Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
ELSEVIER
Bioorganic & Medicinal Chemistry Letters 14 (2004) 707–711

QSAR Studies of PC-3 cell line inhibition activity of TSA and SAHA-like hydroxamic acids

Di-Fei Wang,[a] Olaf Wiest,[a,*] Paul Helquist,[a] Hsuan-Yin Lan-Hargest[b] and Norbert L. Wiech[b]

[a]*Walther Cancer Research Center and Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA*
[b]*Beacon Laboratories, Phoenix, MD 21131, USA*

Received 25 September 2003; revised 6 November 2003; accepted 14 November 2003

Abstract—Quantitative structure–activity relationships (QSAR) for a series of new trichostatin A (TSA)-like hydroxamic acids for the inhibition of cell proliferation of the PC-3 cell line have been developed using molecular descriptors from Qikprop and electronic structure calculations. The best regression model shows that the PM3 atomic charge on the carbonyl carbon in the CONHOH moiety(Qco), globularity (Glob), and the hydrophilic component of the solvent-accessible surface area (FISA) describe the IC_{50} of 19 inhibitors of the PC-3 cell line with activities ranging over five orders of magnitude with an $R^2 = 0.92$ and $F = 59.2$. This information will be helpful in the further design of novel anticancer drugs for treatment of prostate cancer and other diseases affected by HDAC inhibition.

1. Introduction

Histone deacetylases are key enzymes for regulating transcription process by catalyzing the hydrolysis of the ε-acetylated lysine chain in histones. This modification results in changes of the ionization states of the lysine chains and causes chromatin remodeling during the transcription process. Recent studies show that inhibition of histone deacetylases (HDACs) elicits anti-cancer effects in several tumor cell lines by inhibition of cell growth and inducing cell differentiation.[1] For example, prostate cancer, one of the major cancers in USA, has been found to correlate with aberrant HDAC1 activity.[2] Although natural products such as cyclic tetrapeptides[3,4] or depudecin[5] as well as synthetic inhibitors,[6] shown in Figure 1, have been studied for this purpose in cancer cell lines and in tumor animal models (Fig. 1), the vast majority of the compounds under investigation as HDAC in most of the known HDAC inhibitors are hydroxamic acids such as trichostatin A (TSA),[7] or suberoylanilide hydroxamic acid (SAHA).[8] Although the availability of an X-ray structure for a histone deacetylase-like protein (HDLP)[9] facilitates structure-based design, no quantitative structure–activity (QSAR) study has been reported so far. This is unfortunate because factors other than the binding constant to HDAC1, for which HDLP is a model, might affect the biological activity of HDAC inhibitors even when the primary mode of action is the inhibition of this enzyme.

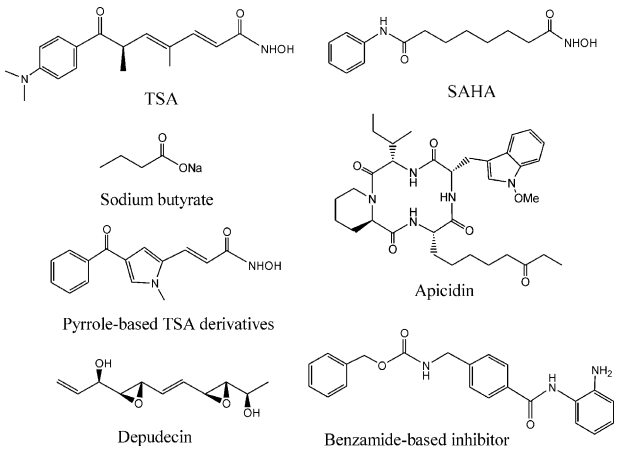


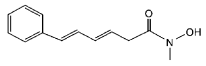
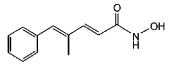
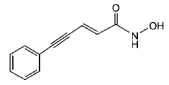
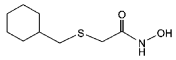
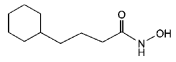
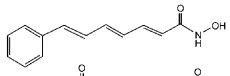
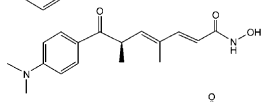
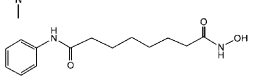
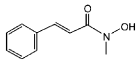
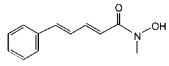
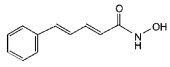
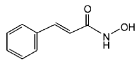
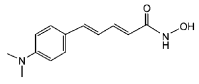
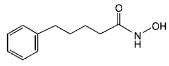
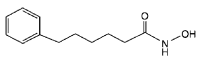
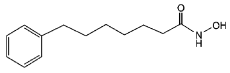
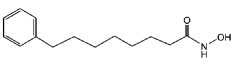
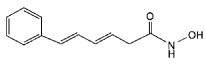
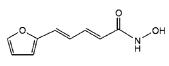
Figure 1. Selected known HDAC1 inhibitors.

