

Ứng dụng phễu lọc hóa dược và mô hình docking phân tử sàng lọc chất ức chế STAT3

Tô Văn Thịnh¹, Trần Nguyễn Ngọc Tâm¹, Nguyễn Văn Phước Chung¹, Trương Ngọc Tuyền¹,
Nguyễn Thị Hương Giang^{1,*}

¹Trường Dược, Đại học Y Dược Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, Việt Nam

Tóm tắt

Đặt vấn đề: Ung thư là một trong những nguyên nhân gây tử vong hàng đầu trên toàn thế giới với gần 10 triệu ca tử vong được ghi nhận vào năm 2022, theo báo cáo của Tổ chức Y tế Thế giới. Mặc dù đích tác động STAT3 (Truyền tín hiệu và kích hoạt phiên mã 3) đã được biết đến như một con đường đầy hứa hẹn để kiểm soát sự phát triển của khối u từ lâu. Nhưng chỉ đến gần đây, cucurbitacin mới được nghiên cứu rộng rãi như là chất tiềm năng để điều trị ung thư bằng cách kích hoạt apoptosis thông qua ức chế chọn lọc đường dẫn JAK / STAT.

Mục tiêu: Ứng dụng phễu lọc hóa dược và mô hình docking phân tử sàng lọc chất ức chế STAT3 đối với các dẫn chất cucurbitacin E.

Đối tượng và phương pháp nghiên cứu: Cơ sở dữ liệu gồm 1056 chất ức chế STAT3 tiềm năng, dựa trên cấu trúc cucurbitacin E, được sàng lọc thông qua các bộ lọc hóa dược, và sau đó trải qua quá trình docking phân tử trên protein STAT3 (PDB ID: 6NJS).

Kết quả: Thông qua bộ lọc hóa dược, đã xác định được 85 hợp chất tiềm năng ức chế STAT3. Các hợp chất này tiếp tục được đánh giá bằng phương pháp docking phân tử, từ đó lựa chọn được 66 hợp chất có ái lực liên kết mạnh đối với STAT3. Dữ liệu và code được công khai trên Github: <https://github.com/phuocchung123/stat3/tree/main>

Kết luận: Thông qua việc sử dụng kết hợp phễu lọc hóa dược và quá trình gắn kết phân tử, 66 dẫn chất cucurbitacin E được xác định là những chất ức chế STAT3 tiềm năng. Mô phỏng động lực học phân tử, tổng hợp và thử nghiệm intro sẽ được tiến hành để đánh giá hoạt tính sinh học của những chất này.

Từ khóa: phễu lọc hóa dược; chất ức chế STAT3; STAT3 protein; phễu lọc ADMET; gắn kết phân tử; Cucurbitacin E

Abstract

APPLICATION OF MEDICINAL CHEMISTRY FILTRATION AND MOLECULAR DOCKING IN SELECTING STAT3 INHIBITORS FROM CUCURBITACIN E DERIVATIVES

To Van Thinh, Tran Nguyen Ngoc Tam, Nguyen Van Phuoc Chung, Truong Ngoc Tuyen, Nguyen Thi Huong Giang

Background: Cancer is one of the leading causes of mortality worldwide with nearly 10 million deaths recorded in 2022, according to the World Health

Ngày nhận bài: 14-07-2025 / Ngày chấp nhận đăng bài: 17-12-2025 / Ngày đăng bài: 15-01-2026

*Tác giả liên hệ: Nguyễn Thị Hương Giang. Trường Dược, Đại học Y Dược Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, Việt Nam.

E-mail: huonggiang@ump.edu.vn

© 2026 Bản quyền thuộc về Tạp chí Y học Thành phố Hồ Chí Minh.

Organization report. While targeting signal transducer and activator of transcription 3 (STAT3) has long been known as a promising pathway for controlling tumor growth. It is only in recent years that cucurbitacin is considered potential for anti-cancer therapy by triggering apoptosis via selectively inhibiting the JAK/STAT path.

Objectives: To select STAT3 inhibitors from cucurbitacin E derivatives by the application of medicinal chemistry filtration and molecular docking.

Methods: The database of 1056 potential STAT3 inhibitors, which were based on cucurbitacin E scaffold, was screened through the medicinal chemistry filters, and then experienced molecular docking on STAT3 protein (PDB ID: 6NJS).

Results: Through medicinal chemistry filters, 85 potential STAT3 inhibitors were identified. These compounds were subsequently evaluated using molecular docking, resulting in the selection of 66 compounds with strong binding affinity toward STAT3. All code and data are publicly accessible in our repository: <https://github.com/phuocchung123/stat3/tree/main>

Conclusion: 66 cucurbitacin E derivatives were identified as potential STAT3 inhibitors using medicinal chemistry filters and molecular docking. Further molecular dynamics simulations, synthesis, and in vitro assays will be conducted to validate their biological activity.

Keywords: medicinal chemistry filtration, STAT3 inhibitors, STAT3 protein, ADMET filter, docking, Cucurbitacin E

1. ĐẶT VẤN ĐỀ

Ung thư là một trong những nguyên nhân gây tử vong hàng đầu trên toàn thế giới với gần 10 triệu ca tử vong được ghi nhận vào năm 2022, theo báo cáo của Tổ chức Y tế Thế giới [1]. Mặc dù đích tác động STAT3 (Truyền tín hiệu và kích hoạt phiên mã 3) đã được biết đến như một con đường đầy hứa hẹn để kiểm soát sự phát triển của khối u từ lâu. Nhưng chỉ đến gần đây, cucurbitacin mới được nghiên cứu rộng rãi như là chất tiềm năng để điều trị ung thư bằng cách kích hoạt apoptosis thông qua ức chế chọn lọc đường dẫn JAK / STAT [2].

