

KHẢO SÁT *in silico* MỘT SỐ ĐẶC ĐIỂM CỦA BETA-GLUCOSIDASE TỪ DỮ LIỆU TRÌNH TỰ DNA METAGENOME 8.4 GB CỦA VI KHUẨN TRONG DẠ CỎ DÊ

Đỗ Thị Huyền^{1,2*}, Nguyễn Minh Chiến^{2,3}, Trương Nam Hải¹

¹Viện Công nghệ sinh học - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

³Công ty TNHH Dược Hanvet

TÓM TẮT

Trong nghiên cứu trước, metagenomic DNA của vi khuẩn trong dạ cỏ dê đã được giải trình tự với kích thước dữ liệu là 8,4 Gb và 164,644 gen đã được dự đoán. Sử dụng các bộ dữ liệu CAZy, KEGG, COG, 384 trình tự mã hóa cho beta-glucosidase đã được khai thác trong đó có 131 trình tự được ước đoán là trình tự hoàn thiện. Phần lớn (93,1%) trình tự có kích thước lớn trên 700 amino acid, có chỉ số pI trong khoảng 5,0-6,5 mang vùng trung tâm xúc tác thuộc họ GH3 và có vùng phụ trợ FN3. Các enzyme được ước đoán có chỉ số alkaline cao, trong đó 67,9% enzyme có khả năng hoạt động ở vùng pH kiềm (chỉ số alkaline ≥ 0.6). Bốn enzyme có chỉ số pI cao nhưng được ước đoán hoạt động tốt ở vùng pH axit và 3 enzyme có chỉ số pI cao được ước đoán hoạt động ở vùng pH kiềm. Trong 7 enzyme này, 5 enzyme được xác định là enzyme tiết, có trình tự tín hiệu tiết ở phía đầu N. Hơn nửa số enzyme (69 enzyme) được ước đoán bền ở nhiệt độ trên 65°C. Hai enzyme có kích thước thấp nhất được ước đoán có khả năng hoạt động ở pH kiềm. Các biến số độ dài gen, giá trị pI, giá trị alkaline, khả năng bền nhiệt, cấu trúc enzyme là các biến độc lập, chưa xác định được mối liên quan với nhau. Nghiên cứu này đã cung cấp những thông tin chung về một số đặc điểm cơ bản của beta-glucosidase được ước đoán dựa trên phân tích tin sinh học, có giá trị cho định hướng khai thác gen và biểu hiện gen.

Từ khóa: Beta-glucosidase, chỉ số alkaline, dữ liệu 8,4 Gb, *in silico*, vi khuẩn trong dạ cỏ dê.

MỞ ĐẦU

Beta-glucosidase (BGL, EC 3.2.1.21) là enzyme quan trọng được ứng dụng nhiều trong nhiều ngành công nghiệp, nông nghiệp, thực phẩm và y dược. BGL có khả năng xúc tác chuyển hóa gốc glucosyl từ đầu không khử của cả ba loại liên kết aryl-, alkyl và beta-glucoside trong disaccharide và cả các chuỗi oligosaccharide ngắn thành glucose và các chuỗi polysaccharide ngắn hơn. Enzyme được dùng trong chuyển hóa sinh khối lignocellulose cho sản xuất cồn sinh học. Ngoài ra, enzyme có vai trò quan trọng trong nhiều quá trình sinh học ở các cơ thể sống khác nhau và được ứng dụng rộng rãi trong công nghiệp thực phẩm, công nghiệp hương liệu, tái chế giấy, khử độc bột sắn và tăng tốc độ đường hóa, chuyển hóa chất xơ trong thức ăn gia súc, gia cầm, tăng hiệu quả sử dụng thức ăn, hấp thụ chất dinh dưỡng từ thức ăn cho động vật. Ở trong một số điều kiện nhất định, beta-glucosidase còn được sử dụng trong tổng hợp oligosaccharide, aryl-, alkyl-beta-glucoside. Đây là các hợp chất có tính ổn định cao, thân thiện với môi trường, bề mặt không bị ion hóa và tự phân hủy sinh học, được ứng dụng nhiều trong chế tạo thuốc, vaccine, sản phẩm làm đẹp, công cụ trong chẩn đoán,... (Seeberger, Werz, 2007; Tran *et al.*, 2023; Ma *et al.*, 2024). Vì vậy enzyme này rất được các nhóm nghiên cứu trên thế giới quan tâm nghiên cứu, tìm kiếm để ứng dụng vào công nghiệp.

