



(12) **BẢN MÔ TẢ GIẢI PHÁP HỮU ÍCH THUỘC BẰNG ĐỘC QUYỀN
GIẢI PHÁP HỮU ÍCH**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ**

(11)



2-0004429

(51) **C12N 15/00**
2022.01

(13) **Y**

(21) 2-2022-00244

(22) 16/06/2022

(45) 27/10/2025 451

(43) 25/12/2023 429A

(73) Viện Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (VN)
18 Hoàng Quốc Việt, phường Nghĩa Đô, thành phố Hà Nội

(72) TRƯƠNG NAM HẢI (VN); NGUYỄN THỊ QUÝ (VN); ĐỖ THỊ HUYỀN (VN); LÊ
THỊ THU HỒNG (VN); ĐÀO TRỌNG KHOA (VN); NGUYỄN HỒNG DƯƠNG
(VN).

(54) TRÌNH TỰ GEN BGC MÃ HÓA BETA-GLUCOXIDAZA CÓ NGUỒN GỐC TỪ
VI KHUẨN TRONG DẠ CỎ DÊ VÀ CẤU TRÚC BETA-GLUCOXIDAZA GH3-
FN3-GH31 MÃ HÓA BỞI GEN NÀY

(21) 2-2022-00244

(57) Giải pháp hữu ích đề cập đến trình tự gen *bgc* mã hóa beta-glucoxidaza có nguồn gốc từ vi khuẩn trong dạ cỏ dê và trình tự axit amin suy diễn từ gen, có cấu trúc mô đun GH3-FN3-GH31, mã hóa beta-glucoxidaza. Bằng công nghệ giải trình tự gen thế hệ mới phân tích đa hệ gen, gen *bgc* mới, mã hóa cho beta-glucoxidaza đã được khai thác từ dữ liệu giải trình tự DNA đa hệ gen của vi khuẩn trong dạ cỏ dê. Việc phân tích cấu trúc vùng bảo tồn, vùng hoạt tính dựa trên các phần mềm tin sinh học đã cho thấy beta-glucoxidaza có cấu trúc mô đun mới gồm ba vùng GH3-FN3-GH31 được sắp xếp tuần tự từ đầu N tới đầu C. Kết hợp với công nghệ DNA tái tổ hợp, gen *bgc* đã được biểu hiện mạnh trong tế bào biểu hiện *Escherichia coli* Rosseta 1. Sản phẩm beta-glucoxidaza tái tổ hợp sau tinh sạch có hoạt tính cao trung bình đạt 35 U/mg trên cơ chất pNPG.

Lĩnh vực kỹ thuật được đề cập

- 5 Giải pháp hữu ích đề cập đến trình tự gen *bgc* mã hóa beta-glucoxidaza có nguồn gốc từ vi khuẩn trong dạ cỏ dê và trình tự axit amin suy diễn từ gen, có cấu trúc mô đun GH3-FN3-GH31, mã hóa beta-glucoxidaza.

Tình trạng kỹ thuật của giải pháp hữu ích

- 10 Việt Nam là đất nước có nền nông nghiệp phát triển. Trữ lượng sinh khối phế phụ phẩm của nông nghiệp hàng năm lên tới trên 150 triệu tấn. Trong đó phế phụ phẩm từ sinh khối thực vật gồm rơm, rạ, bã mía, trấu, vỏ hạt các loại trái cây có trữ lượng khoảng 89 triệu tấn, lâm nghiệp khoảng 5,5 triệu tấn. Nguồn sinh khối thực vật này được đánh giá là nguồn sinh khối xanh rất có giá trị cho phát triển kinh tế, đặc biệt là cho sản xuất điện và năng lượng sinh học điển hình như cồn sinh học và một
- 15 số sản phẩm có giá trị như nhựa sinh học, xylitol, dược phẩm,... (Patel & Shah, 2021). Ở nước ta, theo thông tin của Bộ Công thương, tại thời điểm tháng 12 năm 2021 sinh khối được tận dụng để sản xuất điện sinh học với tỷ lệ rất thấp (chỉ chiếm 0,14% điện thương phẩm và 0,94% công suất lắp đặt trên toàn quốc). Một phần nhỏ sinh khối được tận dụng làm phân bón hữu cơ, trồng nấm, nấu cơm,... Tổng sinh khối
- 20 nông nghiệp được sử dụng chỉ chiếm 11%, 89% lượng còn lại đang bị vứt bỏ, gây lãng phí, ô nhiễm môi trường sinh thái.

- Trong khoảng 7 năm trở lại đây, nhiên liệu sinh học được sản xuất và ứng dụng rất mạnh mẽ ở nhiều nước trên thế giới. Năm 2018, cồn sinh học sản xuất từ sinh khối thực vật lignoxenluloza đạt 3 tỷ lít. Dự kiến, đến năm 2030, sản lượng cồn sinh
- 25 học sẽ đạt khoảng 233 tỷ lít (Morales, Arvesen, & Cherubini, 2021). Ở nước ta, việc khai thác, chuyển hóa nguồn sinh khối này còn rất hạn chế do trở ngại lớn nhất là hiệu quả chuyển hóa sinh khối thành đường để lên men còn rất thấp. Lignoxenluloza được cấu tạo từ ba thành phần chủ yếu là xenluloza (sợi polyme mạch thẳng của đường glucoza liên kết với nhau bởi liên kết 1,4-beta glucosit), hemixenluloza

(polyme phân nhánh của các đường 5 các bon và đường 6 các bon) và lignin. Cấu trúc của lignoxenluloza là cấu trúc vững chắc. Xenluloza là thành phần chính, chiếm khoảng 40-60% sinh khối lignoxenluloza (Su *et al.*, 2020). Mặc dù là phân tử có cấu trúc đơn giản nhất trong lignoxenluloza nhưng việc chuyển hóa xenluloza thành đường cũng cần ít nhất ba loại enzym đó là exo-glucanaza, endo-glucanaza và beta-glucosidaza. Ba enzym này hoạt động theo kiểu hiệp đồng: endo-glucanaza (EC 3.2.1.4) phân cắt ngẫu nhiên sợi xenluloza thành các đoạn oligosaccharit ngắn, exo-glucanaza (1,4- β -xylanobiosidaza, xylanohydrolaza, EC 3.2.1.91, EC 3.2.1.176) cắt sợi oligosaccharit ngắn ở hai đầu để tạo các phân tử xylanbioza (có hai gốc đường glucoza) và cuối cùng beta-glucosidaza (EC 3.2.1.21) phân cắt xylanbioza thành đường đơn glucoza (Suwan, Arthornthurasuk, & Kongsaree, 2017). Beta-glucosidaza là enzym tham gia vào bước cuối cùng trong con đường chuyển hóa xenluloza thành đường nên nó quyết định hiệu quả của cả con đường chuyển hóa (Zhang *et al.*, 2019; Srivastava *et al.*, 2019). Do beta-glucosidaza có vai trò đặc biệt quan trọng trong nền công nghiệp nhiên liệu sinh học nên việc nghiên cứu tìm kiếm enzym có hoạt tính cao, có khả năng sản xuất lượng lớn để đáp ứng nhu cầu enzym trong công nghiệp đang là vấn đề được các nhà nghiên cứu đặc biệt quan tâm (Srivastava *et al.*, 2019).

Beta-glucosidaza công nghiệp được sản xuất từ nhiều vi sinh vật như nấm *Trichoderma reesei*, *Aspergillus niger*, *A. phoenicis*, *A. saccharolyticus*,...; vi khuẩn *Bacillus subtilis*, *Acidothermus cellulyticus* (Srivastava *et al.*, 2019). Bằng công nghệ DNA tái tổ hợp, beta-glucosidaza từ chủng *B. licheniformis* được sản xuất từ chủng *E. coli* tái tổ hợp với hoạt tính cao hiện nay đạt 45,44 U/mL (Liu *et al.*, 2021). Beta-glucosidaza tái tổ hợp được sản xuất từ *E. coli* có sản lượng cao nhất đạt 15 g/L (Ferreira, Azzoni, & Freitas, 2018).

