Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 19 (3) 1-18: Fall 2022
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

QSAR study of camptothecin derivatives as anticancer drugs using genetic algorithm and multiple linear regression analysis

Shahaboddin Mohebbi[1]. Fatemeh Shafiei[1]*, Tahereh Momeni Isfahani[1], Mehdi Ahmadi Sabegh [2]

[1] Department of Chemistry, Science Faculty, Islamic Azad University, Arak, Iran
[2] Department of Chemistry, Science Faculty, Islamic Azad University, Ahar, Iran

Received January 2023; Accepted March 2023

ABSTRACT

A quantitative structure- activity relationship (QSAR) has been widely used to investigation a correlation between chemical structures of molecules to their activities. In the present study, QSAR models have been carried out on 76 camptothecin (CPT) derivatives as anticancer drugs to determine the ^{14}N nucleus quadrupole coupling constants (QCC). These quantum chemical properties have been calculated using Density Functional Theory (DFT) and B3LYP/6-311G (d, p) method in the gas phase. A training set of 60 CPT derivatives were used to construct QSAR models and a test set of 16 compounds were used to evaluate the build models that were made using multiple linear regression (MLR) analysis. Molecular descriptors were calculated by Dragon software, and the stepwise multiple linear regression and the Genetic algorithm (GA) techniques were used to select the best descriptors and build QSAR models respectively. QSAR models were used to delineate the important descriptors responsible for the properties of the CPT derivatives. The statistically significant QSAR models derived by GA-MLR analysis were validated by Leave-One-Out Cross-Validation (LOOCV) and external validation methods. The multicollinearity of the descriptors contributed in the models was tested by calculating the variance inflation factor (VIF) and the Durbin–Watson (DW) statistics. The predictive ability of the models was found to be satisfactory. The results of QSAR study show that quantum parameters, 2D autocorrelations and Walk and path counts descriptors contains important structural information sufficient to develop useful predictive models for the studied activities.

Keywords: Camptothecin (CPT) derivatives; QSAR; Quantum parameters; GA-MLR; molecular descriptors; Leave-One-Out Cross-Validation; Nuclear quadrupole coupling constants (QCC)

1. INTRODUCTION

Camptothecin (CPT) is the first topoisomerase I inhibitory drug isolated from the Chinese tree, *Camptotheca acuminate* [1], a tree native to China used

*Corresponding author: f-shafiei@iau-arak.ac.ir

In traditional Chinese medicine. Camptothecin and some of its analogues are used to design new compounds for the development of anticancer drugs. CPT is a unique pentacyclic quinolone alkaloid (Figure 1) with an asymmetrical center at ring E. The pentacyclic rings A, B, and C includes a pyrrolo-quinoline moiety, ring D is a conjugated pyridone and ring E is a six-membered lactone with a α-hydroxyl group [2,3]. CPT showed anticancer activity in preliminary clinical trials, especially against breast, ovarian, colon, lung, and stomach cancers [4,5].The CPT keto analogues with a five-membered E-ring lacking the lactone oxygen show improved cytotoxicity toward leukemia cells, breast, and prostate cancer cells compared with CPT [6-8].

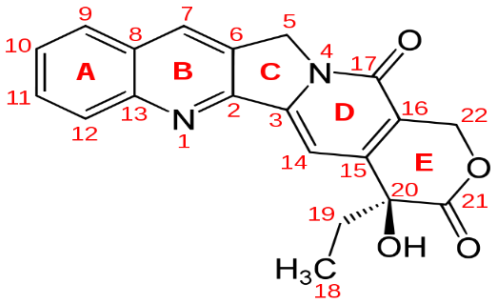


Fig. 1. The template structure of Camptothecin used in the present study.

CPT and its analogs such as Topotecan and Irinotecan bind to a complex formed by DNA with the DNA Topo-I enzyme, which inhibits tumor growth [9-11].

Quantitative Structure-Activity Relationships (QSARs) methods are focus on the quantitative relation between chemical properties and biological activity for prediction of biological activity untested and sometimes not available for the combination [12,13]. The QSAR models based on this calculation of theoretical parameters are very vast, which are known as molecular structure descriptors. Molecular descriptors play a fundamental role in making models for chemistry, pharmaceutical sciences, health research, quality control, toxicology, etc. A molecular descriptor transforms chemical information encoded within a symbolic representation of a molecule into a useful number [14]. There are a large number of molecular descriptors that can be used in QSAR studies, but only some of them are meaningful and associated with biological activity and chemical property [15].

The highest occupied molecular orbital energy (HOMO) and the lowest unoccupied molecular orbital energy (LUMO) and differences between them, the dipole moment, electron affinity, average polarizability, ionization potential, electronegativity, total energy and electronegativity were used to calculate the quantum mechanical descriptors for established QSAR models for prediction the biological activity of sulfonamide compounds [16].

The partial atomic charge of the main nucleus as descriptor has been used to predict the activity of drugs [17-20].

The function and the role of the nitrogen nucleus of aminoadamantane have been study clinically used for Alzheimer illness [21,22].

The nuclear quadrupole resonance (NQR) spectroscopy provides a sensitive technique for evaluating the electric

field gradient (EFG) [23,24] and can be applied for better characterization of structural properties [25,26]. In addition, the NQR spectroscopy was interested due to its ability to straightly record the variations in the electronic environment of the nucleus [21,27,28]. Furthermore, in NQR spectroscopy, the interaction between nuclear electric quadrupole moments, (eQ) of the quadrupole nuclei (having spin I > 1/2) and the local molecular EFG has a special characteristic role [29, 30].

The nuclear quadrupole coupling constant (NQCC) is calculated through the

NQR frequency and it provides information about the electron distribution in the molecule [31]. Also, the local distribution of the electron density around the nucleus using the NQCC is possible, when compared with the NMR chemical shift [32,33].

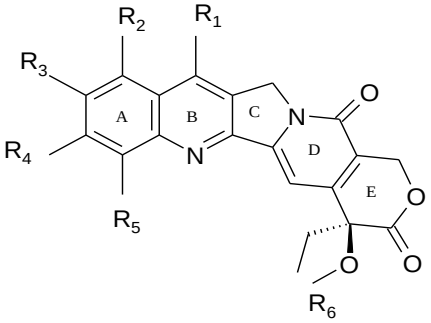
In this research the electronic structures of some pharmaceutical derivatives of 76 Camptothecin are investigated by density functional theory (DFT). The nuclear quadrupole coupling constant for N nucleus are evaluated, and correlations between molecular descriptors and NQCC are investigated.

2. MATERIALS AND METHODS

Computational details

The camptothecin discussed in the present study consist of 76 CPT derivatives with substitution at 1, 5 and 6 position. These substitutions of the 76 congeners are listed in Table **1**. In order to build and test QSAR models, a data set of 76 CPT derivatives was randomly separated into 2 groups: a training set of 60 compounds, which was used to build model and a test set of 16 compounds, which was applied to evaluate the built model.

Table 1. Substitutions at the 1-6 positions in the template structure of camptothecin derivatives



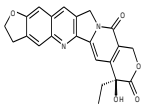
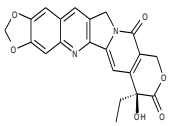
Comp No	R1	R2	R3	R4	R5	R6
1 (CPT)	H	H	H	H	H	H
2	H	H	OH	H	H	H
3	H	H	NH_2	H	H	H
4	H	H	OH	F	H	H
5	H	H	CH_2OH	H	H	H
6	C_2H_5	H	H	H	H	H
7	C_2H_5	H	OH	H	H	H
8	H	NO_2	H	H	H	H
9		--	--	-	-	-
10	H	H	OCH_3	OH	H	H
11	H	H	H	OCH_3	OH	H
12		--	--	--	--	--
13	NO_2	H	H	H	H	H
14	NH_2	H	OCH_3	F	H	H
15	CH_3	H	OCH_3	F	H	H
16	$CH_2N(CH_3)_3$	H	OH	H	H	H
17	$Si(CH_3)_3$	H	OH	H	H	H
18	$CHO_2(CH3)_2$	H	H	H	H	H

