

Ứng dụng phương pháp thiết kế và phân tích dữ liệu *In silico* (ISIDA) trong thiết kế dẫn chất acid hydroxamic mới ức chế HDAC2

Phạm Thế Hải¹, Đào Thị Kim Oanh¹,
Đoàn Việt Nga², Bùi Thanh Tùng², Lê Thị Thu Hương^{2,*}

¹Đại học Dược Hà Nội, 13-15 Lê Thánh Tông, Hà Nội, Việt Nam

²Khoa Y dược, Đại học Quốc gia Hà Nội, 144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam

Nhận ngày 15 tháng 9 năm 2017

Chỉnh sửa ngày 20 tháng 10 năm 2017; Chấp nhận đăng ngày 01 tháng 12 năm 2017

Tóm tắt: Nghiên cứu thuốc điều trị ung thư mới đang là mối quan tâm hàng đầu của thế giới. Một trong số các mục tiêu phân tử đáng chú ý là enzym Histon Deacetylase 2 (HDAC2). Hiện nay, khoa học máy tính ngày càng được sử dụng rộng rãi trong tìm kiếm và phát triển thuốc mới. Trong nghiên cứu này, mô tả mảnh cấu trúc ISIDA được tính toán dựa trên khung cấu trúc acid hydroxamic, nhằm xây dựng mô hình toán học mô tả mối tương quan định lượng giữa cấu trúc và hoạt tính ức chế HDAC2 (QSAR) áp dụng kỹ thuật hồi quy tuyến tính đa biến (MLR) và thuật giải di truyền. Kết quả, chúng tôi đã xây dựng thành công ba mô hình QSAR trên ba tập hợp mảnh cấu trúc, có hệ số xác định $R^2 > 0,7$. Từ các mô hình này, các mảnh cấu trúc quan trọng được xác định và ứng dụng trong thiết kế các dẫn chất acid hydroxamic mới có hoạt tính ức chế HDAC2. Kết quả của nghiên cứu góp phần định hướng cho quá trình tổng hợp và thử hoạt tính trong tìm kiếm liệu pháp mới điều trị ung thư.

Từ khóa: Histon deacetylase 2; mối tương quan định lượng giữa cấu trúc và hoạt tính (QSAR); mô tả phân tử ISIDA; hồi quy tuyến tính đa biến (MLR); Acid hydroxamic.

1. Đặt vấn đề

HDAC2 là enzym quan trọng thuộc HDAC nhóm I tham gia vào quá trình sao chép, kiểm soát trạng thái acetyl hoá và chức năng của nhiều protein trong tế bào chất và trong những yếu tố phiên mã khác [1]. Nhiều nghiên cứu cho thấy HDAC2 biểu hiện quá mức trên nhiều loại ung thư: ung thư vú [2], ung thư đại tràng [3], ung thư dạ dày, biểu mô tế bào [4, 5], ung thư biểu mô phổi [6]. Chính vì vậy, HDAC2 đã

trở thành một mục tiêu phân tử tiềm năng trong tìm kiếm thuốc điều trị ung thư mới [1].

Trong các nhóm chất ức chế HDAC2, các acid hydroxamic là nhóm chất được quan tâm nhiều nhất vì có cấu trúc đơn giản, dễ tổng hợp và có nhóm -NHOH tạo được phức bền với Zn^{2+} ở trung tâm hoạt động của HDAC2, nhờ đó có hoạt tính ức chế enzym tốt. Cục quản lý thực phẩm và dược phẩm Hoa Kỳ (FDA) đã phê duyệt một acid hydroxamic ức chế HDAC điển hình là vorinostat (thường gọi là SAHA) vào năm 2006 sử dụng trong điều trị u lympho da tế bào [7].

Trong nghiên cứu này khung hydroxamic được lựa chọn làm định hướng thiết kế các hợp chất ức chế HDAC2 mới. Mô hình về tương

* Tác giả liên hệ. ĐT.: 84-24-38387949.