Cucurbitacin (cucur) là những triterpene bốn vòng thiên nhiên chiết xuất từ họ Bầu bí, được chia thành 12 loại từ A đến T [3]. Trong đó, vai trò đối kháng STAT3 được biết đến đầu tiên là ở cucur I với nghiên cứu của Blaskovich và cộng sự, đã chứng minh cucur I là một chất ức chế JAK-STAT mạnh trong nhiều loại ung thư ở người [2]. Bên cạnh đó, cucur E – một trong hai cucurbitacin chính được phân lập từ họ Bầu bí [4] – là dạng acetyl hóa của cucur I, cũng thể hiện hoạt tính ức chế STAT3 trên nhiều dòng tế bào ung thư, bao gồm buồng trứng [5] và ung thư vú [6].

Hiện nay, việc ứng dụng khoa học máy tính vào lĩnh vực thiết kế thuốc ngày càng được áp dụng rộng rãi, các phương pháp sàng lọc *in silico* đã cho thấy nhiều thành tựu trong việc dự đoán khả năng đáp ứng thuốc, dự đoán độc tính, hoạt tính kháng ung thư, ... Bên cạnh đó, một hoạt chất muốn được lưu hành rộng rãi trên thị trường Dược phẩm thì các yếu tố được

động học (bao gồm hấp thu, phân bố, chuyển hóa, thải trừ, độc tính - ADMET) cũng cần được quan tâm. Chính vì vậy, các chất cần được sàng lọc trước thông qua bộ lọc hóa dược. Ngoài ra, phương pháp mô phỏng gắn kết phân tử (molecular docking) cũng đóng một vai trò quan trọng trong các phương pháp sàng lọc ảo nhằm dự đoán cấu dạng, liên kết và định hướng không gian để hình thành phức hợp gắn kết thụ thể-phối tử tối ưu. Từ các cơ sở trên, phễu lọc hóa dược và mô hình mô phỏng gắn kết phân tử được ứng dụng để sàng lọc các ứng cử viên tiềm năng thể hiện hoạt tính ức chế STAT3 từ bộ thư viện các dẫn chất cucurbitacin E.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng nghiên cứu

2.1.1. Tài nguyên tính toán

Nghiên cứu *in silico* được tiến hành trên máy tính với hệ điều hành Linux 20.04, CPU AMD Ryzen 9 3900X, 12 nhân, tốc độ xử lý là 3,79 GHz, 500 GB bộ nhớ, 32,0 GB RAM tại Bộ môn Hóa hữu cơ, Trường Dược, Đại Học Y Dược Thành phố Hồ Chí Minh. Ngôn ngữ lập trình được sử dụng là Python 2.7 và 3.7, được thực thi bằng giao diện JupyterLab.

2.1.2. Dữ liệu kiểm tra hồi chứng

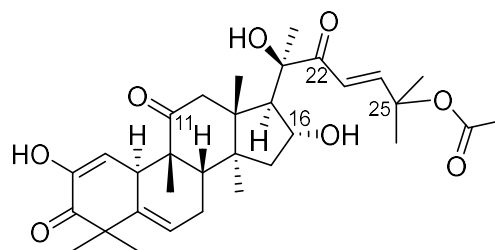
Bộ dữ liệu gồm 142 chất ức chế STAT3 có cùng phương pháp đo hoặc phương pháp đo tương đồng nhau được trích xuất từ Google Patents (US20130225621, US0172429,

US8741932B2, US8466290B2, US129675A1 và US20130172340) cũng như nhiều nguồn khoa học đã được công bố [7-9]. Hoạt tính của các cấu trúc này có giá trị IC_{50} dao động trong khoảng từ 68 nM đến gần 1000 nM. Các chất này sau đó được phân cụm bằng thuật toán Butina để chọn ra tập cấu trúc đa dạng nhất trong không gian hóa học (hệ số Tanimoto $\leq 0,35$) [10]. Kết quả thu được 21 chất cùng với ligand đồng kết tinh của cấu trúc mục tiêu STAT3 (PDB ID: 6NJS) là SD-36 để tạo thành tập có hoạt tính, và tiến hành tạo tập mồi nhử (decoy) bằng thư viện học sâu Deepcoy [11] với tỷ lệ active:decoy là 1:50. Kết quả thu được 22 chất có hoạt tính và 1100 mồi nhử.

2.1.3. Dữ liệu sàng lọc

Cucurbitacin (cucur) là những triterpene bốn vòng thiên nhiên chiết xuất từ họ Bầu bí, được chia thành 12 loại từ A đến T với hơn 200 dẫn xuất [3], vốn được sử dụng lâu đời trong “Guadi” (tiếng Trung), một loại thảo mộc dùng làm thuốc bổ trợ điều trị viêm gan mãn tính và ung thư gan nguyên phát ở Trung Quốc [12].

