DOI: 10.1002/cmdc.200700138

20-(*S*)-Camptothecin Analogues as DNA Topoisomerase I Inhibitors: A QSAR Study

Corwin Hansch and Rajeshwar P. Verma*[a]

The interest in the application of the quantitative structure–activity relationships (QSAR) has steadily increased in recent decades because it has repeatedly proven itself to be a low-cost, high-return investment. 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. We hope it may also be useful in the design and development of new camptothecin derivatives as DNA topoisomerase I (topo I) inhibitors. In this paper, two series of camptothecin derivatives were undertaken to correlate DNA topo I inhibition with their hydrophobic and steric properties, to understand their chemical–biological interactions. The resulting QSAR have shown that the inhibitory activity of camptothecin analogues ***4*** *toward DNA topo I is mainly dependent on their hydrophobic and steric descriptors, whereas the same activity of 10,11-methylenedioxy- camptothecin analogues* ***5*** *is largely dependent on their hydrophobicity at position-7. Using internal [cross-validation, quality factor (Q), Fischer statistics (F), and Y-randomization tests] and external validation tests both of these QSAR models have been validated. Both series of these camptothecin derivatives are also filtered by Lipinski's rule of five to check their oral bioavailability. On the basis of these QSAR models, five compounds (****4-35****,* ***4-36****,* ***5-20****,* ***5-21****, and* ***5-22****) have been predicted that may be the next synthetic target. These molecules also fulfill the conditions of Lipinski's rule of five.*

Introduction

Camptothecin (CPT, **1**), a pentacyclic alkaloid isolated from a native tree of Tibet and China called *Camptotheca acuminata* in latin and xi shu in Chinese, is one of the prominent lead compounds in anticancer drug development.[1,2] It has already been discovered that the cytotoxicity of camptothecin is due to a novel mechanism of action involving selective inhibition of DNA topoisomerase I (topo I).[3,4] This has led to the emergence of CPT analogues as a new class of antitumor agents with impressive activity.[5,6] DNA topoisomerases are ubiquitous enzymes that can manipulate DNA by changing the number of topological links between two strands of the same or different DNA molecules.[7] These enzymes are involved in many cellular processes, such as replication, recombination, transcription, and chromosome segregation. There are two major classes of topoisomerases, topo I, which break and reseal one strand of DNA, and topo II that alter DNA topology by catalyzing the passage of an intact DNA double helix through a transient double-stranded break made in a second helix.[8]

DNA topo I is an essential human enzyme, which can be trapped by anticancer drugs as it cleaves DNA. Moreover, topo I can be trapped by endogenous alterations to DNA (mismatches, abasic sites, nicks, and adducts) and apoptotic alterations to chromatin. Camptothecin is a natural product of which topo I is the only cellular target. Recently, two anticancer camptothecin derivatives (Figure 1), topotecan (for ovarian and lung cancers) and irinotecan (for colorectal cancer) have already been approved by the US Food and Drug Administration.[9] These drugs bind to a transient topo I–DNA covalent complex and inhibit the resealing of a single-strand break that the enzyme creates to relieve superhelical tension in the duplex DNA.[10]

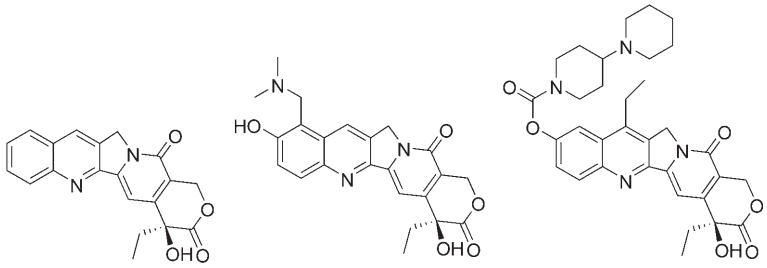


Figure 1. Camptothecin and its derivatives as anticancer drugs.

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

In a recent study on the inhibition of topo I by a series of camptothecin analogues, the quantitative structure–activity relationship (QSAR) described by Equation (1) was obtained.[11]

$$\log 1/C = 0.43\mathrm{Clog}P - 0.43\sigma^{+} + 1.11I - 0.89\mathrm{MR}_9 + 6.37 \quad (1)$$

$n = 17$, $r^2 = 0.862$, $s = 0.226$
$q^2 = 0.681$, $Q = 4.108$, $F_{4,12} = 18.739$

In this equation, C is the molar concentration of camptothecin analogues that cause 50% inhibition of topo I. Clog$P$ is the calculated partition coefficient of each compound in *n*-octanol/water system and is a measure of hydrophobicity of the whole molecule. σ^{+} is the sum of the Hammett parameters of substituents at positions 9, 10, and 11. It represents a substituent ability to delocalize a positive charge. Indicator variable I takes the value of 1 for the presence of 10-OCH_2O-11 substituent and zero for the other substituents. MR_9 is the calculated molar refractivity of substituents at position 9. The number of data points in the study is represented by n, the goodness of fit by r^2, the standard deviation by s, and the cross-validated r^2 by q^2.