Keywords: Histone deacetylase; QSAR; Hydroxamic acids; Enzyme inhibition.
* Corresponding author. Tel.: +1-574-631-5786; fax: +1-574-631-6652; e-mail: owiest@nd.edu

0960-894X/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2003.11.062

Here, we first report three QSAR models that were constructed by multiple linear regression analysis of the pIC_{50} data from a standard PC3 cell line anti-proliferation assay of a series of TSA-like hydroxamic acids that were described previously by two of us.[10] The predictive ability of the proposed models is investigated using a cross-validation method.

Table 1. Structured details and comparison of the experimental data and the calculated data data derived derived from QSAR equations

No.	Structure	PC-3 cell line, pIC_{50} (μM)		
		Exp.	Calcd	
			Forward	Backward
1		−2.26717	−2.44957	−2.58129
2		−1.00000	−0.57751	−0.44419
3		−0.34242	−0.98832	−0.98832
4		−2.33445	−1.65410	−1.67876
5		−1.69897	−1.77775	−1.78650
6		0.22185	−0.26763	−0.12185
7		2.30103	1.82309	2.01205
8		1.88606	1.91131	1.62534
9		−2.36173	−2.47684	−2.50749
10		−1.92942	−1.88757	−2.01772
11		−0.76343	−0.25848	−0.09883
12		−0.77815	−1.19942	−1.16819
13		−0.60206	0.041372	−0.25283
14		−1.62325	−1.71255	−1.56070
15		1.07918	−1.43909	−1.33229
16		1.32222	−1.07829	−1.11556
17		−1.07918	−0.72398	−0.92140
18		−1.35218	−1.40562	−1.18182
19		−0.85773	−0.84192	−1.18505

2. Computation

3-D structures of all compounds were built and minimized using the MM2 force field in MacroModel 7.0. The minimized structures were transferred to QikProp,[11] which was used to calculate a number of the molecular descriptors. Atomic partial charges were obtained from single point PM3[12] calculations using G98.[13] These calculations were carried out on a SGI Octane workstation. Multiple linear regression analysis was performed using SYSTA 10 and Microsoft Office Excel 2000 programs on an IBM-240X laptop.

3. Results and discussion

Table 1 gives all the structures used in this study and their biological activities, as well as the calculated activities from our models generated by forward and backward stepping strategies to the PC-3 cell line and HDAC1. The activity of the compounds spanned approximately five orders of magnitude from the most active compound, TSA **7**, to the least active compound, N-methyl cinnamyl hydroxamic acid **9**. The derived QSAR will reflect a PC3 cellular response driven by a net contribution of a mixture of the different HDAC enzyme in the cell as well as a contribution from the transport properties of the compounds. The descriptors we used are molecular weight (MW); the total solvent-accessible surface area (SASA); the hydrophilic component of the solvent-accessible surface area (FISA); the weak polar component of the solvent-accessible surface area (WPSA); an index of cohesive interactions in solids (ACxDN); the globularity of the compound (Glob), which is $4\pi r^2/SASA$, r is the radius of a sphere with a volume equal to the molecular volume; the calculated octanol/water partition coefficient (LogP); and the PM3 atomic charge on the carbon atom of the CONHOH moiety (Qco).