Email: ltthuong1017@gmail.com

<https://doi.org/10.25073/2588-1132/vnumps.4073>

quan định lượng giữa cấu trúc - hoạt tính (QSAR) được xây dựng giúp dự đoán chính xác hoạt tính ức chế HDAC2 của hợp chất mới thiết kế. Chúng tôi ứng dụng phương pháp thiết kế dựa trên mảnh cấu trúc (FBDD - *Fragment based drug design*) dựa trên mô tả phân tử ISIDA (*In Silico design and Data Analysis*) để tổ hợp ra công thức hoá học của acid hydroxamic mới, dự đoán hoạt tính của chúng từ đó tìm ra một số hợp chất có hoạt tính mạnh nhất, là tiền đề cho các nghiên cứu tổng hợp trong tương lai.

2. Nguyên liệu và phương pháp nghiên cứu

2.1. Nguyên liệu

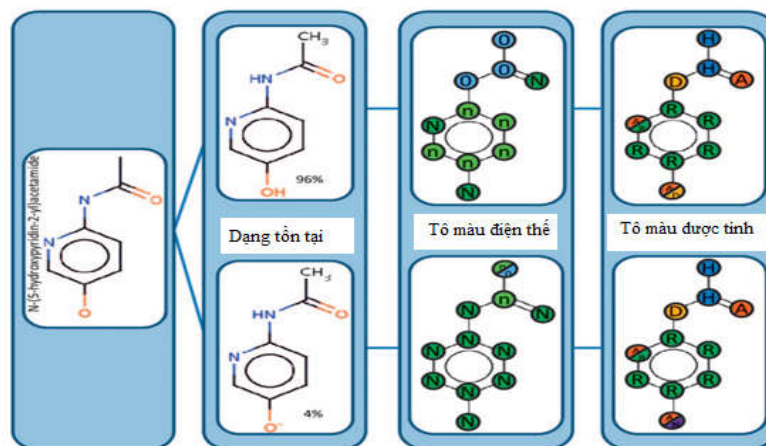
Cơ sở dữ liệu (CSDL) cho xây dựng mô hình bao gồm 45 hợp chất có cấu trúc acid hydroxamic ức chế HDAC2 với giá trị IC_{50} (được chuyển đổi thành $lgIC_{50}$, đơn vị μM) thu thập từ các bài báo khoa học trong nước và quốc tế đã công bố [7-12]. Hoạt tính được xác định trong cùng điều kiện thí nghiệm (sử dụng Kit định lượng Bioscience [13]). Các giá trị IC_{50} ứng với mỗi hợp chất được hiệu chỉnh, lấy giá trị trung bình sau khi loại bỏ các giá trị nằm ngoài khoảng giới hạn [giá trị trung bình \pm độ lệch chuẩn (SD)].

2.2. Phương pháp

Mô hình QSAR biểu diễn mối tương quan định lượng cấu trúc và tác dụng sinh học của các hợp chất dưới biểu thức toán học, có công thức toán học: $Y = a_1x_1 + a_2x_2 + \dots + a_nx_n$ [14]. Trong nghiên cứu này, Y là biến phụ thuộc ($lgIC_{50}$), tương ứng với hoạt tính; x_1, x_2, \dots, x_n là các mô tả phân tử đặc trưng cho cấu trúc; a_1, a_2, \dots, a_n là các trọng số tương ứng với các biến độc lập.

2.2.1. Tính toán mô tả phân tử ISIDA

Các cấu trúc trong CSDL được lưu dưới dạng tệp cấu trúc (SDF-*Structure Data File*). Phương pháp ISIDA được sử dụng trong tính toán mô tả phân tử và phân tích dữ liệu. Phần mềm ISIDA Fragmentor2015 được áp dụng [15], có ưu điểm là tích hợp số lượng lớn và đa dạng mảnh cấu trúc có mặt trong phân tử, cho phép tùy biến tạo mảnh cấu trúc mới tùy theo mục đích. Phương pháp này sử dụng các kí hiệu dán nhãn mang đặc tính mảnh cấu trúc để xây dựng mô hình QSAR và thiết kế hợp chất mới. Các mảnh cấu trúc được biểu diễn thông qua cơ chế mô tả theo màu (kí hiệu, tính thân dầu, độ pH,...) đảm bảo liên quan giữa màu sắc và tính chất hóa học; gán nhãn dựa trên giá trị điện thế và đặc tính dược học (Hình 1).



