



Original Article

Towards Developing Quantitative Structure-activity Relationship Models for the Design of Novel Influenza A Inhibitors Targeting Neuraminidase

Ly Cong Thanh*

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

Received 07 July 2022

Revised 25 July 2022; Accepted 10 March 2023

Abstract: Anti-influenza-A drugs targeting viral neuraminidase have been in use for two decades. In this study, the quantitative structure-activity relationships (QSAR) of oseltamivir derivatives as influenza neuraminidase (IN) inhibitors have been explored using the Monte Carlo method based on the target function involving the index of the ideality of correlation. Three best-obtained models showed appropriate performance with R^2 values of training and test sets ranging from 0.71 to 0.86, respectively. Based on the structural information extracted from the models, new inhibitors were designed and predicted for IN activities. Finally, protein docking was applied to confirm their target binding ability.

Keyword: Anti-influenza A; neuramidase inhibitor; drug design; QSAR; docking.

* Corresponding author.

E-mail address: thanhlc@hup.edu.vn

<https://doi.org/10.25073/2588-1132/vnumps.4412>

Nghiên cứu xây dựng mô hình liên quan cấu trúc - tác dụng dự đoán hoạt tính ức chế neuraminidase hướng điều trị cúm A

Lý Công Thành*

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

Nhận ngày 07 tháng 7 năm 2022

Chỉnh sửa ngày 25 tháng 7 năm 2022; Chấp nhận đăng ngày 10 tháng 3 năm 2023

Tóm tắt: Thuốc chống virus cúm A H1N1 hướng ức chế neuraminidase đã được nghiên cứu và ứng dụng trong hai thập kỷ qua. Trong nghiên cứu này, mô hình liên quan định lượng giữa cấu trúc và hoạt tính (QSAR) của các dẫn xuất oseltamivir hướng ức chế neuraminidase (IN) của virus cúm A đã được phát triển sử dụng phương pháp Monte Carlo dựa trên hàm mục tiêu xác định bởi hệ số tương quan lý tưởng. Kết quả thu được ba mô hình tốt nhất với hệ số tương quan R^2 xác định cho tập huấn luyện và tập kiểm tra lần lượt nằm trong khoảng từ 0,71 đến 0,86. Dựa trên thông tin cấu trúc trích xuất từ các mô hình, một số dẫn xuất oseltamivir mới được thiết kế và dự đoán hoạt tính ức chế neuraminidase. Cuối cùng, mô phỏng protein docking đã được áp dụng để xác nhận khả năng gắn kết với mục tiêu phân tử của chúng.

Từ khóa: Anti-influenza A; neuraminidase inhibitor; drug design; QSAR; docking.

1. Mở đầu

Bệnh cúm A là bệnh nhiễm virus cấp tính đường hô hấp phổ biến nhất hiện nay với đặc điểm dễ lây lan, nhiều biến chứng và có nguy cơ trở thành đại dịch khi xuất hiện các biến chủng nguy hiểm [1]. Virus cúm A thuộc họ orthomyxoviridae, sở hữu bộ gen RNA gồm tám sợi đơn mã hóa 11 protein khác nhau và không có gen mã hóa enzym sửa chữa RNA, tạo điều kiện thuận lợi cho sự xuất hiện các đột biến điểm để hình thành biến chủng mới. Virus cúm A được phân loại dựa trên kháng nguyên bề mặt ngưng kết hồng cầu H (Hemagglutin) và trung hòa N (Neuraminidase) [1]. Có 15 loại kháng nguyên H và 9 loại kháng nguyên N, khi tổ hợp với nhau tạo nên hàng trăm chủng mới. Khoảng hơn 130 chủng cúm A đã được phát hiện, trong đó phải kể đến dịch H1N1 và H5N1 trong 2 thập kỷ qua,

