

XÂY DỰNG MÔ HÌNH 2D-QSAR SÀNG LỌC CÁC CHẤT KHÁNG HEMOZOIN

Kim Ngọc Sơn^{1*}, Nguyễn Thị Tuyết Hà², Hà Thị Phương Linh³, Đỗ Thành Trung⁴, Huỳnh Hiếu Nghĩa⁵

¹Trường Đại học Lạc Hồng, số 10 Huỳnh Văn Nghệ, Bửu Long, TP. Biên Hòa, Đồng Nai, Việt Nam

²Bệnh viện Hoàn Mỹ Bình Phước, ấp 3, xã Tiến Hưng, TP. Đồng Xoài, Bình Phước, Việt Nam

³PKĐK Phúc An Khang, Số 5/678 ĐLBD, Kp. Hoà Lân 1, P. Thuận Giao, TP. Thuận An, Bình Dương, Việt Nam

⁴Trung tâm y tế huyện Hớn Quản, ấp 02, TT. Tân Khai, huyện Hớn Quản, Bình Phước, Việt Nam

⁵Nhà thuốc Thảo My, Việt Nam

*Tác giả liên hệ: sonkim@lhu.edu.vn

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận: 25/7/2024	Tập dữ liệu 592 chất, trong đó: 132 dẫn chất của quinolin, 75 dẫn chất của benzimidazol và 100 chất từ thư viện đại học Tokyo được sử dụng để xây dựng mô hình 2D-QSAR bằng phần mềm MOE 2014.09 sử dụng thuật toán bình phương tối thiểu từng phần. Kết quả xây dựng được 3 mô hình 2D-QSAR với các giá trị R ² , Q ² , RMSE, CCC lần lượt là 0,79; 0,76; 0,20; 0,88 cho mô hình QSAR dẫn chất quinolin (1); 0,79; 0,73; 0,10; 0,88 cho mô hình QSAR dẫn chất benzimidazol (2) và 0,79; 0,76; 0,13; 0,88 cho mô hình QSAR các chất từ thư viện đại học Tokyo (3). Kết quả sàng lọc ảo trong tập được liệu cổ truyền Trung Hoa cho thấy mô hình (1) có 228 chất (IC ₅₀ ≤ 10 μM), mô hình (2) có 5252 chất (IC ₅₀ ≤ 10 μM), mô hình (3) có 313 chất (IC ₅₀ ≤ 15 μM) có hoạt tính kháng hemozoin tốt.
Ngày hoàn thiện: 12/9/2024	
Ngày chấp nhận: 25/9/2024	
Ngày đăng: 11/10/2024	
TỪ KHÓA	
Kháng sốt rét;	
Kháng hemozoin;	
In silico;	
2D-QSAR.	

2D-QSAR MODEL USING PLS METHOD FOR ANTI-MALARIAL ACTIVITIES OF ANTI-HAEMOZOIN COMPOUNDS

Kim Ngọc Sơn^{1*}, Nguyen Thi Tuyen Ha², Ha Thi Phuong Linh³, Do Thanh Trung⁴, Huynh Hieu Nghia⁵

¹Lac Hong University, No.10, Huynh Van Nghe St., Buu Long Ward, Bien Hoa City, Dong Nai, Vietnam

²Hoan My Binh Phuoc Hospital, 3 Hamlet, Tien Hung Commune, Dong Xoai City, Binh Phuoc, Vietnam

³Phuc An Khang Clinic, No. 5/678 DLBD St., Hoa Lan 1 Town, Thuan Giao Ward, Thuan An City, Binh Duong, Vietnam

⁴Hon Quan District Medical Center, 2 Hamlet, Tan Khai Town, Hon Quan district, Binh Phuoc, Vietnam

⁵Thao My Pharmacy, Vietnam

*Corresponding Author: sonkim@lhu.edu.vn

ARTICLE INFO	ABSTRACT
Received: Jul 25 th , 2024	Databases used in 2D-QSAR model included 592 compounds with a wide variety of derivatives. Of these, 132 compounds were derivatives of quinolin, 75 compounds of benzimidazole and 100 compounds were from the University of Tokyo Library. Building 2D-QSAR models used the partial least squares algorithm (PLS) in MOE 2014.09 software. Three 2D-QSAR models were built based on 2D molecular descriptors parameters with R ² , Q ² , RMSE, CCC values which were respectively 0,79; 0,76; 0,20; 0,88 for QSAR model derivatives of quinolin (1); 0,79; 0,73; 0,10; 0,88 for QSAR model derivatives of benzimidazole (2) and 0,79; 0,76; 0,13; 0,88 for QSAR model of compounds from the University of Tokyo Library (3). Virtual screening from 57423 compounds of TCM database by three 2D-QSAR models showed that 228 compounds (IC ₅₀ ≤ 10 μM) corresponded with the model (1), 5252 compounds (IC ₅₀ ≤ 10 μM) corresponded with the model (2), 313 compounds (IC ₅₀ ≤ 15 μM) corresponded with the model (3) having highly anti-haemozoin activities.
Revised: Sep 12 nd , 2024	
Accepted: Sep 25 th , 2024	
Published: Oct 11 st , 2024	
KEYWORDS	
Antimalarial;	
Anti-haemozoin;	
In silico;	
2D-QSAR.	

Doi: <https://doi.org/10.61591/jslhu.18.290>

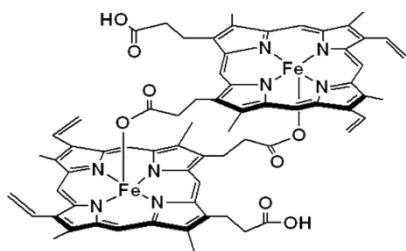
Avaiable online at: <https://js.lhu.edu.vn/index.php/lachong>

1. GIỚI THIỆU

Sốt rét là bệnh xảy ra nhiều ở các nước nhiệt đới, bao gồm Việt Nam, do muỗi *Anopheles* truyền bệnh. Theo báo cáo năm 2022 của Tổ Chức Y Tế Thế Giới (WHO), có 247 triệu ca mắc bệnh sốt rét trên thế giới, 619.000 ca tử vong, chủ yếu do ký sinh trùng *Plasmodium falciparum* (*P. falciparum*) gây ra [1]. Năm 2023, Việt Nam có 448 ca mắc sốt rét. Trong đó, 2 ca tử vong, gần 7 triệu người sống trong vùng bệnh sốt rét lưu hành, xuất hiện nhiều ổ sốt rét mới có nguy cơ lan rộng [2].

