DOI: 10.15625/vap.2020.00130

PHƯƠNG PHÁP SÀNG LỌC ẢO TRONG NGHIÊN CỨU PHÁT TRIỂN THUỐC

Phạm Minh Quân*, Lê Thị Thùy Hương, Trần Quốc Toàn, Phạm Thị Hồng Minh, Phạm Quốc Long

Viện Hóa học Các hợp chất thiên nhiên, Viện Hàn lâm Khoa học và Công nghệ Việt Nam
Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam
**Email: pham-minh.quan@inpc.vast.vn*

Tóm tắt

Ngày nay sàng lọc ảo (Virtual Screening) là một kỹ thuật thường xuyên được sử dụng để xác định các hợp chất tiềm năng trong nghiên cứu hóa dược. Số lượng các phương pháp và phần mềm sử dụng cách nghiên cứu tiếp cận hợp chất và đích mục tiêu đang được phát triển với tốc độ nhanh chóng. Tổng quan này sẽ trình bày ngắn gọn về những tiến bộ và tình hình ứng dụng của công nghệ hóa - sinh - tin trong nghiên cứu phát triển thuốc qua hai hướng nghiên cứu: sàng lọc trên nền tảng cấu trúc chất (Structure Based Virtual Screening - SBVS) và nền tảng hợp chất (Ligand Based Virtual Screening - LBVS).

1. Tình hình sử dụng công nghệ thông tin trong nghiên cứu hoá - sinh - y học

Phương pháp ứng dụng công nghệ thông tin trong nghiên cứu hoá - sinh - y học đã được phát triển từ cuối những năm 1950 trên thế giới. Trong những năm 1960, những chương trình máy tính đơn giản đã có thể sử dụng để mô phỏng phổ NMR. Sử dụng mô hình phân tích mối tương quan hoạt tính - cấu trúc Hansch, nhiều máy tính có thể được kết nối để giải quyết những phương trình hồi quy phức tạp. Tuy nhiên, các phân tử thực tế là khá phức tạp để có thể giải quyết các vấn đề liên quan đến cấu trúc không gian vào thời điểm đó (John & Herbert, 2005).

Trong những năm 1970, với sự cải thiện về tốc độ xử lý cộng với giao diện sử dụng thân thiện, công nghệ tin học đã có những đóng góp đáng kể hơn. Khó khăn chính trong thời gian này là chưa có các chương trình máy tính có thể mô tả chính xác các phân tử cùng các tính chất của chúng từ các kết quả lý thuyết. Rào cản này sau đó được tháo gỡ với sự xuất hiện của các máy tính được trang bị các chương trình đồ hoạ mạnh đủ để có thể miêu tả các HOMO, LUMO, MUP (*molecular electrostatic potential*), các véctơ môмen lưỡng cực,… chồng lên cấu trúc 3D của phân tử. Đầu những năm 1990, các máy tính lớn đa nhân (*cluster*) đã đủ mạnh để thực hiện các tính toán trên các phân tử thực trong thời gian đủ nhỏ, kết quả này cũng góp phần tăng cường sự quan tâm của các nhà hoá học vào sử dụng các ứng dụng của công nghệ thông tin trong nghiên cứu hoá học của các phân tử hữu cơ (Tame, 1999).

Trong các nghiên cứu hoá học các hợp chất thiên nhiên trước kia, các hoạt chất mới được phân lập chủ yếu là ngẫu nhiên và thông qua việc sàng lọc hoạt tính sinh học đơn giản bao gồm các hoạt tính kháng sinh, độc tế bào,… Hiện nay, tại các nước phát triển, các loại thuốc thế hệ mới được phát hiện và phát triển thông qua các *công cụ sàng lọc mạnh về di truyền học và hoá sinh*, trong đó, sử dụng các dòng tế bào thay thế quan trọng,

các trung gian điều hoà, hay sử dụng sự tương tác *thụ thể - hợp chất* (Receptor - Ligand). Các sàng lọc này sẽ cho phép phát hiện chính xác các hợp chất có chứa hoạt tính mong muốn trong rất nhiều các dịch chiết khác nhau. Quan trọng hơn, các thử nghiệm này cung cấp những thông tin ban đầu về cơ chế hoạt động của hoạt chất trong quá trình phát triển thuốc (Reddy & Pati et al., 2007).