làm chết hàng chục triệu gia cầm và là nguyên nhân tử vong của hơn 10.000 người tại 160 quốc gia [2]. Nước ta thuộc nhóm các nước bị ảnh hưởng nặng nề bởi hàng loạt các đợt dịch cúm A như H1N1, H5N1 và H7N9 trong những năm qua. Hiện nay đã có nhiều tiến bộ trong phòng ngừa virus cúm A với sự ra đời của các loại vắc-xin mới. Việc sử dụng thuốc kháng virus đơn độc hoặc kết hợp trên cơ sở Oseltamivir (Tamiflu) được xem là khá hiệu quả trong phòng và điều trị virus cúm A [3]. Tuy nhiên tốc độ sản sinh biến chủng mới, vấn đề kháng thuốc và những biến chứng của nó vẫn là một thách thức lớn, đòi hỏi khoa học cần tăng tốc hơn nữa trong quá trình phát triển các liệu pháp điều trị mới giá rẻ và hiệu quả hơn.

Một trong các phương pháp mới, được ứng dụng phổ biến trong nghiên cứu sàng lọc, tìm

* Tác giả liên hệ.

Địa chỉ email: thanhlc@hup.edu.vn

<https://doi.org/10.25073/2588-1132/vnumps.4412>

kiểm thuốc mới hiện nay là các mô hình tương quan định lượng giữa cấu trúc và hoạt tính (*Quantitative Structure–Activity Relationship*, QSAR). Nhiều mô hình QSAR đã được công bố và áp dụng thành công trong thiết kế chất ức chế neuraminidase hướng điều trị cúm A H1N1, như mô hình của Mercader và Pomilio cho flavonoid [4], mô hình của Yaeghoobi và cộng sự cho cấu trúc carbocyclic và chalcon [5], hay mô hình 3D-QSAR của Lai và cộng sự trong thiết kế dẫn xuất của curcumin [6]. Từ các phân tích nêu trên, nghiên cứu này được tiến hành nhằm phát triển một số mô hình QSAR và ứng dụng trong thiết kế dẫn chất của oseltamivir có tác dụng ức chế H1N1 influenza neuraminidase (IN). Dựa trên mô phỏng Monte Carlo (MCS), ảnh hưởng của các nhóm cấu trúc lên hoạt tính được xác định. Cuối cùng, mô phỏng docking phân tử được áp dụng nhằm khẳng định lại khả năng gắn kết của các chất tốt nhất với đích phân tử IN.

2. Nguyên liệu và phương pháp nghiên cứu

2.1. Nguyên liệu

Cơ sở dữ liệu (CSDL): hoạt tính ức chế neuraminidase của virus cúm A H1N1 (IC_{50} , nM) và cấu trúc hóa học của 73 dẫn xuất của oseltamivir được thu thập từ các công trình đã

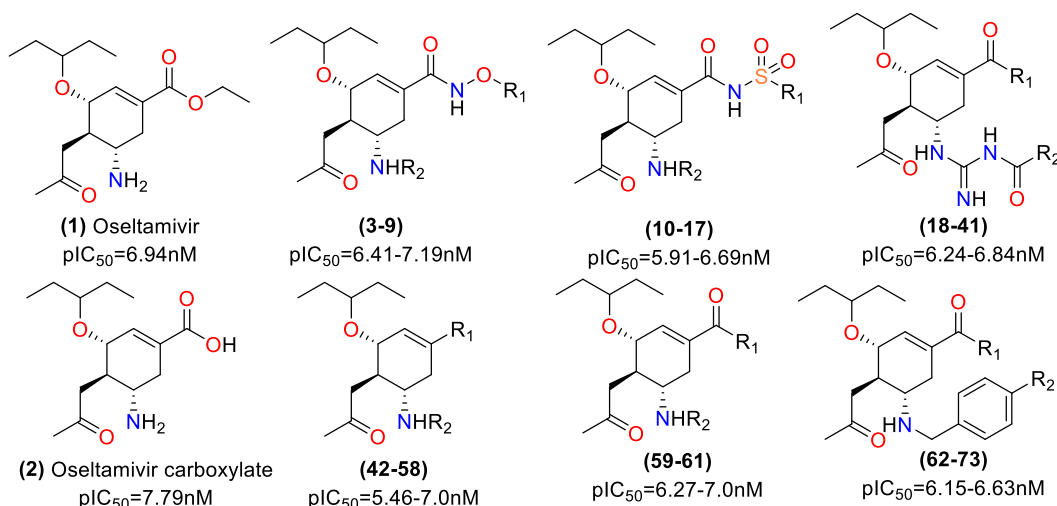
được công bố [7-9]. Logarit hóa giá trị IC_{50} để thu được biến đáp ứng là $pIC_{50} = \log(IC_{50})$. Cấu trúc khung của các chất này và khoảng giá trị của pIC_{50} được biểu diễn trong Hình 1.

