Fragment based G-QSAR and molecular dynamics based mechanistic simulations into hydroxamic-based HDAC inhibitors against spinocerebellar ataxia

Siddharth Sinha, Chetna Tyagi, Sukriti Goyal, Salma Jamal, Pallavi Somvanshi & Abhinav Grover

To cite this article: Siddharth Sinha, Chetna Tyagi, Sukriti Goyal, Salma Jamal, Pallavi Somvanshi & Abhinav Grover (2016) Fragment based G-QSAR and molecular dynamics based mechanistic simulations into hydroxamic-based HDAC inhibitors against spinocerebellar ataxia, Journal of Biomolecular Structure and Dynamics, 34:10, 2281-2295, DOI: 10.1080/07391102.2015.1113386

To link to this article: https://doi.org/10.1080/07391102.2015.1113386
Fragment based G-QSAR and molecular dynamics based mechanistic simulations into hydroxamic-based HDAC inhibitors against spinocerebellar ataxia

Siddharth Sinha, Chetna Tyagi, Sukriti Goyal, Salma Jamal, Pallavi Somvanshi and Abhinav Grover

© 2016 Informa UK Limited, trading as Taylor & Francis Group

Expansion of polyglutamine (CAG) triplets within the coding gene ataxin 2 results in transcriptional repression, forming the molecular basis of the neurodegenerative disorder named spinocerebellar ataxia type-2 (SCA2). HDAC inhibitors (HDACi) have been elements of great interest in polyglutamine disorders such as Huntington’s and Ataxia’s. In this study, we have selected hydroxamic acid derivatives as HDACi and performed fragment-based G-QSAR, molecular docking studies and molecular dynamics simulations for elucidating the dynamic mode of action of HDACi with His-Asp catalytic dyad of HDAC4. The model was statistically validated to establish its predictive robustness. The model was statistically significant with $r^2$ value of .6297, cross-validated co-relation coefficient $q^2$ value of .5905 and $F$-test value of 56.11 confirms absolute robustness of the model. Two combinatorial libraries comprising of 3180 compounds were created with hydroxamate moiety as the template and their pIC50 activities were predicted based on the G-QSAR model. The combinatorial library created was screened on the basis of predicted activity (pIC50), with two resultant top scoring compounds, HIC and DHC. The interaction of the compounds with His-Asp dyad in terms of H-bond interactions with His802, Asp840, Pro942, and Gly975 residues of HDAC4 was evaluated by docking and 20 ns long molecular dynamics simulations. This study provides valuable leads for structural substitutions required for hydroxamate moiety to exhibit enhanced inhibitory activity against HDAC4. The reported compounds demonstrated good binding and thus can be considered as potent therapeutic leads against ataxia.

Keywords: spinocerebellar ataxia type-2; ataxin 2; G-QSAR; HDAC inhibitors

Introduction

Spinocerebellar ataxia type-2 (SCA-2), a prevalent polyglutamine disorder generally originates through mutation that leads to an expansion of polyglutamine (CAG) triplet in ataxin 2, located in the coding region of chromosome 12q23-24.1 (Armstrong et al., 2005). Functional inadequacy of ataxin gene results in progressive spinocerebellar neurodegeneration. The initial symptoms are inability to coordinate walking and hand and eye movements which later lead to slurred speech, muscle weakness, and other symptoms of gait and body incoordination (Sanpei et al., 1996). At present there is no treatment of spinocerebellar ataxia, and the affected patients are restricted to wheel chair after 5–7 years of the onset of symptoms. Later, with disease progression they become completely disabled. The normal SCA-2 allele comprises 15–32 CAG repeats interrupted by 2–4 repeats of CAA, whereas reverted allele of the SCA-2 gene expands to 37–58 CAG repeats and comprises a pure stretch of CAG (Pulst et al., 1996). Recent studies have demonstrated that expanded polyglutamine domains result from mutations that enhance the functions of the protein and are the center for pathogenesis. The mutant protein having CAG trinucleotide repeats affect the gene expression through interaction with several transcription factors thereby altering gene transcription (Butler & Bates, 2006). Gene expression is regulated through the remodeling of chromatin, the DNA complex and the histone proteins. DNA wraps around the histones and brings change in gene expression according to the change in wrap density of the DNA around the histone. It is mediated through histone modification such as acetylation, methylation, ubiquitination, and phosphorylation, which plays a dynamic role in gene expression regulation through transcription factor binding (Riley & Orr, 2006).

Histone acetylation is regulated by interplay of two enzyme classes, i.e. acetylation of histone through Histone acetylases (HAT) and removal of acetyl group through Histone de-acylase (HDAC). Histone acetylation aids transcription factors in accessing specific regions
of DNA, thereby increasing gene transcription (Didonna & Opal, 2015). Alternation of gene transcription has also been evident as an early feature of transgenic mouse model of other polyglutamine disorders such as Huntington’s disease (HD) (Gardian et al., 2005; Hockly et al., 2003). Recent studies have demonstrated that transcriptional repression affects through decreasing several functions including histone acetyl transferases and transcription factors such as A2BP1, TDP-43, DDX6, and PABPC1 (Weber, Sowa, Binder, & Hubener, 2014) in ataxia pathogenesis. Although, the precise molecular pathology for transcription dysregulation in ataxia is yet to be explored and explained, it has been well documented that transcriptional repression mainly results from dysregulation. HDAC’s have been grouped in four different classes based on structural similarity, phylogenetic comity with yeast de-acetylases, expression pattern, catalytic mechanisms, and localization (nucleus or cytoplasm) (Katsuno, Watanabe, Yamamoto, & Sobue, 2014). They have been categorized into: class I comprising of HDAC’s 1, 3, and 8. Class IIa comprising of HDAC’s 4, 5, 7, and 9. Class IIb comprising of HDAC’s 6 and 10. Class III comprising of silent information regulators (SIRT’s 1–7) and class IV consisting of HDAC 11.