Hiện nay, việc ứng dụng BGL vào thực tế còn nhiều hạn chế do enzyme thường có hoạt độ thấp, chỉ hoạt động tốt nhất ở nhiệt độ 40–70°C và pH 4,5–5, nhưng các ứng dụng công nghiệp thường cần enzyme hoạt động ở phổ pH rộng và nhiệt độ nằm ngoài giới hạn này. Ví dụ, các bước tiền xử lý sinh khối trong quá trình sản xuất nhiên liệu sinh học thường hoạt động ở nhiệt độ trên 80°C (Rocha *et al.*, 2023). Để tăng hương vị cho nước ép trái cây, các enzyme có hoạt tính tối ưu ở pH acid 2,8–3,8 sẽ thích hợp để giải phóng các chất bay hơi có chứa kết glycoside (Godse *et al.*, 2021). BGL thường rất nhạy cảm với glucose nên dễ bị ức chế nếu sản phẩm chuyển hóa tăng cao (Ahmed *et al.*, 2017). Vì vậy, việc nghiên cứu mở rộng tìm kiếm các nguồn gen mã hóa cho các BGL có đặc tính quý, được ước đoán từ dữ liệu DNA metagenome từ các hệ sinh thái tiềm năng là rất cần thiết. Hiện nay, công nghệ Metagenomics đã được ứng dụng để giải trình tự đa hệ gen của vi sinh vật trong hệ sinh thái đã cho ra những bộ dữ liệu gen rất lớn từ nguồn vi sinh vật chưa nuôi cấy được. Đây là nguồn gen rất mới, chưa được khai thác, mang những thông tin mới về cấu trúc của các loại protein/enzyme khác nhau. Cùng với sự phát triển của công nghệ thông tin, việc khảo sát cấu trúc và tính chất của enzyme từ dữ liệu lớn này có thể thực hiện được để đưa ra những thông tin mới về cấu trúc cũng như đặc điểm của nhóm enzyme trong hệ sinh thái, cho phép lựa chọn enzyme có đặc tính quý cho nghiên cứu ứng dụng.

NGUYÊN VẬT LIỆU VÀ PHƯƠNG PHÁP**Vật liệu nghiên cứu**

Dữ liệu DNA metagenome có kích thước 8,4 Gb của vi khuẩn trong dạ cỏ dê thu được từ giải trình tự bằng hệ thống HiSeq Illumina 2500 (Illumina, San Diego, Hoa Kỳ). Từ 8,4 Gb dữ liệu, chúng tôi đã lắp ráp và ước đoán được 164,644 gene. Sau đó sử dụng công cụ BLASTall, các gen này đã được so sánh với cơ sở dữ liệu CAZY (carbohydrate-active enzymes database- CAZY, <http://www.cazy.org>), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2008), Cluster of Orthologous Groups (COG) (Powell *et al.*, 2011) để chú giải chức năng của từng gen (Do *et al.*, 2018a, 2018b).

Nghiên cứu khai thác các trình tự mã hóa beta-glucosidase từ bộ dữ liệu DNA metagenome

Từ các kết quả chú giải chức năng dựa trên bộ dữ liệu CAZY, KEGG, COG, các trình tự mã hóa cho BGL sẽ được trích xuất ra một file riêng. Các thông tin về mã gen, trình tự gen, trình tự amino acid suy diễn từ gen được trích xuất bằng công thức Vlookup trong Microsoft Excel. Độ tương đồng của trình tự amino acid với các trình tự trong dữ liệu KEGG/CAZY/NR NCBI sẽ được trích xuất theo cùng trình tự.

Nghiên cứu đánh giá đa dạng phổ pI và chỉ số axit/kiềm dự đoán của beta-glucosidase

Khối lượng phân tử và giá trị pI của mỗi trình tự enzyme BGL được ước đoán bằng công cụ ước đoán trong EXPASY (https://web.expasy.org/compute_pi/). Khoảng pH hoạt động của enzyme được ước đoán dựa trên phần mềm AcalPred (<http://lin.uestc.edu.cn/server/AcalPred>). Đây là công cụ có thể đạt được độ chính xác tổng thể là 94,01% (Fan *et al.*, 2013). Khi một trình tự amino acid nhập vào, phần mềm sẽ tự động tính toán và cho ra kết quả cuối cùng về khả năng chịu kiềm/axit của protein sau vài phút. Kết quả cuối là một giá trị trong khoảng giới hạn từ 0 đến 1 theo mức tăng của chỉ số truy vấn. Nếu việc khảo sát mức độ hoạt động trong vùng alkaline thì trình tự nào có chỉ số càng gần 1 là trình tự có khả năng hoạt động tốt ở vùng alkaline. Do vậy, đối với 1 trình tự, tổng giá trị alkaline và acid sẽ bằng 1.