Phần mềm tính toán: phần mềm tính toán tham số phân tử và xây dựng mô hình QSAR CoralSea2019 [10]. Phần mềm mô phỏng docking: Molegro Virtual Docker 2013.6.0.0.

2.2. Phương pháp nghiên cứu

2.2.1. Tính toán các tham số phân tử đặc trưng cho cấu trúc (TSPT)

Tham số phân tử lai hóa giữa SMILES và 3 sub-graph (HSG, HFG và GAO) được tính toán sử dụng phần mềm CoralSea2019 (descriptor of correlation weight, DCW). Công thức tính tham số DCW có dạng: $DCW(T^*, N^*) = SMILES DCW(T^*, N^*) + Graph DCW(T^*, N^*)$, trong đó, T^* là số mảnh cấu trúc ($1 < T^* < 3$) và N^* là số vòng (epoch) của MCS ($1 < N^* < 25$) [10]. Tổng hợp thu được hơn 200 tham số phân tử. Quá trình chọn biến được thực hiện sử dụng thuật giải di truyền, chia làm 3 cụm chạy song song. 3 mô hình tốt nhất được chọn và kết quả dự đoán là trung bình cộng của 3 mô hình này. Dữ liệu sau khi tính toán TSPT được chia ngẫu nhiên thành tập huấn luyện (54 chất dùng để xây dựng mô hình QSAR) và tập kiểm tra (19 chất dùng để đánh giá khả năng dự đoán của mô hình).



Hình 1. Cấu trúc khung và hoạt tính của các dẫn xuất của oseltamivir.

2.2.2. Phát triển các mô hình QSAR

Mô hình QSAR có dạng: $pIC50 = C_0 + C_1 \times DWC(T^*, N^*)$. Mô hình QSAR được xây dựng dựa trên 2 hàm đích (TF) tích hợp trong phần mềm Coral, có công thức sau: $TF_1 = R_{ATRN} + R_{PTRN} - |R_{ATRN} - R_{PTRN}| \times C$ và $TF_2 = TF_1 + II_C$, trong đó R_{ATRN} và R_{PTRN} là hệ số tương quan cho tập huấn luyện, C là hằng số empirical. II_C là chỉ số lý tưởng cho tương quan, có công thức: $II_C = r_{CAL} \times [\min(-MAE_{CAL}, +MAE_{CAL}) / \max(-MAE_{CAL}, +MAE_{CAL})]$, với r_{CAL} là hệ số tương quan giữa giá trị dự đoán ($Calc_y$) và giá trị thực (Obs_y) của hoạt tính sinh học trong tập huấn luyện. Tính sai số tuyệt đối trung bình MAE (Mean Absolute Error) như sau: $MAE_{CAL} = -(1/N) \sum_{y=1}^{-N} |\Delta_y| \Delta_y$ và $+MAE_{CAL} = -(1/N) \sum_{y=1}^{+N} |\Delta_y| \Delta_y$, với $-N$ và $+N$ là số các giá trị $\Delta_y < 0$ và $\Delta_y \geq 0$ tương ứng, và $\Delta_y = Obs_y - Calc_y$. Đánh giá mô hình QSAR thông qua 3 phương pháp: i) *k-fold cross validation* (R^2_{cv}); ii) dự đoán trên tập kiểm tra (Q^2_{test}); và iii) dự đoán trên dãy giả tự nhiên (data randomization) và xáo trộn (*Y-scrambling*). Các tham số thống kê tiêu chuẩn được sử dụng để đánh giá mô hình bao gồm: hệ số xác định (R^2_{cv} , Q^2_{test}), sai số chuẩn ước tính (s), sai số tuyệt đối trung bình (MAE), hệ số Fischer (F), sai số root-mean-square (RMSE) và một số thông số khác. Miền ứng dụng (applicability domain) của mô hình QSAR được xác định theo hướng dẫn của OECD, dựa trên hàm Defect_{AK} tích hợp trong CoralSea2019 [11].

