molecular pharmaceutics

articles

Understanding Tubulin/Microtubule−Taxane Interactions: A Quantitative Structure–Activity Relationship Study

Corwin Hansch and Rajeshwar P. Verma*

Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, California 91711

Received September 6, 2007; Revised Manuscript Received October 16, 2007; Accepted October 26, 2007

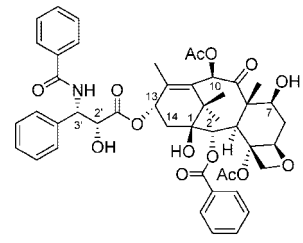
Abstract: For years, the microtubule-stabilizing agents paclitaxel and docetaxel (progenitors of the family of taxanes) have been the most successful anticancer drugs currently used in clinics. However, both drugs are associated with notorious side effects, drug resistance, and cross resistance with other chemotherapeutic agents. These limitations have led to the search for new drugs with improved biological activity. In the present paper, we discuss the interaction of taxanes with the tubulin/microtubule system by the formulation of 6 QSARs. Hydrophobicity of the substituents (π) is found to be one of the most important determinants of the activity followed by steric parameters. Parabolic correlations (eqs 3 and 7) with *B*5 and π are the most encouraging examples, where the optimum values of these parameters are well defined. We believe that these two QSARs may prove to be adequate predictive models that can help to provide guidance in design/synthesis and subsequently yield very specific compounds (**IV** and **VIII**) that may have high biological activities. On the basis of these two QSARs 3 and 7, 18 compounds (**IV-12**−**IV-22** and **VIII-16**−**VIII-22**) are suggested as potential synthetic targets. Cross-validation, quality factor (*Q*), Fischer statistics (*F*), and Y-randomization tests have validated all the QSAR models.

Keywords: Hydrophobicity; microtubule; molar refractivity; QSAR; taxanes; tubulin; Verloop's sterimol parameters

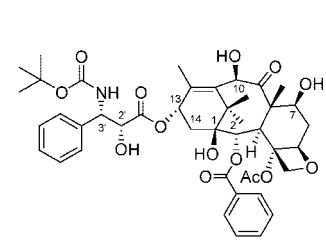
Introduction

Cancer has become a very serious public health problem in the United States and other developed countries. Every year, the American Cancer Society estimates the number of new cancer cases and deaths expected in the United States and compiles the most recent data on cancer incidence, mortality, and survival. The data indicate that at least 1 444 920 new cancer cases and 559 650 deaths from cancers are expected to occur in the United States in 2007. Currently, about 25% of deaths in the United States are due to cancer.[1] Various groups of compounds that interfere with the cell cycle have attracted major interest in the cancer research field because they inhibit the proliferation of cancer cells. These types of compounds (antimitotic) target the microtubules of the cells and generally fall into three distinct classes, which are based on their binding site on β-tubulin, namely the vincas, the colchicines, and the taxanes.[2]

The taxanes such as paclitaxel (Taxol, **I**) and docetaxel (taxotere, **II**) are the two important antimitotic drugs currently used in clinics for the treatment of various types of cancer, including ovarian, breast, lung, and prostate. These drugs



Paclitaxel (**I**)



Docetaxel (**II**)

* To whom correspondence should be addressed. Mailing address: Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, CA 91711. Phone: (909) 607-4249. Fax: (909) 607-7726. E-mail: rverma@pomona.edu.

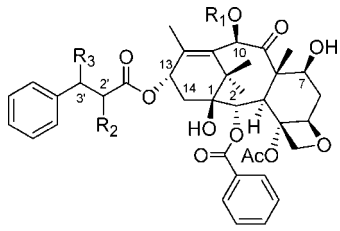
(1) Jemal, A.; Siegel, R.; Ward, E.; Murray, T.; Xu, J.; Thun, M. J. Cancer statistics, 2007. *Ca−Cancer J. Clin.* **2007**, *57*, 43–66.

10.1021/mp700119e CCC: $40.75
Published on Web 01/03/2008

act as microtubule stabilizers and disrupt microtubule dynamics, thus inducing mitotic arrest and, ultimately, cell death by appoptosis.

The major common drawbacks of these antimitotic drugs are drug resistance, neurotoxicity, substrate for drug transporter P-gp, cross-resistance against different chemotherapeutic agents (such as vinblastin, etoposide, and doxorubicin), low oral bioavailability, and no penetration in the blood−brain barrier (BBB). The two main mechanisms involved in resistance to taxanes are the expression of the multidrug resistance (MDR) phenotype and the alterations of their cellular target, namely the tubulin/microtubule system.[3,4] To prevent these disadvantages and improve the clinical application of the taxanes, several new agents have entered into the clinical trials. A large number of microtubule-stabilizing agents, which bind to the taxane site of β-tubulin, that are presently in various stages of clinical trials are discodermolides, eleutherobin, epothilones, laulimalides, peloruside, sarcodictyins, and taxane derivatives.[4–6]

The quantitative structure–activity relationship (QSAR) paradigm may be helpful in the design and development of novel tubulin-binding molecules as new anticancer agents, which will be expected to achieve improvements in anticancer activity, toxicity profile, pharmacology, and drug formulation. In a QSAR study for the inhibition of pig brain microtubule disassembly by two isomeric series of Taxol and its derivatives (**III**), two QSAR models (eqs 1 and 2) were developed.[7]



III

The QSAR model for the (2′*R*,3′*S*) isomeric series (**III**) as developed by Liu et al.:[7]

$$\log 1/\mathrm{ID}_{50} = 0.033(\pm 0.005)\mathrm{MR}_3 + 5.35(\pm 0.076) \quad (1)$$

where $n = 11$, $r = 0.91$, $s = 0.25$, and $F = 42.15$.

The QSAR model for the (2′*S*,3′*R*) isomeric series (**III**) as developed by Liu et al.:[7]

$$\log 1/\mathrm{ID}_{50} = 0.041(\pm 0.007)\mathrm{MR}_3 - 4.14(\pm 0.110) \quad (2)$$

where $n = 9$, $r = 0.90$, $s = 0.32$, and $F = 30.97$.

In these QSAR, ID_{50} is the concentration of the drugs leading to a 50% inhibition of the rate of microtubule disassembly. MR_3 is the molar refractivity of C3′ substituents, n is the number of data points, r is the correlation coefficient, s is the standard deviation, and F is the F ratio between the variances of calculated and observed activities. The parameter MR_3 is significant at 99% confidence level and accounts for 83% ($r^2 = 0.83$) and 81% ($r^2 = 0.81$) of the variance in the activity as represented by eqs 1 and 2, respectively. The coefficient of MR_3 in both eqs 1 and 2 is positive, which indicates that the higher the molar refractivity of C3′ substituent, the higher the activity.

In the present paper, we demonstrate the QSAR (quantitative structure–activity relationship) studies on various sets of taxanes with respect to their interaction with tubulin/microtubule system to understand the chemical−biological interactions. Microtubules are structures composed of polymerized tubulin heterodimers and play fundamental roles in vital cell processes such as chromosome segregation and intracellular transport. In the past 44 years, the use of QSAR (one of the well-developed areas in computational chemistry), since the advent of this methodology,[8] has become increasingly helpful in understanding many aspects of chemical−biological interactions in drug-design process and pesticide research as well as in the areas of toxicology. This method is useful in elucidating the mechanisms of chemical−biological interaction in various biomolecules, particularly enzymes, membranes, organelles, and cells, as well as in human.[9–11] It has also been utilized for the evaluation of absorption, distribution, metabolism, and excretion (ADME) phenomena in many organisms and whole animal studies.[12] The QSAR approach employs extra-thermodynamically derived and computational-based descriptors to correlate biological activity in isolated receptors, cellular systems, and in vivo. Four

