

# QUY TRÌNH VỀ PHƯƠNG PHÁP DỰ ĐOÁN VÀ SO SÁNH CẤU TRÚC PROTEIN SỬ DỤNG NỀN TẢNG ALPHAFOLD

Trần Ánh Thống Trình\*, Nguyễn Minh Thái, Nguyễn Thanh Huy

Trường Đại học Y Dược Thành Phố Hồ Chí Minh

**Tóm tắt:** Kể từ khi được phát hiện vào những năm 1990, kháng thể đơn miền (single-domain antibody, sdAb, nanobody) đã tạo ra một cuộc cách mạng mới trong chẩn đoán và điều trị bệnh liên quan đến vi khuẩn đề kháng. Hiện nay, việc dự đoán và xây dựng cấu trúc của các phân tử nanobody bằng các phần mềm khoa học máy tính là điều rất cần thiết cho việc sàng lọc các kháng thể có tính đặc hiệu cao. Công trình nghiên cứu này tập trung hướng dẫn sử dụng các công cụ trí tuệ nhân tạo và biểu thị hình ảnh như PyMOL, Dali, ColabFold để dự đoán và so sánh kết cấu không gian của nhiều chuỗi amino acid trong cấu trúc tiền thân của nanobody - nhóm thụ thể kháng nguyên thế hệ mới (Variable domain new antigen receptor, VNAR). Các thao tác trong quy trình sẽ tập trung vào 2 chuỗi amino acid trong cấu trúc của VNAR. Kết quả nhận được sẽ là cấu trúc “xếp chồng” (superimposed structure) giúp kiểm tra sự tương đồng về cấu trúc của các protein có cùng nguồn gốc với các tiêu chí liên quan về cấu trúc dự kiến của nanobody. Quy trình *in silico* dự đoán cấu trúc của VNAR sẽ góp vai trò đáng kể trong việc thay đổi VNAR từ loài cá mập để tích hợp vào cơ thể người.

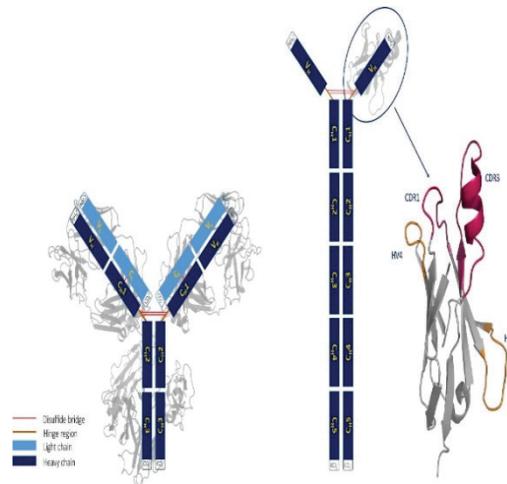
**Từ khóa:** ColabFold, Dali, KABAT position, PyMOL, VNAR

## 1. GIỚI THIỆU

Kháng thể là những phân tử protein đặc hiệu do hệ miễn dịch của động vật có xương sống sản xuất ra, đóng vai trò thiết yếu trong việc nhận diện và trung hòa các tác nhân gây bệnh như vi khuẩn, vi-rút và độc tố. Ở người và hầu hết các loài động vật có vú, kháng thể (hay còn gọi là immunoglobulin) thường có cấu trúc điển hình gồm bốn chuỗi polypeptide: hai chuỗi nặng và hai chuỗi nhẹ, kết hợp với nhau để hình thành vùng gắn kết kháng nguyên (antigen-binding site) và vùng hằng định (constant region) chịu trách nhiệm điều hòa đáp ứng miễn dịch.

Tuy nhiên, ở một số loài thuộc lớp cá sụn (cartilaginous fish) như cá mập và cá đuối, hệ miễn dịch đã tiến hóa theo một hướng khác biệt, tạo ra một loại kháng thể đặc biệt chỉ bao gồm hai chuỗi nặng, gọi là IgNAR (Immunoglobulin New Antigen Receptor) (Flajnik, 2018; Greenberg và

cộng sự, 1995; Hamers-Casterman và cộng sự, 1993). Loại kháng thể này có cấu trúc gồm năm vùng hằng định (constant domains) và một vùng biến đổi (Variable domain new antigen receptor, VNAR) ở đầu N-terminal, trong đó VNAR chịu trách nhiệm nhận diện và liên kết đặc hiệu với kháng nguyên (xem Hình 1). Việc trích xuất VNAR đã mở ra nhiều tiềm năng cho việc xét nghiệm cũng như điều trị các “căn bệnh thế kỷ”, đặc biệt là ung thư; vì cấu trúc nhỏ gọn, chúng có thể dễ dàng đi qua hàng rào máu não (Yang & Shah, 2020)—điều mà các kháng thể với cấu trúc cổ điển không thể làm được. Bên cạnh đó, nanobody còn có thể chịu được điều kiện pH phi sinh lý (3.0 - 9.0)—phù hợp cho việc tiếp cận các quần thể tế bào ung thư, nơi có môi trường acid được sinh ra do hiệu ứng Warburg (Gatenby & Gillies, 2004). Điều này càng chứng minh cho việc nanobody có thể mở ra kỷ nguyên mới cho việc xét nghiệm và điều trị bệnh ung thư.



