

Full-length article

3D-QSAR study of 20 (*S*)-camptothecin analogs

Ai-jun LU^{1,2}, Zhen-shan ZHANG², Ming-yue ZHENG², Han-jun ZOU², Xiao-min LUO², Hua-liang JIANG^{2,3,4}

¹Jiangsu Sincere Pharmaceutical Research Company Ltd, Nanjing 210042, China; ²Drug Discovery and Design Center, State Key Lab of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; ³School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

Key words

comparative molecular field analysis (CoMFA); camptothecin (CPT); topoisomerase I (Top I)

⁴Correspondence to Prof Hua-liang JIANG.
Phn/Fax 86-21-5080-7188.
E-mail hljiang@mail.shenc.ac.cn

Received 2006-06-12

Accepted 2006-08-14

doi 10.1111/j.1745-7254.2007.00477.x

Abstract

Aim: To build up a quantitative structure-activity relationship (QSAR) model of 20 (*S*)-camptothecin (CPT) analogs for the prediction of the activity of new CPT analogs for drug design. **Methods:** A training set of 43 structurally diverse CPT analogs which were inhibitors of topoisomerase I were used to construct a quantitative structure-activity relationship model with a comparative molecular field analysis (CoMFA). The QSAR model was optimized using partial least squares (PLS) analysis. A test set of 10 compounds was evaluated using the model. **Results:** The CoMFA model was constructed successfully, and a good cross-validated correlation was obtained in which q^2 was 0.495. Then, the analysis of the non-cross-validated PLS model in which r^2 was 0.935 was built and permitted demonstrations of high predictability for the activities of the 10 CPT analogs in the test set selected in random. **Conclusion:** The CoMFA model indicated that bulky negative-charged group at position 9, 10 and 11 of CPT would increase activity, but excessively increasing bulky group at position 10 is adverse to inhibitory activity; substituents that occupy position 7 with the bulky positive group will enhance the inhibitive activity. The model can be used to design new CPT analogs and understand the mechanism of action.

Introduction

Camptothecin (CPT) is an alkaloid isolated from the Chinese tree *Camptotheca acuminata* (Figure 1) by Wall Me and co-workers in 1966^[1]. The molecule has a pentacyclic ring structure embodying pyrrolo[3,4-*b*]-quinoline moiety (rings A, B and C), conjugated pyridone moiety (ring D) and a chiral centre within a 6 membered α -hydroxy lactone (ring

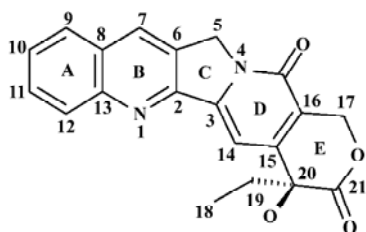


Figure 1. Structure of CPT.

E). Although showing impressive activity in a number of experimental tumors^[2,3] including the human colon, lung and mammary tumor lines, its clinical development was halted in the early 1970s owing to negligible water solubility and unpredictable and formidable toxicities, such as myelosuppression, diarrhoea and hemorrhagic cystitis^[4].

The renewed interest in CPT relies mainly on the recognition of its novel and unique mechanism of action. It was not until 1985^[5] that the nuclear enzyme topoisomerase I (TOP I) was identified as its molecular target reported by Liu and co-workers. The enzyme has been implicated in various DNA transactions such as replication, transcription and recombination. CPT and its analogs bind to a complex formed by DNA with the Top I enzyme, which inhibit tumor growth. Because of the unique mechanism of action, the research interests in CPT have become manifold and lead to the production of several hundreds of synthetic and semi-synthetic analogs to date. Now, 2 compounds in this class, topotecan

(hycamtin) and irinotecan (camptosar), have been approved for clinical use as anticancer drugs in the United States by the Food and Drug Administration (FDA). Topotecan is presently indicated as a second-line therapy for advanced ovarian cancer and small-cell lung cancer. Irinotecan is approved for use in the treatment of advanced colorectal cancer, both as first-line therapy in combination with 5-FU and as salvage treatment in 5-FU refractory disease. There are more than 10 other CPT analogs in various stages of clinical evaluation^[6], including 9-AC, 9-NC, GI-147211, exatecan mesylate, and karenitecin.

