

ỨNG DỤNG PHẪU LỌC HÓA DƯỢC VÀ MÔ HÌNH DOCKING PHÂN TỬ SÀNG LỌC CHẤT ỨC CHẾ STAT3 ĐỐI VỚI CÁC DẪN CHẤT CUCURBITACIN E

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

SUMMARY

Cancer is one of the leading causes of death worldwide with nearly 10 million deaths recorded in 2020, according to the World Health Organization report. While targeting STAT3 (Signal transducer and activator of transcription 3) has been known as a promising pathway for controlling tumor growth, cucurbitacin is considered potential for anti-cancer therapy by triggering apoptosis via selectively inhibiting the JAK/STAT path. The database of 1056 STAT3 inhibitors, which were based on cucurbitacin E scaffold, was screened through the pharmacological filters, and then experienced molecular docking on STAT3 protein (PDB ID: 6NJS). The Natural Product-likeness score (NPcore) resulted from ADMET filter ranged from 1.31 to 3.15; alongside 85 potential substances were chosen to be screened through molecular docking model. Eventually, 66 substances were proved to have adequate binding affinity toward STAT3.

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

ĐẶT VẤN ĐỀ

Hiện nay, việc ứng dụng khoa học máy tính vào lĩnh vực thiết kế thuốc ngày càng được áp dụng rộng rãi, giúp rút ngắn thời gian, tài nguyên và chi phí nghiên cứu, đồng thời cải thiện tỷ lệ thành công trong lĩnh vực khám phá thuốc mới. 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 ngoài việc đáp ứng các yêu cầu về dược lực, khả năng chọn lọc đích tác động thì các yếu tố dược động học (bao gồm hấp thu, phân bố, chuyển hóa, thải trừ, độc tính – ADMET) cũng cần được quan tâm. Chính vì vậy, các đặc tính “giống thuốc” (druglikeness) cần được sàng lọc trước thông qua bộ lọc ADMET. 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.

Bên cạnh đó, điểm kết soát miễn dịch trong ung thư (I-O), đang trở thành xu hướng nghiên cứu được thế giới quan tâm do tầm quan trọng trong việc tăng cường hệ thống miễn dịch của cơ thể để chống lại tế bào ung thư []. Yếu tố chuyển đổi và hoạt hóa quá trình phiên mã (STAT) - điểm hội tụ của nhiều con đường tín hiệu gây ung thư, được xác định là trung tâm trong việc điều chỉnh phản ứng miễn dịch chống khối u. Trong đó, STAT3 được kích hoạt rộng rãi ở cả tế bào ung thư và không ung thư trong hệ sinh thái khối u, đóng vai trò quan trọng trong việc ức chế sự biểu hiện của các chất điều hòa miễn dịch hay thúc đẩy sản xuất các yếu tố ức chế miễn dịch.⁴ Do đó, nhắm mục tiêu vào con đường tín hiệu STAT3 đã nổi lên như một chiến lược điều trị đầy hứa hẹn cho nhiều loại ung thư. Vai trò đối kháng STAT3 được biết đến đầu tiên là ở cucur I với nghiên cứu của Blaskovich và cộng sự [], qua đó tạo tiền đề cho một loạt nghiên cứu khác, đã chứng minh cucur I là một chất ức chế JAK-STAT mạnh trong nhiều loại ung thư ở người. 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.

ĐỐI TƯỢNG-PHƯƠNG PHÁP NGHIÊN CỨU

Đối tượng nghiên cứu

Tài nguyên tính toán

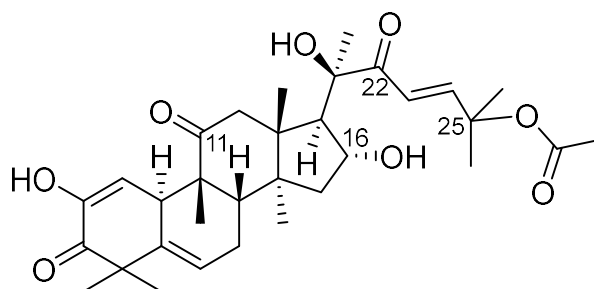
Nghiên cứu *in silico* được tiến hành trên máy tính với hệ điều hành Linux 20.04, CPU AMD Ryzen 9 3900X, 12 nhân, tốc độ xử lý là 3,79 GHz, 500 GB bộ nhớ, 32,0 GB RAM tại Bộ môn Hóa hữu cơ, Khoa Dược, Đại Học Y Dược Thành phố Hồ Chí Minh. Ngôn ngữ lập trình được sử dụng là Python 2.7 và 3.7, được thực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