Tình hình đề kháng các thuốc điều trị sốt rét ngày càng gia tăng. Đầu thế kỉ XX đã xuất hiện các chủng sốt rét kháng quinin. Những năm 1950, các chủng kháng cloroquin đã xuất hiện và lan nhanh khắp thế giới. Một số nước Đông Nam Á đã xuất hiện chủng đề kháng với dihydroartemisinin – piperaquin. Năm 2018 – 2019 đã phát hiện các chủng *P. falciparum* mang gen đột biến kháng dihydroartemisinin – piperaquin là *Pfkelch13 C580Y* và *plasmepsin 2/3* với tỉ lệ 17,9% [3].

Hemozoin là sản phẩm từ sự tiêu hóa hồng cầu của *P. falciparum*. Hemoglobin trong không bào tiêu hóa của *P. falciparum* dưới xúc tác của enzym *Aspartat Protease* được chuyển thành globin và hem. Hem tự do độc với ký sinh trùng, vì vậy nhân hem được polymer hóa thành hemozoin không độc. Vì sự chuyển hóa hemozoin là điều thiết yếu cho sự sống còn của *P. falciparum*, nên ức chế quá trình giải độc hem – kháng hemozoin là đích nghiên cứu đang được quan tâm hiện nay. Một số thuốc sốt rét được sử dụng như: cloroquin, quinin và mefloquin được chứng minh có hoạt tính diệt *P. falciparum* do ức chế quá trình giải độc hem [4].



Hình 1. Cấu trúc phân tử hemozoin [4]

Nhiều công trình nghiên cứu trên thế giới về các hợp chất kháng hemozoin được công bố. Năm 2018, NT Huy và cộng sự đã xây dựng được mô hình dự đoán hoạt tính kháng hemozoin bằng BMA (Bayesian Modeling Average) có phương trình: $\text{logit (probability)} = -23,62 + 0,592 \cdot \text{logP} + 0,351 \cdot N_1 + 1,322 \cdot N_3$ (probability là xác suất có hoạt tính kháng hemozoin của các chất, N_1 là số nguyên tử cho liên kết hydro và N_3 là chỉ số khúc xạ phân tử) [5]. Năm 2023, Bharvi Sharma và cộng sự đã tổng hợp được một số dẫn chất quinolin hydrazone. Các chất này được tiến hành thử hoạt tính kháng hemozoin trên chủng *P. falciparum* W2, kết quả cho thấy nhóm thế axit pyrazin-2-carboxylic có hoạt tính tốt nhất với $\text{IC}_{50} = 0,26$ (μM) [6].

Để hỗ trợ cho nghiên cứu thuốc, phương pháp nghiên cứu trên máy tính (*in silico*) thông qua các mô hình sàng lọc như mô hình gắn kết phân tử (molecular docking), mô hình các điểm gắn kết (pharmacophore) hay mô hình liên quan định tính hoặc định lượng giữa cấu trúc và tác động (quantitative or qualitative structure-activity

relationships) dạng hai chiều (two dimension-2D) hay ba chiều (three dimension-3D),...được xây dựng đã giúp tìm kiếm nhanh các hợp chất tiềm năng trong thời gian ngắn nhất với chi phí thấp và hiệu quả cao.

Đề tài “Xây dựng mô hình 2D-QSAR sàng lọc các chất kháng hemozoin” được thực hiện với mục tiêu chính là dự đoán hoạt tính kháng hemozoin có tác dụng điều trị sốt rét thông qua mô hình 2D-QSAR của các chất kháng hemozoin đã biết, đặc biệt trên cơ sở dữ liệu là những chất có nguồn gốc tự nhiên như các thuốc cổ truyền Trung Hoa (TCM database) [7]. Các chất tiềm năng thu được có thể sử dụng để phát triển thành thuốc điều trị sốt rét mới trong tương lai.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1 Cơ sở dữ liệu

Cơ sở dữ liệu 592 chất có giá trị IC_{50} kháng hemozoin gồm tập dữ liệu 224 chất sàng lọc bằng HTS từ thư viện đại học Tokyo [5] và 368 chất được thu thập từ các bài báo [8-14].

Dựa vào chất đối chiếu có thể quy đổi giá trị hoạt tính kháng hemozoin của các chất theo quy tắc tam suất nếu chênh lệch giữa các giá trị hoạt tính của chất đối chiếu (như cloroquin) trong các bài báo khác nhau từ năm lần trở lên. Do hoạt tính sinh học IC_{50} có khoảng chênh lệch với nhau nhiều nên khi tính toán được đổi thành pIC_{50} với $\text{pIC}_{50} = -\log(\text{IC}_{50})$ để phù hợp cho việc xây dựng phương trình tuyến tính quan hệ cấu trúc – tác dụng.

Cơ sở dữ liệu tập chất và giá trị pIC_{50} được upload <https://drive.google.com/file/d/1bHqcCqYiGwmsYOarx6LJoVnEKZY2gHR/view?usp=sharing>

2.2 Xây dựng mô hình 2D-QSAR

Mô hình 2D-QSAR được xây dựng bằng phần mềm MOE 2014.09. Quá trình gồm 6 bước: [15]

- Xây dựng cơ sở dữ liệu bằng công cụ *Builder* (MOE 2014.09) và lưu dưới dạng *Database* với các định dạng: mol: molecule; name: char; pIC_{50} : float. Mỗi bài báo được lưu dưới dạng mỗi *Database* riêng và cơ sở dữ liệu là tổng tất cả các *Database* được gộp lại thông qua công cụ *Merge Databases* (MOE 2014.09).
- Tính toán các thông số mô tả phân tử 2D (MOE 2014.09): các cấu trúc được tối thiểu hóa năng lượng nhờ công cụ *Energy minimize* với thông số gradient là 0,0001 mV. Công cụ *Compute/Descriptors* phân tích 192 thông số mô tả 2D thể hiện cấu trúc hóa học, tính chất vật lý của một phân tử bằng các giá trị số, có ảnh hưởng tới tác dụng sinh học.
- Phân chia tập hợp thành 2 tập: tập huấn luyện (80%) và tập kiểm tra (20%) dựa trên mức độ đa dạng về cấu trúc của các chất trong tập dữ liệu ban đầu. Sử dụng công cụ *Diverse subset* hoặc *Calculator/RAND* (MOE 2014.09).
- Lựa chọn thông số mô tả phân tử: loại các thông số có > 20% giá trị bằng 0 (Excel 2010), lọc các thông số có độ lệch chuẩn bằng 0 (Rapid Miner 5.3.013), loại thông số tương quan $\geq 0,40$ (Rapid Miner 5.3.013), lựa chọn các thông số mô tả phù hợp nhất để xây dựng mô hình (Weka 3.7.1);
- Các chất có Z-score > 2 được loại. Xây dựng mô hình bằng MOE 2014.09. Công cụ *QSAR-Model: Activity*

