

SÀNG LỌC VÀ KHÁM PHÁ DẪN CHẤT PHLORIZIN HƯỚNG ĐẾN ỨC CHẾ SGLT2: KẾT HỢP SÀNG LỌC ẢO VÀ TRÍ TUỆ NHÂN TẠO TẠO SINH

SCREENING AND DISCOVERY OF PHLORIZIN DERIVATIVES AS SGLT2 INHIBITORS: AN INTEGRATION OF VIRTUAL SCREENING AND GENERATIVE ARTIFICIAL INTELLIGENCE

Tô Văn Thịnh¹, Nguyễn Quý Hiền¹,
Nguyễn Thị Hương Giang¹, Huỳnh Hoàng Thúc^{1,*}

DOI: <http://doi.org/10.57001/huih5804.2025.175>

TÓM TẮT

Glucose được tái hấp thu chủ yếu từ dịch lọc cầu thận vào máu thông qua kênh đồng vận glucose-natri 2 (SGLT2) tại ống lượn gần. Cơ chế này giải thích tiềm năng hạ đường huyết của các chất ức chế SGLT2 (SGLT2i), mở ra một hướng tiếp cận trong điều trị đái tháo đường type 2. Các dẫn chất chứa khung phlorizin đã được chứng minh có khả năng ức chế SGLT2. Trong nghiên cứu này, 91 dẫn chất phlorizin được sàng lọc *in silico* bằng phương pháp docking phân tử trên protein SGLT2-MAP17 (7VSI). Kết quả sàng lọc đã xác định được 5 hợp chất tiềm năng dựa trên năng lượng liên kết cao nhất (từ -11,7 đến -11,8kcal.mol⁻¹) và vị trí tương tác giữa ligand với protein. Từ đó, các mô hình được dùng để phân tích tính chất vật lý và đặc tính giống thuốc của 5 hoạt chất này. Cuối cùng, một mô hình khuếch tán tạo sinh đã được sử dụng để tạo thêm 529 hợp chất mới dựa trên khoang gắn kết của protein SGLT2-MAP17, tìm được 63 chất có năng lượng liên kết cao hơn so với 5 chất tiềm năng nhất ban đầu (-11,9 đến -14,6kcal.mol⁻¹). Nghiên cứu đã thiết lập bộ dữ liệu các hợp chất tiềm năng ức chế SGLT2, làm nền tảng định hướng cho quá trình tổng hợp và đánh giá hoạt tính sinh học.

Từ khóa: SGLT2; phlorizin; docking phân tử; AI tạo sinh; sàng lọc và khám phá thuốc.

ABSTRACT

Glucose is reabsorbed primarily from the glomerular filtrate into the bloodstream via the sodium-glucose cotransporter 2 (SGLT2) channel in the proximal tubule. This mechanism explains the hypoglycemic potential of SGLT2 inhibitors (SGLT2i), marking a new therapeutic approach in the treatment of type 2 diabetes. Phlorizin-containing derivatives have been shown to inhibit SGLT2. In this study, 91 phlorizin derivatives were screened *in silico* using molecular docking on the SGLT2-MAP17 protein (7VSI). The screening identified 5 potential compounds based on the highest affinity (-11.7 to -11.8kcal.mol⁻¹) and binding-site of ligand-protein interactions. Several models were sequentially employed to analyze physicochemical and drug-like properties of these 5 active substances. Finally, a diffusion-based generative model was implemented to generate 529 novel compounds optimized for the SGLT2-MAP17's binding site. Of these, 63 exhibited predicted binding affinities that surpassed those of the five reference ligands (ranging from -11.9 to -14.6kcal mol⁻¹). The study thus delivers a curated dataset of prospective SGLT2 inhibitors that will navigate forthcoming synthetic efforts and biological activity assessments

Keywords: SGLT2; phlorizin; molecule docking; generative AI; screening and discovery drug.

¹Khoa Dược, Trường Đại học Y dược TP. HCM

*Email: hhthuc@ump.edu.vn

Ngày nhận bài: 23/02/2025

Ngày nhận bài sửa sau phản biện: 19/5/2025

Ngày chấp nhận đăng: 28/5/2025

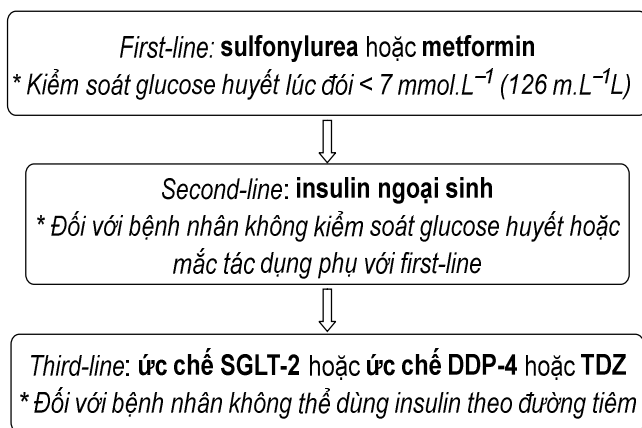
1. GIỚI THIỆU

Bệnh lý đái tháo đường (ĐTĐ) là một rối loạn mãn tính được biểu hiện bằng tình trạng tăng lượng glucose trong

máu do tuyến tụy sản xuất không đủ insulin – đái tháo đường type 1 (ĐTĐ-1) hoặc cơ thể không có khả năng sử dụng hiệu quả insulin được sản xuất – đái tháo đường

type 2 (ĐTĐ-2), với insulin là một loại hormone điều chỉnh lượng đường trong máu [1]. Khi tình trạng bệnh tiến triển hoặc không nhận được điều trị đúng cách, lượng đường trong máu cao có thể dẫn đến tổn thương một số hệ thống cơ thể, từ đó làm tăng nguy cơ mù lòa, suy thận, đau tim, đột quỵ và cắt cụt chi dưới [2]. Theo IDF (Liên đoàn Đái tháo đường thế giới), số bệnh nhân đái tháo đường (độ tuổi từ 20 - 79) được chuẩn đoán năm 2024 là 589 triệu và sẽ tăng trên 850 triệu đến năm 2050, trong đó tỷ lệ tăng cao nhất được ghi nhận lần lượt tại Châu Phi, Trung Đông - Bắc Phi và Đông Nam Á với tỷ lệ tương ứng 142%, 92% và 73% [3]. Trong số các dạng của bệnh lý ĐTĐ, ĐTĐ-2 chiếm hơn 90%, hiện đang là gánh nặng bệnh tật thứ tám trên toàn cầu và được dự đoán tăng lên thứ hai năm 2050 [4, 5].