2.2.3. Thiết kế một số dẫn chất mới của oseltamivir dựa trên mô hình QSAR

Thông qua số lần chạy tối ưu hóa Monte Carlo, các TSPT của phần mềm Coral được phân làm 3 lớp: i) Các TSPT có hệ số tương quan (CW) ≥ 0 có ảnh hưởng tích cực lên hoạt tính; ii) Ngược lại, các TSPT có CW < 0 có ảnh hưởng tiêu cực lên hoạt tính; và iii) Có CW nhận cả giá trị $>$ và < 0 là các TSPT với ảnh hưởng chưa rõ lên hoạt tính. Dựa trên phân loại này, cấu trúc các chất tốt nhất trong tập huấn luyện sẽ được thay đổi và dự đoán tác dụng nhằm tìm kiếm ứng viên mới, có thể có hoạt tính tốt hơn.

2.2.4. Đánh giá khả năng gắn kết với đích của các chất mới thiết kế bằng mô phỏng protein docking

Cấu trúc tinh thể nhiễu xạ tia X của phức hợp N1 neuraminidase và oseltamivir carboxylat (PDB ID: 4B7R) được thu thập từ ngân hàng dữ liệu protein [12]. Chuỗi A được chọn cho nghiên cứu, các bước chuẩn bị protein bao gồm: thêm hydro phân cực, loại nước, gắn trường lực, kiểm tra trạng thái ion hóa. Cấu tử đồng kết tinh sau đó được tách ra và dock lại vào trung tâm hoạt động của protein. Sau khi mô phỏng được trạng thái gắn kết như ban đầu của oseltamivir carboxylat (với độ lệch RMSD $< 1,00$ Å), các chất mới thiết kế được dock vào neuraminidase. Phần mềm Molegro Virtual Docker (MVD 2013.6.0.0 [Win32]) được sử dụng để dock các chất mới thiết kế vào trung tâm hoạt động của đích. Các tham số cài đặt bao gồm: số vòng lặp là 1500 với 20 quần thể tìm kiếm có kích thước là 50 cho mỗi lần chạy. Hàm tính điểm MolDock Score và tối ưu hóa Rerank Score được áp dụng tuần tự để dự đoán năng lượng gắn kết (-kcal/mol) giữa cấu tử với đích.

3. Kết quả và bàn luận

3.1. Kết quả xây dựng mô hình QSAR

Mô hình QSAR dự đoán hoạt tính ức chế enzym H1N1 influenza neuraminidase (IN) được xây dựng từ các hàm TF_i. Cấu trúc của 3 mô hình tốt nhất cùng các thông số thống kê được biểu diễn như Bảng 1.

$$pIC50 = 0,69(\pm 0,07) + 0,07(\pm 0,00) \times DCW(1,7) \quad (M1)$$

$$pIC50 = 3,43(\pm 0,08) + 0,04(\pm 0,00) \times DCW(1,5) \quad (M2)$$

$$pIC50 = 3,32(\pm 0,11) + 0,07(\pm 0,00) \times DCW(1,6) \quad (M3)$$

Độ chính xác của 3 mô hình phản ánh qua các thông số thống kê như trong Bảng 1. Có thể thấy, các mô hình có chất lượng khá tương đồng với độ chính xác tốt trong quá trình huấn luyện.