Table 1 continued

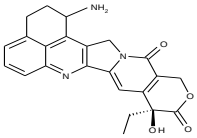
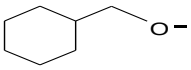
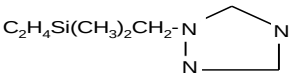
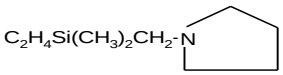
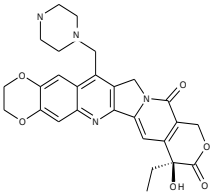
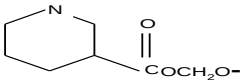
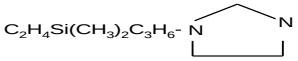
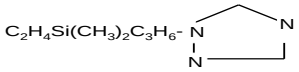
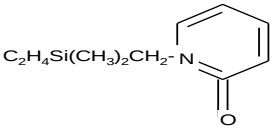
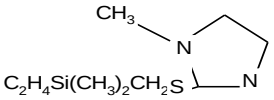
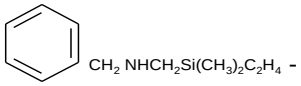
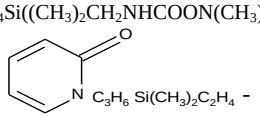
19	$Si(CH_3)_3$	H	H	F	H	H
20	$Si(CH_3)_3$	H	NH_2	H	H	H
21	$Si(CH_3)_3$	H	H	NH_2	H	H
22	$C_2H_4N(CH_3)_2$	H	H	H	H	H
23		--	--	--	--	--
24	C_2H_4 $Si(CH_3)_3$	H	H	H	H	H
25	H	$N(CH_3)_2$	OH	H	H	H
26	$Si(CH_3)_3$	H	NH2	F	H	H
27	$Si(CH_3)_3$	F	NH2	H	H	H
28	$Si(CH_3)_3$	H	H	F	F	H
29	$Si(CH_3)_3$	F	F	H	H	H
30	$C_2H_4Si((CH_3)_2OH)$	H	H	H	H	H
31	$C_2H_4Si((CH_3)_2C_2H_3)$	H	H	H	H	H
32	$C_2H_4Si((CH_3)_2CH_3I)$	H	H	H	H	H
33	$C_2H_4Si((CH_3)_2CH_3OH)$	H	H	H	H	H
34	H	H	O(CH2CF3)	H	H	H
35	C_2H_4 $Si(CH_3)_3$	H	H	H	H	H
36	$Si(CH_3)_3$	H	OCOOCH3	H	H	H
37	$Si(CH_3)_3$	H	H	F	$Si(CH_3)_3$	H
38	$C_2H_4Si((CH_3)_2C_3H_6OH)$	H	H	H	H	H
39	$C_2H_4Si((CH_3)_2C_3H_6Br)$	H	H	H	H	H
40	$C_2H_4Si((CH_3)_2C_3H_6I)$	H	H	H	H	H
41	$C_2H_4Si((CH_3)_2CH_2N_3)$	H	H	H	H	H
42	C_2D_4 $Si(CH_3)_3$	H	H	H	H	H
43	C_2H_4 $Si(CH_3)_3$	H	H	H	H	OCH_2COOCH_3
44	$C_2H_4Si((CH_3)_2CH_2OCOOCH_3)$	H	H	H	H	OCH_2COOCH_3
45	$C_2H_4Si((CH_3)_2NH_2CH_3)$	H	H	H	H	OCH_2COOCH_3
46	C_2H_5	H		H	H	H
47	$C_2H_4Si((CH_3)_2C_3H_6N(CH_3)_2$	H	H	H	H	H
48		H	H	H	H	H
49		H	H	H	H	H
50		--	--	--	--	--
51	H	H	H	H	H	$COO(CH(CH_3)(NH_2))$
52	$Si(CH_3)_3$	H	$NHCOO(O(CH_3)_3)$	H	H	H
53	$C_2H_4Si((CH_3)_2C_3H_6Br)$	H	H	H	H	$OCOOCH_3$
54	$C_2H_4Si((CH_3)_2CH_2N_3)$					$OCOOCH_3$
55	$C_2H_4Si((CH_3)_2CH_2NHCOON(CH_3)_2)$	H	H	H	H	H
56	C_2H_5	H		H	H	H
57	$Si((CH_3)_3$	H	$NHCO_2C(CH_3)_3$	F	H	H
58	$Si((CH_3)_3$	F	$NHCO_2C(CH_3)_3$	H	H	H
59		H	H	H	H	H

Table 1 continued

60		H	H	H	H	H
61		H	H	H	H	H
62		H	H	H	H	H
63		H	H	H	H	H
64	$C_2H_4Si((CH_3)_2C_6H_6PH(OCH_3)_2$	H	H	H	H	H
65	$C_2H_4Si((CH_3)_2C_6H_6N_3)$	H	H	H	H	$OCOOCH_3$
66	$C_2H_4Si((CH_3)_2C_6H_6NHCOON(CH_3)_2)$	H	H	H	H	H
67	$C_2H_4Si((CH_3)_2CH_2NHCOON(CH_3)_2)$	H	H	H	H	H
68		H	H	H	H	H
69	$C_2H_4Si((CH_3)_2CH_2NHCOOC(F)_3)$	H	H	H	H	H
70	$C_2H_4Si((CH_3)_2CH_2NHCOON(CH_3)_2)$	H	H	H	H	COOCH3
71	$C_2H_4Si((CH_3)_2CH_2N(C2H5)COONH(CH_3))$	H	H	H	H	H
72	$C_2H_4Si((CH_3)_2C_3H_6NHCOON(CH_3)_2)$	H	H	H	H	COOCH3
73	$C_2H_4Si((CH_3)_2NHN(CH_3))$	H	H	H	H	COOCH3
74	$C_2H_4Si((CH_3)_2CH_2NHCOON(C_2H_5)_2)$	H	H	H	H	COOCH3
75	$C_2H_4Si((CH_3)_2C_3H_6NHCOOCF)_3)$	H	H	H	H	COOCH3
76	$C_2H_4Si((CH_3)_2C_3H_6NHCOON(CH_3)_2$	H	H	H	H	COOCH3

In this study, the Gauss View software [34] was used to draw all the CPT derivatives. The geometric optimizations compounds were carried out using GAUSSIAN 09W program [35] at the B3LYP [36,37] level of theory and by employing 6-311G (d,p) [38,39] standard basis sets. These methods have been used to optimized many other similar compounds successfully and have published good results [40].

The nuclear quadrupole coupling constants of N_1 and N_4 were calculated using EFG tensor (qxx, qyy and qzz) obtained from DFT [36] computations by the following equation [41].

$$NQCC = \frac{e^2 Qq_{zz}}{h} \quad (1)$$

where, NQCC, e, Q and h are nuclear quadrupole coupling constant (in MHz), the electronic charge (in atomic unit), the nuclear quadrupole moment of the ^{14}N nuclei (in millibarn) and the Plank's constant, respectively. The values of NQCCs for N1 and N4 are listed in Table **2**.

In order to calculate molecular descriptors, the optimized structures were entered in Dragon5.4- 2006 package. A total of 1822 theoretical descriptors were calculated for each compound by Dragon software.

The energies of frontier orbitals (in eV), HOMO and LUMO were used to determine several chemical reactivity parameters as quantum chemical descriptors [42]. These descriptors include ionization potential (I)[43], electron affinity (A)[44], chemical hardness (η)[45], chemical softness (S)[46], electronic chemical potentials (μ)[47], and electrophilicity (ω) [48]. The values of

these descriptors are determined through the following expressions and are recorded in Table **3**,

$I = -E\text{HOMO}$ (2)

$A = -E\text{LUMO}$ (3)

$\eta = (I - A)/2$ (4)

$\mu = (I + A)/2$ (5)

$S = 1/\eta$ (6)

$\omega = \mu^2/2\eta$ (7)

The Genetic algorithms (GA) are written in MATLAB (version 2010a) and backward multiple linear regression analysis (BW-MLR) method using the Statistical Package for the Social Science (SPSS) software was used to reduce the number of molecular descriptors [49-54].

The MLR models were obtained by the SPSS [55,56] statistics version 22 . In this type of regression, the molecular descriptors and quantum descriptors consider as the independent variable and the NQCC parameters considered as the dependent variable.

