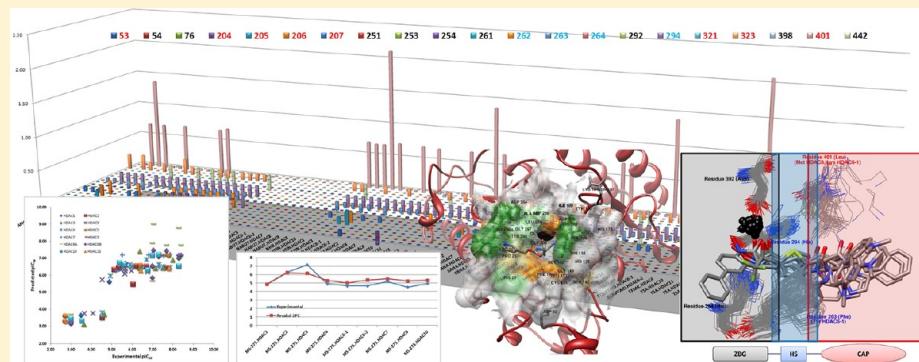


# Histone Deacetylase Inhibitors: Structure-Based Modeling and Isoform-Selectivity Prediction

Laura Silvestri,<sup>†</sup> Flavio Ballante,<sup>†</sup> Antonello Mai,<sup>‡</sup> Garland R. Marshall,<sup>†,§</sup> and Rino Ragno\*,<sup>†</sup>

<sup>†</sup>Rome Center for Molecular Design Dipartimento di Chimica e Tecnologie del Farmaco, Facoltà di Farmacia e Medicina, <sup>‡</sup>Istituto Pasteur—Fondazione Cenci Bolognetti Dipartimento di Chimica e Tecnologie del Farmaco, Facoltà di Farmacia e Medicina, Sapienza Università di Roma, P.le A. Moro 5, 00185 Rome, Italy

Supporting Information



**ABSTRACT:** An enhanced version of comparative binding energy (COMBINE) analysis, named COMBINEr, based on both ligand-based and structure-based alignments has been used to build several 3-D QSAR models for the eleven human zinc-based histone deacetylases (HDACs). When faced with an abundance of data from diverse structure–activity sources, choosing the best paradigm for an integrative analysis is difficult. A common example from studies on enzyme–inhibitors is the abundance of crystal structures characterized by diverse ligands complexed with different enzyme isoforms. A novel comprehensive tool for data mining on such inhomogeneous set of structure–activity data was developed based on the original approach of Ortiz, Gago, and Wade, and applied to predict HDAC inhibitors’ isoform selectivity. The COMBINEr approach (apart from the AMBER programs) has been developed to use only software freely available to academics.

## 1. INTRODUCTION

Histone deacetylases (HDACs) are a family of protein modifying-enzymes found in bacteria, fungi, plants and animals. In the human, 18 different isoforms have been identified and divided into 4 classes according to size, cellular localization, number of active sites and homology with yeast deacetylases.<sup>1</sup> This paper focuses on the eleven zinc-based HDACs isoforms of class I and II found in humans. Class I, that includes HDAC-1, -2, -3, and -8, is related to yeast RPD3, shares nuclear localization with the exception of HDAC3, and has ubiquitous expression. Instead, class II shows domains with similarity to yeast Hda1 and can be further divided into class IIa, which includes HDAC-4, -5, -7, and -9, and class IIb (HDAC-6 and -10) that contains two catalytic sites. HDAC-3 and members of class II have been shown to shuttle between the cytoplasm and nucleus and have tissue-specific expression. HDAC-11 is the only member of class IV. HDAC classes I, II, and IV are zinc-dependent proteases; unlike those of class III, called sirtuins, which require NAD<sup>+</sup> as cofactor.

HDACs play a key role in epigenetics—controlling gene expression involved in all aspects of biology—cell proliferation, chromosome remodeling, and gene transcription.<sup>2</sup> They regulate the acetylated state of histone proteins removing the acetyl

moiety from the *ε*-amino group of lysine residues on the *N*-terminal extension of the core histones, this leads to changes in the structure of histones and therefore modifies the accessibility of transcription enzymes with gene-promoter regions. In addition, HDACs dynamically modify the activity of diverse types of nonhistone proteins.<sup>3</sup> These include transcription factors, signal-transduction mediators, microtubules, and a molecular chaperone. In particular, distinct HDACs class I and II are overexpressed in several types of cancer. For these reasons, HDAC inhibitors (HDACIs) have been developed and approved for the treatment of cutaneous T-cell lymphoma: Merck’s Zolinza (SAHA) and Celgene’s Istodax (Romidepsin, FK228).<sup>4</sup> More recently, HDACIs have emerged as potential therapeutics for the stimulation of viral expression from infected cells in the hope of eradication of HIV infection.<sup>5</sup>

HDAC inhibitors are classified according to their chemical structure as follows: short-chain fatty acids, hydroxamic acids, benzamides, ketones, and cyclic peptides with a pendant functional group. Each HDACI shows variability in its ability

Received: March 26, 2012

Published: July 4, 2012

to inhibit particular isoforms. Unfortunately, as for SAHA and TSA, the majority of HDACIs inhibit most HDAC isoforms nonspecifically. Others, such as MS-275, a benzamide, are more selective for class I. Selective HDAC inhibitors, which affect either a single HDAC isoform or only a few isoforms within a single class, would be ideal molecular scalpels to help elucidate the individual functions of each HDAC isoform in the complexity of epigenetics.

This paper focuses on a predictive tool for the anti-HDAC activity and selectivity prediction of potential new inhibitors to enhance isoform specificity. To this aim, structure-based three-dimensional structure–activity relationship (3-D QSAR) models were derived by a comparative binding energy (COMBINEr) analysis on a series of inhibitors for which biological activities against the 11 human zinc-based HDACs isoforms were available.

## 2. OVERVIEW

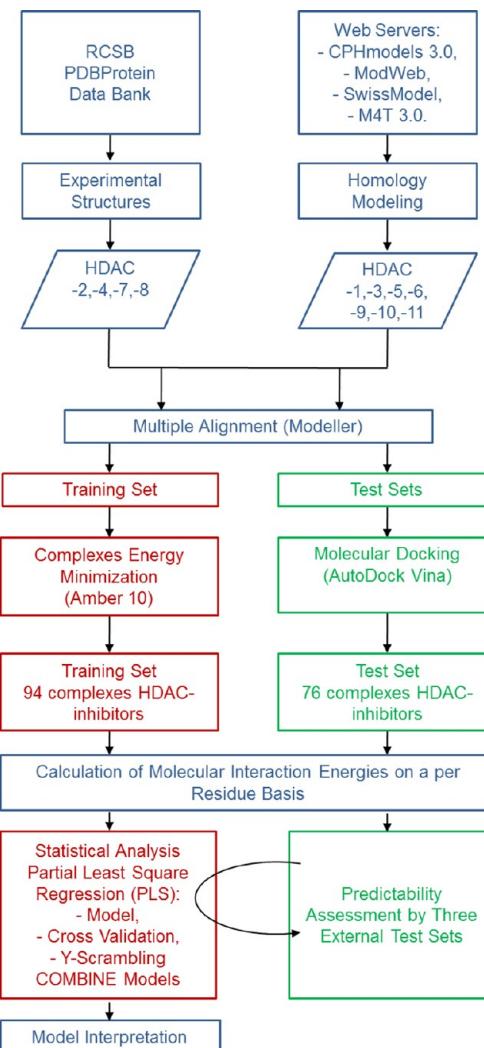
**Modified COMBINE Approach (COMBINEr).** The comparative binding energy (COMBINE) approach is a structure-based 3-D QSAR method that uses a series of receptor–ligand complexes to quantify interaction energies by molecular mechanics.<sup>6</sup> The fundamental idea of a COMBINE analysis is that a simple expression for the differences in binding affinity of a series of related ligand–receptor complexes can be derived by using multivariate statistics to correlate experimental data on binding affinities with per residue ligand–receptor interactions, computed from 3-D structures. The basis of the COMBINE method is the assumption that the protein–receptor binding free energy,  $\Delta G$ , can be approximated by a weighted sum of  $n$  terms,  $\Delta U$ , each describing the change in property  $u$  upon binding as described by the following equation:

$$\Delta G = \sum_{i=1}^n w_i \Delta u_i + C$$

From this expression, biological activities may be derived by assuming that these quantities are linear functions of  $\Delta G$ . The expression is derived by analyzing the interaction of a set of ligands with experimentally known binding affinities for a target receptor.<sup>6b</sup>

In order to apply this approach to predict the selective inhibition of HDAC isozymes, a modified protocol, called COMBINEr,<sup>7</sup> (Figure 1) used the AutoDock's AutoGrid engine to compute the components of the ligand–residues interaction energies for each ligand/enzyme complex. The PLS (*partial least squares for latent variables*) paradigm, as implemented in the R<sup>8</sup> environment, was used to derive robust, predictive COMBINEr models. Although the original COMBINE (gCOMBINE)<sup>9</sup> was available, it was decided to develop COMBINEr because it allows direct calculation of ligand/enzyme per residue interaction from docking results without further complex parametrization as required in the original COMBINE.

**Training Set.** Nine experimental 3-D structures of HDAC-2, -4, -7, and -8 cocrystallized with different ligands were retrieved from the Protein Data Bank<sup>10</sup> (Table 1). The remaining HDAC isoforms whose experimental structures were not experimentally available (HDAC-1, -3, -5, -6-1, -6-2, -9, -10, and -11) were built by homology modeling. In the case of HDAC-6, both the histone- and tubulin-catalytic domains were built (histones: HDAC-6-1 and tubulin: HDAC-6-2) with the same experimental inhibitory activities assigned to each complex.



**Figure 1.** Flowchart of the COMBINEr model generation. The red contoured boxes identify the model preparation, while the green contoured boxes refer to the test set. The blue contoured boxes refer to either training set or test set.

In addition to cocrystallized inhibitors, other compounds (Table 2) reported simultaneously from the same laboratory by Blackwell et al.<sup>13</sup> were selected. The data set composed by 15 different inhibitors and 12 HDAC isoforms was reduced from the theoretical number of 180 to 94 due to lacking of complete isozyme-inhibitory data. Therefore, the final training set summarized in Table 3 comprised 39 complexes derived with crystallized structures, built according to structural similarity of modeled inhibitors with cocrystallized compounds, and 55 complexes derived with homology models. The latter are generated according to the web-servers used for producing the homology models (see the Experimental Section).

The training sets complexes were energy minimized with Amber 10<sup>21</sup> and multiply aligned using Modeler<sup>22</sup> to establish structure-based residue equivalence. This alignment provided the structural basis for computing the molecular-interaction fields with a corresponding per-residue basis for all enzyme isoforms. Because different isoforms of HDACs show structural diversity in terms of amino-acid sequences and differed in numbers of amino acids (multitarget study), all HDAC residues were renumbered in an arbitrarily way: the same numbering were assigned to those residues showing spatial superimposition;

Table 1. PDB Codes, Ligand Names, Chemical Structures, and HDAC Inhibitory Activities of Complexes Downloaded from Protein Data Bank<sup>a</sup>

| PDB code           | HDAC Class | Number | Ligandstructure | IUPAC name   | IC <sub>50</sub> (μM) |
|--------------------|------------|--------|-----------------|--|-----------------------|
| 3MAX <sup>11</sup> | I          | 2      |                 | <i>N</i> -(4-aminobiphenyl-3-yl)benzamide (LLX)  | 0.9 <sup>11</sup>     |
| 3F07 <sup>12</sup> | I          | 8      |                 | (2E)-N-hydroxy-3-[1-methyl-4-(phenylacetyl)-1H-pyrrol-2-yl]prop-2-enamide (APHA8)  | 2.8 <sup>13</sup>     |
| 1T64 <sup>14</sup> | I          | 8      |                 | 7-(4-(dimethylamino)phenyl)-N-hydroxy-4,6-dimethyl-7-oxo-2,4-heptadienamide (TSA)  | 1.1 <sup>13</sup>     |
| 1T67 <sup>14</sup> | I          | 8      |                 | 4-dimethylamino-n-(6-hydroxycarbamoyethyl)benzamide-n-hydroxy-7-(4-dimethylaminobenzoyl)aminoheptanamide (MS-344)        | 0.249 <sup>15</sup>   |
| 1T69 <sup>14</sup> | I          | 8      |                 | octanedioic acid hydroxyamidephenylamide (SAHA)  | 2.2 <sup>13</sup>     |
| 1W22 <sup>14</sup> | I          | 8      |                 | <i>N</i> -hydroxy-4-{methyl[(5-pyridin-2-ylthiophen-2-yl)sulfonyl]amino}benzamide (NHB)                                  | 0.175 <sup>16</sup>   |
| 2VQM <sup>17</sup> | II a       | 4      |                 | <i>N</i> -hydroxy-5-[(3-phenyl-5,6-dihydroimidazo[1,2-a]pyrazin-7(8H)-yl)carbonyl]thiophene-2-carboxamide (HA3)          | 0.978 <sup>17</sup>   |
| 2VQJ <sup>17</sup> | II a       | 4      |                 | 2,2,2-trifluoro-1-{5-[(3-phenyl-5,6-dihydroimidazo[1,2-a]pyrazin-7(8H)-yl)carbonyl]thiophene-2-yl}ethane-1,1-diol (TFMK) | 0.367 <sup>17</sup>   |
| 3C0Z <sup>18</sup> | II a       | 7      |                 | octanedioic acid hydroxyamidephenylamide (SAHA)  | 0.05 <sup>13</sup>    |
| 3C10 <sup>18</sup> | II a       | 7      |                 | 7-(4-(dimethylamino)phenyl)-N-hydroxy-4,6-dimethyl-7-oxo-2,4-heptadienamide (TSA)  | 0.014 <sup>13</sup>   |

<sup>a</sup>IC<sub>50</sub>s were all evaluated in a similar way using a fluorescently labeled acetylated peptide as substrate.

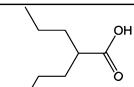
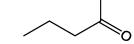
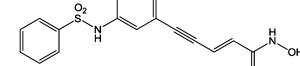
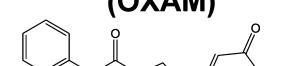
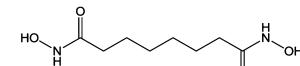
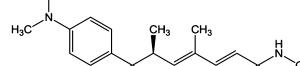
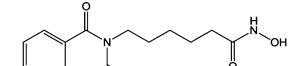
conversely, a “ghost” residue was attributed in the regions which presented structural diversity (see the Supporting Information for the multiple sequence alignment of all proteins, Supporting Info File 6). In this way, a total of 571 amino-acid residues, 12-fragmented HDAC isoform structures, were obtained. The calculation of the ligand/residues was conducted similarly as previously reported.<sup>7</sup> The calculated molecular descriptors were imported in R<sup>23</sup> to generate structure-based 3-D QSAR models. The purpose of training-set complex minimization was to generate not only 94 optimized complexes, but also to have several conformation for each HDAC useful in the subsequent preparation of test-set complexes by ligand cross-docking (see below).

Each derived COMBINEr model was subjected to internal (cross-validation) and external (test-set) assessments. Cross-validation was done using both the leave-one-out (LOO) and random five groups leave-some-out (R5G-LSO) techniques. For external validation, a series of molecules with known inhibitory activity against HDAC isozymes was selected as an external test set for the model’s predictability assessment.

#### External Test Sets for the COMBINEr Model Validation.

Three different test sets were used for external validation. The first one (modeled test set, MTS) contained a series of molecules, docked with AutoDockVina,<sup>24</sup> that showed inhibitory activity against several HDAC isoforms (Table 4).