As a lot of CPT analogs have been synthesized and evaluated, several 3D QSAR studies of CPT were reported from laboratories^[10,18,19]. Among them, Yoon and colleges developed the QSAR model with CPT-11 and other prodrugs with their hydrolysis percent of serum esterase activity to design a new, easily activated SN-38 prodrug^[18]. In two other papers^[10,19], the authors built the QSAR models with little CPT dataset and few substitutes. Furthermore, due to lack of crystal structure, they all align the compounds with the partial least squares (PLS) fit, and have no detailed interpretation of crystal structure. In this paper, in order to further understand the inhibition mechanism of CPT and to guide structural modification and rational drug design of CPT analog, a 3D QSAR study of CPT analogs with larger dataset using CoMFA was implemented. The final 3-D contour maps were compared with the available X-ray crystal structure (PDB entry 1K4T)^[8] from the Protein Databank (PDB, <http://www.rcsb.org>).

Materials and methods

Computational methods All computations are supported by corresponding modeling suite in SYBYL6.8^[7], operated on a SGI Octane-2 graphic workstation (Mountain View, CA, USA).

Structure preparation The 53 compounds for this study and their bioactive data are listed in Figures 2–6 and Tables 1–5. The potency data (IC_{50}) was assessed by the minimum concentration (mol/L) that inhibited the cleavable complex

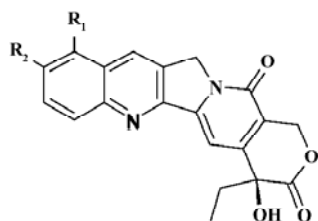


Figure 2. Structures of 9, 10-substituted CPT analogs.

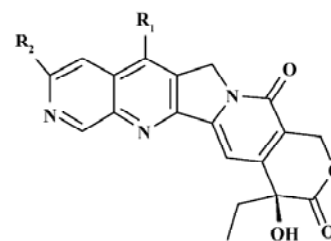


Figure 3. Structures of 7, 10-substituted 11-azaCPT analogs.

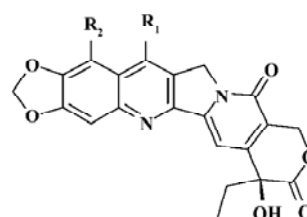


Figure 4. Structures of 7, 9-substituted 10, 11-(methylenedioxy) CPT analogs.

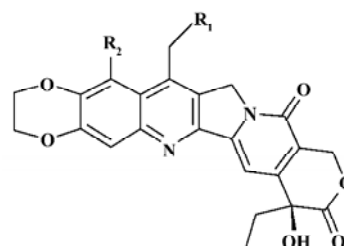


Figure 5. Structures of 7, 9-substituted 10, 11-(ethylenedioxy) CPT analogs.

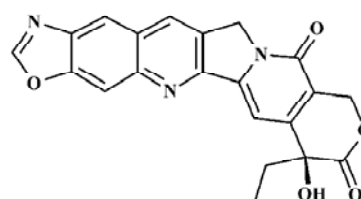
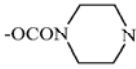
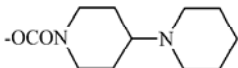


Figure 6. Structures of 10, 11-oxazole CPT analogs.

formation by 50%.

Because the TOP I/DNA/topotecan ternary complex (1K4T)^[8] with CPT analog (topotecan) was resolved, topotecan was selected as the starting structure; its starting geometry was extracted from the available X-ray crystal structure (PDB entry 1K4T)^[8] from the PDB. After correcting the atom type and bond type by using SYBYL6.8, it was minimized with the TRIPOS force field after being charged with the Gasteiger-Hückel charge. Then the prepared topotecan

Table 1. 9,10-Substituted CPT analogs.