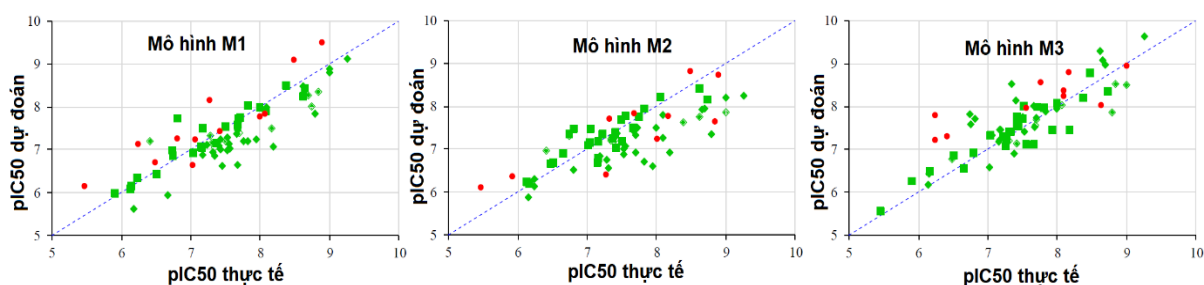
Tương ứng với hệ số xác định R^2 và Q^2 , thứ tự các mô hình như sau: M1 > M3 > M2. Bên cạnh đó, theo tham số Delta K (0,1), mô hình có tương quan tốt giữa các TSPT và hoạt tính với sai số dự đoán nhỏ ($s = 0,31-0,70$ và $MAE = 0,23-0,45$). Kết quả đánh giá bằng phương pháp Y-randomization (khi hoán đổi biến Y, R^2 giảm đột ngột về 0,03-0,06) và Y-scrambling ($R^2_{Y-scr} = 0,76-0,85$ cao tương đương $R^2_{training}$ và Q^2_{test}) cho thấy mô hình ổn định và tương quan tìm được hoàn toàn không phải là ngẫu nhiên. Từ đó, có thể nhận định 3 mô hình M1-3 đã đảm bảo yêu cầu về độ ổn định, chính xác và có thể ứng dụng trong dự đoán cũng như thiết kế hoạt chất mới [11].

Do 3 mô hình M1, M2 và M3 là các mô hình hợp với 5-7 biến, trong đó có sự trùng lặp về TSPT. Các tham số có tần số xuất hiện cao hơn được phân tích tương quan cấu trúc-tác dụng. Nhận xét về các TSPT được lựa chọn có thể thấy chúng là các tham số $SMILES_{DCW}(T^*, N^*)$ [9]. Các đặc điểm cấu trúc đáng chú ý bao gồm: số lượng nguyên tố oxy (-OCO-, -O-) và liên kết đôi (tham số +++O---B2==), liên kết nitơ với oxy (HN=CO-, -NO-, >NCO-) +++N---O==, sự hiện diện của dị vòng (1...., C...@....) và số lượng trung tâm lập thể (BOND101000) có tương quan dương với pIC₅₀. Kết quả này có nhiều điểm tương đồng với nghiên cứu của Yaeghoobi và

cộng sự khi chứng minh các nhóm thế có xu hướng cho liên kết hydro (NH, NH₂, OH) có tác động tích cực lên tác dụng ức chế enzym H1N1 neuraminidase [5]. Mercader và Pomilio cũng nhấn mạnh sự hiện diện của dị vòng với mạch nhánh có nhiều liên kết đôi liên hợp cũng giúp cơ chất dễ dàng gắn kết với trung tâm hoạt động của đích [4].

