2D Quantitative structure–activity relationship studies on a series of cholesteryl ester transfer protein inhibitors

Marcelo S. Castilho,[a,*] Rafael V. C. Guido[b] and Adriano D. Andricopulo[b]

[a]*Laboratório de Bioinformática e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, 40170-290 Salvador, BA, Brazil*
[b]*Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-Carlense 400, 13560-970 São Carlos, SP, Brazil*

Abstract—Coronary heart disease (CHD) is one of the major causes of human death. The most successful therapeutic approach available is based on the reduction of low density-lipoprotein cholesterol (LDL-C). However, it is believed that the next paradigm in CHD treatment will rely on increased HDL-C levels. One of the most promising strategies for this goal is the inhibition of cholesteryl ester transfer protein (CETP). In the present work, robust classical 2D QSAR ($r^2 = 0.76$, $q^2 = 0.72$) and hologram QSAR ($r^2 = 0.88$, $q^2 = 0.70$) models were developed for a series of 85 CETP inhibitors (*N-N*-disubstituted trifluoro-3-amino-2-propanol derivatives). These models are complementary in nature and highlight important structural features for the design of novel CETP inhibitors with improved potency.
© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Although coronary heart disease (CHD) mortality has been diminishing in Western Europe and North America for the past decades, it remains one of the major causes of human death.[1,2] In the past two decades, the most successful therapeutic approach for the treatment of this disease is based on the reduction of low density-lipoprotein cholesterol (LDL-C) levels.[3] Accordingly, statins have become the gold standard treatment for patients with, or at risk for, CHD.[4] However, despite the great benefits of statins, the outcome of many treated patients is still unsatisfactory, requiring immediate attention.[5] Furthermore, epidemiologic studies have identified that low levels of high density-lipoprotein cholesterol (HDL-C) are a higher risk for CHD than LDL-C, total cholesterol or plasma triglycerides (TG).[6] Therefore, it is believed that the next frontier in CHD treatment should rely on increased HDL-C levels.[7]

Nevertheless, the available drugs for CHD management have poor effects on HDL-C levels (Fig. 1): niacin has by far the best profile, raising the level of HDL-C by as much as 30%, however, its cutaneous flushing side-effect limits the patient compliance[8]; Fibrates (e.g., gemfibrozil, fenofibrate) increase HDL-C by 15–25% in patients with hypertriglyceridemia, but have little effect (<10%) on other patients; statins (e.g., atorvastatin, simvastatin) produce only modest effect on HDL-C levels (approximately 5%).[9–11] This fact highlights the importance of probing more effective strategies to increase HCL-C levels.[5,6,8] This goal can be achieved through the inhibition of cholesteryl ester transfer protein (CETP). However, recent clinical trials moderated the perspectives on CETP inhibition as a good target for CHD patients. Although torcetrapib, an irreversible inhibitor of CETP, in association with atorvastatin substantially increased HDL-C and decreased LDL-C when compared with atorvastatin monotherapy, no significant reduction in the progression of coronary atherosclerosis was observed.[12] Although the use of torcetrapib was associated with increased blood pressure, there are a number of CETP inhibitors that do not show similar effects. Therefore, there is an increasing interest in the study of new classes of CETP inhibitors aimed at developing therapies to raise HDL cholesterol levels.[12]
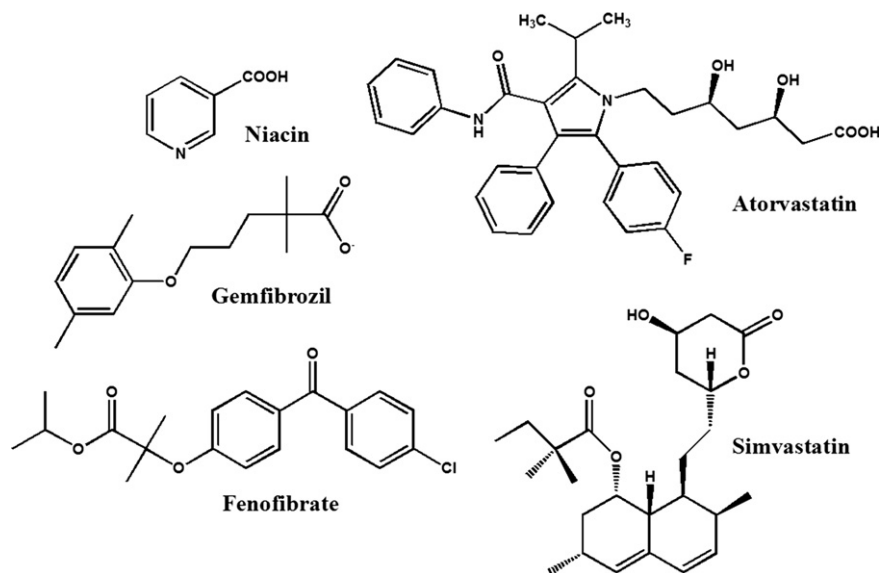
Figure 1. Drugs currently used for CHD management.