3.1. Stepwise multiple linear regression

We used two different approaches to build our QSAR models. In the 'backward stepping' approach, a QSAR equation is developed where all the independent variables are included to describe the dependent variable. Then the independent variable with the lowest *t*-value is removed. This process is continued until a stable, statistically significant equation is obtained. This regression process yields an eq 1 containing three independent variables. The three variables are atomic partial charge on the carbon atom of the CONHOH moiety (Qco), the globularity of the compound (Glob), and the hydrophilic component of the solvent-accessible surface area (FISA).

$$pIC_{50}(\mu M) = 1.96 + 14.08Qco - 15.73Glob + 0.05FISA$$
$$R = 0.96,\ R^2 = 0.92,\ N = 19,\ F = 59.25,$$
$$t\text{-value}:\ 4.33(Qco),\ -4.02(Glob),\ 9.70(FISA) \quad (1)$$

The squared correlation coefficient (R^2) shows a very good linear relationship between pIC_{50} for the PC-3 cell

line and the three variables. The *F*-statistic shows the overall fit and the *t*-values for the three independent variables to be significant. It is clear that high charge reflecting C=O bond is more polarized is good for activity. For the second term Glob, the lower globularity the stronger of the activity. The experimental and calculated pIC_{50} are plotted in Figure 2. However, no data points with pIC_{50} in the range of 0.5 to 1.5 were available for our data set. Also, the activities of **7** and **8** are distinct from all of the other compounds. It is therefore not clear whether there is a 'points and cluster' effect or what would be the result if we were to remove **7** or **8** from our data set.

The second method that we used to build a QSAR equation is 'forward stepping'. This method starts with an equation containing only one descriptor and adds additional ones in turn. If the added descriptor increases the statistical significance of the equation, it remains in the equation; otherwise it is removed. This procedure continues until a statistically significant equation is obtained. Using this method, we obtained the three-term eq 2. It also shows a very good linear relationship having a squared correlation coefficient R^2 of 0.90. Again, *F*-statistic shows the overall fit as well as the *t*-values for the three independent variables to be significant. It can be seen from the plot of the experimental and calculated pIC_{50} values in Figure 2 that the two equations have a similar correlation with the experimental data.

$$\begin{aligned} pIC_{50}\ (\mu M) &= 0.44 + 422.85ACxDN \\ &\quad - 10.03Glob + 1.26LogP \\ R^2 &= 0.90,\ N = 19,\ F = 46.63, \\ t\text{-value:}&\ 8.377(ACxDN),\ -1.551\ (Glob), \\ &\ 3.884(LogP) \end{aligned} \tag{2}$$

The differences of the descriptors in eqs 1 and 2 imply that there might be some colinearity or multicolinearity of those variables.[14] To investigate this question, it is necessary to check the colinearity and multicolinearity of the variables used in the models. However, it should be pointed out that despite potential colinearity or multicolinearity between the equations, both models are chemically reasonable.

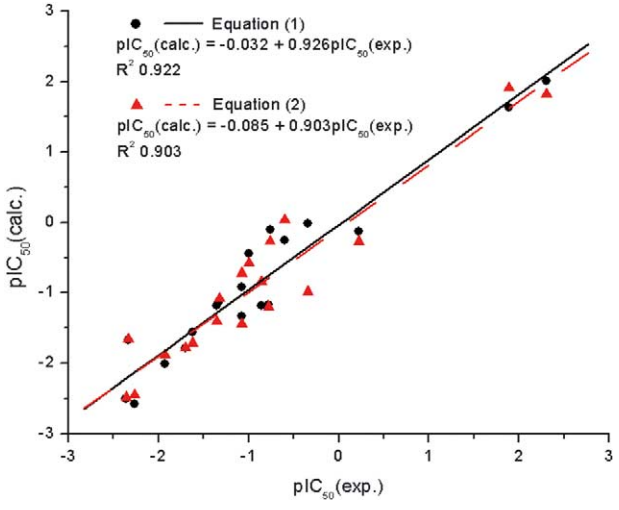


Figure 2. Correlation of experimental versus calculated pIC_{50} using eqs 1 and 2.

3.2. Colinearity and multicolinearity test

Eqs 1 and 2 were assessed for colinearity and multicolinearity. The results are shown in Table 2. While the three variables Qco, Glob and FISA from eq 1 are orthogonal to each other, a larger cross-correlation is found for the descriptors used in eq 2. There is also significant cross-correlation between the descriptors from the two equations, indicating that both equations describe the same physical processes.

After the models passed the colinearity test, a multicolinearity test, shown in Table 3, was performed. It is obvious that eq 1 survives while eq 2 fails this test due to high multicolinearity.