Để thực hiện được các sàng lọc trên nhất định phải có cấu trúc của các "protein đích" quy định bệnh, phương pháp trên ngoài sự chính xác và là nguồn cung cấp cơ chế tác động của thuốc, còn là cơ sở quan trọng để phát triển các loại thuốc mới khi bệnh đã kháng thuốc. Khi sử dụng thuốc không đúng chỉ định hoặc do các điều kiện môi trường, các tác nhân hoá học có thể dẫn đến tình trạng bệnh kháng thuốc do một sự đột biến nào đó trong cấu trúc của ADN, tức là cấu trúc của protein đích có biến đổi. Nếu chỉ dựa trên các sàng lọc hoá học - hoạt tính sinh học thông thường không thể phát hiện ra các biến đổi này. Tuy nhiên, với công nghệ sinh học kết hợp hoá học thì vấn đề có thể được giải quyết bằng việc nghiên cứu những thay đổi trong cấu trúc ADN, sự sai khác giữa tương quan thụ thể - thuốc và biến đổi cấu trúc các thuốc đang sử dụng làm cho hiệu quả của thuốc trở lại. Lĩnh vực sàng lọc trên đòi hỏi sự kết hợp chặt chẽ của các nhà nghiên cứu trong ba lĩnh vực sinh học, hoá học và y dược học.

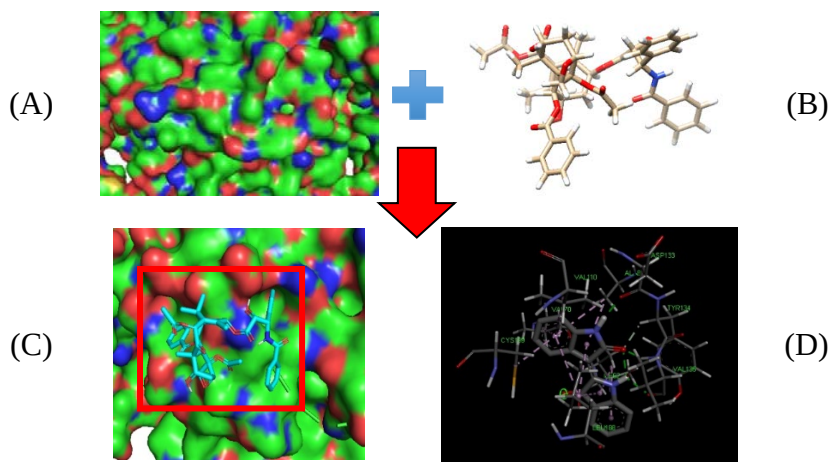
Trong các mô hình sàng lọc hoạt chất hiện đại, mới đây xuất hiện phương pháp sàng lọc ảo *in silico* (*virtual screening*) và ngay lập tức đã đóng một vai trò hết sức quan trọng. Phương pháp trên sử dụng các tiến bộ trong tin học để sàng lọc ảo, mô tả và dự đoán các cấu trúc mới được cho là có hoạt tính mạnh. Ưu điểm của phương pháp là giảm thiểu chi phí và thời gian trong quá trình phát hiện và phát triển thuốc. Nó thường được mô tả là một phương pháp gồm nhiều bước theo tuần tự thông qua các tiêu chí sàng lọc khác nhau để từ đó thu hẹp dần để lựa chọn các hợp chất có tiềm năng phát triển làm thuốc với những hoạt tính sinh học mong muốn. Hợp chất được nghiên cứu không nhất thiết phải có sẵn và việc thử nghiệm chúng là mô phỏng ảo nên không gây tốn kém về nguyên vật liệu. Dựa vào nguyên lý này, bất kỳ hợp chất nào cũng có thể được đánh giá thông qua sàng lọc ảo. Tùy thuộc vào quy mô nghiên cứu, cơ sở dữ liệu hợp chất cho sàng lọc ảo có thể lên tới hàng chục triệu hợp chất và toàn bộ những chất này có thể được phân tích chỉ sàng một lần sàng lọc.

Thông thường, mỗi loại thuốc mới được đưa ra thị trường phải tốn kém khoảng 800 triệu euro và tốn thời gian 10-15 năm (Song & Lim et al., 2009). Trong khi đó, với các hệ thống máy tính nối mạng hiện đại (ví dụ tính toán lưới - Grid) thì hàng triệu cấu trúc có thể được sàng lọc ảo chỉ trong thời gian vài tuần (Mullard, 2014).

