Computational evaluation of cymenes: substituent group effect, pharmacokinetics, and drug-likeness

G. Serdaroğlu

¹Sivas Cumhuriyet University, Faculty of Education, Math. and Sci. Edu., 58140, Sivas, Turkey

Accepted: August 17, 2024

In this work, the *p*-cymene (CYM) and its less common isomers, ortho- and meta-cymene (OCYM and MCYM), were investigated using *in silico* tools to predict and elucidate the physicochemical, electronic, and pharmacokinetic properties, which would be helpful in early-stage drug-design. First, the −CH₃, -NH₂, and -C≡N functionalized cymenes were optimized and verified using frequency computations, at B3LYP/6-311 G** level. Then, lipophilicity, water solubility, pharmacokinetics, and drug-likeness scores of the compounds were evaluated in light of the *in silico* computations. FMO and MEP analyses of the dataset were performed to depict the possible reactivity directions and sites.

Keywords: cymene, substituent effect, DFT, pharmacokinetics, drug likeness

INTRODUCTION

Cymene is known as the type of monoterpene with the chemical formula of CH3C6H4CH(CH3)2 and is found in essential oils of various plants Thymus, Protium heptaphyllum, including Eucalyptus, Protium, etc [1,2]. Until now, they have under the spotlight due neurodegenerative potential in CNS diseases such as anxiety, Alzheimer's disease, oxidative stress, etc [3,4]. Moreover, they have been considered natural protective agents with capabilities of antioxidant, antimicrobial, anticancer, etc due to the lipophilic character allowing them to interact with cell membranes. In addition to the bio-medicinal superiorities, p-cymene is used as a precursor in the organic synthesis of bio-based solvents, green chemicals, and agrochemicals [5-7]. As well known, the optimized physicochemical properties like water solubility and hydrophobicity should be in balance with each other in designing the smart agents for biomedicinal applications, which are crucial in early-level drug-design [8,9]. Herein, functionalized ortho-, para-, and meta-cymene isomers have been investigated using computational tools to evaluate the relationship between the structure and pharmacokinetic characteristics. In this regard, the quantum mechanics simulations are employed to determine the optimized and confirmed structures, and then elucidate the thermochemical and physical properties.

Computational details

The DFT simulations of the cymene isomers were performed by the G16W [10] package at

B3LYP/6-311G** [11,12] level, and optimized geometries, FMO amplitudes and MEP plots were visualized by the GaussView 6.0.16 [13] package. The thermochemical data obtained from the frequency computations were evaluated using the basis of statistical mechanics principles [14, 15]. The HOMO and LUMO energies were used to predict the I ionization energy and A electron affinity [16] that were used to calculated the global reactivity indices, which were $\chi \rightarrow$ electronic chemical potential $\eta \rightarrow$ global hardness, $\omega \rightarrow$ electrophilicity, $\Delta N_{max} \rightarrow$ maximum charge transfer capability index [17, 18], $\omega^{-} \rightarrow$ electrodonating power, $\omega^{+} \rightarrow$ electroaccepting power [19], and $\Delta E_{back-donat.} \rightarrow back-donation energy$ [20]. The SwissADME [21] tools were used to predict the [22-26], water-solubility [27, 28], pharmacokinetics, and drug-likeness profiles of the dataset.

RESULTS AND DISCUSSION

Physicochemistry, pharmacokinetics, and drug likeness

Table 1 depicts the thermodynamic and physical parameters obtained from DFT computations. Accordingly, the ΔE , ΔH , and ΔG quantities of CYM structure were calculated at -389.399730, -389.388182, and -389.436934 au, whereas these values for MCYM and OCYM were predicted at -389.399727, -389.388194, and -389.436823 au and -389.392086, -389.380915, and -389.427065 au, respectively: the para-positioned CH3 group on the main structure lowered these quantities. Also, the E_{therm} . values of CYM, C-1, C-2, and C-3 were calculated as 139.184, 157.462, 150.560, and 139.558 kcal/mol, respectively: the -CH3 substitu-