In this paper, we developed QSAR models on two series of camptothecin derivatives with respect to their inhibitory activities toward topo I. In the past 44 years, the use of QSAR (one of the well-developed areas in computational chemistry)[12] has become increasingly helpful in understanding many aspects of chemical–biological interactions in drug and pesticide research and in 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 humans.[13,14] It has also been utilized for the evaluation of absorption, distribution, metabolism, and excretion (ADME) phenomena in many organisms and in whole animal studies.[15] The QSAR approach employs extrathermodynamically derived and computational-based descriptors to correlate biological activity in isolated receptors, cellular systems, and in vivo. Four standard molecular descriptors routinely used in QSAR analysis are electronic, hydrophobic, steric, and topological indices which are invaluable in helping to delineate a

Table 1. Biological, physicochemical, and structural parameters used to derive QSAR Equation (2) and the predicted molecules.

No.	X	Y	Z	log1/C [Eq. (2)] obsd.	calcd.	Δ	Clog$P$	$B5_X$	MR_Y	MR_Z	I_1	I_2	Ref.
4-1	H	H	H	6.46	6.40	0.06	0.90	1.00	0.10	0.10	0	0	[17]
4-2	CH_3	H	H	5.97	5.66	0.31	1.40	2.04	0.10	0.10	0	1	[18]
4-3	C_2H_5	H	H	5.35	5.41	−0.06	1.93	3.17	0.10	0.10	0	1	[18]
4-4	$(CH_2)_2CH_3$	H	H	5.14	5.65	−0.51	2.45	3.49	0.10	0.10	0	1	[18]
4-5	$(CH_2)_3CH_3$	H	H	5.14	5.45	−0.31	2.98	4.54	0.10	0.10	0	1	[18]
4-6	$CH_2CH(CH_3)_2$	H	H	6.35	5.92	0.43	2.85	4.45	0.10	0.10	0	0	[17]
4-7	$C(CH_3)_3$	H	H	6.62	6.60	0.02	2.72	3.17	0.10	0.10	0	0	[17]
4-8[a]	$CH_2C_6H_5$	H	H	5.34	4.52	0.82	2.96	6.02	0.10	0.10	0	1	[17]
4-9[a]	H	CH_3	H	7.42	6.55	0.87	1.40	1.00	0.56	0.10	0	0	[19]
4-10	H	Cl	H	7.07	6.75	0.32	1.65	1.00	0.60	0.10	0	0	[19]
4-11	H	F	H	6.80	6.57	0.23	1.08	1.00	0.09	0.10	0	0	[19]
4-12[a]	H	NH_2	H	6.96	5.68	1.28	0.35	1.00	0.54	0.10	0	0	[19]
4-13	H	OH	H	6.06	6.34	−0.28	0.95	1.00	0.29	0.10	0	0	[19]
4-14	H	H	CH_3	6.52	6.61	−0.09	1.40	1.00	0.10	0.56	0	0	[19]
4-15	H	H	F	6.43	6.56	−0.13	1.08	1.00	0.10	0.09	0	0	[19]
4-16	H	H	Cl	6.85	6.81	0.04	1.65	1.00	0.10	0.60	0	0	[19]
4-17	H	H	Br	6.89	6.81	0.08	1.80	1.00	0.10	0.89	0	0	[19]
4-18	H	H	NO_2	6.19	6.00	0.19	0.76	1.00	0.10	0.74	0	0	[19]
4-19[a]	H	H	NH_2	6.44	5.74	0.70	0.35	1.00	0.10	0.54	0	0	[20]
4-20	H	H	OH	6.96	7.24	−0.28	0.95	1.00	0.10	0.29	1	0	[19]
4-21	H	H	CN	5.72	5.79	−0.07	0.46	1.00	0.10	0.63	0	0	[19]
4-22	H	H	COOH	6.00	6.16	−0.16	0.93	1.00	0.10	0.69	0	0	[19]
4-23	H	H	OCH_3	6.14	6.33	−0.19	1.17	1.00	0.10	0.79	0	0	[18]
4-24	CH_3	H	OH	6.89	6.49	0.40	1.44	2.04	0.10	0.29	1	1	[18]
4-25	C_2H_5	H	OH	6.49	6.24	0.25	1.97	3.17	0.10	0.29	1	1	[18]
4-26	$(CH_2)_2CH_3$	H	OH	6.22	6.49	−0.27	2.50	3.49	0.10	0.29	1	1	[18]
4-27	$(CH_2)_3CH_3$	H	OH	6.28	6.29	−0.01	3.03	4.54	0.10	0.29	1	1	[18]
4-28	CH_3	H	OCH_3	5.62	5.58	0.04	1.67	2.04	0.10	0.79	0	1	[18]
4-29	C_2H_5	H	OCH_3	5.55	5.33	0.22	2.20	3.17	0.10	0.79	0	1	[18]
4-30	$(CH_2)_2CH_3$	H	OCH_3	5.35	5.58	−0.23	2.73	3.49	0.10	0.79	0	1	[18]
4-31	$(CH_2)_3CH_3$	H	OCH_3	5.55	5.38	0.17	3.26	4.54	0.10	0.79	0	1	[18]
4-32	$CH_2N(CH_3)_2$	H	OH	5.15	5.20	−0.05	0.78	4.08	0.10	0.29	1	0	[21]
4-33	H	Cl	Cl	6.96	7.03	−0.07	2.26	1.00	0.60	0.60	0	0	[19]
4-34	H	$CH_2N(CH_3)_2$	OH	5.98	6.02	−0.04	0.73	1.00	1.87	0.29	1	0	[22]
4-35[b]	H	Br	Br	ND	7.29	ND	2.92	1.00	0.89	0.89	0	0	NA
4-36[b]	H	OH	OH	ND	7.58	ND	1.48	1.00	0.29	0.29	1	0	NA