Về cấu trúc, beta-glucosidaza có nguồn gốc từ nấm, thực vật, vi khuẩn đều có chung trình tự bảo thủ. Việc nhận biết và phân loại enzym hiện nay đều dựa trên hai cách đó là kiểm tra tính đặc hiệu cơ chất của enzym và dựa trên độ tương đồng về trình tự nucleotit và axit amin (Singhania *et al.*, 2013; Singh, Verma, & Kumar, 2016). Việc phân loại enzym dựa trên cả độ tương đồng về trình tự và kiểu cấu trúc

(mô đun) được các nhà khoa học chấp nhận nhiều nhất (Henrissat & Davies, 1997; Li *et al.*, 2012). Theo cách phân loại này, hiện nay (tại thời điểm ngày 15/5/2022), các enzym có hoạt tính phân cắt cacbon hydrat (carbohydrate active enzyme - CAZy) được phân vào 173 họ trong đó beta-glucosidaza chủ yếu thuộc họ GH1, GH3 và một số họ GH khác như GH2, GH5, GH9, GH16, GH30, GH39, GH116 (Chen *et al.*, 2021). Beta-glucosidaza GH1 chủ yếu được sản xuất từ cổ khuẩn, thực vật, động vật, trong khi đó họ GH3 được tổng hợp chủ yếu từ vi khuẩn, nấm men, nấm mốc. Mỗi họ lại được phân thành các dòng (clan) dựa trên sự tương đồng về nguồn gốc và cơ chế xúc tác (Singhanian *et al.*, 2013). Beta-glucosidaza có trình tự rất đa dạng nhưng vùng xúc tác của mỗi họ GH lại rất giống nhau. Beta-glucosidaza GH1, 5, 30 có cấu trúc ống được tạo thành từ 8 phiến beta/xoắn anpha gọi là cấu trúc ống (β/α)₈ và có hai gốc axit cacboxylic bảo thủ nằm ở vị trí xúc tác (Ketudat Cairns & Esen, 2010). Các beta-glucosidaza GH3 lại có hai vùng cấu trúc được tạo thành từ cấu trúc ống (β/α)₈ tiếp theo là sáu phiến beta xếp kẹp hai phía của ba xoắn anpha được gọi là cấu trúc (β/α)₆. Vị trí xúc tác của enzym nằm giữa hai vùng cấu trúc (β/α)₈ và (β/α)₆ (Varghese, Hrmova, & Fincher, 1999; Ketudat Cairns & Esen, 2010).

Các enzym thuộc họ xenlulaza, hemixenlulaza thường có cấu trúc gồm nhiều vùng, các vùng có cấu trúc riêng biệt, có chức năng khác nhau, giúp đảm bảo cho enzym hoạt động được hiệu quả nhất. Dạng cấu trúc này được gọi là cấu trúc mô đun (modular structure). Các enzym thường có ít nhất một mô đun xúc tác. Có enzym có nhiều mô đun xúc tác, chúng thuộc cùng một họ hoặc nhiều họ GH khác nhau. Ngoài vùng xúc tác ra, enzym cũng chứa nhiều các mô đun hỗ trợ cho hoạt tính xúc tác như các vùng bám cơ chất cacbon hydrat (carbohydrate binding module - CBM), vùng tương tự fibronectin-3 (FN3), vùng tương tự globulin miễn dịch (immunoglobulin - Ig) và vùng không biết chức năng (vùng X). Beta-glucosidaza thường không có vùng CBM (Alvarez-Gonzalez *et al.*, 2022). Trong những năm gần đây, ứng dụng kỹ thuật giải trình tự đa hệ gen (metagenomics) các nhà khoa học đã tìm kiếm được rất nhiều cấu trúc enzym mới, có trình tự tương đồng thấp với trình tự gen đã biết. Các nguồn gen mới này chủ yếu được khai thác từ vi sinh vật chưa nuôi cấy được, được cho là chiếm tỷ lệ rất lớn tới 99% trong mỗi hệ sinh thái nhỏ. Dựa trên phân tích trình tự

beta-glucosidaza từ dữ liệu đa hệ gen, cấu trúc thông thường nhất của enzym họ GH3 của vi khuẩn trong dạ cỏ dê là có chứa thêm vùng FN3. Diễn hình, dữ liệu giải trình tự đa hệ gen của vi khuẩn trong dạ cỏ dê với dung lượng 8 Gb, trên 90% trình tự beta-glucosidaza của vi khuẩn được khai thác có chứa FN3 và gần 10% còn lại chỉ chứa vùng xúc tác GH3 (Nguyen *et al.*, 2021). Mô đun FN3 giúp tăng khả năng bám dính của enzym vào cơ chất, giúp ổn định phân tử, tăng hoạt tính riêng của enzym (Mohsin *et al.*, 2019; Nguyen *et al.*, 2021). Ngoài vùng FN3 ra, cấu trúc beta-glucosidaza mới cũng được phát hiện có thêm vùng CBM6. Đây là dạng cấu trúc được tìm thấy trong dữ liệu giải trình tự đa hệ gen vi sinh vật trong dạ cỏ bò Tây Tạng (Yak) (Dai *et al.*, 2012). Cấu trúc này cũng được phát hiện thấy trong dữ liệu DNA đa hệ gen của vi khuẩn trong dạ cỏ dê khi giải trình tự sâu với dung tích lên tới 48,66 Gb (Dao *et al.*, 2021). Trong bộ dữ liệu này, các trình tự được ước đoán mã hóa cho beta-glucosidaza được khai thác là 859 trình tự trong đó có 638 trình tự mã hóa cho enzym thuộc họ GH3. Các cấu trúc được phát hiện thấy nhiều nhất vẫn là GH3-FN3 (500 trình tự). Ngoài ra, các cấu trúc khác được phát hiện là GH3-FN3-CE8, DUF-GH3-FN3, GH3-FN3-GH5, GH3-FN3-Ig và một trình tự có cấu trúc domain là GH3-FN3-GH31. Cấu trúc enzym GH3-FN3-GH31 là một cấu trúc enzym mới, chưa từng được phát hiện và chưa được khẳng định hoạt tính enzym thông qua hoạt tính phân cắt đặc hiệu. Các bài báo thường đề cập đến beta-glucosidaza có vùng xúc tác là GH3-FN3 nhưng không có thêm vùng GH31. Bằng độc quyền sáng chế EP2995687B1, US 9,708,594 B2 mô tả về trình tự beta-glucosidaza chịu nhiệt GH3 và có đề cập đến beta-glucosidaza có vùng GH3 hoặc GH31 nhưng không đề cập đến enzym có hai vùng xúc tác họ GH3-GH31 (Mort *et al.*, 2015; Okuma *et al.*, 2016). Gen mã hóa cho enzym này là gen mới, có mức độ tương đồng thấp với các gen đã từng được nghiên cứu và cấu trúc enzym là hoàn toàn mới.

Bản chất kỹ thuật của giải pháp hữu ích

Mục đích của giải pháp hữu ích là đề xuất trình tự gen *bgc* mã hóa enzym có cấu trúc mô đun GH3-FN3-GH31 như được thể hiện trong SEQ ID NO:1. Trình tự gen này có nguồn gốc từ vi khuẩn trong dạ cỏ dê.

Theo mục đích khác, giải pháp hữu ích đề xuất trình tự axit amin như được thể hiện trong SEQ ID NO:2 mã hóa beta-glucosidaza có cấu trúc mô đun GH3-FN3-GH31. Enzim tái tổ hợp có hoạt tính beta-glucosidaza chuyển hóa esculin, 4-nitrophenyl- β -D-glucopyranosit (pNPG, cơ chất đặc hiệu cho enzym beta-glucosidaza) được mã hóa từ trình tự gen nêu trong SEQ ID NO:1.

Mô tả vắn tắt các hình vẽ

Hình 1 thể hiện cấu trúc vùng (mô đun) trên phân tử enzym có trình tự axit amin nêu tại SEQ ID NO:2 được phân tích dựa trên phần mềm ước đoán cấu trúc PFAM (<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>) (A) và cấu trúc bậc 3 của beta-glucosidaza được ước đoán dựa trên trình tự nêu tại SEQ ID NO:2 bằng công cụ Phyre2, phân tích chuyên sâu (intensive). Cả hai phần mềm đều chỉ ra enzym gồm có 3 vùng chức năng là GH3, FN3 và GH31. Trong đó vùng chức năng GH3 bao gồm hai vùng GH3N và GH3C. Tận cùng đầu N có một trình tự kéo dài từ axit amin số 1 đến số 22 là trình tự tín hiệu tiết. Các chữ số trên hình chỉ vị trí gốc axit amin tính từ đầu N tới đầu C của phân tử enzym.

Hình 2 thể hiện điện di đồ beta-glucosidaza được biểu hiện trong chủng *E. coli* Rosetta (A) và beta-glucosidaza tái tổ hợp được tinh chế (B) trên gel polyacrylamit 12,6%, và hình ảnh beta-glucosidaza tái tổ hợp tinh sạch phân cắt esculin tạo màu nâu xung quanh giếng thạch chứa enzym (C), trong đó các chữ viết tắt được thể hiện trên hình vẽ là M: protein chuẩn (Fermentas); TS: protein tổng số; T: protein pha tan; KT: protein pha không tan; (-): đối chứng âm (PBS); (+): đối chứng dương (xenlulaza, Sigma 0,05 U).

