

SÀNG LỌC ẢO CÁC HỢP CHẤT ỨNG CHẾ ENZYME KINESIN EG5 Ở NGƯỜI DỰA TRÊN MÔ HÌNH 2D-QSAR

Đến tòa soạn 09-02-2023

Nguyễn Ngọc Lê¹, Hứa Hữu Bằng², Lê Đăng Quang³, Nguyễn Cường Quốc⁴, Nguyễn Trọng Tuấn⁴,
Bùi Thị Bửu Huệ⁴, Trần Thanh Mên⁴, Trần Quang Đệ^{4*}

1. Khoa Dược, Trường Đại học Nam Cần Thơ

2. Khoa Dược, Trường Đại học Võ Trường Toản

3. Viện Kỹ thuật nhiệt đới, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

4. Khoa Khoa học Tự Nhiên, Trường Đại học Cần Thơ

*Email: tqde@ctu.edu.vn

SUMMARY

2D-QSAR MODEL BASED VIRTUAL SCREENING FOR KINESIN EG5 INHIBITORS IN HUMANS

Human mitotic kinesin Eg5 which function is helping the formation of bipolar mitotic spindle, and has been identified as a potential target for new drug development in cancer chemotherapy. In this study, the 2D-QSAR model was built to find the correlation between Eg5 and its inhibitor, which was performed using previously reported Eg5 inhibitors such as dihydropyrazol, dihydropyrrol, 4-phenyltetrahydroisoquinolin, and thiophen derivatives. 2D-QSAR model based on 5 descriptors of 92 compounds was built with $R^2 = 0,86$, $RMSE = 0,47$; $\overline{r}_m^2 = 0.81$; $\Delta r_m^2 = 0.09$. Applying the model to the screening process, 296 drugs were obtained from the DrugBank library, and 2464 substances from the ChemDiv library. These compounds are potential candidates for potent inhibition of the Eg5 enzyme. The analyses may be used to design more potent Eg5 inhibitors and predict their activities prior to synthesis.

Keywords: 2D-QSAR, cancer, Eg5 enzyme, virtual screening.

1. GIỚI THIỆU

Kinesin là một protein thuộc loại protein vận động, được tìm thấy trong nhân tế bào. Kinesin di chuyển dọc theo sợi trục vi ống, và được thúc đẩy bởi quá trình thủy phân của adenosin triphosphat (ATP). Enzyme kinesin Eg5 cũng giống như các protein kinesin vận động khác, khi thủy phân ATP thành ADP và phosphat vô cơ các protein kinesin Eg5 sẽ sử dụng một phân tử nước và chuyển năng lượng hóa học tạo thành động lực để di chuyển dọc theo vi ống [1]. Enzym Eg5 có vai trò quan trọng trong quá trình nguyên phân [2], khi các enzyme Eg5 bị ức chế, sẽ kéo theo làm cho các trục vi ống không phát

triển và không thể gắn vào kinetochore, từ đó làm cho quá trình nguyên phân không xảy ra được. Kết quả cho thấy tế bào bị ngừng phát triển và tế bào sẽ chết theo chương trình apoptosis [3]. Vì vậy, phát hiện ra các chất hóa học có khả năng ức chế enzyme Eg5 là mục tiêu nghiên cứu triển vọng trong công cuộc phát triển các liệu pháp điều trị ung thư hiệu quả hiện nay. Chất ức chế enzyme Eg5 đầu tiên được nghiên cứu là Monastrol [4], sau đó các dẫn xuất quinazolinone (ispinesib), thiazolidines (filanesib, litronesib, K85830) và S-trityl-L-cysteine cũng cho thấy khả năng ức chế mạnh mẽ Eg5 ở nồng độ thấp (dưới 1 μ M).