Class Ia HDAC’s demonstrate tissue-specific expression pattern (Finnin et al., 1999) as such in muscle and heart tissue, as compared to class I HDAC’s which have been reported to be expressed ubiquitously, i.e. are expressed globally. Tissue-specific expression of class II HDAC’s indicates their importance in regulating gene expression through interaction with transcription factors, are also highly expressed in brain indicating an additional role in controlling gene transcription in neurons (Kazantsev & Thompson, 2008). Nuclear localization of class II HDAC viz. HDAC’s 4, 5, 6, 7, 9, and 10 is required for histone de-acetylation and it occurs through nuclear localization signal and also through co-localization with other nuclear proteins, i.e. HDAC3. Phosphorylation of the serine residues, a target site for kinases signals the export of the HDAC’s from the nucleus into the cytoplasm where it binds to the protein 14-3-3 and can only shuttle back to nucleus with the loss of phosphorylation and consecutive dissociation with the 14-3-3 protein. This regulation limits the accessibility of nuclear transition protein, including the myocyte enhancer factor 2 (MEF2) family and thereby down regulating the transcriptional process. Shuttling of HDAC4 and other class IIa enzymes depends upon synaptic activity in hippocampal neurons and provides a mechanism towards specific gene expression. Recent in vivo studies on mouse model of polyQ disorder have shown significant improvement in neurological phenotypes and motor coordination when crossed with HDAC4 heterozygous model (Herman et al., 2006). The HDAC active site consists of well-characterized catalytic domains surrounding the Zn$^{2+}$ ions, having two histidine aspartic dyads (His802-Asp838 and His803-Asn845), a large N-terminal regulatory domain, in addition to a C-terminal catalytic domain (Majdzadeh, Morrison, & D’Mello, 2008; Wang et al., 2010) which constitutes the charge relay system involved in the removal of acetyl groups and thereby resulting in the state of hypo-acetylation. The de-acetylase catalytic domain comprises approximately 400 residues arranged into 21 alpha helix and 10 beta strands, structured around a central “catalytic” Zn$^{2+}$ ion. Despite this similarity with class I HDAC’s, class Ia HDAC’s possess a bigger active site, which is mainly caused due to the mutation of a tyrosine into a histidine Y967H in HDAC4 (Lahm et al., 2007). Trifluoroacetyl-lysine has been identified as best characterized substrate for probing the elusive catalytic activity of vertebrate class Ia histone deacetylases, thus validating the regiospecificity of the inhibitors towards specific class Ia HDAC’s, the activity of class I HDAC’s towards trifluoroacetyl-lysine is not clearly distinguished, thereby serve as off targets (Di Giorgio, Gagliostro, & Brancolini, 2015). HDAC inhibitors (HDACi) functions by binding to the Zn$^{2+}$ at the bottom of the tubular pocket, making the charge relay system dysfunctional.

HDACi belonging to different classes are currently under clinical trials, demonstrating significant anti-proliferative activity. These include aliphatic acids, $\alpha$-aminoalnilides, electrophilic ketones, and cyclic peptides (Juvalle et al., 2006), out of which most extensively exemplified class of inhibitors is based upon the hydroxamic acid moiety, which is a widely recognized zinc binding group (Bürlí et al., 2013). HDACi comprises a metal-binding group, a hydrophobic cap that interacts with the amino acid residues and an aliphatic spacer that forms a connection between the hydrophobic cap and the metal binding group. Studies have reported that slight modification of the described structure impact both on the specificity and potency of the inhibitor, crystal structure of the class Ia deacetylase domain have stimulated the development and synthesis of many hydroxamte-based HDACi, with the aim of selectively influencing class Ia HDAC’s. Recent studies have shown that modification in the linker region of the hydroxamic moiety enhances the selectivity towards the Class Iia and Iib HDAC’s (Lobera et al., 2013). Synthetic inhibitors such as sodium phenyl butyrate, sodium valproate, suberanilo hydroxamic acid (SAHA), straight chain Trichostatin A (TSA), and oxamflatin have been reported. The HDAC inhibitor SAHA ameliorates motor deficits in the R6/2 transgenic mouse model of HD, whereas HDAC inhibitor phenyl butyrate exerts significant effects on the survival and ameliorates histopathologic degeneration in the N171–82Q transgenic mouse model of HD when administrated after the onset of the symptoms (Thomas, 2014). These findings have also been validated through the pre-clinical trials with
SAHA, in the R6/2HD mouse model, where it crosses the blood–brain barrier and enhances the histone acetylation. SAHA have been demonstrated to be administered orally in drinking water when complexed with cyclodextrins resulting in improved motor impairment in R6/2 mice clearly validating the pursuit of this class of compounds towards polyglutamine disorder. Hydroxamic acid-based HDACi have broad application in various neurological disorders, such as in Alzheimer’s disease (AD) the HDAC6-specific inhibitor tubastatin A was shown to rescue mitochondrial axonal transport in primary hippocampal neurons exposed to the neurotoxic Aβ peptide and impressively was also effective in rescuing cognitive deficits and reducing tau levels in a mouse models of AD (Bassett & Barnett, 2014). In case of Parkinsons disease HDACi rescue α-synuclein-dependent cytotoxicity both in cellular and fly models of the disease.