Mô tả chi tiết giải pháp hữu ích

Theo một phương án, giải pháp hữu ích đề cập đến trình tự gen *bgc* mã hóa enzym có cấu trúc mô đun GH3-FN3-GH31, có trình tự như nêu trong SEQ ID NO:1.

Trình tự gen *bgc* mã hóa beta-glucosidaza nêu trên được khai thác từ dữ liệu giải trình tự ADN đa hệ gen của các vi khuẩn trong dạ cỏ dê. Dê được sử dụng trong giải pháp hữu ích là 10 con dê gồm 5 dê Cỏ, 5 dê Bách Thảo sống chăn thả tại các vùng núi cao ở Ninh Bình, Quảng Ninh. Mẫu hỗn dịch trong dạ cỏ của 10 con dê sau

khi thu thập được bảo quản lạnh và đưa về phòng thí nghiệm. Mẫu được lọc nhiều lần bằng vải xô để loại bỏ thức ăn và các mảnh thức ăn, sau đó được ly tâm phân đoạn để loại bỏ các sinh vật đơn bào, nấm và thu vi khuẩn. Vi khuẩn này là vi khuẩn tổng số từ mẫu dạ cỏ. Sau đó ADN của các vi khuẩn được tách chiết, giải trình tự trên hệ thống giải trình tự thế hệ mới Illumina 2500 (BGI, Hồng Kông). Tổng kích thước đa hệ gen giải được là 48,66 Gb được lắp ghép bằng công cụ tin sinh học thành các đoạn trình tự dài (contig). Dựa trên các contig, 5.367.270 gen đã được phát hiện. Chức năng của các protein mã hóa từ các khung đọc mở đã được chú giải dựa trên cơ sở dữ liệu CAZy (Cantarel *et al.*, 2009), KEGG (Kanehisa *et al.*, 2008). Bên cạnh đó, nguồn gốc của các gen cũng được ước đoán dựa vào việc so sánh với cơ sở dữ liệu NR (non-redundant) của NCBI kết hợp với phần mềm MEGAN 4.6 (Huson *et al.*, 2007).

Dựa trên việc chú giải chức năng của các gen, từ 5.367.270 gen, các tác giả giải pháp hữu ích đã tìm kiếm được một trình tự mã số GL0036730 (gọi tắt là *bgc*) có kích thước 3780 nucleotit như đã nêu trong SEQ ID NO:1. Phân tích chi tiết khung đọc mở cho thấy 66 nucleotit đầu 5' mã hóa cho trình tự peptit tín hiệu có vai trò trong việc tiết enzym ra ngoài môi trường, 1 mã bộ ba tận cùng đầu 3' không mã hóa axit amin, qui định mã kết thúc và 3711 nucleotit mã hóa cho beta-glucosidaza.

Trình tự nucleotit của gen *bgc* trong SEQ ID NO:1 không tương đồng với bất cứ trình tự gen nào trên NCBI (truy cập ngày 13/05/2022). Toàn bộ trình tự axit amin dọc theo phân tử suy diễn từ gen trong SEQ ID NO:2 không tương đồng với bất cứ trình tự axit amin nào. Một phần trình tự (chiếm 56% kích thước phân tử về phía đầu N) tương đồng cao nhất là 61,07% với protein chứa vùng GH3C từ *Bacteroidales* (mã số MBR4478618.1), tương đồng 61,49% với beta-glucosidaza từ *Bacteroidales* (mã số MBE6253225.1). Vùng tận cùng đầu C có kích thước khoảng 42% phân tử có độ tương đồng cao nhất 44,17% với protein GH31 từ *Bacteroides fragilis* (mã số WP_084819186.1). Việc tìm kiếm trình tự tương đồng với trình tự SEQ ID NO:2 trong dữ liệu các trình tự từ bằng đọc quyền sáng chế và dữ liệu trình tự phân tích đa hệ gen cũng cho thấy trình tự trong SEQ ID NO:1 tương đồng thấp (dưới 56%) và độ bao phủ thấp (dưới 56%) với các trình tự trong cả hai cơ sở dữ liệu này. Điều đó

chứng tỏ, trình tự gen *bgc* trong SEQ ID NO:1 và trình tự axit amin trong SEQ ID NO:2 là mới (Bảng 1).

Bảng 1. Trình tự axit amin của beta-glucosidaza GH3-GN3-GH31 là mới, có độ tương đồng thấp (dưới 65%) so với các trình tự protein/enzym trong ngân hàng gen quốc tế NCBI, trình tự protein từ bằng độc quyền sáng chế quốc tế, và trình tự dữ liệu đa hệ gen quốc tế truy cập ngày 13/5/2022.

Protein	Độ bao phủ	Giá trị E	Độ tương đồng	Mã số
Trình tự của các protein/enzym trong ngân hàng gen quốc tế NCBI				
Protein chứa vùng GH3C từ <i>Bacteroidales</i>	56%	0.0	61.07%	MBR4478618.1
Protein chứa vùng GH3C từ <i>Bacteroidales</i>	56%	0.0	60.81%	MBO4924023.1
Beta-glucosidaza từ <i>Bacteroidales</i>	55%	0.0	61.49%	MBE6253225.1
Beta-glucosidaza từ <i>Bacteroidales</i>	56%	0.0	59.77%	NLZ18893.1
Beta-glucosidaza từ vi sinh vật không nuôi cấy được	56%	0.0	60.21%	AFX98011.1
Protein GH31 từ <i>Bacteroides fragilis</i>	42%	5e-145	44.17%	WP_084819186.1
Protein GH31 từ <i>Parabacteroides pacaensis</i>	42%	7e-145	43.55%	WP_106830034.1
Protein GH31 từ <i>Bacteroides fragilis</i>	42%	1e-144	44.09%	WP_080767459.1
Trình tự của các protein từ các bằng độc quyền sáng chế quốc tế				
Trình tự 380 từ bằng độc quyền US 8101393	55%	0.0	56.64%	AFA67567.1
Trình tự 19 từ bằng độc quyền US 8772010	54%	0.0	47.14%	AJM34198.1
Trình tự 1068 từ bằng độc quyền US 10676751	56%	0.0	46.30%	QPT77886.1
Trình tự 6 từ bằng độc quyền US 8772010	52%	0.0	46.20%	AJM34185.1
Trình tự 14 từ bằng độc quyền US 8847031	52%	0.0	46.46%	AJN18527.1
Trình tự 9149 từ bằng độc quyền US 7090973	42%	7e-146	43.98%	ABJ24181.1
Protein GH3 từ dữ liệu đa hệ gen ruột mới	53%	0.0	51.89%	CCO21731.1
Protein GH3 từ dữ liệu đa hệ gen ruột mới	53%	0.0	51.89%	CCO21842.1

Beta-glucosidaza từ dữ liệu đa hệ gen ruột mồi	56%	0.0	46.68%	KAA6339589.1
Trình tự của các enzym từ dữ liệu đa hệ gen quốc tế				
Beta-glucosidaza B bền nhiệt từ dữ liệu đa hệ gen của đồng ủ	54%	0.0	45.19%	MPL64029.1
Beta-glucosidaza B bền nhiệt từ dữ liệu đa hệ gen của đồng ủ	55%	0.0	46.05%	MNQ28629.1
Beta-glucosidaza B bền nhiệt từ dữ liệu đa hệ gen của ruột mồi	53%	0.0	46.67%	KAA6335917.1
Beta-glucosidaza từ dữ liệu đa hệ gen lỗ thông hơi thủy nhiệt	53%	1e-154	39.70%	VAX25371.1
Alpha-xylosidaza từ dữ liệu đa hệ gen đồng ủ	42%	2e-138	41.31%	MPL89159.1
Alpha-glucosidaza từ dữ liệu đa hệ gen của ruột mồi	42%	3e-131	41.01%	KAA6352008.1
Alpha-glucosidaza từ dữ liệu đa hệ gen của ruột mồi	42%	1e-129	40.26%	KAA6339479.1

Theo một phương án khác, giải pháp hữu ích đề cập đến cấu trúc của trình tự axit amin trong SEQ ID NO:2 suy diễn từ gen *bgc* trong SEQ ID NO:1 có cấu trúc mới. Khảo sát vùng bảo thủ trong BLASTP và cấu trúc mô đun bằng phần mềm ước đoán cấu trúc PFAM (<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>) dựa trên

5 trình tự trong SEQ ID NO:2 cho thấy protein có chứa ba vùng cấu trúc trong đó vùng mô đun GH3N và GH3C hợp thành một vùng xúc tác đặc thù của enzym thuộc họ GH3, tiếp theo là vùng FN3 và cuối cùng là vùng GH31 (Hình 1A). Bằng công cụ phân tích dự đoán cấu trúc chuyên sâu của phần mềm Phyre2, trình tự trong SEQ ID NO:2 có cấu trúc rất điển hình của beta-glucosidaza GH3, trong đó vùng GH3N có

10 cấu trúc điển hình $(\alpha/\beta)_8$, tiếp giáp là vùng GH3C có cấu trúc điển hình $(\alpha/\beta)_6$. Vùng FN3 nằm giữa có cấu trúc dạng phiến gấp nếp beta và cuối cùng là cấu trúc điển hình của GH31 (Hình 1B).