3.2. Thiết kế hợp chất mới

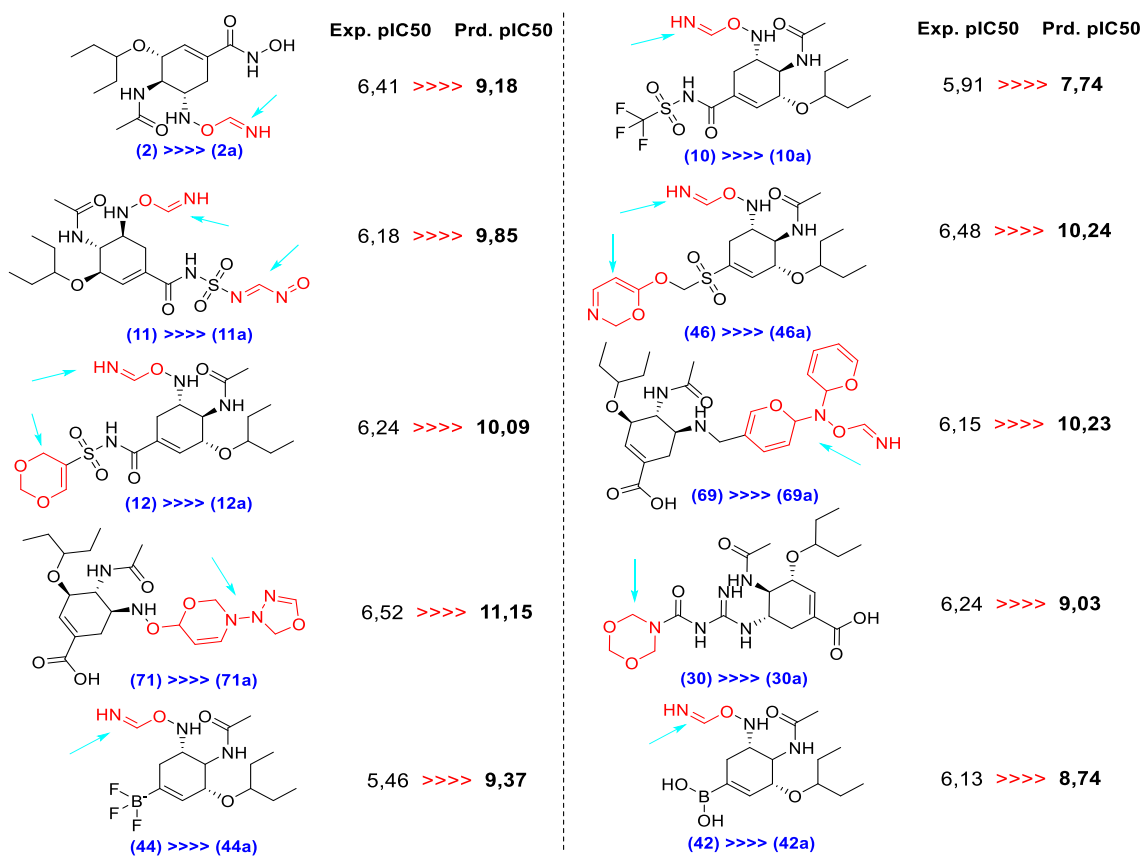
Dựa trên các đặc điểm cấu trúc rút ra từ các TSPT được chọn, 10 dẫn xuất mới của oseltamivir được thiết kế dựa trên một số chất có hoạt tính tốt trong CSDL (Hình 2). So sánh với giá trị pIC₅₀ của hợp chất gốc, hoạt tính dự đoán cho các chất mới đều tốt hơn đáng kể. Qua dự đoán, nhóm thế HN=CO- và hệ dị vòng chứa cấu trúc -OCO hoặc -OCN- có thể có đóng góp quan trọng giúp tăng hoạt tính. Một số hợp chất có cải thiện rõ rệt như **11a**, **12a**, **46a**, **69a** và **71a** có sự kết hợp giữa 2 nhóm cấu trúc nêu trên. Ví dụ như chất **69a**, việc áp dụng hệ dị vòng kết hợp cả nhóm cấu trúc NHCO và OCN giúp tăng hoạt tính >80% so với chất gốc **69** (pIC₅₀ thực nghiệm = 6.15). Ngoài ra, các chất khác có mức thay đổi thấp hơn khi chỉ áp dụng 1 nhóm thế (Hình 2).



Hình 2. Đồ thị biểu diễn tương quan giữa giá trị IC₅₀ thực tế và dự đoán của 3 mô hình.

Bảng 1. Các thông số phản ánh chất lượng của mô hình M1, M2 và M3

Mô hình	$R^2_{training}$	Q^2_{test}	s	MAE	F	Y-randomization	Y-scrambling
M1	0,869	0,854	0,547	0,454	167	0,030	0,854
M2	0,779	0,742	0,706	0,230	85	0,031	0,764
M3	0,828	0,804	0,313	0,247	121	0,069	0,793



Hình 3. Cấu trúc và cải thiện về hoạt tính của các chất mới thiết kế so với các chất gốc.

