Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry
Bioorganic & Medicinal Chemistry 14 (2006) 6924–6932

Comparative molecular field analysis and comparative molecular similarity indices analysis of boron-containing human thymidine kinase 1 substrates☆

Achintya K. Bandyopadhyaya,* Rohit Tiwari and Werner Tjarks

Division of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, The Ohio State University, Columbus, OH 43210, USA

Received 17 April 2006; revised 14 June 2006; accepted 19 June 2006
Available online 7 July 2006

Abstract—Three-dimensional quantitative structure–activity relationship (3D-QSAR) using CoMFA and CoMSIA techniques was applied to evaluate 56 pyrimidine nucleosides as substrates of human thymidine kinase 1 (hTK1), 27 of them containing a carborane substituent either at the 3-, 5-, or 3′-position of the 2′-deoxyuridine scaffold. This is the first report describing 3D-QSAR studies of compounds containing boron atoms. Both CoMFA and CoMSIA models were derived from a training set of 47 molecules and the predictive capacity of the CoMSIA model was successfully validated by accurately calculating known phosphorylation rates of both boronated and non-boron hTK1 substrates that were not included in the training set. The optimal CoMSIA model provided the following values: q^2 0.622, r^2 0.983, s 0.165, and F 187.5. Contour maps obtained from the CoMSIA model were in agreement with the experimentally determined biological data.

1. Introduction

Boron neutron capture therapy (BNCT) is a binary system for the treatment of cancer. In order for this therapy to be effective, the targeted cancer cells must attain a sufficient concentration of ^{10}B, a stable isotope that makes up approximately 20% of natural elemental boron. At the time of treatment, this localized boron content is activated by a suitable flux of low energy thermal neutrons. The nuclear reaction [$^{10}B(n,\alpha)^{7}Li$] that is initiated by neutron capture yields high linear energy transfer (LET) particles, $^{4}He^{2+}$ and $^{7}Li^{3+}$ions. These particles have a range of 9 μm and 5 μm, respectively, in biological tissue and can cause cell death primarily by inflicting DNA damage. However, the damage is largely restricted to the tumor because of the limited range of these particles. Provided that ^{10}B is selectively taken up by the tumor, the amount of this nuclide required to sustain lethal tumor cell damage has been calculated to be in the range of 15–30 μg per gram of tumor, assuming a homogeneous distribution throughout all intra- and extracellular compartments.[1,2] Of pivotal importance for BNCT is the development of ^{10}B delivery systems that selectively target tumor cells. In recent years, our group has developed 3-carboranyl thymidine analogues (3CTAs), a novel class of nucleoside analogue prodrugs, for the treatment of cancer by BNCT.[3–9] The biochemical basis for the tumor cell selectivity of 3CTAs appears to be their capacity to function as substrates for human thymidine kinase 1 (hTK1), a deoxynucleoside kinase that is widely expressed only in proliferating cells[10] and that presumably causes selective intracellular entrapment of 3CTAs in tumor cells by 5′-monophosphorylation.[4,6]

Recently, crystal structures of hTK1,[11,12] *Ureaplasma urealyticum* TK (*Uu*TK),[12,13] and *Clostridium acetobutylicum* TK (*Ca*TK) [PDB ID: 1XX6] were determined and are now available for computational structure-based design of improved 3CTAs. Computational studies carried out in our laboratories,[3] however, indicated

Abbreviations: BNCT, boron neutron capture therapy; 3CTAs, 3-carboranyl thymidine analogues; hTK1, human thymidine kinase 1; Thd, thymidine; PR, phosphorylation rate; CoMFA, comparative molecular field analysis; CoMSIA, comparative molecular similarity indices analysis; 3D-QSAR, three-dimensional quantitative structure–activity relationship; dUrd, deoxyuridine; MMplus, molecular mechanics force field; PLS, partial least square analysis; H-bond, hydrogen bond.

Keywords: Boron neutron capture therapy (BNCT); 3D-QSAR, CoMFA; CoMSIA; 3-carboranyl thymidine analogues (3CTAs); Human thymidine kinase 1 (hTK1); Phosphorylation.

☆ The results described in this paper were partially presented at the 229th ACS National Meeting, San Diego, CA, USA, March 13–17, 2005.

* Corresponding author. Tel.: +614 688 8618; fax: +614 292 2435; e-mail: bandyopadhyaya.1@osu.edu

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.06.037

that binding of thymidine (Thd) and adenosine triphosphate (ATP) in TKs may be associated with a conformational change from an open unoccupied form, over a partially closed form involving Thd or ATP binding, to a closed form binding both Thd and ATP. This type of conformational changes in nucleoside kinases has been identified as a major obstacle in the structure-based in silico design of novel agents.[14] Most commercial in silico software packages do not provide default parameter for calculations involving molecules with hexavalent boron atoms in cage structures, and therefore, only very rudimentary structure-based strategies for the in silico design of carborane containing agents have been developed in recent years.[15–22] This could limit the application of structure-based design of novel 3CTAs because these in silico techniques are probably inadequate for sophisticated computational molecular dynamic studies of different stages of binding of 3CTAs to various dynamically changing hTK1 conformers.

Another major problem for BNCT drug discovery processes is the lack of databases with biological and structural information of boron-containing agents that would allow the ligand-based in silico design of BNCT agents.[23] In recent years, however, several 3CTA libraries have been prepared and were evaluated in phosphoryl transfer assays (PTAs) with recombinant hTK1 in order to obtain phosphorylation rates (PRs), a measure for the substrate characteristics of 3CTAs.[3,8,24,25]

In the present study, these PRs have been compiled into a database, which was then used for the development of innovative three-dimensional quantitative structure–activity relationship (3D-QSAR)-based in silico techniques,[26–33] including comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). These should be suitable for the ligand-based design of 3CTAs with improved biological properties. However, the general 3D-QSAR concept developed in this paper should also be applicable for the in silico design of a wide variety of other boron-containing pharmaceuticals.

