

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,*}



Use your smartphone to scan this QR code and download this article

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: tvhieue@hcmus.edu.vn

Lịch sử

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

DOI:



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.



1 MỞ ĐẦU

2 Protein là đại phân tử sinh học quan trọng, đóng vai
3 trò kiến tạo nên hệ thống sống. Các báo cáo chỉ ra
4 rằng hơn 80% protein không hoạt động đơn lẻ mà
5 thường tương tác với các phân tử khác như DNA,
6 RNA hoặc protein để thực hiện các chức năng chuyên
7 biệt trong tế bào¹. Thuật ngữ “tương tác protein”
8 mô tả sự tiếp xúc vật lý giữa protein và đối tác tương
9 tác, đồng thời mô tả các tương tác của chúng khi
10 thực hiện chức năng². Việc xác định các tương tác
11 giữa protein-protein (protein-protein interactions –
12 PPIs) là vô cùng quan trọng trong việc đào sâu vào
13 những cơ chế của sinh học tế bào, khám phá và thiết
14 kế các phương pháp trị liệu mới, kỹ thuật protein và
15 nghiên cứu đột biến gen³. Một số phương pháp thực
16 nghiệm hiệu năng cao như hệ thống lai đôi ở nấm
17 men (Yeast two-hybrid – Y2H), đồng tủa miễn dịch
18 (co-immunoprecipitation – co-IP), phương pháp khối
19 phổ xác định phức hợp protein (Mass spectrometric
20 protein complex identification – MS-PCI) đã và đang
21 được sử dụng để phát hiện tương tác giữa các pro-
22 tein⁴⁻⁶. Các phương pháp này đóng góp nhiều vào
23 bộ dữ liệu PPI ở nhiều loài nhưng lại tốn kém, mất
24 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.* 2025; (1):1-10.

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.

67 MÔ HÌNH HÓA CẤU TRÚC TƯƠNG 68 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).

85 Mô hình hóa cấu trúc tương tác không 86 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

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

161 **Mô hình hóa cấu trúc tương tác dựa trên** 162 **khuôn mẫu**

163 Trong phương pháp hình hóa dựa trên khuôn mẫu
164 (template-based modeling – TBM), phức hợp cấu trúc
165 tương tác được xây dựng bằng dựa trên cấu trúc phức
166 hợp khuôn mẫu đã được xác định từ cơ sở dữ liệu.
167 Phương pháp này đã được dùng để dự đoán cấu trúc
168 bậc bốn của protein chuỗi đơn dựa trên nguyên tắc
169 sự tương đồng trình tự có thể dẫn đến tương đồng về
170 cấu trúc²⁷. Các bước thực hiện tổng quát như sau:
171 (1) Tìm kiếm một hoặc nhiều khuôn mẫu phù hợp;
172 (2) Sắp giống cột trình tự mục tiêu với trình tự khuôn
173 mẫu; (3) Xây dựng mô hình ban đầu cho cấu trúc truy
174 vấn bằng cách giữ nguyên các phân đoạn cấu trúc từ
175 các vùng sắp giống cột của khuôn mẫu; (4) Thay thế
176 các chuỗi bên phù hợp với trình tự truy vấn; (5) Xây
177 dựng các vùng loop và vùng kết thúc chuỗi; (6) Tinh
178 chỉnh mô hình để thu được toàn bộ cấu trúc. Dù TBM
179 đã thể hiện được một số tiến bộ so với TFM, phương
180 pháp này vẫn tồn tại những hạn chế nổi bật như độ
181 chính xác của mô hình sẽ giảm nếu độ tương đồng của
182 trình tự mục tiêu và trình tự khuôn mẫu thấp hơn 40%
183 cũng như thiếu phương pháp tinh chỉnh phức hợp
184 protein^{28–30}. Hai trong những công cụ được sử dụng
185 để dự đoán PPI theo phương pháp TBM là SWISS-
186 MODEL và AlphaFold2.

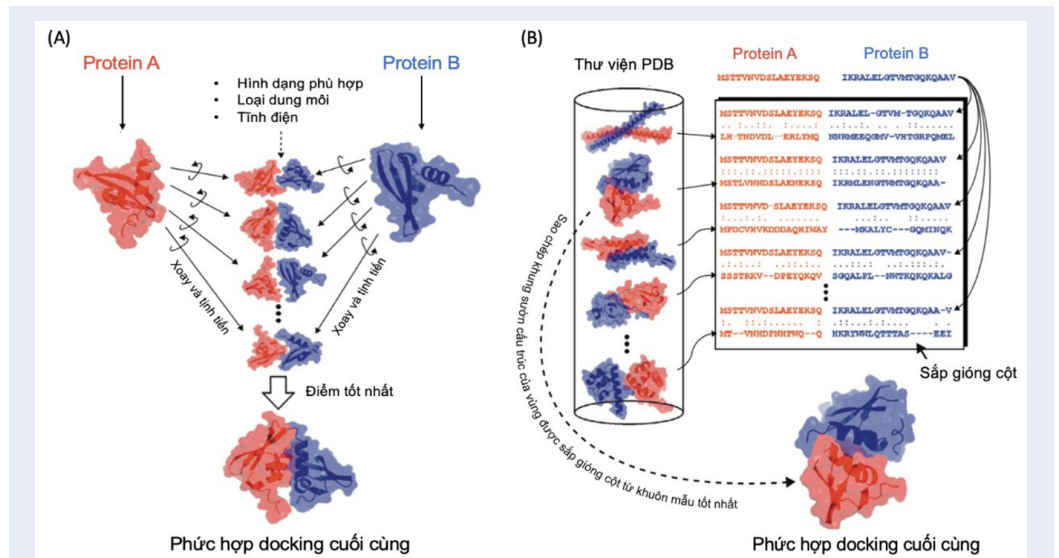
187 - SWISS-MODEL (<https://swissmodel.expasy.org>) là
188 công cụ mô phỏng tự động tiên phong trong quá
189 trình xây dựng mô hình 3D của protein từ năm 1993
190 cho đến nay³¹. Cấu trúc phức hợp protein-protein
191 thông qua thuật toán máy học (machine learning) có
192 giám sát support vector Machines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