In recent years, a substantial amount of computational studies have been performed towards HDACi (Tambunan, Bramantya, & Parikesit, 2011), but not much towards their role in polyglutamine disorders; to the best of our knowledge. HDACi have been largely used to rectify the role of HDAC’s in neurodegeneration (Okazawa, 2003). Recent studies on Caenorhabditis elegans, Drosophila melanogaster, and mouse model of polyglutamine expansion disease identified the clinical characteristics of the disease (Volmar & Wahlestedt, 2015). Studies have reported that inhibitors of HDAC such as TSA, SAHA, and others act on gene expression and are potent persuaders of apoptotic cell death, differentiation, growth arrest in vitro, and in vivo (Wang et al., 2013). HDACi have also been seen as effective compounds to improve transcriptional changes, as observed in yeast, cell culture, and Drosophila models (Bertrand, 2010). HDACi have emerged as favorable drug candidates to restore the balance between the HAT and HDACs in disease state. In order to further exploit these derivatives, we have applied an innovative approach, i.e. G-QSAR (Abdullahi et al., 2014), in which descriptors are calculated on molecular fragments of the derivative data-set. Modifications at specific substitution sites in terms of addition or removal of the substituents may lead to identification of compounds with enhanced activity (Goyal et al., 2014). G-QSAR model provides valuable insights on each fragment towards activity variation and drug inhibitor interactions, thereby providing pragmatic and achievable interpretations. Recent studies have reported successful application of G-QSAR in drug discovery and lead optimization (Estrada & Molina, 2001), especially in cancer therapeutics (Tyagi, Gupta, Goyal, Dhanjal, & Grover, 2014). Thus, G-QSAR model provides leads regarding the substitution of a particular chemical entity at the specific site for activity enhancement (Nair, Teli, Pradeep, & Rajanikant, 2012).

Present study focuses on HDACi having hydroxamic moiety (Zhang, Fang, Zhu, Wang, & Xu, 2009) as its role towards ataxia has not been reported till date to the best of our knowledge. We attempt to identify the structural modifications required for hydroxamic derivatives as anti-ataxia compounds through G-QSAR model-based prediction. The selected compounds were validated by elucidating their binding pattern with HDAC through molecular docking analysis. Current study highlights the contribution of hydroxamic derivatives as anti-ataxia compounds and estimation of lead candidates using the generated G-QSAR model.

**Methodology**

**Protein selection and identification**

In total there are 18 HDAC’s grouped into four classes according to the phylogeny and sequence homology. We have selected class IIa HDAC4 protein (PDB ID: 2VQM) for the current work since it provides the molecular basis of the low enzymatic activity of class IIa HDAC towards acetylated lysine and reveal active site features that can help in identification and design of class IIa specific inhibitors. The structure reveals a class IIa specific conformational flexible structural zinc binding domain conserved in all class IIa enzymes, where in it adopts a different conformation in inhibitor free “closed” and inhibitor bound “open” form of the enzyme (Bottomley et al., 2008). Thus, suggesting a key role of the structural zinc binding domain in the regulation of class IIa HDAC functions. Structural studies have well demonstrated that these class IIa protein have novel features: firstly, class IIa HDAC’s possess a bigger active site than class I HDAC, due to mutation of a tyrosine into histidine, Y967H in HDAC4 (Lobera et al., 2013). Secondly, they possess a N and the C terminal regions comprising of the glutamine-rich domain and catalytic de-acetylase domain, involved in various signaling pathways through specific post translational modifications including nuclear and cytoplasmic shuttling.

**Selection of data-set for model generation**

Forty-four inhibitors belonging to the hydroxamate class of compounds (Price et al., 2007) were selected for G-QSAR model generation. Two-dimensional (2D) structure for compounds were drawn using Marvin Sketch (Wold, Sjöström, & Eriksson, 2001), and later converted from 2D to 3D using Vlife Engine platform of VLife MDS (Friesner et al., 2004). The energy minimization for all the compounds was carried out using force field batch minimization. Template is used as structural moiety common to all compounds of the congeneric data-set and was created using the Modify utility of VLife by substituting dummy atoms on the common moiety thus...
marking the substitution sites. The template had only one substitution site marked as R1.

**Molecular descriptors for G-QSAR model**

After selecting the congenic set of molecules and the template, G-QSAR model was built taking into account a number of 2D descriptors. Activities (in terms of pIC\textsubscript{50} value) of the data-set compounds were incorporated and the physio-chemical descriptors were calculated using the calculate descriptors command of the G-QSAR module. In G-QSAR, a variety of descriptors include retention index (chi), atomic valence connectivity index (chiv), path count, chi chain, chiv chain, path cluster, kappa, element count, estate numbers, and polar surface area. Out of the total 1148 descriptors calculated for the substitution site, 288 descriptors (independent variables) were used in the QSAR analysis, after removing the invariable columns. Invariable descriptors have the same value for each data-point and should be discarded for G-QSAR model development, as they have same quantitative value for each data point and are inefficient.

**G-QSAR model generation and its validation**

In order to compare the biological activities of the set of compounds which have a wide range of chemical structures (i.e. different fragment-based descriptors), the data-set was divided into representative training and test sets using advanced data selection, the training and test set compounds were chosen after selecting the activity pIC\textsubscript{50} as dependent variable and all the calculated descriptors as independent variables. Eight compounds namely 5, 10, 14, 22, 25, 29, 34, and 40 (Supplementary File 1) were selected as the test set and the remaining 36 compounds for the training set using random selection method, with the criteria to keep 80\% compounds in the training set. Partial least square regression (Sastry, Adzhigirey, Day, Annabhimoju, & Sherman, 2013) along with the stepwise forward selection method was used for generating G-QSAR model after validating the test set selection by unicolmn statistics. Advance variable selection procedure was utilized with cross co-relation limit kept as .5, F-test as .5, cut-off variance as .1 and random iterations as 150. Different statistical parameters were used to examine the goodness-of-fit for the model obtained such as cross-validated correlation coefficient ($q^2$), squared correlation coefficient ($r^2$), degree of freedoms (df), chances of model failure (F-test), number of variables ($k$), number of compounds in regression ($n$), pred $r^2$, $Z$-score, and highest $q^2$ value in the randomization test (best_ran $q^2$) need to be taken into account.