2. Results and discussion

The training set for CoMFA and CoMSIA contained 47 compounds. The logarithms of the 'relative' hTK1 PRs of compounds **1**–**47** (Tables 1 and 2) were used as the dependent variable in the 3D-QSAR analyses. High PRs are indicative of good hTK1 substrate characteristics, while low PRs are indicative of poor hTK1 substrate characteristics. They are currently the only biological data available for a large number of hTK1 substrates, have been used previously by us in similar 3D-QSAR studies,[24] and were all obtained by Eriksson and coworkers applying the same experimental protocol.[3,24,39,40]

2.1. CoMFA results

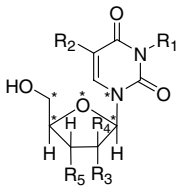
The statistical data obtained from the standard CoMFA model constructed with steric and electrostatic fields are depicted in Table 3. The optimal number of components (6) was determined using the SAMPLS analysis implemented in Sybyl 7.1 with a leave-one-out (LOO) cross-validated q^2 of 0.576. A non-cross-validated correlation coefficient (r^2) of 0.922 was obtained for this model. Predictions for the log(PR)s with residuals less than 0.4 log unit for 8 holdout test compounds out of 9 were obtained. However, in the training set the calculated ΔPR values for compound **7** and endogenous hTK1 substrates, dUrd (**10**) and Thd (**13**), were significantly higher (>50), while the ΔPR values for compounds **2** and **26** were 109.0 and 244.0, respectively (see supplementary data). This model also failed to predict the poor hTK1 substrate characteristics of **27** and produced more than 50% deviation in the prediction of the PR values of **24** and **36** of the test set (Table 5). Based on the quality of the predicted data for both the training set and the test set, CoMFA was not further pursued for the characterization of the library of hTK1 substrates.

A crucial factor for the success of a CoMFA modeling is correct structural alignment. The optimal CoMFA alignment of a ligand is widely assumed to be its experimental *receptor-bound conformation*.[34] As already indicated in Section 2.3, the energy-minimized ribose portion of thymidine (**13**) was used for structural alignment of compounds **1**–**56** (Table 1). However, the extreme heterogeneity among carboranyl and other N3-substituentes in the majority of compounds in our library made an appropriate alignment of this structural component impossible. Also, alignment of molecules **1**–**56** using the coordinates of the ribose moiety of thymidine triphosphate (TTP) bound to the active site of the hTK1 crystal structure (PDB ID#1W4R)[11] caused even further deterioration of the CoMFA model (data not shown). The obtained results clearly indicate that alignment-dependent CoMFA is not suitable for the evaluation of a compound library encompassing a large number of molecules with a highly diverse N3 substitution pattern. Klebe et al.[35] reported that CoMSIA produced significantly better results than alignment-sensitive CoMFA for their studies with inhibitors of the benzamidine type with respect to their binding affinities toward thrombin, trypsin, and factor Xa. Similar observation have also been reported by Sutherland et al.[36] for their QSAR studies on acetylcholineesterase (AchE) inhibitors, benzodiazepine (BZR) receptor binding, and cyclooxygenase-2 (COX2) inhibitors.

2.2. CoMSIA results

Eight models were constructed by varying the steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen-bond acceptor (A) descriptor fields (Table 3). The optimal components of each of these models were determined by SAMPLS analysis implemented in Sybyl 7.1 with maximum q^2 values. The model developed by using only steric and electrostatic fields (SE model) produced a cross-validated q^2 of 0.341 with 2 latent variables. The corresponding conventional r^2 value (0.491) was very low. The quality of the SE model improved by adding H-bond donor/acceptor fields. Both the q^2 and r^2 values increased when these

Table 1. Non-boronated molecules for CoMFA and CoMSIA modeling



Compound	R_1	R_2	R_3	R_4	R_5	PR[a]	log(PR)
1	Me	Me	H	H	OH	44.3	1.65
2	Et	Me	H	H	OH	61.4	1.79
3	$CH_2C{\equiv}CH$	Me	H	H	OH	102.1	2.01
4	*n*-Bu	Me	H	H	OH	84.9	1.93
5	CH_2Ph	Me	H	H	OH	82.3	1.92
6	$CH_2CHOHCH_2OH$	Me	H	H	OH	27.5	1.44
7	H	H	H	H	N_3	70.0	1.85
8	H	Me	H	H	N_3	52.0	1.72
9	H	Me	H	H	CH_2N_3	15.0	1.18
10	H	H	H	H	OH	77.0	1.89
11	H	Me	H	F	OH	45.0	1.65
12	H	Et	H	H	OH	80.0	1.90
13(Thd)	H	Me	H	H	OH	100.0	2.00
14	H	I	H	F	OH	42.0	1.62
15	H	Br	H	H	OH	80.0	1.90
16	H	H	H	OH	OH	1.0	0.00
17	H	CH=CHBr	H	H	OH	1.0	0.00
18	H	CH_2CH_2Cl	H	H	OH	5.0	0.70
19	H	NH_2	H	H	OH	3.0	0.48
20	H	H	OH	H	OH	0.1	−1.00
21	H	H	H	H	OMe	0.1	−1.00
22	H	H	H	H	OEt	0.1	−1.00
23	H	H	F	F	OH	0.1	−1.00
24	*i*-Pr	Me	H	H	OH	69.6	1.84
25	H	Me	H	H	F	30.0	1.48
26	H	F	H	H	OH	95.0	1.98
27	H	Me	H	OH	OH	1.0	0.00
28	H	Me	OH	H	OH	2.0	0.30
29	H	Me	$=CH_2$		OH	0.1	−1.00

