

Article

Not peer-reviewed version

Application of Ligand- and Structure-Based Prediction Models for the Design of Alkylhydrazide-Based HDAC3 Inhibitors as Novel Anti-cancer Compounds

[Emre f Bülbül](#) , [Dina Robaa](#) , [Ping Sun](#) , Fereshteh Mahmoudi , Jelena Melesina , [Matthes Zessin](#) , [Mike Schutkowski](#) , [Wolfgang Sippl](#) *

Posted Date: 16 May 2023

doi: 10.20944/preprints202305.1124.v1

Keywords: docking; binding free energy; pharmacophore; KPLS; QSAR; alkylhydrazide; histone deacetylases (HDAC)



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Application of Ligand- and Structure-Based Prediction Models for the Design of Alkylhydrazide-Based HDAC3 Inhibitors as Novel Anti-Cancer Compounds

Emre F. Bülbül,¹ Dina Robaa¹, Ping Sun¹, Fereshteh Mahmoudi¹, Jelena Melesina,¹ Matthes Zessin², Mike Schutkowski² and Wolfgang Sippl^{1,*}

¹ Department of Medicinal Chemistry, Institute of Pharmacy, Martin-Luther University of Halle-Wittenberg, Halle (Saale), Germany, efbulbul@gmail.com (E.F.B); dina.robbaa@pharmayie.uni-halle.de (D.R); sunpingmc@gmail.com (P.S); fereshteh.mahmoudi85@gmail.com (F.M); jenamelesina@gmail.com (JM); wolfgang.sippl@pharmazie.uni-halle.de (W.S)

² Department of Enzymology, Institute of Biotechnology, Martin-Luther University of Halle-Wittenberg, Halle (Saale), Germany, matthes.zessin@gmail.com (M.Z); mike.schutkowski@biochemtech.uni-halle.de (M.S)

* Correspondence: wolfgang.sippl@pharmazie.uni-halle.de

Abstract: Histone deacetylases (HDAC) represent promising epigenetic targets for several diseases including different cancer types. The HDAC inhibitors approved to date are pan-HDAC inhibitors and most show a poor selectivity profile, side effects and especially hydroxamic acid based inhibitors lack good pharmacokinetic profiles. Therefore, the development of isoform-selective non-hydroxamic acid HDAC inhibitors is a highly regarded field in medicinal chemistry. In this study, we analyzed different ligand-based and structure-based drug design techniques to predict the binding mode and inhibitory activity of recently developed alkylhydrazide HDAC inhibitors. Alkylhydrazides have recently attracted more attention as they have shown promising effects in various cancer cell lines. In this work, pharmacophore models and atom-based quantitative structure-activity relationship (QSAR) models were generated and evaluated. The binding mode of the studied compounds was determined using molecular docking as well as molecular dynamics simulations and compared with known crystal structures. Calculated free energies of binding were also considered to generate QSAR models. The models created show a good explanation of in vitro data and were used to develop novel HDAC3 inhibitors.

Keywords: docking; binding free energy; pharmacophore; KPLS; QSAR; alkylhydrazide; histone deacetylases (HDAC)

1. Introduction

Epigenetics refer to reversible alterations on the gene expressions that do not modify the DNA sequence [1]. Post translational modifications such as methylation, acetylation and others make changes of the N-terminal tails of histones [2]. Histone acetylation and deacetylation modifications are controlled by different class of enzymes, histone acetyltransferases (HAT) and histone deacetylases (HDAC) [3,4]. Thus, the chemical modification is reversible [5,6].

To date, 18 human HDACs have been characterized. HDACs are separated into two groups and four classes depending on sequence similarity to yeast HDAC [7]. Zinc dependent HDACs are class I, class II and class IV HDACs, while nicotinamide adenine dinucleotide (NAD⁺) dependent enzymes are class III HDACs which are also known as sirtuins [8–10]. Class I HDACs (HDAC1, 2, 3 and 8) are located in the nucleus [11]. HDAC1 and HDAC2 interact with the nucleosome remodeling and deacetylase complex (NuRD), transcriptional regulatory protein sin3A, corepressor of REST

(CoREST) and mitotic deacetylase complex (MIDAC) [12–16], while HDAC3 form a complex only with the silencing mediator for retinoid and thyroid receptors (SMRT) and nuclear receptor corepressor (NCoR) [17,18]. HDAC8 does not need to form a complex and works alone [19,20].

HDACs are involved in signal transduction, cell growth and cell death [21]. So far, several inhibitors including SAHA, FK228, belinostat and panobinostat have been approved by the FDA against T-cell lymphoma [22–25]. However, due to reported side effects and unfavorable pharmacokinetics, much effort is made to develop novel selective and better bioavailable HDAC inhibitors against several diseases such as cancer, parasitic diseases, inflammation, and others [26–29].

The majority of HDAC inhibitors consists of three pharmacophoric features: a zinc binding group (ZBG) which chelates the zinc ion at the bottom of the catalytic pocket, a linker group which is located at the substrate binding tunnel, and a cap group which is solvent-exposed at the rim of the pocket [7,30]. Some HDACs can be selectively inhibited by compounds addressing available subpockets of HDACs such as side-pockets, lower-pocket and foot-pocket (FP) [31–39]. For example, class I HDACs show an additional foot pocket [40]. Targeting this foot pocket resulted in class I HDAC-selective inhibitors [31,34,35,41–44]. However, the development of selective inhibitors for the class I members of HDACs, particularly HDAC1-3, remains a critical challenge to overcome. The zinc binding group is part of most of the HDAC inhibitors. Till now, hydroxamic acid, 2-aminobenzamide, 2-substituted benzamide, alky/arylketone and thiol groups are often used as warheads for inhibitors of class I HDACs [35,37,38,41–45]. Recently, the alkylhydrazide scaffold attracted attention for the development of HDAC3 inhibitors due to its high potency, selectivity and good bioavailability [46–52]. Their general structure has been represented in Figure 1.

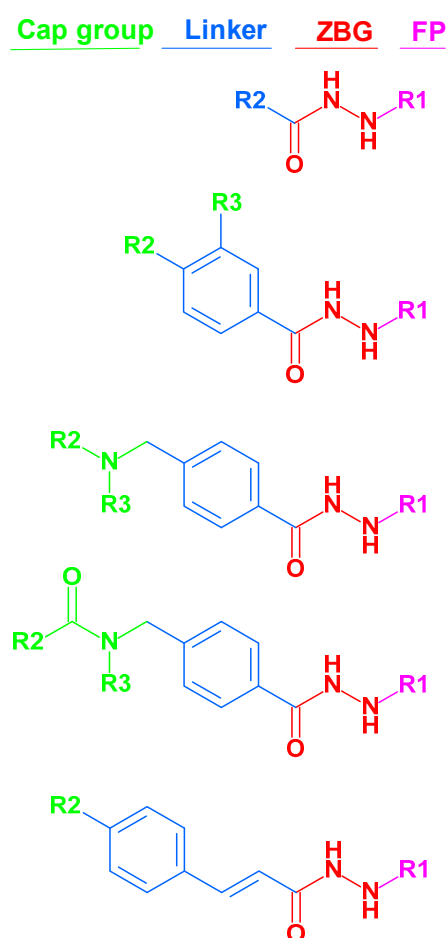


Figure 1. Structures of alkylated hydrazide based HDAC inhibitors that have been used in this article. (The molecular structures of all compounds under study are shown in Table S1, Supplement.) FP: Foot pocket.

Interestingly, by varying the length of the alkyl group (from *n*-propyl to *n*-hexyl) resulted in a shift of the selectivity towards HDAC8 and provided substrate competitive and highly potent inhibitors [53].

In the current study, we performed docking and molecular dynamics studies of alkyldiazides as HDAC3 inhibitors. Available data were compiled to apply ligand-based and structure-based methods to understand the structure-activity relationship of this class of inhibitors. Various quantitative structure-activity relationship (QSAR) methods were evaluated for this purpose, including pharmacophore models, atom-based 3D QSAR models, and binding free energy-based QSAR models. In addition, we tested the models on novel designed alkyldiazides.

2. Results and Discussion

2.1. Diversity analysis of the datasets

A dataset containing 63 compounds having an *N*-monosubstituted hydrazide scaffold were collected from the literature [46,48,50]. The 2D structures and IC₅₀ values of all alkyldiazides are shown in Table S1 (Supplement). The selected compounds cover a reasonable biological activity range (Table 1). All compounds were measured in vitro using recombinant human HDAC3 and the peptidic substrate (Boc-Lys(acetyl)-AMC) [46,48,50].

Table 1. Distribution of inhibitors in training and test set according to their HDAC3 IC₅₀ values.

