

BỘ GIÁO DỤC VÀ ĐÀO TẠO **VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM**
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Nguyễn Minh Chiến

**NGHIÊN CỨU ĐẶC ĐIỂM BETA-GLUCOSIDASE
DỰA TRÊN PHÂN TÍCH DỮ LIỆU DNA METAGENOME CỦA VI
KHUẨN TRONG DẠ CỎ DÊ**

LUẬN VĂN THẠC SĨ SINH HỌC

Hà Nội - 2024

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



Nguyễn Minh Chiến

**NGHIÊN CỨU ĐẶC ĐIỂM BETA-GLUCOSIDASE
DỰA TRÊN PHÂN TÍCH DỮ LIỆU DNA METAGENOME CỦA VI
KHUẨN TRONG DẠ CỎ DÊ**

LUẬN VĂN THẠC SĨ SINH HỌC

Ngành: Sinh học thực nghiệm

Mã số: 8 42 01 14

NGƯỜI HƯỚNG DẪN KHOA HỌC:

PGS. TS Đỗ Thị Huyền

A handwritten signature in blue ink, appearing to be 'Duy Hoang', written over a horizontal line.

Hà Nội - 2024

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài nghiên cứu trong luận văn này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu dưới sự hướng dẫn của PGS. TS Đỗ Thị Huyền. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Đồng thời, kết quả này chưa từng xuất hiện trong bất cứ một nghiên cứu nào. Các số liệu, kết quả nêu trong luận văn là trung thực nếu sai tôi hoàn toàn chịu trách nhiệm trước pháp luật.

Tác giả luận văn



Nguyễn Minh Chiến

LỜI CẢM ƠN

Đầu tiên, em xin bày tỏ lòng biết ơn chân thành nhất đến giáo viên hướng dẫn của em là PGS. TS Đỗ Thị Huyền – người cô giàu kinh nghiệm và đầy nhiệt huyết đã hỗ trợ, định hướng, tận tình chỉ bảo để em có thể hoàn thành luận án trong những điều kiện tốt nhất.

Em xin gửi lời cảm ơn đến TS. Nguyễn Thị Quý cùng các anh, chị Phòng Kỹ thuật di truyền, Viện Công nghệ sinh học, Viện Hàn lâm khoa học và Công nghệ Việt Nam đã giúp đỡ tạo điều kiện thuận lợi để em có thể hoàn thành thực nghiệm của đề tài. Luận văn được thực hiện dựa trên nguồn kinh phí của Đề tài NCVCC08.05/24-25 do PGS. TS Đỗ Thị Huyền làm chủ nhiệm. Em xin cảm ơn về sự hỗ trợ đó.

Em xin chân thành cảm ơn ban Lãnh đạo, phòng Đào tạo và các Phòng chức năng của Học viện Khoa học và Công nghệ cùng toàn thể các thầy cô giáo đã truyền đạt cho em các kiến thức, kinh nghiệm, hỗ trợ em hoàn thành các học phần và mọi thủ tục cần thiết khác để em hoàn thành việc học tập của mình.

Cuối cùng, em xin bày tỏ lời cảm ơn sâu sắc đến gia đình, bạn bè, đồng nghiệp đã luôn chia sẻ, động viên tinh thần và là nguồn cổ vũ, giúp đỡ em vượt qua mọi khó khăn trong suốt quá trình thực hiện luận án.

Tác giả luận văn

Nguyễn Minh Chiến

MỤC LỤC

DANH MỤC KÍ HIỆU, CHỮ VIẾT TẮT.....	iii
DANH MỤC BẢNG.....	v
DANH MỤC HÌNH VẼ, ĐỒ THỊ.....	vi
MỞ ĐẦU.....	1
Chương 1. TỔNG QUAN NGHIÊN CỨU.....	3
1.1. TỔNG QUAN VỀ BETA-GLUCOSIDASE.....	3
1.1.1. Vị trí của beta-glucosidase trong phức hệ cellulase.....	3
1.1.2. Phân loại beta-glucosidase.....	4
1.1.3. Cơ chế xúc tác của beta-glucosidase.....	4
1.1.4. Ứng dụng của beta-glucosidase.....	6
1.2. MỘT SỐ ĐẶC ĐIỂM CỦA BETA-GLUCOSIDASE.....	9
1.2.1. Cấu trúc module của cellulase.....	9
1.2.2. Một số tính chất của beta-glucosidase từ vi khuẩn.....	12
1.3. ỨNG DỤNG CỦA CÔNG NGHỆ METAGENOMICS TRONG KHAI THÁC NGUỒN ENZYME MỚI.....	14
1.3.1. Tổng quan về metagenome.....	14
1.3.2. Công nghệ Metagenomics tạo dữ liệu lớn cho khai thác gen.....	15
1.3.3. Khai thác beta-glucosidase từ dữ liệu metagenome.....	16
Chương 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU.....	18
2.1. ĐỐI TƯỢNG NGHIÊN CỨU.....	18
2.2. PHƯƠNG PHÁP NGHIÊN CỨU.....	19
2.2.1. Phương pháp khai thác các trình tự beta-glucosidase từ bộ dữ liệu DNA metagenome.....	19
2.2.2. Phương pháp ước đoán phổ kích thước phân tử, pI của beta-glucosidase.....	19
2.2.3. Phương pháp ước đoán vùng pH hoạt động của các beta-glucosidase.....	19
2.2.4. Phương pháp ước đoán khả năng chịu nhiệt của enzyme.....	20
2.2.5. Phương pháp xác định cấu trúc module của các beta-glucosidase.....	20
2.2.6. Phương pháp dự đoán cấu trúc không gian của các beta-glucosidase.....	21

2.2.7. Phương pháp xác định hoạt tính beta-glucosidase trên cơ chất esculin trên đĩa thạch	21
2.2.8. Phương pháp xác định ảnh hưởng của pH, nhiệt độ lên hoạt tính beta-glucosidase	21
2.2.9. Phương pháp xử lý số liệu và xác định mối liên quan đặc trưng của beta-glucosidase	23
Chương 3. KẾT QUẢ VÀ THẢO LUẬN	24
3.1. NGHIÊN CỨU KHAI THÁC VÀ ƯỚC ĐOÁN MỘT SỐ TÍNH CHẤT CỦA ENZYME BETA-GLUCOSIDASE	24
3.1.1. Nghiên cứu khai thác các gen mã hóa beta-glucosidase.....	24
3.1.2. Nghiên cứu sự phân bố độ dài của các trình tự beta-glucosidase.....	25
3.1.3. Nghiên cứu phân bố điểm đẳng điện của các trình tự beta-glucosidase.....	27
3.1.4. Ước đoán khoảng pH thích hợp cho hoạt động của beta-glucosidase..	29
3.1.5. Ước đoán khả năng bền nhiệt của beta-glucosidase	32
3.2. NGHIÊN CỨU CẤU TRÚC MODULE CỦA BETA-GLUCOSIDASE TỪ VI KHUẨN TRONG DẠ CỎ DÊ	34
3.2.1. Khảo sát họ GH của beta-glucosidase.....	34
3.2.2. Cấu trúc module beta-glucosidase họ GH1	35
3.2.3. Cấu trúc module beta-glucosidase họ GH16	36
3.2.4. Cấu trúc module beta-glucosidase họ GH3	37
3.3. NGHIÊN CỨU CÁC ĐẶC ĐIỂM CỦA ENZYME ĐƯỢC ƯỚC ĐOÁN VÀ TRÊN THỰC NGHIỆM	44
3.3.1. Đánh giá hoạt tính beta-glucosidase	44
3.3.2. Ảnh hưởng của pH tới hoạt tính beta-glucosidase.....	45
3.3.3. Ảnh hưởng của nhiệt độ tới hoạt tính beta-glucosidase.....	48
KẾT LUẬN VÀ KIẾN NGHỊ.....	49
DANH MỤC TÀI LIỆU THAM KHẢO.....	51

DANH MỤC KÍ HIỆU, CHỮ VIẾT TẮT

Tên viết tắt	Tên tiếng Anh	Nghĩa tiếng Việt
AA	Amino acid	amino acid
BGL	Beta-glucosidase	
BLAST	Basic local alignment search tools	Công cụ so sánh mức độ tương đồng về trình tự nucleotide/amino acid
bp	Base pair	Cặp base
CAZY	Carbohydrate-Active enZymes	Cơ sở dữ liệu về các enzyme tham gia chuyển hóa carbohydrate
CSDL		Cơ sở dữ liệu
eggNOG	Evolutionary genalogy of genes: Non-supervised orthologous groups	Cơ sở dữ liệu chứa các nhóm Orthologous
Expasy	Expert protein analysis system	Hệ thống phân tích protein chuyên sâu
Gb	Gigabyte	Gigabyte
GH	Glycoside hydrolase	Enzyme thủy phân liên kết glycoside
KEGG	Kyoto Encyclopedia of genes and genomes	Cơ sở dữ liệu về gen và hệ gen Kyoto
MEGAN	MEtaGenomic Analyser	Phần mềm phân tích trình tự đa hệ gen
NCBI	National Center for Biotechnology Information	Trung tâm thông tin về Công nghệ Sinh học Quốc gia
pI	Isoelectric point	Điểm đẳng điện
pNP	p-Nitrophenol	p-Nitrophenol
pNPG	p-Nitrophenyl β -D-glucopyranoside	
PHYRE	Protein Homology/analogy Recognition Engine	Protein tương đồng/tương tự

SIB	Swiss Institute of Bioinformatics	Viện nghiên cứu Tin sinh học Thụy Sĩ
SWISS-PROT	Swiss protein	Dữ liệu các trình tự đã được xác định chức năng qua thực nghiệm
TBI	Taiwan Bioinformatic Institute	Viện nghiên cứu Tin sinh học Đài Loan
Tm	Temperature melting	Nhiệt độ nóng chảy

DANH MỤC BẢNG

Bảng 3.1: Số lượng gen mã hóa beta-glucosidase đã được khai thác từ hai bộ dữ liệu dựa trên cơ sở dữ liệu KEGG, CAZy	24
Bảng 3.2: Số lượng các họ GH của beta-glucosidase của vi khuẩn trong dạ cỏ dê	35
Bảng 3.3: Các dạng cấu trúc module beta-glucosidase GH3 của vi khuẩn trong dạ cỏ dê được khai thác từ bộ dữ liệu lớn.....	37
Bảng 3.4: Giá trị pH và pI của một số enzyme đã được nghiên cứu.....	46