**Model validation**

Many statistical parameters such as $n$ (number of compounds in regression), $k$ (number of variables), degree of freedom, optimum component (number of optimum PLS components in the model), $r^2$ (squared correlation coefficient), F-test (Fischer’s value), $q^2$ (cross-validated correlation coefficient), pred $r^2$ ($r^2$ for external test set), $Z$-score (randomization test), best_ran $q^2$ (highest $q^2$ value in the randomization test), and best_ran $r^2$ (highest $r^2$ value in the randomization test) need to be taken into account to consider the model as a robust one. For a model to be statistically significant, the following conditions should be satisfied: $r^2$, $q^2 > .6$ and pred $r^2 > .5$. Since F-test gives an idea of the chances of failure of the model, a value greater than 30 is considered to be good. On the other hand, low standard error values establish absolute quality of the model.

**Cross validation for GQSAR model**

Leave-one-out (LOO) method was used with $q^2$ (cross-validated correlation) for internal validation of the GQSAR model created. $q^2$ is the correlation calculated with respect to randomly created training set compounds using the given equation:

$$q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

where $y_i$ and $\hat{y}_i$ are the actual and predicted activities of the $i$th molecule in the training set, respectively, and $\bar{y}$ is the average activity of all the molecules in the training set.

$p$IC\textsubscript{50} values of each compound in test set were predicted for external validation using the model created through training set, is given by pred $r^2$ and is calculated using the given equation:

$$\text{pred } r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

where $y_i$ and $\hat{y}_i$ are the actual and predicted activities of the $i$th molecule in the test set, respectively, and $\bar{y}$ is the average activity of all the molecules in the training set.

$Y$ randomization test was performed by comparing the resultant linear model with those derived from random data-set (Rücker, Rücker, & Meringer, 2007), in order to eliminate the risk of chance correlation. $Z$-score was used as a basis for comparing models build on random data-set generated by rearranging the compounds in the training set in order to compare them with the G-QSAR model. A $Z$-score value is obtained through following equation:

$$Z \text{ score} = \frac{\bar{y} - \mu}{\sigma}$$
where \( h \) is the \( q^2 \) value calculated for the actual data-set, \( \mu \) is the average \( q^2 \), and \( \sigma \) is the standard deviation calculated for various models built on different random data-sets.

**Generation of combinatorial libraries**

The Leadgrow module in VLife MDS was utilized for creating combinatorial libraries, based on the previous template with the substitution site R1. Two libraries (S1, S2) were generated by substituting aromatic rings, acetate, carbomethoxy, and 2-oxazoyl, carbonic groups at R1 substitution site in library S1, whereas substitution with cyclopentane, vinyl, naphthalene, 2-thiazolyl, alkyl, and cyclic groups in library S2. G-QSAR model generated was used for predicting their biological activity.

**Molecular docking of HDAC4 with combinatorial library compounds**

Fifty-three compounds with highest predicted activity ranging from 6.14 to 7.28, were selected from the libraries and docked to the active site of HDAC4 comprising of histidine-aspartic acid dyads (His802-Asp840 and His803-Asp845) and a histidine (HIS842) residue. All the selected molecules were prepared using LigPrep module in Schrodinger’s suite for generating stable conformers followed by energy minimization. A cubic grid with dimensions of 20*20*20 Å was created around the active site residues (Tandon & Sinha, 2011). Extra precision (Friesner et al., 2006) protocol in the glide module (Dhanjal, Goyal, Sharma, Hamid, & Grover, 2014; Dhanjal, Grover, Sharma, Singh, & Grover, 2014) was used for docking in order to identify compounds with high affinity for HDAC4. Docking has been a widespread technique for screening of potential compounds in computational drug design. The interactions involved between the identified compounds and protein structure were evaluated using LigPlot (Wallace, Laskowski, & Thornton, 1995) and UCSF Chimera (Pettersen et al., 2004).

**Comparison of the proposed HDACi with known inhibitors**

Although great work is being done to develop highly specific HDACi, only three drugs, SAHA, Panobinostat, and FK-228 are in the phase III clinical trial. The identified HDACi were compared to these known potent drugs in terms of their docking score to predict their action potential. The three-dimensional structure of SAHA, panobinostat was retrieved from PubChem and prepared using LigPrep. After preparation the structures were docked against the active site of HDAC’s using the same protocol as of identified HDACi.

**Molecular dynamics simulations of the crystal protein (2VQM) and docked complexes**

MD simulations were carried out using the GROMACS version 5.0 in order to affirm the conformation and stability of the HDAC4-inhibitor complex using GROMOS 53A6 force field (Sehrawat, Sinha, & Saxena, 2015; Van Der Spoel et al., 2005). GROMACS topology files for the identified inhibitors were created using the PRODRG2 server. The protein-inhibitor complex was at the center of a 90*90*90 Å cubic grid, solvated through SPCE water model, which have been shown to be compatible with the GROMOS 53A6 force field (Hess & van der Vegt, 2006). Energy minimization of the protein-inhibitor complex was carried out using steepest descent method for 5000 steps, until an upper limit of 28 kcal/mol was obtained keeping other parameters as default. The complex was optimized using the Schrodinger’s protein preparation module and energy minimized with OPLS_2005 force field. Long-range electrostatic interaction was carried out using the PME electrostatics, whereas for coulombic short-range interactions cut-off scheme was applied, with cut-off radius of 9 Å which was occurring during MD simulation. Finally, 20 ns long MD simulation for the complex was performed at a fixed temperature of 300 K, pressure 1 bar, and at time step of 2 femto second.