Cucur có nhiều tác dụng dược lý khác nhau như khả năng hạ sốt, giảm đau, chống viêm, kháng khuẩn và chống khối u [12]. Trong số đó, cucurbitacin B, D, E, I và các dẫn xuất của chúng được nghiên cứu rộng rãi về các hoạt tính chống khối u mạnh mẽ, từ chống tăng sinh, bắt giữ chu kỳ tế bào đến khởi phát quá trình apoptosis [12]. Trong đó, chỉ có hai cucur chính được tổng hợp từ bầu bí là cucur E và B với nhóm chức acetyl ở C_{25} - có liên quan đến tác động dược lý của chúng [4]. Mặc dù thể hiện hoạt tính sinh học đa dạng, cucur đã được báo cáo có liều hoạt tính và liều gây độc xấp xỉ nhau [2]. Nghiên cứu của Dantas và cộng sự đã đưa ra kết luận cho rằng độc tính của cucur có thể do nhóm carbonyl ở C_{11} [13]. Trong một nghiên cứu liên quan đến các thử nghiệm *in vitro* và mối quan hệ định lượng giữa cấu trúc và tác dụng (QSAR) của cucur, các tác giả đã nhận định mức độ phân cực cao hoặc sự hiện diện nhóm hydroxyl trên C_{16} giúp làm giảm độc tính tế bào của cucur [14]. Hay trong một nghiên cứu vào năm 2005, nhóm ceton ở vị trí C_{22} và gốc acetoxyl ở C_{25} được chứng minh có liên quan đến hiệu quả tác động của cucur [4]. Do đó, thư viện nội bộ gồm 1056 chất được thiết kế dựa trên khung cấu trúc của cucurbitacin E bằng cách thay đổi bốn thành phần chính: nhóm carbonyl ở C_{11} ; nhóm hydroxyl ở C_{16} ; nhóm ceton ở C_{22} và nhóm ester ở C_{25} . Các biến đổi về mặt hóa học các dẫn chất tương đồng được thể hiện ở Hình 1.



C-11: =O, CH₃; **C-16:** -OH, =O, -OCH₃, -OC₂H₅, -OC₆H₅, -OC₆H₄-4-Cl, -OC₆H₄-3-Cl, -OC₆H₄-2-Cl; **C-22:** =O, -OH, NHCH(R)COCOOH (R: mạch nhánh của 20 acid amin); **C-25:** -OCOCH₃, -OCOC₂H₅, -OCOC₆H₅

Hình 1. Chiến lược thiết kế dẫn chất từ cucurbitacin E

2.1.4. Protein STAT3

Protein STAT3 liên kết với ligand đồng kết tinh (SD-36) được lấy từ ngân hàng dữ liệu Protein Data Bank (<https://www.rcsb.org>) với mã protein là 6NJS, chụp bằng phương pháp nhiễu xạ tia X với độ phân giải 2,7 Å. Ligand và protein trong phức hợp sau đó được tách riêng và lưu ở định dạng *.pdb, trong đó protein tiếp tục được tối thiểu hoá năng lượng bằng UCSF Chimera và xác định thông số cho khoảng gắn kết dựa trên vị trí gắn kết của ligand đồng kết tinh (SD-36) trong khoang gắn bằng phần mềm UCSF Chimera [15].

2.2. Phương pháp nghiên cứu

Trong nghiên cứu, các phần mềm và thư viện được sử dụng cho các bước chính gồm: vẽ và chuẩn bị cấu trúc phân tử (ChemDraw), xây dựng tập decoy (Deepcoy), tính toán dấu vân tay và thông số mô tả hóa dược (RDKit) và thực hiện mô phỏng gắn kết phân tử (AutoDock Vina, QVina2, Smina).

2.2.1. Phổ lọc Hóa dược

Tính toán thông số mô tả (LogP: tính kỵ nước, HBA: số liên kết nhận hydro, HBD: số liên kết cho hydro, MW: khối lượng phân tử, TPSA: diện tích bề mặt, Fsp3: số carbon lai hóa sp^3 / tổng số carbon) bằng thư viện hoá tin RDKit [16].

Áp dụng hai quy tắc (quy tắc Lipinski: tính thấm màng, quy tắc Pfizer: mối quan hệ giữa các đặc tính hóa lý và độc tính), bốn phép đo định lượng (Phép đo định lượng Fsp3: độ bão hòa carbon, phép đo định lượng MCE-18: công thức tỉ lệ nhóm chức, SA score: Điểm đánh giá khả năng tiếp cận tổng hợp, NP Score: Điểm giống sản phẩm tự nhiên) và ba bộ lọc (Bộ lọc PAINS, BMS, Chelator: sàng lọc các lần truy cập dương tính giả).

Yêu cầu các quy tắc: Lipinski: Không vượt quá 1 trong 4 yếu tố sau: HBA>10, HBD>5, MW> 600, LogP> 5; Pfizer: Không đồng thời thỏa cả 2 yếu tố: LogP≥ 3 và TPSA ≤ 75; Fsp 3: Fsp 3 ≥ 0,42 là phù hợp; MCE-18: Điểm số MCE-18 ≥ 45; SA Score: SA score ≤ 6; NP Score: Thường nằm trong khoảng từ -5 đến 5; PAINS, BMS, Chelator: Không trùng lặp với bất cứ cấu trúc con nào trong bộ lọc, đồng nghĩa với việc không vi phạm các nguyên tắc tương ứng.