Table 2. The observed, predicted and residual values of the N1 and N4 nucleus QCC by GA- MLR method

No	QCC(N1) (Obs)	QCC (N1) (Pred)	QCC (N1) (Res)	QCC (N4) (Obs)	QCC (N4) (pred)	QCC (N4) (Res)
		Hz			Hz	
1	7.1450	7.1404	0.0046	7.4996	7.4385	0.0611
2	6.8109	6.8208	-0.0099	7.5712	7.5944	-0.0232
3	7.1428	7.1496	-0.0068	10.3512	10.7203	-0.3691
4	7.2085	7.2706	-0.0622	10.1516	10.3594	-0.2078
5	7.1390	7.0948	0.0443	10.1597	10.1213	0.0384
6	6.9545	6.9402	0.0143	9.8543	9.7537	0.1007
7	6.9643	6.9633	0.0010	9.9567	9.9930	-0.0362
8	7.2273	7.2871	-0.0597	10.0431	10.0631	-.0.0200
9	7.1534	7.1992	-0.0458	10.2159	10.2537	-0.0378
10	7.1549	7.1972	-0.0423	10.1154	10.1129	0.0026
11	7.1446	7.1963	-0.0517	10.5299	10.5507	-0.0208
12	7.1801	7.1403	0.0398	10.1175	10.1196	-0.0021
13	7.1986	7.1982	0.0004	10.7568	10.7496	0.0072
14	7.1019	7.1625	-0.0606	8.8343	8.7235	0.1108
15	7.1221	7.1876	-0.0655	9.9351	9.9768	-0.0418
16	7.1050	7.1873	-0.0824	10.1425	10.1677	-0.0252
17	7.0333	7.0106	0.0228	10.2521	10.2619	-0.0098
18	6.8019	6.8335	-0.0316	10.1153	10.1248	-0.0095
19	7.0313	7.0328	-0.0015	9.9928	9.9727	0.0201
20	6.9413	6.9680	-0.0267	10.2871	10.2743	0.0128
21	6.2887	6.2934	-0.0047	9.2042	9.2062	-0.0020
22	6.6532	6.6607	-0.0075	9.5249	9.5195	0.0054
23	6.1767	6.1811	-0.0044	9.0201	9.0578	-0.0376
24	7.0486	7.0430	0.0056	9.8285	9.8961	-0.0676
25	7.1225	7.1298	-0.0073	10.1855	10.1513	0.0342
26	6.9967	6.9768	0.0199	10.2179	10.2189	-0.0011
27	6.9344	6.9382	-0.0038	10.1710	10.1612	0.0098
28	7.0655	7.0690	-0.0035	10.2541	10.2484	0.0058
29	6.9839	6.9389	0.0450	9.9462	9.9539	-0.0077
30	6.8439	6.8434	0.0004	9.7845	9.7526	0.0319
31	7.0483	7.0489	-0.0006	9.8289	9.8269	0.0019
32	7.0418	7.0533	-0.0115	9.8175	9.8666	-0.0491
33	7.0175	7.0208	-0.0034	9.8036	9.8316	-0.0280
34	7.1523	7.1611	-0.0087	10.1950	10.1944	0.0006

Table 2 continued

35	6.8379	6.8125	0.0255	9.7777	9.7613	0.0164
36	6.9331	6.9196	0.0135	10.0396	10.0623	-0.0228
37	6.9890	6.9715	0.0175	9.5131	9.5177	-.00046
38	7.0341	7.0381	-0.0040	9.8321	9.8690	-0.0369
39	7.0497	7.0426	0.0071	9.8411	9.8655	-0.0244
40	7.0309	7.0168	0.0412	9.8229	9.8960	-0.0731
41	6.9057	6.9108	-0.0051	9.8354	9.8586	-0.0232
42	6.8501	6.8598	-0.0097	9.7899	9.7613	0.0286
43	6.9812	6.9897	-0.0085	9.7942	9.7830	0.0112
44	6.9732	6.9846	-0.0114	9.7867	9.7544	0.0323
45	6.9511	6.9411	0.0100	9.7672	9.7585	0.0087
46	7.0731	6.8393	0.2338	9.9809	9.9822	-0.0013
47	6.8227	6.8515	-0.0288	9.7674	9.7417	0.0257
48	6.9005	6.9085	-0.0080	9.8296	9.8529	-0.0234
49	6.8409	6.8561	-0.0151	9.7830	9.7852	-0.0022
50	7.0575	7.0462	0.0113	9.9790	9.9845	-0.0054
51	10.1357	7.1422	2.9936	-0.9454	9.9212	-10.8666
52	7.0461	7.0401	0.0060	10.2677	10.2501	0.0176
53	7.1316	7.1219	0.0098	9.7477	9.7691	-0.0214
54	6.9938	6.9912	0.0026	9.7586	9.7797	-0.0211
55	6.9184	6.9235	-0.0051	9.7329	9.7524	-0.0195
56	7.0838	7.0756	0.0082	9.9358	9.9800	-0.0442
57	7.0075	7.0343	-0.0268	9.9832	9.9604	0.0228
58	6.9536	6.9597	-0.0060	10.0629	10.0419	0.0210
59	6.9267	6.9275	-0.0008	9.8543	9.8920	-0.0377
60	6.9566	6.9560	0.0007	9.8828	9.8612	0.0216
61	6.8043	6.8086	-0.0043	9.7671	9.7403	0.0268
62	6.9991	6.9837	0.0154	9.7969	9.7621	0.0348
63	6.8943	6.8937	0.0006	9.8295	9.8456	-0.0161
64	6.9374	6.9351	0.0023	9.8714	9.8612	0.0102
65	6.7850	6.7932	-0.0082	9.8652	9.8706	-0.0054
66	6.8462	6.8185	0.0277	9.7999	9.7763	0.0236
67	7.0521	7.0688	-0.0167	9.8134	9.8200	-0.0065
68	6.9391	6.9342	0.0049	9.8650	9.8696	-0.0046
69	7.1243	7.1242	0.0001	9.8916	9.8968	-0.0052
70	6.4227	6.4131	0.0096	9.7661	9.7858	-0.0198
71	6.7044	6.7245	-0.0200	9.5038	9.5452	-0.0414
72	7.1413	7.1699	-0.0287	9.7679	9.7973	-0.0295
73	6.7143	6.7197	-0.0055	9.8160	9.8104	0.0056
74	6.6056	6.6032	0.0025	9.7078	9.7081	-0.0003
75	7.1821	7.1856	-0.0035	9.8073	9.9411	-0.1338
76	7.1190	7.1150	0.0040	9.7602	9.7393	0.0209

Table 3. The calculated values of 76 CPT for the energies of frontier orbitals HOMO and LUMO, ionization potential (I); electron affinity (A); chemical hardness (η), chemical softness (S); electronic chemical potentials (μ), and electrophilicity (ω) all in electron volt (eV) units

