CHEMMEDCHEM

DOI: 10.1002/cmdc.200700278

Taxane Analogues against Breast Cancer: A Quantitative Structure–Activity Relationship Study

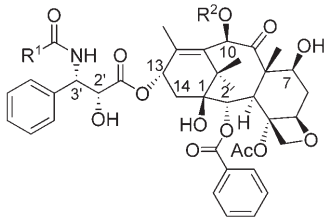
Rajeshwar P. Verma* and Corwin Hansch[a]

Breast cancer is the second leading cause of cancer death among women in the United States. Two taxane analogues, taxol and taxotere, are the most important antimitotic drugs currently in clinical use for the treatment of breast cancers. However, recent reports have indicated that the use of these drugs often results in various undesired side effects as well as multi-drug resistance. These limitations have led to the development of new taxane derivatives with fewer side effects, superior pharmacological properties, and improved anticancer activity to maximize the induced benefits for breast cancer patients. Herein, four series of taxane derivatives were used to correlate their inhibitory activities against breast cancer cells with their hydrophobic and steric properties in order to understand their chemical–biological interactions. The resulting QSARs show that the inhibitory activities of taxane analogues against breast cancers are mainly dependent either on their hydrophobicity or the hydrophobic/molar refractivity descriptor of their substituents. A parabolic correlation with MR_Y is the most encouraging example, in which the optimum value of this parameter is well defined. We believe this correlation may prove to be an adequate predictive model that can help provide guidance in design and synthesis and subsequently yield highly specific compounds that may have high anti-breast-cancer activity with fewer side effects and superior pharmacological properties. On the basis of this QSAR model, five compounds are suggested as potential synthetic targets. Internal (cross-validation ($LOO-q^2$ and $LMO-q^2$), quality factor (Q), Fischer statistics (F), and Y-randomization) and external validation tests have validated all the QSAR models.

Introduction

The term breast cancer refers to a malignant tumor that has developed from cells in the breast. Breast cancer is the most common cancer and the second leading cause of cancer death among women in the United States; annual breast cancer deaths are exceeded only by those from lung cancer. A woman has a lifetime risk of developing breast cancer of about one in eight. In 2007, the American Cancer Society estimated that about 178480 new cases of invasive breast cancer and about 62030 new cases of in situ (noninvasive) breast cancer would be diagnosed among women in the United States, and that about 40460 women will die from this disease. Breast cancer is primarily a disease of women, but about 1% of this disease also occurs in men. The American Cancer Society estimated that during 2007 about 2030 men in the US will develop invasive breast cancer, and that about 450 men will die from the disease this year. There are mainly two common types of breast cancer: 1) Ductal carcinomas start in the tubes (ducts) that move milk from the breast to the nipple. Most breast cancers are of this type that account for about 80% of invasive breast cancer. 2) Lobular carcinoma starts in parts of the breast called lobules, which produce milk. It accounts for about 10% of invasive breast cancer. In rare cases, breast cancer can start in other areas of the breast. Many breast cancers are sensitive to the hormone estrogen, suggesting that estrogen causes the breast cancer tumor to grow. This type of breast cancer is known as estrogen-receptor-positive or ER-positive breast cancer.[1,2]

Among novel chemotherapeutic agents, the taxanes have emerged as the most powerful group of compounds. Taxanes such as taxol (paclitaxel; **1**) and taxotere (docetaxel; **2**) are two important antimitotic drugs currently in clinical use for the treatment of various types of cancers, including breast, lung, ovarian, and prostate cancer. These drugs act as microtubule stabilizers and disrupt microtubule dynamics, thus inducing mitotic arrest and ultimately, cell death by apoptosis. Paclitaxel and docetaxel were approved by the FDA for the treatment of breast cancer in April 1994 and May 1996, respectively.[3–8]



R^1 = Ph, R^2 = Ac, taxol (paclitaxel); **1**
R^1 = *t*BuO, R^2 = H, taxotere (docetaxel); **2**

Taxanes **1** and **2** share some characteristics and also have a number of significant differences, both in terms of preclinical and pharmacokinetic profiles and, most importantly, clinical consequences. The taxanes are now standard therapy in the

[a] *Dr. R. P. Verma, Dr. C. Hansch*
Department of Chemistry, Pomona College
645 North College Avenue, Claremont, CA 91711 (USA)
Fax: (+1) 909-607-7726
E-mail: rverma@pomona.edu

clinical treatment of metastatic breast cancer. Their role as monochemotherapy or in combination with anthracyclines in advanced breast cancer has suggested their potential therapeutic impact in the treatment of patients with early breast cancer. Recent results demonstrate that taxanes, used either in combination with other chemotherapeutic agents or in sequential therapy, can yield significant improvements, particularly in terms of patient survival, confirming the positive impact of taxanes on the natural history of breast cancer.[9] Although drugs **1** and **2** have made significant progress in the treatment of breast cancer, recent reports have indicated that their use often results in various undesired side effects as well as multi-drug resistance (MDR).[10,11] Therefore, it is important to develop new taxanes with fewer side effects, superior pharmacological properties, and improved anticancer activity to maximize the induced benefits for breast cancer patients. The quantitative structure–activity relationship (QSAR) paradigm may be helpful in the design and development of novel taxane molecules as new anticancer agents, which are expected to show improvements in activity against breast cancer and in their toxicity profiles, pharmacology, and drug formulation.

Herein we demonstrate QSAR studies on various sets of taxane analogues with respect to their activities against breast cancer to understand their chemical–biological interactions. The QSAR approach employs extra-thermodynamically derived and computational-based descriptors to correlate biological activity in isolated receptors, cellular systems, and in vivo. Four standard molecular descriptors are routinely used in the development of QSAR: electronic, hydrophobic, steric, and topological indices. These descriptors are invaluable in helping to delineate a large number of receptor–ligand interactions that are critical in biological processes. The quality of a QSAR model, however, depends strictly on the type and quality of the data, and is valid only for the compound structure analogues used to build the model. QSAR models can stand alone to augment other computational approaches or can be examined in tandem with equations of a similar mechanistic genre to establish their authenticity and reliability.[12]

Since the advent of QSAR methodology (one of the well-developed areas in computational chemistry) about 45 years ago,[13] it has become increasingly helpful in understanding many aspects of chemical–biological interactions in drug and pesticide research as well as in the field of toxicology. QSAR is useful in elucidating the mechanisms of chemical–biological interactions in various biomolecules, particularly enzymes, as well as membranes, organelles, cells, and in humans.[14–16] It has also been used for the evaluation of absorption, distribution, metabolism, and excretion (ADME) phenomena in many organisms and whole-animal studies.[17,18] The potential use of QSAR models in screening chemical databases or virtual libraries before their synthesis appears equally attractive to chemical manufacturers, pharmaceutical companies, and government agencies.