Field chọn pIC_{50} và Method chọn PLS. Nhấn Fit để thực hiện các lệnh hồi quy, các giá trị RMSE và R^2 được hiển thị. Phương trình hồi quy được ghi nhận bằng lệnh Report trong cửa sổ QSAR-Model. Mô hình được lưu bằng Save dưới dạng *.fit;

- Đánh giá mô hình 2D-QSAR thông qua đánh giá nội LOO (Validate trong MOE 2014.09), đánh giá ngoại (Model evaluate trong MOE 2014.09), đánh giá Roy (Excel 2010), Hệ số tương quan phù hợp CCC (Excel 2010).

2.3 Ứng dụng mô hình 2D-QSAR

Mô hình xây dựng được dùng để dự đoán hoạt tính kháng *P. falciparum* trên cơ sở dữ liệu TCM [7]. Đầu tiên, các chất được loại theo quy luật Năm Lipinski: tối thiểu hóa năng lượng bằng công cụ Energy minimize (MOE 2014.09) và tính toán 4 thông số mô tả: weight (khối lượng phân tử), a_don (số lượng các nhóm cho liên kết hydro), a_acc (số lượng các nhóm nhận liên kết hydro) và logP (hệ số phân bố giữa octanol/ nước). Các chất không thỏa được loại bỏ, các chất còn lại được chia tỷ lệ thông số mô tả bằng công cụ Normalize (Rapidminer 5.3.013). Sau đó, dùng công cụ model evaluate (MOE 2014.09) để dự đoán hoạt tính sinh học.

3. KẾT QUẢ VÀ BÀN LUẬN

3.1 Khảo sát xây dựng mô hình 2D-QSAR trên các chất kháng hemozoin

Cơ sở dữ liệu 592 chất, sau khi chọn lọc thông số mô tả, từ 192 thông số giữ lại 4 thông số là a_aro, b_1rotR, slogP_VSA0, Q_VSA_NEG. Dùng các thông số này để xây dựng mô hình 2D-QSAR. Kết quả sơ bộ, $R^2 = 0,20$; RMSE = 0,39. Kết quả không đạt cho một mô hình QSAR. Nghiên cứu được tiếp tục bằng cách chia nhỏ tập dữ liệu ban đầu thành các tập:

- Dữ liệu 224 chất sàng lọc bằng nghiên cứu sàng lọc High Throughput Screening (HTS) từ thư viện đại học Tokyo (a).
- Dữ liệu các chất có hoạt tính kháng hemozoin có cấu trúc giống nhau hoặc tương tự nhau và xây dựng từng mô hình 2D-QSAR trên mỗi nhóm cấu trúc này (b).

Với cách chia a), tập xây dựng gồm 179 chất theo phân chia đa dạng, chọn được 6 thông số: BCUT_SMR_2, b_double, b_1rotR, logS, PEOE_VSA-0, PEOE_VSA-6. Kết quả sơ bộ, $R^2 = 0,28$; RMSE = 0,29. Như vậy, với cách chia này, mô hình 2D-QSAR xây dựng được không đạt. Tiếp tục tiến hành chia tập dữ liệu thành nhiều tập chất theo giá trị IC_{50} trong đó có một tập chất thỏa mô hình 2D-QSAR.

Với cách chia b), trên nhóm dẫn chất quinolin (132 chất), 5 thông số mô tả được chọn: a_aro, a_don, a_nCl, b_double, PEOE_VSA+2, vsa_acc. Kết quả sơ bộ $R^2 = 0,69$; RMSE = 0,24. Trên nhóm dẫn chất indol (77 chất), chọn được 7 thông số mô tả: GCUT_PEOE_1, PEOE_PC-, PEOE_VSA-2, PEOE_VSA-5, petitjean, SlogP_VSA4, SMR_VSA3; $R^2 = 0,38$; RMSE = 0,36. Trên nhóm dẫn chất benzimidazol (75 chất), chọn được 5 thông số mô tả: apol, PEOE_RPC+, Q_VSA-FHYD, slogP+VSA4, SMR_VSA7; $R^2 = 0,65$; RMSE = 0,11. Các chất còn lại thuộc nhiều nhóm hóa học khác nhau có

số chất trong nhóm ít nên không xây dựng mô hình. Các mô hình trên có 2 mô hình đạt yêu cầu là mô hình có tính dự đoán tương đối tốt nhưng R^2 chưa cao là mô hình trên các dẫn chất quinolin và benzimidazol.

Nghiên cứu được tiến hành tiếp tục trên mô hình 2D-QSAR của 2 nhóm cấu trúc này.

3.2 Mô hình 2D-QSAR trên các chất kháng hemozoin nhóm dẫn chất quinolin

Tập dữ liệu bao gồm 132 chất từ các bài báo khoa học [8-11] và cùng phương pháp thử nghiệm hoạt tính Pandey. 125 Chất được dùng để xây dựng mô hình 2D-QSAR, 7 chất gây nhiễu được loại. 6 thông số mô tả được lựa chọn và tương quan giữa các thông số này với nhau và với pIC_{50} được xác định (bảng 1).

Từ 6 thông số mô tả được chọn, tiến hành xây dựng mô hình 2D-QSAR bằng phương pháp bình phương tối thiểu từng phần. Kết quả thu được một mô hình có khả năng dự đoán (1).

$$pIC_{50} = 4,86898 - 0,53568*a_{aro} - 0,49008*a_{don} + 0,50595*a_{nCl} + 0,42972*b_{double} - 0,53935*PEOE_VSA+2 - 0,31186*vsa_acc \quad (1)$$

Kết quả chi tiết về mô hình được trình bày ở bảng 2.

