**Mô tả phân tử ISIDA-Ứng dụng trong thiết kế thuốc ức chế HDAC2**

**Phạm Thế Hải1, Đoàn Việt Nga2, Nguyễn Thị Ngọc1, Bùi Thanh Tùng2, Lê Thị Thu Hường2,\***

*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 Na*

Liên hệ:

E-mail: ltthuong1017@gmail.com

Số điện thoại: +84 04 38387949

+84 967300779

**Tóm tắt**

###  Nghiên cứu thuốc điều trị ung thư hiện nay đang là mối quan tâm hàng đầu của nhiều nước với mong muốn đạt hiệu quả cao trong điều trị về cả tác dụng và chi phí. Một trong số các đích phân tử tác dụng của thuốc đáng chú ý là enzyme Histone Deacetylase 2 (HDAC2). Ngày nay, sự phát triển của khoa học máy tính đã hỗ trợ ngành dược trong rất nhiều nghiên cứu tìm kiếm và phát triển thuốc. Trong nghiên cứu này, mô tả mảnh cấu trúc ISIDA được sử dụng để tính toán tham số phân tử, áp dụng kỹ thuật hồi quy tuyến tính đa biến (MLR- *multi linear regression analysis*) trong phần mềm STATISTICA 8.1 giúp xây dựng mô hình toán học định lượng giữa cấu trúc và hoạt tính (QSAR-*Quantitive structure activity relationship*). Kết quả, chúng tôi đã xây dựng thành công ba mô hình hồi quy tuyến tính đa biến trên ba tập hợp mảnh cấu trúc cụ thể, rõ ràng có giá trị phương sai thực nghiệm đều lớn hơn 0.7, độ chính xác và khả năng dự đoán cao. Từ các mô hình, tìm ra được mảnh cấu trúc quan trọng trong thiết kế các chất mới được ứng dụng để hình thành nên các khung cấu trúc có hoạt tính ức chế HDAC2, góp phần tiết kiệm thời gian công sức và nâng cao hiệu quả điều trị ung thư.

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

**I. ĐẶT VẤN ĐỀ**

HDAC2 là enzyme thuộc nhóm I trong họ enzyme HDAC có nhiệm vụ loại bỏ nhóm acetyl (O=C-CH3) từ amino acid ε-N-acetyl lysine trên phân tử histone, cho phép ADN cuộn chặt với histone trong 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 [1]. Trong mối tương quan với sự hình thành ung thư, nhiều nghiên cứu đã cho thấy biểu hiện quá mức của HDAC2 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, 7]. Chính vì vậy, HDAC2 được lựa chọn là đích phân tử và đòi hỏi cần phải thiết kế các hợp chất ức chế hoạt tính enzyme này trong ung thư. Hiện nay đã có rất nhiều hợp chất có nguồn gốc tự nhiên hoặc tổng hợp hóa học có khả năng ức chế HDAC2 nhưng vẫn còn nhiều hạn chế. Mặc dù đã có những tiến bộ trong nghiên cứu và phát triển nhưng đây là một quá trình kéo dài và rất tốn kém, do đó trong nghiên cứu này để vừa tiết kiệm về mọi mặt và nâng cao hiệu quả cũng như tính đặc hiệu, chúng tôi đã thực hiện nghiên cứu theo phương pháp thiết kế thuốc hợp lý dựa trên hiểu biết về đích phân tử. Nghiên cứu đã ứng dụng phương pháp *In silico* với sự hỗ trợ của máy tính trong xây dựng mô hình toán học giúp sàng lọc và thiết kế hợp chất mới. Đặc biệt, ở đây chúng tôi đã ứng dụng phương pháp dựa trên mảnh cấu trúc (FBDD- *Fragment based drug discovery*) thay vì phải sàng lọc hàng triệu hợp chất từ các thư viện để tìm ra hợp chất dẫn đường-điểm khởi đầu cho thiết kế thuốc, FBDD bắt đầu với tập hợp các phân tử cấu trúc nhỏ dựa trên cấu tạo của đích. Một trong số các FBDD hiện nay đó chính là mô tả phân tử ISIDA-*In Silico design and Data Analysis*.

