Heliyon 8 (2022) e10067

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Anticancer activity and QSAR study of sulfur-containing thiourea and sulfonamide derivatives

Ratchanok Pingaew [a,*], Veda Prachayasittikul [b,**], Apilak Worachartcheewan [c], Anusit Thongnum [d], Supaluk Prachayasittikul [b], Somsak Ruchirawat [e,f,g], Virapong Prachayasittikul [h]

[a] Department of Chemistry, Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand
[b] Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[c] Department of Community Medical Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[d] Department of Physics, Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand
[e] Laboratory of Medicinal Chemistry, Chulabhorn Research Institute, Bangkok 10210, Thailand
[f] Program in Chemical Sciences, Chulabhorn Graduate Institute, Bangkok 10210, Thailand
[g] Center of Excellence on Environmental Health and Toxicology (EHT), Commission on Higher Education, Ministry of Education, Bangkok 10400, Thailand
[h] Department of Clinical Microbiology and Applied Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

ARTICLE INFO

Keywords:
Thiourea
Sulfonamide
Anticancer activity
QSAR

ABSTRACT

Sulfur-containing compounds are considered as attractive pharmacophores for discovery of new drugs regarding their versatile properties to interact with various biological targets. Quantitative structure-activity relationship (QSAR) modeling is one of well-recognized *in silico* tools for successful drug discovery. In this work, a set of 38 sulfur-containing derivatives (Types I–VI) were evaluated for their *in vitro* anticancer activities against 6 cancer cell lines. *In vitro* findings indicated that compound **13** was the most potent cytotoxic agent toward HuCCA-1 cell line (IC_{50} = 14.47 μM). Compound **14** exhibited the most potent activities against 3 investigated cell lines (i.e., HepG2, A549, and MDA-MB-231: IC_{50} range = 1.50–16.67 μM). Compound **10** showed the best activity for MOLT-3 (IC_{50} = 1.20 μM) whereas compound **22** was noted for T47D (IC_{50} = 7.10 μM). Subsequently, six QSAR models were built using multiple linear regression (MLR) algorithm. All constructed QSAR models provided reliable predictive performance (training sets: R_{tr} range = 0.8301–0.9636 and $RMSE_{tr}$ = 0.0666–0.2680; leave-one-out cross validation sets: R_{CV} range = 0.7628–0.9290 and $RMSE_{CV}$ = 0.0926–0.3188). From QSAR modeling, chemical properties such as mass, polarizability, electronegativity, van der Waals volume, octanol-water partition coefficient, as well as frequency/presence of C–N, F–F, and N–N bonds in the molecule are essential key predictors for anticancer activities of the compounds. In summary, a series of promising fluoro-thiourea derivatives (**10**, **13**, **14**, **22**) were suggested as potential molecules for future development as anti-cancer agents. Key structure-activity knowledge obtained from the QSAR modeling was suggested to be advantageous for suggesting the effective rational design of the related sulfur-containing anticancer compounds with improved bioactivities and properties.

1. Introduction

Cancer is one of common leading causes of death worldwide. It is also well-recognized as chronic disease which affecting human well-being. New cancer cases and mortality rate have been continuously reported and are predicted to be topped up for approximately 50% in the next two decades [1]. Many clinically available anticancer drugs are problematically concerned for their adverse side effects and drug resistance [2]. Development of novel anticancer agents with desirable effectiveness and minimized toxicities is, therefore, considered to be one of research area to serve the needs for global health [3, 4, 5].

In the field of medicinal chemistry, sulfur containing molecules can be found in many pharmaceuticals and natural products [6, 7]. Among these, thiourea (–NHCSNH–) and sulfonamide (–SO_2NH–) have been

* Corresponding author.
** Corresponding author.
E-mail addresses: ratchanok@g.swu.ac.th (R. Pingaew), veda.pra@mahidol.ac.th (V. Prachayasittikul).

https://doi.org/10.1016/j.heliyon.2022.e10067
Received 9 July 2021; Received in revised form 1 April 2022; Accepted 19 July 2022
2405-8440/© 2022 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

noted as important motifs with anticancer activity and others (i.e., antimicrobial, antimalarial, and antiviral activities). These pharmacologically attractive characters were noted to be due to their structural characteristics, which are capable of participating in various biomolecular targets and eliciting multiple interactions [7, 8, 9, 10, 11, 12, 13, 14]. Furthermore, thiourea and sulfonamide derivatives are generally stable and are well-recognized for their synthetic feasibilities. Sulfonamide derivatives are also well-known for their preferable bioavailability (i.e., good oral absorption) and safety (low side effects) [13].

In an area of anticancer drug discovery, thiourea and sulfonamide derivatives were reported to exhibit their anticancer actions *via* inhibiting diverse molecular targets such as tubulin [14, 15], carbonic anhydrase [16, 17], topoisomerase II [18], aromatase [19, 20], cyclin-dependent kinase (CDK) [21], epidermal growth factor receptor (EGFR) [22, 23], sirtuin [24], nucleotide pyrophosphatase/phosphodiesterase [25], v-Raf murine sarcoma viral oncogene homolog B1 (BRAF) [26], and others. These suggested the potentials of these scaffolds to be promising pharmacophores for discovery of novel anticancer therapeutics.

One of current problems in drug development is the failures that occur in the late stage of development. Accordingly, many computational tools have been included as supportive tools for increasing success rate and time-saving issue. Success story using *in silico* tools in anticancer drug discovery have been reported [27, 28]. One of which, quantitative structure-activity relationship (QSAR) modeling is well-known as a *silico* tool used to gain informative knowledge regarding the key physiochemical features which are required for good potency of the bioactive molecules. Accordingly, QSAR modeling is widely used for guiding design and structural modification of compounds to give new derivatives with improved bioactivity as well as desirable pharmacokinetic properties [28, 29, 30, 31].

In this work, a series of in-house synthesized thiourea and sulfonamide derivatives (Figure 1) were evaluated for their *in vitro* anticancer activities against 6 cancer cell lines to suggest a set of for potential development. Additionally, the experimentally obtained bioactivity values along with the chemical structures of compounds were used as input data sets for QSAR model construction using multiple linear regression (MLR) algorithm to give six predictive models revealing key chemical features, as key predictors, for potent anticancer effects of the compounds which would be of benefit for future successful discovery of novel sulfur-based compounds as anticancer agents.

2. Results and discussion

2.1. Compounds

Anticancer investigations were performed towards a total library of 38 in-house synthesized sulfur-containing compounds. The studied compounds were categorized into 6 types according to their core scaffolds (Figure 1). These thioureas/sulfonamides were substituted with R groups (i.e., phenyl bearing F, Cl, Br, CN, NO_2, CF_3, CH_3, OCH_3, NH_2, including naphthalenyl). Three types of thiourea derivatives (Types **I**–**III**, compounds **1**–**23**) were synthesized by treatment of benzylamine or *meta (m)*- or *para (p)*-xylylenediamine with the corresponding isothiocyanates [19, 32, 33] while another three sets of bis-sulfonamide derivatives (Types **IV**–**VI**, compounds **24**–**38**) were prepared by reaction of *m*- or *p*-xylylenediamine or *m*-phenylenediamine with the corresponding benzenesulfonyl chlorides [20]. The detailed synthesis of these compounds **1**–**38** was provided in the literature [19, 20, 32, 33].

2.2. Cytotoxic activity

Thiourea and sulfonamide derivatives (**1–38**) were investigated for their cytotoxic effects against 6 human cancer cell lines; cholangiocarcinoma (HuCCA-1), hepatocellular carcinoma (HepG2), lung carcinoma (A549) lymphoblastic leukemia (MOLT-3), hormone-independent breast cancer (MDA-MB-231), and hormone-dependent breast cancer (T47D) as summarized in Table 1. Etoposide, and/or doxorubicin were used as reference drugs. Cytotoxicity against normal cell lines (i.e., MRC-5 and Vero cell lines) of the investigated compounds have been reported [20, 32] and their selectivity index (SI) values are summarized in Table S1 in the supplementary file.

For mono-thiourea derivatives (Type I, **1–7**), most of the tested thioureas (**3**–**7**) exerted a wide range of anticancer activities against all tested cell lines (IC_{50} = 5.07–119.51 μM), except for compounds **1** (R = 4-OCH_3Ph) and **2** (R = 4-FPh) which showed low activities against only 2 cell lines (i.e., HepG2 and MOLT-3). Thiourea **6** containing 3,5-diCF_3Ph was the most potent compound of the set. It displayed highly potent cytotoxic effect against MOLT-3 cell line with IC_{50} value of 5.07 μM. Additionally, it exhibited better activity toward HepG2 cell line (IC_{50} = 16.28 μM) when compared with the reference etoposide (IC_{50} = 26.05 μM).