DANH MỤC HÌNH VẼ, ĐỒ THỊ

Hình 1.1: <i>Quá trình chuyển hóa cellulose bằng phức hệ cellulase</i>	3
Hình 1.2: <i>Cơ chế xúc tác của beta-glucosidase</i>	5
Hình 1.3: <i>Cấu trúc bậc 3 của beta-glucosidase họ GH1, GH3</i>	10
Hình 1.4: <i>Cấu trúc β-glucosidase A từ vi khuẩn Clostridium cellulovorans</i> . 10	10
Hình 1.5: <i>Cấu trúc không gian vùng xúc tác của cellulase (A): Dạng túi; (B): Dạng khe hở; (C): Dạng khe ngầm</i>	11
Hình 2.1: <i>Điện di đồ phân tích beta-glucosidase được biểu hiện và tinh chế từ tế bào E. coli Rosetta tái tổ hợp</i>	18
Hình 2.2: <i>Đường chuẩn biểu thị sự phụ thuộc giữa nồng độ pNP với cường độ màu đo được ở bước sóng 410 nm</i>	22
Hình 3.1: <i>Phân bố độ dài các trình tự beta-glucosidase trong hai bộ dữ liệu</i>	26
Hình 3.2: <i>Phân bố pI của các trình tự beta-glucosidase trong hai bộ dữ liệu tương quan với độ dài trình tự</i>	28
Hình 3.3: <i>Phân bố giá trị alkaline của các trình tự beta-glucosidase trong hai bộ dữ liệu</i>	30
Hình 3.4: <i>Mối quan hệ giữa giá trị alkaline ước đoán và độ dài trình tự</i>	31
Hình 3.5: <i>Phân bố nhiệt độ nóng chảy của các trình tự beta-glucosidase của bộ dữ liệu nhỏ trong mối tương quan với độ dài trình tự và giá trị alkaline</i>	32
Hình 3.6: <i>Phân bố nhiệt độ nóng chảy của các trình tự beta-glucosidase của bộ dữ liệu lớn trong mối tương quan với độ dài trình tự và giá trị alkaline</i>	33
Hình 3.7: <i>Cấu trúc module của 30 trình tự beta-glucosidase GH1</i>	35
Hình 3.8: <i>Cấu trúc module của 35 trình tự beta-glucosidase GH16</i>	36
Hình 3.9: <i>Cấu trúc module của 15 trình tự beta-glucosidase GH3 có CBM6</i>	38
Hình 3.10: <i>Cấu trúc module của 21 trình tự beta-glucosidase có PA14/GH31/GH5</i>	40
Hình 3.11: <i>Cấu trúc module của 81 trình tự beta-glucosidase GH3N-FN3 - GH3C</i>	41
Hình 3.12: <i>Cấu trúc module GH3N-GH3C/GH3N-GH3C-FN3 của 341 trình tự beta-glucosidase</i>	42

Hình 3.13: <i>Cấu trúc bậc ba và module của GL0036730 (GH3N-GH3C-FN3-GH31) và GL0203190 (GH3C-FN3-GH3N)</i>	43
Hình 3.14: <i>Định tính hoạt tính enzyme tinh sạch trên cơ chất esculin</i>	45
Hình 3.15: <i>Ảnh hưởng của pH tới hoạt tính enzyme beta-glucosidase trên trình tự GL0036730</i>	46
Hình 3.16: <i>Ảnh hưởng của nhiệt độ lên hoạt tính enzyme beta-glucosidase trên trình tự GL0036730</i>	48

MỞ ĐẦU

β -Glucosidase (BGL, EC 3.2.1.21) là enzyme quan trọng được ứng dụng 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. Với hoạt tính chủ yếu là phân cắt liên kết beta-glucoside, enzyme được dùng trong chuyển hóa sinh khối lignocellulose cho sản xuất cồn sinh học. Ngoài ra, enzyme còn 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 [1], [2], [3]. Vì vậy, enzyme đã thu hút sự quan tâm của các nhà khoa học ở trong và ngoài nước.

Hiện nay, việc ứng dụng BGL vào thực tế còn nhiều hạn chế do enzyme thường có hoạt độ thấp, thường 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 diễn ra ở nhiệt độ trên 80°C [4]. Để tăng hương vị cho nước ép trái cây, các enzyme có hoạt tính tối ưu ở pH axit 2,8–3,8 sẽ thích hợp để giải phóng các chất bay hơi có chứa liên kết glycoside [5]. BGL thường rất nhạy cảm với glucose nên dễ bị ức chế nếu sản phẩm chuyển hóa tăng cao [6]. 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. Môi trường như ruột mối, dạ cỏ dê,... là nơi có sự phân giải cellulose mạnh mẽ bởi vi khuẩn. Nguồn vi khuẩn trong dạ cỏ dê, ruột mối mang nguồn gen tiềm năng cho khai thác các enzyme BGL có giá trị [7], [8].

Tại phòng Kỹ thuật di truyền, hệ vi khuẩn trong dạ cỏ của hai giống dê là dê Cỏ và dê Bách Thảo đã được giải trình tự với hai cấp độ là dữ liệu nhỏ (8,4 Gb) và dữ liệu lớn (giải mã sâu với dung lượng 48,6 Gb). Đây là nguồn gen phục vụ cho việc khai thác các beta-glucosidase có giá trị trong công nghiệp. Từ những lý do trên, chúng tôi đã tiến hành đề tài "*Nghiên cứu đặc điểm beta-glucosidase dựa trên phân tích dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê*"

❖ Mục đích nghiên cứu

Đề tài được tiến hành với hai mục đích chính:

- Ước đoán một số tính chất của beta-glucosidase của vi khuẩn trong dạ cỏ dê.
- Đánh giá được đặc điểm cấu trúc module của beta-glucosidase của vi khuẩn trong dạ cỏ dê.

❖ Nội dung nghiên cứu

Đề tài được tiến hành với một số nội dung chính sau:

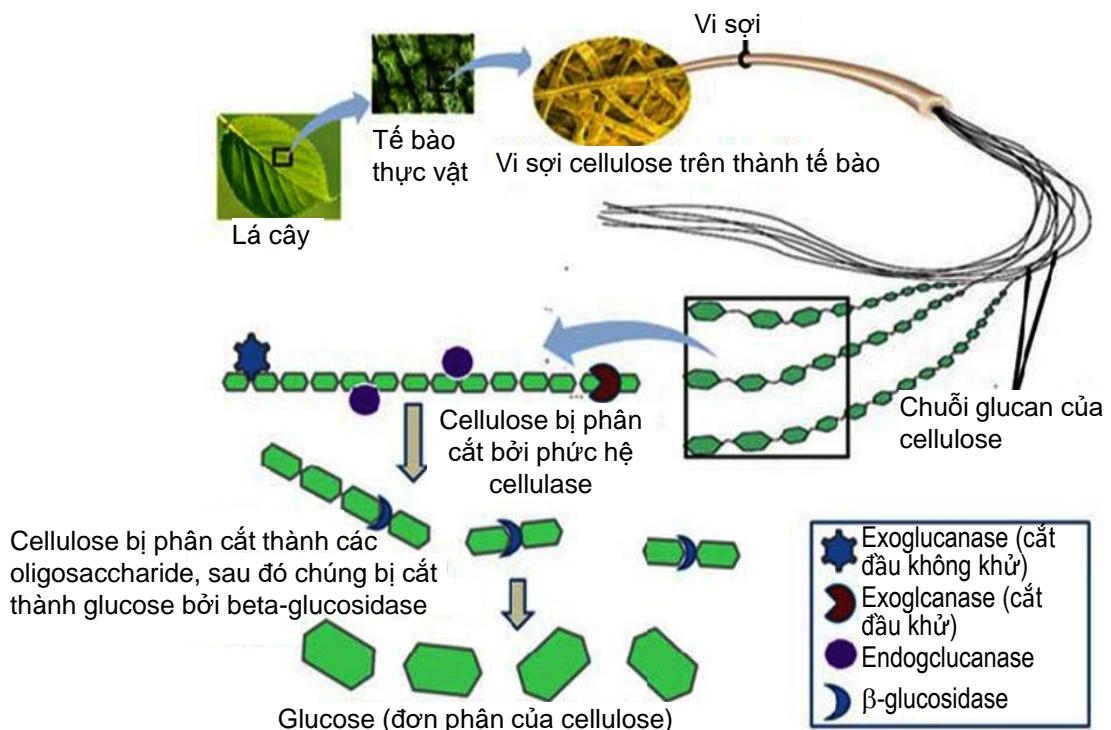
- Ước đoán một số tính chất của beta-glucosidase từ vi khuẩn trong dạ cỏ dê.
- Phân tích đặc điểm cấu trúc module của beta-glucosidase từ vi khuẩn trong dạ cỏ dê.
- So sánh đặc điểm của enzyme được ước đoán và trên thực nghiệm.

Chương 1. TỔNG QUAN NGHIÊN CỨU

1.1. TỔNG QUAN VỀ BETA-GLUCOSIDASE

1.1.1. Vị trí của beta-glucosidase trong phức hệ cellulase

Trong tự nhiên, sinh khối cellulose được thủy phân nhờ ba enzyme chính, được nhóm thành nhóm cellulase gồm endoglucanase (endo-1, 4- β -glucanase EC.3.2.1.4), exoglucanase (cellobiohydrolase) (exo-1, 4- β -glucanase, EC.3.2.1.91) và β -glucosidase (β -D-glucoside glycohydrolase, EC.3.2.1.21) [9]. Endoglucanase thủy phân ngẫu nhiên các liên kết beta-glucoside ở phần giữa của phân tử cellulose. Exoglucanase hoạt động ở các đầu khử và không khử để giải phóng cellobiose và các oligosaccharide khác. Trong quá trình thủy phân cellulose, exo và endoglucanase cùng hoạt động, hỗ trợ cho nhau tuy nhiên lại dễ bị ức chế bởi sản phẩm cellobiose. Cuối cùng, các oligosaccharide này được chuyển hóa thành glucose nhờ β -glucosidase (Hình 1.1). Beta-Glucosidase (BGL) là nhóm enzyme đa dạng, thủy phân mỗi liên kết glucoside để giải phóng gốc glucosyl ở tận cùng đầu không khử của glycoside và oligosaccharide theo cơ chế xúc tác (Hình 1.1).



Hình 1.1: Quá trình chuyển hóa cellulose bằng phức hệ cellulase [9].

1.1.2. Phân loại beta-glucosidase

β -Glucosidase cắt liên kết β -D-glucoside từ nhiều hợp chất khác nhau giải phóng glucose là sản phẩm cuối cùng. Do đó, enzyme này có sự khác biệt lớn về tính đặc hiệu cơ chất, đặc biệt là liên quan đến thành phần aglycone, khiến việc phân loại chúng trở thành một thách thức [9]. Hai phương pháp được chấp nhận rộng rãi để phân loại hiện nay là: (1) phân loại dựa trên tính đặc hiệu của cơ chất; (2) phân loại dựa trên nhận dạng trình tự amino acid và phân tích cụm kỵ nước [8].

Theo tính đặc hiệu của cơ chất, BGL được phân thành ba loại là aryl-BGL (chỉ thủy phân liên kết aryl- β -glucoside), cellobiase (chỉ thủy phân cellobiose) và BGL có tính đặc hiệu cơ chất rộng (thủy phân nhiều loại cơ chất có liên kết khác nhau, ví dụ: β (1 \rightarrow 4), β (1 \rightarrow 3), β (1 \rightarrow 6), và liên kết α (1 \rightarrow 6) [6].

