

# TỔNG QUAN VÀ ĐÁNH GIÁ CÁC PHẦN MỀM SỬ DỤNG TRONG NGHIÊN CỨU PHÁT TRIỂN THUỐC MỚI IN SILICO

## TÓM TẮT

Nguyễn Hoàng Tuấn<sup>1</sup>, Nguyễn Nhật Hoàng<sup>2</sup>, Nguyễn Hoàng Ngân<sup>3\*</sup>

**Mục tiêu:** Tổng quan và đánh giá các phần mềm in silico theo nhóm chức năng, phục vụ thiết kế, sàng lọc và tối ưu hóa ứng viên thuốc.

**Phương pháp:** Nghiên cứu tổng quan có hệ thống các tài liệu khoa học công bố giai đoạn 2010–2024 trên PubMed, Scopus, ScienceDirect, và Google Scholar. Các bài báo được lựa chọn theo tiêu chí: mô tả/ứng dụng phần mềm in silico trong phát triển thuốc, có thông tin về chức năng, ưu điểm và hạn chế. Dữ liệu được trích xuất, tổng hợp và phân loại theo chức năng phần mềm trong chu trình R&D được phẩm.

**Kết quả:** Ghi nhận 6 nhóm phần mềm chính: (1) dự đoán đặc tính ADMET; (2) docking phân tử; (3) mô phỏng động học phân tử; (4) mô hình QSAR; (5) thiết kế hợp chất de novo; và (6) trí tuệ nhân tạo ứng dụng trong sàng lọc thuốc. Việc kết hợp các phần mềm trong một quy trình thống nhất giúp tăng độ chính xác và hiệu quả sàng lọc.

**Kết luận:** Các phần mềm in silico giúp rút ngắn thời gian, giảm chi phí và nâng cao hiệu quả sàng lọc thuốc. Đóng góp mới của nghiên cứu là xây dựng hệ thống phân loại rõ ràng, kèm đánh giá so sánh theo chức năng và khả năng ứng dụng, hỗ trợ lựa chọn công cụ phù hợp cho từng giai đoạn phát triển thuốc. Tuy vậy, chuẩn hóa dữ liệu và hiệu chỉnh mô hình vẫn là thách thức cần tiếp tục giải quyết.

**Từ khóa:** in silico, docking, ADMET, mô phỏng, trí tuệ nhân tạo, phát triển thuốc.

## OVERVIEW AND EVALUATION OF SOFTWARE TOOLS USED IN IN SILICO RESEARCH FOR NEW DRUG DEVELOPMENT

### ABSTRACT

**Objective:** To review and evaluate *in silico* software tools by functional categories, supporting

the design, screening, and optimization of potential drug candidates.

**Methods:** A systematic review of scientific literature published from 2010 to 2024 was conducted using PubMed, Scopus, ScienceDirect, and Google Scholar. Articles were selected if they described or applied *in silico* software in drug development and provided information on functions, advantages, and limitations. Data were extracted, synthesized, and classified according to software functions within the pharmaceutical R&D pipeline.

**Results:** Six major categories of software were identified: (1) ADMET property prediction; (2) molecular docking; (3) molecular dynamics simulation; (4) QSAR modeling; (5) de novo compound design; and (6) artificial intelligence-based drug screening. Integrating multiple tools into a unified workflow improved accuracy and screening efficiency.

**Conclusion:** *In silico* software helps shorten development time, reduce costs, and enhance drug screening efficiency. The novel contribution of this study is the development of a clear classification system with comparative evaluation by function and applicability, providing guidance for selecting appropriate tools at each stage of drug development. Nevertheless, data standardization and model calibration remain challenges that require continued attention.

**Keywords:** in silico, docking, ADMET, simulation, artificial intelligence, drug development.

### I. ĐẶT VẤN ĐỀ

Sự phát triển mạnh mẽ của khoa học máy tính, công nghệ tin sinh học và trí tuệ nhân tạo trong những thập niên gần đây đã mở ra một kỷ nguyên mới trong nghiên cứu phát triển thuốc. Thay vì phụ thuộc hoàn toàn vào các phương pháp truyền thống như sàng lọc thực nghiệm hay thử - sai tốn kém và mất nhiều thời gian, các công cụ tính toán *in silico* cho phép mô phỏng, dự đoán, và sàng lọc hàng nghìn hợp chất tiềm năng chỉ trong thời gian ngắn với chi phí thấp hơn đáng kể [1], [2]. Thực tế cho thấy, chỉ khoảng 10-12% ứng viên thuốc được phát triển thành công, trong khi thời gian trung bình để đưa một thuốc mới ra thị trường có thể kéo dài

1. Trường THPT chuyên Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

2. Trường Đại học Dược Hà Nội

3. Học viện Quân y

\* Tác giả liên hệ: Nguyễn Hoàng Ngân

Email: nganvnu@gmail.com

Ngày nhận bài: 15/10/2025

Ngày phản biện: 15/11/2025

Ngày duyệt bài: 30/11/2025

10–15 năm với chi phí lên tới hàng tỷ USD. Nhờ tích hợp *in silico* vào quy trình R&D, nhiều nghiên cứu ghi nhận chi phí có thể giảm 40-60% và thời gian phát hiện - tối ưu hóa hợp chất được rút ngắn rõ rệt [3].