**II. 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 ức chế HDAC2 với giá trị IC50 và logIC50 đã được nghiên cứu từ các bài báo khoa học trong nước và quốc tế đã công bố trước đó được xác định bằng phương pháp giống nhau (sử dụng Kit định lượng Bioscence [8]).

*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 = a1(x1) + a2(x2)+…+ an(xn**).** Trong nghiên cứu này, Y là giá trị IC50, là biến phụ thuộc, nhận giá trị tương ứng với hoạt tính (đơn vị μg/ml) được đo trong quá trình thực nghiệm; x1, x2..., xn là các tham số phân tử đặc trưng cho cấu trúc; a1, a2,...an là các trọng số tương ứng với từng tham số phân tử xi đối với tác dụng sinh học.

2.2.1 Tính toán mô tả phân tử

 Đây là bước chuyển đổi thông tin từ một mảnh cấu trúc thành một dãy số có khả năng thể hiện đặc tính của phân tử ấy thông qua quá trình biến đổi toán học logic [9]. Các cấu trúc trong CSDL được lưu dưới dạng tệp cấu trúc (SDF- *Structure Data File*). Chương trình ISIDA (Mô tả phân tử thiết kế *in silico* và phân tích dữ liệu) tính mô các mảnh cấu trúc từ SDF. Phần mềm ISIDA Fragmentor2015 được phát triển bởi Phòng thí nghiệm Hóa học, Viện Hóa học, Đại học Strasbourg, Pháp [10]*.* Chương trình mô tả ISIDA ưu điểm hơn các mô tả khác ở khả năng đếm số lượng mảnh cấu trúc có hoạt tính có mặt trong phân tử, tính trình tự đặc trưng và nhóm ở nhánh (nguyên tử tăng cường) thậm chí là một mảnh cấu trúc mới tùy theo mục đích, sử dụng các kí hiệu dán nhãn mang đặc tính mảnh cấu trúc gán trên các vector để xây dựng mô hình sàng lọc và thiết kế hợp chất mới.

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 IC50 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* [11] để xác định tương quan hoạt tính ức chế với tính toán mô tả ISIDA.

2.2.4 Đánh giá mô hình

 Một mô hình QSAR nói chung cần đạt yêu cầu là có đích xác định, có thuật toán xây dựng rõ ràng để tái sử dụng với những hợp chất mới, thỏa mãn các đánh giá chứng minh độ khớp, độ mạnh và độ ổn định; đặc biệt mô hình phải giải thích cơ chế một cách rõ ràng, cụ thể. Đối với riêng mô hình MLR được xây dựng với nhiều tập hợp tham số phân tử khác nhau và các mô hình tốt nhất được đánh giá dựa trên các thông số thống kê và đánh giá chéo [12]. Đánh giá nội trên tập TS cho biết độ khớp, độ mạnh của mô hình thông qua các thông số thống kê: r2- hệ số tương quan, 0 < r2 < 1, trong phương trình hồi quy tuyến tính đa biến giá trị r2 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 [13**]**; F là phân phối Fisher, r2 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 q2LOO (hệ số leave-one-out), q2LMO (hệ số leave-more-out) [14].

 Đánh giá ngoại mô hình MLR trên tập PS để kiểm tra khả năng dự đoán chính xác: RMSEext và MAEext, Q2F1, Q2F2, và Q2F3 (giá trị ngưỡng cho phép là 0.7), CCCext và rm2 (giá trị ngưỡng thấp nhất là 0.5) [15] [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 r2scr và q2scr. Nếu đỉnh trục Y không vượt quá 0.3–0.4 đối với r2scr và 0.05 đối với q2scr 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 r2scr và q2scr được lặp lại 2000 lần [17].

**III. 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**

$$ logIC\_{50}= -0.130 x V\_{1}+5.408 x V\_{2}+2.047 x V\_{3}-6.466 x V\_{4}-2.011 x V\_{5}-2.179 x V\_{6}-0.246 x V\_{7}+0.162 x V\_{8}$$

**M2**

$$ logIC\_{50}= -1.573+5.589 x V\_{1}+1.021 x V\_{2}+0.907 x V\_{3}-6.190 x V\_{4}-2.501 x V\_{5} +1.542 x V\_{6}-1.056 x V\_{7}-2.036 x V\_{8}$$

**M3**

$$logIC\_{50}=4.160+4.348 x V\_{1}+1.371 x V\_{2}-3.336 x V\_{3}-1.558 x V\_{4}-0.487 x V\_{5}+2.330 x V\_{6}-2.416 x V\_{7}-0.332 x V\_{8}$$

Các biến *V1* – *V8*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.