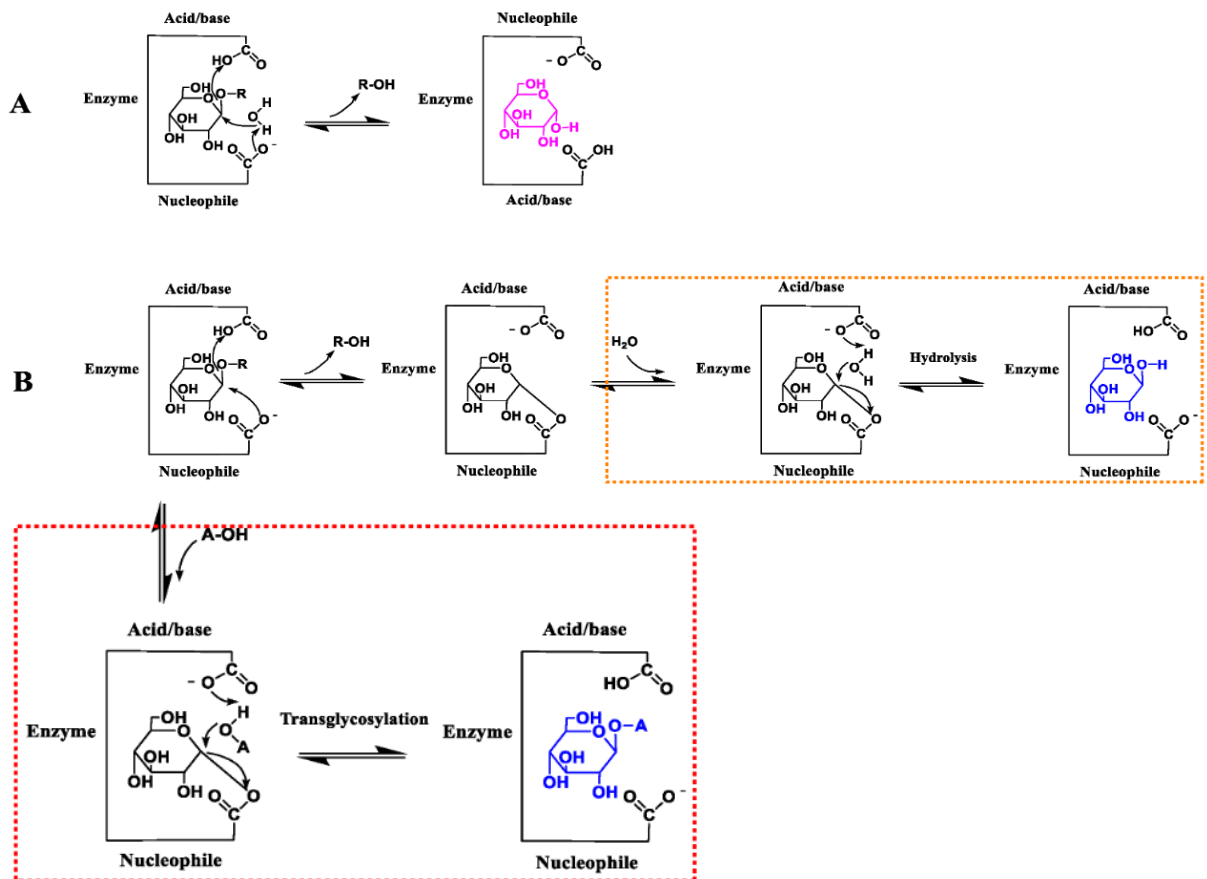
Xét về đặc điểm cấu trúc, trình tự amino acid của enzyme, và theo cơ sở dữ liệu CAZY, β -glucosidase được phân thành 10 họ glycoside hydrolase (GH) bao gồm GH1, 2, 3, 4, 5, 9, 30, 39, 11 và 116 [10], trong đó BGL phổ biến thuộc GH1 và GH3. BGL GH1 thường được tìm thấy rộng rãi ở vi khuẩn cổ, thực vật và động vật, trong khi đó BGL GH3 thường được tìm thấy chủ yếu ở vi khuẩn, nấm và nấm men [11]. β -Glucosidase của vi sinh vật được định vị dưới dạng enzyme nội bào, ngoại bào hoặc liên kết với tế bào [12]. Beta-Glucosidase GH1 bao gồm 62 loại β -glucosidase đã được nghiên cứu kỹ và có hoạt tính 6-phospho- β -glycosidase và thioglucosidase [13]. Hầu hết các BGL GH1 cũng thể hiện hoạt tính β -galactosidase.

β -Glucosidase GH3 thường là các enzyme ngoại bào hoặc liên kết với tế bào trong khi những enzyme thuộc GH1 chủ yếu là enzyme nội bào [10]. Họ GH3 có khoảng 44 β -glucosidase đã được nghiên cứu và thường có hoạt tính hexosaminidase. Phần lớn các β -glucosidase của nấm được báo cáo là enzyme GH3 và là enzyme tiết ra ngoại bào trong khi phần lớn các β -glucosidase GH1 của vi khuẩn được báo cáo là nội bào [6].

1.1.3. Cơ chế xúc tác của beta-glucosidase

BGL xúc tác phản ứng thủy phân liên kết β -1,4-glucoside bằng hai cơ chế chính là ‘giữ lại’ và ‘đảo ngược’. Cơ chế ‘đảo ngược’ chỉ bao gồm một phản ứng duy nhất khi đó các gốc ái nhân của BGL thủy phân một proton của

phân tử nước. Phân tử nước này trở nên được kích hoạt và tấn công vào liên kết glucoside để thay thế aglycone và giải phóng phần đường với sự đảo ngược cấu hình của carbon anomeric. Cơ chế ‘giữ lại’ xảy ra dưới dạng chuyển vị kép bao gồm hai bước: bước glycosyl hóa và khử glycosyl hóa. Bước đầu tiên, enzyme được glycosyl hóa bởi hoạt động phối hợp của các gốc ái nhân (thường là asparagine hoặc glutamate) tấn công carbon dị lập thể của glucose với vai trò là chất cho glycosyl và tạo thành chất trung gian enzyme-glucose [14]. Bước thứ hai liên quan đến việc một phân tử nước tấn công liên kết carbohydrate-enzyme, chuyển proton đến vị trí hoạt động của carboxylate axit/bazơ và giải phóng glucose (Hình 1.2).



Hình 1.2: Cơ chế xúc tác của beta-glucosidase [14].

A: Cơ chế đảo ngược; B: Cơ chế giữ lại. Cả hai nhóm cho (R) và nhận (A) điện tử là các phân tử glycosyl hóa khác nhau.

1.1.4. Ứng dụng của beta-glucosidase

1.1.4.1. Các ứng dụng dựa trên phản ứng thủy phân

- *Sản xuất nhiên liệu sinh học*

Sản xuất nhiên liệu sinh học, ví dụ ethanol sinh học, từ sinh khối thực vật, liên quan đến việc sử dụng nhiều enzyme hoạt động hiệp đồng để phân hủy vật liệu lignocellulose thành đường pentose và hexose, sau đó nguồn đường này được lên men để tạo nhiên liệu sinh học. Cùng với enzyme khác, β -glucosidase tham gia trong khâu cuối cùng tạo ra glucose, nguồn đường có trữ lượng lớn nhất trong sinh khối và dễ được lên men nhất để tổng hợp hiệu quả nhiên liệu sinh học. Việc phân giải các cellobiose và các chuỗi polysaccharide ngăn không những tạo đường mà còn giải ức chế theo cơ chế cạnh tranh cho endoglucanase, exoglucanase do sự tích tụ của sản phẩm lớn. Chính vì vậy β -glucosidase còn được coi là tác nhân giải ức chế trong quá trình phân giải cellulose và điểm thắt cổ chai trong sản xuất nhiên liệu sinh học. Tuy nhiên, BGL lại bị ức chế bởi sản phẩm cuối cùng của chúng, tức là glucose, do đó làm hạn chế tốc độ thủy phân cellulose. Nấm sợi *Trichoderma reesei* có hệ cellulase hoạt tính mạnh nhưng thiếu nguồn β -glucosidase cần thiết nên việc ứng dụng enzyme từ nấm cho chuyển hóa sinh khối bị hạn chế [15]. Do đó, phần lớn β -glucosidase được nghiên cứu ứng dụng cần đánh giá độ nhạy glucose [16].

- *Thủy phân isoflavone glycoside để ứng dụng trong dược phẩm*

Các hợp chất phenolic (flavonoid, flavonone, flavone và isoflavone) là chất chuyển hóa thứ cấp của thực vật, chúng khác nhau về cấu trúc hóa học và chức năng sinh học. Các hợp chất này gần đây đã là tâm điểm của nhiều nhà nghiên cứu, đặc biệt là trong lĩnh vực y tế và công nghệ thực phẩm vì hoạt tính sinh học của chúng như chất chống oxy hóa, chống ung thư, chống dị ứng, chống viêm, hạ huyết áp, v.v [17], [18], [19]. Phần lớn các hợp chất này tồn tại dưới dạng glycoside bao gồm đơn vị monoglucose liên hợp với các loại đường khác như galactose, arabinose hoặc xylose làm giảm mức độ khả dụng [20]. Vì vậy, để giải phóng phần aglycone, các enzyme arabinosidase và β -glucosidase đã được sử dụng để phân cắt các gốc đường ra khỏi aglycone. Aglycone được giải phóng có thể được hấp thụ dễ dàng do đó làm tăng hiệu

lực sinh học của chúng [21].

- *Cải thiện hương vị cho rượu, trái cây*

Trong vài thập kỷ gần đây, các nghiên cứu đã cho thấy rằng hầu hết các hợp chất hương vị trong thực vật và mô trái cây đều tồn tại dưới dạng glycoconjugate khiến chúng trở thành hợp chất không có hương vị và không bay hơi [22]. Các hợp chất hương vị cấu trúc glycoside đã được ghi nhận trong nhiều loại trái cây như nho [23], mận vàng [24], xoài [25], và dâu tây [26]. Các glycoside này rất phức tạp và đa dạng về cấu trúc, đặc biệt là phần aglycone. Phần glycone thường bao gồm đơn vị glucose liên hợp với nhiều glycoside khác nhau như 6-O- α -larabinofuranosyl- β -D-glucopyranoside và 6-O- α -L-arabinopyranosyl- β -D-glucopyranoside. Để chuyển hóa các hợp chất không hương vị này thành chất có hương vị cao, chúng phải được thủy phân để giải phóng phần aglycone. Quá trình thủy phân có thể được thực hiện bằng cách sử dụng axit hoặc enzyme [27]. Đầu tiên, các enzyme như α -L-rhamnosidase hoặc α -L-arabinosidase phân tách giải phóng arabinose và rhamnose, sau đó β -glucosidase tác động lên glucose giải phóng β -D-glucoside tương ứng và phần aglycone như monoterpeneol [28]. Tuy nhiên, β -glucosidase từ thực vật như nho có hoạt tính thấp và không ổn định trong điều kiện sản xuất rượu vang, do đó việc bổ sung β -glucosidase từ vi khuẩn có hoạt tính cao và ổn định là bắt buộc để thủy phân hoàn toàn các hợp chất hương vị. β -Glucosidase có hiệu suất thủy phân cao đối với terpenyl glycoside đã được báo cáo có nguồn gốc từ *Soridiobolus pararoseus* [29]. β -Glucosidase từ *Oenococcus oeni* ATCC BAA-1163 đã được sử dụng làm tăng hương vị trong rượu xạ hương [30], β -glucosidase từ *Lactobacillus brevis* có thêm hoạt tính xylosidase, arabinosidase và cellobiosidase bền và hoạt động tốt ở pH 4,0-7,0 được ứng dụng trong việc tăng cường mùi thơm của rượu vang [31]. Các BGL hoạt động tốt ở pH nghiêng acid (pH 3-4,5) được quan tâm trong nghiên cứu và ứng dụng trong sản xuất rượu [32]. Ngoài ra, enzyme có khả năng hoạt động tốt ở môi trường có ethanol, glucose và các chất khác như β -glucosidase ngoại bào từ *Issatchenkia terricola* cũng được chú ý nhiều trong công nghệ tạo mùi thơm cho rượu.

- *Giải độc sản*

β -Glucosidase có tiềm năng được sử dụng để giải độc sản. Sản là loại

cây giàu carbohydrate mọc ở nhiều nơi trên thế giới và là loại lương thực chính cho 500 triệu người trên thế giới. Tuy nhiên, sử dụng sản phẩm làm thực phẩm có hại cho sức khỏe con người do sự hiện diện của glycoside cyanogen như linamarin và lotaustralin [33]. Hơn nữa, mối tương quan giữa hội chứng nhiễm độc hệ thần kinh trung ương của người “Konzo” và việc tiêu thụ các sản phẩm sản kéo dài đã được chứng minh. Sản tự nhiên được khử độc trong quá trình chế biến và nghiền nhờ β -glucosidase nội sinh và linamarase có trong củ. Tuy nhiên, các enzyme nội sinh không đủ để phân giải hết glycoside cyanogen trong sản do vậy mà việc bổ sung enzyme từ bên ngoài là hết sức cần thiết. Do đó, người ta đề xuất rằng linamarase ngoại sinh và β -glucosidase từ các nguồn vi sinh vật có thể được sử dụng để tăng cường quá trình thủy phân glycoside cyanogen từ thực phẩm quan trọng này [33].