For *meta*-bis-thiourea derivatives (Type II, **8–15**), it was found that all the tested compounds displayed cytotoxicity toward MOLT-3 cell line (IC_{50} = 1.20–32.32 μM). Notably, the substitution with strong electron withdrawing groups leading to thioureas **10** (with 4-F), **13** (with 4-CF_3), **14** (with 3,5-diCF_3), and **15** (with 4-NO_2) which elicited high cytotoxic effects against a broad range of tested cancer cells. Apparently, analog **10** bearing 4-fluoro phenyl group showed the most potent cytotoxic effect against MOLT-3 cells (IC_{50} = 1.20 μM). Interestingly, derivatives **10**, **11**, and **13–15** exhibited more potent activities against HepG2 cells than the etoposide. Among these, thiourea **14** with 3,5-diCF_3 substituents was the most potent cytotoxic agent (IC_{50} = 1.50 μM against HepG2) performing 17.4-fold higher activity than the etoposide, but with 2.6-fold weaker than the doxorubicin.

For *p*-bis-thiourea derivatives (Type III, **16–23**), the *p*-derivatives seem to be less active against all the tested cells than the corresponding *m*-counterparts. This was observed for compounds **10** > **18** (with 4-F), **11** > **19** (with 4-Cl), **13** > **21** (with 4-CF_3), and **15** > **23** (with 4-NO_2). Remarkably, thiourea **22** with 3,5-diCF_3 moieties displayed a diverse range of anticancer activities toward all the tested cells (IC_{50} value range of 2.49–30.95 μM). Notably, the compounds **16**, **21** and **22** showed more potent activity against HepG2 cell (IC_{50} = 8.81–21.67) when compared with the etoposide.

Considering the potent activities of thioureas **13** (IC_{50} = 17.33 μM), **14** (IC_{50} = 21.12 μM), **15** (IC_{50} = 27.59 μM), and **22** (IC_{50} = 7.10 μM) against T47D hormone-dependent breast cancer cell line, their potent cytotoxic effects observed herein might be due to their abilities to act as aromatase inhibitors and inhibit estrogen production, as previously reported by our group [19].

For the bis-sulfonamide derivatives (Type IV–VI, **24**–**38**), 2,3,5,6-tetramethylphenyl derivative **25** as well as 4-chlorophenyl compounds **28** and **38** displayed a broad array of cytotoxic activities toward all the tested cells. Along the line, the 4-chloro derivative **28** was shown to be the most potent cytotoxic compound against HuCCA-1 (IC_{50} = 34.51 μM), A549 (IC_{50} = 39.14 μM), and MOLT-3 (IC_{50} = 16.52 μM) cells. The 2,3,5,6-tetramethylphenyl compound **25**, 2-naphthalenyl compound **26**, and 4-trifluoromethylphenyl derivative **29** were noted as the most potent compounds against T47D (IC_{50} = 30.92 μM), MDA-MB-231 (IC_{50} = 31.14 μM), and HepG2 (IC_{50} = 16.38 μM) cell lines, respectively. In addition, the *meta*-sulfonamide **29** exhibited greater anticancer activity against HepG2 cell line (IC_{50} = 16.38 μM) than the etoposide.

All of the studied bis-sulfonamide derivatives (**24**–**38**) have been disclosed to act as aromatase inhibitors except for compound **35** [20]. Additionally, some of these *meta*-bis-sulfonamides (i.e., **24, 27, 29**, and **31**) have been reported to induce apoptosis in various cancer cell lines. The mechanism underlying the apoptosis of these compounds was anticipated to be induced by their abilities to act as artificial Cl^- ion transporters, which lead to the disruption of ionic homeostasis and excessive production of harmful reactive oxygen species (ROS) [34].

In overview, Type II *meta*-bis-thiourea CF_3 compounds (**13** and **14**) were noted as promising cytotoxic compounds against most of the

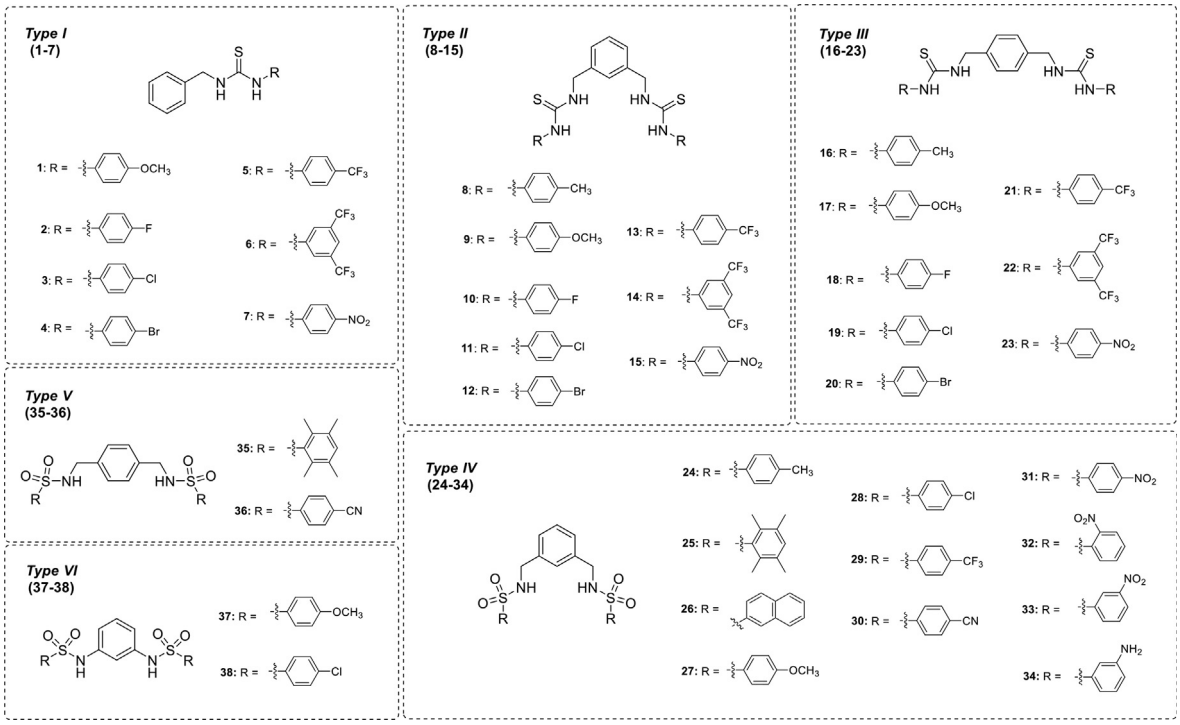

Figure 1. Chemical structures of compounds **1**–**38**.

investigated cancer cell lines. Type III-*p*-bis-thiourea di-CF_3 compound **22** was noted as the most potent anticancer agent against T47D whereas type II *meta*-bis-thiourea fluoro derivative **10** showed the most potent activity against MOLT-3 (Table 1). It should be noted that all these promising compounds are fluoro-containing derivatives. Fluorine has been noted as an attractive atom in medicinal chemistry and drug discovery. Introduction of single fluoro (F) or trifluoromethyl groups (CF_3) have been noted to increase metabolic/chemical stability, improve pharmacokinetic profile (i.e., membrane permeability and bioavailability), as well as increase binding affinity of compounds to target protein [35]. Many anticancer drugs are fluoro derivatives [35]. Additionally, approximately 40% of the marketed drugs recently approved by FDA in the year 2019 are fluorine-containing drugs and insertion of fluorine in bioactive molecules has been noted as one of strategies to lower risk of failure [36].

It should be noted that most of the studied compounds exhibited more potent activity against non-solid cancer MOLT-3 cell line with low IC_{50} values (<10 μM) whereas those of activities against other cell lines showed higher IC_{50} values (>10 μM), except for compounds **14** against HepG2 cell line, and compound **22** against T47D cell line. These highly potent compounds against MOLT-3 cell line included compounds from Type II (i.e., mono-halogenated compounds (**10**–**12**) and NO_2-derivatives (**15**)), Type I (i.e., di-CF_3 derivative **6**), and Type III (i.e., CF_3 derivatives **21** and **22**). This finding suggested the potential of these compounds to be further developed for management of non-solid cancer, a cancer type which mainly relies on chemotherapy. Additionally, compound **10** showed high selectivity index (SI = 57.49, Table S1 in the supplementary file) which suggested its high selectivity against cancer cells with lower side effects.

2.3. QSAR study

QSAR modeling is a computational method to find the relationship between chemical structure of the compounds and their activities. Conceptually, properties of every compound can be described as a set of numerical values called descriptors. These descriptors can be obtained by structural calculation and can be used as independent variables (X variables or predictors) to predict biological activities (Y variable) of the compounds. Prior to model construction, data preprocessing was performed including i) the preparation of bioactivity values (Y variables) by taking the −log10 to convert the IC_{50} values into pIC_{50} values and ii) the selection of informative descriptors (X) from the whole set obtained by calculation to give a final set of predictors. Bioactivity values of all compounds expressed in pIC_{50}, and selected descriptor values are provided in Tables S2 and S3 in the supplementary file, respectively.