Các hợp chất này hoạt động theo cơ chế ức chế hoạt động ATP kinase của Eg5 dẫn đến làm chậm quá trình giải phóng ADP và gây chết tế bào ung thư theo con đường apoptotic, từ đó đã cho thấy hiệu quả đáng khích lệ trong việc điều trị các bệnh ung thư [5, 6]. Chính vì vậy, trong nghiên cứu này, một mô hình 2D-QSAR về mối liên hệ giữa các chất ức chế và thụ thể Eg5 đã được chúng tôi xây dựng dựa trên thuật toán bình phương tối thiểu từng phần (PLS), với tập xây dựng gồm 74 chất và tập kiểm tra bao gồm 18 chất. Mô hình 2D-QSAR có thể ứng dụng vào mục đích sàng lọc ảo, định hướng, khám phá và thiết kế các cấu trúc có tác dụng ức chế Eg5. Từ đó, giúp cho tiết kiệm được rất nhiều thời gian, chi phí trong quá trình nghiên cứu và phát triển thành các loại thuốc điều trị ung thư trong tương lai.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Chuẩn bị cơ sở dữ liệu

Trong nghiên cứu này, mô hình 2D-QSAR được xây dựng với cơ sở dữ liệu gồm 95 chất được thu thập từ các nghiên cứu [7-11] với 4 nhóm khung cấu trúc bao gồm dihydropyrazol, 4-phenyltetrahydroisoquinolin và thiophen. Cơ sở dữ liệu gồm 95 chất có giá trị thử nghiệm hoạt tính sinh học IC_{50} xác định từ 0,0012 μ M đến 76 μ M.

2.1.1. Chuẩn bị cấu trúc hóa học

Công thức hóa học của các chất được vẽ dưới dạng 2D bằng phần mềm ChemBioDraw Ultra 13.0. Sau đó, các công thức đã vẽ sẽ được tổng hợp lại thành một cơ sở dữ liệu dưới dạng MDL MOL format. Tập hợp các chất này sẽ được chuẩn hoá cấu dạng và tối thiểu hoá năng lượng bằng MOE 2008.

2.1.2. Chuẩn bị giá trị pIC_{50}

Tập chất vừa xử lý sẽ được nhập vào các giá trị IC_{50} với đơn vị là μ M. Do giá trị IC_{50} sẽ có khoảng biến đổi lớn nên được quy đổi thành giá trị pIC_{50} bằng công thức (1), với IC_{50} được đổi sang đơn vị Mol để việc xây dựng mô hình có kết quả chính xác hơn [12].

$$pIC_{50} = -\log_{10}(IC_{50}) \quad (1)$$

2.2. Tính toán thông số mô tả phân tử

Thông số mô tả phân tử của các chất hóa học trong cơ sở dữ liệu được tính toán trong phần mềm MOE 2008 bằng công cụ Computer QuaSAR – Descriptor. Các thông số bao gồm: tính chất vật lý, diện tích bề mặt được phân chia, số nguyên tử, số liên kết, chỉ số hình dạng Kappa, chỉ số liên kết Kier&Hall, thông số ma trận khoảng cách và phụ cận, thông số đặc trưng pharmacophore, thông số tích điện từng phần.

2.3. Phân chia tập hợp cơ sở dữ liệu

Cơ sở dữ liệu được chia thành hai tập. Tập huấn luyện (Training Set) chứa 80% số chất trong cơ sở dữ liệu, được dùng để xây dựng mô hình; Tập kiểm tra (Test Set) chứa 20% số chất còn lại trong cơ sở dữ liệu để đánh giá mô hình đã xây dựng. Trong nghiên cứu này, cơ sở dữ liệu được phân chia tập hợp theo hai phương pháp là phân phối đa dạng (diverse) và phân phối ngẫu nhiên (random).

2.4. Loại thông số mô tả

Để loại các thông số không có ý nghĩa hoặc làm cho mô hình tốt giả tạo, các thông số mô tả của tập huấn luyện được loại thô bằng cách:

Loại bỏ thông số có > 20% chất chứa giá trị 0 bằng phần mềm Excel 2007. Loại bỏ thông số có giá trị của tất cả các chất giống nhau bằng phần mềm RapidMiner Studio 7.0. Loại bỏ một trong hai thông số có tương quan chéo với nhau $\geq 90\%$ bằng phần mềm RapidMiner Studio 7.0.