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

tion increased the thermal energy higher than the other substituent groups (-NH₂ and -C \equiv N). A similar trend in Etherm was calculated for the less common isomers MCYM and OCYM. On the other hand, the -C \equiv N group for p-CYM derivatives increased the heat capacity and entropy more than the other substituent groups, while the -CH3 group for the MCYM and OCYM could cause an increase in the heat capacity and entropy. From Table 1, the -C≡N substituted isomers would have the highest dipole moment and polarizability index. Namely, the μ (D) and α (au) orders of CYM structures were calculated as CYM (0.087) < C-1 (0.423) < C-2 (1.666) < C-3(4.538) and CYM (110.811) < C-2 (120.859) < C-1(123.636) < C-3 (126.757), respectively. On the other hand, the μ (D) and α (au) orders for MCYM were determined as M-1< MCYM< M-2< M-3 and MCYM< M-2< M-1< M-3, respectively.

Table 2 shows the lipophilicity and solubility in water properties of the dataset. As expected, the double −CH₃ substituted derivatives would be most lipophilic among the other derivatives for all cymene isomers, as well as for all methods. Namely, the XLOGP3 and MLOGP methods for CYM and its derivatives revealed the order of lipophilicity as C-1 (4.39) > CYM (4.10) > C-3 (3.56)> C-2 (2.59) and C-1 (4.77) > CYM (4.47) > C-2 (2.76)> C-3 (2.69), respectively. Moreover, the iLOGP and WLOGP methods for MYCM derivatives were calculated as the following orders of M-1 (2.77)> MCYM (2.52)≥ M-3 (2.52) > M-2 (2.15) and M-1 (3.43)> MCYM

(3.12) > M-3 (2.99) > M-2 (2.71), respectively. From the mean lipophilicity results, the -NH₂ substituted derivatives C-2, M-2, and O-2 would exhibit less lipophilicity with the logPo/w values of 2.55, 2.58, and 2.54, respectively. As expected, the -CH₃ substitution decreased the water solubility, whereas the -NH₂ functionalization increased the water solubility. Namely, the solubility values of the CYM derivatives were calculated as C-2> C-3> C-1> C-2, depending on ESOL and Ali methods, whereas the water-solubility for OCYM derivatives based on ESOL and Ali methods was determined as O-2> O-3> OCYM> O-1 and O-2> O-3> O-1> OCYM, respectively.

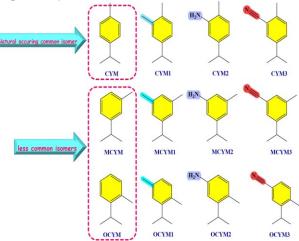


Fig. 1. Optimized chemical structures of the dataset

Table 1. Thermochemical and physical values of the data set

Comp.	ΔE (au)	ΔH (au)	ΔG (au)	E _{therm.} (kcal/mol)	Cv (cal.mol/K)	S (cal.mol/K)	μ (D)	α (au)
CYM	-389.399730	-389.388182	-389.436934	139.184	39.618	102.607	0.087	110.811
C-1	-428.698818	-428.685861	-428.736544	157.462	45.501	106.671	0.423	123.636
C-2	-444.755728	-444.743036	-444.792871	150.560	45.409	104.886	1.666	120.859
C-3	-481.667217	-481.653999	-481.706033	139.558	45.581	109.515	4.538	126.757
MCYM	-389.399727	-389.388194	-389.436823	139.187	39.631	102.349	0.248	109.946
M-1	-428.700113	-428.686603	-428.741403	157.410	45.761	115.337	0.123	123.540
M-2	-444.756239	-444.743220	-444.794609	150.463	45.724	108.158	1.495	120.701
M-3	-481.666852	-481.653494	-481.706760	139.532	45.672	112.108	5.136	126.957
OCYM	-389.392086	-389.380915	-389.427065	139.421	39.469	97.132	0.454	108.847
O-1	-428.692419	-428.679299	-428.730773	157.637	45.595	108.337	0.236	122.804
O-2	-444.747960	-444.735255	-444.784605	150.689	45.543	103.866	1.478	120.029
O-3	-481.659594	-481.646603	-481.697295	139.764	45.492	106.690	5.370	126.714