Multiple linear regression (MLR) is one of the most commonly machine learning algorithm used in the field of drug discovery [37]. A set of actual bioactivity values (experimental pIC_{50} values, Table S2 in the supplementary file) along with selected descriptor values (Table S3 in the supplementary file) were prepared as an input dataset for MLR QSAR modeling, subsequently, the model is returned in a form of linear equation representing Y variable (bioactivity) as a function of multiple X variables (descriptors). This interpretable characteristic makes the MLR an attractive algorithm suitable for QSAR-driven rational design. Several MLR QSAR models for anticancer activity prediction and guiding rational design of many classes of compounds have been reported by our group [38, 39, 40, 41].

Herein, experimental cytotoxic activities (IC_{50} values) along with chemical structures of the investigated compounds (**1**–**38**) were used as input datasets for QSAR modeling. According to cytotoxic activities against 6 cancer cell lines, 6 datasets were separately prepared. For each dataset, inactive compounds were excluded and the remaining active ones were included in the dataset. Chemical structures of compounds (**1**–**38**) were generated, geometrically optimized, and calculated using computer software to obtain a large set of 1,574 descriptor values. Subsequently, correlation-based feature selection followed by multiple linear regression (MLR) algorithm were performed to obtain a set of selected informative descriptors as predictors for bioactivity prediction (i.e., HuCCA-1 model = 2 descriptors, HepG2 = 4 descriptors, A549 = 3 descriptors, MOLT-3 = 3 descriptors, MDA-MB-231 = 3 descriptors, and T47D = 3 descriptors), in which their definitions and values are shown in Table 2 and Table S3 in the supplementary file, respectively).

For each QSAR model, the dataset was randomly divided into two main sets using the leave-one-out cross validation (LOOCV) sampling

Table 1. Cytotoxic activity (IC_{50}, μM) of S-containing derivatives (**1–38**) against 6 cancer cell lines.

Compound	Cancer cell lines[a]					
	HuCCA-1	HepG2	A549	MOLT-3	MDA-MB-231	T47D
1	NC	139.66 ± 4.97	NC	128.72 ± 1.74	NC	NC
2	NC	141.40 ± 2.51	NC	104.91 ± 1.01	NC	NC
3	80.79 ± 1.74	84.51 ± 2.34	77.50 ± 0.97	31.43 ± 0.62	105.43 ± 4.81	80.97 ± 1.68
4	98.59 ± 1.57	71.72 ± 0.95	101.61 ± 1.54	24.41 ± 0.97	99.05 ± 6.47	69.42 ± 0.30
5	72.34 ± 2.87	102.11 ± 2.84	68.63 ± 0.64	14.66 ± 0.20	115.45 ± 2.61	62.45 ± 4.85
6	67.51 ± 0.83	16.28 ± 0.67	61.69 ± 2.35	5.07 ± 0.21	20.35 ± 0.25	52.86 ± 2.24
7	99.29 ± 2.21	22.80 ± 0.19	59.16 ± 1.67	11.07 ± 0.62	119.51 ± 4.42	70.44 ± 1.58
8[b]	NC	NC	NC	13.90 ± 1.09	91.83 ± 2.19	NC
9[b]	NC	NC	NC	32.32 ± 6.92	NC	NC
10	39.11 ± 2.88	14.94 ± 0.29	52.99 ± 2.40	**1.20 ± 0.02**	43.34 ± 0.04	95.38 ± 4.80
11[b]	30.22 ± 1.25	13.29 ± 1.05	NC	2.23 ± 0.26	14.74 ± 0.33	31.54 ± 2.16
12[b]	NC	41.30 ± 0.03	NC	2.62 ± 0.28	13.70 ± 0.03	NC
13[b]	**14.47 ± 1.38**	8.40 ± 0.43	17.97 ± 2.96	1.55 ± 0.17	9.68 ± 0.84	17.33 ± 1.77
14[b]	18.82 ± 2.34	**1.50 ± 0.23**	**16.67 ± 1.11**	3.63 ± 0.46	**8.05 ± 0.68**	21.12 ± 0.08
15[b]	14.84 ± 0.50	10.53 ± 0.58	44.71 ± 3.49	3.40 ± 0.44	12.55 ± 0.82	27.59 ± 0.42
16[b]	82.83 ± 8.48	21.67 ± 2.24	NC	NC	69.46 ± 2.05	NC
17[b]	NC	NC	NC	NC	90.12 ± 2.52	74.92 ± 0.96
18	99.54 ± 5.96	86.88 ± 4.58	NC	NC	NC	106.45 ± 3.67
19[b]	NC	NC	NC	NC	NC	73.53 ± 0.96
20[b]	NC	NC	NC	NC	60.81 ± 4.62	83.48 ± 3.67
21[b]	NC	15.52 ± 1.12	NC	6.06 ± 0.79	NC	71.60 ± 5.30
22[b]	30.95 ± 4.58	8.81 ± 0.71	26.16 ± 0.35	2.49 ± 0.19	12.70 ± 0.40	**7.10 ± 0.18**
23[b]	NC	61.02 ± 2.56	NC	11.40 ± 0.45	59.01 ± 4.10	80.65 ± 7.55
24[c]	43.23 ± 3.20	58.28 ± 1.62	NC	30.07 ± 0.67	61.79 ± 0.77	43.93 ± 0.85
25[c]	77.73 ± 2.06	46.94 ± 3.37	78.30 ± 3.45	31.85 ± 14.34	42.21 ± 2.32	30.92 ± 0.71
26[c]	NC	74.52 ± 5.55	NC	NC	31.14 ± 0.57	NC
27[c]	NC	72.08 ± 1.30	NC	43.06 ± 0.98	68.30 ± 0.35	75.27 ± 2.04
28[c]	34.51 ± 1.13	49.69 ± 1.56	39.14 ± 1.17	16.52 ± 0.62	54.33 ± 0.64	37.47 ± 2.15
29[c]	81.16 ± 4.98	16.38 ± 1.25	NC	NC	57.48 ± 5.48	75.38 ± 5.29
30[c]	NC	NC	NC	38.41 ± 2.08	NC	69.28 ± 3.55
31[c]	38.52 ± 0.55	51.77 ± 0.31	NC	16.92 ± 0.38	54.39 ± 1.77	39.29 ± 2.02
32[c]	NC	NC	NC	34.04 ± 4.34	NC	75.36 ± 2.15
33[c]	NC	NC	NC	24.97 ± 0.50	63.67 ± 0.33	NC
34[c]	NC	NC	NC	50.61 ± 1.01	NC	90.61 ± 0.95
35[c]	NC	74.37 ± 5.31	NC	NC	NC	NC
36[c]	97.96 ± 0.67	NC	NC	NC	95.32 ± 2.56	NC
37[c]	NC	NC	NC	48.00 ± 1.31	NC	NC
38[c]	97.08 ± 2.38	62.03 ± 1.70	97.96 ± 3.20	35.33 ± 0.55	58.97 ± 0.79	87.22 ± 4.39
Doxorubicin[d]	0.42 ± 0.02	0.57 ± 0.05	0.37 ± 0.02	-	1.97 ± 0.30	0.88 ± 0.02
Etoposide[d]	-	26.05 ± 0.50	-	0.041 ± 0.003	-	-

NC: IC_{50} > 50 μg/mL denoted as non-cytotoxic. IC_{50} is a concentration of compound required to produce 50% of inhibitory effect.
The most potent compounds against each cell line displaying the lowest IC_{50} values were highlighted in bold.

[a] Cancer cell lines comprise the following: HuCCA-1 cholangiocarcinoma cancer cell line, HepG2 hepatocellular carcinoma cell line, A549 lung carcinoma cell line, MOLT-3 lymphoblastic leukemia cell line, MDA-MB-231 hormone-independent breast cancer, and T47D hormone-dependent breast cancer.

[b] Cytotoxic activities against HuCCA-1, HepG2, A549 and MOLT-3 have been reported in [32, 33].

[c] Cytotoxic activity against T47D has been reported in [20].

[d] Doxorubicin and etoposide were used as reference drugs.

method. LOOCV is a way to validate the model by excluding one sample from the whole dataset (*N*) to be used as a LOOCV testing set while the remaining samples (*N*-1) were used as a training dataset. The training dataset, containing values of both descriptor (*X*) and pIC_{50} (*Y*) variables, was used to train the MLR for finding the linear relationships between these *X* and *Y* variable, and the result was returned as a predictive equation. Subsequently, the built model (equation) was used for predicting the pIC_{50} (*Y*) value of the leaved out sample (as a LOOCV testing set) by its available descriptor (*X*) values. The accuracy of the prediction was assessed by considering the difference between predicted pIC_{50} and actual pIC_{50} values obtained from the experiment in which the lower gap between these two values indicated low error of the prediction (as shown by nearness of the dots in the plots of Figure 2, low error values in Tables S3 and S4 in the supplementary file). The same sampling process was continued until all samples in the dataset were leaved out to be used as a testing set. After the sampling is completed, the average values of several rounds were calculated as correlation coefficient (R) and root mean squared error (RMSE).

The constructed QSAR models are shown in Eqs. (1), (2), (3), (4), (5), (6). All models provided preferable predictive performance as shown by their statistical parameters (i.e., high correlation coefficient (R_{tr}: 0.8301–0.9636 and R_{CV} = 0.7628–0.9290) but low root mean squared

Table 2. Definitions of informative descriptors for QSAR modeling.