CETP is a glycoprotein that binds to HDL and is involved in the transfer of lipoprotein particles and neutral lipids, including cholesteryl ester, phospholipids, and triglyceride. Accordingly, CETP crystallographic structure revealed a noteworthy size and highly hydrophobic binding site capable of simultaneously binding four lipid molecules.[13] Among the several inhibitors described in the literature,[8] *N*-*N*-disubstituted trifluoro-3-amino-2-propanol derivatives pose themselves as interesting small-molecule targets for drug design due to their synthetic accessibility and high potency. The first crystallographic structure of CETP was made available very recently,[13] thus allowing new possibilities for QSAR studies associated with structure-based design strategies. This was not possible previously, leading to the appearance of QSAR studies employing 3D alignments based on spatial conformations of minimum energy.[14] In the present study, a data set of *N*-*N*-disubstituted trifluoro-3-amino-2-propanol derivatives, considered as CETP inhibitors, was used to develop 2D QSAR models employing easily derived topological descriptors and hologram QSAR (HQSAR), which afforded reliable and predictive QSAR models.

2. Results and discussion

The data set of 85 *N*-*N*-disubstituted trifluoro-3-amino-2-propanol derivatives used for the QSAR analyses was selected from the literature.[15–18] Chemical structures and biological properties for the complete set of compounds (divided in the corresponding training and test sets) are listed in Table 1. The IC$_{50}$ values employed in this work (varying from 20 to 65,000 nM), measured under the same experimental conditions, are acceptably distributed across the range of values (Fig. 2). Thus, the data set is appropriate for the purposes of QSAR model development. The IC$_{50}$ values were converted to the corresponding pIC$_{50}$ ($-\log$IC$_{50}$) and used as dependent variables in the QSAR investigations.

From the original data set of 85 CETP inhibitors, 68 compounds (compounds **1**–**68**, Table 1) were selected as members of the training set for QSAR model development, and the remaining 17 compounds (compounds **69**–**85**, Table 1) were considered as members of the test set for external validation. A statistical cluster analysis carried out with PIROUETTE confirmed that structurally diverse molecules possessing activities of wide range were included in both training and test sets (data not shown).

Classical 2D QSAR studies require the calculation of molecular descriptors, such as connectivity indices, 2D autocorrelation descriptors, and Burden eigenvalues, which are used as independent variables in QSAR modeling. The DRAGON 5.4 software was used to generate the descriptors for the QSAR studies. This procedure afforded 929 descriptors which were subjected to the following selection strategy. Descriptors possessing constant values as well as those with poor correlation to biological property ($r^2 < 0.10$) or that are more than 0.99 correlated were discarded. To further reduce the number of descriptors, BuildQSAR software was employed to systematically search for multiple linear regression (MLR) models of up to four variables with correlation coefficients $r^2 > 0.70$.

The best model shows moderate statistical values ($r^2 = 0.73$, $q^2 = 0.69$, $s = 0.50$, $F = 43.60$, SPRESS = 0.54) but lacked predictive ability when challenged against test set compounds (data not shown). This scenario could be the result of an incomplete description of the binding event by the selected descriptors or an optimistic evaluation of leave-one-out cross-validation procedure. At this point, none of the obtained MLR models succeed to correlate the biological property for the training set compounds within 1.0 log unit errors. These results lead us to consider exploring the statistical principal component analysis (PCA) and partial least squares (PLS) methods. For that purpose, all

6244

M. S. Castillo et al. / Bioorg. Med. Chem. 15 (2007) 6242–6252

Table 1. Chemical structures and corresponding IC$_{50}$ values for a series of CETP inhibitors

