

# Xây dựng mô hình liên quan định lượng giữa cấu trúc và tác dụng ứng dụng trong sàng lọc tìm kiếm chất ức chế histon deacetylase

Phạm Thế Hải<sup>1</sup>, Đoàn Việt Nga<sup>2</sup>, Nguyễn Thị Ngọc<sup>1</sup>,  
Bùi Thanh Tùng<sup>2</sup>, Lê Thị Thu Hương<sup>2,\*</sup>

<sup>1</sup>Đại học Dược Hà Nội, 13-15 Lê Thánh Tông, Hà Nội, Việt Nam

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

---

## Tóm tắt

Ung thư hiện nay là mối quan tâm hàng đầu trên thế giới với rất nhiều nghiên cứu đã được thực hiện nhằm tìm kiếm hợp chất có khả năng điều trị trúng đích. Enzyme histon deacetylase (HDAC) được đánh giá là một trong những đích phân tử quan trọng nhất. Trong nghiên cứu này, chúng tôi đã phát triển và đánh giá mô hình QSAR để xác định hoạt tính của các hợp chất có khả năng ức chế HDAC2 chỉ dựa trên cấu trúc phân tử. Nghiên cứu sử dụng mô tả phân tử DRAGON, kỹ thuật LDA với phần mềm STATISTICA 10.0 để xây dựng mô hình. Mô hình này được đánh giá với độ chính xác hơn 86% và đảm bảo các nguyên tắc OECD. Phương pháp này giúp cho quá trình nghiên cứu phân loại, sàng lọc được nhanh, giảm chi phí và tăng hiệu quả. Ngoài ra, các kết quả đạt được tạo cơ sở cho việc phát hiện và thiết kế hợp chất mới ức chế HDAC2, hướng tới phát triển thành thuốc điều trị ung thư nhắm đích.

Nhận ngày 16 tháng 9 năm 2015, Chính sửa ngày 07 tháng 10 năm 2015, Chấp nhận đăng ngày 05 tháng 12 năm 2016

Từ khóa: Histone deacetylase 2; định lượng mối quan hệ giữa cấu trúc và hoạt tính; mô tả phân tử DRAGON; LDA; STATISTICA.

---

## 1. Đặt vấn đề

Trên thế giới hiện nay, đẩy lùi căn bệnh ung thư luôn là vấn đề được đặt lên hàng đầu. Đã có rất nhiều nghiên cứu trước đây nhằm tìm kiếm các hợp chất điều trị nhắm đích phân tử. Trong nghiên cứu này, chúng tôi đã xác định và lựa chọn đích phân tử là enzyme histone deacetylase 2 (HDAC2), một enzyme quan trọng thuộc họ HDAC có chức năng loại bỏ nhóm nhóm acetyl từ  $\epsilon$ -N-acetyl lysin amino acid của phần histon trong cấu trúc của gen

[1]. Việc ức chế enzyme HDAC2 sẽ giúp ngăn chặn quá trình tổng hợp ADN từ đó ngăn chặn tế bào ung thư phát triển [2]. Tại Việt Nam, đã có một số nghiên cứu tìm kiếm các hợp chất mới có khả năng ức chế HDAC2 và đưa ra một hướng mới trong tìm kiếm các thuốc điều trị ung thư [3-7]. Tuy nhiên, các phương pháp này còn tốn nhiều thời gian và kinh phí. Vì vậy, để tiết kiệm và nâng cao hiệu quả nghiên cứu, chúng tôi đã tính toán các tham số phân tử dựa vào đặc điểm cấu trúc và từ đó xây dựng thành công mô hình định lượng giữa cấu trúc và hoạt tính sinh học (QSAR - *quantitative structure activity relationship*). Trên cơ sở đó, việc sàng lọc và định hướng phát triển hợp chất ức chế

---

\* Tác giả liên hệ. ĐT: 84-4-37450188  
Email: ltthuong1017@gmail.com

HDAC2 trở nên dễ dàng và hiệu quả hơn, tạo tiền đề cho bước nghiên cứu hợp chất chữa ung thư nhắm đích tiếp theo.