Descriptor	Type	Definition
QZZm[a]	Geometrical descriptors	Quadrupole z-component value/weighted by mass
B08 [N–N][a]	2D Atom Pairs	Presence/absence of N–N at topological distance 8
Mor03m[b]	3D-MoRSE descriptors	Signal 03/weighted by mass
Mor07u[b]	3D-MoRSE descriptors	Signal 07/unweighted
Mor29v[b]	3D-MoRSE descriptors	Signal 29/weighted by van der Waals volume
Mor07e[b]	3D-MoRSE descriptors	Signal 07/weighted by Sanderson electronegativity
IC2[c]	Information indices	Information Content index (neighborhood symmetry of 2-order)
Mor11p[c]	3D-MoRSE descriptors	Signal 11/weighted by polarizability
cRo5[c]	Drug-like indices	Complementary Lipinski Alert index
F01[C–N][d]	2D Atom Pairs	Frequency of C–N at topological distance 1
E2m[d]	WHIM descriptors	2nd component accessibility directional WHIM index/weighted by mass
MLOGP[d]	Molecular properties	Moriguchi octanol-water partition coefficient (logP)
ATS7m[e]	2D autocorrelations	Broto-Moreau autocorrelation of lag 7 (log function) weighted by mass
RDF090m[e]	RDF descriptors	Radial Distribution Function - 090/weighted by mass
MLOGP2[e]	Molecular properties	Squared Moriguchi octanol-water partition coefficient (logP2)
F06[F–F][f]	2D Atom Pairs	Frequency of F–F at topological distance 6
RDF095m[f]	RDF descriptors	Radial Distribution Function - 095/weighted by mass
RDF150m[f]	RDF descriptors	Radial Distribution Function - 150/weighted by mass

[a] Predictors of HuCCA-1 model (Eq. 1).
[b] Predictors of HepG2 model (Eq. 2).
[c] Predictors of A549 model (Eq. 3).
[d] Predictors of MOLT-3 model (Eq. 4).
[e] Predictors of MDA-MB-231 model (Eq. 5).
[f] Predictors of T47D model (Eq. 6).

error ($RMSE_{tr}$ 0.0666–0.2680 and $RMSE_{CV}$ = 0.0926–0.3188), Table 3). Plots between experimental and predicted pIC_{50} values of the 6 constructed QSAR models are shown in Figure 2, and values are given in Tables S4 and S5.

2.3.1. HuCCA-1 model

$$pIC_{50} = 0.0002\ (QZZm) + 0.3623\ (B08[N–N]) - 2.0124 \quad (1)$$

Descriptors presented in the QSAR model indicated that the presence/absence of N–N and mass are influencing properties for anticancer effect against HuCCA-1 cell line. Positive regression coefficient values of both descriptors indicated that high values of both descriptors are required for potent activity of the compounds. B08[N–N] is a predictor with higher influence on the prediction, as shown by its higher regression coefficient when compared to that of QZZm.

Among 20 active compounds in all types, compounds in *meta*-bis-thiourea derivatives Type II (**11**, **13**, **14**, and **15**) were shown to be the most potent class because they are the only class having B08[N–N] values = 1, while those of others having B08[N–N] values = 0 (Table S3 in the supplementary file). This is suggested that position of the linked thioureas on the central aromatic ring at *meta*-position seems to be crucial for potent activity against HuCCA-1 cell line. This was also observed for the decreased activity of *para*-bis-thiourea di-CF_3 and F derivatives when compared with their *meta*-analogs (**14** > **22** and **10** > **18**). However, it was observed that an addition of another thiourea group on the molecules of Type I thiourea derivatives (**3**–**6**) improved activities of the obtained bis-thiourea derivatives, both for *meta*-series Type II (**11** > **3**, **13** > **5**, **14** > **6**), and *para*-series Type III (**22** > **6**).

Replacement of thiourea groups on both sides of the central ring with sulfonamides lead to decreased activities of the compounds *via* decreasing the values of both descriptors (QZZm and B08[N–N]), as shown when comparing activities of CF_3, Cl and NO_2 bis-sulfonamide derivatives with their bis-thioureas (**11** > **28**, **13** > **29**, and **15** > **31**). This indicated that thiourea functional group is required for potent activity.

Additionally, shorten the length of alkyl chain linker from 2C to 1C affected values of mass QZZm descriptor and impaired activity of the compound, **28** > **38** (**28**: QZZm = 1133.55, pIC_{50} = −1.538, **38**: QZZm = 574.71, pIC_{50} = −1.987, Tables S2 and S3).

2.3.2. HepG2 model

$$pIC_{50} = \text{-}0.0597\ (Mor03m) + 0.0388\ (Mor07u) + 1.3986\ (Mor29v) + 0.0267\ (Mor07e) - 2.6129 \quad (2)$$

The QSAR model indicated that van der Waal volume (Mor29v), mass (Mor03m), and electronegativity (Mor07e) play role in cytotoxic activity against HepG2 cell line.

Similar with the HuCCA-1 model, active compounds of Type II exhibited the most potent activity than that of the others. It was noticed that the most potent compound **14** exhibited a promising activity approximately 5.25 folds higher than that of the second most potent compound **13**. This demonstrated that the presence of di-CF_3 moieties on two phenyl groups of the central aromatic rings can notably increase activity of the compound. Especially, the Mor03m value of the *meta*-bisthiourea **14** was increased for 1.66 folds when compared with that of compound **13** with mono-CF_3 group (Table S3 in the supplementary file). The significance of aromatic ring bearing di-CF_3 groups was also noted for the compound **22** of *para*-series. The compound **22**, which displayed the highest Mor29v (0.485) but the lowest Mor03m (−12.9) values among other compounds in Type III series, exhibited the highest activity among all compounds in Type III and was ranked as the three most potent compounds among all the investigated compounds (Table 1).

Bis-thioureas (Type II and Type III) exhibited improved activities than their parent thioureas (Type I). This effect was observed as results of increasing values of three descriptors (i.e., Mor07u, Mor29v, and Mor07e) for mono- and di-CF_3 bis-thiourea derivatives of both series (Type II *meta*-series: **13** > **5** and **14** > **6**, Type III para-series **21** > **5** and **22** > **6**). Values of three descriptors for the compared compounds are i) Mor07u: **5** = 5.449, **6** = 5.461, **13** = 8.26, **14** = 8.373, **21** = 8.283, **22** = 7.916, ii) Mor29v: **5** = 0.228, **6** = 0.22, **13** = 0.512, **14** = 0.474, **21** = 0.329, **22** = 0.485, and iii) Mor07e: **5** = 5.775, **6** = 5.921, **13** = 8.936, **14** = 9.455, **21** = 9.005, **22** = 9.042, Table S3 in the supplementary file. The high values of van der Waals volume (Mor29v) could be due to the molecular size of bis-thiourea compared with its parent thiourea. In addition, the di-CF_3 compound had higher electronegativity (Mor07e) than its mono-CF_3 as noted for i.e., compounds **6** > **5**, **14** > **13**, and **22** > **21**. This is resulted from the electronegativity of F atom.

2.3.3. A549 model

$$pIC_{50} = 0.5377\ (IC2) + 0.2234\ (Mor11p) + 0.3085\ (cRo5) - 4.0243 \quad (3)$$

The QSAR results showed that symmetry of the molecule (IC2), polarizability (Mor11p), and drug-like related properties (cRo5) are influencing characters for potent activity against A549 cell line. Although the highest regression coefficient value was not obviously noted, the drug-like descriptor cRo5 seems to be a highly influencing property. This was observed for a set of three top ranked highly potent compounds **14**, **13,** and **22**, which are the only three compounds whose cR05 value = 1 while most of the rest are 0 (Table S3 in the supplementary file). Moreover, these three compounds are bis-thiourea derivatives bearing mono-