Hình 1. Biểu diễn công thức

N-(5-hydroxypyridin-2-yl) acetamide theo chương trình tô màu và gán nhãn. Phân tử đầu tiên được xác định các dạng đồng phân, tô màu theo nguyên tố ở các đỉnh; sau đó kết hợp với dán nhãn dựa trên độ âm điện; cuối cùng là kết hợp với đặc tính dược học.

Bước cuối cùng là sử dụng ngôn ngữ lập trình Pascal tìm kiếm và đếm sự có mặt các mảnh cấu trúc đã đánh dấu trong phân tử.

2.2.2. Thiết kế tập huấn luyện và kiểm tra

Tập huấn luyện (TS-training set) bao gồm 30 phân tử được lựa chọn theo tiêu chí giá trị IC_{50} nằm trong khoảng 0,008 đến 35 μM được sử dụng xây dựng mô hình và tập kiểm tra (PS-test set) với 15 phân tử còn lại để đánh giá khả năng dự đoán chính xác của các mô hình đã xây dựng được.

2.2.3 Phát triển mô hình

Sử dụng kỹ thuật phân tích hồi quy tuyến tính đa biến (MLR-Multiple Linear Regression) trong chương trình STATISTICA v8.1 để xây dựng mô hình tương quan hoạt tính ức chế với mô tả phân tử ISIDA.

2.2.4. Đánh giá mô hình

Đánh giá nội trên tập TS cho biết độ khớp, độ mạnh của mô hình thông qua các thông số thống kê: r^2 - hệ số tương quan, $0 < r^2 < 1$, trong phương trình hồi quy tuyến tính đa biến giá trị r^2 cần đạt trên 0,70; sai số toàn phương trung bình (RMSE-root mean square errors); sai số tuyệt đối (MAE-root mean square errors); tính phù hợp của hệ số tương quan (CCC-concordance correlation coefficient), giá trị có

ý nghĩa thống kê phải cao hơn 0.85 [14]; F là phân phối Fisher, r^2 càng lớn thì F càng lớn.

Bên cạnh đó, độ ổn định của mô hình thông qua đánh giá chéo trên tập TS bằng các hệ số q^2_{LOO} (hệ số leave-one-out), q^2_{LMO} (hệ số leave-more-out) [14].

Đánh giá ngoại mô hình MLR trên tập PS để kiểm tra khả năng dự đoán chính xác: $RMSE_{ext}$ và MAE_{ext} , Q^2F_1 , Q^2F_2 và Q^2F_3 (giá trị ngưỡng cho phép là 0,7), CCC_{ext} và rm^2 (giá trị ngưỡng thấp nhất là 0,5) [16].

Cuối cùng, mô hình MLR được đánh giá tương quan ngẫu nhiên tương phản bằng kiểm tra ngẫu nhiên Y với 2 hệ số r^2_{scr} và q^2_{scr} . Nếu đỉnh trục Y không vượt quá 0,3-0,4 đối với r^2_{scr} và 0.05 đối với q^2_{scr} thì mô hình tự do trong tương quan, các biến độc lập không ảnh hưởng lẫn nhau và ảnh hưởng đến kết quả dự đoán của mô hình. Quá trình tính toán giá trị r^2_{scr} và q^2_{scr} được lặp lại 2000 lần.