## 2. Nguyên liệu và phương pháp nghiên cứu

### 2.1. Nguyên liệu

Cơ sở dữ liệu ChemBL và thông tin từ các bài báo cáo khoa học có liên quan trước đây được khai thác và kiểm tra để chuẩn bị cơ sở dữ liệu (CSDL) đủ lớn và đa dạng các chất có hoạt tính ức chế HDAC2. CSDL đầu tiên gồm 1181 hợp chất có hoạt tính sinh học được xác định qua nồng độ ức chế 50% đối tượng thử ( $IC_{50}$  - *inhibitor concentration 50%*), hằng số ức chế ( $K_i$ ) hoặc thông số khác. Sau đó, 184 hợp chất được thêm vào có hoạt tính ức chế cả HDAC1 và HDAC2. Cuối cùng, CSDL được xây dựng bao gồm 1471 hợp chất trong đó 715 hợp chất có hoạt tính ức chế HDAC2 và 756 hợp chất giống thuốc [8] nhưng không có hoạt tính ức chế.

### 2.2. Phương pháp

Mô hình QSAR biểu diễn mối liên hệ giữa cấu trúc và tác dụng sinh học của các chất dưới biểu thức toán học. Để xây dựng được các mô hình này thì cả cấu trúc và tác dụng sinh học đều phải được định lượng hóa. Mô hình QSAR có dạng:  $Y = a_1(x_1) + a_2(x_2) + \dots + a_n(x_n)$ . Trong đó, Y là biến đáp ứng (mang 2 giá trị: +1 đối với các hợp chất có tác dụng ức chế HDAC2 và -1 đối với các hợp chất không có tác dụng ức chế HDAC2);  $x_1, x_2, \dots, x_n$  là các tham số phân tử đặc trưng cho cấu trúc;  $a_1, a_2, \dots, a_n$  là các hằng số thể hiện trọng số của từng tham số phân tử  $x_i$  đối với tác dụng sinh học.

#### 2.2.1. Tính toán tham số mô tả phân tử

Tham số mô tả phân tử là kết quả của một quá trình toán học và logic chuyển đổi thông tin được mã hóa trong cấu trúc hóa học thành một số có khả năng đặc trưng cho phân tử đầy [9]. Để thực hiện tính toán, tất cả các cấu trúc của CSDL được chuyển đổi thành hệ thống đơn giản hóa với đầu vào được biểu diễn dưới dạng

dòng SMILES (*Simplified Molecular Input Line Entry System*) bằng việc sử dụng phần mềm ChemDraw v.10.0 [10]. Căn cứ vào các mã code, hơn 2000 mô tả phân tử thuộc họ 0D, 1D và 2D được tính toán bằng phần mềm Dragon v.6.0 [11]. Để tính toán mô tả phân tử, tối ưu hóa cấu trúc bán thực nghiệm đã được thực hiện cho mỗi hợp chất. Quy trình này được thực hiện dựa trên các thuật toán Polak - Ribiere sử dụng trong HyperChem v.8.0 [12]. Thuật toán tối thiểu hóa năng lượng thông qua việc thay đổi cấu trúc không gian phân tử để đạt năng lượng thấp nhất, tăng độ bền cho hợp chất [13].

#### 2.2.2. Thiết kế tập huấn luyện và tập kiểm tra

Sau khi tính toán mô tả phân tử, CSDL được chia thành tập huấn luyện (TS - *training set*) và tập kiểm tra (PS - *prediction set*), sử dụng phương pháp phân tích cụm k-MCA (*k-mean cluster analysis*) với phần mềm STATISTICA 10.0 [14]. Tập huấn luyện (chiếm 75% CSDL) được sử dụng cho việc xây dựng (huấn luyện) mô hình. Tập huấn luyện cho phép phát hiện mối quan hệ có khả năng dự đoán. Còn tập kiểm tra bao gồm các phân tử mà mô hình chưa bao giờ gặp được sử dụng để đánh giá độ mạnh và khả năng ngoại suy của các mô hình xây dựng được.

#### 2.2.3. Phát triển mô hình

Nghiên cứu sử dụng kỹ thuật phân tích sự khác biệt tuyến tính (LDA - *linear discriminant analysis*) để xác định mối quan hệ giữa tham số phân tử và hoạt tính ức chế.

*Lựa chọn tham số phân tử:* Kỹ thuật đưa vào dần từng bước các biến (FS - *Forward stepwise*) được sử dụng để lựa chọn bước đầu tập hợp các tham số phân tử. Đây là thuật toán kết hợp dần dần với một biến mới dựa trên trị số  $p$  và tỷ lệ phân phối Fisher (F). Sau đó, nghiên cứu đã dùng phương pháp tập hợp con tốt nhất (BS - *best subset*), tính giá trị lambda Wilks ( $\lambda$ ) để lựa chọn các tham số phân tử có khả năng dự đoán hoạt tính tốt.