**Bảng 1:** Mô tả phân tử cho mỗi mô hình

|  |  |  |  |
| --- | --- | --- | --- |
|  | M1 | M2 | M3 |
| V1 | (N-C),(N-C),(N-H),xN | (N-C),(N-C),(N-H),xN | (N-C),(N-C),(N-H),xN |
| V2 | Cl-C-C-N | Cl-C-C-N-C | H-C=C-C=C-Cl |
| V3 | Cl-C=C-N | Cl-C=C-N-C | C-C=C-O-C-N |
| V4 | C-O-C-N | C=C-O-C-N | H-C-C-N-C=C |
| V5 | H-N-C=C | H-N-C=C-C | C-C=C |
| V6 | O-C-C-O | H-C=C-C-N | Cl-C-C=C-C=N |
| V7 | C-C-C | C=C-C-N | N-S-C-C=C-O |
| V8 | H-C-N-C | S-C-C=C-O | C-C-C-N |

***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

Như chúng tôi đã nói, cơ sở dữ liệu được chia ra thành TS và PS. 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.

 **Bảng 2**: Các thông số đánh giá chéo và độ thích hợp của các mô hình tuyến tính trên tập TS

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | r2 | Q2LOO | RMSE | MAE | CCC | r2(scr) | q2(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 r2 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 phương sai thực nghiêm cao nhất với 0.82. Giá trị q2LOO ở mô hình M1 là 0.76 và khi thay đổi các biến từ V2 –V8 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 q2LOO 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 r2(scr) cao nhất là 0.14 (M2) và q2(scr) cao nhất là -0.34 với mô hình M1; kết quả này chứng tỏ cả 3 mô hình không có tương quan, ảnh hưởng lẫn nhau giữa các biến. 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 3:** Tính toán khả năng dự đoán của mô hình tuyến tính trên tập PS

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Q2F1 | Q2F2 | Q2F3 | RMSEext | MAEext | 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 Q2 rất cao, mô hình M3 cho kết quả tốt nhất với Q2F1, Q2F2, Q2F3 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. Ở đây, nghiên cứu sẽ chỉ ra, giải thích cụ thể vai trò của mảnh cấu trúc mang hoạt tính và chức năng thực sự của mô tả mảnh cấu trúc ISIDA.

Dựa trên các mô hình MLR đã được xây dựng và đánh giá, việc tìm ra chính xác các mảnh cấu trúc có đóng góp như thế nào tới hoạt tính ức chế HDAC2 đều dựa trên các con số toán học. Bảng 1 là các mô tả phân tử khác nhau được sử dụng trong các mô hình tương quan cấu trúc và giá trị logIC50 xây dựng được. Các phân tử được bôi xám có hệ số giá trị âm (CNV- *coefficient of negatives values*) trong mô hình, đồng nghĩa đây là các yếu tố tăng cường hoạt động ức chế (giảm giá trị IC50). Có thể thấy được từ các mô hình, CNV luôn xuất hiện cùng các mảnh cấu trúc có O trong đó. Chính vì vậy, khi thiết kế hợp chất mới, cần tập trung sử dụng nhiều các mảnh cấu trúc này.

Ngược lại, hệ số có giá trị dương (CPV- *coefficients with positive value*) làm tăng giá trị IC50  do đó giúp chỉ ra các mảnh cấu trúc làm giảm hoạt tính ức chế. Xét các giá trị CPV này có thể thấy được rằng trong ba mô hình M1, M2 và M3 có một số mô tả phân tử có giá trị ngang bằng hoặc xấp xỉ đó là biến V1: {(N-C), (N-C), (N-H), xN}, biến V2 và biến V3 mang hệ số CPV ở mô hình M1, M2. Dựa vào biến V2, V3 có thể dễ dàng nhận thấy, sự có mặt của Cl làm giảm hoạt tính vì chỉ xuất hiện với CPV, do đó tránh các mảnh cấu trúc có Cl là rất cần thiết khi thiết kế hợp chất mới.