**Hình 1. So sánh cấu trúc của kháng thể IgG ở người (trái) và kháng thể IgNAR ở cá mập (phải) (Fernández-Quintero và cộng sự, 2021)**

Mặt khác, trong nghiên cứu sinh học phân tử, việc xác định cấu trúc không gian ba chiều của VNAR là một bước then chốt để hiểu rõ cơ chế hoạt động sinh học, cũng như để thiết kế các liệu pháp điều trị bệnh lý một cách chính xác. Tuy nhiên, các phương pháp truyền thống như tinh thể học tia X (X-ray crystallography), cộng hưởng từ hạt nhân (NMR spectroscopy) hay kính hiển vi điện tử cryo-EM đều đòi hỏi quy trình phức tạp, chi phí cao và thời gian thực hiện kéo dài. Chính vì vậy, nhu cầu về những công cụ dự đoán cấu trúc VNAR nhanh chóng, chính xác và tiết kiệm trở nên cấp thiết trong cộng đồng khoa học.

Mô phỏng sinh học bằng máy tính (*in silico*) là một phương pháp trong lĩnh vực sinh học phân tử và dược học, trong đó các mô hình máy tính được sử dụng để mô phỏng quá trình sinh học hoặc phân tích tương tác giữa các hệ protein. Nhờ vào sự phát triển vượt bậc của công nghệ máy tính cũng như ngành khoa học tính toán trong những năm gần đây mà mô phỏng sinh học bằng máy tính ngày càng trở nên quan trọng và được ứng dụng rộng rãi trong ngành dược học và nghiên cứu sinh học. Công dụng nổi bật của nghiên cứu mô phỏng sinh học phải kể đến là dự đoán và mô hình hóa các hệ protein, giúp các nhà nghiên cứu hiểu rõ hơn về cơ

chế hoạt động của các protein, từ đó có thể phát triển các phác đồ điều trị mới. Ngoài ra, mô phỏng *in silico* cũng có thể giúp dự đoán cách mà các phân tử hoạt động và tương tác với các hệ protein cụ thể, từ đó làm tăng hiệu suất nghiên cứu dược học. Những tiến bộ của lĩnh vực mô phỏng sinh học bằng máy tính đã giúp cho nghiên cứu sinh học trở nên hiệu quả và tiết kiệm thời gian so với các phương pháp truyền thống.

AlphaFold, một hệ thống trí tuệ nhân tạo phát triển bởi DeepMind, đã đánh dấu bước đột phá trong lĩnh vực này (Jumper và cộng sự, 2021). Bằng cách kết hợp các kỹ thuật học sâu (deep learning) với dữ liệu sinh học thực nghiệm, AlphaFold có khả năng dự đoán cấu trúc ba chiều của protein chỉ từ trình tự amino acid với độ chính xác tiệm cận các phương pháp thực nghiệm truyền thống. Công dụng của AlphaFold không chỉ dừng lại ở việc dự đoán cấu trúc tự nhiên của protein, mà còn mở rộng sang nhiều lĩnh vực ứng dụng khác như: thiết kế protein mới, dự đoán ảnh hưởng của đột biến lên cấu trúc và chức năng protein, phát triển thuốc và hiểu biết sâu hơn về các bệnh liên quan đến sai lệch cấu trúc protein. Nhờ những tiến bộ vượt bậc của AlphaFold, quá trình nghiên cứu và thao tác trên protein—vốn trước đây tốn kém và mất nhiều năm—giờ đây có

thể được thực hiện một cách nhanh chóng, chính xác và khả thi hơn bao giờ hết. Đối với nghiên cứu kháng thể, AlphaFold được kỳ vọng sẽ cung cấp một công cụ mạnh mẽ để mô phỏng nhanh chóng cấu trúc các vùng biến đổi như VNAR, giúp định hướng thiết kế và tối ưu hóa các biến thể kháng thể với độ ái lực và tính đặc hiệu cao hơn.

Được thúc đẩy bởi những lý do trên, chúng tôi tiến hành nghiên cứu mô phỏng tính toán này. Bằng cách sử dụng công cụ AlphaFold, nghiên cứu tập trung vào việc dự đoán và so sánh cấu trúc không gian VNAR nhằm khám phá và tối ưu hóa các đặc tính sinh học của loại kháng thể đặc biệt này.

## 2. PHƯƠNG PHÁP NGHIÊN CỨU

Nghiên cứu được thực hiện trên nguyên tắc so sánh độ tin cậy về vị trí của các amino acid trong chuỗi polypeptit dự đoán so với thực nghiệm qua các thông số pLDDT (prediction Local Distance Different Test), bên cạnh đó, ta còn có thể đánh giá được độ tin cậy của cấu trúc “xếp chồng” thông qua các thông số mang ý nghĩa thống kê, điển hình là giá trị Z-score (Holm, 2020; Tunyasuvunakool và cộng sự, 2021).