**Principal component analysis**

In order to obtain dominant and collective modes of the protein from the overall dynamics of the MD trajectory, we performed principal component analysis (PCA). PCA is based on the construction of a mass-weighted covariance matrix of the atom displacement. This covariance matrix is diagonalized to extract a set of eigenvectors and eigenvalues that reflect concerted motion of the molecule (Amadei, Linssen, & Berendsen, 1993). Eigenvectors represent the direction of motion, whereas the corresponding eigenvalues represent the amplitudes in those directions (Yamaguchi, van Aalten, Pinak, Furukawa, & Osman, 1998). The Gromacs inbuilt tool gmxcovar was used to yield the eigenvalues and eigenvectors by calculating and diagonalizing the covariance matrix, whereas the gmxcovar tool was used to analyze and plot the eigenvectors (Aalten, Findlay, Amadei, & Berendsen, 1995).

**Calculation of ADMET properties**

QikProp module (Goyal, Dhanjal, Tyagi, Goyal, & Grover, 2014) of Schrodinger suite was used for evaluating the toxicity of final two compounds. It predicts various molecular properties and also provides ranges for comparing these properties with 95% of already available drugs. It predicts absorption, distribution, metabolism,
and elimination properties of the selected compounds through certain parameters based on the similarity index of the available drug information (Natarajan, Sugumar, Bitragunta, & Balasubramanyan, 2015).

Results and discussion

G-QSAR model and validation of the data-set

Forty-four compounds selected for G-QSAR studies were grouped into test and training set having 8 and 36 compounds, respectively. The compounds in the test and training set were validated through unicolumn statistics (Table 1) indicating that the classification of test set is in compliance with the maximum, minimum, and standard deviation values of the test and training sets being derived within maximum–minimum range of the training set.

The G-QSAR model created on the basis of the given parameter is:

$$pIC_{50} = (0.8665548 \times R1 - \text{chi3 Cluster}) + 4.9551 \quad (1)$$

Different statistical parameters are given as:

- $N = 36$, $F$ test $= 56.112$, $r^2 = .6297$, $q^2 = .5905$, $\text{pred } r^2 = .8524$, degree of freedom $= 33$, $r^2 \text{ se} = .1702$, $\text{pred } r^2 \text{ se} = .2845$, $q^2 \text{ se} = .1944$, $Z\text{-score}_{r^2} = 2.10604$, $Z\text{-score}_{q^2} = 4.84859$.

- $n$ indicates the number of molecules for regression analysis, $r^2$ is squared correlation coefficient indicating quality of fit, a high $F$-test value implied that the model is 99.999% statistically valid with less than 1 in 10,000 chance of failure, $q^2$ represents cross-validated squared correlation coefficient, and predicted squared correlation coefficient is represented by $\text{pred } r^2$. Z-Score evaluates the distance of the observed value from the mean (Table 2). The above-mentioned statistical parameters indicate towards the robustness of the G-QSAR model developed. Exploratory statistical analysis was applied between the selected 2D descriptors and it was observed that the majority of 2D descriptors were formed to be statistically significant except the few having a $p$-value greater than .025, analyzed through the correlation matrix between the selected descriptors.

A high correlation coefficient and low standard error value of .6297 and .1702 signify the accuracy of the model. High $F$ test score (56.112) indicates the robustness of the model, i.e. the chance for random selection or failure is negligible. LOO method was used to calculate the internal prediction, i.e. $q^2$ value of .5905 and the predictive power was indicated through $\text{pred } r^2$ value of .8524. Radar graph signifies the difference in activities between predicted and observed activities shown in red and blue color, respectively (Figure 1). Radar graph demonstrates the high $r^2$ value by the extent of overlap between observed and predicted activity for the training set, whereas, and indicates a high $\text{pred } r^2$ coefficient for the test set. The fitness plot represents observed vs. predicted inhibitory activity (Figure 2) and demonstrates the difference in activities in terms of distance points from the regression line. The contribution of the selected descriptor was 74.84% in enhancing the activity (positive contribution) of the hydroxamate derivatives.

The G-QSAR model obtained from 44 hydroxamate derivatives was built upon R1 chi3 Cluster physio-chemical descriptor chosen amongst the other 178 2D descriptors. The QSAR model equation (i) defines the coefficient of the respective descriptor, i.e. chi3 Cluster at substitution site R1 and the last term as the regression constant. It predicts the biological activity ($pIC_{50}$) of novel compounds if the value for this descriptor is known.

$$R1\text{-chi3 Cluster}$$

It defines the constraint index (third order), inferred mainly from gradient holding time. It defines the normalized retention time required for a solute to migrate from the column in chromatographic analysis. Substitution site R1 in the hydroxamic core was substituted by atoms such as hydrogen, chlorine, bromine, and iodine, and groups such as cyclopentane, vinyl, naphthalene, and 2-thiazolyl for descriptor calculation. The descriptor’s percentage of contribution in enhancing the biological activity was 74.28% indicating that the compounds having three carbon atoms (high molecular weight) attached between the hydroxamic derivatives and the cluster chi index enhance the inhibitory activity in comparison to the compounds having less than three carbon atoms attached.

Combinatorial library activity prediction using GQSAR model

About 3180 compounds were obtained through two combinatorial libraries ($S_1$ and $S_2$). Library $S_1$ constituted of

<p>| Table 1. Unicolumn statistics for the training and test set compounds. |
|-------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Standard deviation</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>7.3974</td>
<td>8.4000</td>
<td>5.6000</td>
<td>.7672</td>
<td>258.9100</td>
</tr>
<tr>
<td>Test</td>
<td>7.7067</td>
<td>8.2200</td>
<td>6.0500</td>
<td>.6640</td>
<td>69.3600</td>
</tr>
</tbody>
</table>


2240 compounds by substituting R1 site with aromatic rings, acetate, carbomethoxy, 2-oxazoyl, and carbonic groups, whereas library S2 constituted of 940 compounds through substituting groups such as cyclopentane, vinyl, naphthalene, 2-thiazolyl, alkyl, and cyclic groups. Out of all compounds from both the libraries, top 53 hydroxamate derivatives in terms of high predicted activity were selected for molecular docking and dynamics simulations (20 ns) in order to gain structural information of their inhibitory role. These compounds had an extrapolation value of 0 or close to 0.

