European Journal of Medicinal Chemistry 45 (2010) 1470–1477

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

QSAR modeling of taxane analogues against colon cancer

Rajeshwar P. Verma*, Corwin Hansch

Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, CA 91711, USA

ARTICLE INFO

Article history:
Received 7 May 2009
Received in revised form
16 December 2009
Accepted 23 December 2009
Available online 13 January 2010

Keywords:
Colon cancer
Hydrophobicity
Molar refractivity
QSAR
Taxanes

ABSTRACT

Although taxanes such as paclitaxel and docetaxel are the two most important clinically available anticancer drugs for the treatment of various cancers (including colon cancer), the success of these two drugs has been tempered by the development of various unbearable side effects as well as multi-drug resistance. Therefore, it is essential to search new taxane analogues with improved anticancer activity and fewer side effects to gain the maximum benefits for colon cancer patients. In this paper, four series of taxane derivatives were used to correlate their inhibitory activities against colon cancers mainly with the hydrophobic and steric descriptors of their substituents in order to gain a better understanding of their chemical–biological interactions. QSAR results from this paper have suggested that the steric and hydrophobic parameters of the substituents are the two most important determinants for the activities of taxane analogues (under consideration) against colon cancers, with a major contribution coming from the molar refractivity of the substituents. Statistical diagnostics, internal validation, and external validation tests have validated all the QSAR models.

1. Introduction

Colorectal (colon and rectum) cancer is the third most common cancer in both men and women. An estimate from the American Cancer Society has insinuated that about 108,070 and 40,740, respectively, the new cases of colon and rectal cancers, and about 49,960 deaths from these cancers were expected to occur in the United States during the year of 2008, accounting for about 11% of all the cancer-related deaths [1]. Recent studies have suggested that dominant mutations in the tumor suppressor gene adenomatous polyposis coli (APC) derive tetraploid formation causing the failures in cytokinesis before the earliest steps associated with colorectal cancer progression [2]. These mutations generally cause the loss of C-terminal functions of the APC protein – possibly due to the involvement of microtubule binding, cell polarity, chromosome segregation, and deletion of the SAMP (serine, alanine, methionine, and proline) repeats that are the most important factors for binding to axin and formation of the β-catenin phosphorylation complex. The APC proteins are usually stable and retain some activity in β-catenin binding [3]. The loss of APC immediately induces chromosomal instability as a result of combined mitotic and apoptotic defects [4]. Deregulation of β-catenin signaling is an important consequence of the APC loss, which was strongly supported by the genetic and molecular data. However, APC has functions independent of β-catenin regulation, e.g. in microtubule plus-end binding, stabilization, and efficient spindle checkpoint activation [5,6]. Although the genetic defect in familial adenomatous polyposis affects the rate of tumor initiation by targeting the gatekeeper function of the APC gene, the defect in hereditary nonpolyposis colorectal cancer largely affects tumor progression by targeting the genome guardian function of DNA mismatch repair [7].

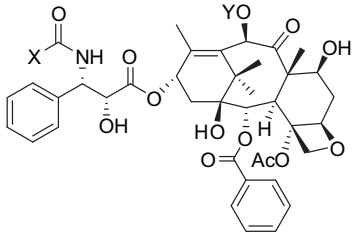
Experimental and epidemiological evidences have illustrated that the diet and nutrition are the two key factors in modulating colon cancer attack and progression [8]. New Western diet (NWD) is one of the major dietary risk factors for the human colon cancer. Colonic tumors can be prevented by elevating dietary calcium and vitamin D3 to the upper level limit as consumed by humans, but tumorigenesis cannot be altered by similarly increasing folate, choline, methionine, or fiber; each of these are also present at the lower levels in the NWD and are associated with the high risk for colon cancer [9]. In a very recent study, it had been demonstrated that the low folate and vitamin B6 intake was associated with an increased risk of p53-overexpressing colon cancers but not wild-type tumors [10]. Selenium is proved to be an important micronutrient engaged in the protection of colonic cells against a wide range of external and internal stressors. It inhibits actively the growth of malignant colonic cells as well as to induce their demise. Thus, the use of selected selenium compounds may be suggested in conjunction with established antineoplastic drugs [11].

* Corresponding author. Tel.: +1 909 607 4249; fax: +1 909 607 7726.
E-mail address: rverma@pomona.edu (R.P. Verma).

0223-5234/$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.12.054

Strong evidence has been generated from the cell culture and animal studies that the tea has a protective effect against carcinogenesis. Tea is an aqueous infusion from the dried leaves of *Camellia sinensis*. It is one of the most widely used beverages throughout the world. It possesses various beneficial properties that may affect carcinogen metabolism, free radical scavenging, or the formation of DNA adducts [12,13]. Epidemiological and laboratory studies have identified epigallocatechin gallate (EGCG), in green tea polyphenols, as one of the most potent chemopreventive agents that can induce apoptosis, suppress the formation and growth of human cancers including colorectal cancer [14]. The treatment of colon cancers by EGCG has resulted in a strong activation of AMP-activated protein kinase (AMPK) and an inhibition of COX-2 expression. Furthermore, it has also been found that the reactive oxygen species (ROS) is an upstream signal of AMPK, and the combined treatment of EGCG with other chemotherapeutic agents such as 5-Fluorouracil and Etoposide, exert a novel therapeutic effect on chemo-resistant colon cancers [15]. In the past 15 years, 5-Fluorouracil and leucovorin have been considered as the standard drugs for the treatment of colorectal cancer that have now been replaced by the combination chemotherapy, at least in the stage III disease. The treatment of stage II disease is still somewhat less established [16]. Although oxaliplatin-based chemotherapy is now considered as the standard care in node-positive colon cancer; it remains controversial for the patients with node-negative disease. It is interesting to note here that fluoropyrimidines play an increasing role in the management of colorectal cancer and can be considered as an alternative to 5-fluorouracil. Despite the progress achieved with the introduction of new cytotoxic agents, recurrence rates for the colon cancer patients with respect to stage II/III disease remain >20% [17].

Taxanes are considered to be the most powerful group of compounds among the current novel chemotherapeutic drugs. Taxane analogues such as paclitaxel (**1**) and docetaxel (**2**) are already in use for the treatment of various types of cancers; including breast, colon, lung, ovarian, and prostate [18–21]. These two drugs bind to the β-subunit of tubulin and promote tubulin polymerization, leading to the inhibition of microtubule dynamics, cell cycle arrest, and ultimately cell death by apoptosis [22–25]. Although these two drugs (**1** and **2**) have made great milestones in the treatment of various cancers, clinical reports have revealed that their use often results in a number of undesirable side effects and multi-drug resistance (MDR). These side effects, along with the onset of MDR, clearly suggest the necessity of developing new taxane derivatives with much improved anticancer activity, fewer side effects, and superior pharmacological properties to maximize the induced benefits for colon cancer patients. This may also reduce the large economic and disease burden worldwide. The quantitative structure–activity relationship (QSAR) paradigm may provide helpful suggestions in the design and development of novel taxane analogues that are expected to achieve the improvements in their anticancer activity against colon cancer, as well as in their toxicity profile, and pharmacology.



