

Nghiên cứu sàng lọc *in silico* các phân tử nhỏ có tiềm năng gắn kết chọn lọc trên CXCR2 so với CXCR1

Thái Khắc Minh¹, Nguyễn Thị Ánh Tuyết¹, Huỳnh Nguyễn Hoài Phương¹
Cao Hoàng Nhi¹, Nguyễn Đắc Nhân¹, Trần Huệ Tịnh,² Lê Minh Trí^{1,3*}

¹Khoa Dược, Đại học Y Dược TP. Hồ Chí Minh

²Bệnh viện Quận 1, Thành phố Hồ Chí Minh

³Khoa Y, Đại học Quốc gia TP. Hồ Chí Minh

Summary

Reducing expression of CXCR1/2 may lead to a decrease in the inflammatory response that in turn provides the therapeutic effect of autoimmune diseases. The selective binding on CXCR2 can provide a broader therapeutic application and reduce undesirable effects on the immune system. Therefore, in this study 3D-pharmacophore, 2D-QSAR, docking models are used for *in silico* screening process of small molecules with selective binding potential on CXCR2 compared to CXCR1. With data from various articles, the study has built a virtual screening process through the following steps: 3D-pharmacophore, 2D-QSAR, ADMET assessment, docking on CXCR2 and CXCR1 protein. In this study, 21.777.093 substances from ZINC database were used for screening potential inhibitors. As a result, 51 substances with the selective binding affinity to CXCR2 could be used as candidates for further *in vitro* and *in vivo* testing.

Keywords: CXCR2, CXCR1, 3D-Pharmacophore, 2D-QSAR, docking.

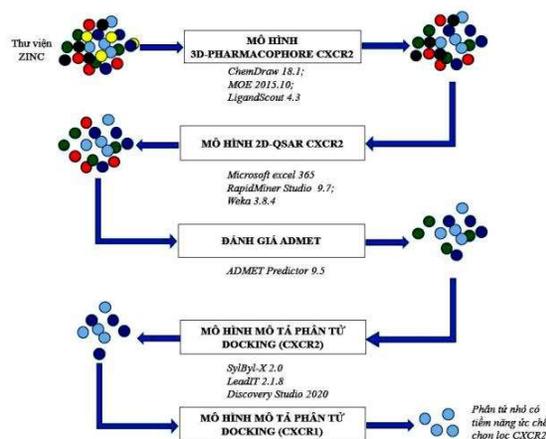
Đặt vấn đề

CXCR1/2 đóng vai trò quan trọng trong hệ miễn dịch của con người và có liên quan đến các bệnh lý như viêm khớp, bệnh phổi tắc nghẽn mãn tính, lupus, hen suyễn, xơ nang, ung thư... Việc giảm biểu hiện của CXCR1/2 có thể dẫn đến giảm đáp ứng viêm, từ đó mang lại tác dụng điều trị đối với các bệnh lý nói trên. Trong điều kiện miễn dịch bình thường, chemokin được tiết ra một lượng nhỏ ở mô hạch bạch huyết và tăng lên cao nhất khi có đáp ứng viêm. Trong cơ thể, chemokin gắn kết với thụ thể CXCR1/2 nhưng ái lực gắn kết của chemokin với thụ thể CXCR1 và CXCR2 có sự khác nhau. CXCR2 gắn kết ái lực cao với tất cả ELR⁺ chemokin, trong khi CXCR1 gắn kết ái lực yếu với các ELR⁺ chemokin ngoại trừ CXCL8. Từ đó, thiết kế thuốc có tác dụng đối kháng chọn lọc trên CXCR2 có thể cho ứng dụng điều trị rộng hơn so với ức chế chọn lọc CXCR, đồng thời hạn chế được tác động không mong muốn đến đáp ứng miễn dịch so với ức chế kép

CXCR1 và CXCR2. Vì vậy, việc xây dựng mô hình *in silico* để sàng lọc các phân tử nhỏ có tiềm năng gắn kết chọn lọc trên CXCR2 so với CXCR1, dựa trên tương tác của các chất với các thụ thể CXCR1/2 là cần thiết.

Phương pháp nghiên cứu

Quá trình sàng lọc các phân tử nhỏ có tiềm năng gắn kết chọn lọc trên CXCR2 so với CXCR1 thực hiện qua các giai đoạn được mô tả trong hình 1.