Training set
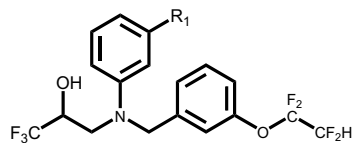
Compound	R$_1$	IC$_{50}$ (µM)	Compound	R$_1$	IC$_{50}$ (µM)	Compound	R$_1$	IC$_{50}$ (µM)	Compound	R$_1$	IC$_{50}$ (µM)
1		0.14	10		0.20	19		0.10	28		0.40
2		0.02	11		0.20	20		0.50	29		2.60
3		0.03	12		0.30	21		1.30	30		0.40
4		0.03	13		0.08	22		0.06	31		0.09
5		0.40	14		0.10	23		0.10	32		0.09
6		1.00	15		0.08	24		0.20	33		0.20
7		2.40	16		0.08	25		0.30	34		0.20
8		3.50	17		0.10	26		0.30	35		0.20
9		0.20	18		0.08	27		0.30			
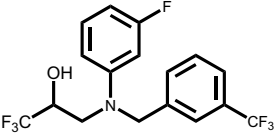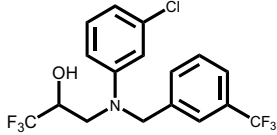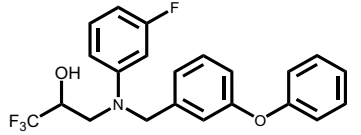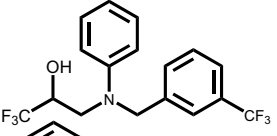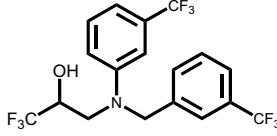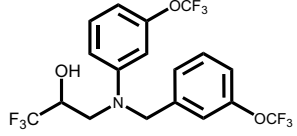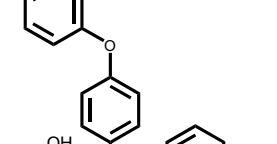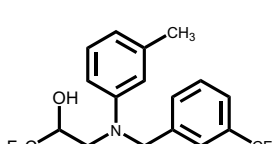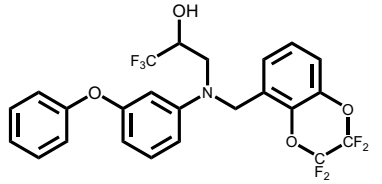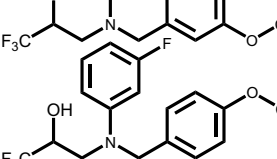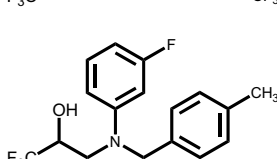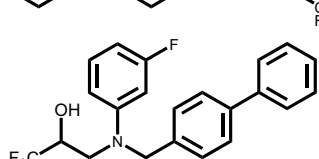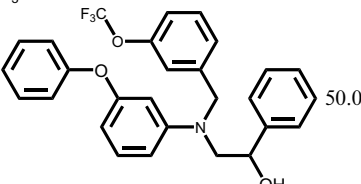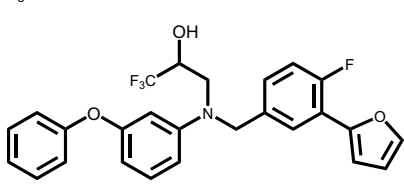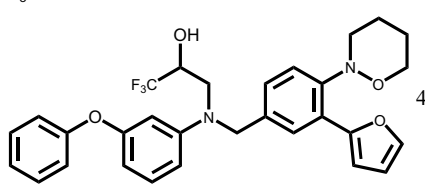
Training set
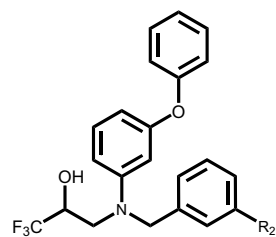
Compound	R$_2$	IC$_{50}$ (μM)	Compound	R$_2$	IC$_{50}$ (μM)	Compound	R$_2$	IC$_{50}$ (μM)	Compound	R$_2$	IC$_{50}$ (μM)
36		0.48	**41**		0.30	**46**		1.00	**50**		5.20
37		0.66	**42**		22.90	**47**		15.00	**51**		0.42
38		4.08	**43**		6.72	**48**		6.79	**52**		16.00
39		0.91	**44**		1.12	**49**		4.42	**53**		14.30
40	OCH$_2$CH$_3$	1.60	**45**	OCF$_3$	1.00						

Table 1 (*continued*)

	Training set				
Compound	Structure	IC$_{50}$ (µM)	Compound Structure	IC$_{50}$ (µM)	Compound Structure	IC$_{50}$ (µM)

Compound	Structure	IC$_{50}$ (µM)	Compound	Structure	IC$_{50}$ (µM)	Compound	Structure	IC$_{50}$ (µM)
54		40.0	**59**		12.00	**64**		15.0
55		15.0	**60**		10.00	**65**		1.50
56		25.0	**61**		10.00	**66**		12.0
57		15.0	**62**		65.00	**67**		20.0
58		50.0	**63**		0.72	**68**		45.8

Compound	R₁	IC$_{50}$ (μM)	Compound	R₁	IC$_{50}$ (μM)
69		0.70	75		0.30
70		2.50	76		0.10
71		0.40	77		0.08
72		0.09	78		7.80
73		0.20	79		6.24
74		0.06			

Compound	Structure	IC$_{50}$ (μM)
80		1.30
81		1.00
82		0.07
83		12
84		3.60
85		3.96

descriptors from the MLR models with $r^2 > 0.70$ (Table 2) were pooled together, autoscaled, and employed for PCA and PLS analysis within the PIROUETTE software.

According to the PCA results, four principal components accounted for 85% of total variance, whereas further components contribute with less than 5%. The first PC explains molecular shape features, whereas the relative potency varies according to both first and second PC. Less bulk compounds have negative values in PC1, while bulkier compounds have positive values. The most potent compounds are grouped in the far right-hand side of PC1 and have negative values in PC2. Thus, training set compounds can be broadly separated into three groups. These preliminary results prove that selected descriptors have good discriminating power and can be used for QSAR modeling studies.

In addition, PLS QSAR models were created using the leave-many-out (LMO) cross-validation procedure. As seen in Figure 3, the best statistical model shows moderate correlation within the training set ($r^2 = 0.76$, $q^2 = 0.72$, with two components).

