

# SỬ DỤNG PHƯƠNG PHÁP MÁY VECTOR HỖ TRỢ TRONG DỰ ĐOÁN HOẠT TÍNH KHÁNG SÓT RÉT MỘT SỐ DẪN CHẤT CHALCON

Thái Khắc Minh\*, Trần Thành Đạo\*, Đặng Trường Luân\*, Nguyễn Đắc Chí\*

## TÓM TẮT

**Mở đầu:** Công bố của tổ chức y tế thế giới WHO 2008 ước tính có khoảng 247 triệu ca sốt rét trong số 3,3 tỉ người nằm trong vùng nguy cơ và bệnh sốt rét gây ra gần 1 triệu ca tử vong, hầu hết là trẻ em dưới 5 tuổi. Nhu cầu cấp thiết tìm ra thuốc mới cho phòng chống sốt rét là một trong những chiến lược chống sốt rét toàn cầu. Ngày càng có nhiều nghiên cứu trong lĩnh vực hóa dược công bố các dẫn chất mới có khả năng tác dụng tốt trên chủng *P. falciparum* để kháng thuốc cũng như các mục tiêu phát triển thuốc mới.

**Mục tiêu:** Xây dựng mô hình dự đoán hoạt tính kháng sốt rét của một số dẫn chất chalcon nhằm thiết kế những công thức có hoạt tính tốt.

**Phương pháp:** Phương pháp máy vector hỗ trợ SVM hồi qui (support vector machine regression) được sử dụng để xây dựng mô hình trên 67 dẫn chất chalcon với hoạt tính kháng sốt rét.

**Kết quả:** Mô hình SVM xây dựng dựa trên tập hợp gồm 67 dẫn chất chalcon cho hệ số tương quan  $r^2$  giữa giá trị thực nghiệm và giá trị dự đoán nằm trong khoảng 0,61-0,70, SE = 0,13-0,14 và  $r^2$  của tập kiểm tra đánh giá chéo 5-lần-cắt-20% = 0,57. Mô hình SVM với  $r^2 = 0,68$  và SE = 0,13 được sử dụng để dự đoán hoạt tính kháng sốt rét trên chủng *P. falciparum* để kháng cloroquin của 18 dẫn chất 2'-hydroxychalcon mới được tổng hợp. Mối quan hệ giữa cấu trúc và tác dụng kháng sốt rét dự đoán của dẫn chất 2'-hydroxychalcon được phân tích trong đó khẳng định vai trò của nhóm methoxy ở vị trí 4 trên vòng B.

**Kết luận:** Mô hình SVM xây dựng được sử dụng để dự đoán hoạt tính kháng sốt rét của các dẫn chất chalcon mới và ứng dụng mô hình này có thể tiết kiệm chi phí thử nghiệm và hạn chế tổng hợp các dẫn chất có tác dụng kém. Kết quả này được sử dụng trong định hướng thiết kế và tổng hợp các phân tử chalcon có hoạt tính kháng sốt rét mạnh hơn.

**Từ khóa:** Thiết kế thuốc, máy vector hỗ trợ, SVM, sốt rét, hồi qui, chalcon, *Plasmodium falciparum*.

## ABSTRACT

### SUPPORT VECTOR MACHINE BASED PREDICTION MODEL FOR ANTIMALARIAL ACTIVITY OF CHALCONE DERIVATIVES

Khac Minh Thai, Thanh Dao Tran, Dang Truong Luan, Nguyen Dac Chi

\* Y Hoc TP. Ho Chi Minh \* Vol. 14 - Supplement of No 1 - 2010: 15 – 22

**Background:** Malaria has been one of the most important diseases of the developing world, killing about 1 million people and causing disease in 247 million people annually based on WHO World Malaria Report 2008. It affects many tropical and subtropical regions of the world. The increasing resistance of *Plasmodium* spp. to existing therapies has heightened alarms about malaria in the international health community. Nowadays, there is a pressing need for identifying and developing new drug-based antimalarial therapies.

**Objective:** The aim of this study is the development of a relevant computational model to predict antimalarial activity of chalcone derivatives. This model could be applied to screen and design new antimalarial drugs.

\* Bộ môn Hóa Dược – Khoa Dược - Đại học Y Dược Thành phố Hồ Chí Minh

Địa chỉ liên hệ: TS Thái Khắc Minh

ĐT: 0909 680 385

Email: thaikhacminh@gmail.com

**Method:** In this study, non-linear support vector machine (SVM) regression approach was applied on sixty seven chalcone analogues with in vitro antimalarial activities.

**Results:** The SVM model was obtained with regression coefficient  $r^2$  in range 0.61-0.70, SE = 0.13-0.14 for training set and  $r^2$  5-fold-leave-20%-out = 0.57 for test set. The best model with  $r^2 = 0.68$  and SE = 0.13 was used to predict the chloroquine-resistant *P. falciparum* inhibitory activity of 18 new synthesized 2'-hydroxychalcone derivatives. The relationship between chemical structure of 2'-hydroxychalcone series and antimalarial property was analysed and it is stress out the important role of 4-methoxy group at B-ring.

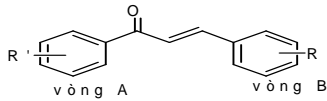
**Conclusion:** The SVM model could be applied to predict antimalarial activity of new chalcone compounds and it may be used as an in silico tool to design and develop the new potent antimalarial novels.