Table 2. Training Set—Chemical Structures and HDACs Inhibitory Activities<sup>a</sup>

| HDAC                        | CLASS   | I                       |       |       |        | IIa    |       |                      | IIb   |                      |       | IV    |                      |
|-----------------------------|---|-------------------------|-------|-------|--------|--------|-------|----------------------|-------|----------------------|-------|-------|----------------------|
|                             |   | Number                  | 1     | 2     | 3      | 8      | 4     | 5                    | 7     | 9                    | 6     | 10    | 11                   |
| Chemical Structures and IDs |    | valproic acid<br>(VALP) | 1000  | 1000  | 226.08 | 228.85 | -     | 2000 <sup>19</sup>   | -     | 2000 <sup>19</sup>   | 1000  | 1000  | -                    |
|                             |    | Butyrate<br>(NABUT)     | 319   | 28.9  | 22.5   | 85.6   | 30    | 2000 <sup>19</sup>   | 30    | 2000 <sup>19</sup>   | 1000  | 292   | -                    |
|                             |    | Oxamflatin<br>(OXAM)    | 0.05  | 0.2   | 0.01   | 2.2    | 0.03  | -                    | 0.03  | -                    | 0.09  | 0.05  | -                    |
|                             |    | APHA8                   | 3.7   | 7.4   | 0.42   | 2.8    | 3.1   | -                    | 3.1   | -                    | 0.1   | 4.2   | -                    |
|                             |   | SAHA                    | 0.1   | 0.44  | 0.02   | 2.2    | 0.05  | 0.378 <sup>20</sup>  | 0.05  | 0.316 <sup>20</sup>  | 0.02  | 0.1   | 0.362 <sup>20</sup>  |
|                             |  | SBHA                    | 2.1   | 4.6   | 0.41   | 3.7    | 1.4   | -                    | 1.3   | -                    | 0.1   | 2.3   | -                    |
|                             |  | TSA                     | 0.005 | 0.021 | 0.005  | 1.1    | 0.014 | 0.0165 <sup>20</sup> | 0.014 | 0.0381 <sup>20</sup> | 0.005 | 0.005 | 0.0152 <sup>20</sup> |
|                             |  | SCRIPTAID<br>(SCRIP)    | 0.17  | 0.64  | 0.03   | 2.3    | 0.2   | -                    | 0.16  | -                    | 0.004 | 0.17  | -                    |
|                             |  | MS-275                  | 13    | 0.51  | 0.07   | 30     | 12    | -                    | 6.2   | -                    | 21    | 11.5  | -                    |

<sup>a</sup>IC<sub>50</sub>s (expressed in micromolar) were all evaluated in a similar way using a fluorescent-labeled acetylated peptide as substrate.

The second test set was comprised of a series of cocrystallized complexes structures (crystal test set, CTS) containing two HDAC-8 complexes (not available from the PDB during model development) and four bacterial HDAC homologues (Table 5). The third test set was also modeled, using largazole (a cyclotetrapeptide-containing HDAC inhibitor, largazole test set, LTS) whose crystal structure with HDAC-8 was reported,<sup>31</sup> but whose inhibitory activity was available only for four HDAC isoforms (Table 6). For LTS, largazole was docked with HDAC-1, HDAC-2, HDAC-3, and HDAC-6-1. The bacterial HDACs

complexes with hydroxamic acids were available from the PDB (Table 5).

### 3. RESULTS AND DISCUSSION

**COMBINER Models—Overall Analysis.** All final models contained 94-inhibitor/enzyme complexes spanning an activity range, expressed as pIC<sub>50</sub>, between 2.7 (NABUT against HDAC-5) and 8.4 (SCRIPTAID against HDAC-6). The statistical results of the final models are summarized in Table 7. Genetic algorithm (GA) variable-selection was applied, but provided little

**Table 3. Training Set Composition: Inhibitors' Names, Corresponding HDAC Used in the Complex, Information on Source of Protein Structure, and pIC<sub>50</sub> Values**

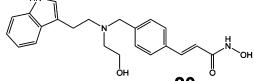
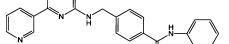
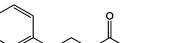
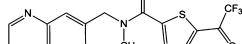
| no. | compound name | HDAC isoform | protein source | pIC <sub>50</sub> | no. | compound name | HDAC isoform | protein source | pIC <sub>50</sub> |
|-----|---------------|--------------|----------------|-------------------|-----|---------------|--------------|----------------|-------------------|
| 1   | VALP          | HDAC1        | ModWeb         | 3.00              | 48  | SAHA          | HDAC6-1      | SwissModel     | 7.70              |
| 2   | NABUT         | HDAC1        | ModWeb         | 3.50              | 49  | TSA           | HDAC6-1      | SwissModel     | 8.30              |
| 3   | MS-275        | HDAC1        | M4T            | 4.89              | 50  | SCRIP         | HDAC6-1      | SwissModel     | 8.40              |
| 4   | APHA8         | HDAC1        | SwissModel     | 5.43              | 51  | NABUT         | HDAC6-2      | CPH            | 3.00              |
| 5   | SBHA          | HDAC1        | CPH            | 5.68              | 52  | VALP          | HDAC6-2      | ModWeb         | 3.00              |
| 6   | SCRIP         | HDAC1        | ModWeb         | 6.77              | 53  | APHA8         | HDAC6-2      | CPH            | 7.00              |
| 7   | SAHA          | HDAC1        | M4T            | 7.00              | 54  | MS-275        | HDAC6-2      | M4T            | 7.00              |
| 8   | OXAM          | HDAC1        | ModWeb         | 7.30              | 55  | SBHA          | HDAC6-2      | ModWeb         | 7.00              |
| 9   | TSA           | HDAC1        | CPH            | 8.30              | 56  | OXAM          | HDAC6-2      | CPH            | 7.05              |
| 10  | VALP          | HDAC2        | Crystal        | 3.00              | 57  | SAHA          | HDAC6-2      | CPH            | 7.70              |
| 11  | NABUT         | HDAC2        | Crystal        | 4.54              | 58  | TSA           | HDAC6-2      | M4T            | 8.30              |
| 12  | APHA8         | HDAC2        | Crystal        | 5.13              | 59  | SCRIPTAID     | HDAC6-2      | M4T            | 8.40              |
| 13  | SBHA          | HDAC2        | Crystal        | 5.34              | 60  | NABUT         | HDAC7        | Crystal        | 4.52              |
| 14  | LLX           | HDAC2        | Crystal        | 6.05              | 61  | MS-275        | HDAC7        | Crystal        | 5.21              |
| 15  | SCRIP         | HDAC2        | Crystal        | 6.19              | 62  | APHA8         | HDAC7        | Crystal        | 5.51              |
| 16  | MS-275        | HDAC2        | Crystal        | 6.29              | 63  | SBHA          | HDAC7        | Crystal        | 5.89              |
| 17  | SAHA          | HDAC2        | Crystal        | 6.36              | 64  | SCRIP         | HDAC7        | Crystal        | 6.80              |
| 18  | OXAM          | HDAC2        | Crystal        | 6.70              | 65  | SAHA          | HDAC7        | Crystal        | 7.30              |
| 19  | TSA           | HDAC2        | Crystal        | 7.68              | 66  | OXAM          | HDAC7        | Crystal        | 7.52              |
| 20  | VALP          | HDAC3        | CPH            | 3.65              | 67  | TSA           | HDAC7        | Crystal        | 7.85              |
| 21  | NABUT         | HDAC3        | SwissModel     | 4.65              | 68  | VALP          | HDAC8        | Crystal        | 3.64              |
| 22  | APHA8         | HDAC3        | CPH            | 6.38              | 69  | NABUT         | HDAC8        | Crystal        | 4.07              |
| 23  | SBHA          | HDAC3        | SwissModel     | 6.39              | 70  | MS-275        | HDAC8        | Crystal        | 4.52              |
| 24  | MS-275        | HDAC3        | CPH            | 7.16              | 71  | SBHA          | HDAC8        | Crystal        | 5.43              |
| 25  | SCRIP         | HDAC3        | SwissModel     | 7.52              | 72  | APHA8         | HDAC8        | Crystal        | 5.55              |
| 26  | SAHA          | HDAC3        | CPH            | 7.70              | 73  | SCRIP         | HDAC8        | Crystal        | 5.64              |
| 27  | OXAM          | HDAC3        | SwissModel     | 8.00              | 74  | OXAM          | HDAC8        | Crystal        | 5.66              |
| 28  | TSA           | HDAC3        | SwissModel     | 8.30              | 75  | SAHA          | HDAC8        | Crystal        | 5.66              |
| 29  | NABUT         | HDAC4        | Crystal        | 4.52              | 76  | TSA           | HDAC8        | Crystal        | 5.96              |
| 30  | MS-275        | HDAC4        | Crystal        | 4.92              | 77  | MS344         | HDAC8        | Crystal        | 6.60              |
| 31  | APHA8         | HDAC4        | Crystal        | 5.51              | 78  | NHB           | HDAC8        | Crystal        | 6.76              |
| 32  | SBHA          | HDAC4        | Crystal        | 5.89              | 79  | NABUT         | HDAC9        | ModWeb         | 2.70              |
| 33  | HA3           | HDAC4        | Crystal        | 6.01              | 80  | VALP          | HDAC9        | CPH            | 2.70              |
| 34  | TFMK          | HDAC4        | Crystal        | 6.44              | 81  | SAHA          | HDAC9        | ModWeb         | 6.50              |
| 35  | SCRIP         | HDAC4        | Crystal        | 6.70              | 82  | TSA           | HDAC9        | ModWeb         | 7.42              |
| 36  | SAHA          | HDAC4        | Crystal        | 7.30              | 83  | VALP          | HDAC10       | M4T            | 3.00              |
| 37  | OXAM          | HDAC4        | Crystal        | 7.52              | 84  | NABUT         | HDAC10       | CPH            | 3.54              |
| 38  | TSA           | HDAC4        | Crystal        | 7.85              | 85  | MS-275        | HDAC10       | ModWeb         | 4.94              |
| 39  | NABUT         | HDAC5        | ModWeb         | 2.70              | 86  | APHA8         | HDAC10       | SwissModel     | 5.38              |
| 40  | VALP          | HDAC5        | ModWeb         | 2.70              | 87  | SBHA          | HDAC10       | M4T            | 5.64              |
| 41  | SAHA          | HDAC5        | CPH            | 6.42              | 88  | SCRIP         | HDAC10       | ModWeb         | 6.77              |
| 42  | TSA           | HDAC5        | CPH            | 7.80              | 89  | SAHA          | HDAC10       | ModWeb         | 7.00              |
| 43  | NABUT         | HDAC6-1      | M4T            | 3.00              | 90  | OXAM          | HDAC10       | CPH            | 7.30              |
| 44  | VALP          | HDAC6-1      | CPH            | 3.00              | 91  | TSA           | HDAC10       | CPH            | 8.30              |
| 45  | APHA8         | HDAC6-1      | SwissModel     | 7.00              | 92  | SAHA          | HDAC11       | ModWeb         | 6.44              |
| 46  | MS-275        | HDAC6-1      | ModWeb         | 7.00              | 93  | TSA           | HDAC11       | ModWeb         | 7.82              |
| 47  | SBHA          | HDAC6-1      | CPH            | 7.00              | 94  | SAHA          | HDAC6-1      | SwissModel     | 7.70              |

improvement in either descriptive or predictive performance; hence, the non-GA-optimize models were used.

Structure–activity relationships of the various HDAC inhibitors have previously been described in other studies.<sup>36</sup> Crystal structures of receptor–ligand complexes have been analyzed qualitatively or by comparison of bound ligands.<sup>37</sup> COMBINEr analysis permits quantification of structure–activity relationships through the electrostatic (Coulombic) and van der Waals interaction energies as well as additional parameters, such as solvation energy. Distinguished from the original COMBINE procedure of Ortiz,<sup>6b</sup> COMBINEr computes enzyme/ligand

interactions using the AutoGrid program based on the AMBER united-atom force field and chosen for its simpler molecular format (PDBQT). The data in Table 7 refer to the monoprobe fields (ELE, STE, DRY) and the multiprobes ones: electrostatic-steric (ELE+STE), electrostatic-desolvation (ELE+DRY), and electrostatic-steric-desolvation (ELE+STE+DRY). The reported statistical coefficients allowed estimates of goodness and robustness of each model. Results indicated the ELE+DRY model as the best. In fact, the overall generated model showed the highest conventional squared correlation coefficient ( $r^2$ ) and lowest standard deviation error of calculation (SDEC) values:

Table 4. MTS Chemical Structures and Reported HDAC Inhibitory Activities<sup>a</sup>

| HDAC                        | Class   | Number                 | I       |         |         | IIa     |         |         | IIb     |         |         | IV      |         |  |
|-----------------------------|---|------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--|
|                             |   |                        | 1       | 2       | 3       | 8       | 4       | 5       | 7       | 9       | 6       | 10      | 11      |  |
| Chemical Structures and IDs |    | LAQ824 <sup>20</sup>   | 0.00323 | 0.01570 | 0.01050 | 0.00384 | 0.00582 | 0.00558 | 0.00611 | 0.00824 | 0.00593 | 0.00841 | 0.00558 |  |
|                             |   | CI-994 <sup>25</sup>   | 0.41    | -       | 0.75    | 100     | -       | -       | -       | -       | 100     | -       | -       |  |
|                             |    | MGCD0103 <sup>26</sup> | 0.15    | 0.29    | 1.66    | -       | -       | -       | -       | -       | -       | -       | 0.59    |  |
|                             |   | JMC-23 <sup>27</sup>   | 19.3    | 69.7    | 1.99    | 100     | 58.9    | 21.0    | 29.7    | 13.3    | 93.5    | 23.1    | 34.1    |  |
|                             |    | MCL-3 <sup>28</sup>    | 64      | 65      | 260     | 93      | 2000    | 2000    | 2000    | 2000    | 240     | -       | -       |  |
|                             |   | MCL-4 <sup>28</sup>    | 0.6     | 0.6     | 2       | 4       | 140     | 25      | 150     | 430     | 0.5     | -       | -       |  |
|                             |  | MCL08-3i <sup>29</sup> | 0.58    | -       | 0.67    | -       | 0.098   | -       | -       | -       | 0.089   | -       | -       |  |
|                             |   | MCL08-3d <sup>29</sup> | 0.32    | -       | 0.23    | -       | 0.076   | -       | -       | -       | 0.36    | -       | -       |  |
|                             |  | CMC-25b <sup>30</sup>  | 0.004   | 0.021   | 0.002   | 2.58    | -       | -       | -       | -       | 0.0002  | 0.002   | -       |  |
|                             |   | CMC-7f <sup>30</sup>   | 0.057   | 0.074   | 0.018   | 1.72    | -       | -       | -       | -       | 0.011   | 0.083   | -       |  |

<sup>a</sup>IC<sub>50</sub> expressed in micromolar.

0.80 and 0.73, respectively (Figure 2A), comparable to those reported by Wade et al. in a similar application.<sup>38</sup> To assess the models' internal predictive power and robustness, two validation methods were used as follows: cross-validation (CV, internal validation) and Y-scrambling. LOO and RSG-LSO methods were chosen for cross-validation, obtaining for both  $q^2$  values of 0.76 for the ELE+DRY probe, using only 2 principal components (Figure 2B). These results suggested good internal predictability (CV) of the model. Furthermore, SDEP (standard deviation error-of-prediction) provided an estimation of model internal predictivity by means of cross-validation; values less than 1 are generally considered indexes of good predictions. Upon further inspection, a high level of inverse correlation between the DRY and STE fields was found; more than 84 out of 94 complexes

(~90%) showed a correlation coefficient between -0.60 and -0.99, rationalizing the similar statistical coefficients among models 4, 5, and 7 (Table 7). Therefore, the DRY field may be interpreted here as a probable estimation of steric interactions as well.