Hình 1. Mô tả các bước tiến hành nghiên cứu

Chịu trách nhiệm: Lê Minh Trí

Email: leminhtri@ump.edu.vn

Ngày nhận: 08/12/2020

Ngày phản biện: 14/01/2021

Ngày duyệt bài: 22/01/2021

Xây dựng mô hình 3D-pharmacophore CXCR2: Mô hình 3D-pharmacophore xây dựng bằng phần mềm LigandScout dựa trên phối tử, do chưa có cấu trúc tinh thể của CXCR2 cũng như phức hợp giữa CXCR2 và ligand trên Ngân hàng dữ liệu Protein (PDB). Cơ sở dữ liệu được xây dựng từ 252 chất có các giá trị IC₅₀ đều được quy đổi về phương pháp thử thay thế CXCL8 được đánh dấu ¹²⁵I ở CXCR2 tái tổ hợp, được biểu hiện bởi tế bào CHO từ 8 bài báo. Tập xây dựng bao gồm 4 chất có tác dụng đối kháng chọn lọc trên CXCR2 tại vùng allosteric. Trong đó, navarixin là một chất đối kháng chọn lọc trên CXCR2 gấp 80 lần so với CXCR1 và có IC₅₀ là 2,6 nM^[1]. Lấy ngưỡng 26 nM (gấp 10 lần hoạt tính của navarixin) để chia cơ sở dữ liệu thành: Tập hoạt tính chứa các chất có IC₅₀ ≤ 26 nM và tập không hoạt tính chứa các chất có IC₅₀ > 26 nM. Các mô hình cần được đánh giá dựa trên các tiêu chuẩn bao gồm độ nhạy (Se), độ đặc hiệu (Sp), độ đúng (Acc), tỷ suất hoạt tính (Ya) và điểm GH^[2].

Xây dựng mô hình 2D-QSAR: Cơ sở dữ liệu ban đầu là 110 chất có IC₅₀ dao động trong khoảng từ 3,98 nM đến 15840 nM được chọn ra từ tập sau bước xây dựng và sàng lọc mô hình 3D-pharmacophore. Vì giá trị IC₅₀ của các chất này có khoảng biến thiên rất lớn nên được quy đổi thành giá trị pIC₅₀. Các chất được nhập vào MOE và tối thiểu hóa năng lượng bằng công cụ *Energy Minimize*. Sau khi tính toán, các thông số mô tả không có ý nghĩa của tập xây dựng mô hình được loại bỏ bằng các phần mềm MS Excel 365 và RapidMiner 9.7. Sau đó, sử dụng phần mềm Weka 3.8.4 để tìm ra các thông số mô tả có giá trị liên quan nhất với pIC₅₀ thực nghiệm. Phương trình QSAR được xây dựng bằng phương pháp PLS, loại nhiễu thông qua phân tích thành phần chính PCA, thông qua Z-score và Zx-score. Mô hình được đánh giá bằng các đánh giá nội (đánh giá trên tập xây dựng), đánh giá ngoại (dùng mô hình xây dựng trên tập xây dựng để đánh giá trên tập kiểm tra), đánh giá Roy.

Đánh giá ADMET: Nghiên cứu sử dụng phần mềm ADMET Predictor 9.5 của Công ty Simulation Plus (Mỹ)^[3]. Các thông số ADMET được tính toán các rủi ro (risk) nhằm hạn chế thất bại cho việc phát triển thuốc sau này. Các giá trị rủi ro thì không có đơn vị và có khoảng từ 0 đến vô cùng. Nếu giá trị của rủi ro càng cao thì khả năng vượt qua các đánh giá trên tiền lâm sàng và lâm sàng càng thấp. Giá trị ngưỡng

của rủi ro chung và rủi ro hấp thu, chuyển hóa, độc tính được trình bày trong bảng 1.

Bảng 1. Giá trị ngưỡng của các rủi ro ADMET theo ADMET Predictor^[3]