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). 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 $\leq 0,35$). 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 []. Tập decoy được tạo thành dựa trên phương pháp property-matched và sử dụng mô hình DEKOIS (không chứa phospho) đã được huấn luyện sẵn để tạo thành các decoy với tỷ lệ active:decoy là 1:50.

Dữ liệu sàng lọc

Thư viện nội bộ gồm 1056 chất được thiết kế dựa trên khung cấu trúc của cucurbitacin E bằng cách thay đổi 4 thành phần: nhóm carbonyl ở C₁₁ (độc tính của cucurbitacin E có thể do nhóm carbonyl ở C₁₁); nhóm hydroxyl ở C₁₆ (mức độ phân cực cao hoặc sự hiện diện nhóm ester trên C₁₆ giúp làm giảm độc tính tế bào của cucurbitacin E); nhóm ceton ở C₂₂ và nhóm ester ở C₂₅ (nhóm ceton ở vị trí C₂₂ và gốc acetoxy ở C₂₅ được chứng minh có liên quan đến hiệu quả tác động của cucurbitacin E). 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

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, sử dụng trường lực Amber với tham số conjugate gradient minimization được để mặc định là 0,001 KJ/mol, và xác định thông số cho khoang gắn kết dựa trên vị trí gắn kết của ligand đồng kết tinh (SD-36) trong khoang gắn bằng phần mềm UCSF Chimera.

Phương pháp nghiên cứu

Phễu lọc Hóa dược

Tính toán thông số mô tả (**LogP**: hệ số phân bố dầu nước, **HBA**: số liên kết nhận hydro, **HBD**: số liên kết cho hydro, **MW**: khối lượng phân tử, **TPSA**: tổng diện tích bề mặt, **Fsp3**: số carbon lai hóa sp^3 / tổng số carbon) bằng thư viện hoá tin RDKit 2022.09.1 và Python 3.7.

Áp dụng hai quy tắc (**quy tắc Lipinski**: ước tính khả năng hấp thu đường uống, **quy tắc Pfizer**: mối quan hệ giữa các đặc tính hóa lý và độc tính), bốn phép đo định lượng (**Fsp3**: độ bão hòa carbon, **MCE-18 (Medicinal Chemistry Evolution 2018)**: công thức tỉ lệ nhóm chức, **Sascore (Synthetic accessibility 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 loại bỏ các chất có nguy cơ dương tính giả). Riêng yêu cầu về MW của Lipinski, ngưỡng vi phạm được điều chỉnh từ 500 g/mol lên 600 g/mol để phù hợp với cơ sở dữ liệu của nghiên cứu.

Yêu cầu các quy tắc: **Lipinski**: Không vượt quá 2 trong 4 yếu tố sau: $HBA \geq 10$, $HBD \geq 5$, $MW \geq 600$, $LogP \geq 5$; **Pfizer**: Không đồng thời thỏa cả 2 yếu tố: $LogP \geq 3$ và $TPSA \leq 75$; **Fsp 3**: $Fsp\ 3 \geq 0,42$ là phù hợp; **MCE-18**: Điểm số MCE-18 ≥ 45 , **SAScore**: 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 [].

Mô hình gắn kết phân tử (Molecular Docking)

- Chuẩn bị và protein và ligand:

Protein 6NJS sau khi tách ligand được mở bằng MGLTools (Python 2.7) trên giao diện của Jupyterlab, sử dụng công cụ *prepare_receptor* để 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 tương tự nhau bằng MGLTools (Python 2.7) trên giao diện của Jupyterlab. Sử dụng công cụ *prepare_ligand* để thêm điện tích Gasteiger và được lưu dưới dạng *.pdbqt.