Table 2. Lipophilicity and water solubility

	CYM	C-1	C-2	C-3	M-CYM	M-1	M-2	M-3	O-CYM	O-1	O-2	O-3
Lipophilicity												
iLOGP	2.51	2.70	2.17	2.49	2.52	2.77	2.15	2.52	2.43	2.70	2.08	2.44
XLOGP3	4.10	4.39	2.59	3.56	4.50	3.78	2.73	3.13	4.38	4.39	2.60	3.13
WLOGP	3.12	3.43	2.71	2.99	3.12	3.43	2.71	2.99	3.12	3.43	2.71	2.99
MLOGP	4.47	4.77	2.76	2.69	4.47	4.77	2.76	2.69	4.47	4.77	2.76	2.69
SILICOS-IT	3.29	3.77	2.55	3.27	3.29	3.77	2.55	3.27	3.29	3.77	2.55	3.27
Avg. LogPo/w	3.50	3.81	2.55	3.00	3.58	3.70	2.58	2.92	3.54	3.81	2.54	2.90
Water solubility												
Log S (ESOL)	-3.63	-3.86	-2.73	-3.37	-3.89	-3.48	-2.82	-3.10	-3.81	-3.86	-2.74	-3.10
Sol.(mg/mL)×10 ⁻²	3.12	2.03	27.5	6.73	1.75	4.93	22.4	12.6	2.08	2.03	27.1	12.6
Class	S	S	S	S	S	S	S	S	S	S	S	S
Log S (Ali)	-3.81	-4.11	-2.79	-3.75	-4.22	-3.47	-2.93	-3.30	-4.10	-4.11	-2.80	-3.30
Sol.(mg/mL)×10 ⁻²	2.10	1.16	24.5	2.86	0.807	4.98	17.5	8.00	1.08	1.16	23.9	8.00
Class	S	MS	S	S	MS	S	S	S	MS	MS	S	S
SILICOS-IT)	-3.57	-3.97	-3.23	-3.68	-3.57	-3.97	-3.23	-3.68	-3.57	-3.97	-3.23	-3.68
Sol. (mg/mL)×10 ⁻²	3.58	1.57	8.81	3.32	3.58	1.57	8.81	3.32	3.58	1.57	8.81	3.32
Class	S	S	S	S	S	S	S	S	S	S	S	S

Table 3. Pharmacokinetics

	GI Abs.	BBB	P-gp susbt.	CYP1A2 inh.	CYP2C19 inh.	CYP2C9 inh.	CYP2D6 inh.	CYP3A4 inh.	Log Kp (skin per.) cm/s
CYM	Low	Yes	No	No	No	No	Yes	No	-4.21
C-1	Low	Yes	No	No	No	No	Yes	No	-4.09
C-2	High	Yes	No	Yes	No	No	No	No	-5.37
C-3	High	Yes	No	Yes	No	No	No	No	-4.74
MCYM	Low	Yes	No	No	No	No	Yes	No	-3.92
M-1	Low	Yes	No	No	No	No	Yes	No	-4.52
M-2	High	Yes	No	Yes	No	No	No	No	-5.27
M-3	High	Yes	No	Yes	No	No	No	No	-5.05
OCYM	Low	Yes	No	No	No	No	Yes	No	-4.01
O-1	Low	Yes	No	No	No	No	Yes	No	-4.09
O-2	High	Yes	No	Yes	No	No	No	No	-5.36
O-3	High	Yes	No	Yes	No	No	No	No	-5.05

According to Table 3, the -NH₂ and -C≡N substituted structures would exhibit high GI-absorption potency, while the -CH₃ substituted structures would have less potency in terms of it. Also, all compounds would have suitable structural and physicochemical properties for passive permeation through the BBB, which could be seen from Figure 2 as well. On the other hand, the studied derivatives would not be effluated from the CNS by the glycoprotein and thus exhibit P-gp substrate. Also, the -NH₂ and -C≡N substituted structures would have potency in terms of CYP1A2 inhibition, while the -CH3 substituted structures would not have. Moreover, none of the compounds would have a potency in terms of CYP2C19, CYP2C9, and