The charts in Figure 2 highlight the results of Table 7 and show linearity between experimental and recalculated/predicted data, expressed as pIC<sub>50</sub>. Two views of experimental versus the RSG-LSO cross-validation predictions, indicating with different symbols each inhibitor and each HDAC isoform, are shown in Figure 3. This double representation emphasizes how the COMBINER model retains the correlation within various subgroups, either considering all the training-set inhibitors versus each HDAC (correlation of anti-HDAC inhibitors

Table 5. CTS: PDB Codes, Ligand Names, Chemical Structures, and HDAC Inhibitory Activities

| PDB code           | Receptor name | Ligand structure | Ligand name   | IC <sub>50</sub> (μM) |
|--------------------|---------------|------------------|---|-----------------------|
| 3SFF <sup>32</sup> | HDAC8         |                  | (2R)-2-amino-3-(3-chlorophenyl)-1-[4-(2,5-difluorobenzoyl)piperazin-1-yl]propan-1-one (ODI) | 0.2                   |
| 3SFH <sup>32</sup> | HDAC8         |                  | (2R)-2-amino-3-(2,4-dichlorophenyl)-1-(1,3-dihydro-2H-isoindol-2-yl)propan-1-one (1DI)      | 0.09                  |
| 1C3R <sup>33</sup> | HDLP          |                  | TSA   | 0.4                   |
| 2GH6 <sup>34</sup> | HDAH          |                  | 9,9,9-trifluoro-8-oxo-n-phenylnonanamide(CF3)   | 11.19                 |
| 1ZZ1 <sup>35</sup> | HDAH          |                  | SAHA  | 0.95                  |
| 1ZZ3 <sup>35</sup> | HDAH          |                  | 3-cyclopentyl-n-hydroxypropanamide (3YP)  | 0.29                  |

Table 6. LTS: PDB Code, Ligand Name, Chemical Structure, and HDAC Inhibitory Activities

| PDB code               | Ligand structure | Ligand name    | IC <sub>50</sub> (μM) |        |        |        |        |
|------------------------|------------------|----------------|-----------------------|--------|--------|--------|--------|
|                        |                  |                | HDAC 1                | HDAC 2 | HDAC 3 | HDAC 6 | HDAC 8 |
| 3RQD <sup>3</sup><br>1 |                  | Largazolethiol | 0.0012                | 0.0035 | 0.0034 | 0.049  | -      |

Table 7. Statistical Results of the COMBINEr Models

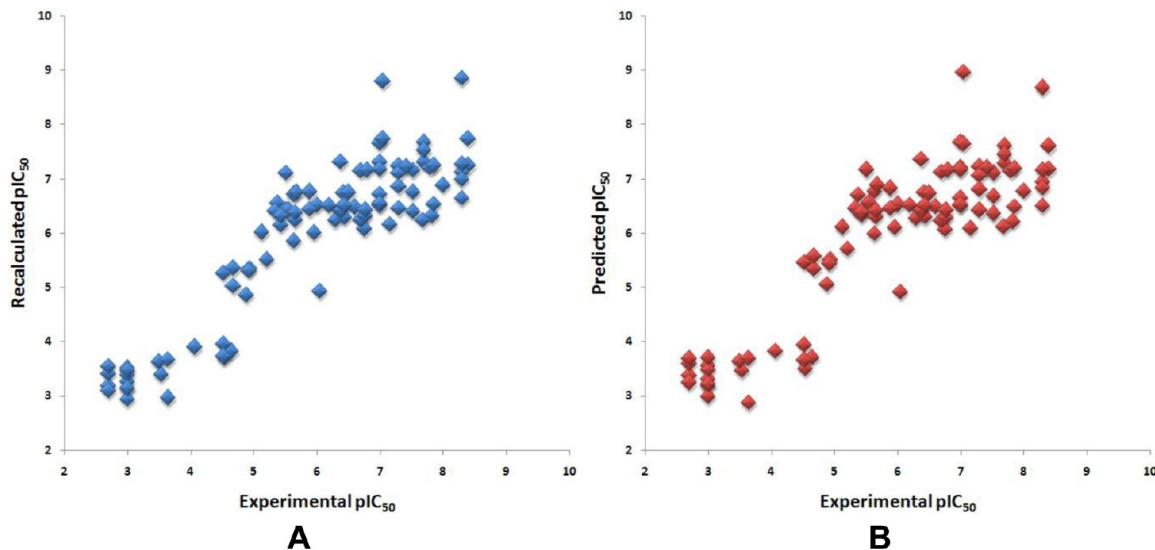
| # | Field       | PC <sup>a</sup> | r <sup>2b</sup> | SDEC <sup>c</sup> | q <sup>2</sup> <sub>RSG-LSO</sub> <sup>d</sup> | SDEP <sub>RSG-LSO</sub> <sup>e</sup> | q <sup>2</sup> <sub>LOO</sub> <sup>f</sup> | SDEP <sub>LOO</sub> <sup>g</sup> | scrambled q <sup>2h</sup> | % positive values | Max. value |
|---|-------------|-----------------|-----------------|-------------------|--|--------------------------------------|--|----------------------------------|---------------------------|-------------------|------------|
| 1 | ELE         | 2               | 0.69            | 0.91              | 0.67   | 0.94                                 | 0.68                                       | 0.93                             | 5                         | 0.07              |            |
| 2 | STE         | 2               | 0.27            | 1.40              | 0.14   | 1.52                                 | 0.15                                       | 1.51                             | n.d.                      | n.d.              |            |
| 3 | DRY         | 2               | 0.46            | 1.21              | 0.34   | 1.33                                 | 0.36                                       | 1.32                             | n.d.                      | n.d.              |            |
| 4 | ELE+STE     | 2               | 0.74            | 0.84              | 0.68   | 0.93                                 | 0.68                                       | 0.93                             | 2                         | 0.05              |            |
| 5 | ELE+DRY     | 2               | 0.80            | 0.73              | 0.76   | 0.81                                 | 0.76                                       | 0.81                             | 6                         | 0.08              |            |
| 6 | STE+DRY     | 3               | 0.54            | 1.11              | 0.33   | 1.34                                 | 0.35                                       | 1.33                             | n.d.                      | n.d.              |            |
| 7 | ELE+DRY+STE | 2               | 0.77            | 0.78              | 0.72   | 0.87                                 | 0.72                                       | 0.87                             | 4                         | 0.04              |            |

<sup>a</sup>Number of principal components used. <sup>b</sup>Conventional squared correlation coefficient. <sup>c</sup>Standard deviation error of recalculations. <sup>d</sup>Cross-validation coefficient for the RSG-LSO method. <sup>e</sup>Standard deviation error of prediction for the RSG-LSO cross-validation. <sup>f</sup>Cross-validation coefficient for the LOO method. <sup>g</sup>Standard deviation error of prediction for the LOO cross-validation. <sup>h</sup>Percentage of positive cross-validation coefficients for the scrambling method and its maximum values.

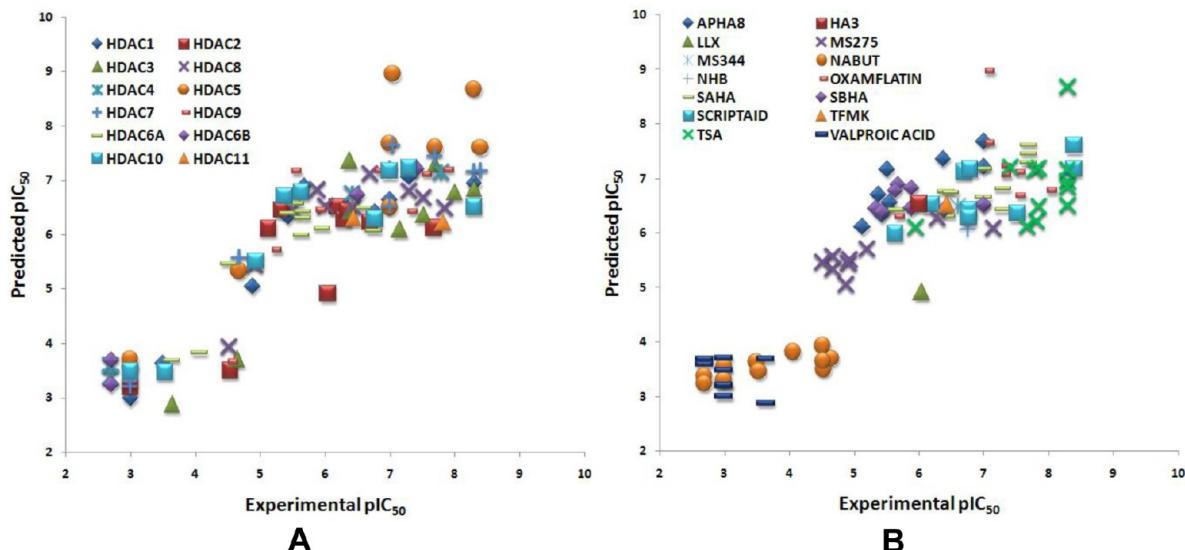
potency, Figure 3) or considering the each inhibitor binding into different HDAC isoforms (correlation of selectivity, Figure 3). This latter consideration is consistent and supported the fact that the LOO and RSG-LSO cross-validation  $q^2$ 's showed the same values. Furthermore, to check for methodological self-consistency, reduced COMBINEr models built for several inhibitors against each HDAC isoform (inhibition potencies) and for each

inhibitor against several HDAC isoforms (selectivity issue) revealed the existence of relationships with  $r^2$  ranging from 0.7 to 0.8.

Finally, both robustness and absence-of-chance correlation of the COMBINEr models listed in Table 7 were checked by random scrambling (Y-scrambling). Through this approach, a random reassignment of inhibitory activity to compounds of the



**Figure 2.** Fitting (A) and RSG-LSO cross-validation (B) plots, for the ELE+DRY model (Table 7).



**Figure 3.** RSG-LSO cross-validation predictions depicted by HDAC isoforms (A) and by inhibitor (B).

data set was achieved to generate numerous data sets; for each scrambled data set, a RSG-LSO cross-validation was run. One-hundred Y-scrambling runs were examined; their analysis revealed that only 6% of all Y vectors had a correlation with the original Y values with maximum scrambled  $q^2$  of only 0.08 in the case of ELE+DRY probe. Regarding the other models, in the case of ELE and ELE+STE+DRY, a chance correlation of 4% and 5% with a  $q^2$  maximum value of 0.04 and 0.07 were observed, respectively. The ELE+STE probe showed a chance correlation of 2% with a  $q^2$  maximum value of 0.05. These correlations appear random and excluded possible correlations between the original Y vector and the scrambled Y vectors. For the best model (ELE+DRY) in 100 random scrambled models, the number of positive  $q^2$  values were only 6 leading to a probability of chance correlation lower than 1% with a  $q^2$  value of 0.1, quite acceptable results considering the cross-validation coefficients of 0.76 of the model. Cross-validation runs using the most stringent leave-half-out method confirmed the robustness of the models (data not shown).

**ELE and DRY Model Interpretation.** Interpretation of COMBINEr models can identify the residues relevant for differences in activity and quantify their relative importance. To this aim, the PLS-coefficients (Figure 4) and activity-contribution plots (Figure 5) are useful. The former provides a global view and gives information on all of the training set. The sign and the magnitude of PLS coefficient of an energy term multiplied by the corresponding energy term (field) shows the influence of the corresponding residue on ligand binding.<sup>6c</sup> Interpretation of the PLS coefficients can lead, however, to possible misconceptions. A positive PLS coefficient for an attractive, negative energy term indicates a term that contributes favorably to binding affinity (resulting in a more negative  $\Delta G$  value). A positive PLS coefficient for a repulsive, positive energy term indicates a term that is unfavorable for binding affinity (resulting in a more positive  $\Delta G$  value). On the other hand, a negative PLS coefficient will result in an energy term favoring binding when the energy term is positive (repulsive) and disfavoring binding when the energy term is negative (attractive).<sup>38</sup> The PLS-coefficient plot is shown in Figure 4A.

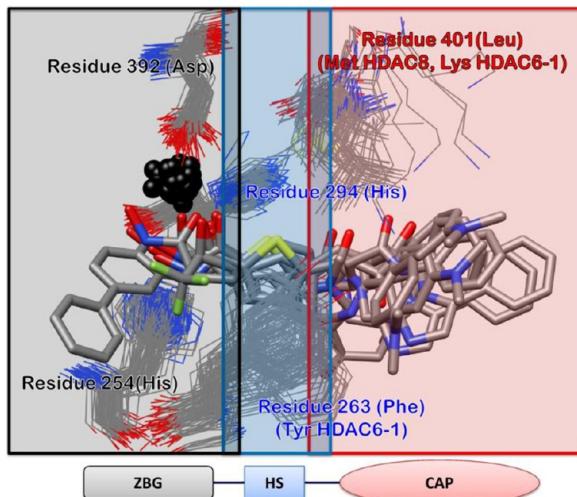


Figure 4. PLS Coeff (A), StDev (B), and PLS Coeff  $\times$  StDev plots for the ELE+DRY COMBINER model (C). The residue were selected using a PLS-coefficient threshold value of 0.001. Residue numbers are color-coded according to Table 8. The residue numbers reported correspond to those in Supporting Information File 6.

By multiplying the PLS coefficients with the field values, the activity-contribution plots were obtained for each training-set compound (see below). As can be seen (Table 8 and Figure 4), the COMBINER model can explain isoform selectivity considering only 34 residues of the enzymes (Table 8) even

though all residues of the eleven HDAC isoforms with a PLS coefficient greater than 0.001 were included in the analyses.

To analyze the significance of the fields (ELE and DRY) and the contribution for each ligand–residue interaction, the residues were color-coded in Table 8. The residues located in the rim region are colored red, while the residues forming the central



**Figure 5.** Four most important residues (MIRs, see text) from the COMBINEr model analysis. The labels and regions are color-coded: in red are the residues in the HDAC's rim region; in blue are those forming the central tube channel; and in black are those in the proximity of the catalytic Zn ion. The zinc binding region (black line box), the connection region (blue line box), and the CAP region (red line box) are also highlighted to recall the HDAC pharmacophore model depicted at the bottom. ZBG: Zn-binding group. HS: hydrophobic spacer. CAP: hydrophobic capping group.

channel are blue, and those in proximity to the catalytic Zn ion are black (Supporting Information File 2). In Figure 4B are reported the ligand/residue/interactions standard deviations (StDev) used to produce the PLS Coeff  $\times$  StDev plot (Figure 4C) in which the PLS coefficients are weighted so that the global importance of the interactions can be understood similar to a standard 3-D QSAR model.<sup>39</sup> The variables reported in Figure 4 and Table 8 are significant for the model; however, the most important residues that modulate the inhibitory activities are as follows: 254 (His for all the HDACs, in the Zn-binding site), 294 (His for all the HDACs, either in the Zn- or tube-binding sites), and 392 (Asp for all the HDACs, in the Zn-binding site) mainly for the ELE field while 263 (Tyr for HDAC-6-1 and Phe for all the others in the tube-binding sites) and 401 (Met for HDAC-8, Lys for HDAC-6-1 and Leu for all the others, in the rim-binding site) for the DRY field (Figure 5). Residue 254 has also some negative modulating factor in the DRY field. These five residues account for 95% of the explained variance ( $\sim 80\%$ ) of the model indicating that interactions of ligands with these four residues are of major importance in determining the inhibitor potencies (coarse tuning, Figure 6). Fine tuning of both potency and selectivity result from other contributions, and therefore, each isoform needs to be inspected individually.

Regarding the importance of the overall interactions, the sums for either the ELE or DRY activity contributions for each training-set complex are shown in Figure 7. While the DRY field contribution mostly modulates the activities (bigger red bars on bulkier compounds), the ELE contribution becomes more important in modulating the low activities of the smaller inhibitors (bigger blue bars on short fatty acid inhibitors), NABUT and VALPROIC ACID (VA), due to missing interactions with residue 401 and others at the enzymes' rims (Figure 8). Indeed, the COMBINEr model correctly indicates that NABUT and VA miss residue 401's contributions so activity contributions from other main residues (254, 294, and 392 of the

ELE field) are highly negative ranging from  $-0.27$  to  $-1.02$  and from  $-0.14$  to  $-1.02$  for NABUT and VA, respectively.