Đánh giá trên mô hình 2D-QSAR cho thấy phương trình đạt mức độ tương quan cao ($R^2 = 0,79 > 0,5$), độ lệch trung bình giữa giá trị dự đoán và giá trị thực nghiệm thấp (RMSE = 0,2 < 0,5). Theo đánh giá Roy, mô hình đạt ($\overline{r_m^2} = 0,70 \geq 0,5$; $\Delta r_m^2 = 0,18 \leq 0,2$). Đánh giá thông qua hệ số tương quan phù hợp CCC = 0,88 > 0,85 lý tưởng cho một mô hình QSAR.

Giá trị dự đoán của các chất hầu như đều nằm trong khoảng 95% dự đoán đúng của mô hình (hình 2).

Bảng 1. Mối tương quan giữa các thông số xây dựng mô hình 2D-QSAR trên các chất kháng hemozoin nhóm quinolin

ST T	Tên	1	2	3	4	5	6
1	a_aro	1					
2	a_don	0,33	1				
3	a_nCl	0,16	-0,33	1			
4	b_double	-0,22	-0,13	-0,07	1		
5	PEOE_VSA+2	-0,08	0,36	-0,03	-0,25	1	
6	vsa_acc	0,25	0,02	-0,38	0,03	0,14	1
7	pIC_{50}	-0,41	-0,59	0,46	0,5	-0,4	-0,41

Chú thích:

a_aro: tổng số nhân thơm

a_don: tổng số nguyên tử cho liên kết hidro

a_nCl: tổng số nguyên tử clor

b_double: tổng số liên kết đôi (ngoài trừ nhân thơm)

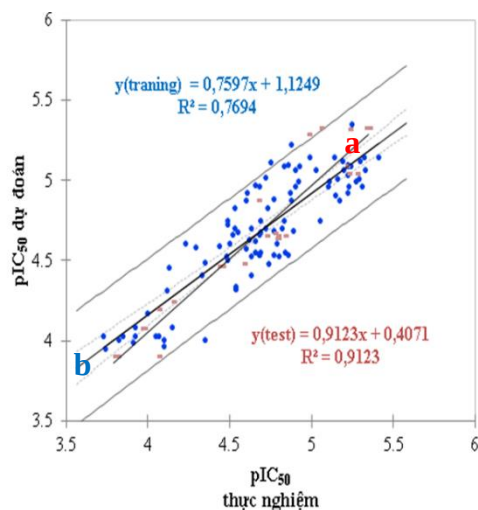
PEOE_VSA+2: tổng diện tích bề mặt Van der Waals có diện tích trong vùng [0,10; 0,15)

vsa_acc: bề mặt Van der Waals có nguyên tử nhận H, trừ acid và nguyên tử cho nhận (OH)

Nhận xét mô hình

Các thông số đánh giá đều đạt cho thấy đây là một mô hình có khả năng dự đoán hoạt tính kháng hemozoin tốt.

Các thông số mô tả quan trọng: b_double, a_aro liên quan đến số liên kết đôi và nhân thơm; a_nCl liên quan đến số nguyên tử clor và a_don liên quan đến khả năng cho liên kết hidro trong phân tử. Cho thấy tầm quan trọng của số liên kết đôi, nhân thơm, số nguyên tử clo và khả năng cho liên kết hidro của hợp chất với hoạt tính kháng hemozoin. Điều này phù hợp với công thức cấu tạo của cloroquin (chất kháng sốt rét có hoạt tính kháng hemozoin). PEOE_VSA+2, vas_acc liên quan đến diện tích bề mặt Van der Waals của phân tử cho thấy tầm quan trọng của liên kết Van der Waals đối với hoạt tính kháng hemozoin. Từ phương trình pIC_{50} cho thấy hoạt tính kháng hemozoin của phân tử càng mạnh khi có nhiều nhóm thế clor và nhiều nối đôi.



Hình 2. Đồ thị tương quan giá trị pIC_{50} thực nghiệm và pIC_{50} dự đoán từ mô hình 2D-QSAR trên các chất kháng hemozoin nhóm quinolin

Chú thích: (a) – tập kiểm tra

(b) – tập xây dựng

Bảng 2. Kết quả mô hình 2D-QSAR trên các chất kháng hemozoin nhóm quinolin

	Tập huấn luyện	Đánh giá chéo	Tập kiểm tra	Toàn tập
N (số chất)	99	99	26	125

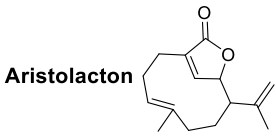
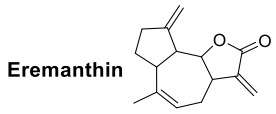
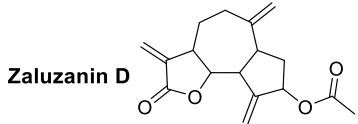
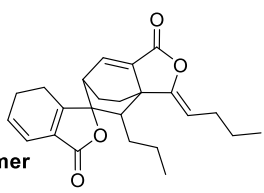
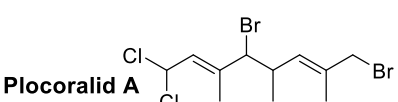
RMSE	0,20	0,21	0,14	0,20
R ²	0,76	0,73	0,91	0,79
r_m^2	0,76		0,91	0,79
$r_m'^2$	0,57		0,83	0,61
$\overline{r_m^2}$	0,67		0,87	0,70
Δr_m^2	0,19		0,08	0,18
CCC			0,94	0,88

Dự đoán hoạt tính kháng hemozoin nhóm quinolin thông qua mô hình 2D-QSAR

Ứng dụng mô hình 2D-QSAR xây dựng được vào dự đoán các chất trong thư viện 57.423 chất từ cơ sở dữ liệu TCM [7]. Kết quả có 228 chất được dự đoán có hoạt tính kháng hemozoin với $IC_{50} \leq 10 \mu M$.

Cấu trúc của một số chất có hoạt tính dự đoán tốt được trình bày ở **bảng 3**.