- Kiểm tra hồi chứng:

Nghiên cứu tiến hành thực hiện re-docking để kiểm tra năng lực gắn kết (docking power) và kiểm tra hồi chứng để kiểm tra năng lực sàng lọc (screening power) trên cả ba phần mềm khác nhau bao gồm Smina, Qvina2 và Autodock Vina 1.2.3 (Vina), 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 ($size_x = 25 \text{ \AA}$, $size_y = 25 \text{ \AA}$, $size_z = 25 \text{ \AA}$), 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. Phần mềm được cho là phù hợp cho việc docking phân tử khi RMSD dưới 2 Å (RMSD biểu thị khoảng cách trung bình giữa các nguyên tử tương ứng trước và sau quá trình docking xếp chồng lên nhau), đồng thời các tương tác trọng yếu trong khoang gắn được tái tạo đầy đủ.

- 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 AUC-ROC; bên cạnh đó đại lượng G-mean cũng được sử dụng để xác định điểm gãy của đường cong, từ đó định hướng chọn được ngưỡng điểm docking phù hợp nhất cho việc sàng lọc ảo. Trong đó:
 - o Đường cong ROC minh họa trực quan về sự chính xác và lỗi được quan sát trong mô hình phân loại, với trục Y là độ chính xác (tỷ lệ dương tính thật) và trục X là tỷ lệ dương tính giả. Diện tích dưới đường cong ROC (AUC-ROC) tính toán hiệu năng mô hình theo các ngưỡng phân loại khác nhau, chỉ số AUC càng cao thì mô hình càng chính xác.
 - o G-mean là đại lượng đo lường sự cân bằng giữa hiệu năng phân loại trên cả lớp đa số và thiểu số. G-mean thấp là dấu hiệu của hiệu năng kém trong việc phân loại các chất có hoạt tính ngay cả khi các trường hợp không hoạt tính được phân loại rất chính xác.
- G-mean** = $\sqrt{\text{Độ nhạy} \times \text{Độ đặc hiệu}}$
- So sánh phần mềm: dựa vào diện tích dưới đường cong AUC-ROC của kiểm tra hồi chứng, lựa chọn được phần mềm phù hợp nhất cho quá trình sàng lọc ảo.

Sàng lọc ảo

1056 chất tương đồng cucurbitacin E được sàng lọc lần lượt qua các phễu lọc hoá dược 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.

KẾT QUẢ

Mô hình gắn kết phân tử

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. Kết quả này được đánh giá dựa trên ba yếu tố bao gồm:

- (1) Định hướng các cấu dạng của ligand trong khoang gắn kết, cụ thể là xem xét vị trí của ligand SD-36 sau khi dock có sai lệch với ligand đồng kết tinh. Tiêu chí chứng tỏ phần mềm docking có khả năng sàng lọc trên protein này là $\text{RMSD} \leq 2$ Å
- (2) Ái lực gắn kết thông qua năng lượng liên kết
- (3) Sự phù hợp về mặt hóa học của các tương tác ligand-protein

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. Từ kết quả re-docking được trình bày trong **Bảng 1**, có thể nhận thấy RMSD ở cả ba phần mềm đều xấp xỉ nhau, gần bằng 2Å nên đều có thể ứng dụng sàng lọc ảo. Ái lực gắn kết của cả ba phần mềm cũng khác biệt không đáng kể, điều này có thể giải thích là do cả ba phần mềm docking đều sử dụng cùng loại hàm tính điểm là dựa theo kinh nghiệm.

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
Ái lực gắn kết (kcal/mol)	-7,12	-7,20	-7,14