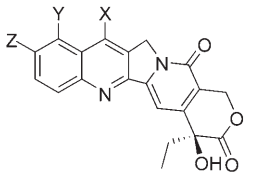
[a] Not used in the derivation of QSAR Equation (2). [b] Predicted molecules on the basis of QSAR Equation (2) for the next synthetic target. ND = Not determined. NA = Not applicable.

large number of receptor–ligand interactions that are critical in biological processes. The quality of a QSAR model, however, depends mainly on the type and quality of the data, and is valid only for the compound structures analogues to those used to build the model. QSAR models can stand alone, augment other computational approaches, or be examined in tandem with equations of a similar mechanistic genre to establish the authenticity and reliability.[16] The 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.

Results and Discussion

Inhibition of DNA topo I by camptothecin analogues 4

Based on data obtained from Manikumar et al.,[17] Vladu et al.,[18] Wall et al.,[19] Wadkins et al.,[20,21] and Carrigan et al.[22] (see Table 1).



4

$$\log 1/C = 0.84(\pm 0.22)\mathrm{Clog}P - 0.62(\pm 0.14)B5_X - 0.59(\pm 0.33)\mathrm{MR}_Y - 0.45(\pm 0.36)\mathrm{MR}_Z + 0.88(\pm 0.28)I_1 - 0.52(\pm 0.28)I_2 + 6.37(\pm 0.31) \quad (2)$$

$n = 30$, $r^2 = 0.854$, $s = 0.258$,
$q^2 = 0.738$, $Q = 3.581$, $F_{6,23} = 22.422$

range in log $1/C = 5.14$–7.42
outliers: $X = CH_2C_6H_5$, $Y = Z = H$; $X = Z = H$, $Y = CH_3$; $X = Z = H$, $Y = NH_2$; $X = Y = H$, $Z = NH_2$

Four compounds in Table 1 were deemed to be outliers on the basis of their deviations (2×SD). This Equation (2) explains 85.4% of variance in log 1/*C*. Clog*P* is the calculated partition coefficient in *n*-octanol/water and is a measure of hydrophobicity for the whole molecule. Positive coefficient of Clog*P* suggests that, at all the parts where substituents have been entered, hydrophobic contacts have been made. Thus, the highly hydrophobic molecules will be more active. $B5_X$ is a Verloop's sterimol descriptor, which measures the maximum width of the X-substituents. MR_Y and MR_Z are the calculated molar refractivities of Y and Z substituents, respectively. A negative sign associated with $B5_X$, MR_Y, and MR_Z brings out steric effects for these substituents. The indicator variable (I_1) takes the value of 1 and 0 for the presence and absence of hydroxyl group at position 10, respectively. Similarly, I_2 takes the value of 1 and 0 for the presence and absence of *n*-alkyl groups at position 7, respectively. The presence of the hydroxyl group at position 10 increases the activity as evidenced by the positive coefficient of the indicator variable (I_1). The negative coefficient of the indicator variable (I_2) suggests that the presence of branched alkyl groups at position 7 is preferred over *n*-alkyl groups. A comparison between observed and calculated log1/*C* of camptothecin (**4**) used in the development of QSAR Equation (2) is shown in Figure 2.

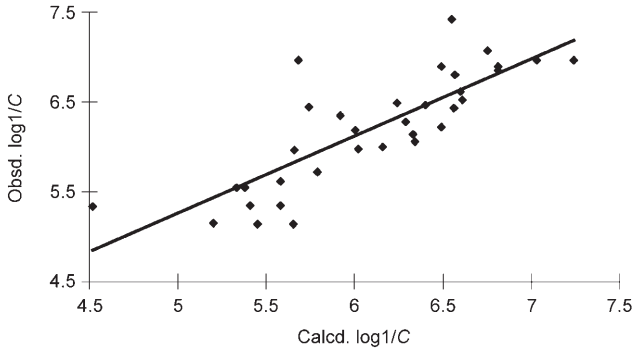


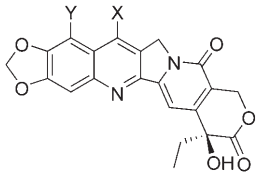
Figure 2. Plot of observed versus calculated log1/*C* [Eq. (2)].

The outlier ($X = CH_2C_6H_5$, $Y = Z = H$) is much more active than expected, by three times the standard deviation. Possible reasons for its unusual activity are not obvious although its bulk and/or geometry due to the presence of phenyl group may reduce the coplanarity with X-group and increase the activity.[23] The other outlier ($X = Z = H$, $Y = CH_3$) is also more active than expected by three times the standard deviation. The reason for this outlier is not very clear because the derivative ($X = Y = H$, $Z = CH_3$) is well predicted. It may be attributed to some experimental error. Two amino analogues ($X = Z = H$, $Y = NH_2$ and $X = Y = H$, $Z = NH_2$) were considered outliers because of being much more active than expected by five and three times the standard deviation, respectively. This anomalous behavior may be ascribed to its aniline nature, which could result in hydrogen abstraction or involve microsomal N-oxidation.[14a,24] It is interesting to note that camptothecin derivatives (**4**) fulfill all conditions of the rule of five.