Các công cụ *in silico* hiện đã được ứng dụng rộng rãi trong hầu hết các giai đoạn phát triển thuốc, bao gồm thiết kế cấu trúc phân tử, dự đoán đặc tính ADMET, docking phân tử, xây dựng mô hình QSAR, mô phỏng động lực học phân tử, cũng như ứng dụng trí tuệ nhân tạo trong sàng lọc và tối ưu hóa hợp chất. Tuy nhiên, sự đa dạng và số lượng lớn phần mềm hiện có - với khác biệt về thuật toán, độ chính xác, chi phí và phạm vi áp dụng - đang đặt ra thách thức lớn cho các nhà nghiên cứu trong việc lựa chọn công cụ phù hợp cho từng mục tiêu cụ thể. Đáng chú ý, các báo cáo tổng hợp trong nước hiện còn rời rạc, thiếu những tổng quan có hệ thống giúp so sánh, phân loại và đánh giá một cách toàn diện các phần mềm theo chức năng sử dụng. Điều này dẫn đến nguy cơ lựa chọn công cụ chưa tối ưu, lặp lại thử nghiệm hoặc khai thác chưa đầy đủ nguồn lực tính toán.

Trong bối cảnh các phương pháp *in silico*, đặc biệt là trí tuệ nhân tạo và mô phỏng quy mô lớn, phát triển rất nhanh, việc cập nhật và hệ thống hóa thông tin về các phần mềm phục vụ phát triển thuốc trở nên cấp thiết. Nghiên cứu này nhằm tổng hợp, phân loại và đánh giá các phần mềm *in silico* được sử dụng phổ biến trong giai đoạn 2010-2024, ở cả trong nước và trên thế giới. Trên cơ sở phân tích các tài liệu khoa học đã công bố, nghiên cứu làm rõ xu hướng phát triển, ưu điểm, hạn chế và tiềm năng ứng dụng của từng nhóm công cụ, đồng thời đề xuất định hướng lựa chọn và tích hợp phần mềm phù hợp trong thực hành nghiên cứu. Mục tiêu cuối cùng là góp phần tối ưu hóa quy trình thiết kế và sàng lọc thuốc, nâng cao hiệu quả ứng dụng công nghệ *in silico* trong nghiên cứu dược học.

## II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Đối tượng, địa điểm và thời gian nghiên cứu

**Đối tượng nghiên cứu:** Các phần mềm, cơ sở dữ liệu và công cụ tính toán được sử dụng trong thiết kế và phát triển thuốc *in silico*.

**Tiêu chuẩn lựa chọn:** Bao gồm các bài báo khoa học mô tả, đánh giá hoặc ứng dụng phần mềm *in silico* trong các giai đoạn phát triển thuốc, được công bố giai đoạn 2010-2024, có trình bày rõ chức năng, phạm vi ứng dụng và đặc tính kỹ thuật.

**Tiêu chuẩn loại trừ:** Loại các bài viết không có toàn văn, thiếu thông tin về phần mềm, không phải bài có phản biện khoa học, hoặc chỉ mang tính lý thuyết mà không có ứng dụng/đánh giá cụ thể.

**Địa điểm nghiên cứu:** Học viện Quân y.

**Thời gian thực hiện:** Từ tháng 01 năm 2024 đến tháng 09 năm 2025.

### 2.2. Phương pháp nghiên cứu

**Thiết kế nghiên cứu:** Tổng quan có hệ thống (systematic review), tuân thủ khuyến cáo PRISMA 2020.

**Cỡ mẫu và chọn mẫu:** Không áp dụng chọn mẫu định lượng; các tài liệu được lựa chọn có chủ đích theo tiêu chí xác định.

**Nguồn dữ liệu và truy vấn tìm kiếm:** Tài liệu được tìm kiếm trên PubMed, Scopus, ScienceDirect và Google Scholar. Chuỗi truy vấn được xây dựng dựa trên tổ hợp từ khóa MeSH/keywords:

("in silico drug design" OR "computer-aided drug design")

AND ("software" OR "tool" OR "platform")

AND ("docking" OR "QSAR" OR "molecular dynamics" OR "ADMET" OR "AI")

AND (Publication years: 2010–2024).

Bộ lọc áp dụng: bài báo có phản biện, tiếng Anh/Việt, lĩnh vực dược học - sinh học - tin sinh học. Dữ liệu được tìm kiếm đến ngày 15/10/2024; các bước sàng lọc và phân tích được thực hiện đến 09/2025.

### Quy trình sàng lọc tài liệu

Tài liệu thu được được xử lý theo ba bước: 1. Loại trùng lặp bằng EndNote/Mendeley; 2. Sàng lọc tiêu đề và tóm tắt theo tiêu chí đề ra; 3. Đọc toàn văn, loại bài không đáp ứng nội dung hoặc thiếu dữ liệu. Qua quá trình tìm kiếm và sàng lọc tài liệu (842 bài ban đầu, còn 693 bài sau loại trùng lặp, 262 bài được đọc toàn văn và cuối cùng lựa chọn 112 bài - gồm 84 bài đánh giá phần mềm chính và 28 bài bổ trợ), tổng số bài đủ tiêu chuẩn được đưa vào phân tích là 112.

### Trích xuất và tổng hợp dữ liệu

Dữ liệu từ mỗi bài được trích xuất theo mẫu chuẩn gồm: tên phần mềm/công cụ; mục đích sử dụng; thuật toán/cơ chế; ưu điểm – hạn chế; lĩnh vực ứng dụng; minh chứng hiệu quả (nếu có).

Sau đó, các công cụ được phân loại theo chức năng: thiết kế cấu trúc, docking, mô phỏng MD, QSAR, dự đoán ADMET, AI và thiết kế *de novo*.

### Đánh giá chất lượng & kiểm soát thiên lệch

Mỗi bài được đánh giá độc lập bởi hai nhà nghiên cứu, theo các tiêu chí: tính phù hợp chủ đề; độ rõ ràng phương pháp; tính minh bạch phần mềm và dữ liệu; giá trị ứng dụng.

Các bất đồng được thảo luận và thống nhất bởi người thứ ba. Phân tích tổng hợp mang tính định

tính, nhằm nhận diện xu hướng, ưu/nhược điểm và khoảng trống nghiên cứu.

**2.3. Đạo đức nghiên cứu:** Nghiên cứu là bài tổng quan sử dụng hoàn toàn dữ liệu công khai từ các nguồn khoa học hợp pháp, không can thiệp

vào người hay động vật thí nghiệm. Toàn bộ quá trình thực hiện tuân thủ quy định về đạo đức trong nghiên cứu khoa học, đảm bảo tính minh bạch, khách quan, trung thực trong thu thập, trích dẫn và phân tích dữ liệu.

### III. KẾT QUẢ NGHIÊN CỨU

#### 3.1. Khái niệm, đặc điểm và phạm vi ứng dụng của nghiên cứu *in silico*

Nghiên cứu thuốc *in silico* (hay Computer-Aided Drug Design/Discovery – CADD) là lĩnh vực ứng dụng các mô hình tính toán, thuật toán và phần mềm chuyên dụng để mô phỏng, dự đoán và tối ưu hóa tương tác thuốc-đích sinh học, hỗ trợ toàn bộ quá trình phát hiện và phát triển thuốc mới [4]. Nhờ khả năng khảo sát không gian hóa học lớn và sàng lọc đồng thời hàng nghìn hợp chất, *in silico* giúp giảm đáng kể số lượng thí nghiệm tiền lâm sàng cần thiết. Một số phân tích kinh tế - kỹ thuật cho thấy việc tích hợp các bước sàng lọc ảo, QSAR và dự đoán ADMET có thể loại bỏ sớm 30-50% hợp chất không phù hợp trước khi tiến hành thí nghiệm *in vitro*, qua đó rút ngắn thời gian nghiên cứu từ giai đoạn phát hiện đến tối ưu hóa thuốc (lead optimization) khoảng 1-3 năm trong nhiều dự án phát triển thuốc. Trong chu trình phát triển thuốc, các công cụ *in silico* được ứng dụng tại nhiều điểm nút quan trọng: (i) sàng lọc ảo để xác định các hợp chất có khả năng gắn kết tốt; (ii) mô phỏng động lực học phân tử nhằm đánh giá độ ổn định phức hợp; (iii) mô hình QSAR để dự đoán hoạt tính sinh học dựa trên cấu trúc; (iv) dự đoán ADMET để phát hiện sớm nguy cơ độc tính hoặc hấp thu kém [4], [5]. Nhờ sự phát triển của thuật toán và dữ liệu sinh học - tin sinh học quy mô lớn, *in silico* ngày càng đóng vai trò trung tâm, hướng tới mục tiêu thiết kế thuốc chính xác và cá thể hóa.

#### 3.2. Các hướng tiếp cận chính trong thiết kế và phát triển thuốc *in silico*

Thiết kế thuốc *in silico* hiện tập trung vào bốn hướng tiếp cận chính. (1) Thiết kế dựa trên cấu trúc (SBDD): dựa trên cấu trúc 3D của protein đích để dự đoán tư thế gắn kết và năng lượng liên kết của ligand, hỗ trợ phát hiện hợp chất dẫn với độ chọn lọc cao [6]. (2) Thiết kế dựa trên phối tử (LBDD): được áp dụng khi cấu trúc đích chưa rõ; sử dụng tương đồng cấu trúc, QSAR và phân lớp học máy để dự đoán hợp chất mới có hoạt tính tương tự phối tử đã biết. (3) Các công cụ mô tả đặc tính dược lý: bao gồm pharmacophore modeling, mô phỏng động lực học phân tử và các mô hình dự đoán hoạt tính, giúp làm rõ đặc điểm không gian và tính ổn định của phức hợp ligand-protein. (4) Trí tuệ nhân tạo (AI) và học máy (ML): ngày càng đóng vai trò then chốt. Các mô hình deep learning (graph neural networks, transformer-based) hỗ trợ dự đoán cấu trúc protein (AlphaFold), sàng lọc hợp chất (DeepChem) và thiết kế thuốc *de novo* (ChemBERTa, REINVENT) [5], [7]. Một số nghiên cứu gần đây ghi nhận AI có thể tăng độ chính xác dự đoán hoạt tính 10-20% so với mô hình truyền thống trong một số bộ dữ liệu chuẩn, đồng thời giảm đáng kể số thí nghiệm cần xác nhận [5], [7].

#### 3.3. Phân loại các nhóm phần mềm và công cụ được sử dụng trong nghiên cứu phát triển thuốc

Tổng hợp 112 tài liệu cho thấy các phần mềm *in silico* có thể chia thành sáu nhóm chức năng chính.

##### 3.3.1. Phần mềm thiết kế cấu trúc phân tử và tối ưu hóa hợp chất

Đây là nhóm phần mềm được sử dụng ở giai đoạn đầu của quá trình thiết kế thuốc, giúp xây dựng, hiển thị và tối ưu hóa cấu trúc phân tử ba chiều (3D). ChemDraw và Avogadro hỗ trợ vẽ cấu trúc 2D, chuyển đổi sang mô hình 3D và tối ưu hình học; OpenBabel giúp chuyển đổi định dạng tệp và tối ưu cấu trúc tự động; trong khi Gaussian và GaussView thực hiện phân tích hóa lượng tử, tính toán năng lượng và đặc tính điện tử [6]. Nhóm công cụ này đóng vai trò cung cấp dữ liệu đầu vào chuẩn hóa; nhiều nghiên cứu ghi nhận rằng việc tối ưu hình học trước docking giúp cải thiện điểm số gắn kết từ 5–15% so với cấu trúc chưa chuẩn hóa [4], [6].

##### 3.3.2. Phần mềm sàng lọc ảo (Virtual Screening) và Docking phân tử

Sàng lọc ảo và docking là bước quan trọng nhất trong thiết kế thuốc dựa trên cấu trúc (SBDD), giúp nhanh chóng xác định các hợp chất có khả năng gắn kết mạnh và chọn lọc với đích sinh học. Các phần mềm như AutoDock và AutoDock Vina (mã nguồn mở) hỗ trợ dự đoán vị trí và ái lực liên kết ligand-protein, trong khi Glide và GOLD (thương mại) sử dụng thuật toán tiên tiến để xác định tư thế gắn kết tối ưu. SwissDock và DOCK là các nền tảng trực tuyến thuận tiện, cho phép sàng lọc quy mô lớn mà không

cần cấu hình máy tính phức tạp [4], [6]. Nhờ đó, nhóm công cụ này giúp rút ngắn thời gian phát hiện hợp chất dẫn và giảm số thí nghiệm *in vitro*, *in vivo*. Một số nghiên cứu so sánh cho thấy phần mềm thương mại (Glide, GOLD) thường cho hệ số tương quan tốt hơn ( $r = 0,55-0,70$ ) so với Vina ( $r = 0,40-0,55$ ), nhưng chi phí cao và cần giấy phép sử dụng [4], [6].

### 3.3.3. Phần mềm mô phỏng động lực học phân tử (Molecular Dynamics - MD)

Các phần mềm mô phỏng động lực học phân tử (MD) cho phép theo dõi sự chuyển động và độ ổn định của phức hợp ligand-protein theo thời gian, giúp làm rõ cơ chế gắn kết trong điều kiện gần với môi trường sinh lý. Các công cụ phổ biến gồm GROMACS (mã nguồn mở, hiệu năng cao), AMBER và CHARMM (cung cấp bộ trường lực chính xác cho protein, acid nucleic và phân tử nhỏ), cùng NAMD (tối ưu cho tính toán song song) [5], [6]. Nhiều nghiên cứu cho thấy MD giúp giảm tỷ lệ “dương giả” sau docking nhờ xem xét thêm tác động của dung môi và chuyển động phân tử [4].

### 3.3.4. Phần mềm dự đoán dược động học và độc tính (ADMET Prediction)

Dự đoán ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) là bước quan trọng trong sàng lọc ảo, giúp loại trừ sớm các hợp chất có đặc tính dược động học kém hoặc độc tính cao. Các công cụ phổ biến như SwissADME và pkCSM dựa trên mô hình học máy cho phép dự đoán khả năng hấp thu, phân bố và chuyển hóa; ADMETlab mở rộng thêm dự đoán về độc tính, độ ổn định hóa học và tương tác protein huyết tương; trong khi Toxtree tập trung phân tích cảnh báo độc tính và cơ chế gây độc dựa trên cấu trúc hóa học [4], [5]. Trong một số nghiên cứu, việc áp dụng bộ công cụ ADMET trước khi thử nghiệm *in vitro* đã giúp giảm 25-35% số hợp chất cần kiểm tra, qua đó tiết kiệm chi phí đáng kể [4], [8].

### 3.3.5. Phần mềm phân tích và khai thác dữ liệu sinh học - tin sinh học

Cơ sở dữ liệu tin sinh học là nguồn đầu vào quan trọng cho mô hình hóa và huấn luyện thuật toán trong thiết kế thuốc. PubChem, ChEMBL, BindingDB cung cấp dữ liệu khổng lồ về cấu trúc hóa học, hoạt tính sinh học và tương tác ligand-protein, trong khi UniProt chứa thông tin toàn diện về protein (chức năng, chuỗi amino acid, vị trí hoạt động) và STRING hỗ trợ phân tích mạng lưới tương tác protein-protein, giúp xác định đích tác dụng tiềm năng [6]. Việc tích hợp các cơ sở dữ liệu này không chỉ nâng cao độ chính xác và tính toàn diện của mô hình *in silico* mà còn tạo nền tảng dữ liệu huấn luyện thiết yếu cho các thuật toán học máy và trí tuệ nhân tạo trong phát hiện và phát triển thuốc [5].

### 3.3.6. Phần mềm ứng dụng trí tuệ nhân tạo trong thiết kế thuốc

Sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) và học máy (Machine Learning – ML) đang tạo nên bước chuyển lớn trong lĩnh vực phát hiện và phát triển thuốc *in silico*. Các nền tảng như DeepChem và Chemprop cho phép xây dựng mô hình học sâu để dự đoán hoạt tính sinh học, đặc tính ADMET và tối ưu cấu trúc hợp chất [5]. AlphaFold của DeepMind đánh dấu bước ngoặt trong dự đoán cấu trúc protein 3D với độ chính xác gần tương đương tinh thể học [7], trong khi các mô hình sinh tổng hợp như MolGPT, REINVENT có khả năng thiết kế *de novo* hợp chất mới, rút ngắn đáng kể chu trình khám phá thuốc. Sự tích hợp giữa AI với các mô hình docking, động lực học phân tử (MD) và QSAR đang hướng tới hệ thống tự động hóa toàn diện, giúp nâng cao độ chính xác, hiệu quả và tốc độ trong nghiên cứu dược học hiện đại. Nhiều dự án báo cáo thời gian sàng lọc hàng triệu hợp chất có thể rút từ vài tháng xuống còn vài ngày nhờ tích hợp AI với các pipeline docking và QSAR tự động [4], [5].

**Bảng 1. Phân loại và chức năng các nhóm phần mềm trong phát triển thuốc *in silico***

| Nhóm phần mềm                                      | Chức năng chính                                                                                                                                                         | Một số phần mềm tiêu biểu                             | Giai đoạn ứng dụng trong phát triển thuốc    |
|----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|----------------------------------------------|
| Thiết kế cấu trúc phân tử và tối ưu hóa hợp chất   | Xây dựng, hiển thị và tối ưu hóa cấu trúc 3D của phân tử; tính toán năng lượng, phân bố điện tử và đặc tính hóa lượng tử; tạo dữ liệu đầu vào cho mô phỏng và sàng lọc. | ChemDraw, Avogadro, OpenBabel, Gaussian, GaussView    | Thiết kế ban đầu hợp chất (lead design)      |
| Sàng lọc ảo (Virtual Screening) và Docking phân tử | Dự đoán vị trí, tư thế và ái lực liên kết ligand-protein; xác định hợp chất có khả năng gắn kết mạnh và chọn lọc cao với đích sinh học.                                 | AutoDock, AutoDock Vina, GOLD, Glide, SwissDock, DOCK | Phát hiện hợp chất dẫn (lead identification) |

| Nhóm phần mềm                                           | Chức năng chính                                                                                                                         | Một số phần mềm tiêu biểu                       | Giai đoạn ứng dụng trong phát triển thuốc                         |
|---------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|-------------------------------------------------------------------|
| Mô phỏng động lực học phân tử (Molecular Dynamics – MD) | Mô phỏng sự chuyển động, ổn định và biến đổi cấu trúc của phức hợp ligand–protein; đánh giá độ bền và tính linh động của tương tác.     | GROMACS, AMBER, CHARMM, NAMD                    | Tối ưu hóa tương tác thuốc–đích (lead optimization)               |
| Dự đoán dược động học và độc tính (ADMET Prediction)    | Dự đoán hấp thu, phân bố, chuyển hóa, thải trừ và độc tính; loại trừ sớm các hợp chất có đặc tính không phù hợp hoặc độc tính cao.      | SwissADME, pkCSM, ADMETlab, Toxtree             | Sàng lọc và đánh giá an toàn tiền lâm sàng                        |
| Phân tích và khai thác dữ liệu sinh học – tin sinh học  | Cung cấp và quản lý dữ liệu hóa học – sinh học; hỗ trợ phân tích tương tác protein–protein, ligand–protein và huấn luyện mô hình AI/ML. | PubChem, ChEMBL, BindingDB, UniProt, STRING     | Xác định đích sinh học và xây dựng mô hình dự đoán                |
| Ứng dụng trí tuệ nhân tạo trong thiết kế thuốc          | Xây dựng mô hình học sâu để dự đoán hoạt tính, ADMET, thiết kế de novo hợp chất và tự động hóa quy trình nghiên cứu.                    | DeepChem, Chemprop, AlphaFold, MolGPT, REINVENT | Tối ưu hóa hợp chất và thiết kế thuốc mới (AI-driven drug design) |

(Nguồn: Tổng hợp của nhóm nghiên cứu từ 112 tài liệu công bố giai đoạn 2010-2024).

### 3.4. Tiêu chí đánh giá phần mềm

Việc lựa chọn phần mềm *in silico* phù hợp có vai trò quyết định đối với độ chính xác và hiệu quả của nghiên cứu, dựa trên bốn tiêu chí chính [4], [6]. Thứ nhất, tính dễ sử dụng và yêu cầu kỹ thuật, bao gồm giao diện trực quan, hỗ trợ đa nền tảng và hướng dẫn chi tiết, giúp thao tác thuận tiện ngay cả với người không chuyên lập trình. Thứ hai, độ chính xác và độ tin cậy của mô hình, phụ thuộc vào thuật toán tính toán và chất lượng dữ liệu đầu vào; các phần mềm thương mại như Glide hay GOLD có độ chính xác cao, trong khi AutoDock Vina được đánh giá tốt nhờ khả năng tái lập kết quả [4]. Thứ ba, tính sẵn có và chi phí, khi các phần mềm mã nguồn mở như GROMACS, AutoDock, SwissADME dễ tiếp cận hơn so với các công cụ thương mại như Schrödinger Suite hay Discovery Studio [5]. Cuối cùng, khả năng tương thích và tích hợp giữa các công cụ (định dạng pdbqt, mol2, sdf) là yếu tố then chốt để xây dựng pipeline thống nhất; các nền tảng tích hợp học máy và cơ sở dữ liệu sinh học hiện được ưu tiên trong dược học tính toán hiện đại [7]. Ở nhiều nghiên cứu so sánh, phần mềm mã nguồn mở (như Vina, GROMACS, SwissADME) có lợi thế chi phí, trong khi một số phần mềm thương mại ghi nhận độ chính xác cao hơn trong một số bài toán đặc thù [6], [7].

### 3.5. So sánh ưu và nhược điểm giữa các nhóm phần mềm

Mỗi nhóm phần mềm *in silico* đều có ưu thế và hạn chế riêng, phụ thuộc vào mục tiêu nghiên cứu và năng lực tính toán sẵn có. Nhóm thiết kế cấu trúc và tối ưu hóa hợp chất (ChemDraw, Avogadro, Gaussian) có ưu điểm dễ sử dụng, phù hợp cho việc tạo mô hình ban đầu nhưng hạn chế khi mô phỏng các tương tác phân tử phức tạp [6]. Nhóm sàng lọc ảo và docking phân tử (AutoDock, Glide, GOLD) có khả năng xử lý hàng nghìn hợp chất, song độ chính xác phụ thuộc vào thuật toán và điều kiện docking [4]. Nhóm mô phỏng động lực học phân tử (MD) (GROMACS, AMBER, NAMD) đạt độ chính xác cao, mô phỏng được tương tác ở cấp độ nguyên tử, nhưng yêu cầu phần cứng mạnh và thời gian tính toán dài [5]. Nhóm dự đoán ADMET (SwissADME, pkCSM, ADMETlab) có ưu điểm nhanh, dễ sử dụng, phù hợp cho sàng lọc ban đầu, nhưng còn hạn chế về độ chính xác do phụ thuộc vào mô hình học máy [8]. Trong khi đó, nhóm thiết kế dựa trên trí tuệ nhân tạo (AI) (DeepChem, Chemprop, AlphaFold, MolGPT) thể hiện tiềm năng vượt trội trong dự đoán cấu trúc và thiết kế hợp chất de novo, song còn đối mặt với thách thức về tính minh bạch của thuật toán và nhu cầu dữ liệu huấn luyện lớn [5], [7]. Sự kết hợp nhiều công cụ đang chứng minh hiệu quả vượt trội; ví dụ, chuỗi tích hợp AutoDock Vina - GROMACS - SwissADME đã giúp xác định hợp chất kháng SARS-CoV-2 tiềm năng ngay trong giai đoạn sàng lọc ban đầu [6]. Do đó, xu hướng sử dụng chuỗi công cụ tích hợp (integrated pipeline) thay vì phần mềm đơn lẻ đang trở thành hướng tiếp cận chuẩn trong dược lý tính toán hiện nay.

**Bảng 2. So sánh các nhóm phần mềm in silico theo mục tiêu và đặc điểm kỹ thuật**

| Nhóm phần mềm                                                                              | Độ chính xác                                                  | Chi phí                                                    | Mức độ phổ biến                        | Khả năng tích hợp |
|--------------------------------------------------------------------------------------------|---------------------------------------------------------------|------------------------------------------------------------|----------------------------------------|-------------------|
| Thiết kế cấu trúc phân tử và tối ưu hóa hợp chất (ChemDraw, Avogadro, OpenBabel, Gaussian) | Trung bình - Cao (phụ thuộc độ phức tạp của mô hình lượng tử) | Thấp - Trung bình (đa số mã nguồn mở, Gaussian thương mại) | Rộng rãi trong giới học thuật          | Tốt               |
| Sàng lọc ảo và docking phân tử (AutoDock, Vina, GOLD, Glide, SwissDock)                    | Cao (Glide, GOLD) / Trung bình (Vina, SwissDock)              | Thấp (mã nguồn mở); Cao (thương mại)                       | Rất phổ biến                           | Rất tốt           |
| Mô phỏng động lực học phân tử (MD) (GROMACS, AMBER, NAMD, CHARMM)                          | Rất cao, gần thực nghiệm                                      | Trung bình – Cao (yêu cầu phần cứng mạnh)                  | Phổ biến chuyên sâu                    | Tốt               |
| Dự đoán dược động học và độc tính (ADMET) (SwissADME, pkCSM, ADMETlab, Toxtree)            | Trung bình (phụ thuộc vào mô hình ML)                         | Miễn phí (đa số trực tuyến)                                | Rất phổ biến trong sàng lọc ban đầu    | Tốt               |
| Phân tích dữ liệu sinh học - tin sinh học (PubChem, ChEMBL, BindingDB, UniProt, STRING)    | Cao (dữ liệu được xác thực)                                   | Miễn phí                                                   | Rộng rãi, là nền tảng dữ liệu bắt buộc | Rất tốt           |
| Ứng dụng trí tuệ nhân tạo trong thiết kế thuốc (DeepChem, Chemprop, AlphaFold, MolGPT)     | Cao - Rất cao (phụ thuộc dữ liệu huấn luyện)                  | Thấp – Trung bình (đa số mã nguồn mở)                      | Tăng nhanh, xu hướng toàn cầu          | Rất tốt           |

(Nguồn: Tổng hợp của nhóm nghiên cứu từ 112 tài liệu công bố giai đoạn 2010-2024).

#### IV. BÀN LUẬN

##### 4.1. Hạn chế và thách thức

Mặc dù công nghệ *in silico* mang lại nhiều lợi ích trong nghiên cứu và phát triển thuốc, song vẫn tồn tại những thách thức cần khắc phục. Trước hết, sự khác biệt giữa mô hình tính toán và thực nghiệm sinh học là vấn đề cốt lõi, do các mô hình dựa trên giả định và dữ liệu mô phỏng nên có thể sai lệch so với kết quả *in vitro* hoặc *in vivo* [4]. Bên cạnh đó, yêu cầu cao về năng lực tính toán và chuyên môn tin sinh học là rào cản lớn, nhất là ở các nước đang phát triển, khi việc vận hành mô phỏng hoặc mô hình AI đòi hỏi phần cứng mạnh và nhân lực được đào tạo chuyên sâu [5]. Ngoài ra, chi phí bản quyền và hạn chế truy cập dữ liệu vẫn là trở ngại, vì nhiều phần mềm thương mại có giá cao và dữ liệu huấn luyện AI thường thuộc sở hữu tư nhân, làm giảm tính minh bạch khoa học [7]. Tại Việt Nam, việc thiếu hạ tầng tính toán mạnh và đào tạo chuyên sâu là yếu tố cần được ưu tiên để mở rộng ứng dụng thực tiễn của công nghệ *in silico* trong tương lai.

##### 4.2. Xu hướng và triển vọng tương lai

Trong những năm gần đây, trí tuệ nhân tạo (AI), học sâu và điện toán đám mây đã trở thành xu hướng chủ

đạo trong nghiên cứu phát hiện và phát triển thuốc. AI giúp tăng tốc sàng lọc, tối ưu thiết kế *de novo* và dự báo đáp ứng lâm sàng sớm, với bước ngoặt quan trọng từ AlphaFold 3 (2024) mở rộng dự đoán tương tác protein-ligand-acid nucleic-ion, tạo đột phá cho thiết kế thuốc dựa trên cấu trúc (SBDD) [9]. Bên cạnh đó, điện toán đám mây và nền tảng dữ liệu mở (như Open Targets, Terra, Seven Bridges) cho phép tích hợp và xử lý dữ liệu quy mô lớn, hỗ trợ xác định đích tác dụng tiềm năng theo nguyên tắc FAIR [10], [11]. Xu hướng tích hợp dữ liệu đa tầng (multi-omics) giúp làm rõ cơ chế bệnh học và phản ứng thuốc, mở đường cho y học cá thể hóa, trong đó liệu pháp được tối ưu theo đặc điểm di truyền từng bệnh nhân [12-15]. Đồng thời, mô hình hợp tác học thuật - công nghiệp và học liên kết (federated learning) đang góp phần nâng cao độ chính xác, bảo mật dữ liệu và khả năng ứng dụng thực tiễn của các mô hình AI trong dược học tính toán [16].