Dự đoán khả năng chịu nhiệt của enzyme

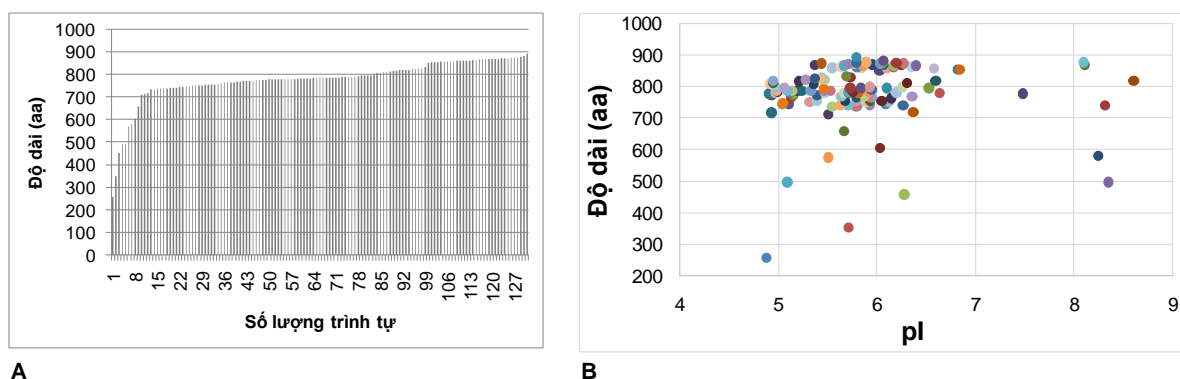
Dựa trên thành phần và trình tự của các amino acid, liên kết hydro, liên kết Vanderwaals, tương tác kỵ nước và đặc điểm của các enzyme từ các sinh vật sống trong điều kiện môi trường có nhiệt độ cao (Ebrahimi *et al.*, 2011), Phần mềm của TBI (www.tbi.org.tw/tools) xây dựng trên số liệu của 150.000 protein chịu nhiệt độ khác nhau trong ngân hàng NCBI để dự đoán khả năng chịu nhiệt, dựa trên nguyên tắc tương đồng. Khả năng chịu nhiệt được dự đoán ở ba mức là trên 65°C, 55-65°C và dưới 55°C.

Dự đoán cấu trúc của enzyme

Một số công cụ dự đoán cấu trúc được ứng dụng nhiều trong những năm gần đây là Swiss-Prot (có tích hợp AlphaFold), Phyre2 hay Comparative modelling. Mỗi công cụ có ưu điểm riêng, phù hợp với từng nhóm trình tự có độ tương đồng với cấu trúc có sẵn khác nhau. Công cụ Swiss-Prot cho phép ước đoán cấu trúc phân tử của các trình tự mới, có độ tương đồng thấp, khó ước đoán được bằng Phyre2. Phyre2 là công cụ dự đoán cấu trúc bậc ba của các trình tự dựa trên nguyên lý nhận biết vùng cuộn xoắn theo các tương tác phân tử đối với các trình tự có độ tương đồng với cấu trúc đã biết trên 30%. Phương pháp Comparative modelling có thể giúp dự đoán cấu trúc phân tử đối với các trình tự có độ tương đồng cao trên 50% so với trình tự đã biết. Trong nghiên cứu này, các enzyme được dự đoán cấu trúc dựa trên trình tự amino acid bằng công cụ Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2>, truy cập tháng 6/2024) với chế độ mô hình hóa mở rộng (intensive) do chúng có độ tương đồng phù hợp với với cấu trúc đã biết khoảng trên 30%. Các vùng bảo thủ của enzyme được ước đoán dựa trên công cụ ước đoán vùng bảo thủ của gói công cụ trực tuyến BLAST.

KẾT QUẢ VÀ THẢO LUẬN**Khai thác các trình tự mã hóa beta-glucosidase**

Từ 164,644 gene, chúng tôi đã thu được 387 trình tự mã hóa betaglucosidase. Trong số các trình tự này chỉ có 131 trình tự hoàn thiện (chiếm 29,2%) 88 trình tự thiếu đầu 3', 60 trình tự thiếu đầu 5' và 103 trình tự thiếu cả hai đầu. Trong số các trình tự hoàn thiện có 43 trình tự được chú giải chức năng dựa trên dữ liệu COG, 113 trình tự chú giải chức năng bởi KEGG (đạt 86,25%), 131 trình tự được chú giải bởi CAZY (đạt 100%). Như vậy, dữ liệu CAZY cho đạt tỉ lệ chú giải gen thành công cao nhất, đạt 100%.