3. Kết quả và bàn luận

Mô hình MLR: Đối với mỗi tập hợp mô tả phân tử, chúng tôi thu được một mô hình tương ứng như sau:

M1

$$\log IC_{50} = 0.130 \times V_1 - 5.408 \times V_2 - 2.047 \times V_3 + 6.466 \times V_4 + 2.011 \times V_5 + 2.179 \times V_6 + 0.246 \times V_7 - 0.162 \times V_8$$

M2

$$\log IC_{50} = 1.573 - 5.589 \times V_9 - 1.021 \times V_{10} - 0.907 \times V_{11} + 6.190 \times V_{12} + 2.501 \times V_{13} - 1.542 \times V_{14} + 1.056 \times V_{15} + 2.036 \times V_{16}$$

M3

$$\log IC_{50} = -4.160 - 4.348 \times V_{17} - 1.371 \times V_{18} + 3.336 \times V_{19} + 1.558 \times V_{20} + 0.487 \times V_{21} - 2.330 \times V_{22} + 2.416 \times V_{23} + 0.332 \times V_{24}$$

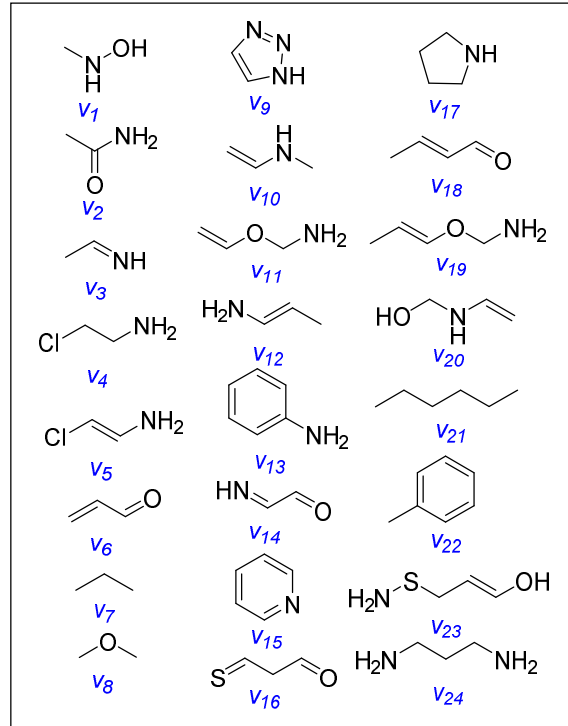
Các biến $V_1 - V_{24}$ tương ứng với từng mô hình là vân tay cấu trúc đại diện cho các mảnh cấu trúc được mã hoá thành tham số nhận các giá trị 0 và 1 ứng với sự xuất hiện của chúng trong phân tử.

Đánh giá mô hình QSAR: Đầu tiên mô hình QSAR xây dựng được thỏa mãn yêu cầu đầu tiên là dựa trên đích phân tử xác định đó chính là HDAC2, thuật toán xây dựng và đánh

giá cụ thể rõ ràng hoàn toàn tái xây dựng được cho hợp chất mới. Về đánh giá độ ổn định, chính xác của mô hình ta xét các thông số thống kê dưới đây:

Đánh giá nội trên tập TS

Các giá trị thực nghiệm và kết quả dự đoán của mô hình MLR cho tập TS được đưa ra trong bảng 2 dưới đây.



Hình 2. Các mảnh cấu trúc quan trọng được lựa chọn trong mô hình **M1**, **M2** và **M3** phản ánh mối tương quan định lượng giữa cấu trúc và hoạt tính sinh học ức chế enzyme HDAC2.