No.	R ₁	R ₂	IC ₅₀ (10 ⁻⁹ ×mol/L)	PIC ₅₀	Reference
1 (CPT)	H	H	677	6.170	9
2 (TOPOTECAN)	-CH ₂ N(CH ₃) ₂	-OH	1046	5.980	9
3	H		4401	5.360	9
4 ^a	H	-Br	126	6.900	9
5	H	-COOH	1022	5.990	9
6	H	-Cl	141	6.850	9
7 ^a	H	-CN	1939	5.710	9
8	H		20546	4.690	9
9	H	-CH ₃	295	6.530	9
10	H	-F	368	6.430	9
11	H	-NH ₂	140	6.850	9
12	H	-NO ₂	635	6.200	9
13	H	-OH	106	6.970	9
14	-Cl	-Cl	108	6.970	9
15 ^a	-Cl	H	86	7.070	9
16	-F	H	163	6.790	9
17	-Me	H	38	7.420	9
18	-NH ₂	H	111	6.950	9
19	-OH	H	873	6.060	9

^aTest set compounds.

Table 2. 7, 10-Substituted 11-azaCPT analogs.

No.	R ₁	R ₂	IC ₅₀ (10 ⁻⁹ ×mol/L)	PIC ₅₀	Reference
20	H	H	383	6.420	11
21	-CH ₂ CH ₃	-COOC ₂ H ₅	9328	5.030	11
22	-CH ₂ CH ₃	-Br	122	6.910	11
23	-CH ₂ CH ₃	-CN	268	6.570	11
24 ^a	-CH ₂ CH ₃	-C ₃ H ₆ NC ₂ H ₅	949	6.020	11
25	-CH ₂ CH ₃	-CH ₂ NH ₂	167	6.780	11
26 ^a	-CH ₂ CH ₃	H	336	6.470	11
27 ^a	-CH ₂ CH ₃	-C(NH ₂)=NH	33	7.480	11
28	-CH ₂ CH ₃	-C(NH ₂)=NOH	71	7.150	11
29	-CH ₃	-Br	28	7.550	11
30	-CH ₃	-CN	143	6.840	11
31	-CH ₃	-CH ₂ NH ₂	232	6.630	11

^aTest set compounds.

was used as the template compound to construct other molecules by modifying and assembling fragments from the SYBYL standard library. Other molecules also constructed

were charged by the Gasteiger-Hückel charge and further minimized by the minimize 2 module in SYBYL6.8. As for CPT analog, 20S-isomer is much more potent than 20R form

Table 3. 7, 9-substituted 10, 11-(methylenedioxy) CTP analogs.

No.	R ₁	R ₂	IC ₅₀ (10 ⁻⁹ ×mol/L)	PIC ₅₀	Reference
32		H	416	6.380	12
33	H	H	27	7.570	9
34	-CH ₃ Cl	H	15	7.820	12
35	H	Cl	150	6.820	9
36	H	-NH ₂	61	7.210	9
37 ^a	H	-NO ₂	150	6.820	9
38		H	210	6.680	13
39		H	160	6.800	13
40		H	220	6.660	13
41		H	140	6.850	13
42		H	100	7.000	13
43		H	190	6.720	13

^a Test set compounds.

and is approximately twice as potent as the 20*RS* form, so only 20*S*-isomer CPT analogs with bioactivity were selected. All structures remained 20*S* configured after minimization.

Alignment rules The molecular alignment is another critical factor in the CoMFA calculation. Based on the known ternary complex, the docked structure was used for CoMFA study after the above-mentioned compounds were docked into the complex model using molecular simulation program Dock4.0 (Department of Pharmaceutical Chemistry, University of California, San Francisco, USA). The enzyme/DNA/topotecan complex, 1K4T^[8], was used to generate the receptor site and the energetic grid for the following docking calculations for molecule alignment. The docking procedure was applied as follows: the sphere centers (a set of overlapping spheres) of the topotecan binding site at a radius of 8 Å were identified by the sphgen program; then a box was created to enclose the spheres to be used for docking. Afterward, the energetic grid was created by a grid program using an all-

atom model and a distance-dependent dielectric function with a 10 Å cutoff; an anchor fragment orientation method was performed subsequently, and 50 conformations were produced per cycle. Finally, the dock energy score was used as the scoring function in the docking runs. Only the best scoring pose from each docking run was considered.