HDAC3 dataset	Number of compounds	7 < pIC ₅₀ Highly active	5.3 < pIC ₅₀ < 7 Moderately active	pIC ₅₀ < 5.3 Inactive
Training	39	30	9	-
Test	17	11	6	-
Inactive	7	-	-	7
Total	63	41	15	7

We first grouped the compounds into three activity classes according to their HDAC3 data (Table 1, Table S1 Supplement);

- Highly active inhibitors showing pIC₅₀ > 7
- Moderately active inhibitors showing pIC₅₀ between 5.30 and 7
- Inactive inhibitors showing pIC₅₀ < 5.30

The compounds were randomly divided into a training set (70 %; 39 compounds) and test set (30 %; 17 compounds) using the "RAND" function in MOE program (MOE – Molecular database calculator – RAND) [54]. The compounds either having no exact IC₅₀ values or showing an IC₅₀ value higher than 5 μM were classed as inactives (Table 1). The same training and test sets were used for the ligand and structure based model development studies. The training set was used to generate the models, while the independent test set was utilized to evaluate the predictive accuracy of the selected best models. The inactive set was only utilized for the validation of the pharmacophore models by calculating the inactive-survival score (detail in Section 2.2).

The applicability domains of the training and external test sets were analyzed by plotting the three most important principal components (PCA1, PCA2, and PCA3) [55,56] of the calculated descriptors (PEOE_VSA_HYD, GCUT_SLOGP_0, TPSA, b_single, lip_acc, lip_don, vsa_hyd - explained in Table 2). The most important PCA of the molecular descriptors can explain about 100 % of the original space. The PCA analysis showed that the training set and external test set were homogeneously distributed in the chemical space (Figure 2).

Table 2. List of selected molecular descriptors for PCA analysis.

Abbreviations	Molecular descriptors
PEOE_VSA_HYD	The partial equalization of orbital electronegativity (PEOE). Total hydrophobic van der Waals surface area
GCUT_SLOGP_0	The GCUT descriptors using atomic contribution to logP
TPSA	Polar surface area
b_single	Number of single bonds (including implicit hydrogens). Aromatic bonds are not considered to be single bonds.
lip_acc	The number of O and N atoms
lip_don	The number of OH and NH atoms
vsa_hyd	Approximation of the sum of VDW surface areas of hydrophobic atoms.

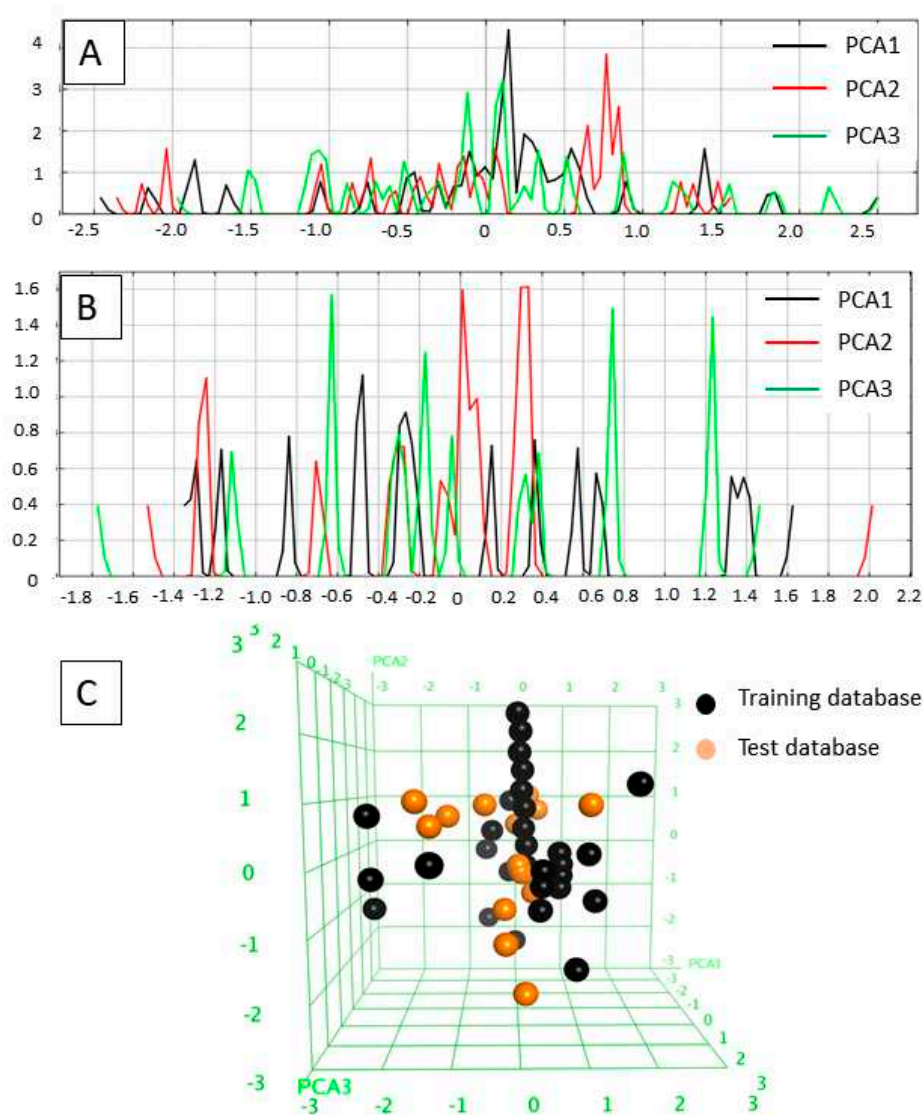


Figure 2. A) Field histogram to visualize the variation of the three most important principal components for the training set. B) Field histogram to visualize the variation of the three most important principal components for the test set. C) 3D plot of the first three PCAs. Training set is colored as black, test set is colored orange.

2.2. Analysis of the pharmacophore model

An important step in establishing a 3D-QSAR model is the development of the correct alignment, usually based on a generated pharmacophore model. In the current work, the pharmacophore model was generated using the Phase module implemented in Schrödinger considering 30 active compounds ($pIC_{50} > 7$) [57]. Then, seven inactive compounds were used to analyze the ability of the generated models for discriminating the actives from inactives.

The pharmacophore model shows the 3D (three dimensional) structural features which might be essential for the biological activity [58,59]. Hence, the pharmacophores that are common to the 30 active compounds showing pIC_{50} more than 7 were obtained. In total, 38 pharmacophore hypothesis were generated. Then, the generated pharmacophore hypotheses were scored according to the survival score. The survival score was generated by evaluating how well the selected pharmacophore hypothesis fits to the most active inhibitors. Additionally, the Phase module penalizes the generated pharmacophore hypothesis that cannot discriminate the actives from inactives. Thus, the developed hypotheses were mapped onto the inactive compounds and scored as inactive-survival score. If the inactive score is better than the survival score, it means that the pharmacophore hypothesis does not discriminate the actives from inactives. For the selected pharmacophore hypothesis, all inactive compounds should show low fitness to the pharmacophores.

After scoring the generated pharmacophore hypothesis, the best-scored pharmacophore model consisting of seven pharmacophoric features (ADDDHRR – Figure 3A) were selected. The survival score (6.923) and the inactive score (1.688) of the hypothesis were shown in Table 3. The survival score is higher than the inactive-survival score meaning that the selected pharmacophore hypothesis can discriminate actives and inactives. Pharmacophore features were specified as hydrogen-bond acceptor (A), hydrogen bond donor (D), hydrophobic (H), negative ionic (N), positive ionic (P) and aromatic ring (R).

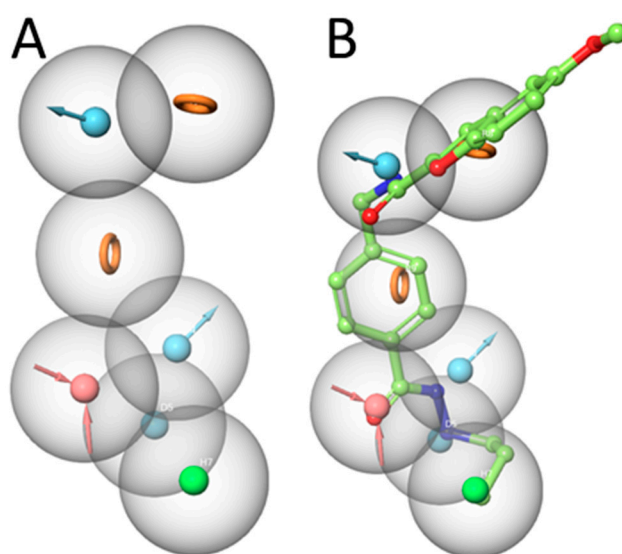


Figure 3. A) Developed pharmacophore model (ADDDHRR) for HDAC3 inhibitors. B) Superposition of compound 1 and the pharmacophore hypothesis (acceptor –red color, donor – cyan color, hydrophobic – green color, ring orange color).

Table 3. Calculated values of the best performing pharmacophore hypothesis.