The predictive power of the best QSAR model derived using the 68 training set molecules was assessed by predicting pIC_{50} values for 17 test set compounds (**69–85**, Table 1), not used for QSAR model development. The external validation process can be considered the most reliable validation method, as cross-validation procedures may lead to very optimistic statistics.[19–21] The results of the external validation are listed in Table 3, and the graphic results for the experimental versus predicted activities of both training set and test set are displayed in Figure 3.

Besides demonstrating statistical significance, QSAR models should also provide useful chemical insights for drug design. For this reason, an acceptable interpretation of the QSAR results is provided below. There is some criticism about using topological descriptors in QSAR modeling, as some of them lack adequate physi-
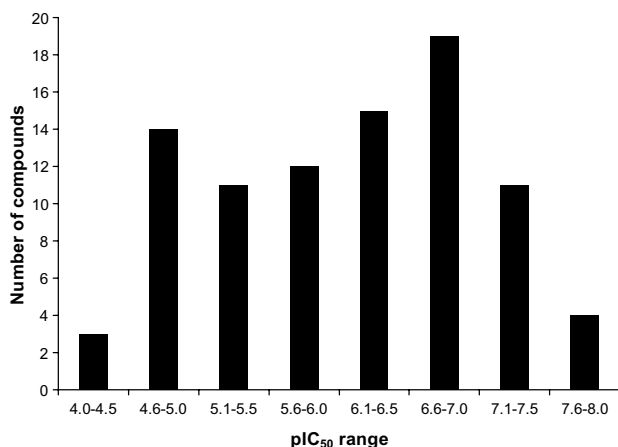
Table 2. Descriptors selected for QSAR model development

Symbol	Definition and description
SPI	Superpendentic index
S1K	1-Path Kier alpha-modified shape index
D/Dr06	Distance/detour ring index of order 6
T(N...O)	Sum of topological distance between N...O
MATS3p	Moran autocorrelation—lag3/weighted by atomic polarizabilities
EEig04r	Eigenvalue 04 from edge adjacency matrix weighted by resonance integrals
MATS3v	Moran autocorrelation—lag3/weighted by atomic van der Waals volumes
BELe4	Lowest eigenvalue N.4 of Burden matrix/weighted by atomic Sanderson electronegativities
GGI10	Topological charge index of order 10
VEA1	Eigenvector coefficient sum from adjacency matrix
BEHm5	Highest eigenvalue N.5 of Burden matrix/weighted by atomic masses
QZZe	Qzz COMMA2 value/weighted by Sanderson electronegativities
RDF060v	Radial distribution function −6.0/weighted by atomic van der Walls volumes
RDF065p	Radial distribution function −6.5/weighted by atomic polarizabilities
SPAN	Span R
RDF065v	Radial distribution function −6.5/weighted by atomic van der Walls volumes
RDF060p	Radial distribution function −6.0/weighted by atomic polarizabilities
RDF030m	Radial distribution function −3.0/weighted by atomic masses
RDF060e	Radial distribution function −6.0/weighted by atomic Sanderson electronegativities
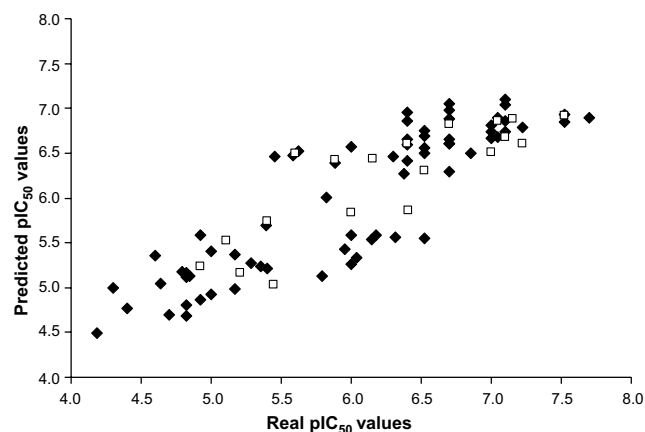
Figure 3. Predicted versus observed values of pIC_{50} values (filled diamonds, training set compounds; open squares, test set compounds).

cal interpretation. However, traditional descriptors (e.g., $\log P$ and pK_a) have also their own limitations, because some may be more correlated with pharmacokinetic properties, whereas others may be more related to affinity. Besides, traditional descriptors often are whole-molecule descriptors that convey little information on particular structural differences relevant for enzyme affinity and inhibitory potency. Nevertheless, it would be cumbersome to explain the influence of all topological descriptors over CETP inhibitors, potency. Instead,

Figure 2. Distribution of the pIC_{50} values for the data set compounds.