2.5. Chia tỉ lệ thông số mô tả

Chia tỉ lệ thông số mô tả là một vấn đề tương tự như vẽ đồ thị [13]. Với đồ thị, các trục được chia tỉ lệ để thể hiện giá trị của biến (trong nghiên cứu QSAR chính là các giá trị thông số ứng với mỗi chất) trong một hệ tham chiếu khác với thực tế. Đây là một bước quan trọng trước khi xây dựng mô hình. Mục đích của việc chia tỉ lệ là để tránh sự ảnh hưởng của các thông số có giá trị lớn đối với các thông số có giá trị nhỏ và giảm bớt khó khăn cho việc tính toán.

Trong nghiên cứu này áp dụng phương pháp chia tỉ lệ khoảng. Các thông số mô tả được chia tỉ lệ trong khoảng [0–1] bằng công cụ

Normalize trong phần mềm RapidMiner Studio 7.0. Công thức chia tỉ lệ như sau:

$$X_n = \frac{X_0 - Min_0}{Max_0 - Min_0} \quad (2)$$

Trong đó: X_n là giá trị mới ; X_0 là giá trị ban đầu ; Min_0 và Max_0 là giá trị nhỏ nhất và lớn nhất của những giá trị ban đầu.

2.6. Lựa chọn thông số mô tả phân tử

Tập huấn luyện sau khi được chia tỉ lệ trong khoảng [0–1], sẽ được lựa chọn thông số mô tả phân tử để xây dựng mô hình dự đoán bằng phần mềm Weka 3.6 và công cụ QSAR-Contignecy trong phần mềm MOE. Các thông số mô tả phù hợp với phương trình QSAR khi hệ số ngẫu nhiên $C > 0,6$ và $V > 0,2$; hệ số bất định $U > 0,2$ và hệ số tương quan $R^2 > 0,2$.

2.7. Xây dựng mô hình

Mô hình các chất có hoạt tính ức chế enzyme kinesin Eg5 được xây dựng bằng phương pháp PLS - Partial Least Square (bình phương tối thiểu từng phần). Mô hình được xây dựng bằng công cụ QuaSAR-Model trong phần mềm MOE 2008.

2.8. Loại chất gây nhiễu

Trong nghiên cứu này các chất gây nhiễu được loại bỏ bằng công cụ Z-Score và công cụ PCA trong MOE. Tập huấn luyện sau khi xây dựng mô hình sẽ có được giá trị Z-Score. Các chất gây nhiễu được loại bỏ khi giá trị Z-Score $> 2,5$. Loại nhiễu bằng phân tích thành phần chính PCA với ưu điểm là loại nhanh và chính xác các chất gây nhiễu do tập dữ liệu được biểu diễn trực quan trong không gian 3 chiều, mức độ tập trung của các chất và các chất gây nhiễu sẽ nằm rời rạc so với vùng tuyến tính của tập dữ liệu.

2.9. Đánh giá mô hình

Mục tiêu của QSAR là xây dựng được mô hình có khả năng dự đoán tốt, có nghĩa là giá trị dự đoán và giá trị thực tế khác nhau không ý nghĩa. Để đánh giá được điều này cần dựa trên các giá trị p, bình phương hệ số tương quan R^2 và căn bậc hai của tổng bình phương phần dư RMSE của tập huấn luyện. Các giá trị này được thẩm

định bằng cách đánh giá nội (đánh giá chéo *bỏ ra một - LOO*) và đánh giá trên tập ngoại. Đánh giá nội được tiến hành đồng thời trong quá trình xây dựng mô hình bằng công cụ *Validate* trong cửa sổ *QSAR-Model* trong phần mềm MOE. Các thông số đánh giá là giá trị trung bình bình phương hệ số tương quan của các mô hình thứ cấp Q^2 (XR^2) và RMSE (XR_{RMSE}) đánh giá chéo. Các thông số RMSE, Q^2 , R^2_{pred} thường được sử dụng để đánh giá mô hình hồi quy:

$$Q^2 = 1 - \frac{\sum(Y_{pred} - Y)^2}{\sum(Y - \bar{Y})^2} \quad (3)$$

$$R^2_{pred}$$

$$= 1 - \frac{\sum(Y_{pred(test)} - Y_{test})^2}{\sum(Y_{test} - \bar{Y}_{train})^2} \quad (4)$$