**Field ELE.** All residues selected having PLS Coeff higher than 0.001, except for 398, showed positive values, indicating that all the electrostatic interaction are attractive (Figure 4A). Indeed the PLS Coeff  $\times$  StDev plot clearly indicates that all electrostatic interactions are positively contributing to the model. In particular, plots in Figure 4 show that the ELE field is definitively more important in the inner part (black-labeled residues) of the HDACs catalytic domains than for residues forming the channel (blue-labeled residues in Figure 4) and those at the entrance rim (red-labeled residues in Figure 4) where only four and five out of 27 residues displayed PLS coefficients higher than the chosen threshold value.

In the outer part of the enzymes, the five selected residues (Figure 4) do not show appreciable activity contributions highlighting that these parts are not associated with high variation in ligand/enzyme electrostatic interactions. Detectable negative values relate to small compounds (NABUT and VA) for which the model correctly records the missing contribution.

Regarding the channel-forming residues, 294 (at the edge between the channel and the bottom of the HDAC-binding sites) displayed the highest values in all three plots of Figure 4. Indeed, this residue (a conserved histidine for all HDACs) is primarily involved in modulating the potency between small inhibitors (NABUT and VA) and channel-filling inhibitors (i.e., SAHA and TSA). For NABUT and VA, diminished interactions with residue 294 account for 0.8 to 1.0 decrement in activity. To some extent, the fact that either NABUT or VA are carboxylic acids indicates that higher negative charge (NABUT and VA were modeled as carboxylates, thus bearing a discrete negative formal charge) in proximity to residue 294 is unproductive. Analogous to a CoMFA analysis, the high PLS Coeff  $\times$  StDev values for residue 294 represent a blue polyhedron, placed in the same space of 294, indicating that an enhanced negative charge decreases the overall activity, while a positive-charged group (or a less negative one) is preferred to maintain the activity (the maximum contribution associated with 294 is lower than 0.01). Among the other channel-forming selected residues, 262 (always a Gly), 263 (mostly a Phe), and 264 (always a Cys), the most interesting is residue 263 involved in modulating the activity decrement for small compounds, in particular for VA.

Most of the ELE-selected residues (18 out of 27) are in the deep part of the channels around the catalytic Zn. Of particular interest are residues, involved in HDAC catalytic process conserved among the 12 isoforms, as follows: residues 253 (His), 254 (His), 292 (Asp), 392 (Asp), and 571 (Zn). In general the activity contribution associated with these five residues modulates the activity decrement for carboxylate-based zinc-binding groups. As examples, residues 253 (SAHA in HDAC-1) and 254 (SAHA in HDAC-3, HDAC-4, and HDAC-6-2; and SBHA in HDAC-4 and HDAC-8) are associated with a positive activity contribution of about 0.1.

**Field DRY.** The DRY field gives a rough estimation of steric interactions. Between ELE and DRY selected residues about 35% of these are shared (12 out of 34) in significance, nevertheless, for the DRY field a totally different and more complicated scenario can be observed on the relative importance of each residue. In general, the most important modulating interaction relates to 401Leu, replaced by Met in HDAC-8 or by Lys in HDAC-6-1 (Table 8). Upon deeper inspection (not considering the small-molecule complexes, NABUT, VA, and NHB), only 27 of 94 activities are modulated by residue 401 with activity contribu-

Table 8. List of the Most Important Residues to Interpret the COMBINER Model<sup>a</sup>

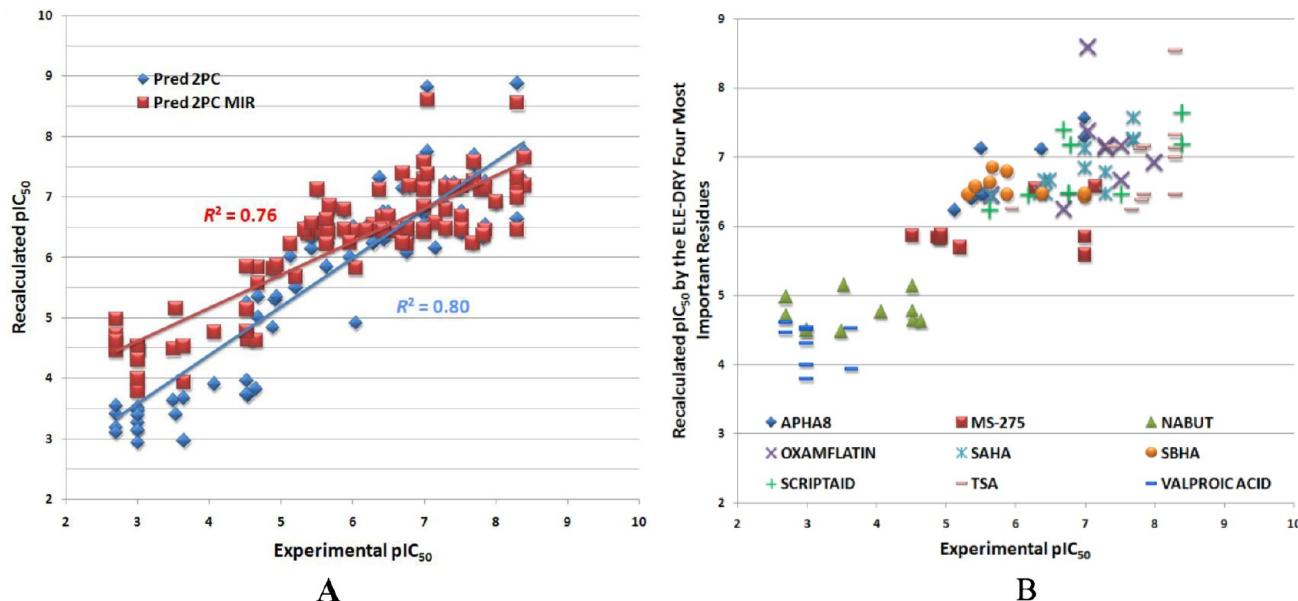
| N. of residuals | 53      | 54     | 76     | 204    | 205    | 206    | 250    | 251    | 253     | 254     | 261     |         |
|-----------------|---------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|---------|
| CLASS I         | HDAC1   | HIS28  | PRO29  | ARG34  | GLU98  | -      | -      | GLY138 | LEU139  | HIS140  | HIS141  | SER148  |
|                 | HDAC2   | HIE22  | PRO23  | ARG28  | GLU92  | -      | -      | GLY132 | LEU133  | HIE134  | HIE135  | SER142  |
|                 | HDAC3   | HIS22  | PRO23  | ARG28  | ASP92  | -      | -      | GLY132 | LEU133  | HIS134  | HIS135  | SER142  |
|                 | HDAC8   | -      | -      | ARG37  | TYR100 | -      | -      | GLY140 | TRP141  | HIS142  | HIS143  | MET161  |
| Class IIa       | HDAC4   | -      | -      | ARG32  | -      | -      | -      | PRO151 | GLY152  | HIE153  | HIE154  | MET840  |
|                 | HDAC5   | HIS704 | PRO705 | ARG710 | -      | -      | -      | PRO830 | GLY831  | HIS832  | HIS833  | ASP137  |
|                 | HDAC7   | HIE27  | PRO28  | ARG33  | -      | -      | -      | PRO153 | GLY154  | HIE155  | HIE156  | CYS137  |
|                 | HDAC9   | -      | -      | ARG660 | -      | -      | -      | PRO780 | GLY781  | HIS782  | HIS783  | MET163  |
| Class IIb       | HDAC6-1 | PHE19  | PRO20  | ARG25  | THR84  | TYR85  | -      | PRO127 | GLY128  | HIS129  | HIS130  | SER150  |
|                 | HDAC6-2 | HIS19  | PRO20  | ARG25  | -      | -      | PHE85  | PRO127 | GLY128  | HIS129  | HIS130  | MET790  |
|                 | HDAC10  | GLU24  | ILE25  | ARG30  | -      | -      | -      | PRO132 | GLY133  | HIS134  | HIS135  | ASN142  |
| Class IV        | HDAC11  | HIS35  | PRO36  | LYS41  | PRO102 | -      | -      | GLY140 | PHE141  | HIS142  | HIS143  | GLY150  |
| N. of residuals | 262     | 263    | 264    | 291    | 292    | 293    | 294    | 295    | 316     | 321     | 322     |         |
| CLASS I         | HDAC1   | GLY149 | PHE150 | CYS151 | ILE175 | ASP176 | ILE177 | HIS178 | HIS179  | LYS200  | GLU203  | TYR204  |
|                 | HDAC2   | GLY143 | PHE144 | CYS145 | ILE169 | ASP170 | ILE171 | HIE172 | HIE173  | LYS194  | TYR198  | -       |
|                 | HDAC3   | GLY143 | PHE144 | CYS145 | ILE169 | ASP170 | ILE171 | HIS172 | HIS173  | LYS194  | ASN197  | TYR198  |
|                 | HDAC8   | GLY151 | PHE152 | CYS153 | LEU177 | ASP178 | LEU179 | HIE180 | HIS181  | LYS202  | GLY206  | PHE207  |
| Class IIa       | HDAC4   | GLY162 | PHE163 | CYS164 | TRP190 | ASP191 | VAL192 | HIE193 | HIE194  | ARG215  | ASN220  | PHE221  |
|                 | HDAC5   | GLY841 | PHE842 | CYS843 | TRP869 | ASP870 | ILE871 | HIS872 | HIS873  | ARG894  | ASN899  | PHE900  |
|                 | HDAC7   | GLY164 | PHE165 | CYS166 | TRP192 | ASP193 | VAL194 | HIE195 | HIE196  | ARG217  | ASN222  | PHE223  |
|                 | HDAC9   | GLY791 | PHE792 | CYS793 | LEU819 | ASP820 | VAL821 | HIS822 | HIS823  | ARG844  | ASN849  | PHE850  |
| Class IIb       | HDAC6-1 | GLY138 | TYR139 | CYS140 | TRP166 | ASP167 | VAL168 | HIS169 | HIS170  | ARG191  | ARG196  | PHE197  |
|                 | HDAC6-2 | GLY138 | PHE139 | CYS140 | TRP167 | ASP168 | VAL169 | HIS170 | HIS171  | ARG192  | THR197  | PHE198  |
|                 | HDAC10  | GLY143 | PHE144 | CYS145 | TRP171 | ASP172 | VAL173 | HIS174 | HIS175  | ARG196  | ARG201  | PHE202  |
| Class IV        | HDAC11  | GLY151 | PHE152 | CYS153 | LEU180 | ASP181 | ALA182 | HIS183 | GLN184  | ASN205  | ILE208  | TYR209  |
| N. of residuals | 323     | 391    | 392    | 397    | 398    | 399    | 401    | 439    | 440     | 441     | 442     |         |
| CLASS I         | HDAC1   | -      | SER263 | ASP264 | ASP269 | ARG270 | -      | LEU271 | GLY300  | GLY301  | GLY302  | TYR303  |
|                 | HDAC2   | PHE199 | ALA257 | ASP258 | ASP263 | ARG264 | -      | LEU265 | GLY294  | GLY295  | GLY296  | TYR297  |
|                 | HDAC3   | PHE199 | ALA258 | ASP259 | ASP264 | ARG265 | -      | LEU266 | GLY295  | GLY296  | GLY297  | TYR298  |
|                 | HDAC8   | PHE208 | ALA266 | ASP267 | ASP272 | PRO273 | -      | MET274 | GLY303  | GLY304  | GLY305  | TYR306  |
| Class IIa       | HDAC4   | PHE222 | PHE284 | ASP285 | HIE290 | PRO291 | THR292 | LEU294 | GLU324  | GLY325  | GLY326  | HIE327  |
|                 | HDAC5   | PHE901 | PHE963 | ASP964 | HIS969 | LEU970 | SER971 | LEU973 | GLU1003 | GLY1004 | GLY1005 | HIS1006 |
|                 | HDAC7   | PHE224 | PHE286 | ASP287 | HIE292 | PRO293 | ALA294 | LEU296 | GLU326  | GLY327  | GLY328  | HIE329  |
|                 | HDAC9   | PHE851 | PHE913 | ASP914 | HIS919 | THR920 | PRO921 | LEU923 | GLU953  | GLY954  | GLY955  | HIS956  |
| Class IIb       | HDAC6-1 | TRP198 | PHE259 | ASP260 | ASP265 | PRO266 | -      | LYS267 | GLU297  | GLY298  | GLY299  | TYR300  |
|                 | HDAC6-2 | PHE199 | PHE260 | ASP261 | ASP266 | PRO267 | -      | LEU268 | GLU298  | GLY299  | GLY300  | TYR301  |
|                 | HDAC10  | TRP203 | PHE264 | ASP265 | ASP270 | PRO271 | GLU272 | -      | GLU302  | GLY303  | GLY304  | TYR305  |
| Class IV        | HDAC11  | -      | THR260 | ASP261 | ASP266 | ARG267 | -      | LEU268 | SER301  | GLY302  | GLY303  | TYR304  |

<sup>a</sup>The labels are color-coded: in red are the residues in the HDAC's rim region; in blue are those forming the central tube channel; and in black are those in the proximity of the catalytic Zn. The residues labels were color-coded according to the reported pharmacophoric model.<sup>1</sup> The residues were selected using a PLS-coefficient threshold value of 0.001. See Supporting Information File 001 for 3-D graphical disposition of the listed residues in each HDAC isoform.

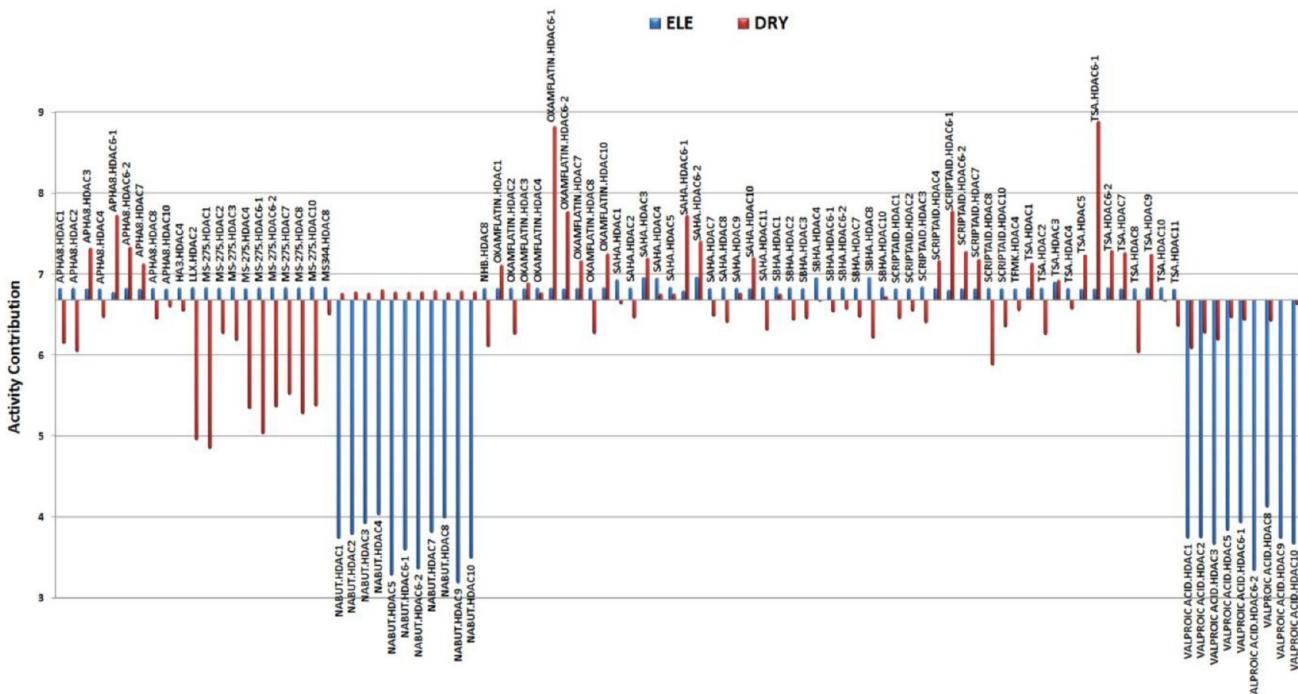
tions ranging between 0.7 and 2.13 (Supporting Information File 1 Figure SI-1).