Table 3. Classical 2D QSAR experimental and predicted activities (pIC_{50}) with residual values for the 17 test set compounds

Compound	pIC$_{50}$		
	Experimental	Predicted	Residual[a]
69	6.15	6.44	−0.28
70	5.60	6.54	−0.94
71	6.40	6.61	−0.21
72	7.05	6.86	0.18
73	6.70	6.83	−0.13
74	7.22	6.64	0.58
75	6.52	6.32	0.21
76	7.00	6.53	0.47
77	7.10	6.71	0.39
78	5.11	5.53	−0.42
79	5.20	5.16	0.05
80	5.89	6.45	−0.56
81	6.00	5.84	0.16
82	7.15	6.91	0.24
83	4.92	5.24	−0.32
84	5.44	5.02	0.43
85	5.40	5.75	−0.35

[a] The difference between experimental and predicted values.
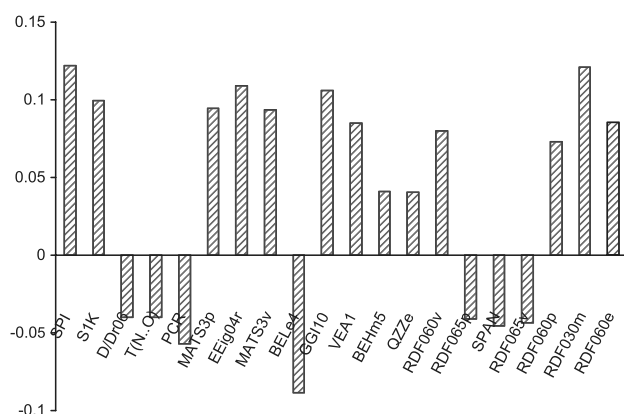
Figure 4. Regression vectors to PC1.

we decided to focus on the three most significant ones to PC1 (50.38% of variance), which are SPI, BELe4 and RDF030m (Fig. 4).

- BELe4[22,23] is a low-dimensional metric, derived from the Burden matrix, which accounts for electrostatic interactions. Analysis of BELe4 values reveals a bell-shaped correlation to potency, weak inhibitors lie in the left (BELe4 value < 1.20) and right (BELe4 value > 1.50) tails and strong potent inhibitors are placed in the center (BELe4 value = 1.46). This result suggests that CETP inhibition depends on a delicate charge balance for interaction with key residues in the CETP active site.
- SPI[24] is a topological descriptor derived from a submatrix of distance matrix that accounts for shape features in CETP inhibitors. This descriptor is linearly correlated to potency and indicates that inhibitors that fit in the active site have a better inhibitory profile. However, some compounds have already exceeded the optimal volume (compounds **64**, **67**, **104** (SPI value > 30) vs compound **2** (SPI value = 27.10)).

- RDF030m[25] is a 2D descriptor calculated from radial distribution function code and represents the probability distribution to find an atom in a spherical volume of radius *r*. Thus, this descriptor also accounts for steric and shape aspects of CETP inhibitors. As a consequence, it is correlated to SPI ($r^2 = 0.59$) and behaves accordingly.

Although the classical 2D QSAR model provided some useful information and showed a good predictive ability, topological descriptors used for model development convey little information on which moieties are particularly important to CETP inhibition. In order to shed some light on this subject, we resorted to a modern approach named hologram QSAR (HQSAR), a 2D QSAR method that has shown predictive ability comparable to those of more sophisticated 3D QSAR techniques.[26–29] HQSAR generates specialized molecular holograms that incorporate information about each 2D fragment (i.e., linear, branched, and overlapping), and each of its constituent subfragments, implicitly encoding 3D structural information that is important for binding affinity.[26–29]

HQSAR results are strongly linked to parameters concerning hologram generation, including hologram length, fragment size, and fragment distinction.[26] The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). Several combinations of these parameters were considered using the fragment size default (4–7), as follows: A/B, A/B/C, A/B/C/ H, A/B/C/H/Ch, A/B/C/H/Ch/DA, A/B/H, A/B/C/Ch, A/B/DA, A/B/C/DA, A/B/H/DA, A/B/C/H/DA, and A/B/H/Ch/DA. HQSAR analysis was performed over the 12 default series of hologram length values ranging from 53 to 401 bins. The patterns of fragment counts from the 68 training set inhibitors were then related to the experimental biological data using PLS analysis. The statistical results are summarized in Table 4.

As shown in Table 4, fragment distinction parameters have considerable effects on the quality of the models as expected. For instance, models 8 and 10 were derived using A/B/DA and A/B/H/DA, respectively. Accordingly, the inclusion of hydrogen as distinction parameters was detrimental for biological activity prediction of excluded molecules during the internal validation procedure (LOO) as indicated by the decrease of ~5% in q^2 value. This result is in good agreement with previous results[26,28,30] which suggest that both options should not be used together due to the substantial increase in the number of fragments generated when both fragment distinctions are considered in the model construction.

The best statistical result among all models (Table 4) was obtained for model 8 ($r^2 = 0.87$, $q^2 = 0.70$, with five components), derived using A/B/DA as fragment distinction. The influence of different fragment sizes, which control the minimum and maximum length of fragments to be included in the hologram, was further investigated for the best HQSAR model as depicted in Table 5.