Bảng 1. Thông tin về một số dự án sàng lọc *in silico* trên thế giới

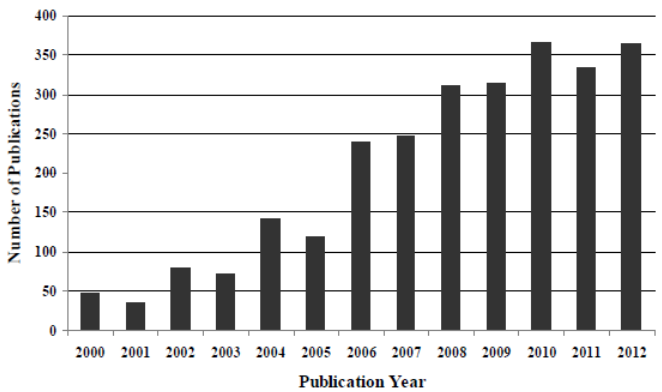
Tên dự án	Protein	Số lượng ligand	TLTK
Malaria	Plasmepsin PMII	10 triệu	(de Beer & Wells et al., 2009)
Avian flu	Neuraminidase	300 triệu	(Lee & Salzemann et al., 2006)
Diabetes	Amylase/Glucoamylase	300 triệu	(Roy & Kumar et al., 2013)
SARS-CoV-2	Chymotrypsin-like cysteine protease -$3CL^{pro}$	1 tỷ	(Ton, Gentile et al., 2020)

Các sàng lọc *in silico* sử dụng các tương tác giữa Receptor - Ligand để tìm ra các hợp chất (Ligand) có cấu trúc được dự đoán liên kết với thụ thể tốt nhất - ở đây là có mức năng lượng ΔG thấp nhất (hình 1). Cấu trúc các protein đích ở mô hình 3 chiều (3D) đối với mỗi bệnh được cung cấp bởi các nhà sinh học, các ligand được phát triển dựa theo cấu trúc của các hợp chất hoá học, đặc biệt là các bộ khung cacbon đã được biết rõ ràng và có nguồn cung cấp, ngoài ra các sàng lọc này yêu cầu các phần mềm máy tính bản quyền và hệ thống máy tính với tốc độ rất nhanh (Pagadala & Syed et al., 2017).



Hình 1. Tương tác protein - ligand. (A) Bề mặt vùng hoạt động của protein; (B) Cấu trúc ba chiều của ligand; (C) Trạng thái liên kết bề mặt protein - ligand; (D) Cấu hình tương tác ba chiều protein - ligand

Công trình nghiên cứu sử dụng phương pháp sàng lọc ảo được ghi nhận công bố quốc tế lần đầu tiên vào năm 1997. Kể từ đó cho tới nay, việc ứng dụng mô hình này ngày càng trở nên phổ biến và trở thành một xu thế nghiên cứu mới trong ngành dược học, đi kèm đó là số lượng các nghiên cứu công bố liên quan tới lĩnh vực này ngày càng tăng mạnh (hình 2).



Hình 2. Tổng số công bố liên quan tới sàng lọc ảo trong giai đoạn từ năm 2000-2012 ở 12 tạp chí lĩnh vực hóa - sinh - tin (Lavecchia và Giovanni, 2013)

2. Các mô hình sàng lọc ảo trên thế giới hiện nay

Phương pháp sàng lọc ảo có thể được chia thành 2 hướng chính bao gồm sàng lọc trên nền tảng hợp chất (LBVS) và sàng lọc trên nền tảng cấu trúc (SBVS). Hướng sàng lọc LBVS sử dụng các dữ liệu tương quan cấu trúc - hoạt tính từ một tệp cơ sở dữ liệu các chất đã biết để lựa chọn chất tiềm năng cho đánh giá thực nghiệm. Hướng nghiên cứu này bao gồm việc tìm kiếm các hợp chất có cấu trúc tương đồng, dẫn xuất, nghiên cứu tương quan cấu trúc - hoạt tính (QSAR) và dược học. Hướng sàng lọc SBVS, theo một cách khác, sử dụng cấu trúc ba chiều của đích sinh học để mô phỏng tương tác ảo với các hợp chất tiềm năng và xếp hạng chúng dựa trên ái lực liên kết hoặc vùng liên kết.