Nhìn chung, việc dự đoán và so sánh cấu trúc protein chú trọng vào việc đi tìm cấu trúc không gian phù hợp nhất cho chuỗi polypeptit. Hệ thống máy tính và trí tuệ nhân tạo sẽ tiến hành so sánh các cấu trúc protein trong thực nghiệm có tính khả thi nhất để đối chiếu với trình tự amino acid có sẵn trong kho dữ liệu, từ đó sẽ xây dựng cấu trúc phù hợp nhất về mặt không gian của phân tử protein.

Bên cạnh đó, dựa vào hệ thống đánh số KABAT giúp ta xác định các khu vực khác nhau của phân tử protein. Cụ thể, trong công trình này, dựa vào việc đánh dấu các “phần thừa” còn lại của các phản ứng peptide hóa, ta sẽ xác định được các vùng quan trọng trong cấu trúc của phân tử kháng thể.

### 2.1 Đối tượng nghiên cứu và phần mềm ứng dụng:

Quy trình dự đoán và so sánh cấu trúc protein cần sử dụng:

- 2 chuỗi amino acid cần so sánh, đối chiếu, trong đó 1 chuỗi sẽ được lấy từ Ngân hàng Protein (Protein Data Bank, PDB, ID: 1SQ2) (5A7), chuỗi còn lại là h5A7 (humanized 5A7), được tạo ra thông qua ghép vùng CDR của chuỗi 5A7 vào một khung kháng thể tương thích của người, DPK9-JK1 (Kovalenko và cộng sự, 2013). Công trình này không tập trung vào việc hướng dẫn cách lắp ghép vùng CDR để tạo ra trình tự biểu hiện mới, mà chỉ tập trung vào việc dự đoán và so sánh cấu trúc không gian của 2 chuỗi amino acid nói trên.

- ColabFold v1.5.5 (Mirdita và cộng sự, 2022) là một phần mềm được phát triển bởi Google Research, có khả năng dự đoán cấu trúc của protein từ chuỗi amino acid. Phần mềm này sử dụng mô hình DeepMind AlphaFold để dự đoán cấu trúc 3D của protein một cách chính xác và nhanh chóng. ColabFold cung cấp một giao diện sử dụng trực tuyến, giúp người dùng dễ dàng tiếp cận và sử dụng công nghệ dự đoán cấu trúc protein một cách hiệu quả. Đây là một công cụ hữu ích trong lĩnh vực nghiên cứu sinh học phân tử và y học.

- PyMOL—phần mềm trực quan hóa cấu trúc không gian ba chiều của phân tử protein sau khi được phân tích và dự đoán bởi ColabFold.

- Công cụ so sánh các chuỗi amino acid để kiểm tra tính tương đồng và xây dựng cấu trúc “xếp chồng”. Trong công trình này, nhóm nghiên cứu sẽ sử dụng công cụ Dali.

### 2.2 Tiến hành:

Sau đây là các bước tiến hành quy trình dự đoán và mô phỏng cấu trúc protein. Các bước từ 1 đến 6 sẽ hướng dẫn quy trình dự đoán cấu trúc không gian protein, các bước

từ 7 đến 12 mô tả các giai đoạn trong quy trình biểu hiện cấu trúc không gian và so sánh sự tương đồng của protein.

1. Mở trình duyệt web, chẳng hạn như Chrome, Firefox, v.v. Truy cập trang web của PDB (Protein data bank)—ngân hàng dữ liệu protein thông qua đường dẫn sau (<https://www.rcsb.org/>), tiếp theo đó, nhập mã định danh của chuỗi amino acid cần phân tích vào thanh tìm kiếm của web. Trong nghiên cứu này, ta sử dụng chuỗi 5A7 (ID: 1SQ2), sau khi thực hiện các thao tác trên, hệ thống sẽ trả cho ta một số kết quả > nhấp chọn cấu trúc có mã 1SQ2 > chọn “Display Files”> chọn “FASTA sequence”, lúc này màn hình sẽ trả kết quả như bên dưới:

```
>1SQ2_1|Chain A[auth L]|Lysozyme C|Gallus gallus (9031)
KVFGRCELAAAMKRHGLDNYGYS LGNWW
CAAKFESNFNTQATNRNTDGDY GILQIN
SRWWCNDGRTPGSRNLCNIPCS ALLSSDIT
ASVNCAKKIVSDGNGMNAWVAWRNRCKGTD
VQAWIRGCRL
>1SQ2_2|Chain B[auth N]|novel
antigen receptor|Ginglymostoma
cirratum (7801)
```

2. Sao chép trình tự cần phân tích, từ kết quả của bước đầu tiên, chúng ta có thể thấy rằng, phân tử protein có ID 1SQ2 được hình thành từ 2 chuỗi amino acid từ hai loài khác nhau, chuỗi chúng ta cần phân tích đến từ loài *Ginglymostoma cirratum*, lưu ý: khi sao chép chỉ cần sao chép trình tự amino acid (đoạn được tô đỏ), không cần phải sao chép toàn bộ định dạng FASTA.

3. Truy cập ColabFold v1 .5 .5 thông qua đường dẫn: <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=kOb-lAo-xetgx>

4. Dán trình tự amino acid ở bước 2 vào thanh “query\_sequence” > nhấp chọn “runtime” > “run all” > “run anyway”, lúc này, ColabFold sẽ bắt đầu chạy chương trình phân tích và dự đoán cấu trúc của chuỗi amino acid trên, tùy thuộc vào độ dài của chuỗi mà thời gian phân tích có thể nhanh hoặc chậm.