**Key words:** Drug design, Support vector machine, SVM, malaria, regression, chalcone, *Plasmodium falciparum*.

## ĐẶT VẤN ĐỀ

Công bố của tổ chức y tế thế giới WHO 2008 (1) ước tính có khoảng 247 triệu ca sốt rét trong số 3.3 tỉ người nằm trong vùng nguy cơ và bệnh sốt rét gây ra gần 1 triệu ca tử vong, hầu hết là trẻ em dưới 5 tuổi. Tổng cộng 109 quốc gia nằm trong vùng dịch tễ sốt rét tính tới năm 2008, trong đó có 45 quốc gia thuộc khu vực châu phi. Tình trạng đề kháng thuốc đang ngày một gia tăng trên diện rộng, khiến cho việc kiểm soát dịch sốt rét ngày càng trở nên khó khăn, đặc biệt tại các các quốc gia nghèo đói thuộc khu vực châu phi (1). “Nguyên tắc vàng” trong điều trị sốt rét trước đây là chloroquin, cùng với các thuốc kháng folat ra đời sau đó, hiện nay chỉ còn nhạy cảm cho một vài khu vực (4). Đa số các vùng còn lại phải dùng chiến lược điều trị kết hợp thuốc. Sự kết hợp các công cụ và phương pháp trong cuộc chiến chống lại bệnh sốt rét hiện tại bao gồm phát triển ý thức cộng đồng, phun thuốc diệt muỗi có tác dụng dài và liệu pháp kết hợp thuốc dựa trên nền tảng là artemisinin. Cho tới nay có rất ít trường hợp báo cáo về tình trạng đề kháng của ký sinh trùng sốt rét với artemisinin, nhưng điểm hạn chế rất lớn của giải pháp này là thời gian bán thải của artemisinin và các dẫn chất rất ngắn. Hiện nay, artemisinin chỉ dùng để cắt cơn trong chiến lược điều trị sốt rét và sự phối hợp artemisinin với các thuốc khác có tác dụng dài hiện đang là chìa khóa trong điều trị sốt rét (12, 13). Tuy nhiên, chỉ khoảng 25% số người mắc bệnh sốt rét được

chăm sóc thuốc men đầy đủ (1). Nguyên nhân có thể do (i) mạng lưới chăm sóc y tế chưa được rộng rãi tới các vùng hẻo lánh, (ii) tình trạng đề kháng thuốc ngày càng lan rộng, và (iii) thu nhập quá thấp ở các quốc gia nghèo dẫn đến việc chăm sóc thuốc men trở nên quá khả năng... Những tổn thất do kí sinh trùng sốt rét gây ra vẫn đang tăng theo từng năm, đặc biệt ở các quốc gia nghèo đói. Tình trạng đề kháng thuốc của ký sinh trùng sốt rét *P. falciparum* ngày càng gia tăng trên diện rộng (1). Thực tiễn đòi hỏi thế giới phải có một chiến lược hiệu quả và dài hạn trong cuộc chiến chống lại bệnh sốt rét trên toàn cầu (13), trong đó việc nghiên cứu tìm ra những thuốc mới có hiệu quả trên các dòng *Plasmodium* đề kháng là nhu cầu cấp bách nhằm tìm ra giải pháp điều trị với các tiêu chí: hiệu quả, an toàn, độc tính thấp, và giá rẻ để mọi cá nhân nằm trong vùng nguy cơ sốt rét đều được chăm sóc (4, 13). Trong nghiên cứu này, thuật toán máy vector hỗ trợ SVM hồi qui (support vector machine) được sử dụng để xây dựng mô hình dự đoán hoạt tính kháng sốt rét trên 67 dẫn chất chalcon (Hình 1). Thông tin về mối liên hệ giữa cấu trúc – tác dụng kháng sốt rét của các dẫn chất có được từ mô hình SVM hồi qui cũng như khả năng dự đoán của mô hình có thể ứng dụng trong định hướng nghiên cứu tổng hợp các hoạt chất có thể phát triển thành thuốc kháng sốt rét mới.

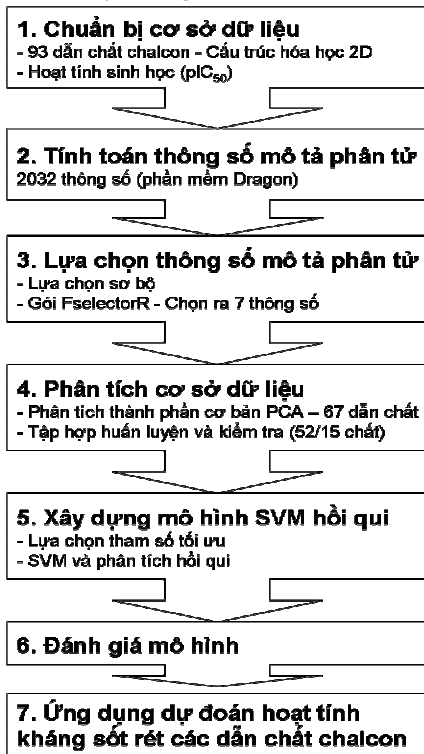


Hình 1. Cấu trúc hóa học của các dẫn chất chalcon

## ĐỐI TƯỢNG - PHƯƠNG PHÁP NGHIÊN CỨU

### Quy trình xây dựng mô hình máy vector hỗ trợ SVM hồi qui

Mô hình SVM hồi qui dự đoán hoạt tính kháng sốt rét được tiến hành theo các bước được trình bày trong hình 2.