HYPO ID	Survival Score	Inactive Score
ADDDHRR	6.923	1.688

The generated pharmacophore model mapped onto the most active compound 1. This pharmacophore model shows the importance of the hydrogen bond donor and acceptor functions of the hydrazide moiety (Figure 3B). The carbonyl oxygen of the hydrazide serves as acceptor, while the two nitrogen atoms serve as donor groups. The alkyl chain shows hydrophobic features. The two ring systems are able to exhibit aromatic interactions. The amide scaffold between linker and cap group donates the hydrogen of nitrogen (details shown in the docking part). Accordingly, the selected ADDDHRR pharmacophore model shows the important structural features which can interact with the HDAC3 pocket.

2.3. Analysis of atom-based 3D-QSAR model

The atom-based 3D-QSAR model was built in Schrödinger19 using the 39 compounds in the training set [57–59]. The 3D-QSAR model enable us to consider all relevant structural features such as steric clashes as well as pharmacophores for HDAC3 activity. In this step, the previously selected seven-featured pharmacophore model (ADDDHRR) was used as an alignment rule for the generation of an atom-based QSAR model. First, 39 compounds were aligned by the pharmacophore model and then, the atom based 3D-QSAR models were generated, and these models were cross-validated. The best atom-based 3D-QSAR model was obtained with a good correlation coefficient (R^2 : 0.95), and cross-validated correlation coefficient (Q^2 : 0.88) (Table 4).

Table 4. The best performing ligand-based 3D_QSAR model.

HDAC3 model	N	SD	R^2	RMSE	Q^2
1	39	0.27	0.95	0.39	0.88

*SD (standard deviation of the regression), R^2 (correlation coefficient of the regression), RMSE (root mean square error of test set prediction), Q^2_{LOO} (leave one-out cross validation for the prediction values).

The atom-based 3D-QSAR techniques visualize the compounds as 3D (three-dimensional) based on the non-covalent protein-ligand interactions such as hydrogen bond acceptor and donor, hydrophobic, positive and negative ionic bonds. The model marks the favorable structural features with green cubes showing and unfavorable structural features with red cubes showing. To understand the most favorable and less favorable interactions, we have analyzed the all compounds from the training set. As examples, three compounds having low activity (Compounds 35, 36, 38) and three compounds having good activity (Compounds 1, 2, 3) from the training set were chosen to analyze the atom-based QSAR model (Figure S1, Supplement). According to the atom-based QSAR model, compound 35 exhibited poor activity due to its heptyl alkyl chain (Figure S1A, Supplement). In Figure S1B (Supplement), the meta-substituent on the phenyl linker group as exemplified with compound 36 showed unfavorable effect on the HDAC3 activity. In case of compound 38, the thiophene ring showed unfavorable structural features decreasing the HDAC3 activity (Figure S1C, Supplement). On the other hand, the propyl alkyl chain attached to the hydrazide is favored for three active compounds (Figure S1D–F, Supplement). In addition to that, para-substituted cap groups are also favored and covered by green cubes. According to the model visualization, the least active compounds (Figure S1A–C, Supplement) are mainly covered by red cubes, while the more active compounds especially the cap groups (Figure S1D–F, Supplement) are mostly covered by green cubes.

The external validation was performed using a test set which was not used for model generation, with the aim of evaluating the predictive accuracy and reliability of the generated atom-based QSAR model. The scatter plot of the training and test set was shown in Figure 4. The prediction results of the training, test and inactive databases are shown in Table S2.

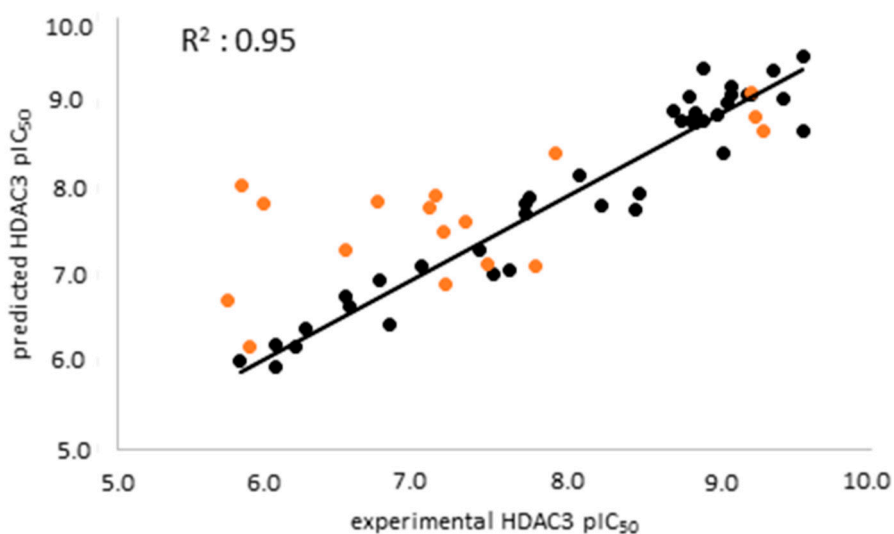


Figure 4. Scatter plot of the best ligand-based QSAR model for training set (black) and test set (orange).

Analysis of the test set revealed that the atom-based QSAR model well predicts the active inhibitors with differences less than 1 log unit. However, several of the moderately active inhibitors (compounds 51, 53, 55, 56) in the external test set as well as the seven inactive compounds were predicted with differences more than 1 log unit (Table 5, Table S2 Supplement). The atom-based QSAR model classified most of the moderately active and inactive inhibitors into the active class. Due to limited accuracy of the ligand-based models in correctly predicting the inactives/weakly actives, we tried to overcome this by generating structure-based prediction models. For this we applied docking and binding free energy calculation techniques discussed in the next section.

Table 5. Prediction results of the test set compounds (atom-based QSAR model).

Compound number	pIC ₅₀ HDAC3	Prediction by atom-based QSAR	Difference (experimental – predicted activity) atom-based	References
40	9.29	8.69	0.60	[46]
41	9.24	8.87	0.37	[46]
42	9.21	9.14	0.07	[46]
43	7.90	8.43	0.53	[48]
44	7.78	7.13	0.64	[48]
45	7.46	7.17	0.29	[48]
46	7.31	7.65	0.34	[48]
47	7.17	6.92	0.24	[50]
48	7.16	7.54	0.38	[50]
49	7.11	7.96	0.85	[48]
50	7.07	7.81	0.74	[50]
51	6.73	7.88	1.16	[50]
52	6.51	7.32	0.81	[50]
53	5.96	7.85	1.89	[50]
54	5.87	6.19	0.33	[50]
55	5.81	8.08	2.26	[50]
56	5.72	6.74	1.01	[50]

2.4. Analyzing the binding mode of alkylhydrazides at HDAC3

We started with docking of all inhibitors to HDAC3 (PDB ID: 4A69 [60]) (Figure S2, Supplement). We used a docking setup in Glide that we previously validated for HDAC inhibitors from different chemical series [31,34,53]. To date, there is no crystal structure of an HDAC in complex with an alkylhydrazide but we could recently show that alkylhydrazides similar to inhibitors 1 and 47 (Table S1, Supplement) [46,50] are reversible and substrate competitive inhibitors of HDACs [53]. Thus, we docked the alkylhydrazides into the catalytic pocket of HDAC3 and analyzed whether they are able to chelate the catalytic zinc ion. The analysis of the docking results of active inhibitors, as exemplified by compounds 1 and 2 from the training set (Figure 5), showed that the hydrazide moiety chelates the zinc ion in a bidentate manner through its nitrogen and carbonyl oxygen and exhibits conserved hydrogen bond interactions with H134, H135 and Y298 at the bottom of the catalytic pocket. The aromatic linker group were placed into the hydrophobic tunnel consisting of F144, H172, F200, L266, where it makes aromatic pi-pi interactions with F144 and F200. The cap group is interacting with residues at the surface by forming hydrogen bond interaction with D93 as well as occupies the hydrophobic pocket comprised of H22, P23 in HDAC3. A structural difference which influences the potency and selectivity on HDAC3 is observed in the foot pocket region. According to the docking results, propyl and butyl chains of the alkylhydrazides fit well into the foot pocket of HDAC3. However, replacing the propyl or butyl chains by pentyl or longer side chains resulted in dramatic decrease in the HDAC3 activity due to the steric hindrance observed in HDAC3. The Y107 residue pushes L133 resulting in a narrower foot pocket region [31,34]. The pentyl and longer alkyl chains in the foot pocket region show steric clashing with M24 and L133 causing a decrease or loss in the HDAC3 activity.