CoMFA^[20] CoMFA interaction potential energy fields were evaluated on a region (lattice) extending 4 Å in the X, Y and Z axes beyond the volume defined by the union of all molecules with a grid spacing of 2 Å. Standard CoMFA steric and electrostatic fields were calculated using distance-dependent dielectric constant with energy truncation values of 30 kcal/mol. The minimum-sigma (column filtering) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. PLS was carried out for compounds in the training set, except 10 compounds randomly selected as the test set, cross-validated with leave-one-out.

Table 4. 7, 9-Substituted 10, 11-(ethylenedioxy) CPT analogs.

No	R ₁	R ₂	IC ₅₀ (10 ⁻⁹ ×mol/L)	PIC ₅₀	Reference
44		H	300	6.520	12
45	-CH ₃ Cl	H	11	7.960	12
46 ^a		H	190	6.720	13
47		H	200	6.700	13
48		H	180	6.740	13
49		H	250	6.600	13
50		H	600	6.220	13
51 ^a		H	520	6.280	13
52		H	340	6.470	13

^a Test set compounds.

Table 5. 10, 11-Oxazole CPT analogs.

No.	IC ₅₀ (10 ⁻⁹ ×mol/L)	PIC ₅₀	Reference
53	150	6.820	10

Results and discussion

The CoMFA model was constructed with 43 compounds in the training set, which was validated by the remaining molecules in the test set.

PLS model The PLS technique^[21-23] was employed to generate a linear relationship that correlates changes in the computed steric and electrostatic potential fields with changes in the corresponding experimental values of the bioactivity (PIC₅₀) for the data set of ligands. Employing the CoMFA potential energy fields for each molecule as the independent variable and the corresponding PIC₅₀ values as the dependent variable, PLS converts the steric and electro-

static field descriptors to so-called latent variables or principle components (PC) that consist of linear combinations of the original independent variables.

To assess the internal predictive ability of the CoMFA models, we employed a 'leave-one-out' cross-validation procedure. In this procedure, each compound is excluded one at a time, after which its bioactivity is predicted by the model constructed from the remaining compounds in the data set. Cross-validation determines the optimum number of PC, corresponding to the smallest error of prediction and the highest cross-validated q^2 (or r^2_{cv}). PLS analysis was repeated without validation using the optimum number of PC to generate a final CoMFA model from which the conventional r^2 , a measure of the internal consistency of the model, was derived. The result using this procedure is described as follows.

After cross-validated PLS analysis was carried out for the training set, the number of optimal PC at 6 was obtained with cross-validated q^2 of 0.495 (Table 6). The non-cross-validated PLS analysis with the optimum PC revealed a con-

ventional r^2 value of 0.935, $F=86.748$, and an estimated standard error of 0.176. Furthermore, the obtained model showed high predictability. As illustrated in Table 7, the predicted values were very close to the observed values. Nine of 10 compounds in the test set were predicted well in terms of activity, except compound 7. The predicted r^2 of 9 compounds was 0.58, which showed good correlation. So the 3D QSAR model is predictable, and results of the obtained model are summarized in Tables 6 and 7.

Table 6. Computation results of CoMFA model.

Q^2	R^2	NOC ^a	SE ^b	F^c	RC-S ^d	RC-E ^e
0.495	0.935	6	0.176	86.748	0.651	0.349

Table 7. Predicted activity vs experimental activity of CPT analogs in the test set.