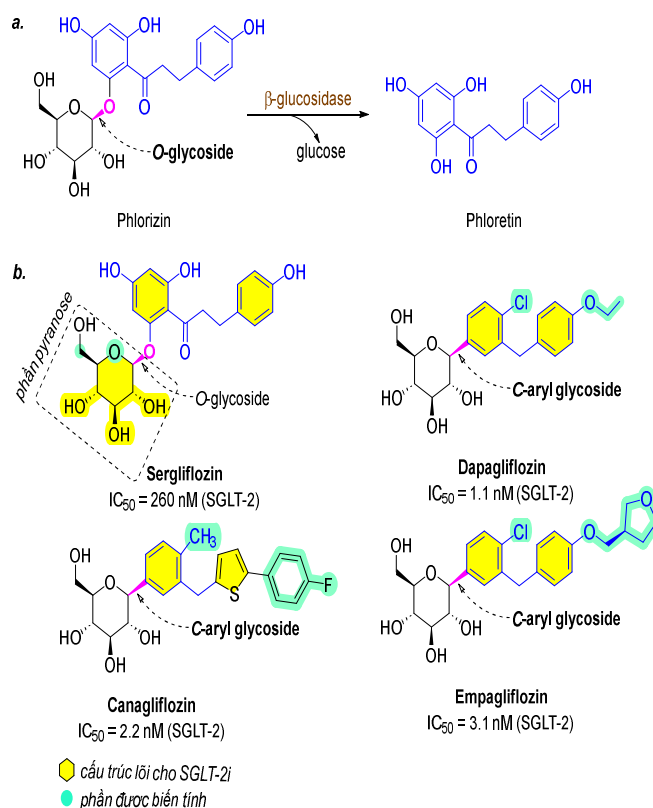
Hiện nay, nguyên nhân tiên phát của bệnh ĐTĐ-2 vẫn chưa được làm rõ nhưng một số yếu tố nguy cơ làm gia tăng tỷ lệ mắc bệnh đã được ghi nhận như béo phì, lười vận động, tuổi tác, các bệnh lý tim mạch và chủng tộc (Nam Á, Phi-Carribe và Tây Ban Nha) [6]. Trong đó, bệnh nhân béo phì, tình trạng kháng insulin tăng nhanh dẫn đến quá trình bất dung nạp glucose, hệ quả là nồng độ glucose huyết tăng và gây ra rối loạn chức năng của tế bào β tụy [7]. Phác đồ điều trị do WHO khuyến cáo năm 2018 dựa trên sự kiểm soát nồng độ glucose huyết [8].



Hình 1. Phác đồ điều trị đái tháo đường type 2 theo WHO PEN 2018

Tuy nhiên, trong khuyến cáo của ADA (Hiệp hội Đái tháo đường Mỹ) năm 2024, đối với các bệnh nhân có bệnh đồng mắc như tim mạch, suy tim hay bệnh thận mãn tính, các thuốc nhóm SGLT-2i được sử dụng như thuốc đầu tay (*first-line*) giúp kiểm soát glucose huyết tốt và giảm các biến chứng với nhóm bệnh nhân này [9]. Bên cạnh đó, việc sử dụng SGLT-2i theo đường uống dưới chỉ định một đơn vị liều/ngày vừa có sinh khả dụng đường uống tốt vừa thuận lợi cho các bệnh nhân không tự sử dụng insulin bằng đường tiêm [10].

SGLT-2 là một protein đồng vận chuyển natri-glucose tại đoạn S1 của ống lượn gần trên nephron cầu thận, có chức năng tái hấp thu khoảng 90% glucose trở lại vào máu [11]. Phlorizin được biết đến là hoạt chất đầu tiên có khả năng ức chế SGLT-2 (SGLT-2i), một glycosid của dihydrochalcon được phân lập từ vỏ cây táo [12]. Tuy nhiên, sự kém chọn lọc lên thụ thể SGLT-2 ($IC_{50} = 35,6nM$) so với thụ thể SGLT-1 ($IC_{50} = 330nM$) và sự thủy phân cấu nối *O*-glycosid bởi β -glucosidase tại ruột non tạo thành hợp chất phloretin (một hợp chất ức chế hấp thu glucose chủ động của tế bào, kể cả tế bào não) được minh họa ở hình 2^a đã làm cho các dẫn xuất nhóm *O*-glycosid phlorizin sớm bị thay thế [13]. Vì vậy, các dẫn xuất C-aryl glycoside hình 2b cho thấy khả năng chọn lọc trên SGLT-2 và bền vững trong quá trình chuyển hóa khi sử dụng theo đường uống [14-16].



Hình 2. a) phản ứng thủy phân phlorizin; b) Các cấu trúc tương đồng phlorizin

Dapagliflozin (Dapagliflozin Viartis®, viên uống 5/10mg, 2023 [17]) là hoạt chất đầu tiên được phát hiện như là một thuốc điều trị ĐTĐ theo cơ chế SGLT-2i, với sự thay thế nhóm glycosyl bằng nhóm C-aryl và rút ngắn khoảng cách giữa hai vòng benzen làm tăng tính chọn lọc lên thụ thể SGLT2 thể hiện ở giá trị IC_{50} chỉ 1,1nM [18]. Tương tự, canagliflozin (Invokana®, viên uống 100/300 mg, 2013 [18]) với thay thế vòng benzen bằng dị vòng

thiophen cũng cho thấy khả năng ức chế SGLT-2, tuy nhiên, nó có khả năng giảm đáng kể HbA1c (1,77 - 1,78%) trong máu khi dùng phối hợp với metformin [19, 20]. Cuối cùng, empagliflozin (Jardiance®, viên uống 10/25 mg, 2014 [21]) cũng được điều chỉnh cấu trúc so với dapagliflozin khi thay thế nhóm ethyl trong nhóm ether bằng một nhánh tetrahydrofuranomethyl. Mặc dù, empagliflozin có giá trị IC₅₀ thấp hơn pagliflozin nhưng hoạt chất này dễ dàng trong phối hợp điều trị với metformin, sulfonilurea và làm giảm đáng kể mức độ HbA1c trong máu (0,77 – 0,78%) [21-23].