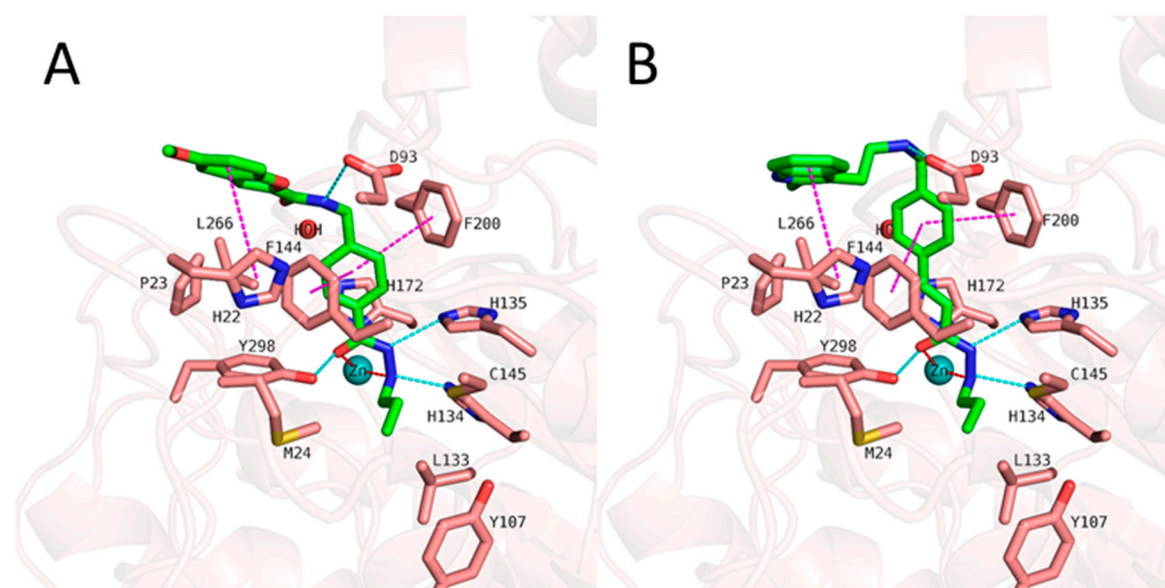


Figure 5. Docking poses of 1 (A, green colored sticks), 2 (B, orange colored sticks) in HDAC3 (PDB ID: 4A69). Hydrogen bonds (cyan dashed lines), hydrophobic interactions (magenta dashed lines) and metal coordination (red dashed lines) between inhibitors and the protein are shown. Relevant residues are shown in stick representation with salmon carbon atoms in HDAC3. The zinc ion is shown as cyan colored sphere. The conserved water molecule is shown as red sphere.

Although the docking poses show reasonable binding in the HDAC3 active site, the correlation between docking scores and pIC_{50} was poor ($R^2 = 0.28$ for HDAC3). Thus, we rescored the docking poses by calculating binding free energies.

In addition to the docking results, we have checked the interaction of the potent inhibitors 1 and 2 using 50 ns MD simulation (Figures 5 and 6, Figures S3 and S4, Supplement). MD simulation analysis of compound 1 and 2 revealed that the *n*-propyl chain attached to the hydrazide fit into the foot pocket consisting of M24 and L133. Notably, M24 and L133 play a key role as a gate keeper in

the foot pocket region of HDAC3. M24 and L133 closed the foot pocket and make the volume narrower which only propyl or butyl side chain favorably fit. This conformational changes of M24 and L133 might explain the decrease or loss of the HDAC3 activity of the compounds having longer alkyl chain than butyl and propyl. The zinc binding group which is the common part of the compounds 1 and 2 preserves its bidentate coordination and makes hydrogen bond interactions with H134, H135 and Y298 throughout the MD simulation. Furthermore, the linker groups of compounds 1 and 2 are sandwiched between F144 and F200. In addition to these similar protein-ligand interactions of compounds 1 and 2, MD simulation analysis exerted some differences in the cap region of compounds 1 and 2.

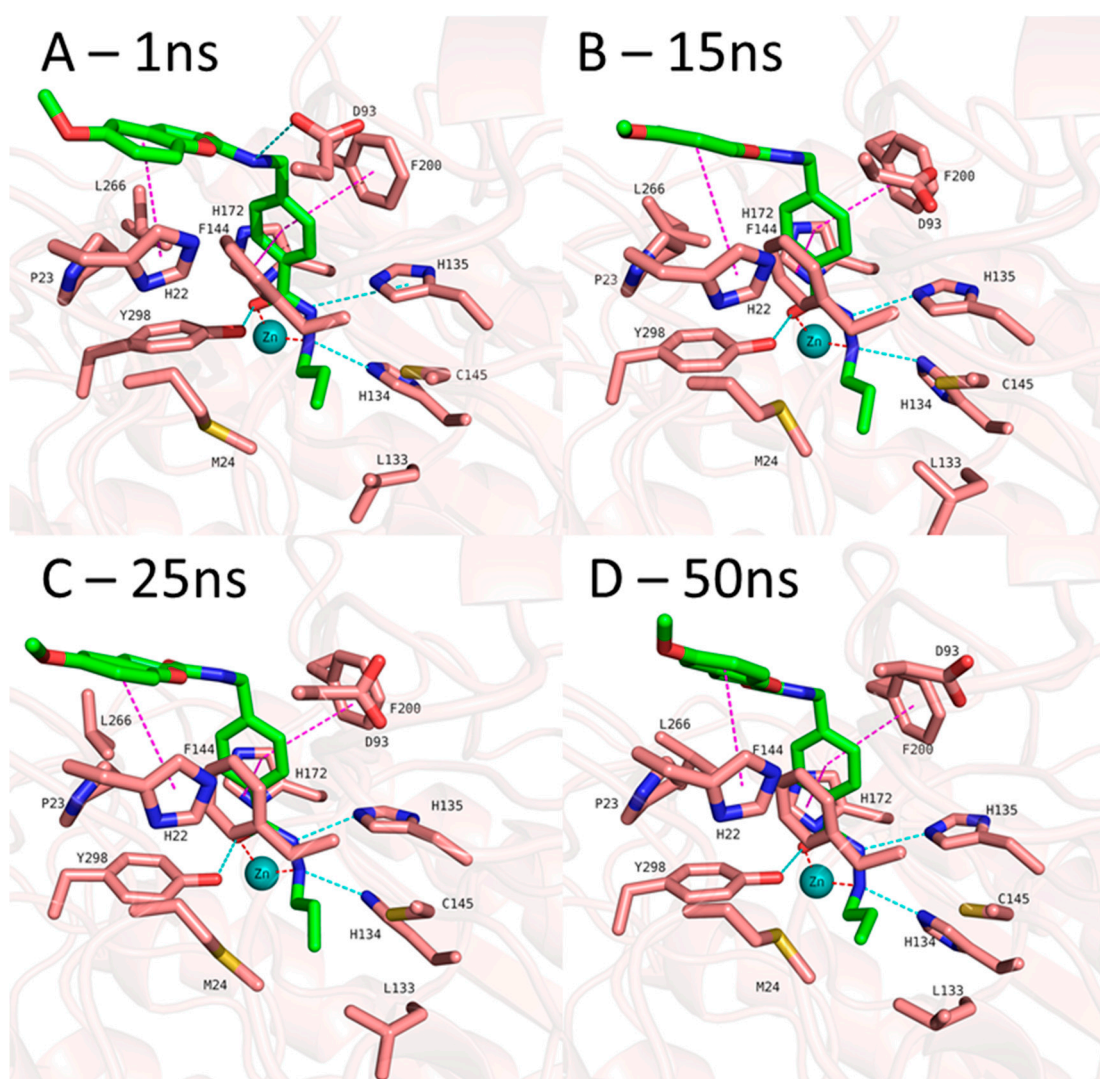


Figure 6. MD frames of HDAC3-1 complex. A) Frame at 1ns MD simulation, B) Frame at 15ns MD simulation, C) Frame at 25ns MD simulation, D) Frame at 50ns MD simulation, Hydrogen bonds (cyan dashed lines), hydrophobic interactions (magenta dashed lines) and metal coordination (red dashed lines) between inhibitors and the protein are shown. Relevant residues are shown in stick representation with salmon carbon atoms in HDAC3. Ligand is shown in stick representation with green carbon atoms. The zinc ion is shown as cyan colored sphere.

According to the MD simulation of compound 1, the selected docking pose as well as the HDAC3-inhibitor complex are stable during the 50 ns MD simulation (Figure 6 and Figure S3, Supplement). In case of the compound 1 (Figure 5), the amide group lost its hydrogen bond interaction with D93, although the ligand maintained the binding conformation. The D93 side chain which is located at the surface of the protein moved away from the compound 1. The benzofuran cap

group of compound 1 is occupying the hydrophobic pocket by making aromatic interaction with H22 at the surface of the protein.

In case of compound 2 (Figure 7), the flexible 2-methyl-1*H*-indole attached to phenyl via (ethylamino)methyl group showed some conformational changes. Hence, the ligand RMSD of compound 2 showed higher fluctuations (Figure S4, Supplement). The 2-methyl-1*H*-indole group changed its conformation and turned down and, then, directed towards D92 and D93 at the surface of the HDAC3. Throughout 50 ns MD simulation, the indole group kept its hydrophobic interactions with H22. The zinc bind group (alkylhydrazide) and the linker group ((prop-1-en-1yl)benzene) of the compound 2 showed a stable binding during the 50 ns MD simulation.