Bảng 1. Các thông số đánh giá chéo và độ thích hợp của các mô hình tuyến tính trên tập TS

	r^2	Q^2_{LOO}	RMSE	MAE	CCC	$r^2(scr)$	$q^2(scr)$	F
M1	0,81	0,76	0,52	0,41	0,89	0,13	-0,34	17,8
M2	0,79	0,77	0,49	0,38	0,85	0,14	-0,31	18,5
M3	0,82	0,80	0,54	0,43	0,90	0,13	-0,35	19,0

Xét tổng quát ở cả 3 mô hình đều cho giá trị r^2 lớn hơn 0,70 do đó các mô hình xây dựng được đều đạt độ tuyến tính tốt trong đó mô hình **M3** cho giá trị cao nhất là 0,82. Giá trị q^2_{LOO} ở mô hình **M1** là 0,76 và khi thay đổi các biến từ V_2-V_8 theo xu hướng kéo dài mạch C, tăng dần độ bất bão hòa, thêm các nguyên tố có độ âm điện cao như Cl, O, N, S ta được mô hình **M2**, **M3** với q^2_{LOO} tương ứng tăng 0,01 và 0,04. Bên cạnh đó, thông số kiểm tra ngẫu nhiên Y cho các mô hình **M1**, **M2**, **M3** với giá trị $r^2(scr)$ cao nhất là 0,14 (**M2**) và $q^2(scr)$ cao nhất là -0,34 với mô hình **M1**; kết quả này chứng tỏ cả 3 mô hình không mô tả tương quan ngẫu nhiên giữa biến độc lập và biến phụ thuộc. Ngoài ra, các

mô hình đều có ý nghĩa thống kê khi giá trị CCC đều không thấp hơn 0,85. Từ đó chứng tỏ độ mạnh, ổn định và độ khớp cao của cả 3 mô hình MLR.

Đánh giá ngoại trên tập PS

Bảng 2. Tính toán khả năng dự đoán của mô hình tuyến tính trên tập PS

	Q^2F_1	Q^2F_2	Q^2F_3	RMSE _{ext}	MAE _{ext}	CCC
M1	0,77	0,78	0,81	0,39	0,31	0,84
M2	0,76	0,79	0,76	0,40	0,32	0,80
M3	0,80	0,79	0,82	0,37	0,29	0,85

Dựa trên các chỉ tiêu thông số đánh giá đã nêu, cả ba mô hình đều đạt yêu cầu thậm chí giá trị Q^2 rất cao, mô hình **M3** cho kết quả tốt nhất

với Q^2F_1 , Q^2F_2 , Q^2F_3 lần lượt là 0,80, 0,79, 0,82. Các giá trị này chứng tỏ khả năng dự đoán tốt của cả 3 mô hình nhất là mô hình **M3**.

Giải thích cơ chế

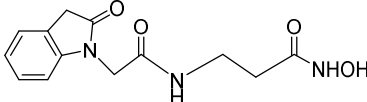
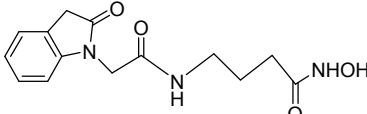
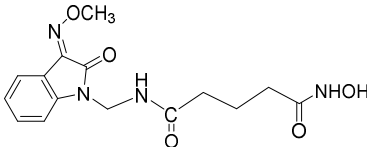
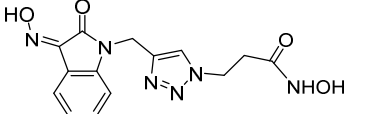
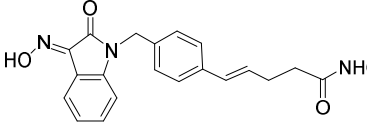
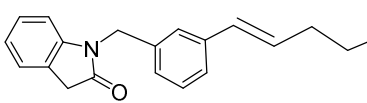
Đây là bước quan trọng để từ đó định hướng thiết kế hợp chất mới. Cả ba mô hình **M1**, **M2**, **M3** xây dựng được trên 3 tập mảnh cấu trúc khác nhau đều cho kết quả tốt, từ đó ta hoàn toàn tìm ra được các biến có ảnh hưởng tới hoạt tính sinh học. Hình 2 là các mô tả mảnh cấu trúc được sử dụng trong các mô hình QSAR. Các mảnh có hệ số nhận giá trị âm là các yếu tố tăng cường hoạt động ức chế (giảm giá trị IC_{50}). Ngược lại, hệ số có giá trị dương ứng với các mảnh cấu trúc làm giảm hoạt tính ức chế.