No	Homo ev	Lumo ev	IP ev	EA ev	η ev	σ ev	χ ev	μ ev	ω ev
1	-0.2272	-0.0927	0.2272	0.0927	0.0673	14.8666	0.1599	-0.1599	0.1901
2	-0.2220	-0.0892	0.2220	0.0892	0.0664	15.0614	0.1556	-0.1556	0.1824
3	-0.2275	-0.0907	0.2275	0.0907	0.0675	14.8192	0.1582	-0.1582	0.1855
4	-0.2258	-0.0940	0.2258	0.0940	0.0659	15.1642	0.1599	-0.1599	0.1938
5	-0.2247	-0.0900	0.2247	0.0900	0.0674	14.8467	0.1573	-0.1573	0.1837
6	-0.2242	-0.0893	0.2242	0.0893	0.0674	14.8291	0.1568	-0.1568	0.1822
7	-0.2189	-0.0856	0.2189	0.0856	0.0667	15.0015	0.1523	-0.1523	0.1739
8	-0.2214	-0.0877	0.2214	0.0877	0.0668	14.9645	0.1545	-0.1545	0.1787
9	-0.2108	-0.0848	0.2108	0.0848	0.0630	15.8730	0.1478	-0.1478	0.1733
10	-0.2228	-0.0882	0.2228	0.0882	0.0673	14.8577	0.1555	-0.1555	0.1796
11	-0.2217	-0.0940	0.2217	0.0940	0.0638	15.6629	0.1578	-0.1578	0.1951
12	-0.2311	-0.0992	0.2311	0.0992	0.0660	15.1619	0.1652	-0.1652	0.2068
13	-0.2258	-0.0939	0.2258	0.0939	0.0660	15.1550	0.1599	-0.1599	0.1936
14	-0.2238	-0.0879	0.2238	0.0879	0.0680	14.7070	0.1558	-0.1558	0.1786
15	-0.2239	-0.0909	0.2239	0.0909	0.0665	15.0387	0.1574	-0.1574	0.1862
16	-0.2174	-0.0853	0.2174	0.0853	0.0660	15.1435	0.1513	-0.1513	0.1734
17	-0.2189	-0.0856	0.2189	0.0856	0.0667	15.0015	0.1523	-0.1523	0.1739
18	-0.2260	-0.0908	0.2260	0.0908	0.0676	14.7973	0.1584	-0.1584	0.1856
19	-0.2276	-0.0969	0.2276	0.0969	0.0654	15.2975	0.1622	-0.1622	0.2013
20	-0.2286	-0.0626	0.2286	0.0626	0.0830	12.0496	0.1456	-0.1456	0.1277
21	-0.2175	-0.0850	0.2175	0.0850	0.0663	15.0932	0.1512	-0.1512	0.1726
22	-0.0935	-0.0489	0.0935	0.0489	0.0223	44.8632	0.0712	-0.0712	0.1137
23	-0.2154	-0.0900	0.2154	0.0900	0.0627	15.9473	0.1527	-0.1527	0.1858
24	-0.2231	-0.0880	0.2231	0.0880	0.0676	14.8038	0.1555	-0.1555	0.1790
25	-0.2199	-0.0859	0.2199	0.0859	0.0670	14.9287	0.1529	-0.1529	0.1744
26	-0.2152	-0.0864	0.2152	0.0864	0.0644	15.5316	0.1508	-0.1508	0.1767
27	-0.2146	-0.0879	0.2146	0.0879	0.0634	15.7853	0.1512	-0.1512	0.1805
28	-0.2294	-0.1001	0.2294	0.1001	0.0647	15.4655	0.1648	-0.1648	0.2099
29	-0.2293	-0.1001	0.2293	0.1001	0.0646	15.4811	0.1647	-0.1647	0.2099
30	-0.2218	-0.0857	0.2218	0.0857	0.0681	14.6951	0.1537	-0.1537	0.1736
31	-0.2228	-0.0872	0.2228	0.0872	0.0678	14.7438	0.1550	-0.1550	0.1771
32	-0.2056	-0.0955	0.2056	0.0955	0.0551	18.1571	0.1506	-0.1506	0.2058
33	-0.2208	-0.0840	0.2208	0.0840	0.0684	14.6210	0.1526	-0.1526	0.1698
34	-0.2265	-0.0955	0.2265	0.0955	0.0655	15.2672	0.1610	-0.1610	0.1979
35	-0.2231	-0.0880	0.2231	0.0880	0.0676	14.8006	0.1555	-0.1555	0.1790
36	-0.2196	-0.0900	0.2196	0.0900	0.0648	15.4273	0.1548	-0.1548	0.1848
37	-0.2280	-0.0969	0.2280	0.0969	0.0656	15.2451	0.1624	-0.1624	0.2011
38	-0.2233	-0.0869	0.2233	0.0869	0.0682	14.6649	0.1551	-0.1551	0.1763
39	-0.2263	-0.0995	0.2263	0.0995	0.0634	15.7701	0.1629	-0.1629	0.2093
40	-0.2160	-0.0847	0.2160	0.0847	0.0656	15.2352	0.1503	-0.1503	0.1722
41	-0.2370	-0.0695	0.2370	0.0695	0.0837	11.9462	0.1532	-0.1532	0.1403
42	-0.2160	-0.0890	0.2160	0.0890	0.0635	15.7451	0.1525	-0.1525	0.1830

Table 2 continued

43	-0.2289	-0.0790	0.2289	0.0790	0.0750	13.3404	0.1539	-0.1539	0.1580
44	-0.2255	-0.0996	0.2255	0.0996	0.0630	15.8839	0.1625	-0.1625	0.2098
45	-0.2237	-0.0852	0.2237	0.0852	0.0692	14.4485	0.1544	-0.1544	0.1723
46	-0.2214	-0.0746	0.2214	0.0746	0.0734	13.6166	0.1480	-0.1480	0.1491
47	-0.2596	-0.0622	0.2596	0.0622	0.0987	10.1269	0.1609	-0.1609	0.1311
48	-0.2299	-0.0900	0.2299	0.0900	0.0700	14.2959	0.1599	-0.1599	0.1828
49	-0.2166	-0.0986	0.2166	0.0986	0.0590	16.9460	0.1576	-0.1576	0.2104
50	-0.2036	-0.0906	0.2036	0.0906	0.0565	17.7112	0.1471	-0.1471	0.1916
51	-0.2138	-0.0993	0.2138	0.0993	0.0572	17.4842	0.1565	-0.1565	0.2140
52	-0.2237	-0.0903	0.2237	0.0903	0.0667	15.0034	0.1570	-0.1570	0.1849
53	-0.2265	-0.0910	0.2265	0.0910	0.0678	14.7545	0.1588	-0.1588	0.1860
54	-0.2137	-0.0993	0.2137	0.0993	0.0572	17.4937	0.1565	-0.1565	0.2142
55	-0.2037	-0.0994	0.2037	0.0994	0.0521	19.1830	0.1515	-0.1515	0.2202
56	-0.2253	-0.0790	0.2253	0.0790	0.0732	13.6659	0.1521	-0.1521	0.1581
57	-0.2187	-0.0903	0.2187	0.0903	0.0642	15.5807	0.1545	-0.1545	0.1860
58	-0.2366	-0.9050	0.2366	0.9050	0.3342	2.9922	0.5708	-0.5708	--0.4874
59	-0.2066	-0.9471	0.2066	0.9471	0.3703	2.7008	0.5769	-0.5769	-0.4494
60	-0.2279	-0.0913	0.2279	0.0913	0.0683	14.6410	0.1596	-0.1596	0.1865
61	-0.2200	-0.0882	0.2200	0.0882	0.0659	15.1791	0.1541	-0.1541	0.1803
62	-0.2265	-0.0795	0.2265	0.0795	0.0735	13.6047	0.1530	-0.1530	0.1593
63	-0.2157	-0.0695	0.2157	0.0695	0.0731	13.6769	0.1426	-0.1426	0.1390
64	-0.2237	-0.0952	0.2237	0.0952	0.0642	15.5756	0.1595	-0.1595	0.1980
65	-0.2100	-0.0757	0.2100	0.0757	0.0671	14.8930	0.1428	-0.1428	0.1519
66	-0.2137	-0.0777	0.2137	0.0777	0.0680	14.7108	0.1457	-0.1457	0.1561
67	-0.2187	-0.0897	0.2187	0.0897	0.0645	15.4990	0.1542	-0.1542	0.1842
68	-0.2279	-0.0750	0.2279	0.0750	0.1514	6.6034	0.0765	-0.0765	0.0193
69	-0.2066	-0.0732	0.2066	0.0732	0.0667	14.9904	0.1399	-0.1399	0.1467
70	-0.2237	-0.0671	0.2237	0.0671	0.0783	12.7789	0.1454	-0.1454	0.1351
71	-0.2037	-0.0785	0.2037	0.0785	0.0626	15.9818	0.1411	-0.1411	0.1591
72	-0.2297	-0.0883	0.2297	0.0883	0.0707	14.1443	0.1590	-0.1590	0.1787
73	-0.2267	-0.0790	0.2267	0.0790	0.0738	13.5415	0.1528	-0.1528	0.1581
74	-0.2255	-0.0983	0.2255	0.0983	0.0636	15.7195	0.1619	-0.1619	0.2060
75	-0.2037	-0.0758	0.2037	0.0758	0.0639	15.6471	0.1398	-0.1398	0.1528
76	-0.2279	-0.0668	0.2279	0.0668	0.0805	12.4184	0.1474	-0.1474	0.1348

3. RESULTS AND DISCUSSION

3.1. QSAR models

The GA–MLR analysis led to the derivation of 6 models for the NQCCs, with 13-18 descriptors (Table 4).