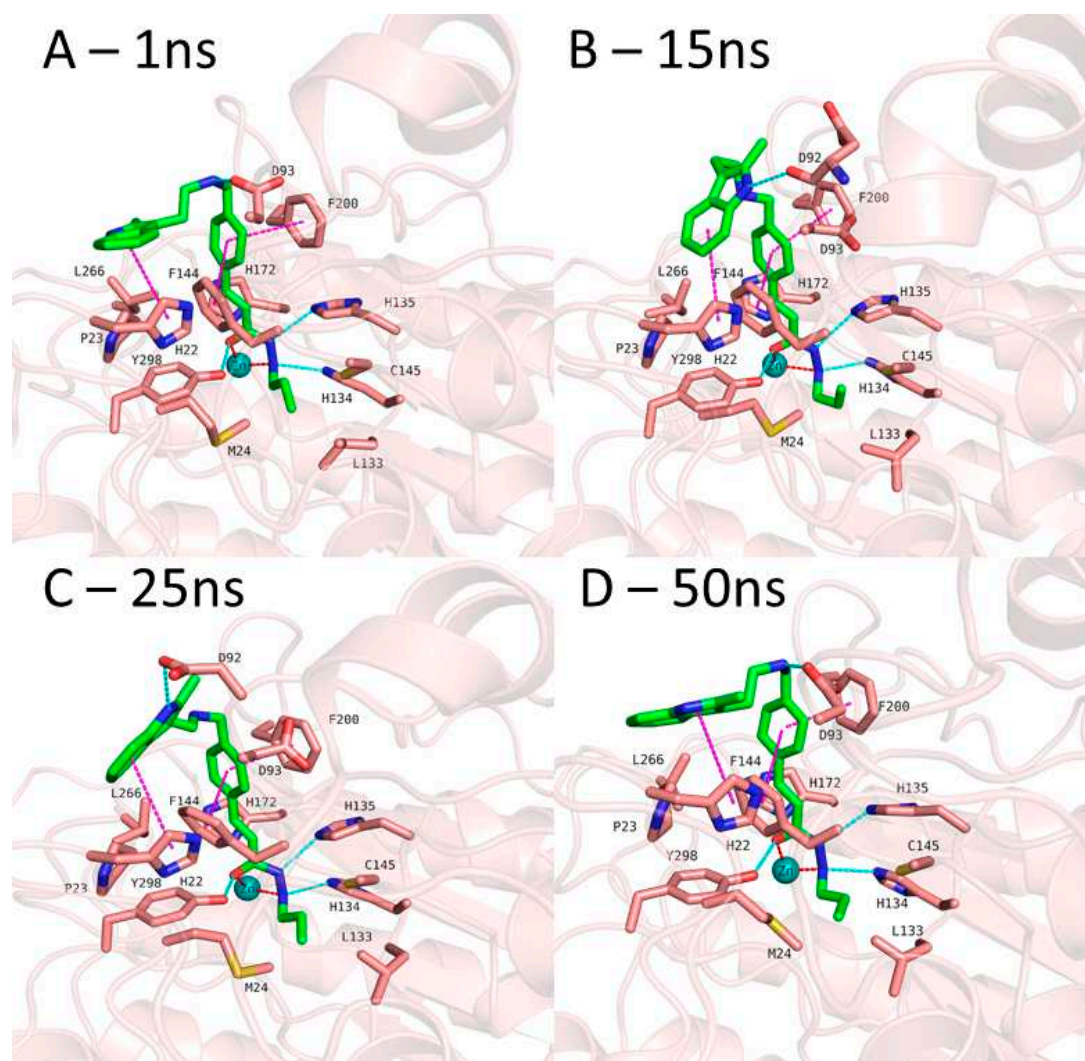


Figure 7. MD frames of HDAC3-2 complex. A) Frame at 1ns MD simulation, B) Frame at 15ns MD simulation, C) Frame at 25ns MD simulation, D) Frame at 50ns MD simulation, Hydrogen bonds (cyan dashed lines), hydrophobic interactions (magenta dashed lines) and metal coordination (red dashed lines) between inhibitors and the protein are shown. Relevant residues are shown in stick representation with salmon carbon atoms in HDAC3. Ligand is shown in stick representation with green carbon atoms. The zinc ion is shown as cyan colored sphere.

2.5. Binding free energy calculation

Due to the low correlation between docking scores and pIC_{50} values, rescoring of the selected docking poses was performed using the MM/GBSA method in AMBER16 [61]. The total energies of HDAC3-inhibitor complexes were calculated using four different parameter settings (solvation models) and six different frame settings (see Methods part for details). The training set including the

39 compounds was used again for model generation. In total, 24 models were generated. The models were assessed based on the correlation coefficients (Q^2 , R^2) between biological data and calculated energy values (Figure 8). The prediction results of the training, test set and inactive compounds are shown in Table S3 (Supplement).

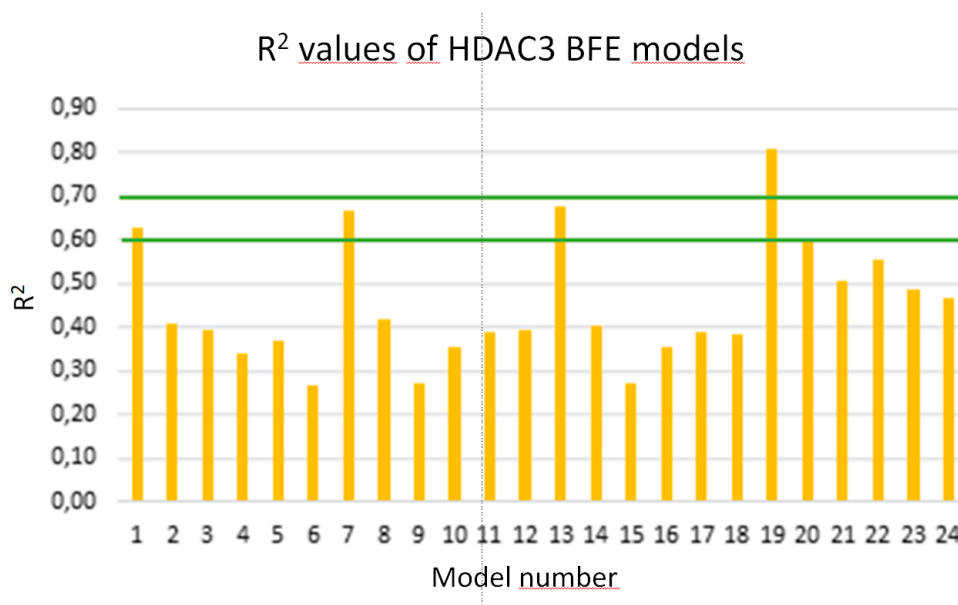


Figure 8. Calculated R^2 values of the established models for alkylhydrazides (training set, orange bars).

Four models showing an R^2 value > 0.6 were considered for further statistical analysis (MODEL1, 7, 13, 19). Interestingly, the selected four models are all based on the protein-ligand complex obtained with one minimization step (Emin1, Table 6).

Table 6. Best performing BFE models.

Model number	N	Method	Frame	2D Descriptor	R^2	RMSE	Q^2_{LOO}	QMSE
MODEL1	39	GB1	Emin1	-	0.63	0.69	0.58	0.73
MODEL7	39	GB2	Emin1	-	0.66	0.65	0.63	0.69
MODEL13	39	GB5	Emin1	-	0.67	0.65	0.64	0.68
MODEL19	39	GB8	Emin1	-	0.81	0.49	0.78	0.52
MODEL19_1	39	GB8	Emin1	PEOE_VSA_HYD	0.87	0.40	0.84	0.44

* R^2 (correlation coefficient), RMSE (root mean square error), Q^2_{LOO} (leave one-out cross validation), QMSE (crossed-root mean square error), Emin1 (single frame after the first energy minimization step).

Among the selected four models, Model 19 based on the GB8 solvation model showed the lowest RMSE and QMSE values with 0.49 and 0.52, respectively (Figure 9 and Table 6). In addition, we tested whether the inclusion a 2D descriptor describing the shape/electronic properties of the inhibitors could improve the models. 2D descriptors were computed for all compounds in MOE [54]. All available 2D descriptors were tried to assess if they improve the model. The total hydrophobic van der Waals surface area (PEOE_VSA_HYD) gave the best improvement. The combination of the 2D descriptor and selected energy term improved R^2 from 0.81 to 0.87 and Q^2 from 0.78 to 0.84. In addition, MODEL19_1 exhibits lower RMSE and QMSE values compared to MODEL19 (Table 6).