(2) Mohanraj, S.; Doble, M. 3-D QSAR studies of microtubule stabilizing antimitotic agents towards six cancer cell lines. *QSAR Comb. Sci.* **2006**, *25*, 952–960.
(3) Galletti, E.; Magnani, M.; Renzulli, M. L.; Botta, M. Paclitaxel and docetaxel resistance: Molecular mechanisms and development of new generation taxanes. *ChemMedChem* **2007**, *2*, 920–942.
(4) Kuppens, I. E. L. M. Current state of the art of new tubulin inhibitors in the clinic. *Curr. Clin. Pharmacol.* **2006**, *1*, 57–70.
(5) He, L.; Orr, G. A.; Horwitz, S. B. Novel molecules that interact with microtubules and have functional activity similar to taxol. *DDT* **2001**, *6*, 1153–1164.
(6) Pineda, O.; Farràs, J.; Maccari, L.; Manetti, F.; Botta, M.; Vilarrasa, J. Computational comparison of microtubule-stabilising agents laulimalide and peloruside with taxol and colchicine. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4825–4829.
(7) Liu, A.-L.; Chi, H.-L. Study on QSAR of taxol and its derivatives based on sterwise multivariate linear regression analysis. *J. Chin. Pharm. Sci.* **1997**, *6*, 18–21.
(8) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* **1962**, *194*, 178–180.
(9) Hansch, C.; Leo, A. In *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington D.C., 1995.
(10) Selassie, C. D.; Garg, R.; Kapur, S.; Kurup, A.; Verma, R. P.; Mekapati, S. B.; Hansch, C. Comparative QSAR and the radical toxicity of various functional groups. *Chem. Rev.* **2002**, *102*, 2585–2605.
(11) Verma, R. P.; Kurup, A.; Mekapati, S. B.; Hansch, C. Chemical-biological interactions in human. *Bioorg. Med. Chem.* **2005**, *13*, 933–948.
(12) Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A. QSAR and ADME. *Bioorg. Med. Chem.* **2004**, *12*, 3391–3400.

standard molecular descriptors are routinely used in QSAR analysis: electronic, hydrophobic, steric, and topological indices. These descriptors are invaluable in helping to delineate a large number of receptor–ligand interactions that are critical to biological processes. The quality of a QSAR model depends strictly on the type and quality of the data, and not on the hypotheses, and is valid only for the compound structure analogues to those 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.[13] Potential use of QSAR models for screening of chemical databases or virtual libraries before their synthesis appears equally attractive to chemical manufacturers, pharmaceutical companies, and government agencies.

Materials and Methods

All of the data has been collected from the literature (see individual QSAR for respective references). ED_{50} or ID_{50} is the molar concentration of a compound which induces 50% tubulin/microtubule stability; log $1/ED_{50}$ or log $1/ID_{50}$ is the subsequent dependent variable that defines the biological parameter for QSAR development. Physicochemical descriptors are autoloaded, and multiregression analyses (MRA) are used to derive the QSAR by utilizing the C-QSAR program.[14] Selection of descriptors is 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, and the choice of parameters and their use in the development of QSAR models have already been discussed.[15,16]

The parameters used in this paper have also been discussed in detail along with their application.[9] Briefly, *C* log *P* is the calculated partition coefficient of a compound in *n*-octanol/water and is a measure of its hydrophobicity, while π is the hydrophobic parameter for substituents only. *B*1, *B*5, and *L* are Verloop's sterimol parameters for substituents.[17] *B*1 is a measure of the minimum width of a substituent, *B*5 defines the maximum width of the substituent, and *L* is the substituent length. Calculated molar refractivity (CMR) is the calculated molar refractivity for the whole molecule. Molar refractivity (MR) is calculated from the Lorentz−Lorenz equation and is described as $[(n^2 - 1)/(n^2 + 2)](MW/\delta)$, where n is the refractive index, MW is the molecular weight, and δ is the density of the substance. MR is dependent on both the volume and polarizability. It can be used for a substituent or for the whole molecule.

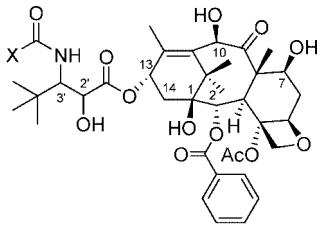
In QSAR equations, n is the number of data points, r is the correlation coefficient between observed values of the dependent and the values calculated from the equation, r^2 is the square of the correlation coefficient and represents the goodness of fit, q^2 is the cross-validated r^2 (a measure of the quality of the QSAR model), and s is the standard deviation. The cross-validated r^2 (q^2) is obtained by using a leave-one-out (LOO) procedure as described by Cramer et al.[18] Q is the quality factor (quality ratio), where $Q = r/s$. Chance correlation, due to the excessive number of parameters (which increases the r and s values also), can thus 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 "overfitting". F is the Fischer statistic (Fischer ratio), $F = fr^2/[(1 - r^2)m]$, where f is the number of degree 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 the MLR models. The modeling was taken to be optimal when Q reached a maximum together with F, even if slightly nonoptimal 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; i.e., 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.[19–21] Compounds were deemed to be outliers on the basis of their deviation between observed and calculated activities from the equation ($>2s$).[22–26] Each regression

(13) Selassie, C. D.; Mekapati, S. B.; Verma, R. P. QSAR: then and now. *Curr. Top. Med. Chem.* **2002**, *2*, 1357–1379.
(14) C-QSAR Program, BioByte Corp., 201 W. 4th St., Suite 204, Claremont, CA 91711, www.biobyte.com.
(15) Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C. D. Chem-Bioinformatics: Comparative QSAR at the interface between chemistry and biology. *Chem. Rev.* **2002**, *102*, 783–812.
(16) Verma, R. P.; Hansch, C. Development of QSAR models using C-QSAR program: a regression program that has dual databases of over 21,000 QSAR models. *Nat. Protoc.* advance online publication (DOI: 10.1038/nprot.2007.125). Freely available at: http://www.natureprotocols.com/2007/03/05/development_of_qsar_models_usi_1.php.
(17) Verloop, A. *The Sterimol Approach to Drug Design*; Marcel Dekker: New York, 1987.
(18) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Cross validation, Bootstrapping and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.
(19) Pogliani, L. Modeling with special descriptors derived from a medium-sized set of connectivity indices. *J. Phys. Chem.* **1996**, *100*, 18065–18077.
(20) Pogliani, L. From molecular connectivity indices to semiempirical connectivity terms: Recent trends in graph theoretical descriptors. *Chem. Rev.* **2000**, *100*, 3827–3858.
(21) Agrawal, V. K.; Singh, J.; Khadikar, P. V.; Supuran, C. T. QSAR study on topically acting sulfonamides incorporating GABA moieties: A molecular connectivity approach. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2044–2051.
(22) Selassie, C. D.; Kapur, S.; Verma, R. P.; Rosario, M. Cellular apoptosis and cytotoxicity of phenolic compounds: a quantitative structure-activity relationship study. *J. Med. Chem.* **2005**, *48*, 7234–7242.
(23) Verma, R. P.; Hansch, C. Cytotoxicity of organic compounds against ovarian cancer cells: A quantitative structure-activity relationship study. *Mol. Pharmaceutics* **2006**, *3*, 441–450.

equation includes 95% confidence limits for each term in parentheses.

Results and Discussion

1. 3′-*tert*-Butyl-3′-dephenyl Analogues of Paclitaxel (IV).[27] Ali and co-workers[27] studied the inhibition of bovine brain tubulin by a novel series of 3′-*tert*-butyl-3′-dephenylpaclitaxels (**IV**) in the microtubule-assembly assay. From the data in Table 1, we derived a parabolic correlation 3 in terms of $B5_X$ (Verloop's sterimol width parameter of X substituents). It suggests that the activities of paclitaxel derivatives (**IV**) first increase with an increase in Verloop's sterimol width parameter of X substituents up to an optimum $B5_X$ value of 4.56 and then decreases.



IV

$$\log 1/ED_{50} = 2.10(\pm 0.74)B5_X - 0.23(\pm 0.08)B5_X^2 + 1.68(\pm 1.51) \quad (3)$$

where $n = 10$, $r^2 = 0.878$, $s = 0.109$, $q^2 = 0.804$, $Q = 8.596$, and $F_{2,7} = 25.189$; optimum $B5_X = 4.56(4.34–4.75)$; outlier X = NHC_6H_5; and π_X vs $B5_X$; $r = 0.419$.