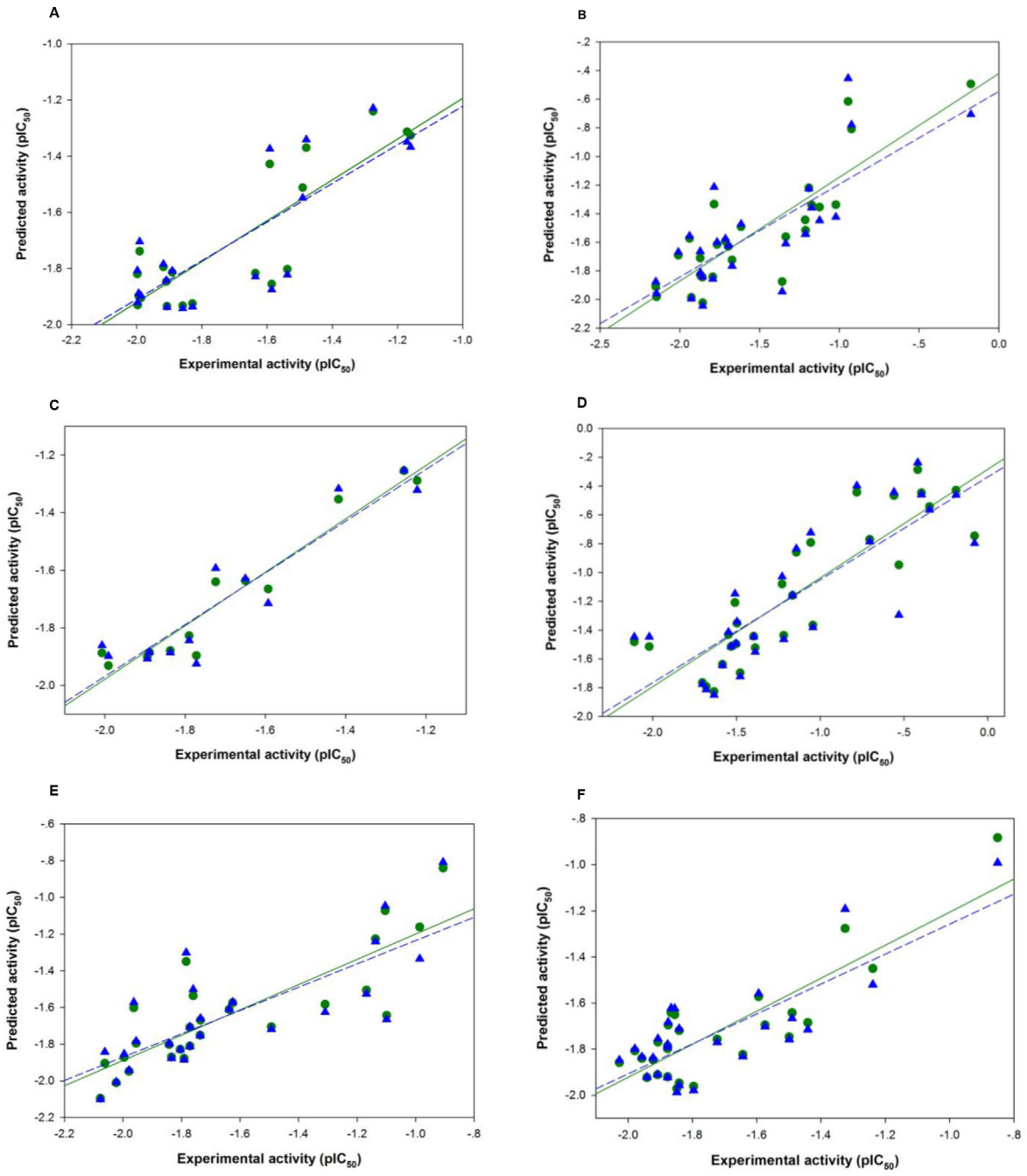

Figure 2. Plots of experimental VS predicted activities from 6 QSAR models (Eqs. (1), (2), (3), (4), (5), and (6)). Plots of training set are presented as circle symbols and solid regression lines whereas those of testing set (leave-one-out cross validation) are presented as triangles and dashed regression lines. **A**: HuCCA-1 model, Eq. (1) (N = 20), **B**: HepG2 model, Eq. (2) (N = 27), **C**: A549 model, Eq. (3) (N = 13), **D**: MOLT-3 model, Eq. (4) (N = 29), **E**: MDA-MB-231 model, Eq. (5) (N = 27), **F**: T47D model, Eq. (6) (N = 27).

(**13**) and di-CF_3 (**14** and **22**) phenyl groups on the central aromatic ring. This indicated that the CF_3 moiety is essential for potent activity. However, it was found that *meta*-Type II di-CF_3 bis-thiourea derivative **14** exhibited more potent activity than its *para*-Type III analog **22**, which suggested that the position of substituted bis-thiourea groups on the central aromatic ring or the isomeric effect impacted the decreases of both IC2 and Mor11p descriptor values (**14**: IC2 = 3.809, Mor11p = 1.702 and **22**: IC2 = 3.757, Mor11p = 1.535). This might be due to *meta* isomeric form of bis-thiourea is an appropriate molecule for interacting with the site of action compared with its *para*-isomer. Similar effect was observed for chloro-bis-sulfonamide derivatives, in which the shortening of the alkyl chain linker of compound **28** gave the less active compound **38** with lower descriptor values of IC2 and Mor11p (Table S3 in the supplementary file).

2.3.4. MOLT3 model

$$pIC_{50} = 0.0626\ (F01[C–N]) + 1.0473\ (E2m) + 0.1427\ (MLOGP) - 2.4164 \quad (4)$$

The constructed model demonstrated that mass (E2m), octanol-water partition property (MLOGP), and the frequency of C–N bonds presented in the molecule (F01[C–N]) are influential factors for activity against MOLT3 cell line. In overview, the compounds in Type II series were noted as the most potent class, in which the most potent one (compound **10**)

Table 3. Predictive performance of the constructed* QSAR models.

Cell line	N	Training[1]		LOOCV[2]	
		R_{tr}	$RMSE_{tr}$	R_{CV}	$RMSE_{CV}$
HuCCA-1	20	0.8528	0.1449	0.8528	0.1646
HepG2	27	0.8520	0.2367	0.7628	0.2952
A549	13	0.9636	0.0666	0.9290	0.0926
MOLT-3	29	0.8690	0.2680	0.8114	0.3188
MDA-MB-231	27	0.8301	0.1958	0.7766	0.2217
T47D	27	0.8465	0.1424	0.7966	0.1618

R_{tr}: Correlation coefficient of the training set.
$RMSE_{tr}$: Root mean square error of the training set.
R_{CV}: Correlation coefficient of the leave-one-out cross validation set.
$RMSE_{CV}$: Root mean square error of the leave-one-out cross validation set.
[1] Training set is a dataset used for construction of the model. The dataset includes a set of both descriptor values (*X* variables) and activities (*Y* variables).
[2] LOOCV set is an excluded sample in which its activity (*Y* variable) was predicted using the relationship model constructed by the training set.

containing F atom displayed its pIC_{50} value approximately 0.42 folds greater than that of the second most potent CF_3 analog **13**.

All of the compounds in *meta*-series (Type II) and *para*-series (Type III) had the same value of F01[C–N] value = 8, except for those of NO_2 derivatives **15** and **23**. It was observed that although these NO_2 analogs possessed higher frequency of C–N bonds (F01[C–N] value = 10), but they elicited lower activities than others due to their lower MLOGP and E2m values (Type II **15**: MLOGP = 3.922, E2m = 0.27/Type III **23**: MLOGP = 3.922, E2m = 0.419), Table S3 in the supplementary file.

For Type IV bis-sulfonamide compounds, it was found that the *meta-/ortho-/para*-positions of NO_2 attached to distal aromatic rings affected activity of the compounds (activity: *para* > *meta* > *ortho*; **31** > **33** > **32**) by altering the value of mass descriptor (E2m; **31** = 0.667, **33** = 0.321, **32** = 0.253, Table S3 in the supplementary file), in which the higher descriptor value was observed for the higher activity.

2.3.5. MDA-MB-231 model

$$pIC_{50} = 0.2508\ (ATS7m) + 0.0153\ (RDF090m) + 0.0151\ (MLOGP2) - 3.0427 \quad (5)$$

The QSAR equation indicated that mass and octanol-water partition property are important properties contributing to the anticancer effect against MDA-MB-231 cell line. The most potent compound **14** possessed the highest ATS7m (4.732) value. The effect of *meta-/para*-position of the attached bis-thiourea on the central ring was also observed. It was shown that Type III *para*-bis-thiourea compounds exhibited lower activity than its Type II *meta*-bis-thiourea analogs (*meta* > *para*: **14** > **22**, **13** > **21**, **12** > **20**) due to their lower values of mass descriptors (i.e., ATS7m and RDF090m, Table S3 in the supplementary file).

Considering the Type IV bis-sulfonamide analogs, naphthalene-containing bis-sulfonamide **26**, performed the best activity due to its highest ATS7m value (ATS7m = 3.981). Additionally, the compound bearing *para*-NO_2 on the distal aromatic rings provided better activity than the *meta*-NO_2 compound (**31** > **33**) because of its higher mass descriptors values (i.e., ATS7m and RDF090m, Table S3 in the supplementary file).

Type I thiourea compounds exhibited the poorest activity and ranked last among all the studied compounds, except for di–CF_3–derivative **6**, which possessed the highest values of two predictors (i.e., ATS7m and MLOGP2) among all others in the same class, Table S3 in the supplementary file).