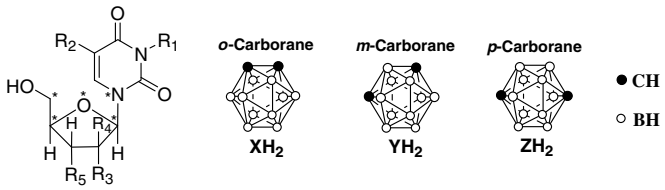
*These atoms were used for the alignment of the molecules. Molecules **18**, **24**, **27**, and **28** were used as test set structures.
[a] The values are given in % relative to that of Thd, which is set at 100%. PR values of all the compounds in this table were previously published.[24]

fields were added. Addition of the H-bond acceptor fields (SEA model) to the SE model produced q^2 and r^2 values of 0.426 and 0.878, respectively, whereas the SE model including H-bond donor fields generated q^2 and r^2 values of 0.349 and 0.548, respectively. It was also observed that the quality of the model slightly improved when hydrophobic fields were included as one of the descriptors to the SE model (SEH model). Adding the hydrophobic field descriptor to SEA and SED models showed the same trend by increasing q^2 and r^2 values. The model with the entire set of five descriptors (ALL) showed a q^2 of 0.518 at 6 components and the corresponding r^2 was 0.922. The SEDA model gave the highest cross-validated coefficient q^2 (0.637) with 13 components, thereby indicating the highest predictive capacity with the lowest SEE (0.145) and the highest *F*-test value (206.0). The highest squared correlation coefficient r^2 (0.984) was also obtained for the SEDA model showing a strong internal consistency, Therefore, this model was chosen for further analysis.

2.3. Progressive scrambling

The chosen CoMSIA model was further refined and validated by progressive scrambling, which probes the dependence of the model on chance correlations.[37,38] The standard method of cross-validation (SAMPLS) encompasses redundant data sets, which potentially can overstate q^2 values. Progressive scrambling determines the sensitivity of the generated PLS model to small systematic perturbations of the response variables at a critical level of perturbation. After removal of these redundancies, the introduced noise in form of small systematic perturbations solidifies the statistics with the result that q^2 values as low as 0.35 at the critical point are indicative of the stability of the original unperturbed model. The obtained q^2 values (evaluated at the critical point) are plotted against the correlation of the original dependent variables versus the perturbed dependent variables ($r^2_{yy'}$). The effective slope ($dq^{2'}/dr^2_{yy'}$) should be near unity (0.8–1.2) for a stable model that changes

Table 2. Carborane-containing molecules for CoMFA and CoMSIA modeling



Compound	R_1	R_2	R_5	PR[a]	log(PR)
30	H	Me	XH	8.1	0.91
31	H	Me	CH_2XH	0.8	−0.10
32	H	XH	OH	0.01	−2.00
33	H	$XCH_2CH(OH)CH_2OH$	OH	0.01	−2.00
34	CH_2XH	Me	OH	10.0	1.00
35	$(CH_2)_2XH$	Me	OH	39.0	1.59
36	$(CH_2)_3XH$	Me	OH	30.0	1.48
37	$(CH_2)_4XH$	Me	OH	13.0	1.11
38	$(CH_2)_5XH$	Me	OH	41.0	1.61
39	$(CH_2)_6XH$	Me	OH	28.0	1.45
40	$(CH_2)_7XH$	Me	OH	11.0	1.04
41	$(CH_2)_2YNH_2$	Me	OH	71.9	1.86
42	$(CH_2)_3YNH_2$	Me	OH	58.2	1.76
43	$(CH_2)_4YNH_2$	Me	OH	58.0	1.76
44	$(CH_2)_5YNH_2$	Me	OH	45.1	1.65
45	$(CH_2)_6YNH_2$	Me	OH	14.2	1.15
46	$[(CH_2)_2O](CH_2)_2XH$	Me	OH	39.0	1.59
47	$[(CH_2)_2O]_2(CH_2)_2XH$	Me	OH	38.0	1.58
48	$[(CH_2)_2O]_3(CH_2)_2XH$	Me	OH	42.0	1.62
49	$[(CH_2)_2O]_3(CH_2)_2ZH$	Me	OH	40.0	1.60
50	$[(CH_2)_2O]_4(CH_2)_2XH$	Me	OH	37.0	1.57
51	$(CH_2)_2XCH_2CH(OH)CH_2OH$	Me	OH	45.0	1.65
52	$(CH_2)_3XCH_2CH(OH)CH_2OH$	Me	OH	40.0	1.60
53	$(CH_2)_4XCH_2CH(OH)CH_2OH$	Me	OH	21.0	1.32
54	$(CH_2)_5XCH_2CH(OH)CH_2OH$	Me	OH	41.0	1.61
55	$(CH_2)_6XCH_2CH(OH)CH_2OH$	Me	OH	32.0	1.51
56	$(CH_2)_7XCH_2CH(OH)CH_2OH$	Me	OH	13.0	1.11

[a] The values are given in % relative to that of Thd, which is set at 100%. PRs of compounds **30**–**31**,[39] **32**–**33**,[40] and **34**–**56**[3] were obtained from the literature as indicated. Molecules **36**, **39**, **44**, **49**, and **54** were used as test set structures.