Results and Discussion

Inhibition of MCF-7 breast cancer cell growth by taxane derivatives 3

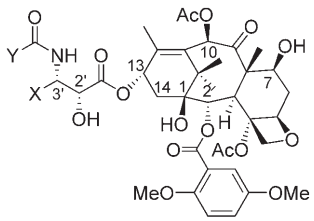
Equation (1)is based on the data obtained from Baloglu et al.[19] (see Table 1):

$$\log 1/IC_{50} = 5.19(\pm 3.25)MR_Y - 1.14(\pm 0.79)MR_Y{}^2 + 4.52(\pm 3.32)$$
$$n = 20,\ r^2 = 0.622,\ s = 0.150,\ q^2 = 0.514,\ q_m{}^2 = 0.557,$$
$$Q = 5.254,\ F_{2,17} = 13.987 \qquad (1)$$
$$\text{optimum } MR_Y = 2.29\ (2.16 - 2.85)$$
$$\text{outliers}: X = CH{=}C(CH_3)_2,\ Y = CH{=}CHCH_3;$$
$$X = \text{2-furyl},\ Y = C_6H_5$$

IC_{50} is the molar concentration of taxane analogues **3** that blocks 50% of the growth of MCF-7 cancer cell population after 72 h of exposure. The above QSAR model is a parabolic correlation in terms of MR_Y (molar refractivity of Y substituents), which suggests that the cytotoxic activities of taxane derivatives **3** first increase with an increase in molar refractivity of Y substituents up to an optimum MR_Y value of 2.29 and then decreases. Two compounds ($X = CH{=}C(CH_3)_2$, $Y = CH{=}CHCH_3$ and X = 2-furyl, $Y = C_6H_5$) were deemed to be outliers on the basis of their deviations (obsd − pred $> 2 \times s$). One derivative (X = 2-furyl, $Y = C_6H_5$) was considered to be an outlier because it was less active than expected, by 4.7 times the standard deviation. Possible reasons for its unusually low activity are not clear, although its steric bulk or geometry due to the presence of a phenyl group at the Y position may decrease coplanarity with the X group (2-furyl) and minimize activity.[20] Placing the same aromatic (phenyl) group at both the X and Y positions increases the activity of the compound (**3**-21; $X = Y = C_6H_5$), and this is well predicted. To assess the effects of excluding outliers, QSAR models were examined before and after the removal of compound. The statistically significant QSAR models were not obtained by either considering all the compounds ($n = 22$, $r^2 = 0.097$, $q^2 = -0.542$) or with one outlier [either **3**-4 ($n = 21$, $r^2 = 0.460$, $q^2 = 0.189$) or **3**-12 ($n = 21$, $r^2 = 0.122$, $q^2 = -0.505$)]. Thus the consideration of two outliers, **3**-4 and **3**-12, in the development of QSAR Equation (1) is justified. There is a high correlation between $\log P$ and MR_Y ($r = 0.671$). Therefore, $\log P$ can replace MR_Y and by doing so in Equation (1), we can develop Equation (2) (see Table 1).

$$\log 1/IC_{50} = 2.51(\pm 1.54)\log P - 0.28(\pm 0.19)\log P^2 + 4.80(\pm 3.06)$$
$$n = 19,\ r^2 = 0.671,\ s = 0.144,\ q^2 = 0.501,\ q_m{}^2 = 0.548,$$
$$Q = 5.688,\ F_{2,16} = 16.316 \qquad (2)$$
$$\text{optimum } \log P = 4.50\ (4.24 - 5.59)$$
$$\text{outliers}: X = CH{=}C(CH_3)_2,\ Y = CH{=}CHCH_3;\ X = \text{2-furyl},$$
$$Y = CH{=}C(CH_3)_2;$$
$$X = \text{2-furyl}, Y = C_6H_5$$

Table 1. Biological and physicochemical parameters used to derive QSAR Equations (1) and (2) as well as the calculated ΔG values.



3

Compd	X	Y	log 1/IC_{50}					MR_Y	log *P*	ΔG
			Obsd	Eq. (1)		Eq. (2)				
				Pred	Δ	Pred	Δ			
3-1	CH=C$(CH_3)_2$	OC$(CH_3)_3$	10.37	10.41	−0.04	10.43	−0.06	2.10	4.77	−6.51
3-2	CH=C$(CH_3)_2$	CH=C$(CH_3)_2$	10.70	10.38	0.32	10.42	0.28	2.03	4.16	−5.67
3-3	CH=C$(CH_3)_2$	2-furyl	10.42	10.17	0.25	10.29	0.13	1.79	3.73	−5.09
3-4[a,b]	CH=C$(CH_3)_2$	CH=CHCH_3	10.92	9.86	1.06	10.30	0.62	1.56	3.76	−5.13
3-5	CH=C$(CH_3)_2$	C_6H_5	10.39	10.38	0.01	10.45	−0.06	2.54	4.56	−6.22
3-6	CH=C$(CH_3)_2$	O$(CH_2)_3CH_3$	10.49	10.44	0.05	10.35	0.14	2.17	5.12	−6.98
3-7	CH=C$(CH_3)_2$	2-thiophenyl	10.49	10.44	0.05	10.45	0.04	2.40	4.40	−6.00
3-8[b]	2-furyl	CH=C$(CH_3)_2$	10.39	10.38	0.01	10.08	0.31	2.03	3.35	−4.57
3-9	2-furyl	OC$(CH_3)_3$	10.38	10.41	−0.03	10.37	0.01	2.10	3.96	−5.40
3-10	2-furyl	2-furyl	9.92	10.17	−0.25	9.88	0.04	1.79	3.07	−4.19
3-11	2-furyl	CH=CHCH_3	9.88	9.86	0.02	9.78	0.10	1.56	2.95	−4.02
3-12[a,b]	2-furyl	C_6H_5	9.68	10.38	−0.70	10.35	−0.67	2.54	3.90	−5.31
3-13	2-furyl	O$(CH_2)_3CH_3$	10.51	10.44	0.07	10.44	0.07	2.17	4.31	−5.88
3-14	2-furyl	2-thiophenyl	10.21	10.44	−0.23	10.29	−0.08	2.40	3.73	−5.09
3-15	2-thiophenyl	CH=C$(CH_3)_2$	10.30	10.38	−0.08	10.32	−0.02	2.03	3.82	−5.21
3-16	2-thiophenyl	OC$(CH_3)_3$	10.51	10.41	0.10	10.45	0.06	2.10	4.43	−6.04
3-17	2-thiophenyl	2-furyl	10.19	10.17	0.02	10.20	−0.01	1.79	3.54	−4.83
3-18	2-thiophenyl	CH=CHCH_3	9.79	9.86	−0.07	10.13	−0.34	1.56	3.42	−4.66
3-19	2-thiophenyl	O$(CH_2)_3CH_3$	10.32	10.44	−0.12	10.43	−0.11	2.17	4.78	−6.52
3-20	2-thiophenyl	2-thiophenyl	10.44	10.44	0.00	10.43	0.01	2.40	4.20	−5.73
3-21	C_6H_5	C_6H_5	10.49	10.38	0.11	10.44	0.05	2.54	4.72	−6.44
3-22	C_6H_5	OC$(CH_3)_3$	10.20	10.41	−0.21	10.43	−0.23	2.10	4.78	−6.53

[a] Not used in the derivation of QSAR Equation (1). [b] Not used in the derivation of QSAR Equation (2).

This is a parabolic correlation in terms of log *P* (calculated hydrophobicity of the whole molecule), which suggests that the cytotoxic activities of taxane derivatives **3** first increase with an increase in hydrophobicity of the whole molecule up to an optimum log *P* value of 4.50 and then decreases. The statistics of this [Eq. (2)] is better than that of Equation (1), but it has one more outlier, namely **3**-8. We kept this equation because it speaks very well about the compound's ability to cross the cancer cell membrane.

It is well known that the taxane analogues bind to the β subunits of tubulin polymers in a stoichiometric ratio and promote tubulin polymerization. This phenomenon disrupts tubulin polymerization dynamics, leading to cell-cycle arrest and ultimately cell death by apoptosis. Microtubules are structures composed of polymerized tubulin heterodimers, which play fundamental roles in vital cell processes such as chromosome segregation and intracellular transport. Thus, it will be interesting to formulate a QSAR model for the interaction of these taxane analogues with the tubulin/microtubule system and compare it with the present QSAR models [Eq. (1) and Eq. (2)] to understand their chemical–biological interactions. Unfortunately, we do not have the binding affinity data for these taxane analogues **3** with the tubulin/microtubule system.