2.2.2. Mô hình gắn kết phân tử (Molecular Docking)

Chuẩn bị và ligand

MGLTools (Python 2.7) trên giao diện của Jupyterlab được sử dụng để mở tệp chứa cấu trúc của STAT3 thu được sau khi sàng lọc qua phễu lọc hóa được, tiến hành thêm hydro có thể có của cấu trúc, bao gồm cả hydro phân cực (polar hydrogen) và hydro không phân cực (non-polar hydrogen) và điện tích Gasteiger vào toàn bộ nguyên tử. Cấu trúc cuối cùng của protein được lưu dưới dạng *.pdbqt.

Ligand đồng kết tinh, 21 chất có hoạt tính thu được sau quá trình phân cụm dữ liệu cùng 1100 decoy và 1056 chất trong thư viện sàng lọc, đều được xử lý giống nhau thông qua MGLTools (Python 2.7) trên giao diện của Jupyterlab.

Kiểm tra hồi chứng

Nghiên cứu tiến hành thực hiện re-docking để đánh giá năng lực gắn kết (docking power) và kiểm tra hồi chứng để đánh giá năng lực sàng lọc (screening power) trên cả ba phần mềm khác nhau bao gồm Smina [17], Qvina2 [18] và Autodock Vina 1.2.3 (Vina) [19], từ đó phân tích và so sánh để lựa chọn ra phần mềm phù hợp nhất cho đích tác động STAT3.

Khoang gắn kết: kích thước hộp Gridbox (size_x = 25 Å, size_y = 25 Å, size_z = 25 Å), tọa độ (center_x = 12,599, center_y = 54,028, center_z = 1,242), random seed được mặc định là 42, exhaustiveness bằng 150, num_modes bằng 10.

Re-docking: tiến hành gắn kết phân tử giữa ligand đồng kết tinh (SD-36) và cấu trúc protein STAT3 đã được xử lý. Kết quả re-docking được đánh giá dựa trên sự so sánh định hướng các cấu dạng của ligand trong khoang gắn kết trước và sau quá trình docking (được thể hiện qua giá trị RMSD), ái lực gắn kết thông qua năng lượng gắn kết (kcal/mol) và sự phù hợp về mặt hóa học của các tương tác trong phức hợp thụ thể-phối tử.

Kiểm tra hồi chứng: tiến hành gắn kết phân tử 22 chất tập hoạt tính (bao gồm ligand đồng kết tinh) và 1100 decoy được tạo thành, tiến hành trích xuất điểm số docking âm nhất vào tệp *.csv để xây dựng đường cong ROC và tính diện tích dưới đường cong ROC.

So sánh phần mềm: dựa vào diện tích dưới đường cong ROC của kiểm tra hồi chứng, lựa chọn được phần mềm phù hợp nhất cho quá trình sàng lọc ảo.

2.2.3. Sàng lọc ảo

1056 chất tương đồng cucurbitacin E được sàng lọc lần lượt qua các phễu lọc hoá được và sau đó được tiến hành gắn kết phân tử với phần mềm docking phù hợp nhất. Sau đó, tiến hành phân tích tương tác giữa các chất tiềm năng và protein (thường gặp là liên kết hydro; liên kết pi-pi, carbon-pi; liên kết *Van der Waals*; tĩnh điện;...) để định hướng tổng hợp hữu cơ.

3. KẾT QUẢ

3.1. Mô hình gắn kết phân tử

3.1.1. Kết quả Re-docking

Để đánh giá năng lực gắn kết và chứng minh tính phù hợp của tất cả các thông số được sử dụng trên ba phần mềm (Vina, Qvina2 và Smina), ligand đồng kết tinh SD-36 được sử dụng làm chứng dương cho quá trình re-docking.

Về mặt định hướng cấu dạng và ái lực gắn kết, ligand SD-36 sau quá trình re-docking được nằm gọn trong khoang gắn kết được xác định từ trước bởi phần mềm UCSF Chimera. Kết quả được ghi nhận cụ thể qua Bảng 1.

Bảng 1. Kết quả re-docking của ba phần mềm docking phân tử

Đại lượng đánh giá	Vina	Qvina2	Smina
RMSD (Å)	1,98	2,06	2,05
Điểm số gắn kết phân tử (kcal/mol)	-7,12	-7,20	-7,14

Về mặt tương tác hóa học, có sự tương đồng giữa ligand đồng kết tinh ban đầu và ligand được re-docking trên các acid amin Arg609, Ser611, Glu612, Ser613, Ser636 được tổng hợp cụ thể trong Bảng 2 dưới đây. Có thể thấy cả ba phần mềm Vina, Qvina2 và Smina đều đã tái tạo hầu hết các tương tác quan trọng trong khoang SH2 của STAT3 và cho kết quả phù hợp để tiến hành gắn kết phân tử.

Bảng 2. Các tương tác giữa ligand đồng kết tinh SD-36 với các acid amin của protein STAT3 trước và sau khi re-docking

Trước re-docking	Sau re-docking		
	Qvina2	Smina	Vina
Arg 609: H – acceptor và ionic	Arg 609: H-acceptor và ionic	Arg 609: H-acceptor và ionic	Arg 609: H – acceptor và ionic
Ser 611: H-acceptor	Ser 611: H-acceptor	Ser 611: H-acceptor	Ser 611: H-acceptor
Glu 612: H-acceptor	Glu 612: H-acceptor	Glu 612: H-acceptor	Glu 612: H-acceptor
Ser 613: H-acceptor	Ser 613: H-acceptor	Ser 613: H-acceptor	Ser 613: H-acceptor
Ser 636: H-donor	Ser 636: H-donor	Ser 636: H-donor	Ser 636: H-donor
Glu 638: H-donor và H-acceptor Gln 644: H-acceptor Tyr 657: H-acceptor	Glu 638: H-acceptor Tyr 657: H-acceptor	Glu 638: H-acceptor	Glu 638: H-donor và H-acceptor Gln 644: H-acceptor