- *Tái chế giấy thải*

Ngành giấy và bột giấy là một trong những ngành tiêu thụ nhiều gỗ nhất và dự kiến sẽ còn mở rộng hơn nữa do sự gia tăng của nền kinh tế và dân số thế giới. Giấy thải là một trong những chất gây ô nhiễm môi trường lớn. Tái chế giấy thải đang thu hút nhiều sự quan tâm hơn trong thời điểm hiện nay để giải quyết vấn đề hai chiều này: tiêu thụ tiết kiệm gỗ rừng và hạn chế ô nhiễm bãi chôn lấp. Tái chế giấy thải hiện nay có thể được thực hiện bằng phương pháp hóa học hoặc enzyme. Rào cản lớn nhất đối với việc tái chế giấy thải là việc loại bỏ mực. Loại bỏ mực khỏi giấy thải bằng phương pháp thông thường sử dụng một số hóa chất có hại cho môi trường và làm giảm độ sáng của giấy. Phương pháp enzyme để tái chế giấy thải đã được báo cáo là có hiệu quả trong việc giải quyết những vấn đề này. Các chế phẩm enzyme để tái chế giấy thải có chứa cellulase, β -glucosidase và hemiaellulase [34], [35].

1.1.4.2. Ứng dụng beta-glucosidase dựa trên hoạt động sinh tổng hợp

β -glucosidase được biết là có hoạt tính tổng hợp glycosyl tạo ra các oligosaccharide, aryl- và alkyl- β -D-glycoside với nhiều ứng dụng. Sự tổng hợp oligosaccharide bằng β -glucosidase có nhiều ưu điểm hơn so với glycosyl transferase vì enzyme có tính đặc hiệu cơ chất tốt hơn và đơn giản hơn so với glycosyl transferase [36]. Alkyl glycoside có nhiều ứng dụng vì chúng là chất hoạt động bề mặt không ion có khả năng phân hủy sinh học, có đặc tính nhũ hóa và kháng khuẩn tốt nhờ nhóm đầu carbohydrate của chúng. Este nalkyl

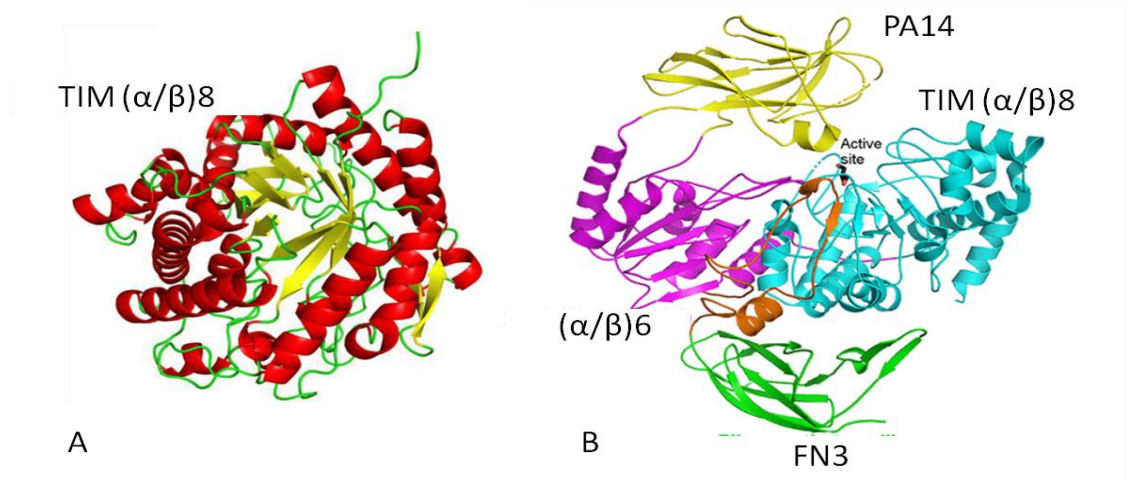
glucoside được hình thành do phản ứng của axit phenyl butyric và n-alkyl butyl glucoside bởi β -glucosidase –lipase được sử dụng trong điều trị sốt [37]. Mặt khác, oligosaccharide tổng hợp có thể được sử dụng trong nhiều ứng dụng khác nhau: (1) tác nhân trị liệu như heparin và acarbose; (2) các kỹ thuật dựa trên carbohydrate như vắc xin kháng khuẩn, chống ký sinh trùng và kháng virus; và (3) tạo các prebiotics tăng cường sự phát triển của các vi sinh vật có lợi trong hệ thực vật đường ruột của con người [3], [36].

1.2. MỘT SỐ ĐẶC ĐIỂM CỦA BETA-GLUCOSIDASE

1.2.1. Cấu trúc module của cellulase

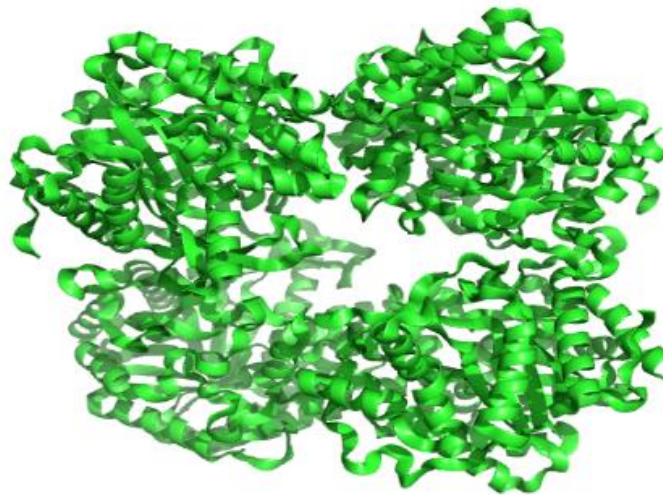
Cellulase tồn tại ở hai dạng (1) enzyme không có cấu trúc module là phân tử enzyme chỉ chứa duy nhất vùng xúc tác, có hoạt tính xúc tác; (2) enzyme có cấu trúc module (modular enzyme) là trong phân tử, ngoài vùng xúc tác ra enzyme còn chứa các vùng hỗ trợ khác, có cấu trúc bậc 3 tách biệt được gọi là các module. Các module không xúc tác trên phân tử enzyme được gọi là module phụ trợ. Ngoài ra, nhiều enzyme có hơn một vùng xúc tác.

Beta-Glucosidase được chia thành 10 họ gồm GH 1, 2, 3, 4, 5, 9, 30, 39, 11, 116 [10]. Cấu trúc điển hình của tất cả các beta-glucosidase là dạng ống TIM (α/β)₈ gồm tám phiến gấp nếp beta song song tạo thành ống và được bao bọc bởi 8 xoắn ốc alpha ưa nước nằm ở phía ngoài. Điển hình là BGL GH1 có cấu trúc điển hình dạng ống TIM (α/β)₈ và trung tâm xúc tác dạng túi [11] (Hình 1.3 A). Họ GH3, có cấu trúc phức tạp hơn bao gồm cấu trúc TIM (β/α)₈ barrel, (α/β)₆ barrel và vùng giống fibronectin III (FN3). Ngoài hai vùng này ra, GH3 còn có thêm các vùng module khác như PA14 (Hình 1.3 B). Ngoài ra, mặc dù cấu trúc dạng ống (α/β)₈ là cấu trúc điển hình của beta-glucosidase nhưng gần đây BGL GH116 được phát hiện không có cấu trúc điển hình này mà lại có cấu trúc (α/β)₆ ở tận cùng đầu C và cấu trúc chủ yếu phiến beta nằm ở phía đầu N xen kẽ với chuỗi xoắn α ngắn [38]. Các họ BGL khác vẫn chưa được nghiên cứu cấu trúc.



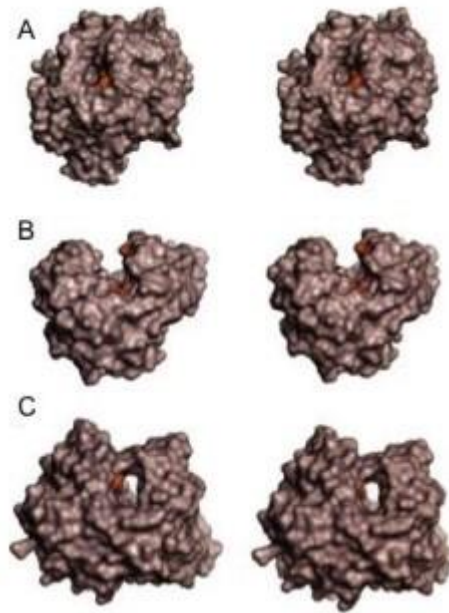
Hình 1.3: Cấu trúc bậc 3 của beta-glucosidase họ GH1, GH3.

(A) Cấu trúc của *PcBGL1A* (PDB ID: 2e3z) thuộc họ GH1 có nguồn gốc từ *Phanerodontia chrysosporium* [39] và cấu trúc *BcpE2* GH3 của *Sarcoptes scabiei* [40].



Hình 1.4: Cấu trúc β -glucosidase A từ vi khuẩn *Clostridium cellulovorans* [41, <https://en.wikipedia.org/wiki/%CE%92-Glucosidase>].

β -Glucosidase thường có cấu trúc bậc 4 ở dạng dimer gồm 2 chuỗi homopolyptide [42]. Mỗi chuỗi gồm 438 amino acid tạo thành một tiểu đơn vị enzyme, mỗi tiểu đơn vị này chứa một vị trí hoạt động [43]. Vị trí này thường có cấu trúc không gian dạng túi, khe hở và khe ngàm [44] (Hình 1.5). Cấu trúc túi có lợi cho việc nhận biết monosaccharide như glucose. Cấu trúc khe hở cho phép liên kết với đường để tạo thành cấu trúc polysaccharide. Cấu trúc khe ngàm cho phép enzyme gắn vào polysaccharit và sau đó giải phóng sản phẩm trong khi vẫn gắn vào đường [44].



Hình 1.5: Cấu trúc không gian vùng xúc tác của cellulase (A): Dạng túi; (B): Dạng khe hở; (C): Dạng khe ngàm [44].

Module phụ trợ của cellulase khá đa dạng bao gồm module liên kết carbohydrate (carbohydrate-binding module-CBM), vùng giống fibronectin 3 (fibronectin like domain-FN3, có độ đa dạng cao, tương đồng về trình tự 33,7-83,2% [45], nhưng giống nhau về mặt cấu trúc [46]), vùng tương tự globulin miễn dịch (immunoglobulin-like domain-Ig, tương đồng chủ yếu về mặt cấu trúc ở vùng bảo thủ [47]), vùng X, nhân tố kết tập type C (F5/6 type C coagulation factor), dockerin type I, S-layer homology, hoặc tác nhân dị ứng phấn hoa,... CBM có ái lực với các polysaccharide sợi đơn hoặc kết thành bó để neo giữ hoặc định hướng cho các enzyme nằm trên cùng phân tử tiếp cận được cơ chất [48]. Trong một số trường hợp, CBM có khả năng phá vỡ cấu trúc tinh thể của vi sợi cellulose để giúp cho cellulase hoạt động [48].