The binding of organic compounds by proteins has long been recognized as an important concern for biochemists and pharmacologists. The protein surface is covered by water molecules with varying degrees in affinity. When a ligand binds to a site on a given protein surface, it must displace one or more of these water molecules. Therefore, the nature of water molecule interactions with the protein is important in the ligand binding process.[21] This is supported by the fact that the binding of organic compounds to bovine serum albumin (BSA) is well correlated with their hydrophobicity; this is quite similar to the case with other proteins such as hemoglobin and ribonuclease. These findings are important for enzyme chemistry and pharmacology because they show that a linear free-energy relationship exists between the binding of organic compounds or drugs to proteins and their hydrophobicity.[22] The hydrophobic parameter log *P* is a free-energy-related term from the van 't Hoff isotherm:

$$\Delta G = -2.303\,RT \times \log P = -1.364 \times \log P\ (\text{kcal mol}^{-1}) \qquad (3)$$

and thus represents the free-energy change during the transfer of the solute from water to the nonaqueous (*n*-octanol) phase[23] (*R* is the gas constant, *T* is the absolute temperature,

and ΔG is the Gibbs free energy). The $\log P$ values of taxane derivatives **3** have been converted into ΔG values by the application of Equation (3) (see Table 1). It is interesting to note here that the ΔG value of the parent compound, taxol, is calculated as $-6.45\ \text{kcal mol}^{-1}$ ($\Delta G = -1.364 \times \log P = -1.364 \times 4.73 = -6.45$), which is in very good agreement with the published value ($\Delta G_{exp} = -6.46\ \text{kcal mol}^{-1}$). In contrast, the same is not true for taxotere ($\Delta\Delta G = \Delta G_{exp} - \Delta G = -6.90 + 5.57 = -1.33$). In addition to the above, a very poor correlation for a set of nine taxanes was obtained between their ΔG_{exp} and ΔG values ($r = 0.300$), in which ΔG_{exp} is the published value of ΔG obtained from K_{exp} (experimental binding energy in M).[24] On the basis of the above reasons, the calculated values of ΔG from Equation (3) were not considered in the development of a QSAR model. On the other hand, the correlation between $\log P$ and ΔG ($r = 1.00$) does not assist in the development of another QSAR between $\log 1/IC_{50}$ and ΔG, because the correlation between $\log 1/IC_{50}$ and $\log P$ is already in hand [Eq. (2)]. Thus, the change in free energy of ligand binding, which includes protein–ligand interactions (when moving from an aqueous solution to a protein binding pocket that is frequently quite hydrophobic in character) cannot be calculated from the hydrophobic parameter of the ligands ($\log P$) alone. This may suggest that the binding affinity of the ligands does not depend only on the interaction with the target protein, but also on (de)solvation and entropic effects as well as induced fit: the adaptation of the protein conformation to the ligand topology.

Over the past three decades, a major trend in the evolution of QSAR has been the development of 3D QSAR. The main reason for this trend is due to an in-depth understanding of protein–ligand interactions at the atomic level supported by a wealth of experimental evidence. In 3D QSAR, the structures of the molecules are represented by three-dimensional entities, which allow the quantitation of electrostatic forces, hydrogen bonds, and hydrophobic interactions at the atomic level. Models based on 3D QSAR typically represent a binding site surrogate with physicochemical properties mapped onto their surface or a grid surrounding the ligand molecules, superimposed in 3D space. Unfortunately, the ability of 3D QSAR methodology such as CoMFA to describe the entropic contribution to the free energy of binding has not yet been well sorted out. Nevertheless, new insight could be obtained considering the waters surrounding macromolecules along with the traditional QSAR results.[21] The mini-receptor modeling approach has been the most recent advance in the field of 3D QSAR studies of taxanes and the corresponding protein environment. This type of modeling approach is useful for identifying the pharmacophore features of the ligands and the respective counterparts on the tubulin binding site that mainly interact by hydrogen bonds and hydrophobic contacts.[24]

Inhibition of MCF-7 breast cancer cell growth by taxane derivatives 4

The biological data of taxane derivatives **4** was used by Islam et al.[25] for CoMFA and CoMSIA analysis to define the important interacting regions at the paclitaxel–tubulin binding site. The CoMFA analysis showed that the steric and electrostatic fields are more prominent than the lipophilic field in contributing toward binding and biological activity. In contrast, CoMSIA analysis highlighted the steric, electrostatic, hydrophobic, and H-bonding properties of taxane analogues **4**.

Based on the data of Islam et al.[25] (see Table 2), the following QSAR [Eq. (4)] was developed:

$$\begin{aligned}
&\log 1/IC_{50} = 3.58(\pm 2.52)\log P - 0.33(\pm 0.25)\log P^2 \\
&+0.37(\pm 0.22)I_1 - 0.79(\pm 0.27)I_2 - 6.30(\pm 0.21) \\
&n = 16,\ r^2 = 0.907,\ s = 0.192,\ q^2 = 0.740, \\
&q_m{}^2 = 0.752,\ Q = 4.958,\ F_{4,11} = 26.819 \qquad (4)\\
&\text{optimum } \log P = 5.42\ (5.11 - 6.78) \\
&\text{outliers: X = Ac, Y = } t\text{BuO, Z = 4-fluorophenyl;} \\
&\text{X = H, Y = } t\text{BuO, Z = } CH_2CH(CH_3)_2
\end{aligned}$$

This is a parabolic correlation in terms of $\log P$, which suggests that the cytotoxic activities of taxane derivatives **4** first increase with an increase in the hydrophobicity of the whole molecule up to an optimum $\log P$ value of 5.42 and then decreases. I_1 and I_2 are indicator variables that pinpoint the unusual activities of Z = CH=C(CH$_3$)$_2$ and X = Ac, respectively. The positive coefficient of I_1 suggests that the presence of Z = CH=C(CH$_3$)$_2$ will promote cytotoxic activity. On the other hand, the presence of X = Ac will be detrimental to the activity, as evidenced by the negative coefficient of the indicator variable I_2. Two compounds (X = Ac, Y = *t*BuO, Z = 4-fluorophenyl and X = H, Y = *t*BuO, Z = $CH_2CH(CH_3)_2$) were deemed to be outliers on the basis of their deviations (obsd − pred $> 2 \times s$).

Because both series of taxane derivatives **3** and **4** represent inhibitors to the growth of MCF-7 breast cancer cells, we combined these data and developed Equation (5), which may apply in the evaluation of the activities of the predicted molecules from Equation (1).

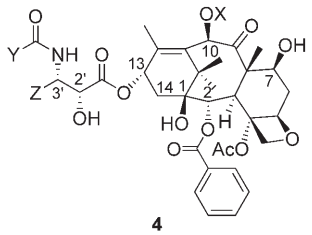
Inhibition of MCF-7 breast cancer cell growth by taxane derivatives 3 and 4

Equation (5) is based on the data obtained from Baloglu et al.[19] and Islam et al.[25] (see Table 3):

$$\begin{aligned}
&\log 1/IC_{50} = 0.52(\pm 0.14)\text{CMR} - 3.11(\pm 1.32) \\
&n = 36,\ r^2 = 0.613,\ s = 0.334,\ q^2 = 0.558,\ q_m{}^2 = 0.593, \\
&Q = 2.344,\ F_{1,34} = 53.855 \qquad (5)\\
&\text{outliers: } \mathbf{3}\text{-4, } \mathbf{4}\text{-1, } \mathbf{4}\text{-9, and } \mathbf{4}\text{-12}
\end{aligned}$$

This QSAR [Eq. (5)] reveals that steric features influence the inhibitory activity in a linear model. Positive CMR (calculated molar refractivity of the whole molecules) suggests that the increase in the molar refractivity of the whole molecule increases the inhibitory activities of the combined sets of taxane derivatives **3** and **4** against MCF-7 breast cancer cells. Four compounds (**3**-4, **4**-1, **4**-9, and **4**-12) were not used in the deriva-

Table 2. Biological, physicochemical, and structural parameters used to derive QSAR Equation (4).