5. Sau khi thực hiện với chuỗi amino acid 5A7 (id 1SQ2), ta dán trình tự của chuỗi h5A7 như sau vào trong “query\_sequence”:

```
ARVDQSPSSLSASVGDRTITCVLR
DASYALGSTCWYQQKPGKAPKSISKG
GRYSESVNSGSKSFTLTISLQPEDFA
TYYCGLGVAGGYCDYALCSSRYAECGQ
GTKVEIK
```

6. Sau khi hoàn tất quá trình dự đoán cấu trúc protein, hệ thống sẽ tự động tải về 1 file kết quả, trong đó là toàn bộ hình ảnh và dữ liệu dự đoán các cấu trúc không gian khả thi nhất của phân tử protein. Tìm và chọn file định dạng .pdb có tên “rank\_001”, mở file bằng phần mềm PyMOL.

7. Sau khi hiển thị được 2 cấu trúc như trên, tiến hành hiển thị đánh dấu các vùng của cấu trúc VNAR như sau: chọn “Display”> “Sequence”> chuyển màu từng phần của cấu trúc bằng cách nhấp lên các amino acid theo hướng dẫn sau (trình tự màu có thể thay đổi tùy vào cảm quan của người làm):

- **FW1**: từ axit amin đầu tiên (Ala-1, vị trí KABAT 1) đến axit amin thứ 21 (vị trí Cys-21, KABAT), amino acid tiếp theo là điểm bắt đầu của CDR1.

- **CDR1**: Từ kí tự V chính tắc (Val-22, vị trí KABAT 22), điểm cuối của CDR1 là kí tự C chính tắc (Cys-35, vị trí KABAT 35).

- **FW2**: Từ kí tự W chính tắc (Tryp-36, vị trí KABAT 36) và vùng này sẽ kết thúc ở amino acid 44 (Gly-44, vị trí KABAT 44).

- **HV2**: Từ kí tự N chính tắc (Asn-45, vị trí KABAT 45) đến cuối HV2 là kí tự G chính tắc (Gly-52, vị trí KABAT 52).

- **FW3a**: Từ G chính tắc (Gly-53, vị trí KABAT 53) và vùng này sẽ kết thúc ở amino acid 60 (Asn-60, vị trí KABAT 60).

- **HV4**: Từ canonical S (Ser-61, KABAT vị trí 61) và khu vực này sẽ kết thúc tại canonical S (Ser-65, KABAT vị trí 65).

- **FW3b**: Từ canonical F (Phe-66, vị trí KABAT 66), vùng này sẽ kết thúc ở amino acid 82 (ARG-82, vị trí KABAT 82).

- **CDR3**: Từ C chính tắc (Cys-83, vị trí KABAT 83) và vùng này sẽ kết thúc ở amino acid 106 (Asp-106, vị trí KABAT 106).

- **FW4**: Các amino acid còn lại của cấu trúc.

- Tiếp theo đó, thể hiện các liên kết disulfide hình thành từ sự tương tác giữa các amino acid Cystenin bằng cách chọn hộp thoại “S” gần tên của protein, chọn “Disulfide” > “Stick”.

8. Tiến hành kiểm tra độ tin cậy và sự tương đồng giữa 2 cấu trúc bằng công cụ Dali.

9. Truy cập Dali thông qua đường link: <http://ekhidna2.biocenter.helsinki.fi/dali/>, chọn “Pairwise”.

10. Ở khung “STEP 1” > “choose file” > chọn file .pdb “rank\_001” của phân tử protein thứ 1, làm tương tự với protein thứ hai ở khung “STEP 2”> Có thể đặt tên gọi nhớ ở “STEP 3”> bấm chọn “Submit”.

11. Sau khi chạy, Dali sẽ trả cho người dùng kết quả, nhấn chọn “Interactive

(html)” > Tải file tên “PDB” trong mục Summary dưới định dạng .pdb về máy, trong file này là thông tin về chuỗi amino acid của cấu trúc xếp chồng để một phần chứng minh được tính tương thích giữa 2 chuỗi amino acid.

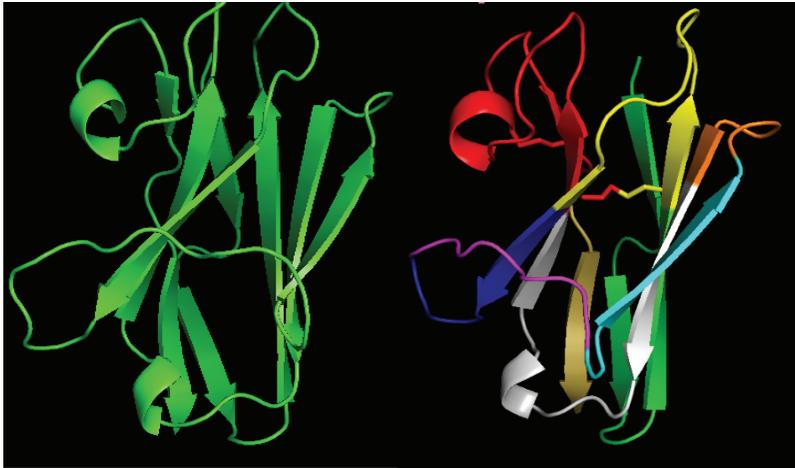
12. Sau khi đã tải file về, ta tiếp tục sử dụng PyMOL để tạo cấu trúc “xếp chồng”, Cấu trúc vừa được tải về sẽ có trình tự giống 5A7 hoặc h5A7, ta cần chọn chuỗi amino acid còn lại để hình thành cấu trúc xếp chồng. Trong trường hợp này, ta chọn h5A7.

