

### 3-D QSAR Studies on Histone Deacetylase Inhibitors. A GOLPE/GRID Approach on Different Series of Compounds

Rino Ragno,<sup>\*,§</sup> Silvia Simeoni,<sup>§</sup> Sergio Valente,<sup>§</sup> Silvio Massa,<sup>#</sup> and Antonello Mai<sup>\*,§</sup>

Istituto Pasteur - Fondazione Cenci Bolognetti, Dipartimento di Studi Farmaceutici, Università degli Studi di Roma "La Sapienza", P.le A. Moro 5, 00185 Roma, Italy, and Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, via A. Moro, 53100 Siena, Italy

Received December 28, 2005

Docking simulation and three-dimensional quantitative structure–activity relationships (3D-QSARs) analyses were conducted on four series of HDAC inhibitors. The studies were performed using the GRID/GOLPE combination using structure-based alignment. Twelve 3-D QSAR models were derived and discussed. Compared to previous studies on similar inhibitors, the present 3-D QSAR investigation proved to be of higher statistical value, displaying for the best global model  $r^2$ ,  $q^2$ , and cross-validated SDEP values of 0.94, 0.83, and 0.41, respectively. A comparison of the 3-D QSAR maps with the structural features of the binding site showed good correlation. The results of 3D-QSAR and docking studies validated each other and provided insight into the structural requirements for anti-HDAC activity. To our knowledge this is the first 3-D QSAR application on a broad molecular diversity training set of HDACIs.

#### INTRODUCTION

Histone acetyltransferases (HATs) and histone deacetylases (HDACs) are two families of enzymes that play a crucial role in gene expression. HATs acetylate the lysine  $\epsilon$ -amino termini of histone tails allowing transcription factors to access the DNA. On the contrary, HDACs catalyze the reverse reaction deacetylating the lysine residues.<sup>1–4</sup>

Disruption of this equilibrium by aberrant HDAC activity leads to hypoacetylated histones, condensed chromatin structure, and access restriction to transcription factors with consequent transcriptional repression.<sup>5, 6</sup>

This epigenetic mechanism has been proved to cause inappropriate gene expression as the cause of many forms of cancer.<sup>7–11</sup> In fact, HDAC inhibition has been related with cell growth arrest, differentiation, and/or apoptosis in several cancer cell lines.<sup>12–16</sup>

According to molecular mass and their sequence similarity with RPD3,<sup>17</sup> HDA1,<sup>18,19</sup> and Sir2<sup>20,21</sup> enzymes, first identified in yeast, mammalian HDACs have been recently distinct in classes I, II, and III, respectively.<sup>22</sup>

Classes I and II HDACs have a  $\text{Zn}^{2+}$ -dependent deacetylation mechanism and are mainly inhibited by hydroxamic acid-containing (i.e. trichostatin A (TSA)<sup>23</sup> and suberoylanilide hydroxamic acid (SAHA)<sup>24,25</sup>) inhibitors. Instead, a different  $\text{NAD}^+$ -dependent deacetylation mechanism was found for class III HDACs, that are insensitive to class I/II HDAC inhibitors.<sup>26</sup> Another eukaryote  $\text{Zn}^{2+}$ -dependent deacetylase, maize HD2, structurally different from mammalian HDACs, has been attributed to a class of its own.<sup>27</sup> Interestingly, HD2 seems to have some biological similarity to class I/II HDACs. In fact, HD2 is inhibited by a number

of reported HDAC inhibitors (HDACIs) and because of its easily affordable isolation and purification has been widely used for the screening of potential new HDACIs.<sup>27–37</sup> Moreover, a satisfactory linear correlation between HD2 and HDAC1 inhibitory data has been recently described<sup>38</sup> and used to convert HD2 to HDAC1  $\text{pIC}_{50}$  values to be taken in a QSAR study.

Three-dimensional quantitative structure–activity relationship (3-D QSAR) methods, such as comparative molecular field analysis (CoMFA)<sup>39</sup> and the GRID/GOLPE combination,<sup>40</sup> have been successfully applied in many instances to guide the design of new bioactive molecules. The application of 3-D QSAR methodology to the design of HDACIs has received little attention to date, and to our knowledge only three examples have been reported.<sup>41–43</sup> Nevertheless, the studies were conducted on congeneric series of compounds using the CoMFA<sup>39</sup> and/or CoMSIA<sup>44</sup> procedures. The present study is the application of the GRID/GOLPE procedure to different series of HDAC inhibitors that were all tested against HD2 (Chart 1) to gain insights into how steric, electrostatic, hydrophobic, and hydrogen-bonding interactions could influence the inhibitory enzyme activity and to derive predictive 3-D QSAR models for designing and forecasting the activity of untested HDACIs.

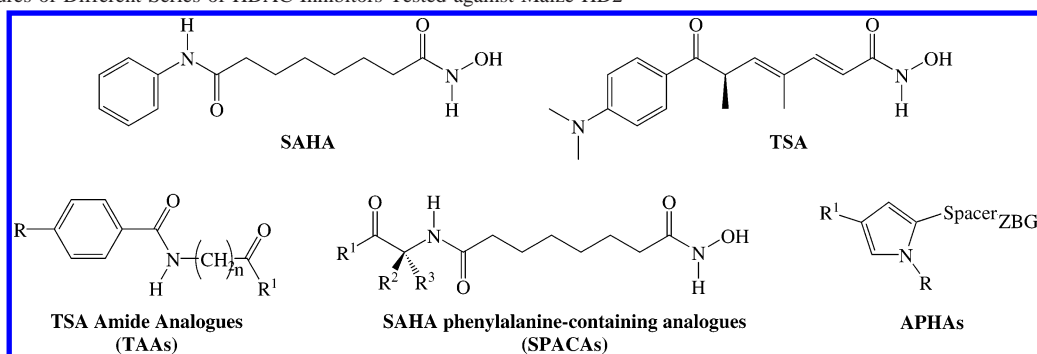
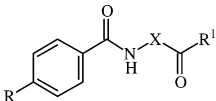
#### RESULTS AND DISCUSSION

The training set used comprises TSA, SAHA, and 101 compounds taken from four different molecular series previously tested against HD2 enzyme. The first series contains TSA amide analogues (TAAs, Table 1) reported in 1999 by Jung et al.,<sup>29</sup> the second series belongs to SAHA phenylalanine-containing analogues (SPACAs, Table 2) reported by Wittich et al. in 2002,<sup>30</sup> and in the third and fourth series some aroyl-pyrrolyl-hydroxy-amides (APHA1 and APHA2, Tables 3 and 4) recently reported by us<sup>28,33,34,36</sup> have been included.

\* Corresponding authors phone: +396-4991-3937; fax: +396-491491; e-mail: rino.ragno@uniroma1.it (R.R.); phone: +396-4991-3392; fax: +396-491491; e-mail: antonello.mai@uniroma1.it (A.M.).

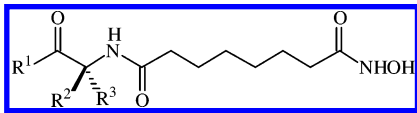
<sup>§</sup> Università degli Studi di Roma "La Sapienza".

<sup>#</sup> Università degli Studi di Siena.

**Chart 1.** Structures of Different Series of HDAC Inhibitors Tested against Maize HD2**Table 1.** TSA Amide Analogues (TAAs) Included in the Training Set<sup>29</sup>


entry	paper no.	R	R <sub>1</sub>	X	IC <sub>50</sub> (μM)
1	5a	Me <sub>2</sub> N	NHOH	(CH <sub>2</sub> ) <sub>5</sub>	0.100
2	7a	Me <sub>2</sub> N	OH	(CH <sub>2</sub> ) <sub>5</sub>	NA <sup>a</sup>
3	5b	Me <sub>2</sub> N	NHOH	(CH <sub>2</sub> ) <sub>3</sub>	NA <sup>a</sup>
4	7b	Me <sub>2</sub> N	OH	(CH <sub>2</sub> ) <sub>3</sub>	NA <sup>a</sup>
5	5c	Me <sub>2</sub> N	NHOH	(CH <sub>2</sub> ) <sub>4</sub>	2
6	7c	Me <sub>2</sub> N	OH	(CH <sub>2</sub> ) <sub>4</sub>	NA <sup>a</sup>
7	5d	Me <sub>2</sub> N	NHOH	(CH <sub>2</sub> ) <sub>6</sub>	0.100
8	7d	Me <sub>2</sub> N	OH	(CH <sub>2</sub> ) <sub>6</sub>	NA <sup>a</sup>
9	5e	Me <sub>2</sub> N	NHOH	(CH <sub>2</sub> ) <sub>7</sub>	0.3
10	7e	Me <sub>2</sub> N	OH	(CH <sub>2</sub> ) <sub>7</sub>	NA <sup>a</sup>
11	5f	Me <sub>2</sub> N	NHOH	4-Ph-CH <sub>2</sub>	NA <sup>a</sup>
12	7f	Me <sub>2</sub> N	OH	4-Ph-CH <sub>2</sub>	NA <sup>a</sup>
13	5g	Me <sub>2</sub> N	NHOH	CH <sub>2</sub> -4-Ph	0.180
14	7g	Me <sub>2</sub> N	OH	CH <sub>2</sub> -4-Ph	NA <sup>a</sup>
15	5h	MeO	NHOH	(CH <sub>2</sub> ) <sub>5</sub>	0.140
16	7h	MeO	OH	(CH <sub>2</sub> ) <sub>5</sub>	NA
17	5i	H	NHOH	(CH <sub>2</sub> ) <sub>5</sub>	0.900
18	7i	H	OH	(CH <sub>2</sub> ) <sub>5</sub>	NA <sup>a</sup>
19	5j	Cl	NHOH	(CH <sub>2</sub> ) <sub>5</sub>	0.150
20	7j	Cl	OH	(CH <sub>2</sub> ) <sub>5</sub>	NA <sup>a</sup>
21	5k	O <sub>2</sub> N	NHOH	(CH <sub>2</sub> ) <sub>5</sub>	0.180
22	7k	O <sub>2</sub> N	OH	(CH <sub>2</sub> ) <sub>5</sub>	NA <sup>a</sup>
23	5l	Ph	NHOH	(CH <sub>2</sub> ) <sub>5</sub>	0.100
24	7l	Ph	OH	(CH <sub>2</sub> ) <sub>5</sub>	NA <sup>a</sup>