Without considering the contribution of residue 401, it is evident from the plot in Figure 8B that the other 10 residues play a major role in modulating the overall biological activities (Supporting Information File 1 Figure SI-2, Table SI-1). Seven out of 10 residues (204, 253, 254, 262, 263, 294, and 442) are related to negative modulating values, while the other three (205, 206, and 323) are positive modulators. Residue 263 (Tyr for HDAC-6-1 and Phe for the others) located in the wall of the channel shows the largest range with larger negative values. No

specific pattern is detected for residue 263 in modulating regarding the different enzyme classes or inhibitor structures (Supporting Information File 1 Figure SI-3). The small inhibitor NABUT is not influenced by residue 263, likely due to the fact that there are no direct contacts. Residue 442 (His for Class IIa and Tyr for the others) located in the bottom of the binding sites shows the largest range with larger negative values associated mainly with class I complexes, with particular reference to HDAC-8 (Supporting Information File 1 Figure SI-4) thus suggesting that interaction with this residues might be used to selectively avoid inhibition of HDAC-8.



**Figure 6.** Comparison between the cross-validation predictions for the full model (blue squares) and with only the four most-important residues (MIRs). The coarse tuning of the relationships by the MIRs is indicated by the red squares in panel A. The differences between the red and blue squares indicate the importance of fine-tuning determined by relatively minor interactions. In B are reported the MIR predictions classified by inhibitor type. For comparison purposes, only inhibitors for which isozyme profiles of inhibition data were available are shown.



**Figure 7.** ELE and DRY total-activity contributions. The constant (PLS intercept) of the COMBINER equation takes the value of 6.68. The sum of ELE and DRY contribution is obtained by the algebraic sum of all per-residue contributions.

Residue 254 (His in the zinc-binding region) is second with the higher StDev value and from Supporting Information File 1 Figure SI-5 clearly negatively modulates mainly nonhydroxamate inhibitors making complexes (LLX, MS-275, and VA) consistently with that reported for the ELE field. Residue 204 (of various nature present on the rim of 6 out of 12 HDACs) and 294 (His, a channel-forming residue) are also negative-modulating residues, but the associated low standard deviation indicates that no selectivity can be attributed to the DRY

interactions (Supporting Information File 1 Figure SI-6 and 7); residue 204 seems to specifically modulate the inhibitory activity for HDAC-8 complexes (Supporting Information File 1 Figure SI-7). Considering the high correlation between DRY and STE, interactions with residues 263 and 294 are of crucial importance for optimal fitting of inhibitors in the HDAC channels.

Among the three DRY positive-modulating residues, 323, an aromatic side-chain-bearing residue missing in HDAC-1 and HDAC-11, shows the highest maximum-activity contribution

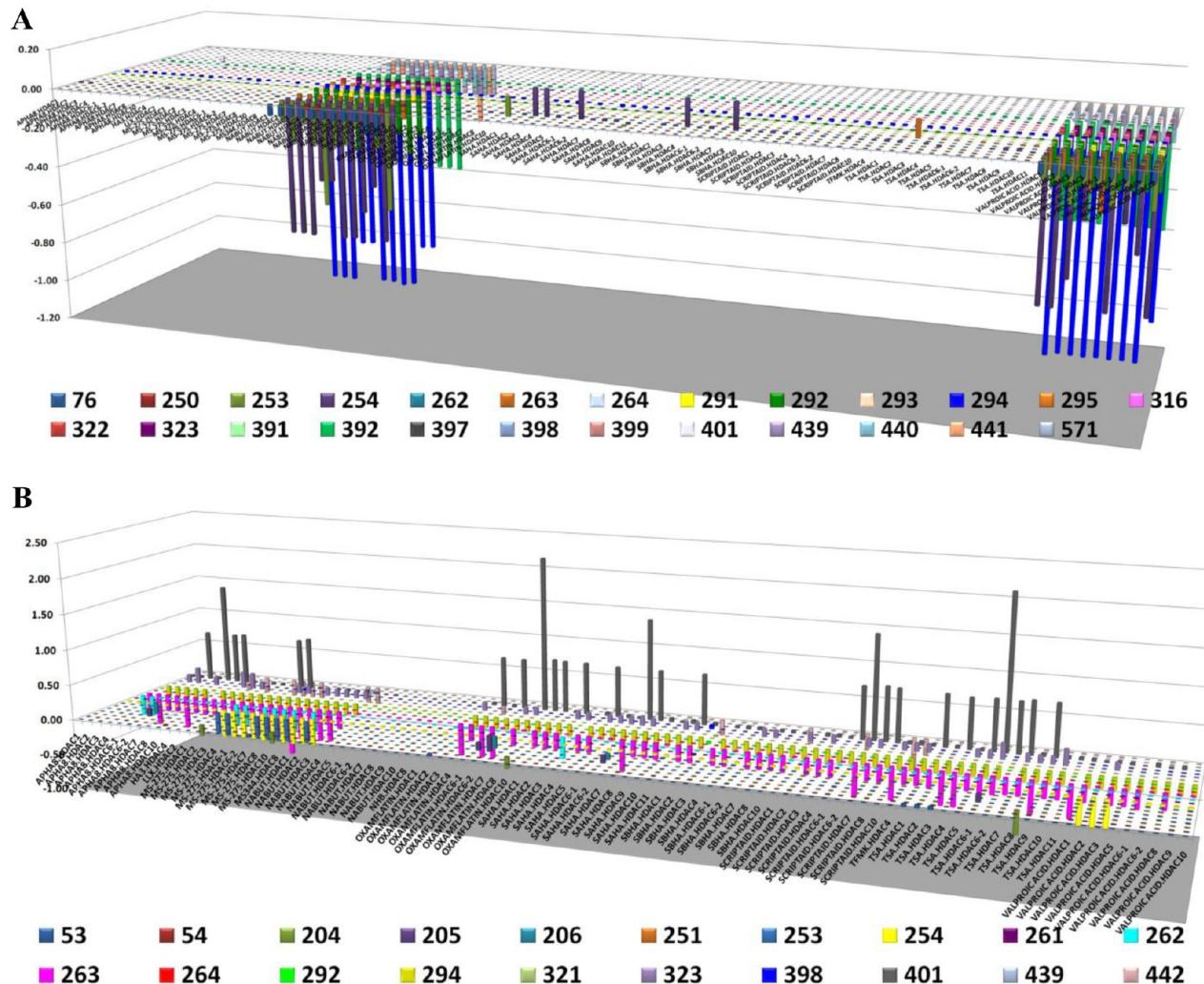


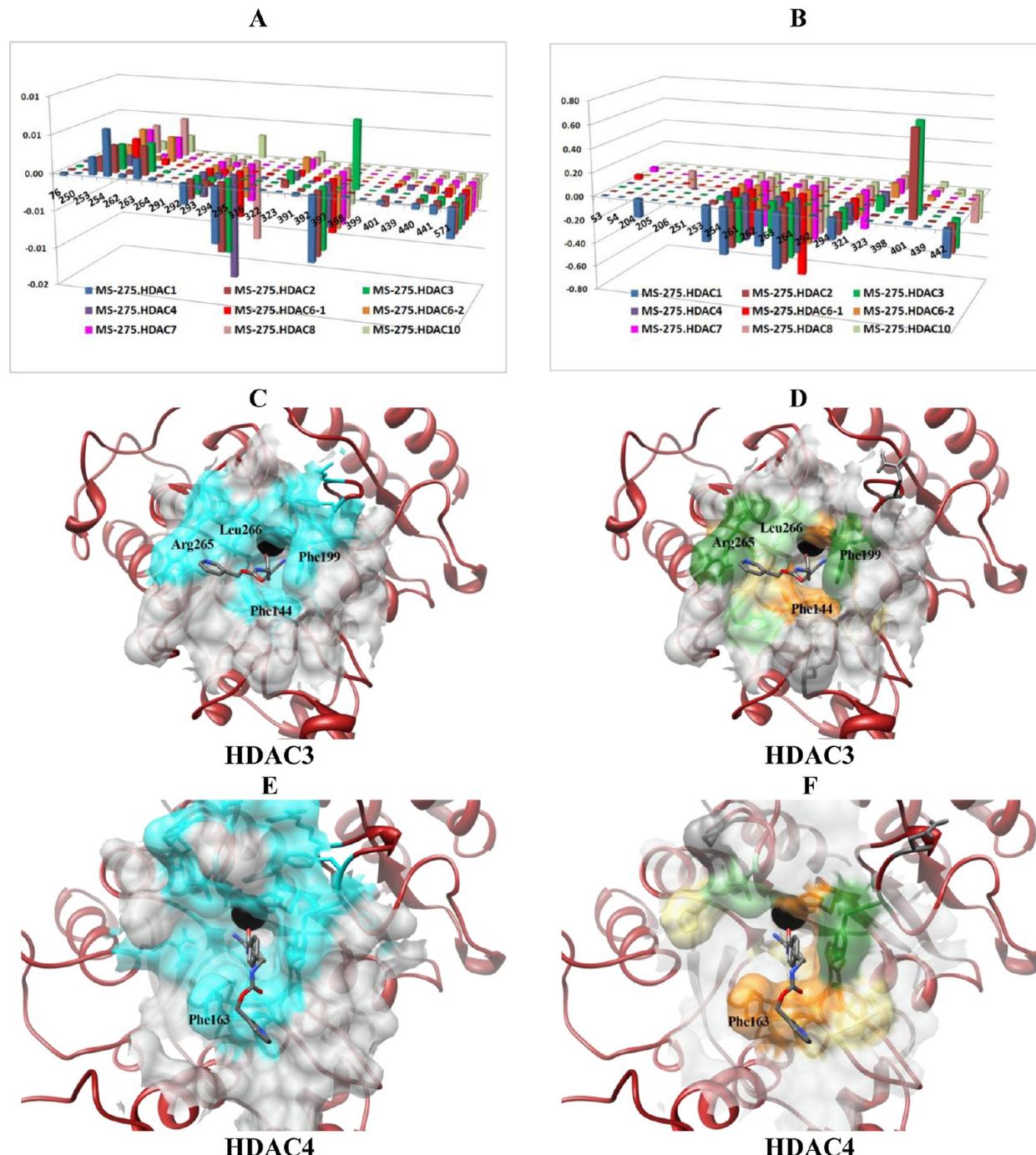
Figure 8. Per-residue activity-contribution plots for the ELE (A) and DRY (B) fields.

and larger variability; maximum-activity contributions occur with APHA8 and TSA binding to either class I or class II enzymes (Supporting Information File 1 Figure SI-8). The other highly positively contributing residue 205 is peculiar for HDAC-6-1 (Tyr85) and thus uniquely modulates inhibition of this enzyme.

**Analysis of Interactions Contributing to Isoform Selectivity.** Interaction- and activity-contribution analyses suggest that useful insight into structural determinants exists for both HDAC isoforms and their inhibitors to help optimize isoform-specific inhibitors using the derived COMBINEr model. Derivation of rules to guide the structural basis for isoform selectivity required single analysis for each specific isoform model. For nine of the inhibitors used in the training set (Table 2), at least 9 out of 12 isoform-inhibition profiles were available (Supporting Information File 1 Table SI-2). In Supporting Information File 3 are reported the recalculated activity profiles for each of the nine inhibitors of Table 2 showing the models sensitivity to HDAC-isoform inhibition by different compounds. To illustrate the COMBINEr model's potential use, two inhibitors were selected seeking potential structure determinants for isoform selectivity. Among the training set, analysis on the activity range indicated MS-275 and SCRIPTAID as good examples. From Supporting Information File 1 Table SI-2, MS-275 and SCRIPTAID display large variability, and from Table 2, MS-275 results partially selective for class I HDACs (particularly

for HDAC-3  $IC_{50} = 0.07 \mu\text{M}$  and HDAC-2  $IC_{50} = 0.5 \mu\text{M}$ ), while SCRIPTAID is partially selective for class II displaying submicromolar activities against these enzymes.

**MS-275.** This inhibitor is specifically selective for class I HDAC-3 over class IIa HDAC-4 and comparison of data belonging to the relative complexes shows how the model helps rationalize the higher activity of MS-275 for HDAC-3 versus HDAC-4. As shown in Figure 9, it is possible to indicate, either numerically or graphically, the residues responsible for this activity difference. Considering electrostatic interactions, it is evident that, as already above highlighted, there is very low correlation with activity, and only gray or light blue surfaces can be observed in Figures 10C and E (see Figure 9 footnotes for color-coding). On the other hand, the DRY field seems very sensitive as shown in Figure 9D and F; there is a high color variation clearly indicating those residues responsible for the higher activity of MS-275 against HDAC-3 (Phe199 and Arg265 are dark green). Other green-colored residues are also located around the rim, for example, Leu266. A few residues are colored yellow, residue 263 (Phe144 in Figure 9D) indicating that MS-275 anti-HDAC-3 activity could be improved by optimizing the interactions in the enzyme channel. Going to the MS-275/HDAC-4 complex, many DRY surfaces have turned from green to yellow thus highlighting that residue 263 (HDAC4-Phe163)