In the above QSAR model, ED_{50} is the molar concentration which causes polymerization of 50% of the tubulin present in 15 min at 37 °C. One derivative (X = NHC_6H_5) was considered to be an outlier because it was six times more active than expected to the standard deviation. Possible reasons for its unusual activity are not obvious, although its geometry due to the presence of a phenyl group may reduce the coplanarity with the X group and increase the activity.[28] But this reason is not sufficient because the other derivative (X = OC_6H_5) is very well predicted. It may be attributed to some experimental error. According to this QSAR (eq 3), the activities of paclitaxel derivatives (**IV**) are mainly dependent on the Verloop's sterimol width parameters of the X substituents and should be a maximum at $B5_X = 4.56$.

Table 1. Biological and Physicochemical Parameters Used To Derive QSAR 3 as Well as for That of Predicted Compounds

		log 1/ED_{50} (eq *3*)			
no.	X	obsd	pred	Δ	$B5_X$
IV-1	C_6H_5	5.77	5.82	−0.05	3.11
IV-2	$OC(CH_3)_3$	6.46	6.29	0.17	4.35
IV-3	$OCH_2CH(CH_3)_2$	6.30	6.29	0.01	4.42
IV-4	$O(CH_2)_3CH_3$	6.10	6.29	−0.19	4.79
IV-5	$O(CH_2)_5CH_3$	5.68	5.65	0.03	6.23
IV-6	$OC(CH_3)_2CH_2CH_3$	6.38	6.29	0.09	4.42
IV-7	$C(CH_3)_3$	5.87	5.86	0.01	3.17
IV-8	$CH2C(CH_3)_3$	6.30	6.27	0.03	4.18
IV-9	$NHC(CH_3)_3$	6.20	6.29	−0.09	4.39
IV-10	OC_6H_5	5.89	5.89	0.00	5.89
IV-11[a]	NHC_6H_5	6.48	5.85	0.63	5.95
IV-12[b]	$OCBr_3$	ND[c]	6.29	ND[c]	4.69
IV-13[b]	$N = CCl_2$	ND[c]	6.30	ND[c]	4.54
IV-14[b]	OCHBrCl	ND[c]	6.29	ND[c]	4.69
IV-15[b]	OCHBrF	ND[c]	6.29	ND[c]	4.69
IV-16[b]	$OCHBr_2$	ND[c]	6.29	ND[c]	4.69
IV-17[b]	SO_2CHCl_2	ND[c]	6.30	ND[c]	4.50
IV-18[b]	$CH=CCl_2$	ND[c]	6.30	ND[c]	4.52
IV-19[b]	CH_2CCl_3	ND[c]	6.30	ND[c]	4.50
IV-20[b]	$OCHBrCH_3$	ND[c]	6.29	ND[c]	4.69
IV-21[b]	$OCONHCH_3$	ND[c]	6.30	ND[c]	4.55
IV-22[b]	SCN	ND[c]	6.30	ND[c]	4.45
IV-23[b]	NH_2	ND[c]	4.75	ND[c]	1.97
IV-24[b]	NH(OH)	ND[c]	5.44	ND[c]	2.63
IV-25[b]	$NHNH_2$	ND[c]	5.72	ND[c]	2.97
IV-26[b]	SCF_3	ND[c]	6.21	ND[c]	3.94
IV-27[b]	OCN	ND[c]	6.23	ND[c]	4.01
IV-28[b]	$SCONHNH_2$	ND[c]	6.26	ND[c]	4.97
IV-29[b]	$CH_2C(NO_2)_3$	ND[c]	6.25	ND[c]	5.00
IV-30[b]	$SCH_2CH_2NO_2$	ND[c]	6.11	ND[c]	5.47
IV-31[b]	$CH=NNHCSNH_2$	ND[c]	6.13	ND[c]	5.41
IV-32[b]	SC_2H_5	ND[c]	6.22	ND[c]	3.97
IV-33[b]	$N(C_6H_5)_2$	ND[c]	5.85	ND[c]	5.95
IV-34[b]	$NHC_6H_3(3,4\text{-}Cl_2)$	ND[c]	5.23	ND[c]	6.71
IV-35[b]	$OC_6H_4(4\text{-}NO_2)$	ND[c]	4.55	ND[c]	7.31
IV-36[b]	SC_6H_5	ND[c]	5.50	ND[c]	6.42
IV-37[b]	$(CH_2)_4SC_2H_5$	ND[c]	5.06	ND[c]	6.87
IV-38[b]	$OC_6H_4(4\text{-}CF_3)$	ND[c]	4.32	ND[c]	7.48

[a] Not included in the derivation of QSAR 3. [b] Predicted compounds from QSAR 3. [c] Not determined.

2. C-4-Modified Paclitaxel Analogues (V).[29] Chen et al.[29] synthesized a series of C-4-modified paclitaxel analogues (**V**) and evaluated their biological activity in tubulin

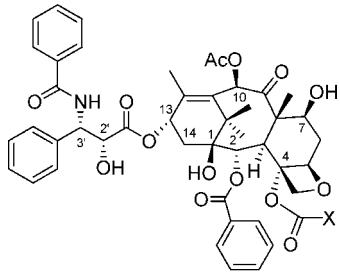
(24) Verma, R. P.; Hansch, C. Matrix metalloproteinases (MMPs): Chemical-biological functions and QSARs. *Bioorg. Med. Chem.* **2007**, *15*, 2223–2268.
(25) Verma, R. P.; Hansch, C. Understanding human rhinovirus infections in terms of QSAR. *Virology* **2007**, *359*, 152–161.
(26) Verma, R. P.; Hansch, C. Investigation of DNA-binding properties of organic molecules using quantitative structure-activity relationship (QSAR) models. *J. Pharm. Sci.* **2008,** *97*, 88–110.
(27) Ali, S. M.; Hoemann, M. Z.; Aube, J.; Mitscher, L. A.; Georg, G. I.; McCall, R.; Jayasinghe, L. R. Novel cytotoxic 3′-(tert-butyl) 3′-dephenyl analogs of paclitaxel and docetaxel. *J. Med. Chem.* **1995**, *38*, 3821–3828.
(28) Verma, R. P.; Kapur, S.; Barberena, O.; Shusterman, A.; Hansch, C. H.; Selassie, C. D. Synthesis, cytotoxicity, and QSAR analysis of X-thiophenols in rapidly dividing cells. *Chem. Res. Toxicol.* **2003**, *16*, 276–284.
(29) Chen, S.-H.; Wei, J.-M.; Long, B. H.; Fairchild, C. A.; Carboni, J.; Mamber, S. W.; Rose, W. C.; Johnston, K.; Casazza, A. M.; Kadow, J. F.; Farina, V.; Vyas, D. M.; Doyle, T. W. Novel C-4 paclitaxel (Taxol) analogs: potent antitumor agents. Bioorg. *Med. Chem. Lett.* **1995**, *5*, 2741–2746.

Table 2. Biological and Physicochemical Parameters Used To Derive QSAR 4

		log 1/ED_{50} (eq *4*)				
no.	X	obsd	pred	Δ	$C\pi_X$	CMR
V-1[a]	CH_3	6.03	6.94	−0.91	0.70	22.09
V-2	C_6H_5	ND[b]	4.47	ND[b]	1.91	24.14
V-3	4-F-C_6H_4	4.25	4.62	−0.37	2.05	24.16
V-4	CH_2F	5.95	6.51	−0.56	0.37	22.11
V-5	CCl_3	6.20	6.39	−0.19	2.58	23.57
V-6[a]	C_2H_5	5.86	6.69	−0.83	1.23	22.56
V-7	$CH{=}CH_2$	6.41	6.14	0.27	0.85	22.61
V-8	$(CH_2)_2CH_3$	6.25	6.44	−0.19	1.76	23.02
V-9	$CH(CH_3)_2$	6.37	6.18	0.19	1.54	23.02
V-10	$C(CH_3){=}CH_2$	6.18	5.62	0.56	1.16	23.07
V-11	*trans*-$CH{=}CHCH_3$	6.19	5.89	0.30	1.38	23.07
V-12	Cy-C_3H_5	6.65	6.13	0.52	1.28	22.88
V-13	$(CH_2)_3CH_3$	6.01	6.19	−0.18	2.28	23.48
V-14	Cy-C_4H_7	6.39	5.72	0.67	1.61	23.31
V-15	$(CH_2)_4CH_3$	5.50	5.95	−0.45	2.81	23.95
V-16	Cy-C_5H_9	5.84	5.50	0.34	2.17	23.77
V-17	OCH_3	6.42	6.63	−0.21	0.68	22.25
V-18	OCH_2CH_3	6.23	6.38	−0.15	1.21	22.71
V-19	$O(CH_2)_2CH_3$	6.15	6.13	0.02	1.74	23.17
V-20[a]	NH-Cy-C_4H_7	6.15	4.92	1.23	1.55	23.68
V-21	Imidazole	3.61	3.88	−0.27	0.18	23.36
V-22	Aziridine	5.58	5.89	−0.31	0.87	22.75

[a] Not included in the derivation of QSAR 4. [b] Not determined.

polymerization assay. The activity was given in the tubulin polymerization ratio ($ED_{50\,(analogues)}/ED_{50\,(paclitaxel)}$), which gives the activity with regard to paclitaxel itself (paclitaxel: ED_{50} = 0.93 μmol/L[27]). These values were converted into log 1/ED_{50} of the analogues (**V**) in molar concentration and are given in Table 2. We derived eq 4 from the data in Table 2.