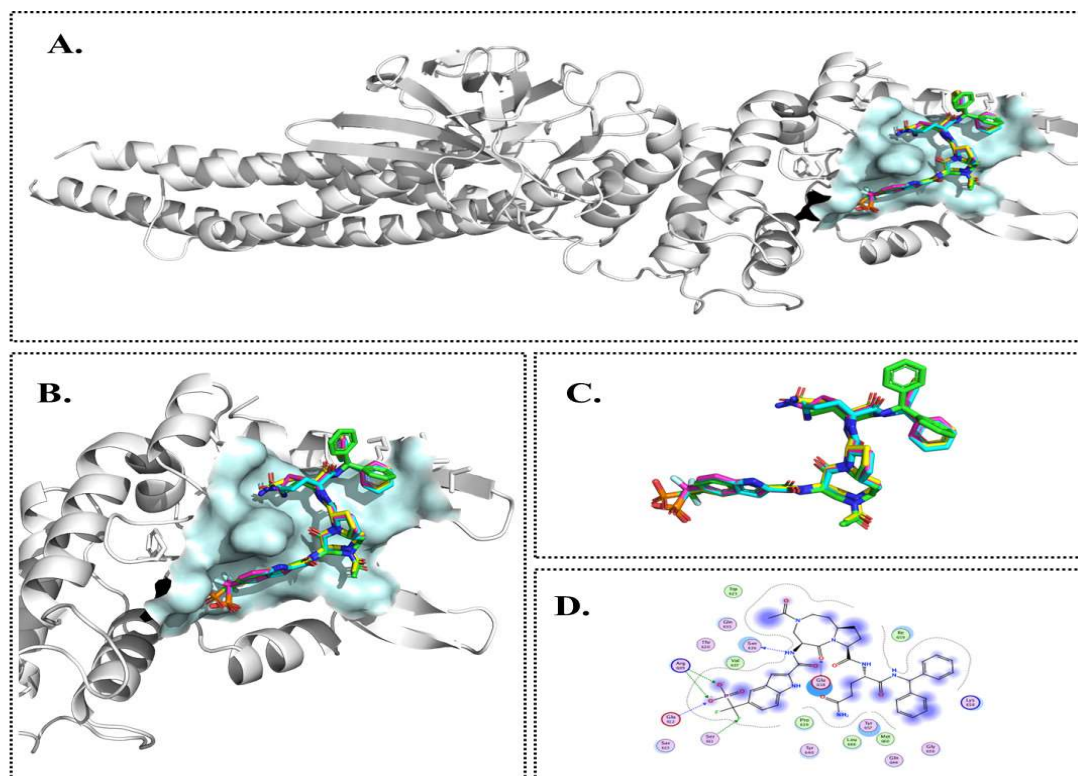
Về mặt tương tác hóa học, có sự tương đồng giữa ligand đồng kết tinh ban đầu và ligand được re-docking trên các acid amin Arg609, Ser611, Glu612, Ser613, Ser636 được tổng hợp cụ thể trong **Bảng 2** dưới đây. Có thể thấy, phần mềm docking đã tái tạo hầu hết các tương tác quan trọng trong khoang SH2 của STAT3.

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	Glu 638: H-acceptor Tyr 657: H-acceptor	Glu 638: H-acceptor	Glu 638: H-donor và H-acceptor
Gln 644: H-acceptor			Gln 644: H-acceptor
Tyr 657: H-acceptor			

(Với H-acceptor là nhận liên kết hydro từ acid amin và H-donor là liên kết hydro cho acid amin, ionic là liên kết ion).

Thông qua quá trình re-docking, có thể thấy cả ba phần mềm Vina, Qvina2 và Smina đều cho kết quả phù hợp để tiến hành gắn kết phân tử.

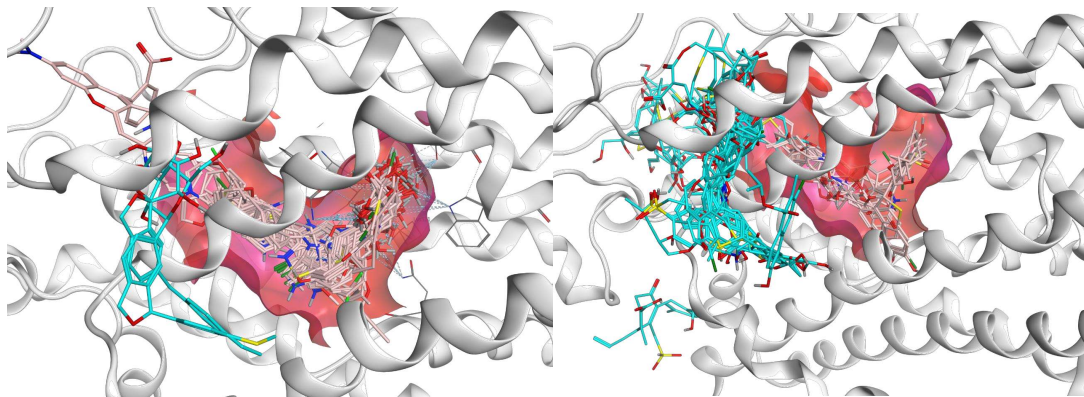


Hình 2. **A.** Ligand đồng kết tinh nằm trong khoang gắn kết của protein STAT3. **B.** Ligand trước và sau khi re-docking đều nằm trong khoang gắn kết của protein STAT3. **C.** Cấu dạng của ligand đồng kết tinh SD-36 trước và sau khi re-docking. **D.** Tương tác giữa ligand trước và sau khi re-docking với các acid amin có trong khoang gắn kết của protein STAT3 của Qvina2.