2.1. Hướng nghiên cứu sàng lọc trên nền tảng cấu trúc chất (SBVS)

Đối với hướng nghiên cứu này, dữ liệu đầu vào bao gồm: cấu trúc của đích protein nghiên cứu đã được làm rõ đi kèm với tệp cơ sở dữ liệu các hợp chất nghiên cứu. Các hợp chất này sẽ được nghiên cứu thông qua mô phỏng docking chúng trên các vùng hoạt động (*active sites*) của đích sinh học (*protein/enzyme*) sử dụng những thuật toán tính toán khác nhau. Tiếp theo, một thuật toán khác sẽ tính điểm để xếp hạng sự gắn kết giữa hợp chất với đích sinh học. Đây thường là một quy trình nhiều bước trong đó hợp chất được xếp hạng và lựa chọn dựa trên điểm tương tác và một số tiêu chí khác. Thông thường, chỉ một số ít các hợp chất có điểm cao nhất mới được đem thử nghiệm thực tế.

Vào những năm đầu tiên khi mô hình sàng lọc mới phát triển, phần mềm thuật toán được sử dụng thời điểm này có tên UCSF Dock ((Irwin D. K., 1982), kể từ đó đến nay rất nhiều các phần mềm khác đã được phát triển, ví dụ: Gold (Gareth J., 1997), Dock (Ewing T. J., 2001), Glide (Thomas A. H., 2004; Richard A. F., 2004), FlexX (Bernd K., 1999), AutoDock (Oleksandr V. B., 2002) (bảng 2) (Pagadala & Syed et al., 2017).

Bảng 2. Thống kê một số phần mềm sàng lọc ảo đang có trên thế giới

Software	Website
AutoDock	http://autodock.scripps.edu/
Dock	http://dock.compbio.ucsf.edu/
FlexX	http://www.biosolveit.de/flexx/
Glide	http://www.schrodinger.com/
Gold	http://www.ccdc.cam.ac.uk/products/life_sciences/gold/

Một trong những bước quyết định trong mô hình SBVS là việc xếp hạng điểm của các hợp chất. Ngày nay, cho dù việc dự đoán cấu hình tương tác giữa hợp chất với đích sinh học có thể được thực hiện dễ dàng với nhiều phần mềm khác nhau, tuy nhiên, việc tính điểm và xếp hạng chúng vẫn là một bài toán hóc búa và nhiều thách thức. Sự khó khăn này xuất phát từ thực tế rằng trong một số tình huống, một số tương tác rất khó để tham số hóa. Việc tính điểm được sử dụng cho những mục tiêu sau: a) Đánh giá các cấu hình tương tác của một hợp chất được tạo ra bởi các thuật toán khác nhau để chọn ra được tương tác khả dĩ nhất; b) Xếp hạng các hợp chất từ đó lọc ra hợp chất có tiềm năng nhất.

Các phương pháp tính điểm đã được phát triển liên tục trong nhiều năm qua, chúng được phân ra thành 3 mô hình chính: trường lực (*force field-based*), cơ sở kiến thức (*knowledge-based*) và thực nghiệm (*empirical*). Một số mô hình tính điểm sử dụng kết hợp hai mô hình force field-based và empirical (Krovat & Steindl et al., 2005).

Mô hình force field-based dự đoán năng lượng liên kết tự do là tổng của các trường năng lượng cơ học phân tử như: Coulomb, Van der Waals, liên kết hydrogen (Meng, Shoichet et al., 1992). Năng lượng solvat hóa và entropy cũng có thể được tính đến. Mô hình tính điểm empirical coi năng lượng liên kết tự do là tổng của các liên kết gồm: liên kết hydrogen, liên kết kỵ nước bằng cách khớp điểm tính toán với số liệu ái lực liên kết thực nghiệm đối với các bộ phức hợp protein-ligand. Mô hình knowledge-based dựa trên số liệu thống kê phân tích tần số cặp nguyên tử trong phức hợp phối tử protein-ligand với cấu trúc ba chiều đã biết.

Trong hai thập kỷ qua, nhiều nỗ lực đáng kể đã được thực hiện để tinh chỉnh các chức năng tính điểm để dự đoán chính xác năng lượng liên kết tự do, do đó chúng có thể được sử dụng để xếp hạng trừ trường hợp định lượng về hoạt tính. Tuy nhiên, do sự phức tạp của quá trình liên kết protein-ligand và các phép tính gần đúng được thực hiện khi tính toán các quá trình desolvat hóa và entropy, điểm docking vẫn chưa chứng tỏ được độ chính xác trong dự đoán ái lực liên kết. Một số biện pháp đã được đưa ra nhằm cải thiện khả năng tính điểm bao gồm thêm vào các yếu tố để tính hiệu ứng solvat hóa và entropy để cho ra các thuật ngữ chính xác bằng các phép tính lượng tử cao cấp, các hàm tính điểm cụ thể theo mục tiêu và tính điểm đồng thời bằng cách kết hợp nhiều mô hình tính điểm. Mặt khác, có một cách hiệu quả hơn là sử dụng điểm docking làm định hướng để xác định mức độ phù hợp của tương tác kết hợp với các thông số đo khác như khả năng vừa khớp đối với từng chất riêng biệt. Những thông số này có thể thu được thông qua việc quan sát các liên kết hydrogen, đây là một tham số rất quan trọng trong docking, cấu hình trong không gian của liên kết π-π và/hoặc độ chiếm dụng không gian của vùng kị nước trước vị trí của ligand trong vùng liên kết.