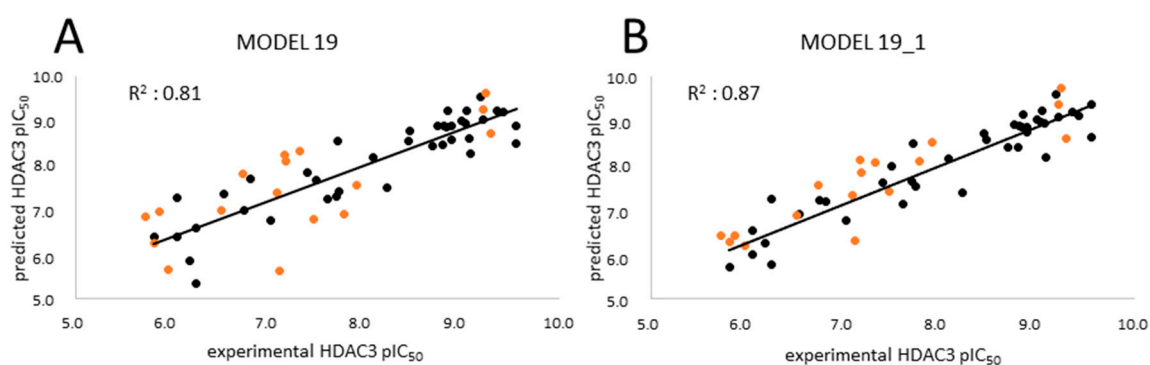


Figure 9. Correlation plots of the A) MODEL19 and B) MODEL19_1 showing correlation between the predicted data and experimental data for HDAC3. Training set (black), external test set (orange).

The test was used to evaluate the accuracy of the best generated model (Model19_1). In the test set (Table 7), all of the compounds were predicted with less than 1 log unit differences. Additionally, the prediction of the inactive compounds was more satisfying compared to the ligand-based QSAR models, with less than 1 log unit except compound 61 (Table S3, Supplement). For compound 61, the docking poses could not explain the incorrect prediction. The scatter plot and prediction results were shown in Figure 7, Table 7 and Table S3 (Supplement), respectively.

Table 7. Prediction results of the test set compounds using the BFE model (MODEL19_1).

Compound number	pIC ₅₀ HDAC3	Prediction of BFE	Difference (experimental – predicted activity)	References
40	9.29	8.60	0.70	[46]
41	9.24	9.75	0.51	[46]
42	9.21	9.37	0.17	[46]
43	7.90	8.53	0.63	[48]
44	7.78	8.10	0.33	[48]
45	7.46	7.42	0.03	[48]
46	7.31	8.06	0.75	[48]
47	7.17	7.85	0.68	[50]
48	7.16	8.12	0.96	[50]
49	7.11	6.30	0.81	[48]
50	7.07	7.32	0.25	[50]
51	6.73	7.56	0.83	[50]
52	6.51	6.89	0.38	[50]
53	5.96	6.19	0.22	[50]
54	5.87	6.41	0.55	[50]
55	5.81	6.27	0.46	[50]
56	5.72	6.42	0.69	[50]

2.6. Evaluation of the generated models on newly designed compounds.

The created models; atom-based 3D QSAR model, Model 19_1, and generated docking poses were then used to predict alkyl hydrazides with other linkers and capping groups that we synthesized (chemistry and in vitro testing has been published elsewhere [53]). The compounds have been designed starting from compound 47 (Table S1). Several structural modifications have been carried out to extend the SAR on this series of HDAC inhibitors. In the first series of compounds, the effect of a different length of the alkyl side chain has been evaluated. In the second series of compounds, different substitutions on position 3 or 4 of the phenyl ring have been analyzed. In the

next series of compounds, the aminopyrimidine linker group with different *N*-arylmethyl, *N*-arylethyl, or *N*-ethylpiperazynl moieties has been evaluated. In the last series of compounds, piperazynl-piperidine linker group has been attached to the phenylalkylhydrazide core of the compounds. The general structures of the new inhibitors are summarized in Figure 10. The experimental HDAC3 IC₅₀ values and the prediction results are shown in Table 8 and Tables S5 and S6 (Supplement, all 2D structures are summarized in Table S4 Supplement).

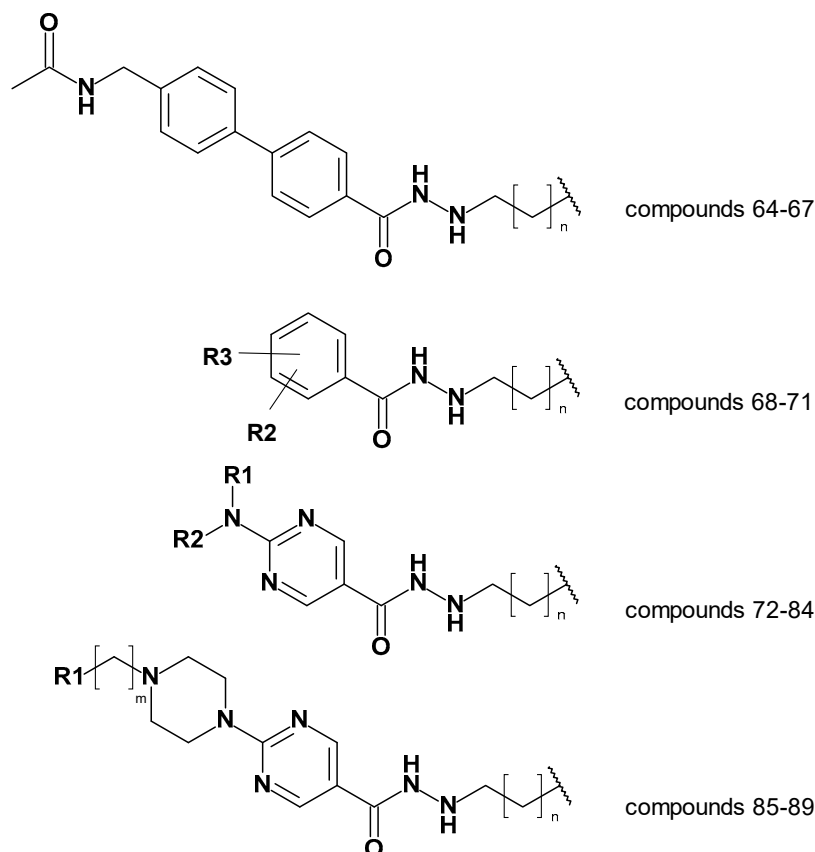


Figure 10. General structures of the tested compounds [53].

First, the ligand based models; pharmacophore models and 3D_atom-based QSAR models, were tested. The developed pharmacophore model was used to align the 26 new compounds to apply the atom-based QSAR model.

Table 8. Experimental and predicted activities by BFE-based and atom-based models.

Compound number	pIC ₅₀ HDAC3	Prediction by atom-based QSAR	Difference (experimental – predicted activity) atom-based	Prediction by BFE model	Difference (experimental – predicted activity) BFE
64	7.04	7.03	0.01	7.16	-0.12
65	6.46	6.59	-0.13	6.89	-0.43
66	5.82	6.57	-0.75	6.36	-0.54
67	<5.00	6.43	-	5.41	-
68	5.80	7.25	-1.46	6.69	-0.89
69	<5.00	7.05	-	5.07	-
70	<5.00	6.98	-	3.92	-
71	9%@1uM	6.82	-	0.24	-

72	7.37	7.85	-0.49	7.60	-0.23
73	6.70	7.29	-0.59	6.68	0.02
74	41%@1uM	7.39	-	7.29	-
75	7.09	7.82	-0.73	8.06	-0.97
76	7.22	7.12	0.10	6.86	0.36
77	7.43	7.96	-0.53	7.06	0.38
78	7.24	7.03	0.20	7.98	-0.75
79	6.92	6.80	0.12	6.32	0.60
80	6.96	8.10	-1.14	7.42	-0.46
81	<5.00	6.54	-	1.65	-
82	<5.00	6.86	-	1.61	-
83	5.52	7.25	-1.73	3.26	2.26
84	<5.00	7.61	-	2.69	-
85	7.52	7.63	-0.10	7.64	-0.11
86	7.00	7.18	-0.18	7.29	-0.29
87	6.52	6.52	0.00	6.86	-0.33
88	6.00	7.40	-1.40	6.13	-0.13
89	5.85	7.13	-1.28	6.11	-0.26

Analyzing the atom-based prediction results of the 26 new compounds revealed interesting results (Table 8). Based on the atom-based QSAR model prediction results, the difference between experimental and predicted pIC₅₀ of 9 compounds, which are moderate or inactive compounds, was more than >1 log unit. This is a weak point of the atom-based QSAR model. Although atom-based QSAR prediction has correctly worked for most of the active and highly compounds, this model showed weakness to discriminate the moderate or inactive compounds from active inhibitors.