X = C_6H_5, Y = $COCH_3$; Paclitaxel (1)
X = $OC(CH_3)_3$, Y = H; Docetaxel (2)

In the present study, we developed four QSAR models on four different sets of taxane analogues (**3–6**) with respect to their anticancer activities against colon cancers. The QSAR methodology is a powerful mathematical tool that has been used successfully since past four decades in understanding many aspects of chemical–biological interactions in drug-design process, pesticide research, and in the areas of toxicology [26–31]. This method can shed light on mechanisms of action of various ligands with enzymes, membranes, organelles, cells, virus, bacteria, and human [32–41]. It has also been frequently utilized for the evaluation of absorption, distribution, metabolism, and excretion phenomena in many organisms and whole animal studies [40,41]. The QSAR method utilizes computational-based and extra-thermodynamically derived descriptors to correlate biological activity in isolated receptors, in cellular systems, and in vivo. There are mainly four standard molecular descriptors such as electronic, hydrophobic, steric, and topological indices, routinely used in the QSAR analysis. QSAR models can stand alone to augment other computational approaches. It can be examined in tandem with equations of a similar mechanistic genre to establish its authenticity and reliability [42].

2. Experimental

All the biological data were taken from the literature and the reference has been cited with their respective QSAR. IC_{50} is the molar concentration of a compound that inhibits 50% of growth of the cancer cell population; log $1/IC_{50}$ is the subsequent dependent variable that defines the biological parameter for the QSAR model. With the aid of the C-QSAR program, multiregression analyses were applied to derive appropriate QSAR models using autoloaded descriptors [43]. In order to avoid the collinearity problems, the selection of descriptors was carried out with the aid of permutation and correlation matrices among the descriptors. Details about the use of C-QSAR program, in the development of QSAR models, have already been discussed in refs. [44,45].

2.1. Molecular descriptors

Clog *P* is the calculated partition coefficient of a compound in 1-octanol/water system and is a measure of its hydrophobicity, whereas π is the hydrophobic parameter for the substituent only. Similarly, CMR is the calculated molar refractivity for the whole molecule, whereas MR is the molar refractivity for the substituent only. MR has been scaled at 0.1 to make it equiscalar with π. CMR is calculated by using the Lorentz–Lorenz equation: $[(n^2-1)/(n^2+2)](MW/\delta)$, where n is the refractive index, MW is the molecular weight, and δ is the density of the substance. *B*5 and *L* are the Verloop's sterimol parameters for substituents, where *B*5 is a measure of the maximum width of the substituent and *L* is the substituent length [46]. The indicator variable *I* is assigned the value of 1 and 0 for the presence and absence, respectively, of the certain structural features with unusual effects that cannot be parameterized and has been explained wherever it comes to play.

2.2. Statistical parameters

In all the QSAR models, n is the number of data points, r^2 is the square of the correlation coefficient, q^2 is the cross-validated r^2, s is the standard deviation, and the data within the parentheses represents the 95% confidence intervals. The leave-one-out (LOO) cross-validated r^2 (q^2 or LOO–q^2) was obtained by using the LOO procedure of Cramer III et al. [47]. Similarly, the leave-five-out (LFO) cross-validated r^2 (q^2_f or LFO–q^2) was obtained by using the similar LFO procedure. *Q* is the quality factor, for which $Q = r/s$ (where, r is

the correlation coefficient and s is the standard deviation). F is the Fischer statistics, for which $F = fr^2/[(1-r^2)m]$, where f is the number of degrees of freedom [$f = n - (m + 1)$], n is the number of data points, and m is the number of variables. The value of F within parentheses represents the literature F-value at 95% level [48].

2.3. Outliers

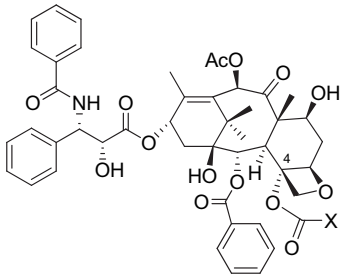
Only those compounds, which have unexpected biological activities and unable to fit in a QSAR model, are considered as outliers. Although the presence of outliers is mainly due to the possibility that the molecules may act by different mechanisms or interact with the receptor in different binding modes, it may also due to the intrinsic noise associated with both the original data and methodological aspects involved in the development of QSAR models [49–51]. Compounds were considered to be outliers on the basis of their deviation between observed and predicted activities from the model (observed activity – predicted activity $> 2s$, where s is the standard deviation) [52–55].

2.4. Model validation

The QSAR model validation was carried out in three steps such as statistical diagnostics, internal validation, and external validation.

3. Results and discussion

3.1. QSAR for the inhibition of growth of HCT-116 (human colon cancer cells) by paclitaxel analogues (3)



Chen et al. [56] synthesized a series of C-4 modified paclitaxel analogues (**3**) and evaluated for their biological activities in a cytotoxicity assay against HCT-116 cell line and results were given in IC_{50} (nM). Those values of IC_{50} were converted into log $1/IC_{50}$ in molar concentration and given in Table 1. From the data in Table 1, QSAR Eq. (1) was developed.

$$\log 1/IC_{50} = 1.34(\pm 0.36)\pi_X - 1.97(\pm 0.42)MR_X + 0.89(\pm 0.43)I_{CYALK} + 9.05(\pm 0.44) \quad (1)$$

$$n = 22, r^2 = 0.855, s = 0.360, q^2 = 0.786, q_f^2 = 0.772, Q = 2.569, F_{3,18} = 35.379 (3.159)$$

IC_{50} is the molar concentration of paclitaxel analogues (**3**) that measures the drug concentration required for the inhibition of HCT-116 cell proliferation by 50% after 72 h of incubation. π_X and MR_X are the calculated hydrophobicity and molar refractivity of the X substituents, respectively. According to this QSAR model, the paclitaxel derivative (**3**) must have a more hydrophobic but smaller or less polarizable X substituent for improved cytotoxicity. I_{CYALK} is an indicator variable for the unusual activity of the cycloalkyl containing X substituents, where $I_{CYALK} = 1$ for the presence of cycloalkyl containing X substituent otherwise $I_{CYALK} = 0$. The positive coefficient of I_{CYALK} suggests that the presence of cycloalkyl containing X substituents will be more favorable to the activity. A comparison between observed and predicted values of log $1/IC_{50}$ for paclitaxel analogues (**3**) used in the development of QSAR Eq. (1) is shown in Fig. 1.

Table 1
Biological (IC_{50}, mol L^{-1}) [56], physicochemical, and structural parameters of paclitaxel analogues (**3**) used to derive QSAR (Eq. (1)).