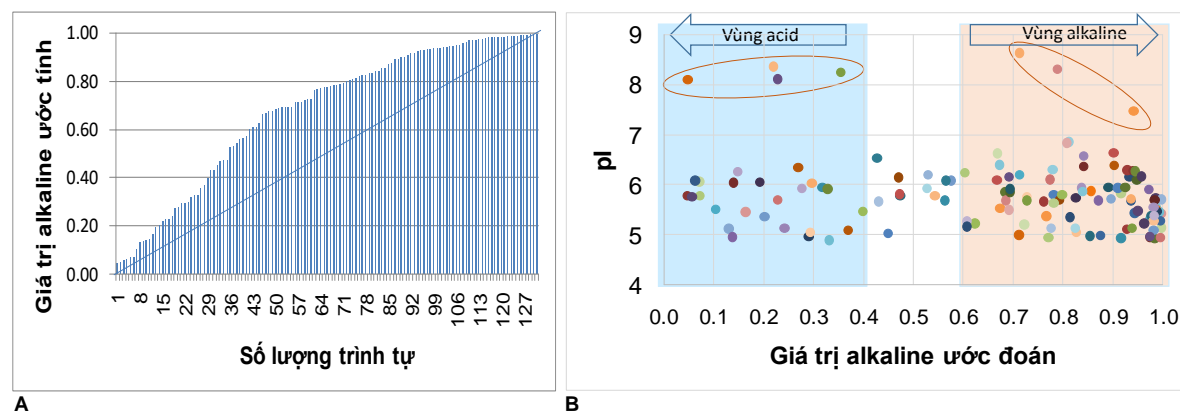
Hình 1. Biểu đồ phân bố độ dài (A) và pI (B) của 131 trình tự amino acid của beta-glucosidase từ bộ dữ liệu DNA metagenome trong dạ cỏ dê

Kết quả khảo sát độ dài các trình tự cho thấy, trình tự gen có độ dài ngắn nhất là 257 amino acid và dài nhất là 894 amino acid. Trong đó có 122 trình tự (chiếm 93,1%) có độ dài trên 700 amino acid. Phần lớn các trình tự có chỉ số pI nghiêng về phía axit nhẹ, nằm trong khoảng 5,0-6,5 chỉ có 1 trình tự có chỉ số pI thấp dưới 5 và 6 trình tự có pI cao trên 7 (Hình 1B). Trình tự có kích thước thấp nhất được ước đoán có chỉ số pI thấp nhất. Không có mối liên hệ nào giữa độ dài và giá trị pI của beta-glucosidase.

Thông thường các trình tự beta-glucosidase GH1 thường có kích thước khoảng 400-500 amino acid. Điển hình như BGL GH1 của *Trichoderma* spp. thường có kích thước trong khoảng 455 đến 466 amino acid (Mohamad Sobri *et al.*, 2020). Các beta-glucosidase GH3 thường có kích thước lớn hơn do phân tử thường có 2 module cho hoạt tính xúc tác là GH3N và GH3C cùng với các vùng phụ trợ khác giúp enzyme bám lên cơ chất (Nguyen *et al.*, 2021). Trong nghiên cứu này, tất cả các trình tự đều được CAZy ước đoán là beta-glucosidase GH3 và phần lớn có thêm vùng FN3 ở phía sau vùng GH3C. Đây là cấu trúc khá điển hình của enzyme beta-glucosidase trong dạ cỏ dê. Các nghiên cứu trước cho thấy điểm đẳng điện của các beta-glucosidase trong bộ dữ liệu nghiêng về vùng axit nhẹ, trong khi đó các enzyme beta-glucosidase GH1 của *Trichoderma* spp. có điểm đẳng điện dao động trong khoảng 5,10-5,53 (Mohamad Sobri *et al.*, 2020).