3.3. Predictive ability of the models: leave-one-out test

In the next step, we evaluated the predictive ability of the equations. This test is necessary because a high correlation coefficient *R* only indicates how well the equations fits the data rather than how well it can predict the data. We used the most common scheme of cross-validation, the 'leave-one-out' (LOO) test, that can also test the stability of the model.[12,13] The cross-validated coefficient R^2_{loo}, an indicator of the predictive performance and stability of a model, is recalculated

Table 2. Cross correlation matrix for the independent variables of eqs 1 and 2

	Qco	Glob	FISA	ACxDN	LogP
Qco	1.000				
Glob	0.221	1.000			
FISA	0.241	0.462	1.000		
ACxDN	0.130	0.130	0.821	1.000	
LogP	0.248	0.692	0.043	0.523	1.000

Table 3. Multicollinerearity test for eqs 1 and 2

Equation (1)			R^2
Qco	=	0.181 + 0.153(±0.299)Glob + 0.0FISA	0.074
Glob	=	0.896 + 0.106(btc0.206) − 0.001(±0.000)FISA	0.226
FISA	=	441.182 − 101.971(±156.214)Qco − 329.494(±172.010)Glob	0.230
Equation (2)			
ACxDN	=	0.081 − 0.073(±0.026)Glob − 0.005(±0.001)LogP	0.509
Glob	=	0.962 − 4.447(±1.603)ACxDN − 0.040(±0.007)LogP	0.648
LogP	=	16.467 − 111.321(±27.334)ACxDN − 16.095(±2.966)Glob	0.744

each time after leaving out one data point. R^2_{loo} is defined as:

$$R^2_{loo} = 1 - \sum_{i=1}^{N}(y_i - \tilde{y}_i)^2 / \sum_{i=1}^{N}(y_i - \bar{y})^2 = 1 - \text{PRESS/SSD}$$

Here SSD is the Sum of Squares Deviation for each actual activity value y_i from the average activity y over the entire data set. PRESS, the Predictive Sum of Squares, is the sum of the squared differences between the actual activity y_i and the predicted activity $\tilde{y}_i$ when compound i is not included in the regression. If R^2_{loo} is close to 1, the equation has predictive ability.[15] The calculated values of PRESS, SSD, and R^2_{loo} for eqs 1 and 2 are given in Table 4. It is obvious that both models have good predictive ability. In both cases, R^2_{loo} is larger than 0.7. It is also worth mentioning that the signs of all of the descriptors in these equations remain the same. No large numerical changes of the coefficients of these descriptors are observed. Thus, our models have significant predictive ability.

3.4. Points and cluster

As mentioned above, no compounds with pIC_{50} values between 0.5 and 1.5 were available for the training set. Thus, the two points with pIC_{50} values above 1.5 are quite isolated. In order to test whether elimination of these two points would significantly change the QSAR equation due to a 'point and cluster' effect, we recalculated the equation without **7** and **8**. This modification caused compounds **4** and **6** to become outliers that were then excluded from the data set also. This resulted in eq 3 based on 15 data points.

$$pIC_{50}\ (\mu M) = -3.818 + 9.191\text{Qco} - 5.697\text{Glob} + 0.038\text{FISA}$$

$$R = 0.952,\ R^2 = 0.907,\ N = 15,\ F = 35.81,$$

$$t\text{-value}:\ 4.660(\text{Qco}),\ -2.167(\text{Glob}),\ 10.063(\text{FISA}) \quad (3)$$

Eq 3 is similar to eq 1 but with some changes of coefficients. These changes are not surprising considering almost 30% of the data were eliminated in the regression. A plot correlating experimental versus calculated pIC_{50} using eqs 1 and 3 shows that there is excellent agreement between the predictive abilities of eqs 1 and 3 (Fig. 3). It is also obvious that one may predict compounds having pIC_{50} larger than zero by extrapolating through eq 3. New compounds in this activity range are currently under investigation and will improve the test set to further evaluate the usefulness and limitations of the models.