Kết hợp cả 3 mô hình, chúng tôi đã tìm ra 9 mảnh cấu trúc giữ vai trò quan trọng nhất đóng góp vào hoạt tính ức chế HDAC2 (Hình 2).



**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**.**

#### Sau đó, tổ hợp các mảnh cấu trúc này để tạo thành các hợp chất mới một các tự động bằng Thuật giải di truyền [18] chỉ trong một thời gian ngắn giúp tạo ra hàng nghìn hợp chất. 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 (IC50) của các hợp chất được thiết kế bằng chính những mô hình đã xây dựng được. Sử dụng phần mềm ISIDA phiên bản 2.0 tính toán các tham số phân tử của các biến 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 IC50 dự đoán tốt và quan trọng là các cấu trúc thiết kế này phải có khả năng tổng hợp phù hợp điều kiện thực tế (bảng 4).

**Bảng 4:** Các khung cấu trúc mới được đề xuất cho thiết kế

|  |  |  |
| --- | --- | --- |
| STT | Cấu trúc | Dự đoán IC50 (μg/ml) |
| **1a** |  | 1,60125 |
| **2a** |  | 3.10026 |
| **3a** |  | 0,96005 |
| **4a** |  | 0,30018 |
| **5a** |  | 1,11231 |
| **6a** |  | 4,62709 |

 Từ 6 khung mới được thiết kế, nhận thấy khung có tác dụng ức chế HDAC2 tốt nhất, có giá trị IC50 được dự đoán là thấp nhất là **4a.**



**Hình 3.** Khung dẫn chất 4a

***Mô tả mảnh cấu trúc ISIDA*** [19]: Mục đích để hiểu rõ hơn cơ chế tính toán và ứng dụng của mô tả này trong xây dựng mô hình.

a. Cơ chế mô tả

*Cơ chế tô màu*

Cho phép viết các nguyên tố dựa theo bất kì cách nào (kí hiệu, tính chất tĩnh điện, tính thân dầu, độ pH…) đảm bảo sự liên quan giữa màu sắc và tính chất hóa học. Đối với một cấu trúc thường biểu diễn bằng hình ảnh tô màu, các đỉnh được đánh dấu theo tính chất của các nguyên tố tại đỉnh. Đặc biệt, thể hiện tính dược học dựa trên tô màu được sử dụng rộng rãi để nắm bắt được bản chất của sự tương tác giả định (liên kết kỵ nước, liên kết hydro, cầu muối) với một nhóm chức năng trong một phối tử có thể xảy ra khi liên kết với đích.

*Gán nhãn dựa trên điện thế*

Đầu tiên tính toán trên ChemAxon tất cả điện tích dựa trên phương trình obital điện tích. Điện thế Vi  của mỗi nguyên tố i được tính như sau

$$Vi=\sum\_{i \ne j }^{}\frac{qj}{dij}+\frac{qi}{do}$$

Ở đây: qj là điện tích một phần của nguyên tố i

 dij là là khoảng cách topo giữa nguyên tố i và j

 do là khoảng cách ảo để tính điện tích của các nguyên tố liên quan (do = 0.4 sau khi được điều chỉnh theo kinh nghiệm, để đảm bảo các nguyên tố cực dương và cực âm trong các chất hữu cơ được phân loại phù hợp với phán đoán hóa học thông thường).

Giá trị V sau đó được chia thành 5 nhóm:

* N: điện tích âm ( Vi ≤ -0.28)
* n: âm nhẹ (-0.32< Vi ≤ -0.08)
* 0: trung tính ( -0.12 < Vi ≤ 0.12)
* p: dương nhẹ ( 0.08 < Vi ≤ +0.32)
* P: điện tích dương ( Vi ≥ + 0.28)

Trong các trường hợp không rõ ràng có thể kí hiệu cả hai, ví dụ nếu Vi = -0.3 thì nguyên tố i được kí hiệu N/n.