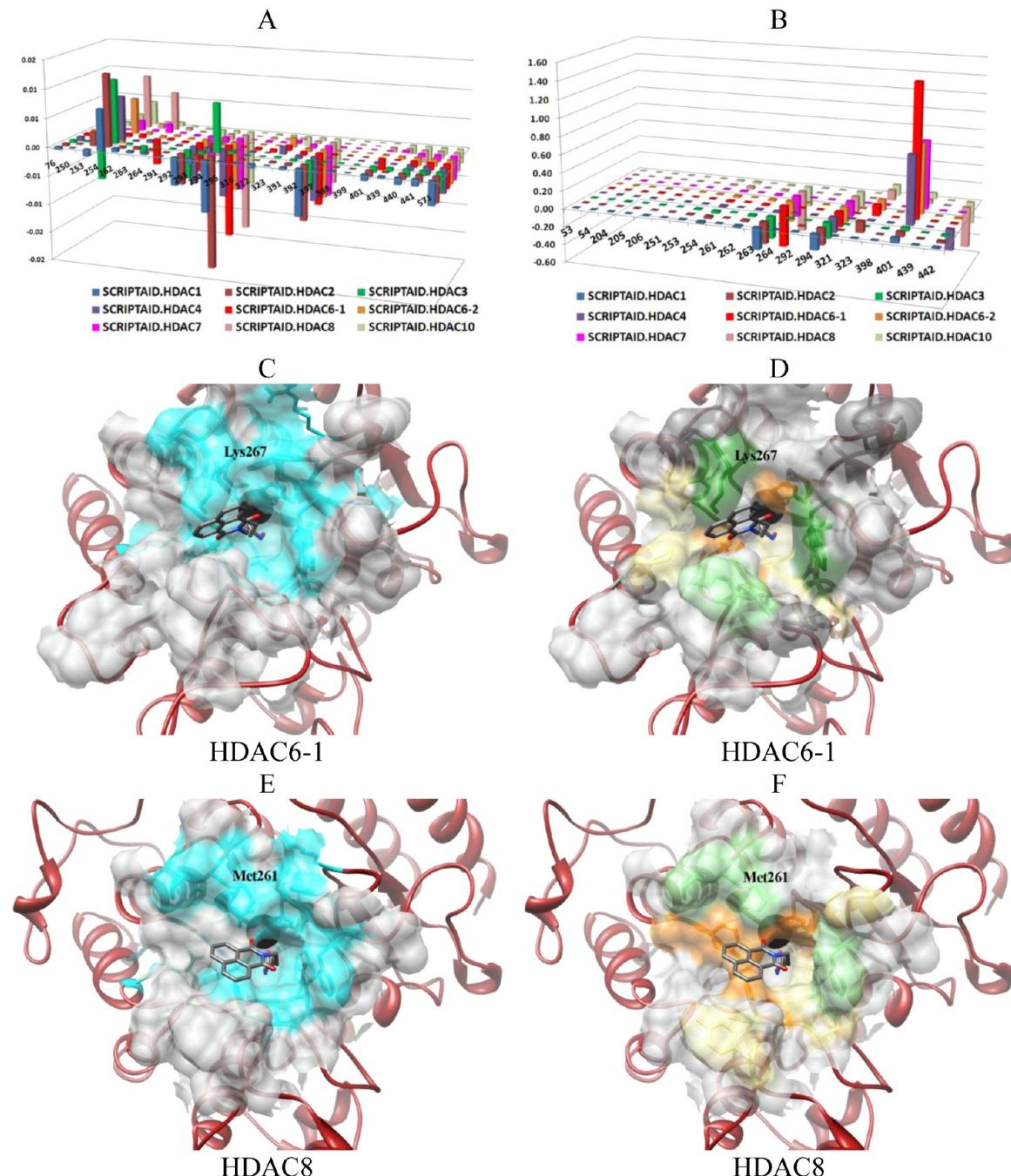


**Figure 9.** Activity contributions (A and B) for MS-275 and their graphical representations (C–F). A, C, and E in the left side, account for the ELE field. The DRY is depicted on the right (B, D, and F). Residue surfaces are color-coded: for ELE blue-based surfaces indicate a positive contribution (light blue if the contribution is less than 50% of maximum contribution for a given residue; dark blue indicate areas with higher contributions); red-based surfaces indicate negative contributions (light red for absolute contribution less than 50% of the corresponding residue; dark red for higher percentage of negative contribution). For the DRY field, positive contributions are indicated in green (dark green for contribution higher than 50% of the maximum activity contribution; light green for less contribution); yellow colors are used to indicate negative DRY contributions (dark yellow for absolute contribution higher than 50% of the maximum activity contribution; light yellow for low negative contributions). Dark gray surfaces indicate zero contribution, while light gray are residues with PLS coefficients lower than 0.001 (see above in the model interpretation section). Only residues cited in the text are labeled.

plays a major role in decreasing activity with many residues showing zero activity contribution.

**SCRIPTAID.** SCRIPTAID was chosen as a selective class II inhibitor. Similarly to MS-275, the electrostatic interactions differentiated when comparing the activity contributions of HDAC-6 and HDAC-8 (Figure 10). Indeed, Figure 10A clearly indicates that the ELE contributions are below 0.02. So

analogously to MS-275, DRY terms help rationalize the inhibitory activities of SCRIPTAID with HDAC-6 and HDAC-8. Most differences are located in the rim zone. Specifically, Lys267 in HDAC-6 is responsible of a strong positive contribution, while Met261, its counterpart in HDAC-8, displays a much smaller contribution.



**Figure 10.** Activity contributions (A and B) for SCRIPTAID and their graphical representations (C–F). A, C, and E account for the ELE field. The DRY field is depicted on the right (B, D, and F). Residue surfaces are color-coded: for ELE, blue-based surfaces indicate positive contributions (light blue if the contribution is less than 50% of maximum contribution for a given residue; dark blue indicate areas with higher contributions); red-based surfaces indicate negative contributions (light red for absolute contributions less than 50% of the corresponding residue; dark red for higher percentage of negative contributions). For the DRY field, positive contributions are indicated in green (dark green for contribution higher than 50% of the maximum activity contribution; light green for less contribution); yellow colors are used to indicate negative DRY contribution (dark yellow for absolute contribution higher than 50% of the maximum activity absolute contribution; light yellow for low negative contributions). Dark gray surfaces indicate zero contributions, while light gray are residue with PLS coefficients lower than 0.001 (see above in the model interpretation section). Activity contribution plots and associated graphicals for all the training set are reported in Supporting Information Files 4 and 5. Only residues cited in the text are labeled.

**Docking Assessment.** X-ray structures of HDAC-inhibitors were used to evaluate the ability of a docking program to predict the correct geometry of protein–ligand complex (redocking). To this aim, two different docking programs were tested: AutoDock Ver. 4.2 and AutoDockVina Ver. 1.1. Docking results were assessed with RMSD (root-mean-square deviation) of the predicted ligand configuration versus the crystal structure. Table

9 and 10 show RMSD values for best docked (the lowest energy docked conformation of the first cluster generated), best cluster (the lowest energy docked conformation of the most populated cluster), and best fit (the lowest energy conformation of the cluster showing the lowest RMSD value),<sup>40</sup> obtained with the two programs. In all cases, AutoDockVina was found to be more accurate displaying a docking accuracy (DA) of 75% for the best

**Table 9.** Redocking Results (RMSD) with AutoDock Program

| complex name | best docked | best cluster | best fit |
|--------------|-------------|--------------|----------|
| LLX.HDAC2    | 0.48        | 0.48         | 0.48     |
| HA3.HDAC4    | 5.25        | 4.76         | 4.4      |
| TMFK.HDAC4   | 3.46        | 5.75         | 3.46     |
| SAHA.HDAC7   | 10.36       | 10.36        | 2.18     |
| TSA.HDAC7    | 6.06        | 6.06         | 1.4      |
| APHA.HDAC8   | 5.4         | 2.26         | 2.26     |
| SAHA.HDAC8   | 5.84        | 7.29         | 4.1      |
| TSA.HDAC8    | 5.1         | 5.52         | 1.45     |
| DA %         | 12.5        | 18.75        | 50       |

**Table 10.** Redocking Results (RMSD) with AutoDockVina Program

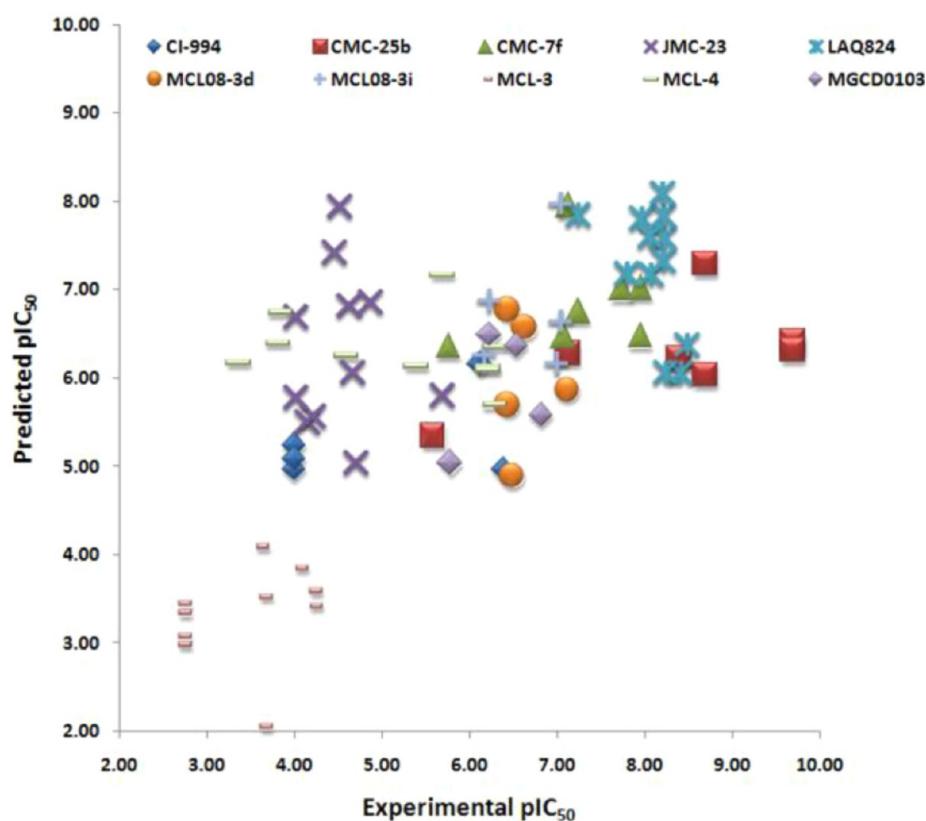
| complex name | best docked | best cluster | best fit |
|--------------|-------------|--------------|----------|
| LLX.HDAC2    | 0.24        | 0.24         | 0.24     |
| HA3.HDAC4    | 3.87        | 2.34         | 1.93     |
| TMFK.HDAC4   | 4.02        | 1.9          | 1.46     |
| SAHA.HDAC7   | 2.45        | 2.45         | 1.88     |
| TSA.HDAC7    | 2.19        | 2.19         | 1.21     |
| APHA.HDAC8   | 1.43        | 1.43         | 1.43     |
| SAHA.HDAC8   | 2.49        | 2.49         | 1.72     |
| TSA.HDAC8    | 2.09        | 1.22         | 1.22     |
| DA %         | 50          | 75           | 100      |

cluster poses (Tables 9 and 10). AutoDockVina was able to predict the right binding disposition of all ligands with a RMSD <3 Å. From Tables 9 and 10, the best cluster conformation displayed the lowest RMSD values. For subsequent dockings, therefore, only the AutoDockVina program was used considering

the best cluster conformation as the first choice. Considering the Best Fit pose, AutoDockVina proved to be able to find the correct binding mode with a DA of 100%. Although the Best Fit poses are irrelevant for the docking applicability, they further supported that AutoDockVina is quite good in searching for the right conformation, but the scoring function is not able to select it. For docking, the side-chain flexibility features of AutoDock and AutoDockVina were not used as the results were always worse than in fixed receptor dockings in preliminary docking studies.

**Model Predictivity.** Once the docking protocols were assessed, a cross-docking approach was applied to the MTS, CTS, and LTS test sets of inhibitors to prepare the HDAC- $\alpha$  complexes.

**Modeled Test Set.** Regarding the MTS, all minimized HDAC structures were used as templates for docking simulations. Thus, each inhibitor of Table 4 was docked into all receptor binding sites, for a total of 304 individual docking simulations. For each isoform, all poses were collected in a bin and the output poses clustered by means of the AutoDock program. It was found that AutoDockVina had the ability to reproduce the experimental binding modes with modest errors (Table 10); in some cases (not shown), the best cluster conformation was found in a nonactive pose (i.e., the zinc-binding group rotated away from the Zn ion). This clearly indicated the limitations of the docking protocol in selecting the correct poses. In these cases, either the best-docked pose or an arbitrary-chosen conformation on the basis of Zn chelation that mimicking the binding mode of closest-related experimentally bound inhibitor was used. This approach is consistent with the fact that AutoDock Vina proved to be able to find the right binding mode (see comments for the Best Fit pose in the Docking Assessment section). For MTS, a total of 76 HDAC-inhibitors complexes were compiled, and the ELE+DRY

**Figure 11.** Experimental/predicted pIC<sub>50</sub> for the MTS.

**Table 11.** Experimental/Predicted  $\text{pIC}_{50}$  for the CTS Test Set<sup>a</sup>

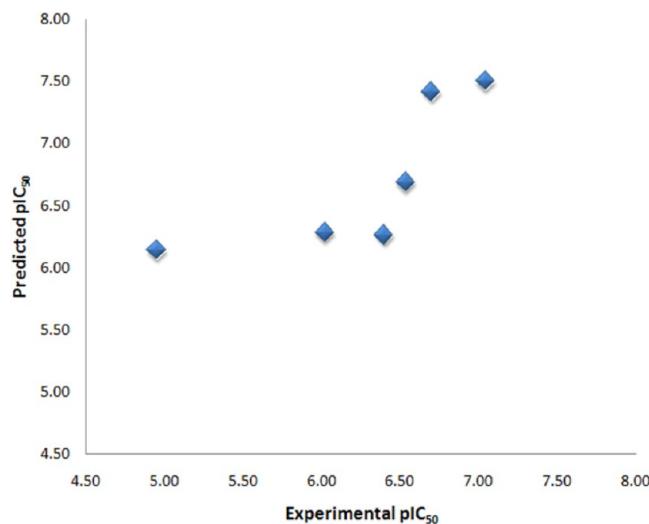
| PDB code           | HDAC  | molecule name | experimental | PC1  | PC2  | PC3  | PC4  | PC5  |
|--------------------|-------|---------------|--------------|------|------|------|------|------|
| 3SFF <sup>32</sup> | HDAC8 | 1DI           | 7.05         | 8.81 | 7.50 | 7.34 | 7.19 | 7.08 |
| 3SFH <sup>32</sup> | HDAC8 | 0DI           | 6.70         | 8.90 | 7.41 | 7.21 | 6.90 | 6.96 |
| 1ZZ3 <sup>35</sup> | HDAH  | 3YP           | 6.54         | 6.46 | 6.69 | 6.39 | 6.48 | 6.34 |
| 2GH6 <sup>34</sup> | HDAH  | CF3           | 4.95         | 6.53 | 6.14 | 5.99 | 6.02 | 6.05 |
| 1ZZ1 <sup>35</sup> | HDAH  | SAHA          | 6.02         | 6.72 | 6.28 | 6.01 | 5.82 | 5.76 |
| 1C3R <sup>33</sup> | HDLP  | TSA           | 6.40         | 6.72 | 6.26 | 6.45 | 6.58 | 6.76 |

<sup>a</sup>PDB codes are used for indicating the structure of inhibitor from Table 5. The predicted values at different principal components (PCs) are reported.

COMBINEr model was used to predict inhibitors activities. Figure 11 and Supporting Information File 1 Table SI-3 show the  $\text{pIC}_{50}$  predicted for the MTS external test set and statistical results ( $\text{SDEP}_{\text{ext}}$  and AAEP). The model showed a good external predictivity with SDEP of 1.41 for the optimal two principal components. Figure 11 reveals that JMC-23 and MCL-4 are the worst predicted compounds. JMC-23 contains an oxime amide as a ZBG (Zn binding group) that can be interpreted as a modified version of the efficient hydroxamate moiety. As reported by Botta et al.,<sup>27</sup> this compound is a poor pan-HDAC inhibitor; the COMBINEr model fails in predict correctly 5 out of 11 activities. Regarding MCL-4, this is the hydroxamate version of MCL-3, while the latter is recognized as a very poor inhibitor with the correct trend, MCL-4 is highly over predicted in HDAC-4, HDAC-5, HDAC-7, and HDAC-9 complexes. Nevertheless the average  $\text{pIC}_{50}$  value for MCL-4 (exp = 5.18, pred = 6.31) was correctly calculated to be higher than that for MCL-3 (exp = 3.40, pred = 3.33).

Comparisons of predictions for single HDAC isoforms reveal that complexes of HDAC-2 and HDAC-3 were the best predicted with an average absolute error of prediction (AAEP) of 0.53 and 0.65, respectively. Complexes related with HDAC-7, HDAC-9, HDAC-10, and HDAC-11 showed the highest AAEP values. For HDAC-9, HDAC-10, and HDAC-11, the worst predictions were associated with a lower number of complexes in the training set. In general, the model was able to reproduce the activity of class I HDACs better than class II. Regarding HDAC-10 and HDAC-11, the smaller amount of experimental data in the training set was the probable cause for the failed activity-trend predictions (Supporting Information File 1 Figure SI-10, Panels K and L). Notably the external SDEP value confirmed that the model at 2 PCs was indeed the most predictive as correctly indicated by the cross-validation runs (Supporting Information File 1 Table SI-3). The application of the COMBINEr model to the MTS proved the ability of the model in predicting the relative potency and the correct activity trend of a given series of inhibitors for 10 out of 12 HDAC isoforms (Supporting Information File 1 Table SI-3 and Figure SI-10) even when the binding conformations of the test set inhibitors were obtained from docking. Furthermore the lowest  $\text{SDEP}_{\text{ext}}$  and AAEP values obtained from the MTS analysis fully supported the optimal number of PCs as indicated by cross-validation.