*Phát triển mô hình:* Mô hình LDA được tối ưu hóa và phát triển bằng phương pháp phân tích sự khác biệt tuyến tính tổng quát (GLDA - *General Linear Discriminant Analysis*) trong STATISTICA 10.0 dựa trên lý thuyết xác suất

thống kê. Ở đây, phương pháp phân bố tuyến tính Fisher được sử dụng để giải quyết vấn đề phân loại.

Mô hình tốt nhất đối với mỗi kỹ thuật được lựa chọn dựa vào các thông số thống kê khi đánh giá trên TS và PS là: độ chính xác hay tỉ lệ phần trăm các chẩn đoán chính xác ( $Ac$  - Accuracy), độ nhạy ( $Se$  - Sensitivity), độ đặc hiệu ( $Sp$  - Specificity), hệ số tương quan Matthews (MCC- matthews correlation coefficient), tỉ lệ dương tính giả ( $FP_{rate}$ - False Positive Rate), đường cong đặc trưng ROC (Receiver operating characteristic) [15] thông qua thông số về giá trị diện tích dưới đường cong (AURC-area under the ROC curve). Trong đó:

$$Ac = (Tp + Tn)/(Tp + Tn + Fp + Fn)$$

$$Se = Tp/(Tp + Fn)$$

$$Sp = Tn/(Tn + Fp)$$

$$MCC = (Tp \times Tn - Fn \times Fp) / [(Tp + Fn)(Tp + Fp)(Tn + Fn)(Tn + Fp)]^{1/2}$$

$$ROC = ((Su - br * Se) / (1 - br); Se);$$

Ở đây:  $Su = (Tp + Fn)/(Tp + Tn + Fp + Fn)$ ;  
 $br = (Tp + Fp)/(Tp + Tn + Fp + Fn)$

$Tp$ : Số dương tính thật;  $Tn$ : số dương tính giả;  $Fp$ : số âm tính thật;  $Fn$ : số âm tính giả.

#### 2.2.4 Đánh giá mô hình

Mô hình QSAR được đánh giá theo 5 nguyên tắc của OECD như sau:

(1) Có đích xác định.

(2) Các thuật toán được sử dụng để xây dựng mô hình rõ ràng (mô hình có thể được tái xây dựng lại những hợp chất mới).

(3) Có miền cấu trúc ứng dụng (AD - applicability domain) được xác định (sử dụng phương pháp tròn bẫy).

(4) Thỏa mãn các đánh giá nội trên tập TS để kiểm tra mức độ khớp, độ mạnh và ổn định của mô hình.

(5) Giải thích được cơ chế (nếu có thể).

### 3. Kết quả và bàn luận

Đầu tiên, CSDL đã được chia ra thành TS và PS. Các hợp chất có hoạt tính được chia thành 10 cụm và các hợp chất không có hoạt tính thành 12 cụm. Từ mỗi cụm, khoảng 75%

các hợp chất được lựa chọn để xây dựng thành TS. Kết quả, 536 hợp chất có hoạt tính và 676 hợp chất không có hoạt tính được chọn vào TS; 368 hợp chất còn lại được xếp vào PS.

**Mô hình LDA (M1):** Kết quả đã lựa chọn được 9 tham số mô tả phân tử tốt nhất. Mô hình có dạng:

$$Class (+/-) = 29.75 - 1.87 \times GATS2e + 2.19 \times nRNHO + 2.29 \times Uc - 8.11 \times SpMax1\_Bh(e) + 1.13 \times C038 - 0.42 \times N-067 - 0.53 \times GGI3 - 7.20 \times Eta\_epsi\_A + 0.89 \times nCRX3 \quad (M1)$$