An attractive feature of the compounds used in the training set is that the anti-HD2 activities were all measured in the same laboratory,<sup>28–30,33,34,36</sup> thus minimizing the error always present in comparing compounds assayed by different research groups, as outlined in a recently reported classical QSAR study on HDAC inhibitors.<sup>38</sup> Moreover, the four series span over different chemical scaffolds (Chart 1) introducing in the 3-D QSAR study the needed chemical diversity ensuring a wide range of applicability and covering a range of more than 5 log units of pIC<sub>50</sub>. The cocrystallized structure of either TSA or SAHA into the histone deacetylase-like protein (HDLP)<sup>45</sup> provided a suitable template for modeling the training set compounds. As previously described,<sup>34,36,45</sup> TSA and SAHA showed into the HDLP catalytic pocket a different orientation of their ketone/amide groups, thus using a manual docking procedure either TSA-based or a SAHA-based alignments for each compound were performed. Two complexes for each modeled inhibitor were built and minimized. During the minimization a shell of 8 Å from the cocrystallized inhibitors (SAHA and TSA) was allowed to relax; this implies that in this structure-based alignment

**Table 2.** SAHA Phenylalanine-Containing Analogues (SPACAs) Included in the Training Set<sup>30</sup>


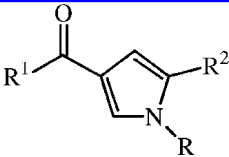
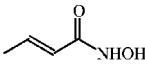
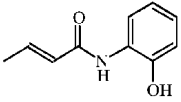
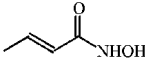
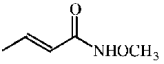
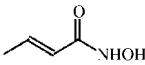
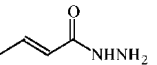
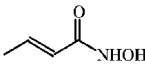
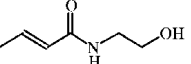
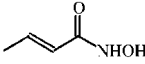
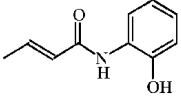
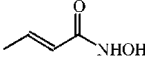
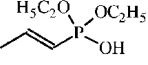
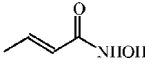
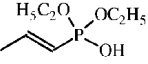
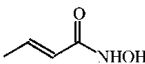
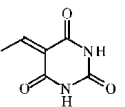
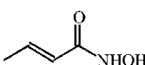
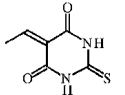
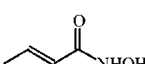
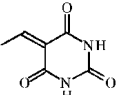
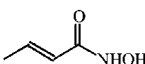
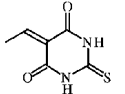
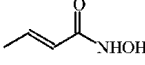
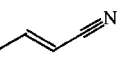
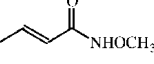
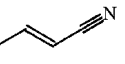
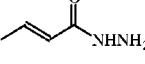
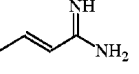
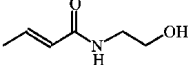
entry	paper no.	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	IC <sub>50</sub> (μM)
25	11a	NH <sub>2</sub>	PhCH <sub>2</sub>	H	1.100
26	11b	NHPh	PhCH <sub>2</sub>	H	0.08
27	11c	NHCH <sub>2</sub> Ph	PhCH <sub>2</sub>	H	0.14
28	11d	NH(CH <sub>2</sub> ) <sub>2</sub> Ph	PhCH <sub>2</sub>	H	0.33
29	11e	NH(CH <sub>2</sub> ) <sub>3</sub> Ph	H	PhCH <sub>2</sub>	0.34
30	11f	NH(CH <sub>2</sub> ) <sub>3</sub> Ph	PhCH <sub>2</sub>	H	0.38
31	11g	pyrrolyl	PhCH <sub>2</sub>	H	0.52
32	11h	OCH <sub>3</sub>	H	PhCH <sub>2</sub>	0.076
33	11i	OH	H	PhCH <sub>2</sub>	0.075
34	11j	OCH <sub>3</sub>	4-MeO-PhCH <sub>2</sub>	H	0.025
35	11k	OCH <sub>3</sub>	4-NO <sub>2</sub> -PhCH <sub>2</sub>	H	0.085
36	11l	OCH <sub>3</sub>	4-Ph-PhCH <sub>2</sub>	H	0.021
37	11m	OCH <sub>3</sub>	iPr	H	0.009
38	11n	OCH <sub>3</sub>	1-naphthylCH <sub>2</sub>	H	0.036
39	11o	OCH <sub>3</sub>	2-naphthylCH <sub>2</sub>	H	0.004
40	11p	OCH <sub>3</sub>	2-thenyl	H	0.007
41	11q	OCH <sub>3</sub>	3-indolylCH <sub>2</sub>	H	0.012
42	11r	NH(CH <sub>2</sub> ) <sub>2</sub> Ph	1-naphthylCH <sub>2</sub>	H	0.035
43	5	OCH <sub>3</sub>	PhCH <sub>2</sub>	H	0.800

procedure some level of induced-fit is also considered. The inhibitor structures were extracted from the minimized complexes while retaining their coordinates allowing to obtain the aligned training set.

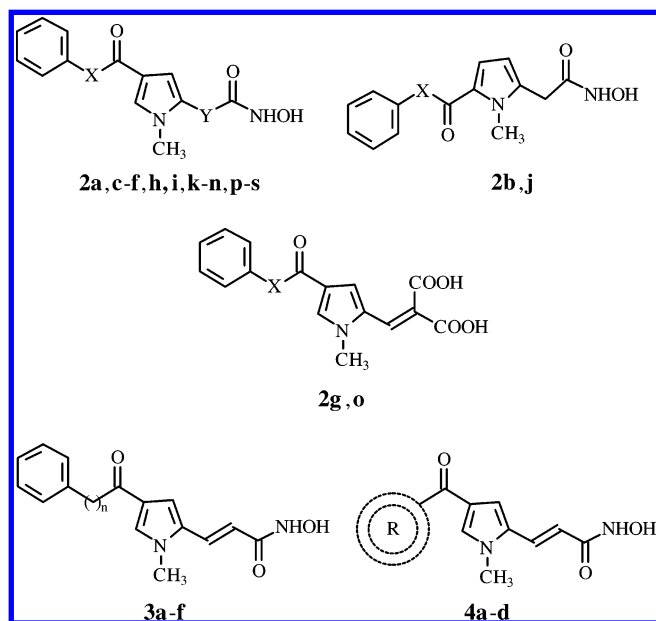
Moreover, to develop a 3-D QSAR model able to discriminate active versus inactive derivatives, the latter were included in the training set. To inactive compounds were arbitrarily assigned IC<sub>50</sub> values equal to 50% of the least active one.<sup>46–48</sup> Although the use of arbitrarily values could be argued, a 3-D QSAR model containing data regarding inactive compounds could be of valuable use in discriminating inactive from active compounds.

Finally, the predictive capability of the 3-D QSAR models was evaluated through an external test set composed of 7 newly derived APHA compounds taken from a recent report<sup>31,32</sup> (test set-1, Table 5) and against a second external test set composed of 12 new uracil-based hydroxamates (UBHAs) recently reported from our group<sup>35</sup> as novel HD2 inhibitors (test set-2, Table 6). Test set-1, being formed of derivatives that recall APHA1 and APHA2 group, would confirm the goodness of the developed models. On the other hand, the second test set-2 being formed of a different chemical scaffold is a real test in proving the models predictability toward new chemical entities.