Then we evaluated the binding free energy based prediction results (Table 8). Initially, the 26 compounds have been docked to the HDAC3 using the same protocol as for the training set. Similar HDAC3 docking poses as obtained for compound 1 and 2 has been obtained for the new 26 compounds (exemplified in Figure 11). Among the 26 compounds, the common parts; alkylhydrazide moiety and various aromatic linker groups, exhibited similar interactions as observed for compound 1 and 2. A bidentate chelation between zinc ion and hydrazide moiety was observed for all compounds. In addition, the hydrazide moiety showed hydrogen bonds with H134, H135 and Y298 in the zinc binding region of the HDAC3 catalytic pocket. The aromatic linker groups are accommodated into the hydrophobic tunnel and interact with F144 and F200 showing pi-pi interactions. Besides, the cap groups and foot pocket targeting groups showed significant differences which has an impact on the HDAC3 activity.

In the first series exemplified by compound 64 (Figure 11), the acetoamidomethyl cap group was placed at the entrance of the pocket and showed hydrogen bond interaction with D93. The different length of the hydrazide alkyl side chain resulted in a crucial difference on HDAC3 activity. Compound 64 possessing a propyl side chain has been predicted as having better pIC₅₀ than 65, 66, and 67. The difference between experimental and predicted values is less than < 1 log unit in this series. Moreover, compound 67 having a hexyl side chain has been predicted to be inactive. This result confirms that the BFE model is sensible to the side chain effects on this dataset.

In the second series, only compound 68 (Figure 11) bearing the acetamidomethyl cap group and propyl side chain in foot pocket targeting region showed moderate activity on HDAC3. The other compounds 69, 70 and 71 having a hexyl side chain, did not show significant activity on HDAC3. Similar to the experimental activity, the BFE model also predicted the compound 68 as moderate inhibitor, while 69, 70 and 71 have been predicted as inactive compounds.

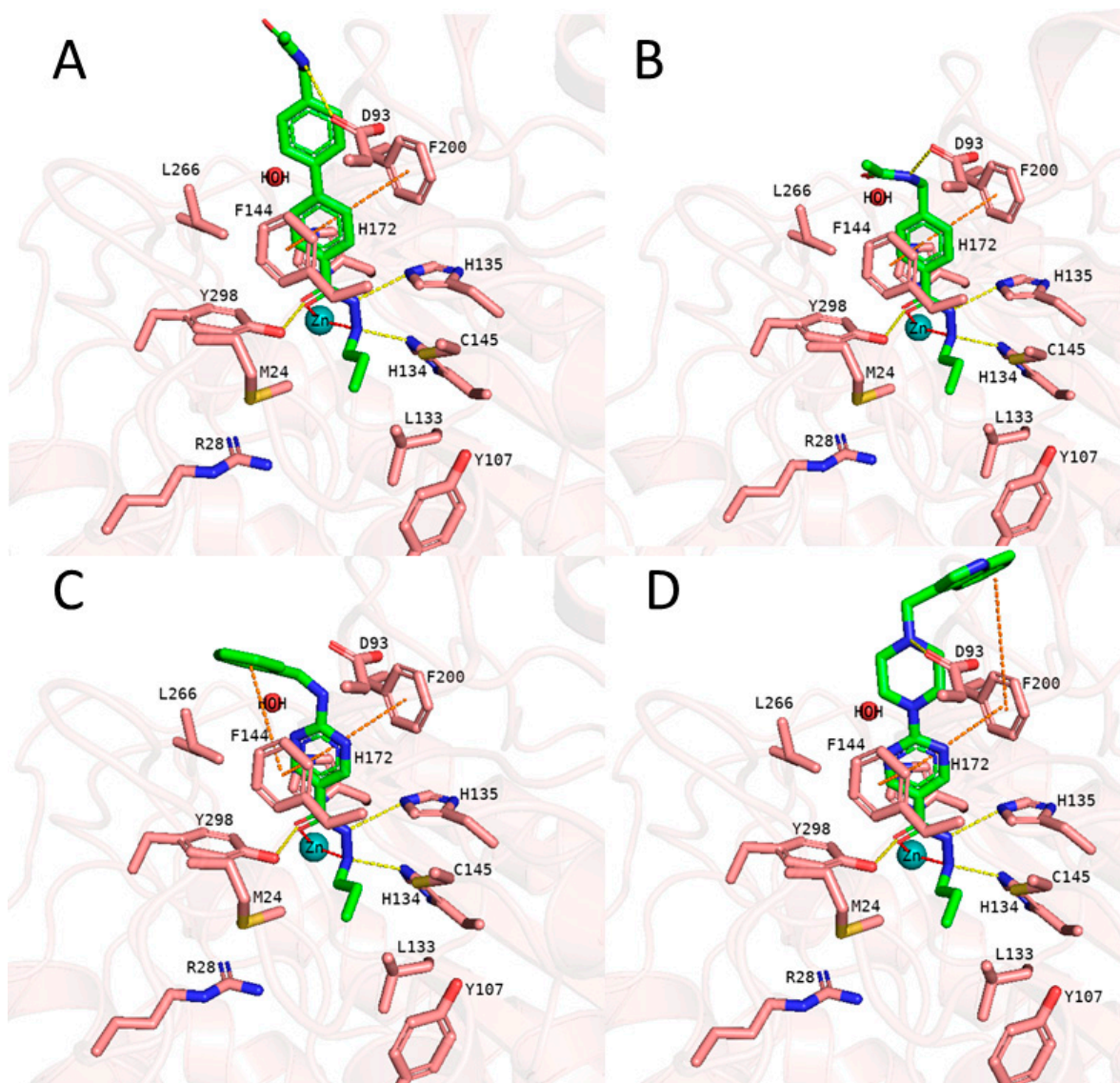


Figure 11. A) Docking pose of compound 64, B) Docking pose of compound 68, C) Docking pose of compound 72, D) Docking pose of compound 85. Hydrogen bonds (yellow dashed lines), hydrophobic interactions (orange dashed lines) and metal coordination (red dashed lines) between inhibitors and the protein are shown. Relevant residues are shown in stick representation with salmon carbon atoms in HDAC3. Ligand is shown in stick representation with green carbon atoms. The zinc ion is shown as cyan colored sphere.

In the third series of compounds bearing an aminopyrimidine group as linker group, different *N*-arylmethyl, *N*-arylethyl, or *N*-methylpiperaziniyl moieties as capping groups have been tested. This series have been exemplified by compound 72 (Figure 11). All compounds bearing a propyl side chain except compound 74 have been predicted with less than <1 log unit compared to experimental activities. Interestingly, compound 74 bearing a *N*-arylethyl cap group did not show significant inhibitory activity (41%@1 μ M) on HDAC3. Flexible cap group could cause for the reduced activity on HDAC3. However, the model predicted compound 74 as active inhibitor. Besides, compounds 81-84 possessing a hexyl side chain have been predicted as inactive inhibitors.

The last series of compounds bearing indole or *N*-methylindole groups attached via methyl or ethyl bridge to the piperazinyipyrimidine scaffold were predicted close to the experimental activities. The difference between experimental and predicted values is less than <1 log unit. The docking poses of this series have been exemplified by compound 85 (Figure 11).

In conclusion we applied the validated BFE models to the further development of the alkylhydrazide based class I HDAC inhibitors. The best inhibitors from this series were also tested for their immunomodulatory effects in Jurkat cells and showed promising cellular effects [53]. As we have recently demonstrated potent T cell memory response by combined class I HDAC inhibition and immune-checkpoint blockade in hepatocellular carcinoma (HCC) therapy the new alkylhydrazides represent an interesting class of inhibitors to explore their potential for cancer therapy.

3. Materials and Methods

3.1. Ligand database preparation

A ligand data set of 63 compounds having hydrazide as zinc binding group (ZBG) was collected from the literature [46,48,50]. Only compounds having an *N*-monosubstituted hydrazide scaffold were considered. The IC₅₀ values of the selected compounds were retrieved from three publications and they all were determined against HDAC3 using the same fluorogenic substrate (Boc-Lys(acyl)-AMC (amino methyl coumarin)). The same human recombinant HDAC3 enzyme was used for the in vitro studies [46,48,50]. The compounds were prepared in ligprep tool using the OPLS3e forcefield in Schrödinger suite [62]. Subsequently, the output of the ligprep step was submitted the Confgen to generate 64 conformers per ligand while minimizing the output conformers using the OPLS3e forcefield [62,63]. The compounds were automatically divided into training (70 %) and external test set (30 %) using the "RAND" function in MOE program (MOE – Molecular database calculator-RAND) [54]. The same training set and external test set were used for the model development studies. The compounds having no exact IC₅₀ values were considered as inactive. The QSAR models were built using the most active and moderate inhibitors for which exact IC₅₀ values were available.

The diversity analysis of the compounds was performed by analyzing the three most important principal components using the principal component analysis (PCA) implemented in MOE [54–56]. The 2D descriptors were computed in MOE [54]. Several 2D descriptors were selected using the Contingency tool in MOE. The three most important principal components (PCA1, PCA2, PCA3) were calculated using the selected 2D descriptors. These principal components were used to check the diversity of the compounds.