V

$$\log 1/ED_{50} = 1.22(\pm 0.43)C\pi_X - 1.92(\pm 0.58)CMR + 48.53(\pm 12.87) \quad (4)$$

where n = 18, r^2 = 0.774, s = 0.395, q^2 = 0.628, Q = 2.228, and $F_{2,15}$ = 25.686; outliers X = CH_3; C_2H_5; NH-Cy-C_4H_7.

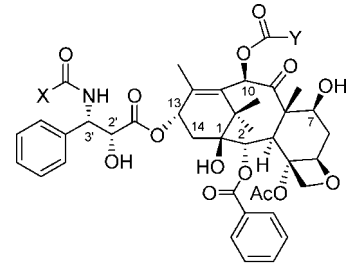
In this equation, $C\pi_X$ is the calculated hydrophobic parameter of the X substituents and CMR is the calculated molar refractivity of the whole molecules. Positive $C\pi_X$ suggests that tubulin polymerization activity of these molecule increases with the increase of hydrophobicity of the X substituents. On the contrary, the increase in the molar refractivity of the whole molecule (CMR) decreases the tubulin polymerization activity of these compounds (negative coefficient). Three compounds (X = CH_3; C_2H_5; NH-Cy-C_4H_7) were not used in the derivation of eq 4 due to their high deviation from the observed activity (obsd − pred > 2s). There is a high correlation between CMR and MR_X (r = 0.997), where MR_X is the calculated molar refractivity of X substituents. Thus, MR_X can replace CMR. By substituting MR_X for CMR in eq 4, we can develop eq 4a . Finally, we preferred eq 4 on the basis of their statistics, which is better than that of eq 4a.

$$\log 1/ED_{50} = 1.22(\pm 0.45)C\pi_X - 2.01(\pm 0.63)MR_X + 7.33(\pm 0.70) \quad (4a)$$

where n = 18, r^2 = 0.756, s = 0.410, q^2 = 0.569, Q = 2.122, and $F_{2,15}$ = 23.234; outliers X = CH_3; C_2H_5; NH-Cy-C_4H_7.

Thus, the tubulin polymerization activity of C-4-modified paclitaxel analogues (**V**) correlates with the hydrophobic parameter of X-substituents and the molar refractivity of the whole molecules.

3. C-3′NH/C-10-Modified Paclitaxel Analogues (VI).[30] A series of C-3′NH/C-10-modified paclitaxel analogues (**VI**) was synthesized and evaluated for their biological activities in a microtubule assembly by Baloglu et al.[30] Equation 5 was developed from the data in Table 3.



VI

$$\log 1/ID_{50} = -16.43(\pm 7.16)C\pi + 4.28(\pm 1.91)C\pi_X^2 + 3.54(\pm 1.22)MR_X + 13.22(\pm 4.11) \quad (5)$$

where n = 18, r^2 = 0.740, s = 0.312, q^2 = 0.559, Q = 2.756, and $F_{3,14}$ = 13.282; inversion point for $C\pi_X$ = 1.92(1.88 – 1.99); outliers X = $(CH_2)_4CH_3$, Y = $CH(CH_3)_2$; X = 2-furyl, Y = $CH(CH_3)_2$; $C\pi_X$ vs MR_X; r = 0.535.

This is an inverted parabolic correlation in terms of $C\pi_X$ (calculated hydrophobicity of X-substituents), which suggests that activity of these compounds first decreases as the hydrophobicity of the X-substituents increases and that after a certain point (inversion point; $C\pi_X$ = 1.92) activity begins to increase. This may correspond to an allosteric reac-

(30) Baloglu, E.; Hoch, J. M.; Chatterjee, S. K.; Ravindra, R.; Bane, S.; Kingston, D. G. I. Synthesis and biological evaluation of C-3′NH/C-10 and C-2/C-10 modified paclitaxel analogues. *Bioorg. Med. Chem.* **2003**, *11*, 1557–1568.

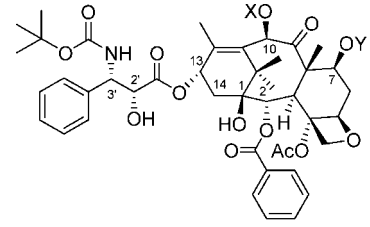
Table 3. Biological and Physicochemical Parameters Used To Derive QSAR 5

			log 1/ID_{50} (eq *5*)				
no.	X	Y	obsd	pred	Δ	$C\pi_X$	MR_X
VI-1	C_6H_5	CH_3	6.36	6.36	0.00	1.97	2.51
VI-2	C_6H_5	$CH(CH_3)_2$	6.85	6.36	0.49	1.97	2.51
VI-3	C_6H_5	$CH{=}CHCH_3$	6.27	6.36	−0.09	1.97	2.51
VI-4	C_6H_5	CH_2CH_3	6.15	6.36	−0.21	1.97	2.51
VI-5	C_6H_5	$CH_2CH_2CH_3$	6.17	6.36	−0.19	1.97	2.51
VI-6	$(CH2)_4CH_3$	CH_3	6.96	6.77	0.19	2.43	2.32
VI-7	$O(CH_2)_3CH_3$	CH_3	5.42	5.50	−0.08	2.39	2.01
VI-8	2-Furyl	CH_3	5.70	6.10	−0.40	1.15	1.73
VI-9[a]	$(CH2)_4CH_3$	$CH(CH_3)_2$	5.92	6.77	−0.85	2.43	2.32
VI-10	$(CH2)_4CH_3$	$CH{=}CHCH_3$	7.10	6.77	0.33	2.43	2.32
VI-11	$(CH2)_4CH_3$	CH_2CH_3	6.19	6.77	−0.58	2.43	2.32
VI-12	$(CH2)_4CH_3$	$CH_2CH_2CH_3$	6.85	6.77	0.08	2.43	2.32
VI-13	$O(CH_2)_3CH_3$	$CH(CH_3)_2$	5.69	5.50	0.19	2.39	2.01
VI-14	$O(CH_2)_3CH_3$	$CH{=}CHCH_3$	5.31	5.50	−0.19	2.39	2.01
VI-15	$O(CH_2)_3CH_3$	CH_2CH_3	5.79	5.50	0.29	2.39	2.01
VI-16	$O(CH_2)_3CH_3$	$CH_2CH_2CH_3$	5.30	5.50	−0.20	2.39	2.01
VI-17[a]	2-Furyl	$CH(CH_3)_2$	7.16	6.10	1.06	1.15	1.73
VI-18	2-Furyl	$CH{=}CHCH_3$	5.96	6.10	−0.14	1.15	1.73
VI-19	2-Furyl	CH_2CH_3	6.31	6.10	0.21	1.15	1.73
VI-20	2-Furyl	$CH_2CH_2CH_3$	6.44	6.10	0.34	1.15	1.73

[a] Not included in the derivation of QSAR 5.

tion.[15,31–33] MR_X is the calculated molar refractivity of X substituents, and its positive coefficient suggests that an increase in molar refractivity of X substituents should result in the improvement of activity. Thus, the activity of C-3′NH/C-10-modified paclitaxel analogues (**VI**) correlates with the hydrophobicity of X substituents (inverted parabolic fashion) as well as the molar refractivity of X substituents. There is no effect on the Y substituents. Two compounds in Table 3 were not used in the derivation of eq 5 due to their deviations from the observed activity (obsd − pred > 2*s*).