In the process of drug discovery and development, the estimation of molecular transport properties, particularly intestinal absorption and blood-brain barrier penetration, is one of the important key factors. Traditionally, calculated values of *n*-octanol/water partition coefficient have been used for this purpose.[25] A set of rules, which imposes the limitations on log *P*, molecular weight, and the number of hydrogen bond donors and acceptors is known as the rule of five and was introduced by Lipinski et al.[26] According to Lipinski's rule of five, drug-like molecules should have log $P \leq 5$; molecular weight ≤ 500; number of hydrogen bond acceptors (expressed as the sum of Ns and Os) ≤ 10; and number of hydrogen bond donors (expressed as the sum of OHs and NHs) ≤ 5. Molecules violating more than one of these rules may have problems with oral bioavailability.

Inhibition of DNA topo I by 10,11-methylenedioxycamptothecin analogues 5.

Based on data obtained from Manikumar et al.,[17] Wall et al.,[19] and Wadkins et al.,[21] (see Table 2).



5

$$\log 1/C = -0.71(\pm 0.20)\pi_X - 0.88(\pm 0.39)I_3 + 0.73(\pm 0.44)I_4 + 7.19(\pm 0.22) \quad (3)$$

$n = 16$, $r^2 = 0.873$, $s = 0.215$,
$q^2 = 0.788$, $Q = 4.349$, $F_{3,12} = 27.496$

range in log $1/C = 5.52–7.40$

outlier: $X = CH_2NHCH_2CH_2OH$, $Y = H$

One compound in Table 2 was deemed to be an outlier on the basis of its deviation (2×SD). Equation (3) explains 87.3% of variance in log1/*C*. π_X is the calculated hydrophobic parameter for X-substituents, and is the most important parameter for this dataset. The negative coefficient with π_X suggests that molecules (**5**) with highly hydrophilic X-substituents would present better inhibitory activity. I_3 and I_4 are indicator variables, which pinpoint the unusual activities of X = aminomethyl groups and X = cyclic groups, respectively. Negative coefficient of I_3 suggests that the presence of X = aminomethyl groups will be detrimental to the activity. On the other hand, the presence of X = cyclic groups will promote the inhibitory activity as shown by the positive coefficient of I_4. A comparison between observed and calculated log1/*C* of camptothecin (**5**) used in the development of QSAR Equation (3) has been shown in Figure 3.

It is interesting to note that all the compounds of this data set except two ($X = C_6H_4$-4-Cl, $Y = H$ and $X = CH_2NHC(CH_2OH)_3$, $Y = H$) fulfill all conditions of the rule of five. One compound ($X = C_6H_4$-4-Cl, $Y = H$) violates only one rule with very minor deviation (molecular weight = 502.93) and was kept in the data set. The other compound ($X = CH_2NHC(CH_2OH)_3$, $Y = H$) violates two rules (*MW* = 525.56, HBA = 12), thus we removed this compound from the data set and developed Equation (4).

$$\log 1/C = -0.71(\pm 0.21)\pi_X - 0.91(\pm 0.43)I_3 + 0.72(\pm 0.46)I_4 + 7.19(\pm 0.23) \quad (4)$$

$n = 15$, $r^2 = 0.873$, $s = 0.222$
$q^2 = 0.763$, $Q = 4.207$, $F_{3,11} = 25.205$

range in log$1/C = 5.52–7.40$

outlier: $X = CH_2NHCH_2CH_2OH$, $Y = H$

removed compound: $X = CH_2NHC(CH_2OH)_3$, $Y = H$

Statistics of the QSAR [Eq. (3) and (4)] are very much identical. This may be due to the fact that the compound ($X = CH_2NHC(CH_2OH)_3$, $Y = H$) is very well predicted from Equation (3). This is the reason; we preferred Equation (3) and kept all the compounds of this series (**5**).

Table 2. Biological, physicochemical, and structural parameters used to derive QSAR Equation (3), the validation molecules, and predicticed molecules.

No.	X	Y	log1/*C* [Eq. (3)] obsd.	calcd.	Δ	π_X	I_3	I_4	Ref.
5-1	H	H	7.30	7.19	0.11	0.00	0	0	[17]
5-2	C_2H_5	H	6.57	6.46	0.11	1.03	0	0	[17]
5-3	Cyclo-C_5H_9	H	6.72	6.46	0.26	2.06	0	1	[17]
5-4	Cyclo-C_6H_{11}	H	6.00	6.06	−0.06	2.62	0	1	[17]
5-5	$CH_2C_6H_5$	H	5.71	5.73	−0.02	2.07	0	0	[17]
5-6	C_6H_5	H	6.82	6.58	0.24	1.89	0	1	[17]
5-7	C_6H_4(4-F)	H	6.41	6.48	−0.07	2.03	0	1	[17]
5-8	C_6H_4(4-Cl)	H	5.94	6.07	−0.13	2.60	0	1	[17]
5-9	C_6H_4(4-CH_3)	H	5.98	6.23	−0.25	2.39	0	1	[17]
5-10	CH_2NH_2	H	6.70	7.05	−0.35	−1.05	1	0	[21]
5-11	$CH_2NHCH(CH_3)_2$	H	6.29	6.16	0.13	0.21	1	0	[21]
5-12[a]	$CH_2NHCH_2CH_2OH$	H	5.52	7.17	−1.65	−1.22	1	0	[21]
5-13	$CH_2NHCH(CH_2OH)_2$	H	7.40	7.28	0.12	−1.38	1	0	[21]
5-14	$CH_2NHC(CH_2OH)_3$	H	6.96	6.86	0.10	−0.79	1	0	[21]
5-15	H	NO_2	6.82	7.19	−0.37	0.00	0	0	[19]
5-16	H	NH_2	7.32	7.19	0.13	0.00	0	0	[19]
5-17	H	Cl	7.22	7.19	0.03	0.00	0	0	[19]
5-18[b]	CH_2(–NCH_2CH_2N(Me)CH_2CH_2–)	H	6.52	6.17	0.35	0.19	1	0	[36]
5-19[b]	CH_2(-NCH=CHN=CH-)	H	6.80	6.76	0.04	−0.64	1	0	[37]
5-20[c]	CH_2(–$NCH_2CH_2NHCH_2CH_2$–)	H	ND	6.50	ND	−0.27	1	0	NA
5-21[c]	CH_2NHCH_3	H	ND	6.75	ND	−0.63	1	0	NA
5-22[c]	Cyclo-C_4H_7	H	ND	7.01	ND	1.28	0	1	NA