### 3. KẾT QUẢ VÀ THẢO LUẬN

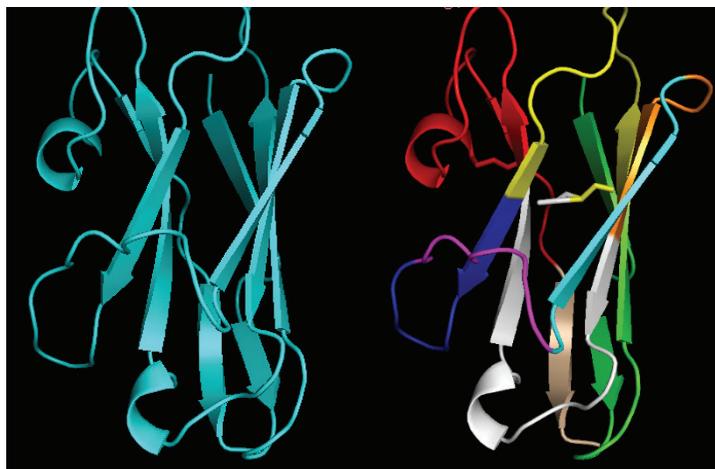
Sau khi hoàn thành bước số 6 trong quy trình trên, kết quả thu được sẽ là cấu trúc không gian của protein 5A7 và h5A7 (Hình 2 và Hình 3), việc thay đổi màu sắc giúp cho việc nhận diện các vùng cấu trúc của protein và vị trí của cầu nối disulfide trong không gian trở nên dễ dàng hơn.

Cấu trúc xếp chồng phản ánh sự tương đồng của 2 cấu trúc trước và sau khi được chỉnh sửa (Hình 4). Các vùng có trình tự tương đồng nhau sẽ không có sự chồng chéo màu sắc, ngược lại, các phần có màu giao nhau sẽ có sự khác nhau về trình tự amino acid và cấu trúc không gian 3 chiều.

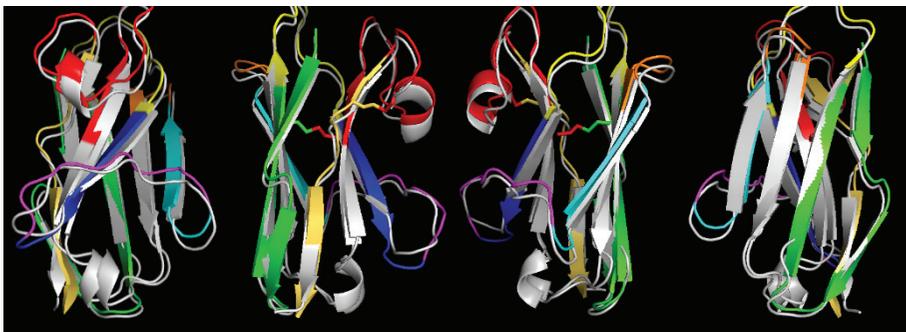
Ngoài ra, khi hoàn thành xong bước số 5, trong file kết quả tải về còn có các biểu đồ biểu hiện độ phủ trình tự (Hình 5) và độ tin cậy về trình tự của các amino acid trong chuỗi (Hình 6).



**Hình 2.** Cấu trúc của protein 5A7 trước (bên trái) và sau (bên phải) khi được chuyển màu và vùng được khoanh đỏ thể hiện vị trí của cầu nối disulfide



**Hình 3.** Cấu trúc của protein 5A7 sau khi được chỉnh sửa đoạn CDR, trước (bên trái) và sau (bên phải) được chuyển màu phân vùng cấu trúc và thể hiện các cầu nối disulfide (vùng được khoanh đỏ)