2.3.6. T47D model

$$pIC_{50} = 0.0168\ (RDF095m) + 0.0142\ (F06[F–F]) + 0.0292\ (RDF150m) - 1.9864 \quad (6)$$

QSAR analysis revealed that cytotoxic activity of the investigated compounds was influenced by mass (RDF150m and RDF095m) and distance between fluoro groups in the molecule (F06[F–F]). It was observed that only two di-CF_3 bis-thiourea compounds (*meta*-series **14** and *para*-series **22**) possessed high F06[F–F] value of 18, while those of others were absence (F06[F–F] = 0, Table S3 in the supplementary file). In addition, mono-diCF_3 derivative of Type I thiourea **6** with F06[F–F] = 9 was noted. Unlike other cell lines that the most potent compounds were members of the Type II series, the Type III *para*-compound **22** was noted to be the most effective anticancer agent against the T47D cell line. Compound **22** displayed the highest value of RDF150m = 20.658, which is 5.96 folds greater than its *meta*-analog **14**. Although it had the lower RDF095m value than that of the compound **14** (RDF095m: **14** = 21.091 and **22** = 14.504), but this descriptor had lower influence on the activity, as shown by its lower regression coefficient value when compared with that of RDF150m (Eq. 6). With an exception for the most potent compound **22**, most of the derivatives in *meta*-Type II series (i.e., F, Cl, CF_3, and NO_2) provided better activities than their *para*-analogs in Type III series (*meta* > *para*: **10** > **18**, **11** > **19**, **13** > **21**, and **15** > **23**).

In overview, QSAR analysis provided information for insight structure-activity relationship analysis, which is summarized in Figure S1 and Table S6 in the supplementary file. Generally, an introduction of another thiourea into the molecules provide bis-thioureas with improved activities (Type II *meta*-bis-thiourea/Type III *para*-bis-thiourea > Type I thiourea). However, it was found that the attachment of another thiourea group should be *via meta*-position on the central aromatic ring to provide potent activity. The replacement of thioureas with sulfonamides led to decreased activity of the compounds (Type II > Type IV). Types of substitutions on the distal benzene ring influenced activities of the derivatives, in which the di-CF_3 group tend to be the most effective one for potent activity. Additionally, appropriate length of alkyl chain linker attached to the central aromatic ring affected activities of the compounds. Shortening the linker chain, mostly, leads to compounds with impaired activity.

3. Materials and methods

3.1. General procedure for the synthesis of thioureas (1–23) [19,32,33]

Benzylamine (4 mmol) and appropriate phenylisothiocyanate (4 mmol) were mixed and stirred at room temperature for 3–16 h (monitored by TLC). The formed thiourea derivatives **1**–**7** as solid products were filtered and recrystallized [19]. In case of bis-thioureas **8**–**23**, they were synthesized using xylylenediamine (2 mmol) and phenylisothiocyanate (4 mmol) [32, 33].

Chemical structures of all thiourea derivatives **1**–**23** were confirmed by ^{1}H NMR, ^{13}C NMR, HRMS and IR analysis. Their spectral data have been reported in the literature [19, 32, 33].

3.2. General procedure for the synthesis of sulfonamides (24–36) [20]

Appropriate benzenesulfonyl chloride (5 mmol) and sodium carbonate (10 mmol) in dichloromethane (20 mL) were mixed and stirred. A solution of xylylenediamine (2.5 mmol) in dichloromethane (50 mL) was subsequently added in dropwise manner to the stirred mixture then stirred at room temperature for 15–24 h (monitored by TLC). Distilled water (20 mL) was added, organic phase was separated, and the aqueous phase was extracted with dichloromethane (2 × 30 mL). Subsequently, the organic extracts were combined and washed with 30 mL of water. The organic layer then was dried over anhydrous sodium sulfate, filtered, and evaporated under reduced pressure to give dry crude product. The crude product then was purified by column chromatography on silica gel or recrystallization to give sulfonamide derivatives (**24**–**33**, **35** and **36**). In case of sulfonamide **34**, it was synthesized from the reduction of sulfonamide **33** using stannous chloride in refluxing ethanol.

3.3. General procedure for the synthesis of sulfonamides (37–38) [20]

A solution of *meta*-phenylenediamine (2.5 mmol) in pyridine (5 mL) was prepared, then 5 mmol of benzenesulfonyl chloride was added and stirred for 6 h under reflux. Then the mixture was concentrated under reduced pressure and 20 mL of water was added. The extraction then was performed with dichloromethane (3 × 20 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and evaporated under reduced pressure to give dry crude product. The purification of the crude product was carried out by column chromatography on silica gel to provide the sulfonamide derivatives.

Chemical structures of all sulfonamide derivatives **24**–**38** were confirmed by ^{1}H NMR, ^{13}C NMR, HRMS and IR analysis. Spectral data of the compounds have been reported in the literature [20].

4. Cytotoxic assay: cancer cell lines

Six types of commercially available human cancer cell lines were used for cytotoxic investigations including i) cholangiocarcinoma (HuCCA-1, Immunology laboratory Siriraj Hospital), ii) hepatocellular carcinoma (HepG2, ATCC: HB-8065), iii) lung carcinoma (A549, ATCC: CCL-185), iv) acute lymphoblastic leukemia (MOLT-3, ATTC: CRL-1552), v) hormone-independent breast cancer (MDA-MB-231, ATCC: HTB-26), vi) hormone-dependent breast cancer (T47D, ATCC: HTB-133).

Cells were suspended in the culture medium and then inoculated on 96-well microtiter plates (Corning Inc., NY, USA) at a density of 10,000–20,000 cells per well, then incubated at 37 °C in a humidified atmosphere with 95% air and 5% CO_2 for 24 h. Additional medium containing serial dilutions of either the tested compounds, positive control (etoposide and/or doxorubicin), or negative control (DMSO) were equally added to achieve desired final concentrations and the prepared microtiter plates were further incubated for 48 h. After the incubation complete, the number of survived cells in each well were counted using MTT assay [42, 43] (for HuCCA-1, HepG2, A549, MDA-MB-231 and T47D cells) and XTT assay [44] (for MOLT-3 cells). Then the IC_{50} value was calculated to define a drug (or compound) concentration that required for 50% inhibition of cell growth (relative to negative control). The compounds exhibited $IC_{50} >$ 50 μg/mL were considered as non-cytotoxic.

5. QSAR study

The general workflow of the QSAR modeling is simply explained herein. Only active compounds were included in the QSAR study. Datasets were initially prepared from a series of experimental bioactivity values (IC_{50} values as dependent Y variables) and calculated parameters of chemical structures (descriptor values as X variables). All chemical structures were geometrically optimized to provide most stable conformations for descriptor calculation. The optimized structures were calculated to obtain a large set of descriptor variables. Feature selection was performed to finally select a set of informative descriptor variables as predictors for the QSAR modeling. QSAR models were constructed using MLR algorithm to find the relationships between these set of predictors (selected descriptors) and bioactivity potency (IC_{50}). Reliability (predictive performance) of the built models was assessed using statistical parameters. Details of each step are provided in following subsections.

5.1. Datasets

Datasets were prepared using experimental bioactivities (IC_{50} values) and chemical structures of the tested compounds (**1**–**38**, Figure 1). Only experimentally active compounds were included for QSAR analysis while inactive ones were excluded from the datasets. Six datasets were separately prepared according to anticancer activity against 6 cancer cell lines. To normalize the data points, the bioactivity IC_{50} values were converted to pIC_{50} values by taking the negative logarithm to the base of 10 (-log IC_{50}), Table S2 in the supplementary file.

5.2. Molecular structure optimization and descriptor calculation

Chemical structures of the tested compounds (**1**–**38**) were generated using the GaussView software [45]. All drawn structures were geometrically optimized to obtain low-energy conformers for subsequence descriptor calculations. The optimization process was initially performed using Gaussian 09 [46] at the semi-empirical Austin Model 1 (AM1) level followed by density functional theory (DFT) calculation using the Becke's three-parameter hybrid method with the Lee-Yang-Parr correlation functional (B3LYP) together with the 6–31g(d) level. The optimized structures then were subjected to extraction of quantum chemical descriptors and calculation of molecular descriptors.

In-house developed script was used to extract a set of 13 quantum chemical descriptors. The in-house developed script is a python-based coding written by our research group. The code was written to facilitate the manipulation of quantum chemical descriptor calculation by allowing the calculation and extraction of a set of multiple types of descriptor values of several compounds within the same time. The optimized chemical structures were used as input files for extraction running. All input files were arranged to be located within a single assigned folder to allow the commanding, running, and returning the results. Finally, a single text-type file compiling calculated descriptor values of all input compounds was obtained for further processing. A set of obtained quantum chemical descriptors included Mulliken electronegativity (χ), electrophilic index (ω_i), electrophilicity (ω), mean absolute atomic charge (Q_m), electron affinity (EA), ionization potential (IP), total energy (E_{total}), total dipole moment (μ), hardness (η), softness (S), lowest unoccupied molecular orbital energy (E_{LUMO}), highest occupied molecular orbital energy (E_{HOMO}), and energy difference of HOMO and LUMO ($HOMO\text{-}LUMO_{Gap}$).