$$N = 1103; \lambda = 0.41; \Delta^2 = 5.77; F = 130.73; p < 0.0001$$

$$TS: MCC=0.76 \quad Ac = 0.88 \quad Se = 0.84 \\ Sp = 0.86 \quad AURC=0.94$$

$$PS: MCC=0.74 \quad Ac = 0.86 \quad Se = 0.81 \\ Sp = 0.84 \quad AURC=0.91$$

Từ kết quả đánh giá trên có thể thấy mô hình M1 có giá trị  $F$  lớn,  $p$  nhỏ cùng với Wilks  $\lambda$  thấp có ý nghĩa trong thống kê. Hơn nữa, mô hình M1 cho kết quả thực hiện phân loại với giá trị  $Ac$  là 88.1% và 86.7% tương ứng trên tập TS và PS; hệ số MCC trên cả 2 tập  $\sim 75\%$  chứng tỏ mức độ cao và cân bằng của tỷ lệ giá trị dương tính và âm tính thật. Ngoài ra,  $AURC > 0.9$  mà giá trị AURC càng gần 1 càng tốt; do vậy giá trị này cũng cho thấy mô hình có độ chính xác cao. Tóm lại, chúng ta đã có thể chứng minh được rằng mô hình LDA - QSAR (M1) được phát triển từ 9 mô tả phân tử được lựa chọn có độ mạnh, khả năng dự đoán và có ý nghĩa thống [1-22] kê cao.

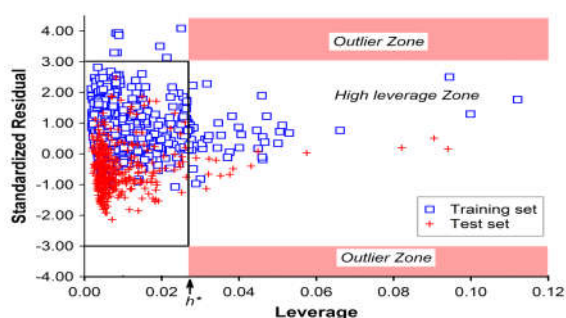
**Đánh giá mô hình QSAR:** dựa vào 5 nguyên tắc của OECD đã nêu ở trên

**Tiêu chí 1.** Ở đây đích xác định là phân loại các hợp chất thành: nhóm có hoạt tính ức chế và nhóm không có hoạt tính ức chế HDAC2 dựa vào nhiều thông số khác nhau như  $IC_{50}$ , phần trăm ức chế ( $I\%$ ) hay  $K_i$ .

**Tiêu chí 2.** Thuật toán áp dụng rõ ràng: Các phương pháp xây dựng mô hình trong nghiên cứu chúng tôi đã nêu ở phần 2. Do tính rõ ràng và đúng đắn của phương pháp mà các mô hình có thể được sử dụng lặp lại để phát hiện các hợp chất mới ức chế HDAC2.

**Tiêu chí 3.** Để đạt được khả năng dự đoán tốt trong sàng lọc ảo, việc xác định được AD

của mô hình QSAR là vô cùng cần thiết. AD có thể ứng dụng với tất cả các mô hình đã được đánh giá bằng phương pháp điểm Williams. Biểu đồ dưới đây (Hình 1) cho thấy mối quan hệ giữa phần dư chuẩn (*Standardized Residual* - xác định bởi LDA) và giá trị “đòn bẩy” *Leverage* ( $h$ ). Chỉ số  $h$  cho biết một chất có ảnh hưởng tới không gian cấu trúc của biến như thế nào và  $h$  được định nghĩa như sau:  $h_i = x_i^T (X^T X)^{-1} x_i$ ; trong đó  $x_i$  là vectơ mô tả phân tử, trị số đòn bẩy càng cao thì hiệu quả mô hình càng kém. Giá trị đòn bẩy cảnh báo  $h^*$  là một ngưỡng quan trọng để tham chiếu các dự đoán bởi mô hình QSAR cho hợp chất trong TS và PS;  $h^* = 3 \times p' / n$  trong đó  $n$  là kích thước của tập TS và  $p'$  là số lượng các biến mô hình.



Hình 1. Miền ứng dụng của mô hình QSAR xác định hợp chất có khả năng ức chế HDAC2.