No.	Exp	Pre	Res
4	6.900	6.894	0.006
7	5.710	6.815	-1.105
15	7.070	6.524	0.546
24	6.020	5.700	0.320
26	6.470	6.997	-0.527
27	7.480	6.810	0.670
37	6.820	6.980	-0.160
40	6.660	6.508	0.152
46	6.720	6.402	0.318
51	6.280	6.529	-0.249

The QSAR model also gave the relative electrostatic contribution and relative steric contribution. In this model, the steric contribution was much higher than the electrostatic contribution in which the steric field descriptors explained 65.1% of the variance, while the electrostatic descriptors explained 34.9%. These are in accordance with the 'drug-stacking' hypothesis.

Visualization of 3-D contour maps To visualize the information content of the derived 3D QSAR models, CoMFA contour maps were generated by interpolating the products between the 3D QSAR coefficients and their associated standard deviations. The green contours represented the regions of high steric tolerance, while the yellow contours represented regions of unfavorable steric effects. The blue contours described regions where a positively charged group enhanced activity, while the red contours described regions

where a negatively charged group enhanced activity. To aid in visualization, topotecan was displayed in the maps. 3-D contours were mostly found around 7, 9, 10, and 11 positions of CPT analogs, because most substituents of CPT analogs are focused on this position. However, there was a contour at the southeast region in Figure 7A, where there was no substituent near the 14-position. Because of docking CPT analogs into complexes according to the alignment rule, the compounds with methylenedioxy and enthylenedioxy substituents with high activity rotated a little to right down. It seemed that there was a substitute in the vicinity of the 14-position. So it is an artifact in which more bulky substituents in the 14-position will improve activity. The steric CoMFA map (Figure 7 A) illustrates a favorable region of steric interaction at the 10 and 11 positions, but much more bulky substituents will decrease activity^[15]. The compounds with methylenedioxy and enthylenedioxy substituents (compounds 32–52) exhibit considerable increase of activity, while compounds 3, 8 and 21 decrease their activity. At the same time, there is another favorable steric region at the 7-position, indicating that more bulky substituents are needed in this position to improve the inhibitive activity, which is consistent with the fact that silyl-substituted compounds display more strong potency^[14]. The prediction is confirmed by the molecular surface of the binding pocket which is big enough to accommodate more bulky substituents which will conflict with Lys425 if they are too large. More negative substituents will increase inhibitive activity at positions 9, 10, and 11, which is in agreement with previous observations^[9,10,16]. It is also in agreement with the mapped enzyme/DNA/topotecan complex because positively-charged residue Lys425 is in the direction of positions 9 and 10. The hydrogen bond between 10-hydroxyl of topotecan and a water molecule^[8] also contributed to the negative contour map. Furthermore, CPT analogues were stacked into base pairs, like C112 (Figure 7B) and A113 (not shown in Figure for clarity) and other base pairs (not shown) in this ternary complex, which might show slight positive charge. In addition, more positive substituents at position 7 would strengthen the binding of the inhibitors to top the I-DNA complex, so alkyl-substituents at position 7 possess considerable activity, and silyl-substituents show more inhibitive activity^[14], which should be correlative with the hydroxyl of base T10. There is a green contour on the left because of the negative charged phosphodiester. The CoMFA contour maps are illustrated in Figure 7.

Furthermore, the steric interaction at positions 7, 9, 10, and 11 can be interpreted using the enzyme/DNA/topotecan ternary complex with CPT analogs (topotecan)^[8] in Figure 8.