CYP3A4 inhibition. As is well known, Kp (skin permeability) is defined as the penetration rate of a chemical substance relevant across the stratum corneum, and lipophilicity has a critical role in the skin absorption of a specific molecular system [29-31]. Herein, the log P values of CYM, MCYM, and OCYM derivatives were calculated in the ranges of (-4.09)-(-5.37), (-3.92)-(-5.27), and (-4.01)-(-5.36)cm/s, respectively: the most lipophilic structures would have higher skin permeation where as the most water-soluble structures would have the less potency in terms of skin permeation. From Table 4, the Veber and Egan rules revealed that all compounds could have structural physicochemical necessities for drug-like potency. Also, the Lipinski rules implied that all compounds would have proper properties for drug likeness, even though the MLOGP indexes of the –CH₃ substituted structures could be higher than 4.15. On the other hand, the MW of the compounds would be lower

than 160 g/mol, which would be a violation of drug likeness potency depending on the Ghose approach. Moreover, Muegge's approach gave two violations: MW<200 g/mol and the number of heteroatoms<2.

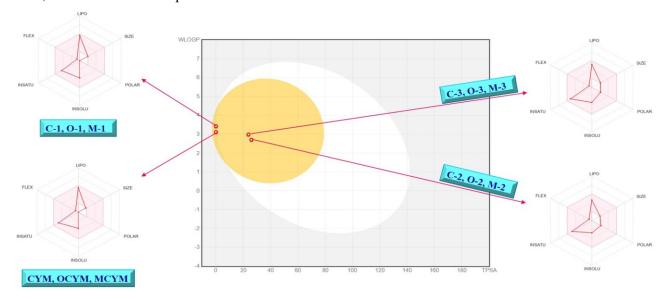


Fig. 2. Boiled-egg model and radar graphs

Table 4. Drug likeness and bioavailability scores

	Lipinski	Ghose	Veber	Egan	Muegge	Bioavail. score
CYM	Yes; MLOGP>4.15	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
C-1	Yes; MLOGP>4.15	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
C-2	YES	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
C-3	YES	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
M-CYN	M Yes; MLOGP>4.15	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
M-1	Yes; MLOGP>4.15	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
M-2	YES	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
M-3	YES	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
O-CYM	Yes; MLOGP>4.15	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
O-1	Yes; MLOGP>4.15	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
O-2	YES	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55
O-3	YES	No; MW<160	YES	YES	No; MW<200, Heteroatoms<2	0.55

Table 5. 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	ΔE _{back} . / eV
CYM	-6.394	-0.176	6.218	-3.285	3.109	0.064	0.018	0.138	1.057	-0.777
C-1	-6.265	-0.079	6.186	-3.172	3.093	0.060	0.016	0.132	1.026	-0.773
C-2	-5.472	0.122	5.595	-2.675	2.797	0.047	0.011	0.109	0.956	-0.699
C-3	-7.015	-1.534	5.480	-4.274	2.740	0.123	0.057	0.214	1.560	-0.685
MCYM	-6.507	-0.168	6.340	-3.337	3.170	0.065	0.018	0.140	1.053	-0.792
M-1	-6.422	-0.080	6.342	-3.251	3.171	0.061	0.016	0.136	1.025	-0.793
M-2	-5.479	0.041	5.520	-2.719	2.760	0.049	0.012	0.112	0.985	-0.690
M-3	-7.199	-1.528	5.671	-4.363	2.836	0.123	0.056	0.217	1.539	-0.709
OCYM	-6.508	-0.167	6.341	-3.337	3.170	0.065	0.018	0.140	1.053	-0.793
O-1	-6.275	-0.089	6.185	-3.182	3.093	0.060	0.016	0.133	1.029	-0.773
O-2	-5.395	0.003	5.398	-2.696	2.699	0.049	0.012	0.111	0.999	-0.675
O-3	-7.062	-1.477	5.585	-4.270	2.792	0.120	0.054	0.211	1.529	-0.698