Hình 2. Quy trình xây dựng mô hình máy vector hỗ trợ SVM hồi qui.

### Cơ sở dữ liệu

Cấu trúc của 93 dẫn chất chalcon (8) được thu thập dựa trên một vài tiêu chí chung như: (i) cấu trúc có triển vọng, (ii) giá trị  $IC_{50}$  in vitro có được từ phương pháp đo độ hấp thụ  $^3H$ -hypoxanthin (3, 8), (iii) giá trị hoạt tính sinh học xác định trên chủng *P. falciparum* K1 đề kháng chloroquin (CQ), và (iv) các cấu trúc có cùng xương sống trên cùng 1 mô hình tiến hành. Nghiên cứu được tiến hành trên 93 dẫn chất chalcon với khung cơ bản của các cấu trúc nghiên cứu được trình bày ở hình 1 (8). Giá trị hoạt tính

sinh học  $IC_{50}$  được quy đổi thành giá trị  $pIC_{50} = -\log(IC_{50})$  được sử dụng trong nghiên cứu.

### Thông số mô tả phân tử

Thông số mô tả phân tử là những thông số mô tả tính chất của các chất có ảnh hưởng tới tác dụng sinh học, bao gồm những tính chất về lý, hóa. Cấu trúc 2D của các dẫn chất được xây dựng và tính toán thông số mô tả phân tử bằng phần mềm Dragon® (2) phiên bản Evaluation 5.5. Tổng số 2032 thông số mô tả phân tử 2D thuộc 11 nhóm chính được tính toán.

### Lựa chọn thông số mô tả phân tử

Để loại bỏ đi các thông số không quan trọng và lựa chọn thông số tốt nhất cho xây dựng mô hình, một số phương pháp và thuật toán được áp dụng, bao gồm quá trình loại thông số mô tả thô và quá trình lựa chọn thông số thích hợp sau cùng cho mô hình (5, 14). Một vài qui tắc cho việc loại thô ban đầu bao gồm: (i) các thông số có  $\geq 80\%$  giá trị = 0; (ii) các thông số có độ lệch chuẩn  $\leq 0,5$ ; và (iii) các thông số có tương quan với giá trị  $pIC_{50} \leq 0,07$ . Các thông số còn lại được phân chia tỷ lệ giá trị trong khoảng (0-1) bằng phương pháp chuẩn hóa cực tiểu – cực đại.<sup>11</sup> Công thức cụ thể như sau theo công thức:

$$V_n = \left[ \frac{V_0 - MIN_0}{MAX_0 - MIN_0} \right] \times (MAX_n - MIN_n) + MIN_n$$

Với:  $V_n$ : giá trị mới

$V_0$ : giá trị hiện tại

$MAX_0, MIN_0$ : giá trị lớn nhất và nhỏ nhất của dãy giá trị hiện tại

$MAX_n, MIN_n$ : giá trị lớn nhất và nhỏ nhất của khoảng giá trị mới cần qui đổi

Các thông số sau khi được phân chia tỷ lệ sẽ được áp dụng vào thuật toán “rừng ngẫu nhiên RF” trong gói Fselector trong R với hàm “tầm quan trọng rừng ngẫu nhiên” (“random.forest.importance”) để chọn lọc những thông số mô tả đặc trưng nhất cho hoạt tính sinh học (10, 11). Hàm “cắt xén . k” (“cutoff.k”) cũng trong gói Fselector hỗ trợ để cắt ra “k” thông số được đánh giá có liên quan nhất với giá trị tham

chiều là giá trị  $pIC_{50}$ . Giá trị “k” được xác định phù hợp (11).

### Phân tích thành phần cơ bản

Các thông số được sử dụng để xây dựng mô hình SVM hồi quy phi tuyến tính trên toàn bộ cơ sở dữ liệu cho kết quả  $r^2$  khá thấp. Để xây dựng mô hình có khả năng mô tả được hoạt tính sinh học bằng các thông số mô tả phân tử và loại bỏ các chất gây nhiễu ảnh hưởng đến kết quả của mô hình, thuật toán phân tích thành phần cơ bản (principal component analysis - PCA) được sử dụng. Trong nghiên cứu này, thuật toán phân tích thành phần cơ bản không tuyến tính (nonlinear principal component analysis - NL-PCA) dựa trên lý thuyết mạng thần kinh trong gói pcaMethod của R được sử dụng (7).