4

Compd	X	Y	Z	log 1/IC_{50} [Eq. (4)]			log *P*	I_1	I_2
				Obsd	Pred	Δ			
4-1	Ac	C_6H_5	C_6H_5	8.77	8.55	0.22	4.73	0	1
4-2	H	*t*BuO	C_6H_5	9.00	8.91	0.09	4.08	0	0
4-3[a]	Ac	*t*BuO	4-fluorophenyl	9.38	8.63	0.75	4.94	0	1
4-4	COC_2H_5	*t*BuO	$CH_2CH(CH_3)_2$	9.46	9.49	−0.03	5.59	0	0
4-5[a]	H	*t*BuO	$CH_2CH(CH_3)_2$	8.40	9.12	−0.72	4.35	0	0
4-6	H	*t*BuO	$CH{=}C(CH_3)_2$	9.26	9.27	−0.01	4.07	1	0
4-7	COC_2H_5	*t*BuO	$CH{=}C(CH_3)_2$	9.74	9.86	−0.12	5.31	1	0
4-8	CO(*c*-C_3H_5)	*t*BuO	$CH{=}C(CH_3)_2$	9.70	9.87	−0.17	5.37	1	0
4-9	Ac	*t*BuO	$CH_2CH_2CF_3$	8.03	8.20	−0.17	4.18	0	1
4-10	$COOCH_3$	*t*BuO	$CH_2CH(CH_3)_2$	9.48	9.40	0.08	4.87	0	0
4-11	CH_3	*t*BuO	$CH_2CH(CH_3)_2$	9.55	9.43	0.12	4.95	0	0
4-12	Ac	C_6H_5	4-fluorophenyl	8.29	8.61	−0.32	4.87	0	1
4-13	Ac	*t*BuO	$CH{=}C(CH_3)_2$	9.22	8.95	0.27	4.78	1	1
4-14	$CON(CH_3)_2$	*t*BuO	$CH{=}C(CH_3)_2$	9.89	9.85	0.04	5.20	1	0
4-15	$CONHC_2H_5$	*t*BuO	$CH_2CH(CH_3)_2$	9.51	9.50	0.01	5.36	0	0
4-16	$CONHCH_3$	*t*BuO	$CH{=}C(CH_3)_2$	9.40	9.62	−0.22	4.55	1	0
4-17	$CON(C_2H_5)_2$	*t*BuO	$CH{=}C(CH_3)_2$	9.70	9.64	0.06	6.26	1	0
4-18	$COOCH_3$	*t*BuO	$CH{=}C(CH_3)_2$	9.85	9.71	0.14	4.73	1	0

[a] Not used in the derivation of QSAR Equation (4).

tion of Equation (5) due to their significant deviation from the observed activity (obsd−pred $> 2\times s$). There is not a good correlation between log*P* and CMR ($r = 0.022$). A comparison between observed and predicted values of log 1/IC_{50} for taxane analogues **3** and **4** used in the development of QSAR Equation (5) is shown in Figure 1.

Inhibition of growth of MCF-7-R (doxorubicin-resistant human breast cancer cells) by taxane analogues 5

Equation (6) is based on the data obtained from Ojima et al.[8] (see Table 4):

$$\log 1/IC_{50} = 0.55(\pm 0.17)\pi_X + 1.19(\pm 0.47)MR_Y - 5.41(\pm 1.19)\pi_Z + 15.69(\pm 2.51)$$
$$n = 20,\ r^2 = 0.910,\ s = 0.160,\ q^2 = 0.836,\ q_m{}^2 = 0.761,\ Q = 5.963,\ F_{3,16} = 53.926 \quad (6)$$

in which π_X and π_Z are the calculated hydrophobic parameters of X and Z substituents, respectively, whereas MR_Y is the calculated molar refractivity of Y substituents. According to this QSAR model, the taxane derivative **5** must have a more hydrophobic X substituent, a more hydrophilic Z substituent, and a bulkier or more polarizable Y substituent for improved cytotoxicity against MCF-7-R cancer cells. A comparison between observed and predicted log 1/IC_{50} values of taxane analogues **5** used in the development of QSAR Equation (6) is shown in Figure 2.

Inhibition of growth of LCC6-MDR (MDR1-transduced human breast carcinoma) by paclitaxel, docetaxel, doxorubicin, and other taxane derivatives 7 and 8

Equation (7) is based on the data obtained from Ojima et al.[26] (see Table 5):

$$\log 1/IC_{50} = 0.54(\pm 0.21)\log P + 5.46(\pm 0.80)$$
$$n = 11,\ r^2 = 0.792,\ s = 0.379,\ q^2 = 0.714,\ q_m{}^2 = 0.689,\ Q = 2.348,\ F_{1,9} = 34.269$$
outliers : paclitaxel and **8**-5 (7)

Hydrophobicity is found to be the single most important parameter for this dataset, which shows that hydrophobic contacts are made at all positions where substituents have been entered. The linear log*P* model suggests that highly hydrophobic molecules will be more active. It is interesting to note here that a different class of chemotherapeutic agent such as doxorubicin is very well predicted by this QSAR model [Eq. (7)]. Two compounds, paclitaxel and **8**-5, were deemed to be outliers on the basis of their deviations (obsd−pred $> 2\times s$).

Table 3. Biological and physicochemical parameters used to derive QSAR Equation (5).

Compd	log 1/IC_{50} [Eq. (5)]			CMR
	Obsd	Pred	Δ	
3-1	10.37	10.09	0.28	22.18
3-2	10.70	10.03	0.67	22.08
3-3	10.42	9.94	0.48	21.90
3-4[a]	10.92	9.80	1.12	21.61
3-5	10.39	10.35	0.04	22.68
3-6	10.49	10.09	0.40	22.18
3-7	10.49	10.25	0.24	22.49
3-8	10.39	9.98	0.41	21.97
3-9	10.38	10.03	0.35	22.07
3-10	9.92	9.89	0.03	21.79
3-11	9.88	9.74	0.14	21.51
3-12	9.68	10.29	−0.61	22.58
3-13	10.51	10.03	0.48	22.07
3-14	10.21	10.19	0.02	22.39
3-15	10.30	10.29	0.01	22.57
3-16	10.51	10.34	0.17	22.67
3-17	10.19	10.20	−0.01	22.39
3-18	9.79	10.05	−0.26	22.10
3-19	10.32	10.34	−0.02	22.67
3-20	10.44	10.50	−0.06	22.98
3-21	10.49	10.70	−0.21	23.36
3-22	10.20	10.44	−0.24	22.86
4-1[a]	8.77	10.06	−1.29	22.13
4-2	9.0	9.31	−0.31	20.66
4-3	9.38	9.81	−0.43	21.64
4-4	9.46	9.70	−0.24	21.43
4-5	8.40	8.97	−0.57	20.01
4-6	9.26	8.96	0.30	19.98
4-7	9.74	9.69	0.05	21.41
4-8	9.70	9.86	−0.16	21.73
4-9[a]	8.03	9.25	−1.22	20.55
4-10	9.48	9.61	−0.13	21.25
4-11	9.55	9.21	0.34	20.47
4-12[a]	8.29	10.07	−1.78	22.14
4-13	9.22	9.45	−0.23	20.94
4-14	9.89	9.88	0.01	21.78
4-15	9.51	9.89	−0.38	21.80
4-16	9.40	9.64	−0.24	21.31
4-17	9.70	10.36	−0.66	22.70
4-18	9.85	9.53	0.32	21.10

[a] Not used in the derivation of QSAR Equation (5)

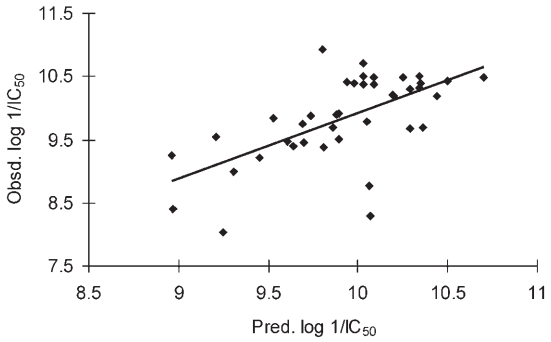


Figure 1. Plot of observed versus predicted log 1/IC_{50} from Equation (5).