**Hình 4.** Cấu trúc “xếp chồng” của protein 5A7 và h5A7 sau khi được phân tích bởi Dali

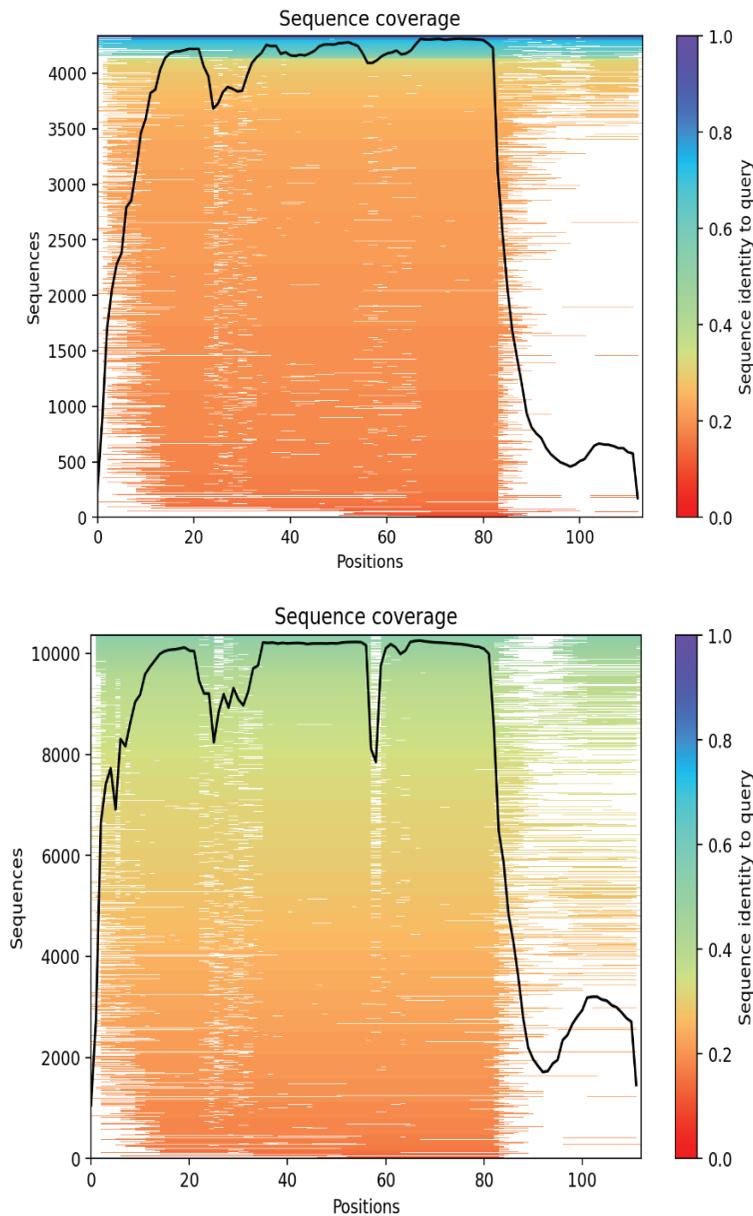
Đối với biểu đồ phản ánh độ phủ trình tự (Hình 5), ta có:

**Trục X (Positions):** Biểu thị các vị trí trong chuỗi protein hoặc peptide.

**Trục Y (Sequences):** Số lượng trình tự được căn chỉnh tại mỗi vị trí.

**Thanh màu (Sequence identity to query):** Biểu thị mức độ tương đồng giữa các trình tự so với trình tự tham chiếu.

**Đường màu đen:** Biểu diễn đường chuẩn của chuỗi amino acid cần dự đoán.



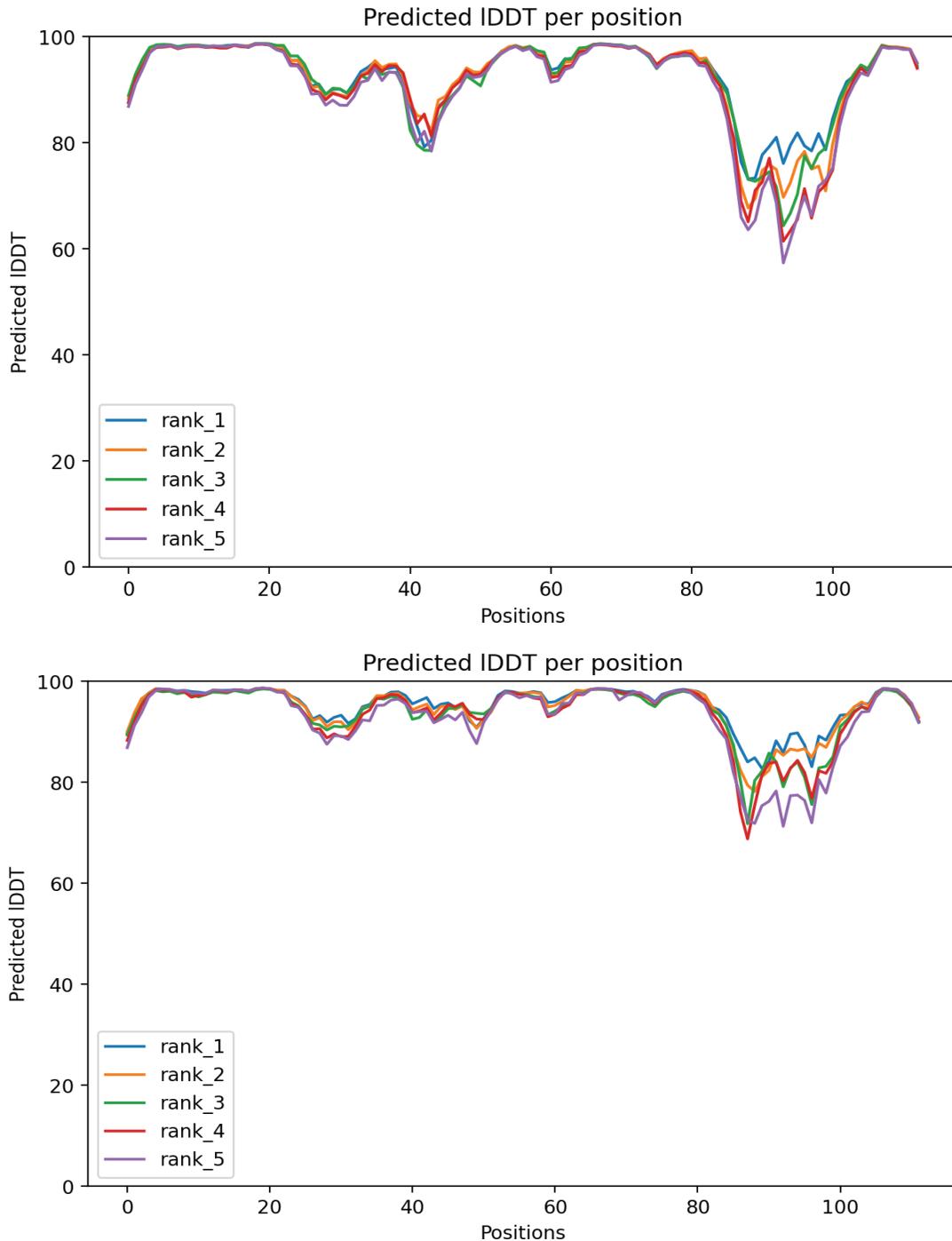
**Hình 5. Biểu đồ đánh giá độ phủ trình tự trong cấu trúc protein 5A7 (ID: 1SQ2) trước (phía trên) và sau (phía dưới) khi chỉnh sửa vùng CDR tính theo từng vị trí.**