Trong đó: Y_{pred} là giá trị dự đoán trên tập huấn luyện bởi mô hình đã được xây dựng. Y là giá trị sinh học (pIC_{50}) thực các chất trong tập huấn luyện, \bar{Y} là giá trị hoạt tính trung bình. $Y_{pres(test)}$ là giá trị dự đoán trên tập kiểm tra bởi mô hình đã được xây dựng. Y_{test} là giá trị sinh học (pIC_{50}) thực các chất trong tập kiểm tra. \bar{Y}_{train} là giá trị hoạt tính trung bình.

Mô hình được chấp nhận khi Q^2 và $R^2 > 0,5$ [14].

Ngoài ra, giá trị r_m^2 cũng được sử dụng để đánh giá khả năng dự đoán của mô hình được xây dựng với ưu điểm là tránh được sai số dự đoán do khoảng rộng của giá trị trở về (Y) [14].

$$r_m^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (5)$$

Trong đó: r^2 là hệ số tương quan. r_0^2 là hệ số tương quan với hệ số chặn bằng 0. Với trục x là giá trị hoạt tính dự đoán Y_{pred} và trục y là giá trị hoạt tính thực Y .

$$r_m'^2 = r'^2 \left(1 - \sqrt{r'^2 - r_0'^2}\right) \quad (6)$$

Trong đó: r'^2 là hệ số tương quan. $r_0'^2$ là hệ số tương quan với hệ số chặn bằng 0. Với trục x là giá trị hoạt tính thực Y và trục y là giá trị hoạt tính dự đoán Y_{pred} . Từ (5) và (6) tính được:

$$\overline{r_m^2} = \frac{r_m^2 + r_m'^2}{2} \text{ và } \Delta r_m^2 = |r_m^2 - r_m'^2|.$$

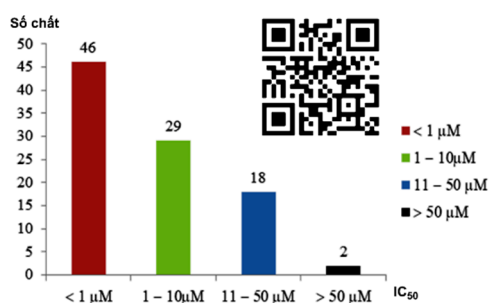
Mô hình có khả năng dự đoán tốt khi $\overline{r_m^2} \geq 0,5$ và $\Delta r_m^2 \leq 0,2$ [14].

Tập ngoại (tập kiểm tra) được tách riêng từ đầu không tham gia vào quá trình xây dựng mô hình để đảm bảo tính khách quan cho khả năng dự đoán của mô hình. Tập ngoại được ứng dụng để đánh giá mô hình bằng công cụ *Computer/Model-Evaluate* trong MOE 2008. Phần lớn các giá trị dự đoán của tập ngoại phải nằm trong khoảng tin cậy 95%. Và còn có các thông số đánh giá là bình phương hệ số tương quan R^2 và căn của tổng bình phương phần dư RMSE giữa giá trị dự đoán và giá trị thực nghiệm.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Xây dựng và đánh giá mô hình

Cơ sở dữ liệu bao gồm 95 hợp chất với khả năng ức chế enzyme Eg5 có giá trị IC_{50} trong khoảng từ 0,0012 μM đến 76 μM . Hình 1 cho thấy sự phân bố IC_{50} theo số lượng các hợp chất, trong đó mã QR-Code dẫn đến tập dữ liệu xây dựng mô hình 2D-QSAR bao gồm 95 hợp chất với công thức hoá học, các thông số IC_{50} thực nghiệm, pIC_{50} thực nghiệm và pIC_{50} được dự đoán.