Fibronectin type III thường được tìm thấy ở động vật với chức năng chính của chúng là làm trung gian cho các tương tác giữa protein-protein, là vùng nối (linker) giúp ổn định cấu trúc protein, đảm bảo hoạt tính sinh học chính xác của enzyme. FN3 cũng được tìm thấy ở các protein màng hay protein nội, ngoại bào. Trình tự FN3 của vi khuẩn chỉ được tìm thấy trên các phân tử enzyme glycohydrolase ngoại bào [49]. Ngược lại với CBM, FN3 phân bố ngẫu nhiên và thường có nhiều bản sao trên các glycohydrolase thuộc

các họ khác nhau của vi khuẩn (polygalacturonidase, cellulase, chitinase, pullulanase, amylase,...) [50]. Các nghiên cứu đưa ra giả thiết rằng FN3 đóng vai trò như vùng nối dài giúp cho phản ứng enzyme và cơ chất xảy ra chính xác. FN3 ở vi khuẩn được tìm thấy lần đầu tiên trong cấu trúc CBHA của *Clostridium thermocellum* [51]. Các nhà khoa học cho rằng rất có thể module này làm nối lỏng bề mặt cellulose, làm bong tróc, tách sợi cellulose và bộc lộ các vùng cho enzyme thủy phân tiếp cận phân giải cellulose. FN3 cũng định hướng chuỗi cellulose vào trung tâm xúc tác để enzyme dễ dàng chuyển hóa cơ chất. Trình tự module FN3 và Ig đều tương đồng với FN3 và globulin miễn dịch của người nhưng chức năng sinh học thì chưa được biết [52].

1.2.2. Một số tính chất của beta-glucosidase từ vi khuẩn

Beta-Glucosidase là một enzyme phổ biến và được phát hiện ở tất cả các thực thể sống như động vật, thực vật, và ở tất cả các dạng vi sinh vật như: vi khuẩn [53], vi khuẩn cổ [35], và nấm [54]. Nó cũng đã được tìm thấy ở thực vật và động vật [55]. Đối với enzyme ứng dụng trong công nghiệp, vi sinh vật được coi là sự lựa chọn tốt nhất để sản xuất enzyme, việc ưu tiên sử dụng vi sinh vật làm nguồn sản xuất enzyme quy mô công nghiệp là do một số nguyên nhân sau: (1) vi sinh vật phát triển nhanh chóng giúp đẩy nhanh quá trình sản xuất enzyme; (2) vi sinh vật dễ nuôi cấy ở quy mô lớn, dễ kiểm soát; (3) vi sinh vật có thể dễ dàng được điều khiển với sự trợ giúp của kỹ thuật di truyền, gây đột biến và tiến hóa trực tiếp; (4) một số vi sinh vật tạo ra enzyme có đặc tính đặc biệt như khả năng chịu nhiệt và tính kiềm có thể được sử dụng trong nhiều ngành công nghiệp đòi hỏi điều kiện khắc nghiệt như vậy.

Đến nay, nhiều nghiên cứu đã được công bố tập trung vào các đối tượng vi khuẩn sản sinh BGL trong đó bao gồm các đối tượng chính như *Clostridium thermocellum* [56], *Pyrococcus furiosus* [57], *Bacillus circulans* sub spp. *Alkalophilus* [11], *Flavobacterium johnsoniae* [11], *Actinomycete*, *Thermobifida fusca* [58], *Paenibacillus* sp. C5 [12], *Lactobacillus brevis* [31], *Caldicellulosiruptor saccharolyticus* [59], và *Terrabacter ginsenosidimutans* spp [60] Nhìn chung các loài vi khuẩn sản sinh lượng beta-glucosidase thấp, tuy nhiên chúng có tiềm năng lớn để ứng dụng sản xuất trong quy mô công nghiệp.

- *Nhiệt độ và pH tối ưu*

β -Glucosidase của vi khuẩn thể hiện hoạt độ tối ưu ở pH 6-7 tuy nhiên cũng có BGL có hoạt tính tối ưu ở pH 5, 8 và 10 đã được báo cáo như BGL tách từ *Caldicellulosiruptor saccharolyticus* [59], *Bacillus halodurans* [61]. β -Glucosidase của vi khuẩn được khai thác từ các bộ dữ liệu metagenomics cũng thể hiện hoạt độ tối ưu ở pH 4-10 có tiềm năng ứng dụng trong công nghiệp [62], [63].

Trong tự nhiên, nhiệt độ tối ưu cho hoạt động của beta-glucosidase được tìm thấy nằm trong khoảng 25°C tới 105°C nhưng phần lớn các BGL GH1 có hoạt tính tối ưu ở xấp xỉ 50°C và GH3 thường có hoạt tính tối ưu ở xấp xỉ 60°C. β -Glucosidase từ vi khuẩn thích nghi lạnh *Micrococcus antarcticus* (MaBGL) có nhiệt độ tối ưu ở 25°C và có thể hoạt động được tốt ở nhiệt độ 4-25°C.

- *Tính đặc hiệu cơ chất*

Dựa trên tính đặc hiệu cơ chất β -glucosidase được phân loại thành ba loại: (1) cellobiase xúc tác quá trình thủy phân cellobiose và oligosaccharide ngắn; (2) aryl- β -D-glucosidase chỉ thủy phân aryl- β -D-glucoside như p-NPG và (3) β -glucosidase có tính đặc hiệu cơ chất rộng, xúc tác quá trình thủy phân nhiều loại cơ chất với các liên kết β (1-4), β (1-2), β (1-3), α (1-4), α (1-2), α (1-6), v.v. [64]. β -Glucosidase với tính đặc hiệu cơ chất rộng đã được xác định từ nhiều loại nấm. Các nghiên cứu sâu về cấu trúc, đặc điểm cấu trúc liên quan đến tính đặc hiệu cơ chất sẽ giúp khai thác được các enzyme có tính đặc hiệu cơ chất như mong muốn.

- *Khả năng dung nạp glucose*

β -Glucosidase của vi sinh vật bị ức chế bởi glucose - sản phẩm quá trình chuyên hóa. Ví dụ, β -glucosidase của *Fusarium oxysporum* bị ức chế cạnh tranh bởi glucose có giá trị hằng số ức chế (K_i) là 2,05 mM và β -glucosidase của *Penicillium italicum* bị ức chế không cạnh tranh với K_i là 8,9 mM. Trong tự nhiên cũng có nhiều enzyme BGL dung nạp glucose (glucose-tolerant BGL) có thể tránh được sự tích tụ của sản phẩm glucose trong quá

trình chuyển hóa sinh khối, đặc biệt BGL có nguồn gốc từ nấm sợi [65]. BGL GH1 được ghi nhận có khả năng dung nạp glucose tốt hơn so với BGL GH3 [66]. Tất cả các BGL dung nạp glucose đều là enzyme chịu nhiệt [67], [68]. Tuy nhiên, các enzyme có khả năng dung nạp glucose càng cao thì hiệu quả chuyển hóa lại thấp [69], [70], [71]. Do đó, việc xác định β -glucosidase chịu nhiệt và chịu glucose có ý nghĩa rất lớn để xử lý nhiệt độ tăng cao và tăng nồng độ glucose trong quá trình xử lý sinh học trong công nghiệp.

1.3. ỨNG DỤNG CỦA CÔNG NGHỆ METAGENOMICS TRONG KHAI THÁC NGUỒN ENZYME MỚI

1.3.1. Tổng quan về metagenome

Theo phương pháp truyền thống, nghiên cứu về vi sinh vật thường dựa trên việc nuôi cấy, tuy nhiên việc này có nhược điểm khi số lượng vi khuẩn có thể nuôi cấy được với tổng số vi sinh vật của hệ sinh thái thường không nhiều [72]. Metagenomics là một phương pháp mới để nghiên cứu về tổng số bộ gen của quần xã vi sinh vật trong một môi trường cụ thể bằng cách sử dụng sàng lọc chức năng gen hoặc sàng lọc trình tự. Trong metagenomics, việc nghiên cứu các hệ gen của các vi sinh vật trong một quần xã không chỉ cho biết về di truyền, sinh lý và hóa sinh của các vi sinh vật mà còn cung cấp thông tin chi tiết về vòng tuần hoàn dinh dưỡng và năng lượng trong quần xã, cấu trúc bộ gen, chức năng gen, di truyền quần thể và chuyển gen giữa các thành viên của một quần thể sinh vật không thể nuôi cấy. Nghiên cứu metagenomics đang phát triển nhanh chóng trong y học, nông nghiệp, bảo vệ môi trường và các lĩnh vực khác.

Metagenomics cung cấp thông tin về chức năng gen của các quần xã vi sinh vật và do đó đưa ra mô tả rộng hơn nhiều so với các khảo sát về nguồn gốc gen thường chỉ dựa trên sự đa dạng của một gen chẳng hạn như gen 16S rRNA. Bằng kỹ thuật metagenomics, các thông tin về chất xúc tác sinh học hoặc các enzyme mới, mối liên kết giữa chức năng và phát sinh loài đối với các sinh vật chưa được nuôi cấy có thể được phát hiện. Metagenomics cũng là

một công cụ mạnh mẽ để tạo ra các giả thuyết mới về chức năng của vi sinh vật như quang dị dưỡng dựa trên sinh vật quang dị dưỡng hoặc vi khuẩn cố oxy hóa amoniac [73].

1.3.2. Công nghệ Metagenomics tạo dữ liệu lớn cho khai thác gen

Nghiên cứu hiện tại ước tính rằng đất và đại dương của Trái Đất chứa lần lượt khoảng $4-5 \times 10^{30}$ và $3,6 \times 10^{29}$ tế bào vi sinh vật. Chỉ có 1% số vi sinh vật này có thể nuôi cấy được bằng các kỹ thuật trong phòng thí nghiệm và phần lớn còn lại, khoảng 99%, không thể nuôi cấy được trong điều kiện phòng thí nghiệm, do đó khiến chúng chưa được nghiên cứu và sử dụng để sản xuất nhiều sản phẩm có giá trị gia tăng [72]. Metagenomic, một thuật ngữ do Handelsma đặt ra vào năm 1998, là kỹ thuật không phụ thuộc vào nuôi cấy được sử dụng để phân tích vật liệu di truyền có trong mẫu môi trường [74]. Cách tiếp cận này bắt đầu bằng việc tách chiết DNA trong môi trường, xây dựng thư viện DNA metagenome và sàng lọc các thư viện để phân lập gen và trình tự quan tâm [72]. Việc sàng lọc thư viện metagenomic có thể được thực hiện bằng cách sàng lọc dựa trên chức năng, sàng lọc cụ thể gen hoặc giải trình tự trực tiếp [72]. Phương pháp metagenomics có thể được sử dụng để tìm ra gen mới mã hóa cho protein mới, ví dụ như các enzyme có đặc tính đặc biệt. Số lượng β -glucosidase đã được xác định bằng cách sử dụng phương pháp metagenomics từ các mẫu môi trường khác nhau như metagenome đất ngập nước [75], metagenome dạ cỏ bò [62], metagenome manh tràng thỏ [76], metagenome dạ cỏ trâu [65],... ngày càng tăng. Metagenomic sẽ cung cấp lượng thông tin lớn hơn về chức năng, trình tự gen của một quần xã vi sinh vật trong môi trường đánh giá. Điều này cho thấy sự ưu việt và tiềm năng ứng dụng lớn của kỹ thuật này so với phương pháp truyền thống [77].