No.	X	log $1/IC_{50}$ (Eq. (1))			π_X	MR_X	I_{CYALK}
		Obsd.	Pred.	Δ			
1	CH_3	8.62	9.06	−0.44	0.70	0.46	0
2	C_6H_5	6.39	6.65	−0.26	1.91	2.51	0
3	4-F-C_6H_4	6.10	6.81	−0.71	2.05	2.53	0
4	CH_2F	8.15	8.60	−0.45	0.37	0.48	0
5	CCl_3	8.40	8.68	−0.28	2.58	1.94	0
6	C_2H_5	8.70	8.86	−0.16	1.23	0.93	0
7	$CH{=}CH_2$	8.22	8.26	−0.04	0.85	0.98	0
8	$(CH_2)_2CH_3$	8.96	8.65	0.31	1.76	1.39	0
9	$CH(CH_3)_2$	8.30	8.36	−0.06	1.54	1.39	0
10	$C(CH_3) = CH_2$	8.35	7.76	0.59	1.16	1.44	0
11	*trans*-$CH = CHCH_3$	8.64	8.05	0.59	1.38	1.44	0
12	Cy-C_3H_5	9.00	9.18	−0.18	1.28	1.25	1
13	$(CH_2)_3CH_3$	8.70	8.45	0.25	2.28	1.86	0
14	Cy-C_4H_7	8.82	8.79	0.03	1.61	1.68	1
15	$(CH_2)_4CH_3$	8.22	8.24	−0.02	2.81	2.32	0
16	Cy-C_5H_9	8.70	8.62	0.08	2.17	2.14	1
17	OCH_3	8.70	8.74	−0.04	0.68	0.62	0
18	OCH_2CH_3	9.00	8.54	0.46	1.21	1.08	0
19	$O(CH_2)_2CH_3$	8.59	8.33	0.26	1.74	1.54	0
20	NH-Cy-C_4H_7	8.04	7.97	0.07	1.55	2.05	1
21	Imidazole	6.10	5.88	0.22	0.18	1.73	0
22	Aziridine	7.81	8.01	−0.20	0.87	1.12	0

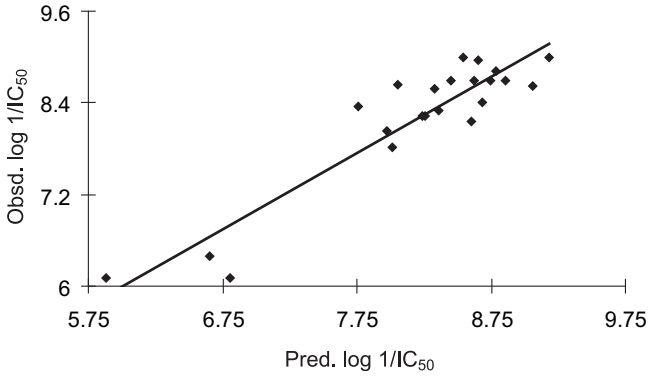
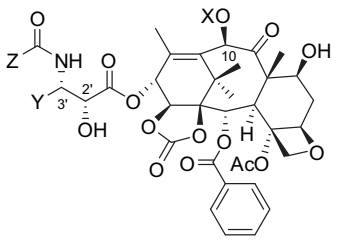


Fig. 1. Plot of observed versus predicted log $1/IC_{50}$ (Eq. (1)).

3.2. QSAR for the inhibition of growth of HT-29 (human colon carcinoma cells) by taxane analogues (4)



QSAR Eq. (2) is based on the data obtained from Ojima et al. [21], for which the original data IC_{50} (nM) was converted into log $1/IC_{50}$ in molar concentration and given in Table 2.

Table 2
Biological (IC_{50}, mol L^{-1}) [21] and physicochemical parameters of taxane analogues (**4**) used to derive QSAR Eq. (2).

No.	X	Y	Z	log 1/IC_{50} (Eq. (2))			MR_Y	π_Z	Clog *P*
				Obsd.	Pred.	Δ			
1	H	2-furyl	$OC(CH_3)_3$	9.22	9.33	−0.11	1.81	2.04	2.68
2	$COCH_3$	2-furyl	$OC(CH_3)_3$	9.22	9.11	0.11	1.81	2.04	3.39
3	H	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	9.30	9.24	0.06	1.92	2.04	3.49
4	$COCH_3$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	8.62	9.02	−0.40	1.92	2.04	4.20
5	$COCH_2CH_3$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	9.22	8.86	0.36	1.92	2.04	4.73
6	CO(Cy-C_3H_5)	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	9.00	8.84	0.16	1.92	2.04	4.79
7	$CON(CH_3)_2$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	8.92	8.89	0.03	1.92	2.04	4.62
8	$COCH=CHCH_3$ (*E*)	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	8.82	8.72	0.10	1.92	2.04	5.17
9	$COOCH_3$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	9.15	9.04	0.11	1.92	2.04	4.15
10	H	$CH=CHCH_3$ (*E*)	$OC(CH_3)_3$	8.72	8.64	0.08	1.45	2.04	3.09
11	$COCH_3$	$CH=CHCH_3$ (*E*)	$OC(CH_3)_3$	8.52	8.42	0.10	1.45	2.04	3.81
12	CO(Cy-C_3H_5)	$CH{=}C(CH_3)_2$	C_5H_{11}	7.80	7.90	−0.10	1.92	2.24	4.99
13	$CON(CH_3)_2$	$CH{=}C(CH_3)_2$	C_5H_{11}	8.29	7.95	0.34	1.92	2.24	4.82
14	$COCH=CHCH_3$ (*E*)	$CH{=}C(CH_3)_2$	C_5H_{11}	7.55	7.79	−0.24	1.92	2.24	5.36
15[a]	H	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	8.38	9.19	−0.81	1.94	2.04	3.78
16	$COCH_3$	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	8.82	8.97	−0.15	1.94	2.04	4.49
17	CO(Cy-C_3H_5)	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	8.48	8.79	−0.31	1.94	2.04	5.07
18	$CON(CH_3)_2$	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	8.89	8.84	0.05	1.94	2.04	4.90
19	H	$(CH_2)_2CH_3$	$OC(CH_3)_3$	8.42	8.59	−0.17	1.48	2.04	3.38
20	$COCH_3$	$(CH_2)_2CH_3$	$OC(CH_3)_3$	8.32	8.37	−0.05	1.48	2.04	4.09

[a] Not included in deriving QSAR Eq. (2).

$$\log 1/IC_{50} = 1.56(\pm 0.72)MR_Y - 4.42(\pm 1.66)\pi_Z - 0.31(\pm 0.19)\text{Clog}P + 16.33(\pm 3.33) \quad (2)$$

$$n = 19, r^2 = 0.828, s = 0.219, q^2 = 0.724, q_f^2 = 0.796, Q = 4.155, F_{3,15} = 24.070(3.287)$$

$$\text{outlier}: X = H, Y = CH_2CH(CH_3)_2, Z = OC(CH_3)_3$$

IC_{50} is the molar concentration of taxane analogues (**4**) that inhibits 50% of the growth of HT-29 cancer cell. MR_Y is the calculated molar refractivity of Y substituents, whereas π_Z is the calculated hydrophobicity of Z substituents. Clog *P* is the calculated hydrophobic parameter for the whole molecule. According to this QSAR model, the taxane derivative (**4**) must have a more hydrophilic Z substituent, a bulkier or more polarizable Y substituent, and lower hydrophobicity of the whole molecule for improved cytotoxicity against HT-29 cancer cells. One compound (X = H, Y = $CH_2CH(CH_3)_2$, Z = $OC(CH_3)_3$) was deemed to be an outlier because it was less active than expected, by 3.7 times the standard deviation. A comparison between observed and predicted values of log 1/IC_{50} for paclitaxel analogues (**4**) used in the development of QSAR Eq. (2) is shown in Fig. 2.