Conclusions

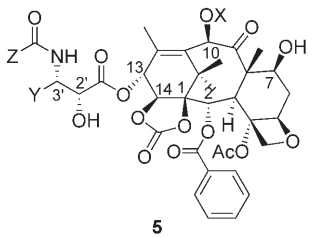
Among novel chemotherapeutic agents, taxane analogues have emerged as the most powerful group of compounds. The analogues taxol and taxotere are two important antimitotic drugs in current clinical use for the treatment of breast cancers. These drugs act as microtubule stabilizers and disrupt microtubule dynamics, thus inducing mitotic arrest and ultimately cell death by apoptosis. Although taxol and taxotere have made significant progress in the treatment of breast cancers, recent reports indicate that their use often results in various undesired side effects as well as multi-drug resistance. These limitations have led to the development of new taxane derivatives that have fewer side effects, superior pharmacological properties, and improved anticancer activity, thus maximizing the induced benefits for breast cancer patients.

The QSAR paradigm may be helpful in the design and development of novel taxane molecules that are expected to have improved anti-breast-cancer activities, toxicity profiles, and pharmacology, as well as better drug formulation. Our QSAR results suggest that the inhibitory activities of taxane analogues against breast cancers are mainly dependent on either the hydrophobicity or the hydrophobic/molar refractivity descriptors of their substituents. Parabolic correlation [Eq. (1)] with MR_Y is the most encouraging example, in which the optimum value of this parameter is well defined. We believe that this correlation may prove to be an adequate predictive model that can provide guidance in design and synthesis and subsequently yield very specific compounds (compounds **3**) that may have high anticancer activity with fewer side effects and superior pharmacological properties. On the basis of this QSAR model, five compounds (**3**-23, **3**-24, **3**-25, **3**-26, and **3**-27) are suggested as potential synthetic targets. Further QSAR studies should not only enlarge the areas of their application, but also increase our understanding of the mechanisms of chemical–biological interactions.

Experimental Section

All data were collected from published sources (see individual QSAR for specific literature references). IC_{50} is the molar concentration of a compound that inhibits 50% of the growth of a given cancer cell population; log 1/IC_{50} is the subsequent dependent variable that defines the biological parameter for QSAR development. Physicochemical descriptors were auto-loaded, and multi-regression analyses (MRA) was used to derive the QSAR by using the C-QSAR program.[27] Selection of descriptors was made on the basis of permutation and correlation matrices among the descriptors in order to avoid collinearity problems. Details about the C-QSAR program, the search engine, the choice of parameters, and their use in the development of QSAR models have been discussed previously.[28] The log *P* value is the calculated partition coefficient of a compound in *n*-octanol/water and is a measure of its hydrophobicity, whereas π is the hydrophobic parameter for the substituents only. CMR is the calculated molar refractivity for the whole molecule. Molar refractivity (MR) is calculated from the Lorentz–Lorenz equation, and is described as follows: $[(n^2-1)/(n^2+2)](M_w/\delta)$, for which n is the refractive index, M_w is the molecular weight, and δ is the density of a substance. MR is dependent on volume and polariza-

Table 4. Biological and physicochemical parameters used to derive QSAR Equation (6).



5

Compd	X	Y	Z	log 1/IC_{50} [Eq. (6)]			π_X	MR_Y	π_Z
				Obsd	Pred	Δ			
5-1	H	2-furyl	$OC(CH_3)_3$	6.87	6.82	0.05	0.00	1.81	2.04
5-2	$COCH_3$	2-furyl	$OC(CH_3)_3$	7.31	7.22	0.09	0.71	1.81	2.04
5-3	H	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	7.14	6.95	0.19	0.00	1.92	2.04
5-4	$COCH_3$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	7.44	7.34	0.10	0.71	1.92	2.04
5-5	$COCH_2CH_3$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	7.59	7.63	−0.04	1.24	1.92	2.04
5-6	CO(*c*-C_3H_5)	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	7.55	7.66	−0.11	1.30	1.92	2.04
5-7	$CON(CH_3)_2$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	7.48	7.57	−0.09	1.13	1.92	2.04
5-8	$COCH{=}CHCH_3$ (*E*)	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	7.77	7.87	−0.10	1.67	1.92	2.04
5-9	$COOCH_3$	$CH{=}C(CH_3)_2$	$OC(CH_3)_3$	7.42	7.31	0.11	0.66	1.92	2.04
5-10	H	$CH{=}CHCH_3$ (*E*)	$OC(CH_3)_3$	6.12	6.39	−0.27	0.00	1.45	2.04
5-11	$COCH_3$	$CH{=}CHCH_3$ (*E*)	$OC(CH_3)_3$	7.14	6.79	0.35	0.71	1.45	2.04
5-12	CO(*c*-C_3H_5)	$CH{=}C(CH_3)_2$	C_5H_{11}	6.49	6.59	−0.10	1.30	1.92	2.24
5-13	$CON(CH_3)_2$	$CH{=}C(CH_3)_2$	C_5H_{11}	6.53	6.50	0.03	1.13	1.92	2.24
5-14	$COCH{=}CHCH_3$ (*E*)	$CH{=}C(CH_3)_2$	C_5H_{11}	6.86	6.80	0.06	1.67	1.92	2.24
5-15	H	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	6.72	6.98	−0.26	0.00	1.94	2.04
5-16	$COCH_3$	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	7.44	7.37	0.07	0.71	1.94	2.04
5-17	CO(*c*-C_3H_5)	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	7.66	7.69	−0.03	1.30	1.94	2.04
5-18	$CON(CH_3)_2$	$CH_2CH(CH_3)_2$	$OC(CH_3)_3$	7.66	7.60	0.06	1.13	1.94	2.04
5-19	H	$(CH_2)_2CH_3$	$OC(CH_3)_3$	6.41	6.42	−0.01	0.00	1.48	2.04
5-20	$COCH_3$	$(CH_2)_2CH_3$	$OC(CH_3)_3$	6.70	6.82	−0.12	0.71	1.48	2.04

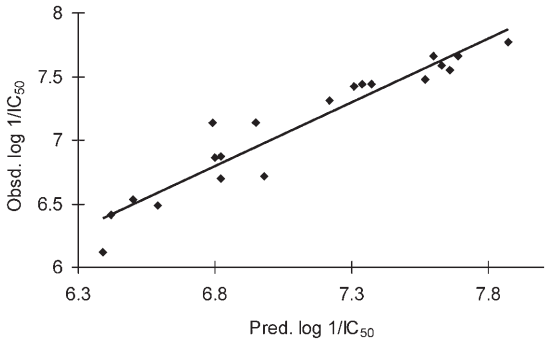
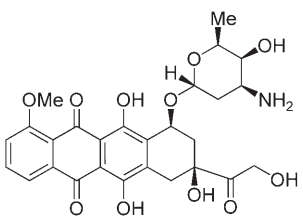


Figure 2. Plot of observed versus predicted log 1/IC_{50} from Equation (6).

