

MÔ HÌNH HÓA TƯƠNG ĐỒNG VÀ ỨNG DỤNG TRONG THIẾT KẾ THUỐC

Lê Anh Vũ^{1,*}, Phan Thị Cẩm Quyên², Nguyễn Thúy Hương¹

¹Trường Đại học Bách Khoa TP. HCM,

²Trung tâm Giống Kiên Giang

TÓM TẮT

Trong nhiều thập kỷ qua, mô hình hóa tương đồng đã trở thành công cụ phổ biến để mô hình hóa cấu trúc ba chiều (3D) giả thuyết của các protein quan tâm. Mục đích của quá trình này là để tạo ra cấu trúc 3D của các protein mà cấu trúc vẫn chưa được xác định bằng các phương pháp thực nghiệm. Cơ sở cho mô hình hóa tương đồng dựa trên sự quan sát rằng các homolog protein với chuỗi acid amin tương tự sẽ có cấu trúc 3D giống nhau. Mô hình hóa tương đồng sử dụng các phương pháp dựa trên máy tính (*in silico*) để tạo ra các mô hình cấu trúc 3D cho protein mục tiêu dựa trên một protein khuôn mẫu. Cách gấp cuộn của mô hình được tạo ra dựa trên sự tương ứng giữa mô-típ cấu trúc của protein khuôn mẫu và mục tiêu. Mô hình hóa tương đồng hiện đang là phương pháp đáng tin cậy nhất bên cạnh các phương pháp thực nghiệm để tạo ra mô hình chất lượng cho nhiều ứng dụng khác nhau trong thiết kế thuốc. Cho đến nay, mô hình hóa tương đồng đã được sử dụng thành công trong xác định phân tử khởi nguồn bằng docking phân tử, đề xuất cơ chế tương tác giữa thụ thể - phối tử, để tạo điều kiện cho các thí nghiệm gây đột biến và để làm cơ sở cho tối ưu hóa các phối tử tiềm năng. Trong bài tổng quan này, chúng tôi báo cáo những phát triển hiện tại trong lĩnh vực này, thảo luận về những hạn chế của mô hình hóa tương đồng và giới thiệu các ứng dụng mới nhất của kỹ thuật này vào quy trình thiết kế thuốc hiện đại.

Từ khóa: công nghệ sinh học; cấu trúc protein; thiết kế thuốc; mô hình hóa tương đồng; thiết kế thuốc dựa trên cấu trúc.

Ngày nhận bài: 17/5/2019; Ngày hoàn thiện: 21/7/2019; Ngày đăng: 27/7/2019

HOMOLOGY MODELING AND ITS APPLICATIONS TO DRUG DESIGN

Le Anh Vu^{1,*}, Phan Thi Cam Quyen², Nguyen Thuy Huong¹

¹Ho Chi Minh City University of Technology,

²KienGiang Seed Research Center

ABSTRACT

In the last decades, homology modeling has become a popular tool to build hypothetical three-dimensional (3D) structures of interested proteins. The aim of this process is to model 3D structure of proteins that have not been structured from experimental methods. The basis for homology modeling is based on the observation that homolog proteins with similar amino acid sequences will have the same 3D structure. Homology modeling uses computer-based methods (*in silico*) to create 3D structures of target protein based on a template protein. The folding pattern of the model is created based on the correspondence between the structural motif of template and target proteins. Homology modeling is currently the most reliable method besides experimental methods to create quality models for many different applications in drug design. Homology modeling so far has been successfully used to identify lead molecule by molecular docking, to propose ligand-receptor interactions, to facilitate mutagenesis experiments, and to guide optimization of potential ligands. In this review, we report current developments in this area, discuss the limitations of homology modeling, and address the latest applications of this technique to the modern drug design.

Keywords: biotechnology; protein structure; drug design; homology modeling; structure based drug design.

Received: 17/5/2019; Revised: 21/7/2019; Published: 27/7/2019

* Corresponding author. Email: lavu68@gmail.com

1. Giới thiệu

Hiện nay, sự phát triển của các kỹ thuật phân tích cấu trúc như tinh thể học tia X và phổ cộng hưởng từ hạt nhân (NMR) đã cải thiện quá trình xác định cấu trúc lập thể (3D) của protein [1]. Tuy nhiên, những kỹ thuật này hiện vẫn khó có thể được áp dụng cho toàn bộ các protein vì nhiều lý do. Cụ thể, kỹ thuật NMR thường chỉ được áp dụng với các protein có kích thước phân tử nhỏ. Trong trường hợp sử dụng tinh thể học tia X, phân tử protein nên được phân tích dưới dạng kết tinh. Bên cạnh đó, một nhược điểm nữa của các kỹ thuật này là tốn nhiều thời gian và chi phí nghiên cứu. Điều này đã dẫn đến khó khăn trong việc xác định cấu trúc protein với độ phân giải cao, đặc biệt là đối với các protein màng, do những khó khăn trong quá trình tinh chế và kết tinh các protein đó so với các protein hòa tan khác [2]. Vì protein màng chiếm tỷ lệ quan trọng trong các mục tiêu thuốc, những tiến bộ trong việc xác định cấu trúc các protein này dự kiến sẽ đẩy nhanh quá trình thiết kế thuốc. Gần đây, việc dự đoán cấu trúc 3D của protein sử dụng các công cụ *in silico* đã được phát triển [3].

Mô hình hóa tương đồng là một trong những phương pháp dự đoán cấu trúc *in silico* được sử dụng để xác định cấu trúc 3D của protein từ trình tự acid amin của nó dựa trên một khuôn mẫu đã biết. Cơ sở cho mô hình hóa tương đồng dựa trên hai luận điểm chính. Đầu tiên, cấu trúc 3D của protein được xác định bởi trình tự acid amin của nó. Thứ hai, cấu trúc của protein được bảo tồn nhiều hơn và sự thay đổi thường xảy ra với tốc độ chậm hơn nhiều so với trình tự acid amin trong quá trình tiến hóa. Kết quả là các trình tự tương đồng thường gấp cuộn thành các cấu trúc tương tự nhau và thậm chí các trình tự có mức liên quan thấp vẫn có thể có cấu trúc 3D tương tự [4]. Do đó, mô hình hóa tương đồng đã được sử dụng để mô phỏng các cấu trúc 3D với độ chính xác cao [1]. Ngoài ra, kỹ thuật này còn có ưu điểm là cần ít thời gian và chi phí thấp hơn so với các phương pháp truyền thống. Điều này đã làm thay đổi cách thức tiến hành docking và thúc đẩy ứng dụng của sàng lọc ảo dựa trên cấu trúc trong quy trình thiết kế thuốc hiện đại [5]. Đã có đề xuất rằng các mô hình được xây dựng trên khuôn mẫu có mức độ tương đồng >50% là đủ chính xác cho mục

đích thiết kế thuốc dựa trên cấu trúc [6]. Vì thuốc thể hiện hoạt tính khi tương tác với các thụ thể trong đó chủ yếu là protein, mô hình hóa tương đồng có nhiều ứng dụng trong quá trình thiết kế thuốc, ví dụ như xác định các tương tác giữa protein và phối tử, từ đó góp phần xác định các ứng viên thuốc tiềm năng [3].

Tổng quan này tóm tắt đặc điểm những bước chính trong quá trình thực hiện mô hình hóa tương đồng. Bên cạnh đó, các công cụ phổ biến được phát triển cho mục đích mô hình hóa trong những năm gần đây cũng được trình bày. Bài viết này cũng cung cấp những đánh giá về các vấn đề có thể gặp trong mô hình hóa và tiềm năng ứng dụng của mô hình hóa trong thiết kế thuốc.

2. Mô hình hóa tương đồng

Các phương pháp mô hình hóa thường được phân loại thành mô hình hóa dựa trên khuôn mẫu (mô hình hóa tương đồng) và mô hình hóa *de novo* [6]. Hiện nay, mô hình hóa tương đồng được coi là chính xác hơn so với mô hình hóa *de novo*, và do đó nó được áp dụng phổ biến hơn trong nghiên cứu cấu trúc protein [7]. Mô hình hóa tương đồng là một phương pháp dự đoán cấu trúc bao gồm nhiều bước và có thể có những thay đổi để phù hợp với từng mục đích nghiên cứu. Các bước đặc trưng của mô hình hóa tương đồng được tóm tắt như trong Hình 1 và chi tiết được đề cập dưới đây.



Hình 1. Các bước cơ bản trong quá trình mô hình hóa cấu trúc 3D của protein [3]

2.1 Xác định khuôn mẫu

Trong bước này, trình tự acid amin của protein mục tiêu được sử dụng để xác định cấu trúc khuôn mẫu có trong các cơ sở dữ liệu (CSDL) như NCBI Reference Sequences (RefSeq) [8], UniProt [9], Protein Data Bank (PDB) [10], Worldwide Protein Data Bank

(wwPDB) [11], Protein Data Bank in Europe (PDBe) [12], Protein Data Bank Japan (PDBj) [13] và một số CSDL với quy mô nhỏ hơn [14]. Hiện đã có nhiều công cụ với nhiều cách tiếp cận khác nhau được phát triển cho việc tìm kiếm khuôn mẫu thích hợp với trình tự mục tiêu. Trong đó, Basic Local Alignment Search Tool (BLAST) [15] là công cụ được sử dụng phổ biến nhất cho mục đích này bằng cách sắp giống cột trình tự mục tiêu với trình tự các protein có sẵn trong CSDL. Bên cạnh đó, một số cách tiếp cận khác được sử dụng trong xác định khuôn mẫu bao gồm “profile-profile alignments” [16] và “Hidden Markov models” [17].

Mức độ tương đồng của trình tự khuôn mẫu so với trình tự mục tiêu có ảnh hưởng lớn trong việc tạo cấu trúc 3D với độ chính xác cao. Tuy nhiên, độ tương đồng trình tự không phải là yếu tố duy nhất quyết định độ chính xác của các cấu trúc được tạo thành. Về giới hạn tương đồng trình tự tối thiểu trong mô hình hóa tương đồng, có nhiều ý kiến về mức độ nhưng ở các mức độ tương đồng lớn hơn 25% thường cho thấy khuôn mẫu và mục tiêu sẽ có cấu trúc 3D tương đồng [18]. Ngoài mức độ tương đồng trình tự, các yếu tố khác được xem xét trong việc chọn một khuôn mẫu đủ điều kiện bao gồm sự tương ứng về kiểu gen giữa trình tự khuôn mẫu và trình tự mục tiêu. Các khuôn mẫu từ cây phát sinh giống nhau hoặc tương ứng với trình tự mục tiêu có thể dẫn đến cấu trúc 3D với độ chính xác cao [3]. Bên cạnh đó, các yếu tố môi trường như pH, loại dung môi và sự tồn tại của phối tử ràng buộc cũng quan trọng trong việc chọn mẫu đủ điều kiện vì chúng có vai trò đảm bảo các điều kiện tối ưu nhất trong việc xây dựng cấu trúc mục tiêu chính xác. Độ phân giải của cấu trúc thử nghiệm đang được xem xét cũng là một yếu tố trong việc chọn lựa khuôn mẫu đủ điều kiện [19].

2.2 Sắp giống cột trình tự và hiệu chỉnh

Sau khi trình tự khuôn mẫu phù hợp nhất được chọn, đôi khi cần thiết phải sắp xếp và hiệu chỉnh chúng. Sự sắp xếp có thể là giữa mục tiêu – khuôn mẫu hoặc khuôn mẫu – khuôn mẫu khi sử dụng nhiều hơn một khuôn mẫu. Lỗi trong sự sắp xếp của các residue gây ra sự dịch chuyển của α carbon. Một khoảng trống residue trong phần xoắn α (α helix) sẽ kích hoạt sự xoay phần còn lại của residue

trong xoắn α . Do đó, sự sắp xếp của các residue theo đúng cách là rất quan trọng trong mô hình hóa tương đồng [20].

2.3 Xây dựng mô hình

Với mỗi khuôn mẫu được căn chỉnh, thông tin chứa trong đó phải được sử dụng để tạo ra mô hình cấu trúc 3D của mục tiêu và được biểu diễn dưới dạng tập hợp tọa độ Cartesian cho mỗi nguyên tử trong cấu trúc protein. Có nhiều phương pháp khác nhau được sử dụng để tạo mô hình 3D cho trình tự mục tiêu dựa trên các khuôn mẫu của nó. Nhìn chung, các phương pháp này có thể được phân loại thành phương pháp tổ hợp phần cứng (rigid-body assembly methods), phương pháp khớp phân đoạn (segment matching methods), phương pháp thỏa mãn hạn chế không gian (satisfactions of spatial restraint methods) và phương pháp tiến hóa nhân tạo (artificial evolution methods). Trong phương pháp tổ hợp phần cứng, mô hình tương đồng được xây dựng dựa vào việc lắp ráp các phân đoạn cấu trúc được bảo tồn. Các phân đoạn này được xác định bằng cách xem xét các cấu trúc có liên quan đã được làm sáng tỏ. Do đó, các protein chưa được làm rõ có thể được mô hình hóa bằng cách xây dựng vùng lõi được bảo tồn và sau đó thay thế các vùng khác nhau từ các protein khác trong tập hợp các cấu trúc đã được làm rõ. Việc ứng dụng phương pháp này khác nhau chủ yếu ở cách chúng xử lý các khu vực không được bảo tồn hoặc thiếu khuôn mẫu. Các vùng biến thiên thường được xây dựng với sự trợ giúp của các thư viện phân đoạn [21]. Phương pháp khớp phân đoạn chia mục tiêu thành một chuỗi các phân đoạn ngắn, mỗi phân đoạn được khớp với mẫu riêng được trích xuất từ CSDL. Do đó, việc căn chỉnh trình tự được thực hiện trên các phân đoạn chứ không phải trên toàn bộ protein. Việc lựa chọn mẫu cho từng phân đoạn dựa trên sự tương đồng về trình tự, so sánh các tọa độ α carbon và dự đoán các xung đột không gian phát sinh từ bán kính van der Waals của các nguyên tử phân kỳ giữa mục tiêu và khuôn mẫu [22]. Hạn chế không gian là phương pháp mô hình hóa tương đồng phổ biến nhất hiện nay dựa trên các tính toán cần thiết để xây dựng cấu trúc 3D từ dữ liệu được tạo ra bởi phổ NMR. Một hoặc nhiều sắp xếp mục tiêu – khuôn mẫu được sử dụng để xây dựng tập hợp các tiêu chí hình học mà sau đó

được chuyển đổi thành các hàm mật độ xác suất cho mỗi hạn chế không gian. Các hạn chế áp dụng cho khoảng cách xương sống protein và góc nhị diện, từ đó làm cơ sở cho quy trình tối ưu hóa vị trí nguyên tử. Quy trình này sử dụng phương pháp tối thiểu hóa năng lượng gradient liên hợp để tinh chỉnh vị trí của tất cả các nguyên tử trong protein [23]. Cuối cùng, phương pháp tiến hóa nhân tạo sử dụng mô phỏng quá trình tiến hóa tự nhiên cho đến khi trình tự khuôn mẫu giống với mục tiêu trình tự. Ví dụ, sự căn chỉnh trình tự - cấu trúc có thể tách ra như là một chuỗi các sự kiện tiến hóa như đột biến, thêm hoặc bớt đoạn. Sau đó, mô hình cấu trúc có thể được xây dựng từ cấu trúc khuôn mẫu bằng cách thay đổi một sự kiện tiến hóa tại một thời điểm nhất định [24].

2.4 Mô hình hóa điểm loop

Các khoảng trống hoặc đoạn chèn vào được gọi là các điểm loop xuất hiện trong trình tự các protein tương đồng. Cấu trúc của các điểm loop không được bảo tồn trong quá trình tiến hóa. Ngay cả khi không có đoạn xóa hoặc chèn, vẫn có thể tìm thấy sự phù hợp cấu dạng điểm loop khác nhau trong trình tự truy vấn và mẫu. Tính đặc hiệu chức năng của protein thường được xác định bởi các điểm loop. Vì vậy, độ chính xác của mô hình điểm loop là một yếu tố quan trọng quyết định giá trị của các mô hình được tạo cho các ứng dụng theo sau. Do các điểm loop cho thấy sự biến đổi cấu trúc cao hơn các chuỗi bên và xoắn, nên việc dự đoán cấu trúc của chúng thường khó khăn hơn [25]. Có hai phương pháp quan trọng được sử dụng trong việc phát triển các điểm loop. Một là phương pháp tìm kiếm cơ sở dữ liệu và hai là phương pháp tìm kiếm cấu dạng. Phương pháp tìm kiếm cơ sở dữ liệu sàng lọc tất cả các cấu trúc protein đã biết để phát hiện các phân đoạn cung cấp các vùng lõi quan trọng [26]. Trong khi đó, phương pháp tìm kiếm cấu dạng phụ thuộc vào tối ưu hóa chức năng cho điểm [27]. Hiện nay, mô hình hóa điểm loop được thực hiện ở mức 4 – 7 residue. Điều này là do sự thay đổi về cấu dạng tăng khi chiều dài của điểm loop tăng lên. Để giải quyết những hạn chế trên, các phương pháp *de novo* được sử dụng cho các dự đoán về hình dạng điểm loop bằng cách tìm kiếm không gian cấu dạng đã được phát triển. Mô phỏng Monte Carlo, mô phỏng annealing, thuật toán di truyền và mô phỏng

động lực phân tử là những ví dụ cho phương pháp này. Trong các phương pháp như vậy, độ dài của điểm loop có thể được mô hình hóa không bị giới hạn nhưng khi độ dài tăng số lượng hình dạng có thể tăng lên nhanh chóng khiến cho việc mô hình hóa rất tốn thời gian [28].

2.5 Mô hình hóa chuỗi bên

Mô hình hóa chuỗi bên thường được thực hiện bằng cách đặt chuỗi bên lên tọa độ xương sống có nguồn gốc từ cấu trúc khuôn mẫu và/hoặc từ mô phỏng ban đầu (*ab initio*). Trong thực tế, dự đoán chuỗi bên chỉ hiệu quả ở mức độ cao của tương đồng trình tự. Chuỗi bên có mặt trong một số cấu trúc hạn chế với năng lượng thấp được gọi là rotamer. Tùy thuộc vào chức năng năng lượng được xác định và chiến lược tìm kiếm, rotamer được chọn theo trình tự protein được ưu tiên và tọa độ xương sống đã cho. Độ chính xác của dự đoán thường cao đối với rotamer của lõi kỵ nước nhưng thấp đối với rotamer trên bề mặt tiếp xúc với nước [29].

2.6 Tối ưu hóa mô hình

Tối ưu hóa mô hình thường bắt đầu bằng việc giảm thiểu năng lượng bằng cách sử dụng các trường lực cơ học phân tử. Ở mỗi lần giảm thiểu năng lượng, một vài lỗi lớn được loại bỏ nhưng nhiều lỗi nhỏ khác được đưa ra cùng lúc và bắt đầu tích lũy [30]. Do đó, hạn chế vị trí nguyên tử, thực hiện giảm thiểu năng lượng và sử dụng các trường lực chính xác hơn như trường lực lượng tử [31] và trường lực tự tham số hóa [32] có thể được sử dụng để giảm sai số trong tối ưu hóa mô hình. Để tối ưu hóa mô hình hơn nữa, các phương pháp như động học phân tử và Monte Carlo có thể được sử dụng [33, 34].

Việc đánh giá mô hình tương đồng mà không tham chiếu đến cấu trúc tự nhiên thường được thực hiện bằng hai phương pháp: thống kê tiềm năng hoặc tính toán năng lượng vật lý. Cả hai phương pháp đều ước tính năng lượng cho mô hình và tiêu chí độc lập là cần thiết để xác định mức chấp nhận được. Nhược điểm của hai phương pháp này là không tương quan tốt với độ chính xác của cấu trúc thực sự, đặc biệt là về các nhóm protein ít được đề cập trong CSDL, chẳng hạn như protein màng [35].

Thống kê tiềm năng là các phương pháp thực nghiệm dựa trên quan sát tần số tiếp xúc

residue - residue giữa các protein có cấu trúc đã biết trong CSDL. Phương pháp này chỉ định một điểm xác suất hoặc năng lượng cho từng tương tác có thể có giữa các acid amin và kết hợp các điểm tương tác theo cặp này thành một điểm duy nhất cho toàn bộ mô hình. Một số phương pháp như vậy cũng có thể tạo ra đánh giá residue-by-residue xác định các khu vực có điểm kém trong mô hình, mặc dù mô hình có thể có điểm số tổng thể hợp lý. Những phương pháp này chú ý nhiều đến lõi kỵ nước và acid amin phân cực tiếp xúc với dung môi [36].

Tính toán năng lượng vật lý nhằm mục đích nắm bắt các tương tác liên nguyên tử chịu trách nhiệm vật lý cho sự ổn định protein trong dung dịch, đặc biệt là lực van der Waals và các tương tác tĩnh điện. Những tính toán này được thực hiện bằng cách sử dụng trường lực cơ học phân tử, do protein thường quá lớn ngay cả đối với các tính toán dựa trên cơ học lượng tử bán thực nghiệm. Việc sử dụng các phương pháp này dựa trên giả thuyết mặt bằng năng lượng của việc gấp cuộn protein, giả định rằng trạng thái tự nhiên của protein là mức tối thiểu năng lượng của nó. Các phương pháp như vậy thường sử dụng phương pháp solvat hóa liên tục, cung cấp liên tục gần đúng dung môi solvat cho một phân tử protein duy nhất mà không cần sự biểu diễn rõ ràng của các phân tử dung môi riêng lẻ [37].

Gần đây, một phương pháp mới hơn để đánh giá mô hình dựa trên các kỹ thuật học máy như mạng lưới thần kinh nhân tạo, có thể được đào tạo để đánh giá trực tiếp cấu trúc hoặc hình thành sự đồng quy giữa nhiều phương pháp dựa trên thống kê và năng lượng. Kết quả sử dụng hồi quy “support vector machine” cho kết quả đánh giá cao hơn so với các phương pháp thống kê, tính toán năng lượng [38].

3. Phần mềm mô hình hóa tương đồng

Trong hai thập kỷ qua, nhiều phần mềm và máy chủ đã được phát triển cho tác vụ mô hình hóa tương đồng một mô hình hoàn chỉnh từ các chuỗi truy vấn (hay còn gọi là phần mềm/máy chủ đa tác vụ) (Bảng 1). Ngoài các phần mềm/máy chủ kể trên, nhiều công cụ được phát triển cho mỗi bước cụ thể trong quy trình mô hình hóa cũng đã được ghi nhận trong các công bố và/hoặc địa chỉ Internet để các nhà nghiên cứu điều chỉnh, cải thiện và xác minh mô hình cho phù hợp với từng trường hợp cụ thể. Một số công cụ đóng vai trò là thành phần của các nền tảng lớn hơn để hình thành các quy trình mô hình hóa tổng thể, tùy thuộc vào lựa chọn của nhà nghiên cứu (Bảng 2). Ngoài ra, các nghiên cứu hướng đến so sánh đặc điểm của các phần mềm/máy chủ có độ chính xác cao cũng đã được báo cáo [39-41].

Bảng 1. Các phần mềm và máy chủ đa tác vụ cho mục đích mô hình hóa.

Phần mềm/ Máy chủ	Địa chỉ truy cập	Tài liệu tham khảo (TLTK)
MODELLER	http://www.salilab.org/modeller/	[42]
I-TASSER	https://zhanglab.ccmb.med.umich.edu/I-TASSER/	[43]
SWISS-MODEL	http://swissmodel.expasy.org/	[44]
Molecular Operating Environment (MOE)	https://www.chemcomp.com/MOEMolecular_Operating_Environment.htm	[45]
PHYRE2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	[46]
HHPRED	http://toolkit.tuebingen.mpg.de/hhpred	[47]
ROBETTA	http://www.robetta.org/	[48]
PRIME	https://www.schrodinger.com/prime	[49]

Bảng 2. Các phần mềm/công cụ dùng cho các bước trong mô hình hóa tương đồng.

Phần mềm/Công cụ	Địa chỉ truy cập
Sắp giống cột và căn chỉnh trình tự	
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
PSI-BLAST	http://www.ncbi.nlm.nih.gov/BLAST/newblast.html
MUSCLE	http://www.drive5.com/muscle/
ClustalW	http://www.ebi.ac.uk/clustalw/
PROBCONS	http://probcons.stanford.edu/
T-Coffee	http://www.tcoffee.org/
PROMALS	http://prodata.swmed.edu/promals/
Mô hình hóa điểm loop	
Swiss-PDB Viewer	http://spdbv.vital-it.ch/
CONGEN	http://www.congenomics.com/congen/doc/index.html
RAMP	http://www.ram.org/computing/ramp/ramp.html
BTPRED	http://www.biochem.ucl.ac.uk/bsm/btpred/
BRAGI	http://bragi.gbf.de/index.html
Mô hình hóa chuỗi bên	
RAMP	http://www.ram.org/computing/ramp/ramp.html
SCWRL	http://www.fccc.edu/research/labs/dunbrack/scwrl
Segmod/CARA	http://www.bioinformatics.ucla.edu/~genemine
SMD	http://condor.urbb.jussieu.fr/Smd.html
Tối ưu hóa và Đánh giá mô hình	
PROCHECK	http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
WHATCHECK	http://www.sander.embl-heidelberg.de/whatcheck/
ProsaII	http://www.came.sbg.ac
VERIFY3D	http://www.doe-mbi.ucla.edu/Services/Verify_3D/
ERRAT	http://www.doe-mbi.ucla.edu/Services/Errat.html
ANOLEA	http://www.fundp.ac.be/pub/ANOLEA.html
Probe	http://kinemage.biochem.duke.edu/software/probe.php

4. Ứng dụng trong thiết kế thuốc

Hiện nay, mô hình hóa tương đồng đã được sử dụng rộng rãi trong lĩnh vực thiết kế thuốc với sự trợ giúp của máy tính, đặc biệt là trong quá trình thiết kế thuốc dựa trên cấu trúc mục tiêu. Tầm quan trọng của mô hình hóa tương đồng ngày càng tăng khi số lượng cấu trúc tinh thể được xác định tăng lên. Một số ứng dụng phổ biến khác của mô hình hóa tương đồng bao gồm: (a) nghiên cứu ảnh hưởng của các đột biến; (b) xác định vị trí hoạt động và liên kết của protein; (c) tìm kiếm các phối tử cho một vị trí bám nhất định (khai thác cơ sở dữ liệu); (d) thiết kế các phối tử mới cho một vị trí bám nhất định; (e) mô hình hóa tính đặc hiệu cơ chất; (f) dự đoán cấu trúc epitope; (g) mô phỏng docking protein – protein và (h) thay thế phân tử trong tinh thể cấu trúc tia X [4]. Các ứng dụng điển hình của mô hình hóa tương đồng trong thiết kế thuốc đòi hỏi độ chính xác rất cao vị trí chuỗi bên tại điểm gắn kết. Một số lượng lớn các mô hình tương đồng đã được xây dựng trong những năm qua

bao gồm các kháng thể và nhiều protein liên quan đến y sinh học [6].

Trong một ví dụ điển hình, mô hình tương đồng đã được sử dụng để thiết kế các chất ức chế bơm NorA efflux ở vi khuẩn *Staphylococcus aureus*. Một số nghiên cứu đã chứng minh rằng tình trạng kháng thuốc ở các chủng *S. aureus* trên lâm sàng có liên quan đến sự biểu hiện quá mức của bơm NorA efflux. Tuy nhiên, cấu trúc 3D của *S. aureus* NorA vẫn chưa được xác định rõ. Do đó, phương pháp mô hình hóa tương đồng được thực hiện dựa trên cấu trúc tinh thể của chất vận chuyển glycerol-3-phosphate (PDB ID: 1PW4) để xây dựng cấu trúc 3D của *S. aureus* NorA. Thông qua docking và tìm kiếm tương đồng, 14 phân tử khởi nguồn mới đã được xác định, trong đó các tính chất được lý và đánh giá rủi ro độc tính cũng đã được thực hiện. Do đó, các phân tử khởi nguồn này sẽ hữu ích trong việc thiết kế và tổng hợp thuốc ức chế bơm NorA efflux mới nhằm khôi phục tính nhạy cảm của các hợp chất thuốc [50]. Trong một nghiên cứu khác, Thái Khắc Minh và

cộng sự [51] cũng đã sử dụng mô hình cấu trúc bơm NorA efflux để sàng lọc các hoạt chất tự nhiên nhằm tìm kiếm các chất ức chế bơm NorA tiềm năng. Cũng trên đối tượng *S. aureus*, Lê Anh Vũ và Nguyễn Thúy Hương [52] đã sử dụng khuôn mẫu protein SarR để mô hình hóa protein SarA, một enzyme quan trọng trong quá trình sản xuất màng sinh học của *S. aureus*, từ đó sàng lọc các chất ức chế tiềm năng dựa trên cấu trúc của protein SarA.

Trong những năm gần đây, cấu trúc 3D của các mục tiêu trong bệnh ung thư được tạo ra bằng mô hình tương đồng có thể được sử dụng để thiết kế các tác nhân hóa trị liệu hiệu quả [53]. Các cấu trúc 3D đáng tin cậy của các thụ thể kết hợp G-protein (G-protein-coupled receptors) mà là mục tiêu của gần một phần ba các loại thuốc được FDA chấp thuận đã được xây dựng bằng mô hình hóa tương đồng [54]. Gần đây, Armando et al. [55] đã tiến hành một nghiên cứu để phát triển các chất ức chế tiêu đơn vị dyskerin (DKC1) của telomerase, vì các tế bào khối u có tiềm năng sao chép không giới hạn chủ yếu là do holoenzyme telomerase. Mô hình hóa cấu trúc 3D được thực hiện bởi máy chủ I-TASSER. Theo dự đoán về cấu trúc tương đồng, protein có PDB ID: 3UAI (cấu trúc tinh thể của phức hợp Shq1-Cbf5-Nop10-Gar1 từ *Sacharomyces cerevisiae*) đã được sử dụng

làm khuôn mẫu. Chất lượng hóa học lập thể của các mô hình thu được được đánh giá bằng phần mềm PROCHECK. Mô hình DKC1 đã được sàng lọc dựa trên thư viện gồm 450.000 phân tử “giống như thuốc”. Sau đó, 10 phân tử cho thấy giá trị ái lực cao nhất đã được chọn để kiểm tra hoạt tính ức chế của chúng trên dòng tế bào MDA MB 231 (Monroe Dunaway Anderson Metastasis Breast cancer 231), từ đó thu được ba hợp chất cho thấy hoạt tính ức chế. Trong một ví dụ khác, Trần Thành Đạo và cộng sự [56] đã nghiên cứu khả năng gắn kết giữa aromatase với các chất ức chế aromatase bằng mô hình docking phân tử. Aromatase là enzyme quan trọng trong quá trình chuyển hóa androgen thành estrogen, yếu tố góp phần cho sự phát triển của tế bào ung thư vú. Nhóm tác giả đã xây dựng mô hình để dự đoán mối liên hệ giữa cấu trúc và tác động chất ức chế aromatase - ứng dụng mô hình sàng lọc ảo trên ngân hàng cơ sở dữ liệu để giải thích cơ chế tác động toàn diện của nhóm ức chế aromatase.

Một số ứng dụng khác của mô hình hóa tương đồng là xác định cấu trúc 3D RNA polymerase của virus Ebola và cấu trúc 3D protein NS5 của virus Zika, từ đó thiết kế các chất ức chế tiềm năng [57, 58]. Các ứng dụng gần đây của mô hình hóa tương đồng trong thiết kế thuốc được tóm tắt trong Bảng 3.

Bảng 3. Một số nghiên cứu ứng dụng mô hình protein trong thiết kế thuốc.

Mô hình protein	Ứng dụng	Công cụ sử dụng	TLTK
Bơm NorA efflux	Thiết kế nhóm thuốc ức chế bơm NorA efflux để phục hồi tính miễn cảm với kháng sinh ở <i>S. aureus</i>	BLAST, CLUSTALX, MODELLER, PROCHECK, WHATIF, VERIFY3D	[50]
Bơm NorA efflux	Sàng lọc các hoạt chất tự nhiên nhằm tìm kiếm các chất ức chế bơm NorA	MODELLER, PROCHECK, AUTODOCK	[51]
Enzyme SarA	Sàng lọc các chất ức chế enzyme SarA nhằm tìm kiếm các chất ức chế sự hình thành màng sinh học	SWISS-MODEL, AUTODOCK	[52]
Dyskerin pseudouridine synthase (DKC1)	Phát triển các chất ức chế telomerase, chọn sự tương tác giữa RNA template - DKC1 làm mục tiêu	I-TASSER, PROCHECK	[55]
Enzyme aromatase	Giải thích cơ chế tác động toàn diện của nhóm ức chế aromatase	SWISS-MODEL, MOE	[56]
RNA polymerase ở virus Ebola	Thiết kế thuốc ức chế virus Ebola, các loại thuốc có thể được tái sử dụng để chống lại virus Ebola cũng như nghiên cứu về phương thức hoạt động của virus Ebola	BLAST, SWISS-MODEL, PROCHECK	[57]
Protein Ns5 ở virus Zika	Phát hiện hai chất ức chế tiềm năng của ZIKV Methyltransferase và RNA polymerase phụ thuộc RNA	BLAST, CLUSTALX, MODELLER, PROCHECK	[58]

Acetohydroxy acid synthase (AHAS)	Thiết kế các chất ức chế mới chống lại cơ chế gây bệnh của vi khuẩn lao <i>Mycobacterium tuberculosis</i>	BLAST, MODELLER, PROCHECK	[59]
CD20 antigen	Nghiên cứu về cấu trúc của kháng nguyên CD20, là mục tiêu phát triển các kháng thể đơn dòng mới	PSI-BLAST, T-COFFEE, SWISS-MODEL, I-TASSER, PHYRE2, MUSTER, RAMPAGE	[60]
GABA transporter 1 (GAT1)	Thiết kế các chất ức chế GAT1 nhằm phát triển thuốc chống co giật và thuốc chống trầm cảm	CLUSTALW, PRIME, GLIDE XP, SCHRODINGER	[61]
Hsp70	Xác định cấu trúc 3D của hsp70 chaperone protein sử dụng làm mục tiêu phổ rộng mới trong trị liệu ung thư	BLAST, SWISS-MODEL, QMEAN, PSVS	[62]
Parkinson's linked mutant leucine-rich repeat kinase 2 (LRRK2)	Xác định điểm mới trong con đường báo hiệu cái chết của tế bào thần kinh	MOE, GLIDE 1, MAESTRO, CHARMM	[63]
Peroxisome proliferator-activated receptor gamma (PPAR γ)	Xác định các phối tử mới làm giảm thụ thể PPAR γ trong các biến chứng tiểu đường tuýp 2	PRIME, GLIDE XP, SCHRODINGER	[64]
Ribonucleotide reductase ở vi khuẩn <i>Mycobacterium leprae</i>	Sàng lọc các loại thuốc mới trong điều trị bệnh phong kháng thuốc	SWISS-MODEL, HHPRED, PROFUNC, ERRAT, WHATIF, PROSA, GLIDE XP, SCHRODINGER	[65]
Thụ thể histamine H ₂	Phát triển thuốc mới chống loét dạ dày bằng cách nhắm mục tiêu thụ thể histamine H ₂	BLAST, CLUSTALX, MODELLER, PROCHECK, AUTODOCK, STRING	[66]
Thụ thể tuýp 1 của enzyme angiotensin II	Thiết kế các tác nhân đối kháng thụ thể angiotensin	BLAST, CLUSTALW, SYBYL, MODELLER, I-TASSER, PROCHECK, SURFLEXDOCK	[67]
α -glucosidase	Thiết kế các nhóm thuốc ức chế α -glucosidase mới	BLAST, PRIME, PROCHECK, SITEMAP, GLIDE XP, SCHRODINGER, MAESTRO	[68]
Protease tái tổ hợp của HIV-1 (HIV-1PrHis)	Sàng lọc các chất ức chế của enzyme HIV-1PrHis	SWISS-MODEL, MODELLER, PROCHECK	[69]

5. Giới hạn của phương pháp

Số lượng cấu trúc 3D protein chất lượng cao được xác định đã tăng lên trong những thập kỷ qua. Việc áp dụng các phương pháp thí nghiệm mới như kính hiển vi điện tử Cryo (Cryo-Electron Microscopy) sẽ làm tăng số lượng cấu trúc 3D được xác định bằng thực nghiệm [70]. Tuy nhiên đến nay cấu trúc 3D của tất cả các gấp cuộn của protein trong tự nhiên vẫn chưa được biết đến. Điều này dẫn đến khó khăn trong việc xây dựng cấu trúc 3D của protein khi cấu trúc của họ protein chưa được xác định.

Có nhiều phương pháp được áp dụng để xây dựng mô hình trong mô hình hóa tương đồng. Bên cạnh các phương pháp đã được áp dụng từ lâu thì một số phương pháp với các thuật toán mới đã được phát triển. Nhiều nghiên cứu đã chỉ ra rằng không có chương trình mô hình hóa nào vượt trội về mọi đặc tính so với các chương trình khác [18]. Vì vậy, việc lựa chọn phương pháp được sử dụng tùy theo đặc điểm của protein quan tâm và mục đích ứng dụng trong tương lai của mô hình là rất quan trọng. Hiện nay mô hình được xây dựng chủ yếu dựa trên sự tương đồng trình tự. Trong

xác định cấu trúc thực nghiệm, phối tử ít được xem xét vì chúng thường bị mất trong quá trình tinh chế. Thiếu sót này đã được xử lý với việc giới thiệu các hướng tiếp cận đề cập đến trạng thái phối tử. Tuy nhiên, cách tiếp cận như vậy cần chuyên môn và can thiệp thủ công mất thời gian. Do đó, việc giới thiệu các công cụ mô hình tương đồng hoàn toàn tự động có thể giải quyết các vấn đề như vậy là một vấn đề quan trọng [6].

Một hạn chế khác của mô hình tương đồng là sự hiện diện của các điểm loop và rotamer, vì rất khó để mô hình hóa chúng mà không có dữ liệu mẫu. Để có một mô hình với độ chính xác cao, tối ưu hóa vùng điểm loop và chuỗi bên là rất quan trọng. Trong trường hợp có mức độ tương tự trình tự thấp giữa mục tiêu và mẫu, việc sử dụng nhiều mẫu là thuận lợi. Nhưng việc sử dụng nhiều mẫu có thể dẫn đến quang sai trong căn chỉnh trừ khi các mẫu từ cây phát sinh giống hệt hoặc tương tự được sử dụng làm chuỗi mục tiêu. Ngoài ra, vào cuối quá trình mô hình hóa tương đồng, nhiều mô hình của một mục tiêu được xây dựng nói chung. Có nhiều mô hình được tạo ra là tín hiệu tốt, nhưng việc xác định mô hình tốt nhất cần được nghiên cứu thêm. Để xác định mô hình tốt nhất, các mô hình được xây dựng được so sánh bằng cách sử dụng các tham số khác nhau, chẳng hạn như điểm số năng lượng protein được tối ưu hóa riêng biệt (discrete optimized protein energy), điểm số mô hình mẫu (template modeling) và giá trị root-mean-square deviation (RMSD) được sử dụng để so sánh [3]. Tham số xác định được quyết định tùy thuộc vào mục đích của kết quả mô hình hóa.

6. Kết luận

Tóm lại, để bổ sung cho các phương pháp thực nghiệm vốn tốn nhiều thời gian, chi phí và nhân lực, các phương pháp mô hình hóa cấu trúc protein sử dụng công cụ *in silico* với khả năng dự đoán cấu trúc 3D đáng tin cậy sẽ được tiếp tục ứng dụng trong những năm tiếp theo. Khi độ chính xác của các mô hình được tạo ra tăng lên, ứng dụng của chúng trong quá trình thiết kế thuốc với sự hỗ trợ của máy tính cũng tăng theo. Như đã đề cập, các mô hình thu được bằng kỹ thuật này đã đóng góp không chỉ trực tiếp cho quá trình thiết kế thuốc mà còn cung cấp kiến thức cho nhiều lĩnh vực khác, từ đó góp phần vào việc thiết

kế thuốc. Có thể kể đến một số lĩnh vực như: xác định vị trí tác động và chức năng protein; đánh giá vai trò sinh học của các đột biến ở vị trí gắn kết; giải thích các chế độ liên kết; tối ưu hóa hợp chất khởi nguồn; sàng lọc ảo dựa trên cấu trúc, docking phân tử.

Những tiến bộ trong sinh học cấu trúc thu được bằng cách sử dụng các mô hình tương đồng đã chứng minh độ tin cậy của các phần mềm hiện có. Sự phát triển gần đây của các phần mềm này với những cải tiến về thuật toán căn chỉnh, mô hình hóa điểm loop và chuỗi bên, phát hiện lỗi và xác nhận mô hình đã giúp cải thiện độ chính xác của mô hình tạo ra. Ngày nay, với các phần mềm và khuôn mẫu thích hợp, các mô hình lý thuyết có thể được xây dựng với độ chính xác gần với các mô hình thu được bằng các phương pháp thực nghiệm. Những mô hình này đã đóng góp hiệu quả, và dự kiến sẽ tiếp tục đóng vai trò quan trọng trong quá trình thiết kế thuốc hiện đại.

Lời cảm ơn

Nghiên cứu được tài trợ bởi Trường Đại học Bách khoa – Đại học Quốc gia TP. HCM trong khuôn khổ đề tài mã số TNCS-KTHH-2017-12.

TÀI LIỆU THAM KHẢO

- [1]. H. Deng, Y. Jia, and Y. Zhang, "Protein structure prediction", *International Journal of Modern Physics B*, Vol. 32, No. 18, pp. 1840009 (17 pages), 2017.
- [2]. J. G. Almeida, A. J. Preto, P. I. Koukos, A. M. Bonvin, and I. S. Moreira, "Membrane proteins structures: A review on computational modeling tools", *Biochimica et Biophysica Acta – Biomembranes*, Vol. 1859, No. 10, pp. 2021-2039, 2017.
- [3]. M. T. Muhammed and E. Aki- Yalcin, "Homology modeling in drug discovery: Overview, current applications, and future perspectives", *Chemical Biology & Drug Design*, Vol. 93, pp. 12-20, 2019.
- [4]. V. K. Vyas, R. D. Ukawala, M. Ghate, and C. Chintha, "Homology modeling a fast tool for drug discovery: current perspectives", *Indian Journal of Pharmaceutical Sciences*, Vol. 74, No. 1, pp. 1-17, 2012.
- [5]. T. C. França, "Homology modeling: an important tool for the drug discovery", *Journal of Biomolecular Structure and Dynamics*, Vol. 33, No. 8, pp. 1780-1793, 2015.

- [6]. T. Schmidt, A. Bergner, and T. Schwede, "Modelling three-dimensional protein structures for applications in drug design", *Drug Discovery Today*, Vol. 19, No. 7, pp. 890-897, 2014.
- [7]. S. D. Lam, S. Das, I. Sillitoe, and C. Orengo, "An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences", *Acta Crystallographica Section D Structural Biology*, Vol. 73, No. 8, pp. 628-640, 2017.
- [8]. K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins", *Nucleic Acids Research*, Vol. 35 (Database issue), pp. D61-D65, 2006.
- [9]. The UniProt Consortium, "UniProt: the universal protein knowledgebase", *Nucleic Acids Research*, Vol. 46, No. 5, pp. 2699, 2018.
- [10]. S. K. Burley, H. M. Berman, C. Bhikadiya, et al., "RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy", *Nucleic Acids Research*, Vol. 47, No. D1, pp. D464-D474, 2019.
- [11]. H. Berman, K. Henrick, H. Nakamura, and J. L. Markley, "The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data", *Nucleic Acids Research*, Vol. 35 (Database issue), pp. D301-D303, 2006.
- [12]. S. Velankar, Y. Alhroub, A. Alili, et al., "PDBe: Protein Data Bank in Europe", *Nucleic Acids Research*, Vol. 39 (Database issue), pp. D402-D410, 2011.
- [13]. A. R. Kinjo, H. Suzuki, R. Yamashita, et al., "Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format", *Nucleic Acids Research*, Vol. 40 (Database issue), pp. D453-D460, 2011.
- [14]. D. Xu and Y. Xu, "Protein databases on the internet", *Current Protocols in Molecular Biology*, Chapter 19, Unit 19.4, 2004.
- [15]. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool", *Journal of Molecular Biology*, Vol. 215, No. 3, pp. 403-410, 1990.
- [16]. G. Wang and R. L. Dunbrack, "Scoring profile-to-profile sequence alignments", *Protein Science*, Vol. 13, No. 6, pp. 1612-1626, 2004.
- [17]. J. Söding, "Protein homology detection by HMM-HMM comparison", *Bioinformatics*, Vol. 21, No. 7, pp. 951-960, 2005.
- [18]. Z. Xiang, "Advances in homology protein structure modeling", *Current Protein & Peptide Science*, Vol. 7, No. 3, pp. 217-227, 2006.
- [19]. T. Schwede, "Protein modeling: what happened to the protein structure gap?", *Structure*, Vol. 21, No. 9, pp. 1531-1540, 2013.
- [20]. S. Andrea and W. Hans-Joachim, "Sequence alignment and homology modelling", In: *Modelling of GPCRs: A practical handbook*, Springer, 2013.
- [21]. A. Szilagyi and Y. Zhang, "Template-based structure modeling of protein-protein interactions", *Current Opinion in Structural Biology*, Vol. 24, pp. 10-23, 2013.
- [22]. M. Levitt, "Accurate modeling of protein conformation by automatic segment matching", *Journal of Molecular Biology*, Vol. 226, No. 2, pp. 507-533, 1992.
- [23]. A. Šali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints", *Journal of Molecular Biology*, Vol. 234, No. 3, pp. 779-815, 1993.
- [24]. P. R. Daga, R. Y. Patel, and R. J. Doerksen, "Template-based protein modeling: recent methodological advances", *Current Topics in Medicinal Chemistry*, Vol. 10, No. 1, pp. 84-94, 2010.
- [25]. A. Fiser, R. K. Do, and A. Sali, "Modeling of loops in protein structures", *Protein Science*, Vol. 9, No. 9, pp. 1753-1773, 2000.
- [26]. N. Fernandez-Fuentes, B. Oliva, and A. Fiser, "A supersecondary structure library and search algorithm for modeling loops in protein structures", *Nucleic Acids Research*, Vol. 34, No. 7, pp. 2085-2097, 2006.
- [27]. Y. Li, "Conformational sampling in template-free protein loop structure modeling: an overview", *Computational and Structural Biotechnology Journal*, Vol. 5, e201302003, 2013.
- [28]. M. Jamroz and A. Kolinski, "Modeling of loops in proteins: a multi-method approach", *BMC Structural Biology*, Vol. 10, No. 5, 9 pages, 2010.
- [29]. S. Liang and N. V. Grishin, "Side-chain modeling with an optimized scoring function", *Protein Science*, Vol. 11, No. 2, pp. 322-333, 2002.
- [30]. K. Joo, J. Lee, and J. Lee, "Methods for accurate homology modeling by global optimization", In: *Homology modeling. Methods in Molecular Biology (Methods and Protocols)*, A. Orry and R. Abagyan (eds), Vol. 857, Humana Press, 2011.
- [31]. H. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hermans, and W. Yang, "Quantum mechanics simulation of protein dynamics on long timescale", *Proteins*, Vol. 44, No. 4, pp. 484-489, 2001.