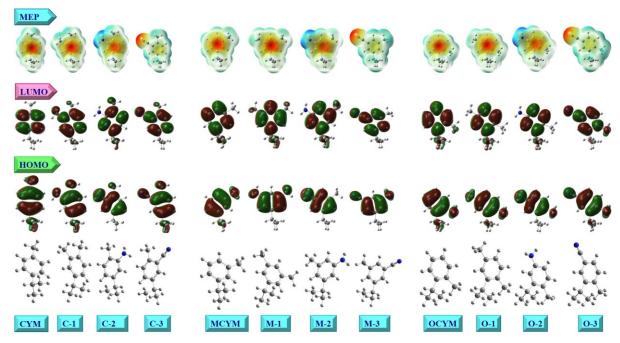


Fig. 3. Optimized structures, HOMO, LUMO, and MEP diagrams of the data set

FMO and MEP analyses

The FMO analyses provide an insight into chemical reactivity trends and sites for the molecular systems relevant. According to Table 5, the energy gap order of CYM and its derivatives was calculated as CYM (6.218) > C-1 (6.186) > C-2 (5.595) > C-3(5.480): the core cymene structure would prefer the intermolecular actions instead of the intramolecular charge movement from HOMO to LUMO due to having biggest energy gap value. Moreover, the ΔE_{L} H orders of MCYM and OCYM derivatives were determined as M-1 (6.342) > MCYM (6.340) > M-3(5.671) > M-2 (5.520) and OCYM (6.341) > O-1(6.185) > O-3 (5.585) > O-2 (5.398), respectively. Among all structures, the M-1 would have the highest energy gap, while the O-2 could have the lowest ΔE_{L-H} value. Moreover, the μ (eV) values implied that the -C≡N substituted compounds C-3 (-4.274), M-3 (-4.363), and O-3 (-4.270) could be more stable than the other substituted structures due to having the lowest values, and vice versa for -NH2 substituted counterparts. The η (eV) order of the compounds was calculated as η (eV): M-1 (3.171) > MCYM = OCYM (3.170) > CYM (3.109) > C-1 = O-1 (3.093) > M-3 (2.836) > C-2 (2.797) > O-3(2.792)> M-2 (2.760) > C-3 (2.740) > O-2 (2.699): the M-1 structure would be the hardest molecule, while the O-2 would be the softer one among the compounds. The ω (au) values of the -C \equiv N substituted compounds C-3 (0.123), M-3 (0.123), and O-3 (0.120) would be the highest value among their groups, due to the lone pair of the N atom. Moreover, the C-3 (1.560), M-3 (1.539), and O-3

(1.529) compounds would have the highest charge transfer capability (eV) among the compounds. Last, CYM (-0.777 eV), MCYM (-0.792 eV), and OCYM (-0.793 eV) structures would gain more stability via back donation due to having the lowest ΔE_{back} values.

Also, the HOMO of CYM, C-1, and C-3 derivatives covered the whole surface, whereas the HOMO of C-2 was densified on the substituted aromatic ring mostly except for the isopropyl group. Also, the isopropyl group for the M-1, M-2, M-3, and O-2 structures could not be host to HOMO. while the HOMO for MCYM, OCYM, O-1, and O-3 structures was expanded on the isopropyl group, more or less, in addition to the aromatic ring. The LUMO for C-2, M-2, and O-2 molecules would not distributed on the -NH2 group, which implied that this group could not have a role in electrophilic attack reactions. Except for the -C≡N functionalized derivatives, the aromatic ring for all compounds would be covered by red color (V<0) as a sign of the electron rich region that depicted the suitable region for electrophiles. Also, the Hs of –NH2 group were covered by blue (V>0) as a marker of the electronpoor region that implied the suitable site for nucleophiles.

CONCLUSION

In this work, the DFT computations revealed that the -C≡N substitution on the cymene isomers could gain the core structure more polarizability as well as responsible for the highest dipole moment of C-3, M-3, and, O-3 structures. On the other hand, -NH₂ functionalized structures would exhibit more

solubility in water, whereas the $-CH_3$ substitution gain to the structure relevant more lipophilicity, as expected. MEP plots of the dataset implied that the aromatic ring for all compounds would be covered by red color (V<0) that depicted the suitable region for electrophiles, except for $-C\equiv N$ decorated derivatives. Moreover, the Hs of $-NH_2$ group were covered by blue (V>0) that showed the suitable site for nucleophiles.