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

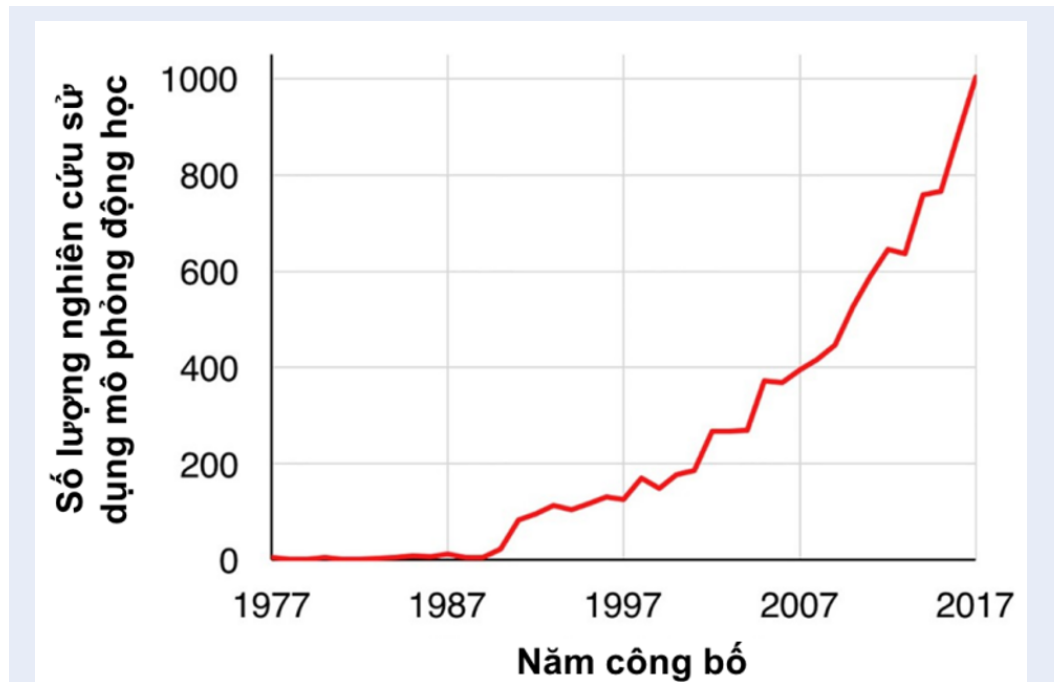
213 Khuyết điểm lớn nhất của SWISS-MODEL hay Al-
214 phaFold2 là cả hai đều bị ảnh hưởng bởi thông tin về
215 các phức hợp protein-protein có sẵn để dự đoán chính
216 xác. Kết quả của SWISS-MODEL hoàn toàn được mô
217 phỏng dựa trên cấu trúc khuôn mà người dùng chọn;
218 vì thế, việc lựa chọn được cấu trúc khuôn mẫu đáng
219 tin cậy cũng như kiến thức về tương tác là vô cùng cần
220 thiết. Trong khi đó, AlphaFold2 mặc dù đã cải tiến
221 hơn với việc chỉ dựa vào trình tự và dự đoán tương
222 tác thông qua thông tin di truyền nhưng vẫn còn hạn
223 chế trong việc đưa ra cấu trúc protein-protein chính
224 xác. Việc này chủ yếu đến từ bộ dữ liệu để huấn luyện
225 cho AlphaFold-Multimer vẫn còn ít và chưa bao quát
226 được toàn bộ các trường hợp có thể xảy ra. Tuy nhiên,
227 các nhà nghiên cứu vẫn đang tiếp tục phát triển, cải
228 thiện AlphaFold và đã đạt được một số kết quả hứa
229 hẹn^{35–38}.