Hình 1. Đồ thị phân bố IC_{50} theo số lượng các hợp chất

Mã QR-Code dẫn đến tập dữ liệu xây dựng mô hình 2D-QSAR bao gồm 95 hợp chất

Dữ liệu 95 chất được chia thành hai tập nhỏ là tập huấn luyện 76 chất (80%) và tập kiểm tra 19 chất (20%) được lựa chọn ngẫu nhiên và phân chia đa dạng. Thông số mô tả được lựa chọn

bằng phần mềm Weka 3.6 gồm 7 thông số bao gồm chiral, PEOE_VSA+2, Q_VSA_NEG, vsa_don, SlogP_VSA2, SlogP_VSA3 và SMR_VSA1. Mối tương quan giữa các thông số được tính toán bằng công cụ Correlation Matrix trong phần mềm MOE 2008 và được thể hiện trong Bảng 1.

Các thông số được lựa chọn tiếp tục bằng công cụ *Qsar-contingency* với các thông số và điều kiện $C > 0,6$; $V > 0,2$; $U > 0,2$; $R^2 > 0,2$ là tốt, đồng thời dựa trên phân tích thành phần cơ bản (PCA) của các thông số trong việc xây dựng mô hình, giá trị các thông số được trình bày trong Bảng 1.

Trong đó có 5 thông số mô tả được lựa chọn để xây dựng mô hình gồm: PEOE_VSA+2, Q_VSA_NEG, vsa_don, SlogP_VSA2, SMR_VSA1 và loại ra 2 thông số là chiral và SlogP_VSA3. Mô hình toàn tập được xây dựng với 5 thông số lựa chọn có kết quả như Bảng 2.

Trong nghiên cứu này có 3 chất gây nhiễu được xác định là MOL-2012-17-2015-B10, BMCL-2007-17-3652-33 và BMCL-2006-16-2095-21 (xem cấu trúc các hợp chất trong mã QR-Code-Hình 1), sau đó chúng đã được loại bỏ bằng công cụ *Z-score* và PCA.

Mô hình được xây dựng sau khi loại nhiễu, được tiến hành phân chia đa dạng và phân chia ngẫu nhiên ($k = 5$) với kết quả được thể hiện như Bảng 3.

Mức độ tương quan giữa giá trị dự đoán và giá trị thực nghiệm khi đánh giá chéo là 0,84 cho thấy mô hình có tính tương quan cao. Giá trị RMSE thể hiện sai số trung bình giữa pIC_{50} dự đoán và thực nghiệm là 0,46. Tập huấn luyện có giá trị $R^2 = 0,86 > 0,5$; giá trị $\overline{r_m^2} = 0,81 > 0,5$ và giá trị $\Delta r_m^2 = 0,09 < 0,2$. Như vậy mô hình xây dựng đạt điều kiện đánh giá và có khả năng dự đoán tốt. Đường thẳng tương quan pIC_{50} dự đoán và thực nghiệm của toàn bộ tập dữ liệu 92 chất xây dựng mô hình (đã loại nhiễu) được biểu diễn trong Hình 2a. Mô hình được đánh giá với tập kiểm tra gồm 18 chất theo phân chia đa dạng, đảm bảo tính khách quan khi ứng dụng mô hình.