### Máy vector hỗ trợ hồi qui phi tuyến tính

Tính toán máy vector hỗ trợ SVM là một nhóm các phương pháp học có sự giám sát dùng trong phân loại hay phương trình hồi quy (6, 7, 9, 11). Trong nghiên cứu này, gói e1071 trong môi trường R được sử dụng (7, 11). Máy vector hỗ trợ SVM trong gói e1071 bao gồm cả hai chức năng phân loại cũng như hồi qui (7). SVM hồi qui trong gói e1071 bao gồm 2 thuật toán cho xây dựng mô hình hồi qui là hồi qui epsilon ( $\epsilon$ -regression) và hồi qui nu (nu-regression). Trong đó hồi qui epsilon dùng để xây dựng mô hình hồi qui cho dữ liệu phi tuyến tính và thuật toán này được sử dụng trong nghiên cứu. Khi tiến hành xây dựng mô hình máy vector hỗ trợ hồi qui phi tuyến tính (7, 9) thì 3 tham số quan trọng cần xác định là tham số  $\epsilon$  (tham số của hàm tổn thất  $\epsilon$ -insensitive) và tham số của hàm nhân Kernel RBF (C,  $\gamma$ ). Tham số  $\epsilon$  qui định biên

độ của mặt phẳng hồi quy và 2 tham số C và  $\gamma$  qui định bề mặt phẳng hồi quy. Sự thay đổi giá trị của 3 tham số này ảnh hưởng lớn tới kết quả dự đoán. Hàm “tune.svm” (gói e1071) được sử dụng để dò tìm giá trị tham số tối ưu cho mô hình bằng phương pháp đánh giá chéo k nhóm (k-folds cross validation)(11). Trong đánh giá chéo, dữ liệu được chia làm k nhóm, 1 nhóm dùng để thử, và (k-1) nhóm còn lại dùng để huấn luyện. Hàm “svm” và “predict” (gói e1071) lần lượt dùng để huấn luyện và dự đoán cho mô hình SVM hồi qui phi tuyến tính.

### Đánh giá mô hình

Sau khi huấn luyện và dự đoán hoạt tính sinh học bằng mô hình SVM hồi qui phi tuyến tính, giá trị  $pIC_{50}$  dự đoán của các chất nghiên cứu được xử lý bằng hàm “lm” trong R (hàm hồi qui tuyến tính đơn giản) để đánh giá mức độ tương quan giữa giá trị dự đoán và giá trị thực nghiệm (10, 11). Các tiêu chí đánh giá bao gồm: (i) hệ số tương quan pearson r, (ii) sai số chuẩn (SE), (iii) trị số P, (iv) và hệ số xác định bội  $r^2$ .

## KẾT QUẢ VÀ BÀN LUẬN

### Lựa chọn thông số mô tả phân tử

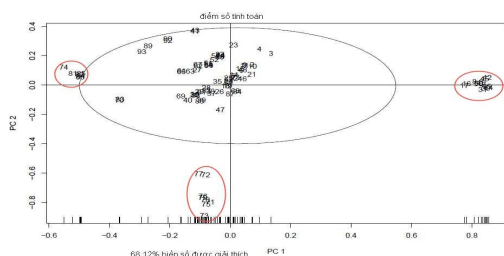
Sau khi tiến hành loại thô ban đầu, tổng cộng 65 thông số đáp ứng điều kiện và được áp dụng vào thuật toán “rừng ngẫu nhiên RF” bằng gói FSelector với hàm “tầm quan trọng rừng ngẫu nhiên” (“random.forest.importance”). Từ kết quả FSelector, nhóm 7 thông số được lựa chọn để xây dựng mô hình vì có giá trị đóng góp vào giá trị hoạt tính sinh học  $pIC_{50}$  cao và nhóm thông số này được trình bày ở bảng 2.

**Bảng 2.** Nhóm 7 thông số mô tả được lựa chọn từ FSelector dùng để xây dựng mô hình

Tên	Định nghĩa	Lớp mô tả
ESpm15u	Moment phổ 15 từ mạng các đỉnh gần kề	Chỉ số các đỉnh liền kề
ESpm14r	Moment phổ 14 từ mạng các đỉnh gần kề điều chỉnh bởi tích phân cộng hưởng	Chỉ số các đỉnh liền kề
ESpm15r	Moment phổ 15 từ mạng các đỉnh gần kề điều chỉnh bởi tích phân cộng hưởng	Chỉ số các đỉnh liền kề
EEig14x	Giá trị riêng 14 từ mạng các đỉnh gần kề hiệu chỉnh bởi các mức độ đỉnh	Chỉ số các đỉnh liền kề
VRD1	Chỉ số dựa vào vector riêng loại ngẫu nhiên mạng khoảng cách	Chỉ số dựa vào giá trị riêng
EEig09x	Giá trị riêng 09 từ mạng các đỉnh gần kề hiệu chỉnh bởi các mức độ đỉnh	Chỉ số các đỉnh liền kề
BAC	Chỉ số balaban trung tâm	Thông số hình học topo

### Phân tích thành phần cơ bản

NL-PCA được thực hiện trên cơ sở dữ liệu 93 chất với số thành phần cơ bản là 2 (number of components = 2). Trong phân tích PCA này, 68% giá trị hoạt tính sinh học có thể giải thích dựa vào 2 thành phần cơ bản xây dựng từ 7 thông số mô tả. Hình 3 biểu diễn sự phân bố của các dẫn chất dựa trên mặt phẳng tạo bởi 2 PC trong phân tích PCA. Kết quả phân tích cho thấy có 26 chất nằm ngoài và có thể là yếu tố gây nhiễu cho cơ sở dữ liệu. Do đó, 26 chất này được loại bỏ khỏi cơ sở dữ liệu và 67 chất còn lại được sử dụng trong xây dựng mô hình máy vector hỗ trợ SVM hồi quy.