Bảng 3. Một số chất có hoạt tính dự đoán tốt từ mô hình 2D-QSAR dẫn chất quinolin trong thư viện TCM

STT	Công thức cấu tạo	IC_{50} dự đoán (μM)
1	 Aristolacton	10
2	 Eremanthin	10
3	 Zaluzanin D	9,5
4	 Ligustilid dimer	8,71
5	 Plocoralid A	5,25

Các chất có IC_{50} dự đoán tốt từ mô hình 2D-QSAR nhóm dẫn chất quinolin chủ yếu thuộc nhóm cấu trúc: alkaloid,

lacton và terpenoid, trong cấu trúc có một hoặc nhiều nhóm halogen (brom, clor). Trong đó, một số chất đã được nghiên cứu thử nghiệm hoạt tính kháng sốt rét. Plocoralid A và một số monoterpen khác được chiết từ *Plocamium corallorhiza* (Plocamiaceae) có hoạt tính ức chế *P. falciparum* với IC_{50} thấp hơn cloroquin. Cao chiết ethanol từ lá *Vernonia amygdalina* (Compositae) (thành phần hóa học có zaluzanin D) có hoạt tính ức chế *P. falciparum* đề kháng cloroquin 3D7 với $IC_{50} = 9,82 \mu\text{g/ml}$. Ligustilid dimer được tìm thấy trong rễ đương quy *Angelica sinensis* (Apiaceae), đương quy được chứng minh có nhiều tác dụng như kháng viêm, tăng huyết áp, kháng oxy hóa, kháng kết tập tiểu cầu [16].

3.3 Mô hình 2D-QSAR trên các chất kháng hemozoin nhóm dẫn chất benzimidazol

Tập dữ liệu bao gồm 75 chất từ các bài báo khoa học [12-14] và cùng phương pháp thử nghiệm hoạt tính là phương pháp Sandlin. 71 Chất được dùng để xây dựng mô hình 2D-QSAR, 4 chất gây nhiễu được loại. 5 thông số mô tả được lựa chọn và tương quan giữa các thông số này với nhau và với pIC_{50} được xác định (bảng 4).

Từ 5 thông số mô tả được chọn, tiến hành xây dựng mô hình 2D-QSAR bằng phương pháp bình phương tối thiểu từng phần. Kết quả thu được một mô hình có khả năng dự đoán tốt (2).

$$pIC_{50} = 4,08183 + 0,54961*apol + 0,59011*Q_VSA_FHYD - 0,43199*PEOE_RPC+ + 0,17921*SlogP_VSA4 + 0,09700*SMR_VSA7 \quad (2)$$

Kết quả chi tiết về mô hình được trình bày ở bảng 5.

Đánh giá trên mô hình 2D-QSAR cho thấy phương trình đạt mức độ tương quan cao ($R^2 = 0,79 > 0,5$), độ lệch trung bình giữa giá trị dự đoán và giá trị thực nghiệm thấp ($RMSE = 0,1 < 0,5$). Theo đánh giá Roy, mô hình đạt ($r_m^2 = 0,70 \geq 0,5$; $\Delta r_m^2 = 0,18 \leq 0,2$). Đánh giá thông qua hệ số tương quan phù hợp CCC = 0,88 > 0,85 lý tưởng cho một mô hình QSAR.

Bảng 4. Mối tương quan giữa các thông số xây dựng mô hình 2D-QSAR trên các chất kháng hemozoin nhóm benzimidazol

STT	Tên	1	2	3	4	5	6
1	apol	1					
2	PEOE_ RPC+	-0,31	1				
3	Q_VSA_ FHYD	0,23	-0,16	1			
4	SlogP_ VSA4	0,08	-0,08	0,13	1		
5	SMR_V SA7	0,04	-0,10	0,33	-0,09	1	
6	pIC_{50}	0,60	-0,59	0,64	0,28	0,29	1

Chú thích:

apol: tổng số nguyên tử phân cực

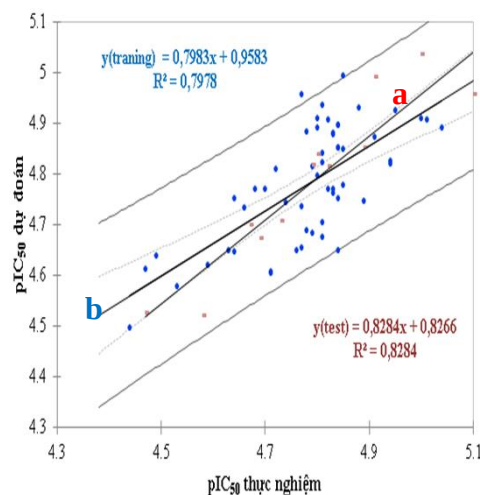
PEOE_RPC+: tổng diện tích dương phân cực: tỷ số giữa diện tích dương lớn nhất q_i với diện tích dương q_i

Q_VSA_FHYD: tổng diện tích bề mặt kỵ nước Van der Waals

SMR_VSA7: tổng diện tích bề mặt Van der Waals có độ khúc xạ trong vùng > 0,56

SlogP_VSA4: tổng diện tích bề mặt có thể tạo liên kết Van der Waals khi $\log P(o/w)$ nằm trong khoảng (0,1; 0,15]

Giá trị dự đoán của các chất hầu như đều nằm trong khoảng 95% dự đoán đúng của mô hình (hình 3).



Hình 3. Đồ thị tương quan giá trị pIC_{50} thực nghiệm và pIC_{50} dự đoán từ mô hình 2D-QSAR trên các chất kháng hemozoin nhóm benzimidazol

Chú thích: (a) – tập kiểm tra

(b) – tập xây dựng

Bảng 5. Kết quả mô hình 2D-QSAR trên các chất kháng hemozoin nhóm benzimidazol

	Tập huấn luyện	Đánh giá chéo	Tập kiểm tra	Toàn tập
N (số chất)	56	56	15	71
RMSE	0,09	0,11	0,26	0,10
R^2	0,80	0,72	0,94	0,79
r_m^2	0,80		0,94	0,79
$r_m'^2$	0,62		0,88	0,61
$\overline{r_m^2}$	0,71		0,91	0,70
Δr_m^2	0,18		0,06	0,18
CCC			0,97	0,88

Nhận xét mô hình

Các thông số đánh giá đều đạt cho thấy đây là một mô hình có khả năng dự đoán hoạt tính kháng hemozoin tốt.