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 Cumhuriyet University (Project No: EĞT-2023-098).

REFERENCES

- A. Marchese, C.R. Arciola, R. Barbieri, A.S. Silva, S.F. Nabavi, A.J. Tsetegho Sokeng, M. Izadi, N.J. Jafari, I. Suntar, M. Daglia, S.M. Nabavi, *Materials*, 10(8), 947 (2017).
- A. Balahbib, N. El Omari, N. El Hachlafi, F. Lakhdar, N. El Menyiy, N. Salhi, H.N. Mrabti, S. Bakrim, G. Zengin, A. Bouyahya, Food Chem. Toxicol., 153, 112259 (2021).
- 3. T.M. de Oliveira, R.B.F. de Carvalho, I.H.F. da Costa, G.A.L. de Oliveira, A.A. de Souza, S.G. de Lima, R.M. de Freitas, *Pharm. Biol.*, **53**(3), 423 (2015).
- 4. B. Seifi-Nahavandi, P. Yaghmaei, S. Ahmadian, M. Ghobeh, A. Ebrahim-Habibi, *J. Diabetes Metab. Disord.*, **19**, 1381 (2020).
- 5. L. Caputo, G. Amato, L. De Martino, V. De Feo, F. Nazzaro, *Int. J. Mol. Sci.*, **24**(7), 6073 (2023).
- 6. M. Kamitsou, G.D. Panagiotou, K.S. Triantafyllidis, K. Bourikas, A. Lycourghiotis, C. Kordulis, *Appl. Catal. A Gen.*, **474**, 224 (2014).
- 7. M.A. Martín-Luengo, M. Yates, M.J. Martínez Domingo, B. Casal, M. Iglesias, M. Esteban, E. Ruiz-Hitzky, *Appl. Catal. B. Environ.*, **81**(3–4), 218 (2008).
- 8. B. Testa, P. Crivori, M. Reist, P.A. Carrupt, *Perspect. Drug Discov. Des.*, **19**, 179 (2000).

- 9. T. Chmiel, A. Mieszkowska, D. Kempińska-Kupczyk, A. Kot-Wasik, J. Namieśnik, Z. Mazerska, *Microchem J.*, **146**, 393 (2019).
- 10. M. J. Frisch et al. Gaussian 09W, Revision D.01, Gaussian, Inc, Wallingford CT, 2013.
- 11. A. D. Becke, J. Chem. Phys., 98, 1372 (1993).
- 12. C. Lee, W. Yang, R.G. Parr, *Phys. Rev.*, **B37**, 785 (1988).
- 13. GaussView 6.0.16, Gaussian, Inc, Wallingford CT, (2016).
- 14. D. A. McQuarrie, Statistical Thermodynamics, Harper & Row Publishers, New York, **1973**.
- 15. G. Serdaroglu, S. Durmaz, *Indian J. Chem.*, **49**, 861 (2010)
- 16. T. Koopmans, *Physica*, 1, 104 (1934).
- 17. R. G. Pearson, *Proc. Natl. Acad. Sci. USA.*, **83**, 8440 (1986).
- 18. R. G. Parr, L.V. Szentpaly, S. Liu, *J. Am. Chem. Soc.*, **121**, 1922 (1999).
- 19. J. L. Gazquez, A. Cedillo, A. Vela, *J. Phys. Chem. A.*, **111**(10), 1966 (2007).
- 20. 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).
- A. Daina, O. Michielin, V. Zoete, Sci Rep., 7, 42717 (2017).
- 22. A. Daina, O. Michielin, V. Zoete, *J. Chem. Inf. Model.* **54**(12), 3284 (2014).
- 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).
- Silicos-IT. [Online]. Available: https://www.silicosit.be.
- 27. J. 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. Y. Chang, C. Chen, C. Chen, Procedia Eng., **45**, 875 (2011).
- 30. P. Chen, C. Chen, W. Huang, C. Chang, Molecules, **23**, 911 (2018).
- 31. S. Soriano-Meseguer, E. Fuguet, A. Port, M. Rosés, *ADMET DMPK*, **8**, 16 (2020).