**Hình 3.** Sự phân bố của 93 dẫn chất chalcon dựa trên mặt phẳng tạo bởi 2 PC trong phân tích NL-PCA. Các chất được đánh dấu bằng hình elip nhỏ: là các dữ liệu gây nhiễu và được loại bỏ

### Phân chia dữ liệu

Tập dữ liệu gồm 67 dẫn chất chalcon có được từ phân tích NL-PCA được phân chia ngẫu nhiên 5 lần theo tỷ lệ 4:1 (tập huấn luyện : tập kiểm tra) bằng hàm sample trong R (11). Kết quả tạo ra 5 tập hợp huấn luyện có 52 chất và 5 tập hợp kiểm tra gồm 15 chất. Phương pháp máy vector hỗ trợ hồi quy phi tuyến tính được tiến hành trên 5 tập huấn luyện có được từ phân chia ngẫu nhiên. Kết quả trung bình từ 5 mô hình này tương ứng với quá trình đánh giá chéo 5-lần-cắt-20% (5-time Leave-20%-out). Đồng thời

**Bảng 3.** Kết quả SVM hồi quy phi tuyến tính giữa mô hình thứ cấp A-E và mô hình F với toàn bộ cơ sở dữ liệu

Tiêu chí đánh giá	Mô hình A-E		Mô hình F với toàn bộ 67 dẫn chất chalcon
	Tập huấn luyện	Tập kiểm tra	
$r^2$	0,66 <sup>a</sup>	0,57	0,68
Sai số chuẩn SE	0,13 <sup>a</sup>	0,15	0,13
Trị số P	$1.57 \times 10^{-12}$	0,001	$2.2 \times 10^{-16}$

<sup>a</sup> đánh giá chéo 5-lần-cắt-20%

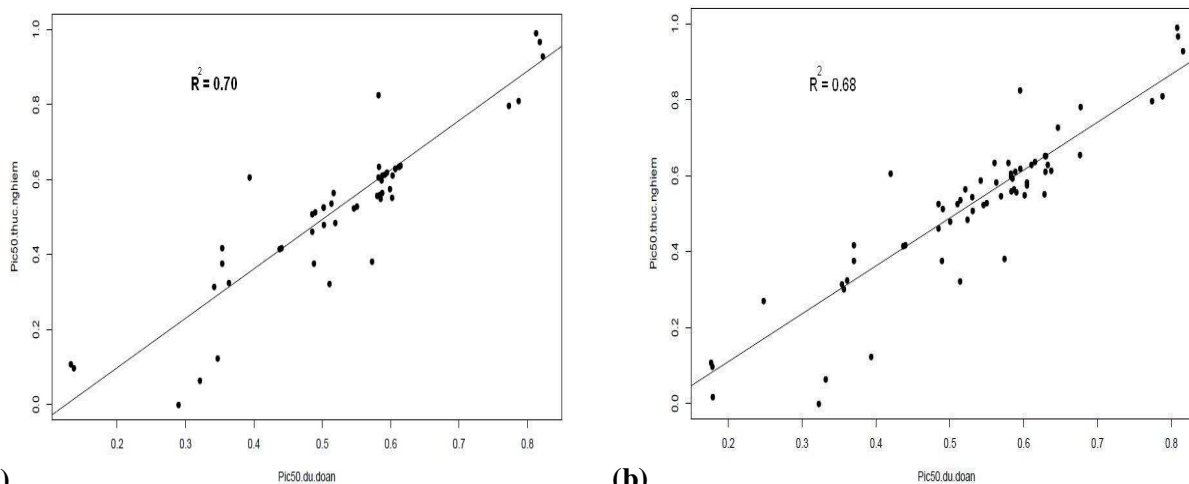
SVM hồi quy được tiến hành trên tập dữ liệu toàn bộ 67 chất.

### Chọn lựa tham số tối ưu

Hàm “tune.svm” (gói e1071) được sử dụng để lựa chọn tham số tối ưu cho máy vector hỗ trợ hồi quy phi tuyến tính trên toàn bộ cơ sở dữ liệu (6, 7, 11). Kết giống nhau cho các lần tiến hành lặp lại khi sử dụng tham số tối ưu thu được là  $C = 1$  (trong khoảng dò  $10^0 : 3$ ),  $\gamma = 1$  (trong khoảng dò  $10^{(-6 : 0)}$ ), và  $\epsilon = 0,1$  (cố định). Sử dụng thông số tối ưu này cho kết quả sai số trung bình của mô hình = 0,034 với đánh giá chéo trên k=10 nhóm phân chia.

### Mô hình máy vector hỗ trợ hồi quy phi tuyến tính trên dẫn chất chalcon

SVM hồi quy phi tuyến tính với các tham số tối ưu được phân tích trên 5 tập hợp huấn luyện (gồm 52 dẫn chất chalcon) tương ứng với 5 mô hình thứ cấp A-E và kết quả trung bình sẽ ứng với đánh giá chéo 5-lần-cắt-20%. Như trình bày ở Bảng 3, mô hình cho kết quả  $r^2$  đánh giá chéo 5-lần-cắt-20% là 0,66 và sai số chuẩn là 0,13. Đối với tập hợp kiểm tra tương ứng, mô hình A-E cho giá trị  $r^2 = 0,57$  và SE=0,15. Đồng thời, mô hình F cũng được xây dựng bằng phương pháp SVM hồi quy phi tuyến tính trên toàn bộ cơ sở dữ liệu gồm 67 dẫn chất chalcon và kết quả trình bày ở Bảng 3. Mô hình F cho kết quả  $r^2 = 0,68$ , SE = 0,13, trị số P =  $2.2 \times 10^{-16}$  và kết quả này giống nhau ở các lần huấn luyện lặp lại. Giá trị  $r^2 = 0,68$  của mô hình F tương đương với giá trị của mô hình đánh giá chéo 5-lần-cắt-20%. Đường thẳng tuyến tính giữa giá trị thực nghiệm và giá trị dự đoán của mô hình thứ cấp E và mô hình F trên toàn bộ với dữ liệu toàn bộ được trình bày ở Hình 4.