Table 4. Results of HQSAR analyses for various fragment distinctions on the statistical parameters using fragment size default (4–7)

Model	Fragment distinction	q^2	r^2	SEE	HL	N
1	A/B	0.64	0.87	0.35	257	5
2	A/B/C	0.57	0.87	0.33	401	5
3	A/B/C/H	0.59	0.83	0.38	401	5
4	A/B/C/H/Ch	0.62	0.83	0.39	151	5
5	A/B/C/H/Ch/DA	0.67	0.88	0.33	307	5
6	A/B/H/	0.59	0.85	0.37	353	5
7	A/B/C/Ch	0.58	0.86	0.35	151	5
8	A/B/DA	0.70	0.87	0.33	257	5
9	A/B/C/DA	0.67	0.90	0.30	401	5
10	A/B/H/DA	0.65	0.89	0.32	257	5
11	A/B/C/H/DA	0.63	0.81	0.41	61	5
12	A/B/H/Ch/DA	0.67	0.89	0.32	307	5

Table 5. HQSAR analyses for the influence of various fragment sizes on the key statistical parameters

Model	Fragment size	q^2	r^2	SEE	HL	N
13	2–5	0.62	0.83	0.39	199	5
14	3–6	0.66	0.86	0.35	257	5
15	4–7	0.70	0.87	0.33	257	5
16	5–8	0.69	0.80	0.42	59	4
17	6–9	0.70	0.88	0.32	353	4
18	7–10	0.70	0.87	0.34	353	4

This analysis revealed that, fragment size variation did not improve the statistical parameters. However, slightly simpler models were generated when fragment sizes of 6–9 ($r^2 = 0.88$, $q^2 = 0.70$ and SEE = 0.32) and 7–10 ($r^2 = 0.87$, $q^2 = 0.70$, and SEE = 0.34) were investigated. In order to select the best and more robust model, an external validation procedure was carried out. Accordingly, the predictive ability of models 17 and 18 was assessed by predicting pIC$_{50}$ values for the same test set molecules previously used in classical 2D QSAR studies (compounds **69**–**85**, Table 1). Model 17 presented predictive-r^2 value of 0.71, whilst model 18 presented predictive-r^2 value of 0.61 (data not shown). The results of both training and test set for the best generated model (model 17) are displayed in Figure 5 and Table 6. The good agreement between experimental and predicted values for the test set compounds indicates the reliability of the HQSAR model ($r^2 = 0.88$
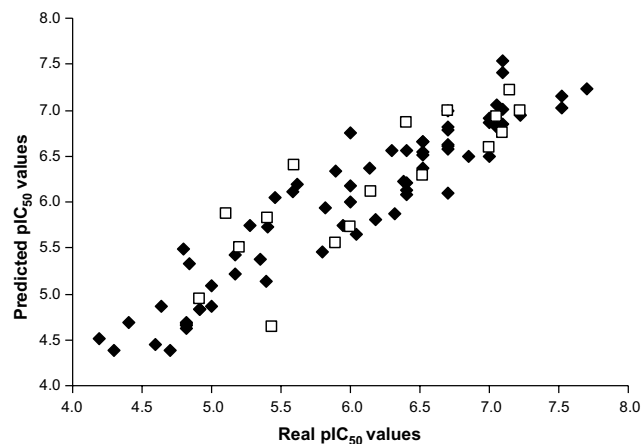
Figure 5. Predicted versus experimental values of pIC$_{50}$ for the training and test of CETP inhibitors obtained by the best HQSAR model. Filled diamonds, training set compounds; open squares, test set compounds.

Table 6. HQSAR experimental and predicted activities (pIC$_{50}$) with residual values for the 17 test set compounds

Compound	pIC$_{50}$		
	Experimental	Predicted	Residual[a]
69	6.15	6.11	0.04
70	5.60	6.40	−0.80
71	6.40	6.87	−0.47
72	7.05	6.93	0.12
73	6.70	7.00	−0.30
74	7.22	6.99	0.23
75	6.52	6.28	0.24
76	7.00	6.59	0.41
77	7.10	6.76	0.34
78	5.11	5.88	−0.77
79	5.20	5.50	−0.30
80	5.89	5.56	0.33
81	6.00	5.73	0.27
82	7.15	7.21	−0.06
83	4.92	4.94	−0.02
84	5.44	4.64	0.80
85	5.40	5.82	−0.42

[a] The difference between experimental and predicted values.

and SEE = 0.32). The predicted values fall close to the experimental pIC50 values, within 0.80 log units. This result is slightly better than those obtained for the classical 2D QSAR model, which fails to predict compound **70** by almost a log unit.

Besides improved statistical robustness, the HQSAR model is able to provide useful insights into the relationship between structural fragments and biological activity, which can be visualized through contribution maps. For example, analysis of the HQSAR contribution maps (Fig. 6) for compounds **56** (IC$_{50}$ = 25.0 μM) and **45** (IC$_{50}$ = 1.0 μM) clearly revealed different fragment contributions to biological activity.
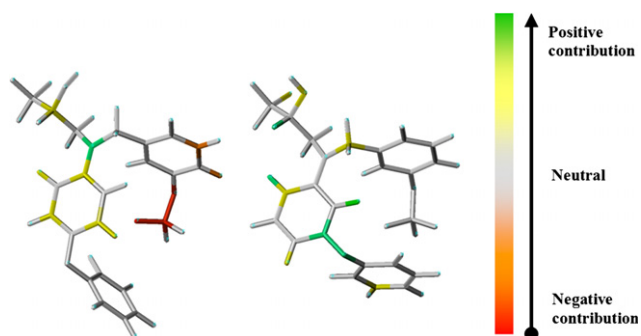
Figure 6. Contribution maps of compounds **56** (left panel) and **45** (right panel). The color scheme represents fragment contribution to CETP inhibitors, potency.