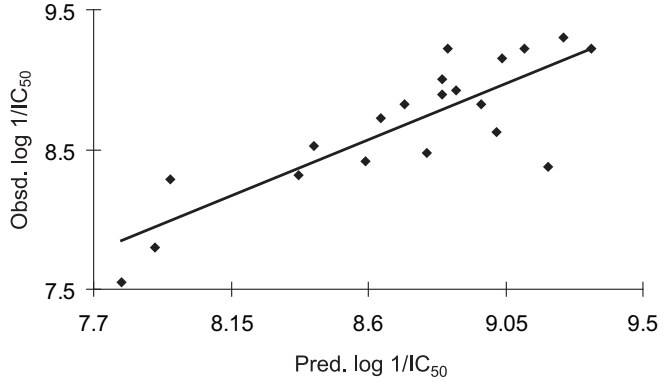
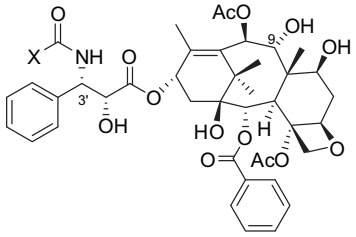


Fig. 2. Plot of observed versus predicted log 1/IC_{50} (Eq. (2)).

3.3. QSAR for the inhibition of growth of HT-29 (human colon adenocarcinoma cells) by taxane analogues (**5**)



Maring et al. [57] synthesized a series of C-3′-*N*-acyl analogues of 9(*R*)-dihydrotaxol (**5**) and tested for their cytotoxic activity against HT-29 cell line and results were given in IC_{50} (μg/mL). Those values were converted into log 1/IC_{50} in molar concentration and given in Table 3. From the data in Table 3, we developed QSAR Eq. (3).

$$\log 1/IC_{50} = 0.97(\pm 0.29)B5_X - 0.75(\pm 0.55)I_{CMe3} + 7.69(\pm 1.10) \quad (3)$$

$$n = 10, r^2 = 0.897, s = 0.277, q^2 = 0.793, q_f^2 = 0.727, Q = 3.419, F_{2,7} = 30.481(4.737)$$

$$\text{outlier}: X = OC(CH_3)_3$$

This is a linear equation in terms of $B5_X$ (Verloop's sterimol width parameter of X substituents) and I_{CMe3} (an indicator variable that pinpoints the unusual activity of $C(CH_3)_3$ containing X substituents, where $I_{CMe3} = 1$ for the presence of $C(CH_3)_3$ fragment in the X group otherwise $I_{CMe3} = 0$). Positive coefficient associated with $B5_X$ suggests that the cytotoxic activity of these compounds against HT-29 cells increases with the increase of $B5_X$. On the contrary, the presence of $C(CH_3)_3$ containing X substituent (negative coefficient of I_{CMe3}) decreases the cytotoxic activity of the analogues. One compound (X = $OC(CH_3)_3$) was deemed to be an outlier because it was more active than expected, by 5.7 times the standard deviation. A comparison between observed and predicted

Table 3
Biological (IC_{50}, mol L^{-1}) [57], physicochemical, and structural parameters of taxane analogues (**5**) used to derive QSAR Eq. (3).

No.	X	log 1/IC_{50} (Eq. (3))			$B5_X$	I_{CMe3}
		Obsd.	Pred.	Δ		
1	4-CH_3-C_6H_4	10.86	10.69	0.17	3.11	0
2	CH_3	9.47	9.66	−0.19	2.04	0
3	$C(CH_3)_3$	ND[b]	10.00	ND[b]	3.17	1
4	$CH_2C(CH_3)_3$	10.98	10.98	0.00	4.18	1
5	$NHC(CH_3)_3$	11.18	11.18	0.00	4.39	1
6[a]	$OC(CH_3)_3$	12.73	11.14	1.59	4.35	1
7	$OC(CH_3)_2CH_2CH_3$	12.11	11.96	0.15	4.42	0
8	$OCH(CH_3)_2$	12.02	11.65	0.37	4.10	0
9	$OCH_2C(CH_3)_3$	11.76	11.96	−0.20	4.42	0
10	$OCH_2CH(CH_3)_2$	11.57	11.96	−0.39	4.42	0
11	OCH_2CH_3	11.22	10.93	0.28	3.36	0
12	$OCH_2C_6H_5$	10.87	11.07	−0.20	3.50	0

[a] Not included in deriving QSAR Eq. (3).
[b] Not determined.

values of log 1/IC_{50} for paclitaxel analogues **V** used in the development of QSAR Eq. (3) is shown in Fig. 3.

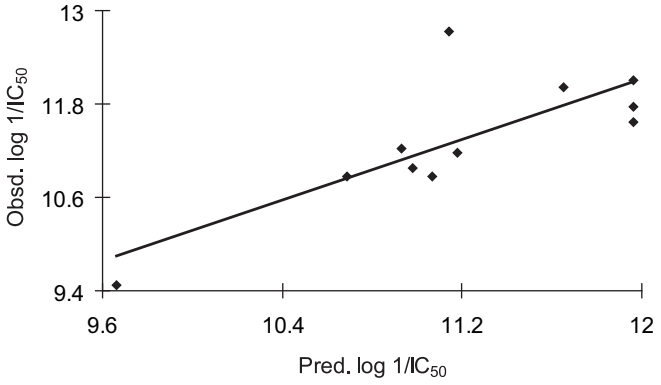
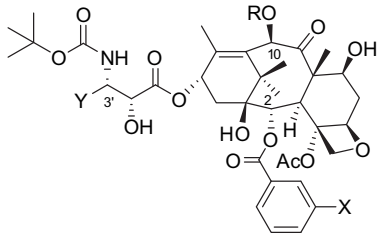


Fig. 3. Plot of observed versus predicted log 1/IC_{50} (Eq. (3)).

3.4. QSAR for the inhibition of growth of HT-29 (human caucasian colon adenocarcinoma cells) by taxane analogues (*6*)



Kuznetsova et al. [58] synthesized a series of 3′-difluoromethyl/trifluoromethyl-taxanes (**6**) with modifications at the C-2 and C-10 positions and evaluated for their *in vitro* cytotoxicities against human breast carcinoma (MCF7-S, MCF7-R, LCC6-WT, LCC6-MDR), non-small cell lung carcinoma (H460) and colon adenocarcinoma (HT-29) cell lines. The cytotoxicity data of these analogues (**6**) against HT-29 cell line (IC_{50}, nM) was converted into log 1/IC_{50} in molar concentration and given in Table 4. From the data in Table 4, QSAR Eq. (4) was developed.