Trong những năm gần đây, với sự phát triển nhanh chóng của công nghệ giải trình tự gen thế hệ mới và những tiến bộ vượt bậc của công nghệ 4.0 trong việc ứng dụng công nghệ thông tin vào lắp ghép, chú giải, khai thác gen, việc giải trình tự metagenome của vi sinh vật để cung cấp bộ dữ liệu lớn cho khai thác gen được các nhà khoa học đặc biệt quan tâm. Nghiên cứu gen từ DNA đa hệ gen thường trải qua ba bước chính như sau: (1) tách chiết và giải trình tự mẫu DNA đa hệ gen; (2) tập hợp các đoạn read ngắn thành các contig dài; (3) sử dụng những phần mềm tinh sinh chuyên dụng để ước đoán

chức năng của gen. Hiện nay, các công cụ giải trình tự thế hệ mới có nhiều ưu việt hơn phương pháp Sanger truyền thống với chi phí thấp hơn, độ dài đọc lớn hơn và khả năng ứng dụng cao để tạo bộ dữ liệu metagenome. Sau khi thu nhận trình tự từ bộ dữ liệu, các gen được dự đoán về đơn vị phân loại và chức năng gen. Dựa vào kết quả đánh giá mức độ tương đồng của trình tự với các trình tự trên các cơ sở dữ liệu có sẵn để ước đoán chức năng và đơn vị phân loại. Hiện nay, cơ sở dữ liệu về đơn vị phân loại của gen thường dùng là cơ sở dữ liệu NR (là CSDL chứa các trình tự non – redundant từ ngân hàng gen cùng với các trình tự từ các dữ liệu ngân hàng khác như Refseq, PDB, SwissProt, PIR và PRF), các CSDL về chức năng gen đáng tin cậy như: Kyoto Encyclopedia of Genes and Geneomes (KEGG) (<https://www.kegg.jp>) là CSDL phân loại chức năng gen theo con đường chuyển hóa, evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) là CSDL phân loại chức năng theo tiến hóa của gen, Clusters of Orthologous Group (COG) là một hệ thống của các họ gen từ các genome hoàn chỉnh [78]. Từ việc dự đoán chức năng, các gen mã hóa cho các enzyme tiềm năng được khai thác hiệu quả.

1.3.3. Khai thác beta-glucosidase từ dữ liệu metagenome

Enzyme beta-glucosidase là một enzyme quan trọng và có nhiều ứng dụng trong cuộc sống và sản xuất. Vì vậy, việc nghiên cứu, tìm kiếm và phân lập vi sinh vật sinh enzyme tốt mang các đặc tính sinh học quý được chú trọng nghiên cứu. Năm 2006, một nhóm tác giả Trung Quốc phân tích trình tự DNA thu được từ cặn bột giấy và đã xác định được hai gen mã hóa endoglucanase, ba gen mã hóa exoglucanase và hai gen β -glucosidase [73]. Các trình tự thu được từ bộ dữ liệu từ manh tràng thỏ cũng đã được dự đoán chức năng. Kết quả cho thấy mười một gen mã hóa cellulase gồm bốn gen endo- β -1,4-glucanase và bảy gen β -glucosidase đã được phân lập [69]. Một công trình khác công bố năm 2008 với kết quả đã phân lập được 118 trình tự mã hóa enzyme beta-glucosidase từ dạ cỏ trâu. Việc sàng lọc các gen này cho thấy tám gen có hoạt tính β -glucosidase cao ở pH 5,0 và 37°C và một trong số tám gen đó tiếp tục được khảo sát sâu hơn, thu được kết quả gen chọn lọc có độ tương đồng cao với gen mã hóa β -glucosidase từ *Bacillus spp.* [79]. Ngoài ra còn có rất nhiều công bố từ các nhóm nghiên cứu khác nhau trên thế giới

thực hiện đánh giá phân tích sàng lọc gen mã hóa enzyme beta-glucosidase từ nhiều môi trường khác nhau như: đất ngập nước [75], metagenome dạ cỏ bò [62], ...

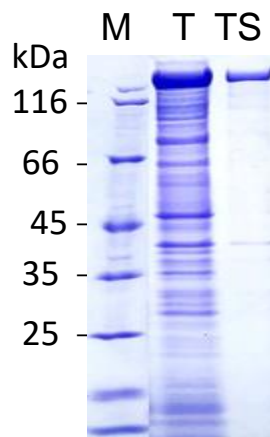
Ở Việt Nam, các nghiên cứu về các loài vi sinh vật và phân lập, tìm kiếm các gen mã hóa cellulase đã được tiến hành từ những năm 2000. Nhóm nghiên cứu của phòng Kỹ thuật di truyền, Viện Công nghệ sinh học đã giải trình tự metagenome của vi khuẩn trong ruột mối và đã khai thác 316 gen mã hóa enzyme tham gia chuyển hóa cellulose bao gồm β -glucosidase, licheninase, endoglucanase, cellobiosidase, và phosphorylase cellobiose [80]. Với kỹ thuật tương tự, DNA đa hệ gen của vi sinh vật trong dạ cỏ dê đã được nhóm nghiên cứu giải mã với dung lượng thấp (8,4 Gb) năm 2018 và giải mã sâu năm 2019 (dung lượng 48,6 Gb). Từ bộ dữ liệu nhỏ, 816 gen mã hóa 11 họ GH của cellulase đã được khai thác [8]. Đồng thời, đa hệ gen vi khuẩn trong suối nước nóng Bình Châu đã được giải mã với dung lượng 9,4 Gb. Qua nghiên cứu khai thác, nhóm nghiên cứu của TS. Nguyễn Kim Thoa đã thu được 82 trình tự mã hóa cho endoglucanase, exoglucanase và β -glucosidase [75]. Như vậy, kỹ thuật metagenomic cho thấy tiềm năng to lớn ứng dụng và thu thập được lượng lớn dữ liệu có giá trị trên nhiều môi trường tiềm năng khác nhau. Mặc dù các nghiên cứu khai thác được lượng enzyme lớn nhưng việc nghiên cứu sâu về đặc điểm cấu trúc và ước đoán một số tính chất của enzyme chưa được nghiên cứu. Vì vậy, trong luận án này, dựa trên dữ liệu metagenome của vi khuẩn trong dạ cỏ dê đã được giải mã, chúng tôi sẽ khai thác gen, ước đoán cấu trúc và dự đoán một số đặc điểm của beta-glucosidase của vi khuẩn trong dạ cỏ dê.

Việc áp dụng β -glucosidase một cách hiệu quả đòi hỏi kiến thức chi tiết về cấu trúc, cơ chế phản ứng của enzyme. Những nghiên cứu về module mã enzyme cho phép tìm kiếm được các enzyme có phổ cơ chất rộng, enzyme nhiều hoạt tính giúp chuyển hóa hiệu quả cơ chất. Đồng thời, việc tìm hiểu các module phụ trợ còn giúp đưa ra cơ sở khoa học cho các nghiên cứu cải biến trình tự, thiết kế enzyme, gập đột biến enzyme giúp làm tăng ái lực enzyme-cơ chất và phân cắt cơ chất hiệu quả.

Chương 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. ĐỐI TƯỢNG NGHIÊN CỨU

Bộ dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê đã được giải trình tự với dung lượng 8,4 Gb, với 164644 gene đã được ước đoán. Sau khi giải mã sâu đa hệ gen vi khuẩn trong dạ cỏ dê thu được bộ dữ liệu lớn có dung lượng dữ liệu là 48,66 Gb, tổng cộng có 5367270 gene đã được dự đoán. Các gene trong hai bộ dữ liệu đã được chú giải chức năng dựa trên dữ liệu KEGG, CAZy. Cả hai bộ dữ liệu này được sử dụng để khai thác các gen mã hóa beta-glucosidase, nghiên cứu đặc điểm của enzyme dựa trên công cụ tin sinh học và so sánh đặc điểm enzyme từ hai bộ dữ liệu.



Hình 2.1: Điện di đồ phân tích beta-glucosidase được biểu hiện và tinh chế từ tế bào *E. coli* Rosetta tái tổ hợp.

M: Protein chuẩn (Fermentas); T: Protein pha tan được biểu hiện trong E. coli Rosetta mang gen GL0036730; TS: Beta-glucosidase tinh sạch được sử dụng để kiểm tra ảnh hưởng của pH, nhiệt độ lên hoạt tính enzyme.

Để kiểm chứng lại các kết quả dự đoán, trình tự GL0036730 mã hóa cho beta-glucosidase GH3-GH31 đã được Phòng Kỹ thuật di truyền tối ưu mã bộ ba và biểu hiện trong *Escherichia coli*. Beta-Glucosidase đã được tinh chế thành công bằng cột sắc ký ái lực His-tag (Hình 2.1). Enzyme này sau đó được sử dụng để nghiên cứu ảnh hưởng của pH, nhiệt độ tới hoạt tính phân giải cơ chất đặc hiệu pNPG.

2.2. PHƯƠNG PHÁP NGHIÊN CỨU

2.2.1. Phương pháp khai thác các trình tự beta-glucosidase từ bộ dữ liệu DNA metagenome

Từ 2 bộ dữ liệu metagenome của vi khuẩn sống tự do trong dạ cỏ dê và kết quả chú giải chức năng bằng công cụ KEGG/CAZy, các trình tự mã hóa cho BGL sẽ được trích xuất. Các thông tin Mã trình tự, trình tự amino acid sẽ 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. Các bộ dữ liệu trình tự của các hệ vi sinh khác nhau được phân tích riêng rẽ và sử dụng để dự đoán phổ pI, chỉ số alkaline, nhiệt độ nóng chảy (T_m) của enzyme và nghiên cứu cấu trúc module của mỗi enzyme.

2.2.2. Phương pháp ước đoán phổ kích thước phân tử, pI của beta-glucosidase

Khối lượng phân tử và giá trị pI là hai giá trị cơ bản của enzyme và có khả năng liên quan đến cấu trúc phân tử, nhiệt độ nóng chảy và khả năng tan của enzyme ở các điều kiện pH khác nhau. Các enzyme đơn giản, mang một vùng xúc tác thường có kích thước phân tử bé hơn các enzyme có cấu trúc phức tạp gồm nhiều module khác nhau. Các phân tử phức tạp thường dễ bị ảnh hưởng bởi nhiệt độ. Do vậy, trong nghiên cứu này các trình tự BGL đã được đưa vào công cụ ước đoán pI (https://web.expasy.org/compute_pi/) trong gói công cụ EXPASY. Trình tự sau khi đưa vào, phần mềm ngay lập tức sẽ tính ra chỉ số pI của enzyme. Độ dài trình tự amino acid của mỗi enzyme được tính toán dựa trên công thức đếm số lượng ký tự trong Excel 2010.

2.2.3. Phương pháp ước đoán vùng pH hoạt động của các beta-glucosidase

Năm 2013, Fan và đồng tác giả đã xây dựng công cụ Pred-enzyme để dự đoán các enzyme có tính axit và kiềm, trong đó các enzyme kiềm có khả năng hoạt động tốt trong vùng pH kiềm và ngược lại enzyme có tính axit sẽ có khả năng hoạt động tốt trong vùng axit. Công cụ có độ chính xác tổng thể là 94,01% [81]. Việc dự đoán pH hoạt động của enzyme được thực hiện trực tuyến theo đường link <http://lin-group.cn/server/AcalPred>. Khi có một trình

tự, hoặc một dãy các trình tự amino acid được định dạng fasta được nhập vào phần mềm, phần mềm sẽ tự động tính toán và cho ra kết quả cuối cùng là protein thuộc alkaline hay acid sau vài phút với giá trị alkaline hay acid trong khoảng 0-1 [82]. Trong đó tổng điểm của alkaline và acid là 1. Chỉ số alkaline càng cao thì enzyme có khả năng hoạt động càng tốt ở vùng pH kiềm và ngược lại.