[a] Not used in the derivation of QSAR Equation (3). [b] Used for the validation of QSAR Equation (3). [c] Predicted molecules on the basis of QSAR Equation (3) for the next synthetic target. ND = Not determined. NA = Not applicable.

Conclusion

The studies on the mechanism of action of camptothecin and the discovery of DNA topo I as a therapeutic target opened a new area for cancer drug development, including approval for clinical use of topotecan and irinotecan, and discovery of several analogues that are currently in various stages of clinical trials. This area has become a part of the multimillion dollar industry that is dedicated to finding better chemotherapeutic agents with excellent antitumor activity and less normal tissue toxicity. Thus, we have high expectations for the next generation and further development of new and better antitumor camptothecin derivatives.

The application of the QSAR paradigm has already been

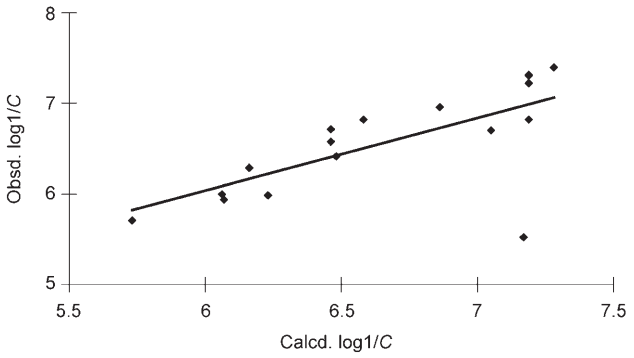


Figure 3. Plot of observed versus calculated log1/*C* [Eq. (3)].

proved to be useful in elucidating the mechanisms of chemical–biological interactions. Our QSAR results suggest that the inhibitory activity of camptothecin analogues **4** toward topo I is mainly dependent on their hydrophobic and steric descriptors, whereas the same activity of 10,11-methylenedioxycamptothecin analogues **5** is largely dependent on their hydrophobic descriptors at position 7. On the basis of these two QSAR models (2 and 3), we can predict at least five compounds (**4-35**, **4-36**, **5-20**, **5-21**, and **5-22**) that may be the next synthetic target. These proposed molecules also fulfill the conditions of Lipinski's rule of five for oral bioavailability.

Further development of QSAR studies should not only enlarge the areas of their application, but also increase our understanding towards the mechanisms of chemical–biological interactions. This may assist in the development of new CPT with excellent antitumor activity and low toxicity.

Experimental Section

All the data of DNA topoisomerase I inhibition for camptothecin derivatives have been collected from the literature.[17–22] *C* is the molar concentration of a compound and log1/*C* is the dependent variable that defines the biological parameter for QSAR equations. Physicochemical descriptors are autoloaded, and multiregression analyses (MRA) used to derive QSAR were executed with the C-QSAR program.[27] The parameters used in this paper have already been discussed in detail along with their application.[13] Briefly, Clog*P* is the calculated partition coefficient in *n*-octanol/water and is a measure of hydrophobicity, and π is the hydrophobic parameter for the substituents. CMR is the calculated molar refractivity for the whole molecule. *MR* is calculated from the Lorentz-Lorenz equation and is described as follows: $[(n^2-1)/(n^2+2)](MW/\delta)$, where *n* is the refractive index, *MW* is the molecular weight, and δ is the density of a substance. MR is dependent on volume and polarizability. 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/compound. Although it contains no information about the shape, it has found considerable usage in biological QSAR where intermolecular effects predominate. MR is usually scaled at 0.1 to make it equiscalar with π. *B*1, *B*5, and *L* are the Verloop's sterimol parameters for substituents.[28] *B*1 is a measure of the minimum width of a substituent, *B*5 is an attempt to define maximum width of the substituent, and *L* is the substituent length. The indicator variable *I* is assigned the value of 1 or 0 for special features with special effects that cannot be parametrized 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) is obtained by using leave-one-out (LOO) procedure as described by Cramer et al.[29] *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. The high values of *Q* showed the high predictive power of the QSAR models and also no over fitting. *F* is the Fischer statistics (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 degree 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 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 good 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 assigned to be outliers on the basis of their deviation between observed and calculated activities from the equation ($>2\,s$).[31] Each regression equation includes 95% confidence limits for each term in parentheses. The QSAR models reported here are derived by us and were not formulated by the original authors. These QSAR are found to be statistically significant, which fulfill the conditions of acceptable models as given by Golbraikh and Tropsha.[32]