231 **MÔ PHỎNG ĐỘNG LỰC HỌC PHÂN** 232 **TỬ**

233 Mặc dù được xem như là một công cụ vô cùng hữu
234 dụng trong những nghiên cứu về mối quan hệ giữa
235 cấu trúc và chức năng, docking phân tử vẫn tồn tại
236 nhiều nhược điểm. Điều này được bắt nguồn từ việc
237 docking phân tử chỉ được xem như một phương pháp
238 lấy mẫu cứng, trong khi protein lại vô cùng linh động,
239 và không hề phản ánh được sự thay đổi của phân tử
240 khi bị ảnh hưởng bởi phân tử khác (ion, dung môi,...)
241 hay môi trường (pH, điện trường,...)³⁹. Một số tương
242 tác phân tử không thể được dự đoán chính xác, bao
243 gồm tác động của dung môi hay thay đổi entropy^{40,41}.
244 Từ đó, docking phân tử cần được kết hợp với một số
245 kỹ thuật khác để có thể cung cấp những kết quả đáng
246 tin cậy hơn, và một trong số đó chính là mô phỏng
247 động lực học phân tử (Molecular Dynamics Simula-
248 tions – MDs). Từ những năm 50 của thế kỷ trước,
249 các nhà khoa học đã thể hiện mối quan tâm đến cách
250 thức hoạt động ở cấp độ nguyên tử liên quan đến chức
251 năng của một phân tử⁴². Kể từ đó mô phỏng động
252 lực học phân tử được xem như một giải pháp hữu hiệu
253 để nghiên cứu sự chuyển động không ngừng của các
254 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 2017 ⁴⁶

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

265 Nguyên lý và nhiệm vụ của mô phỏng động 266 học phân tử

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

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

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

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

283 Trường lực

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

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

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

($E_{dihedral}$) tương ứng với n góc nhị diện có mặt trong
298 chuỗi⁴⁸. Trong khi đó, năng lượng liên phân tử, đóng
299 vai trò quan trọng trong tương tác protein-protein,
300 để cập đến các tương tác tĩnh điện và Van der Waals
301 (VdW) giữa những nguyên tử không tạo liên kết trực
302 tiếp. Năng lượng từ tương tác VdW được tính dựa
303 trên thế năng Lennard-Jones 12-6, mô hình lực đẩy
304 điện tử và các tương tác phân tán⁵¹. Cuối cùng, định
305 luật Coulomb được sử dụng để mô tả tương tác tĩnh
306 điện giữa một cặp nguyên tử đại diện bởi đại lượng
307 $E_{coulomb}$ ⁵¹.

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

320 Các chương trình mô phỏng động lực học 321 phân tử

322 - CHARMM (Chemistry at HARvard Molecular Me-
323 chanic, <https://www.charmm.org/>) được phát triển từ
324 40 năm trước, căn bản tập trung vào các đại phân tử
325 sinh học như protein, nucleic acid, lipid, carbohydrate
326 v.v.. Chương trình cung cấp một lượng lớn bộ công
327 cụ tính toán bao gồm nhiều phương pháp lấy mẫu
328 cấu hình và ước tính năng lượng tự do, tối thiểu năng
329 lượng, động học và kỹ thuật phân tích cũng như khả
330 năng xây dựng mô hình^{63,64}. CHARMM được nhà
331 phát triển xây dựng với ba yếu tố giúp làm nên độ hiệu
332 quả của gói mô phỏng này là (1) hoàn tất công việc
333 trong khoảng thời gian ngắn nhất có thể, (2) dữ liệu
334 đầu vào được đơn giản hóa tối đa, (3) nhà phát triển
335 quan tâm đến việc thiết kế giao diện chương trình để
336 hiểu và dễ chỉnh sửa⁶⁴.

337 - AMBER (Assisted Model Building with Energy Re-
338 finement, <https://ambermd.org/>) là chương trình mô
339 phỏng giúp phân tích hiệu quả động học của các phân
340 tử như protein, nucleic acid và carbohydrate. AM-
341 BER không phải là một chương trình đơn lẻ mà là
342 một tập hợp các gói chương trình được thiết kế để
343 hoạt động cùng nhau. Quy trình thực hiện gồm ba
344 bước chính: chuẩn bị hệ thống, mô phỏng và phân
345 tích kết quả. Hình 3 tóm tắt quy trình thông tin trong
346 gói mô phỏng AMBER. Quy trình này được phân tích
347 chi tiết trong bài tổng quan “Chương trình mô phỏng
348 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.