The models were evaluated with regression parameters: squared regression coefficient (R^2), adjusted correlation coefficient (R^2_{adj}), root mean squared error (RMSE), Fisher ratio (F) and Significance (Sig) [57-59].

The results of models are very satisfactory. The statistical coefficients of the six models are almost similar; so, the model 6, which has the lowest number of descriptors, has been selected. QSAR model and statistical parameters for thirty molecular descriptors is shown as follows:

Q_{CC} (N_1)/Hz = -125.979+ 0.216 (AMW) + 0.134 (VDA) -0.001 (SMTI)- 0.064(**ω**) + 0.000021 (WW) + 22.433(MWC08) - 14.447 (piPC06) - 1.523 (IDDE) - 2.236 (IC2) -3.127 (CIC5) +2.376 (GATS7v) + 2.063(η) + 1.413(GATS3e) (8)

N=60 R= 0.984 R^2=0.968 R^2_{adj}=0.963 DW=2.037 RMSE=0.273 F= 855.798 Sig=0.000

As can be seen in Table **4**, the thirty descriptors are useful to predict the NQCCs, which are: AMW, VDA, SMTI, **ω**, WW, MWC08, piPC06, IDDE, IC2, CIC5, GATS7v, η and GATS3e.

These descriptors are classified in Constitutional indices (AMW, Se, WW), topological indices (SMTI, VDA), Walk and path counts (MWC08, piPC06), information indices (IDDE, IC2, CIC5), 2D autocorrelations (GATS7v, GATS3e), and quantum descriptors (**ω**,).

The suitable linear model for the N_4 nuclear quadrupole coupling constant (QCC) includes four molecular descriptors which are: AMW, Se, **ω**and **η** (see Table **5**). These descriptors are classified in Constitutional indices (AMW,Se), and quantum descriptors (**ω**, **η**). QSAR model and statistical parameters for four molecular descriptors is shown in Equation 9.

Q_{CC} (N_4) /Hz = 15.994+1.603(AMW) + 0.513(Se) - 0.057(**ω**) - 7.369(**η**) (9)

N = 60 R= 0.937 R^2=0.878 R^2_{adj} = 0.907 DW=1.81 RMSE=1.405 F=222.148 Sig=0.000

In present study, to find the best model for predicting the mentioned properties, we will use the following sections.

3.2. Multicollinearity

The reliability and robustness of the models are checked using the autocorrelation and multicollinearity properties of the descriptors contributed in the models using the variance inflation factor (VIF) and the Pearson coefficient correlation (PCC). These parameters were calculated by SPSS program. [60].

Table 4. Statistical parameters of the models calculated with the SPSS software for the N_1 nuclear quadrupole coupling constant (QCC) of 60 CPTs

Model	Independent variables	R	R^2	R^2_{adj}	RMSE	F	Sig
1	GATS8e, CIC5, GATS3e, GATS6e, η, IDDE, AMW, GATS7v, MATS5e, SMTI, IC2, ATS3m, MWC08, ww, piPC06, ω, MW, VDA	0.986	0.972	0.963	0.273	662.623	0.000
2	GATS8e, CIC5, GATS3e, GATS6e, η, IDDE, AMW, GATS7v, MATS5e, SMTI, IC2, ATS3m, MWC08, ww, piPC06, ω, VDA	0.986	0.972	0.964	0.270	709.115	0.000
3	GATS8e, CIC5, GATS3e, GATS6e, η, IDDE, AMW, GATS7v, SMTI, IC2, ATS3m, MWC08, ww, piPC06, ω, VDA	0.986	0.972	0.964	0.269	726.628	0.000
4	GATS8e, CIC5, GATS3e, GATS6e, η, IDDE, AMW, GATS7v, SMTI, IC2, MWC08, ww, piPC06, ω, VDA	0.986	0.972	0.965	0.267	801.199	0.000
5	GATS8e, CIC5, GATS3e, η, IDDE, AMW, GATS7v, SMTI, IC2, MWC08, ww, piPC06, ω, VDA	0.985	0.970	0.964	0.268	834.629	0.000
6	CIC5, GATS3e, η, IDDE, AMW, GATS7v, SMTI, IC2, MWC08, ww, piPC06, ω, VDA	0.984	0.968	0.963	0.273	855.798	0.000

If the VIF value lies between 1 and 10, there is no multicollinearity; if VIF<1 or >10, there is multicollinearity [61, 62].

The suitable linear model for QSAR study of the N_4 nuclear quadrupole coupling constant (QCC) (Equation 1) includes thirty descriptors. The results of the correlation between these descriptors are listed in Table **6**.

From Table **6**, we can infer that, the VIF for descriptors (VDA, SMTI), (MWC08, piPC06), (**ω**, MWC08), and (SMTI, WW) are bigger than 10, and the PCC between them is close to unity, therefore there is linearity between these descriptors. After removing SMTI, WW, VDA and **ω** from this model, the new model recorded by SPSS as follows:

Q_{CC} (N_1)/Hz = 16.278 - 1.487(piPC06) + 1.136(GATS3e) (10)

N=60 R= 0.937 R^2= 0.878 R^2_{adj} = 0.910 DW=1.673 RMSE=0.426 F= 665.784 Sig=0.000

Table 5. Statistical parameters of the models calculated with the SPSS software for the N4 nuclear quadrupole coupling constant (QCC) of 60 CPTs

Model	Independent variables	R	R^2	R^2_{adj}	RMSE	F	Sig
1	MATS7p, AMW, **η**, MW, X2, SMTI, **ω**, Sp, Se, TWC, MWC07	0.954	0.910	0.913	1.357	62.830	0.000
2	MATS7p, AMW, **η**, MW, X2, SMTI, **ω**, Sp, Se, TWC	0.953	0.908	0.914	1.355	75.932	0.000
3	MATS7p, AMW, **η**, X2, SMTI, **ω**, Sp, Se, TWC	0.951	0.904	0.912	1.365	81.919	0.000
4	MATS7p, AMW, **η**, X2, **ω**, Sp, Se, TWC	0.948	0.899	0.911	1.373	101.948	0.000
5	AMW, **η**, X2, **ω**, Sp, Se, TWC	0.945	0.893	0.909	1.390	113.852	0.000
6	AMW, **η**, **ω**, Sp, Se, TWC	0.941	0.885	0.908	1.400	131.833	0.000
7	AMW, **η**, **ω**, Sp, Se	0.940	0.884	0.908	1.397	201.054	0.000
8	AMW, **η**, **ω**, Se	0.937	0.878	0.907	1.405	222.148	0.000

Table 6. Correlation between the molecular descriptors in Equation (8) for the N_1 QCC

Descriptor	GATS3e	μ	IDDE	AMW	GATS7v	SMTI	IC2	MWC08	ww	CIC5	piPC06	ω	VDA
GATS3e	1.00												
μ	-0.198	1.00											
IDDE	-0.236	-0.079	1.00										
AMW	0.349	0.155	-0.11	1.00									
GATS7v	0.096	0.255	-0.117	0.151	1.00								
SMTI	-0.044	-0.187	0.232	-0.076	-0.568	1.00							
IC2	0.361	-0.338	-0.251	0.104	-0.15	0.392	1.00						
MWC08	-0.006	0.348	-0.251	0.037	0.291	-0.551	-0.268	1.00					
ww	0.054	0.046	0.071	0.211	0.491	-0.727	-0.308	0.366	1.00				
CIC5	0.277	-0.384	0.148	0.287	-0.263	0.472	0.682	-0.439	-0.068	1.00			
piPC06	-0.058	-0.295	0.179	-0.168	-0.453	0.548	0.329	-0.968	-0.463	0.382	1.00		
ω	0.072	-0.234	0.950	-0.084	-0.195	0.469	0.138	-0.767	-0.539	0.018	0.569	1.00	
VDA	0.014	0.23	-0.295	0.069	0.516	-0.969	-0.376	0.653	0.665	-0.452	-0.578	-0.626	1.00

Descriptor	Tolerance	VIF1	VIF2	VIF3
GATS3e	0.586	2.066	1.348	1.038
μ	0.453	2.205	-	-
IDDE	0.194	5.165	3.560	-
AMW	0.484	2.066	-	-
GATS7v	0.417	2.397	1.452	-
SMTI	0.001	1215.339	-	-
IC2	0.134	7.471	4.823	-
MWC08	0.017	57.396	11.170	-
ww	0.014	72.697	-	-
CIC5	0.1	10.001	4.352	-
piPC06	0.087	11.507	5.191	1.038
ω	0.005	199.382	-	-
VDA	0.001	1263.979	-	-

For the the N_4 QCC (Equation 2), the Pearson correlation between (Se, **ω**) and (**ω**, η) is close to the unit and VIF for **ω** and Se descriptors are bigger than 10(see Table **7**). After removing Se index, we obtained the best model of one descriptor. The regression parameters for new model are provided in Equation (9).