2.2.4. Phương pháp ước đoán khả năng chịu nhiệt của enzyme

Khả năng chịu nhiệt của enzyme được tính toán theo phần mềm miễn phí TBI do Viện nghiên cứu Tin sinh học của Đài Loan thiết kế dựa trên tổng hợp đặc điểm chịu nhiệt của protein đã nghiên cứu như trình tự của các amino acid, liên kết hidro, 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 [83]. Phần mềm được xây dựng trên số liệu của 150000 protein chịu nhiệt độ khác nhau để dự đoán nhiệt độ nóng chảy T_m của các trình tự, dựa trên nguyên tắc tương đồng. Các trình tự amino acid của beta-glucosidase được nạp vào phần mềm của TBI (<http://tm.life.nthu.edu.tw/>). Kết quả, phần mềm sẽ đưa ra giá trị T_m của từng trình tự ở ba mức là trên 65°C, 55 - 65°C và dưới 55°C. Như vậy, nếu trình tự có $T_m < 55^\circ\text{C}$ thì enzyme này có khả năng bị biến tính ở nhiệt độ dưới 55°C. Enzyme có $T_m = 55 - 65^\circ\text{C}$ thì sẽ bị biến tính ở khoảng nhiệt độ này. Enzyme có thể hoạt động tốt ở nhiệt độ dưới 55°C.

2.2.5. Phương pháp xác định cấu trúc module của các beta-glucosidase

Hiện nay, có nhiều công cụ trợ giúp nghiên cứu cấu trúc module của các enzyme, trong đó đầu tiên, để dự đoán vùng bảo thủ các trình tự beta-glucosidase, các trình tự mã hóa enzyme được nhập vào phần mềm BLASTP [84]. Tiếp đó, các phần mềm bao gồm Blast2go [85], [86], INTERPRO [87] có tích hợp HMMER được sử dụng để khảo sát vị trí chính xác của các vùng module trên phân tử. Một khi trình tự được đưa vào phần mềm, dựa trên nguyên tắc tương đồng trình tự, các vùng module và vị trí module sẽ được phát hiện, đánh dấu và đưa ra kết quả cho từng vùng. Dựa trên các kết quả này, các trình tự có cấu trúc module tương đồng nhau được trích vào một nhóm. Các cấu trúc module của các trình tự được vẽ bằng Excel 2010. Ngoài ra, các vùng module của mỗi phân tử có thể được khảo sát dựa trên phần mềm

trực tuyến theo đường link

<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>.

2.2.6. Phương pháp dự đoán cấu trúc không gian của các beta-glucosidase

Một số công cụ dự đoán cấu trúc không gian của enzyme 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>) 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% [88], [89]. Ngoài ra, các trình tự mới còn được ước đoán cấu trúc phân tử dựa trên Swiss-Prot theo đường link <https://swissmodel.expasy.org/interactive/tzvMNY/models>.

2.2.7. Phương pháp xác định hoạt tính beta-glucosidase trên cơ chất esculin trên đĩa thạch

Beta-glucosidase có khả năng phân cắt esculin thành glucose và esculetin. Trong môi trường có sắt, ion sắt sẽ bị esculetin khử tạo màu nâu hoặc nâu thẫm. 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 [90]. Đĩa thạch môi trường LBA có bổ sung thêm esculin (3 g/l) và ferric ammonium citrate (0,2 g/l) được chuẩn bị sẵn. Sau đó, dùng dụng cụ tạo các giếng có đường kính khoảng 0,5 cm trên đĩa thạch. Hút 50 µl protein tổng số pha tan nhỏ vào một giếng. Sau đó, đĩa được ủ ở 37°C trong khoảng 16-20 giờ. Đối chứng âm là 50 µl đệm PBS 50 mM, pH 7 và đối chứng dương là 50 µl cellulase (Sigma, 0,05 U).

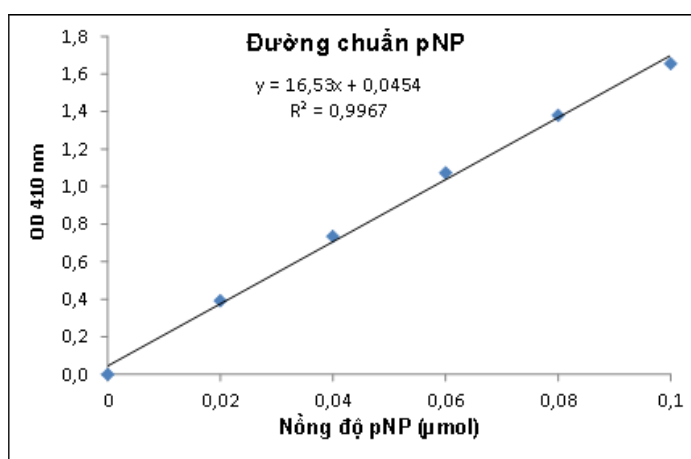
2.2.8. Phương pháp xác định ảnh hưởng của pH, nhiệt độ lên hoạt tính

beta-glucosidase

Để đánh giá sơ bộ được độ chính xác của các công cụ ước đoán nhiệt độ nóng chảy hay pH hoạt động của enzyme, beta-glucosidase GL0036730 đã được phòng Kỹ thuật di truyền biểu hiện, tinh chế được sử dụng làm nguyên liệu cho nghiên cứu xác định ảnh hưởng của pH nhiệt độ lên hoạt tính enzyme và so sánh đặc điểm được ước đoán và trên thực nghiệm.

Hoạt độ beta-glucosidase tái tổ hợp được xác định trên cơ chất 4-nitrophenyl beta-D-glucopyranoside (pNPG). Dưới tác dụng của enzyme, cơ chất sẽ giải phóng ra một phân tử đường và p-nitrophenol hấp thụ bước sóng 410 nm. Vì vậy, để đo được hoạt độ enzyme, đầu tiên đường chuẩn pNP đã được xây dựng.

Xây dựng đường chuẩn p-nitrophenol (pNP): Pha loãng chất chuẩn pNP (code 1048, Sigma) từ nồng độ 1 $\mu\text{mol/ml}$ bằng đệm PBS 20 mM, pH 7 về các nồng độ từ 0 - 0,02 - 0,04 - 0,06 - 0,08 - 0,1 μmol trong tổng thể tích là 200 μl (mỗi nồng độ lặp lại 3 lần để lấy giá trị trung bình). Tiếp theo, bổ sung 800 μl dung dịch Na_2CO_3 0,2 M vào mỗi ống, trộn đều. Sau đó đo độ hấp phụ trong các ống ở bước sóng 410 nm. Lập đồ thị thể hiện mối tương quan tuyến tính giữa nồng độ pNP và cường độ màu đo được ở bước sóng 410 nm (Hình 2.2).



Hình 2.2: Đường chuẩn biểu thị sự phụ thuộc giữa nồng độ pNP với cường độ màu đo được ở bước sóng 410 nm

Xác định hoạt độ beta-glucosidase: Để xác định hoạt độ beta-glucosidase trên cơ chất pNPG, phản ứng giữa enzyme và cơ chất được thực hiện trong tổng thể tích 200 μl bao gồm 20 μl enzyme tổng số pha tan (pha loãng 12-30 lần bằng

đệm PBS 20 mM có pH khác nhau, pH của dung dịch được chuẩn bằng HCl); 180 μ l pNPG 5 mM. Ống đối chứng có thành phần tương tự ống phản ứng, chỉ thay 20 μ l enzyme bằng 20 μ l đệm PBS 20 mM, pH 5,5. Phản ứng được ủ ở dải nhiệt độ khác nhau trong 15 phút, sau đó dừng phản ứng bằng cách bổ sung 800 μ l Na_2CO_3 0,2M. Màu sau phản ứng được đo ở bước sóng 410 nm. Mỗi mẫu được làm lặp lại ba lần. *Một đơn vị hoạt tính beta-glucosidase được định nghĩa là lượng enzyme cần thiết để giải phóng u 1 phút ở điều kiện thí nghiệm.*

2.2.9. Phương pháp xử lý số liệu và xác định mối liên quan đặc trưng của beta-glucosidase

Đặc điểm của enzyme trong các bộ dữ liệu sẽ được phân tích xu thế, mối liên quan dựa trên phần mềm SPSS. Sự khác biệt được kiểm tra bằng các thông số thống kê T-test và mô hình hóa mối tương quan bằng biểu đồ trong Microsoft Excel.

Chương 3. KẾT QUẢ VÀ THẢO LUẬN

3.1. NGHIÊN CỨU KHAI THÁC VÀ ƯỚC ĐOÁN MỘT SỐ TÍNH CHẤT CỦA ENZYME BETA-GLUCOSIDASE

3.1.1. Nghiên cứu khai thác các gen mã hóa beta-glucosidase

Theo nghiên cứu trước, đa hệ gen vi khuẩn trong dạ cỏ dê Cỏ và dê Bách Thảo đã được giải trình tự với dung lượng 8,46 Gb và 48,66. Các gene trong hai bộ dữ liệu đã được chú giải chức năng dựa trên dữ liệu KEGG, CAZy. Cả hai bộ dữ liệu này được sử dụng để khai thác các gen mã hóa beta-glucosidase, nghiên cứu đặc điểm của enzyme dựa trên công cụ tin sinh học và so sánh đặc điểm enzyme từ hai bộ dữ liệu

Bảng 3.1: Số lượng gen mã hóa beta-glucosidase đã được khai thác từ hai bộ dữ liệu dựa trên cơ sở dữ liệu KEGG, CAZy

Đặc điểm	Bộ dữ liệu nhỏ		Bộ dữ liệu lớn	
	Số lượng	Tỷ lệ (%)	Số lượng	Tỷ lệ (%)
Kích thước dữ liệu (Gb)	8,46		48,66	
Tổng số gen	164644		5367270	
Tổng số gen mã hóa beta-glucosidase	387	100	10444	100
Tổng số gene hoàn thiện	131	33,9	961	9,2
Tổng số gene thiếu 3'	88	22,7	3001	28,7
Tổng số gene thiếu 5'	60	15,5	1780	17,0
Tổng số gene thiếu cả 2 đầu	103	26,6	4702	45,0

Từ kết quả bảng 3,1 cho thấy ở bộ dữ liệu nhỏ, 164444 gen đã được nhận biết, trong đó có 387 trình tự mã hóa enzyme beta-glucosidase (chiếm 0,23% tổng số gen). Các gen mã hóa enzyme được chia thành 4 nhóm là : (1) trình tự hoàn thiện, (2) trình tự thiếu 3', (3) trình tự thiếu 5', (4) trình tự thiếu cả 2 đầu. Kết quả đánh giá cho thấy trong số 387 trình tự mã hóa enzyme thu được, có 131/387 trình tự hoàn thiện (chiếm tỉ lệ 33,9%), đây là nhóm có số lượng trình tự lớn nhất. Tiếp đến lần lượt là các nhóm thiếu cả 2 đầu (103/387 trình tự - 26,6%), thiếu đầu 3' (88/387 trình tự - 22,7%) và nhóm thiếu đầu 5'

(60/387 trình tự - 15,5%) là nhóm có số lượng trình tự nhỏ nhất trong bộ dữ liệu nhỏ thu được.

Đánh giá tương tự với bộ dữ liệu lớn với dung lượng dữ liệu 48,66 Gb kết quả cho thấy bộ dữ liệu thu được 5367270 gen. Trong đó, có 10444 gen mã hóa cho enzyme beta-glucosidase chiếm tỉ lệ 0,19%. trong đó nhóm có số lượng trình tự gen lớn nhất là nhóm trình tự có cấu trúc thiếu cả 2 đầu (4702/10444 trình tự - 45,0%), tiếp đến lần lượt là nhóm thiếu đầu 3' (3001/10444 trình tự - 28,7%), nhóm thiếu đầu 5' (1780/10444 trình tự - 17,0%) và nhóm có số lượng trình tự thấp nhất là nhóm trình tự có cấu trúc hoàn thiện với 961/10444 trình tự, chiếm tỉ lệ 9,2% tổng số trình tự enzyme mã hóa thu được.