*Gán nhãn theo đặc tính dược học*

Nhập tính chất dược học của nguyên tố bằng công cụ Pmapper, thực hiện theo các nguyên tắc sau:

* Các vòng thơm sẽ kí hiệu “R”
* Mang điện tích dương là “P”
* Trung tâm của nhóm mang điện âm là “N”
* Bất kỳ O hay N liên kết với H kí hiệu là “D” ( hoặc HB)
* Bất kỳ O hay N hay nhóm sulphide âm điện hoặc có thio (=S) kí hiệu là “A”
* Bất kỳ C hoặc halogen trừ khi nếu liên quan đến các quy tắc trên sẽ được kí hiệu là “H” ( chất kị nước)
* Các nguyên tử không phù hợp với các quy tắc trên sẽ được kí hiệu “F”



**Hình 4:** Biểu diễn công thức N-(5-hydroxypyridin-2-yl) acetamide theo chương trình tô màu và gắn nhãn [20]. 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

b. Tính toán mảnh cấu trúc

Sử dụng ngôn ngữ lập trình Pascal, tìm kiếm sự có mặt các tính năng trong các chuỗi trình tự, các cặp nguyên tố, nguyên tố tăng cường (nhánh) [21] [22] hoặc cấu trúc cây.

* Chuỗi: chuỗi liên kết các nguyên tử mang chức năng. Kích thước kiểm soát chiều dài của chuỗi thường có giới hạn dưới là 3 và giới hạn trên 5 do đó sẽ xác định các trình tự có 3,4,5 nguyên tố trong cấu tạo. Ví dụ ở hình 3 đếm chuỗi R\*R\*R-D thấy có 4 lần trong cấu tạo 1 và 2 lần trong cấu tạo 2 do đó giá trị tính toán cho đếm mảnh cấu trúc: 4 x 96 + 2x4 = 392.
* Cặp: là các mảnh được phân tách ra, bao gồm bất kỳ 2 nguyên tử nào, ở đó có khoảng cách topo giữa chúng và đếm mảnh cấu trúc theo 2 nguyên tố đó. Ví dụ, trình tự chuỗi R\*R\*R-D là một trong những phương án có thể có của các cặp biểu diễn như R3D, ở đây không quan tâm nguyên tố ở giữa. Đếm số cặp được tính bằng tổng số lần xuất hiện của chuỗi.
* Nguyên tử tăng cường: biểu diễn chuỗi và đếm số lượng sẽ tương tự kiểu mô tả phân tử dạng SMILES.
* Cấu trúc cây: chính là nguyên tử tăng cường (cấu trúc có nhánh) mà ở đó các tính năng cốt lõi trung tâm và bên cạnh được thể hiện. Nguyên tử ở giữa được che đậy bằng việc thay thế bới “?”[23] [24]*.*

**IV. KẾT LUẬN VÀ KIẾN NGHỊ**

Như vậy, sử dụng mô tả mảnh cấu trúc ISIDA đã giúp xây dựng thành công 3 mô hình toán học MLR định lượng mối quan hệ giữa giá trị IC50 trong ức chế hoạt động của enzyme HDAC2 và mảnh cấu trúc có độ tuyến tính và khả năng dự đoán cao thông qua các đánh giá dựa trên các hệ số, giá trị rõ ràng. Nghiên cứu đã giúp định hướng một cách nhanh chóng, tập trung được các mảnh cấu trúc đảm nhận vai trò hoạt tính chính, loại bỏ được các dạng cấu trúc làm giảm sự ức chế HDAC2, từ đó giúp hình thành được các khung cấu trúc cho bước thiết kế hợp chất mới hiệu quả, đặc hiệu hơn, góp phần rút ngắn thời gian và chi phí cho điều trị ung thư trúng đích. Hơn nữa, kết quả của nghiên cứu sẽ là tiền đề cho bước tổng hợp hóa học và thử nghiệm tiền lâm sàng và lâm sàng tiếp theo.

**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] Choudhary C, et al., *Lysine acetylation targets protein complexes and co-regulates major cellular functions.* Science, (2009) 325(5942): p. 834-40.

[2] Milos Dokmanovic, Cathy Clarke, and P.A. Marks, *Molecular Cancer Research*. Vol. 67. (2007): American Association for Cancer Research. 3145-3152.

[3] Ping Zhu, et al., *Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis.* Cancer cell, (2004) 5: p. 455–463.

[4] JAEHWI SONG, et al., *Increased expression of histone deacetylase 2 is found in human gastric cancer.* APMIS (2005). 113: p. 264–268.

[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, (2005). 12: p. 395-404.