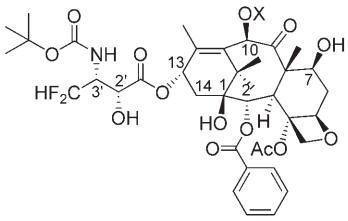
bility. It can be used for a substituent or for the whole molecule. MR is thus a means of characterizing the bulk and polarizability of a substituent or compound. Although it contains no information about the shape, it has found considerable use in biological QSAR, where intermolecular effects predominate. MR is usually scaled at 0.1 to make it equiscalar with π. The indicator variable *I* is assigned the value of 1 or 0 for special features with special effects that cannot be parameterized, and has been explained wherever used.

In QSAR equations, *n* is the number of data points, *r* is the correlation coefficient between observed values of the dependent and the values calculated from the equation, r^2 is the square of the correlation coefficient and represents the goodness of fit, q^2 is the cross-validated r^2 (a measure of the quality of the QSAR model), and *s* is the standard deviation. The cross-validated r^2 (q^2) or LOO−q^2 is obtained by using the leave-one-out (LOO) procedure as described by Cramer III et al.[29] Similarly, the cross-validated r_m^2 (q_m^2) or LMO−q^2 is obtained by using the leave-many-out (LMO) procedure. *Q* is the quality factor (quality ratio), for which $Q = r/s$. Chance correlation due to the excessive number of parameters (which also increases the *r* and *s* values) can, therefore, be detected by the examination of the *Q* value. High values of *Q* indicate the high predictive power of the QSAR models and the lack of "over-fitting". *F* represents the Fischer statistics (Fischer ratio), $F = fr^2/[(1-r^2)m]$, 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 *F* value is actually the ratio between explained and unexplained variance for a given number of degrees of freedom. Thus, it indicates a true relationship, or the significance level for MLR models. The modeling was taken to be optimal when *Q* reached a maximum together with *F*, even if slightly non-optimal *F* values have normally been accepted. A significant decrease in *F* with the introduction of one additional variable (with increasing *Q* and decreasing *s*) could mean that the new descriptor is not as significant as expected, that is, its introduction has endangered the statistical quality of the combination. However, the statistical quality could be improved by the introduction of a more convincing descriptor.[30] Compounds were deemed to be outliers on the basis of their deviation between observed (obsd) and calculated (pred) activities if obsd−pred > 2*s*.[31] Each regression equation includes 95% confidence limits for each term in parentheses. All QSAR models reported herein are derived by us and were not formulated by the original authors. These QSARs are statistically significant, which fulfill the conditions of acceptable models as given by Golbraikh and Tropsha.[32]

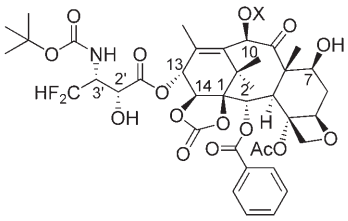
Table 5. Biological and physicochemical parameters used to derive QSAR Equation (7).



6



7



8

Compd	X	log 1/IC_{50} [Eq. (7)] Obsd	Pred	Δ	log *P*
paclitaxel (**1**)[a]		6.46	8.0	−1.54	4.73
docetaxel (**2**)		6.92	7.65	−0.73	4.08
doxorubicin (**6**)		5.54	5.63	−0.09	0.32
7-1	H	7.36	7.15	0.21	3.14
7-2	$COCH_3$	8.06	7.53	0.53	3.86
7-3	$CON(CH_3)_2$	8.09	7.75	0.34	4.27
7-4	COC_2H_5	8.24	7.82	0.42	4.39
7-5	$COCH_2C(CH_3)_3$	8.30	8.53	−0.23	5.71
8-1	H	ND[b]	6.83	ND[b]	2.54
8-2	$COCH_3$	7.19	7.21	−0.02	3.25
8-3	$CON(CH_3)_2$	7.28	7.43	−0.15	3.67
8-4	COC_2H_5	7.20	7.24	−0.04	3.32
8-5[a]	$COCH_2C(CH_3)_3$	6.88	8.23	−1.35	5.15
8-6	$CO(c\text{-}C_3H_5)$	7.31	7.54	−0.23	3.88

[a] Not used in the derivation of QSAR Equation (7). [b] Not determined.

Validation of QSAR models

QSAR model validation becomes an essential part to understand statistically robust models capable of making accurate and reliable predictions of biological activities of new compounds not present in the data set. The following approaches have been used to validate QSAR Equations (1), (2), and (4)–(7):

Internal validation

- **Fraction of the variance (r^2):** It is believed that the closer the value of r^2 to unity, the better the QSAR model. The values of r^2 for these six QSAR models [Eqs. (1), (2), (4)–(7)] range from 0.613 to 0.910, which suggests that these QSAR models explain 61.3–91.0% of the variance in the data. According to the literature, the predictive QSAR model must have $r^2 > 0.6$.[32,33]
- **Cross-validation test:** The cross-validated r^2 [q^2 (LOO−q^2)] and [$q_m{}^2$ (LMO−q^2)] values for these QSAR models range from 0.501 to 0.836, and 0.548 to 0.761, respectively. High values of q^2 and $q_m{}^2$ validate the QSAR models. According to the literature, the predictive QSAR model must have $q^2 > 0.5$.[32,33]
- **Standard deviation (*s*):** The smaller the value of *s*, the better the QSAR model. The values of *s* for these QSAR models range from 0.144 to 0.379.
- **Quality factor (*Q*)**: High values of *Q* (2.344–5.963) for these QSAR models suggest their high predictive power.
- **Fischer statistics (*F*)**: The larger the value of *F*, the greater the probability that the QSAR equation is significant. The *F* values for these QSAR models range from 13.987 to 53.926, which are statistically significant at the 95% level.
- **Y-randomization test:** In this test, the dependent-variable vector (Y vector) is randomly shuffled, and a new QSAR

Table 6. Y-Randomization data for QSAR Equations (1), (2), and (4)–(7).[a]

Eq.	NOR-1		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	0.248	−0.055	0.198	−0.278	0.225	−0.323	0.063	−0.427	0.088	−0.339
2	0.095	−0.528	0.131	−0.150	0.102	−0.314	0.306	0.025	0.052	−0.379
4	0.346	−0.230	0.044	−1.218	0.332	−1.466	0.400	−0.392	0.335	−1.386
5	0.166	0.094	0.004	−0.102	0.152	0.081	0.108	0.018	0.035	−0.099
6	0.186	−0.288	0.185	−0.133	0.187	−0.130	0.248	−0.202	0.185	−0.208
7	0.001	−0.517	0.013	−0.412	0.148	−0.489	0.244	−0.074	0.144	−0.164

[a] NOR = number of Y randomization.

model is developed using the original independent variable matrix. The process is repeated several times. It is expected that the resulting QSAR models should have low r^2 and low q^2 values. This is a widely used technique to ensure the robustness of a QSAR model. The statistical data of r^2 and q^2 for five runs are listed in Table 6 [Eqs. (1), (2), (4)–(7)]. The poor values of r^2 and q^2 in the Y-randomization test ensure the robustness of QSAR models.[31a,33–35]