**Table 3.** Aryl-Pyrrolyl-Hydroxy-Amide Series 1 (APHA1) Included in the Training Set<sup>33</sup>

											
entry	paper no.	R	R <sup>1</sup>	R <sup>2</sup>	IC <sub>50</sub> (μM)	entry	paper no.	R	R <sup>1</sup>	R <sup>2</sup>	IC <sub>50</sub> (μM)
44	1	CH <sub>3</sub>	Ph		3.8	59	16	CH <sub>3</sub>	Ph		NA <sup>a</sup>
45	2	CH <sub>3</sub>	4-Cl-Ph		2.4	60	17	CH <sub>3</sub>	PhCH <sub>2</sub>		NA
46	3	CH <sub>3</sub>	4-F-Ph		3.8	61	18	CH <sub>3</sub>	PhCH <sub>2</sub>		NA
47	4	CH <sub>3</sub>	4-O <sub>2</sub> N-Ph		3.9	62	19	CH <sub>3</sub>	PhCH <sub>2</sub>		NA
48	5	CH <sub>3</sub>	4-CH <sub>3</sub> -Ph		1.9	63	20	CH <sub>3</sub>	PhCH <sub>2</sub>		NA
49	6	CH <sub>3</sub>	4-CH <sub>3</sub> O-Ph		2.9	64	21	CH <sub>3</sub>	Ph		NA
50	7	CH <sub>3</sub>	4-Me <sub>2</sub> N-Ph		2.4	65	22	CH <sub>3</sub>	PhCH <sub>2</sub>		NA
51	8	CH <sub>3</sub>	PhCH <sub>2</sub>		0.1	66	23	CH <sub>3</sub>	Ph		NA
52	9	CH <sub>3</sub>	PhCH=CH		1	67	24	CH <sub>3</sub>	Ph		NA
53	10	H	Ph		5	68	25	CH <sub>3</sub>	PhCH <sub>2</sub>		104
54	11	iPr	Ph		53	69	26	CH <sub>3</sub>	PhCH <sub>2</sub>		85
55	12	Ph	Ph		110	70	27	CH <sub>3</sub>	Ph		27.4
56	13	CH <sub>3</sub>	Ph		NA <sup>b</sup>	71	28	CH <sub>3</sub>	PhCH <sub>2</sub>		NA
57	14	CH <sub>3</sub>	Ph		NA	72	29	CH <sub>3</sub>	Ph		23.3
58	15	CH <sub>3</sub>	Ph		NA						

<sup>a</sup> NA, not active at starting concentration of 30 μM.

**Table 4.** Aroyl-Pyrrolyl-Hydroxy-Alkylamide Series 2 (APHA2) Included in the Training Set<sup>28</sup>

entry	paper no.	X	n	R	Y	IC <sub>50</sub> (μM)
73	2a	none			CH <sub>2</sub>	34.7
74	2b	none				20.5
75	2c	none			CH <sub>2</sub> CH <sub>2</sub>	36.8
76	2d	none			CH=C(CH <sub>3</sub> )	2.7
77	2e	none			CH=C(C <sub>2</sub> H <sub>5</sub> )	NA <sup>a</sup>
78	2f	none			CH=CH-CH=CH	0.77
79	2g	none				NA <sup>a</sup>
80	2h	none			CH=C(CONHOH)	66
81	2i	CH <sub>2</sub>			CH <sub>2</sub>	10.8
82	2j	CH <sub>2</sub>				17.1
83	2k	CH <sub>2</sub>			CH <sub>2</sub> CH <sub>2</sub>	1.2
84	2l	CH <sub>2</sub>			CH=C(CH <sub>3</sub> )	1.4
85	2m	CH <sub>2</sub>			CH=C(C <sub>2</sub> H <sub>5</sub> )	NA <sup>a</sup>
86	2n	CH <sub>2</sub>			CH=CH-CH=CH	1.48
87	2o	CH <sub>2</sub>				NA <sup>a</sup>
88	2p	CH <sub>2</sub>			CH=C(CONHOH)	0.2
89	2q	CH=CH			CH=C(CH <sub>3</sub> )	1.3
90	2r	CH=CH			CH=C(C <sub>2</sub> H <sub>5</sub> )	NA <sup>a</sup>
91	2s	CH=CH			CH=CH-CH=CH	9.9
92	3a		2			0.043
93	3b		3			0.13
94	3c		4			0.11
95	3d		5			0.071
96	3e		6			0.266
97	3f		7			0.209
98	4a			1-Ph-c-Pr		0.101
99	4b			1-naphthyl		0.144
100	4c			2-Ph-c-Pr		0.31
101	4d			2-naphthyl		0.115

<sup>a</sup> NA, not active at starting concentration of 30 μM.

**Definition of the 3-D QSAR Models.** Partial least-squares<sup>49</sup> (PLS), the statistical method used in deriving the 3-D QSAR models, is a variation of principal component regression in which the original variables are replaced by a small set of linear combinations thereof. The latent variables so generated are used for multivariate regression, maximizing the communality of predictor, and response variable blocks. Several attractive features of PLS need mention, such as (i) the ability to handle multivariate regression analysis in cases where the number of independent variables is greater than the number of samples as found in CoMFA<sup>39</sup> and GRID/

**Table 5.** Test Set-1 Composition<sup>31,32</sup>

entry	paper no.	R	IC <sub>50</sub> (μM)
102	3a	H	0.28
103	3b	2-Cl	5.2
104	3c	3-Cl	24.2
105	3d	4-Cl	25.2
106	3e	2-Me	0.27
107	3f	3-Me	7.9
108	3g	4-Me	10.2

**Table 6.** Test Set-2 Composition<sup>35</sup>

entry	paper no.	R	X	IC <sub>50</sub> (μM)
109	1a	Bz	(CH <sub>2</sub> ) <sub>2</sub>	0.038
110	1b	Bz	(CH <sub>2</sub> ) <sub>3</sub>	0.229
111	1c	Bz	(CH <sub>2</sub> ) <sub>4</sub>	0.125
112	1d	Bz	(CH <sub>2</sub> ) <sub>5</sub>	0.018
113	1e	Bz	(CH <sub>2</sub> ) <sub>6</sub>	0.037
114	1f	Bz	(CH <sub>2</sub> ) <sub>7</sub>	0.061
115	1g	Bz	CH=CH	9.000
116	1i	H	(CH <sub>2</sub> ) <sub>5</sub>	0.213
117	1j	CH <sub>3</sub>	(CH <sub>2</sub> ) <sub>5</sub>	0.110
118	1k	n-Pr	(CH <sub>2</sub> ) <sub>5</sub>	0.135
119	1l	Ph	(CH <sub>2</sub> ) <sub>5</sub>	0.012
120	1m	2-Ph-ethyl	(CH <sub>2</sub> ) <sub>5</sub>	0.035

**Table 7.** Statistical Results of the 3D QSAR Models 1–8

model	series	alignment	N <sup>a</sup>	Vars <sup>b</sup>	PC <sup>c</sup>	r <sup>2</sup>	q <sup>2</sup>	SDEP <sub>CV</sub>
1	TAAAs	TSA-based	26	622	2	0.98	0.85	0.62
2	TAAAs	SAHA-based	26	930	2	0.96	0.73	0.85
3	SPACAs	TSA-based	21	780	3	0.98	0.72	0.29
4	SPACAs	SAHA-based	21	805	3	0.99	0.66	0.32
5	APHA1	TSA-based	31	609	3	0.98	0.78	0.57
6	APHA1	SAHA-based	31	885	3	0.98	0.78	0.57
7	APHA2	TSA-based	31	745	3	0.97	0.76	0.62
8	APHA2	SAHA-based	31	753	2	0.95	0.78	0.62

<sup>a</sup> Number of compounds used in the model. <sup>b</sup> Number of selected variables. <sup>c</sup> Number of principal components which showed the maximum q<sup>2</sup> value.

GOLPE 3-D QSAR analyses;<sup>40</sup> (ii) the ability to work well even when interdescriptor correlations (a potential source of problem for traditional multiple linear regression) exist; and (iii) the reduction of the risk of chance correlations.<sup>50</sup>

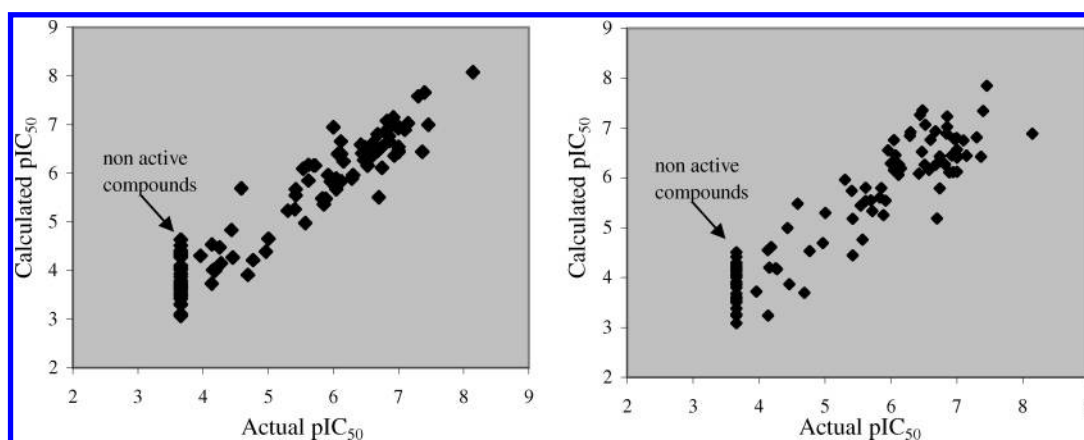
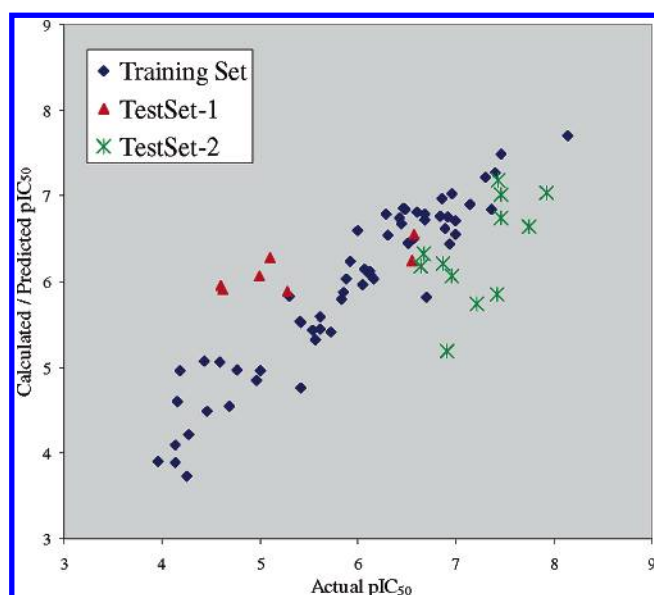
Group cross-validated PLS analyses were used to determine the optimum number of components to be used in the final QSAR models. PLS results are summarized in Tables 7 and 8 and in Figures 1 and 2.