Validation of QSAR

QSAR model validation becomes an essential part in developing 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 to validate the QSAR Equations (2) and (3):

Internal Validation

- Fraction of the Variance: The fraction of the variance of an MRA model is expressed by 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 two QSAR models [Eq. (2) and (3)] are 0.854 and 0.873, which suggests that these QSAR models explain 85.4 and 87.3% of the variance, respectively in the data set. According to the literature, the predictive QSAR model must have $r^2>0.6$.[32,33]
- Cross-validation Test: The values of q^2 for these two QSAR models are 0.738 and 0.788. The high values of q^2 validate the QSAR models. According to the literature, the predictive QSAR model must have $q^2>0.5$.[32,33]
- 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 0.258 and 0.215.
- Quality factor (*Q*): Chance correlation, due to the excessive numbers of parameter (which also increases the *r* and *s* values), is detected by the examination of *Q* value.[30] The high values of *Q* (3.581 and 4.349) for these QSAR models suggest their high predictive power as well as no overfitting.

- 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 equation is significant. The *F*-values for these QSAR models are 22.422 and 27.496, 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 10 runs have been listed in Table 3 [Eq. (2) and (3)]. The poor values of r^2 and q^2 in the Y-randomization test ensure the robustness of QSAR models.[31a,33–35]
- Both QSAR models also fulfill the rule of thumb condition that (number of data points)/(number of descriptors) ≥ 4.

Table 3. Y-randomization data for Equations (2) and (3).

NOR	Equation (2)		Equation (3)	
	r^2	q^2	r^2	q^2
1	0.274	−0.423	0.238	−0.643
2	0.477	−0.442	0.251	−0.251
3	0.480	−0.780	0.397	−0.082
4	0.520	−0.084	0.277	−0.427
5	0.317	−0.421	0.112	−0.711
6	0.315	−0.191	0.232	−0.456
7	0.345	−0.085	0.082	−0.567
8	0.234	−0.796	0.243	−0.450
9	0.459	0.135	0.191	−0.375
10	0.305	−0.157	0.164	−1.007

NOR = Number of Y-randomization.

External Validation

- Selection of the Training and Test Sets for Camptothecin Analogues (**4**) Used in the Development of QSAR Equation (2): The data set was divided into training ($n = 18$) and test ($n = 12$) sets in a random manner (10 trials). The QSAR models for both of the sets (training and test) were generated by using the same descriptors as those of Equation (2) and validated on the basis of their statistics. A random selection pattern of the training and test sets with their statistical parameters, used for the external validation of QSAR Equation (2), is given in Table 4. The statistical parameters of randominally generated training and test sets are in acceptable range ($r^2 > 0.6$ and $q^2 > 0.5$).
- Selection of the Training and Test Sets for Camptothecin Analogs (**5**) Used in the Development of QSAR Equation (3): It is difficult to divide the small data set ($n = 16$) into training and test sets to validate the Equation (3). Thus, we used this Equation (3) to calculate the activity of two known compounds (**5-18** and **5-19**) and found that the calculated values are in very good agreement to their experimental (observed) values (see Table 2). These observation has validated the QSAR Equation (3).

Table 4. A random selection pattern of training and test sets with the statistical parameters used for the external validation of QSAR Equation (2).