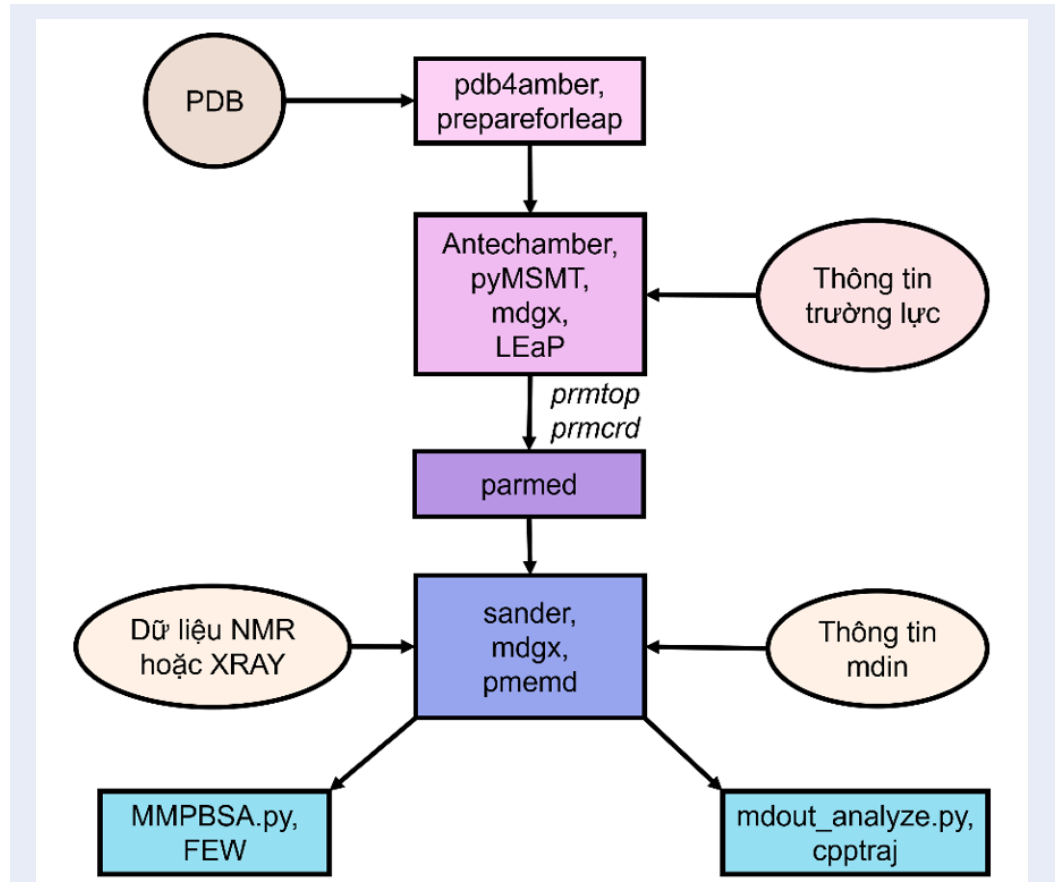
- NAMD (Nanoscale Molecular Dynamics, <http://www.ks.uiuc.edu/Research/namd/>) là gói chương trình được thiết kế để mô phỏng hiệu suất cao của các hệ thống phân tử sinh học lớn. Một mục tiêu quan trọng của NAMD là cung cấp cho người dùng một giao diện thống nhất trên tất cả các nền tảng, từ laptop đến 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 nhằm 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.