To check for self-consistency, 3-D QSAR models were first developed for each molecular series (TAAAs, SPACAs, APHA1, and APHA2) included in the training set. According to the above-reported alignment rules, eight highly statistical significant 3-D QSAR models were obtained (named models 1–8). Such eight models were characterized by correlation coefficient (r<sup>2</sup>), cross-validated correlation coefficient (q<sup>2</sup>), and cross-validated Standard Deviation of Errors of Predic-

**Table 8.** Statistical Results of the 3D QSAR Models 9–12

model	series	alignment	$N^a$	Vars <sup>b</sup>	PC <sup>c</sup>	$r^2$	$q^2$	SDEP <sub>CV</sub>	SDEP <sub>test set-1</sub>	SDEP <sub>test set-2</sub>
9	United	TSA-based	103	726	3	0.90	0.75	0.69	1.20	1.46
10	United	SAHA-based	103	829	2	0.83	0.71	0.74	1.00	1.39
11	United OA <sup>d</sup>	TSA-based	71	659	3	0.94	0.83	0.41	1.30	1.06
12	United OA <sup>d</sup>	SAHA-based	71	770	3	0.91	0.75	0.51	0.96	0.99

<sup>a</sup> Number of compounds used in the model. <sup>b</sup> Number of selected variables. <sup>c</sup> Number of principal components which showed the maximum  $q^2$  value. <sup>d</sup> OA: only active compounds were used in the training set.

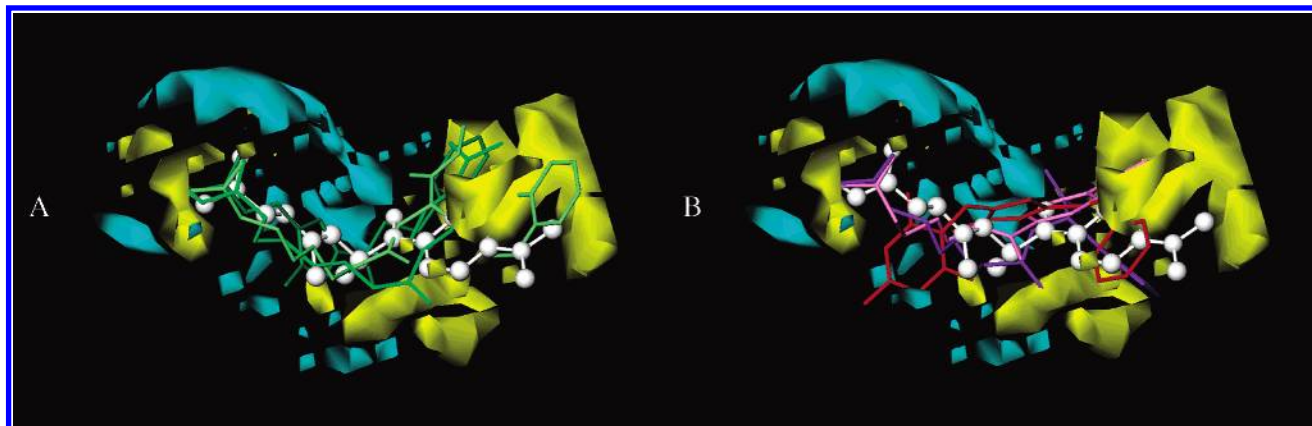
**Figure 1.** Experimental vs calculated  $pIC_{50}$  plots for the models 9 (left) and 10 (right).**Figure 2.** Experimental vs calculated/predicted  $pIC_{50}$  plot for the 3-D QSAR model 12. Blue diamonds represent the training set, and red triangles represent the test set predictions.

tion<sup>51</sup> (SDEP<sub>CV</sub>) values falling in the ranges 0.95–0.99, 0.66–0.85, and 0.29–0.85, respectively (Table 7). The statistical results listed in Table 7 are of outstanding interest if one compares them with those reported<sup>41–43</sup> for the 3-D QSAR (CoMFA and CoMSIA) applications on other series of HDAC inhibitors. In these papers the training set composition ranged from 21 to 36 leading to  $r^2$  and  $q^2$  values in the ranges of 0.939–0.980 and 0.342–0.713, respectively, in the average considerably lower to those obtained in the present study. Furthermore the robustness of the present models are even of higher values considering that the cross-validations were performed using the group method available in the GOLPE program (the leave-one-out method was used in the reported studies) and that the number of optimal latent

variable ranged from 2 to 3, while 4–6 principal components were used to define the 3-D QSAR models for the reported CoMFA or CoMSIA applications.

The self-consistency of the above models allowed us to develop two further global models (model 9 and model 10) by merging all together the four molecular series TAAs, SPACAs, APHA1, and APHA2. Again, 3-D QSAR models endowed with high statistical values were obtained showing  $r^2$ ,  $q^2$ , and SDEP<sub>CV</sub> values in the ranges 0.83–0.90, 0.71–0.75, and 0.69–0.74, respectively (Table 8, Figure 1). These two models as well as those developed for each series (models 1–8) incorporated also inactive compounds, to which was arbitrarily assigned an  $IC_{50}$  value equal to 50% of the least active compound or even a  $pIC_{50} = 0$ , led to minimal differences in the models statistics (data not shown) and predictive capability. The use of models containing information from inactive derivatives would certainly be of high utility in the design of new compounds, especially in the case of a virtual high throughput screening, allowing the medicinal chemist to directly filter only compounds that should have some level of activity and discarding those predicted in the same activity range of nonactive derivatives. On the other hand, the inclusion of nonactive compounds, to which arbitrary values were assigned, is a source of spurious data making the model less precise in detecting the highly active ones. To fill this gap and to avoid the application of arbitrarily activity values, two further 3-D QSAR models (model 11 and model 12) were also derived using only compounds displaying some level of activity. A comparison between models 9,10 and models 11,12 showed the four models to have similar statistical profiles (compare  $r^2$ ,  $q^2$ , and SDEP<sub>CV</sub> values of models 9,10 with those of models 11,12 in Table 8).





**Figure 3.** GRID/GOLPE PLS coefficient countour maps for the 3D QSAR model 12 (countour levels 0.0040 yellow,  $-0.0040$  cyan, for color code see text). To aid interpretation high active compounds (TAA\_5a in green, SPACA\_11o in light green, and APHA2\_3a in dark green) are displayed in A and low active compounds (TAA\_5c in pink, SPACA\_11a in purple, and APHA1\_25 in red) are in B. For comparison purposes TSA is also displayed in white. For the sake of clarity hydrogens are not displayed.