This result shows structural features that are related to activity, but does not provide a straightforward physico-chemical explanation for the differences in potency of these compounds, for instance, the different color scheme might be a consequence of steric or electrostatic contribution or even both. However, when classical QSAR descriptors are considered as well, it can be seen that both molecules have optimal electrostatic proper-ties (BELe4 value = 1.46), but different steric properties, as observed for their RDF30m values (compound **45** = 8.23 and compound **56** = 12.1). Accordingly, the data suggest that the lower inhibitory activity of com-pound **56** is possibly associated to steric restraints within the active site. This confirms previous results from Dur-ley and colleagues[15,17] indicating that the *para* substitu-tion pattern of the benzamine moiety is detrimental to activity. In addition, our QSAR models afforded a quantitative stand point of view on this subject and underscored the facts that due to sterical restraints com-pound **56** is prevented from binding properly in CETP active site.

3. Conclusion

The use of both molecular holograms and topological descriptors, along with the standard and reproducible high quality biological data, allowed the generation of robust QSAR models for this series of CETP inhibitors. A synergic approach, using information from the HQSAR contribution maps and also from a careful interpretation of the topological descriptors from classi-cal 2D QSAR, shed some light on the effect of the sub-stitution pattern over fragments that should be sterically constrained. Therefore, the QSAR models described herein are complementary in nature and shall be useful for the design of structurally related CETP inhibitors with improved potency.

4. Materials and methods

4.1. Data set

The data set of 85 CETP inhibitors used in the QSAR analyses was collected from the literature.[15–18] A statis-tical cluster analysis was carried out with PIROUETTE software (Infometrix, Washington, USA). Accordingly, the data set was split into training (compounds **1**–**68**, Table 1) and test (compounds **69**–**85**, Table 1) sets in the ratio of 5:1 (20%). The test set of 17 compounds was selected in such a way to provide an appropriate representation of the training set in terms of structural composition and activity range. The pIC_{50} values were treated as dependent variables during the Partial Least Squares (PLS) analyses.

4.2. Classical QSAR studies

The 2D molecular descriptors, such as topological descriptors, connectivity indices, 2D autocorrelation descriptors, Burden eigenvalue indices, and so on, were computed using the software DRAGON 5.4 (Talette SRL, Milan, Italy) and used as independent variables in the classical QSAR studies. A total of 929 molecular descriptors was then subjected to the following selection criteria; descriptors with constant values or found to have poor correlation to biological property were discarded ($r^2 < 0.10$). This strategy afforded 536 descrip-tors. BuildQSAR[19] software was employed to systemat-ically search for models of up to four variables that give rise to multiple linear regression (MLR) models with $r^2 > 0.70$. All descriptors present in MLR models were pooled together, autoscaled, and used for PLS analysis in PIROUETTE software.

4.3. HQSAR modeling

The HQSAR modeling analyses, calculations, and visualizations were performed with SYBYL 7.2 pack-age (Tripos Inc., St. Louis, USA), running on Red Hat Enterprise Linux workstations. Several combina-tions of fragment distinction were considered during the QSAR modeling runs. Holograms were generated using six distinct fragment sizes over the 12 default series of hologram lengths (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins). The molecular holograms generated were used as independent vari-able during the PLS regression analyses to derive the HQSAR models.

4.4. Statistical analyses

The several statistical models generated in our studies were investigated using the PLS leave-many-out (LMO) method. The predictive ability of the models was assessed by their full cross-validated r^2 (q^2) values. Internal validation methods leave-one-out and LMO (with 10 groups of compounds) were used to determine the optimum number of PLS components and the sta-bility of the models. The number of components used in the final nonvalidated model was optimized to give the highest q^2 value and the lowest standard error of prediction. The noncross-validated models were as-sessed by the conventional correlation coefficient r^2 and F-values. External validation was performed with a test set of 17 compounds, which were not included in the training set during the process of QSAR model generation.