3.3. Mô phỏng protein Docking dự đoán khả năng gắn kết

Các dẫn chất mới được thiết kế sau đó được tiến hành dock vào trung tâm hoạt động của enzym H1N1 neuraminidase (PDB ID: 4B7R) nhằm xác định khả năng gắn kết của chúng với đích phân tử này. Bước đầu tiên trong đánh giá mô hình protein docking là dock lại cơ chất đồng tinh thể (oseltamivir carboxylat (OC)) vào đích. Kết quả của quá trình này thu được cấu trúc dock lại của OC gần như chồng lấn hoàn toàn trên cơ chất đồng tinh thể (độ lệch RMSD = 0,05Å) và các tương tác quan trọng với neuraminidase như tương tác hydro với Arg371, Tyr347 và Arg292 được bảo toàn. Bước tiếp theo là dock các chất mới vào vị trí gắn kết tương tự oseltamivir. Dựa trên kết quả docking có thể thấy 4 chất **11a**, **12a**, **46a** và **71a** có khả năng gắn kết chặt hơn với đích so với OC. Các chất này không chỉ có năng

lượng gắn kết tốt hơn với đích so với OC mà còn có khả năng tạo được nhiều tương tác với các amino acid đích hơn OC. Tất cả 4 chất đều tạo được 5 liên kết hydro với Arg152, Arg371, Arg292 và Tyr374 tương tự OC. Ngoài ra chúng còn có khả năng tạo tương tác pi-pi stacking với Trp178 và pi-alkyl stacking với Asp151, Glu277, Asn294 và Ser246. Bảng 2 mô tả kết quả docking của 4 chất này so với OC. Dựa trên các kết quả dự đoán bởi mô hình QSAR và protein docking, các dẫn xuất **11a**, **12a**, **46a** và **71a** có sự cải tiến về hoạt tính so với chất gốc dùng để thiết kế, chúng cũng thể hiện khả năng gắn đích tốt hơn so với oseltamivir carboxylate.

4. Kết luận và kiến nghị

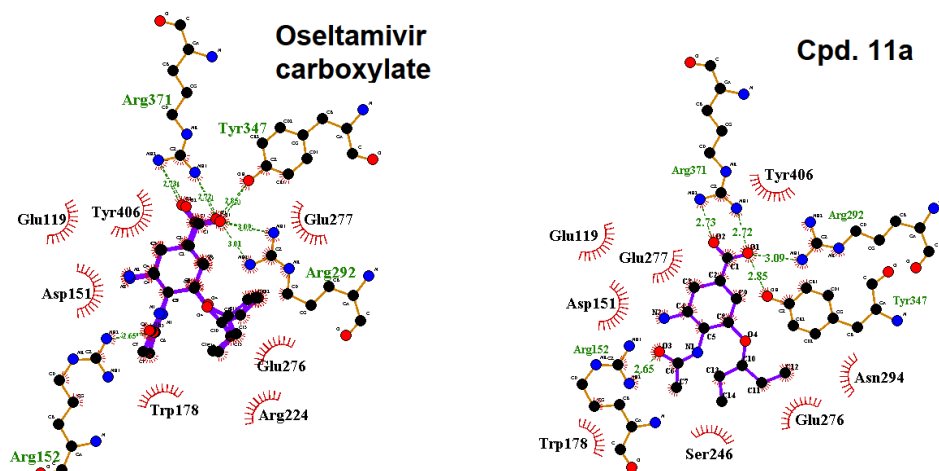
Trong nghiên cứu này, 3 mô hình QSAR dự đoán hoạt tính ức chế H1N1 influenza

neuraminidase đã được xây dựng thành công. Ảnh hưởng của các yếu tố cấu trúc lên hoạt tính sinh học đã được phân tích dựa trên các tham số $SMILES_{DCW}(T^*, N^*)$ tích hợp trong phần mềm Coralsea. Nghiên cứu cũng đã thiết được một số dẫn xuất mới của oseltamivir và dự đoán hoạt tính của chúng sử dụng các mô hình QSAR đã

xây dựng. Các chất mới thiết kế, đặc biệt là **11a**, **12a**, **46a** và **71a** được dự đoán với hoạt tính cải thiện hơn rõ rệt so với các chất gốc. Dựa trên mô phỏng protein docking, các chất này cũng cho thấy khả năng gắn kết tốt hơn với đích neuraminidase so với oseltamivir carboxylat.