Table 4. Cross-validation results of eqs 1 and 2

Equation	Descriptors	N	PRESS	SSD	R^2_{loo}
1	Qco, Glob, FISA	19	3.690	28.443	0.870
2	ACxDN, Glob, LogP	19	4.364	28.443	0.847

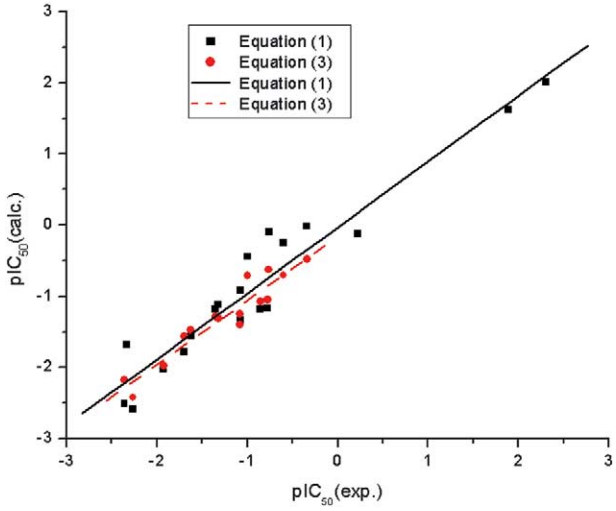


Figure 3. Comparison of correlation of experimental versus calculated pIC_{50} using eqs 1 and 3.

4. Conclusion

The first QSAR model, eq 1, for inhibition of proliferation of the PC-3 cell line by TSA-like hydroxamic acids has been developed. It indicates that the atomic charge value on the carbonyl carbon in the CONHOH moiety Qco, the shape of the compound Glob, and hydrophilicity of the compound have significant contributions to the PC-3 cell line inhibition activity. The cross-validation test and leave-one-out test show that the model does have good predictive ability and is statistically significant. This simple equation can be used to estimate the PC-3 inhibition activity for compounds prior to synthesis by calculating these three independent terms. Based on the results shown here, compound **6** and analogues have been chosen for further preclinical development.

Acknowledgements

Additional support for computational resources from the National Science Foundation (DMR0079647) and the Walther Cancer Research Center is gratefully acknowledged.

References and notes

1. (a) Johnstone, R. W. *Nat. Rev. Drug Discov.* **2002**, *1*, 287. (b) Kramer, O. H.; Gottlicher, M.; Heinzel, T. *Trends in Endocrinol. and Metabolism* **2001**, *12*, 294. (c) Meinke, P.; Liberator, P. *Curr. Med. Chem.* **2001**, *8*, 211.
2. Patra, S. K.; Patra, A.; Dahiya, R. *Biochem. Biophy. Res. Commun.* **2001**, *287*, 705.
3. Han, J. W.; Ahn, S. H.; Park, S. H.; Wang, S. Y.; Bae, G. U.; Seo, D. W.; Known, H. K.; Hong, S.; Lee, Y. W.; Lee, H. W. *Cancer Res.* **2000**, *60*, 6068.
4. Kijima, M.; Yoshida, M.; Suguta, K.; Horinouchi, S.; Beppu, T. *J. Biol. Chem.* **1993**, *268*, 22429.
5. Kwon, H. J.; Owa, T.; Hassig, C. A.; Shimada, J.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3356.

6. A recent review of HDAC inhibitors: Miller, T. A.; Witter, D. J.; Belvedere, S. *J. Med. Chem.* **2003**, *46*, 5096 and references therein.
7. Yoshida, M.; Kijima, M.; Akita, M.; Beppu, T. *J. Biol. Chem.* **1990**, *265*, 17174.
8. Richon, V. M.; Emiliani, S.; Verdin, E.; Webb, Y.; Breslow, R.; Rifkind, R. A.; Marks, P. A. *Proc. Natl. Acad. Sci. U.S.A* **1998**, *95*, 3003.
9. Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. *Nature* **1999**, *401*, 188.
10. Lan-Hargest, H.-Y.; Kaufman, R. J.; Wiech, N. L. US Patent Appl. US2002143196.
11. (a) Qikprop 1.6, Schrodinger Inc. 2001. For publications, see Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878. (b) Jorgensen, W. L.; Duffy, E. M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155.
12. (a) Stewart, J. J. P. *J. Comp. Chem.* **1989**, *10*, 209. (b) Stewart, J. J. P. *J. Comp. Chem.* **1989**, *10*, 221.
13. Gaussian 98, Revision A.9, M. J. Frisch, et al. Gaussian, Inc., Pittsburgh PA, 1998.
14. Livingstone, D. *Data Analysis for Chemists: Applications to QSAR and Chemical Product Design*; Oxford University Press Inc: New York, 1995.
15. Chatterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*; Wiley: New York, 2000.