Q_{CC} (N_4)/Hz= 9.245 +0.001(**ω**) (11)

N=60 R= 0.900 R^2=0.810 R^2_{adj}= 0.900 D=1.576 RMSE = 1.485 F=544.590 Sig=0.000

3.3. Durbin-Watson statistic

In order to verify and validity of the regression models, we will focus on the Durbin-Watson (DW) statistics. The Durbin–Watson statistic is a test statistic used to detect the presence of autocorrelation at lag 1 in the residuals (prediction errors) from a regression analysis. The Durbin-Watson statistic is always between 0 and 4. Values from 0 to less than 2 indicate positive autocorrelation and values from more than 2 to 4 indicate negative autocorrelation [63]. A value near 2 indicates non-autocorrelation in the sample.

The Durbin-Watson statistics for training and test sets are listed in Table **9**. Almost all values of the DW are less than 2, therefore, there are evidence of a positive serial correlation (non-autocorrelation).

3.4. Validation

The verification and validity of the regression QSAR models were tested using squared cross- validation coefficient for leave –one- out (Q^2LOO) as internal validation [64-67].

The Q^2LOO value (Equation **12**) computed from 50 % of randomly chosen data was found to be positive and smaller than one.

$$Q^2 = 1 - \frac{PRESS}{TSS} \qquad Q^2 \leq 1 \qquad (12)$$

3.4. Validation

The verification and validity of the regression QSAR models were tested using squared cross- validation coefficient for leave –one- out (Q^2LOO) as internal validation [64-67].

The Q^2LOO value (Equation **12**) computed from 50 % of randomly chosen data was found to be positive and smaller than one.

$$Q^2 = 1 - \frac{PRESS}{TSS} \qquad Q^2 \leq 1 \qquad (12)$$

In the Equation (10), PRESS refers to the Predictive residual sum of squares and TSS represents the total sum of square [68]. The Q^2_{LOO} values of the NQCC for $N_1$and N_2 nucleus were calculated 0.869 and 0.871 respectively.

The external validation of QSAR models for judgment of reliability of predictions of models also has been checked using regression parameters: R^2, R^2adj, RMSE, DW, F and Sig of 16 compounds as test set. The comparison between these parameters for the training and test sets are listed in Table **8**.

3.5. Residual

Another method to validate the QSAR model is comparison between the observed and predicted values and the difference between them (residual). These values are listed in Table **2**.

Table 7. Correlation between the molecular descriptors in Equation (9) for the N_4 QCC

Descriptor	AMW	η	ω	Se	Tolerance	VIF1	VIF2
AMW	1.00				0.321	3.118	-
η	-0.649	1.00			0.134	7.479	-
ω	-0.437	0.663	1.00		0.013	78.479	1.00
Se	0.583	-0.796	-0.972	1.00	0.008	122.415	-

Table 8. Statistical parameters for the training and test sets based on Equations 10,11

Data set	Activity	N	R	R^2	R^2_{adj}	RMSE	DW	F	sig
training	N_1 QCC	60	0.984	0.968	0.963	0.273	2.037	855.798	0.00
testing	N_1 QCC	16	0.937	0.878	0.910	0.426	1.673	665.784	0.00
training	N_4 QCC	60	0.937	0.878	0.907	1.405	1.810	222.148	0.00
testing	N_4 QCC	16	0.900	0.810	0.900	1.485	1.576	544.590	0.00

The correlation between predicted values of N_1QCC and N_4QCC versus observed values for the training and test sets is shown in Figs. **2, 3** respectively. These data sets are on the line that it shows the predicted values proposed based on GA-MLR method are in good agreement with the observed values.

Figs **3,4** present the linear correlation between the observed and residual values of the above mentioned activity that were obtained using Equations **10,11**.

3.6. Interpretation of the descriptors

The results and discussion lead us to conclude that combining the two descriptors (piPC06, GATS3e) can be used successfully for predicting the N_1QCC. These descriptors are classified into 2D autocorrelations and Walk and path counts descriptors, respectively.

2D autocorrelations descriptor (GATS3e) weighted by Sanderson electronegativity molecules is calculated from molecular graph by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag) [69-72].

Walk and path counts descriptors obtained from the molecular graph, counting paths, walks and self-returning walks of different lengths [73].

Electrophilicity (ω) is a Quantum chemical descriptor that calculated based on chemical potential (μ) and chemical hardness (**η**)(Equation**7**). This descriptor is useful to predict the N4QCC of CPT derivatives than the other descriptors (see Table **9**).

4. CONCLUSION

GA-MLR technique was used to study a linear relationship between molecular and quantum chemical descriptors and the N_1 and N_4 nucleus QCC of 76 CPTs. These descriptors calculated by the DRAGON software and Gaussian program. The genetic algorithm (GA) and backward multiple linear regression analysis (GA-MLR) were used to select and reduce descriptors.

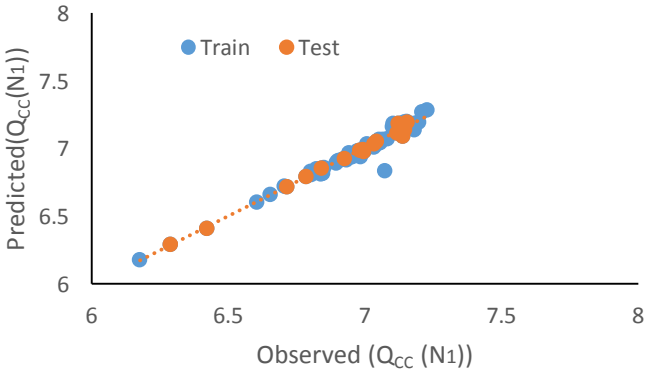


Fig. 2. Comparison between predicted and observed values of the N_1QCC calculated by the GA-MLR method.

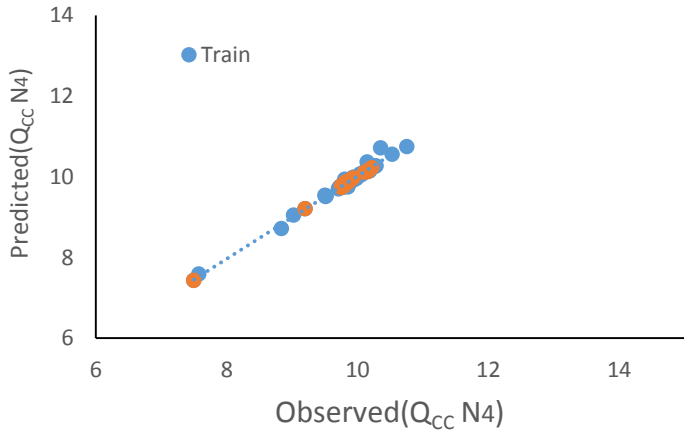


Fig. 3. Comparison between predicted and observed values of the N_4QCC calculated by the GA-MLR method.

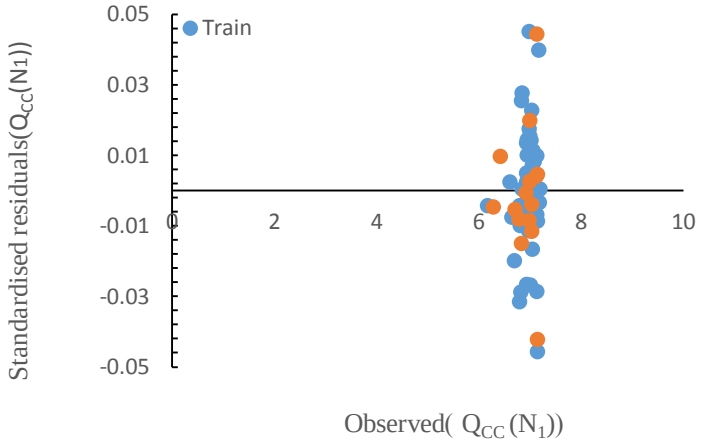


Fig. 4. Plot of residuals against the observed values of the N_1QCC of 76 CPTs for training and test sets using GA-MLR method.