3.1.2. Kết quả kiểm tra hồi chứng

Tập decoy được tạo từ thư viện Deepcoy dựa trên phương pháp property-matched sử dụng mô hình DEKOIS có sẵn với tỉ lệ active:decoy là 1:50. Kết thúc quá trình trên, 22 chất có hoạt tính tạo ra được 1100 decoy. Cụ thể, giá trị diện tích

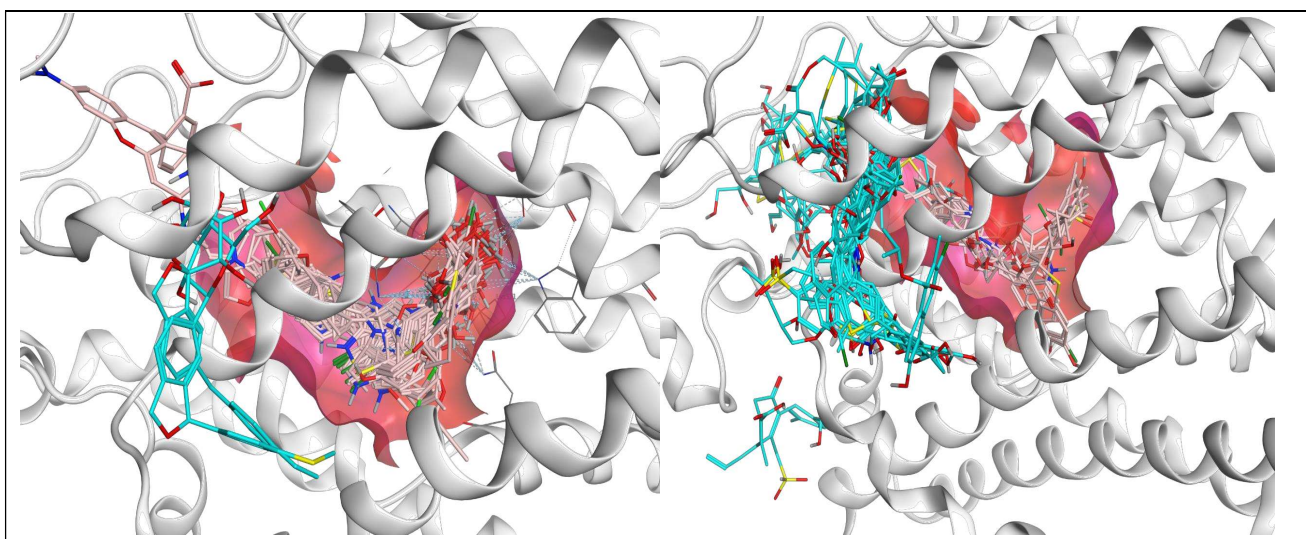
dưới đường cong ROC tiến về 0,5 cho thấy khả năng phân biệt chất có hoạt tính và decoy rất khó khăn. Giá trị diện tích dưới đường cong ROC INN ở nghiên cứu này nằm ở mức chấp nhận được (0,7328).

Các hợp chất nằm ngoài khoang gắn kết (Hình 2) được điều chỉnh giá trị điểm số docking tương ứng bằng với giá trị của hợp chất có số điểm ít âm nhất của từng phần mềm (-3,0; -3,0 và -3,1 kcal/mol đối với Vina, Qvina2 và Smina). Mục đích của việc này nhằm giảm tối đa sai số của mô hình, đồng thời giúp cho bộ dữ liệu điểm số docking sau quá trình chuẩn hóa bằng MinMaxScaler không bị mất cân bằng.

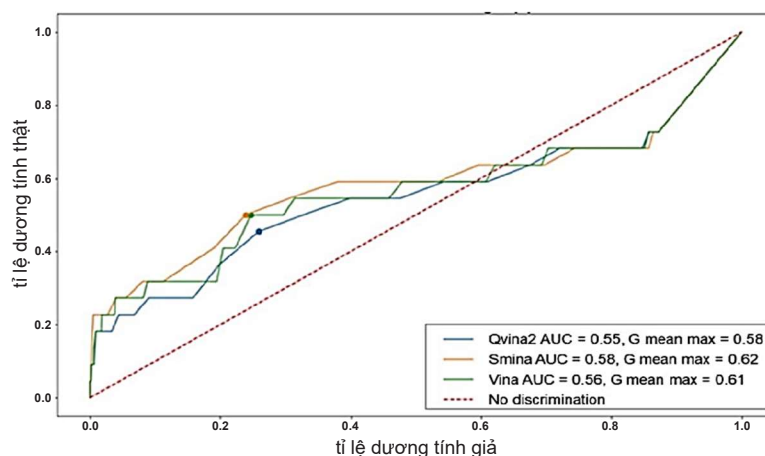
Tiến hành chuyển đổi điểm số docking theo phương pháp MinMaxScaler của thư viện Scikit-learn với khoảng giá trị [0;1], từ đó thiết lập đồ thị diện tích dưới đường cong ROC để đánh giá hiệu năng của kiểm tra hồi chứng các mô hình docking (Hình 3).