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

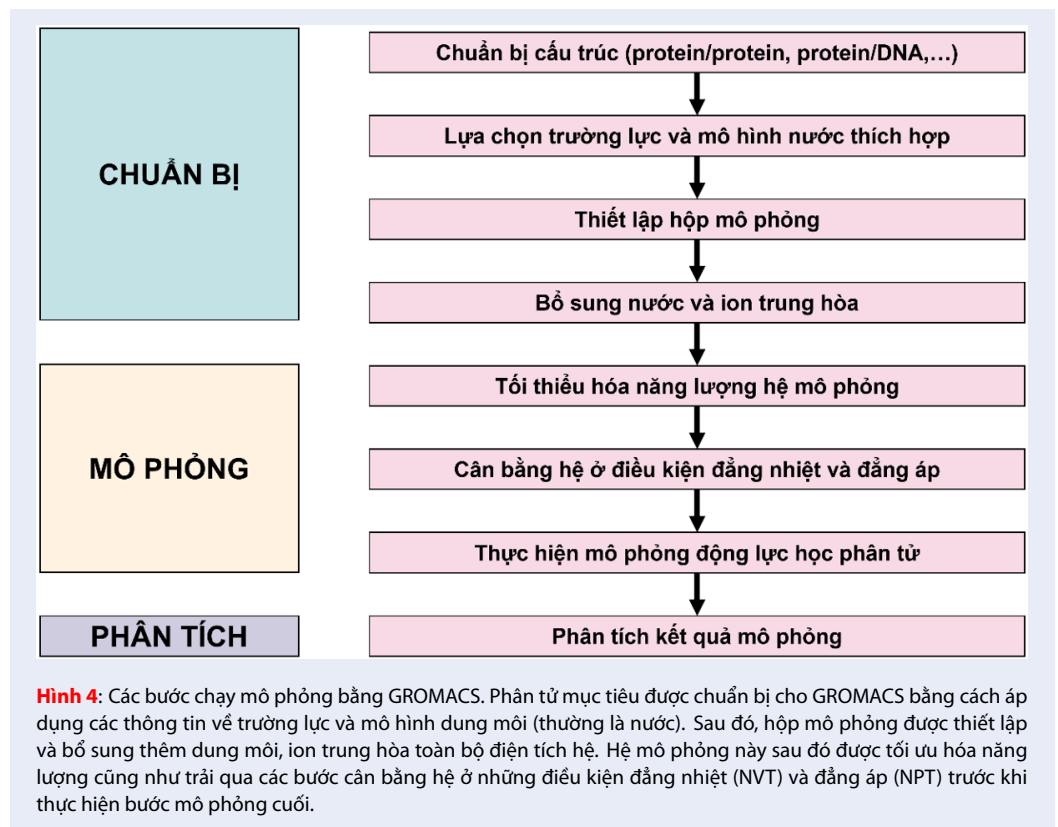
425 KẾT LUẬN

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

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

443 DANH MỤC TỪ VIẾT TẮT

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



448 MS-PCI: mass spectrometric protein complex
 449 AFM: atomic force microscopy
 450 TFM: template-free modeling
 451 TBM: template-based modeling
 452 VdW: Van der Waals
 453 CHARMM: chemistry at Harvard molecular me-
 454 chanic
 455 AMBER: assisted model building with energy refine-
 456 ment
 457 GROMACS: groningen machine for chemical simu-
 458 lations
 459 NAMD: nanoscale molecular dynamics
 460 MCA: multiple copy algorithms
 461 GPCR: G protein-coupled receptor

462 XUNG ĐỘT LỢI ÍCH

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

464 ĐÓNG GÓP CỦA CÁC TÁC GIẢ

465 Nguyễn Văn Minh Thường: viết, tổng hợp và chỉnh
 466 sửa bản thảo
 467 Lý Cẩm Tú: viết, tổng hợp và chỉnh sửa bản thảo
 468 Đinh Thuận Thiên: lên ý tưởng, viết, tổng hợp và
 469 chỉnh sửa bản thảo
 470 Trần Văn Hiếu: lên ý tưởng, tham gia chỉnh sửa bản
 471 thảo và chấp thuận bản thảo

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

TÀI LIỆU THAM KHẢO

- 474 1. Berggård T, Linse S, James P. Methods for the detection
 475 and analysis of protein-protein interactions. *Proteomics*.
 476 2007;7(16):2833–2842. Available from: <https://doi.org/10.1002/pmic.200700131>.
 477
- 478 2. Gonzalez MW, Kann MG. Chapter 4: Protein interactions and
 479 disease. *PLoS computational biology*. 2012;8(12):1002819–
 480 1002819. Available from: <https://doi.org/10.1371/journal.pcbi.1002819>.
 481
- 482 3. Abbasi WA, Asif A, Ben-Hur A. Learning protein binding
 483 affinity using privileged information. *BMC bioinformatics*.
 484 2018;19(1):1–12. Available from: <https://doi.org/10.1186/s12859-018-2448-z>.
 485
- 486 4. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki YJPOT-
 487 NAOS; 2001. Available from: <https://doi.org/10.1073/pnas.061034498>.
 488
- 489 5. Tang Z, Takahashi Y. Analysis of Protein-Protein Interaction by
 490 Co-IP in Human Cells. *Methods Mol Biol*. 2018;1794:289–296.
 491 Available from: https://doi.org/10.1007/978-1-4939-7871-7_20.
 492
- 493 6. Ho Y; 2002.
- 494 7. Mrowka R, Patzak A, Herzel HJGR. Is there a bias in proteome
 495 research?; 2001. Available from: <https://doi.org/10.1101/gr.206701>.
 496
- 497 8. You ZH, et al. Highly efficient framework for predicting inter-
 498 actions between proteins. vol. 47, no. 3, pp. 731–743, 2016;.
- 499 9. Bourgeat L, et al. Experimental Protein Molecular
 500 Dynamics: Broadband Dielectric Spectroscopy cou-
 501 pled with nanoconfinement,” *Scientific Reports*, vol.
 502 9, no. 1, p. 17988, 2019/11/29 2019; Available from:
 503 <https://doi.org/10.1038/s41598-019-54562-8>.