Khảo sát khả năng hoạt động ở vùng pH kiềm của enzyme

Theo ước đoán bằng phần mềm AcalPred, các enzyme beta-glucosidase trong dạ cỏ dê có khả năng hoạt động trong vùng pH rất rộng, từ vùng pH rất axit (với chỉ số alkaline thấp nhất (0,05)) đến vùng hoạt động rất kiềm (chỉ số alkaline cao nhất là 1). Nếu phân loại chỉ số alkaline thành 3 vùng là vùng axit (chỉ số alkaline $\leq 0,4$) vùng trung tính (lớn hơn 0,4 và nhỏ hơn 0,6) và vùng kiềm (từ 0,6 đến 1,0) thì 30 enzyme hoạt động ở vùng axit, 12 enzyme hoạt động ở vùng trung tính và 89 enzyme (chiếm 67,9%) hoạt động ở vùng pH kiềm (Hình 2). Quan sát trên biểu đồ (Hình 2B) có thể thấy, sự phân bố enzyme hoạt động ở các vùng pH khác nhau là ngẫu nhiên, không có mối liên quan tới điểm đẳng điện của enzyme. Tuy nhiên, có 4 enzyme có giá trị pI cao lại hoạt động tốt ở vùng pH axit, trong khi đó có 4 enzyme có pI cao hoạt động ở vùng pH cao, không có enzyme pI thấp hoạt động ở vùng pH cao và cũng không nhìn thấy enzyme có pI thấp hoạt động ở vùng pH thấp (Bảng 1). Kết quả khảo sát mối liên quan giữa độ dài trình tự và giá trị alkaline của enzym cũng cho thấy hai đại lượng này không có mối liên hệ với nhau (kết quả không được trình bày).

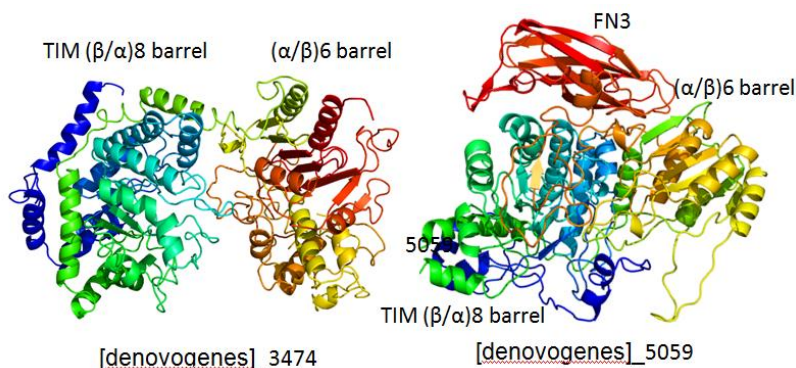


Hình 2. Biểu đồ phân bố số lượng các trình tự beta-glucosidase theo chỉ số alkaline (A) và sự tương quan giữa chỉ số alkaline với pI của 131 trình tự

Bảng 1. Một số đặc điểm của 4 trình tự có pI cao hoạt động trong vùng pH thấp và 3 trình tự pI cao hoạt động trong vùng pH cao trong bộ dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê

| Mã trình tự | Độ dài (amino acid) | Khối lượng (kDa) | pI | Vùng xúc tác | Vùng đầu C | Độ tương đồng (%) | Mã CaZy/KEGG | Điểm số acid | Điểm số alkaline | Tm |
|---------------------|---------------------|------------------|------|--------------|------------|-------------------|--------------|--------------|------------------|-------|
| [denovogenes]_3474 | 819 | 90,6 | 8,62 | GH3 | | 40,93 | CBK66838 | 0,29 | 0,71 | >65 |
| [denovogenes]_17198 | 493 | 54,1 | 8,35 | GH3 | FN3 | 46,36 | K05349 | 0,78 | 0,22 | >65 |
| [denovogenes]_5059 | 741 | 81,7 | 8,32 | GH3 | FN3 | 44,48 | K05349 | 0,21 | 0,79 | 55-65 |
| [denovogenes]_11003 | 580 | 64,1 | 8,25 | GH3 | FN3 | 86,14 | K05349 | 0,65 | 0,35 | 55-65 |
| [denovogenes]_2683 | 869 | 97,0 | 8,11 | GH3 | FN3 | 76,46 | K01188 | 0,95 | 0,05 | 55-65 |
| [denovogenes]_2632 | 874 | 96,9 | 8,1 | GH3 | FN3 | 76,62 | K01188 | 0,77 | 0,23 | >65 |
| [denovogenes]_4363 | 775 | 84,4 | 7,48 | GH3 | FN3 | 60,57 | K05349 | 0,06 | 0,94 | 55-65 |