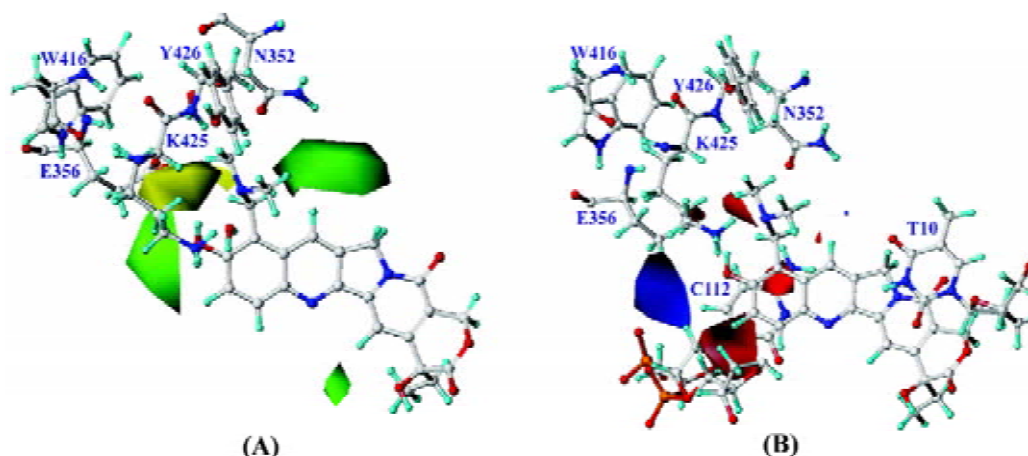


Figure 7. CoMFA contour maps displayed with compound 2 (topotecan) and related residues in the binding site of the Top I-DNA complex (A), steric field distribution, and (B) electrostatic field distribution.

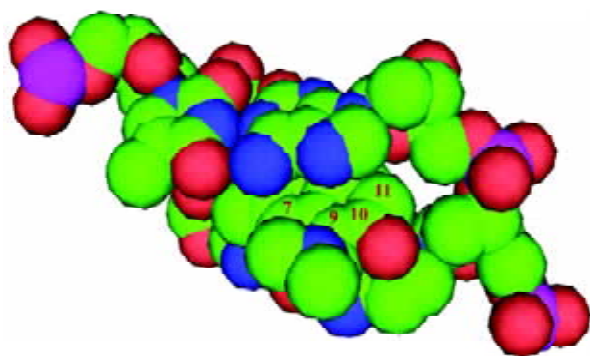


Figure 8. Corey-Pauling-Koltun (CPK) model of the ternary cleavable complex structure viewed facing the DNA major groove (topotecan).

The topotecan molecule is oriented with the E ring near the DNA break, and the concave portion of the drug molecule faces the DNA major groove^[17]. The structural model demonstrates that 7, 9, 10, and 11 positions face the major groove of the DNA, and the substituents on these positions point toward the sugar backbone. So the model suggests that electron-rich groups of moderate or large size would be favored for these positions.

Conclusion

A study on the quantitative structure–activity relationship with CoMFA for a series of CPT analogs was performed successfully, and a good cross-validated correlation was obtained with q^2 of 0.495. Then the PLS model non-cross-validated was built and permitted demonstrations of high

predictability for the activities of the CPT analogs in the test set selected randomly. The CoMFA contour maps illustrated that more negative-charged large group at positions 9, 10, and 11 would increase activity, but excessively increasing the bulky group at position 10 would decrease activity; substituents that occupy position 7 with the bulky positively-charged group will enhance inhibitive activity.

References

- 1 Wall ME, Wani MC, Cook CE, Palmer KH, McPhail AT, Sim GA. Plant antitumor agents I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibitor from *camptotheca acuminata*. *J Am Chem Soc* 1966; 88: 3888–90.
- 2 Tanizawa A, Fujimori A, Fujimori Y, Pommier Y. Comparison of topoisomerase I inhibition, DNA damage, and cytotoxicity of camptothecin derivatives presently in clinical trials. *J Natl Cancer Inst* 1994; 86: 836–42.
- 3 Verweij J. Topoisomerase I inhibitors and other new cytotoxic drugs. *Eur J Cancer* 1995; 31: 828–30.
- 4 Moertel CG, Schutt AJ, Reitemeier RJ, Hahn RG. Phase II study of camptothecin (NSC-100880) in the treatment of advanced gastrointestinal cancer. *Cancer Chemother Rep* 1972; 56: 95–101.
- 5 Hsiang YH, Hertzberg R, Hecht S, Liu LF. Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J Biol Chem* 1985; 260: 14873–8.
- 6 Cragg GM, Newman DJ. A tale of two tumor targets: topoisomerase I and tubulin. The Wall and Wani contribution to cancer chemotherapy. *J Nat Prod* 2004; 67: 232–44.
- 7 SYBYL. Version 6.8. St Louis (MO): Tripos Associates; 2000.
- 8 Staker BL, Hjerrild K, Feese MD, Behnke CA, Burgin AB, Stewart LJ. The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc Natl Acad Sci USA* 2002; 99: 15 387–92.
- 9 Wall ME, Wani MC, Nicholas AW, Manikumar G, Tele C, Moore

- L, *et al.* Plant antitumor agents 30. Synthesis and structure activity of novel camptothecin analogs. *J Med Chem* 1993; 36: 2689–700.
- 10 Carrigan SW, Fox PC, Wall ME, Wani MC, Bowena JP. Comparative molecular field analysis and molecular modeling studies of 20-(*S*)-camptothecin analogs as inhibitors of DNA topoisomerase I and anticancer/antitumor agents. *J Comput Aided Mol Des* 1997; 11: 71–8.
- 11 Uehling DE, Nanthakumar SS, Croom D, Emerson DL, Leitner PP, Luzzio MJ, *et al.* Synthesis, topoisomerase I inhibitory activity, and *in vivo* evaluation of 11-azacampthotecin analogs. *J Med Chem* 1995; 38: 1106–18.
- 12 Luzzio MJ, Besterman JM, Emerson DL, Evans MG, Lackey K, Leitner PL, *et al.* Synthesis and camptothecin antitumor activity of novel water soluble derivatives of camptothecin as specific inhibitors of topoisomerase I. *J Med Chem* 1995; 38: 395–401.
- 13 Lackey K, Sternbach DD, Croom DK, Emerson DL, Evans MG, Leitner PL, *et al.* Water soluble inhibitors of topoisomerase I: quaternary salt derivatives of camptothecin. *J Med Chem* 1996; 39: 713–9.
- 14 Bom D, Curran DP, Kruszewski S, Zimmer SG, Strode JT, Kohlhaagen G, *et al.* The novel silatecan 7-*tert*-butyldimethylsilyl-10-hydroxycamptothecin displays high lipophilicity, improved human blood stability, and potent anticancer activity. *J Med Chem* 2000; 43: 3970–80.
- 15 Thomas CJ, Rahier NJ, Hecht SM. Camptothecin: current perspectives. *Bioorg Med Chem* 2004; 12: 1585–604.
- 16 Kim DK, Ryu DH, Lee JY, Lee N, Kim YW, Kim JS, *et al.* Synthesis and biological evaluation of novel A-ring modified hexacyclic camptothecin analogues. *J Med Chem* 2001; 44: 1594–602.
- 17 Fan Y, Weinstein JN, Kohn KW, Shi LM, Pommier Y. Molecular modeling studies of the DNA-topoisomerase I ternary cleavable complex with camptothecin. *J Med Chem* 1998; 41: 2216–26.
- 18 Yoon KJP, Krull EJ, Morton CL, Bornmann WG, Lee RE, Potter PM, *et al.* Activation of a camptothecin prodrug by specific carboxylesterases as predicted by quantitative structure-activity relationship and molecular docking studies. *Mol Cancer Ther* 2003; 2: 1171–81.
- 19 Zhu J, Zhu W, Ji H, Zhou Y, Zhu J, Lü J. Comparative molecular field analysis (CoMFA) of camptothecin analogs. *Chin J Med Chem* 1999; 9: 277–84.
- 20 Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA) 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 1988; 110: 5959–67.
- 21 Wold S, Albano C, Dunn WJ, Edlund U, Esbensen K, Geladi P, *et al.* Multivariate data analysis in chemistry. In: Kowalski B, editors. *CHEMOMETRICS: Mathematics and statistics in chemistry*. Dordrecht, Netherlands: Reidel; 1984.
- 22 Stahle L, Wold S. Multivariate data analysis and experimental design in biomedical research. *Prog Med Chem* 1988; 25: 292–338.
- 23 Esbensen KH, Wold S, Geladi P. Relationships between higher-order data array configurations and problem formulations in multivariate data analysis. *J Chemometrics* 1988; 3: 33–48.