Bảng 1. Mối tương quan giữa các thông số mô tả với pIC_{50} và giá trị thông số theo C, U, V, R²

	chiral	PEOE_VSA+2	Q_VSA_NEG	vsa_don	Slop_VSA2	SlogP_VSA3	SMR_VSA1	pIC_{50}	C	V	U	R ²
chiral	1,00								0,71177	0,29252	0,19869	0,25101
PEOE_VSA+2	0,41	1,00							0,80079	0,38569	0,42865	0,54862
Q_VSA_NEG	0,37	0,54	1,00						0,75087	0,32820	0,35423	0,41505
vsa_don	-0,38	-0,007	0,04	1,00					0,50818	0,17033	0,11921	0,09452
SlogP_VSA2	0,46	0,67	0,58	-0,21	1,00				0,86189	0,49064	0,47278	0,72844
SlogP_VSA3	0,10	0,33	0,22	-0,10	0,55	1,00			0,78484	0,36561	0,31607	0,17982
SMR_VSA1	0,49	0,72	0,67	-0,13	0,69	0,21	1,00		0,82620	0,42335	0,46396	0,56250
pIC_{50}	0,5	0,74	0,64	-0,31	0,85	0,42	0,75	1,00				

Bảng 2. Kết quả mô hình toàn tập được xây dựng với 5 thông số lựa chọn

Giá trị	RMSE	R ²	RMSE đánh giá chéo	R ² (Q ²) đánh giá chéo
Trước loại nhiễu (N = 95)	0,51	0,83	0,54	0,81
Sau loại nhiễu (N = 92)	0,47	0,86	0,50	0,84

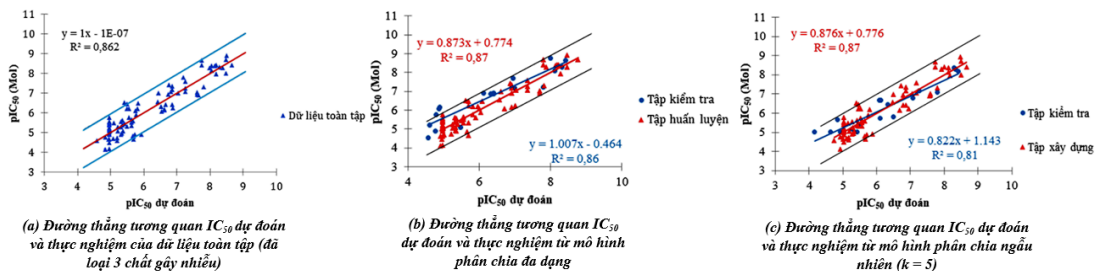
Bảng 3. Thông số đánh giá kết quả khả năng dự đoán tập huấn luyện và tập kiểm tra của mô hình

Chỉ tiêu	Phân chia đa dạng			Phân chia ngẫu nhiên		Toàn tập
	Tập huấn luyện	Tập kiểm tra	Đánh giá chéo	Tập huấn luyện	Tập kiểm tra	
Số chất	74	18	92	74	18	92
RMSE	0,44	0,30	0,50	0,46	0,38	0,46
R ² /Q ²	0,87	-	0,84	0,87	-	0,86
R ² _{pred}	-	0,86	-	-	0,81	-
r ² _m	0,75	0,69	-	0,76	0,64	0,74
r' ² _m	0,87	0,77	-	0,85	0,69	0,86
r ² _m	0,81	0,72	-	0,81	0,67	0,80
Δr ² _m	0,12	0,11	-	0,09	0,07	0,12

Tất cả giá trị $\overline{r_m^2} > 0,5$ và giá trị $\Delta r_m^2 < 0,2$, cho thấy mô hình được xây dựng có khả năng dự đoán tốt.

Kết quả mô hình thu được theo phân chia đa dạng của tập kiểm tra với R² = 0,86, $\overline{r_m^2}$ = 0,72 và Δr_m^2 = 0,11. Giá trị dự đoán các chất đều nằm trong khoảng 95% dự đoán đúng của mô hình được thể hiện trong Hình 2b. Ngoài ra, mô hình được đánh giá trên tập ngoại gồm 18 chất

được lấy theo phân phối ngẫu nhiên (k = 5). Kết quả tập kiểm tra theo phân phối ngẫu nhiên với các giá trị R² = 0,81, $\overline{r_m^2}$ = 0,67 và Δr_m^2 = 0,07. Giá trị dự đoán của các chất đều nằm trong khoảng 95% dự đoán đúng của mô hình, kết quả được trình bày trong Hình 2c.