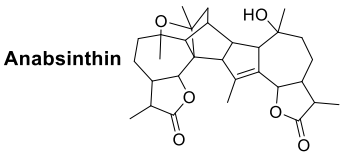
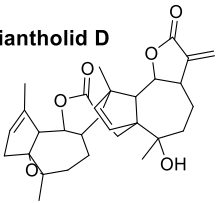
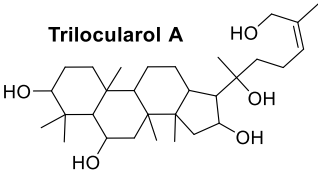
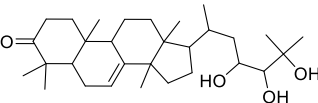
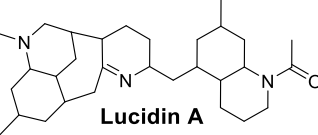
Các thông số mô tả Q_VSA_FHYD, SMR_VSA7, SlogP_VSA4 liên quan đến diện tích bề mặt Van der Waals của phân tử (độ khúc xạ, kỵ nước, chỉ số octan/nước) cho thấy tầm quan trọng của liên kết Van der Waals đối với hoạt tính kháng hemozoin. Các thông số apol, PEOE_RPC+ liên quan đến tính phân cực, tích điện tích dương của phân tử. Từ phương trình pIC_{50} cho thấy phân tử có hoạt tính kháng hemozoin tăng khi có nhiều nguyên tử phân cực và khả năng tạo liên kết Van der Waals.

Dự đoán hoạt tính kháng hemozoin nhóm benzimidazol thông qua mô hình 2D-QSAR

Ứng dụng mô hình 2D-QSAR xây dựng được vào dự đoán các chất trong thư viện 57.423 chất từ cơ sở dữ liệu TCM [7]. Kết quả có 5252 chất được dự đoán có hoạt tính kháng hemozoin với $IC_{50} \leq 10 \mu M$.

Cấu trúc của một số chất có hoạt tính dự đoán tốt nhất được trình bày ở **bảng 6**.

Bảng 6. Một số chất có hoạt tính dự đoán tốt từ mô hình 2D-QSAR dẫn chất benzimidazol trong thư viện TCM

STT	Công thức cấu tạo	IC_{50} dự đoán (μM)
1	 Anabsinthin	5,89
2	 Artemyriantholid D	6,76
3	 Trilocularol A	5,37
4	 Piscidinol A	5,5
5	 Lucidin A	5,89

Các chất có IC_{50} dự đoán tốt từ mô hình 2D-QSAR nhóm dẫn chất benzimidazol chủ yếu thuộc nhóm cấu trúc: alkaloid, steroid, triterpen và sesquiterpen. Một số cây được sử dụng để trị sốt rét ở châu Phi là *Turraea robusta* và *Turraea nilotica* (trong thành phần hóa học có piscidinol A và một số triterpen có cấu trúc tương tự) được chứng minh có hoạt tính kháng sốt rét [17].

3.4 Mô hình 2D-QSAR các chất kháng hemozoin bằng HTS từ thư viện đại học Tokyo

Tập dữ liệu bao gồm 100 chất từ cơ sở dữ liệu thử nghiệm hoạt tính kháng hemozoin với giá trị $IC_{50} > 20 \mu M$ của đại học Tokyo bằng phương pháp Sandlin. 4 thông số mô tả được chọn và tương quan giữa các thông số này với nhau và với pIC_{50} được xác định (**bảng 7**).

Từ 4 thông số mô tả được chọn, tiến hành xây dựng mô hình 2D-QSAR bằng phương pháp bình phương tối thiểu từng phần. Kết quả thu được một mô hình có khả năng dự đoán tốt (3).

$$pIC_{50} = 4,07221 - 0,49118*b_{1rotR} + 1,04030*logS - 0,00654*PEOE_VSA-0 - 0,47361*PEOE_VSA-6 \quad (3)$$

Kết quả chi tiết về mô hình được trình bày ở **bảng 8**.

Đánh giá trên mô hình 2D-QSAR cho thấy phương trình đạt mức độ tương quan cao ($R^2 = 0,79 > 0,5$; $RMSE = 0,13 < 0,5$). Theo đánh giá Roy, mô hình đạt ($r_m^2 = 0,70 \geq 0,5$; $\Delta r_m^2 = 0,19 \leq 0,2$). Đánh giá CCC = $0,88 > 0,85$ lý tưởng cho một mô hình QSAR.

Bảng 7. Mối tương quan giữa các thông số xây dựng mô hình 2D-QSAR trên các chất kháng hemozoin từ thư viện ĐH Tokyo

ST T	Tên	1	2	3	4	5
1	b_1rotR	1				
2	logS	-0,04	1			
3	PEOE_VSA-0	0,10	0,37	1		
4	PEOE_VSA-6	-0,04	-0,02	-0,09	1	
5	pIC_{50}	-0,31	0,76	0,31	-0,33	1

Chú thích:

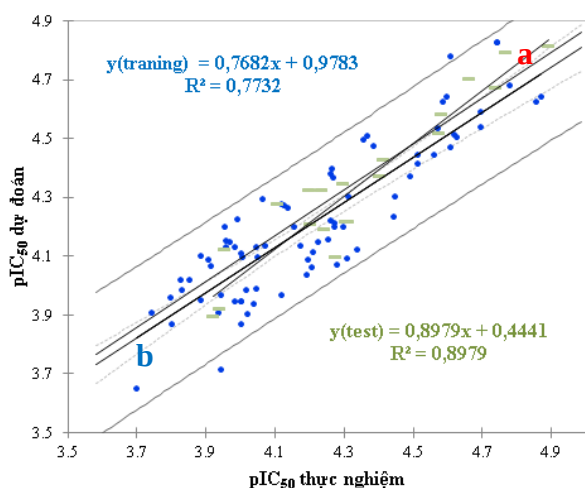
b_1rotR: tỉ lệ liên kết đơn quay được

logS: Log của độ tan trong nước có giá trị (-1, 1)

PEOE_VSA-0: diện tích bề mặt VDW với chỉ số điện tích từng phần trong khoảng [-0,05; 0,00]

PEOE_VSA-6: diện tích bề mặt VDW với chỉ số điện tích từng phần trong khoảng [-0,05; 0,00]

Giá trị dự đoán của các chất hầu như đều nằm trong khoảng 95% dự đoán đúng của mô hình (**hình 4**).