References and notes

1. Abildstrom, S. Z.; Rasmussen, S.; Rosén, M.; Madsen, M. *Heart* **2003**, *89*, 507.
2. Jemal, A.; Ward, E.; Hao, Y.; Thun, M. *JAMA* **2005**, *294*, 1255–1259.
3. Ong, H. T. *Q. J. Med.* **2005**, *98*, 599.
4. Shah, P. K. *Eur. Heart J.* **2007**, *28*, 5.
5. Shinkai, H. *Mini Rev. Med. Chem.* **2002**, *2*, 271.
6. Barter, P. J.; Kastelein, J. J. P. *J. Am. Coll. Cardiol.* **2006**, *47*, 492.
7. Milani, R. V.; Lavie, C. J. O. *J. Am. Coll. Cardiol.* **2006**, *48*, 1791.
8. Sikorski, J. A. *J. Med. Chem.* **2006**, *49*, 2.
9. Milani, R. V.; Lavie, C. J. In *Cardiovascular Drug Therapy*; Messerli, F. H., Ed.; Saunders: Philadelphia, PA, 1996; pp 1098–1110.
10. Jones, P. H.; Davidson, M. H.; Stein, E. A.; Bays, H. E.; McKenney, J. M.; Miller, E.; Cain, V. A.; Blasetto, J. W. *Am. J. Cardiol.* **2003**, *92*, 152.
11. Deedwania, P. C.; Hunninghale, D. B.; Bays, H. E.; Jones, P. H.; Cain, V. A.; Blassetto, J. W. *Am. J. Cardiol.* **2005**, *95*, 360.
12. Nissen, S. E.; Tardif, J. C.; Nicholls, S. J.; Revkin, J. H.; Shear, C. L.; Duggan, W. T.; Ruzyllo, W.; Bachinsky, W. B.; Lasala, G. P.; Tuzcu, E. M. ILLUSTRATE Investigators. *N. Engl. J. Med.* **2007**, *356*, 1304.
13. Qiu, X.; Mistry, A.; Ammirati, M. J.; Chrunyk, B. A.; Clark, R. W.; Cong, Y.; Culp, J. S.; Danley, D. E.; Freeman, T. B.; Geoghegan, K. F.; Griffor, M. C.; Hawrylik, S. J.; Hayward, C. M.; Hensley, P.; Hoth, L. R.; Karam, G. A.; Lira, M. E.; Lloyd, D. B.; McGrath, K. M.; Stutzman-Engwall, K. J.; Subashi, A. K.; Subashi, T. A.; Thompson, J. F.; Wang, I. K.; Zhao, H.; Seddon, A. P. *Nat. Struct. Mol. Biol.* **2007**, *14*, 106.
14. Kelkar, M. A.; Pednekar, D. V.; Pimple, S. R.; Akamanchi, K. G. *Med. Chem. Res.* **2004**, *13*, 590.
15. Durley, R. C.; Grapperhaus, M. L.; Massa, M. A.; Mischke, D. A.; Parnas, B. L.; Fobian, Y. M.; Rath, N. P.; Honda, D. D.; Zeng, M.; Connolly, D. T.; Heuvelman, D. M.; Witherbee, B. J.; Glenn, K. C.; Krul, E. S.; Smith, M. E.; Sikorski, J. A. *J. Med. Chem.* **2000**, *43*, 4576.
16. Massa, M. A.; Spangler, D. P.; Durley, R. C.; Hickory, B. S.; Connolly, D. T.; Witherbee, B. J.; Smith, M. E.; Sikorski, J. A. *Bioorganic. Med. Chem. Lett.* **2001**, *11*, 1625.
17. Durley, R. C.; Grapperhaus, M. L.; Hickory, B. S.; Massa, M. A.; Wang, J. L.; Spangler, D. P.; Mischke, D. A.; Parnas, B. L.; Fobian, Y. M.; Rath, N. P.; Honda, D. D.; Zeng, M.; Connolly, D. T.; Heuvelman, D. M.; Witherbee, B. J.; Melton, M. A.; Glenn, K. C.; Krul, E. S.; Smith, M. E.; Sikorski, J. A. *J. Med. Chem.* **2002**, *45*, 3891.
18. Reinhard, E. J.; Wang, J. L.; Durley, R. C.; Fobian, Y. M.; Grapperhaus, M. L.; Hickory, B. S.; Massa, M. A.; Norton, M. B.; Promo, M. A.; Tollefson, M. B.; Vernier, W. F.; Connolly, D. T.; Witherbee, B. J.; Melton, M. A.; Regina, K. J.; Smith, M. E.; Sikorski, J. A. *J. Med. Chem.* **2003**, *46*, 2152.
19. Oliveira, D. B.; Gaudio, A. C. *Quant. Struct.–Activ. Relat.* **2003**, *19*, 599.
20. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, *20*, 269.
21. Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O. *J. Comput. Aided Mol. Des.* **2006**, *20*, 83.
22. Pearlman, R. S.; Smith, K. M. In 3D-QSAR in drug design, Kubinyi, H., Ed.; Kluwer/Escom, 1998; Vol. 2, pp 339–353.
23. Pearlman, R. S.; Smith, K. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 28.
24. Gupta, S.; Singh, M.; Madan, A. K. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 272.
25. Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. *Vibrat. Spectrosc.* **1999**, *19*, 151.
26. HQSAR™ Manual, SYBYL 6.9.2, Tripos Inc., St. Louis, MO, 2003.
27. Tong, W.; Lowis, D. R.; Perkins, R.; Chen, Y.; Welsh, W.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669.
28. Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119.
29. Castilho, M. S.; Guido, R. V. C.; Andricopulo, A. D. *Lett. Drug. Des. Discov.* **2007**, *4*, 106.
30. Valadares, N. F.; Castilho, M. S.; Polikarpov, I.; Garratt, R. C. *Bioorg. Med. Chem.* **2007**, *15*, 4609.