- 504 10. Chong WL. Protein-protein interactions: Insight from molec-
505 ular dynamics simulations and nanoparticle tracking analysis.
506 *Molecules*, vol. 26, no. 18, p. 5696, 2021;.
- 507 11. Rocha JJ, Whitaker JW, O'Connell JD, et al. Functional un-
508 knownomics: Systematic screening of conserved genes of un-
509 known function. *PLoS Biol.* 2023;21(8):e3002222;.
- 510 12. Jansen R, Yu H, Greenbaum D, et al. A Bayesian networks ap-
511 proach for predicting protein-protein interactions from ge-
512 nomic data. *Science.* 2003;302(5644):449-453;.
- 513 13. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available
514 to model nearly all complexes of structurally characterized
515 proteins. *Proc Natl Acad Sci U S A.* 2012;109(24):9438-9441;.
- 516 14. Cai W, Shao X, Maigret B. Protein-ligand recognition using
517 spherical harmonic molecular surfaces: towards a fast and
518 efficient filter for large virtual throughput screening. *J Mol*
519 *Graph Model.* 2002;20(4):313-328;.
- 520 15. Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM. Real
521 spherical harmonic expansion coefficients as 3D shape de-
522 scriptors for protein binding pocket and ligand compar-
523 isons. *Bioinformatics.* 2005;21(10):2347-2355; Available from:
524 :10.1093/bioinformatics/bti355.
- 525 16. Zhang Q, Sanner M, Olson AJ. Shape complementarity of
526 protein-protein complexes at multiple resolutions. *Proteins.*
527 2009;75(2):453-467;.
- 528 17. Vajda S, Camacho CJ. Protein-protein docking: is the glass
529 half-full or half-empty? *Trends Biotechnol.* 2004;22(3):110-
530 116;.
- 531 18. Hiếu TV, Thiên ĐT, Thường NVM, Dũng TD. Protein struc-
532 ture modeling using cloud-based servers. *J Sci Technol.*
533 2024;8(1):2828-2837;.
- 534 19. Huang S-Y. Search strategies and evaluation in protein-
535 protein docking: principles, advances and challenges. *Drug*
536 *Discov Today.* 2014;19(8):1081-1096;.
- 537 20. Zhang Z, Schindler CEM, Lange OF, Zacharias M. Applica-
538 tion of enhanced sampling Monte Carlo methods for high-
539 resolution protein-protein docking in Rosetta. *PLoS One.*
540 2015;10(6):e0125941;.
- 541 21. Torchala M, Moal IH, Chaleil RAG, Fernandez-Recio J, Bates
542 PA. Enhanced sampling of protein conformational states for
543 dynamic cross-docking within the protein-protein docking
544 server SwarmDock. *Proteins.* 2020;88(8):962-972;.
- 545 22. Kaynak BT, et al. Sampling of protein conformational space us-
546 ing hybrid simulations: A critical assessment of recent meth-
547 ods. *Front Mol Biosci.* 2022;9:821314;.
- 548 23. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a
549 fully automated algorithm for protein-protein docking. *Nucleic*
550 *Acids Res.* 2004;32(Web Server issue):W96-W99;.
- 551 24. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for
552 protein-protein docking. *Nat Protoc.* 2017;12(2):255-278;.
- 553 25. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al. The HAD-
554 DOCK2.2 Web Server: User-Friendly Integrative Modeling of
555 Biomolecular Complexes. *J Mol Biol.* 2016;428(4):720-725;.
- 556 26. Moreira IS, Fernandes PA, Ramos MJ. Protein-protein docking
557 dealing with the unknown. *J Comput Chem.* 2010;31(2):317-
558 342;.
- 559 27. Chothia C, Lesk AM. The relation between the divergence of
560 sequence and structure in proteins. *EMBO J.* 1986;5(4):823-
561 826;.
- 562 28. Kundrotas PJ, Vakser IA, Janin J. Structural templates for mod-
563 eling homodimers. *Protein Sci.* 2013;22(11):1655-1663;.
- 564 29. Chen H, Skolnick J. M-TASSER: an algorithm for protein quar-
565 ternary structure prediction. *Biophys J.* 2008;94(3):918-928;.
- 566 30. Szilagyí A, Zhang Y. Template-based structure model-
567 ing of protein-protein interactions. *Curr Opin Struct Biol.*
568 2014;24:10-23;.
- 569 31. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello
570 G, Gumienny R, et al. SWISS-MODEL: homology modelling
571 of protein structures and complexes. *Nucleic Acids Res.*
572 2018;46(W1):W296-303;.
- 573 32. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Mod-
574 eling protein quaternary structure of homo- and hetero-