Hình 4. Đồ thị tương quan giá trị pIC_{50} thực nghiệm và pIC_{50} dự đoán từ mô hình 2D-QSAR trên các chất kháng hemozoin từ thư viện đại học Tokyo

Chú thích: (a) – tập kiểm tra

(b) – tập xây dựng

Bảng 8. Kết quả mô hình 2D-QSAR trên các chất kháng hemozoin từ thư viện đại học Tokyo

	Tập huấn luyện	Đánh giá chéo	Tập kiểm tra	Toàn tập
N (số chất)	80	80	20	100
RMSE	0,14	0,15	0,09	0,13
R^2	0,76	0,73	0,90	0,79
r_m^2	0,76		0,90	0,79
$r_m'^2$	0,56		0,80	0,60
$\overline{r_m^2}$	0,66		0,85	0,70
Δr_m^2	0,20		0,10	0,19
CCC			0,94	0,88

Nhận xét mô hình

Các thông số đánh giá đều đạt cho thấy đây là một mô hình có khả năng dự đoán hoạt tính kháng hemozoin tốt.

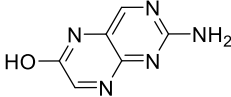
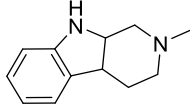
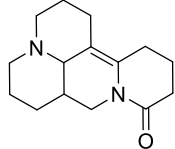
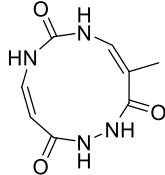
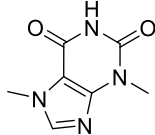
Các thông số mô tả quan trọng: PEOE_VSA-6, logS, PEOE_VSA-0, b_1rotR có liên quan đến diện tích bề mặt Van der Waals của phân tử. Điều này cho thấy tầm quan trọng của việc tạo liên kết Van der Waals đối với hoạt tính kháng hemozoin. Thông số b_1rotR liên quan đến số liên kết đơn quay được, logS liên quan đến độ tan của phân tử trong nước. Từ phương trình pIC_{50} cho thấy phân tử càng có ít liên kết đơn quay được và kém tan trong nước thì hoạt tính kháng hemozoin càng mạnh.

Dự đoán hoạt tính kháng hemozoin thông qua mô hình 2D-QSAR

Ứng dụng mô hình 2D-QSAR xây dựng được vào dự đoán các chất trong thư viện 57.423 chất từ cơ sở dữ liệu TCM [7]. Kết quả có 313 chất được dự đoán có hoạt tính kháng hemozoin với $IC_{50} \leq 15 \mu M$.

Cấu trúc của một số chất có hoạt tính dự đoán tốt nhất được trình bày ở **bảng 9**.

Bảng 9. Một số chất có hoạt tính dự đoán tốt từ mô hình 2D-QSAR đại học Tokyo trong thư viện TCM

STT	Công thức cấu tạo	IC_{50} dự đoán (μM)
1		14,79
2		14,45
3		13,8
4		12,3
5		12,59

Các chất có IC_{50} dự đoán tốt từ mô hình 2D-QSAR từ thư viện Tokyo chủ yếu là các phân tử nhỏ, trong cấu trúc có oxy hoặc nitơ.

4. KẾT LUẬN VÀ KIẾN NGHỊ

4.1 Kết luận

Đề tài đã nghiên cứu ứng dụng thành công 3 mô hình 2D-QSAR trên các chất kháng hemozoin gồm tập hợp 125 dẫn chất của quinolin, 71 dẫn chất của benzimidazol và 100 chất sàng lọc bằng HTS từ thư viện đại học Tokyo bằng phương pháp PLS. Kết quả mô hình dẫn chất của quinolin cho ra phương trình với 6 thông số mô tả, có các giá trị $R^2 = 0,79$; $RMSE = 0,2$; $Q^2_{LOO} = 0,76$. Mô hình dẫn chất của benzimidazol cho ra phương trình với 5 thông số mô tả, có các giá trị $R^2 = 0,79$; $RMSE = 0,1$; $Q^2_{LOO} = 0,73$. Mô hình

từ thư viện đại học Tokyo cho ra phương trình với 4 thông số mô tả, có các giá trị $R^2 = 0,79$; $RMSE = 0,13$; $Q^2_{LOO} = 0,76$. Cả 3 mô hình đều đạt các yêu cầu theo đánh giá Roy và hệ số tương quan phù hợp CCC. Từ kết quả đánh giá 3 mô hình 2D-QSAR cho thấy cả 3 mô hình đều đạt yêu cầu của một mô hình dự đoán tốt với sai số giữa giá trị thực nghiệm và giá trị dự đoán là chấp nhận được và có thể ứng dụng để dự đoán hoạt tính kháng hemozoin trên ký sinh trùng sốt rét.

Kết quả sàng lọc ảo trong tập TCM bằng 3 mô hình dự đoán 2D-QSAR cho thấy:

- Với mô hình 2D-QSAR các dẫn chất quinolin, sàng lọc được 228 chất có hoạt tính kháng hemozoin tốt, với $IC_{50} \leq 10 \mu M$.
- Với mô hình 2D-QSAR các dẫn chất benzimidazol, sàng lọc được 5252 chất có hoạt tính kháng hemozoin tốt, với $IC_{50} \leq 10 \mu M$.
- Với mô hình 2D-QSAR các chất từ thư viện đại học Tokyo, sàng lọc được 313 chất có hoạt tính kháng hemozoin tốt, với $IC_{50} \leq 15 \mu M$.

Trong đó, một số dẫn chất được xác định có trong các cây ở Việt Nam như cây lá đắng *Vernonia amygdalina* (Compositae), cây đương quy di thực *Angelica sinensis* (Apiaceae)...định hướng tiến hành thử nghiệm *in vitro* kháng sốt rét trên dịch chiết các cây này.

4.2 Kiến nghị

Tuy nhiên nghiên cứu không thành công trong việc xây dựng mô hình 2D-QSAR trên toàn bộ cơ sở dữ liệu thu thập được. Do tập dữ liệu bao gồm nhiều chất với cấu trúc đa dạng, thuật toán PLS còn hạn chế trong việc giải quyết vấn đề trên.