**Crystal Test Set.** The CTS was compiled using only experimental bound inhibitors. The usefulness of this test set was twofold. First, from Table 11, the training-set model-binding conformations were confirmed to be self-consistent with only 2 PCs (Figure 12), the COMBINEr model predicted the correct trend and activity potencies with an AAEP values of only 0.71 (not shown). Second, the inclusion of bacterial HDACs (HDAH and HDLP) indicates that the derived COMBINEr model might be used to predict activities against nonhuman HDACs,

**Figure 12.** Experimental/predicted  $\text{pIC}_{50}$  for the CTS.

potentially useful in the search for antiparasitic, antifungal, and antibacterial therapeutics.

**Largazole Test Set.** Finally the third test set comprised a cyclotetrapeptide-like inhibitor (largazole).<sup>31</sup> In this case the model was tested for its predictive ability against a class of inhibitor (peptide-like) totally different from those included in the training set. To some extent, the COMBINEr model was able to recognize the relative potency of largazole for HDAC-1, HDAC-2, and HDAC-6-1; while for HDAC-3, the predicted  $\text{pIC}_{50}$  was underestimated, indicating that further modeling of this class of inhibitor is needed (Table 12 and Figure 13). As a

**Table 12.** Experimental/Predicted  $\text{pIC}_{50}$  for the LTS Test Set<sup>a</sup>

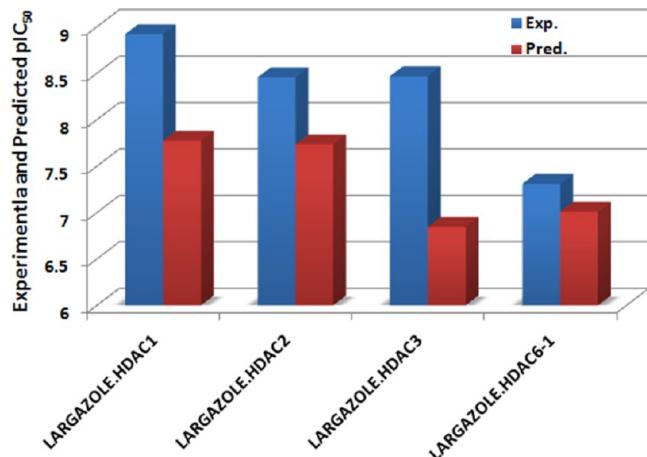
|          | exp  | PC1  | PC2  | PC3  | PC4  | PC5  |
|----------|------|------|------|------|------|------|
| HDAC-1   | 8.92 | 6.98 | 7.64 | 8.03 | 7.88 | 8.09 |
| HDAC-2   | 8.46 | 6.94 | 7.72 | 7.59 | 7.23 | 7.33 |
| HDAC-3   | 8.47 | 6.80 | 6.73 | 6.97 | 6.80 | 6.86 |
| HDAC-6-1 | 7.31 | 7.12 | 6.47 | 6.26 | 5.77 | 6.35 |

<sup>a</sup>The predicted values at different principal components (PC) are reported.

matter of fact, the docking approach used did not allow flexibility of the largazole cyclic headgroup; thus, full ligand flexible docking and smaller error of prediction should be expected with enhanced docking and inclusion of more inhibitors that interact with the headgroup region.

## CONCLUSION

A structure-based 3-D QSAR model using comparative binding-energy analysis that focused on the selectivity of the 11 human



**Figure 13.** LTS predictions at two PCs. The X-axis represents HDACs complexes with largazole, and the y-axis represents biological activity values measured as  $pIC_{50}$ .

zinc-based histone deacetylase isoforms has been developed through a modified protocol called COMBINEr. The derived COMBINEr model shows good statistical coefficients, was predictive for the compounds in the test sets, and robust to cross-validation while omitting multiple data. The model was able to rationalize the different activity profiles of the HDAC inhibitors studied. This model should provide a useful tool for the *a priori* prediction of activity of compounds yet to be synthesized in order to improve their selectivity profiles. The role of dynamic acetylation in epigenetics and other signaling pathways<sup>3</sup> provides strong motivation for the development of molecular scalpels, specific inhibitors of histone deacetylases, to dissect the complexities of epigenetic control of gene expression and other signaling pathways. The COMBINEr model should prove useful in this endeavor.

## ■ EXPERIMENTAL SECTION

All molecular graphics images were produced using UCSF Chimera package ([www.cgl.ucsf.edu/chimera/](http://www.cgl.ucsf.edu/chimera/)) from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, on a 3 Ghz AMD CPU-equipped, IBM-compatible workstation using the Debian 5.0 version of the Linux operating system. For all calculations, a Beowulf cluster of 12 quadcore Xeon CPUs was used.

**Complex Preparation. Inhibitor Structures.** All ligands were generated with Chemaxon Marvin molecular mechanics software (<http://www.chemaxon.com/>) and used without further optimization. The protonation and tautomer states were assigned considering a physiological pH and the more common tautomer according to basic organic chemistry and structural information reported in the corresponding ligand referenced papers.

**HDAC Homology Models.** Those HDAC isoforms whose experimental structures were not available (HDAC-1, -3, -5, -6-1, -6-2, -9, -10, and -11), were built by homology modeling using four automated web servers:

- CPHmodels-3.0 Server<sup>41</sup> (<http://www.cbs.dtu.dk/services/CPHmodels/>),
- M4T Server ver.3.0<sup>42</sup> (<http://manaslu.aecom.yu.edu/M4T/>),
- SwissModel<sup>43</sup> (<http://swissmodel.expasy.org/>),

- ModWeb Server<sup>44</sup> (<http://modbase.compbio.ucsf.edu/ModWeb20-html/modweb.html>).

Using several protein conformations for each HDAC isoform was done to include some target flexibility in the subsequent training set, test set cross-docking runs. For each HDAC isoform, four homology models were generated. All inhibitors were modeled into each of the four homology models, and the resulting complexes energy minimized to supply four complexes for each inhibitor leading to 220 complexes. The servers were used with their default parameters and in a total automatic way to avoid human intervention and to allow maximum reproducibility.

To compile the final training set of 94 complexes (see the Training Set section above), one homology complex per inhibitor was chosen using the preliminary COMBINEr models derived with only crystallized HDAC complexes, (data not shown). For each inhibitor, the HDAC/inhibitor complex whose predicted  $pIC_{50}$ s had the best fit to the experimental  $pIC_{50}$ s for that isoform was selected and utilized in the final training set (Table 13).

**Complex Minimization.** Training set complexes were submitted to a single-point minimization using a protocol described earlier.<sup>40</sup> Briefly, the minimization protocol was applied as follows. (1) ANTECHAMBER with AM1-BCC charges was used to determine missing ligand parameters; (2) the tLeap module was used to solvate the complexes with water molecules in a octahedral box extending 10 Å and to neutralize them with Na<sup>+</sup> and Cl<sup>-</sup> ions; (3) the structures were minimized with the Amber 2003 force field by energy minimization with the SANDER modules: 1000 steps of steepest-descent energy minimization followed by 4000 steps of conjugate-gradient energy minimization, with a nonbonded cutoff of 5 Å. Trials for longer nonbonded cutoff values were done without substantial differences; therefore, 5 Å was chosen for faster calculations. The Zn ion was treated as nonbonded, similarly as in several other applications where HDACs were reported.

**COMBINEr. Ligand/Residues Interactions.** The calculation of the ligand/residue interactions was conducted similarly as previously reported.<sup>7</sup> The AutoGrid module of AutoDock was used with its default setting to compute the interaction energies between each amino-acid residue of the enzymes and an inhibitor. AutoGrid used the united-atom AMBER force field and returned an energy value combining Lennard-Jones (LJ) and hydrogen-bonding (HB) energies between a target and each atom type (probe). The electrostatic interactions were calculated using a distant-dependent Coulombic function, and finally, a third score for hydrophobic interactions was also estimated. In its original use, AutoGrid calculated the interaction energies of a probe atom that was placed on a regularly spaced grid in which a molecular target (the protein) or a portion of it was buried. In this way AutoGrid returns what is called the molecular interaction field (MIF) of a given target, where at each grid point estimates the interactions value for LJ and HB (STE), electrostatic (ELE) and desolvation (DRY), saved in three distinct map files. In the COMBINEr approach, the target was the inhibitor in the complex and the STE, ELE, and DRY interactions were calculated using a grid box centered, at each step, on each atom of the protein (the probe). To the grid is given a step size, so that the whole complex was contained within it, and thus only one value was returned (the center) for each field. The interaction energy for each amino acid of the enzyme was simply obtained by summing all the values for all residue atoms. The

**Table 13. Predicted pIC<sub>50</sub> for the Modeled Complexes Inserted in the Final Training Set**

| HDAC     | complex name         | homology server | pIC <sub>50</sub> exp | pIC <sub>50</sub> pred |
|----------|----------------------|-----------------|-----------------------|------------------------|
| HDAC-1   | APHA8/HDAC1          | SwissModel      | 5.432                 | 6.13                   |
|          | MS-275/HDAC1         | M4T             | 4.886                 | 5.2                    |
|          | SAHA/HDAC1           | M4T             | 7                     | 6.69                   |
|          | SBHA/HDAC1           | CPH             | 5.678                 | 6.61                   |
|          | TSA/HDAC1            | CPH             | 8.301                 | 6.78                   |
|          | OXAMFLATIN/HDAC1     | ModWeb          | 7.301                 | 6.92                   |
|          | NABUT/HDAC1          | ModWeb          | 3.496                 | 3.7                    |
|          | VALPROICACID/HDAC1   | ModWeb          | 3                     | 3.2                    |
|          | SCRIPTAID/HDAC1      | ModWeb          | 6.77                  | 6.2                    |
|          |                      |                 |                       |                        |
| HDAC-3   | APHA8/HDAC3          | CPH             | 6.377                 | 6.8                    |
|          | MS-275/HDAC3         | CPH             | 7.155                 | 6.4                    |
|          | SAHA/HDAC3           | CPH             | 7.699                 | 6.92                   |
|          | SBHA/HDAC3           | SwissModel      | 6.387                 | 6.2                    |
|          | TSA/HDAC3            | SwissModel      | 8.301                 | 6.64                   |
|          | OXAMFLATIN/HDAC3     | SwissModel      | 8                     | 6.43                   |
|          | NABUT/HDAC3          | SwissModel      | 4.648                 | 4.34                   |
|          | VALPROICACID/HDAC3   | CPH             | 3.646                 | 3.2                    |
|          | SCRIPTAID/HDAC3      | SwissModel      | 7.523                 | 6.17                   |
|          |                      |                 |                       |                        |
| HDAC-5   | SAHA/HDAC5           | CPH             | 6.423                 | 6.6                    |
|          | TSA/HDAC5            | CPH             | 7.796                 | 6.97                   |
|          | NABUT/HDAC5          | ModWeb.2        | 2.699                 | 4.19                   |
|          | VALPROICACID/HDAC5   | ModWeb.3        | 2.699                 | 3.43                   |
|          |                      |                 |                       |                        |
| HDAC-6-1 | APHA8/HDAC6-1        | SwissModel      | 7                     | 6.65                   |
|          | MS-275/HDAC6-1       | ModWeb.1        | 4.678                 | 5.32                   |
|          | SAHA/HDAC6-1         | SwissModel      | 7.699                 | 6.77                   |
|          | SBHA/HDAC6-1         | CPH             | 7                     | 6.25                   |
|          | TSA/HDAC6-1          | SwissModel      | 8.301                 | 7.62                   |
|          | OXAMFLATIN/HDAC6-1   | SwissModel      | 7.046                 | 7.68                   |
|          | NABUT/HDAC6-1        | M4T             | 3                     | 3.65                   |
|          | VALPROICACID/HDAC6-1 | CPH             | 3                     | 3.13                   |
|          | SCRIPTAID/HDAC6-1    | SwissModel      | 8.398                 | 7.63                   |
|          |                      |                 |                       |                        |
| HDAC-6-2 | APHA8/HDAC6-2        | CPH             | 7                     | 6.44                   |
|          | MS-275/HDAC6-2       | M4T             | 4.678                 | 5.68                   |
|          | SAHA/HDAC6-2         | CPH             | 7.699                 | 6.44                   |
|          | SBHA/HDAC6-2         | ModWeb.1        | 7                     | 6.2                    |
|          | TSA/HDAC6-2          | M4T             | 8.301                 | 7.02                   |
|          | OXAMFLATIN/HDAC6-2   | CPH             | 7.046                 | 7.1                    |
|          | NABUT/HDAC6-2        | CPH             | 3                     | 4.7                    |
|          | VALPROICACID/HDAC6-2 | ModWeb.1        | 3                     | 3.84                   |
|          | SCRIPTAID/HDAC6-2    | M4T             | 8.398                 | 7.13                   |
|          |                      |                 |                       |                        |
| HDAC-9   | SAHA/HDAC9           | ModWeb.1        | 6.5                   | 6.7                    |
|          | TSA/HDAC9            | ModWeb.1        | 7.419                 | 7                      |
|          | NABUT/HDAC9          | ModWeb.1        | 2.699                 | 4.03                   |
|          | VALPROICACID/HDAC9   | CPH             | 2.699                 | 4.05                   |
|          |                      |                 |                       |                        |
| HDAC-10  | APHA8/HDAC10         | SwissModel      | 5.377                 | 6.24                   |
|          | MS-275/HDAC10        | ModWeb.1        | 4.939                 | 5.67                   |
|          | SAHA/HDAC10          | ModWeb.1        | 7                     | 6.96                   |
|          | SBHA/HDAC10          | M4T             | 5.638                 | 6.6                    |
|          | TSA/HDAC10           | CPH             | 8.301                 | 6.21                   |
|          | OXAMFLATIN/HDAC10    | CPH             | 7.301                 | 6.8                    |
|          | NABUT/HDAC10         | CPH             | 3.535                 | 4.1                    |
|          | VALPROICACID/HDAC10  | M4T             | 3                     | 4.25                   |
|          | SCRIPTAID/HDAC10     | ModWeb.2        | 6.77                  | 6.23                   |
|          |                      |                 |                       |                        |
| HDAC-11  | SAHA/HDAC11          | ModWeb.3        | 6.441                 | 6.21                   |
|          | TSA/HDAC11           | ModWeb.1        | 7.824                 | 5.64                   |

calculations were performed in a box with dimensions of 70 Å × 128 Å × 74 Å. This procedure allowed the decomposition of the enzymes/inhibitor interactions energies into three main contributions (fields) as follows: steric, electrostatic, and hydrophobic. The default parameters for Zn in AutoGrid were used and no attempts to include intramolecular terms were done.

**Statistical Analysis.** All statistical calculations were performed with R, a free software environment for statistical computing and graphics. For the final training set, seven different combinations of the fields previously calculated were tried: the single fields (STE, ELE, and DRY) and the multifield ELE+STE, ELE+DRY, STE+DRY, and ELE+STE+DRY.

**PLS.** All calculations were conducted using the PLS and cross-validation features of the PLS package described by Mevik.<sup>45</sup> An in-house R script was compiled to import the data and carry out all calculations.

**BUW.** Furthermore, in the case of multiple probes, a scaling procedure, called Block Unscaled Weights (BUW), was applied as data pretreatment. This procedure enforces the same importance to each interaction type within the model, normalizing the energy distribution of the X-variables as described by Kastenholz et al.<sup>46</sup> BUW coefficients are reported in Supporting Information File 1 Table SI-4.