4. C-7- and/or C-10-Modified Docetaxel Analogues (VII).[34] For the purpose of understanding the role of lipophilicity on the binding of taxoids to microtubules, Guénard et al.[34] synthesized a series of docetaxel derivatives (**VII**) possessing alkyl side chains of different lengths at C-7 and/or C-10 in order to compare the contribution of hydrophilicity/hydrophobicity on tubulin binding. On the basis of their observations, the authors suggested that the presence of an alkyl chain at the C-7 and/or C-10 positions decrease the activity while their corresponding more polar analogues restore the activity of these molecules. In this case, with increasing lipophilicity, the decreasing activity could be the result of a destabilization of the drug−tubulin complex due to an unfavorable interaction between the hydrophobic chains and their hydrophilic environment or their aqueous vicinity. It appears that the recognition of taxoids by tubulin depends on the location of the most important hydrophobic area.[34] The binding of taxoids (**VII**) to microtubules was evaluated by microtubule disassembly assay, and the results are given in the ratio ($ID_{50\ (analogues)}/ID_{50\ (paclitaxel)}$), which gives the activity with regard to paclitaxel itself (paclitaxel: ID_{50} = 0.4 μmol/L[35]). These values were converted into log 1/ID_{50} of the analogues (**VII**) in molar concentration and are given in Table 4. We derived eq 6 from these data in Table 4.



VII

$$\log 1/ID_{50} = -1.35(\pm 0.57)\pi_X + 0.36(\pm 0.15)\pi_X^2 - 0.20(\pm 0.13)MR_Y + 6.71(\pm 0.45) \quad (6)$$

where n = 13, r^2 = 0.775, s = 0.202, q^2 = 0.578, Q = 4.356, and $F_{3,9}$ = 10.333; inversion point for π_X = 1.90(1.68

(31) Verma, R. P. Understanding topoisomerase I and II in terms of QSAR. *Bioorg. Med. Chem.* **2005**, *13*, 1059–1067.

(32) Verma, R. P. An approach towards the quantitative structure-activity relationships for sulfamate-based estrone sulfatase inhibitors. *Lett. Drug Design Discov.* **2005**, *2*, 205–218.

(33) Mekapati, S. B.; Kurup, A.; Verma, R. P.; Hansch, C. The role of hydrophobic properties of chemicals in promoting allosteric reactions. *Bioorg. Med. Chem.* **2005**, *13*, 3737–3762.

(34) Guénard, D.; Thoret, S.; Dubois, J.; Adeline, M.-T.; Wang, Q.; Guéritte, F. Effects of the hydrophobicity of taxoids on their interaction with tubulin. *Bioorg. Med. Chem.* **2000**, *8*, 145–156.

(35) Guéritte-Voegelein, F.; Guénard, D.; Lavelle, F.; Le Goff, M.-T.; Mangatal, L.; Potier, P. Relationships between the structure of taxol analogs and their antimitotic activity. *J. Med. Chem.* **1991**, *34*, 992–998.

Table 4. Biological and Physicochemical Parameters Used To Derive QSAR 6

			log 1/ID_{50} (eq *6*)				
no.	X	Y	obsd	pred	Δ	π_X	MR_Y
VII-1	H	$CO(CH_2)_2CH_3$	6.06	6.31	−0.25	0.00	1.98
VII-2	H	$CO(CH_2)_5CH_3$	6.12	6.03	0.09	0.00	3.37
VII-3	H	$CO(CH_2)_6CH_3$	6.12	5.93	0.19	0.00	3.83
VII-4	H	$CO(CH_2)_7CH_3$	6.00	5.84	0.16	0.00	4.30
VII-5	$CO(CH_2)_2CH_3$	H	6.14	6.09	0.05	0.51	0.10
VII-6[a]	$CO(CH_2)_5CH_3$	H	6.08	5.42	0.66	2.10	0.10
VII-7[a]	$CO(CH_2)_6CH_3$	H	6.08	5.59	0.49	2.62	0.10
VII-8	$CO(CH_2)_7CH_3$	H	6.10	5.96	0.14	3.15	0.10
VII-9	$COCH_2CH_3$	$COCH_2CH_3$	6.55	6.31	0.24	0.06	1.58
VII-10	$CO(CH_2)_2CH_3$	$CO(CH_2)_2CH_3$	5.44	5.71	−0.27	0.51	1.98
VII-11	$CO(CH_2)_3CH_3$	$CO(CH_2)_3CH_3$	5.25	5.20	0.05	1.04	2.44
VII-12	H	$COC{\equiv}CC_6H_5$	5.88	5.89	−0.01	0.00	4.04
VII-13	H	$COCH{=}CHC_6H_5$	5.59	5.86	−0.27	0.00	4.22
VII-14	COC_6H_4-4-C_6H_5	H	6.40	6.48	−0.08	3.64	0.10
VII-14	$COCH{=}CHC_6H_5$	H	5.70	5.72	−0.02	2.84	0.10

[a] Not included in the derivation of QSAR 6.

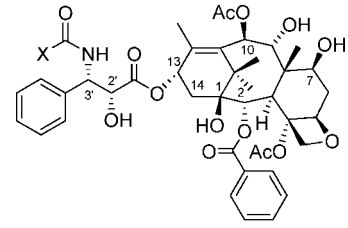
– 2.16); outliers X = $CO(CH_2)_5CH_3$, Y = H; X = $CO(CH_2)_6CH_3$, Y = H.

Similar to eq 5, this is also an inverted parabolic correlation in terms of π_X (measured hydrophobicity of X substituents), with an inversion point of $\pi_X = 1.90$. MR_Y is the calculated molar refractivity of the Y substituents, and its negative sign brings out a steric effect for the Y substituents that does not appear to reach a hydrophobic surface. Thus, the result of our QSAR (eq 6) is not in good agreement with the findings of Guénard et al.[34] Equation 6 suggests that the binding of taxoids (**VII**) to microtubules correlates with hydrophobic descriptors of X substituents (inverted parabolic fashion) followed by molar refractivity of Y substituents. Two compounds in Table 4 were deemed to be outliers on the basis of their deviations from the observed activity (obsd − pred > 2*s*).

5. C-3′-N-Acyl Analogues of 9(*R*)-Dihydrotaxol (VIII).[36] Maring et al.[36] synthesized a series of C-3′-*N*-acyl analogues of 9(*R*)-dihydrotaxol (**VIII**) from 7-triethylsilyl-9(*R*)-dihydrobaccatin III and the corresponding (3*R*,4*S*)-*N*-acyl-3-(1-ethoxyethoxy)-4-phenylazetidin-2-ones. These analogues were tested in a microtubule assembly assay and the results given in the ratio ($ED_{50\ (analogues)}/ED_{50\ (paclitaxel)}$), which gives the activity with regard to paclitaxel itself (paclitaxel: $ED_{50} = 0.93\ \mu mol/L$[27]). These values were converted into log 1/ED_{50} of the analogues (**VIII**) in molar concentration and are given in Table 5. We derived eq 7 from the data in Table 5.

$$\log 1/ED_{50} = 0.70(\pm 0.23)C\pi_X - 0.16(\pm 0.06)C\pi_X^2 + 5.35(\pm 0.23) \quad (7)$$

where $n = 13$, $r^2 = 0.826$, $s = 0.080$, $q^2 = 0.630$, $Q = 11.363$, and $F_{2,10} = 23.736$; optimum $C\pi_X = 2.19(1.97 – 2.50)$; outlier X = 2-furyl; $C(CH_3)_3$.



VIII

This is a parabolic correlation in terms of $C\pi_X$ (calculated hydrophobicity of X substituents), which suggests that the activity of compounds (**VIII**) first increases with an increase in the hydrophobicity of the X substituents up to an optimum $C\pi_X$ of 2.19 and then decreases. Thus, the activities of compounds (**VIII**) are mainly dependent on the hydrophobic descriptors of X substituents and should be a maximum at $\pi_X = 2.19$. Two compounds in Table 5 were not used in the derivation of eq 7 due to their deviations from the observed activity (obsd − pred > 2*s*).