No.	Training Sets (n=18)	Test Sets (n=12)	Training Sets (n=18)			Test Sets (n=12)		
			r^2	q^2	s	r^2	q^2	s
1	**4-3**, **-4**, **-6**, **-11**, **-13**, **-15**, **-17**, **-18**, **-21**, **-22**, **-23**, **-26**, **-27**, **-29**, **-30**, **-31**, **-32**, **-33**	**4-1**, **-2**, **-5**, **-7**, **-10**, **-14**, **-16**, **-20**, **-24**, **-25**, **-28**, **-34**	0.901	0.735	0.225	0.948	0.570	0.200
2	**4-1**, **-3**, **-6**, **-7**, **-11**, **-15**, **-16**, **-17**, **-18**, **-21**, **-23**, **-25**, **-28**, **-29**, **-30**, **-32**, **-33**, **-34**	**4-2**, **-4**, **-5**, **-10**, **-13**, **-14**, **-20**, **-22**, **-24**, **-26**, **-27**, **-31**	0.927	0.634	0.196	0.937	0.655	0.243
3	**4-1**, **-3**, **-6**, **-7**, **-11**, **-14**, **-15**, **-16**, **-18**, **-21**, **-25**, **-26**, **-28**, **-29**, **-31**, **-32**, **-33**, **-34**	**4-2**, **-4**, **-5**, **-10**, **-13**, **-17**, **-20**, **-22**, **-23**, **-24**, **-27**, **-30**	0.925	0.685	0.187	0.942	0.542	0.249
4	**4-1**, **-3**, **-4**, **-5**, **-6**, **-7**, **-11**, **-13**, **-14**, **-15**, **-17**, **-22**, **-23**, **-24**, **-25**, **-27**, **-30**, **-33**	**4-2**, **-10**, **-16**, **-18**, **-20**, **-21**, **-26**, **-28**, **-29**, **-31**, **-32**, **-34**	0.913	0.530	0.221	0.984	0.939	0.114
5	**4-1**, **-2**, **-6**, **-7**, **-11**, **-14**, **-15**, **-16**, **-17**, **-18**, **-24**, **-25**, **-26**, **-28**, **-29**, **-31**, **-33**, **-34**	**4-3**, **-4**, **-5**, **-10**, **-13**, **-20**, **-21**, **-22**, **-23**, **-27**, **-30**, **-32**	0.887	0.534	0.194	0.973	0.645	0.166
6	**4-1**, **2**, **4**, **6**, **10**, **11**, **15**, **16**, **18**, **22**, **24**, **25**, **26**, **28**, **29**, **30**, **31**, **33**	**4-3**, **5**, **7**, **13**, **14**, **17**, **20**, **21**, **23**, **27**, **32**, **34**	0.853	0.543	0.281	0.974	0.725	0.150
7	**4-1**, **-2**, **-3**, **-7**, **-13**, **-14**, **-15**, **-16**, **-17**, **-18**, **-21**, **-22**, **-24**, **-25**, **-28**, **-29**, **-31**, **-33**	**4-4**, **-5**, **-6**, **-10**, **-11**, **-20**, **-23**, **-26**, **-27**, **-30**, **-32**, **-34**	0.952	0.825	0.142	0.957	0.583	0.220
8	**4-1**, **-2**, **-3**, **-7**, **-11**, **-14**, **-16**, **-17**, **-18**, **-20**, **-21**, **-23**, **-26**, **-27**, **-28**, **-29**, **-33**, **-34**	**4-4**, **-5**, **-6**, **-10**, **-13**, **-15**, **-22**, **-24**, **-25**, **-30**, **-31**, **-32**	0.940	0.769	0.156	0.987	0.942	0.119
9	**4-1**, **-2**, **-3**, **-4**, **-7**, **-14**, **-15**, **-16**, **-18**, **-23**, **-24**, **-25**, **-27**, **-28**, **-29**, **-31**, **-33**, **-34**	**4-5**, **-6**, **-10**, **-11**, **-13**, **-17**, **-20**, **-21**, **-22**, **-26**, **-30**, **-32**	0.896	0.696	0.220	0.958	0.626	0.212
10	**4-1**, **-2**, **-3**, **-4**, **-5**, **-6**, **-7**, **-13**, **-15**, **-16**, **-17**, **-18**, **-24**, **-26**, **-29**, **-30**, **-33**, **-34**	**4-10**, **-11**, **-14**, **-20**, **-21**, **-22**, **-23**, **-25**, **-27**, **-28**, **-31**, **-32**	0.867	0.541	0.279	0.990	0.636	0.091

Prediction of Molecules

We believe that the validated QSAR models (2) and (3) may be considered as the predictive models to narrow the synthetic challenges in order to yield very specific topo I inhibitors. On the basis of these models, we can predict at least five compounds (**4-35**, **4-**

Table 5. Rule of five criteria for the proposed molecules.

No.	log1/*C*	Clog*P*	Mol. Wt.	HBA	HBD
4-35[a]	7.29	2.92	506.15	6	1
4-36[a]	7.58	1.48	380.36	8	3
5-20[b]	6.50	1.15	490.52	10	2
5-21[b]	6.75	0.78	435.44	9	2
5-22[b]	7.01	2.92	446.46	8	1

[a] Predicted by using QSAR Equation (2). [b] Predicted by using QSAR Equation (3). HBA = Number of hydrogen bond acceptors. HBD = Number of hydrogen bond donors.

36, **5-20**, **5-21**, and **5-22**) that may be the next synthetic target (see Table 1 and Table 2). The proposed molecules (**4-35**, **4-36**, **5-20**, **5-21**, and **5-22**) also fulfill the conditions of Lipinski's rule of five for oral bioavailability. Predicted log1/*C* for these proposed molecules against topo I along with their Clog*P*, molecular weight, and the number of hydrogen bond acceptors and donors (conditions of Lipinski's "rule of five") are given in Table 5.

Keywords: camptothecin · hydrophobicity · molar refractivity · quantitative structure–activity relationships · topoisomerase I