Chung	Hấp thu	Chuyển hóa	Độc tính
7,50	3,50	2,50	2,00

Xây dựng mô hình mô tả phân tử docking CXCR2: Cấu trúc protein được sử dụng trong nghiên cứu này là cấu trúc homology được xây dựng bằng máy chủ I-TASSER, sử dụng kỹ thuật mô tả tính tương đồng (homology) dựa trên trình tự acid amin được cung cấp (số acid amin từ 10 đến 1500)^[4]. Đề tài tiến hành redocking navarixin và lấy điểm số docking làm ngưỡng để sàng lọc các chất gắn kết mạnh trên CXCR2. Navarixin cần được chuẩn bị bằng phần mềm Sybyl-X 2.0 trước khi tiến hành docking. Thông qua ZINID, 5000 chất có giá trị pIC₅₀ tốt nhất từ trang web <http://zinc15.docking.org/> được tải về dưới định dạng tệp *.sdf. Quá trình docking được tiến hành bằng phần mềm LeadIT 2.1.8. Vùng allosteric được tập trung nghiên cứu do đó khoang gắn kết sẽ được xác định bởi các acid amin quan trọng là: Thr83, Asp84, Asp143, Ala147, Ala249, Tyr314, Lys320, Phe321^[5]. Kết quả docking cho biết ái lực gắn kết của ligand với protein và tương tác giữa acid amin với ligand đó.

Xây dựng mô hình mô tả phân tử docking CXCR1: CXCR1 có sẵn cấu trúc 3D trên ngân hàng PDB (Protein Data Bank) với mã 2LNL. Đề tài sử dụng DF2156A làm phối tử redocking trên CXCR1 từ đó xác định điểm số docking, đây cũng chính là điểm số docking được đề tài chọn làm ngưỡng phân biệt giữa gắn kết chọn lọc trên CXCR2 và không chọn lọc trên CXCR2. DF2156A cũng được chuẩn bị 3 bước bằng phần mềm Sybyl-X 2.0. Quá trình docking được thực hiện trên LeadIT 2.1.8 với khoang gắn kết của CXCR1 bao gồm 5 acid amin quan trọng là Tyr46, Lys99, Asn120, Tyr258, Glu291^[6].

Kết quả

Mô hình 3D-pharmacophore

Sau khi xây dựng mô hình 3D-pharmacophore, kết quả xuất ra 200 mô hình, lựa chọn được 10 mô hình. Tiến hành đánh giá độ nhạy, độ đặc hiệu, tỷ suất hoạt tính, độ đúng và điểm GH của từng mô hình. Kết quả đánh giá 10 mô hình 3D-pharmacophore được trình bày ở bảng 2.

Bảng 2. Kết quả đánh giá các mô hình 3D-pharmacophore

Mô hình	Fit 130		Non-Fit 122		Độ nhạy (Se)	Độ đặc hiệu (Sp)	Tỷ suất hoạt tính (Ya)	Độ đúng (Acc)	Điểm GH
	TP	FN	FP	TN					
Ph001	109	21	48	74	0,8385	0,6066	0,6943	0,7262	0,4430
Ph025	95	35	35	87	0,7308	0,7131	0,7308	0,7222	0,5211
Ph045	117	13	57	65	0,9000	0,5328	0,6724	0,7222	0,3886
Ph047	115	15	50	72	0,8846	0,5902	0,6970	0,7421	0,4390
Ph060	107	23	22	100	0,8231	0,8197	0,8295	0,8214	0,6786
Ph064	88	42	17	105	0,6769	0,8607	0,8381	0,7659	0,6866
Ph079	120	10	49	73	0,9231	0,5984	0,7101	0,7659	0,4567
Ph082	117	13	44	78	0,9000	0,6393	0,7267	0,7738	0,4923
Ph155	110	20	43	79	0,8462	0,6475	0,7190	0,7500	0,4861
Ph156	128	2	95	27	0,9846	0,2213	0,5740	0,6151	0,1497

Trong số các kết quả đánh giá, mô hình Ph060 và Ph064 đều cho kết quả độ đúng và điểm GH cao cho thấy khả năng dự đoán của 2 mô hình này là khá tốt. Mô hình Ph060 có độ nhạy cao hơn so với mô hình Ph064 (0,823 > 0,6769), tuy nhiên mô hình Ph064 lại có tỷ suất hoạt tính cao hơn mô hình Ph060 (0,8381 > 0,8295). Giữa độ nhạy và tỷ suất hoạt tính, đề tài chú trọng tỷ suất hoạt tính hơn vì khi tỷ suất hoạt tính càng cao thì xác suất lấy được chất dương tính thật trong số các chất dương tính càng cao.

Mô hình 2D-QSAR

Từ 206 thông số mô tả 2D ban đầu, đề tài tiến hành lọc thô bằng Microsoft Excel 365, loại các thông số mô tả có tương quan > 0,40 bằng RapidMiner Studio 9.7, loại các thông số không phù hợp bằng phần mềm Weka 3.8.4, thu được 7 thông số: h_pKb, PEOE_RPC-, PEOE_VSA-3, Rsynth, SlogP_VSA0, SMR_VSA3, vsa_acc. Dùng các thông số mô tả này để xây dựng mô hình 2D-QSAR, 3 chất gây nhiễu được loại bỏ bằng PCA, 11 chất gây nhiễu được loại bỏ bằng z-score và zx-score, còn lại 96 chất. Sau khi xây dựng mô hình toàn tập 2D-QSAR, chia tập dữ liệu thành tập xây dựng (80% toàn tập, XD) và tập kiểm tra (20% toàn tập, KT) theo phương pháp phân chia ngẫu nhiên để đánh giá nội, đánh giá ngoại và đánh giá trên mô hình toàn tập. Tiến hành phân chia 5 lần, đánh giá mỗi lần một cách độc lập rồi lấy trung bình cả 5 lần. Kết quả đánh giá cho thấy bình phương hệ số tương quan R² trên các tập trung bình, ngẫu nhiên và trên toàn tập đều lớn hơn 0,70.

Đánh giá nội (Q² > 0,50) và đánh giá ngoại (R²_{pred} > 0,50) đều đạt trên tất cả các tập. Về đánh giá Roy, giá trị γ^2_m đều đạt lớn hơn 0,50 ở tất cả các tập, giá trị γ^2_m mặc dù không đạt ở 2 tập KT4 và KT5 nhưng đạt trên hầu hết các tập khác trong đó có tập trung bình và toàn tập, do đó kết quả này là chấp nhận được.

Phương trình 2D-QSAR xây dựng được có dạng:

$$pIC_{50} = 6,24033 + 1,81004 * h_pKb - 0,86256 * PEOE_RPC - 0,37833 * PEOE_VSA-3 - 0,69499 * rsynth + 0,52543 * SlogP_VSA0 + 0,69728 * SMR_VSA3 + 0,25629 * vsa_acc$$

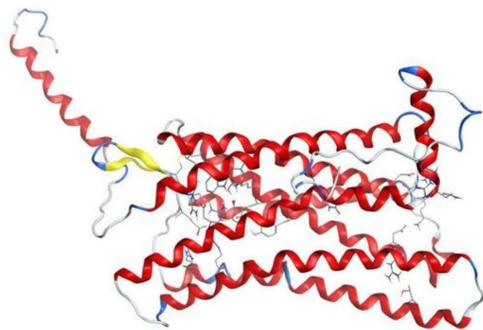
Với N = 96, R2 = 0,79 và RMSE = 0,30.

Đánh giá ADMET

103.772 chất sau khi sàng lọc qua mô hình 2D-QSAR được đưa vào phần mềm ADMET Predictor 9.5 để dự đoán các rủi ro hấp thu, chuyển hóa và độc tính. Kết quả có 86.673 chất vượt qua được đánh giá ADMET và được sử dụng cho các bước sàng lọc tiếp theo.

Mô hình homology CXCR2

Trình tự 360 acid amin của CXCR2 được tải lên máy chủ I-TASSER để xây dựng mô hình protein CXCR2 dựa vào kỹ thuật mô tả tính tương đồng một cách tự động cho ra 5 mô hình homology của CXCR2. Tiếp tục đánh giá 5 mô hình với các thông số bao gồm C-score, TM-score, RMSD và mật độ đám. Dựa vào kết quả đánh giá, mô hình 1 có C-score = -0,57, TM-score = 0,64 ± (-0,13), RMSD = 7,8 ± (-4,4), mật độ đám = 0,1406 cho thấy mô hình 1 là mô hình tốt nhất và đáng tin cậy nhất trong 5 mô hình.



Hình 2. Mô hình homology số 1 của cấu trúc CXCR2 được tải về từ máy chủ I-TASSER và được thể hiện trên MOE

Mô hình mô tả phân tử docking trên CXCR2

Tiến hành redocking tất cả các cấu dạng gắn kết navarixin trên khoang allosteric nội bào. Kết quả có tổng cộng 1264 pose với điểm số docking từ -8,7887 kJ/mol đến -0,0057 kJ/mol. Để có cơ sở so sánh về tiềm năng gắn kết với các chất sau sàng lọc, nghiên cứu tiến hành phân tích sự tương tác giữa navarixin với các acid amin quan trọng trong khoang gắn kết của CXCR2 bằng công cụ Ligand Interactions của phần mềm MOE 2015.10. Kết quả chọn được cấu dạng của navarixin tương tác với 5 acid amin quan trọng Thr83, Ala249, Tyr314, Lys320, Phe321 với điểm số docking là -8,7887 kJ/mol.

Trong số 5000 chất có pIC_{50} tốt nhất được chọn ra từ kết quả sàng lọc sau các mô hình 3D-pharmacophore, 2D-QSAR, đánh giá ADMET, có 4848 chất docking thành công (96,96%) với điểm số docking trải dài từ -25,0510 kJ/mol đến 6,8340 kJ/mol. Lấy điểm số docking âm nhất của navarixin làm ngưỡng (-8,7887 kJ/mol), chọn ra được 2245 chất có điểm số gắn kết bằng hoặc tốt hơn navarixin.

Mô hình mô tả phân tử docking trên CXCR1

Kết quả DF2165A redocking được 567 pose với điểm số docking từ -18,9839 kJ/mol đến 0,7337 kJ/mol. 2245 chất docking thành công trên CXCR2 được tiếp tục tiến hành docking trên CXCR1 tại vị trí gắn kết nội bào allosteric. Lấy trung bình điểm số của pose có điểm âm nhất và pose có điểm dương nhất làm ngưỡng (-9,1251 kJ/mol). Các chất không gắn được trên CXCR1 hoặc có điểm số docking dương hơn ngưỡng này được xem là gắn kết không tốt trên CXCR1. Kết quả có 4 chất không docking được

trên CXCR1 và 47 chất có điểm số docking lớn hơn -9,1251 kJ/mol. Vậy có tổng cộng 51 chất được xem là có tiềm năng gắn kết chọn lọc trên CXCR2.

Kết luận

Từ 4 chất ức chế chọn lọc CXCR2 đang được thử nghiệm trên lâm sàng, nghiên cứu đã xây dựng được mô hình 3D-pharmacophore 5 điểm với tỷ suất hoạt tính $Y_a = 0,83$; độ đúng = 0,77 cho thấy khả năng dự đoán của mô hình là khá tốt. Tiếp đến, mô hình 2D-QSAR được xây dựng từ tập dữ liệu gồm 96 chất với $R^2 = 0,79$ và RMSE = 0,30 được ứng dụng để dự đoán giá trị IC_{50} của các chất. Sau đó, tiến hành đánh giá ADMET và loại đi các chất không thỏa mãn ngưỡng tham chiếu. Vì chưa có cấu trúc tinh thể trên Ngân hàng protein PDB nên cấu trúc protein CXCR2 trong bài được xây dựng bằng kỹ thuật mô tả tính tương đồng homology với điểm số tin cậy C-score = -0,57 và điểm TM-score = 0,64 cho thấy mô hình homology xây dựng được có độ tin cậy cao và cấu trúc hình học không gian đúng so với cấu trúc gốc của CXCR2. Cuối cùng, đề tài thực hiện docking top 5000 chất có giá trị pIC_{50} tốt nhất lên hai thụ thể CXCR1 và CXCR2 nhằm so sánh điểm số gắn kết cũng như tương tác với các acid amin quan trọng để chọn ra các chất có tiềm năng gắn kết chọn lọc trên CXCR2. Mô hình được đánh giá bằng cách redocking navarixin và DF2165A và so sánh với các kết quả đã được nghiên cứu trước đó, kết quả cho thấy mô hình docking khá tin cậy. Kết quả ứng dụng sàng lọc ảo từ 21.777.093 chất thuộc thư viện ZINC đã thu được 51 chất tiềm năng được đoán là có khả năng gắn kết chọn lọc trên CXCR2. Nghiên cứu cần được tiến hành chạy mô phỏng động lực học 51 chất trên để có thể hướng đến việc thử nghiệm *in vitro* và *in vivo*.

Tài liệu tham khảo

1. Lu H. et al. (2018), *J. Med. Chem.*, 61 (6), pp. 2518-2532.
2. Kirchmair J. et al. (2008), *J. Comput. Aided Mol. Des.*, 22 (3), pp. 213-228.
3. SimulationPlus. *ADMET Predictor v9.5*, Inc. Simulations Plus, (2019), Lancaster, CA, USA
4. Hillisch A. et al. (2004), *Drug Dis., Today*, 9 (15), pp. 659-669.
5. Ha H. et al. (2017), *Theranostics*, 7 (6), pp. 1543-1588.
6. Bertini R. et al. (2012), *British J. Pharmacol.*, 165 (2), pp. 436-454.