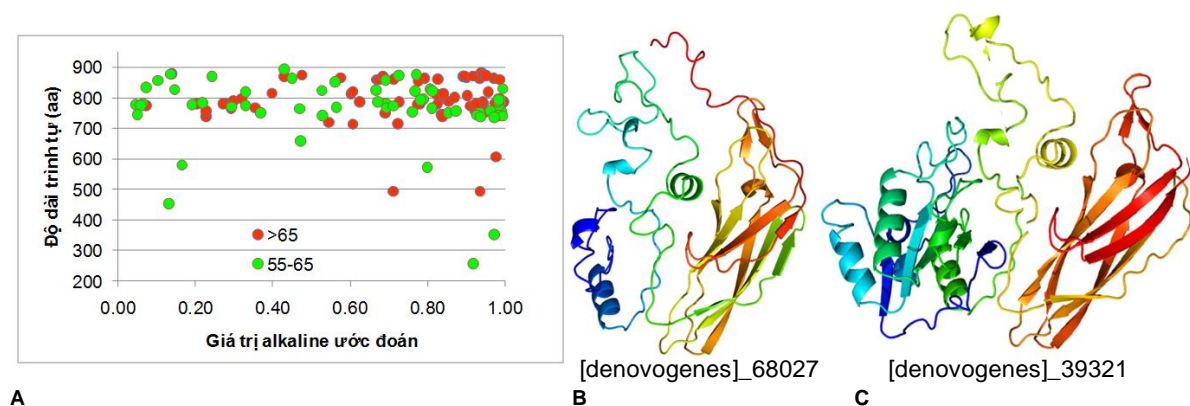
Dựa trên phần mềm SignalP 5.0 của Trường Đại học Kỹ thuật Đan Mạch (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>), trong số 7 trình tự ở trên, các trình tự có kích thước lớn hơn 700 amino acid đều có chứa tín hiệu tiết, còn hai trình tự kích thước nhỏ không chứa tín hiệu tiết. Tất cả các trình tự đều có hai vùng xúc tác của GH3N và GH3C trong đó 5 trình tự có vùng FN3 được cấu tạo bởi phiến beta. Các trình tự đều có cấu trúc đặc thù của beta-glucosidase gồm cấu trúc TIM (β/α)8 barrel, (α/β)6 barrel (Hình 3). Các cấu trúc beta glucosidase đều tương đồng 31-51% so với cấu trúc của beta-glucosidase trong dữ liệu PDB. Trình tự FN3 đã được nghiên cứu có khả năng làm tăng giá trị pI cho enzyme (Nguyen *et al.*, 2021). Enzyme mã hóa từ gen [denovogenes]_3474 có cấu trúc tương đồng 33% với glycosyl hydrolase family 3 (Kim *et al.*, 2011). Enzyme mã hóa từ gen [denovogenes]_5059 có cấu trúc tương đồng 37% với beta-glucosidase chịu nhiệt của *Clostridium thermocellum* (Almeida, Muniz, 2022). Quan sát cấu trúc của các enzyme, chúng tôi không nhận thấy các điểm khác biệt giữa các enzyme có pI cao nhưng có khả năng hoạt động ở vùng pH axit và enzyme có pI cao nhưng dự đoán hoạt động trong vùng pH bazơ.



Hình 3. Cấu trúc dự đoán của hai beta-glucosidase điển hình mã hóa từ hai trình tự gen

Khảo sát khả năng bền nhiệt của các enzyme beta-glucosidase

Dựa trên phần mềm TBI (www.tbi.org.tw/tools), khả năng bền nhiệt của enzyme thông qua chỉ số Tm đã được ước đoán. Kết quả, 100% các trình tự được ước đoán có khả năng bền nhiệt cao từ 55°C trở lên và không có trình tự nào bền nhiệt thấp dưới 55°C. Trong đó 53% số lượng enzyme (69 enzyme) được ước đoán bền ở nhiệt độ trên 65°C, đồng nghĩa là 53% enzyme sẽ bị mất hoạt tính khi nhiệt độ tăng lên trên 65°C và 47% số enzyme còn lại bị mất hoạt tính ở nhiệt độ 55-65°C. Kết quả khảo sát mối liên hệ giữa độ dài trình tự với khả năng bền nhiệt của enzyme cho thấy, hai đại lượng này không có mối quan hệ với nhau. Trong đó đáng chú ý là hai enzyme có trình tự ngắn nhất lại được dự đoán hoạt động ở trong vùng pH kiềm. Trình tự [denovogenes]_68027 tạo thành từ 257 amino acid là trình tự ngắn nhất trong bộ dữ liệu, tiếp theo đó là trình tự [denovogenes]_39321 có 352 amino acid. Kết quả khảo sát cấu trúc của enzyme bằng Phyre2 cho thấy cả hai trình tự có cấu trúc tương đồng cao với beta-glucosidase trong đó enzyme [denovogenes]_68027 được ước đoán có cấu trúc tương đồng 52% so với cấu trúc của beta-glucosidase bglpcc1 của *Prevotella copri* (Zhou *et al.*, 2023). Enzyme mã hóa từ gen còn lại có cấu trúc tương đồng 45% với cấu trúc của beta-glucosidase của *Kluyveromyces marxianus* (Yoshida, Hidaka, 2010). Cả 2 enzyme đều có cấu trúc FN3 ở vùng đầu C và một phần của cấu trúc (α/β)6 barrel ở đầu N. Do vậy rất có thể enzyme bị thiếu đi vùng GH3N. Tuy nhiên, để chắc chắn đây có phải là enzyme hoàn thiện, có thể hiện chức năng sinh học hay không thì cần phải được nghiên cứu tiếp bằng sinh học phân tử tái tổ hợp gen.