oligomers beyond binary interactions by homology. *Sci Rep.* 575
2017;7(1):10480;.
- 576 33. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger
577 O, et al. Highly accurate protein structure prediction with Al-
578 phaFold. *Nature.* 2021;596(7873):583-9;.
- 579 34. Skolnick J, Gao M, Zhou H, Singh S. AlphaFold 2: Why It Works
580 and Its Implications for Understanding the Relationships of
581 Protein Sequence, Structure, and Function. *J Chem Inf Model.*
582 2021;61(10):4827-31;.
- 583 35. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green
584 T, et al. Protein complex prediction with AlphaFold-Multimer.
585 *bioRxiv.* 2022;.
- 586 36. Liu J, Gao M, Skolnick J. Enhancing AlphaFold-Multimer-
587 based protein complex structure prediction with MULTICOM
588 in CASP15. *bioRxiv.* 2023;.
- 589 37. Wallner BJB. AFsample: improving multimer prediction
590 with AlphaFold using massive sampling. *Bioinformatics.*
591 2023;39(9):btad573;.
- 592 38. Bryant P, Noé FJ. Improved protein complex prediction with
593 AlphaFold-multimer by denoising the MSA profile. *bioRxiv.*
594 2023;.
- 595 39. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock
596 Vina 1.2.0: New Docking Methods, Expanded Force Field, and
597 Python Bindings. *J Chem Inf Model.* 2021;61(8):3891-8;.
- 598 40. Sethi A, Joshi K, Sasikala KA, Alvala MJDD. Molecular docking
599 in modern drug discovery: principles and recent applications.
600 *Drug Discov Today.* 2019;24(3):578-88;.
- 601 41. Huang SY, Zou X. Advances and challenges in protein-ligand
602 docking. *Int J Mol Sci.* 2010;11(8):3016-34;.
- 603 42. Alder BJ, Wainwright TE. Phase transition for a hard sphere sys-
604 tem. *J Chem Phys.* 1957;27(5):1208-9;.
- 605 43. Karplus M, McCammon JA. Molecular dynamics simulations of
606 biomolecules. *Nat Struct Biol.* 2002;9(9):646-52;.
- 607 44. McCammon JA, Gelin BR, Karplus M. Dynamics of folded pro-
608 teins. *Nature.* 1977;267(5612):585-90;.
- 609 45. Shi L, Quick M, Zhao Y, Weinstein H, Javitch JA. The mechanism
610 of a neurotransmitter: sodium symporter—inward release of
611 Na⁺ and substrate is triggered by substrate in a second bind-
612 ing site. *Mol Cell.* 2008;30(6):667-77;.
- 613 46. Growth of molecular dynamics simulations in structural bi-
614 ology. *Web of Knowledge [Internet].* 2018 Mar 25 [cited
615 2024]; Available from: <https://webofknowledge.com/>.
- 616 47. Guvench O, MacKerell AD. Comparison of protein force fields
617 for molecular dynamics simulations. In: Kukol A, editor.
618 *Molecular Modeling of Proteins.* Totowa, NJ: Humana Press;
619 2008. p. 63-88;.
- 620 48. Lin F-Y, MacKerell AD. Force fields for small molecules. In:
621 Bonomi M, Camilloni C, editors. *Biomolecular Simulations:*
622 *Methods and Protocols.* New York, NY: Springer New York;
623 2019. p. 21-54;.
- 624 49. Vanommeslaeghe K, Guvench O, MacKerell AD Jr. Molecular
625 mechanics. *Curr Pharm Des.* 2014;20(20):3281-92;.
- 626 50. Burkowski FJ. *Structural bioinformatics: an algorithmic ap-
627 proach.* Boca Raton, FL: Chapman and Hall/CRC; 2008;.
- 628 51. González M. Force fields and molecular dynamics simulations.
629 *École thématique de la Société Française de la Neutronique.*
630 2011;12:169-200;.
- 631 52. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ,
632 Mondragon-Ramirez C, et al. Update of the CHARMM all-atom
633 additive force field for lipids: validation on six lipid types. *J*
634 *Phys Chem B.* 2010;114(23):7830-43;.
- 635 53. Moore PB, Lopez CF, Klein ML. Dynamical properties of a hy-
636 drated lipid bilayer from a multianosecond molecular dy-
637 namics simulation. *Biophys J.* 2001;81(5):2484-93;.
- 638 54. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH.
639 The MARTINI force field: coarse grained model for biomolec-
640 ular simulations. *J Phys Chem B.* 2007;111(27):7812-24;.
- 641 55. Kar P, Gopal SM, Cheng YM, Panahi A, Feig M. Transferring the
642 primo coarse-grained force field to the membrane environ-
643 ment: simulations of membrane proteins and helix-helix as-
644 sociation. *J Chem Theory Comput.* 2014;10(8):3459-72;.