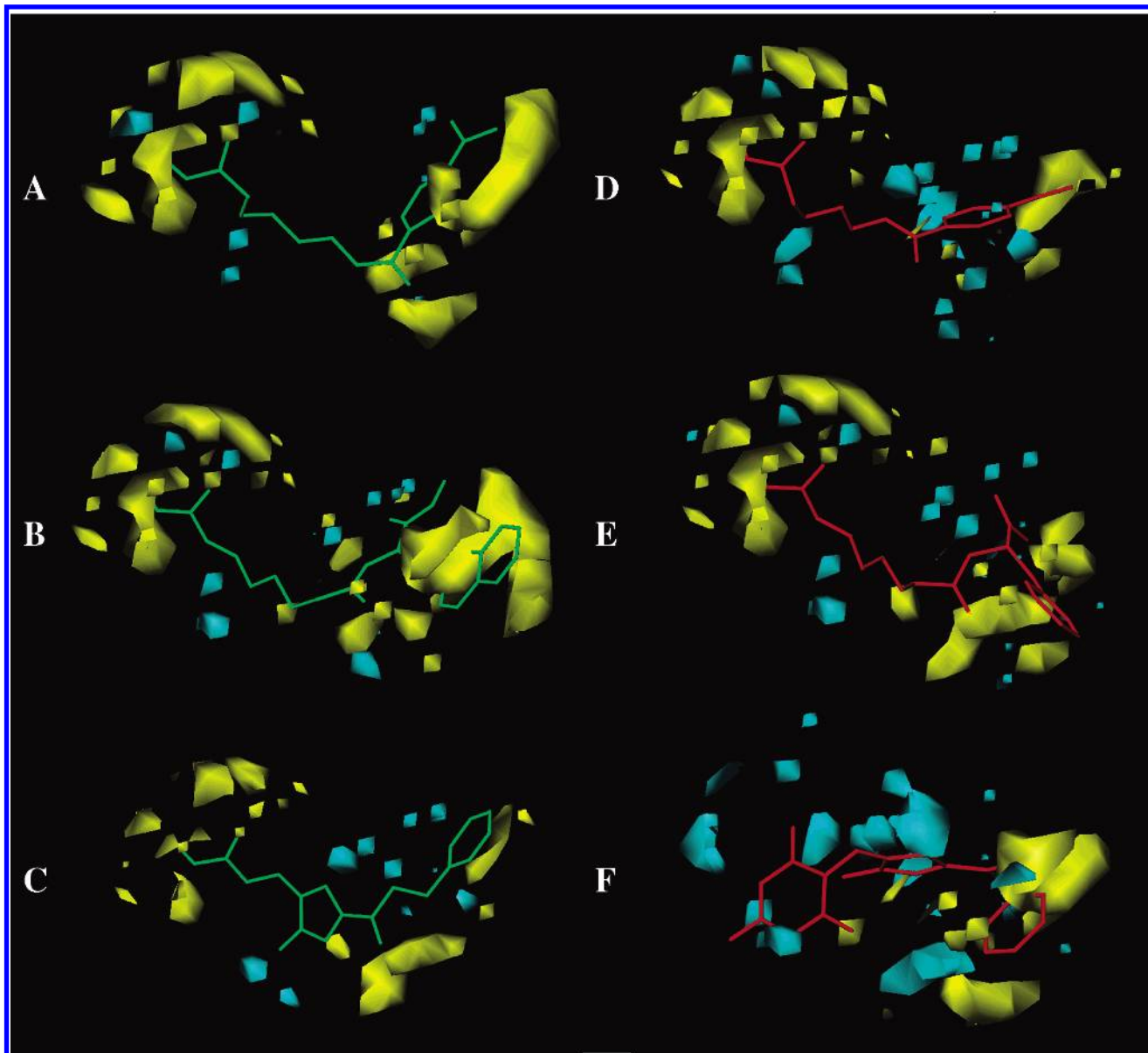
**Prediction Capability of the 3-D QSAR Models.** To test the predictive ability of the obtained models two external test sets (test set-1 and test set-2, Tables 5 and 6) of compounds not included in the training sets were used. The test sets comprised 7 recently reported new APHA derivatives<sup>31,32</sup> and 12 UBHA derivatives<sup>35</sup> both tested against the maize HD2. Molecules in test set-1 differentiated from the previous APHAs in that the cap group contain a vinyl insertion between the benzoyl group and the pyrrole ring as present in the APHA1 series, while molecules in the test set-2 belong to a new series of histone deacetylase inhibitors which display a molecular scaffold different to any of those included in the training set. Application of the 3-D QSAR models 9–12 to the test sets showed a good predictability with a standard deviation error of prediction (SDEP) values ranging from 0.96 to 1.30 (Table 8) and from 0.99 to 1.46 for test set-1 and test set-2, respectively. Regarding test set-1, although slightly overpredicting, either the TSA-based or the SAHA-based models showed to be able to discriminate between more active and less active derivatives (see SDEP values of models 9 and 10 in Table 8 and red triangles of Figure 2). On the other hand, test set-2 predictions resulted to be slightly underpredictive for either TSA-based or SAHA-based models; nevertheless the overall predicted activities trend reflected the experimental one (Figure 1, green crosses). These results suggest that the training set and test sets molecules might adopt binding modes in some way similar to either SAHA or TSA as recently observed for APHA derivatives.<sup>33,34,36</sup> In fact, comparing either the TSA- or SAHA-based fitted and cross-validated  $\text{pIC}_{50}$ s about half of the training set (52% and 44% for the SAHA-based model and 48% and 56% for the TSA-based model) returned a lower standard error of estimation (SEE) and SDEP values for either model 11 or 12 (data not shown). As proof of concept, mixed aligned models were also attempted selecting those training set compounds with the least estimation (fitting  $\rightarrow$  MIXED<sub>FIT</sub> model) and prediction errors (cross-validation  $\rightarrow$  MIXED<sub>CV</sub> model); after the proper variable pretreatment and FFD selection the two mixed models showed similar or slightly increased statistical coefficients (MIXED<sub>FIT</sub>:  $r^2 = 0.95$ ,  $q^2 = 0.83$ ; MIXED<sub>CV</sub>:  $r^2 = 0.96$ ,  $q^2 = 0.85$ ).

Interestingly, the slightly less statistically robust SAHA-based alignment afforded to more predictive 3-D QSAR models than those by TSA-based alignment (in Table 8

compare  $r^2$ ,  $q^2$ ,  $\text{SDEP}_{\text{CV}}$ ,  $\text{SDEP}_{\text{test set-1}}$ , and  $\text{SDEP}_{\text{test set-2}}$  values of models 10 and 12 with those of models 9 and 11), model 12 showing the lowest error of predictions ( $\text{SDEP}_{\text{test set-1}} = 0.96$ ,  $\text{SDEP}_{\text{test set-2}} = 0.99$ , Figure 2 and Tables A and B of the Supporting Information). Moreover, not surprisingly, only active compound containing models 11 and 12 displayed slightly improved statistical coefficients profiles if compared with those of models 9 and 10. Likely, this is due to the elimination of inactive compounds that lead to a reduced training set (from 103 to 71 compounds) with molecules all endowed with some anti-HD2 activity and that in some way do bind the enzyme.

Because of the highest predictive ability and the presence of nonactive derivatives in models 9 and 10 (Table 8), the SAHA-based model 12 was chosen as the final 3-D QSAR, and therefore the interpretation and any further consideration will be referred to this model and the related molecular alignment.

**3-D QSAR Model 12 Interpretation.** One important feature of 3-D QSARs is the graphical representation of the model making its interpretation easier. In the GOLPE software several options are available to display the final model. Among them, PLS pseudocoefficients, present field, and activity contribution plots are very useful. The PLS pseudocoefficients plot allows the visualization of the selected grid points at a determined molecules/probe interaction energy level, indicating graphical information of the whole training set. However, in this plot the signs of the coefficients can induce to errors as coefficients have opposite meaning depending on the fact that the compound produces positive or negative field values in the same area. In Figure 3, the PLS coefficients plot, that is the overall graphical representation of the 3-D QSAR model for the 71 compounds in the training set, is reported. The comparison of PLS coefficients plot to the compound present field plots helps the 3-D QSAR interpretation; in fact, the yellow and cyan areas close to the zinc binding groups (ZBG) of Figure 3, although having opposite signs, contribute both positively to the activity. The big cyan polyhedrons represent significant electrostatic and hydrogen bond interactions between the ZBG with zinc ion or active site residues. The ZBG position is also important: in fact, derivatives with an efficient ZBG (for example APHA1\_11 or APHA1\_12) have low activity because their chelation group is shifted away from zinc ion



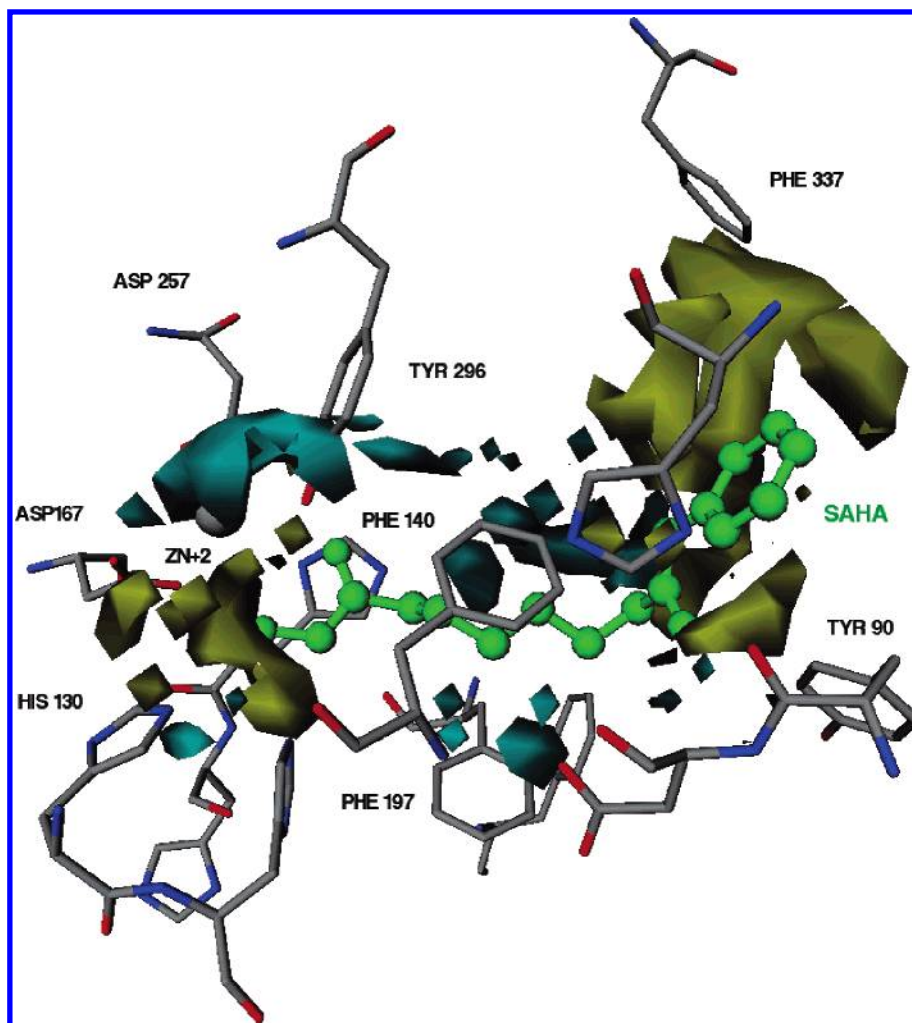
**Figure 4.** Activity contribution plots of selected representatives for each of the molecular series included in the training set. In green are reported the more active compounds and in red the lower active. A: TAA\_5a; B: SPACA\_11r; C: APHA2\_3a; D: TAA\_5c; E: SPACA\_11o; and F: APHA1\_25.