Kết quả kiểm tra hồi chứng

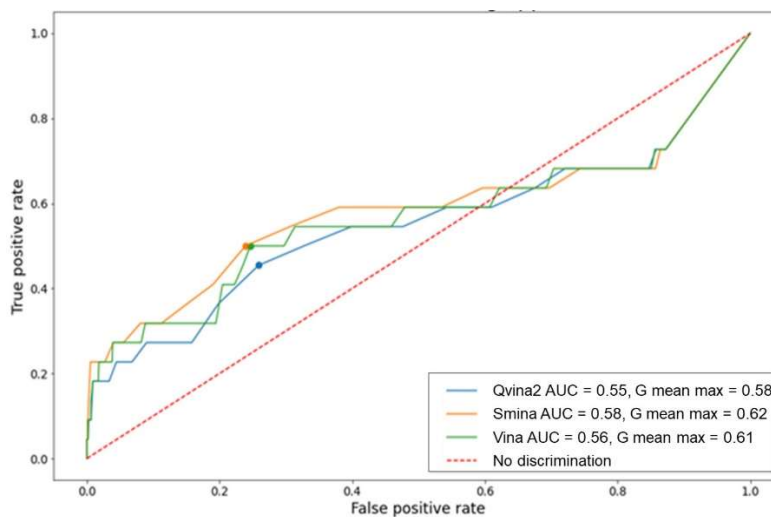
Quá trình tạo decoy bằng thư viện Deepcoy đã đưa ra các thông số đánh giá chất lượng của mỗi nhữ bao gồm điểm số DOE đạt 0,2508; trung bình và giá trị tối đa điểm số Doppleganger lần lượt đạt 0,2659 và 0,3626. Trong đó các chỉ số đều ở mức phù hợp (mức độ tương đồng giữa chất có hoạt tính và chất môi nhữ cao trong không gian hóa lý). Cụ thể, giá trị AUC-ROC tiến về 0,5 cho thấy khả năng phân biệt chất có hoạt tính và decoy rất khó khăn. Giá trị AUC-ROC INN ở nghiên cứu này nằm ở mức chấp nhận được (0,7328).

Cả ba phần mềm cho thấy sự khác biệt về sự phân ái lực gắn kết của hai tập decoy và hoạt tính. Điều này được lý giải do hầu hết các chất có hoạt tính có thể tiếp cận được khoang gắn kết SH2 của STAT3, trong khi tập decoy lại có nhiều cấu trúc nằm ngoài khoang gắn (**Hình 3**). Các hợp chất nằm ngoài khoang gắn kết 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.



Hình 3. Hình ảnh minh họa vị trí tập tập hoạt tính khảo sát trong khoang gắn kết protein STAT3 (trái) và tập decoy (phải) trong khoang gắn kết protein STAT3

Tiến hành chuyển đổi điểm số docking của tất cả các chất có hoạt tính và decoy theo phương pháp MinMaxScaler của thư viện Scikit-learn với khoảng giá trị [0;1], từ đó thiết lập đồ thị ROC-AUC để đánh giá hiệu năng của kiểm tra hồi chứng các mô hình docking.