Hình 4. Biểu đồ phân bố các enzyme beta-glucosidase dựa theo độ dài, giá trị alkaline và khả năng bền nhiệt của các enzyme (A), cấu trúc ước đoán của hai enzyme mã hóa từ hai gen điển hình (B, C)

KẾT LUẬN

Trong số 131 trình tự hoàn thiện mã hóa cho beta-glucosidase được khai thác từ dữ liệu giải trình tự đa hệ gen của vi khuẩn trong dạ cỏ dê, 100% trình tự có vùng xúc tác GH3, 93,1% trình tự có kích thước lớn trên 700 amino acid, 67,9% enzyme có khả năng hoạt động ở vùng pH kiềm (chỉ số alkaline $\geq 0,6$), 52,7% enzyme ước đoán bền ở nhiệt độ trên 65°C. Kết quả này có ý nghĩa cho lựa chọn trình tự mã hóa enzyme có đặc tính mong muốn cho biểu hiện protein dạng tái tổ hợp.

Lời cảm ơn: Công trình được hỗ trợ kinh phí từ đề tài nghiên cứu mã số NCVCC08.05/24-25 do PGS. TS Đỗ Thị Huyền làm chủ nhiệm.

TÀI LIỆU THAM KHẢO

- Ahmed A, Nasim F, ul-Hassan, Batool K, Bibi A (2017). Microbial β -glucosidase: Sources, production and applications. *J Appl Environ Microbiol* 5: 31–46.
- Almeida L, Muniz J (2022). RCSB PDB - 7MS2: Three-dimensional structure of a GH3 beta-glucosidase from *Clostridium thermocellum* in complex with glycerol.
- Do TH, Dao TK, Nguyen KHV, Le NG, Nguyen TMP, Le TL, Phung TN, van Straalen NM, Roelofs D, Truong NH (2018a). Metagenomic analysis of bacterial community structure and diversity of lignocellulolytic bacteria in Vietnamese native goat rumen. *Asian-Australas J Anim Sci* 31: 738–747.
- Do TH, Le NG, Dao TK, Nguyen TMP, Le TL, Luu HL, Nguyen KHV, Nguyen VL, Le LA, Phung TN, van Straalen NM, Roelofs D, Truong NH (2018b). Metagenomic insights into lignocellulose-degrading genes through Illumina-based de novo sequencing of the microbiome in Vietnamese native goats' rumen. *J Gen Appl Microbiol* 64: 108–116.
- Ahmed A, Nasim F ul-Hassan, Batool K, Bibi A (2017). Microbial β -glucosidase: Sources, production and applications. *J Appl Environ Microbiol*, 5: 31–46.
- Almeida L, Muniz J (2022). RCSB PDB - 7MS2: Three-dimensional structure of a GH3 beta-glucosidase from *Clostridium thermocellum* in complex with glycerol.
- Ebrahimi M, Lakizadeh A, Agha-Golzadeh P, Ebrahimie E, Ebrahimi M (2011). Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. *PLoS ONE*,6.
- Fan GL, Li QZ, Zuo YC (2013). Predicting acidic and alkaline enzymes by incorporating the average chemical shift and gene ontology informations into the general form of Chou's PseAAC. *Process Biochem*, 48: 1048–1053.
- Godse R, Bawane H, Tripathi J, Kulkarni R (2021). Unconventional β -glucosidases: A promising biocatalyst for industrial biotechnology. *Appl Biochem Biotechnol*, 193: 2993–3016.
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res*, 36: D480–484.
- Kim Y, Chhor G, Bearden J, Joachimiak A (2011). X-ray diffraction data for the crystal structure of glycoside hydrolase from *Synechococcus* (3SQL). *IRRM*.
- Ma LJ, Liu X, Guo L, Luo Y, Zhang B, Cui X, Yang K, Cai J, Liu F, Ma N, Yang FQ, He X, Shi SP, Wan JB (2024). Discovery of plant chemical defence mediated by a two-component system involving β -glucosidase in *Panax* species. *Nat Commun*, 15: 602.
- Mohamad Sobri MF, Abd-Aziz S, Abu Bakar FD, Ramli N (2020). *In-silico* characterization of glycosyl hydrolase family 1 β -glucosidase from *Trichoderma asperellum* UPM1. *Int J Mol Sci*, 21: 4035.
- Nguyen KHV, Dao TK, Nguyen HD, Nguyen K. H., Nguyen TQ, Nguyen TT, Nguyen TMP, Truong NH, Do TH (2021). Some characters of bacterial cellulases in goats' rumen elucidated by metagenomic DNA analysis and the role of fibronectin 3 module for endoglucanase function. *Anim Biosci*, 34: 867–879.