Temp0135 with an imidazole ring substitution at the R1 site and molecular weight of 68.077 g/mol had the highest predicted activity of 7.2484. In temp0174, the group C3H8O at R1 substitution site is lighter than the imidazole ring, having a molecular weight of 60.10 g/mol and hence a comparatively lower activity value of 6.428. Similarly, temp0133 further goes down on the activity value as R1 site is substituted with the ethyl group having low molecular weight 46.06 g/mol. It was observed as a general trend, as on substitution of a chemical group such as alkyl, alkene, acid comprising of ethyl, isopropyl, isobutyl, carboxyl, and in carbonic acid the compounds showed lower activity (5.24–5.81) as compared to methoxy ketone having high molecular weight of 100.16 g/mol and having an inhibitory activity of 6.29. Thus, it was observed that the presence of a molecule with high molecular weight at the substitution site R1 enhances the inhibitory activity of hydroxamic acid derivatives as in temp1169, temp2848, temp3166, and others. The predicted activity values generated using the G-QSAR model with respect to substitution site R1 have been given for the top 30 compounds (Table 3).

Docking and molecular dynamics simulations carried out for screened compounds of the combinatorial library

Top 53 molecules selected from the combinatorial library of 3180 compounds having high predicted activity were docked at the His-Asp dyad active site of HDAC4. Two top scoring compounds with good binding affinity to the His-Asp active site are reported. His-Asp catalytic dyad active site comprises His 802, Asp 840, Gly 975, and Pro 942 as neighboring residues. The protein-inhibitor complexes were also subjected to MD simulations of 20 ns. The interaction pattern after MD simulation is discussed. The first compound temp0135 named, N-hydroxy 5 (1H Imidazole-4yl) thiopene-2-carboxamide (HIC) (Figure 3(a)), with a glide score of -8.84 kcal/mol, showed hydrophobic interaction with Pro 942, Phe 871, Phe 812, Leu 943, Asp 934, His 842, Asp 840, His 802, Gly 811, Glu 677, Gly 975, His 976, His 803, Pro 800, Gly 974, Arg 681. Important interacting residues have been depicted in Figure 4(a). No H-bonding interactions were observed. The inhibitor seemed to align itself to a more vertical position thereby subsequently losing contact with the catalytic dyad and opening up of the active site cleft. However, at 15 ns time step the inhibitor anchored itself to the dyad site again (Figure 5a–5e).
This occurs by interacting with the side chains of Asp 934 and His 803 and thus closing the active site again. The second compound temp 1864 named, 5\(((1R,4aR,8aR)\) decahydronaphthelene-1yl-3'-Nhydroxy-thiopene-2-carboxamide\) (DHC) (Figure 3(b)), with a glide score of -7.59 kcal/mol, showed hydrogen bonding with

![Fitness Plot](image)