Hình 2. Đường biểu diễn tương quan IC₅₀ dự đoán và thực nghiệm của mô hình trong từng trường hợp

3.2. Ứng dụng mô hình 2D – QSAR

Mô hình sau khi kiểm tra và đáp ứng các điều kiện, được sử dụng để sàng lọc ảo các chất thu thập được từ nguồn thư viện Ngân hàng thuốc (Drug Bank) và ChemDiv.

Kết quả cho thấy từ 12066 hợp chất thu thập từ hai cơ sở dữ liệu (Drug Bank và ChemDiv) sau khi áp dụng mô hình 2D-QSAR đã sàng lọc ra được 349 hợp chất cho thấy tiềm năng trong việc ức chế kinesin Eg5 (với giá trị $IC_{50} < 0,5 \mu M$). Cụ thể, từ 1661 hợp chất của thư viện Drug Bank sau khi lọc theo quy luật 5 lipinski thu được 831 hợp chất, sau khi ứng dụng các thông số của mô hình 2D-QSAR thì thu được 296 hợp chất, trong số đó có 45 hợp chất cho thấy có hoạt tính sinh học cao với $IC_{50} < 0,5 \mu M$. Tương tự, từ 10405 hợp chất từ thư viện ChemDiv thu được kết quả lần lượt là 3642 hợp chất (sau khi lọc qua quy luật 5 lipinski) và 2464 hợp chất (sau khi lọc qua thông số của mô hình 2D-QSAR), trong số này có 304 chất có hoạt tính sinh học rất cao ($IC_{50} < 0,1 \mu M$). Đặc biệt, từ dữ liệu Drugbank cho thấy 4 loại thuốc là Dicloxacillin, Cloxacillin, Tazobactam và Oxacilin cho thấy hoạt tính cao với $IC_{50} < 0,008 \mu M$, các loại thuốc này đang lưu hành trên thị trường và có thể sẵn sàng tiến hành các thử nghiệm trên Eg5 mà không cần bước tái tổng hợp lại các hợp chất.

Tóm lại, trong nghiên cứu này chúng tôi đã tiến hành xây dựng thành công mô hình 2D-QSAR từ cơ sở dữ liệu bao gồm 95 hợp chất ức chế Eg5. Mô hình đã bước đầu được ứng dụng vào sàng lọc các hợp chất từ cơ sở dữ liệu Drugbank và Chemdiv, 349 hợp chất được phát hiện là tiềm năng để ức chế Eg5. Trong tương lai, các hợp chất này tiếp tục được tiến hành đánh giá và sàng lọc về khả năng liên kết giữa enzyme Eg5 và phối tử bằng các phương pháp mô phỏng và động lực học, từ đó chọn ra các hợp chất tốt nhất cho các thử nghiệm *in vitro* và *in vivo*.

4. KẾT LUẬN

Trong nghiên cứu này, mô hình 2D-QSAR đã được thực hiện để sàng lọc các hợp chất cho khả năng ức chế enzyme Eg5. Mô hình này được ứng dụng vào sàng lọc ảo để phát hiện, định hướng thiết kế và phát triển các loại thuốc mới có tác dụng nhắm mục tiêu Eg5 trong điều trị ung thư sau này. Nghiên cứu này là cơ sở cho các nghiên cứu tiếp theo về xây dựng các mô hình 2D-QSAR tương tự cho các mục tiêu điều trị ung thư trong tương lai.