Theo một phương án khác, giải pháp hữu ích đề cập đến protein tái tổ hợp có trình tự axit amin nêu trong SEQ ID NO:2, có hoạt tính của beta-glucosidaza được

15 thể hiện ở khả năng chuyển hóa đặc hiệu cơ chất 4-nitrophenyl- β -D- glucopyranoside (pNPG).

Phương pháp sinh tổng hợp beta-glucosidaza tái tổ hợp theo giải pháp hữu ích được thực hiện trong tế bào *Escherichia coli*. Phần gen *bgc* trong SEQ ID NO:1 không chứa trình tự tín hiệu (được gọi là gen *gh3-fn3-gh31*, gồm 66 nucleotit tận cùng đầu 5' được loại bỏ) được tối ưu mã bộ ba, mã hóa cho phân trình tự axit amin trong SEQ ID NO:2 gồm đầy đủ ba vùng GH3-FN3-GH31 được đưa vào vector mang gen tại vị trí của enzym giới hạn *NcoI* và *XhoI*. Vector dùng để mang gen và chuyển gen *gh3-fn3-gh31* vào *E. coli* là pET22b(+) (Novagen) để tạo plasmid biểu hiện pET22-*gh3fn3gh31*.

Plasmid tái tổ hợp pET22-*gh3fn3gh31* được đưa vào tế bào chất của chủng biểu hiện *E. coli* Rosetta 1 bằng phương pháp sốc nhiệt. Quá trình biểu hiện gen *gh3-fn3-gh31* được thực hiện như sau: chủng vi khuẩn *E. coli* Rosetta 1 mang gen *gh3-fn3-gh31* được nuôi cấy trong môi trường LB có ampicilin (LBA) ở 37°C qua đêm. Sau đó dịch nuôi cấy qua đêm được chuyển sang môi trường TBcb có bổ sung ampicilin mới và tiếp tục nuôi cấy lắc 200 vòng/phút ở 37°C cho đến khi đạt giá trị OD₆₀₀ khoảng 1,5 thì tiến hành cảm ứng với nồng độ IPTG cuối cùng là 0,3 mM. IPTG có mặt trong môi trường sẽ đi vào trong tế bào chất và cảm ứng vùng khởi động *T7lac* để khởi động quá trình phiên mã sinh tổng hợp protein tái tổ hợp. Tế bào được nuôi cấy trong môi trường cảm ứng trên ở 25°C trong 4 giờ thì được thu lại bằng phương pháp ly tâm. Cặn tế bào được hòa lại bằng nước để đạt giá trị OD₆₀₀ cuối cùng là 10. Dung dịch tế bào chứa protein tái tổ hợp được tiến hành siêu âm để thu protein pha tan và pha không tan. Protein tái tổ hợp được kiểm tra bằng phương pháp điện di trên gel polyacrylamit 12,6% (Laemmli, 1970). Trên điện di đồ, protein tái tổ hợp được biểu hiện với kích thước khoảng 136 kDa và tồn tại chủ yếu trong pha tan (Hình 2A).

Dung dịch protein ở pha tan được bổ sung đệm PBS đến nồng độ cuối cùng là 20 mM và imidazol đến nồng độ cuối cùng là 20 mM. Dung dịch protein này sau đó được đưa vào cột HiTrap đã được gắn Ni²⁺. Do protein tái tổ hợp được gắn 6 axit amin histidin nên liên kết chặt với Ni²⁺ được giữ lại trên cột còn các protein khác không có hoặc có ái lực yếu với Ni²⁺ sẽ bị rửa trôi ở các nồng độ imidazol cao hơn (50, 70 mM). Protein tái tổ hợp sau đó được tách ra khỏi cột nhờ đệm có nồng độ imidazol cao 300 mM và được thẩm tách trong đệm PBS 20 mM, pH 7 có 5% glycerol

qua đêm ở 4°C. Protein tái tổ hợp thu được có độ tinh sạch đạt >90% (như được đánh giá bằng QuantityOne) và được bảo quản ở -80°C tại Viện Công nghệ sinh học – Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Protein tái tổ hợp được chứng minh có hoạt tính của beta-glucosidaza, có khả năng phân cắt cơ chất esculin trên đĩa thạch tạo màu nâu đặc trưng trong môi trường có sắt; phân cắt đặc hiệu cơ chất pNPG với hoạt tính đạt khoảng 35 U/mg. Như vậy, trình tự gen trong SEQ ID NO:1 mã hóa cho protein có trình tự trong SEQ ID NO:2 là beta-glucosidaza có cấu trúc mới GH3-FN3F-GH31.

Ví dụ thực hiện giải pháp hữu ích

10 *Ví dụ 1: So sánh tìm mức độ tương đồng của gen bgc, protein với các trình tự trên ngân hàng gen*

Trình tự gen *bgc* trong SEQ ID NO:1, trình tự axit amin trong SEQ ID NO:2 được tìm kiếm các trình tự gen/protein tương đồng từ ngân hàng gen dựa trên phần mềm nucleotit BLAST hoặc protein BLAST trong
15 <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Các database được lựa chọn là dữ liệu của toàn bộ các trình tự có được, dữ liệu trình tự từ các bằng sáng chế, dữ liệu từ giải trình tự metagenome.

Ví dụ 2: Ước đoán cấu trúc protein dựa trên trình tự axit amin

Trình tự axit amin trong SEQ ID NO:2 được chuyển vào phần mềm tìm kiếm
20 vùng chức năng và vị trí của các họ protein dọc trên phân tử (<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>). Cấu trúc bậc ba của phân tử protein được ước đoán dựa trên phần mềm SwissProt (<https://swissmodel.expasy.org/interactive>) và Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) ở chế độ phân tích chuyên sâu (intensive).

25 *Ví dụ 3: Biểu hiện gen gh3-fn3-gh31 trong tế bào E. coli Rosetta 1*

Nuôi cấy một khuẩn lạc *E. coli* Rosetta 1 mang plasmit pET22-*gh3-fn3-gh31* trong 2 ml môi trường LB (0,5% cao nấm men, 1% trypton, 0,5% NaCl) có bổ sung 100 µg/ml ampicillin (LBA) lắc 200 vòng/phút ở 37°C qua đêm. Chuyển 200 µl dịch

nuôi cấy qua đêm sang 50 ml môi trường TBcb (2,4% cao nấm men, 1,2% pepton, 0,24% D-glucoza, 17 mM KH_2PO_4 , 72 mM K_2HPO_4) có bổ sung 100 $\mu\text{g/ml}$ ampicilin (gọi chung là môi trường TBcb-Amp), lắc tiếp trong 2 giờ ở 37°C cho đến khi OD_{600} đạt khoảng 1.5 thì tiến hành cảm ứng với 0,3 mM IPTG, tế bào chuyển sang nuôi cấy ở nhiệt độ cảm ứng 25°C, lắc 200 vòng/phút trong 4 giờ. Tế bào sau nuôi cấy được thu lại bằng cách ly tâm ở 5000 vòng/phút trong 5 phút. Cặn tế bào được hòa lại vào nước để đạt OD_{600} cuối cùng bằng 10. Dung dịch tế bào được tiến hành siêu âm lặp lại 100-200 lần, mỗi lần phóng sóng siêu âm trong khoảng 0,5 giây, cách nhau khoảng 0,5 giây, trong điều kiện nhiệt độ 4°C. Dịch tế bào sau siêu âm được coi là mẫu protein tổng số. Dịch này được ly tâm ở tốc độ 8000 vòng/phút trong 10 phút để tách riêng hai pha protein tan và không tan. Mẫu protein tổng số, protein tan đều được bổ sung dung dịch đệm mẫu 6X (tris 0,375 M, pH8, glycerol 60%, SDS 12%, bromophenol blue 0,06%, 2-mecaptoethanol 10%) theo tỷ lệ thể tích 5 mẫu:1 đệm, trộn đều, biến tính ở 100°C trong 5 phút. Mẫu protein được kiểm tra bằng điện di trên gel polyacrylamit 12,6% (Hình 2A). Beta-glucosidaza tái tổ hợp có kích thước khoảng 136 kDa đã được biểu hiện ở pha protein tan.

Ví dụ 4: Tinh chế beta-glucosidaza tái tổ hợp bằng sắc ký ái lực

Dịch protein tan được bổ sung đệm PBS 30x (NaCl 4 M, KCl 80 mM, Na_2HPO_4 300 mM, KH_2PO_4 53 mM) và imidazol đến nồng độ cuối cùng của PBS1x và imidazol đến nồng độ cuối cùng là 20 mM, sau đó đưa lên cột sắc ký ái lực HisTrap 5 ml đã được gắn Ni^{2+} . Dung dịch beta-glucosidaza được biểu hiện có chứa hexahistidin ở tận cùng đầu C có ái lực cao với niken trên cột và được giữ lại trên cột. Các protein tạp không có hoặc có ái lực yếu với niken sẽ bị đẩy ra khỏi cột bằng đệm rửa PBS 1x có imidazol 50 mM và 70 mM. Protein sau đó được thu lại ở dung dịch đệm chứa nồng độ imidazol 300 mM và được thẩm tách trong đệm PBS 20 mM, pH7 có 5% glycerol qua đêm ở 4°C. Protein sau khi thu được từ các phân đoạn được định lượng bằng dung dịch Bradford và kiểm tra bằng điện di trên gel polyacrylamit 12,6% (Hình 2B). Gel điện di sau khi rửa sạch được scan bằng máy scanner và đưa vào phần mềm QuantityOne để phân tích độ sạch, cho thấy có độ sạch đạt trên 90%.

Ví dụ 5: Đánh giá hoạt tính beta-glucoxidaza trên cơ chất đặc hiệu

Cơ chất esculin

Beta-glucoxidaza có khả năng phân cắt esculin thành glucoza và esculetin. Trong môi trường có sắt, ion sắt sẽ bị esculetin khử tạo màu nâu hoặc đen. Thí nghiệm kiểm tra khả năng thủy phân esculin của protein tái tổ hợp được thực hiện dựa theo phương pháp của Veena và đồng tác giả (Veena *et al.*, 2011). Đĩa thạch môi trường LBA có bổ sung thêm esculin (3 g/l) và ferric ammonium xitrat (0,2 g/l) được chuẩn bị sẵn. Sau đó, dùng dụng cụ tạo các giếng có kích thước khoảng 5 mm trên đĩa thạch. Hút 50 µl protein sau tinh chế và thẩm tích loại muối cho vào mỗi giếng. Sau đó, đĩa được ủ ở 37°C trong 16-20 giờ. Đối chứng âm là 50 µl PBS, pH7 và đối chứng dương là xenlulaza (Sigma) 0,05 U.

Cơ chất pNPG

Hoạt độ beta-glucoxidaza tái tổ hợp được xác định theo phương pháp của Fang và đồng tác giả (Fang *et al.*, 2014). Để xác định hoạt độ beta-glucoxidaza trên cơ chất pNPG, phản ứng giữa enzym và cơ chất được thực hiện trong tổng thể tích 200 µl bao gồm 20 µl enzym (90 ng) trong đệm PBS pH5,5; 2 µl triton X-100, 45 µl pNPG 20 mM, 133 µl PBS 1x, pH5,5 được ủ ở 40°C trong 15 phút. Sau đó phản ứng được dừng bằng cách bổ sung 800 µl Na₂CO₃ 0,2M. Màu sau phản ứng được đo ở bước sóng 410 nm. Một đơn vị hoạt tính beta-glucoxidaza được định nghĩa là lượng enzym cần thiết để giải phóng 1 µmol pNPG trong 1 phút ở điều kiện thí nghiệm. Mỗi thí nghiệm được lặp lại 3 lần. Ống đối chứng có thành phần tương tự ống phản ứng, chỉ thay 20 µl enzym bằng 20 µl PBS 1x, pH5,5.

Hiệu quả đạt được của giải pháp hữu ích

Giải pháp hữu ích nhằm thu được trình tự gen *bgc* mã hóa beta-glucoxidaza từ các vi khuẩn trong dạ cỏ dê mã hóa cho enzym beta-glucoxidaza có cấu trúc mới gồm 2 vùng xúc tác GH3 và GH31 và vùng FN3 nối giữa hai vùng xúc tác. Enzim có hoạt tính phân cắt đặc hiệu cơ chất pNPG với hoạt độ cao, đạt trung bình 35 U/mg.

Các tài liệu tham khảo

- ALVAREZ-GONZALEZ, C., SANTOS, V.E., LADERO, M. & BOLIVAR, J.M. (2022) Immobilization-stabilization of β -glucosidase for implementation of intensified hydrolysis of cellobiose in continuous flow reactors. *Catalysts* **12**, 80. Multidisciplinary Digital Publishing Institute.
- 5 CANTAREL, B.L., COUTINHO, P.M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. (2009) The Carbon hydrat-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**, D233–D238.
- 10 CHEN, A., WANG, D., JI, R., LI, J., GU, S., TANG, R. & JI, C. (2021) Structural and catalytic characterization of TsBGL, a β -glucosidase from *Thermofilum* sp. ex4484_79. *Frontiers in Microbiology* **12**, 723678.
- 15 DAI, X., ZHU, Y., LUO, Y., SONG, L., LIU, D., LIU, L., CHEN, F., WANG, M., LI, J., ZENG, X., DONG, Z., HU, S., LI, L., XU, J., HUANG, L., ET AL. (2012) Metagenomic insights into the fibrolytic microbiome in Yak rumen. *PLoS ONE* **7**, e40430.
- 20 DAO, T.K., DO, T.H., LE, N.G., NGUYEN, H.D., NGUYEN, T.Q., LE, T.T.H. & TRUONG, N.H. (2021) Understanding the role of *Prevotella* genus in the digestion of lignocellulose and other substrates in Vietnamese native goats' rumen by metagenomic deep sequencing. *Animals: an open access journal from MDPI* **11**, 3257.
- FANG, S., CHANG, J., LEE, Y.S., GUO, W., CHOI, Y.L. & ZHOU, Y. (2014) Cloning and characterization of a new broadspecific β -glucosidase from *Lactococcus* sp. FSJ4. *World Journal of Microbiology & Biotechnology* **30**, 213–223.
- 25 FERREIRA, R. DA G., AZZONI, A.R. & FREITAS, S. (2018) Techno-economic analysis of the industrial production of a low-cost enzyme using *E. coli*: the case of recombinant β -glucosidase. *Biotechnology for Biofuels* **11**, 81.
- HENRISSAT, B. & DAVIES, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* **7**, 637–644.
- 30 HUSON, D.H., AUCH, A.F., QI, J. & SCHUSTER, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Research* **17**, 377–386.
- 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 research* **36**, D480-484.
- 35 KETUDAT CAIRNS, J.R. & ESEN, A. (2010) β -Glucosidases. *Cellular and Molecular Life Sciences* **67**, 3389–3405.

- LAEMMLI, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- LI, S., YANG, X., YANG, S., ZHU, M. & WANG, X. (2012) Technology prospecting on enzymes: application, marketing and engineering. *Computational and Structural Biotechnology Journal* **2**, e201209017.
- LIU, C., SUN, D., LIU, J., ZHU, J. & LIU, W. (2021) Recent advances and perspectives in efforts to reduce the production and application cost of microbial flocculants. *Bioresources and Bioprocessing* **8**, 51.
- MOHSIN, I., POUDEL, N., LI, D.-C. & PAPAGEORGIOU, A.C. (2019) Crystal structure of a GH3 β -glucosidase from the thermophilic fungus *Chaetomium thermophilum*. *International Journal of Molecular Sciences* **20**, 5962.
- MORALES, M., ARVESEN, A. & CHERUBINI, F. (2021) Integrated process simulation for bioethanol production: Effects of varying lignocellulosic feedstocks on technical performance. *Bioresource Technology* **328**, 124833.
- MORT, A., RAY, A., SAYKHEDKAR, S.S. & PRADE, R.A. (2015) Systems and methods for production and use of fungal glycosyl hydrolases. <https://patents.google.com/patent/WO2015187697A2/en> [accessed 29 April 2022].
- NGUYEN, K.H.V., DAO, T.K., NGUYEN, H.D., NGUYEN, K.H., NGUYEN, T.Q., NGUYEN, T.T., NGUYEN, T.M.P., TRUONG, N.H. & DO, T.H. (2021) Some characters of bacterial cellulases in goats' rumen elucidated by metagenomic DNA analysis and the role of fibronectin 3 module for endoglucanase function. *Animal Bioscience* **34**, 867–879.
- OKUMA, J., SUDA, M., YAMAGUCHI, A., HIROSE, Y., KONDO, Y., SATO, M. & SHIBATA, D. (2016) Thermostable beta-glucosidase. <https://patents.google.com/patent/US20160168553A1/en> [accessed 29 April 2022].
- PATEL, A. & SHAH, A.R. (2021) Integrated lignocellulosic biorefinery: Gateway for production of second generation ethanol and value added products. *Journal of Bioresources and Bioproducts* **6**, 108–128.
- SINGH, G., VERMA, A.K. & KUMAR, V. (2016) Catalytic properties, functional attributes and industrial applications of β -glucosidases. *3 Biotech* **6**, 3.
- SINGHANIA, R.R., PATEL, A.K., SUKUMARAN, R.K., LARROCHE, C. & PANDEY, A. (2013) Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresource Technology* **127**, 500–507.
- SRIVASTAVA, N., RATHOUR, R., JHA, S., PANDEY, K., SRIVASTAVA, M., THAKUR, V.K., SENGAR, R.S., GUPTA, V.K., MAZUMDER, P.B., KHAN, A.F. & MISHRA,

- P.K. (2019) Microbial beta glucosidase enzymes: Recent advances in biomass conversation for biofuels application. *Biomolecules* **9**, E220.
- 5 SU, T., ZHAO, D., KHODADADI, M. & LEN, C. (2020) Lignocellulosic biomass for bioethanol: Recent advances, technology trends, and barriers to industrial development. *Current Opinion in Green and Sustainable Chemistry* **24**, 56–60.
- SUWAN, E., ARTHORNTHURASUK, S. & KONGSAEREE, P.T. (2017) A metagenomic approach to discover a novel β -glucosidase from bovine rumens. *Pure and Applied Chemistry* **89**, 941–950. De Gruyter.
- 10 VARGHESE, J.N., HRMOVA, M. & FINCHER, G.B. (1999) Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure (London, England: 1993)* **7**, 179–190.
- VEENA, V., POORNIMA, P., PARVATHAM, R., SIVAPRIYADHARSINI & KALALSELVI, K. (2011) Isolation and characterization of beta-glucosidsase producing bacteria from different sources. *African Journal of Biotechnology* **10**, 14907–14912.
- 15 ZHANG, Y., MIN, Z., QIN, Y., YE, D.-Q., SONG, Y.-Y. & LIU, Y.-L. (2019) Efficient display of *Aspergillus niger* β -glucosidase on *Saccharomyces cerevisiae* cell wall for aroma enhancement in wine. *Journal of Agricultural and Food Chemistry* **67**, 5169–5176. American Chemical Society.

YÊU CẦU BẢO HỘ

1. Trình tự gen *bgc* mã hóa beta-glucosidaza, có nguồn gốc từ vi khuẩn trong dạ cỏ dê, có trình tự nêu trong SEQ ID NO:1.
- 5 2. → Cấu trúc beta-glucosidaza GH3-FN3-GH31 có trình tự axit amin nêu trong SEQ ID NO:2, trong đó protein này được mã hóa bởi trình tự gen nêu trong điểm 1 có hoạt tính cao 35 U/mg trên cơ chất đặc hiệu pNPG.

DANH MỤC TRÌNH TỰ

<110> Viện Công nghệ sinh học

<120> Trình tự gen *bgc* mã hóa beta-glucosidaza có nguồn gốc từ vi khuẩn trong dạ cỏ dê và cấu trúc beta-glucosidaza GH3-FN3-GH31 tái tổ hợp có hoạt độ cao.

5 <160> 2

<210> 1

<211> 3780

<212> ADN

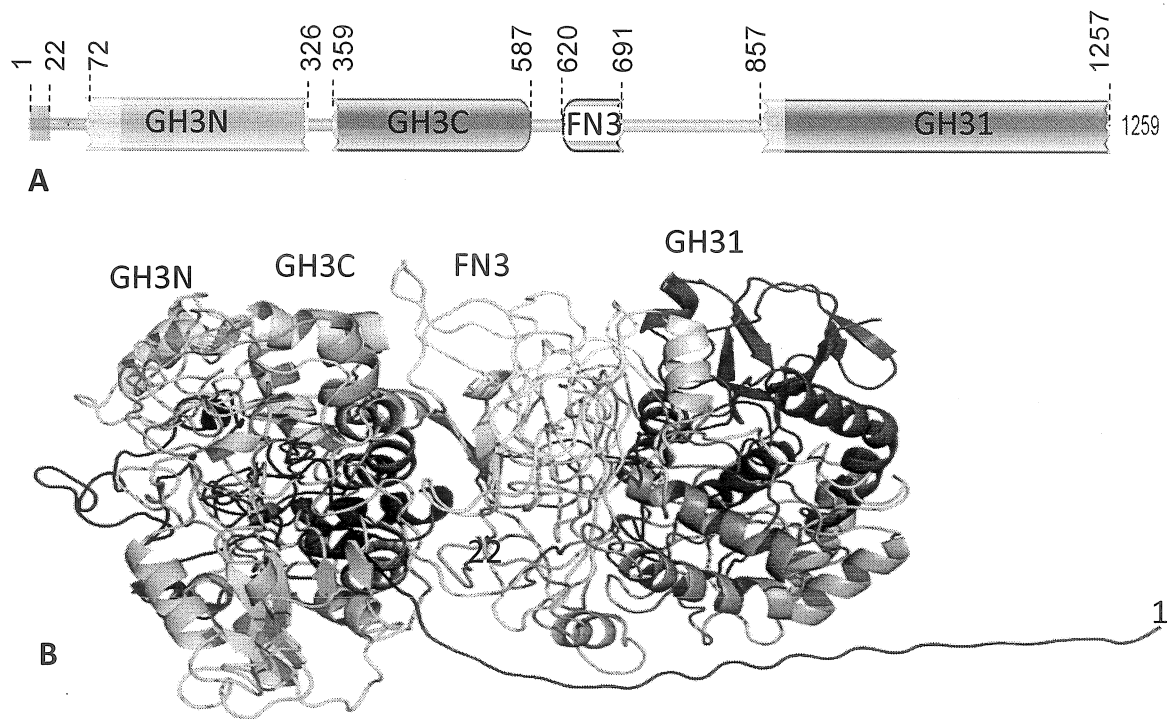
10 <213> Vi khuẩn trong dạ cỏ dê

<400> 1

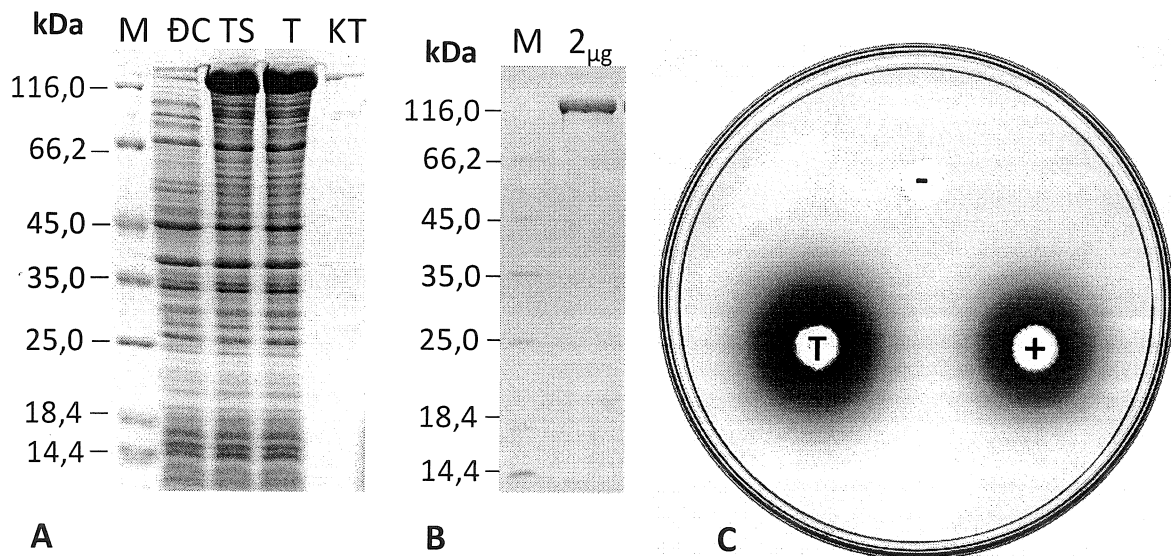
15	1	ATGCATCTGATACGTCATTTTCATCACCCGTGTTCCCTTACCGTCATGCTTTCCTTCGGCGCA	60
	61	GCGGCGCAGCCACAGTTGCGGGAAGACAATATTGATGAGGTCCTTGCCGCCATGACCCCTT	120
	121	CAGGAGAAGGCAACCCTGCTGGTGGGCGGCAGCCGCGCAATCGTTTTTCGAAGGCGTGAAA	180
	181	ACCGGCTATGCCTGGCAGGTGCCCGGTGCTGCGGCAGTACCCGTCCCATAGAGCGCCTG	240
	241	GGCATTCGGTCAACATCCTGGCCGACGCCCCGCGGCCTTCGCATCGCCCCCTACCCGA	300
	301	AAAGATACGGACCAGACCTTCTGGACCACCGGTTTCCCTGCGGAACGCTGCTTGCCAGT	360
20	361	TCCTGGGATACCGCCCTTTTGGAGGAGTGGGCGCCGCTATAGGCGAAGAAGTTCATGAA	420
	421	TACGGGGTGGACGTGCTTCTCGCACCTGGTGTGAACATCCACCGCAATCCCCTCTGCGGC	480
	481	CGCAATTTTCGAGTATTTCTCTGAAGATCCGCTGCTGAGCGGCAGCATGGCCACGGCCTAT	540
	541	ATCAACGGCGTACAAAGCAATGGGGTGGGAGTATCCATCAAACATTTTGCAGCCAACAAC	600
	601	CAGGAGACCGCTCGCGTATGGGGCGATTCCCGGGTAAGCGAAAGGGCGCTACGGGAAATC	660
25	661	TACCTCAAGGCCTTCGAGATGGCGGTGCGGGACGCGCACCCCTGGACGGTAATGTCTGCTG	720
	721	TACAACAAGATAAAACGGCGAATATGCCCAGCAAAACAGGGACCTGCTCACCACCGTGCTG	780
	781	CGCGAAGATGGGGTTTCGACGGCATCGTGATGACCGACTGGGGTAAAAAGGCCGGAACG	840
	841	GTGAAGGCAGTGACGGCCGGGAACGACCTTATGGAACCCGGATACGATGAAGAGATAGAC	900
	901	CGTATCGTAAAGGCGGTGGAAGACGGCACGCTTAACATGGCCGATGTGGACCGCAACGTC	960
30	961	CGCAACATGCTCCGGTACATCGTAAGAACCCCGTCTTTCAAGCACTACGCCTTTTCAAC	1020
	1021	AATCCCGACCTTGAGGCCACGCGCGCCCTCGCCCCGAAAGGCAGCAGCAGAGTCGTGTGTA	1080
	1081	CTCCTGAAGAATGAAGGGGAACTTTGCCGCTGGCGGGAACGGAAACAGTGGCCCTGTAT	1140
	1141	AACATATCTTCCGTCGATTTCTGTCGAGGTGGCACAGGAAGCGGCAGCATCAGCAAAATCC	1200
	1201	CGTTTGGTAGACCTCCGCACCGGCCTGGAAAAAGCCGGATTACCGTGAACGGGAGCCAG	1260
35	1261	GACGCGGAGGTGGCGGTAGTGGTCTTGGCCGCGAGCGCCGCGAGGCCCCGCGACCGCAGC	1320
	1321	GAAGAGAAGGATTTCTCCTGAGCGAAGAGGAGCACCGGCAGCTGGAAAATATCAGCGGG	1380
	1381	GCTTTCCACGGCCAAGGCAAAAAGTAGTAGTGTGCTGAACATCTGCGGGGTGGTGGAA	1440
	1441	ACGGCCTCCTGGAAGCATCTGGCCGATGCCATCCTTCTTCCCTGGGCACCCGGGCAGGAA	1500
	1501	GGGGGCCATGCCGTGGCCGATGTCTTGCCGGCAGGGTCAGTCCTTCAGGCAGGCTCACC	1560
40	1561	ATGACCTTCCCGGTCTCCTATTCCGACGACCTTCTGCCTCCAACCTCCCGGCCACACG	1620
	1621	CAGGAATATGTCAACAAGAGCGACAGCCCTCGCAAGGACATTGAATATGTAAACTATGAA	1680
	1681	GAAGGCATTTTGTGCGTTACCGCCACTATGTAACAGAGGGAAAAGAGGTTTCTATCCG	1740
	1741	TTCGGCTACGGCCTGGGATACACTTCTTTCGCATGGTCCAGGCCTTCGGTAAAGAAAACC	1800
	1801	GCCGACGGTTTCGAAGCGCGATTACCGTTACCAATAAGGGTGACAGGTGCGGCCGCGAT	1860
45	1861	GTGGTGCAGCTTTATGTTTCCGCCCCGGCAGGTGGCATGGAAAAGCCGCTTCGGGAGCTG	1920
	1921	AAAGGCTTTGCCAAAACACGTGAACCTGAATCCAGGAGAGAGCGAAACGGTGACATACGG	1980
	1981	GTGAATACCTACGCCCTGGCATCGTACAATGAAGGTACATCCTGCTGGGAAAACAGCTGCA	2040
	2041	GGGGTTTATAAGCTGCAATTTCGGAGCCAATGCCCGGATATCCGTGCCGGAGCGGGCTGG	2100
	2101	AAGCAGAAGAAGGCTGCCTCATGGAAAGCCAACAAAGCCTTCTCCCCGAAGACGGCCCTG	2160
50	2161	CAGCCGGCCGCGAGCCCTGCAGCGCAGGAGCATGTTACCGTTATCCAGCCGCTGGAAGGA	2220
	2221	GAATACTGGTGGGGCGCTGTGGTGAACAAGGGCTACGACCAGCCGTTACCAATTTTCGAA	2280
	2281	GCATGCGACAACCTTCTATCTGGACTGGCCCCATCCCAAGGGACCTCGGGACGGACGTCCC	2340
	2341	TTGACCTGGGAACGATGAGCGCAAGGGCTACTACCGTGCCGCTGCTGCTCTCCAGCAAA	2400
	2401	GACGCTATGTCTGGGGCAACAGGGCTTTTCTTGTCTTTAAGGATGGGGTTCTCACG	2460
55	2461	GTAACCTCAAAGTATGAGAAGTTGGAGCCGGTAGCGGCGGGAACCACTCTCAAAGAAGCC	2520
	2521	TACAAGGCCGCTGCGCGGCGCATTTCTTCTTCGACGGGAGAGAACCGGCTGAACTCATG	2580
	2581	TTACCATGCCCCAGTTCAACAACCTGGATAGAGACCGTGGTGTGGGCATCAACCAGGAG	2640
	2641	AACGCCGAGACCTTTGTGGACAAGCTTCACGAAAGCGGCTTCCCTGCGGGGTGGTGATG	2700
	2701	ATAGACGGCGGCTGGCAGCGTTACCACGGCTGCCGGGACTTCAATCCCGACACTTTCCC	2760
60	2761	GAACCGCAGCGTCTTTTGGACAAGATACAGACTACGGCTACAGGGCGTGCTTTGGATT	2820
	2821	GCATACTTCTTTTCGCCCCAGCGCGCCGGAATACACCAGCTACAAGCCCGGAATGCAG	2880
	2881	GACATCCTGGTGACAGCAAGCACAGCCAGAAGGACCCGCACTGGTATGGTGGTGGAAAC	2940
	2941	GGCATCAGCGTGACCATGGACCTTACCCGTGCTGACATTTCGGGAAAAGTTTATCGGGGAG	3000

3001	CTGAAGGATATGGCGCAGCGTTTCCACTTCGACGGCTACAAGTTCGACGGAGGTGATCCG	3060
3061	GAATACTTCCGTGGCGACGCCGCATTAGCGAGCCCTGGATGGAAGAGGCCGACTTCGGA	3120
3121	AAGGCCTGGCAGCAGGTGGGGGAAGCTTTCCCATATAATGAATACCGTGCCGGCTATAAC	3180
3181	AACGGCGGCACGCCCTGGTGCTGCGCCACTACGATGTGGGCCACACCTGGACCGAATTC	3240
5 3241	CAGGCCGTGATACCAACCTTACCACCGCCGGCTGCTGGGCTATCCCTACGTTTTTGCC	3300
3301	GACATGATAGGTGGGGCCTCTCGGCAGCTATATGCCCGAAAAGACTTTCTCTCACAAAG	3360
3361	CTGTTCCCTTCGCTCCTGCGAGGCCAGGCCCTTATGCCCATGATGCAGTTCAGCGCAGCC	3420
3421	CCATGGCGCGTGCTCAGCCCCGAAGAATGCGAGGTGTGCCGCCGCTATGCGGAACCTCAC	3480
3481	GTACAGTTCGGCCCCCTATATCATGGAGCAGGTACACCATGCCTCCGCCACCGGAGAACCC	3540
10 3541	ATCATGCGCAGCATGGAATACGAGTCCCCGGATGCGGATACGAGAAGGTGGACCAGCAG	3600
3601	TTCATGCTTGGGCCGCGATATCTGGTGGCTCCCGTTACCAGCGAAGACGACTCAAAGAGC	3660
3661	GTGTATCTTCCCGCCGGCCGCTGGCGCGACGACCAGGGCAAGGTGTGGAGAGGTCCTAAG	3720
3721	GTTCTGGAACCTCAAGGATATCCCTGTAGAAAGACTCCCTTATTACGAAAAGCTCAAATAA	3780
15		
	<210> 2	
	<211> 259	
	<212> axit amin	
	<213> Trình tự suy diễn từ gen	
20	<400> 1	
	1 MetHisLeuIleArgHisPheIleThrLeuPheLeuThrValMetLeuSerLeuGlyAla	20
	21 AlaAlaGlnProGlnLeuArgGluAspAsnIleAspGluValLeuAlaAlaMetThrLeu	40
	41 GlnGluLysAlaThrLeuLeuValGlyGlySerArgAlaIleValPheGluGlyValLys	60
	61 ThrGlyTyrAlaTrpGlnValProGlyAlaCysGlySerThrArgProIleGluArgLeu	80
25	81 GlyIleProValThrIleLeuAlaAspGlyProAlaGlyLeuArgIleAlaProThrArg	100
	101 LysAspThrAspGlnThrPheTrpThrThrGlyPheProCysGlyThrLeuLeuAlaSer	120
	121 SerTrpAspThrAlaLeuLeuGluGluValGlyAlaAlaIleGlyGluGluValHisGlu	140
	141 TyrGlyValAspValLeuLeuAlaProGlyValAsnIleHisArgAsnProLeuCysGly	160
	161 ArgAsnPheGluTyrPheSerGluAspProLeuLeuSerGlySerMetAlaThrAlaTyr	180
30	181 IleAsnGlyValGlnSerAsnGlyValGlyValSerIleLysHisPheAlaAlaAsnAsn	200
	201 GlnGluThrAlaArgValTrpGlyAspSerArgValSerGluArgAlaLeuArgGluIle	220
	221 TyrLeuLysAlaPheGluMetAlaValArgAspAlaAlaProTrpThrValMetSerSer	240
	241 TyrAsnLysIleAsnGlyGluTyrAlaGlnGlnAsnArgAspLeuLeuThrThrValLeu	260
	261 ArgGluGluTrpGlyPheAspGlyIleValMetThrAspTrpGlyLysLysAlaGlyThr	280
35	281 ValLysAlaValGlnAlaGlyAsnAspLeuMetGluProGlyTyrAspGluGluIleAsp	300
	301 ArgIleValLysAlaValGluAspGlyThrLeuAsnMetAlaAspValAspArgAsnVal	320
	321 ArgAsnMetLeuArgTyrIleValArgThrProSerPheLysHisTyrAlaPhePheAsn	340
	341 AsnProAspLeuGluAlaHisAlaAlaLeuAlaArgLysAlaAlaAlaGluSerCysVal	360
	361 LeuLeuLysAsnGluGlyGlyThrLeuProLeuAlaGlyThrGluThrValAlaLeuTyr	380
40	381 AsnIleSerSerValAspPheValAlaGlyGlyThrGlySerGlySerIleSerLysSer	400
	401 ArgLeuValAspLeuArgThrGlyLeuGluLysAlaGlyPheThrValAsnGlySerGln	420
	421 AspAlaGluValAlaValValValLeuGlyArgSerAlaGlyGluAlaArgAspArgSer	440
	441 GluGluLysAspPheLeuLeuSerGluGluGluHisArgGlnLeuGluAsnIleSerGly	460
	461 AlaPheHisGlyGlnGlyLysLysValValValLeuLeuAsnIleCysGlyValValGlu	480
45	481 ThrAlaSerTrpLysHisLeuAlaAspAlaIleLeuLeuProTrpAlaProGlyGlnGlu	500
	501 GlyGlyHisAlaValAlaAspValLeuAlaGlyArgValSerProSerGlyArgLeuThr	520
	521 MetThrPheProValSerTyrSerAspAspProSerAlaSerAsnPheProGlyHisThr	540
	541 GlnGluTyrValAsnLysSerAspSerProArgLysAspIleGluTyrValAsnTyrGlu	560
	561 GluGlyIlePheValGlyTyrArgHisTyrValThrGluGlyLysGluValSerTyrPro	580
50	581 PheGlyTyrGlyLeuGlyTyrThrSerPheAlaTrpSerArgProSerValLysLysThr	600
	601 AlaAspGlyPheGluAlaArgIleThrValThrAsnLysGlyAspArgSerGlyArgAsp	620
	621 ValValGlnLeuTyrValSerAlaProAlaGlyGlyMetGluLysProLeuArgGluLeu	640
	641 LysGlyPheAlaLysThrArgGluLeuAsnProGlyGluSerGluThrValHisIleArg	660
	661 ValAsnThrTyrAlaLeuAlaSerTyrAsnGluGlyThrSerCysTrpGluThrAlaAla	680
55	681 GlyValTyrLysLeuGlnPheGlyAlaAsnAlaArgAspIleArgAlaGlyAlaGlyTrp	700
	701 LysGlnLysLysAlaAlaSerTrpLysAlaAsnLysAlaPheSerProLysThrAlaLeu	720
	721 GlnProAlaAlaSerProAlaAlaGlnGluHisValThrValIleGlnProLeuGluGly	740
	741 GluTyrTrpTrpGlyAlaValValAsnLysGlyTyrAspGlnProPheThrAsnPheGlu	760
	761 AlaCysAspAsnPheTyrLeuAspTrpProHisProLysGlyProArgAspGlyArgPro	780
60	781 PheAspLeuGlyThrMetSerAlaLysGlyThrThrValProLeuLeuLeuSerSerLys	800
	801 GlyArgTyrValTrpGlyAsnArgProPheSerPheValPheLysAspGlyValLeuThr	820
	821 ValThrSerLysTyrGluLysLeuGluProValAlaAlaGlyThrThrLeuLysGluAla	840
	841 TyrLysAlaAlaCysAlaAlaHisPheLeuPheAspGlyArgGluProAlaGluLeuMet	860
	861 PheThrMetProGlnPheAsnAsnTrpIleGluThrValValLeuGlyIleAsnGlnGlu	880

	881	AsnAlaGluThrPheValAspLysLeuHisGluSerGlyPheProCysGlyValValMet	900
	901	IleAspGlyGlyTrpGlnArgTyrHisGlyCysArgAspPheAsnProAspThrPhePro	920
	921	GluProGlnArgLeuPheAspLysIleHisAspTyrGlyTyrLysGlyValLeuTrpIle	940
	941	AlaTyrPheLeuSerProAspSerArgProGluTyrThrSerTyrLysProGlyMetGln	960
5	961	AspIleLeuValHisSerLysHisSerGlnLysAspProAlaLeuValTrpTrpTrpAsn	980
	981	GlyIleSerValThrMetAspLeuThrArgAlaAspIleArgGluLysPheIleGlyGlu	1000
	1001	LeuLysAspMetAlaGlnArgPheHisPheAspGlyTyrLysPheAspGlyGlyAspPro	1020
	1021	GluTyrPheArgGlyAspAlaAlaPheSerGluProTrpMetGluGluAlaAspPheGly	1040
	1041	LysAlaTrpGlnGlnValGlyGluAlaPheProTyrAsnGluTyrArgAlaGlyTyrAsn	1060
10	1061	AsnGlyGlyThrProLeuValLeuArgHisTyrAspValGlyHisThrTrpThrGluPhe	1080
	1081	GlnAlaValIleProAsnLeuThrThrAlaGlyLeuLeuGlyTyrProTyrValPheAla	1100
	1101	AspMetIleGlyGlyGlyLeuSerGlySerTyrMetProGlyLysThrPheSerHisLys	1120
	1121	LeuPheLeuArgSerCysGluAlaGlnAlaLeuMetProMetMetGlnPheSerAlaAla	1140
	1141	ProTrpArgValLeuThrProGluGluCysGluValCysArgArgTyrAlaGluLeuHis	1160
15	1161	ValGlnPheGlyProTyrIleMetGluGlnValHisHisAlaSerAlaThrGlyGluPro	1180
	1181	IleMetArgSerMetGluTyrGluPheProGlyCysGlyTyrGluLysValAspGlnGln	1200
	1201	PheMetLeuGlyProArgTyrLeuValAlaProValThrSerGluAspAspSerLysSer	1220
	1221	ValTyrLeuProAlaGlyArgTrpArgAspAspGlnGlyLysValTrpArgGlyProLys	1240
20	1241	ValLeuGluLeuLysAspIleProValGluArgLeuProTyrTyrGluLysLeuLysOc*	1259



Hình 1



Hình 2