during the complex minimization. The big yellow polyhedron highlights furthermore a hydrogen bond between ZBG and the surrounding HDAC residues. Further big yellow polyhedrons run around the capping groups meaning that in this area hydrophobic interactions contribute positively to the activity. In fact, active compounds place their capping groups superimposed to that of TSA/SAHA (TAA\_5a, SPACA\_11o, and APHA2\_3a in Figure 3A), while less active derivatives (TAA\_5c, SPACA\_11a, and APHA1\_25 in Figure 3B) adopt a different binding conformation contributing negatively to the activity.

While the PLS coefficient plots give a global interpretation of the 3-D QSAR model, the activity contribution plot gives the possibility to display spatial regions that are individually important for the selected molecules. The activity contribution plot, different for every molecule within the training set, results from multiplying the values of the coefficients by the actual values of the field for each molecule. In Figure 4 the activity contribution plots of selected highly active

(Figure 4A: TAA\_5a; Figure 4B: SPACA\_11o; Figure 4C: APHA2\_3a) and poorly active compounds (Figure 4D: TAA\_5c; Figure 4E: SPACA\_11a; Figure 4F: APHA1\_25) are reported as representative samples for each above series. In the activity contribution plots cyan polyhedrons (negative) show areas where a negative contribution to the inhibitory activity is associated, while yellow polyhedrons indicate areas of positive contribution to the anti-HDAC activity.

Of particular evidence is the difference between the plots of APHA derivatives. Comparing Figure 4C (APHA2\_3a,  $IC_{50} = 0.043 \mu M$ ) with Figure 4F (APHA1\_25,  $IC_{50} = 104 \mu M$ ) it is possible to observe the conversion of some yellow colored polyhedrons (Figure 4C) into cyan colored ones (Figure 4F). In this case the most important feature responsible for the decreased activity of APHA1\_25 is the thiobarbituric moiety (see cyan polyhedrons in Figure 4F) as already above outlined. The yellow-to-cyan color conversion is only partially visible comparing the activity contribution plot of TAA\_5a with that of TAA\_5c (compare part A with part D



**Figure 5.** PLS coefficient maps superimposed to a 6 Å core of HDAC1. Bound conformation of SAHA is also displayed to aid interpretation.

of Figure 4) and activity contribution plot of SPACA\_11o with that of SPACA\_11a (compare part B with part E of Figure 4). This is due to the higher activity ratios for TAA\_5a/TAA\_5c ( $IC_{50}$  ratio = 0.05) and SPACA\_11o/SPACA\_11a ( $IC_{50}$  ratio = 0.004) than that for the APHA2\_3a/APHA1\_25 ( $IC_{50}$  ratio = 0.0004). The reduced activity ranges for the TAA and SPACA series is correctly recorded by the 3-D QSAR model 12 producing similar activity contribution plots.

As a rule, the 3-D QSAR maps cannot be employed to extract information about ligand/receptor interactions, as the model has been created just with the training set structures. This implies that any receptor model is extremely affected by a lack in design, correlation in structural features, and alignments. Considering that the alignment of the training set was based on a manual docking procedure by minimizing the complexes obtained using a virtual HDAC1<sup>34</sup> and considering the above cited HD2/HDAC1  $IC_{50}$  relationship,<sup>38</sup> it would be useful to check for matching between the HDAC1 receptor channel residues and the 3-D QSAR maps. In Figure 5, a 6 Å core of HDAC1 residues is overlapped with the 3-D PLS coefficients maps, and the close match between the NNBS residues and the PLS polyhedrons is evident. The biggest, most important polyhedrons are close to crucial residues of the HDAC catalytic core: the big cyan polyhedron is situated near or over the Zn ion, Asp257, Asp167, and Tyr296 residues, likely identifying the region

more involved in electrostatic interactions between ligand and enzyme. If the TSA/HDAC1 complex is analyzed, the cyan big contour expresses a favorable interaction, besides Zn and the ZBG group, between hydroxamate C=O and the Tyr296 hydroxyl group. Moreover the yellow polyhedrons, overlapping the ZBG hydroxyl group, point toward His130 suggesting that a hydrogen bond could occur. This implies that to display an appreciable activity an inhibitor should bear both hydrogen bond acceptor and donor to support the above interactions. Smaller yellow and cyan polyhedrons are spread along the receptor cavity underlying the importance of this region. This enzyme pocket, similar to a tube, has a wideness of 8 Å delimited by Phe197 and Phe140 as possible anchor points where the inhibitor could place insert their linker portions (joining the ZBG and the cap groups). In this context, the APHA derivatives place their pyrrole rings (part of the linker portion) between the two phenylalanines making favorable  $\pi$ - $\pi$  stacking interactions. If the linker moiety is branched or in the case of APHAs the N-pyrrole substituent is bigger than CH<sub>3</sub>, because of steric clash, upon complex minimization the ligand moves outward the entrance leading to a reduced metal chelation and lowered anti-HDAC activity, even if an efficient ZBG group like a hydroxamate is present. Two further big yellow-colored polyhedrons overlap the residues in the HDAC channel entrance, suggesting that important interactions occur between the enzyme rim and the ligand cap region. In fact all the most active compounds



display their hydrophobic caps in this zone to make favorable interactions with either Phe337 or Tyr90. Derivative SPACA\_11r, one of the most active compounds, has two hydrophobic groups in the cap portion that could contemporaneously interact with both the side chains of Phe337 and Tyr90. Furthermore, a yellow-colored polyhedron close to the Tyr90 hydroxyl suggests that molecules bearing a hydrogen bond acceptor close to the cap group could have further significant favorable interactions and thus higher anti-HDAC activity.

In conclusion, taken all together the polyhedrons well define the tubelike shape enzyme cavity confirming the self-consistence of the 3-D QSAR model 12.

## CONCLUSION

In this study we report the first GRID/GOLPE 3-D QSAR study on an enlarged heterogeneous training set of 103 HDAC inhibitors (models 9 and 10). Twelve 3-D QSAR models were developed using four congeneric series of HDACIs. All models displayed high statistical coefficient values, and using two distinct external test sets composed of seven and 12 molecules not included in the training set the global models 9–12 proved to be highly predictive displaying a low standard error of predictions. Compared to previous 3-D QSAR applications the present model was far more robust and self-consistent. Being the most predictive ( $SDEP_{\text{test set-1}} = 0.96$ ,  $SDEP_{\text{test set-2}} = 0.99$ ) and considering the structure-based alignment, the only active compound containing model 12 proved to be a valuable tool for future structure-based drug design of new HDACIs. From the 3-D QSAR model interpretation using the structure of a virtual HDAC1 it was possible to depict the areas around the training set (ligand-based approach) and enzyme residues (structure-based approach) to have a potent HDAC inhibitor.

Further 3-D QSAR studies are in progress considering a wider range of molecular diversity by the inclusion of a greater number of inhibitors in the training set and using mammalian HDAC inhibition activity values.

## EXPERIMENTAL SECTION

**Data Sets.** The training set used comprised TSA, SAHA, and 101 compounds taken from four different molecular series (TAAs,<sup>29</sup> SPACAs,<sup>30</sup> APHA1,<sup>33</sup> and APHA2<sup>28</sup>) was previously tested against HD2 enzyme. The  $IC_{50}$  values for the training set were converted to  $pIC_{50}$  ( $-\log IC_{50}$ ) values and used as dependent variables in the GRID/GOLPE QSAR analyses. The test sets were composed of 7 and 12 recently reported HD2 inhibitors.<sup>31,32,35</sup> Inactive compounds, to which was arbitrarily assigned an  $IC_{50}$  value equal to 50% of the least active compound, were included in the training set.<sup>46–48</sup> Attempts to assign different arbitrary activity values to the inactive derivatives, such as 40% and 60% of the least active compound or even a  $pIC_{50} = 0$ , led to minimal differences in the models statistics (data not shown) and predictive capability.