Một khía cạnh khác chưa khai thác của mô hình SBVS là độ linh động của thụ thể đích, điều này sẽ tiêu tốn nhiều tài nguyên máy tính cũng như phức tạp hơn để xử lý. Trong những năm gần đây, một trong những thách thức lớn nhất của rất nhiều thuật toán docking là xử lý những thụ thể đích linh động. “Soft docking” (có trong mọi phần mềm Docking) cho phép xảy ra những sự chồng chéo nhỏ giữa ligand và thụ thể mà không có khoảng không lớn (Jiang và Kim, 1991). Tuy nhiên, điều này có thể làm tăng tỉ lệ sai kết quả vì nó khiến các chất có cấu trúc đa dạng hơn được liên kết. Nó cũng không cho các chất có cấu hình lớn thay đổi ví dụ như xoay mạch nhánh hay dịch chuyển bộ khung chính protein. Một số phần mềm như Autodock4, Dock, Gold, EADock, IFREDA, FlexE hay GLIDE induced Fit (bảng 3) cho phép mô phỏng xoay quanh vị trí xoắn bậc tự do của mạch nhánh đã chọn (ví dụ các chuỗi thuộc vùng liên kết) áp dụng các phương pháp tương tự để khám phá cấu hình không gian của ligand linh động.

Bảng 3. Các phần mềm Docking bao gồm tính linh động của protein

Tên phần mềm	Tính linh động ligand	Tính linh động protein	Mô hình tính điểm
Autodock	Evolutionary algorithm	Flexible side chain	Force field
Dock	Incremental build	Protein side chain and flexibility	Force field or contact score
Gold	Evolutionary algorithm	Protein side chain and backbone flexibility	Empirical score
EADock	Evolutionary algorithm	Flexible side chain and backbone	Force field

Ngày nay nhiều học thuyết khác đang được phát triển liên tục và ứng dụng của chúng cũng rất tiềm năng cho sàng lọc ảo. Một trong những học thuyết này là Relaxed Complex Scheme (RCS). RCS sử dụng một tập hợp các cấu trúc năng lượng thấp được trích xuất từ mô phỏng động học phân tử (MD) để tìm kiếm trong các cơ sở dữ liệu thông qua docking các hợp chất. Nó kết hợp các ưu điểm của thuật toán docking với thông tin động của cấu trúc có bởi mô phỏng MD, tính toán chi tiết cho cấu trúc động của cả thụ thể và các hợp chất đã dock (Lin & Perryman et al., 2003). Các mô phỏng MD thời gian càng dài thì càng tăng khả năng nghiên cứu cấu hình không gian của thụ thể trước khi dock. Mô hình này đã được phát triển kết hợp với nhiều gói phần mềm MD khác nhau bao gồm: AMBER (Case & Cheatham et al., 2005), NAMD (Phillips & Braun et al., 2005), GROMACS (Van Der Spoel & Lindahl et al., 2005) và AUTODOCK (Morris & Goodsell et al., 1998) để làm dock các ligand.

2.2. Hướng nghiên cứu trên nền tảng hợp chất (LBVS)

Đối với hướng nghiên cứu này, dữ liệu hoạt tính sinh học đã được biết sẵn nhằm xác định được những hợp chất có hoặc không có hoạt tính để từ đó tìm kiếm các hợp chất tiềm năng hơn dựa trên sự tương đồng cấu trúc, dược lý và các tiêu chí khác.