Nguồn: Kết quả xử lý số liệu

Sau khi tính toán, ta thấy phần lớn các hợp chất trong TS nằm trong miền ứng dụng giới hạn bởi khung vuông  $\pm 3$  đối với phần dư chuẩn và  $h^*$  bằng 0.0272. Bảy hợp chất ức chế HDAC2 có giá trị dư chuẩn lớn hơn 3: bao gồm bốn hợp chất được sản xuất hoàn toàn từ vi khuẩn gram âm *C.violaceum* [16]. Các hợp chất khác có cấu trúc hydroxamate tương đồng với nhiều hợp chất khác trong TS. 44 hợp chất (gồm 13 hợp chất có hoạt tính và 31 hợp chất không có hoạt tính) được tìm thấy trong vùng vượt quá so với giá trị  $h^*$ . Các mẫu huấn luyện này có thể khiến mô hình thống kê có ý nghĩa kém và chúng nên được xóa khỏi tập huấn luyện. Hầu hết các hợp chất được phân bổ liên tục với giá trị  $h$  trong khoảng 0.0273 đến

0.0529; bao gồm cấu trúc có trong phân tử liên kết amide nối trifluoromethyl ketone,  $\alpha$ -aminoanilide, polipeptides và steroids. Ngoài ra, 4 hợp chất có giá trị đòn bẩy rất cao bao gồm: pyrrole-imidazole polyme SAHA ( $h = 0.112$ ), cefpodoxime ( $h = 0.099$ ), I-Coeleneterazine ( $h = 0.094$ ) và ticarcillin ( $h = 0.066$ ).

Bên cạnh đó, nhiều biến không có ý nghĩa ở các mô hình dự đoán trên tập PS và 14 hợp chất có giá trị đòn bẩy cao đã bị loại.

**Tiêu chí 4.** Kết quả của quá trình đánh giá đã được trình bày ở trên với độ chính xác và độ ổn định khá tốt trên các mô hình.

**Tiêu chí 5.** Giải thích cơ chế: Giải thích 9 mô tả phân tử của mô hình M1.

*GATS2e*: phân bố nguyên tử mang điện âm của các cầu nối có độ dài xác định ( $\text{lag} = 2$ ).

*SpMax1\_Bh (e)*: cung cấp thông tin môi trường điện tử của các nguyên tử có trong mô phỏng cấu trúc phân tử mà đỉnh hydro bị xóa.

*Eta\_epsi\_A, GGI3*: mô tả được định nghĩa là  $\Sigma \epsilon_i / N$  trong đó  $\epsilon_i$ : nguyên tử thứ  $i$  ( $i = 1, \dots, N$ ) được tính toán dựa trên phần lõi và hóa trị electron liên kết đến một đỉnh trong mô tả cấu trúc có đỉnh Hydro bị xóa.

*C-038* và *N-067*: là hai dạng mô tả phân tử mảnh trung tâm. Thuật ngữ Al đề cập đến nhóm béo (*Aliphatic groups*) và X là nguyên tử có độ âm điện cao (O, S). Tính toán cho thấy các yếu tố này lại cho hiệu quả ngược lại khi góp mặt trong cấu trúc các chất ức chế HDAC trong khi các nhóm Cacbonyl hoặc cacbothionyl giúp tăng hoạt tính ức chế, trong khi đó nhóm ketimine cũng cho hiệu quả đóng góp ngược với hoạt tính một phần.

*nRNHO* và *nCRX3* liên quan đến số lượng nhóm hydroxylamine và nhóm thế halogen (MHSs- *multiple halogen substituents*), thành phần cấu trúc đóng góp khá tốt vào hoạt tính ức chế. Dựa vào CSDL các dẫn xuất của acid hydroxamic với nhóm quan trọng là hydroxylamine, phân tích mối quan hệ cấu trúc và hoạt động (SAR – *Structure activity relationship*), chúng tôi xác định được cơ chế tạo phức với  $\text{Zn}^{2+}$  của các dẫn chất hydroxylamic tại vị trí hoạt động [20]. Việc thêm các nhóm thế halogen (Cl, Br, F,  $\text{CF}_3$  ...)

cũng được nghiên cứu kỹ để thiết kế các hợp chất mới. Ngoài ra, theo các nghiên cứu thì nhóm thế ceton trifluoromethyl cũng được đánh giá có hoạt tính trong cấu trúc acid hydroxamic. Như vậy, MHSs cần được lưu ý trong thiết kế hợp chất mới thông qua đánh giá mô tả phân tử này.

*Uc* là một trong những mô tả ảnh hưởng nhất, có khả năng ức chế HDAC tốt với cấu trúc bất bão hòa, được tính như sau  $Uc = \log 2 \times (1 + nDB + nTB + nAB)$  ở đây nDB, nTB và nAB tương ứng là số liên kết đôi, liên kết ba và vòng thơm.