Table 3. Summary of CoMFA and CoMSIA results for various models

	CoMFA	CoMSIA (SE)	CoMSIA (SEH)	CoMSIA (SEA)	CoMSIA (SEHD)	CoMSIA (ALL)	CoMSI A (SEDA)	CoMSIA (SED)	CoMSIA (SEHA)	CoMSIA (FINAL)
Components:	6	2	2	6	2	6	13	2	6	11
q^2	0.576	0.341	0.421	0.426	0.404	0.518	0.637	0.349	0. 468	0.622
r^2	0.922	0.491	0.547	0.878	0.590	0.922	0.988	0.548	0.908	0.983
SEE	0.333	0.814	0.768	0.418	0.731	0.333	0.145	0.767	0.363	0.165
F	79.3	21.2	26.6	48.1	31.7	79.3	206.0	26.7	65.6	187.5
Field contributions(%)										
S	0.631	0.409	0.253	0.243	0.175	0.126	0.157	0.262	0.181	0.155
E	0.369	0.591	0.362	0.517	0.259	0.215	0.335	.381	0.317	0.322
H			0.385		0.300	0.242			0.318	
D					0.266	0.257	0.340	0.357		0.349
A				0.240		0.161	0.168		0.184	0.174

q^2, squared cross-validated coefficient; r^2, squared conventional coefficient; SEE, standard error of estimate; *F*, *F*-test value; S, E, H, A, D, denote steric, electrostatic, hydrophobic, as well as hydrogen-bond acceptor and donor fields, respectively; ALL denotes the combination of all fields.

proportionally with small perturbations. Results of the progressive scrambling applied to the chosen SEDA model are presented in Table 4. All SEDA models with components 9–14 showed slopes near unity. However, based on the highest q^2 (0.386) and the optimal cSDEP (1.000), the model with 11 components was chosen as

Table 4. Results of progressive scrambling of CoMSIA models

Components	1	10	11	12	13	14
q^2	0.364	0.378	0.386	0.385	0.374	0.353
cSDEP	0.989	0.992	1.000	1.015	1.039	1.072
$dq^{2'}/dr^2_{yy'}$	0.944	1.001	1.012	0.993	0.984	1.001

q^2, predictivity at the critical point (set to 0.88); cSDEP, standard error of prediction at the critical point; $dq^{2'}/dr^2_{yy'}$, sensitivity to perturbation.

the final CoMSIA model. It is noteworthy that the model with 12 components produced comparable data for q^2 (0.385) and cSDEP (1.015).

The scatter plot of experimental versus predicted log(PR)s for the final CoMSIA model is presented in Figure 1. The predicted Δlog(PR) for the training set molecules were within 0.3 log units, except for compounds **11** and **17** (see supplementary data). The predicted Δlog(PR) for the test set compounds were also within 0.3 log units, except for the compound **28** (Table 5). However, the experimentally determined PR of **28** was 2.0, while the predicted PR was 0.10. Thus, the poor hTK1 substrate characteristic of this compound was still accurately predicted. Experimental and predicted PRs for all the compounds are reported in the supplementary data. PRs were calculated from the log(PR)s in order to

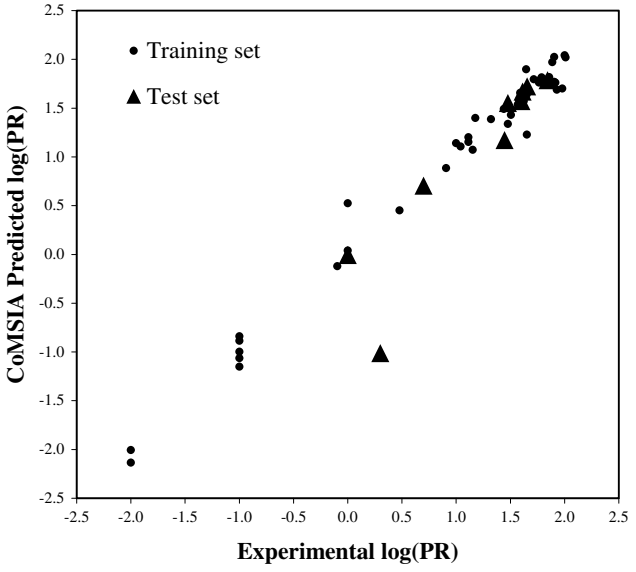


Figure 1. Experimental versus predicted log(PR)s of compounds in the training set and the test set for the CoMSIA model.

better correlate the numbers with the actual experimental PRs. The scatter plot of experimental versus predicted log(PR)s shows higher linearity for the CoMSIA model compared to the abandoned CoMFA model. This result was expected based on the alignment-independent nature of CoMSIA.[35,36]

2.4. CoMSIA contour maps

Figure 2A shows the steric contour maps for the CoMSIA with dUrd (**10**) as a reference structure. A small sterically favorable green contour combined with a significantly larger yellow contour near C-5 indicates that substituents at this position should not exceed significantly the size of the methyl group in Thd (**13**). Consequently, compounds **13**, **15**, and **26** with methyl, bromo and fluoro substitution at C-5 have PRs higher than dUrd (**10**), while compounds with ethyl (**12**), bromovinyl (**17**), chloroethyl (**18**), and *o*-carboranyl (**32**, **33**) substituents at C-5 have lower PRs than **13** (Thd). The 3-benzyl group of **5** and the 3-butyl group of **4** correspond with the sterically favorable green contour near the N-3 position of the pyrimidine ring implying that these substitutions are favored compared with the 3-Me (**1**), 3-Et (**2**), and the 3-^{i}Pr (**27**) groups. The large yellow contour in the vicinity of the 3′-position of the deoxyribose sugar suggests that any bulky substitution at this location results in deteriorated PRs as can be seen by the low PRs of compounds **21**, **22**, **30**, and **31**. A small yellow sterically disfavorable contour is seen at a distance of approximately four C—C bonds from N-3. This may indicate that a certain steric orientation of carboranyl substituents in 3CTAs is important for their biological activities as demonstrated by the low PRs of **37** and **43**, in which cases carborane clusters and yellow isopleth overlap.

The electrostatic contours Figure 2B generated by CoMSIA indicate that small electronegative groups are favored near C-5 of pyrimidine ring (red contour), as demonstrated by relatively high PRs for compounds with fluorine (**26**) and bromine (**15**) at this position. A large blue contour near N-3 indicates that the most favorable substituent at that position should be a hydrogen atom while any other type of substitution may lead to decreased biological activity. In this context, it should be noted that the biological activities of 3CTAs are remarkable, but they do, of course, not compare with that of endogenous Thd. Blue (two) and red (one) con-

Table 5. Experimental versus predicted PR values of the test set molecules

Compound	Experimental data		Predicted CoMFA data				Predicted CoMSIA data			
	PR	log(PR)	log(PR)	Calcd PR	Δlog(PR)	ΔPR	log(PR)	Calcd PR	Δlog(PR)	ΔPR
18	5.00	0.70	0.69	4.84	0.01	0.16	0.70	5.06	−0.01	−0.06
24	69.60	1.84	1.54	34.44	0.31	35.16	1.79	61.21	0.06	8.39
27	1.00	0.00	1.31	20.24	−0.31	−19.24	−0.01	0.98	0. 01	0.02
28	2.00	0.30	0.03	1.06	0.28	0.94	−1.01	0.10	1.31	1.90
36	30.00	1.48	1.83	67.96	−0.36	−38.00	1.55	35. 62	−0.08	−5.62
39	28.00	1.45	1.16	14.28	0.29	13.72	1.17	14.89	0.27	13.11
44	45.10	1.65	1.61	40.47	0.05	4.63	1.72	53.03	−0.07	−7.93
49	40.00	1.60	1.59	38.65	0.02	1.35	1.57	37.40	0.03	2.60
54	41.00	1.61	1.61	40.26	0.01	0.74	1.67	46.63	−0.06	−5.63

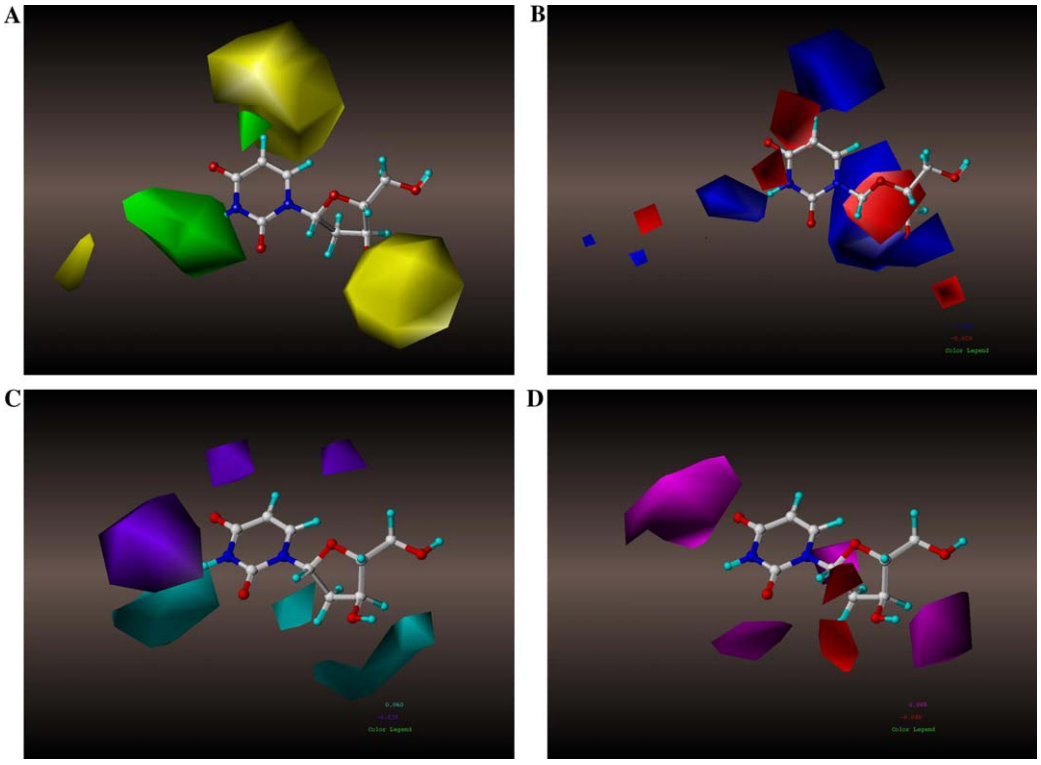


Figure 2. CoMSIA StDev × Coeff contour plots for steric (A), electrostatic (B), H-bond donor (C), and H-bond acceptor (D) fields with **10** (dUrd) as the reference. (A) Sterically favored and disfavored areas are in green and yellow. (B) Positive charge favored and disfavored areas are in blue and red. (C) Favored and disfavored H-bond donor regions are in cyan and purple. (D) Favored and disfavored H-bond acceptor regions are in magenta and red.

tours were observed at a distance of four to six C—C bonds from the N-3 position. Boron atoms as well as the carbon atoms of the carborane cage have negative or close to zero AM1 charges, and therefore, the carborane cage as a whole has a slightly electronegative character. The two blue contours are seen within the dimensions of the carborane cages of compounds **34**, **37**, **40**, and **53**. Thus, they have low PRs as 10.0, 13.0, 11.0, and 21.0, respectively. The red contour coincides with the carborane cages of **38** and **46**, both having higher PRs than the former 3CTAs. Among the 3CTAs with amino-substituted *m*-carborane (**41**–**45**), the analysis of the contour maps was based on the observation that the amino group has an electron withdrawing effect on the entire cage structure increasing its electropositivity compared with unsubstituted carboranes. This observation is supported by the AM1 charges of these molecules as well additional ab initio (3-21 G basis set) calculations on parental *m*-carboranyl 3CTAs with unaltered boron atom types (data not shown). Therefore, in the case of **43** the blue contour inside the amino carborane cage corresponds with higher activity compared with other 3CTAs having a butylene spacer (**37**, **50**, and **53**) The red contour is located outside of the carborane cage **42** in very close proximity to the amino group, indicating that electronegative substituents are favored at this position. The presence of a red contour within the ribose portion indicates the importance of the 2′-deoxy position for the activity. A blue contour surrounding the red contour at the 2′-position covers both the α (ribose) and β (arabinose) faces of the carbohydrate portion, which indicates that 2′-hydroxy substitution is detrimental for any biological activity. This finding is substantiated by low PRs of **11**, **14**, **16**, **20**, and **23** in which the 2′-position is substituted with electronegative groups (fluorine or hydroxyl). In the case of compound **29**, the electron cloud of its olefinic carbons occupies the space covered by the blue contour hence explaining the inactivity of this compound. For the 3′-azido compounds **7** and **8** it was observed that the quaternary nitrogen of the azido group was perfectly aligned with the small blue contour present in this area. However, when there was an additional methylene group between the 3′-carbon and the azido group (compound **9**), the tertiary electronegative nitrogen coincided with the blue contour explaining its relatively low activity.

CoMSIA also revealed that the contributions from H-bond donor fields (34.9%) and electrostatic fields (32.2%) are equally important. CoMSIA contour maps for the H-bond donor field are shown in Figure 2C. The purple contours near the C-5 position of the pyrimidine ring indicate that hydrogen-bond donor groups should not be present near C-5, which explains the very low activity of compound **19** with amino group at C-5. Hydrogen-bond donor groups are favored near the 3′-position emphasizing the importance of an unsubstituted hydroxyl function at this position. A large favored cyan isopleth was observed near the hydrogen bound to N-3, which explains the high activity of Thd (**13**) compared with its N-3-substituted analogues. Compound **6**, having two OH groups in the close proximity to the large purple region, has lower activity compared with other non-boronated N-3-substituted Thds.

The contribution of the acceptor field (17.4%) is slightly higher than that of the steric field (15.5%). However, the combined contribution from both fields is significantly lower than the combined contribution from electrostatic and hydrogen-bond donor fields. The CoMSIA contour map of the H-bond acceptor field Figure 2D revealed the importance of two carbonyl oxygens at the pyrimidine rings for all molecules, as documented by two large favored magenta isopleths in their vicinity. Interestingly, the 3′-OH group is flanked by two red contours. Since the OH group can act both as a H-bond donor as well as an acceptor, one of the red contours may indicate increased activity, whereas the second may predict the opposite. Nevertheless, the activity of the compounds bearing a hydroxyl group at 3′-position seems to be mainly governed by electrostatic and H-bond donor fields. Another favorable magenta region may have been generated around N-1 indicating the capacity of this atom to function as a H-bond acceptor.

3. Conclusions

To the best of our knowledge, we have described in this paper for the first time the use of standard computational 3D-QSAR techniques for boron-containing compounds. The derived CoMSIA model showed predictive capabilities and a high level of internal consistency. Contour maps obtained from CoMSIA model correlated with the experimentally developed SAR. With some additional refinement, the obtained model may be utilized for the development of improved 3CTAs for BNCT. X-ray crystal structures of hTK1 have become available in recent years.[11,12] However, routine structure-based computational design techniques that allow calculations with molecules containing hexavalent boron atoms and that account for conformational changes of nucleoside kinases are not yet available. Such techniques could lead to an effective in silico structure-based design of 3CTAs with superior biological properties than those of the currently available library. Until this time, the novel boron-specific 3D-QSAR analysis described in this paper provides a useful alternative for the in silico design of carborane containing agents in general and of novel improved 3CTAs in particular.

4. Materials and methods

4.1. Biological data

The PRs of hTK1 substrates **1**–**56** (Tables 1 and 2) were reported previously[3,24,39,40] and are expressed relative to the PR value of the endogenous hTK1 substrate Thd. All compounds are derivatives of Thd or dUrd in β-D-configuration with the base in *anti* conformation.

4.2. Molecular modeling

Three-dimensional QSAR studies were carried out on a SGI O_2 workstation using Sybyl 7.1 molecular modeling software (Tripos Inc., St. Louis, MO). Although the 'B' atom type is recognized by the 7.1 release of Sybyl, energy minimization of carborane-containing compounds, containing hexavalent boron atoms, cannot be performed using the Tripos Force Field implemented in Sybyl 7.1 with the hardware system indicated above.[15,16] Therefore, the energy minimizations of compounds **1**–**56** were initially performed with a Dell Inspiron 4100 using Hyperchem release 7.5 (Hypercube, Inc., Gainesville, FL) according to the following protocol: (1) The structures of *o*-, *m*-, and *p*-carborane were pre-minimized using the semi-empirical AM1 method to an energy gradient of $0.005\,\mathrm{kcal\,mol^{-1}\,\AA^{-1}}$ as described previously.[15,16] (2) A structure of Thd was pre-minimized using the MMplus module to an energy gradient of $0.005\,\mathrm{kcal\,mol^{-1}\,\AA^{-1}}$. (3) Non-boronated structures (**1**–**29**) were constructed from the pre-minimized Thd. (4) Carboranyl nucleosides (**30**–**56**) were constructed by adding appropriate spacer units between pre-minimized Thd and the pre-minimized carboranes. (5) Molecules **1**–**56** were minimized by the MMplus method using Polak-Ribiere algorithm until an energy gradient of $0.03\,\mathrm{kcal\,mol^{-1}\,\AA^{-1}}$ was reached. (6) Atomic point charges were calculated using the semi-empirical AM1 method.