Một trong những mô hình nghiên cứu LBVS phổ biến nhất đó là nghiên cứu tương quan hoạt tính cấu trúc (QSAR). Mục tiêu của QSAR là xác định mối tương quan giữa các đặc tính cấu trúc/hóa lý của hoạt chất đã biết với hoạt tính sinh học của chúng. Những thông tin về mức độ hoạt động của hợp chất như ái lực liên kết (K_D) hay nồng độ ức chế tối thiểu (IC_{50}) là rất cần thiết đối với QSAR. Ở đây cấu trúc của hợp chất thường được miêu tả bởi tập hợp các thông tin về cấu trúc, hóa lý được coi là có liên quan tới việc liên kết của chúng. Chất lượng của mô hình QSAR bị ảnh hưởng bởi khả năng tương thích với mỗi trường hợp, dữ liệu đầu vào cấu trúc - hoạt tính, cách miêu tả hợp chất, ảnh hưởng của các dữ liệu ngoại vi, tính phù hợp của các mối tương quan đã phát triển, cấu hình 3D và việc lựa chọn các hướng giải quyết (Verma & Khedkar et al., 2010).

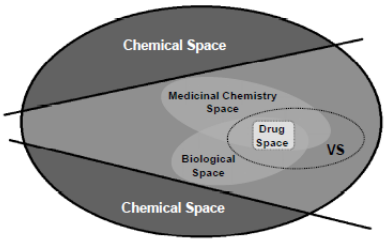
Công nghệ máy tự học (Machine learning) đang ngày càng được sử dụng phổ biến hơn trong nền tảng thuật toán cho hướng nghiên cứu LBVS nhằm xây dựng và tìm kiếm nhanh chóng, chính xác các mối tương quan hoạt tính - cấu trúc. Đã có nhiều công nghệ khác

nhau được phát triển, mỗi công nghệ có những ưu và nhược điểm riêng. Trong số những phương pháp này, các mô hình hồi quy và phân loại ví dụ như: Hồi quy đa tuyến tính, thuật toán láng giềng, phân loại Naïve Bayesian, véctơ hỗ trợ (Support Vector Machines), mạng neuron nhân tạo và thuật toán Decision trees đến nay đã được áp dụng khá thành công. Những thuật toán này dựa trên một số đặc tính nhất định để lọc ra các hợp chất có hoạt tính (Melville & Burke et al., 2009).

Hiệu quả của công nghệ Machine learning phụ thuộc vào nhiều yếu tố như: sự đa dạng của dữ liệu, khả năng xử lý về sự mất cân bằng trong tệp dữ liệu (số hợp chất không hoạt tính thường vượt trội so với các hợp chất có hoạt tính) và các tham số về hoạt tính của các hợp chất.

3. Cơ sở dữ liệu sử dụng trong sàng lọc ảo

Một trong những điều kiện tiên quyết trong phát triển thuốc truyền thống đó là xác định được một đích sinh học đã được xác thực, ví dụ một hợp chất đã được nghiên cứu chứng minh rằng có khả năng tương tác với đích sinh học đó dẫn tới khả năng chữa được bệnh hoặc cải thiện triệu chứng bệnh. Bước đầu tiên này bao gồm xác định đích sinh học tiềm năng và sau đó xác thực chúng. Việc xác định đích sinh học tiềm năng cần tới việc nghiên cứu trong "Vùng Sinh học" (Biological space) (hình 3) thông qua việc giải trình tự gen người, phụ thuộc vào công nghệ giải trình tự tốc độ cao và các thuật toán máy tính để xử lý lượng lớn dữ liệu xuất ra. Sau khi đã tìm và xác thực được đích sinh học, bước tiếp theo là xác định một thực thể có thể tương tác chọn lọc với đích đó theo cách có thể tạo ra hiệu ứng chữa bệnh. Theo khái niệm của lĩnh vực nghiên cứu thuốc, thực thể này là một hợp chất hóa học khối lượng phân tử nhỏ. Việc tìm kiếm một hợp chất liên kết chọn lọc tới đúng vùng hoạt động của protein là không hề dễ dàng. Để tăng cơ hội thành công, cần tìm kiếm kĩ lưỡng chúng trong "Vùng Hóa học" (Chemical space). Về lý thuyết, tổng số hợp chất có trong Vùng Hóa học có thể ước lượng tới 10 triệu hợp chất (Bohacek & McMartin et al., 1996). Đây là một con số rất lớn và vượt ngoài khả năng của các nhà khoa học hiện nay.



Hình 3. Mô hình tìm kiếm trong nghiên cứu dược học

Mặc dù đã có rất nhiều nỗ lực trong việc xây dựng những cơ sở dữ liệu siêu lớn, việc thu thập đầy đủ hợp chất cho "Vùng Hóa học" là điều chưa thể thực hiện được hiện nay, ngoài ra rất ít tập đoàn dược nào có được tệp cơ sở dữ liệu nhiều hơn 2 triệu chất. Tuy nhiên, chỉ một phần nhỏ hợp chất trong các cơ sở dữ liệu đó có được tính ổn định, tan trong nước, có những nhóm chức phù hợp để tạo liên kết với đích sinh học chẳng hạn như các protein hay axit nucleic và đủ đặc điểm cấu trúc để đáp ứng được các tính chất chọn

lọc, chúng được xếp vào vùng "Hợp chất dược học" (Medicinal Chemistry Space) (Selzer & Roth et al., 2005). Có ý kiến cho rằng những hợp chất trong "Vùng Hóa học" có được từ việc thu thập truyền thống là không đủ để đương đầu với những đích sinh học chưa xác thực hoặc chưa có thuốc chữa và cần thiết phải nghiên cứu mở rộng thêm ngoài "Vùng Hóa học" này. Một nguồn hợp chất có thể dùng để nghiên cứu có thể được xây dựng từ các dẫn xuất của hợp chất tự nhiên. Các hợp chất thu được từ vi khuẩn, thực vật, động vật, sinh vật biển thông qua các công nghệ tổng hợp cho các nhà khoa học các tệp cơ sở dữ liệu hợp chất nguồn gốc từ thiên nhiên.

Các cơ sở dữ liệu hợp chất thường được phân phối miễn phí bởi các công ty thương mại hoặc viện nghiên cứu. Chúng bao gồm các loại thuốc, carbohydrate, chất tổng hợp, chất tự nhiên ... (bảng 4). ZINC là một cơ sở dữ liệu miễn phí trên mạng với sức chứa lên tới 13 triệu hợp chất ở phiên bản hiện tại với các thông tin về hoạt tính sinh học đi kèm (khối lượng phân tử, ClogP và số liên kết xoay). Các tệp cơ sở dữ liệu khác như các hợp chất giống thuốc, tiềm năng và các phân mảnh cũng đã được xây dựng.

Bảng 4. Một số cơ sở dữ liệu hợp chất được phân phối bởi các công ty và viện nghiên cứu trên thế giới

Cơ sở dữ liệu	Số hợp chất	Website
ZINC	13 triệu	http://zinc.docking.org
ChemDB	5 triệu	http://cdb.ics.uci.edu
eMolecules	7 triệu	http://www.emolecules.com
ChemSpider	26 triệu	http://www.chemspider.com
PubChem	30 triệu	http://pubchem.ncbi.nlm.nih.gov
ChemBank	1,2 triệu	http://chembank.broadinstitute.org
DrugBank	4.800 thuốc; 2.500 đích sinh học	http://www.drugbank.ca
NCI Open Database	265.000	http://cactus.nci.nih.gov/ncidb2.2/
Chimiothequè Nationale	48.370	http://chimiotheque-nationale.enscm.fr/?lang=fr
Drug Discovery Center Collection	340.000	http://www.drugdiscovery.uc.edu/
ChEMBL	1 triệu	http://www.ebi.ac.uk/chembldb/index.php
WOMBAT	263.000	http://www.sunsetmolecular.com

4. Kết luận

Bài viết tổng quan này đã cung cấp một cái nhìn tổng quan ngắn gọn về tính tiên tiến trong hai phương pháp sàng lọc LBVS và SBVS cũng như trong các kỹ thuật mô phỏng hóa tính toán, nhấn mạnh những tiến bộ và thay đổi to lớn mà sàng lọc ảo (*in silico*) đã đạt được trong những năm qua, khiến nó trở thành một công cụ có giá trị và hiệu quả để tìm kiếm phát triển thuốc mới. Việc sử dụng kết hợp các phương pháp này đã được chứng minh qua các nghiên cứu trên thế giới là đặc biệt hữu ích trong tìm hiểu cơ chế phân tử, tác dụng phụ của thuốc và tái định hướng các dược phẩm để tác dụng mục tiêu qua những đích sinh học mới và điều trị các bệnh khác nhau một cách an toàn.