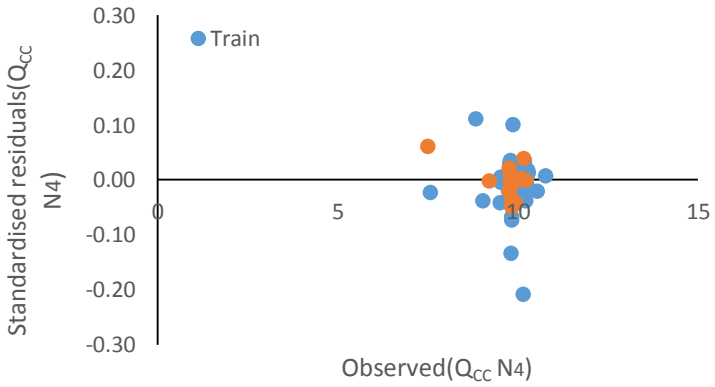


Fig. 5. Plot of residuals against the observed values of the N_4 QCC of 76 CPTs for training and test sets using GA-MLR method.

Table 9. List of the best selected molecular descriptors that appear in the final models.

Activity	Symbol	description	Block description
Q_{CC} (N_1)	piPC06	molecular multiple path count of order 6	Walk and path counts
	GATS3e	Geary autocorrelation of lag 3 weighted by Sanderson electronegativity	2D autocorrelations
Q_{CC} (N_4)	ω	Electrophilicity is a measure of the ability of a molecule to take up electrons, based on chemical potential and chemical hardness	Quantum descriptor

Multicollinearity of the descriptors contributed in the models was examined by calculating the variance inflation factor (VIF) and Pearson correlation coefficient (PCC) statistics. The analysis of the QSAR models based on the CPT derivatives suggest that 2D autocorrelations descriptors(GATS3e), walk and path counts (piPC06) and Quantum chemical descriptor(ω) are equally important as the others descriptors in describing the N_1 and N_4 nucleus of CPTs (see Equation 6,7). The predictive power of QSAR models were checked using squared cross-validation coefficient for leave –one- out (Q^2LOO) as internal validation and external validation. The Q^2LOO values of the N_1 and N_4 nucleus QCC were found to be 0.869 and 0.871 respectively. Results of validations indicate that QSAR models proposed were acceptable. These QSAR models help to delineate the important descriptors responsible for designing some new CPT derivatives and for predicting its activity.

REFERENCES

[1] M. E. Wall, M. C. Wani, C. E. Cook, K.H. Palmer, A.T. McPhail, G. A. Sim, *J. Am. Chem. Soc.*, 88(16) (1966) 3888.

[2] D. K. Kim, N. Lee. *Mini. Rev. Med. Chem*. 2 (6) (2002) 611.

[3] C. J. Thomas, N. J. Rahier, S. M. Hecht, *Bioorg. Med. Chem*., 12 (7) (2004**)** 1585.

[4] M. E. Wall, *Med. Res. Rev*. 18(5) (1998) 299.

[5] K. Legarza, L. X. Yang, *In ViVo*. 19(1) (2005) 283.

[6] Y. Pommier, *Nat. Rev. Cancer*. 6(10) (2006) 789.

[7] P. Hautefaye, B. Cimetiere, A. Pierre, S. Leonce, J. Hickman, W. Laine, C. Bailly, G. Lavielle, Bioorg. *Med. Chem. Lett*. 13(16) (2003) 2731.

[8] Y. H. Hsiang, R. Hertzberg, S. Hecht, L. F. Liu, *J. Biol. Chem.* 260 (1985) 14873.

[9] L. Zhang, D. Ma, Y. Zhang, W. He, J. Yang, C. Li, H. Jiang, *PLOS ONE*. 8(2) (2013) e56458.

[10] C. Hansch, R. P. Verma, *Chem. Med. Chem.* 2(12) (2007) 1807.

[11] Y. H. Hsiang, L. F. Liu, M. E. Wall, M. C. Wani, A. W. Nicholas, G. Manikumar, S. Kirschenbaum, R. Silber, M. Potmesil, *J. Cancer Res.* 49(16)(1989) 4385.

[12] C. Hansch, A. Kurup, *J. Chem. Inf. Comput. Sci.* 43 (2003) 1647.

[13] [13] D. Latifa, D. Villemin, D. Cherqaoui, *Int.J. Mol. Sci*.5 (2004) 48.

[14] K. Vorčáková, M. Májeková, E. Horáková, P. Drabina, M. Sedlák, Š. Štěpánková, Bioorg. Chem.78 (2018) 280.

[15] B. Hu, Z. Kun Kuang, S.Y. Feng, D. Wang, S.B. He, D. Xin Kong, *Molecules. 21* (11) (2016) 1554.

[16] E. Eroglu, H. Türkmen, *J. Mol. Graph. Model*. 26 (2007) 701.

[17] O. Raevsky, V. I. Fetisov, E. P. Trepalina, J. W. McFarland, K. J. Schaper, *Mol. Inform*.19 (2000) 366.

[18] D. Setaki, D. Tataridis, G. Stamatiou, A. Kolocouris, G. B. Foscolos, G. Fytas, N. Kolocouris, E. Padalko, J. Neyts, C.E. De, *Bioorg. Chem*.34 (2006) 248.
[19] Y. Fan, L. M. Shi, K. W. Kohn, Y. Pommier, J. N. Weinstein, *J. Med. Chem*.44 (2001) 3254.
[20] M. Zhu, F. Ge, R. Zhu, X. Wang, X. Zheng, *Chemosphere*. 80 (2010) 46.
[21] S. A. Lipton, *Nat. Rev. Drug. Discov*. 5 (2006) 160.
[22] J. A. Kemp, R. M. McKernan, *Nat. Neurosci*. 5 (2002) 1039.
[23] R. S. Drago, *Physical Methods in Chemistry*, W. B. Saunders Company, London: 1977, p. 245-290.
[24] J. D. Graybeal, *Molecular Spectroscopy*, McGraw - Hill, Singapore: 1988, p. 150-173.
[25] M. A. Rafiee, N. L. Hadipour, H. Naderi - mannesh, H. *J. Comp. Aid. Mole*. 18(3) (2004) 215.
[26] N. L. Hadipour, M. A. Rafiee, M. Javaheri, M. K. Mousavie, *Chem. Phys. Lett*. 356(5) (2002**)** 445.
[27] W. J. Hehre, L. Radom, P. Schleyer, R. von, J. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, 1986.
[28] G. K. Semin, *Use of Nuclear Quadrupole Resonance in Chemistry*, Khimiya, Leningard: 1972, p. 547-602.
[29] D. J. Graybeal, *Molecular Spectroscopy*. McGraw-Hill: Singapore. 1988.
[30] R. Bersohn, *J. Chem. Phys*. 20(10) (1952**)** 1505.
[31] E. A. C. Lucken, *Nuclear Quadrupole Coupling Constant*, Academic Press, London: 1969, p. 204-270.
[32] J. N. Latosinska, *Chem. Phys. Lett*. 398 (2004) 324.
[33] J. N. Latosinska, *J. Pharmaceut. Biomed*. 38 (2005) 577.
[34] R. D. Dennington, T. A. Keith, J. M. Millan, *GaussView, v. 5.0.8,* Semicchem, Inc.: Shawnee KS, 2008.
[35] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, *Gaussian 03, Revision A9*, Gaussian, Inc., Pittsburgh, PA, 2003.
[36] W. Kohn, L. Sham, *J. Phys. Rev*. 140 (1965) 1133.
[37] C.Lee, W. Yang, R.G. Parr, *Phys. Rev. B.* 37 (1998) 785.
[38] A. D. Becke *J. Chem. Phys*. 98 (1993) 5648.
[39] A. D. Mclean, G. S. Chandler, *J. Chem. Phys.* 72 (1980) 5639.
[40] K. Raghavachari, J. S. Binkey, R. Seeger, J. A. Pople, *J. Chem. Phys*. 72 (1980) 650.
[41] M. A. Rafiee, N. L. Hadipour, H. Naderi-manesh, *J. Chem. Inf. Model*. 45 (2005) 366.
[42] M. Talebi Tari, N. Ahmadinejad, *Chem. Sci. Trans*. 6(1) (2017)173.
[43] K. Fukui, *Theory of Orientation and Stereo Selection*. Springer-Verlag, New York, 1975.
[44] B. Kosar, C. Albayrak, Spectrochim. *Acta, Part A.* 78 (2011) 160.
[45] P. Sarmah, *J. Mol. Graph. Model*. 2020 Nov; 100: 107682. doi: 10.1016/j.jmgm.2020.107682. Epub 2020 Jul 12. PMID: 32739639.
[46] A. Luis, R.L. Zárate Hernández, C. Mendoza, S. González-Montiel, J. Cruz-Borbolla, *J. Mol. Graph. Model*. 104 (2021) 107852.
[47] G. Zhang C.B. Musgrave, *J. Phys. Chem. A.* 111 (8) (2007) 1554.
[48] M. A. Safarpour, B. Hemmateenejad, R. Miri, M. Jamali. *QSAR Comb. Sci* . 22 (2004) 997.
[49] I. Dohoo, C. Ducrot, C. Fourichon, A. Donald, D. Hurnik, *Prev. Vet. Med*. 29 (3) (1997) 221.