[6] Alain C. Borczuk, et al., *Non-Small-Cell Lung Cancer Molecular Signatures Recapitulate Lung Developmental Pathways.* American Journal of Pathology, 2003. 163: p. 1949-1960.

[7] Kwang Hwa Jung, et al., *HDAC2 overexpression confers oncogenic potential to human lung cancer cells by deregulating expression of apoptosis and cell cycle proteins.* J. Cell. Biochem, (2012). 113 p. 2167–77.

[8] Naresh Kandakatla and G. Ramakrishnan, *Ligand Based Pharmacophore Modeling and Virtual Screening Studies to Design Novel HDAC2 Inhibitors.* Advances in Bioinformatics, (2014) p. ID 812148.

[9] Todeschini. R, V. Consonni, and H.o.M.D. 1st, *Handbook of Molecular Descriptors 1st*. (2000), Weinheim, Germany.

[10] Fiorella Ruggiu, e.a., *ISIDA Fragmentor2015*. 2015: User Manual.

[11] F. Provost and T. Fawcett, *Analysis and visualization of classifier performance: comparison under imprecise class and cost distribution*. (1997), Third International Conference on Knowledge Discovery and Data Mining., Newport Beach, California.

[12] Castillo-Garit, J.A., et al., *Estimation of ADME properties in drug discovery: Predicting Caco-2 cell permeability using atom-based stochastic and non-stochastic linear indices.* J. Pharm. Sci., (2008) 97: p. 1946-76.

[13] Roberto Todeschini and Viviana Consonni, *Molecular Descriptors for Chemoinformatics*. (2009) 41.

[14] Farzin Hadizadeh, Saadat Vahdani, and Mehrnaz Jafarpou, *Quantitative Structure-Activity Relationship Studies of 4-Imidazolyl- 1,4-dihydropyridines as Calcium Channel Blockers.* Iranian journal of basic medical sciences, (2013) 16(8): p. 910-16.

[15] V Consonni and D Ballabio, *Comments on the definition of the Q 2 parameter for QSAR validation.* J Chem Inf Model., (2009) 49: p. 1669–1678.

[16] M., L., et al., *QSAR Models Using a Large Diverse Set of Estrogens.* J. Chem. Inf. Comput. Sci, (2011) 2001: p. 186-195.

[17] Sain, S.R., *The Nature of Statistical Learning Theory*. (2012) Southern Methdist University.

[18] Nguyễn Thị Thúy 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) p. 266-70.

[19] F. Ruggiu, et al., *Isida property-labelledfragment descriptors,.* Molecular Informatics, (2010) 29: p. 855-68.

[20] Fiorella Ruggiu, et al., *ISIDA Property-labelled fragment descriptors*. (2010) 2nd Strasbourg Summer School on Chemoinformatics, VVF Obernai, France.

[21] A. Varnek, et al., *Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures.* J. Computer-Aided Molecular Design, (2005) 19: p. 693-703.

[22] A. Varnek, et al., *ISIDA - platform for virtual screening based on fragment and pharmacophoric descriptors.* Curr ComputAided Drug Des, (2008) 4: p. 191-98.

[23] F. Bonachera, et al., *Fuzzy tricentric pharmacophore fingerprints. 1. Topological fuzzy pharmacophore tripletsand adapted molecular similarity scoring schemes.* J Chem Inf Model.,, (2006) 46: p. 2457-77.

[24] F. Bonachera and D. Horvath, *Fuzzy tricentric pharmacophore fingerprints. 2. application of topological fuzzy pharmacophore triplets in quantitative structureactivity relationships.* J Chem Inf Model, (2008) 48: p. 409-25.

**Abstract**

Reasearch target cancer treatment medicines is a extremely necessary problem which has interested of many countries. One of several molecular targets is exactly enzyme histone deacetylase 2 (HDAC2). With the development of computer science, pharmacology has a lot of studies that these have approach for finding and developing drugs. In this study, we have used describer molecular fragmentor ISIDA, MLR technology with STATISTICA 8.1 for building QSAR model. Finally, there are 3 successful MLR models with experimental variance values are all over 0.70, high result predictions and accuracy. Therefore, the models help us to identify important structure frame for designing novel inhibitor HDAC2 compounds which mean saving much times and money.

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