- **Lack of over-fitting**: A model over-fits if it includes more descriptors than required. The lack of over-fitting for all the QSAR models [Eqs. (1), (2), (4)–(7)] was confirmed by using the following conditions:
  1) Number of data points/Number of descriptors ≥ 4
  2) High values of Q
  3) All the QSAR models [Eqs. (1), (2), (4)–(7)] were checked for their correlation with fewer descriptors than that of the original. None of them was found to be statistically significant.
  4) Y-randomization test (Table 6) suggests that the high r^2 values of the QSAR models [Eqs. (1), (2), (4)–(7)] are not due to chance correlation or over-fitting.[36]

External validation

Selection of the training and test sets: The original data set of QSAR model (1) was divided into training [$n=15$ (~75%)] and test [$n=5$ (~25%)] sets in a random manner (five trials). The QSAR models for these five training sets were generated by using the same descriptors as those of Equation (1) and validated on the basis of their statistics (acceptance criteria: $r^2>0.6$ and $q^2>0.5$). The predictive capacity of these models is judged from their predictive R^2 (R^2_{pred}) values, which were calculated by the following [Eq. (8)]:

$$R^2_{pred} = 1 - \frac{\sum (Y_{pred(test)} - Y_{test})^2}{\sum (Y_{test} - \overline{Y}_{training})^2} \quad (8)$$

Table 7. Random selection pattern of the test sets and statistical parameters of the QSARs for their respective training sets obtained from the division of the original sets of the QSAR Equations (1), (2), and (4)–(7).

Eq.	NOS[a]	Test sets	Statistical parameters[b]				
			n	r^2	q^2	s	R^2_{pred}
1	1a	**3**-3, **3**-10, **3**-14, **3**-18, **3**-22	15	0.694	0.591	0.109	0.597
	1b	**3**-2, **3**-3, **3**-7, **3**-10, **3**-11	15	0.709	0.505	0.108	0.510
	1c	**3**-2, **3**-3, **3**-10, **3**-18, **3**-22	15	0.733	0.618	0.093	0.579
	1d	**3**-2, **3**-7, **3**-14, **3**-18, **3**-22	15	0.681	0.550	0.120	0.537
	1e	**3**-2, **3**-14, **3**-16, **3**-18, **3**-22	15	0.697	0.570	0.116	0.512
2	2a	**3**-10, **3**-13, **3**-18, **3**-22	15	0.704	0.523	0.109	0.632
	2b	**3**-2, **3**-11, **3**-18, **3**-22	15	0.792	0.605	0.079	0.609
	2c	**3**-3, **3**-10, **3**-18, **3**-22	15	0.730	0.523	0.106	0.634
	2d	**3**-13, **3**-16, **3**-18, **3**-19	15	0.719	0.521	0.125	0.536
	2e	**3**-3, **3**-7, **3**-14, **3**-18	15	0.731	0.524	0.123	0.523
4	4a	**4**-1, **4**-11, **4**-16, **4**-18	12	0.935	0.718	0.185	0.695
	4b	**4**-2, **4**-8, **4**-10, **4**-18	12	0.916	0.765	0.211	0.792
	4c	**4**-4, **4**-7, **4**-8, **4**-10	12	0.914	0.785	0.217	0.831
	4d	**4**-8, **4**-11, **4**-15, **4**-16	12	0.930	0.748	0.200	0.640
	4e	**4**-7, **4**-8, **4**-15, **4**-18	12	0.904	0.742	0.217	0.923
5	5a	**3**-2, **3**-3, **3**-6, **3**-12, **3**-13, **4**-2, **4**-3, **4**-5, **4**-15	27	0.674	0.626	0.241	0.527
	5b	**3**-2, **3**-6, **3**-7, **3**-9, **3**-13, **3**-16, **3**-21, **4**-5, **4**-17	27	0.614	0.552	0.272	0.600
	5c	**3**-2, **3**-7, **3**-12, **3**-16, **3**-21, **4**-2, **4**-3, **4**-5, **4**-17	27	0.633	0.582	0.258	0.566
	5d	**3**-2, **3**-12, **3**-16, **3**-21, **4**-2, **4**-3, **4**-5, **4**-6, **4**-17	27	0.606	0.543	0.258	0.534
	5e	**3**-2, **3**-7, **3**-12, **3**-21, **4**-2, **4**-5, **4**-6, **4**-11, **4**-17	27	0.606	0.551	0.264	0.616
6	6a	**5**-2, **5**-6, **5**-8, **5**-14, **5**-19	15	0.891	0.751	0.182	0.951
	6b	**5**-5, **5**-9, **5**-14, **5**-17, **5**-18	15	0.888	0.779	0.187	0.977
	6c	**5**-1, **5**-4, **5**-8, **5**-13, **5**-20	15	0.898	0.756	0.179	0.921
	6d	**5**-7, **5**-8, **5**-14, **5**-16, **5**-20	15	0.903	0.776	0.179	0.912
	6e	**5**-4, **5**-7, **5**-12, **5**-18, **5**-19	15	0.882	0.743	0.184	0.968
7	7a	**7**-1, **7**-4, **7**-5	8	0.774	0.671	0.406	0.854
	7b	**7**-2, **7**-4, **8**-3	8	0.853	0.769	0.345	0.566
	7c	**7**-5, **8**-2, **8**-6	8	0.790	0.694	0.433	0.700
	7d	**7**-4, **8**-2, **8**-4	8	0.800	0.671	0.425	0.709
	7e	**7**-3, **7**-5, **8**-3	8	0.766	0.641	0.428	0.889

[a] NOS = number of selection of the training/test sets. [b] Parameters of the QSARs for the training sets (the compounds remaining after removal of the test set compounds from the original set).

Table 8. Predicted log $1/IC_{50}$ values of the proposed molecules obtained from Equations (1), (1a), (2), (2c), (5), and (5e) and their mean values along with physicochemical parameters.

Compd	X	Y	Predicted log $1/IC_{50}$							MR_Y	log *P*	CMR
			Eq. (1)	TSBP[a] (1a)	Eq. (2)	TSBP[a] (2c)	Eq. (5)	TSBP[a] (5e)	Mean value			
3-23	3-thiophenyl	$O(CH_2)_3CH_3$	10.44	10.47	10.43	10.46	10.34	10.42	10.43	2.17	4.78	22.67
3-24	3-thiophenyl	$CH_2CH{=}C(CH_3)_2$	10.45	10.48	10.45	10.47	10.48	10.58	10.49	2.25	4.41	22.95
3-25	OC_6H_5	$O(CH_2)_3CH_3$	10.44	10.47	10.42	10.44	10.52	10.61	10.48	2.17	4.16	23.01
3-26	C_6H_5	$CH{=}CHSO_2CF_3$	10.45	10.49	10.44	10.46	10.62	10.73	10.53	2.28	4.28	23.21
3-27	C_6H_5	$CH_2CH{=}C(CH_3)_2$	10.45	10.48	10.44	10.46	10.59	10.69	10.52	2.25	4.76	23.15

[a] TSBP = training set with best predictive capacity.

In Equation (8), $Y_{pred(test)}$ and Y_{test} are the respective predicted and observed activities of the test set compounds, and $\overline{Y}_{training}$ is the mean activity of the training set compounds. A random selection pattern of the test sets and the statistical parameters of the respective training sets for QSAR (1) are given in Table 7. Similarly, the QSAR models (2) and (4)–(7) were also divided into training and test sets for their external validation. Details about the selection pattern of the test sets and the statistical parameters of the respective training sets for the QSAR Equations (2) and (4)–(7) are also listed in Table 7.