Calculation of molecular descriptor values of the optimized structures were subsequently carried out using Dragon software (version 5.5) [47] to obtain an additional set of 1,562 descriptor values. Total 22 classes of dragon descriptors were included in this study such as 3D-MoRSE descriptors, Walk and path counts, 2D autocorrelation, RDF descriptors, 2D frequency fingerprints, 2D binary fingerprints, Topological descriptors, Topological charge indices, Charge descriptors, Information indices, Connectivity indices, Atom-centred fragments, WHIM descriptors, Constitutional descriptors, Edge adjacency indices, GETAWAY descriptors, Burden eigenvalues, Randic molecular profiles, Functional group counts, Geometrical descriptors, Eigenvalue-based indices, and Molecular properties. Finally, both calculated quantum chemical and molecular descriptors were subjected to feature selection process.

5.3. Feature selection

All calculated descriptors were initially filtered to prioritize descriptors (as predictors) which are highly correlated to bioactivity using correlation-based feature selection method. Pearson's pair-correlation values (r) were calculated for each pair of descriptor and bioactivity (pIC_{50} values). Cut-off value of 0.6 ($|r| \geq 0.6$) was used to select highly correlated descriptor variables whose $|r| \geq 0.6$, while those with low correlation ($|r| < 0.6$) were filtered out. The remaining highly correlated descriptors (with ($|r| \geq 0.6$) were subjected to further processes of selection including stepwise MLR and/or attribute selection (CfsSubsetEval, Best First) as implemented in Waikato Environment for Knowledge Analysis (WEKA) version 3.4.5 [48]. After the feature selection process, a set of informative descriptors was finally obtained for multivariate analysis (Table S3 in the supplementary file).

5.4. Multivariate analysis

MLR algorithm was selected as a machine learning algorithm for the QSAR modeling in this study due to its interpretable characteristic. The multivariate analysis was performed by WEKA version 3.4.5 [48] using the MLR algorithm. Values of selected descriptors and pIC_{50} values were prepared as final datasets for the model construction. Descriptor values

were assigned as independent variables (X) while pIC_{50} values were assigned as dependent variable (Y). The MLR model was constructed according to the following equation:

$$Y = B0 + \sum B_n X_n \quad (7)$$

where Y is the pIC_{50} values of compounds, B_0 is the intercept and B_n are the regression coefficient of descriptors X_n.

5.5. Data sampling

During the model construction, the dataset was randomly separated using leave-one-out cross validation (LOO-CV) method to divide the dataset into two sets (i.e., training set and testing LOO-CV set). The LOO-CV is described as a method that one sample was removed from the whole dataset (N) to be used as testing set (in which its bioactivity, as Y value, was predicted by the equation built by the training set) and the remaining samples (N-1) were used as training set (in which the algorithm was learned to find relationship between predictors and dependent variables to give predictive equation). This same sampling process was repeated until every sample in the dataset was chosen as the testing set to predict Y variable (activity).

5.6. Evaluating the performance of QSAR models

Reliability of the constructed models was assessed by statistical parameters calculated by average values of several rounds of the sampling. Two aspects of the model performance were considered by calculated correlation coefficient (R) and root mean square error (RMSE) values to reflect the predictive performance and predictive errors of the built models, respectively. Accordingly, the model with good reliability should provide high R but low RMSE values.

6. Conclusions

Discovery of novel classes of anticancer agent is one of on-going research areas. In this work, the cytotoxic effects of the S-containing compounds (**1–38**) against 6 human cancer cell lines (i.e., HuCCA-1, HepG2, A549, MOLT-3, MDA-MB-231, and T47D) were studied. *In vitro* results indicated that fluorine-containing compounds in Type II *meta*-bis-thioureas (**10**, **13**, and **14**) and Type III (*para*-bis-thiourea **22**) are promising lead compounds for further development. Interestingly, Type II compounds (especially the most potent compound **10**) exhibited the most promising cytotoxic effect against MOLT-3 cell line among all the tested cell lines displaying the lowest range of IC_{50} values. Additionally, some of the studied compounds (**3–7**, **10**, **13–15**, **22, 25**, **28** and **38**) showed broad-spectrum anticancer effects against multiple types of cell lines, which indicated their wide range possibilities for further development. Subsequently, experimental activity (IC_{50} values) and chemical structures of the compounds were used as input datasets for *in silico* QSAR modeling. According to anticancer activities against six cancer cell lines, six interpretable QSAR models were separately constructed using MLR algorithm. The constructed QSAR models showed high correlation coefficient (R) but low RMSE values, which indicated their good predictive performance. In-depth structure-activity relationships analysis was performed by considering the values of key descriptors (predictors) presented in the QSAR equations, as a result, a set of key chemical properties required for potent anticancer activities against each cell lines were revealed. Collectively, the present study demonstrated the facilitating role of *in silico* QSAR modeling for successful drug development by providing insightful knowledge which would be useful for further related research in terms of screening/filtering of potential compounds from large library, guiding effective rational design, as well as facilitating structural modifications to achieve new derivatives with improved properties. However, additional studies to elucidate possible mechanisms of action, pharmacokinetic profiles, as well as supportive *in vivo* models are required for further development as therapeutics.

Declarations

Author contribution statement

Ratchanok Pingaew: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Veda Prachayasittikul: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Apilak Worachartcheewan, Anusit Thongnum: Performed the experiments; Contributed reagents, materials, analysis tools or data.

Supaluk Prachayasittikul, Virapong Prachayasittikul: Conceived and designed the experiments; Analyzed and interpreted the data.

Somsak Ruchirawat: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

Ratchanok Pingaew was supported by Srinakharinwirot University (grant no. 654/2563). Dr. Veda Prachayasittikul was supported by Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation, Research Grant for New Scholar (grant no. "RGNS 64-167"), and by Mahidol University (Basic Research Fund: fiscal year 2022).

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi:10.1016/j.heliyon.2022.e10067.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249.
[2] X. Wang, H. Zhang, X. Chen, Drug resistance and combating drug resistance in cancer, Cancer Drug Resist. 2 (2019) 141–160.
[3] A. Carugo, G.F. Draetta, Academic discovery of anticancer drugs: historic and future perspectives, Annu. Rev. Cell Biol. 3 (2019) 385–408.
[4] S.J. Nass, M.L. Rothenberg, R. Pentz, H. Hricak, A. Abernethy, K. Anderson, A.W. Gee, R.D. Harvey, S. Piantadosi, M.M. Bertagnolli, D. Schrag, R.L. Schilsky, Accelerating anticancer drug development - opportunities and trade-offs, Nat. Rev. Clin. Oncol. 15 (2018) 777–786.
[5] W.N. Hait, P.F. Lebowitz, Moving upstream in anticancer drug development, Nat. Rev. Drug Discov. 18 (2019) 159–160.
[6] M. Feng, B. Tang, S.H. Liang, X. Jiang, Sulfur containing scaffolds in drugs: synthesis and application in medicinal chemistry, Curr. Top. Med. Chem. 16 (2016) 1200–1216.
[7] R. Ronchetti, G. Moroni, A. Carotti, A. Gioiello, E. Camaioni, Recent advances in urea- and thiourea-containing compounds: focus on innovative approaches in medicinal chemistry and organic synthesis, RSC Med. Chem. 12 (2021) 1046–1064.
[8] V. Kumar, S.S. Chimni, Recent developments on thiourea based anticancer chemotherapeutics, Anti Cancer Agents Med. Chem. 15 (2015) 163–175.
[9] A. Mishra, S. Batra, Thiourea and guanidine derivatives as antimalarial and antimicrobial agents, Curr. Top. Med. Chem. 13 (2013) 2011–2025.
[10] C. Zhao, K.P. Rakesh, L. Ravidar, W.Y. Fang, H.L. Qin, Pharmaceutical and medicinal significance of sulfur (S(VI))-containing motifs for drug discovery: a critical review, Eur. J. Med. Chem. 162 (2019) 679–734.
[11] F.A. Khan, S. Mushtaq, S. Naz, U. Farooq, A. Zaidi, S.M. Bukhari, A. Rauf, M.S. Mubarak, Sulfonamides as potential bioactive scaffolds, Curr. Org. Chem. 22 (2018) 818–830.

[12] A. Ammazzalorso, B. De Filippis, L. Giampietro, R. Amoroso, N-acylsulfonamides, Synthetic routes and biological potential in medicinal chemistry, Chem. Biol. Drug Des. 90 (2017) 1094–1105.

[13] C.T. Supuran, Special issue: sulfonamides, Molecules 22 (2017) 1642.

[14] J. Ceramella, A. Mariconda, C. Rosano, D. Iacopetta, A. Caruso, P. Longo, M.S. Sinicropi, C. Saturnino, α–ω Alkenyl-bis-S-guanidine thiourea dihydrobromide affects HeLa cell growth hampering tubulin polymerization, ChemMedChem 15 (2020) 2306–2316.

[15] J.C. Shing, J.W. Choi, R. Chapman, M.A. Schroeder, J.N. Sarkaria, A. Fauq, R.J. Bram, A novel synthetic 1,3-phenyl bis-thiourea compound targets microtubule polymerization to cause cancer cell death, Cancer Biol. Ther. 15 (2014) 895–905.