Như vậy, để tối đa hóa hoạt tính, chúng tôi tập trung tăng các mảnh cấu trúc nRNHO, *Uc*, C-038 và nCRX3 song song với việc giảm thiểu GATS2e, SpMax1\_Bh (e), N-067, GGI3 và Eta\_epsi\_A; cuối cùng điều kiện tốt nhất cho từng mô tả đã được xác định như sau: *GATS2e* = 0.18, *nRNHO* = 1 hoặc 2, *Uc* = 5.33, *SpMax1\_Bh (e)* = 3.80, *C-038* = 1 hoặc 2, *N-067* = 0, *GGI3* = -0.42, *Eta\_epsi\_A* = 0.49 và *nCRX3* = 1. Tuy nhiên thực tế khó có thể xây dựng được cấu trúc hoàn toàn chính xác từ thông số trên nhưng nó cũng đã đưa ra được hướng tìm kiếm và phát triển các hợp chất ức chế HDAC2.

Dựa trên mô hình xây dựng được và vai trò các mảnh cấu trúc đã xác định, trong bước thiết kế các hợp chất mới tiếp theo cần tập trung thêm các nhóm thế halogen (Cl, Br, F...), nhóm chức cacbothynyl, cacbonyl, hydroxylamin trên các sườn cấu trúc, đặc biệt là dẫn chất của acid hydroxamic. Cấu trúc bất bão hòa (liên kết đôi, liên kết ba, vòng thơm) cũng góp phần tăng tính ức chế HDAC2. Đây chính là các nhóm chức, nhóm thế, phần cấu trúc tiềm năng cốt lõi, cơ sở để thiết kế các chất có hoạt tính ức chế mới cần được chú ý.

#### 4. Kết luận và kiến nghị

Nghiên cứu đã xây dựng thành công mô hình toán học biểu diễn mối quan hệ định lượng giữa cấu trúc và tác dụng ức chế enzyme HDAC2 thỏa mãn cả năm nguyên tắc OECD. Mô hình

xây dựng được giúp phân loại các hợp chất có hoạt tính sinh học được nhanh chóng và dễ dàng hơn. Mô hình này còn có thể được ứng dụng để sàng lọc các CSDL lớn nhằm tìm kiếm các hợp chất ức chế HDAC2. Ngoài ra, kết quả nghiên cứu cũng giúp định hướng thiết kế các hợp chất mới có hoạt tính ức chế cao hơn bằng việc tìm ra các mảnh cấu trúc chịu trách nhiệm tăng hoạt tính. Tất cả hướng đến việc tìm kiếm các hợp chất dẫn đường cho quá trình nghiên cứu và phát triển các thuốc chống ung thư hướng đích mới.

#### Lời cảm ơn

Nghiên cứu được thực hiện trong khuôn khổ đề tài 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] Katrina J. Falkenberg and Ricky W. Johnstone, Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nature Reviews Drug Discovery* 13 (2014) 673.
- [3] Đào Thị Kim Oanh, et al., Tổng hợp N1-(benzo[d]thiazol-2-yl)-N6-hydroxyadipamid và dẫn chất hướng ức chế histon deacetylase. *Tạp chí Dược học* 51 (2011) 55.
- [4] Đào Thị Kim Oanh, et al., Tổng hợp N1-(benzo[d]thiazol-2-yl)-N8-hydroxyoctandiamid và dẫn chất hướng ức chế histon deacetylase. *Tạp chí Dược học* 51 (2011) 45.
- [5] Đào Thị Kim Oanh, Sang-Bea Han, and Nguyễn Hải Nam, Tổng hợp N1-(benzo[d]thiazol-2-yl)-N4-hydroxysuccinamid và dẫn chất hướng ức chế histon deacetylase. *Tạp chí Dược học* 51 (2011) 47.
- [6] Nam N-H, et al., Synthesis, bioevaluation and docking study of 5-substitutedphenyl-1,3,4-thiadiazole-based hydroxamic acids as histone deacetylase inhibitors and antitumor agents.