**Figure 2.** Fitness Plot of the G-QSAR model obtained for actual and predicted activities. The closer the points are from the linear trendline, minimum is the difference between the actual and predicted activities.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Compound</th>
<th>R1</th>
<th>Predicted activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Template 113</td>
<td>-C4H9</td>
<td>6.54</td>
</tr>
<tr>
<td>2</td>
<td>Template 133</td>
<td>-4-methyl-1-benzo thiopen</td>
<td>7.04</td>
</tr>
<tr>
<td>3</td>
<td>Template 135</td>
<td>-4-methyl-1-imidazole</td>
<td>7.08</td>
</tr>
<tr>
<td>4</td>
<td>Template 169</td>
<td>-2-methyl-butane</td>
<td>7.04</td>
</tr>
<tr>
<td>5</td>
<td>Template 182</td>
<td>-2,2-dimethyl-propane</td>
<td>6.98</td>
</tr>
<tr>
<td>6</td>
<td>Template 204</td>
<td>-methyl acetate</td>
<td>6.78</td>
</tr>
<tr>
<td>7</td>
<td>Template 453</td>
<td>-carbonic acid</td>
<td>6.64</td>
</tr>
<tr>
<td>8</td>
<td>Template 864</td>
<td>-acetaldehyde</td>
<td>7.28</td>
</tr>
<tr>
<td>9</td>
<td>Template 1298</td>
<td>-methyl formate</td>
<td>7.28</td>
</tr>
<tr>
<td>10</td>
<td>Template 1428</td>
<td>-benzene</td>
<td>7.04</td>
</tr>
<tr>
<td>11</td>
<td>Template 1531</td>
<td>-phenol</td>
<td>7.04</td>
</tr>
<tr>
<td>12</td>
<td>Template 1579</td>
<td>-toluene</td>
<td>7.04</td>
</tr>
<tr>
<td>13</td>
<td>Template 1577</td>
<td>-1-methyl naphthalene</td>
<td>6.65</td>
</tr>
<tr>
<td>14</td>
<td>Template 1587</td>
<td>-1H-pyrole</td>
<td>6.65</td>
</tr>
<tr>
<td>15</td>
<td>Template 1603</td>
<td>-4-methyl-1H-indole</td>
<td>6.41</td>
</tr>
<tr>
<td>16</td>
<td>Template 1639</td>
<td>-furan</td>
<td>6.41</td>
</tr>
<tr>
<td>17</td>
<td>Template 1688</td>
<td>-1-benzofuran</td>
<td>6.54</td>
</tr>
<tr>
<td>18</td>
<td>Template 1708</td>
<td>-thiopene</td>
<td>7.14</td>
</tr>
<tr>
<td>19</td>
<td>Template 1749</td>
<td>-pyridine</td>
<td>7.14</td>
</tr>
<tr>
<td>20</td>
<td>Template 1765</td>
<td>-ethyl acetate</td>
<td>7.14</td>
</tr>
<tr>
<td>21</td>
<td>Template 1793</td>
<td>-1H-imidazole</td>
<td>7.04</td>
</tr>
<tr>
<td>22</td>
<td>Template 1837</td>
<td>-2-methyl-1,3 oxazole</td>
<td>7.04</td>
</tr>
<tr>
<td>23</td>
<td>Template 1864</td>
<td>-1-methyl-naphthalene</td>
<td>7.28</td>
</tr>
<tr>
<td>24</td>
<td>Template 1907</td>
<td>-2-methyl pyridine</td>
<td>6.87</td>
</tr>
<tr>
<td>25</td>
<td>Template 1969</td>
<td>-4-ethyl-pyrimidine</td>
<td>6.87</td>
</tr>
<tr>
<td>26</td>
<td>Template 2036</td>
<td>-4-methyl-1H-pyrazole</td>
<td>6.54</td>
</tr>
<tr>
<td>27</td>
<td>Template 2246</td>
<td>-pyridazine</td>
<td>6.54</td>
</tr>
<tr>
<td>28</td>
<td>Template 2544</td>
<td>-amine</td>
<td>6.89</td>
</tr>
<tr>
<td>29</td>
<td>Template 2845</td>
<td>-phenoxy</td>
<td>6.89</td>
</tr>
<tr>
<td>30</td>
<td>Template 3172</td>
<td>-5-methyl-2H-imidazole</td>
<td>6.42</td>
</tr>
</tbody>
</table>
Gly 975 and Pro 942 and hydrophobic interactions with Asp 934, Pro 676, Glu 677, Arg 681, Leu 943, Gly 811, Gly 974, His 803, His 842, Phe 871, and His 976. The interactions have been depicted in Figure 4(b). The O atom of Pro 942 formed a 3.03Å long H-bond with the N1 atom of DHC and N atom of Gly 975 formed a 2.95Å long H-bond with O2 atom of DHC. The stable structure was obtained between 2 and 20 ns time steps during MD simulations. The RMSD graph (Figure 5a) demonstrates the molecular dynamics simulation course performed with respect to RMSD vs. time for HDAC4 crystal structure shown in red, the HDAC4-HIC complex shown in blue, and the HDAC4-DHC complex shown in green. RMSD plots of individual inhibitors HIC and DHC are shown in Figure 5b. Analyzing the post docking interactions of the hydroxamic acid derivatives and HDAC, it was found that the binding orientation of the docked compounds was slightly different than the experimental binding mode (PDB ID: 2VQM) as demonstrated through the H-bond interaction with Pro 942 and Gly 975. We also analyzed the post dynamics deviation between the RMSD plot of the inhibitor free and inhibitor bound HDAC structures showing considerable difference in the conformational of the structural zinc-binding domain. Two loops numbered 225–342 and 885–900 harboring the residue coordinating the structural zinc ion moved 10–20 Å towards active site, resulting in the closed conformation where the inhibitor clashes with the structural zinc binding domain, thus promoting open conformation. Structural analysis demonstrates a switch with His675 and Cys669 in inhibitor free structure and His665 and His678 in the inhibitor bound structure, comparison of the inhibitor-free and inhibitor-bound HDAC structures suggest that inhibitors effect the conformation of the zinc binding domain, resulting in the deviation between RMSD of the inhibitor-free and bounded HDAC’s. In order to gain further insights of the structure compactness, we performed the RMS fluctuations and radius of gyration study for the for the HDAC4-inhibitor complex (Figure 5c–5e). Radius of gyration (Rg) is defined as the mass weighted root mean square distance of a collection of atoms from their common center of mass, providing an insight of the overall dimensions of the protein. It was observed that Rg value for the HIC and DHC complexes were 1.90 and 1.91 nm, respectively. The curve reveals that DHC complex demonstrates higher Rg value than the HIC complex during the whole simulation time period (Figure 5e). Root-mean-square fluctuation (RMSF) reveals the conformational flexibility of the bound complex for both atoms and protein residue, on analyzing the fluctuations it was observed that fluctuation of amino acid were largest in the region 830–980 in both HIC and DHC. Slight fluctuations were also observed at residue 725–764 in DHC. Amino acid falling in this region (Figure 5c) shows strong binding (hydrogen bonding and hydrophobic interactions) as reported. Figure 5d depicts the RMSF graph for amino acids of the HDAC4-inhibitor complex. The fluctuations observed in the catalytic dyad confirm the important role being played by these residues in ligand binding. Thus, both the screened compounds, i.e. HIC and DHC were also subjected to comparative study with the known HDACi such as SAHA, Panobinostat, and FK-228 all of them in the phase III clinical trials. The compound SAHA demonstrated the highest docking score of −6.82 kcal/mol followed by panobinostat and fk-228 resulted in the docking score of −6.31 and −5.41 kcal/mol. The comparison with respect to mode of interaction for HIC and DHC with commercially available HDACi in terms of hydrogen bond and hydrophobic interaction demonstrates binding to the identical His802-Asp840 catalytic dyad of HDAC4. The results were in line with our structural studies. The predicted activities of HIC and DHC were 7.04 and 7.28 and were in accordance with the docking studies. The predicted activities of HIC and DHC were 7.04 and 7.28 and were in accordance with the docking phase III clinical trials. The compound SAHA demonstrated the highest docking score of −6.82 kcal/mol followed by panobinostat and fk-228 resulted in the docking score of −6.31 and −5.41 kcal/mol. The comparison with respect to mode of interaction for HIC and DHC with commercially available HDACi in terms of hydrogen bond and hydrophobic interaction demonstrates binding to the identical His802-Asp840 catalytic dyad of HDAC4. The results were in line with our structural studies. The predicted activities of HIC and DHC were 7.04 and 7.28 and were in accordance with the docking results in terms of good binding affinity with HDAC4 and similar binding pattern with the His-Asp catalytic dyad.