**Hình 4.** Đường hồi quy tuyến tính giữa giá trị thực nghiệm và giá trị dự đoán của mô hình thứ cấp E (a) và mô hình F xây dựng từ toàn bộ 67 dẫn chất chalcon (b)

**Bàn luận về mô hình trên dẫn chất chalcon**

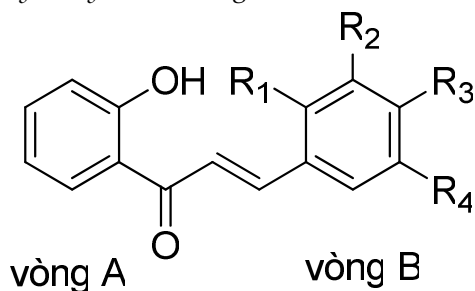
Kết quả mối tương quan giữa giá trị pIC<sub>50</sub> dự đoán và thực nghiệm của các dẫn chất chalcon được tóm tắt ở Bảng 3 cho thấy giá trị của trị số P có được ở tất cả các lần phân tích đều << 0,05, nhưng không bằng 0, điều đó cho thấy sự khác nhau không có ý nghĩa thống kê giữa giá trị dự đoán và giá trị thực nghiệm.<sup>11</sup> Giá trị hệ số xác định bội r<sup>2</sup> nằm trong khoảng 0,61-0,70 và r<sup>2</sup> = 0,68 cho mô hình F với dữ liệu toàn bộ là mức độ chấp nhận được cho tập cơ sở dữ liệu 67 chất. Kết quả r<sup>2</sup> của tập thử nghiệm đánh giá chéo 5-lần-cắt-20% = 0,57. Mô hình F với mức độ đa dạng về cấu trúc hóa học các dẫn chất chalcon được ứng dụng để dự đoán các chất chalcon khác.

**Ứng dụng mô hình trong dự đoán hoạt tính kháng sốt rét các dẫn chất chalcon**

Mô hình F được sử dụng để dự đoán 18 dẫn

chất 2'-hydroxychalcon mới được tổng hợp bởi nhóm nghiên cứu TS. Trần Thành Đạo, Bộ môn Hóa Dược, Đại học Y Dược Thành phố Hồ Chí Minh (14). Nhóm 18 dẫn chất chalcon này có cấu trúc hoá học khác với nhóm dẫn chất chalcon dùng để xây dựng mô hình. Cấu trúc 18 dẫn chất của 2'-hydroxychalcon với nhóm hydroxy tại vị trí 2' trên vòng A và các nhóm thế khác nhau trên vòng B, được trình bày trong Bảng 4. 18 dẫn chất 2'-hydroxychalcon này đã được nghiên cứu về hoạt tính kháng viêm theo cơ chế ức chế enzym cyclooxygenase 2 (COX-2) (14). Một vài chất trong 18 dẫn chất này đã được báo cáo cáo hiệu quả kháng viêm giảm đau tốt. Giá trị của 7 thông số mô tả tính toán của 18 chất này được lấy tỷ lệ chung với 67 chất mô hình F trong khoảng (0-1). Kết quả dự đoán hoạt tính kháng sốt rét trên *P. falciparum* đề kháng CQ-chủng K1 được trình bày trong Bảng 4.

**Bảng 4.** Cấu trúc các dẫn chất 2'-hydroxychalcon và giá trị sinh học dự đoán từ mô hình F



Tên chất	R1	R2	R3	R4	Giá trị pIC <sub>50</sub> dự đoán	Giá trị IC <sub>50</sub> qui đổi (μM) <sub>a</sub>
BMCL2009_19_1650_01	H	H	H	H	-1.35	22.46
BMCL2009_19_1650_02	H	H	Cl	H	-1.34	22.07
BMCL2009_19_1650_03	H	H	Br	H	-1.34	22.07
BMCL2009_19_1650_04	H	H	CH <sub>3</sub>	H	-1.36	22.97
BMCL2009_19_1650_05	H	H	OCH <sub>3</sub>	H	-1.02	10,40
BMCL2009_19_1650_06	H	H	SCH <sub>3</sub>	H	-2.21	160,67
BMCL2009_19_1650_07	H	H	OCF <sub>3</sub>	H	-1.45	28.21
BMCL2009_19_1650_08	H	H	Ph	H	-1.35	22.20
BMCL2009_19_1650_09	H	H	OBn	H	-1.34	21.95
BMCL2009_19_1650_10	H	Br	H	H	-1.47	29.45
BMCL2009_19_1650_11	H	Cl	Cl	H	-1.37	23.36
BMCL2009_19_1650_12	H	Br	OCH <sub>3</sub>	H	-1.36	22.71
BMCL2009_19_1650_13	H	OBn	OBn	H	-1.35	22.45
BMCL2009_19_1650_14	H	OBn	OCH <sub>3</sub>	H	-1.36	22.71
BMCL2009_19_1650_15	OCH <sub>3</sub>	OCH <sub>3</sub>	H	H	-1.41	25.43
BMCL2009_19_1650_16	OCH <sub>3</sub>	H	OCH <sub>3</sub>	H	-1.00	9.89
BMCL2009_19_1650_17	H	OCH <sub>3</sub>	OCH <sub>3</sub>	OCH <sub>3</sub>	-1.00	10,00
BMCL2009_19_1650_18	H	-O-CH <sub>2</sub> -O-		H	-1.36	22.71