6. C-3′-Modified Analogues of 9(*R*)-Dihydrotaxol (IX).[37] The synthesis and biological evaluation of a series of C-3′-modified analogues of 9(*R*)-dihydrotaxol (**IX**) was described by Li and co-workers.[37] This study revealed that the C-3′-phenyl ring was not required for the activity and identified several compounds which had equal or greater *in vitro* and *in vivo* activity than Taxol. The tubulin assembly activity data for these compounds was reported as in the ratio ($ED_{50\ (analogues)}/ED_{50\ (paclitaxel)}$), which gives the activity with regards to paclitaxel itself (paclitaxel: $ED_{50} = 0.70\ \mu mol/$

(36) Maring, C. J.; Grampovnik, D. J.; Yeung, C. M.; Klein, L. L.; Li, L.; Thomas, S. A.; and Plattner, J. J. C-3′-N-acyl analogs of 9(R)-dihydrotaxol: synthesis and structure activity relationships. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 1429–1432.

(37) Li, L.; Thomas, S. A.; Klein, L. L.; Yeung, C. M.; Maring, C. J.; Grampovnik, D. J.; Lartey, P. A.; Plattner, J. J. Synthesis and biological evaluation of C-3′-modified analogs of 9(R)-dihydrotaxol. *J. Med. Chem.* **1994**, *37*, 2655–2663.

Table 5. Biological and Physicochemical Parameters Used To Derive QSAR 7 as Well as for That of Predicted Compounds

		log 1/ED_{50} (eq 7)			
no.	X	obsd	pred	Δ	$C\pi_X$
VIII-1	C_6H_5	6.15	6.11	0.04	1.97
VIII-2	4-CH_3-C_6H_4	6.16	6.10	0.06	2.47
VIII-3	4-OH-C_6H_4	6.00	6.07	−0.07	1.64
VIII-4[a]	2-Furyl	6.17	5.94	0.23	1.15
VIII-5	CH_3	5.53	5.55	−0.02	0.31
VIII-6[a]	$C(CH_3)_3$	5.66	6.05	−0.39	1.55
VIII-7	$CH_2C(CH_3)_3$	6.02	6.12	−0.10	2.17
VIII-8	$NHC(CH_3)_3$	6.01	6.09	−0.08	1.75
VIII-9	$OC(CH_3)_3$	6.09	6.11	−0.02	2.04
VIII-10	$OC(CH_3)_2CH_2CH_3$	6.17	6.09	0.08	2.57
VIII-11	$OCH(CH_3)_2$	6.16	6.07	0.09	1.64
VIII-12	$OCH_2C(CH_3)_3$	6.12	6.08	0.04	2.66
VIII-13	O-Adamantyl	5.84	5.86	−0.02	3.46
VIII-14	$OCH_2CH(CH_3)_2$	6.02	6.12	−0.10	2.26
VIII-15	OC_2H_5	6.12	6.00	0.12	1.33
VIII-16[b]	C_5H_9	ND[c]	6.12	ND[c]	2.14[d]
VIII-17[b]	OC_6H_5	ND[c]	6.12	ND[c]	2.08
VIII-18[b]	SC_6H_5	ND[c]	6.12	ND[c]	2.32[d]
VIII-19[b]	$CH_2N(CH_3)C_6H_5$	ND[c]	6.12	ND[c]	2.09[d]
VIII-20[b]	$N(CH_3)CH_2C_6H_5$	ND[c]	6.12	ND[c]	2.05
VIII-21[b]	2-Benzthiazolyl	ND[c]	6.12	ND[c]	2.13[d]
VIII-22[b]	$CH_2OC_6H_5$	ND[c]	6.12	ND[c]	2.07
VIII-23[b]	OCH_3	ND[c]	5.77	ND[c]	0.71[d]
VIII-24[b]	C_2F_5	ND[c]	6.10	ND[c]	1.89[d]
VIII-25[b]	C_2H_5	ND[c]	5.90	ND[c]	1.02[d]
VIII-26[b]	SC_2H_5	ND[c]	5.92	ND[c]	1.07[d]
VIII-27[b]	$CH(CH_3)_2$	ND[c]	5.97	ND[c]	1.22[d]
VIII-28[b]	$Si(CH_3)_3$	ND[c]	6.09	ND[c]	2.59[d]
VIII-29[b]	$CH_2CH_2CH_2CH_3$	ND[c]	6.12	ND[c]	2.13[d]
VIII-30[b]	C_6H_{11}	ND[c]	6.06	ND[c]	2.76[d]
VIII-31[b]	$NHCH_2C_6H_5$	ND[c]	5.89	ND[c]	1.00[d]
VIII-32[b]	$CH{=}CHC_6H_5$	ND[c]	6.08	ND[c]	2.68[d]
VIII-33[b]	4-C_6H_5-C_6H_4	ND[c]	5.53	ND[c]	4.09[d]
VIII-34[b]	$N(C_6H_5)_2$	ND[c]	5.84	ND[c]	3.50[d]

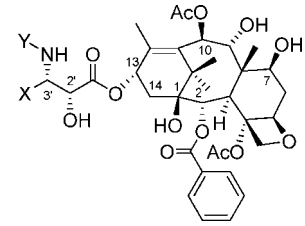
[a] Not included in the derivation of QSAR 7. [b] Predicted compounds from QSAR 7. [c] Not determined. [d] Measured value.

L[38]). These values were converted into log 1/ED_{50} of the analogues (**IX**) in molar concentration and are given in Table 6. Equation 8 was developed from the data in Table 6.

$$\log 1/ED_{50} = 0.19(\pm 0.10)\pi_X - 0.25(\pm 0.11)L_X + 7.14(\pm 0.53) \quad (8)$$

where $n = 11$, $r^2 = 0.800$, $s = 0.162$, $q^2 = 0.520$, $Q = 5.519$, and $F_{2,8} = 16.000$; outliers X = CH_3, Y = OCOC-$(CH_3)_3$; π_X vs L_X; $r = 0.578$.

π_X and L_X are the hydrophobic and Verloop's sterimol length parameters for X substituents. The positive coefficient



IX

of π_X suggests that, at all the parts where substituents have been entered, hydrophobic contacts have been made. Thus, the molecules with highly hydrophobic X substituents will be more active. A negative sign associated with L_X brings out a steric effect for these substituents. No term appears for the Y substituents; it may be due to the fact that all of the analogues (except two) have identical substituents at the Y position. One compound in Table 6 was deemed to be an outlier on the basis of its deviation from the observed activity (obsd − pred > 2*s*). QSAR (eq 8) results support the findings of Li and co-workers;[37] i.e., the C-3′-phenyl ring is not required for the activity but the hydrophobic and Verloop's sterimol length parameters of the X-substituents play a critical role.

A comparison of the statistics of QSARs 3–8 obtained from multiregression analyses (MRA) is shown in Table 7. All of the QSAR are found to be statistically significant.

Validation of QSAR Models

QSAR model validation becomes an essential part in the development of a statistically valid and predictive model because the real utility of a QSAR model is in its ability to predict accurately the modeled property for new compounds. The following approaches have been used for the validation of QSARs 3–8:

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 QSAR models are from 0.740 to 0.878, which suggests that these QSAR models explain 74.0–87.8% of the variance in the data. According to the literature, the predictive QSAR model must have $r^2 > 0.6$.[39,40]

Cross-Validation Test. The values of q^2 for these QSAR models are from 0.520 to 0.804. The high values of q^2 validate these QSAR models. According to the literature, the predictive QSAR model must have $q^2 > 0.5$.[39,40]

Standard Deviation (*s*). *s* is the standard deviation about the regression line. The smaller the value of *s* the better the QSAR model. The values of *s* for these QSAR models are from 0.080 to 0.395.

Quality Factor (*Q*). Chance correlation, due to the excessive number of parameters, is detected by the examina-

(38) Georg, G. I.; Cheruvallath, Z. S.; Himes, R. H.; Mejillano, M. R.; Burke, C. T. Synthesis of biologically active taxol analogues with modified phenylisoserine side chains. *J. Med. Chem.* **1992**, *35*, 4230–4237.