**PCA for HIC and DHC**

The MD trajectory of system was inspected with the principal components to better understand the conformational changes of the HDAC, with both HIC and DHC complexes. Correlated motion plot shows how atoms move relative to each other. Motions can be positively correlated (fluctuation vectors in the same direction, colored in red), anti-correlated (fluctuation vectors in the same direction, colored in blue), and uncorrelated. The...
positive and negative limits for both the compounds are shown in Figure 6.

Covariance analyses revealed that most of the dominant motions were found in loops numbered from 42–93, 225–342, 438–501, and 885–900 which corresponds to residues 664–681, 725–764, 796–840, and 935–950 part of the protein, respectively. DHC-HDAC4 bound complex exhibits both anti-correlated and correlated motions in all the four loops except in the loops having atom number from 225–342 where they are uncorrelated, whereas HIC-HDAC4 in which the correlated and anti-correlated motions are either weak, or are uncorrelated (Figure 6).

**ADMET analysis of the compounds selected, i.e. HIC and DHC**

Qik Prop module in the Schrodinger package calculated 50 ADME properties, first property named #star indicates the number of properties that lies outside the similarity criteria of 95% of known drugs and having a reference range of 0–5. Its value for both HIC and DHC were 0, which signifies that no molecular property of HIC and DHC falls out of the similarity criteria. CNS property having reference range from −2 (inactive) to 2 (active) predicts the central nervous system activity, resulted in a value of −2 for HIC and 0 for DHC. HIC and DHC resulted in values of −1.031 and −.668, respectively, for
Figure 5a. RMSD trajectory of the HDAC4 protein and two docked complexes during 20 ns long MD simulations, HDAC4 crystal structure (2VQM) structure in red, HIC bound HDAC4 in blue and DHC bound HDAC4 in green.

Figure 5b. RMSD trajectory of the inhibitors (HIC and DHC) for the time period of 20 ns.

Figure 5c. Root mean square fluctuation (RMSF) of the backbone atoms of the HDAC4-inhibitor bound with HIC and DHC at 300 K. DHC bound HDAC4 structure is shown in blue; HIC bound HDAC4 structure is shown in orange.
predicted blood–brain partition coefficient QPlogBB, which were well within the reference range of 3.0–1.2. Descriptors such as SASA (solvent accessible surface area) have a reference range of 300–1000, its hydrophilic component FISA (N, O, and H on heteromers) and its hydrophobic component FOSA (saturated carbon and attached hydrogen) have a reference range of 7–33 and 0–750, respectively, the values obtained for these descriptors were within the defined range for both the compounds. Percent human oral absorption descriptor had a value of 65.83% for HIC and 90.38% for DHC, is predicted on a scale from 0–100%. QPlogkp is the predicted skin permeability factor having values between −4.092 and −3.557 for HIC and DHC, respectively, having a reference range from −8 to −1.0. HIC and DHC resulted in high value of coefficient for descriptors demonstrating the dispersion of the compounds within the body, as depicted by the values 7.571 and 8.836 for HIC and DHC, respectively, for descriptor QPlogPC16 indicating free energy of solvation in hexadecane, which were well within the defined range of 4–18. QPlogPw value between 13.256 and 8.939 for HIC and DHC, respectively, indicates free energy of solvation in water, having a reference range of 4–45. QPlogPoct values of 17.354 and 15.686 for HIC and DHC, respectively, indicates free energy of solvation in octanol, have a reference range of 8–35. The predicted octanol/water partition coefficient QPlogPo/w indicates hydrophobic nature of the chemical compounds suggesting easy absorption through lipid bilayer and resulted in values between −.217 and 2.452 for HIC and DHC, respectively, lying well within the reference range of −2–6.5.

Electron linkage (eV) and ionization potential for compounds were reported as 1.175 and 8.885 for HIC and DHC.
1.166 and 9.5838 for DHC, respectively, which also were found to be within the defined range of \(-9.9–1.7\) for electron linkage and from \(7.9–10.5\) for ionization potential. Lastly, both the compounds HIC and DHC were found to be satisfactory in all the parameters, when evaluated on the grounds of Lipinski rule-of-five and Jorgensen rule-of-three infringement.

Conclusion
In present study, we have developed a G-QSAR model on a set of 44 hydroxamate derivatives that play major role in inhibiting HDAC enzymes resulting in up-regulation of histone acetylation in gene expression as a potent avenue in SCA2. Sequestration of histone acetylases in the diseased state leads to the state of hypo-acetylation, thereby highlighting the functional role of HDAC’s in regulating the gene expression. The data-set was grouped into test set (8 compounds) and training set (36 compounds) and the contribution of various 2D descriptors at R1 substitution site was calculated which resulted in chi3Cluster as the most important one. Two combinatorial libraries were constructed comprising of 3180 compounds and their activities were predicted using G-QSAR model. Compounds with high molecular weight or with three carbon atom between the hydroxamate derivative and the cluster chi index contributed positively towards the inhibitory activity. Top 53 scoring molecules (in terms of predicted activity) were docked against the His-Asp catalytic dyad of the HDAC4 protein structure. Finally on the basis of the docking and molecular dynamics simulation study, compounds HIC and DHC having pIC_{50} (activity) between 7.28 and 7.04, respectively, were reported. DHC can be selected for further evaluation as it shows promising results, being less toxic in comparison to HIC. The in silico work provides a valuable insights towards the inhibitory roles of hydroxamic acid derivatives in HDAC inhibition and for consideration of selected compounds as anti-ataxia prospects, as further studies could provide leads for highly selective inhibitors for HDAC4 enzyme in the future.

Supplementary material
The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2015.1113386.

Acknowledgments
Abhinav Grover is thankful to the Jawaharlal Nehru University for usage of all computational facilities. Abhinav Grover is grateful to the University Grants Commission, India for the Faculty Recharge Position.

Disclosure statement
No potential conflict of interest was reported by the authors.

References