#### V. KẾT LUẬN

Các phần mềm *in silico* hiện đóng vai trò trung tâm trong phát triển thuốc, cho phép mô phỏng, dự

đoán và tối ưu hóa quá trình phát hiện thuốc ở cấp độ phân tử mà không phải thực hiện quá nhiều thí nghiệm thực nghiệm. Hệ thống công cụ này rất đa dạng, gồm: thiết kế và tối ưu cấu trúc (ChemDraw, Avogadro, Gaussian), sàng lọc ảo và docking (AutoDock, Glide, GOLD), mô phỏng động lực học phân tử (GROMACS, AMBER), dự đoán ADMET (SwissADME, pkCSM, ADMETlab), khai thác dữ liệu sinh học (PubChem, ChEMBL, UniProt) và các nền tảng AI (DeepChem, AlphaFold, MolGPT). Phần mềm mã nguồn mở có ưu thế chi phí thấp và dễ tiếp cận, trong khi phần mềm thương mại thường có độ chính xác và khả năng tích hợp cao hơn. Việc kết hợp nhiều công cụ trong một pipeline giúp nâng cao hiệu quả từ sàng lọc, dự đoán đến mô phỏng tương tác phân tử. Tuy nhiên, các kết quả *in silico* vẫn cần được kiểm chứng bằng thí nghiệm *in vitro* và *in vivo* để đảm bảo tính tin cậy. Tại Việt Nam, việc ứng dụng công nghệ này là xu hướng tất yếu; cần đầu tư hạ tầng tính toán, dữ liệu sinh học và đào tạo nhân lực liên ngành nhằm xây dựng nền tảng phát triển thuốc hiện đại, an toàn và hiệu quả.

#### TÀI LIỆU THAM KHẢO

1. **Pham, Quan & Hương, Lê & Quoc Toan và cs (2020).** Phương pháp sàng lọc ảo trong nghiên cứu phát triển thuốc. 10.15625/vap.2020.00130.
2. **Wouters OJ, McKee M, Luyten J. 2020.** Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA*, 323(9):844–853. <https://doi.org/10.1001/jama.2020.1166>.
3. **Schuhmacher A, Hinder M, Brief E, et al. 2025.** Benchmarking R&D success rates of leading pharmaceutical companies: An empirical analysis of FDA approvals (2006–2022). *Drug Discov Today*, 30(2):104291. <https://doi.org/10.1016/j.drudis.2025.104291>.
4. **Gupta P, Pal Y, Kumar P, et al. 2024.** A critical review on computational techniques through *in silico* assisted drug design. *Int J Pharm Investig*, 14(4):1035–1041. <https://doi.org/10.5530/ijpi.14.4.113>.
5. **Koirala M, Yan L, Mohamed Z, et al. 2025.** AI-integrated QSAR modeling for enhanced drug discovery: From classical approaches to deep learning and structural insight. *Int J Mol Sci*, 26(19):9384. <https://doi.org/10.3390/ijms26199384>.
6. **Ouma RBO, Ngari SM, Kibet JK. 2024.** A review of the current trends in computational approaches in drug design and metabolism. *Discover Public Health*, 21:108. <https://doi.org/10.1186/s12982-024-00229-3>.
7. **Ocana A, Pandiella A, Privat C, et al. 2025.** Integrating artificial intelligence in drug discovery and early drug development: A transformative approach. *Biomark Res*, 13(1):45. <https://doi.org/10.1186/s40364-025-00758-2>.
8. **Shinde SS, Giram PS, Wakte PS, et al. 2025.** ADMET tools in the digital era: Applications and limitations. In: Narayan CV, Verma S, Grewal AS, Singh N, Nimesh H (Eds). *Advances in Pharmacology*, Vol.103:65–80. Academic Press. <https://doi.org/10.1016/bs.apha.2025.01.004>.
9. **Zhang K, Yang X, Wang Y, et al. 2025.** Artificial intelligence in drug development. *Nat Med*, 31:45–59. <https://doi.org/10.1038/s41591-024-03434-4>.
10. **Degatano K, Awdeh A, Cox RS III, et al. 2025.** Warp analysis research pipelines: cloud-optimized workflows for biological data processing and reproducible analysis. *Bioinformatics*, 41(10):btaf494. <https://doi.org/10.1093/bioinformatics/btaf494>.
11. **Guillot P, Bøgsted M, Vesteghem C. 2023.** FAIR sharing of health data: A systematic review of applicable solutions. *Health Technol*, 13:869–882. <https://doi.org/10.1007/s12553-023-00789-5>.
12. **Abdelaziz EH, Ismail R, Mabrouk MS, et al. 2024.** Multi-omics data integration and analysis pipeline for precision medicine: Systematic review. *Comput Biol Chem*, 113:108254. <https://doi.org/10.1016/j.compbiolchem.2024.108254>.
13. **Baião AR, Cai Z, Poulos RC, et al. 2025.** A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. *Brief Bioinform*, 26(4):bbaf355. <https://doi.org/10.1093/bib/bbaf355>.
14. **Carini C, Seyhan AA. 2024.** Tribulations and future opportunities for artificial intelligence in precision medicine. *J Transl Med*, 22:411. <https://doi.org/10.1186/s12967-024-05067-0>.
15. **Srivastava R. 2025.** Advancing precision oncology with AI-powered genomic analysis. *Front Pharmacol*, 16:1591696. <https://doi.org/10.3389/fphar.2025.1591696>.
16. **Teo ZL, Jin L, Li S, Miao D, et al. 2024.** Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. *Cell Rep Med*, 5(2):101419. <https://doi.org/10.1016/j.xcrm.2024.101419>.