**Molecular Modeling and Alignments.** Three-dimensional structure building and all modeling were performed using the Macromodel<sup>52</sup> 7.1 and its graphical user interface MAESTRO version 3 on a 3 GHz AMD CPU equipped IBM-compatible workstation with the SUSE 9.0 version of the Linux operating system. Two initial conformations of

compounds in the training set and test set were generated using TSA and SAHA, respectively, extracted from the corresponding complexes with HDLP (PDB<sup>53</sup> entry code: 1c3r and 1c3s for SAHA and TSA, respectively). The training alignment was thus carried out in two ways: a SAHA-based and a TSA-based alignment. The final two molecular alignments used for the studies were obtained by minimizing the above modeled conformations into the catalytic core of a virtually mutated HDAC1.<sup>34,36</sup> During the minimization a core of residue comprised within 8 Å from the ligand was allowed to relax. A preminimization of 100 iterations to relax any steric clash was first performed on the obtained complexes. Then by the means of the MO-PAC2000<sup>54</sup> program the AM1 point charges of the whole complexes were calculated and used for the next full minimization.

Energy minimizations were performed using the BATCH-MIN module of the Macromodel program using the all atom AMBER force field<sup>55</sup> with a distance-dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/(mol Å).

**GRID Calculations.** The interaction energies were calculated by using the program GRID<sup>56</sup> (version 21) with a grid spacing of 1 Å and the grid dimensions (Å):  $X_{\min}/X_{\max}$ ,  $-10.0/24.0$ ;  $Y_{\min}/Y_{\max}$ ,  $-13.0/13.0$ ; and  $Z_{\min}/Z_{\max}$ ,  $-13.0/16.0$ .

**GOLPE Analyses.** PLS models were calculated with GOLPE 4.5.12<sup>57,58</sup> running on a SGI O2 R10000 equipped with the IRIX operating system 6.5.11. To measure the goodness of the model the statistical indices  $r^2$ ,  $q^2$ , and SDEP were employed.

$$r^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - Y_{\text{fit}})^2}$$

$$q^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - Y_{\text{CV}})^2}$$

$$SDEP = \sqrt{\frac{\sum_{i=1}^n (Y_i - Y_{\text{CV}})^2}{N}}$$

where  $Y_i$  = experimental value;  $Y_{\text{fit}}$  = recalculated value; = mean value;  $Y_{\text{CV}}$  = predicted value; and  $N$  = number of experiments.

**Probe Selection.** In this study we started with 10 probes (BOTH,  $C_{\text{sp3}}$ , DRY, OH2,  $Zn^{2+}$ ,  $NH_{\text{amide}}$ ,  $N_{\text{sp2}}$ ,  $N_{\text{sp3}}^+$ ,  $O_{\text{sp2}}$ ,  $O_{\text{CO2}}$ ) according to the nature of the HDAC1 active site. For 3-D QSAR, water probes were selected on the basis of the PCA and PLS tools available in GOLPE. Plotting the PCA and PLS scores, we evinced that the water,  $NH_{\text{amide}}$ ,  $N_{\text{sp2}}$ ,  $N_{\text{sp3}}^+$ ,  $O_{\text{sp2}}$ , and  $O_{\text{CO2}}$  probes were those better describing the molecular diversity of the set A (not shown). Selection of the water probe (OH2) was mainly based on the

experimental/recalculated plot (not shown) of a preliminary PLS analysis where the OH2 probe visually performed a better correlation other than a good discrimination between active and inactive/less active derivatives (not shown).

**Variable Preselection.** The resulting probe-target interaction energies for each compound were unfolded to produce one-dimensional vector variables for each compound, which were assembled in the so-called **X** matrix. This matrix was pretreated by first using a cutoff of 5 kcal/mol to produce a more symmetrical distribution of energy values and then zeroing small variable values and removing variables with a small standard deviation, using appropriate cutoffs. In addition, variables taking only two and three distributions were also removed.

**Smart Region Definition (SRD).** A number of seeds (1000) were selected using a D-optimal design criterion in the *weight space*. Structural differences between different molecules in the series will be reflected in groups of variables, and therefore groups were generated around each seed in the 3-D space. Variables with a distance of no more than 2 Å to the seeds were included in the groups. If two neighboring groups (with a distance smaller than 10 Å) contained the same information the groups were collapsed. The groups were used in the variable selection procedure replacing the original variables. The effect of the groups on the predictivity was evaluated, and groups instead of individual variables were removed from the data file.

**Region Selection.** The effect of the grouped variables on the predictivity was evaluated using a fractional factorial design (FFD) procedure. A number of reduced models (twice the number of variables) were built removing some of the variables according to the FFD design. The effect of dummy variables (20%) on the predictivity was calculated, and only if a variable had a positive effect on the predictivity larger than the effect of the average dummy variable was included in the final model. The FFD selection was repeated until the  $r^2$  and  $q^2$  values did not increase significantly. In the FFD selection the cross-validation was conducted using five random groups for 20 times and a maximum of tree principal components.

**Cross-Validation.** The models were validated using random groups. Molecules were assigned in a random way to five groups of equal size. Reduced models were built keeping out one group at a time. The formation of the groups was repeated 100 times and using a maximum model dimensionality of three components.

#### ACKNOWLEDGMENT

Many thanks are due to Prof. Gabriele Cruciani and Prof. Sergio Clementi (Molecular Discovery and MIA srl) for the use the GOLPE program in their chemometric laboratory (University of Perugia, Italy) and for having provided the GRID program. This work was partially supported by PRIN 2004 (A.M.) and AIRC 2005 (A.M.) grants.

**Supporting Information Available:** Model 12 calculated  $pIC_{50}$  for the training set and predicted  $pIC_{50}$  for test set-1 and test set-2 (Tables A and B). This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### REFERENCES AND NOTES

- Doenecke, D.; Gallwitz, D. Acetylation of histones in nucleosomes. *Mol. Cell. Biochem.* **1982**, *44*, 113–28.
- Loidl, P. Histone acetylation: facts and questions. *Chromosoma* **1994**, *103*, 441–9.
- Lopez-Rodas, G.; Brosch, G.; Georgieva, E. I.; Sendra, R.; Franco, L.; Loidl, P. Histone deacetylase. A key enzyme for the binding of regulatory proteins to chromatin. *FEBS Lett.* **1993**, *317*, 175–80.
- Magnaghi-Jaulin, L.; Ait-Si-Ali, S.; Harel-Bellan, A. Histone acetylation and the control of the cell cycle. *Prog. Cell. Cycle Res.* **2000**, *4*, 41–7.
- Kayne, P. S.; Kim, U. J.; Han, M.; Mullen, J. R.; Yoshizaki, F.; Grunstein, M. Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell* **1988**, *55*, 27–39.
- Rusche, L. N.; Kirchmaier, A. L.; Rine, J. The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. *Annu. Rev. Biochem.* **2003**, *72*, 481–516.
- Archer, S. Y.; Hodin, R. A. Histone acetylation and cancer. *Curr. Opin. Genet. Dev.* **1999**, *9*, 171–4.
- Brown, R.; Strathdee, G. Epigenomics and epigenetic therapy of cancer. *Trends Mol. Med.* **2002**, *8*, S43–8.
- Davis, P. K.; Brackmann, R. K. Chromatin remodeling and cancer. *Cancer Biol. Ther.* **2003**, *2*, 22–9.
- Laird, P. W. Cancer epigenetics. *Hum. Mol. Genet.* **2005**, *14 Spec. No. 1*, R65–76.
- Tycko, B. Epigenetic gene silencing in cancer. *J. Clin. Invest.* **2000**, *105*, 401–7.
- Marks, P.; Rifkind, R. A.; Richon, V. M.; Breslow, R.; Miller, T.; Kelly, W. K. Histone deacetylases and cancer: causes and therapies. *Nat. Rev. Cancer* **2001**, *1*, 194–202.
- Saunders, N.; Dicker, A.; Popa, C.; Jones, S.; Dahler, A. Histone deacetylase inhibitors as potential anti-skin cancer agents. *Cancer Res.* **1999**, *59*, 399–404.
- Greenberg, V. L.; Williams, J. M.; Cogswell, J. P.; Mendenhall, M.; Zimmer, S. G. Histone deacetylase inhibitors promote apoptosis and differential cell cycle arrest in anaplastic thyroid cancer cells. *Thyroid* **2001**, *11*, 315–25.
- Jung, M. Inhibitors of histone deacetylase as new anticancer agents. *Curr. Med. Chem.* **2001**, *8*, 1505–11.
- Mai, A.; Massa, S.; Rotili, D.; Cerbara, I.; Valente, S.; Pezzi, R.; Simeoni, S.; Ragno, R. Histone deacetylation in epigenetics: an attractive target for anticancer therapy. *Med. Res. Rev.* **2005**, *25*, 261–309.
- De Rubertis, F.; Kadosh, D.; Henchoz, S.; Pauli, D.; Reuter, G.; Struhl, K.; Spierer, P. The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **1996**, *384*, 589–91.
- Rundlett, S. E.; Carmen, A. A.; Kobayashi, R.; Bavykin, S.; Turner, B. M.; Grunstein, M. HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14503–8.
- Carmen, A. A.; Rundlett, S. E.; Grunstein, M. HDA1 and HDA3 are components of a yeast histone deacetylase (HDA) complex. *J. Biol. Chem.* **1996**, *271*, 15837–44.
- Brachmann, C. B.; Sherman, J. M.; Devine, S. E.; Cameron, E. E.; Pillus, L.; Boeke, J. D. The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes Dev.* **1995**, *9*, 2888–902.
- Gotta, M.; Strahl-Bolsinger, S.; Renauld, H.; Laroche, T.; Kennedy, B. K.; Grunstein, M.; Gasser, S. M. Localization of Sir2p: the nucleolus as a compartment for silent information regulators. *Embo. J.* **1997**, *16*, 3243–55.
- Gregoret, I. V.; Lee, Y. M.; Goodson, H. V. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *J. Mol. Biol.* **2004**, *338*, 17–31.
- Yoshida, M.; Kijima, M.; Akita, M.; Beppu, T. Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J. Biol. Chem.* **1990**, *265*, 17174–9.
- Butler, L. M.; Agus, D. B.; Scher, H. I.; Higgins, B.; Rose, A.; Cordon-Cardo, C.; Thaler, H. T.; Rifkind, R. A.; Marks, P. A.; Richon, V. M. Suberoylanilide hydroxamic acid, an inhibitor of histone deacetylase, suppresses the growth of prostate cancer cells in vitro and in vivo. *Cancer Res.* **2000**, *60*, 5165–70.
- Richon, V. M.; Zhou, X.; Rifkind, R. A.; Marks, P. A. Histone deacetylase inhibitors: development of suberoylanilide hydroxamic acid (SAHA) for the treatment of cancers. *Blood Cells Mol. Dis.* **2001**, *27*, 260–4.
- Bedalov, A.; Gathbonton, T.; Irvine, W. P.; Gottschling, D. E.; Simon, J. A. Identification of a small molecule inhibitor of Sir2p. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 15113–8.
- Brosch, G.; Lusser, A.; Goralik-Schramel, M.; Loidl, P. Purification and characterization of a high molecular weight histone deacetylase complex (HD2) of maize embryos. *Biochemistry* **1996**, *35*, 15907–14.
- Mai, A.; Massa, S.; Cerbara, I.; Valente, S.; Ragno, R.; Bottoni, P.; Scatena, R.; Loidl, P.; Brosch, G. 3-(4-Aroyl-1-methyl-1H-2-pyrrolyl)-