TÀI LIỆU THAM KHẢO

1. Bohacek, R. S., C. McMartin and W. C. Guida (1996). "The art and practice of structure-based drug design: A molecular modeling perspective". Medicinal Research Reviews 16 (1): 3-50.
2. Case, D. A., T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods (2005). "The amber biomolecular simulation programs". Journal of Computational Chemistry 26 (16): 1668-1688.
3. de Beer, T. A. P., G. A. Wells, P. B. Burger, F. Joubert, E. Marechal, L. Birkholtz and A. I. Louw (2009). "Antimalarial drug discovery: in silico structural biology and rational drug design". Infectious Disorders - Drug Targets 9 (3): 304-318.
4. Jiang, F. and S.-H. Kim (1991). "Soft docking": Matching of molecular surface cubes" Journal of Molecular Biology 219 (1): 79-102.
5. John, W. C. and L. S. Herbert (2005). Catalyzing inquiry at the interface of computing and biology. Washington (DC), National Academies Press (US).
6. Krovat, E., T. Steindl and T. Langer (2005). "Recent advances in docking and scoring". Current Computer Aided-Drug Design 1(1): 93-102.
7. Lavecchia, A. and C. Giovanni (2013). "Virtual screening strategies in drug discovery: a critical review". Current Medicinal Chemistry 20 (23): 2839-2860.
8. Lee, H.-C., J. Salzemann, N. Jacq, H.-Y. Chen, L.-Y. Ho, I. Merelli, L. Milanesi, V. Breton, S. C. Lin and Y.-T. Wu (2006). "Grid-enabled high-throughput in silico screening against influenza a neuraminidase". IEEE Transactions on Nanobioscience 5 4): 288-295.
9. Lin, J.-H., A. L. Perryman, J. R. Schames and J. A. McCammon (2003). "The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme". Biopolymers 68 (1): 47-62.
10. Melville, J., E. Burke and J. Hirst (2009). "Machine learning in virtual screening". Combinatorial Chemistry & High Throughput Screening 12 (4): 332-343.
11. Meng, E. C., B. K. Shoichet and I. D. Kuntz (1992). "Automated docking with grid-based energy evaluation". Journal of Computational Chemistry 13 (4): 505-524.
12. Morris, G. M., D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function". Journal of Computational Chemistry 19 (14): 1639-1662.
13. Mullard, A. (2014). "New drugs cost US$2.6 billion to develop". Nature Reviews Drug Discovery 13 (12): 877-877.
14. Pagadala, N. S., K. Syed and J. Tuszynski (2017). "Software for molecular docking: a review". Biophysical Reviews 9 (2): 91-102.

15. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten (2005). "Scalable molecular dynamics with NAMD". Journal of Computational Chemistry 26 (16): 1781-1802.
16. Reddy, A. S., S. P. Pati, P. P. Kumar, H. N. Pradeep and G. N. Sastry (2007). "Virtual screening in drug discovery - a computational perspective". Current Protein & Peptide Science 8 (4): 329-351.
17. Roy, D., V. Kumar, K. K. Acharya and K. Thirumurugan (2013). "Probing the binding of syzygium-derived α-glucosidase inhibitors with n- and c-terminal human maltase glucoamylase by docking and molecular dynamics simulation". Applied Biochemistry and Biotechnology 172 (1): 102-114.
18. Selzer, P., H.-J. Roth, P. Ertl and A. Schuffenhauer (2005). "Complex molecules: do they add value?" Current Opinion in Chemical Biology 9 (3): 310-316.
19. Song, C. M., S. J. Lim and J. C. Tong (2009). "Recent advances in computer-aided drug design". Briefings in Bioinformatics 10 (5): 579-591.
20. Tame, J. R. H. (1999). "Scoring functions: A view from the bench". Journal of Computer-Aided Molecular Design 13 (2): 99-108.
21. Ton, A. T., F. Gentile, M. Hsing, F. Ban and A. Cherkasov (2020). "Rapid identification of potential inhibitors of *Sars-Cov-2* main protease by deep docking of 1.3 billion compounds". Molecular Informatics 39 (8).
22. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen (2005). "GROMACS: Fast, flexible, and free". Journal of Computational Chemistry 26 (16): 1701-1718.
23. Verma, J., V. Khedkar and E. Coutinho (2010). "3D-QSAR in drug design - a review". Current Topics in Medicinal Chemistry 10 (1): 95-115.

VIRTUAL SCREENING IN DRUG DISCOVERY: A REVIEW

Pham Minh Quan, Le Thi Thuy Huong
Tran Quoc Toan, Pham Thi Hong Minh, Pham Quoc Long

Institute of Natural Products Chemistry, VAST
Graduate University of Science and Technology, VAST

Summary

Nowadays, virtual screening is a powerful technique for identifying hit molecules as starting points for medicinal chemistry. The number of methods and softwares which use the ligand and target-based VS approaches is increasing at a rapid pace. This report will present a brief overview of the progress and application of chemoinformatics and bioinformatics in drug development on the basis of two directions of structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS).