

Mô hình hóa cấu trúc và mô phỏng động lực học phân tử trong nghiên cứu tương tác protein-protein

Nguyễn Văn Minh Thường^{1,2}, Lý Cẩm Tú^{1,2}, Đinh Thuận Thiên^{1,2}, Trần Văn Hiếu^{1,2,*}

TÓM TẮT

Protein là đại phân tử sinh học quan trọng cấu thành nên hệ thống sống. Để thực hiện chức năng, các protein không chỉ hoạt động độc lập mà còn hình thành một mạng lưới chặt chẽ thông qua tương tác protein-protein (protein-protein interactions – PPIs), yếu tố mang tính quyết định đối với hầu hết các quá trình sinh học trong cơ thể sinh vật. Tuy nhiên, quá trình nghiên cứu gặp nhiều trở ngại đến từ sự thiếu hụt cấu trúc thực nghiệm của phức hợp hoặc của các protein tham gia vào sự tương tác. Với sự hỗ trợ mạnh mẽ của các công cụ tin sinh học, các phương pháp và thuật toán dự đoán tương tác protein nổi lên như một giải pháp tiềm năng để giải quyết những khó khăn trên. Trong số các phương pháp dự đoán cấu trúc tương tác protein-protein *in silico*, docking phân tử (molecular docking) và mô phỏng động lực học phân tử (molecular dynamics simulation – MDs) là hai phương pháp được sử dụng phổ biến bởi khả năng mô hình hóa cấu trúc và mô phỏng xấp xỉ quá trình sinh học, phù hợp với trạng thái tự nhiên của hệ trong thực tế. Việc ứng dụng những tiến bộ tin sinh học này đã giúp rút ngắn thời gian, công sức thực hiện và kinh phí nghiên cứu cho các nhà khoa học. Bài tổng quan này cung cấp thông tin về phương pháp dự đoán cấu trúc phức hợp protein-protein *in silico*, đặc biệt là phương pháp mô phỏng động lực học phân tử, tập trung vào khía cạnh trường lực, các gói chương mô phỏng đang được sử dụng rộng rãi và ứng dụng trong nghiên cứu phát triển thuốc.

Từ khóa: docking phân tử, dự đoán tương tác protein, mô phỏng động lực học phân tử, TBM, TFM, tin sinh học

¹Khoa Sinh học - Công nghệ Sinh học, Trường Đại học Khoa học Tự nhiên, TP. Hồ Chí Minh,

²Đại học Quốc gia TP. Hồ Chí Minh, Việt Nam

Liên hệ

Trần Văn Hiếu, Khoa Sinh học - Công nghệ Sinh học, Trường Đại học Khoa học Tự nhiên, TP. Hồ Chí Minh,

Đại học Quốc gia TP. Hồ Chí Minh, Việt Nam

Email: tvhieu@hcmus.edu.vn

Lịch sử

- Ngày nhận: 10-8-2023
- Ngày sửa đổi: 08-11-2023
- Ngày chấp nhận: 28-12-2024
- Ngày đăng: 31-12-2024

DOI:

<https://doi.org/10.32508/stdjns.v8i4.1301>



Bản quyền

© ĐHQG TP.HCM. Đây là bài báo công bố mở được phát hành theo các điều khoản của the Creative Commons Attribution 4.0 International license.



MỞ ĐẦU

Protein là đại phân tử sinh học quan trọng, đóng vai trò kiến tạo nên hệ thống sống. Các báo cáo chỉ ra rằng hơn 80% protein không hoạt động đơn lẻ mà thường tương tác với các phân tử khác như DNA, RNA hoặc protein để thực hiện các chức năng chuyên biệt trong tế bào¹. Thuật ngữ “tương tác protein” mô tả sự tiếp xúc vật lý giữa protein và đối tác tương tác, đồng thời mô tả các tương tác của chúng khi thực hiện chức năng². Việc xác định các tương tác giữa protein-protein (protein-protein interactions – PPIs) là vô cùng quan trọng trong việc đào sâu vào những cơ chế của sinh học tế bào, khám phá và thiết kế các phương pháp trị liệu mới, kỹ thuật protein và nghiên cứu đột biến gen³. Một số phương pháp thực nghiệm hiệu năng cao như hệ thống lai đôi ở nấm men (Yeast two-hybrid – Y2H), đồng tủa miễn dịch (co-immunoprecipitation – co-IP), phương pháp khối phổ xác định phức hợp protein (Mass spectrometric protein complex identification – MS-PCI) đã và đang được sử dụng để phát hiện tương tác giữa các protein⁴⁻⁶. Các phương pháp này đóng góp nhiều vào bộ dữ liệu PPI ở nhiều loài nhưng lại tốn kém, mất nhiều thời gian và công sức. Ngoài ra, sự ảnh hưởng

bởi yếu tố ngoại cảnh của điều kiện thí nghiệm cũng làm tăng tỷ lệ dương tính và âm tính giả⁷. Từ đó, bằng cách kết hợp những thông tin có được từ thực nghiệm và sự hỗ trợ của sinh học tính toán, các kết quả tương tác protein-protein hay tương tác phối tử-thụ thể ngày càng có sự chính xác và chất lượng tốt hơn⁸.

Cấu trúc protein đơn lẻ hay phức hợp protein-protein đã đem lại nhiều thông tin và cung cấp kiến thức về quá trình hoạt động của phân tử. Tuy vậy, trong thực tế, protein luôn chuyển động và chức năng sinh học cũng như tương tác protein-phối tử phụ thuộc hoàn toàn vào sự linh động của các phân tử thành phần. Để theo dõi được động học của protein, các phương pháp thực nghiệm đã được áp dụng và nắm bắt được chuyển động của protein ở mức từ femto giây tới nano giây như tinh thể hóa tia X độ phân giải siêu cao (ultra resolution X-ray crystallography), cộng hưởng từ hạt nhân siêu nhanh (ultra fast NMR), kính hiển vi lực nguyên tử (atomic force microscopy – AFM) hay phổ huỳnh quang (fluorescence spectroscopy). Tương tự với quá trình xác định tương tác protein, các phương pháp nêu trên cũng gây khó khăn cho các nhà nghiên cứu vì đòi hỏi nhiều thiết bị tiên tiến hay kỹ thuật có độ khó cao để theo dõi được động học của protein⁹. Mô

Trích dẫn bài báo này: Thường N V M, Tú L C, Thiên D T, Hiếu T V. **Mô hình hóa cấu trúc và mô phỏng động lực học phân tử trong nghiên cứu tương tác protein-protein.** *Sci. Tech. Dev. J. - Nat. Sci.* 2024, 8(4):3168-3178.

phỏng động lực học phân tử (Molecular dynamic simulations – MDs) dự đoán chuyển động của các nguyên tử trong protein theo thời gian dựa trên mô hình vật lý mô tả các lực tác động lên chúng¹⁰. Các mô phỏng này có thể đưa cho chúng ta những góc nhìn mới về các quá trình sinh học quan trọng như thay đổi cấu trúc, tương tác phối tử hay gấp cuộn protein. Vì thế, MDs đang ngày càng được sử dụng rộng rãi kết hợp cùng với các phương pháp thực nghiệm nhằm quan sát được chuyển động với thời gian dài và chính xác hơn.

Bài tổng quan này được thực hiện với mục tiêu giới thiệu về các phương pháp dự đoán cấu trúc tương tác protein-protein *in silico*, đặc biệt là phương pháp mô phỏng động lực học phân tử, tập trung vào khía cạnh trường lực, các gói chương mô phỏng đang được sử dụng rộng rãi và ứng dụng trong nghiên cứu phát triển thuốc.

MÔ HÌNH HÓA CẤU TRÚC TƯƠNG TÁC PROTEIN-PROTEIN

Tương tác protein-protein giữ chức năng quan trọng trong hầu hết các hoạt động sinh học của tế bào, bao gồm dẫn truyền tín hiệu, sự di chuyển của tế bào, nhận diện kháng nguyên và tiêu diệt các tác nhân xâm nhiễm. Tuy đã có những cải tiến về sinh học phân tử cũng như hệ gen học (genomics), chức năng của phần lớn protein vẫn chưa được biết rõ¹¹. Theo Jansen, sự tương tác giữa những protein đã biết hay chưa biết chức năng có thể đóng góp hiệu quả vào việc giải mã những bí ẩn này¹². Hiện nay, hai phương pháp được phát triển và sử dụng cho việc xây dựng mô hình cấu trúc tương tác protein-protein là docking phân tử (molecular docking), hay mô hình hóa không khuôn mẫu (template-free modeling – TFM), và mô hình hóa dựa trên khuôn mẫu (template-based modeling – TBM) (Hình 1).

Mô hình hóa cấu trúc tương tác không khuôn mẫu hay Docking phân tử

Trong mô hình hóa cấu trúc không khuôn mẫu (template-free modeling – TFM), hay docking phân tử (molecular docking), phức hợp cấu trúc tương tác được xây dựng bằng cách lắp ráp các cấu trúc đã biết của các phân tử tương tác, được tìm thấy hoặc dự đoán ở dạng protein đơn, thông qua cách tìm kiếm hệ thống và lựa chọn các hướng gắn khác nhau. Phương pháp này dựa trên sự phù hợp về đặc tính lí hoá và tính bổ sung hình học ở bề mặt của phức hợp¹³. Bề mặt phân tử có thể được mô tả theo diện tích bề mặt có thể tiếp cận với dung môi và mô tả theo bề mặt tiếp xúc với phân tử mục tiêu của nó. Tính bổ sung

giữa hai bề mặt phân tử được đánh giá dựa trên sự phù hợp về hình dạng, hỗ trợ bằng việc xác định các rãnh hoặc hốc bổ sung trên bề mặt phân tử của những mục tiêu cho quá trình docking¹⁴⁻¹⁶. Cách tiếp cận này hiệu quả với phức hợp có bề mặt bổ sung nhau một cách rõ ràng với diện tích bề mặt lớn (>1400 Å²) và có ưu thế về tính kỵ nước¹⁷. Phương pháp này mô phỏng tốt phần lớn các cấu trúc phức hợp; tuy nhiên, sự hạn chế trong các thuật toán lấy mẫu (sampling) không gian có thể tạo ra những kết quả dương tính giả có tính bổ sung bề mặt tốt nhưng lại khác xa với cấu trúc tự nhiên và phụ thuộc vào độ chính xác của hàm chấm điểm¹³⁻¹⁸. Các phương pháp tìm kiếm cho docking protein-protein có thể được phân thành ba hướng tiếp cận bao gồm: tìm kiếm toàn diện tổng thể (exhaustive global search), tương thích hình dạng vị trí cục bộ (local shape feature matching) và tìm kiếm ngẫu nhiên (randomized search). Trong khi đó, các thuật toán lấy mẫu tiên tiến hơn so với docking phân tử nhỏ là mô hình hóa bằng mạng ClusENM kết hợp HADDOCK, lấy mẫu Monte Carlo nâng cao, hay phương pháp Aether đang ngày càng được sử dụng rộng rãi¹⁹⁻²². ClusPro và HADDOCK là hai trong số những sever docking thường được sử dụng.

- ClusPro (<https://cluspro.bu.edu/login.php?redir=/home.ph>) dự đoán phức hợp protein-protein dựa trên phương pháp docking thân cứng (rigid docking)²³. Quá trình docking gồm ba bước bao gồm: (1) Phối tử được docking cứng với thụ thể trong 70000 bước xoay; (2) Trong 70000 bước xoay đó, chọn ra 1000 tổ hợp tịnh tiến/xoay có điểm thấp nhất; (3) Một nghìn tổ hợp đó được phân cụm với bán kính RMSD < 9 Å, nghĩa là vị trí phối tử có nhiều lân cận (neighbor) trong bán kính 9 Å được chọn và trở thành trung tâm cụm (centroid). ClusPro cho ra kết quả được xếp hạng theo điểm tổng số năng lượng thấp nhất được tính dựa trên hàm năng lượng có sẵn và độ lớn của cụm cấu trúc; trong đó, cấu trúc trung tâm được chọn làm đại diện cho từng cụm cấu trúc. ClusPro cho phép người dùng tùy chỉnh giới hạn tìm kiếm để trả kết quả theo số lượng mong muốn²⁴.

- HADDOCK (High ambiguity driven protein-protein docking) (<https://www.bonvinlab.org/software/>) là công cụ docking linh hoạt (flexible docking) nhiều loại đại phân tử từ protein đến glycan²⁵. HADDOCK hoạt động theo quy trình: (1) Docking thân cứng các tiểu phần một cách ngẫu nhiên vào các không gian với khoảng cách tương đối và tối thiểu năng lượng để tạo phức hợp; (2) Docking bán linh hoạt và tinh chỉnh các cấu trúc có số điểm cao; (3) Docking linh hoạt thực hiện việc mô phỏng động lực học phân tử các mô hình từ bước trước. Cách xây dựng mô hình của HADDOCK là tích hợp tổng thể

các thông tin sinh hóa, tin sinh học và thông tin cấu trúc liên quan²⁶. HADDOCK là một trong những công cụ docking đầu tiên áp dụng đến giới hạn tương tác không rõ ràng (ambiguous interaction restraints – AIR). Thông qua các thông tin thực nghiệm có sẵn như dịch chuyển hóa học NMR, đột biến, HDX, các amino acid tham gia vào tương tác sẽ được chọn và từ đó có thể xây dựng phức hợp tương tác một cách chính xác hơn.

Mô hình hóa cấu trúc tương tác dựa trên khuôn mẫu

Trong phương pháp hình hóa dựa trên khuôn mẫu (template-based modeling – TBM), phức hợp cấu trúc tương tác được xây dựng bằng dựa trên cấu trúc phức hợp khuôn mẫu đã được xác định từ cơ sở dữ liệu. Phương pháp này đã được dùng để dự đoán cấu trúc bậc bốn của protein chuỗi đơn dựa trên nguyên tắc sự tương đồng trình tự có thể dẫn đến tương đồng về cấu trúc²⁷. Các bước thực hiện tổng quát như sau: (1) Tìm kiếm một hoặc nhiều khuôn mẫu phù hợp; (2) Sắp giống cột trình tự mục tiêu với trình tự khuôn mẫu; (3) Xây dựng mô hình ban đầu cho cấu trúc truy vấn bằng cách giữ nguyên các phân đoạn cấu trúc từ các vùng sắp giống cột của khuôn mẫu; (4) Thay thế các chuỗi bên phù hợp với trình tự truy vấn; (5) Xây dựng các vùng loop và vùng kết thúc chuỗi; (6) Tinh chỉnh mô hình để thu được toàn bộ cấu trúc. Dù TBM đã thể hiện được một số tiến bộ so với TFM, phương pháp này vẫn tồn tại những hạn chế nổi bật như độ chính xác của mô hình sẽ giảm nếu độ tương đồng của trình tự mục tiêu và trình tự khuôn mẫu thấp hơn 40% cũng như thiếu phương pháp tinh chỉnh phức hợp protein²⁸⁻³⁰. Hai trong những công cụ được sử dụng để dự đoán PPI theo phương pháp TBM là SWISS-MODEL và AlphaFold2.

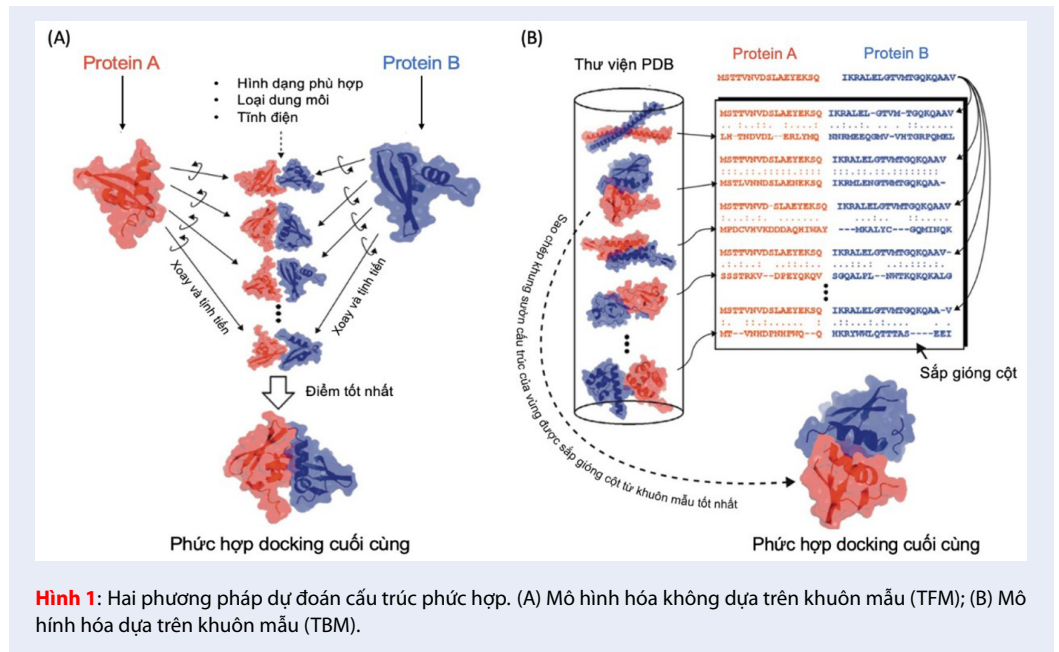
- SWISS-MODEL (<https://swissmodel.expasy.org>) là công cụ mô phỏng động lực học phân tử tiên phong trong quá trình xây dựng mô hình 3D của protein từ năm 1993 cho đến nay³¹. Cấu trúc phức hợp protein-protein thông qua thuật toán máy học (machine learning) có giám sát support vector Machines (SVM), kết hợp bởi những thông tin về vùng tương tác bảo tồn, gom cụm cấu trúc, và những thông số khác để xây dựng giá trị chất lượng cấu trúc bậc bốn (auaternary structure quality estimate – QSQE)³². Những phức hợp có mức độ tương đồng cao thường sở hữu cấu trúc bậc bốn và phương thức tương tác như nhau; qua đó, Bertoni và cộng sự đã đề xuất dấu chỉ PPI (PPI fingerprint), một hàm số thể hiện được sự bảo tồn bề mặt tương tác qua quá trình tiến hóa. Phân tích dấu chỉ PPI giúp cung cấp thêm các thông tin quan trọng cho dự đoán cấu trúc bậc bốn của protein.

- AlphaFold2 bên cạnh dự đoán cấu trúc protein đơn lẻ còn có thể giúp chúng ta xây dựng mô hình tương tác dựa trên các phức hợp có sẵn. Công cụ hoạt động theo cách sử dụng các trình tự có mối liên hệ tiến hóa, sắp giống cột nhiều trình tự; sau đó sử dụng Evoformer để tìm mối liên hệ giữa trình tự và cấu trúc khuôn mẫu³³. Evoformer xem protein như các đám mây đơn phân di chuyển xung quanh mạng lưới để tạo ra mô hình 3D; cuối cùng các tinh chỉnh cục bộ được thực hiện để đưa ra kết quả³⁴.

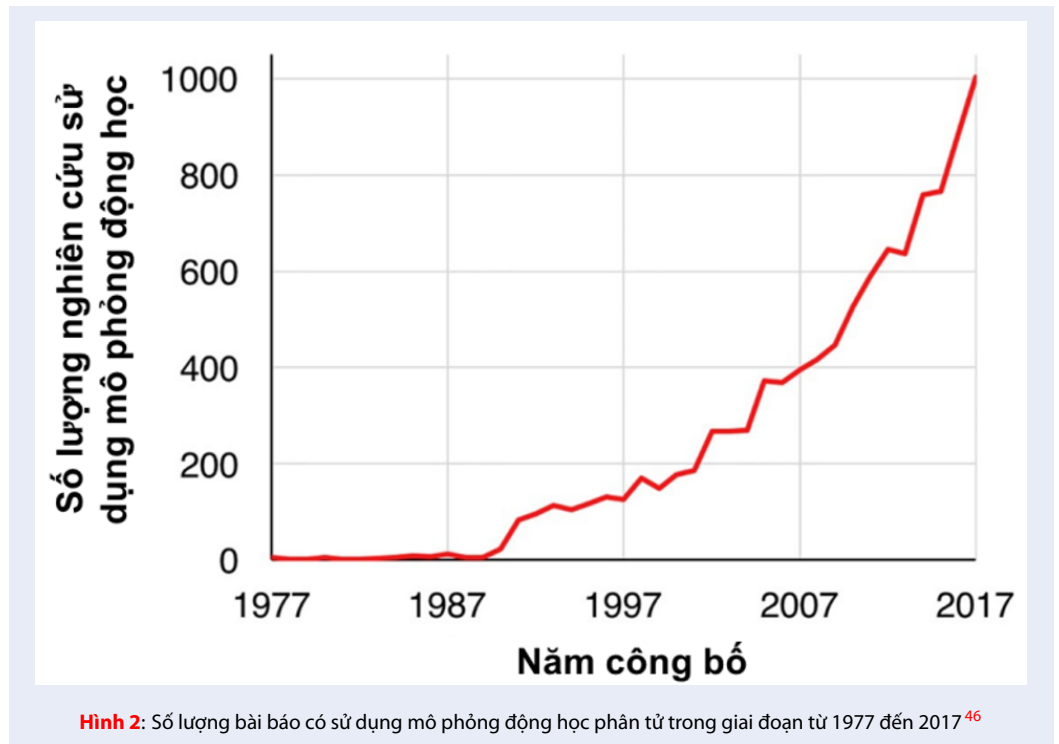
Khuyết điểm lớn nhất của SWISS-MODEL hay AlphaFold2 là cả hai đều bị ảnh hưởng bởi thông tin về các phức hợp protein-protein có sẵn để dự đoán chính xác. Kết quả của SWISS-MODEL hoàn toàn được mô phỏng dựa trên cấu trúc khuôn mẫu mà người dùng chọn; vì thế, việc lựa chọn được cấu trúc khuôn mẫu đáng tin cậy cũng như kiến thức về tương tác là vô cùng cần thiết. Trong khi đó, AlphaFold2 mặc dù đã cải tiến hơn với việc chỉ dựa vào trình tự và dự đoán tương tác thông qua thông tin di truyền nhưng vẫn còn hạn chế trong việc đưa ra cấu trúc protein-protein chính xác. Việc này chủ yếu đến từ bộ dữ liệu để huấn luyện cho AlphaFold-Multimer vẫn còn ít và chưa bao quát được toàn bộ các trường hợp có thể xảy ra. Tuy nhiên, các nhà nghiên cứu vẫn đang tiếp tục phát triển, cải thiện AlphaFold và đã đạt được một số kết quả hứa hẹn³⁵⁻³⁸.

MÔ PHỎNG ĐỘNG LỰC HỌC PHÂN TỬ

Mặc dù được xem như là một công cụ vô cùng hữu dụng trong những nghiên cứu về mối quan hệ giữa cấu trúc và chức năng, docking phân tử vẫn tồn tại nhiều nhược điểm. Điều này được bắt nguồn từ việc docking phân tử chỉ được xem như một phương pháp lấy mẫu cứng, trong khi protein lại vô cùng linh động, và không hề phản ánh được sự thay đổi của phân tử khi bị ảnh hưởng bởi phân tử khác (ion, dung môi,...) hay môi trường (pH, điện trường,...)³⁹. Một số tương tác phân tử không thể được dự đoán chính xác, bao gồm tác động của dung môi hay thay đổi entropy^{40,41}. Từ đó, docking phân tử cần được kết hợp với một số kỹ thuật khác để có thể cung cấp những kết quả đáng tin cậy hơn, và một trong số đó chính là mô phỏng động lực học phân tử (Molecular Dynamics Simulations – MDs). Từ những năm 50 của thế kỷ trước, các nhà khoa học đã thể hiện mối quan tâm đến cách thức hoạt động ở cấp độ nguyên tử liên quan đến chức năng của một phân tử⁴². Kể từ đó mô phỏng động lực học phân tử được xem như một giải pháp hữu hiệu để nghiên cứu sự chuyển động không ngừng của các nguyên tử nhờ vào sự hỗ trợ của máy tính. Năm 1977,



Hình 1: Hai phương pháp dự đoán cấu trúc phức hợp. (A) Mô hình hóa không dựa trên khuôn mẫu (TFM); (B) Mô hình hóa dựa trên khuôn mẫu (TBM).



Hình 2: Số lượng bài báo có sử dụng mô phỏng động học phân tử trong giai đoạn từ 1977 đến 2017⁴⁶

McCammon và cộng sự đã thực hiện nghiên cứu đầu tiên để có cái nhìn sâu sắc về động học của protein ức chế trypsin tuyến tụy bò, nghiên cứu này đã làm thay đổi quan điểm xem protein như một hệ cứng nhắc^{43,44}. Hình 2 thể hiện số lượng các công bố về sinh học cấu trúc có sử dụng mô phỏng động học kể từ năm 1977 cho thấy tính hữu ích của mô phỏng động học phân tử trong nghiên cứu, đặc biệt trong lĩnh vực khoa học thần kinh khi được dùng để nghiên cứu các protein quan trọng với tín hiệu thần kinh⁴⁵.

Nguyên lý và nhiệm vụ của mô phỏng động học phân tử

Mô phỏng động học là kỹ thuật nghiên cứu sự chuyển động của một hệ gồm rất nhiều nguyên tử, được xem như các hạt, trong một khoảng thời gian ngắn bằng việc sử dụng phương trình chuyển động của Newton. Trong khoảng thời gian mô phỏng, sự biến đổi của các hạt trong hệ được tính toán bằng tích phân phương trình chuyển động của chúng:

$$a_i(t) = \frac{F_i(t)}{m_i}$$

Với m_i là khối lượng, a_i là gia tốc, F_i là lực tác dụng lên hạt thứ i .

Điểm mạnh của mô phỏng động học phân tử là cho phép hệ ở trạng thái động, điều này phù hợp với thực tế là các đại phân tử sinh học nói chung hay protein nói riêng tồn tại ở trạng thái động, cung cấp phương tiện để mô hình hóa tính linh hoạt và những thay đổi về hình dạng của protein trong các quá trình sinh học ở các điều kiện khác nhau của môi trường⁴⁴.

Trường lực

Trường lực được định nghĩa là một biểu thức toán học mô tả sự phụ thuộc năng lượng của một hệ thống vào tọa độ của từng nguyên tử cấu thành hệ. Có nhiều cách phân loại và dạng trường lực sẵn có hiện nay⁴⁷⁻⁴⁹, tuy nhiên đều sử dụng cùng một biểu thức điển hình cho trường lực:

$$U = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} V_n [1 + \cos(n\phi - \gamma)] + \sum_i \sum_{i \neq j}^4 \sum_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_i \sum_{i \neq j} \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}}$$

Biểu thức trên, nhìn chung, được tách biệt thành hai dạng là năng lượng nội phân tử và liên phân tử. Năng lượng liên kết bao gồm ba phần tử đầu tiên trong biểu thức trên là E_{bond} được tính bằng tổng độ dài tất cả các liên kết giữa hai nguyên tử và E_{angle} tính toán độ rộng các góc liên kết. Cách khai triển hai phần tử này tương tự như cách tính độ co giãn của lò xo trong định luật Hooke⁵⁰. Năng lượng xoay góc nhị diện ϕ và ψ

($E_{dihedral}$) tương ứng với n góc nhị diện có mặt trong chuỗi⁴⁸. Trong khi đó, năng lượng liên phân tử, đóng vai trò quan trọng trong tương tác protein-protein, đề cập đến các tương tác tĩnh điện và Van der Waals (VdW) giữa những nguyên tử không tạo liên kết trực tiếp. Năng lượng từ tương tác VdW được tính dựa trên thế năng Lennard-Jones 12-6, mô hình lực đẩy điện tử và các tương tác phân tán⁵¹. Cuối cùng, định luật Coulomb được sử dụng để mô tả tương tác tĩnh điện giữa một cặp nguyên tử đại diện bởi đại lượng $E_{coulomb}$ ⁵¹.

Hiện nay có ba mô hình phân tử chính đã được phát triển gồm: trường lực toàn nguyên tử^{52,53}, trường lực thô hóa⁵⁴, và trường lực kết hợp^{55,56}. Các nghiên cứu về động học phân tử đang sử dụng một cách phổ biến bốn họ trường lực dành cho protein là: AMBER, CHARMM, GROMOS, OPLS-AA⁴⁷. Những trường lực này khác nhau về dữ liệu và quy trình sử dụng để tham số hóa. Tuy nhiên, Jorgensen và Tirado đã khẳng định các trường lực hiện có vẫn cung cấp các kết quả đáng tin cậy⁵⁷. Bảng 1 tóm tắt một số trường lực hiện có trong các họ trường lực nêu trên.

Các chương trình mô phỏng động lực học phân tử

- CHARMM (Chemistry at HARvard Molecular Mechanic, <https://www.charmm.org/>) được phát triển từ 40 năm trước, căn bản tập trung vào các đại phân tử sinh học như protein, nucleic acid, lipid, carbohydrate v.v.. Chương trình cung cấp một lượng lớn bộ công cụ tính toán bao gồm nhiều phương pháp lấy mẫu cấu hình và ước tính năng lượng tự do, tối thiểu năng lượng, động học và kỹ thuật phân tích cũng như khả năng xây dựng mô hình^{63,64}. CHARMM được nhà phát triển xây dựng với ba yếu tố giúp làm nên độ hiệu quả của gói mô phỏng này là (1) hoàn tất công việc trong khoảng thời gian ngắn nhất có thể, (2) dữ liệu đầu vào được đơn giản hóa tối đa, (3) nhà phát triển quan tâm đến việc thiết kế giao diện chương trình dễ hiểu và dễ chỉnh sửa⁶⁴.

- AMBER (Assisted Model Building with Energy Refinement, <https://ambermd.org/>) là chương trình mô phỏng giúp phân tích hiệu quả động học của các phân tử như protein, nucleic acid và carbohydrate. AMBER không phải là một chương trình đơn lẻ mà là một tập hợp các gói chương trình được thiết kế để hoạt động cùng nhau. Quy trình thực hiện gồm ba bước chính: chuẩn bị hệ thống, mô phỏng và phân tích kết quả. Hình 3 tóm tắt quy trình thông tin trong gói mô phỏng AMBER. Quy trình này được phân tích chi tiết trong bài tổng quan “Chương trình mô phỏng sinh học phân tử Amber” của David và cộng sự và tài

Bảng 1: Các trường lực hiện có trong mô phỏng động lực học đại phân tử sinh học

Họ trường lực	Trường lực	Mô tả
CHARMM ⁵⁸	CHARMM22	Một trường lực toàn nguyên tử dành cho protein
	CHARMM27	Trường lực dành cho nucleic acid và lipid
AMBER ⁵⁹	ff19SB	Đang được sử dụng rộng rãi để mô phỏng protein
	GLYCAM	Mô phỏng carbohydrate
	ff14SBonlysc	Là trường lực tốt nhất đối với mô phỏng trong dung môi ẩn
GROMOS ^{60,61}	GROMOS53a6	Trường lực dùng trong mô phỏng carbohydrate
	GROMOS45a3	Sử dụng cho các tập hợp lipid như màng tế bào, micelle
	GROMOS45a4	Cải thiện cách tham số hóa nucleic acid
OPLS-AA ⁶²	OPLS3	Mô phỏng động lực học phân tử nhỏ và protein

liệu hướng dẫn sử dụng Amber22⁶⁵. AMBER hỗ trợ một trường lực tổng quát cho các phân tử hữu cơ nhỏ, được dùng để phân tích phức hợp của phân tử nhỏ với protein hoặc nucleic acid⁶⁶.

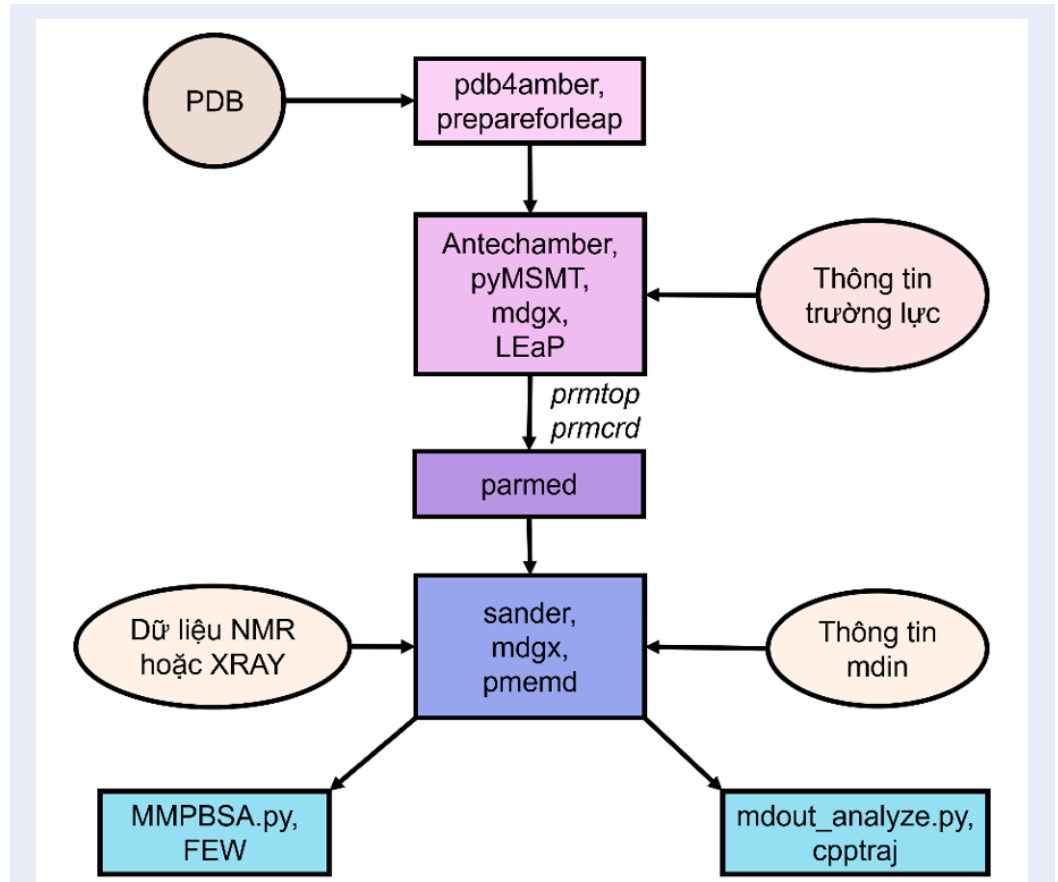
- GROMACS (GRONingen MACHine for Chemical Simulations, <https://www.gromacs.org/>) là một hệ thống mã nguồn mở được sử dụng rộng rãi nhất trong mô phỏng động lực học phân tử sinh học như protein, lipid, nucleic acid, cũng như các phân tử polymer phi sinh học⁶⁷. GROMACS được đánh giá là một phần mềm nhanh, linh hoạt, miễn phí và thân thiện với người dùng^{68,69}. Hiệu suất của chương trình mô phỏng được nâng cao hơn kể từ phiên bản GROMACS 2019 khi có sự hỗ trợ của phần mềm tích hợp gmx_api. Phần mềm này cung cấp một giao diện Python cho phép kết nối các câu lệnh của GROMACS với các công cụ phân tích của bên thứ ba⁷⁰. Chi tiết các bước chạy mô phỏng trên GROMACS được thể hiện ở Hình 4 và bài tổng quan của Smith và cộng sự⁷¹.

- NAMD (Nanoscale Molecular Dynamics, <http://www.ks.uiuc.edu/Research/namd/>) là gói chương trình được thiết kế để mô phỏng hiệu suất cao của các hệ thống phân tử sinh học lớn. Một mục tiêu quan trọng của NAMD là cung cấp cho người dùng một giao diện thống nhất trên tất cả các nền tảng, từ laptop đến máy tính để bàn và cả trên hệ thống siêu máy tính. Điều này được thực hiện dựa trên hệ thống lập trình song song Charm⁺⁺⁷². NAMD sử dụng thuật toán sao chép nhiều lần (Multiple Copy Algorithms – MCA) hỗ trợ tốt hơn trong việc lấy mẫu cấu hình, tính toán năng lượng tự do hay tinh chỉnh các trạng thái chuyển tiếp⁷³. Ngoài ra, một ưu điểm của NAMD là giao diện tập lệnh Tcl cung cấp tính linh hoạt tối đa cho người dùng có thể sửa đổi hoặc triển khai MCA một cách độc lập ngay cả khi đang chạy mô phỏng^{72,73}.

ỨNG DỤNG TRONG NGHIÊN CỨU PHÁT TRIỂN THUỐC

Các bất thường xảy ra trong tương tác protein-protein có thể dẫn đến các loại bệnh khác nhau, trong đó có ung thư⁷⁴. Nghiên cứu tương tác protein-protein bằng công cụ tin sinh học đã hỗ trợ mạnh mẽ trong công cuộc phát triển thuốc chữa bệnh^{75,76}. Chiến lược phòng ngừa và điều trị bằng chất ức chế cộng hóa trị nổi lên như một phương pháp tiềm năng vì ít khả năng thúc đẩy quá trình kháng thuốc⁷⁷. Năm 2019, Khuchtumur và cộng sự phát triển chất ức chế TEAD-347 nhắm vào tương tác giữa protein Yap và TEAD thông qua tương tác với một cysteine bảo tồn trong túi palmitate⁷⁸. Ngoài ra, mạng lưới tương tác protein-protein được xây dựng nhằm nghiên cứu các con đường truyền tín hiệu và dự đoán chức năng của protein chưa biết⁷⁹.

Hiện nay mô phỏng động lực học phân tử được dùng rộng rãi trong nghiên cứu protein như sự tương tác protein-protein hay các con đường truyền tín hiệu⁸⁰. Ví dụ, cặp thụ thể protein G (G protein-coupled receptor – GPCR) là loại thụ thể quan trọng với các loại thuốc, hơn một phần tư số thuốc nhắm mục tiêu vào loại thụ thể này⁸⁰⁻⁸². Năm 2020, 24 loại thuốc mới nhắm mục tiêu GPCR đã được phê duyệt lâm sàng và 44 loại thuốc mới đang trong giai đoạn thử nghiệm⁸³. Mô phỏng động lực học phân tử cung cấp thông tin chi tiết về ba cơ chế của GPCR bao gồm: sự thay đổi cấu hình trong trạng thái hoạt động và không hoạt động, tương tác của GPCR với phối tử/chất ức chế, tác động của lipid đối với cấu trúc của GPCR⁸⁴. Một ví dụ khác là trong nghiên cứu cách thức các thể đột biến của SARS-CoV-2 tương tác với thụ thể ACE2. Cụ thể, bốn biến thể chứa các đột biến đã được nghiên cứu, trong đó đột biến E484K và N501Y làm tăng ái lực liên



Hình 3: Quy trình thông tin đầu vào trong gói mô phỏng AMBER. pdb4amber: chương trình giúp chuyển đổi file ở định dạng khác sang định dạng pdb để tương thích với Leap; prepareforleap: một tùy chọn trong công cụ cpptraj có cùng chức năng như pdb4amber, thích hợp cho carbohydrate; Antechamber: chương trình giúp xây dựng trường lực cho các phân tử hữu cơ nhỏ như thuốc, amino acid được biến đổi; pyMSMT: gói mô hình hóa vị trí gắn kim loại của hệ hỗn hợp; mdgx: chương trình tạo các thông số trường lực liên kết; LeAP: chương trình tạo ra hệ mô phỏng trong Amber; sander: chương trình tối thiểu hóa năng lượng và động lực học phân tử; pmemd: một phiên bản của sander được tối ưu hóa; MMPBSA.py: code Python để phân tích năng lượng từ quá trình mô phỏng động lực học phân tử; FEW: chương trình tính toán năng lượng tự do của hệ protein-phối tử; mdout_analyze.py: file đầu ra của chương trình sander và pmemd; cpptraj: chương trình phân tích quỹ đạo và dữ liệu.

kết, khiến khả năng lây nhiễm cao hơn⁸⁵. Song song đó, Acharya và cộng sự đã thực hiện mô phỏng động lực học trên 8000 hợp chất nhằm sàng lọc các chất ức chế mạnh protein S và xác định được 77 hợp chất thuốc phân tử nhỏ tiềm năng trong điều trị COVID-19⁸⁶.

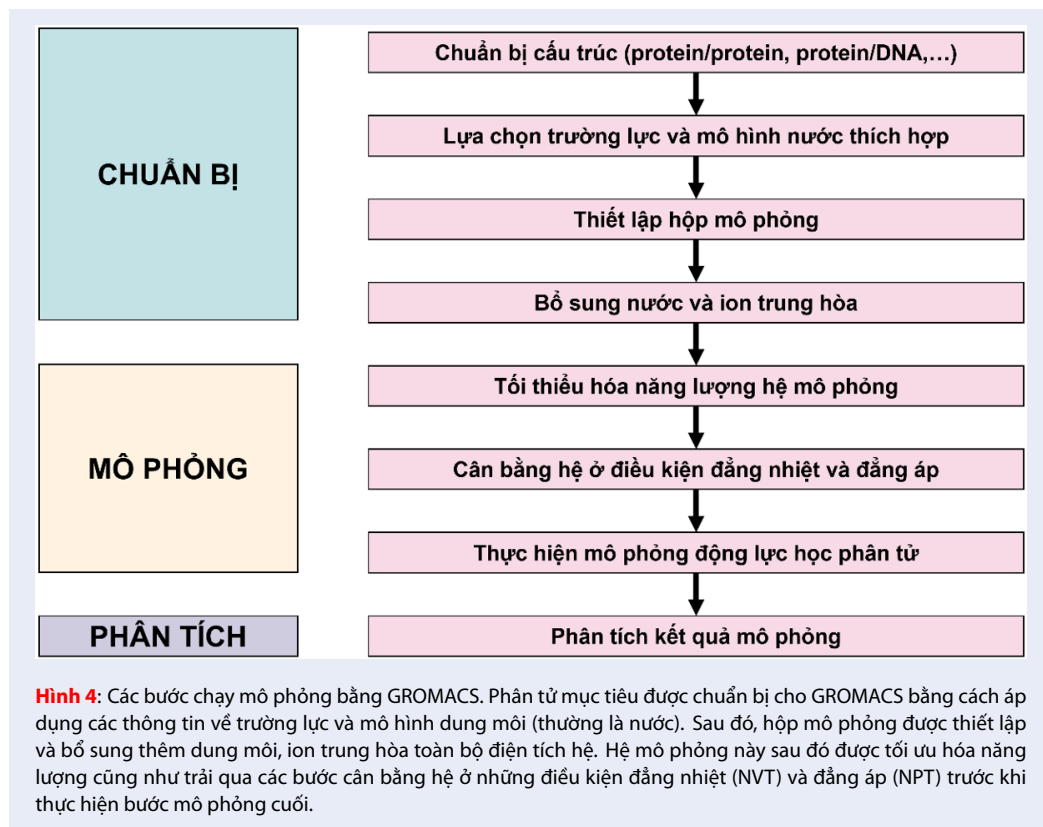
KẾT LUẬN

Mối quan tâm về cấu trúc phức hợp protein-protein được xem như vấn đề nổi trội, giúp các nhà nghiên cứu có cái nhìn sâu hơn vào cơ chế tương tác trong các quá trình sinh học. Do các khó khăn trong việc giải cấu trúc thực nghiệm nên việc dự đoán cấu trúc phức hợp protein-protein bằng công cụ tin sinh học đã trở thành một phương pháp thay thế tối ưu. Tuy nhiên, trong thực tế các phân tử sinh học tham gia vào quá

trình sinh học trong cơ thể luôn ở trạng thái động. Vì vậy, phương pháp mô phỏng động lực học phân tử đã được sử dụng từ những năm 70 của thế kỷ trước để hỗ trợ cho nghiên cứu sự tương tác liên quan đến chức năng của protein. Từ đó việc kết hợp phương pháp mô hình hóa cấu trúc phức hợp protein-protein và mô phỏng động lực học phân tử đã trở thành chìa khóa để mở rộng các nghiên cứu mang tính ứng dụng về sau.

DANH MỤC TỪ VIẾT TẮT

- PPIs: protein-protein interactions
- MDs: molecular dynamics simulation
- Y2H: yeast two-hybrid
- co-IP: co-immunoprecipitation



MS-PCI: mass spectrometric protein complex

AFM: atomic force microscopy

TFM: template-free modeling

TBM: template-based modeling

VdW: Van der Waals

CHARMM: chemistry at Harvard molecular mechanics

AMBER: assisted model building with energy refinement

GROMACS: groningen machine for chemical simulations

NAMD: nanoscale molecular dynamics

MCA: multiple copy algorithms

GPCR: G protein-coupled receptor

XUNG ĐỘT LỢI ÍCH

Các tác giả cam kết không có xung đột lợi ích.

ĐÓNG GÓP CỦA CÁC TÁC GIẢ

Nguyễn Văn Minh Thường: viết, tổng hợp và chỉnh sửa bản thảo

Lý Cẩm Tú: viết, tổng hợp và chỉnh sửa bản thảo

Đình Thuận Thiên: lên ý tưởng, viết, tổng hợp và chỉnh sửa bản thảo

Trần Văn Hiếu: lên ý tưởng, tham gia chỉnh sửa bản thảo và chấp thuận bản thảo

Tất cả các tác giả đồng ý với bản thảo cuối cùng.

TÀI LIỆU THAM KHẢO

1. Berggård T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. *Proteomics*. 2007;7(16):2833–2842. Available from: <https://doi.org/10.1002/pmic.200700131>.
2. Gonzalez MW, Kann MG. Chapter 4: Protein interactions and disease. *PLoS computational biology*. 2012;8(12):1002819–1002819. Available from: <https://doi.org/10.1371/journal.pcbi.1002819>.
3. Abbasi WA, Asif A, Ben-Hur A. Learning protein binding affinity using privileged information. *BMC bioinformatics*. 2018;19(1):1–12. Available from: <https://doi.org/10.1186/s12859-018-2448-z>.
4. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki YJPOT-NAOS; 2001. Available from: <https://doi.org/10.1073/pnas.061034498>.
5. Tang Z, Takahashi Y. Analysis of Protein-Protein Interaction by Co-IP in Human Cells. *Methods Mol Biol*. 2018;1794:289–296. Available from: https://doi.org/10.1007/978-1-4939-7871-7_20.
6. Ho Y; 2002.
7. Mrowka R, Patzak A, Herzel HJGR. Is there a bias in proteome research?; 2001. Available from: <https://doi.org/10.1101/gr.206701>.
8. You ZH, et al. Highly efficient framework for predicting interactions between proteins. vol. 47, no. 3, pp. 731–743, 2016;.
9. Bourgeat L, et al. Experimental Protein Molecular Dynamics: Broadband Dielectric Spectroscopy coupled with nanoconfinement,” *Scientific Reports*, vol. 9, no. 1, p. 17988, 2019/11/29 2019; Available from: <https://doi.org/10.1038/s41598-019-54562-8>.

10. Chong WL. Protein-protein interactions: Insight from molecular dynamics simulations and nanoparticle tracking analysis. *Molecules*, vol. 26, no. 18, p. 5696, 2021;.
11. Rocha JJ, Whitaker JW, O'Connell JD, et al. Functional unknowns: Systematic screening of conserved genes of unknown function. *PLoS Biol.* 2023;21(8):e3002222;.
12. Jansen R, Yu H, Greenbaum D, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science.* 2003;302(5644):449-453;.
13. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc Natl Acad Sci U S A.* 2012;109(24):9438-9441;.
14. Cai W, Shao X, Maigret B. Protein-ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening. *J Mol Graph Model.* 2002;20(4):313-328;.
15. Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics.* 2005;21(10):2347-2355; Available from: [:10.1093/bioinformatics/bti355](https://doi.org/10.1093/bioinformatics/bti355).
16. Zhang Q, Sanner M, Olson AJ. Shape complementarity of protein-protein complexes at multiple resolutions. *Proteins.* 2009;75(2):453-467;.
17. Vajda S, Camacho CJ. Protein-protein docking: is the glass half-full or half-empty? *Trends Biotechnol.* 2004;22(3):110-116;.
18. Hiếu TV, Thiên ĐT, Thường NVM, Dũng TD. Protein structure modeling using cloud-based servers. *J Sci Technol.* 2024;8(1):2828-2837;.
19. Huang S-Y. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. *Drug Discov Today.* 2014;19(8):1081-1096;.
20. Zhang Z, Schindler CEM, Lange OF, Zacharias M. Application of enhanced sampling Monte Carlo methods for high-resolution protein-protein docking in Rosetta. *PLoS One.* 2015;10(6):e0125941;.
21. Torchala M, Moal IH, Chaleil RAG, Fernandez-Recio J, Bates PA. Enhanced sampling of protein conformational states for dynamic cross-docking within the protein-protein docking server SwarmDock. *Proteins.* 2020;88(8):962-972;.
22. Kaynak BT, et al. Sampling of protein conformational space using hybrid simulations: A critical assessment of recent methods. *Front Mol Biosci.* 2022;9:821314;.
23. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* 2004;32(Web Server issue):W96-W99;.
24. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein-protein docking. *Nat Protoc.* 2017;12(2):255-278;.
25. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol.* 2016;428(4):720-725;.
26. Moreira IS, Fernandes PA, Ramos MJ. Protein-protein docking dealing with the unknown. *J Comput Chem.* 2010;31(2):317-342;.
27. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 1986;5(4):823-826;.
28. Kundrotas PJ, Vakser IA, Janin J. Structural templates for modeling homodimers. *Protein Sci.* 2013;22(11):1655-1663;.
29. Chen H, Skolnick J. M-TASSER: an algorithm for protein quaternary structure prediction. *Biophys J.* 2008;94(3):918-928;.
30. Szilagyí A, Zhang Y. Template-based structure modeling of protein-protein interactions. *Curr Opin Struct Biol.* 2014;24:10-23;.
31. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296-303;.
32. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep.* 2017;7(1):10480;.
33. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583-9;.
34. Skolnick J, Gao M, Zhou H, Singh S. AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. *J Chem Inf Model.* 2021;61(10):4827-31;.
35. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv.* 2022;.
36. Liu J, Gao M, Skolnick J. Enhancing AlphaFold-Multimer-based protein complex structure prediction with MULTICOM in CASP15. *bioRxiv.* 2023;.
37. Wallner BJB. AFsample: improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics.* 2023;39(9):btad573;.
38. Bryant P, Noé FJ. Improved protein complex prediction with AlphaFold-multimer by denoising the MSA profile. *bioRxiv.* 2023;.
39. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model.* 2021;61(8):3891-8;.
40. Sethi A, Joshi K, Sasikala KA, Alvala MJDD. Molecular docking in modern drug discovery: principles and recent applications. *Drug Discov Today.* 2019;24(3):578-88;.
41. Huang SY, Zou X. Advances and challenges in protein-ligand docking. *Int J Mol Sci.* 2010;11(8):3016-34;.
42. Alder BJ, Wainwright TE. Phase transition for a hard sphere system. *J Chem Phys.* 1957;27(5):1208-9;.
43. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol.* 2002;9(9):646-52;.
44. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature.* 1977;267(5612):585-90;.
45. Shi L, Quick M, Zhao Y, Weinstein H, Javitch JA. The mechanism of a neurotransmitter: sodium symporter—inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol Cell.* 2008;30(6):667-77;.
46. Growth of molecular dynamics simulations in structural biology. *Web of Knowledge [Internet].* 2018 Mar 25 [cited 2024]; Available from: <https://webofknowledge.com/>.
47. Guvench O, MacKerell AD. Comparison of protein force fields for molecular dynamics simulations. In: Kukol A, editor. *Molecular Modeling of Proteins.* Totowa, NJ: Humana Press; 2008. p. 63-88;.
48. Lin F-Y, MacKerell AD. Force fields for small molecules. In: Bonomi M, Camilloni C, editors. *Biomolecular Simulations: Methods and Protocols.* New York, NY: Springer New York; 2019. p. 21-54;.
49. Vanommeslaeghe K, Guvench O, MacKerell AD Jr. Molecular mechanics. *Curr Pharm Des.* 2014;20(20):3281-92;.
50. Burkowski FJ. *Structural bioinformatics: an algorithmic approach.* Boca Raton, FL: Chapman and Hall/CRC; 2008;.
51. González M. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique.* 2011;12:169-200;.
52. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B.* 2010;114(23):7830-43;.
53. Moore PB, Lopez CF, Klein ML. Dynamical properties of a hydrated lipid bilayer from a multianosecond molecular dynamics simulation. *Biophys J.* 2001;81(5):2484-93;.
54. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B.* 2007;111(27):7812-24;.
55. Kar P, Gopal SM, Cheng YM, Panahi A, Feig M. Transferring the primo coarse-grained force field to the membrane environment: simulations of membrane proteins and helix-helix association. *J Chem Theory Comput.* 2014;10(8):3459-72;.

56. Shi Q, Izvekov S, Voth GA. Mixed atomistic and coarse-grained molecular dynamics: simulation of a membrane-bound ion channel. *J Phys Chem B*. 2006;110(31):15045-8;.
57. Jorgensen WL, Tirado-Rives J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci U S A*. 2005;102(19):6665-70;.
58. Małolepsza E, Strodel B, Khalili M, Trygubenko S, Fejer SN, Wales DJ. Symmetrization of the AMBER and CHARMM force fields. *J Comput Chem*. 2010;31(7):1402-9;.
59. Case DA, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, Cerutti DS, et al. AMBER 2022 Reference Manual. 2022 Mar 24; Available from: <https://ambermd.org/doc12/Amber22.pdf>.
60. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*. 2004;25(13):1656-76;.
61. Soares TA, Hünenberger PH, Kastenholz MA, Kräutler V, Lenz T, Oostenbrink C, et al. An improved nucleic acid parameter set for the GROMOS force field. *J Comput Chem*. 2005;26(7):725-37;.
62. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J Chem Theory Comput*. 2016;12(1):281-96;.
63. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation program. *J Comput Chem*. 2009;30(10):1545-1614;.
64. Brooks BR, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, vol. 4, no. 2, pp. 187-217, 1983;.
65. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. *J Comput Chem*. 2005;26(16):1668-88;.
66. Case DA. The Amber biomolecular simulation programs. *Journal of computational chemistry*, vol. 26, no. 16, pp. 1668-1688, 2005;.
67. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1:19-25;.
68. Sharma S, Kumar P, Chandra R, Singh S, Mandal A, Dondapati R. Overview of BIOVIA materials studio, LAMMPS, and GROMACS. *Mol Dyn Simul Nanocompos BIOVIA Mater Stud Lammps Gromacs*. Amsterdam: Elsevier; 2019. p. 39-100;.
69. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-18;.
70. Irrgang ME, Davis C, Kasson PM. gmxapi: A GROMACS-native Python interface for molecular dynamics with ensemble and plugin support. *PLoS Comput Biol*. 2022;18(2):e1009835;.
71. Smith A, Dong X, Raghavan V. An overview of molecular dynamics simulation for food products and processes. *Processes*. 2022;10(1):119;.
72. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys*. 2020;153(4):044130;.
73. Jiang W, Hardy DJ, Phillips JC, MacKerell AD, Schulten K, Roux B. Generalized Scalable Multiple Copy Algorithms for Molecular Dynamics Simulations in NAMD. *Comput Phys Commun*. 2014;185(3):908-16;.
74. Oláh J, Szénási T, Lehotzky A, Norris V, Ovádi J. Challenges in discovering drugs that target the protein-protein interactions of disordered proteins. *Int J Mol Sci*. 2022;23(3):1550;.
75. Hu L, Wang X, Huang YA, Hu P, You ZH. A survey on computational models for predicting protein-protein interactions. *Brief Bioinform*. 2021;22(5):bbab036;.
76. Soleymani F, Paquet E, Viktor H, Michalowski W, Spinello D. Protein-protein interaction prediction with deep learning: A comprehensive review. *Comput Struct Biotechnol J*. 2022;20:3701-17;.
77. Biji I, Ramharack P, Khan S, Cherqaoui D, Soliman ME. Tracing potential covalent inhibitors of an E3 ubiquitin ligase through target-focused modelling. *Molecules*. 2019;24(17):3125;.
78. Cheng SS, Yang GJ, Wang W, Leung CH, Ma DL. The design and development of covalent protein-protein interaction inhibitors for cancer treatment. *J Hematol Oncol*. 2020;13(1):26;.
79. Shaikat Z, Aiman S, Li CH. Protein-protein interactions: methods, databases, and applications in virus-host study. *World J Virol*. 2021;10(6):288-303;.
80. Dror RO, Dirks RM, Grossman J, Xu H, Shaw DE. Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys*. 2012;41:429-52;.
81. Heilker R, Wolff M, Tautermann CS, Bieler M. G-protein-coupled receptor-focused drug discovery using a target class platform approach. *Drug Discov Today*. 2009;14(5-6):231-40;.
82. Lundstrom K. Latest development in drug discovery on G protein-coupled receptors. *Curr Protein Pept Sci*. 2006;7(5):465-70;.
83. Mizuno H, Kihara Y. Druggable lipid GPCRs: past, present, and prospects. In: *Druggable Lipid Signaling Pathways*. Cham: Springer; 2020. p. 223-58;.
84. Sinha S, Tam B, Wang SM. Applications of molecular dynamics simulation in protein study. *Membranes*. 2022;12(9):844;.
85. Jangra S, Ye C, Rathnasinghe R, Stadlbauer D, Krammer F, Simon V, et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *Lancet Microbe*. 2021;2(7):e283-4;.
86. Acharya A, Agarwal R, Baker MB, Baudry J, Bhowmik D, Boehm S, et al. Supercomputer-based ensemble docking drug discovery pipeline with application to COVID-19. *J Chem Inf Model*. 2020;60(12):5832-52;.

Structure modeling and molecular dynamics simulation in protein-protein interaction study

Nguyen Van Minh Thuong^{1,2}, Ly Cam Tu^{1,2}, Dinh Thuan Thien^{1,2}, Tran Van Hieu^{1,2,*}

ABSTRACT

Protein is a vital macromolecule that contributes to the living system. Functionally, proteins not only work independently, but also form a close network via protein-protein interactions (PPIs), the decisive factor in most biological processes in living organisms. Nevertheless, the study of PPIs has encountered many obstacles due to the lack of experimental structures of the complexes or proteins involved in the interactions. Thanks to the support of bioinformatics tools, methods, and algorithms, the prediction of protein interaction structure has emerged as a potential solution to solve the above difficulties. Molecular docking and molecular dynamics simulation (MDs) are two commonly used methods for predicting PPIs because of their ability to model structure and approximate biological processes, which are consistent with the natural state of the realistic system. Applying these bioinformatics advances has shortened the time, effort, and research costs for scientists. This review provides information on the protein-protein complex structure prediction method and some tools to assist in previous studies. In addition, the force field component of the molecular kinetics simulation method is highlighted, and simulation programs are extensively used.

Key words: Bioinformatics, Molecular docking, Molecular dynamics simulations, Proteins interaction prediction, TBM, TFM

¹Faculty of Biology – Biotechnology,
University of Science, VNU-HCM

²Vietnam National University Ho Chi
Minh City

Correspondence

Tran Van Hieu, Faculty of Biology –
Biotechnology, University of Science,
VNU-HCM

Vietnam National University Ho Chi Minh
City

Email: tvhieu@hcmus.edu.vn

History

- Received: 10-8-2023
- Revised: 08-11-2023
- Accepted: 28-12-2024
- Published Online: 31-12-2024

DOI : <https://doi.org/10.32508/stdjns.v8i4.1301>



Copyright

© VNUHCM Press. This is an open-
access article distributed under the
terms of the Creative Commons
Attribution 4.0 International license.



Cite this article : Thuong NVM, Tu LC, Thien DT, Hieu TV. **Structure modeling and molecular dynamics simulation in protein-protein interaction study.** *Sci. Tech. Dev. J. - Nat. Sci.* 2024, 8(4):3168-3178.