- N-hydroxy-2-propenamides as a new class of synthetic histone deacetylase inhibitors. 2. Effect of pyrrole-C2 and/or -C4 substitutions on biological activity. *J. Med. Chem.* **2004**, *47*, 1098–109.
- (29) Jung, M.; Brosch, G.; Kolle, D.; Scherf, H.; Gerhauser, C.; Loidl, P. Amide analogues of trichostatin A as inhibitors of histone deacetylase and inducers of terminal cell differentiation. *J. Med. Chem.* **1999**, *42*, 4669–79.
- (30) Wittich, S.; Scherf, H.; Xie, C.; Brosch, G.; Loidl, P.; Gerhauser, C.; Jung, M. Structure–activity relationships on phenylalanine-containing inhibitors of histone deacetylase: in vitro enzyme inhibition, induction of differentiation, and inhibition of proliferation in Friend leukemic cells. *J. Med. Chem.* **2002**, *45*, 3296–309.
- (31) Mai, A.; Massa, S.; Pezzi, R.; Rotili, D.; Loidl, P.; Brosch, G. Discovery of (aryloxopropenyl)pyrrolyl hydroxyamides as selective inhibitors of class IIa histone deacetylase homologue HD1-A. *J. Med. Chem.* **2003**, *46*, 4826–9.
- (32) Mai, A.; Massa, S.; Pezzi, R.; Simeoni, S.; Rotili, D.; Nebbioso, A.; Scognamiglio, A.; Altucci, L.; Loidl, P.; Brosch, G. Class II (IIa)-selective histone deacetylase inhibitors. 1. Synthesis and biological evaluation of novel (aryloxopropenyl)pyrrolyl hydroxyamides. *J. Med. Chem.* **2005**, *48*, 3344–53.
- (33) Mai, A.; Massa, S.; Ragno, R.; Cerbara, I.; Jesacher, F.; Loidl, P.; Brosch, G. 3-(4-Aroyl-1-methyl-1H-2-pyrrolyl)-N-hydroxy-2-alkylamides as a new class of synthetic histone deacetylase inhibitors. 1. Design, synthesis, biological evaluation, and binding mode studies performed through three different docking procedures. *J. Med. Chem.* **2003**, *46*, 512–24.
- (34) Mai, A.; Massa, S.; Ragno, R.; Esposito, M.; Sbardella, G.; Nocca, G.; Scatena, R.; Jesacher, F.; Loidl, P.; Brosch, G. Binding mode analysis of 3-(4-benzoyl-1-methyl-1H-2-pyrrolyl)-N-hydroxy-2-propenamide: a new synthetic histone deacetylase inhibitor inducing histone hyperacetylation, growth inhibition, and terminal cell differentiation. *J. Med. Chem.* **2002**, *45*, 1778–84.
- (35) Mai, A.; Massa, S.; Rotili, D.; Pezzi, R.; Bottoni, P.; Scatena, R.; Meraner, J.; Brosch, G. Exploring the connection unit in the HDAC inhibitor pharmacophore model: novel uracil-based hydroxamates. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4656–61.
- (36) Massa, S.; Mai, A.; Sbardella, G.; Esposito, M.; Ragno, R.; Loidl, P.; Brosch, G. 3-(4-aryloyl-1H-pyrrol-2-yl)-N-hydroxy-2-propenamides, a new class of synthetic histone deacetylase inhibitors. *J. Med. Chem.* **2001**, *44*, 2069–72.
- (37) Ragno, R.; Mai, A.; Massa, S.; Cerbara, I.; Valente, S.; Bottoni, P.; Scatena, R.; Jesacher, F.; Loidl, P.; Brosch, G. 3-(4-Aroyl-1-methyl-1H-pyrrol-2-yl)-N-hydroxy-2-propenamides as a new class of synthetic histone deacetylase inhibitors. 3. Discovery of novel lead compounds through structure-based drug design and docking studies. *J. Med. Chem.* **2004**, *47*, 1351–9.
- (38) Xie, A.; Liao, C.; Li, Z.; Ning, Z.; Hu, W.; Lu, X.; Shi, L.; Zhou, J. Quantitative structure–activity relationship study of histone deacetylase inhibitors. *Curr. Med. Chem. Anti-Cancer Agents* **2004**, *4*, 273–99.
- (39) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–67.
- (40) Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J. Med. Chem.* **1994**, *37*, 2589–601.
- (41) Liu, B.; Lu, A.-J.; Liao, C.-Z.; Liu, H.-B.; Zhou, J.-J. 3D-QSAR of sulfonamide hydroxamic acid HDAC inhibitors. *Wuli Huaxue Xuebao* **2005**, *21*, 337.
- (42) Chen, H.-f.; Kang, J.-h.; Li, Q.; Zeng, B.-s.; Yao, X.-j.; Fan, B.-t.; Yuan, S.-g.; Panay, A.; Doucet, J. P. 3D-QSAR study on apicidin inhibit histone deacetylase. *Chin. J. Chem.* **2003**, *21*, 1596–1607.
- (43) Guo, Y.; Xiao, J.; Guo, Z.; Chu, F.; Cheng, Y.; Wu, S. Exploration of a binding mode of indole amide analogues as potent histone deacetylase inhibitors and 3D-QSAR analyses. *Bioorg. Med. Chem.* **2005**, *13*, 5424–34.
- (44) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–46.
- (45) Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **1999**, *401*, 188–93.
- (46) Ragno, R.; Marshall, G. R.; Di Santo, R.; Costi, R.; Massa, S.; Rompei, R.; Artico, M. Antimycobacterial pyrroles: synthesis, anti-*Mycobacterium tuberculosis* activity and QSAR studies. *Bioorg. Med. Chem.* **2000**, *8*, 1423–32.
- (47) Di Santo, R.; Costi, R.; Artico, M.; Massa, S.; Ragno, R.; Marshall, G. R.; La Colla, P. Design, synthesis and QSAR studies on N-aryl heteroarylisopropanolamines, a new class of non-peptidic HIV-1 protease inhibitors. *Bioorg. Med. Chem.* **2002**, *10*, 2511–26.
- (48) Tafi, A.; Anastassopoulou, J.; Theophanides, T.; Botta, M.; Corelli, F.; Massa, S.; Artico, M.; Costi, R.; Di Santo, R.; Ragno, R. Molecular modeling of azole antifungal agents active against *Candida albicans*. 1. A comparative molecular field analysis study. *J. Med. Chem.* **1996**, *39*, 1227–35.
- (49) Wold, S.; Johansson, E.; Cocchi, M. PLS - partial least-squares projections to latent structures. *3D QSAR Drug Des.* **1993**, 523–50.
- (50) Clark, M.; Cramer, R. D., III The probability of chance correlation using partial least squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137–45.
- (51) Cruciani, G.; Baroni, M.; Clementi, S.; Costantino, G.; Riganelli, D.; Skagerberg, B. Predictive ability of regression models. Part I. Standard deviation of prediction errors (SDEP). *J. Chemom.* **1992**, *6*, 335–46.
- (52) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel-An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440–467.
- (53) Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980.
- (54) Stewart, J. J. P. *MOPAC 2000*; Fujitsu Limited: Tokyo, Japan, 1999.
- (55) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, I.; Thomas, E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (56) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–57.
- (57) *GOLPE*; Multivariate Infometric Analysis Srl., Viale dei Castagni 16, Perugia, Italy, 1999.
- (58) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating optimal linear PLS estimations (GOLPE): An advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20.

CI050556B