- 646 56. Shi Q, Izvekov S, Voth GA. Mixed atomistic and coarse-grained
647 molecular dynamics: simulation of a membrane-bound ion
648 channel. *J Phys Chem B*. 2006;110(31):15045-8;.
- 649 57. Jorgensen WL, Tirado-Rives J. Potential energy functions for
650 atomic-level simulations of water and organic and biomolec-
651 ular systems. *Proc Natl Acad Sci U S A*. 2005;102(19):6665-70;.
- 652 58. Małolepsza E, Strodel B, Khalili M, Trygubenko S, Fejer SN,
653 Wales DJ. Symmetrization of the AMBER and CHARMM force
654 fields. *J Comput Chem*. 2010;31(7):1402-9;.
- 655 59. Case DA, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR,
656 Cerutti DS, et al. AMBER 2022 Reference Manual. 2022 Mar
657 24; Available from: <https://ambermd.org/doc12/Amber22.pdf>.
- 658 60. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A
659 biomolecular force field based on the free enthalpy of hydra-
660 tion and solvation: the GROMOS force-field parameter sets
661 53A5 and 53A6. *J Comput Chem*. 2004;25(13):1656-76;.
- 662 61. Soares TA, Hünenberger PH, Kastenholz MA, Kräutler V, Lenz T,
663 Oostenbrink C, et al. An improved nucleic acid parameter set
664 for the GROMOS force field. *J Comput Chem*. 2005;26(7):725-
665 37;.
- 666 62. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et
667 al. OPLS3: a force field providing broad coverage of drug-
668 like small molecules and proteins. *J Chem Theory Comput*.
669 2016;12(1):281-96;.
- 670 63. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux
671 B, et al. CHARMM: The biomolecular simulation program. *J*
672 *Comput Chem*. 2009;30(10):1545-1614;.
- 673 64. Brooks BR, et al. CHARMM: A program for macromolecular
674 energy, minimization, and dynamics calculations. *Journal of*
675 *Computational Chemistry*, vol. 4, no. 2, pp. 187-217, 1983;.
- 676 65. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM,
677 et al. The Amber biomolecular simulation programs. *J Comput*
678 *Chem*. 2005;26(16):1668-88;.
- 679 66. Case DA. The Amber biomolecular simulation programs. *Jour-
680 nal of computational chemistry*, vol. 26, no. 16, pp. 1668-1688,
681 2005;.
- 682 67. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al.
683 GROMACS: High performance molecular simulations through
684 multi-level parallelism from laptops to supercomputers. *Soft-
685 wareX*. 2015;1:19-25;.
- 686 68. Sharma S, Kumar P, Chandra R, Singh S, Mandal A, Donda-
687 pati R. Overview of BIOVIA materials studio, LAMMPS, and
688 GROMACS. *Mol Dyn Simul Nanocompos BIOVIA Mater Stud*
689 *Lammps Gromacs*. Amsterdam: Elsevier; 2019. p. 39-100;.
- 690 69. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE,
691 Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput*
692 *Chem*. 2005;26(16):1701-18;.
- 693 70. Irrgang ME, Davis C, Kasson PM. gmxapi: A GROMACS-native
694 Python interface for molecular dynamics with ensemble and
695 plugin support. *PLoS Comput Biol*. 2022;18(2):e1009835;.
- 696 71. Smith A, Dong X, Raghavan V. An overview of molecular
697 dynamics simulation for food products and processes. *Pro-
698 cesses*. 2022;10(1):119;.
- 699 72. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi
700 RC, et al. Scalable molecular dynamics on CPU and GPU archi-
701 tectures with NAMD. *J Chem Phys*. 2020;153(4):044130;.
- 702 73. Jiang W, Hardy DJ, Phillips JC, MacKerell AD, Schulten K, Roux
703 B. Generalized Scalable Multiple Copy Algorithms for Molec-
704 ular Dynamics Simulations in NAMD. *Comput Phys Commun*.
705 2014;185(3):908-16;.
- 706 74. Oláh J, Szénási T, Lehotzky A, Norris V, Ovádi J. Challenges in
707 discovering drugs that target the protein-protein interactions
708 of disordered proteins. *Int J Mol Sci*. 2022;23(3):1550;.
- 709 75. Hu L, Wang X, Huang YA, Hu P, You ZH. A survey on compu-
710 tational models for predicting protein-protein interactions.
711 *Brief Bioinform*. 2021;22(5):bbab036;.
- 712 76. Soleymani F, Paquet E, Viktor H, Michalowski W, Spinello
713 D. Protein-protein interaction prediction with deep learn-
714 ing: A comprehensive review. *Comput Struct Biotechnol J*.
715 2022;20:3701-17;.
- 716 77. Biji I, Ramharack P, Khan S, Cherqaoui D, Soliman ME. Tracing
717 potential covalent inhibitors of an E3 ubiquitin ligase through
718 target-focused modelling. *Molecules*. 2019;24(17):3125;.
- 719 78. Cheng SS, Yang GJ, Wang W, Leung CH, Ma DL. The design
720 and development of covalent protein-protein interaction in-
721 hibitors for cancer treatment. *J Hematol Oncol*. 2020;13(1):26;.
- 722 79. Shaikat Z, Aiman S, Li CH. Protein-protein interactions: meth-
723 ods, databases, and applications in virus-host study. *World J*
724 *Virol*. 2021;10(6):288-303;.
- 725 80. Dror RO, Dirks RM, Grossman J, Xu H, Shaw DE. Biomolecular
726 simulation: a computational microscope for molecular biol-
727 ogy. *Annu Rev Biophys*. 2012;41:429-52;.
- 728 81. Heilker R, Wolff M, Tautermann CS, Bieler M. G-protein-
729 coupled receptor-focused drug discovery using a target class
730 platform approach. *Drug Discov Today*. 2009;14(5-6):231-40;.
- 731 82. Lundstrom K. Latest development in drug discovery
732 on G protein-coupled receptors. *Curr Protein Pept Sci*.
733 2006;7(5):465-70;.
- 734 83. Mizuno H, Kihara Y. Druggable lipid GPCRs: past, present,
735 and prospects. In: *Druggable Lipid Signaling Pathways*. Cham:
736 Springer; 2020. p. 223-58;.
- 737 84. Sinha S, Tam B, Wang SM. Applications of molecular dynamics
738 simulation in protein study. *Membranes*. 2022;12(9):844;.
- 739 85. Jangra S, Ye C, Rathnasinghe R, Stadlbauer D, Krammer F, Si-
740 mon V, et al. SARS-CoV-2 spike E484K mutation reduces anti-
741 body neutralisation. *Lancet Microbe*. 2021;2(7):e283-4;.
- 742 86. Acharya A, Agarwal R, Baker MB, Baudry J, Bhowmik D, Boehm
743 S, et al. Supercomputer-based ensemble docking drug discov-
744 ery pipeline with application to COVID-19. *J Chem Inf Model*.
745 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,*}



Use your smartphone to scan this QR code and download this article

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:

DOI :



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.* 2025; ():1-1.