Subsequent to this protocol, the energy-minimized structures with AM1 charges were saved as Sybyl-readable mol2 files and imported into Sybyl. The atom type assignments were then changed to the Sybyl compatible format by using the Built/Edit option. In addition to problems with the energy minimization of carborane-containing compounds, the generation of hydrophobic descriptors for COMSIA studies (Section 4.5) with compounds containing a 'B' atom type was also not possible with Sybyl 7.1. Therefore, all compounds were specifically formatted for the CoMFA and CoMSIA studies by changing the 'B' atom type to 'C.3' as described previously by us for the preparation of carboranyl estradiol derivatives in structure-based computational studies.[15,16]

4.3. Alignment rules and 3D molecular database

A database was created with all energy-minimized molecules, which were aligned using the 'align database' option of Sybyl using Thd (**13**, Table 1) as the template. The atoms used for the alignments are indicated by asterisks in the drawing of the generalized molecule in Table 1. The molecules were aligned based on the assumption that they all interact with the substrate-binding site of hTK1 in the same manner.

4.4. CoMFA

The aligned molecules were placed in a three-dimensional grid (2 Å spacing) automatically generated by Sybyl extending 4 Å beyond the volumes of all investigated molecules on all axes. For CoMFA, Lennard–Jones potentials and Coulomb potentials were employed to calculate the CoMFA steric and electrostatic interaction fields, respectively. At each grid point the steric- and electrostatic energies were measured for each molecule using the default probe atom, sp^3 carbon, with a charge of +1. The default cutoff values for steric and electro-

static energies were both set to 30 kcal mol$_{-1}$ and the dielectric field was distance dependent. CoMFA uses partial least squares (PLS) to predict activity of the molecules from the energy values at the grid points.

4.5. CoMSIA

The molecular alignment was placed in a three-dimensional grid (2 Å spacing) as described for the CoMFA. CoMSIA differs from CoMFA in the implementation of the descriptor fields. It calculates steric and electrostatic fields in addition to hydrophobic, hydrogen-bond donor and hydrogen-bond acceptor fields, and uses Gaussian equations for the field calculations. In the CoMSIA, the following standard parameters for the probe atom were used: a radius of 1 Å, a charge of +1, hydrophobicity of +1, and hydrogen-bond donor or acceptor properties of +1. A default value of 0.3 was used as the attenuation factor, α. Five columns were created with steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor descriptors.

4.6. Design of the training and the test sets

The prediction of biological data based on statistical methods depends on a proper design of the training set and the test set. The structure and all relevant properties of the compounds in the test set should be similar to those in the training set. Structures having similar molecular geometries and functionalities are expected to have similarities in their relevant properties. The training set of both CoMFA and CoMSIA models consisted of the same 47 molecules. Nine molecules were intuitively chosen as the test set in both models. Several molecules in Tables 1 and 2 have similar structural features. In order to achieve maximum compatibilities between the training set and the test set molecules, selection criteria for the latter were representative structural features (e.g., 5-substituted deoxyuridines, non-boronated N3-substituted Thd, 3-CTAs with single substitution at carborane cluster, and 3CTAs with substitution at both carbons of the carborane cluster). In addition test set molecules with moderately high to very low PR values were selected.

4.7. PLS analysis and validations

The CoMFA and CoMSIA descriptors were used as independent variables and log(PR)s were used as the dependent variables in the partial least square (PLS) regression analyses to derive 3D-QSAR models. The PLS analysis was carried out using the cross-validated 'Leave-One-Out' option and SAMPLS (SAMple distance PLS) to determine the optimum number of components used in the final non-cross-validated analysis. The number of components used was not larger than 1/3 of the number of rows in the training set. The optimum number of components produces the smallest root mean predictive sum of squared errors, which corresponds to the highest cross-validated coefficient (q^2). Non-cross-validated analyses were then performed using the optimum number of components. Further, the progressive scrambling technique was invoked to eliminate any redundancies in the descriptor data set and the final CoMSIA model was developed based on the optimized number of components.

Acknowledgment

This work was supported by the US Department of Energy Grant DE-FG02-90ER60972.

Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.bmc.2006.06.037.

References and notes

1. Barth, R. F.; Coderre, J. A.; Vicente, M. G. H.; Blue, T. E. *Clin. Cancer Res.* **2005**, *11*, 3987.
2. Soloway, A. H.; Tjarks, W.; Barnum, B. A.; Rong, F.-G.; Barth, R. F.; Codogni, I. M.; Wilson, J. G. *Chem. Rev.* **1998**, *98*, 1515.
3. Byun, Y.; Narayanasamy, S.; Johnsamuel, J.; Bandyopadhyaya, A. K.; Tiwari, R.; Al-Madhoun, A. S.; Barth, R. F.; Eriksson, S.; Tjarks, W. *Anti-Cancer Agents Med. Chem.* **2006**, *6*, 127.
4. Al-Madhoun, A. S.; Johnsamuel, J.; Barth, R. F.; Tjarks, W.; Eriksson, S. *Cancer Res.* **2004**, *64*, 6280.
5. Al-Madhoun, A. S.; Johnsamuel, J.; Yan, J.; Ji, W.; Wang, J.; Zhuo, J.-C.; Lunato, A. J.; Woollard, J. E.; Hawk, A. E.; Cosquer, G. Y.; Blue, T. E.; Eriksson, S.; Tjarks, W. *J. Med. Chem.* **2002**, *45*, 4018.
6. Barth, R. F.; Yang, W.; Al-Madhoun, A. S.; Johnsamuel, J.; Byun, Y.; Chandra, S.; Smith, D. R.; Tjarks, W.; Eriksson, S. *Cancer Res.* **2004**, *64*, 6287.
7. Byun, Y.; Yan, J.; Al-Madhoun, A. S.; Johnsamuel, J.; Yang, W.; Barth, R. F.; Eriksson, S.; Tjarks, W. *J. Med. Chem.* **2005**, *48*, 1188.
8. Johnsamuel, J.; Lakhi, N.; Al-Madhoun, A. S.; Byun, Y.; Yan, J.; Eriksson, S.; Tjarks, W. *Bioorg. Med. Chem.* **2004**, *12*, 4769.
9. Lunato, A. J.; Wang, J.; Woollard, J. E.; Anisuzzaman, A. K. M.; Ji, W.; Rong, F.-G.; Ikeda, S.; Soloway, A. H.; Eriksson, S.; Ives, D. H.; Blue, T. E.; Tjarks, W. *J. Med. Chem.* **1999**, *42*, 3378.
10. Al-Madhoun, A. S.; Tjarks, W.; Eriksson, S. *Mini-Rev. Med. Chem.* **2004**, *4*, 341.
11. Birringer, M. S.; Claus, M. T.; Folkers, G.; Kloer, D. P.; Schulz, G. E.; Scapozza, L. *FEBS Lett.* **2005**, *579*, 1376.
12. Welin, M.; Kosinska, U.; Mikkelsen, N.-E.; Carnrot, C.; Zhu, C.; Wang, L.; Eriksson, S.; Munch-Petersen, B.; Eklund, H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17970.
13. Kosinska, U.; Carnrot, C.; Eriksson, S.; Wang, L.; Eklund, H. *FEBS J.* **2005**, *272*, 6365.
14. Mizutani, M. Y.; Takamatsu, Y.; Ichinose, T.; Nakamura, K.; Itai, A. *Proteins: Struct. Funct. Bioinform.* **2006**, *63*, 878.
15. Johnsamuel, J.; Byun, Y.; Jones, T. P.; Endo, Y.; Tjarks, W. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3213.
16. Johnsamuel, J.; Byun, Y.; Jones, T. P.; Endo, Y.; Tjarks, W. *J. Organomet. Chem.* **2003**, *680*, 223.
17. Endo, Y.; Yaguchi, K.; Tsuji, M.; Yamaguchi, K.; Shudo, K. *Chem. Pharm. Bull.* **1999**, *47*, 699.
18. Endo, Y.; Yoshimi, T.; Kimura, K.; Itai, A. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2561.