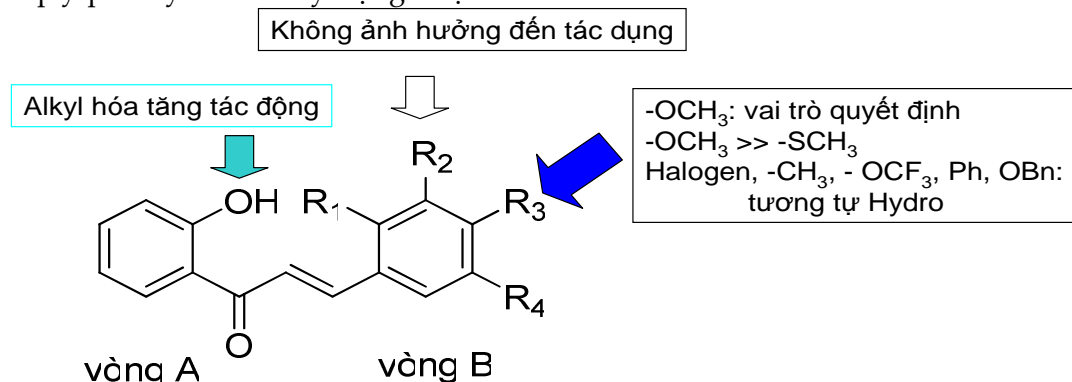
<sup>a</sup>Giá trị dự đoán từ mô hình F trên hoạt tính kháng *P. falciparum* để kháng CQ-chủng K1

Nghiên cứu liên quan cấu trúc tác dụng (8) trên cấu trúc của các 2',4'-hydroxy chalcon, 4'-hydroxychalcon cho thấy mối liên hệ giữa moment lưỡng cực, sự phân nhánh, độ khúc xạ phân tử và logkw (PLS,  $r^2=0,818$ ,  $q^2=0,602$ ,  $n=17$ ; tập kiểm tra  $n=13$ , RMSEP=0,425). Liu et al. (8) cũng chỉ ra rằng nhóm dẫn chất 2',4'-hydroxy chalcon, 4'-hydroxychalcon có hoạt tính tốt là những chất phân cực và chứa các nhóm chức có kích thước nhỏ. Điều này thì ngược lại với dẫn chất 4'-methoxy-, và 2',4'-dimethoxy chalcon.<sup>3</sup> Trong 97 chất được nghiên cứu, dẫn chất 2',3',4'-trimethoxy-3-quinolinylochalcon cho hoạt tính kháng sốt rét mạnh nhất với IC<sub>50</sub> = 2 μM (8). Kết quả dự đoán cho nhóm 18 dẫn chất 2'-hydroxychalcon được tổng hợp bởi nhóm nghiên cứu TS. Trần Thành Đạo cho thấy các dẫn chất đều cho tác động tốt trên ký sinh trùng sốt rét với giá trị pIC<sub>50</sub> = -1.00: -1.47, tương ứng với giá trị IC<sub>50</sub> trong khoảng 9 – 29 μM trên dòng K<sub>1</sub> để kháng CQ. Trên nhóm dẫn chất 2'-hydroxychalcon này cho thấy vai trò quan trọng của nhóm methoxy tại vị trí 4 trên vòng B so với các nhóm thế còn lại. Kết quả cho thấy sự giảm tác dụng khi thay thế ở vị trí 4' bởi các nhóm thế

hút điện tử mạnh (cloro, bromo, trifluoromethoxy), các nhóm đẩy điện tử yếu (-CH<sub>3</sub>, -SCH<sub>3</sub>) hoặc nhóm đẩy điện tử mạnh (benzyloxy). Sự giảm hoạt tính mạnh nhất khi thay thế nhóm 4-methoxy (pIC<sub>50</sub> dự đoán = -1.02) bằng nhóm thế 4-methiol (pIC<sub>50</sub> dự đoán = -1.21). Khi đưa thêm các nhóm thế không phải là methoxy vào các vị trí 2, 3 vào khung 2'-hydroxy-4-methoxy-chalcon cũng làm giảm tác dụng kháng sốt rét của nhóm dẫn chất. Khi đưa thêm 2 nhóm thế methoxy vào vị trí 2,3 của 2'-hydroxychalcon không làm tăng tác dụng kháng sốt rét của khung cơ bản. Điều này khẳng định vai trò quan trọng của nhóm 4-methoxy trong cấu trúc 2'-hydroxychalcon trong hoạt tính sinh học này của nhóm dẫn chất. Sự hiện diện càng nhiều nhóm thế methoxy ở các vị trí 2, 3, 5 trên cấu trúc 2'-hydroxy-4-methoxychalcon làm tăng hoạt tính của nhóm dẫn chất này. Một số nghiên cứu cho thấy alkyl hóa vị trí hydroxy của vòng A làm tăng tác dụng kháng sốt rét của nhóm dẫn chất. Sự hiện diện của một số dị vòng tại vị trí 3 trên vòng B (đặc biệt là quinoliny) cũng làm tăng hoạt tính kháng sốt rét. Tóm tắt liên quan cấu trúc – tác dụng kháng ký sinh trùng sốt rét của

dẫn chất 2'-hydrochalcon được trình bày ở hình 5. Những thông tin này kết hợp với mô hình SVM hồi quy phi tuyến tính xây dựng được có

thể ứng dụng để thiết kế ra các dẫn chất chalcon có hoạt tính sinh học cao hơn.



Hình 5. Mối quan hệ giữa cấu trúc và tác dụng kháng sốt rét dự đoán của dẫn chất 2'-hydroxychalcon

### KẾT LUẬN

Trong nghiên cứu này, phương pháp máy vector hỗ trợ SVM được áp dụng để xây dựng phương trình hồi qui trên tập hợp dữ liệu tuyến tính. Kết quả xây dựng được mô hình SVM hồi qui phi tuyến tính trên tập hợp gồm 67 dẫn chất chalcon. Mô hình thu được có hệ số tương quan  $r^2$  giữa giá trị thực nghiệm và giá trị dự đoán nằm trong khoảng 0,61-0,70, SE = 0,13-0,14 và  $r^2$  của tập thử nghiệm đánh giá chéo 5-lần-cắt-20% = 0,57. Mô hình trên dẫn chất chalcon với  $r^2$  = 0,68 và SE = 0,13 được sử dụng để dự đoán hoạt tính kháng sốt rét trên chủng *P. falciparum* để kháng CQ của 18 dẫn chất 2'-hydroxychalcon mới được tổng hợp. Mối quan hệ giữa cấu trúc và tác dụng kháng sốt rét dự đoán của dẫn chất 2'-hydroxychalcon được phân tích trong đó khẳng định vai trò của nhóm methoxy ở vị trí 4 trên vòng B. Kết quả này được sử dụng trong định hướng thiết kế và tổng hợp các phân tử chalcon có hoạt tính kháng sốt rét mạnh.

### TÀI LIỆU THAM KHẢO

1. Aregawi, M.; Cibulskis, R.; Otten, M.; William, R.; Dye, C. (2008). World Malaria Report 2008, World Health Organization, pp 1-215. WHO Press, Geneva.
2. Dragon® (2009): The software for calculating molecular descriptors, [www.taletale.mi.it/](http://www.taletale.mi.it/)
3. Fattorusso, C.; Campiani, G.; Catalanoti, B.; Prisco, M.; Basillico, N.; Parapini, S.; Tarameli, D.; Campadnuolo, C.; Fattorusso, E.; Romano, A.; Scafatti, O.T. (2006). Endoperoxide derivatives from manne organism : 1,2-dioxans of the plakertin family as novel antimalarial agents. J. Med. Chem., 49:7088-7094.

4. Fidock, D.A.; Rosenthal, P.J.; Croff, S.L.; Brun, R.; Nwaka, S. (2004). Antimalaria drug discovery: efficacy models for compound screening. Nat. Rev. Drug. Dis., 3(6):509-520,
5. Huang, J.; Ma, G.; Muhammad, I.; Cheng, Y. (2007). Identifying P-glycoprotein substrates using support vector machine optimized by a partical swarm. J. Chem. Inf. Model., 47(4):1638-1647.
6. Ivaciuc, O. (2007). Applications of support vector machine in chemistry, Volume 23. In: Lipkowitz, K.B.; Cundari, T.R. Reviews in computatinal chemistry, pp 291-400, Wiley-VCH, Weinheim.
7. Karatzoglou, A.; Meyer, D.; Hornik, K. (2006). Support Vector Machines in R. Journal of Statistical Software, 15(9):1-26.
8. Liu, M.; Wilairat, B.; Go, M.L. (2001). Antimalarial alkoxyated chalcones : structure – activity relationship analysis. J. Med. Chem., 44: 4443-4452.
9. Mayer, D. (2009). LIBSVM : support vector machine. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
10. Nguyễn Văn Tuấn (2007). Phân tích số liệu và tạo biểu đồ bằng R- hướng dẫn thực hành, in lần thứ nhất, 1-340, NXB Khoa Học và Kỹ Thuật, Hà Nội.
11. R (2009): The R project for statistical computing. <http://r-project.org/>.
12. Schlitzer, M. (2007) Malaria chemotherapeutics part I: History of antimalarial drug development, currently used therapeutics, and drugs in clinical development. ChemMedChem, 2(7): 944-986
13. Schlitzer, M. (2008). Antimalarial Drugs – What is in use and what is in the pipeline. Arch. Pharm. Chem. Lifesci., 341:149-163.
14. Thai, K.-M.; Ecker, G. F. (2008). A Binary QSAR Model for Classification of hERG Potassium Channel Blockers. Bioorg. Med. Chem, 16:4107-4119.
15. Tran, T.-D.; Park, H.; Kim, H.P.; Ecker, G.F.; Thai, K.-M. (2009). Inhibitory activity of prostaglandin E2 production by the synthetic 2'-hydroxychalcone analogues: synthesis and SAR study. Bioorg. Med. Chem. Lett., 19: 1650-1653.