Đối với biểu đồ pLDDT (Hình 6):

**Trục X:** Số thứ tự amino acid trong chuỗi protein.

**Trục Y:** Giá trị pLDDT (từ 0 đến 100), thể hiện mức độ tin cậy của từng vị trí.

Nhận xét biểu đồ: các vùng khác của trình tự amino acid thường có trình tự ổn định, trong khi đó, các vùng CDR thì có trình tự thay đổi nhiều hơn.



**Hình 6.** Biểu đồ pLDDT so sánh sự khác nhau về trình tự amino acid trong các chuỗi 5A7 (trên) và h5A7 (dưới), xếp hạng từ rank 1 đến rank 5.

Ngoài việc đánh giá sự tương đồng thông qua PyMOL, ngay từ khi xem kết quả ở Dali ta đã có thể nhận định ban đầu về độ tin cậy qua giá trị Z-score. Dali có riêng một phần đánh giá giá trị Z (Holm và cộng sự, 2008), nếu Z-score < 2 thì chứng tỏ rằng giữa hai cấu trúc trước và sau chỉnh sửa có sự khác biệt lớn và có thể ảnh hưởng đến chức năng của protein. Trong quy trình này, cấu trúc được phân tích có Z-score = 22.5 cho thấy 2 protein có sự tương đồng cao trong cấu trúc không gian từ đó dự đoán rằng sự gắn kết đặc hiệu với kháng nguyên của 5A7 vẫn có thể được giữ lại ở h5A7.

#### 4. KẾT LUẬN VÀ ĐỊNH HƯỚNG NGHIÊN CỨU

Quy trình dự đoán và so sánh cấu trúc nanobody được trình bày trên có thể hỗ trợ đánh giá hiệu quả ban đầu của quá trình chỉnh sửa trình tự vùng CDR trong chuỗi amino acid của nanobody thông qua cấu trúc “xếp chồng” để so sánh sự thay đổi trong cấu trúc của cùng một nanobody trước và sau khi được chỉnh sửa. Bên cạnh đó, dựa vào thông số Z-score trong phần mềm Dali, người dùng có thể đưa ra kết luận rằng việc chỉnh sửa vùng CDR không làm thay đổi đến cấu trúc tổng quan cũng như tính chất của nanobody. Tóm lại, quy trình dự đoán và so sánh cấu trúc protein được mô phỏng bằng phần mềm ColabFold có thể hỗ trợ người làm thực nghiệm đưa ra

những đánh giá mang tính định tính về hiệu quả của việc chỉnh sửa cấu trúc protein.

Trong nghiên cứu hiện tại của nhóm chỉ tập trung vào quy trình *in silico* trong việc dự đoán và so sánh cấu trúc không gian của VNAR được nhân hóa (humanize) thông qua phân tích trình tự, mô hình hóa cấu trúc và đánh giá các chỉ số dự đoán như Z-score và pLDDT. Tuy nhiên, giá trị Z-score chỉ phản ánh về độ tương đồng trong hình học và cấu trúc không gian, không hoàn toàn phản ánh được ái lực gắn kết hoặc hoạt tính sinh học, chính vì điều đó, nhóm nghiên cứu sẽ tiến hành thêm các thử nghiệm mô phỏng docking để so sánh khả năng liên kết với kháng nguyên đặc hiệu của 5A7 và h5A7, từ đó sẽ đưa ra câu trả lời chính xác hơn về việc chỉnh sửa vùng CDR có thay đổi nhiều đến chức năng của VNAR hay không. Bên cạnh đó, nhóm nghiên cứu định hướng tiến hành các thử nghiệm *in vitro* tham khảo theo quy trình đánh giá chức năng của VNAR sau khi được chỉnh sửa (Rossotti và cộng sự, 2022). Các kỹ thuật như ELISA, phân tích dòng tế bào bằng flow cytometry và đo lường ái lực (như xác định hằng số  $K_D$ ) sẽ đóng vai trò quan trọng trong việc kiểm tra chức năng của các biến thể VNAR đã chỉnh sửa, từ đó giúp kiểm định độ tin cậy của các kết quả *in silico* và hỗ trợ định hướng ứng dụng tiền lâm sàng.