Table 4
Biological (IC_{50}, mol L^{-1}) [58], physicochemical, and structural parameters of taxane analogues (**6**) used to derive QSAR Eq. (4).

No.	R	X	Y	log 1/IC_{50} (Eq. (4))			L_R	$B5_X$	I_{HAL}
				Obsd.	Pred.	Δ			
1	$COCH_3$	OCH_3	CF_2H	9.28	9.20	0.08	4.06	3.07	0
2	COC_2H_5	OCH_3	CF_2H	9.23	9.26	−0.03	4.87	3.07	0
3	$CON(CH_3)_2$	OCH_3	CF_2H	9.37	9.25	0.12	4.77	3.07	0
4	$COOCH_3$	OCH_3	CF_2H	9.36	9.25	0.11	4.73	3.07	0
5	$COCH_3$	F	CF_2H	9.46	9.37	0.09	4.06	1.35	1
6[a]	COC_2H_5	F	CF_2H	9.07	9.43	−0.36	4.87	1.35	1
7	$CON(CH_3)_2$	F	CF_2H	9.46	9.43	0.03	4.77	1.35	1
8	$COOCH_3$	F	CF_2H	9.37	9.42	−0.05	4.73	1.35	1
9[a]	$COCH_3$	Cl	CF_2H	8.71	9.42	−0.71	4.06	1.80	1
10	COC_2H_5	Cl	CF_2H	9.37	9.49	−0.12	4.87	1.80	1
11[a]	$CON(CH_3)_2$	Cl	CF_2H	9.24	9.48	−0.24	4.77	1.80	1
12	$COOCH_3$	Cl	CF_2H	9.54	9.48	0.06	4.73	1.80	1
13	$COCH_3$	N_3	CF_2H	9.24	9.33	−0.09	4.06	4.18	0
14	COC_2H_5	N_3	CF_2H	9.43	9.39	0.04	4.87	4.18	0
15	$CON(CH_3)_2$	N_3	CF_2H	9.40	9.39	0.01	4.77	4.18	0
16	$COOCH_3$	N_3	CF_2H	9.44	9.38	0.06	4.73	4.18	0
17	$COCH_3$	OCH_3	CF_3	9.16	9.20	−0.04	4.06	3.07	0
18	COC_2H_5	OCH_3	CF_3	9.27	9.26	0.01	4.87	3.07	0
19	$CON(CH_3)_2$	OCH_3	CF_3	9.17	9.25	−0.08	4.77	3.07	0
20	$COOCH_3$	OCH_3	CF_3	9.28	9.25	0.03	4.73	3.07	0
21	$COCH_3$	F	CF_3	8.95	8.99	−0.04	4.06	1.35	0
22	COC_2H_5	F	CF_3	8.94	9.05	−0.11	4.87	1.35	0
23	$CON(CH_3)_2$	F	CF_3	9.12	9.04	0.08	4.77	1.35	0
24	$COOCH_3$	F	CF_3	9.07	9.04	0.03	4.73	1.35	0
25	$COCH_3$	Cl	CF_3	9.07	9.04	0.03	4.06	1.80	0
26	COC_2H_5	Cl	CF_3	8.95	9.11	−0.16	4.87	1.80	0
27[a]	$CON(CH_3)_2$	Cl	CF_3	9.35	9.10	0.25	4.77	1.80	0
28	$COOCH_3$	Cl	CF_3	9.17	9.10	0.07	4.73	1.80	0
29	$COCH_3$	N_3	CF_3	9.30	9.33	−0.03	4.06	4.18	0
30	COC_2H_5	N_3	CF_3	9.40	9.39	0.01	4.87	4.18	0
31	$CON(CH_3)_2$	N_3	CF_3	9.30	9.39	−0.09	4.77	4.18	0
32	$COOCH_3$	N_3	CF_3	9.40	9.38	0.02	4.73	4.18	0
33	H	N_3	CF_3	9.15	9.18	−0.02	2.06	4.18	0

[a] Not included in deriving QSAR Eq. (4).

$$\log 1/IC_{50} = 0.08(\pm 0.05)L_R + 0.12(\pm 0.03)B5_X + 0.38(\pm 0.09)I_{HAL} + 8.51(\pm 0.27) \quad (4)$$

$$n = 29, r^2 = 0.807, s = 0.076, q^2 = 0.742, q_f^2 = 0.812, Q = 11.816, F_{3,25} = 34.845(2.991)$$

outliers : R = COC_2H_5, X = F, Y = CF_2H;
R = $COCH_3$, X = Cl, Y = CF_2H;
R = $CON(CH_3)_2$, X = Cl, Y = CF_2H;
R = $CON(CH_3)_2$, X = Cl, Y = CF_3

This is a linear equation in terms of L_R (Verloop's sterimol length parameter of R substituents), $B5_X$ (Verloop's sterimol width parameter of X substituents) and I_{HAL} (an indicator variable, where $I_{HAL} = 1$ for X = F or Cl and Y = CF_2H otherwise $I_{HAL} = 0$). Positive coefficient associated with L_R and $B5_X$ suggests that the cytotoxic activity of these compounds against HT-29 cells increases with an increase in L_R and $B5_X$. The presence of either F or Cl at X position and CF_2H at Y position will also be favorable to the activity as evidenced by the positive coefficient of the indicator variable (I_{HAL}). Four compounds (R = COC_2H_5, X = F, Y = CF_2H; R = $COCH_3$, X = Cl, Y = CF_2H; R = $CON(CH_3)_2$, X = Cl, Y = CF_2H; R = $CON(CH_3)_2$, X = Cl, Y = CF_3) were deemed to be outliers because these analogues were either less or more active than expected, by 4.9, 9.3, 3.2, and 3.3 times the standard deviation. A comparison between observed and predicted values of log 1/IC_{50} for paclitaxel analogues (**6**) used in the development of QSAR Eq. (4) is shown in Fig. 4.

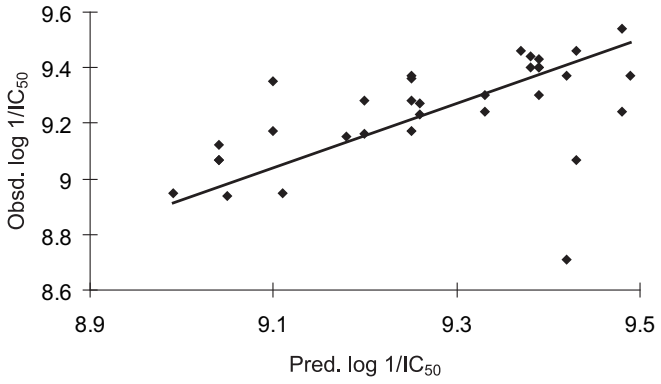


Fig. 4. Plot of observed versus predicted log $1/IC_{50}$ (Eq. (4)).

4. Validation of the QSAR models

QSAR model validation implies quantitative assessment of model robustness and their predictive power. The predictive power of a QSAR model can be defined as its ability to predict accurately the modeled property (e.g. biological activity) of new compounds. The details about the validation criteria for QSAR models have already been discussed previously [59–65]. The regression coefficients/statistics of QSAR models (Eqs. (1)–(4)) are shown in Table 5. The validation of QSAR models (Eqs. (1)–(4)) constitutes in the following steps: (a) Statistical Diagnostics, (b) Internal Validation, and (c) External Validation.

4.1. Statistical diagnostics

(i) *n/p Ratio*: $n/p \geq 4$ or $n \geq 4\ p$, where n is the number of data points and p is the number of descriptors used in the QSAR model. All the four QSAR models (Eqs. (1)–(4)) obey this thumb condition.

(ii) *Fraction of the Variance (r^2)*: The value of r^2 may vary between 0 and 1, where 1 means a perfect model explaining 100% of the variance in the data, and 0 means a model without any explanatory power. It has already been suggested that the only QSAR model having $r^2 > 0.6$ will be considered for validation [63–65]. The values of r^2 for these QSAR models (Eqs. (1)–(4)) are from 0.807 to 0.897.

(iii) *Cross-Validation Test (q^2)*: According to the literature, a QSAR model must have $q^2 > 0.5$ for their predictive ability [63–65]. The values of q^2 for these QSAR models (Eqs. (1)–(4)) are ranging from 0.724 to 0.793.

(iv) *Standard Deviation (s)*: The smaller s value is always required for the predictive QSAR model. The values of s for QSAR models (Eqs. (1)–(4)) are ranged from 0.076 to 0.360.

(v) $r^2 - q^2 < 0.3$: This difference between r^2 and q^2 for a QSAR model should never be exceeded by 0.3. A large difference between r^2 and q^2 suggests the following: (a) over-fitted model, (b) presence of outliers, or (c) presence of irrelevant variables in the data set [60]. The values of $r^2 - q^2$ for the QSAR models (Eqs. (1)–(4)) are from 0.065 to 0.104.

(vi) *Quality Factor (Q)*: Chance correlation and overfitting, due to the excess number of descriptors, can be detected by the Q value. High values of Q (2.569–11.816) for these QSAR models (Eqs. (1)–(4)) suggest their high predictive power and the lack of "overfitting" [66,67].

(vii) *Fischer Statistics (F)*: The F value of each QSAR model was compared to that of their respective literature value at 95% level [48]. The F-values of QSAR models (Eqs. (1)–(4)) ranging from 24.070 to 35.379 (where $F > F_{(lit)}$) suggest that all the QSAR models are statistically significant at the 95% level.

4.2. Internal validation

(i) *Cross-Validations Test*: Cross-Validations (CVs) are the most commonly used techniques for internal validation, in which compounds with different proportions are removed from the original data set and developed new QSAR models in order to verify the internal predictive ability of the original QSAR model, e.g. q^2 (leave-one-out), q_f^2 (leave-five-out), q_m^2 (leave-many-out), etc. Cross-validated r^2 (q^2 or q_f^2 or q_m^2) is calculated by the following Equation (5):

$$q^2\left(\text{or } q_f^2 \text{ or } q_m^2\right) = 1 - \frac{\sum\left(Y_{obs} - Y_{pred}\right)^2}{\sum\left(Y_{obs} - \hat{Y}\right)^2} \quad (5)$$

where Y_{obs}, Y_{pred}, and $\hat{Y}$ are the observed, predicted, and averaged activities, respectively. In the case of leave-one-out (LOO) cross-validation, each member of the original data set in turn is removed,

Table 5
Comparison of the regression coefficients and the statistics obtained from the multiregression analyses (MRA) process for QSAR models (Eqs. (1)–(4)).

QSAR no.	System	n	Regression coefficients				Statistical parameters						
			Hydrophobic	Steric/pol	Indicator variable	Intercept	r^2	q^2	$r^2 - q^2$	q_f^2	s	Q	F[a]
1	HCT-116 cells	22	1.34 π_X	$-1.97MR_X$	0.89 I_{CYALK}	9.05	0.855	0.786	0.069	0.772	0.360	2.569	35.379 (3.159)
2	HT-29 cells	19	$-4.42\pi_Z - 0.31 \text{Clog}P$	1.56 MR_Y		16.33	0.828	0.724	0.104	0.796	0.219	4.155	24.070 (3.287)
3	HT-29 cells	10		0.97 $B5_X$	$-0.75 I_{CMe3}$	7.69	0.897	0.793	0.104	0.727	0.277	3.419	30.481 (4.737)
4	HT-29 cells	29		0.08 L_R + 0.12 $B5_X$	0.38 I_{HAL}	8.51	0.807	0.742	0.065	0.812	0.076	11.816	34.845 (2.991)

[a] The figure within parenthesis referred to the literature F-value at 95% level.

Table 6
Y-randomization data for QSAR models Eqs. (1)–(4).

QSAR no.	NOR-1[a]		NOR-2		NOR-3		NOR-4		NOR-5	
	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2
1	0.390	0.014	0.377	0.137	0.390	0.158	0.383	0.054	0.393	0.157
2	0.584	0.256	0.585	0.270	0.467	0.212	0.482	0.275	0.573	0.371
3	0.255	−0.207	0.062	−0.527	0.426	−0.075	0.308	0.014	0.121	−0.789
4	0.270	0.080	0.295	0.124	0.318	0.158	0.384	0.245	0.446	0.292

[a] NOR = number of Y-randomization.

Table 7
Random selection pattern of the test set compounds as well as the regression coefficients and statistical parameters of the QSAR for their respective training set compounds obtained from the division of the original data of QSAR Models (Eqs. (1)–(4)).

QSAR *no.*[a]	Test set compd	Training set compd	Regression coefficients				Statistical parameters			
			Hydrophobic	Steric/pol	Indicator variable	Intercept	r^2	q^2	s	R^2_{pred}
1	2, 8, 10, 12, 14, 17	Rest of the compd ($n = 16$)	1.32 π_X	$-1.93MR_X$	0.98 I_{CYALK}	8.97	0.840	0.684	0.382	0.880
2	1, 3, 7, 13, 20	Rest of the compd ($n = 14$)	$-5.35\pi_Z - 0.28ClogP$	1.43 MR_Y		18.34	0.834	0.730	0.230	0.720
3	1, 2, 10	Rest of the compd ($n = 7$)		0.92 $B5_X$	$-0.82I_{CMe3}$	7.94	0.816	0.508	0.267	0.922
4	3, 5, 7, 12, 16, 21, 32	Rest of the compd ($n = 22$)		0.08 $L_R + 0.11$ $B5_X$	0.29 I_{HAL}	8.54	0.767	0.681	0.073	0.761

[a] QSAR for the respective training set of the original QSAR.

and the remaining $n - 1$ members are used in the development of new QSAR models. Similarly for leave-five-out (LFO) cross-validation, five members of the original data set in turn are removed, and the remaining $n - 5$ members are used in the development of new QSAR models. On the other hand, a certain data points are removed from the original data set in case of leave-many-out (LMO) cross-validation. A low value of q^2 (in the LOO, LFO and/or LMO test) typically indicates low predictive power of a QSAR model, but its high value does not necessarily suggest the high predictivity. Nevertheless, the cross-validated r^2 (q^2, q^2_f, and/or q^2_m) are frequently used as an important criterion for both robustness and predictive ability of the QSAR model. A high value of q^2, q^2_f, and/or q^2_m (q^2, q^2_f, and/or $q^2_m > 0.5$) is often considered as an ultimate proof for the high predictive power of QSAR model [63–65]. The values of q^2 and q^2_f for these four QSAR models (Eqs. (1)–(4)) are ranging from 0.724–0.793 to 0.727–0.812, respectively.

(ii) *Y-Randomization Test*: This is a widely used technique to establish the QSAR model robustness. In this technique, the dependent-variable vector (**Y** vector) is randomly shuffled, and a new QSAR model is developed using the unchanged independent-variable matrix. This process was repeated for five times at 95% confidence interval. The statistical data of r^2 and q^2 for five runs are listed in Table 6 (Eqs. (1)–(4)). The lower values of r^2 and q^2 in the Y-randomization test confirm the robustness of the QSAR models (Eqs. (1)–(4)) [64,68].

4.3. External validation

The external validation of QSAR models was carried out in two steps: (i) Random selection (Training set), and (ii) Predictive power of the QSAR models.

(i) *Random Selection (Training Set)*: There are several methods for selecting the training set. The simplest one is a random selection. In this method, the original (whole) data set is divided into training (~75%) and test (~25%) sets in a random manner [19]. The QSAR model for the resulting training set is then generated by using the same descriptors as those of the original equation and validated on the basis of their statistics (acceptance criteria: $r^2 > 0.6$ and $q^2 > 0.5$).

(ii) *Predictive power of QSAR models*: The true predictive power of a QSAR model is determined by comparing the predicted and observed activities of the test set compounds that are not used in the QSAR model development of training set. The predictive power of a QSAR model can be estimated by their predictive R^2 (R^2_{pred}), which is calculated by the following Eq. (6):

$$R^2_{pred} = 1 - \frac{\sum\left(Y_{pred(test)} - Y_{test}\right)^2}{\sum\left(Y_{test} - \overline{Y}_{training}\right)^2} \quad (6)$$

where $Y_{pred(test)}$ and Y_{test} are the respective predicted and observed activities of the test set compounds and $\overline{Y}_{training}$ is the observed mean activity of the training set compounds [19,37,64].

A random selection pattern of the test sets as well as the regression coefficients and the statistical parameters of their respective training sets for all the QSAR models (Eqs. (1)–(4)) are given in Table 7.

5. Conclusions

An analysis of QSAR models (Eqs. (1)–(4)) reveals a number of interesting points. The most important of these are π and MR descriptors. QSAR model Eq. (1) suggests that the cytotoxicity of taxane derivatives (**3**) against HCT-116 cells might be improved by the presence of a more hydrophobic/less polarizable substituent at C-4 position. On the other hand a bulkier group at C-2, C-10, C-3′, and/or C-3′-N, and a more hydrophilic substituent at C-3′-N positions might improve the cytotoxicity of taxane analogues (**4–6**) against HT-29 cells as evidenced by the QSAR models (Eqs. (2)–(4)). In short, the inhibitory activities of taxane analogues against colon cancers are mainly dependent on the steric (MR, *B*5, and *L*) and hydrophobic (π) descriptors of their substituents, with a major contribution coming from the molar refractivity (MR) of substituents. Further development of QSAR studies on taxane analogues should not only enlarge the areas of their application, but it may also increase our understanding towards the mechanisms of chemical–biological interactions.

References

[1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M.J. Thun, CA Cancer J. Clin. 58 (2008) 71–96.
[2] C.M. Caldwell, R.A. Green, K.B. Kaplan, J. Cell. Biol. 178 (2007) 1109–1120.
[3] S. Segditsas, I. Tomlinson, Oncogene 25 (2006) 7531–7537.
[4] D. Dikovskaya, D. Schiffmann, I.P. Newton, A. Oakley, K. Kroboth, T. Jamieson, V. Meniel, A. Clarke, O. Sansom, I.S. Näthke, J. Cell Biol. 176 (2007) 183–195.
[5] I. Naethke, Nat. Rev. Cancer 6 (2006) 967–974.
[6] C.J. Ceol, D. Pellman, L.I. Zon, Nat. Med. 13 (2007) 1286–1287.
[7] K.W. Kinzler, B. Vogelstein, Cell 87 (1996) 159–170.
[8] M. Linsalata, F. Russo, Nutrition 24 (2008) 382–389.
[9] K. Yang, N. Kurihara, K. Fan, H. Newmark, B. Rigas, L. Bancroft, G. Corner, E. Livote, M. Lesser, W. Edelmann, A. Velcich, M. Lipkin, L. Augenlicht, Cancer Res. 68 (2008) 7803–7810.
[10] E.S. Schernhammer, S. Ogino, C.S. Fuchs, Gastroenterology 135 (2008) 770–780.
[11] E. Rudolf, V. Kralova, M. Cervinka, Anticancer Agents Med. Chem. 8 (2008) 598–602.
[12] H. Zhang, M.R. Spitz, G.E. Tomlinson, M.B. Schabath, J.D. Minna, X. Wu, Cancer Detect. Prev. 26 (2002) 411–418.
[13] D. Sadava, E. Whitlock, S.E. Kane, Biochem. Biophys. Res. Commun. 360 (2007) 233–237.
[14] N. Kumar, D. Shibata, J. Helm, D. Coppola, M. Malafa, Front. Biosci. 12 (2007) 2309–2315.
[15] J.T. Hwang, J. Ha, I.J. Park, S.K. Lee, H.W. Baik, Y.M. Kim, O.J. Park, Cancer Lett. 247 (2007) 115–121.
[16] J. Wils, J. Chemother. 19 (2007) 115–122.
[17] M.S. Mano, F. Duhoux, Clin. Colorectal Cancer 7 (2008) 178–183.
[18] S.G. Arbuck, B.A. Blaylock, Taxol: clinical results and current issues in development. in: M. Suffness (Ed.), Taxol: Science and Applications. CRC Press, Boca Raton, 1995, pp. 379–415.

[19] R.P. Verma, C. Hansch, ChemMedChem 3 (2008) 642–652.
[20] E.J. Roh, D. Kim, C.O. Lee, S.U. Choi, C.E. Song, Bioorg. Med. Chem. 10 (2002) 3145–3151.
[21] I. Ojima, J.C. Slater, S.D. Kuduk, C.S. Takeuchi, R.H. Gimi, C.M. Sun, Y.H. Park, P. Pera, J.M. Veith, R.J. Bernacki, J. Med. Chem. 40 (1997) 267–278.
[22] P.B. Schiff, J. Fant, S.B. Horwitz, Nature 277 (1979) 665–667.
[23] M.V. Blagosklonny, T. Fojo, Int. J. Cancer 83 (1999) 151–156.
[24] T.-H. Wang, H.-S. Wang, Y.-K. Song, Cancer 88 (2000) 2619–2628.
[25] C. Hansch, R.P. Verma, Mol. Pharmaceutics 5 (2008) 151–161.
[26] C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Nature 194 (1962) 178–180.
[27] J.M. Bermudez-Saldana, M.T.D. Cronin, Pest Manag. Sci. 62 (2006) 819–831.
[28] T.W. Schultz, M. Hewitt, T.I. Netzeva, M.T.D. Cronin, QSAR Comb. Sci. 26 (2007) 238–254.
[29] C. Hansch, A. Leo, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. American Chemical Society, Washington D.C., 1995.
[30] C.D. Selassie, R. Garg, S. Kapur, A. Kurup, R.P. Verma, S.B. Mekapati, C. Hansch, Chem. Rev. 102 (2002) 2585–2605.
[31] R.P. Verma, C. Hansch, Chem. Rev. 109 (2009) 213–235.
[32] R.P. Verma, C. Hansch, Bioorg. Med. Chem. 15 (2007) 2223–2268.
[33] M. Fujikawa, K. Nakao, R. Shimizu, M. Akamatsu, Bioorg. Med. Chem. 15 (2007) 3756–3767.
[34] R.P. Verma, C. Hansch, C.D. Selassie, J. Comput. Aided Mol. Des. 21 (2007) 3–22.
[35] R.P. Verma, Top. Heterocycl. Chem. 9 (2007) 53–86.
[36] R. Garg, S.P. Gupta, H. Gao, M.S. Babu, A.K. Debnath, C. Hansch, Chem. Rev. 99 (1999) 3525–3601.
[37] R.P. Verma, C. Hansch, Mol. Pharmaceutics 5 (2008) 745–759.
[38] R.P. Verma, A. Kurup, S.B. Mekapati, C. Hansch, Bioorg. Med. Chem. 13 (2005) 933–948.
[39] L.H. Hall, L.M. Hall, SAR QSAR Environ. Res. 16 (2005) 13–41.
[40] C. Hansch, A. Leo, S.B. Mekapati, A. Kurup, Bioorg. Med. Chem. 12 (2004) 3391–3400.
[41] M.T.H. Khan, I. Sylte, Curr. Drug Discov. Tech. 4 (2007) 141–149.
[42] C.D. Selassie, S.B. Mekapati, R.P. Verma, Curr. Top. Med. Chem. 2 (2002) 1357–1379.
[43] C-QSAR Program, BioByte Corp., 201W, 4th St., Suit 204, Claremont, CA 91711, USA.www.biobyte.com.
[44] C. Hansch, D. Hoekman, A. Leo, D. Weininger, C.D. Selassie, Chem. Rev. 102 (2002) 783–812.
[45] R.P. Verma, C. Hansch, Nat. Protoc., http://www.natureprotocols.com/2007/03/05/development_of_qsar_models_usi_1.php. doi:10.1038/nprot.2007.125.
[46] A. Verloop, The Sterimol Approach to Drug Design. Marcel Dekker, New York, 1987.
[47] R.D. Cramer III, J.D. Bunce, D.E. Patterson, I.E. Frank, Quant. Struct. Act. Relat. 7 (1988) 18–25.
[48] C. A Bennett, N.L. Franklin, Statistical Analysis in Chemistry and the Chemical Industry. John Wiley & Sons, New York, 1967, pp. 708–709 (5th printing).
[49] R.P. Verma, C. Hansch, Bioorg. Med. Chem. 13 (2005) 4597–4621.
[50] J. Polanski, A. Bak, R. Gieleciak, T. Magdziarz, J. Chem. Inf. Model 46 (2006) 2310–2318.
[51] S. Ekins, J. Mestres, B. Testa, Br. J. Pharmacol. 152 (2007) 9–20.
[52] C.D. Selassie, S. Kapur, R.P. Verma, M. Rosario, J. Med. Chem. 48 (2005) 7234–7242.
[53] R.P. Verma, C. Hansch, Mol. Pharm 3 (2006) 441–450.
[54] C. Hansch, R.P. Verma, ChemMedChem 2 (2007) 1807–1813.
[55] R.P. Verma, C. Hansch, J. Pharm. Sci. 97 (2008) 88–110.
[56] S.-H. Chen, J.-M. Wei, B.H. Long, C.A. Fairchild, J. Carboni, S.W. Mamber, W.C. Rose, K. Johnston, A.M. Casazza, J.F. Kadow, V. Farina, D.M. Vyas, T.W. Doyle, Bioorg. Med. Chem. Lett. 5 (1995) 2741–2746.
[57] C.J. Maring, D.J. Grampovnik, C.M. Yeung, L.L. Klein, L. Li, S.A. Thomas, J.J. Plattner, Bioorg. Med. Chem. Lett. 4 (1994) 1429–1432.
[58] L.V. Kuznetsova, A. Pepe, I.M. Ungureanu, P. Pera, R.J. Bernacki, I. Ojima, J. Fluor. Chem. 129 (2008) 817–828.
[59] J.S. Jaworska, M. Comber, C. Auer, C.J. van Leeuwen, Environ. Health Perspect. 111 (2003) 1358–1360.
[60] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Environ. Health Prospect. 111 (2003) 1361–1375.
[61] M.T.D. Cronin, J.D. Walker, J.S. Jaworska, M.H.I. Comber, C.D. Watts, A.P. Worth, Environ. Health Perspect. 111 (2003) 1376–1390.
[62] M.T.D. Cronin, J.S. Jaworska, J.D. Walker, M.H.I. Comber, C.D. Watts, A.P. Worth, Environ. Health Perspect. 111 (2003) 1391–1401.
[63] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model 20 (2002) 269–276.
[64] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb. Sci. 22 (2003) 69–77.
[65] S. Zhang, L. Wei, K. Bastow, W. Zheng, A. Brossi, K.-H. Lee, A. Tropsha, J. Comput. Aided Mol. Des. 21 (2007) 97–112.
[66] L. Pogliani, Chem. Rev. 100 (2000) 3827–3858.
[67] L. Pogliani, J. Phys. Chem. 100 (1996) 18065–18077.
[68] S. Wold, L. Eriksson, Statistical validation of QSAR results, validation tools. in: H. van de Waterbeemd (Ed.), Chemometrics Methods in Molecular Design. Wiley-VCH, Weinheim, 1995, pp. 309–318.