[16] S. Qaiser, M.S. Mubarak, S. Ashraf, M. Saleem, Z. Ul-Haq, M. Safdar, A. Rauf, T. Abu-Izneid, M.I. Qadri, A. Maalik, Benzilydene and thiourea derivatives as new classes of carbonic anhydrase inhibitors: an in vitro and molecular docking study, Med. Chem. Res. 30 (2021) 552–563.

[17] S.M. Morsy, A.M. Badawi, A. Cecchi, A. Scozzafava, C.T. Supuran, Carbonic anhydrase inhibitors. biphenylsulfonamides with inhibitory action towards the transmembrane, tumor-associated isozymes IX possess cytotoxic activity against human colon, lung and breast cancer cell lines, J. Enzym. Inhib. Med. Chem. 24 (2009) 499–505.

[18] J. Janočková, J. Plšíková, J. Koval, R. Jendželovský, J. Mikeš, J. Kašpárková, V. Brabec, S. Hamuľaková, P. Fedoročko, M. Kožurková, Tacrine derivatives as dual topoisomerase I and II catalytic inhibitors, Bioorg. Chem. 59 (2015) 168–176.

[19] R. Pingaew, V. Prachayasittikul, N. Anuwongcharoen, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Synthesis and molecular docking of N,N'-disubstituted thiourea derivatives as novel aromatase inhibitors, Bioorg. Chem. 79 (2018) 171–178.

[20] R. Leechaisit, R. Pingaew, V. Prachayasittikul, A. Worachartcheewan, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Synthesis, molecular docking, and QSAR study of bis-sulfonamide derivatives as potential aromatase inhibitors, Bioorg. Med. Chem. 27 (2019), 115040.

[21] A.M. Shawky, N.A. Ibrahim, M.A.S. Abourehab, A.N. Abdalla, A.M. Gouda, Pharmacophore-based virtual screening, synthesis, biological evaluation, and molecular docking study of novel pyrrolizines bearing urea/thiourea moieties with potential cytotoxicity and CDK inhibitory activities, J. Enzym. Inhib. Med. Chem. 36 (2021) 15–33.

[22] S.Y. Abbas, R.A.K. Al-Harbi, M.A.M. Sh El-Sharief, Synthesis and anticancer activity of thiourea derivatives bearing a benzodioxole moiety with EGFR inhibitory activity, apoptosis assay and molecular docking study, Eur. J. Med. Chem. 198 (2020), 112363.

[23] H.Q. Li, T. Yan, Y. Yang, L. Shi, C.F. Zhou, H.L. Zhu, Synthesis and structure-activity relationships of N-benzyl-N-(X-2-hydroxybenzyl)-N'-phenylureas and thioureas as antitumor agents, Bioorg. Med. Chem. 18 (2010) 305–313.

[24] A.S. Farooqi, J.Y. Hong, J. Cao, X. Lu, I.R. Price, Q. Zhao, T. Kosciuk, M. Yang, J.J. Bai, H. Lin, Novel lysine-based thioureas as mechanism-based inhibitors of sirtuin 2 (SIRT2) with anticancer activity in a colorectal cancer murine model, J. Med. Chem. 62 (2019) 4131–4141.

[25] S. Ullah, M.I. El-Gamal, S. Zaib, H.S. Anbar, S.O. Zaraei, R.M. Sbenati, J. Pelletier, J. Sévigny, C.H. Oh, J. Iqbal, Synthesis, biological evaluation, and docking studies of new pyrazole-based thiourea and sulfonamide derivatives as inhibitors of nucleotide pyrophosphatase/phosphodiesterase, Bioorg. Chem. 99 (2020), 103783.

[26] E.M.H. Ali, M.S. Abdel-Maksoud, U.M. Ammar, K.I. Mersal, K. Ho Yoo, P. Jooryeong, C.H. Oh, Design, synthesis, and biological evaluation of novel imidazole derivatives possessing terminal sulphonamides as potential BRAF(V600E)inhibitors, Bioorg. Chem. 106 (2021), 104508.

[27] V. Prachayasittikul, A. Worachartcheewan, W. Shoombuatong, N. Songtawee, S. Simeon, V. Prachayasittikul, C. Nantasenamat, Computer-aided drug design of bioactive natural products, Curr. Top. Med. Chem. 15 (2015) 1780–1800.

[28] W. Cui, A. Aouidate, S. Wang, Q. Yu, Y. Li, S. Yuan, Discovering anti-cancer drugs via computational methods, Front. Pharmacol. 11 (2020) 733.

[29] C. Phanus-Umporn, V. Prachayasittikul, C. Nantasenamat, S. Prachayasittikul, V. Prachayasittikul, QSAR-driven rational design of novel DNA methyltransferase 1 inhibitors, EXCLI J 19 (2020) 458–475.

[30] R. Pratiwi, V. Prachayasittikul, S. Prachayasittikul, C. Nantasenamat, Rational design of novel sirtuin 1 activators via structure-activity insights from application of QSAR modeling, EXCLI J 18 (2019) 207–222.

[31] A. Worachartcheewan, V. Prachayasittikul, S. Prachayasittikul, V. Tantivit, C. Yeeyahya, V. Prachayasittikul, Rational design of novel coumarins: a potential trend for antioxidants in cosmetics, EXCLI J. 19 (2020) 209–226.

[32] R. Pingaew, N. Sinthupoom, P. Mandi, V. Prachayasittikul, R. Cherdtrakulkiat, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Synthesis, biological evaluation and in silico study of bis-thiourea derivatives as anticancer, antimalarial and antimicrobial agents, Med. Chem. Res. 26 (2017) 3136–3148.

[33] R. Pingaew, P. Tongraung, A. Worachartcheewan, C. Nantasenamat, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Cytotoxicity and QSAR study of (thio)ureas derived from phenylalkylamines and pyridylalkylamines, Med. Chem. Res. 22 (2013) 4016–4029.

[34] T. Saha, M.S. Hossain, D. Saha, M. Lahiri, P. Talukdar, Chloride-mediated apoptosis-inducing activity of bis(sulfonamide) anionophores, J. Am. Chem. Soc. 138 (2016) 7558–7567.

[35] C. Isanbor, D. O'Hagan, Fluorine in medicinal chemistry: a review of anti-cancer agents, J. Fluor. Chem. 127 (2006) 303–319.

[36] M. Inoue, Y. Sumii, N. Shibata, Contribution of organofluorine compounds to pharmaceuticals, ACS Omega 5 (2020) 10633–10640.

[37] A. Talevi, J.F. Morales, G. Hather, J.T. Podichetty, S. Kim, P.C. Bloomingdale, S. Kim, J. Burton, J.D. Brown, A.G. Winterstein, S. Schmidt, J.K. White, D.J. Conrado, Machine learning in drug discovery and development part 1: a primer, CPT Pharmacometrics Syst. Pharmacol. 9 (2020) 129–142.

[38] V. Prachayasittikul, R. Pingaew, N. Anuwongcharoen, A. Worachartcheewan, C. Nantasenamat, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Discovery of novel 1,2,3-triazole derivatives as anticancer agents using QSAR and in silico structural modification, SpringerPlus 4 (2015) 571.

[39] V. Prachayasittikul, R. Pingaew, A. Worachartcheewan, S. Sitthimonchai, C. Nantasenamat, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Aromatase inhibitory activity of 1,4-naphthoquinone derivatives and QSAR study, EXCLI J. 16 (2017) 714–726.

[40] V. Prachayasittikul, R. Pingaew, A. Worachartcheewan, C. Nantasenamat, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Synthesis, anticancer activity and QSAR study of 1,4-naphthoquinone derivatives, Eur. J. Med. Chem. 84 (2014) 247–263.

[41] R. Pingaew, V. Prachayasittikul, A. Worachartcheewan, C. Nantasenamat, S. Prachayasittikul, S. Ruchirawat, V. Prachayasittikul, Novel 1,4-naphthoquinone-based sulfonamides: synthesis, QSAR, anticancer and antimalarial studies, Eur. J. Med. Chem. 103 (2015) 446–459.

[42] J. Carmichael, W.G. DeGraff, A.F. Gazdar, J.D. Minna, J.B. Mitchell, Evaluation of a tetrazolium-based semiautomated colorimetric assay: assessment of radiosensitivity, Cancer Res. 47 (1987) 943–946.

[43] T. Mosmann, Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays, J. Immunol. Methods 65 (1983) 55–63.

[44] A. Doyly, J.B. Griffiths, Mammalian Cell Culture-Essential Techniques, John Wiley & Sons, Chichester, New York, Weinheim, Brisbane, Toronto., 1997.

[45] I.I.R. Denning ton, T. Keith, J. Millam, K. Eppinnett, W.L. Hovell, R. Gilliland, GaussView, Version 3.09 [Computer Software], Semichem Shawnee Mission, KS, USA, 2003.

[46] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision B.01, Gaussian Inc., Wallingford CT, 2009.

[47] Talete, Dragon for Windows (Software for Molecular Descriptor Calculations), Version 5.5 [Computer Software]. Milan, Italy, 2007.

[48] I.H. Witten, E. Frank, M.A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed., Morgan Kaufmann, San Francisco, CA, 2011.