[1] M. E. Wall, M. C. Wani, C. E. Cook, K. H. Palmer, A. T. McPhail, G. A. Sim, *J. Am. Chem. Soc.* **1966**, *88*, 3888 – 3890.
[2] M. E. Wall, *Med. Res. Rev.* **1998**, *18*, 299 – 314.
[3] Y. H. Hsiang, R. Hertzberg, S. Hecht, L. F. Liu, *J. Biol. Chem.* **1985**, *260*, 14873 – 14878.
[4] Y. H. Hsiang, M. G. Lihou, L. F. Liu, *Cancer Res.* **1989**, *49*, 5077 – 5082.
[5] J. Dancey, E. A. Eisenhauer, *Br. J. Cancer* **1996**, *74*, 327 – 338.
[6] *Camptothecin: New Anticancer Agents* (Eds.: M. Potmesil, H. Pinedo), CRC: Boca Raton, **1995**.
[7] J. J. Champoux, *Annu. Rev. Biochem.* **2001**, *70*, 369 – 413.
[8] a) A. H. Corbett, N. Osheroff, *Chem. Res. Toxicol.* **1993**, *6*, 585 – 597; b) J. C. Wang, *Annu. Rev. Biochem.* **1996**, *65*, 635 – 692; c) J. L. Nitiss, *Biochim. Biophys. Acta* **1998**, *1400*, 63 – 81.
[9] Y. Pommier, *Nat. Rev. Cancer* **2006**, *6*, 789 – 802.
[10] B. L. Staker, M. D. Feese, M. Cushman, Y. Pommier, D. Zembower, L. Stewart, A. B. Burgin, *J. Med. Chem.* **2005**, *48*, 2336 – 2345.
[11] R. P. Verma, *Bioorg. Med. Chem.* **2005**, *13*, 1059 – 1067.
[12] C. Hansch, P. P. Maloney, T. Fujita, R. M. Muir, *Nature* **1962**, *194*, 178 – 180.
[13] C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington D.C., **1995**, pp. 169 – 543.
[14] a) C. D. Selassie, R. Garg, S. Kapur, A. Kurup, R. P. Verma, S. B. Mekapati, C. Hansch, *Chem. Rev.* **2002**, *102*, 2585 – 2605; b) R. P. Verma, A. Kurup, S. B. Mekapati, C. Hansch, *Bioorg. Med. Chem.* **2005**, *13*, 933 – 948.
[15] C. Hansch, A. Leo, S. B. Mekapati, A. Kurup, *Bioorg. Med. Chem.* **2004**, *12*, 3391 – 3400.
[16] C. D. Selassie, S. B. Mekapati, R. P. Verma, *Curr. Top. Med. Chem.* **2002**, *2* 1357 – 1379.
[17] G. Manikumar, R. M. Wadkins, D. Bearss, D. D. von Hoff, M. C. Wani, M. E. Wall, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5377 – 5381.
[18] B. Vladu, J. M. Woynarowski, G. Manikumar, M. C. Wani, M. E. Wall, D. D. von Hoff, R. M. Wadkins, *Mol. Pharm.* **2000**, *57*, 243 – 251.
[19] M. E. Wall, M. C. Wani, A. W. Nicholas, G. Manikumar, C. Tele, L. Moore, A. Truesdale, P. Leitner, J. M. Besterman, *J. Med. Chem.* **1993**, *36*, 2689 – 2700.
[20] R. M. Wadkins, D. Bearss, G. Manikumar, M. C. Wani, M. E. Wall, D. D. von Hoff, *Curr. Med. Chem.: Anti Cancer Agents* **2004**, *4*, 327 – 334.
[21] R. M. Wadkins, D. Bearss, G. Manikumar, M. C. Wani, M. E. Wall, D. D. von Hoff, *Cancer Res.* **2004**, *64*, 6679 – 6683.
[22] S. W. Carrigan, P. C. Fox, M. E. Wall, M. C. Wani, J. P. Bowen, *J. Comput.-Aided Mol. Des.* **1997**, *11*, 71 – 78.
[23] R. P. Verma, S. Kapur, O. Barberena, A. Shusterman, C. H. Hansch, C. D. Selassie, *Chem. Res. Toxicol.* **2003**, *16*, 276 – 284.
[24] S. Kapur, A. Shusterman, R. P. Verma, C. Hansch, C. D. Selassie, *Chemosphere* **2000**, *41*, 1643 – 1649.
[25] M. P. Freitas, *Org. Biomol. Chem.* **2006**, *4*, 1154 – 1159.
[26] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **1997**, *23*, 3 – 25.
[27] C-QSAR Program, BioByte Corporation, Claremont, CA 91711,USA, http://www.biobyte.co
[28] A. Verloop, J. Tipker, *QSAR in Drug Design and Toxicology* (Eds.: D. Hadzi, B. Jorman-Blazic), Elsevier, Amsterdam, **1987**, p. 97.
[29] R. D. Cramer, 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; *J. Phys. Chem.* **1996**, *100*, 18065 – 18077; b) 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; R. P. Verma, C. Hansch, *Virology* **2007**, *359*, 152 – 161; R. P. Verma, C. Hansch, *Bioorg. Med. Chem.* **2007**, *15*, 2223 – 2268.
[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, in *Chemometrics Methods in Molecular Design* (Ed.: H. van de Waterbeemd), Wiley-VCH, Weinheim, **1995**, pp. 309 – 318.
[35] M. F. Melzig, G. D. Tran, K. Henke, C. D. Selassie, R. P. Verma, *Pharmazie* **2005**, *60*, 869 – 873.
[36] M. J. Luzzio, J. M. Besterman, D. L. Emerson, M. G. Evans, K. Lackey, P. L. Leitner, G. McIntyre, B. Morton, P. L. Myers, M. Peel, J. M. Sisco, D. D. Sternbach, W.-Q. Tong, A. Truesdale, D. E. Uehling, A. Vuong, J. Yates, *J. Med. Chem.* **1995**, *38*, 395 – 401.
[37] K. Lackey, D. D. Sternbach, D. K. Croom, D. L. Emerson, M. G. Evans, P. L. Leitner, M. J. Luzzio, G. McIntyre, A. Vuong, J. Yates, J. M. Besterman, *J. Med. Chem.* **1996**, *39*, 713 – 719.

Received: June 6, 2007
Revised: August 22, 2007
Published online on September 20, 2007