#### Thông tin tác giả:

**Trần Ánh Thống Trình** (\*Tác giả liên hệ), Trường Đại học Y Dược Thành Phố Hồ Chí Minh

Email: tattrinh.duoc23@ump.edu.vn

**Nguyễn Minh Thái**, Trường Đại học Y Dược Thành Phố Hồ Chí Minh

Email: minhthai2511@ump.edu.vn

**Nguyễn Thanh Huy**, Trường Đại học Y Dược Thành Phố Hồ Chí Minh

Email: nthuy.d21@ump.edu.vn

## Ghi chú

Các tác giả xác nhận không có tranh chấp về lợi ích đối với bài báo này.

### Thông tin bài báo:

**Ngày nhận bài:** 12/03/2025

**Ngày hoàn thiện biên tập:** 01/06/2025

**Ngày duyệt đăng:** 03/06/2025

## TÀI LIỆU THAM KHẢO

Fernández-Quintero, M. L. và cộng sự (2021). Shark Antibody Variable Domains Rigidify Upon Affinity Maturation-Understanding the Potential of Shark Immunoglobulins as Therapeutics. *Front Mol Biosci*, 8, 639166. doi:10.3389/fmolb.2021.639166.

Gatenby, R. A. & Gillies, R. J. (2004). Why do cancers have high aerobic glycolysis? *Nat Rev Cancer*, 4(11), 891-899. doi:10.1038/nrc1478.

Holm, L. (2020). Using Dali for Protein Structure Comparison. *Methods Mol Biol*, 2112, 29-42. doi:10.1007/978-1-0716-0270-6\_3.

Holm, L. và cộng sự (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics*, 24(23), 2780-2781. doi:10.1093/bioinformatics/btn507.

Jumper, J. và cộng sự (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. doi:10.1038/s41586-021-03819-2.

Kovalenko, O. V. và cộng sự (2013). Atypical antigen recognition mode of a shark immunoglobulin new antigen receptor (IgNAR) variable domain characterized by humanization and structural analysis. *J Biol Chem*, 288(24), 17408-17419. doi:10.1074/jbc.M112.435289.

Mirdita, M. và cộng sự (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679-682. doi:10.1038/s41592-022-01488-1.

Rossotti, M. A. và cộng sự (2022). Immunogenicity and humanization of single-domain antibodies. 289(14), 4304-4327. doi: <https://doi.org/10.1111/febs.15809>.

Tunyasuvunakool, K. và cộng sự (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873), 590-596. doi:10.1038/s41586-021-03828-1.

Yang, E. Y. & Shah, K. (2020). Nanobodies: Next Generation of Cancer Diagnostics and Therapeutics. 10. doi:10.3389/fonc.2020.01182.

## **PROTOCOL: PREDICTING AND COMPARING PROTEIN STRUCTURES USING ALPHAFOLD**

**Tran Anh Thong Trinh\***, Nguyen Minh Thai, Nguyen Thanh Huy

*University of Medicine and Pharmacy at Ho Chi Minh*

**Abstract:** *Since their discovery in the 1990s, single-domain antibodies (sdAbs, nanobodies) have revolutionized the diagnosis and treatment of antibiotic-resistant bacterial diseases. Currently, predicting and modeling nanobody structures using computational tools are essential for screening highly specific antibodies. This research proposal outlines the use of artificial intelligence and visualization tools like PyMOL, Dali, and ColabFold to predict and compare the 3D structures of amino acid chains in the precursor framework of VNAR (Variable domain of new antigen receptor) - a next-generation antigen-binding scaffold. The workflow will focus on two key amino acid chains in the VNAR structure, generating superimposed models to evaluate structural homology among evolutionarily related proteins and assess critical parameters in nanobody design. This in silico VNAR structure prediction pipeline will contribute significantly to the humanization of shark-derived VNARs for therapeutic applications.*

**Keywords:** *ColabFold, Dali, KABAT position, PyMOL, VNAR*

### **Author Information:**

**Tran Anh Thong Trinh** (\*Corresponding author), University of Medicine and Pharmacy at Ho Chi Minh

Email: [tattrinh.duoc23@ump.edu.vn](mailto:tattrinh.duoc23@ump.edu.vn)

**Nguyen Minh Thai**, University of Medicine and Pharmacy at Ho Chi Minh

Email: [minhthai2511@ump.edu.vn](mailto:minhthai2511@ump.edu.vn)

**Nguyen Thanh Huy**, University of Medicine and Pharmacy at Ho Chi Minh

Email: [nthuy.d21@ump.edu.vn](mailto:nthuy.d21@ump.edu.vn)

### **Note**

The authors declare no competing interests.