Hình 4. Đường cong ROC-AUC từ ba phần mềm docking

Từ **Hình 4** có thể nhận thấy đường cong ROC ở cả ba phần mềm khá tương đồng nhau, đều có xu hướng lồi (nằm trên đường giới hạn) khi tăng TPR từ 0 lên 0,6 nhưng lại có xu hướng lõm (nằm dưới đường giới hạn) khi tiếp tục tăng lên 1. Việc đường cong ROC nằm dưới đường giới hạn, cho thấy mô hình phân biệt rất kém giữa chất hoạt tính và môi nhử và không thể ứng dụng được. Ngoài ra, có thể xác định được giá trị AUC từ biểu đồ ROC kiểm tra hồi chứng lần lượt là 0,55; 0,56 và 0,58 cho ba phần mềm Qvina2, Vina và Smina. Điểm gãy của đồ thị ROC-AUC được xác định dựa trên điểm G-mean cao nhất với tỷ lệ dương tính thật (TPR) và dương tính giả (FPR) 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. Nhìn chung, giá trị AUC của cả 3 phần mềm khá thấp, nguyên nhân là do tỷ lệ dương tính giả tại điểm gãy tương đối cao.

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.

Sàng lọc ảo

Phân tích bộ dữ liệu sàng lọc

Thông số mô tả tập dữ liệu sàng lọc gồm 1056 chất được tính toán bằng thư viện RDKit và trình bày ở **Bảng 4**. Các chỉ số khác trong luật Lipinski như LogP, HBA và HBD đều có giá trị trung bình gần với cận trên của phễu lọc. Chỉ số Fsp³ có trung bình là 0,62 cao hơn cận dưới là 0,42 chứng tỏ các cấu trúc có nhiều carbon no, tăng khả năng xoay được trong không gian để tăng khả năng gắn kết với protein. Chỉ số TPSA trung bình quá lớn so với cận dưới 75, dự đoán hầu hết các chất đều khó qua được phễu lọc Pfizer.

Bảng 3. Giá trị trung bình và độ lệch chuẩn của 6 thông số mô tả thường dùng

Đại lượng đánh giá	MW	LogP	HBA	HBD	TPSA	Fsp ³
MEAN ± SD	768,38±80,49	6,00 ±1,40	10,19±0,85	5,07±1,17	188,89±24,0	0,62±0,08

Sàng lọc qua phễu lọc Hoá dược

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 5**.

Bảng 4. Bảng tổng hợp kết quả đánh giá dựa trên yêu cầu phễu lọc hóa dược

Quy tắc, phép đo và bộ lọc	Thỏa	Không thỏa
Lipinski	88	968
Pfizer	88	968
SA Score	1014	42
Fsp ³	1056	0
MCE-18	1056	0
PAINS	1056	0
BMS	1008	48
Chelator	1056	0

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 BMS và chất **P122** bị loại bỏ do nguy cơ khó tổng hợp khi SAScore lớn hơn 6.

Tập cấu trúc tham gia sàng lọc ảo được xây dựng từ cucurbitacin E (chiết xuất từ họ Cucurbitaceae) do đó có giá trị NPscore tương đối cao, trải dài từ 1,31 đến 3,15.

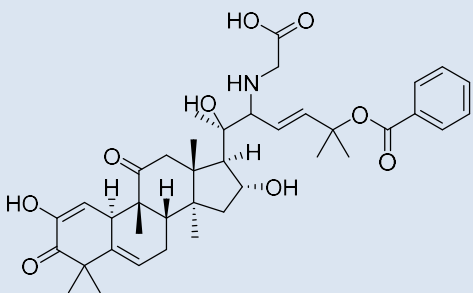
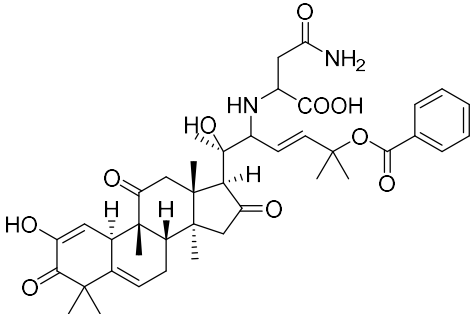
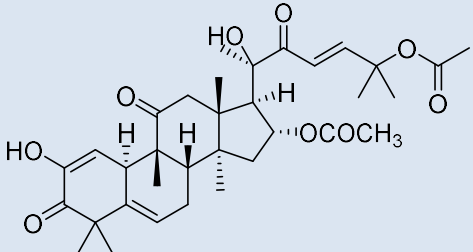
Sàng lọc ảo bằng mô hình gắn kết phân tử

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:

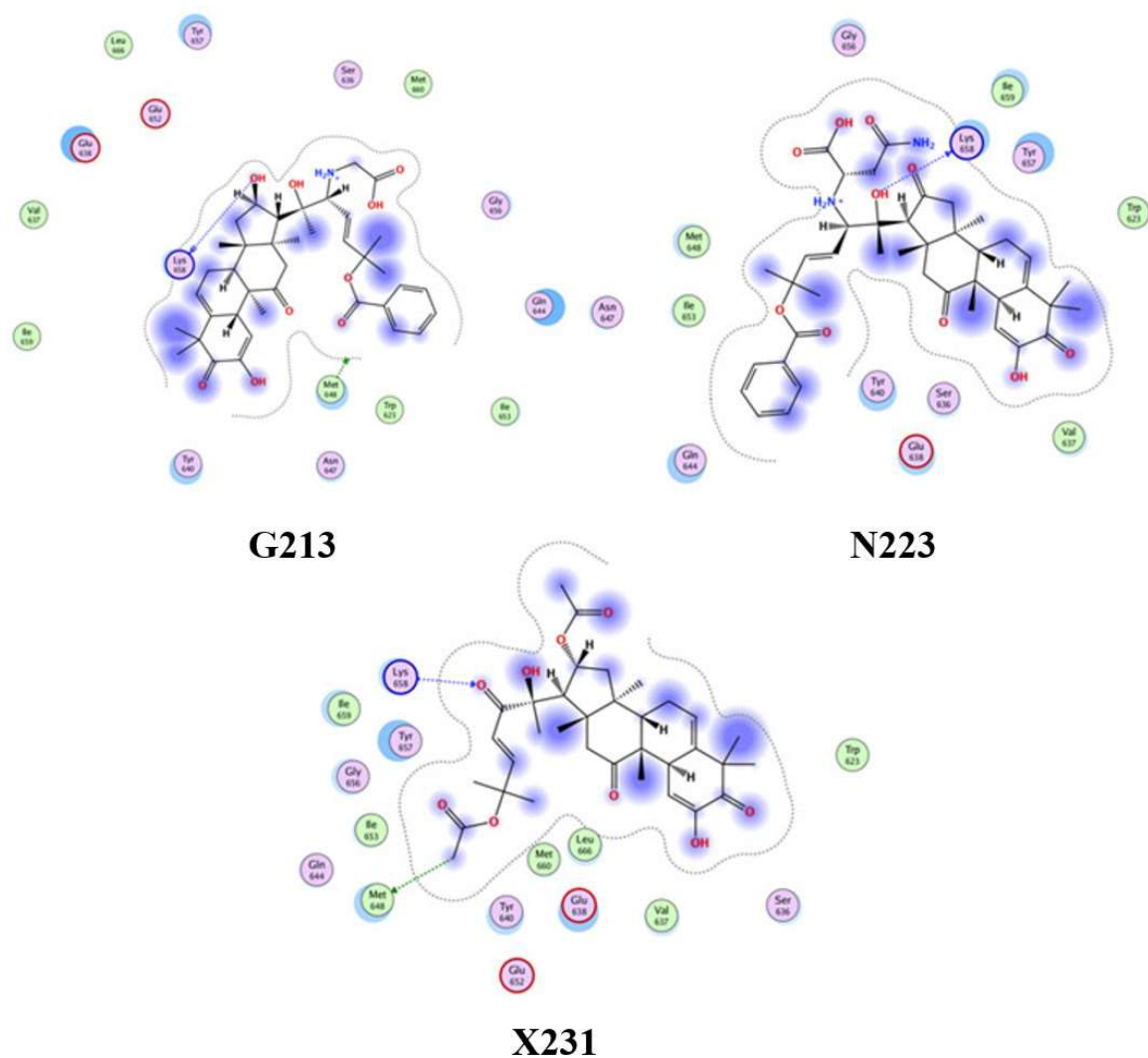
1. Có 17 chất nằm ngoài khoang gắn kết, đồng thời 68 chất còn lại đều nằm gọn trong khoang gắn kết của protein STAT3.
2. Ái lực gắn kết (affinity binding) dao động từ -7,5 kcal/mol đến -5,9 kcal/mol, quá trình thu được 66 chất có điểm số docking thỏa mãn ngưỡng hoạt tính -6,0 kcal/mol.

Kết quả cho thấy, việc thêm nhóm ester vào vị trí C₁₆ với mục đích ban đầu giúp làm giảm độc tính của các hợp chất, nhưng đây lại là nhóm làm cho điểm số docking ít âm nhất, với **X231** có điểm số docking là -7,0 kcal/mol. Do đó, có thể dự đoán, việc thêm nhóm ester vào C₁₆ có thể làm giảm độc tính nhưng đồng thời cũng hạn chế khả năng gắn kết giữa ligand với protein STAT3. Các chất trong nhóm này chủ yếu tương tác với các acid amin như Lys658 – tương đồng hai nhóm trên cùng các acid amin như Tyr640, Gln644, Asn647, Gln644, Met648 và Ile653 thông qua các tương tác kỵ nước. Qua điểm số docking, nhận thấy có ba chất tiềm năng đến từ ba nhóm cấu trúc là **G213**, **N223** và **X231** (**bảng 5**) với kết quả điểm số docking và các loại tương tác trong khoang gắn kết, được thể hiện trong **Bảng 6**.

Bảng 5. Cấu trúc của 3 chất G213, N223, X231

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

Kết quả tương tác của **G213**, **N223** và **X231** được biểu diễn như **Hình 5**.



Hình 5. G213 nằm gọn trong khoang liên kết và có tương tác H-donor với Lys558. N223 nằm gọn trong khoang liên kết và có tương tác H-donor với Lys558. X231 nằm gọn trong khoang liên kết và có tương tác H-donor với Met 648 và H-acceptor với Lys 558

Tiến hành so sánh 3 chất tiềm năng này với SD-36 và cucurbitacin E và tóm tắt ở **Bảng 6**.

Bảng 6. So sánh ba chất tiềm năng nhất với ligand đồng kết tinh SD-36 và cucurbitacin E

Nhóm	Điểm số docking trung bình (kcal/mol)	Tương tác với acid amin
G213	-7,4	Lys558
N223	-7,5	Lys558
X231	-7,0	Lys558, Met648
Cucurbitacin E	- 6,8	Ser636, Gln635
SD-36	- 7,1	Arg609, Ser611, Glu612, Ser613, Ser636, Glu638, Gln644

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. Tuy nhiên, kết quả thu được có sự khác nhau đối với tương tác của

G213, N223, X231 với các acid amin nằm trong khoang hoạt động protein STAT3 so với cucurbitacin E.

Bên cạnh việc so sánh cucurbitacin E – chất dùng để thiết kế thư viện nội bộ, nghiên cứu cũng tiến hành đánh giá các chất tiềm năng nhất ở ba nhóm cấu trúc so với ligand đồng kết tinh của protein STAT3 (với ID PDB 6NJS) là SD-36. 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.

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 kỳ 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 hay các quy tắc sàng lọc hóa dược. 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 đặc tính “giống thuốc” của hợp chất.
- 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.

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**.

TÀI LIỆU THAM KHẢO

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. 2021;71(3):209-249.
2. Debela DT, Muzazu SG, Heraro KD, et al. New approaches and procedures for cancer treatment: Current perspectives. *SAGE open medicine*. 2021;9:20503121211034366.
3. Christofi T, Baritaki S, Falzone L, Libra M, Zaravinos A. Current Perspectives in Cancer Immunotherapy. *Cancers*. 2019;11(10).
4. Zou S, Tong Q, Liu B, Huang W, Tian Y, Fu X. Targeting STAT3 in Cancer Immunotherapy. *Molecular cancer*. 2020;19(1):145.
5. Kaushik U, Aeri V, Mir SR. Cucurbitacins - An insight into medicinal leads from nature. *Pharmacognosy reviews*. 2015;9(17):12-18.
6. Ríos JL, Andújar I, Escandell JM, Giner RM, Recio MC. Cucurbitacins as inducers of cell death and a rich source of potential anticancer compounds. *Current pharmaceutical design*. 2012;18(12):1663-1676.
7. Zhang L, Zhang H, Ai H, et al. Applications of Machine Learning Methods in Drug Toxicity Prediction. *Current topics in medicinal chemistry*. 2018;18(12):987-997.

8. Xie L, Xu L, Kong R, Chang S, Xu X. Improvement of Prediction Performance With Conjoint Molecular Fingerprint in Deep Learning. 2020;11.