TÀI LIỆU THAM KHẢO

- [1] Cochran, J. C., Krzysiak, T. C., & Gilbert, S. P, (2006). Pathway of ATP hydrolysis by monomeric kinesin Eg5. *Biochemistry*, **45(40)**, 12334-12344.
- [2] Wojcik, E. J., Buckley, R. S., Richard, J., Liu, L., Huckaba, T. M., & Kim, S, (2013). Kinesin-5: cross-bridging mechanism to targeted clinical therapy. *Gene*, **531(2)**, 133-149.
- [3] Heck, M. M., Pereira, A., Pesavento, P., Yannoni, Y., Spradling, A. C., & Goldstein, L. S, (1993). The kinesin-like protein KLP61F is essential for mitosis in Drosophila. *The Journal of cell biology*, **123(3)**, 665-679.
- [4] Mayer, T. U., Kapoor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L., & Mitchison, T. J, (1999). Small Molecule Inhibitor of Mitotic Spindle Bipolarity Identified in a Phenotype-Based Screen. *Science*, **286(5441)**, 971-974.
- [5] Luo, X., Shu, M., Wang, Y., Liu, J., Yang, W., & Lin, Z, (2012). 3D-QSAR Studies of Dihydropyrazole and Dihydropyrrole Derivatives as Inhibitors of Human Mitotic Kinesin Eg5 Based on Molecular Docking. *Molecules*, **17(2)**, 2015-2029.
- [6] Chen, G.-Y., Kang, Y. J., Gayek, A. S., Youyen, W., Tüzel, E., Ohi, R., & Hancock, W. O, (2017). Eg5 Inhibitors Have Contrasting Effects on Microtubule Stability and Metaphase Spindle Integrity. *ACS Chemical Biology*, **12(4)**, 1038-1046.

- [7] [7] Cox, C. D., Torrent, M., Breslin, M. J., Mariano, B. J., Whitman, D. B., Coleman, P. J., ... & Hartman, G. D, (2006). Kinesin spindle protein (KSP) inhibitors. Part 4: Structure-based design of 5-alkylamino-3, 5-diaryl-4, 5-dihydropyrazoles as potent, water-soluble inhibitors of the mitotic kinesin KSP. *Bioorganic & medicinal chemistry letters*, **16(12)**, 3175-3179.
- [8] [8] Fraley, M. E., Garbaccio, R. M., Arrington, K. L., Hoffman, W. F., Tasber, E. S., Coleman, P. J., ... & Hartman, G. D, (2006). Kinesin spindle protein (KSP) inhibitors. Part 2: The design, synthesis, and characterization of 2, 4-diaryl-2, 5-dihydropyrrole inhibitors of the mitotic kinesin KSP. *Bioorganic & Medicinal Chemistry Letters*, **16(7)**, 1775-1779.
- [9] [9] Tarby, C. M., Kaltenbach III, R. F., Huynh, T., Pudzianowski, A., Shen, H., Ortega-Nanos, M., ... & Roussell, D. L, (2006). Inhibitors of human mitotic kinesin Eg5: characterization of the 4-phenyl-tetrahydroisoquinoline lead series. *Bioorganic & medicinal chemistry letters*, **16(8)**, 2095-2100.
- [10] [10] Pinkerton, A. B., Lee, T. T., Hoffman, T. Z., Wang, Y., Kahraman, M., Cook, T. G., ... & Davis, R. L, (2007). Synthesis and SAR of thiophene containing kinesin spindle protein (KSP) inhibitors. *Bioorganic & Medicinal Chemistry Letters*, **17(13)**, 3562-3569.
- [11] [11] Roecker, A. J., Coleman, P. J., Mercer, S. P., Schreier, J. D., Buser, C. A., Walsh, E. S., ... & Hartman, G. D, (2007). Kinesin spindle protein (KSP) inhibitors. Part 8: Design and synthesis of 1, 4-diaryl-4, 5-dihydropyrazoles as potent inhibitors of the mitotic kinesin KSP. *Bioorganic & medicinal chemistry letters*, **17(20)**, 5677-5682.
- [12] [12] Stewart, M. J., & Watson, I. D, (1983). Standard units for expressing drug concentrations in biological fluids. *British journal of clinical pharmacology*, **16(1)**, 3.
- [13] [13] Livingstone, D, (1995). Data Analysis for Chemists. *Oxford Science Publications*, Oxford University Press.
- [14] [14] Ojha, P. K., Mitra, I., Das, R. N., & Roy, K, (2011). Further exploring rm2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems*, **107(1)**, 194-205.