**Molecular Docking.** *AutoDock Settings.* The AutoDock-Tools package was employed to generate the docking input files and to analyze the docking results. A grid box size of 57 × 44 × 53 with a spacing of 0.375 Å between the grid points was implemented. A total of 100 runs were generated by using the genetic algorithm, while the remaining run parameters were maintained at their default setting. A cluster analysis was carried out using 2 Å as the RMSD tolerance.

*AutoDockVina Settings.* The same AutoDock grid box was used for its calculations. The docking simulations were carried out with an energy range of 10 kcal/mol and exhaustiveness of 100. The output comprised 20 different conformations for every receptor considered. Although Vina does not include any clustering of the output poses, the clustering feature of the AutoDock program was used to inspect the conformation families using a clustering tolerance set at 2 Å.

## ASSOCIATED CONTENT

### S Supporting Information

Figures SI-1–9, Tables SI-1–4, and supporting files. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: rino.ragno@uniroma1.it.

### Notes

The authors declare no competing financial interest.

<sup>§</sup>Visiting Professor from the Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, United States.

## ACKNOWLEDGMENTS

G.R.M. thanks the Facoltà di Farmacia e Medicina at Sapienza University for supporting his visiting professorship during the fall of 2011 as well as their outstanding hospitality. This study was supported by grants from Italian Ministry of University and Research (MIUR Grant 2008ZTN724\_003 and FIRB Grant RBFR10ZJQT).

## ■ REFERENCES

- (1) Mai, A.; Massa, S.; Rotili, D.; Cerbara, I.; Valente, S.; Pezzi, R.; Simeonia, S.; Ragno, R. Histone Deacetylation in Epigenetics: An Attractive Target for Anticancer Therapy. *Med. Res. Rev.* **2005**, *25* (3), 261–309.
- (2) Hu, E.; Dul, E.; Sung, C. M.; Chen, Z.; Kirkpatrick, R.; Zhang, G. F.; Johanson, K.; Liu, R.; Lago, A.; Hofmann, G.; Macarron, R.; de los Frailes, M.; Perez, P.; Krawiec, J.; Winkler, J.; Jaye, M. Identification of novel isoform-selective inhibitors within class I histone deacetylases. *J. Pharmacol. Exp. Ther.* **2003**, *307*, 720–728.
- (3) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **2009**, *325* (5942), 834–40.
- (4) Zain, J.; Kaminetzky, D.; O'Connor, O. A. Emerging role of epigenetic therapies in cutaneous T-cell lymphomas. *Expert. Rev. Hematol.* **2010**, *3* (2), 187–203.
- (5) (a) Savarino, A.; Mai, A.; Norelli, S.; El Daker, S.; Valente, S.; Rotili, D.; Altucci, L.; Palamara, A. T.; Garaci, E. "Shock and kill" effects of class I-selective histone deacetylase inhibitors in combination with the glutathione synthesis inhibitor buthionine sulfoximine in cell line models for HIV-1 quiescence. *Retrovirology* **2009**, *6*, 52. (b) Choudhary, S. K.; Margolis, D. M. Curing HIV: Pharmacologic Approaches to Target HIV-1 Latency. *Ann. Rev. Pharmacol. Toxicol.* **2011**, *51* (1), 397–418. (c) Matalon, S.; Rasmussen, T. A.; Dinarello, C. A. Histone deacetylase inhibitors for purging HIV-1 from the latent reservoir. *Mol. Med.* **2011**, *17* (5–6), 466–72.
- (6) (a) Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C. Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors. *J. Med. Chem.* **1997**, *40* (7), 1136–1148. (b) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691. (c) Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. Comparative Binding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation as a Powerful Tool in Receptor-Based Drug Design. *J. Med. Chem.* **1998**, *41* (6), 836–852. (d) Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F. 3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14* (4), 341–353.
- (7) Ballante, F.; Musmuca, I.; Marshall, G. R.; Ragno, R. Comprehensive Models of Wild-Type and Mutant HIV-1 Reverse Transcriptases. *J. Comput.-Aided Mol. Design* **2012**, in press.
- (8) Team, R. D. C. *The R Foundation for Statistical Computing*. <http://www.r-project.org/>.
- (9) Gil-Redondo, R.; Klett, J.; Gago, F.; Morreale, A. gCOMBINE: A graphical user interface to perform structure-based comparative binding energy (COMBINE) analysis on a set of ligand-receptor complexes. *Proteins* **2010**, *78* (1), 162–72.
- (10) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112* (3), S35–42.
- (11) Bressi, J. C.; Jennings, A. J.; Skene, R.; Wu, Y.; Melkus, R.; De Jong, R.; O'Connell, S.; Grimshaw, C. E.; Navre, M.; Gangloff, A. R. Exploration of the HDAC2 foot pocket: Synthesis and SAR of substituted N-(2-aminophenyl)benzamides. *Bioorg. Med. Chem. Lett.* **2010**, *20* (10), 3142–3145.
- (12) Dowling, D. P.; Gantt, S. L.; Gattis, S. G.; Fierke, C. A.; Christianson, D. W. Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. *Biochemistry* **2008**, *47* (51), 13554–63.
- (13) Blackwell, L.; Norris, J.; Suto, C. M.; Janzen, W. P. The use of diversity profiling to characterize chemical modulators of the histone deacetylases. *Life Sci.* **2008**, *82* (21–22), 1050–1058.
- (14) Somoza, J. R.; Skene, R. J.; Katz, B. A.; Mol, C.; Ho, J. D.; Jennings, A. J.; Luong, C.; Arvai, A.; Buggy, J. J.; Chi, E.; Tang, J.; Sang, B. C.; Verner, E.; Wynands, R.; Leahy, E. M.; Dougan, D. R.; Snell, G.; Navre, M.; Knuth, M. W.; Swanson, R. V.; McRee, D. E.; Tari, L. W. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure* **2004**, *12* (7), 1325–34.
- (15) Ortore, G.; Di Colo, F.; Martinelli, A. Docking of hydroxamic acids into HDAC1 and HDAC8: a rationalization of activity trends and selectivities. *J. Chem. Inf. Model.* **2009**, *49* (12).
- (16) Vannini, A.; Volpari, C.; Filocamo, G.; Casavola, E. C.; Brunetti, M.; Renzoni, D.; Chakravarty, P.; Paolini, C.; De Francesco, R.; Gallinari, P.; Steinkuhler, C.; Di Marco, S. Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (42), 15064–9.
- (17) Bottomley, M. J.; Lo Surdo, P.; Di Giovine, P.; Cirillo, A.; Scarpelli, R.; Ferrigno, F.; Jones, P.; Neddermann, P.; De Francesco, R.; Steinkuhler, C.; Gallinari, P.; Carfi, A. Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *J. Biol. Chem.* **2008**, *283* (39), 26694–704.
- (18) Schuetz, A.; Min, J.; Allali-Hassani, A.; Schapira, M.; Shuen, M.; Loppnau, P.; Mazitschek, R.; Kwiatkowski, N. P.; Lewis, T. A.; Maglathlin, R. L.; McLean, T. H.; Bochkarev, A.; Plotnikov, A. N.; Vedadi, M.; Arrowsmith, C. H. Human HDAC7 harbors a class IIa histone deacetylase-specific zinc binding motif and cryptic deacetylase activity. *J. Biol. Chem.* **2008**, *283* (17), 11355–63.
- (19) Fass, D. M.; Shah, R.; Ghosh, B.; Hennig, K.; Norton, S.; Zhao, W. N.; Reis, S. A.; Klein, P. S.; Mazitschek, R.; Maglathlin, R. L.; Lewis, T. A.; Haggarty, S. J. Effect of Inhibiting Histone Deacetylase with Short-Chain Carboxylic Acids and Their Hydroxamic Acid Analogs on Vertebrate Development and Neuronal Chromatin. *ACS Med. Chem. Lett.* **2011**, *2* (1), 39–42.
- (20) Hanessian, S.; Auzzas, L.; Larsson, A.; Zhang, J.; Giannini, G.; Gallo, G.; Ciacci, A.; Cabri, W. Vorinostat-Like Molecules as Structural, Stereochemical, and Pharmacological Tools. *ACS Med. Chem. Lett.* **2010**, *1* (2), 70–74.
- (21) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–88.
- (22) Fiser, A.; Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **2003**, *374*, 461–491.
- (23) Ballante, F.; Ragno, R. 3-D QSARgrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications. *J. Chem. Inf. Model.* **2012**, *S2*, 1674–1685.
- (24) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461.
- (25) Beckers, T.; Burkhardt, C.; Wieland, H.; Gimmlrich, P.; Ciossek, T.; Maier, T.; Sanders, K. Distinct pharmacological properties of second generation HDAC inhibitors with the benzamide or hydroxamate head group. *Int. J. Cancer* **2007**, *121* (5), 1138–1148.
- (26) Zhou, N.; Moradei, O.; Raeppl, S.; Leit, S.; Frechette, S.; Gaudette, F.; Paquin, I.; Bernstein, N.; Bouchain, G.; Vaisburg, A.; Jin, Z.; Gillespie, J.; Wang, J.; Fournel, M.; Yan, P. T.; Trachy-Bourget, M. C.; Kalita, A.; Lu, A.; Rahil, J.; MacLeod, A. R.; Li, Z.; Besterman, J. M.; Delorme, D. Discovery of N-(2-aminophenyl)-4-[(4-pyridin-3-ylpyrimidin-2-ylamino)methyl]benzamide (MGCD0103), an orally active histone deacetylase inhibitor. *J. Med. Chem.* **2008**, *51* (14), 4072–4075.
- (27) Botta, C. B.; Cabri, W.; Cini, E.; De Cesare, L.; Fattorusso, C.; Giannini, G.; Persico, M.; Petrella, A.; Rondinelli, F.; Rodriguez, M.; Russo, A.; Taddei, M. Oxime Amides as a Novel Zinc Binding Group in Histone Deacetylase Inhibitors: Synthesis, Biological Activity, and Computational Evaluation. *J. Med. Chem.* **2011**, *54* (7), 2165–2182.
- (28) Fass, D. M.; Shah, R.; Ghosh, B.; Hennig, K.; Norton, S.; Zhao, W. N.; Reis, S. A.; Klein, P. S.; Mazitschek, R.; Maglathlin, R. L.; Lewis, T. A.; Haggarty, S. J. Effect of Inhibiting Histone Deacetylase with Short-Chain Carboxylic Acids and Their Hydroxamic Acid Analogs on Vertebrate Development and Neuronal Chromatin. *ACS Med. Chem. Lett.* **2010**, *2* (1), 39–42.

- (29) (a) Bottomley, M. J.; Lo Surdo, P.; Di Giovine, P.; Cirillo, A.; Scarpelli, R.; Ferrigno, F.; Jones, P.; Neddermann, P.; De Francesco, R.; Steinkuhler, C.; Gallinari, P.; Carfi, A. Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *J. Biol. Chem.* **2008**, *283* (39), 26694–704. (b) Jones, P.; Bottomley, M. J.; Carfi, A.; Cecchetti, O.; Ferrigno, F.; Lo Surdo, P.; Ontoria, J. M.; Rowley, M.; Scarpelli, R.; Schultz-Fademrecht, C.; Steinkuhler, C. 2-Trifluoroacetylthiophenes, a novel series of potent and selective class II histone deacetylase inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18* (11), 3456–3461.
- (30) (a) Kozikowski, A. P.; Chen, Y.; Gaysin, A. M.; Savoy, D. N.; Billadeau, D. D.; Kim, K. H. Chemistry, biology, and QSAR studies of substituted biaryl hydroxamates and mercaptoacetamides as HDAC inhibitors-nanomolar-potency inhibitors of pancreatic cancer cell growth. *ChemMedChem* **2008**, *3* (3), 487–501. (b) Kozikowski, A. P.; Tapadar, S.; Luchini, D. N.; Kim, K. H.; Billadeau, D. D. Use of the nitrile oxide cycloaddition (NOC) reaction for molecular probe generation: a new class of enzyme selective histone deacetylase inhibitors (HDACIs) showing picomolar activity at HDAC6. *J. Med. Chem.* **2008**, *51* (15), 4370–4373.
- (31) Cole, K. E.; Dowling, D. P.; Boone, M. A.; Phillips, A. J.; Christianson, D. W. Structural basis of the antiproliferative activity of largazole, a depsipeptide inhibitor of the histone deacetylases. *J. Am. Chem. Soc.* **2011**, *133* (32), 12474–12477.
- (32) Whitehead, L.; Dobler, M. R.; Radetich, B.; Zhu, Y.; Atadja, P. W.; Claiborne, T.; Grob, J. E.; McRiner, A.; Pancost, M. R.; Patnaik, A.; Shao, W.; Shultz, M.; Tichkule, R.; Tommasi, R. A.; Vash, B.; Wang, P.; Stams, T. Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors. *Bioorg. Med. Chem.* **2011**, *19* (15), 4626–4634.
- (33) 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* (6749), 188–93.
- (34) Nielsen, T. K.; Hildmann, C.; Riester, D.; Wegener, D.; Schwienhorst, A.; Ficner, R. Complex structure of a bacterial class 2 histone deacetylase homologue with a trifluoromethylketone inhibitor. *Acta Crystallogr. Sect. F* **2007**, *63* (Pt 4), 270–3.
- (35) Nielsen, T. K.; Hildmann, C.; Dickmanns, A.; Schwienhorst, A.; Ficner, R. Crystal structure of a bacterial class 2 histone deacetylase homologue. *J. Mol. Biol.* **2005**, *354* (1), 107–120.
- (36) (a) Ragni, R.; Simeoni, S.; Rotili, D.; Caroli, A.; Botta, G.; Brosch, G.; Massa, S.; Mai, A. Class II-selective histone deacetylase inhibitors. Part 2: alignment-independent GRIND 3-D QSAR, homology and docking studies. *Eur. J. Med. Chem.* **2008**, *43* (3), 621–32. (b) Ragni, R.; Simeoni, S.; Valente, S.; Massa, S.; Mai, A. 3-D QSAR studies on histone deacetylase inhibitors. A GOLPE/GRID approach on different series of compounds. *J. Chem. Inf. Model.* **2006**, *46* (3), 1420–30.
- (37) (a) Mai, A.; Massa, S.; Ragni, 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* (4), 512–24. (b) Mai, A.; Massa, S.; Ragni, 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* (9), 1778–84.
- (38) Henrich, S.; Feierberg, I.; Wang, T.; Blomberg, N.; Wade, R. C. Comparative binding energy analysis for binding affinity and target selectivity prediction. *Proteins* **2010**, *78* (1), 135–153.
- (39) Cramer, R. D.; 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* (18), 5959–5967.
- (40) Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragni, R. Combining 3-D Quantitative Structure-Activity Relationship with Ligand Based and Structure Based Alignment Procedures for in Silico Screening of New Hepatitis C Virus NS5B Polymerase Inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 662–676.
- (41) Nielsen, M.; Lundsgaard, C.; Lund, O.; Petersen, T. N. CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W576–581.
- (42) Fernandez-Fuentes, N.; Madrid-Aliste, C. J.; Rai, B. K.; Fajardo, J. E.; Fiser, A. M4T: a comparative protein structure modeling server. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W363–368.
- (43) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22* (2), 195–201.
- (44) Eswar, N.; John, B.; Mirkovic, N.; Fiser, A.; Ilyin, V. A.; Pieper, U.; Stuart, A. C.; Marti-Renom, M. A.; Madhusudhan, M. S.; Yerkovich, B.; Sali, A. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3375–3380.
- (45) Mevik, B.-H.; Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Software* **2007**, *18* (2), 1–24.
- (46) Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E.; Fox, T. GRID/CPCA: a new computational tool to design selective ligands. *J. Med. Chem.* **2000**, *43* (16), 3033–3044.