(39) Golbraikh, A.; Tropsha, A. Beware of q^2. *J. Mol. Graph. Modell.* **2002**, *20*, 269–276.

(40) Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.

Table 6. Biological and Physicochemical Parameters Used To Derive QSAR 8

			log 1/ED_{50} (eq *8*)				
no.	X	Y	obsd	pred	Δ	π_X	L_X
IX-1	$CH(OH)CH_2OH$	$OCOC(CH_3)_3$	5.72	5.56	0.16	−1.93	4.79
IX-2	$CH(OH)CH_3$ (*R* or *S*)	$OCOC(CH_3)_3$	6.25	6.15	0.10	0.26	4.11
IX-3	$CH(OH)CH_3$ (*S* or *R*)	$OCOC(CH_3)_3$	6.08	6.15	−0.07	0.26	4.11
IX-4	CH_2OCH_3	$CH_2C_6H_5$	5.66	5.89	−0.23	−0.21	4.78
IX-5	$CH_2OC_6H_5$	$CH_2C_6H_5$	5.39	5.39	0.00	1.66	8.19
IX-6[a]	CH_3	$OCOC(CH_3)_3$	6.12	6.53	−0.41	0.56	2.87
IX-7	$CH{=}CH_2$	$OCOC(CH_3)_3$	6.19	6.22	−0.03	0.82	4.29
IX-8	CH_2CH_3	$OCOC(CH_3)_3$	6.37	6.30	0.07	1.02	4.11
IX-9	$(CH_2)_3CH_3$	$OCOC(CH_3)_3$	5.82	5.99	−0.17	2.13	6.17
IX-10	$(CH_2)_4CH3$	$OCOC(CH_3)_3$	5.85	5.89	−0.04	2.63	6.97
IX-11	$CH_2CH(CH_3)_2$	$OCOC(CH_3)_3$	6.18	6.23	−0.05	1.70	4.92
IX-12	C_6H_{11}	$OCOC(CH_3)_3$	6.40	6.13	0.27	2.82	6.17

[a] Not included in the derivation of QSAR 8.

Table 7. Comparison of the Statistics Obtained from the Multiregression Analyses (MRA) Process for QSARs 3–8

			descriptors coefficient							
QSAR no.	compd	*n*	hydrophobic	steric/pol.	r^2	q^2	*s*	*Q*	*F*	intercept
3	**IV**	10		2.10 $B5_X$ − 0.23 $B5_X{}^2$	0.878	0.804	0.109	8.596	25.189	1.68
4	**V**	18	1.22 $C\pi_X$	−1.92 CMR	0.774	0.628	0.395	2.228	25.686	48.53
5	**VI**	18	−16.43 $C\pi_X$ + 4.28 $C\pi_X{}^2$	3.54 MR_X	0.740	0.559	0.312	2.756	13.282	13.22
6	**VII**	13	−1.35 π_X + 0.36 $\pi_X{}^2$	−0.20 MR_Y	0.775	0.578	0.202	4.356	10.333	6.71
7	**VIII**	13	0.70 $C\pi_X$ − 0.16 $C\pi_X{}^2$		0.826	0.630	0.080	11.363	23.736	5.35
8	**IX**	11	0.19 π_X	−0.25 L_X	0.800	0.520	0.162	5.519	16.000	7.14

Table 8. Y-Randomization Data for QSARs 3–8

	NOR-1[a]		NOR-2		NOR-3		NOR-4		NOR-5	
QSAR no.	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2
3	0.078	−3.939	0.187	−0.408	0.163	−0.520	0.303	−0.782	0.057	−0.730
4	0.228	−0.219	0.226	−0.102	0.225	−0.045	0.307	0.094	0.257	−0.211
5	0.288	−0.183	0.486	0.196	0.259	−0.296	0.294	−0.159	0.267	−0.183
6	0.318	−0.329	0.102	−1.062	0.459	0.021	0.508	0.115	0.102	−0.682
7	0.096	−0.423	0.135	−0.771	0.169	−1.649	0.089	−0.280	0.246	−4.059
8	0.387	−0.541	0.341	−0.124	0.455	−0.014	0.439	−0.201	0.347	−0.297

[a] NOR = number of Y-randomization.

tion of the *Q* value.[19–21] High values of *Q* (2.228–11.363) for these QSAR models suggest their high predictive power.

Fischer Statistics (*F*). Fischer statistics (*F*) is the ratio between explained and unexplained variance for a given number of degree of freedom. The larger the value of *F* the greater the probability that the QSAR model is significant. The *F* values for these QSAR models are from 10.333 to 25.686, 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 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 8 (eqs 3–8). The poor values of r^2 and q^2 in the Y-randomization test ensure the robustness for these QSAR models.[22,40–42]

Lack of Overfitting. A model overfits if it includes more descriptors than required. The lack of overfitting for all the QSAR models (eqs 3–8) was confirmed by using the following conditions: (i) Number of data points/number of descriptors ≥ 4. (ii) High values of *Q* for these QSAR models indicate the lack of overfitting. (iii) All of the QSARs 3–8 were checked for their correlation with a fewer number of descriptors than that of the original. None of these QSAR

(41) Wold, S.; Eriksson, L. In *Chemometrics Methods in Molecular Design: Statistical Validation of QSAR Results*; van de Waterbeemd, H., Ed.; VCH: Weinheim, Germany, 1995; pp 309–318.

(42) Melzig, M. F.; Tran, G. D.; Henke, K.; Selassie, C. D.; Verma, R. P. Inhibition of neutrophil elastase and thrombin activity by caffeic acid esters. *Pharmazie* **2005**, *60*, 869–873.

Table 9. QSAR Obtained by the Use of Fewer Descriptors than That of the Original QSAR with Their Statistical Parameters

eq no.	NSQ[a]	QSAR with fewer descriptors than that of the original QSAR	*n*	r^2	q^2	*s*
3	3a	log 1/ED_{50} = −0.05(±0.22)$B5_X$ + 6.31(±1.02)	10	0.031	−0.843	0.287
4	4b	log 1/ED_{50} = 0.13(±0.56)Cπ_X + 5.71(±0.92)	18	0.015	−0.430	0.798
	4c	log 1/ED_{50} = −0.68(±0.67)CMR + 21.57(±15.46)	18	0.224	0.013	0.708
5	5a	log 1/ID_{50} = −0.01(±0.58)Cπ_X + 6.18(±1.20)	18	0.000	−0.234	0.572
	5b	log 1/ID_{50} = 0.84(±0.85)MR_X + 4.34(±1.85)	18	0.215	0.037	0.507
	5c	log 1/ID_{50} = −0.40(±0.59)Cπ_X + 1.20(±0.98)MR_X + 4.38(±1.81)	18	0.309	0.062	0.491
	5d	log 1/ID_{50} = 1.66(±6.56)Cπ_X − 0.47(±1.83)Cπ_X^2 + 4.83(±5.43)	18	0.019	−0.304	0.585
6	6a	log 1/ID_{50} = 0.02(±0.18)π_X + 5.93(±0.29)	13	0.008	−0.393	0.384
	6b	log 1/ID_{50} = −0.06(±0.14)MR_Y + 6.07(±0.38)	13	0.062	−0.205	0.374
	6c	log 1/ID_{50} = −0.06(±0.28)π_X − 0.09(±0.23)MR_Y + 6.21(±0.74)	13	0.086	−0.476	0.387
	6d	log 1/ID_{50} = −0.87(±0.70)π_X + 0.27(±0.21)π_X^2 + 6.06(±0.25)	13	0.458	0.209	0.298
7	7a	log 1/ED_{50} = 0.10(±0.14)Cπ_X + 5.83(±0.30)	13	0.179	−1.255	0.166
8	8a	log 1/ED_{50} = 0.05(±0.17)π_X + 5.94(±0.29)	11	0.046	−0.342	0.334
	8b	log 1/ED_{50} = −0.14(±0.15)L_X + 6.73(±0.80)	11	0.341	0.011	0.277

[a] NSQ = number of QSAR models obtained by the use of less number of descriptors than that of the original QSAR.

was found to be statistically significant (Table 9). (iv) Y-Randomization test (Table 8) suggests that the high r^2 of the QSARs 3–8 is not due to a chance correlation or overfitting.[43]

New Molecule Prediction

QSARs 3 and 7 are the parabolic correlations in terms of $B5_X$ (Verloop's sterimol width parameter of X substituents) and Cπ_X (calculated hydrophobicity of X substituents), respectively, where the optimum value of $B5_X$ and Cπ_X are well-defined. We believe that these two equations may prove to be adequate predictive models that can help to provide guidance in design/synthesis and subsequently yield very specific compounds (**IV** and **VIII**) that may have high biological activities and less side effects. On the basis of these models, 18 compounds (**IV-12**−**IV-22**, Table 1, and **VIII-16**−**VIII-22**, Table 5) are suggested as potential synthetic targets. The predicted log 1/ED_{50} of these compounds obtained from QSAR 3 and 7 along with their physicochemical parameters are given in Tables 1 and 5, respectively.