The 26 compounds have been collected from the article published by our group to evaluate the established models and check their reliability on different dataset [53]. The ligands have been prepared using the same protocol with validation set.

3.2. Pharmacophore Model

The pharmacophore model was established using 30 inhibitors having IC₅₀ values lower than 100 nM in the training set and the 7 inactive compounds in Phase module of Schrödinger [57]. The compounds were prepared in ligprep using OPLS3e forcefield in the previous step [62,63]. The conformational search was performed in phase module by adjusting 64 conformers per compound and minimizing the output conformers using the "Develop Pharmacophore model" module in Schrödinger [57]. The common pharmacophore hypotheses were developed, scored and ranked. The selected pharmacophore model was used as an alignment rule for atom-based 3D-QSAR model.

3.3. Atom-based 3D-QSAR model

The ligand-based 3D-QSAR model was generated using the training data set in Phase module of Schrödinger [57]. The 39 compounds in the training database were aligned using the selected pharmacophore hypothesis from the previous step. The QSAR models were built with four latent factor and 1.0 Å grid spacing as well as leave-one-out-cross-validation approach. The generated models were evaluated by means of standard deviation of the regression (SD), R² (correlation coefficient of regression), RMSE (root mean square error of test set prediction), Q² (cross-training of test set prediction).

3.4. Docking study

The hydroxamic acid scaffold and hydrazide scaffold are structurally similar groups. Therefore, X-ray crystal structures of HDAC2 (PDB ID: 4LXZ [35]) and HDAC3 (PDB ID: 4A69 [60]) were retrieved from the Protein Data Bank (PDB, [rcsb.org](https://www.rcsb.org) [64]) and analyzed in MOE [54]. SAHA having hydroxamic acid scaffold in complex with HDAC2 protein (PDB ID: 4LXZ) was defined as pan-HDAC inhibitor and shows activity on HDAC3 [35]. First, the HDAC2 (PDB ID: 4LXZ) and HDAC3 (PDB ID: 4A69) were superposed in MOE [54]. Then, SAHA was transferred from the HDAC2 protein (PDB ID: 4LXZ) to HDAC3 to mimic the induced fit effect of the zinc binding group.

The HDAC3-SAHA complex was prepared in the protein preparation wizard of Schrödinger's suite by adding hydrogen bonds and missing side chains and assigning the bond orders [62]. The water molecules (except W2083) and ions (except Zn^{+2} ion) were deleted. The protonation states and tautomers were optimized at pH 7.4 using PROPKA tool. The optimized complex was minimized using the OPLS3e force field to remove the steric clashes [63].

Molecular docking studies were done applying the standard precision (SP) mode in Glide implemented in Schrödinger Suite [62]. The grid box including the information of active site coordinates of the proteins was defined with 10 Å radius around the ligand. Ten docking poses were employed for further post-docking minimization. The other settings were kept as default. The docking results were visually analyzed in MOE program [54].

3.5. Molecular dynamics simulation

The selected docking poses of the compound 1 and 2 in complex with HDAC3 (PDB ID: 4A69) was subjected to a 50 ns MD simulation in AMBER16 [61]. Antechamber package was used to prepare the topologies, force field parameters, atom types and bond types by applying the semi-empirical Austin Model1 with bond charge correction (AM1-BCC) [65,66]. Then, the tLEaP module was employed to prepare the protein-ligand complexes. General amber force field (GAFF), the Duan force field (ff03.r1) and 12-6-4LJ ionic model were used for ligand, protein and zinc, respectively [67–70]. The system was solvated by TIP3P water model, and a margin of 10 Å. Two minimization steps including the two sub-steps in each minimization were carried out. In the first step, 4000 iterations (2000 cycles of steepest descent and then 2000 conjugate gradient) were performed, while the protein residues, ligand and zinc ion were restrained to their initial geometries (force constant of $10 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$) to relieve the bad contacts. In the second step, 4000 iterations (2000 cycles of steepest descent and then 2000 conjugate gradient) were performed to remove the steric clashes in the entire complex. The restraint on the protein, ligand and zinc were removed during the second minimization. Then, the system was heated at 300 K through 100 ps of MD. The protein-ligand complex was restrained to prevent large structural deviations (force constant of $10 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$). The SHAKE algorithm was activated to constrain bonds involving hydrogens [71]. Finally, the system was equilibrated within a period of 200 ps. Langevin dynamics was applied to keep the temperature of 300 K with a collision frequency of 2 ps [72]. The pressure was kept at 1 bar using isotropic position scaling with a relaxation time of 2 ps. Afterwards, a 50 ns MD simulation was run with a time step of 2 fs using the same conditions as in the equilibration step. A non-bonded cut-off distance was 10 Å was used. The electrostatic interactions were calculated by applying the particle mesh Ewald (PME) method. After the MD simulation, CPPTRAJ module of AMBER was used to analyze the MD snapshots.

3.6. Binding free energy calculation

Binding free energies (BFE) of the prepared protein-ligand complexes were calculated using the AMBER16 program [61]. The MMPBSA.py script was utilized for calculations [73]. Different implicit solvent models (GB_{HCT} (igb = 1), GB_{OBC} (igb = 2), GB_{OBC2} (igb = 5), and GB_n (igb = 8) were tested [74–76]. Molecular mechanics (MM) and solvent models were combined for the MMGBSA calculations [77–79]. The results of BFE were analyzed using the following six different methods: 1) single frame at the first minimization step (Emin1), 2) single frame at the second minimization step (Emin2), 3) single frame at the third minimization after MD (Emin3), 4) 1-50 frames during MD (MD-1) with an

interval of 5, 5) 51-100 frames during MD (MD-2) with an interval of 5, 6) 101-500 frames during MD (MD-3) with an interval of 5. The correlation between biological activity and the energy results was measured by using the QSAR tool in MOE program [54].

4. Conclusion

In the current study, we have evaluated several QSAR models including ligand-based and structure-based techniques to understand the structure-activity relationship of alkylhydrazides developed as HDAC3 inhibitors. Additionally, the binding modes of the most potent two HDAC3 inhibitors (compounds 1 and 2) were verified through 50 ns MD simulation, since there is no X-ray structure crystallized with a alkylhydrazide derivative.

The ligand based models enabled us to obtain a general overview of the binding of the compounds in HDAC3 protein. Established pharmacophore model and atom-based 3D-QSAR model can be used to filter the big databases. However, the predictive power of the ligand-based models was not satisfactory. These models predicted moderate inhibitors and inactive compounds as active compounds, although they predict the actives as actives. Thus, these methods should be used to reduce the number of compounds in the big database.

The weakness of the ligand-based methods directed us to generate the structure-based methods. The binding mode of the alkylhydrazide predicted by docking. The selected binding modes from the validation set (compounds 1 and 2) were verified by the short 50 ns MD simulation. The analysis of the MD simulation revealed that the M24 and L133 are gatekeepers in the foot pocket of HDAC3. Additionally, the alkylhydrazides kept their bidentate chelation to zinc ion during the 50 ns MD simulation. The compounds were rescored by means of BFE calculations. The binding free energies were correlated with experimental derived inhibitory activities. The established binding free energy based QSAR model predicted the all compounds in the test set with less than 1 log differences. Additionally, we have tested the established model on a new dataset containing 26 molecules which have been designed, synthesized, tested taking knowledge of the developed BFE models [53]. The structure based model was able to predict these novel compounds with less than 1 log unit error and showed its value for the chemical optimization. For the different test sets, the structure based model showed better accuracy than the ligand based models.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org., Figure S1: Visualization of Atom-based 3D-QSAR model in training set; Figure S2: Docking poses of training set, external test set and inactive set; Figure S3: 50ns MD results for the HDAC3-1 complex; Figure S4: 50ns MD results for the HDAC3-2 complex; Table S1: 2D chemical structures and biological data of compounds in training set, external test set and inactive set; Table S2: Prediction values of the best atom-based QSAR model generated for training set, test set and inactive set in HDAC3; Table S3: Prediction values of the best BFE model generated for training set, test set and inactive set in HDAC3; Table S4: 2D chemical structures and biological data of test set; Table S5: Prediction values of the best atom-based QSAR model generated for test set in HDAC3 ; Table S6: Prediction values of the best BFE model generated for test set in HDAC3.

Author Contributions: EFB and DR did the computational studies and wrote the manuscript. JM and DR supervised the MD simulations and revised the manuscript. PS, FM and MZ carried out the data correction and statistical tests and wrote part of the manuscript. MS and WS supervised the experiments and revised the manuscript.

Funding: This work was funded by the Deutsche Forschungsgemeinschaft (DFG) SI868/22-1 project number 46995445 (to W.S.).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

81. Cournia, Z., B. Allen, and W. Sherman. "Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations." *Journal of Chemical Information and Modeling* 57, no. 12 (2017): 2911-37. <https://doi.org/10.1021/acs.jcim.7b00564>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.