Powell S, Szklarczyk D, Trachana K, Roth A., Kuhn M, Muller J, Arnold R, Rattei T, Letunic, Doerks T, Jensen LJ, von Mering C, Bork P (2011). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res*, 40: D284–D289.

Rocha REO, Mariano DCB, Almeida TS, CorrêaCosta LS, Fischer PHC, Santos LH, Caffarena ER, da Silveira CH, Lamp LM, Fernandez-Quintero ML, Liedl KR, de Melo-Minardi RC, de Lima LHF (2023). Thermostabilizing mechanisms of canonical single amino acid substitutions at a GH1 β -glucosidase probed by multiple MD and computational approaches. *Proteins*, 91: 218–236.

Seeberger PH, Werz DB (2007). Synthesis and medical applications of oligosaccharides. *Nature*, 446: 1046–1051.

Tran TNA, Son JS, Awais M, Ko JH, Yang DC, Jung SK (2023). β -Glucosidase and its application in bioconversion of ginsenosides in *Panax ginseng*. *Bioeng Basel Switz*, 10: 484.

Yoshida E, Hidaka M (2010). RCSB PDB - 3AC0: Crystal structure of Beta-glucosidase from *Kluyveromyces marxianus* in complex with glucose.

Zhou J, Zhu H, Zhang R, Huang Z (2023). RCSB PDB - 8IDZ: Beta-glucosidase BglPcC1.

***In silico* PREDICTION OF SOME PROPERTIES OF BETA-GLUCOSIDASE MINED FROM METAGENOMIC DNA DATA 8.4 GB OF BACTERIA IN GOATS RUMEN**

Do Thi Huyen^{1,2*}, Nguyen Minh Chien^{2,3}, Truong Nam Hai¹

¹*Institute of Biotechnology–Vietnam Academy of Science and Technology*

²*Graduate University of Science and Technology - Vietnam Academy of Science and Technology*

³*Hanvet Pharmaceutical Company Limited*

SUMMARY

In a previous study, metagenomic DNA of bacteria in goats' rumen was sequenced to yield of 8.4 Gb, from which 164 644 genes were predicted. Using CAZy, KEGG, COG databases, 384 genes coding for beta-glucosidase were mined, of which 131 sequences were estimated to be complete sequences. The majority proportion (93.1%) of the enzymes coded by the complete genes had long size of over 700 amino acids. The significant portion of the enzymes possessed pI value in the range of 5.0-6.5, contained a catalytic domain belonging to GH3 family and an auxiliary region FN3. The enzymes were estimated to have a high alkaline score, of which 67.9% of enzymes were predicted to be capable of exhibiting activity in the alkaline pH region (alkaline index ≥ 0.6). Among them, four enzymes have high pI scores but were estimated to work well in the acidic pH region and three enzymes had high pI scores but were estimated to work well in the alkaline pH region. Of these seven enzymes, five enzymes were identified as secretory enzymes because of containing a signal peptide sequence at the N-terminus. More than half of the enzymes (69 enzymes) were estimated to be stable at temperatures up to 65°C. The two enzymes with the lowest sizes were predicted to be able to function at alkaline pH. The amino acid length, pI value, alkaline value, Tm and enzyme structure were independent variables, no relationship between each others were observed. This study has provided general information about some basic characteristics of beta-glucosidase estimated based on bioinformatics analysis, which is valuable for gene mining and expression.

Keywords: DGGE, PCR tough-up program, bacterial diversity, persistent diarrhea, V3 gene region.

* Author for correspondence: Tel: 0437917021; Email: dohuyen@ibt.ac.vn