[50] M. P. Gonzalez, C. Teran, L. Saiz-Urra, M. Teijeira, *Curr. Top. Med. Chem*. 8(18) (2008) 1606.
[51] B. Lučić, N. Trinajstić, *SAR QSAR Environ. Res*.7 (1997) 45.
[52] G. S. Kapur, A. Ecker, R. Meusinger, *Energy Fuels*.15 (2001) 943.
[53] C. Yin, X. Liu, W. Guo, T. Lin, X. Wang, L. Wang, *Water. Res*. 36 (2002) 2975.
[54] P. Gramatica, P. Pilutti, E. Papa, *SAR QSAR Environ. Res*. 13 (2002) 743.
[55] M. V. Diudea, *QSPR/QSAR studies for molecular descriptors*. Ed Nova Science Hunting don, New York. 2000.
[56] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model*. 20 (4) (2002) 269.
[57] S. Chatterjee, J. S. Simonoff, *Handbook of Regression Analysis*, John Wiley & Sons, Hoboken, NJ, 2013.
[58] A. C. Rencher, G. B. Schaalje, *Linear models in statistics*, 2nd ed., John Wiley & Sonc, Inc., Hoboken, 2008.
[59] R. J. Hyndman, A. B. Koehler, *Int. J. Forecast*. 22 (4)(2006)679.
[60] E. Pourbasheer, S. Ahmadpour, R. Zare-Dorabei, M. M. Nekoei, *Arab. J. Chem*. 10 (1) (2017) 33.
[61] B. Reisfeld, A. N. Mayeno, Computational Toxicology: Volume 21, On the Development and Validation of QSAR Models, Springer: Science+Business Media, LLC, 2013, p. 499-529.
[62] T. A. Craney, J. G. Surles, *Qual. Eng*. 14 (2002) 391.
[63] S. Chatterjee, J. S. Simonoff, *Handbook of regression analysis* (Vol. 5). John Wiley & Sons. 2013.
[64] N. R. Hateka, *Tests for Detecting Autocorrelation. Principles of Econometrics: An Introduction (Using R)*. SAGE Publications, 2010. P 379–82.
[65] R. Benigni, C. Bossa, *J. Chem. Inf. Model*. 48 (2008) 971.
[66] M. Stone, *J. Stat. Soc. B.* 36(1974) 111.
[67] R. D. III. Cramer, J. D. Bunce, D. E. Patterson, *Quant. Struct. Act. Relat*. 7(1988**)** 18.
[68] D. G. Kleinbaum, *Applied regression analysis and other multivariable methods*, Australia; Belmont, CA: Brooks/Cole, 2008.
[69] T. Tarpey, *Am. Stat*. 54(2) (2000) 116.
[70] P. Broto, G. Moreau, C. Vandicke, *Eur. J. Med. Chem*. 19 (1984) 79.
[71] P. Broto, J. Devillers, In: *Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology*; Karcher, W. Devillers, J. Eds., Kluwer Academic publishers, Dordrecht, 1990, p 105-127.
[72] I. Ponzoni, V. Sebastián-Pérez, C. Requena-Triguero, C. Roca, M. J. Martínez, F. Cravero, M. F. Díaz, J. A. Páez, R. G. Arrayás, J. Adrio, N. E. Campillo, *Sci. Rep.*7(2017) *2403*.
[73] *G. Rucker, C. Rucker*, J. Chem. Inf. Comput. Sci. 40 (2000) 99.

مجله شیمی فیزیک و شیمی نظری
دانشگاه آزاد اسلامی واحد علوم و تحقیقات
جلد ۱۹، شماره ۲، پاییز ۱۴۰۱
ISSN ۱۷۳۵-۲۱۲۶

مطالعه ارتباط کمی ساختار- فعالیت مشتقات کمپتوسین به عنوان داروهای ضد سرطان با استفاده از تحلیل های رگرسیون خطی چند گانه و الگوریتم ژنتیک

شهاب الدین محبی[۱]، فاطمه شفیعی[۱*]، طاهره مومنی اصفهانی[۱]، مهدی احمدی سابق[۲]

[۱] گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران
[۲] گروه شیمی، واحد اهر، دانشگاه آزاد اسلامی، اهر، ایران

چکیده

ارتباط کمی ساختار–فعالیت (QSAR) به طور گسترده ای برای بررسی ارتباط بین ساختارهای شیمیایی مولکول ها با فعالیت آنها استفاده شده است. در مطالعه حاضر، مدل‌های QSAR برای تعیین ثابت‌های جفت شدگی چهار قطبی (QCC) هسته نیتروژن ^{14}N ۷۶ مشتق کمپتوتسین (CPT) به عنوان داروهای ضد سرطان به کار گرفته شده است. این خواص شیمیایی کوانتومی با استفاده از نظریه تابع چگالی (DFT) و روش (B3LYP/6-311G (d, p در فاز گاز محاسبه شده‌اند. یک مجموعه آموزشی از ۶۰ مشتق CPT برای ساخت مدل های QSAR و یک مجموعه آزمون از ۱۶ ترکیب برای ارزیابی مدل های ساخته شده با استفاده از تحلیل رگرسیون خطی چندگانه (MLR) استفاده شد. توصیفگرهای مولکولی توسط نرم افزار دراگون محاسبه شده است، و از روش های رگرسیون خطی چندگانه گام به گام و الگوریتم ژنتیک (GA) به ترتیب برای انتخاب بهترین توصیف گرها و ساخت مدل های QSAR استفاده شده است. مدل‌های QSAR برای شناسایی توصیفگرهای مهم مرتبط با ویژگی‌های مشتقات CPT به کار گرفته شده است. مدل‌های QSAR آماری معنا دار به دست آمده از تجزیه و تحلیل GA-MLR، توسط روش‌های اعتبار سنجی متقاطع یکی بیرون ((Leave-One-Out (LOOCV) و روش‌های اعتبار سنجی خارجی تأیید شدند. همبستگی توصیفگرهای ارائه شده در مدل‌ها با محاسبه ضریب نفوذ پذیری (VIF) و آماره دوربین واتسون (DW) مورد آزمایش قرار گرفت. توانایی پیش‌بینی مدل‌ها رضایت بخش بود. نتایج مطالعه QSAR نشان می‌دهد که پارامترهای کوانتومی، خود همبستگی های دو بعدی و توصیف‌گرهای شمارش راه و مسیر حاوی اطلاعات ساختاری مهمی برای توسعه مدل های مفید برای پیش‌بینی فعالیت های مورد مطالعه می‌باشند.

کلید واژه‌ها: مشتقات کمپتوتسین؛ ارتباط کمی ساختار – فعالیت؛ پارامترهای کوانتومی؛ الگوریتم ژنتیک- رگرسیون خطی چندگانه؛ توصیفگرهای مولکولی؛ اعتبار سنجی یکی بیرون؛ ثابت‌های جفت شدگی چهار قطبی.

* مسئول مکاتبات: f-shafiei@iau-arak.ac.ir