Kết hợp cả 3 mô hình, 10 mảnh cấu trúc giữ vai trò quan trọng nhất đóng góp vào hoạt tính ức chế HDAC2 bao gồm *V2*, *V3*, *V8*, *V9*, *V10*, *V11*, *V14*, *V17*, *V18*, *V22* (Hình 2).

Đây là các mảnh cấu trúc được sử dụng trong thiết kế nhóm nhận diện và cầu nối dựa trên cấu trúc cố định nhóm gắn với kẽm acid hydroxamic để tăng cường khả năng gắn kết và hoạt tính ức chế HDAC2. Tổ hợp các mảnh cấu trúc này để tạo thành các hợp chất mới một cách tự động bằng Thuật giải di truyền giúp tạo ra hàng nghìn hợp chất [17]. Các khung cấu trúc này được lọc lần thứ nhất bằng việc kiểm tra xem đã từng được công bố chưa bằng cách tìm cấu trúc đồng dạng theo SMILES trên trang web <https://scifinder.cas.org/>.

Bước lọc thứ hai là dự đoán hoạt tính (IC_{50}) của các hợp chất được thiết kế bằng chính những mô hình đã xây dựng được. Sử dụng phần mềm ISIDA Fragmentor2015 tính toán các mô tả phân tử và thay vào các mô hình **M1**, **M2** và **M3**, lấy trung bình cộng các kết quả dự đoán. Từ các cấu trúc mới đã tính toán, lọc ra các cấu trúc có giá trị IC_{50} dự đoán tốt và khả thi trong tổng hợp (Bảng 3).

Bảng 3. Các khung cấu trúc mới được đề xuất cho thiết kế

STT	Cấu trúc	Dự đoán IC_{50} (μ M)
1a		1,60125
2a		3.10026
3a		0,96005
4a		0,30018
5a		1,11231
6a		4,62709

4. Kết luận và Kiến nghị

Như vậy, dựa trên cơ sở lý thuyết của mô tả mảnh cấu trúc ISIDA, nghiên cứu đã xây dựng thành công 3 mô hình QSAR có khả năng dự đoán giá trị IC_{50} trong ức chế hoạt động của enzyme HDAC2. Ngoài ra nghiên cứu đã tổ hợp 10 mảnh cấu trúc trên nền khung acid hydroxamic nhằm tạo ra một số khung dẫn xuất mới, được dự đoán là có tác dụng ức chế mạnh trên HDAC2. Đặc biệt nghiên cứu này đề xuất khung cấu trúc dẫn chất **4a** được với hoạt tính mạnh, có đặc điểm hoá học không quá khó để tổng hợp, là tiền đề cho các nghiên cứu tìm kiếm hợp chất dẫn đường mới trong tương lai.

Lời cảm ơn

Nghiên cứu được thực hiện trong khuôn khổ Đề tài khoa học và công nghệ cấp ĐHQGHN, mã số QG.16.24.