19. Endo, Y.; Iijima, T.; Yaguchi, K.; Kawachi, E.; Inoue, N.; Kagechika, H.; Kubo, A.; Itai, A.. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1307.
20. Endo, Y.; Iijima, T.; Yamakoshi, Y.; Fukasawa, H.; Miyaura, C.; Inada, M.; Kubo, A.; Itai, A. *Chem. Biol.* **2001**, *8*, 341.
21. Tsuji, M.; Koiso, Y.; Takahashi, H.; Hashimoto, Y.; Endo, Y. *Biol. Pharm. Bull.* **2000**, *23*, 513.
22. Mizutani, M. Y.; Tomioka, N.; Itai, A. *J. Mol. Biol.* **1994**, *243*, 310.
23. Hawthorne, M. F.; Lee Mark, W. *J. Neurooncol.* **2003**, *62*, 33.
24. Bandyopadhyaya, A. K.; Johnsamuel, J.; Al-Madhoun, A. S.; Eriksson, S.; Tjarks, W. *Bioorg. Med. Chem.* **2005**, *13*, 1681.
25. Byun, Y.; Yan, J.; Al-Madhoun, A. S.; Johnsamuel, J.; Yang, W.; Barth, R. F.; Eriksson, S.; Tjarks, W. *Appl. Radiat. Isot.* **2004**, *61*, 1125.
26. Bhongade, B. A.; Gadad, A. K. *Bioorg. Med. Chem.* **2004**, *12*, 2797.
27. Datar, P.; Desai, P.; Coutinho, E.; Iyer, K. *J. Mol. Model.* **2002**, *8*, 290.
28. Doytchinova, I.; Valkova, I.; Natcheva, R. *Quant. Struct.-Act. Relat.* **2001**, *20*, 124.
29. Lepper, E. R.; Ng, S. S. W.; Guetschow, M.; Weiss, M.; Hauschildt, S.; Hecker, T. K.; Luzzio, F. A.; Eger, K.; Figg, W. D. *J. Med. Chem.* **2004**, *47*, 2219.
30. Spadola, L.; Novellino, E.; Folkers, G.; Scapozza, L. *Eur. J. Med. Chem.* **2003**, *38*, 413.
31. Sperandio da Silva, G. M.; Sant'Anna, C. M. R.; Barreiro, E. J. *Bioorg. Med. Chem.* **2004**, *12*, 3159.
32. Ungwitayatorn, J.; Samee, W.; Pimthon, J. *J. Mol. Struc.* **2004**, *689*, 99.
33. Zhu, L.; Hou, T.; Xu, X. *J. Mol. Model.* **2001**, *7*, 223.
34. Cramer, R. D. *J. Med. Chem.* **2003**, *46*, 374.
35. Böhm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458.
36. Sutherland, J. J.; O'Brien, L. A.; Weaver, D. F. *J. Med. Chem.* **2004**, *47*, 5541.
37. Clark, R. D.; Fox, P. C. *J. Comput. Aided Mol. Des.* **2004**, *18*, 563.
38. Ashek, A.; Cho, S. J. *Bioorg. Med. Chem.* **2006**, *14*, 1474.
39. Yan, J.; Naeslund, C.; Al-Madhoun Ashraf, S.; Wang, J.; Ji, W.; Cosquer Guirec, Y.; Johnsamuel, J.; Sjoberg, S.; Eriksson, S.; Tjarks, W. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2209.
40. Wang, J.; Lunato, A. J.; Anisuzzaman, A. K. M.; Ikeda, S.; Ji, W.; Rong, F.-G.; Eriksson, S.; Ives, D. H.; Soloway, A. H.; Tjarks, W. Evaluation of carboranyl 2′-deoxyuridine derivatives as substrates for human thymidine kinases 1 and 2. In *Frontiers in Neutron Capture Therapy*; Hawthorne, M. F., Ed.; Plenum Publishers: New York, 2001; Vol. 2, pp 1009–1013.