Giá trị AUC của đồ thị diện tích dưới đường cong ROC đối với kiểm tra hồi chứng lần lượt là 0,55; 0,56 và 0,58 cho ba phần mềm Qvina2, Vina và Smina. Điểm gây của đồ thị diện tích dưới đường cong ROC được xác định dựa trên điểm G-mean cao nhất đã xác định ngưỡng hoạt tính -5,875 kcal/mol của Qvina2; -5,865 kcal/mol tương ứng Vina và -5,989 kcal/mol là kết quả của phần mềm Smina.

Smina là phần mềm thu được giá trị AUC cao nhất với ngưỡng hoạt tính âm nhất, nên nghiên cứu quyết định lựa chọn Smina để thực hiện mô phỏng gắn kết phân tử dành cho quá trình sàng lọc ảo sau đó.



Hình 2. Hình ảnh minh họa vị trí tập hợp hoạt tính khảo sát trong khoang gắn kết protein STAT3 (trái) và tập decoy (phải) trong khoang gắn kết protein STAT3



Hình 3. Đường cong ROC từ ba phần mềm docking

3.2. Sàng lọc ảo

3.2.1. Sàng lọc qua phễu lọc Hoá dược

Trong khi cucurbitacin E lại có khối lượng phân tử tương đối cao vi phạm luật Lipinski, (556,7 g/mol) nhưng không đồng nghĩa với việc phân tử này không có tiềm năng làm thuốc. Điều này được minh chứng trong nghiên cứu phân tích MW của các thuốc đường uống được FDA chấp thuận từ năm 2008 đến năm 2017 của Shultz và cộng sự [20]. Nghiên cứu chỉ ra rằng nhiều thuốc được FDA phê duyệt vẫn có phân tử khối lớn hơn 500 Da. Do đó, việc điều chỉnh giới hạn trên của MW là cần thiết nhằm giảm thiểu nguy cơ loại bỏ các cấu trúc có hoạt tính tiềm năng. Đồng thời, nghiên cứu này tuân theo tiêu chí của quy tắc Lipinski bằng cách cho phép vi phạm tối đa 1 trong 4 nguyên tắc, nhằm hạn chế tối đa việc loại bỏ các hợp chất có triển vọng hoạt tính sinh học:

Chỉ số NP Score không cần thiết vì các chất được bán tổng hợp từ cucurbitacin E.

Điều chỉnh yêu cầu MW của Lipinski từ 500 g/mol lên 600 g/mol.

Nhận các chất thỏa hoàn toàn 8/8 yêu cầu của quy tắc, phép đo và bộ lọc.

Tập SDF gồm 1056 hợp chất ở định dạng *.mol được sàng lọc qua phễu lọc hóa dược, kết quả được thể hiện ở Bảng 3.

Kết quả, chỉ có 85 chất qua được phễu lọc hóa dược, trong đó có 88 chất đáp ứng yêu cầu Lipinski, trừ hai chất C221, C222 bị loại bỏ do vi phạm bộ lọc MCE-18 và chất P122 bị loại bỏ do nguy cơ khó tổng hợp khi SA score lớn hơn 6.

Bảng 3. Bảng tổng hợp kết quả đánh giá dựa trên yêu cầu phễu lọc hóa dược

Quy tắc, phép đo và bộ lọc	Vượt qua bộ lọc	Không vượt qua bộ lọc
Lipinski	88	968
Pfizer	1055	1
SA Score	1013	43
Fsp ³	1055	1
MCE-18	1053	3
PAINS	1055	1
BMS	1008	48
Chelator	1055	1

3.2.2. Sàng lọc ảo bằng mô hình gắn kết phân tử

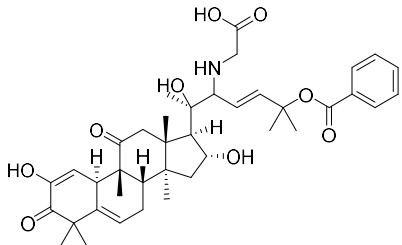
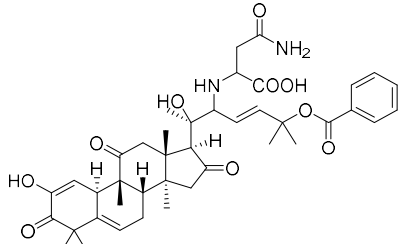
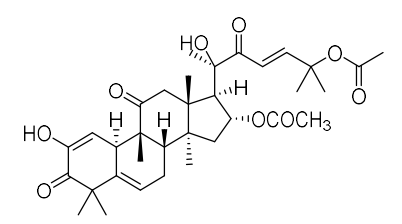
Sau quá trình sàng lọc ảo bằng phễu lọc hóa dược, 85 chất tiềm năng được tiến hành thực hiện quá trình mô phỏng gắn kết phân tử thông qua phần mềm Smina với các thông số tương tự quá trình re-docking thu được kết quả như sau:

Có 17 chất nằm ngoài khoang gắn kết, đồng thời 68 chất còn lại đều nằm gọn trong khoang gắn kết của protein STAT3.

Điểm số gắn kết (docking score) dao động từ -7,5 kcal/mol đến -5,9 kcal/mol, quá trình thu được 66 chất có điểm số docking thỏa mãn ngưỡng hoạt tính -5,989 kcal/mol.

Qua điểm số docking, nhận thấy có ba chất tiềm năng đến từ ba nhóm cấu trúc là G213, N223 và X231 với kết quả điểm số docking và các loại tương tác trong khoang gắn kết được thể hiện qua Bảng 4.

Bảng 4. Cấu trúc của 3 chất G213, N223, X231

STT	Cấu trúc hóa học	ID	Điểm số gắn kết phân tử (kcal/mol)	Liên kết hydro
1		G213	-7,4	Lys658
2		N223	-7,5	Lys658
3		X231	-7,0	Met648, Lys658

Tiến hành so sánh 3 chất tiềm năng này với SD-36 và cucurbitacin E, cả 3 chất có điểm số docking âm nhất của 3 nhóm cấu trúc đều có ái lực với protein STAT3 tốt hơn so với cucurbitacin E (-6,8 kcal/mol). So với ligand đồng kết tinh của protein STAT3 (với ID PDB 6NJS) là SD-36 (-7,1 kcal/mol), hai chất G213 và N223 có điểm số docking âm hơn SD-36 cho thấy chúng có ái lực cao hơn ligand này, trong khi X231 thì cho kết quả điểm số docking hơi cao hơn so với SD-36.

4. BÀN LUẬN

Nghiên cứu đã đánh giá, chọn lọc cẩn thận dựa trên các bước kiểm tra định lượng và sử dụng nhiều phần mềm. Khác với sàng lọc ảo truyền thống chỉ sử dụng một mô hình, không qua bất kì bước thẩm định nào, do đó có thể dẫn đến sai sót trong quá trình sàng lọc. Nghiên cứu chú trọng vào việc lựa chọn công cụ docking và các quy tắc sàng lọc hóa dược. Sự chênh lệch tỷ lệ sàng lọc của sàng lọc hóa dược và mô hình docking xuất phát từ việc 1056 chất trong thư viện sàng lọc

đều là dẫn chất của Cucurbitacin E với khung cấu trúc công kênh, khiến đa số vi phạm giới hạn kích thước của phễu lọc hóa dược. Tuy nhiên, do được thiết kế dựa trên khung sườn vốn đã có hoạt tính kháng STAT3, các dẫn chất vượt qua vòng đầu đều có độ tương thích không gian cao trong mô hình docking. Việc đánh giá, thẩm định mô hình không những lựa chọn được phần mềm phù hợp mà còn giảm thiểu tối đa khả năng bỏ sót các chất tiềm năng.

Nghiên cứu đã đạt được một số kết quả như sau:

Xây dựng phễu lọc hóa dược nhằm sàng lọc các hợp chất mang đặc tính “giống thuốc”.

Xây dựng mô hình docking phân tử, tiến hành re-docking và thử nghiệm hồi chứng (retrospective control) cùng tối ưu hoá ngưỡng lựa chọn.

Sàng lọc ảo các chất từ thư viện nội bộ và chọn ra các chất ưu tiên cho việc tổng hợp.

Trong tương lai, nhóm nghiên cứu dự kiến tiến hành mô phỏng động lực học phân tử (Molecular Dynamics) trên 66

hợp chất tiềm năng nhằm đánh giá khả năng tương tác của các cấu dạng thu được trong điều kiện cấu trúc mục tiêu linh động. Sau khi đánh giá kết quả mô phỏng động lực học phân tử, việc tổng hợp và tiến hành các thử nghiệm *in vitro* sẽ được thực hiện để xác định hoạt tính sinh học, qua đó làm rõ hơn giá trị ứng dụng của các hợp chất này.

5. KẾT LUẬN

Thông qua việc xây dựng mô hình sàng lọc ảo gồm phễu lọc hóa dược và quá trình gắn kết phân tử, nghiên cứu đã sàng lọc thành công một số chất tiềm năng ức chế protein STAT3 từ chất gốc cucurbitacin E. Qua quá trình trên, xác định được 66 chất tiềm năng có hoạt tính ức chế STAT3, đặc biệt là 3 chất tiềm năng nhất là G213, N223 và X231.

Lời cảm ơn

Nhóm nghiên cứu xin trân trọng cảm ơn Đại học Y Dược Thành phố Hồ Chí Minh đã tạo điều kiện thuận lợi để nhóm thực hiện và hoàn thành đề tài nghiên cứu. Nhóm cũng xin ghi nhận sự hỗ trợ về cơ sở vật chất và môi trường nghiên cứu từ Bộ môn Hóa Hữu cơ, Trường Dược Đại học Y Dược Thành phố Hồ Chí Minh trong quá trình triển khai đề tài.

Nguồn tài trợ

Nghiên cứu này từ đề tài cơ sở của Đại học Y Dược TP. Hồ Chí Minh, mã số 217/2025/HĐ-ĐHYD và không nhận bất kỳ nguồn tài trợ nào khác.

Xung đột lợi ích

Không có xung đột lợi ích nào liên quan đến nghiên cứu này.

ORCID

Nguyễn Thị Hương Giang

<https://orcid.org/0000-0001-6007-6312>

Trương Ngọc Tuyền

<https://orcid.org/0000-0002-0952-1633>

Tô Văn Thịnh

<https://orcid.org/0000-0002-7640-0807>

Nguyễn Văn Phước Chung

<https://orcid.org/0009-0000-1825-2781>

Trần Nguyễn Ngọc Tâm

<https://orcid.org/0009-0004-3893-0497>

Đóng góp của các tác giả

Ý tưởng nghiên cứu: Trương Ngọc Tuyền, Nguyễn Thị Hương Giang.

Đề cương và phương pháp nghiên cứu: Tô Văn Thịnh, Trần Nguyễn Ngọc Tâm, Nguyễn Văn Phước Chung

Thu thập dữ liệu: Trần Nguyễn Ngọc Tâm, Nguyễn Văn Phước Chung.

Giám sát nghiên cứu: Trương Ngọc Tuyền, Nguyễn Thị Hương Giang.

Nhập dữ liệu: Tô Văn Thịnh, Trần Nguyễn Ngọc Tâm, Nguyễn Văn Phước Chung

Quản lý dữ liệu: Trương Ngọc Tuyền.

Phân tích dữ liệu: Tô Văn Thịnh.

Viết bản thảo đầu tiên: Tô Văn Thịnh.

Góp ý bản thảo và đồng ý cho đăng bài: Trương Ngọc Tuyền, Nguyễn Thị Hương Giang.

Cung cấp dữ liệu và thông tin nghiên cứu

Tác giả liên hệ sẽ cung cấp dữ liệu nếu có yêu cầu từ Ban biên tập.

Chấp thuận của Hội đồng Đạo đức

Nghiên cứu này miễn trừ Hội đồng Đạo đức.

TÀI LIỆU THAM KHẢO

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2024;74(3):229-63.
2. Blaskovich MA, Sun J, Cantor A, Turkson J, Jove R, Sebt SM. Discovery of JSI-124 (cucurbitacin I), a selective Janus kinase/signal transducer and activator of transcription 3 signaling pathway inhibitor with potent antitumor activity against human and murine cancer cells in mice. Cancer Res. 2003;63(6):1270-9.
3. Cai Y, Fang X, He C, Li P, Xiao F, Wang Y, et al. Cucurbitacins: a systematic review of the phytochemistry and anticancer activity. Am J Chin Med. 2015;43(07):1331-50.

4. Matsuda H, Nakashima S, Abdel-Halim OB, Morikawa T, Yoshikawa M. Cucurbitane-type triterpenes with anti-proliferative effects on U937 cells from an egyptian natural medicine, bryonia cretica: structures of new triterpene glycosides, bryoniaosides A and B. Chemical and Pharmaceutical Bulletin. 2010;58(5):747-51.
5. Shan XL, Zhou XY, Yang J, Wang YL, Deng YH, Zhang MX. Inhibitory effect of cucurbitacin E on the proliferation of ovarian cancer cells and its mechanism. Chin J Cancer. 2010;29(1):20-4.
6. Lan T, Wang L, Xu Q, Liu W, Jin H, Mao W, et al. Growth inhibitory effect of cucurbitacin E on breast cancer cells. Int J Clin Exp Pathol. 2013;6(9):1799-805.
7. Ren X, Duan L, He Q, Zhang Z, Zhou Y, Wu D, et al. Identification of niclosamide as a new small-molecule inhibitor of the STAT3 signaling pathway. ACS Med Chem Lett. 2019;1(9):454-9.
8. Turkson J, Ryan D, Kim JS, Zhang Y, Chen Z, Haura E, et al. Phosphotyrosyl peptides block Stat3-mediated DNA binding activity, gene regulation, and cell transformation. Journal of Biological Chemistry. 2001;276(48):45443-55.
9. Siddiquee K, Zhang S, Guida WC, Blaskovich MA, Greedy B, Lawrence HR, et al. Selective chemical probe inhibitor of Stat3, identified through structure-based virtual screening, induces antitumor activity. Proceedings of the National Academy of Sciences. 2007;104(18):7391-6.
10. Bender BJ, Gahbauer S, Luttens A, Lyu J, Webb CM, Stein RM, et al. A practical guide to large-scale docking. Nat Protoc. 2021;16(10):4799-832.
11. Imrie F, Bradley AR, Deane CM. Generating property-matched decoy molecules using deep learning. Bioinformatics. 2021;37(15):2134-41.
12. Lee DH, Iwanski GB, Thoennissen NH. Cucurbitacin: ancient compound shedding new light on cancer treatment. The Scientific World Journal. 2010;10(1):565972.
13. Dantas INF, Gadelha GCM, Chaves DC, Monte FJQ, Pessoa C, de Moraes MO, et al. Studies on the cytotoxicity of cucurbitacins isolated from *Cayaponia racemosa* (Cucurbitaceae). Z Naturforsch C J Biosci. 2006;61(9-10):643-6.
14. Bartalis J, Halaweish FT. *In vitro* and QSAR studies of cucurbitacins on HepG2 and HSC-T6 liver cell lines. Bioorganic & Medicinal Chemistry. 2011;19(8):2757-66.
15. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF chimera-a visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004;25(13):1605-12.
16. Landrum G. RDKit: open-source cheminformatics software [Internet]. 2016. <https://www.rdkit.org>.
17. Koes DR, Baumgartner MP, Camacho CJ. Lessons Learned in empirical scoring with smina from the csar 2011 benchmarking exercise. J Chem Inf Model. 2013;53(8):1893-904.
18. Alhossary A, Handoko SD, Mu Y, Kwok CK. Fast, accurate, and reliable molecular docking with QuickVina 2. Bioinformatics. 2015;31(13):2214-6.
19. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. J Chem Inf Model. 2021;61(8):3891-8.
20. Shultz MD. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. J Med Chem. 2019;62(4):1701-14.