- Journal of enzyme inhibition and medicinal chemistry 29 (2013) 611.
- [7] Nam N-H, et al., Novel isatin-based hydroxamic acids as histone deacetylase inhibitors and antitumor agents. *European journal of medicinal chemistry* 70 (2013) 477.
- [8] Iskander Yusof and Matthew D. Segall, Considering the Impact of "Drug-like" Properties on the Chance of Success. Optibrium Ltd., 7221 Cambridge Research Park, Beach Drive, Cambridge, CB25 9TL, 2013.
- [9] Todeschini, R. and V. Consonni, and H.o.M.D. 1st, *Handbook of Molecular Descriptors* 1st Wiley VCH; Weinheim, Germany, 2000.
- [10] Prof. Dr. Stefan Bienz, Short Manual to the Chemical Drawing Program ChemDraw®. IfC Department of Chemistry.
- [11] Andrea Mauri, et al., DRAGON SOFTWARE: an easy approach to molecular descriptor and calculator. *MATCH Commun. Math. Comput. Chem.* 56 (2006) 237.[12]. Froimowitz M, HyperChem: a software package for computational chemistry and molecular modeling, *Biotechniques* 14(6) (1993) 1010.
- [12] Anh, N.T.K., Nghiên cứu giải thích phản ứng SEAR bằng phần mềm HyperChem 7.54. Khóa luận tốt nghiệp, Đại học Sư Phạm Hà Nội, 2011.
- [13] STATISTICA (data analysis software system) Version 8.0.1, StatSoft, Inc., Tulsa, OK, 2012.
- [14] F. Provost and T. Fawcett., Analysis and visualization of classifier performance: comparison under imprecise class and cost distribution. *Third International Conference on Knowledge Discovery and Data Mining.*, Newport Beach, California, 1997.
- [15] C. Wang, et al., Thailandepsins: bacterial products with potent histone deacetylase inhibitory activities and broad-spectrum antiproliferative activities. *Journal of Natural products* 74 (2011) 2031.
- [16] Huong Le-Thi-Thu, et al., A Comparative Study of Nonlinear Machine Learning for the "In Silico" Depiction of Tyrosinase Inhibitory Activity from Molecular Structure. *Molecular informatics* 30 (2011) 527.
- [17] Iskander Yusof and Matthew D. Segall, Considering the Impact of 'Drug-like' Properties on the Chance of Success, *Drug Discov Today* 18 (2013 ) 659.
- [18] Huong Le-Thi-Thu, et al., QSAR models for tyrosinase inhibitory activity description applying modern statistical classification techniques: A comparative study. *Chemometrics and Intelligent Laboratory Systems* 104(2) (2010) 249.
- [19] D. Hadjipavlou-Litina and E. Pontiki, Quantitative Structure–Activity Relationship Studies on Hydroxamic Acids Acting as Histone Deacetylase Inhibitors,. *Hydroxamic Acids: A Unique Family of Chemicals with Multiple Biological Activities*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [20] Alexander Tropsha and Alexander Golbraikh, *Handbook of Chemoinformatics Algorithm*, Chapter 6. 173-210, CRC Mathematical and Computational Biology Series, New york, 2010.
- [21] OECD Principles for the validation, for Regulatory purposes, of (Quantitative) structure-activity relationship models, 37<sup>th</sup> Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology, 2004.

## Development of Quantitative Structure - Activity Relationship Model for Discovery of Novel Histone Deacetylase 2 Inhibitors

Pham The Hai<sup>1</sup>, Doan Viet Nga<sup>2</sup>, Nguyen Thi Ngoc<sup>1</sup>,  
Bui Thanh Tung<sup>2</sup>, Le Thi Thu Huong<sup>2</sup>

<sup>1</sup>Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hanoi, Vietnam

<sup>2</sup>VNU School of Medicine and Pharmacy, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

**Abstract:** Treatment cancer has interested the top of many countries with a lot of investigations for finding novel molecularly targeted compounds. Histone deacetylase (HDAC) was known as an

enzyme in epigenetics way; that means it is an important target to explore. In this study, we have developed and validated successfully QSAR model for the identification of HDAC2 inhibitors only from molecular structures. Firstly, we computally described these molecules by DRAGON software, latter we used LDA approach with STATISTICA 10.0 for buiding model. The OECD principles were used to assess the QSAR model validation. The model accuracy was higher than 86% and this model fulfilled all OECD principles. The developed method helps us saving time, money in screening step in discovery of new lead compounds. Moreover, these results were provided the basis for designing new HDAC2 inhibitor compounds towards developing the targeted cancer therapy drug.

*Keywords:* Histone Deacetylase 2; Quantitative Structure-Activity Relationship; Molecular Descriptors DRAGON; LDA; STATISTICA.