Nghiên cứu ứng dụng thêm các mô hình QSAR với những thuật toán cao cấp hơn PLS như máy vector hỗ trợ (SVM), mạng nơ ron nhân tạo (ANN) nhằm xây dựng được các mô hình có ứng dụng cao trong sàng lọc các chất có hoạt tính kháng hemozoin.

Nghiên cứu ứng dụng mô hình 3D-pharmacophore để có thể đưa ra thêm một phương pháp sàng lọc ảo nhằm giúp cho việc sàng lọc nhanh chóng và hiệu quả hơn trên một số lượng lớn hoạt chất.

Tiến hành thử hoạt tính *in vitro* dịch chiết những dược liệu có thành phần hóa học tương tự những chất có hoạt tính dự đoán kháng hemozoin tốt.

5. LỜI CẢM ƠN

Cảm ơn ban biên tập và phản biện Tạp chí Khoa học Lạc Hồng (Journal of science of Lac Hong University) đã góp ý cho bài báo này.

Cảm ơn PGS. TS. Nguyễn Thụy Việt Phương (Khoa Dược, đại học Y Dược TPHCM) đã định hướng cho nghiên cứu và PGS. TS. Nguyễn Tiến Huy (đại học Nagasaki, Nhật Bản) đã cung cấp dữ liệu 224 có hoạt tính kháng hemozoin chất sàng lọc bằng HTS từ thư viện đại học Tokyo.

6. TÀI LIỆU THAM KHẢO

- [1] Marsh, Kevin, et al. "Development of WHO recommendations for the final phase of elimination and prevention of re-establishment of malaria." *The American Journal of Tropical Medicine and Hygiene* 110.4 Suppl (2024): 3.
DOI: <https://doi.org/10.4269/ajtmh.22-0768>
- [2] Gerth-Guyette, Emily, et al. "Assessing the operational feasibility of integrating point-of-care G6PD testing into *Plasmodium vivax* malaria management in Vietnam." *Pathogens* 12.5 (2023): 689.
DOI: <https://doi.org/10.3390/pathogens12050689>
- [3] Quang, Huynh Hong, et al. "Multidrug-resistant *Plasmodium falciparum* parasites in the Central Highlands of Vietnam jeopardize malaria control and elimination strategies." *Antimicrobial Agents and Chemotherapy* 65.4 (2021): 10-1128.
DOI: <https://doi.org/10.1128/aac.01639-20>
- [4] Pandey, Amit V., et al. "Mechanism of malarial haem detoxification inhibition by chloroquine." *Biochemical Journal* 355.2 (2001): 333-338.
DOI: <https://doi.org/10.1042/bj3550333>
- [5] Mosaddeque, Farhana, et al. "Prediction model for antimalarial activities of hemozoin inhibitors by using physicochemical properties." *Antimicrobial Agents and Chemotherapy* 62.5 (2018): 10-1128.
DOI: <https://doi.org/10.1128/aac.02424-17>
- [6] Sharma, Bharvi, et al. "Quinoline-based heterocyclic hydrazones: Design, synthesis, anti-plasmodial assessment, and mechanistic insights." *Chemical Biology & Drug Design* 101.4 (2023): 829-836.
DOI: <https://doi.org/10.1111/cbdd.14185>
- [7] Chen, Calvin Yu-Chian. TCM Database@ Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PloS one*, 2011, 6.1: e15939.
DOI: <https://doi.org/10.1371/journal.pone.0015939>
- [8] Vijayaraghavan, Shilpa, and Supriya Mahajan. "Docking, synthesis and antimalarial activity of novel 4-anilinoquinoline derivatives." *Bioorganic & medicinal chemistry letters* 27.8 (2017): 1693-1697.
DOI: <https://doi.org/10.1016/j.bmcl.2017.03.005>
- [9] Andayi, Warren Andrew, Timothy J. Egan, and Kelly Chibale. "Kojic acid derived hydroxypyridinone-chloroquine hybrids: Synthesis, crystal structure, antiplasmodial activity and β -haematin inhibition." *Bioorganic & Medicinal Chemistry Letters* 24.15 (2014): 3263-3267.
DOI: <https://doi.org/10.1016/j.bmcl.2014.06.012>
- [10] Andayi, Warren A., et al. "Synthesis, antiplasmodial activity, and β -haematin inhibition of hydroxypyridone-chloroquine hybrids." *ACS medicinal chemistry letters* 4.7 (2013): 642-646.
DOI: <https://doi.org/10.1021/ml4001084>
- [11] Fong, Kim Y., Rebecca D. Sandlin, and David W. Wright. "Identification of β -haematin inhibitors in the MMV Malaria Box." *International Journal for Parasitology: Drugs and Drug Resistance* 5.3 (2015): 84-91.
DOI: <https://doi.org/10.1016/j.ijpddr.2015.05.003>
- [12] Saify, Zafar Saied, et al. "New benzimidazole derivatives as antiplasmodial agents and *plasmepsin* inhibitors: synthesis and

analysis of structure–activity relationships." *Bioorganic & medicinal chemistry letters* 22.2 (2012): 1282-1286.

DOI: <https://doi.org/10.1016/j.bmcl.2011.10.018>

[13] Sandlin, Rebecca D., et al. "Identification of β -hematin inhibitors in a high-throughput screening effort reveals scaffolds with *in vitro* antimalarial activity." *International Journal for Parasitology: Drugs and Drug Resistance* 4.3 (2014): 316-325.

DOI: <https://doi.org/10.1016/j.ijpddr.2014.08.002>

[14] Wicht, Kathryn J., et al. "Identification and SAR evaluation of hemozoin-inhibiting benzamides active against *Plasmodium falciparum*." *Journal of medicinal chemistry* 59.13

(2016): 6512-6530.

DOI: <https://doi.org/10.1021/acs.jmedchem.6b00719>

[15] <https://www.chemcomp.com/>

[16] Chao, Wen-Wan, and Bi-Fong Lin. "Bioactivities of major constituents isolated from *Angelica sinensis* (Danggui)." *Chinese medicine* 6.1 (2011): 1-7.

DOI: <https://doi.org/10.1186/1749-8546-6-29>

[17] Irungu, Beatrice N., et al. "Antiplasmodial and cytotoxic activities of the constituents of *Turraea robusta* and *Turraea nilotica*." *Journal of ethnopharmacology* 174 (2015): 419-425.

DOI: <https://doi.org/10.1016/j.jep.2015.08.039>