Bảng 2. Điểm docking của các dẫn xuất của oseltamivir

Hợp chất	Năng lượng liên kết hydro	Năng lượng liên kết van der waals	Điểm docking (MolDock)	Tương tác với các amino acid
	kCal/mol			
11a	-18,75	-97,28	-162,60	Arg152, Arg371, Arg292, Tyr347, Glu119, Trp178, Glu227, Asn294, Ser246
12a	-13,70	-95,74	-141,66	Arg152, Arg371, Arg292, Tyr347, Trp178, Glu227
46a	-16,55	-75,88	-140,80	Arg152, Arg371, Tyr347, Arg292, Glu277, Trp178
71a	-5,64	-110,33	-147,18	Arg152, Arg371, Tyr347, Arg292, Glu227, Trp178,
Oseltamivir carboxylate	-12,39	-102,72	-124,35	Arg152, Arg371, Arg292, Tyr347, Glu119, Glu227



Hình 4. Cấu trúc và cải thiện về hoạt tính của chất **11a** so với các chất đối chứng.

Tài liệu tham khảo

- [1] G. Neumann et al., Emergence and Pandemic Potential of Swine-Origin H1N1 Influenza Virus. *Nature*, Vol. 459, 2009, pp. 931.
- [2] E. Takashima et al., Global Update on the Susceptibility of Human Influenza Viruses to Neuraminidase Inhibitors, 2013-2014. *Antiviral Research*, Vol. 117, 2015, pp. 27.
- [3] A. Moscona, Oseltamivir Resistance-Disabling Our Influenza Defenses. *New England Journal of Medicine*, Vol. 353, 2005, pp. 2633.
- [4] A. G. Mercader, A. B. Pomilio, QSAR Study of Flavonoids and Biflavonoids As Influenza H1N1

- Virus Neuraminidase Inhibitors. *European Journal of Medicinal Chemistry*, Vol. 45, 2010, pp. 1724.
- [5] M. Yaeghoobi et al., QSAR, *In Silico* Docking and *In Vitro* Evaluation of Chalcone Derivatives as Potential Inhibitors for H1N1 Virus Neuraminidase, *Medicinal Chemistry Research*, Vol. 25, 2016, pp. 2133.
- [6] Y. Lai et al., 3D-Quantitative Structure–Activity Relationship and Antiviral Effects of Curcumin Derivatives as Potent Inhibitors of Influenza H1N1 Neuraminidase. *Archives of Pharmacal Research*, Vol. 43, 2020, pp. 489.
- [7] Z. Li et al., Discovery of Acylguanidine Oseltamivir Carboxylate Derivatives as Potent Neuraminidase Inhibitors. *Bioorganic & Medicinal Chemistry*, Vol. 25, 2017, pp. 2772.
- [8] B. T. Hong et al., Boronate, Trifluoroborate, Sulfone, Sulfinat and Sulfonate Congeners of Oseltamivir Carboxylic Acid: Synthesis and Anti-Influenza Activity, *European Journal of Medicinal Chemistry*, Vol. 163, 2019, pp. 710.
- [9] J. Zhang et al., Optimization of N-Substituted Oseltamivir Derivatives As Potent Inhibitors of Group-1 and-2 Influenza A Neuraminidases, Including A Drug-Resistant Variant. *Journal of Medicinal Chemistry*, Vol. 61, 2018, pp. 6379.
- [10] A. A. Toropov, A. P. Toropova, Use of the Index of Ideality of Correlation to Improve Predictive Potential for Biochemical Endpoints, *Toxicology Mechanisms and Methods*, Vol. 29, 2019, pp. 43.
- [11] N. Chirico, P. Gramatica, Real External Predictivity of QSAR Models: How to Evaluate It? Comparison of Different Validation Criteria and Proposal of using the Concordance Correlation Coefficient, *Journal of Chemical Information and Modeling*, Vol. 51, 2011, pp. 2320.