New molecule prediction

QSAR model (1) is the parabolic correlation in terms of MR_Y (molar refractivity of Y substituents), and is the most encouraging example in which the optimum value of MR_Y is well defined. We believe that this model may prove to be an adequate predictive model for providing guidance in design and synthesis, and for yielding very specific compounds (such as **3**) that may have high anti-breast-cancer activity with fewer side effects and superior pharmacological properties. On the basis of this QSAR model, five compounds (**3**-23, **3**-24, **3**-25, **3**-26, and **3**-27) are suggested as potential synthetic targets. The predicted log $1/IC_{50}$ values of these proposed molecules obtained from QSAR models (1), (2), and (5) and also from their best training sets [(1a), (2c), and (5e); Table 7] are listed in Table 8 along with their physicochemical parameters. Good agreement among the log 1/C values of the predicted molecules obtained from different models suggests accuracy in the prediction.

Keywords: breast cancer · hydrophobicity · molar refractivity · QSAR · taxanes

[1] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, M. J. Thun, *CA Cancer J. Clin.* **2007**, *57*, 43–66.

[2] a) Medical Encyclopedia (Breast Cancer), accessed September 12, 2007, available at: http://www.nlm.nih.gov/medlineplus/ency/article/000913.htm; b) American Cancer Society (Overview: Breast Cancer), accessed September 12, 2007, available at: http://www.cancer.org/docroot/CRI/CRI_2_1x.asp?dt=5.

[3] D. Guénard, F. Guéritte-Vogelein, P. Potier, *Acc. Chem. Res.* **1993**, *26*, 160–167.

[4] M. Suffness, *Taxol: From Discovery to Therapeutic Use, Annual Reports in Medicinal Chemistry, Vol. 28* (Ed.: J. A. Bristol), Academic Press, San Diego, **1993**, pp. 305–314.

[5] *Taxane Anticancer Agents: Basic Science and Current Status* (Eds.: G. I. Georg, T. T. Chen, I. Ojima, D. M. Vyas), American Chemical Society, Washington DC, **1995**.

[6] I. E. L. M. Kuppens, *Curr. Clin. Pharmacol.* **2006**, *1*, 57–70.

[7] E. Galletti, M. Magnani, M. L. Renzulli, M. Botta, *ChemMedChem* **2007**, *2*, 920–942.

[8] 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.* **1997**, *40*, 267–278.

[9] J.-M. Nabholtz, J. Gligorov, *Expert Opin. Pharmacother.* **2005**, *6*, 1073–1094.

[10] S. G. Arbuck, B. A. Blaylock, *Taxol: Clinical Results and Current Issues in Development, Taxol: Science and Applications* (Ed.: M. Suffness), CRC, Boca Raton, **1995**, pp. 379–415.

[11] J. Verweij, M. Clavel, B. Chevalier, *Ann. Oncol.* **1994**, *5*, 495–505.

[12] C. D. Selassie, S. B. Mekapati, R. P. Verma, *Curr. Top. Med. Chem.* **2002**, *2*, 1357–1379.

[13] C. Hansch, P. P. Maloney, T. Fujita, R. M. Muir, *Nature* **1962**, *194*, 178–180.

[14] C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington DC, **1995**, pp. 169–543.

[15] C. D. Selassie, R. Garg, S. Kapur, A. Kurup, R. P. Verma, S. B. Mekapati, C. Hansch, *Chem. Rev.* **2002**, *102*, 2585–2605.

[16] R. P. Verma, A. Kurup, S. B. Mekapati, C. Hansch, *Bioorg. Med. Chem.* **2005**, *13*, 933–948.

[17] C. Hansch, A. Leo, S. B. Mekapati, A. Kurup, *Bioorg. Med. Chem.* **2004**, *12*, 3391–3400.

[18] R. P. Verma, C. Hansch, C. D. Selassie, *J. Comput.-Aided Mol. Des.* **2007**, *21*, 3–22.

[19] E. Baloglu, M. L. Miller, E. E. Roller, E. E. Cavanagh, B. A. Leece, V. S. Goldmacher, R. V. J. Chari, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5885–5888.

[20] R. P. Verma, S. Kapur, O. Barberena, A. Shusterman, C. H. Hansch, C. D. Selassie, *Chem. Res. Toxicol.* **2003**, *16*, 276–284.

[21] K. H. Kim, *J. Comput.-Aided Mol. Des.* **2001**, *15*, 367–380.

[22] F. Helmer, K. Kiehs, C. Hansch, *Biochemistry* **1968**, *7*, 2858–2863.

[23] a) J. C. Dearden, *Environ. Health Perspect.* **1985**, *61*, 203–228; b) Y.-Z. Da, K. Ito, H. Fujiwara, *J. Med. Chem.* **1992**, *35*, 3382–3387; c) E. Akaho, G. Morris, D. Goodsell, D. Wong, A. Olson, *J. Chem. Software* **2001**, *7*, 103–114.

[24] F. Manetti, L. Maccari, F. Corelli, M. Botta, *Curr. Top. Med. Chem.* **2004**, *4*, 203–217.

[25] M. N. Islam, Y. Song, M. N. Iskander, *J. Mol. Graphics Modell.* **2003**, *21*, 263–272.

[26] I. Ojima, S. Lin, J. C. Slater, T. Wang, P. Pera, R. J. Bernacki, C. Ferlini, G. Scambia, *Bioorg. Med. Chem.* **2000**, *8*, 1619–1628.

[27] C-QSAR Program, BioByte Corporation, Claremont, CA 91711 (USA), http://www.biobyte.com (latest access: December 17, 2007).

[28] a) C. Hansch, D. Hoekman, A. Leo, D. Weininger, C. D. Selassie, *Chem. Rev.* **2002**, *102*, 783–812; b) R. P. Verma, C. Hansch, *Nat. Protoc.*, DOI: 10.1038/nprot.2007.125. Freely available at: http://www.natureprotocols.com/2007/03/05/development_of_qsar_models_usi_1.php.

[29] R. D. Cramer III, J. D. Bunce, D. E. Patterson, I. E. Frank, *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.

[30] a) L. Pogliani, *Chem. Rev.* **2000**, *100*, 3827–3858; b) L. Pogliani, *J. Phys. Chem.* **1996**, *100*, 18065–18077; c) V. K. Agrawal, J. Singh, P. V. Khadikar, C. T. Supuran, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2044–2051.

[31] a) C. D. Selassie, S. Kapur, R. P. Verma, M. Rosario, *J. Med. Chem.* **2005**, *48*, 7234–7242; b) R. P. Verma, C. Hansch, *Mol. Pharm.* **2006**, *3*, 441–450;

c) R. P. Verma, C. Hansch, *Virology* **2007**, *359*, 152–161; d) R. P. Verma, C. Hansch, *Bioorg. Med. Chem.* **2007**, *15*, 2223–2268; e) R. P. Verma, C. Hansch, *J. Pharm. Sci.* **2008**, *97*, 88–110.
[32] A. Golbraikh, A. Tropsha, *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.
[33] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **2003**, *22*, 69–77.
[34] S. Wold, L. Eriksson, *Statistical Validation of QSAR Results, Chemometrics Methods in Molecular Design* (Ed.: H. van de Waterbeemd), VCH, Weinheim, **1995**, pp. 309–318.
[35] a) M. F. Melzig, G. D. Tran, K. Henke, C. D. Selassie, R. P. Verma, *Pharmazie* **2005**, *60*, 869–873; b) C. Hansch, R. P. Verma, *ChemMedChem* **2007**, *2*, 1807–1813.
[36] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, *J. Chem. Inf. Model.* **2006**, *46*, 1984–1995.

Received: October 3, 2007
Revised: November 28, 2007
Published online on January 14, 2008