The modified compounds (**IV-12**−**IV-22**) in Table 1 have nearly identical predictive activities, even though there is a significant variation among the new X substituents. This was expected because all of the compounds (**IV-12**−**IV-22**) have $B5_X$ values (*B*5 of X substituents) from 4.45 to 4.69, which is in the range of the optimum value of $B5_X$ = 4.56(4.34 – 4.75) for the parabolic correlation 3. The identical predictive activities of the compounds (**IV-12**−**IV-22**) may confuse the predictive ability of this model (eq 3). Since a good QSAR model will also correctly predict the activity of a poor compound, we decided to add 16 more compounds (**IV-23**−**IV-38**) in Table 1 with a large variation in their $B5_X$ values, i.e., from 1.97 to 7.48, which gave the expected predictive activities from eq 3, i.e., log 1/ED_{50} values (4.75 – 6.26 – 4.32). A plot between the predicted log 1/ED_{50} and $B5_X$ of all of the compounds of this series (**IV-1**−**IV-38**) is shown in Figure 1, which is a good parabolic plot suggesting the predictive ability of this model.

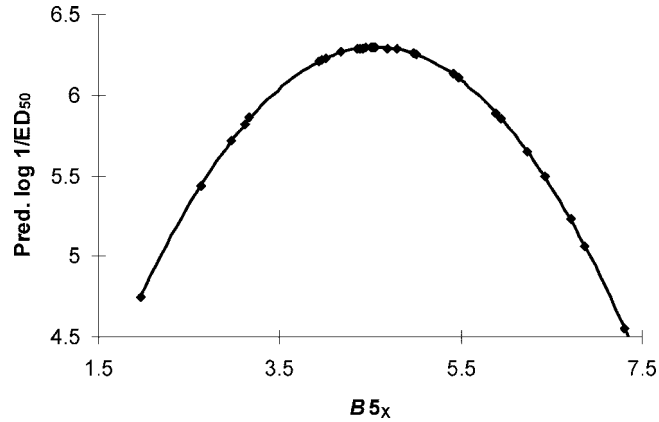


Figure 1. Plot of predicted log 1/ED_{50} vs $B5_X$ from Table 1 (eq 3).

Similarly, the modified compounds (**VIII-16**−**VIII-22**) in Table 5 have identical predictive activities, even though there is a significant variation among the new X substituents. This was also expected because all of the compounds (**VIII-16**−**VIII-22**) have Cπ_X values (calculated hydrophobic descriptors of X-substituents) from 2.05 to 2.32, which is in the range of the optimum value of Cπ_X = 2.19(1.97 – 2.50) for the parabolic correlation 7. To avoid confusion about the predictive ability of this model, we decided to add 12 more compounds (**VIII-23**−**VIII-34**) in Table 5 with a wide variation in their Cπ_X values, i.e., from 0.71 to 4.09, which gave the expected predicted activities from eq 7, i.e., log 1/ED_{50} values (5.77 – 6.12 – 5.53). A plot between predicted log 1/ED_{50} and Cπ_X of all of the compounds of this series (**VIII-1**−**VIII-34**) is shown in Figure 2, which is a good parabolic plot suggesting the predictive ability of this model.

(43) Zhang, S.; Golbraikh, A.; Oloff, S.; Kohn, H.; Tropsha, A. A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J. Chem. Inf. Model.* **2006**, *46*, 1984–1995.

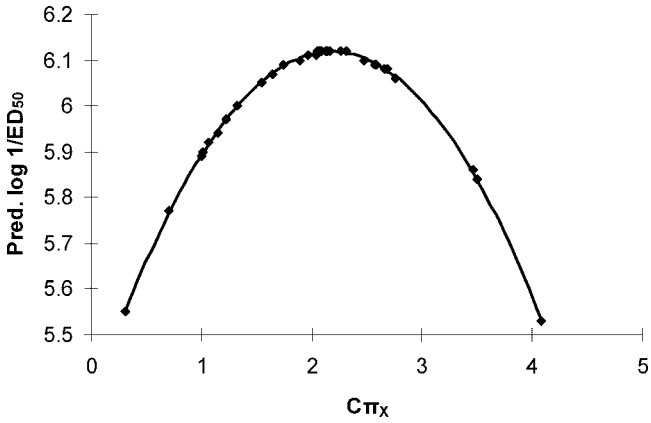


Figure 2. Plot of predicted log $1/ED_{50}$ vs $C\pi_X$ from Table 5 (eq 7).

Overview

An analysis of our QSAR models (eqs 38) reveals a number of interesting points. The most important of these is the hydrophobic parameter of the substituent. Out of 6 QSAR, 5 contain a correlation between activity and hydrophobicity of the substituent. A positive linear correlation is found in two equations (eqs 4 and 8). The coefficients associated with these hydrophobic parameters are 1.22 (eq 4) and 0.19 (eq 8). These data suggest that activity might be improved by increasing substituent hydrophobicity. Parabolic correlations with hydrophobic parameter of the substituents are found in three equations (eqs 5,6,and 7). Two of these (eqs 5 and 6) reflect situations where activity declines with increasing hydrophobicity of the substituents and then changes direction and increases. This may correspond to an allosteric reaction. The other (eq 7) situation shows that the activity is optimal for a particular value, or range of values, of π. The optimal value of π for this equation is 2.19.

The second important parameter is the molar refractivity of the molecules/substituents, which is present in 3 QSAR. A positive linear correlation is found in one equation (eq 5) with coefficient 3.54, which suggests that the activity might be improved by increasing substituent molar refractivity/polarizability. A negative linear correlation is found in two equations (eqs 4 and 6), and the coefficients are −1.92 (eq 4) and −0.20 (eq 6). Less steric congeners in these compound families might display enhanced activity. Parabolic correlation with Verloop's sterimol width parameter of the substituents is found in one equation (eq 3), which shows that the activity is optimal for a particular value, or range of values, of *B*5. The optimal value of *B*5 for this equation is 4.56. Verloop's sterimol length parameter of the substituents is also present in one equation (eq 8) with a negative coefficient (−0.25), which suggests the steric effects of the substituent.

Conclusion

The microtubule-stabilizing agents paclitaxel and docetaxel (progenitors of the family of taxanes) are the most successful anticancer drugs currently used in clinics. However, both drugs are associated with notorious side effects, drug resistance, and cross-resistance with other chemotherapeutic agents. These limitations have led to the search for new drugs with improved biological activity. The QSAR paradigm may be helpful in the design and development of novel tubulin-binding molecules as new anticancer agents, which will be expected to achieve improvements in anticancer activity, toxicity profile, pharmacology, and drug formulation. Our QSAR results suggest that the hydrophobic parameter of the substituents (π) is one of the most important determinants of the activity followed by steric parameters. Parabolic correlations (eqs 3 and 7) with *B*5 and π are found to be the most encouraging examples, where the optimum values of these parameters are well-defined. We believe that these two equations may prove to be adequate predictive models that can help provide guidance in design/synthesis and subsequently yield very specific compounds (**IV** and **VIII**) that may have high biological activities and less toxicity. On the basis of these two QSAR models, 18 compounds (**IV-12**−**IV-22** and **VIII-16**−**VIII-22**) are suggested as potential synthetic targets.

MP700119E