Tài liệu tham khảo

- [1] C. Choudhary, et al., Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325 (2009) 834.
- [2] M. Dokmanovic, et al., Histone deacetylase inhibitors: overview and perspectives. *Molecular Cancer Research* 67 (2007) 3145.
- [3] P. Zhu, et al., Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. *Cancer cell* 5 (2004) 455.
- [4] J. Song, et al., Increased expression of histone deacetylase 2 is found in human gastric cancer. *APMIS* 113 (2005) 264.
- [5] B.H. Huang, et al., Inhibition of histone deacetylase 2 increases apoptosis and p21Cip1/WAF1 expression, independent of histone deacetylase 1. *Cell Death and Differentiation* 12 (2005) 395.
- [6] A.C. Borczuk, et al., Non-Small-Cell Lung Cancer Molecular Signatures Recapitulate Lung Developmental Pathways. *American Journal of Pathology* 163 (2003) 1949.
- [7] D.T.M. Dung, et al., Exploration of novel 5'(7')-substituted-2'-oxospiro[1,3]dioxolane-2,3'-indoline-based N-hydroxypropenamides as histone deacetylase inhibitors and antitumor agents. *Arabian Journal of Chemistry* 10(2015) 465.
- [8] D.T.M. Dung, et al., Novel 3-substituted-2-oxoindoline-based N-hydroxypropenamides as Histone Deacetylase Inhibitors and Antitumor Agents. *Medicinal Chemistry* 11(2015) 725-735
- [9] P. Chun, et al., Synthesis and Importance of Bulky Aromatic Cap of Novel SAHA Analogs for HDAC Inhibition and Anticancer Activity. *Bulletin of the Korean Chemical Society* 32 (2011) 1891.
- [10] M.J. Lai, et al., Synthesis and Biological Evaluation of 1-Arylsulfonyl-5-(N-hydroxyacrylamide)indoles as Potent Histone Deacetylase Inhibitors with Antitumor Activity in Vivo. *Medicinal Chemistry* 55 (2012) 3777.
- [11] F. Thaler, et al., Synthesis and biological characterization of spiro[2H-(1,3)-benzoxazine-2,40-piperidine] based histone deacetylase inhibitors. *European Journal of Medicinal Chemistry* 64 (2013) 273.
- [12] T.T.L. Huong, et al., Novel 2-oxoindoline-based hydroxamic acids: Synthesis, cytotoxicity and inhibition of histone deacetylation. *Tetrahedron Letters* 56 (2015) 6425.
- [13] N. Kandakatla and G. Ramakrishnan, Ligand Based Pharmacophore Modeling and Virtual Screening Studies to Design Novel HDAC2 Inhibitors. *Advances in Bioinformatics*, 2014. ID 812148.
- [14] T. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, WILEY-VCH, Betzdruck GmbH, D-63291 Darmstadt, 2000.
- [15] A. Varnek, et al., ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. *Current Computer-Aided Drug Design* 4 (2008) 191.
- [16] V. Consonni and D. Ballabio, Comments on the definition of the Q^2 parameter for QSAR validation. *Journal of Chemical Information and Modeling* 49 (2009) 1669.
- [17] N.T.T. Hoài and P.T.T.Q. Chiến, Thuật giải di truyền và ứng dụng. *Tuyển tập Báo cáo Hội nghị Sinh viên nghiên cứu khoa học lần thứ 6* (2008) 266.

Application of in Silico Fragment and Data Analysis (ISIDA) Approaches for the Design of Novel Hydroxamic Acids Targeting HDAC2

Pham The Hai¹, Dao Thi Kim Oanh¹, Doan Viet Nga²,
Bui Thanh Tung², Le Thi Thu Huong²

¹*Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hanoi, Vietnam*

²*VNU School of Medicine and Pharmacy, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam*

Abstract: Finding a new treatment for cancer is one of the most interested fields for pharmaceutical research worldwide. Enzyme histone deacetylase 2 (HDAC2), being a member of HDAC class I appear to be an important druggable target. This study focused on rational design of novel HDAC2 inhibitors using molecular descriptors derived from ISIDA fragmentor methodology. Quantitative structure-activity relationship was explored to develop mathematical models able to predict HDAC2 inhibitory bioactivity of acid hydroxamic derivatives. Multiple linear regression (MLR) algorithm implemented in STATISTICA 8.0 was used for model development. Consequently 3 QSAR models were obtained showing acceptable performance $r^2 > 0.70$ for further use. Based on these models 10 important fragments attributing to better inhibitory potency were identified. Finally several novel hydroxamic derivatives were designed and screened for HDAC2 inhibitory activity.

Keywords: Histone Deacetylase 2; Quantitative Structure-Activity Relationship (QSAR); molecular descriptors ISIDA; multiple linear regression (MLR); acid hydroxamic.