- [32]. E. Krieger, G. Koraimann, and G. Vriend, "Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field", *Proteins*, Vol. 47, No. 3, pp. 393-402, 2002.
- [33]. H. Fan and A. E. Mark, "Refinement of homology-based protein structures by molecular dynamics simulation techniques", *Protein Science*, Vol. 13, No. 1, pp. 211-220, 2004.
- [34]. C. Guda, E. D Scheeff, P. E. Bourne, and I. N. Shindyalov, "A new algorithm for the alignment of multiple protein structures using Monte Carlo optimization", *Pacific Symposium on Biocomputing*, pp. 275-286, 2001.
- [35]. A. Kryshchuk and K. Fidelis, "Protein structure prediction and model quality assessment", *Drug Discovery Today*, Vol. 14, No. 7-8, pp. 386-393, 2009.
- [36]. M. Y. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures", *Protein Science*, Vol. 15, No. 11, pp. 2507-2524, 2006.
- [37]. S. Crivelli, E. Eskow, B. Bader, et al., "A physical approach to protein structure prediction", *Biophysical Journal*, Vol. 82, No. 1, pp. 36-49, 2002.
- [38]. J. Cheng, A. N. Tegge, and P. Baldi, "Machine learning methods for protein structure prediction", *IEEE Reviews in Biomedical Engineering*, Vol. 1, pp. 41-49, 2008.
- [39]. A. Nayeem, D. Sitkoff and S. Krystek, "A comparative study of available software for high-accuracy homology modeling: from sequence alignments to structural models", *Protein Science*, Vol. 15, No. 4, pp. 808-824, 2006.
- [40]. A. Saxena, R. S. Sangwan, and S. Mishra, "Fundamentals of homology modeling steps and comparison among important bioinformatics tools: an overview", *Science International*, Vol. 1, pp. 237-252, 2013.
- [41]. H. J. Hasani and K. Barakat, "Homology modeling: an overview of fundamentals and tools", *International Review on Modelling and Simulations (IREMOS)*, Vol. 10, No. 2, pp. 129-145, 2017.
- [42]. A. Webb and A. Sali, "Comparative protein structure modeling using MODELLER", *Current Protocols in Bioinformatics*, Vol. 54, pp. 5.6.1-5.6.37, 2016.
- [43]. J. Yang and Y. Zhang, "Protein structure and function prediction using I-TASSER", *Current Protocols in Bioinformatics*, Vol. 52, pp. 5.8.1-5.8.15, 2015.
- [44]. T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, "SWISS-MODEL: An automated protein homology-modeling server", *Nucleic Acids Research*, Vol. 31, No. 13, pp. 3381-3385, 2003.
- [45]. Molecular Operating Environment (MOE) Montreal, QC: Chemical Computing Group Inc; 2013.
- [46]. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis", *Nature Protocols*, Vol. 10, No. 6, pp. 845-858, 2015.
- [47]. J. Söding, A. Biegert, and A. N. Lupas, "The HHpred interactive server for protein homology detection and structure prediction", *Nucleic Acids Research*, Vol. 33 (Web Server issue), pp. W244-W248, 2005.
- [48]. D. E. Kim, D. Chivian, and D. Baker, "Protein structure prediction and analysis using the Robetta server", *Nucleic Acids Research*, Vol. 32 (Web Server issue), pp. W526-W531, 2004.
- [49]. Schrödinger, Prime. 'Version 3.5.'. New York: LLC; 2014.
- [50]. A. V. Bhaskar, T. M. Babu, N. V. Reddy, and W. Rajendra, "Homology modeling, molecular dynamics, and virtual screening of NorA efflux pump inhibitors of *Staphylococcus aureus*", *Drug Design, Development and Therapy*, Vol. 10, pp. 3237-3252, 2016.
- [51]. Khac-Minh Thai, Trieu-Du Ngo, Thien-Vy Phan, Thanh-Dao Tran, Ngoc-Vinh Nguyen, Thien-Hai Nguyen and Minh-Tri Le, "Virtual Screening for Novel *Staphylococcus aureus* NorA Efflux Pump Inhibitors from Natural Products", *Medicinal Chemistry*, Vol. 11, No. 2, pp. 135-155, 2015.
- [52]. Anh-Vu Le and Thuy-Huong Nguyen, "Structure Based Drug Design of Inhibitors for *Staphylococcus aureus* Biofilm", *International Journal of Modern Engineering Research*, Vol. 5, No. 9, pp. 10-17, 2015.
- [53]. A. N. Cavasotto, S. S. Phatak, "Homology modeling in drug discovery: current trends and applications", *Drug Discovery Today*, Vol. 14, No. 13-14, pp. 676-683, 2009.
- [54]. A. Kufareva, M. Rueda, V. Katritch, R. C. Stevens, and R. Abagyan, "Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment", *Structure*, Vol. 19, No. 8, pp. 1108-1026, 2011.
- [55]. R. G. Armando, D. L. Mengual Gómez, E. I. Juritz, P. Lorenzano Menna, and D. E. Gomez, "Homology model and docking-based virtual screening for ligands of human dyskerin as new inhibitors of telomerase for cancer treatment", *International Journal of Molecular Sciences*, Vol. 19, No. 10, pp. 3216, 2018.

- [56]. Trần Thành Đạo, Lê Minh Trí, Thái Khắc Minh, “Nghiên cứu mô hình QSAR và mô tả phân tử docking các chất ức chế aromatase”, *Tạp chí Dược học*, T. 59, S. 3, trang 43-47, 2019.
- [57]. M. Balmith, M. Faya, and M. E. Soliman, “Ebola virus: A gap in drug design and discovery - experimental and computational perspective”, *Chemical Biology & Drug Design*, Vol. 89, No. 3, pp. 297-308, 2017.
- [58]. P. Ramharack, and M. E. S. Soliman, “Zika virus NS5 protein potential inhibitors: an enhanced in silico approach in drug discovery”, *Journal of Biomolecular Structure and Dynamics*, Vol. 36, No. 5, pp. 1118-1133, 2018.
- [59]. K. Gokhale and B. Tilak, “Mechanisms of bacterial acetohydroxyacid synthase (AHAS) and specific inhibitors of Mycobacterium tuberculosis AHAS as potential drug candidates against tuberculosis”, *Current Drug Targets*, Vol. 16, No. 7, pp. 689-699, 2015.
- [60]. Z. Payandeh, M. Rajabibazl, Y. Mortazavi, A. Rahimpour, and A. H. Taromchi, “Ofatumumab monoclonal antibody affinity maturation through in silico modeling”, *Iranian Biomedical Journal*, Vol. 22, No. 3, pp. 180-192, 2018.
- [61]. R. B. Singh, G. K. Singh, K. Chaturvedi, et al., “Design, synthesis, characterization, and molecular modeling studies of novel oxadiazole derivatives of nipecotic acid as potential anticonvulsant and antidepressant agents”, *Medicinal Chemistry Research*, Vol. 27, No. 1, pp. 137-152, 2018.
- [62]. S. Mishra and V. Gomase, “Computational comparative homology based 3D-structure modelling of the HSp70 Protein from GWD”, *Journal of Health and Medical Informatics*, Vol. 7, No. 3, pp. 233-239, 2016.
- [63]. N. Antoniou, D. Vlachakis, A. Memou, et al., “A motif within the armadillo repeat of Parkinson's-linked LRRK2 interacts with FADD to hijack the extrinsic death pathway”, *Scientific Reports*, Vol. 8, No. 1, pp. 3455, 2018.
- [64]. S. Prabhu, S. Vijayakumar, P. Manogar, G. P. Maniam, and N. Govindan, “Homology modeling and molecular docking studies on Type II diabetes complications reduced PPAR γ receptor with various ligand molecules”, *Biomedicine & Pharmacotherapy*, Vol. 92, pp. 528-535, 2017.
- [65]. P. S. Mohanty, A. K. Bansal, F. Naaz, U. D. Gupta, V. D. Dwivedi, and U. Yadava, “Ribonucleotide reductase as a drug target against drug resistance Mycobacterium leprae: A molecular docking study”, *Infection, Genetics and Evolution*, Vol. 60, pp. 58-65, 2018.
- [66]. V. Singh, N. Gohil, and R. Ramírez- García, “New insight into the control of peptic ulcer by targeting the histamine H₂ receptor”, *Journal of Cellular Biochemistry*, Vol. 119, No. 2, pp. 2003-2011, 2018.
- [67]. V. K. Vyas, M. Ghate, K. Patel, G. Qureshi, and S. Shah, “Homology modeling, binding site identification and docking study of human angiotensin II type I (Ang II-AT1) receptor”, *Biomedicine & Pharmacotherapy*, Vol. 74, pp. 42-48, 2015.
- [68]. N. C. Jadhav, A. R. Pahlkar, N. V. Desai, et al., “Design, synthesis and molecular docking study of novel pyrrole-based α -amylase and α -glucosidase inhibitors”, *Medicinal Chemistry Research*, Vol. 26, No. 10, pp. 2675-2691, 2017.
- [69]. Nguyễn Thị Hồng Loan, Nguyễn Văn Sáng, Trịnh Hồng Thái, Phan Tuấn Nghĩa, Bùi Phương Thuận, “Một số đặc điểm cấu trúc của protease HIV-1 tái tổ hợp”, *Tạp chí Sinh học*, T. 34, S. 4, trang 520 -527, 2016.
- [70]. T. R. D. Costa, A. Ignatiou, and E. V. Orlova, “Structural analysis of protein complexes by cryo electron microscopy”. In: *Bacterial Protein Secretion Systems. Methods in Molecular Biology*, L. Journet and E. Cascales (eds), Vol. 1615, Humana Press, New York, 2017.