Hiện nay, các phương pháp sàng lọc và khám phá thuốc dựa trên máy tính và trí tuệ nhân tạo đã giúp tiết kiệm được thời gian, chi phí nghiên cứu và sự phát thải ô nhiễm. Docking phân tử là một bước quan trọng trong sàng lọc ảo các chất dựa trên nguyên lý ổ khóa - chìa khóa. Các ligand (phân tử thuốc) sẽ được mô phỏng gắn kết vào một đích tác động (protein) và sự tương tác này được thể hiện thông qua vị trí gắn kết (liên kết với amino acid và thông số khoảng gắn kết) cũng như thể hiện qua ái lực gắn kết. Bên cạnh phương pháp docking phân tử truyền thống, gần đây các mô hình khuếch tán tạo sinh (generative diffusion models) đang nổi lên như một hướng tiếp cận mới cho thiết kế phân tử thuốc trong không gian 3D [24]. Tiêu biểu có thể kể đến TargetDiff và Pocket2Mol nổi bật như hai mô hình khuếch tán SE(3)-equivariant có khả năng đồng thời sinh tọa độ 3D và kiểu nguyên tử của phối tử trong khoang gắn kết của protein đích [25, 26]. Cơ chế khử nhiễu phi tuần tự, bảo toàn tính đối xứng hình học giúp các mô hình này tạo ra các hợp chất mới với ái lực gắn kết vượt trội. Do đó, các mô hình tạo sinh khuếch tán cho phép mở rộng nhanh chóng thư viện chất ức chế SGLT-2, định hướng tổng hợp và rút ngắn đáng kể quá trình khám phá thuốc điều trị đái tháo đường type 2.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng nghiên cứu

Cơ sở dữ liệu dùng trong nghiên cứu

Tập hợp gồm 91 ligand đã được tổng hợp trong các nghiên cứu [27-30], các hoạt chất này có khung cấu trúc tương đồng phlorizin nên có tiềm năng SGLT-2i. Thông tin mỗi chất được biểu diễn thông qua cấu trúc 2D (đầy đủ thông tin lập thể) và danh pháp IUPAC.

Phương tiện trong nghiên cứu

Đề tài được thực hiện tại Bộ môn Hóa hữu cơ, Khoa Dược, Đại học Y dược Thành phố Hồ Chí Minh, sử dụng trang thiết bị và phần mềm như liệt kê ở bảng 1.

Bảng 1. Thông số của các thiết bị và phần mềm nghiên cứu

Trang thiết bị/Phần mềm	Cấu hình, đặc tính thiết bị
Máy tính cá nhân	DELL XPS 9320 (12th Gen Intel(R), Core(TM) i7-1260P (16 CPUs), 2.1GHz)
Máy tính PC	Bộ vi xử lý AMD Ryzen Threadripper 7970X (32 nhân, 64 luồng, xung nhịp tối đa 5.3GHz, bộ nhớ đệm 128MB), bộ nhớ RAM 128GB, 2 card đồ họa NVIDIA RTX A6000 48GB GDDR6.
Phần mềm ChemOffice 22.0, Discovery Studio Visualizer 2025, MGLtools 1.5.7, PyMOL (bản dùng thử 30 ngày), Dwimperl-5.14.2.1	
Mô hình tạo sinh Pocket2Mol	https://github.com/pengxingang/Pocket2Mol

2.2. Phương pháp nghiên cứu

Chuẩn bị protein và ligand

Protein SGLT2-MAP17 được tải từ RCSB (<https://www.rcsb.org/structure/7VSI>) có mã là 7VSI (độ phân giải 2.95Å), tiến hành tách ligand đồng kết tinh (empagliflozin) và protein, các cấu trúc sau khi tách được lưu ở dạng *.pdbqt.

Các ligand được trích từ các công bố, được tái lập cấu trúc 3D và lấy mã SMILES từ cửa sổ Chem-3D. Sau đó, sử dụng google collab để chuyển mã SMILES thành tệp *.pdbqt, lưu tại đường dẫn: [Multiple-ligands-docking](#)

Xác định thông số ô mạng và thực hiện docking

Thông số ô mạng (grid box) được xác định thông qua xác định vùng liên kết của ligand empagliflozin với protein SGLT2-MAP17 bằng phần mềm Autodock Tools của MGLtools 1.5.6. Kích thước ô mạng là 20×30×26Å³ với tọa độ x = 39,418; y = 51,846; z = 45,489. Để chuẩn hóa quá trình docking, các cấu dạng của ligand đồng kết tinh sau tách được re-dock vào vùng hoạt động ban đầu của protein để tính toán giá trị độ lệch bình phương trung bình gốc (RMSD) bằng PyMOL. Giá trị RMSD nhỏ hơn 2Å là đáng tin cậy.

Thực hiện docking phân tử 91 ligand đã chuẩn bị lên protein có mã 7VSI với phần mềm Autodock Tools [31].

Xác định ra 05 chất có ái lực gắn kết cao nhất. Từ đó, sử dụng phần mềm PyMOL để biểu diễn khoan liên kết của 5 chất này và empagliflozin [32], sau đó dùng phần mềm Discovery studio tìm vị trí gắn kết và loại gắn kết [33] của 05 chất này với protein đích.

Xác định thông số hóa lý và đặc điểm giống thuốc của 05 chất

Khảo sát một số đặc trưng hóa lý theo quy tắc Lipinski-5 tại website:

<https://scfbio-iitd.res.in/software/drugdesign/lipinski.jsp>

Xác định các đặc điểm giống thuốc tại website: <http://admetlab3.scbdd.com>

Sử dụng trí tuệ nhân tạo tạo sinh bộ chất tương đồng với 05 chất có ái lực cao nhất

Để tài sử dụng mô hình khuếch tán tạo sinh Pocket2Mol [26] để tạo các chất tiềm năng ức chế SGLT-2 dựa trên protein đích với mã 7VSI.

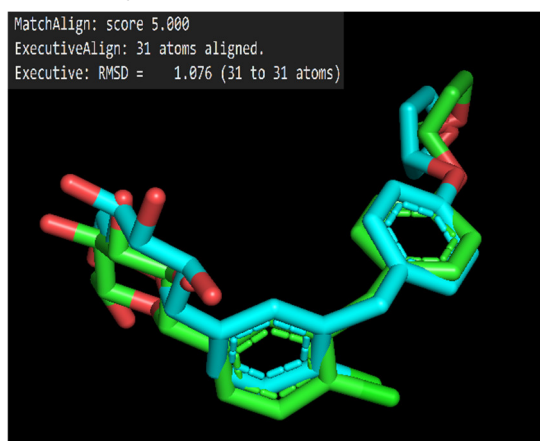
Trên cơ sở của 5 ligand có năng lượng liên kết cao nhất, chúng tôi dùng bộ khuếch tán tạo sinh Pocket2Mo để thiết lập các cấu trúc tương đồng có chứa khung phlorizin nhằm ức chế SGLT-2.

3. KẾT QUẢ VÀ BÀN LUẬN

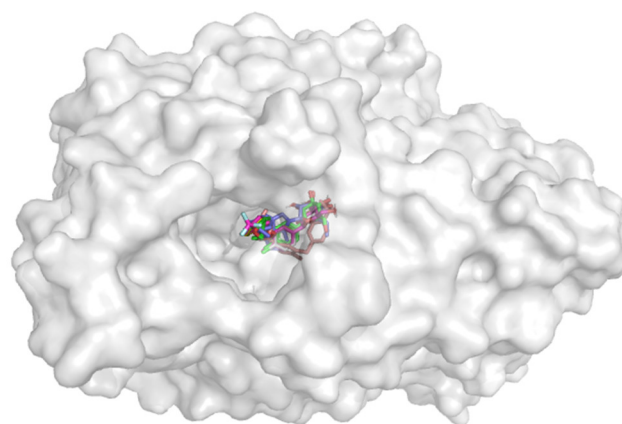
3.1. Kết quả docking phân tử

Năng lượng tương tác giữa ligand đồng kết tinh - empagliflozin - vào protein 7VSI là $-11,5 \text{ kcal.mol}^{-1}$, giá trị RMSD = $1,076 \text{ \AA}$ (nhỏ hơn $2 \text{ \AA}$) của quá trình re-docking được biểu diễn ở **hình 3** cho thấy mô hình docking này đáng tin cậy.

Kết quả docking 91 ligand vào protein thu được 11 chất có năng lượng liên kết cao hơn so với empagliflozin (từ $11,6$ đến $11,8 \text{ kcal.mol}^{-1}$). Trong nghiên cứu này, chúng tôi đã chọn 5 ligand Lig68, Lig75 ($\Delta G = -11,8 \text{ kcal.mol}^{-1}$) và Lig26, Lig40, Lig65 ($\Delta G = -11,7 \text{ kcal.mol}^{-1}$) để so sánh vị trí khoang gắn kết của chúng với empagliflozin. Các kết quả mô phỏng bằng phần mềm PyMOL cho thấy các ligand cùng vị trí khoang gắn kết với ligand đồng kết tinh trong protein 7VSI và trong phạm vi liên kết với các amino acid của khoang như **hình 4**.



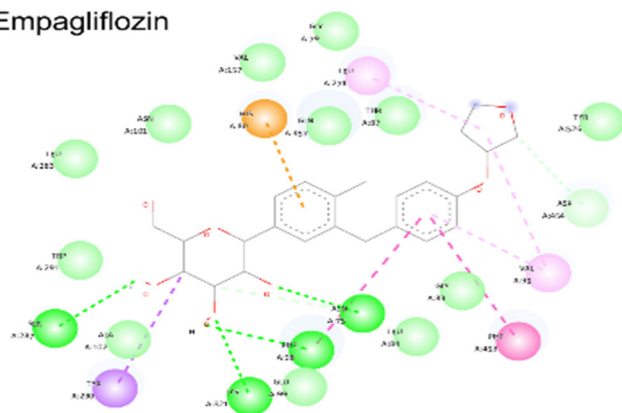
Hình 3. Kết quả re-docking ligand đồng kết tinh với protein



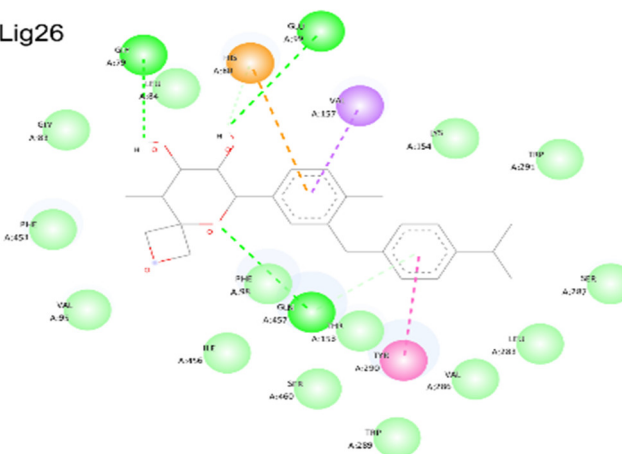
Hình 4. Vị trí khoang liên kết của 5 chất và empagliflozin trên protein

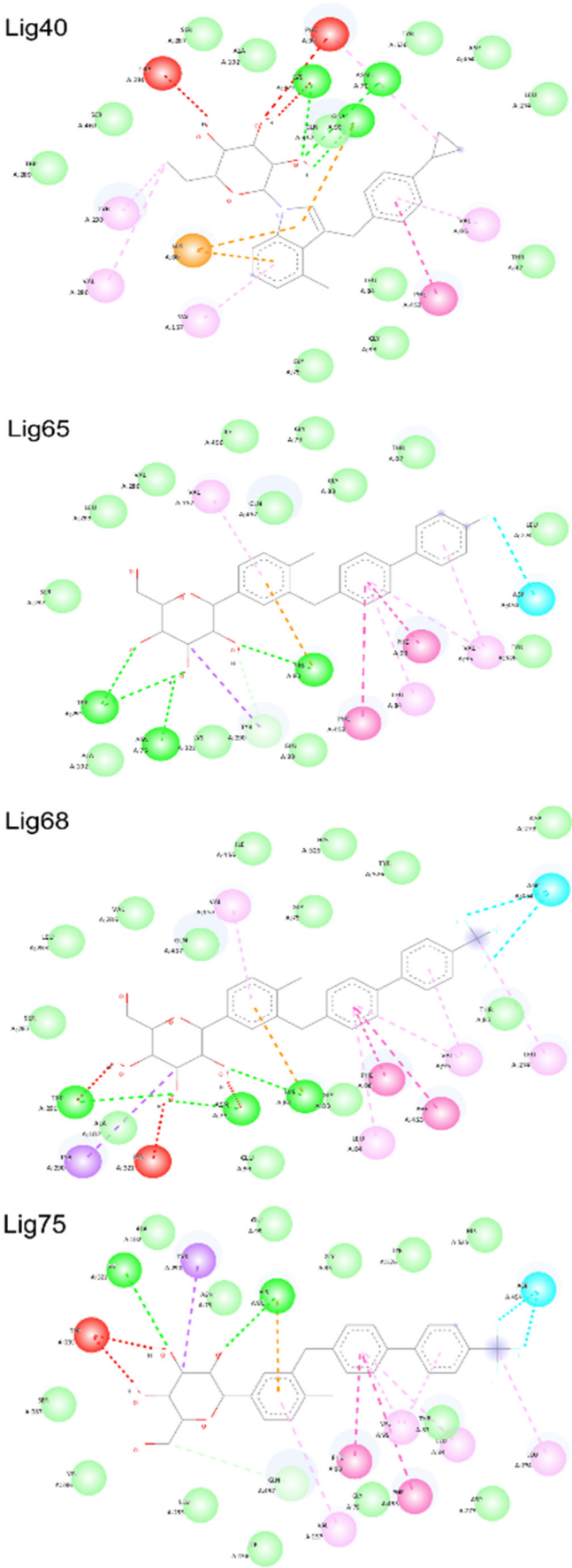
Để làm rõ hơn về mối liên quan giữa cấu trúc và hoạt tính sinh học ức chế SGLT-2, 5 ligand có điểm số docking cao và có những đặc thù về mặt cấu trúc được lựa chọn để tiếp tục phân tích. Phần mềm Discovery studio được sử dụng để biểu diễn tương tác hai chiều giữa 5 ligand trong khoang gắn kết, loại liên kết và vị trí liên kết với các loại amino acid được biểu diễn ở **hình 5**. Năng lượng liên kết và một số loại tương tác chính được trình bày trong **bảng 2**.

Empagliflozin



Lig26





Interactions

van der Waals

Conventional Hydrogen Bond

Carbon Hydrogen Bond

Halogen (Fluorine)

Unfavorable Donor-Donor

Pi-Cation

Pi-Sigma

Pi-Pi Stacked

Pi-Pi T-shaped

Alkyl

Pi-Alkyl

Hình 5. Mô hình tương tác hai chiều của các ligand và emplitflozin với protein 7VSI

Bảng 2. Các loại tương tác giữa 5 ligand và empagliflozin với protein 7VSI

Ligand	Năng lượng liên kết (kcal.mol ⁻¹)	Các loại tương tác chính			
		Liên kết hydroge n	π-σ chồng π-π	π-alkyl	π-cation
Lig26	-11,7	Gly79, Glu99, Phe98	Val157	Tyr290	His80
Lig40	-11,7	Asn75, Glu99, Lys321, Gln457		Phe453	Val95, Phe98, Val157, Val286, Tyr290
Lig65	-11,7	Asn75, His80, Trp291	Tryp290	Phe98, Phe453	Leu84, Val95, Val157
Lig68	-11,8	Asn75, His80, Trp291	Tryp290	Phe98, Phe453	Leu84, Val95, Val157, Leu274
Lig75	-11,8	His80, Lys321	Tryp290	Phe98, Phe453	Leu84, Val95, Val157, Leu274
Empagliflozin	-11,5	Asn75, Phe98, Ser287, Lys321	Tryp290	Phe98, Phe453	Val95, Leu274

Các tương tác giữa 05 ligand với 7VSI đều tương tự empagliflozin. Tuy nhiên, sự khác biệt rõ nhất đến từ tương tác van der Waals (một dạng tương tác kỵ nước), hình 5 cho thấy khi số lượng tương tác van der Waals tăng lên thông qua sự tăng đa phân nhánh (Lig26), thay đổi đơn vòng thơm bằng một vòng ngưng tụ (Lig40) và tăng lên 3 vòng thơm (Lig65, Lig68 và Lig75). Tất cả các yếu tố này đã làm tăng năng lượng liên kết giữa ligand và protein. Trong số 5 ligand này, có 4 ligand thuộc nhóm C-

aryl glycosid (Lig26, Lig65, Lig68, Lig75), trong khi đó Lig40 là một *N*-glycosid. Kết quả này cũng đồng thuận với nghiên cứu của C.H. Yao và cộng sự, khi Lig40 bền vững hơn trong chuyển hóa (so với *O*-glycosid), chọn lọc trên SGLT-2 với $EC_{50} = 161\text{nM}$ và thử nghiệm *in vivo* cho thấy sinh khả dụng đường uống tốt và nó chủ yếu được thải qua nước tiểu [34]. Bên cạnh đó, các Lig65, Lig68 và Lig75 có xuất hiện thêm tương tác halogen của fluor với Asp454. Với sự xuất hiện của nhóm trifluoromethyl (Lig68 và Lig75) giúp hình thành thêm 2 tương tác halogen và alkyl, điều này làm tăng đáng kể năng lượng liên kết so với các ligand còn lại.

3.2. Khảo sát tính chất hóa lý và đặc tính giống thuốc

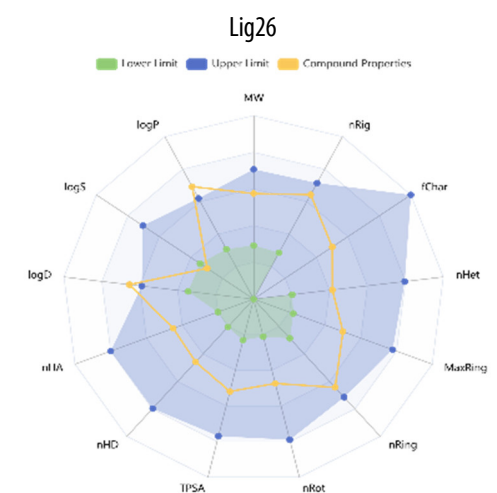
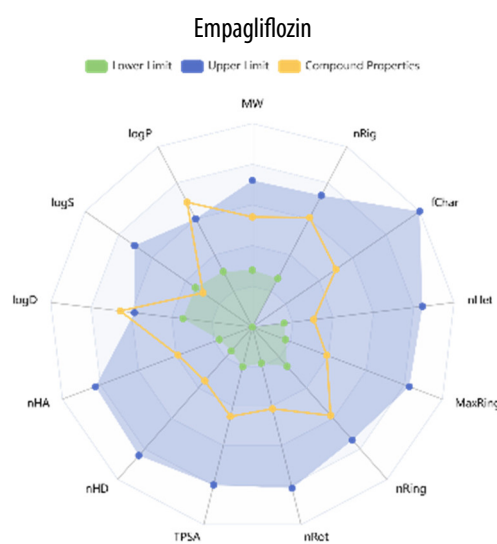
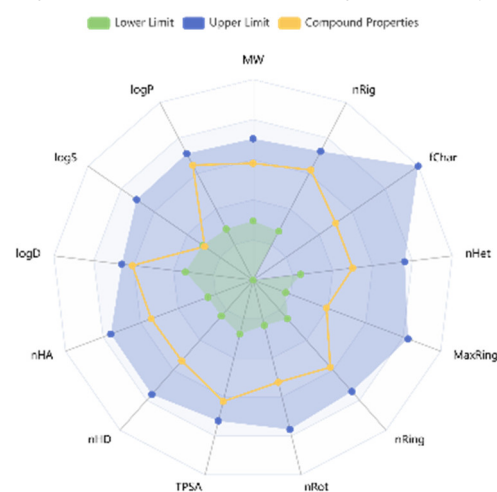
Theo quy tắc Lipinski-5 được đề xuất năm 2001, các chất có độ hấp thu kém khi khối lượng phân tử (Molecular weight - MW) lớn hơn 500Da, hệ số logP lớn hơn 5, có số liên kết hydrogen kiểu cho lớn hơn 5 và số liên kết hydrogen kiểu nhận lớn hơn 10 [35]. Bảng 3 mô tả các thông số tính chất hóa lý của 5 ligand và empagliflozin dựa trên quy tắc Lipinski-5.

Bảng 3. Các thông số tính chất hóa lý của các ligand và empagliflozin theo Lipinski-5

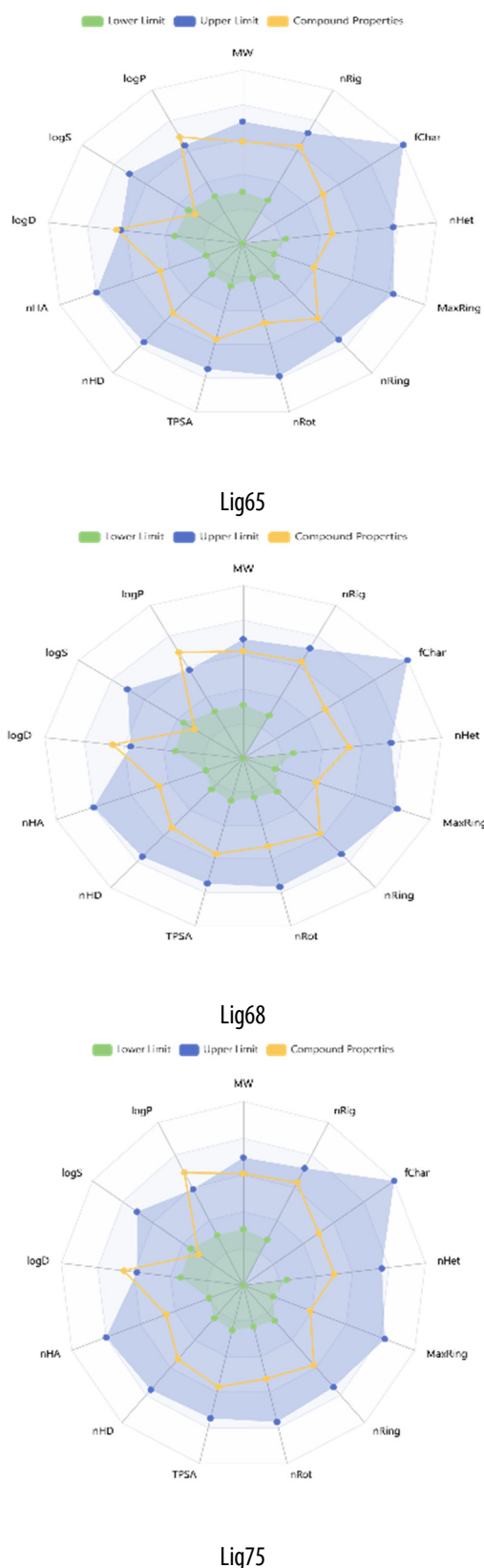
Ligand	Khối lượng phân tử (Da)	Liên kết H kiểu cho	Liên kết H kiểu nhận	LogP
Lig26	396.00	2	4	3.91
Lig40	441.50	3	4	4.15
Lig65	458.50	4	5	3.25
Lig68	508.50	4	5	4.13
Lig75	488.00	4	5	3.79
Empagliflozin	423.50	4	7	0.84

Số liệu cho thấy các ligand đều có các thông số tuân thủ theo quy tắc Lipinski-5, ngoại trừ trường hợp của Lig68 có $MW > 500\text{Da}$. Các Lig65, Lig68 và Lig75 đều có số liên kết hydrogen kiểu cho bằng và liên kết hydrogen kiểu nhận của 5 ligand ít hơn so với empagliflozin, điều này phù hợp với kết quả docking của các ligand với protein. Tương tự, cả 5 ligand đều có giá trị logP cao hơn so với ligand đồng kết tinh. Điều này chỉ ra rằng tính thân dầu của chúng là rất cao và có thể dẫn đến khó khăn khi bào chế thành dược chất, chúng có thể làm gia tăng độc tính của thuốc vì chúng được giữ lâu trong lớp lipid kép của màng tế bào và phân bố rộng khắp trong cơ thể [36]. Mô hình động học - độc tính của 5 ligand nhằm làm rõ khả năng giống thuốc (drug-likeness) được xác định bằng công cụ web của ADMET Lab3.0 [37]. Hình 6 biểu diễn 13 tính chất hóa lý của 5 hợp chất là đường màu vàng, tương

ứng với ngưỡng trên (upper limits) là vùng xanh dương và ngưỡng dưới (lower limits) là vùng xanh lá cây.



Lig40



Hình 6. Radar động học - độc tính của empagliflozin và 5 chất có năng lượng liên kết cao nhất

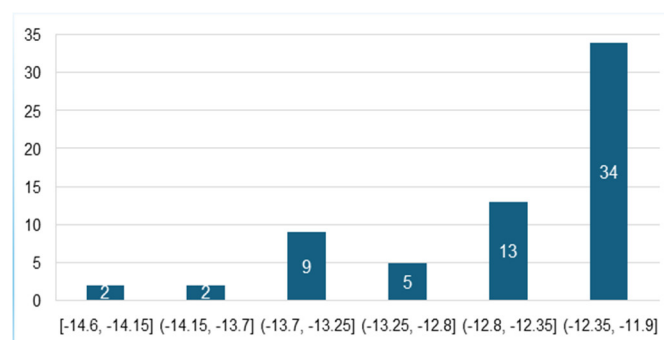
Radar của các ligand cho thấy tính chất của chúng đều thuộc vùng tối ưu giống thuốc và tương tự empagliflozin.

Tuy nhiên, các ligand đều có độ tan, tổng diện tích bề mặt phân cực (logS và TPSA) thấp hơn và độ thân dầu (logP) cao hơn so với empagliflozin. Điều này có thể làm giảm khả năng hấp thu và phân bố của chúng so với empagliflozin. Trong số 5 ligand, Lig68 và Lig75 đều cho thấy độ thân dầu cao và độ hòa tan trong nước cao hơn so với các chất còn lại, điều này có thể làm tăng sinh khả dụng theo đường uống khi thỏa mãn được cả tính thẩm và tính tan trong nước. Trong quá trình chuyển hóa, enzyme CYP450 là một dãy các enzyme chịu trách nhiệm chuyển hóa xenobiotic [39]. Theo kết quả thu được từ ADMET, empagliflozin và Lig75 cho thấy tương tác yếu với CYP, các chất còn lại đều tương tác mạnh với CYP nên dễ bị chuyển hóa. Bên cạnh đó, với cùng năng lượng liên kết ($-11,7\text{kcal.mol}^{-1}$), Lig26 có các số liên kết quay (nRot) khá thấp, sẽ giúp cấu dạng của nó ổn định khi chuyển hóa trong cơ thể hơn [38].

3.3. Bộ dữ liệu tạo sinh từ trí tuệ nhân tạo

Bộ dữ liệu thu được gồm 529 chất với mã SMILES, được đặt tên chung id_#.txt. Sau đó, Google Collab được dùng để chuyển SMILES thành file *.pdbqt để tiếp tục docking phân tử và lưu tại đường dẫn: [Alig docking](#)

Tất cả 529 ligand được thực hiện docking lên protein 7VSI với thông số mạng không thay đổi. Kết quả cho thấy có 63 ligand tạo sinh có năng lượng liên kết cao hơn so với 5 ligand ban đầu, năng lượng của chúng trong khoảng từ $-11,9$ đến $-14,6\text{kcal.mol}^{-1}$ được phân bố như hình 7.

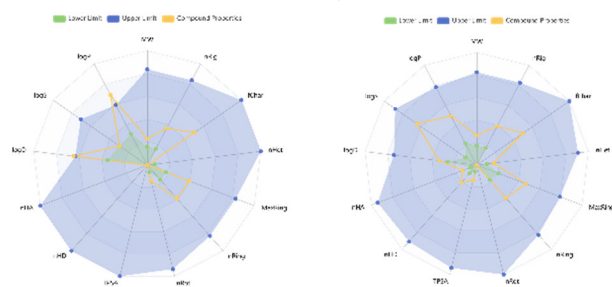


Hình 7. Phân bố số lượng ligand theo năng lượng liên kết

Theo phân bố, các ligand tạo sinh có năng lượng liên kết trong khoảng $(-11,9$ đến $-12,35\text{kcal.mol}^{-1})$ chiếm số lượng cao nhất với 34 chất. Khi năng lượng liên kết càng tăng, số lượng ligand thỏa mãn càng giảm đi rõ rệt. Năng lượng liên kết trong khoảng $-11,5$ (năng lượng liên kết của empagliflozin) đến $-12,5\text{kcal.mol}^{-1}$ (năng lượng liên kết mô phỏng) phù hợp với quá trình docking lên protein 7VSI.

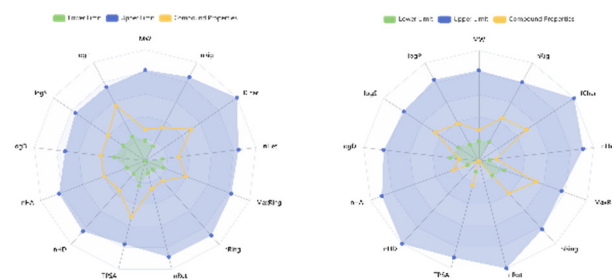
Từ kết quả của quá trình docking, chọn ngẫu nhiên 6 ligand tạo sinh trong sáu vùng phân bố của hình 7 để

thực hiện quá khảo sát đặc tính giống thuốc của phân tử tạo sinh. Kết quả được trình bày ở hình 8.



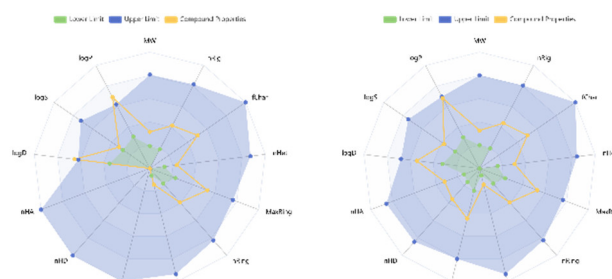
Id_284 ($\Delta G = -13.0 \text{ kcal. .mol}^{-1}$)

Id_289 ($\Delta G = -12.4 \text{ kcal. .mol}^{-1}$)



Id_352 ($\Delta G = -12.0 \text{ kcal. .mol}^{-1}$)

Id_357 ($\Delta G = -13.8 \text{ kcal. .mol}^{-1}$)



Id_425 ($\Delta G = -13.5 \text{ kcal. .mol}^{-1}$)

Id_466 ($\Delta G = -14.6 \text{ kcal. .mol}^{-1}$)

Hình 8. Radar tính chất hóa lý của các ligand tạo sinh

Các chất được lấy ngẫu nhiên chưa thỏa mãn đáp ứng đầy đủ 6 tiêu chí dùng làm thuốc. Đường biểu diễn tính chất của các ligand tạo sinh đều vi phạm vùng tối ưu của đặc tính giống thuốc theo ADMET. Sự gia tăng tính thân dầu và giảm tính thân nước chỉ giúp tăng năng lượng liên kết lên protein như ở các chất id_357, id_435 và id_466. Điều này sẽ không phù hợp cho những thuốc thuộc nhóm SGLT-2i dùng cho đường uống (thấm tốt và tan tốt).

4. KẾT LUẬN

Nghiên cứu đã sàng lọc được 11 chất có năng lượng liên kết lên protein có mã 7VSI cao hơn so với empagliflozin, có khả năng ức chế SGLT-2. Thông qua quá trình docking đã làm rõ được mối liên quan giữa cấu trúc và tác dụng SGLT-2i. Ngoài ra, các khảo sát về tính chất hóa lý và đặc điểm giống thuốc cho thấy 5 ligand có điểm

số docking cao nhất có tiềm năng phát triển thành thuốc với sinh khả dụng đường uống cao cùng với các dữ liệu thử hoạt tính sinh học chính xác như *in vitro* và *in vivo*.

Nghiên cứu đã tạo ra một bộ dữ liệu với 63 chất tạo sinh có điểm số docking hơn 5 hoạt chất ban đầu thông qua trí tuệ nhân tạo tạo sinh. Tuy nhiên, các chất này vẫn chưa thỏa mãn về mặt cấu trúc và đặc tính giống thuốc. Vì vậy, mô hình huấn luyện cần được tiếp tục tinh chỉnh và phát triển để cung cấp những cấu trúc tối ưu hơn.

TÀI LIỆU THAM KHẢO

- [1]. World Health Organization, *Guidance on global monitoring for diabetes prevention and control: framework, indicators and application*. Geneva, Licence: CC BY-NC-SA 3.0 IGO, 2024.
- [2]. World Health Organization, *Classification of diabetes mellitus*. Geneva, 2019. <https://iris.who.int/handle/10665/325182>, truy cập ngày 01/4/2025.
- [3]. International Diabetes Federation, *IDF Diabetes Atlas 2025*. 11, 6, 2025.
- [4]. Institute for Health Metrics and Evaluation (IHME), *GBD Compare Data Visualization*. Seattle, WA: IHME, University of Washington, 2024.
- [5]. Institute for Health Metrics and Evaluation (IHME), *GBD Foresight Visualization*. Seattle, WA: IHME, University of Washington, 2024. Available from <https://vizhub.healthdata.org/gbd-foresight>, truy cập ngày 01/4/2025.
- [6]. World Health Organization, *Diagnosis and management of type 2 diabetes (HEARTS-D)*. Geneva, (WHO/UCN/NCD/20.1), 2020. Licence: CC BY-NC-SA 3.0 IGO.
- [7]. S. Klein, A. Gastaldelli, H. Yki-Järvinen, P. E. Scherer, "Why does obesity cause diabetes?," *Cell metabolism*, 34(1), 11-20, 2022.
- [8]. World Health Organization, *Guidelines on second- and third-line medicines and type of insulin for the control of blood glucose levels in non-pregnant adults with diabetes mellitus*. Geneva. Licence: CC BY-NCSA 3.0 IGO, 2018.
- [9]. American Diabetes Association Professional Practice Committee, *Pharmacologic approaches to glycemic treatment: Standards of Care in Diabetes*. Diabetes Care, 47 (Suppl. 1), S158-S178, 2019.
- [10]. Ernest M. Wright, et al., "SGLT2 Inhibitors: Physiology and Pharmacology," *Kidney* 360, 2(12), 2027-2037, 2021.
- [11]. V. Vallon, H. Koepsell, T. Rieg, et al., "SGLT2 mediates glucose reabsorption in the early proximal tubule," *J. Am. Soc. Nephrol.*, 22, 104-112, 2011.
- [12]. J. R. L. Ehrenkranz, N. G. Lewis, C. R. Kahn, J. Roth, "Phlorizin: a review," *Dia. Meta. Res. Rev.*, 21, 31-38, 2005.
- [13]. Hanefeld M., Forst T., "Dapagliflozin, an SGLT2 inhibitor, for diabetes," *Lancet*, 375, 2196-2198, 2010.

- [14]. C. Flick, S. J. Fink, Chris. J. O'Donnell, et al., "Synthetic Approaches to the New Drugs Approved During 2015," *J. Bioorg. Med. Chem.*, 23, 1895-1922, 2015.
- [15]. J. J. Neumiller, J. R. White Jr, R. K. Campbell, "Sodium-glucose co-transport inhibitors: progress and therapeutic potential in type 2 diabetes mellitus," *Drugs*, 70, 377-385, 2010.
- [16]. Mehta H. R., Patel P. B., Mazhvancherli T. P., et al., "Dapagliflozin: newer molecule for treatment of type 2 diabetes mellitus," *Res. J. Pharm.*, 9(11), 76-84, 2018.
- [17]. Drugbank, *Dapagliflozin*. <https://go.drugbank.com/drugs/DB06292>, truy cập ngày 05/4/2025.
- [18]. A. Manoj, S. Das, A. Joseph et al., "SGLT2 inhibitors, an accomplished development in field of medicinal chemistry: an extensive review," *Future Med. Chem.*, 12, 1961-1990, 2020,
- [19]. Drugbank, *Canagliflozin*. <https://go.drugbank.com/drugs/DB08907>, truy cập ngày 05/4/2025.
- [20]. W. Cai, W. Liu, G. Zhao, et al., "Design of SGLT2 inhibitors for the treatment of type 2 diabetes: a history driven by biology to chemistry," *Med. Chem.*, 11, 317-328, 2015.
- [21]. B. Neal, M. Desai, D. R. Matthews, et al., "Canagliflozin and cardiovascular and renal events in type 2 diabetes," *N. Engl. J. Med.*, 377, 644-657, 2017.
- [22]. Drugbank, *Empagliflozin*. <https://go.drugbank.com/drugs/DB09038>, truy cập ngày 05/4/2025.
- [23]. M.J. Levine, "Empagliflozin for type 2 diabetes mellitus: an overview of phase 3 clinical trials," *Curr. Diabetes Rev.*, 13, 405-423, 2017.
- [24]. A. Alakhdar, B. Poczos, N. Washburn, et al., "Diffusion Models in De Novo Drug Design," *J. Chem. Inf. Model*, 64(19), 7238-7256, 2024.
- [25]. J. Guan, J. Peng, J. Ma, et al., *3D equivariant diffusion for target-aware molecule generation and affinity prediction*. Conference paper at ICLR 2023, 02/02/2023. <https://iclr.cc/virtual/2023/poster/11022>
- [26]. X. Peng, J. Peng, J. Ma, et al., "Pocket2Mol: Efficient Molecular Sampling Based on 3D Protein Pockets," in *International Conference on Machine Learning*, PMLR 162, 17644-17655, 2022.
- [27]. Ding Y., J. Zhao, W. Hu, et al., "C-Aryl glucoside SGLT2 inhibitors containing a biphenyl motif as potential anti-diabetic agents," *Bioorg. Med. Chem. Lett.*, 25(14), 2744-8, 2015.
- [28]. Anderson R. Aguilón, "Synthetic Strategies toward SGLT2 Inhibitors_review," *Org. Process Res. Dev.*, 22(4), 467-488, 2018.
- [29]. J. Ramani, V. K. Vyas, M. Sharma, et al., "A review on the medicinal chemistry of sodium glucose co-transporter 2 inhibitors (SGLT2-I): Update from 2010 to present," *Eur. J. Med. Chem. Rep.*, 6, 100074, 2022.
- [30]. B. Tiwari, R. P. Pandey, N. Hussain, "Recent advances in the synthesis of SGLT2 Inhibitors and natural products from carbohydrates," *Carbohydrate Res.*, 552, 109477, 2025.
- [31]. Morris GM, Huey R., Lindstrom W., et al., "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *J. comput. chem.*, 30(16), 2785-91, 2009.
- [32]. DeLano, Warren L., "Pymol: An open-source molecular graphics tool," *CCP4 Newsl. Protein Crystallogr*, 40(1), 82-92, 2022.
- [33]. Biovia, Dassault Systemes, *Discovery studio modeling environment*. Release San Diego, CA, 2017.
- [34]. C.H. Yao, et al., "Discovery of novel N-β-D-Xylosylindole derivatives as sodium-dependent glucose cotransporter 2 (SGLT2) inhibitors for the management of hyperglycemia in diabetes," *J. Med. Chem.*, 54, 166-178, 2011.
- [35]. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug Deliv. Rev.*, 46(1-3), 3-26, 2001.
- [36]. Usha Subbiah, et al., "Molecular docking and dynamics simulation of Orthosiphon stamineus against SGLT1 and SGLT2," *J. of Biomol. Struc. and Dyn.*, 41(23), 13663-13678, 2023.
- [37]. Fu L., Wang W., Wu C., et al., "ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance," *API functionality and decision support. Nucleic acids research*, 52(W1), W422-W431, 2024.
- [38]. Daina A., O. Michielin, V. Zoete, "SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Scientific Reports*, 7, 42717, 2017.
- [39]. Mingzhe Z., Lin H., Shengying Q., et al., "Cytochrome P450 Enzymes and Drug Metabolism in Humans," *Int. J. Mol. Sci.*, 22(23), 12808, 2021.

AUTHORS INFORMATION

To Van Thinh, Nguyen Qui Hien, Nguyen Thi Huong Giang, Huynh Hoang Thuc

Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam