2D6 subs.	0.001	0.041	0.003	0.067	0.007	0.154	0.001
CYP3A4 inh.	0.153	0.293	0.471	0.135	0.489	0.755	0.266
CYP3A4 subs.	0.969	0.865	0.266	0.703	0.825	0.969	0.168
CYP2B6 inh.	1.000	0.985	0.908	0.993	0.995	0.995	0.998
CYP2B6 subs.	0.534	0.967	0.241	0.842	0.936	0.992	0.007
CYP2C8 inh.	0.998	0.981	0.995	0.999	0.985	0.998	0.997
HLM Stability	0.868	0.967	0.886	0.959	0.958	0.960	0.882

* Permeability, Pe; Penetration, Pen, Inhibitor, Inh; Substrate, subs

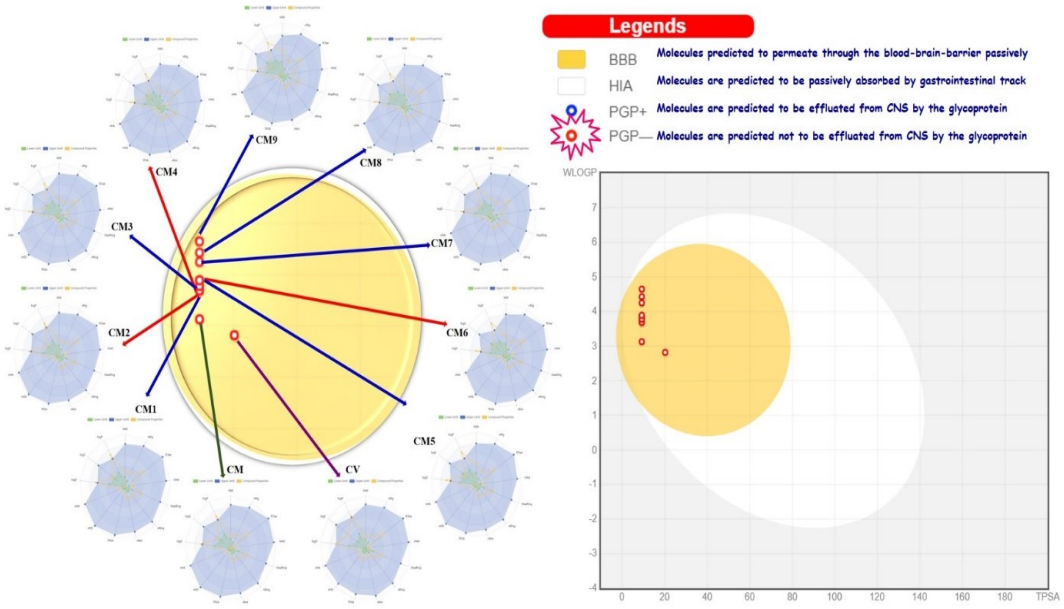


Fig. S2. Boiled-Egg model and radar graphs.

Table S3. Toxicity values of the data set

	CV	***CM***	***CM1***	***CM2***	***CM3***	***CM4***	***CM5***	***CM6***
Medicinal								
hERG Blockers	0.106	0.129	0.127	0.163	0.136	0.188	0.107	0.118
hERG Blockers (10um)	0.633	0.612	0.573	0.566	0.628	0.656	0.608	0.623
DILI	0.202	0.291	0.374	0.436	0.41	0.52	0.629	0.679
AMES Mutagenicity	0.398	0.405	0.396	0.534	0.317	0.356	0.316	0.361
Rat Oral Acute Toxicity	0.417	0.364	0.531	0.479	0.418	0.415	0.553	0.535
FDAMDD	0.359	0.3	0.385	0.405	0.365	0.336	0.674	0.568
Skin Sensitization	0.717	0.637	0.562	0.411	0.716	0.631	0.829	0.716
Carcinogenicity	0.606	0.628	0.552	0.704	0.628	0.678	0.664	0.709
Eye Corrosion	0.968	0.961	0.936	0.844	0.975	0.938	0.993	0.978
Eye Irritation	0.996	0.991	0.988	0.981	0.987	0.979	0.997	0.996
Respiratory	0.675	0.642	0.748	0.746	0.733	0.719	0.712	0.692
Human Hepa totoxicity	0.488	0.55	0.651	0.635	0.56	0.531	0.432	0.421
Drug-induced Nephrotoxicity	0.261	0.447	0.7	0.81	0.505	0.609	0.338	0.372
Ototoxicity	0.334	0.356	0.464	0.472	0.408	0.415	0.271	0.271
Hematotoxicity	0.335	0.475	0.487	0.55	0.499	0.524	0.213	0.244
Genotoxicity	0.119	0.046	0.166	0.254	0.048	0.09	0.194	0.225
RPMI-8226 Immunitoxicity	0.049	0.056	0.062	0.079	0.059	0.067	0.061	0.066
A549 Cytotoxicity	0.11	0.08	0.08	0.116	0.119	0.155	0.103	0.102
Hek293 Cytotoxicity	0.183	0.139	0.151	0.211	0.181	0.249	0.181	0.19
Drug-induced Neurotoxicity	0.487	0.545	0.675	0.72	0.558	0.648	0.598	0.645
Environmental Toxicity								
BCF	1.834	2.719	2.647	2.444	2.972	2.738	2.928	2.738
IGC_{50}	3.75	3.992	3.903	3.716	4.164	4.021	4.188	4.088
$LC_{50}FM$	4.418	4.427	4.399	4.213	4.695	4.535	4.862	4.694
$LC_{50}DM$	4.578	4.282	4.503	4.346	4.761	4.73	4.972	4.977
Tox21 Pathway								
NR-AhR	0.005	0.001	0.001	0.003	0.002	0.002	0.002	0.001
NR-AR	0.0	0.0	0.0	0.001	0.001	0.0	0.001	0.0
NR-AR-LBD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NR-Aromatase	0.0	0.0	0.0	0.0	0.001	0.0	0.0	0.0
NR-ER	0.17	0.273	0.132	0.074	0.241	0.083	0.211	0.098
NR-ER-LBD	0.0	0.0	0.0	0.0	0.002	0.0	0.001	0.0
NR-PPAR-gamma	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SR-ARE	0.01	0.001	0.004	0.001	0.01	0.001	0.007	0.001
SR-ATAD5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SR-HSE	0.006	0.001	0.004	0.0	0.021	0.0	0.007	0.0
SR-MMP	0.866	0.037	0.061	0.019	0.187	0.03	0.08	0.012
SR-p53	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

***** The abbreviations are defined as: Tox, Toxicity; sens, Sensitization; BCF, the unit of bioconcentration factors, IGC_{50}, $LC_{50}FM$, and $LC_{50}DM$ are given in -Log10[(mg/L)/(1000xMW)].