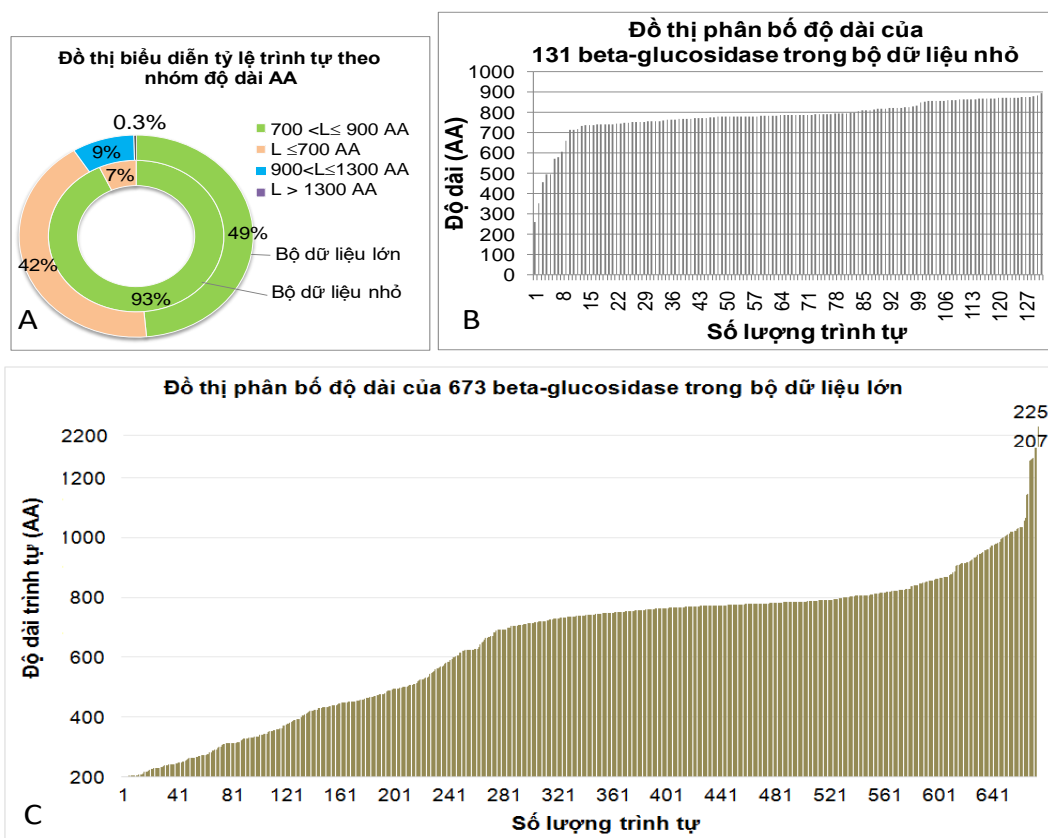
So sánh kết quả khai thác dữ liệu từ 2 bộ dữ liệu trình tự cho thấy mặc dù được giải mã sâu với dung lượng lớn gấp 5,75 lần bộ dữ liệu nhỏ nhưng tỉ lệ gen mã hóa cho enzyme beta-glucosidase thu được là thấp hơn (0,19%) trong khi bộ dữ liệu nhỏ thu được là 0,23%. Tỷ lệ này không chênh nhau nhiều và sai khác không có ý nghĩa thống kê nhưng điểm khác nhau đặc biệt đó là tỷ lệ gen mã hóa BGL hoàn thiện ở bộ dữ liệu nhỏ cao hơn hẳn (bộ dữ liệu nhỏ, tỷ lệ gen hoàn thiện chiếm 33,9% trong khi bộ dữ liệu lớn tỷ lệ gen hoàn thiện chỉ chiếm có 9,2%). Mặc dù theo báo cáo cả hai bộ dữ liệu đều được giải trên hệ thống Illumina Hiseq 2500 nên độ dài đoạn đọc có thể không khác nhau nhưng khả năng lắp ráp tin sinh học có thể khác nhau dẫn đến tỷ lệ gen hoàn thiện được lắp ráp khác nhau.

Dựa trên kết quả chú giải gen trong KEGG/CAZy, trong số 961 trình tự hoàn thiện mã hóa beta-glucosidase thu được từ bộ dữ liệu lớn có tới 206 trình tự không tìm thấy vùng xúc tác; 8 trình tự có vùng xúc tác không liên quan tới hoạt tính beta-glucosidase, 74 trình tự có kích thước ngắn dưới 200 amino acid không phù hợp với kích thước thông thường của các beta-glucosidase đã nghiên cứu. Do vậy, bộ dữ liệu lớn chỉ còn 673 trình tự beta-glucosidase được đưa vào nghiên cứu các đặc điểm của enzyme cũng như cấu trúc module.

3.1.2. Nghiên cứu sự phân bố độ dài của các trình tự beta-glucosidase

Tổng số 131 trình tự hoàn thiện mã hóa enzyme beta glucosidase từ bộ

dữ liệu nhỏ và 673 trình tự trình tự hoàn thiện từ bộ dữ liệu lớn được phân tích đánh giá theo chỉ tiêu về độ dài trình tự thu được (Hình 3,1). Từ kết quả này cho thấy beta-glucosidase từ bộ dữ liệu nhỏ có độ dài ngắn nhất là 257 amino acid và dài nhất là 894 amino acid. Trong đó, 122/131 trình tự (chiếm 93%) trình tự có độ dài từ 700 đến 900 amino acid. Trong khi đó với bộ dữ liệu lớn, trong 673 trình tự hoàn chỉnh, chiều dài trình tự có chiều dài ngắn nhất là 200 amino acid và chiều dài lớn nhất là 2253 amino acid. Khi phân chia theo nhóm, nhóm có chiều dài 700-900 amino acid có 329 trình tự (chiếm 49,0%) là nhóm có số lượng trình tự lớn nhất, tiếp đến là nhóm trình tự có độ dài 200-700 amino acid với 283 trình tự (chiếm tỉ lệ 42%), tiếp theo là nhóm có độ dài trình tự có độ dài 900-1300 amino acid với 59 trình tự chiếm tỉ lệ 9,0% số trình tự. Nhóm có số lượng trình tự nhỏ nhất là nhóm có độ dài trình tự lớn hơn 1300 amino acid có 2 trình tự chiếm 0,3%.



Hình 3.1: Phân bố độ dài các trình tự beta-glucosidase trong hai bộ dữ liệu.

A: Đồ thị biểu diễn tỉ lệ nhóm độ dài trình tự trong hai bộ dữ liệu; B: Đồ thị phân bố độ dài trình tự của 131 beta-glucosidase của bộ dữ liệu nhỏ; C: Đồ thị phân bố độ dài của 673 beta-glucosidase của bộ dữ liệu lớn; L: Độ dài trình tự.

Như vậy, có thể thấy ở bộ dữ liệu lớn do có số lượng trình tự hoàn thiện lớn hơn khoảng 5 lần nên độ dài của các trình tự thu được cũng có sự đa dạng hơn, xuất hiện nhiều nhóm và phân bố giá trị đa dạng hơn. Nghiên cứu khảo sát đa dạng trình tự beta-glucosidase của vi khuẩn trên rạn san hô cũng cho thấy kích thước trình tự beta-glucosidase trên 500 amino acid [91]. Enzyme beta-glucosidase của vi khuẩn *Halothermothrix orenii* được ứng dụng trong công nghiệp cũng có kích thước từ hơn 400 amino acid đến khoảng 750 amino acid [92]. Các enzyme thuộc họ GH3 thường có kích thước dài hơn so với các enzyme thuộc họ GH khác [91]. Do vậy, có thể dự đoán trong bộ dữ liệu lớn, số lượng các trình tự thuộc họ GH3 chiếm tỷ lệ cao.

3.1.3. Nghiên cứu phân bố điểm đẳng điện của các trình tự beta-glucosidase

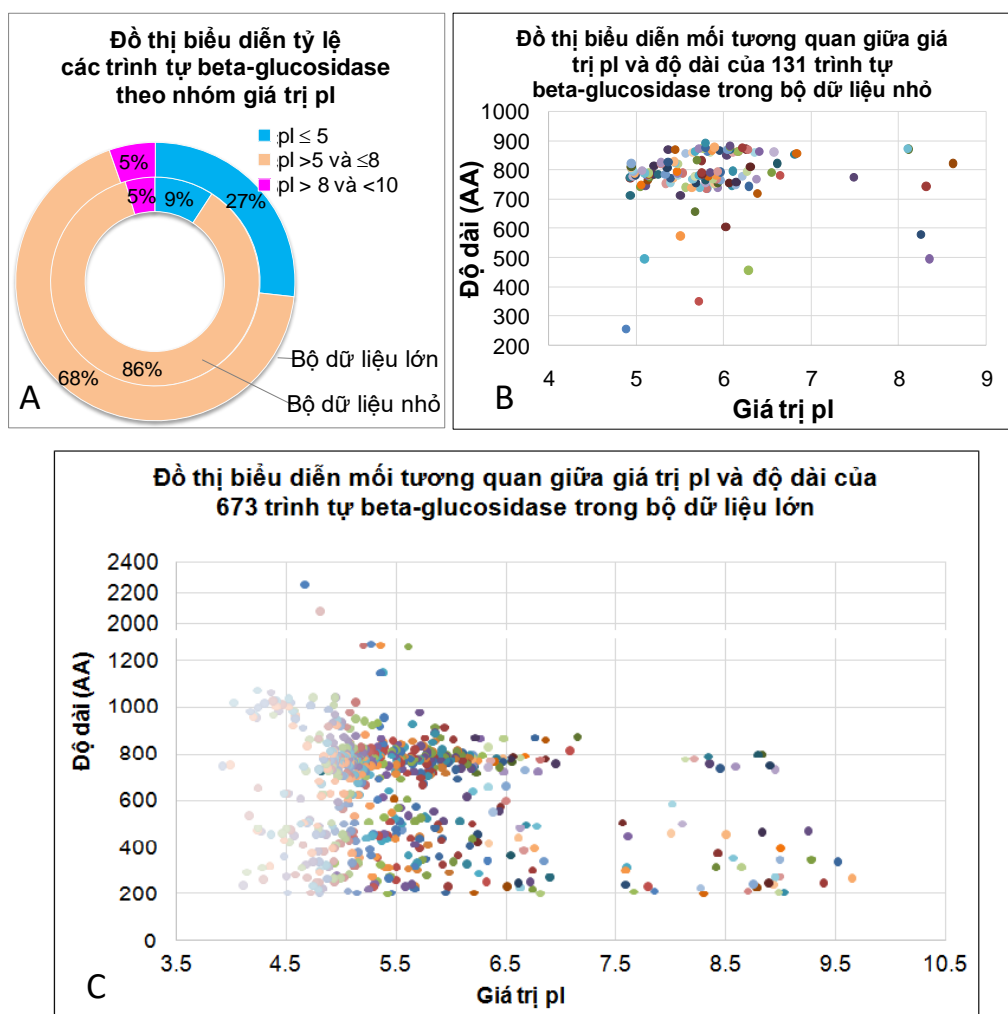
Điểm đẳng điện của enzyme có ý nghĩa lớn trong các nghiên cứu về hóa sinh học enzyme, chúng giúp xác định được khoảng pH mà ở đó enzyme tan hay enzyme ở dạng tinh thể, giúp tạo enzyme dạng tan có hoạt tính sinh học. Tại điểm đẳng điện, enzyme sẽ ở trạng thái không tích điện do vậy enzyme ở trạng thái không tan. Dựa trên điểm đẳng điện, các enzyme có thể được phân tách dễ dàng ra khỏi phức hợp. Điểm đẳng điện giúp các nhà hóa sinh học tìm được các điều kiện để tinh chế enzyme, cố định, bao gói enzyme.

Kết quả hình 3.2 ước đoán, khảo sát điểm đẳng điện của các beta-glucosidase của vi khuẩn trong dạ cỏ dê cho thấy, trong số 131 trình tự của bộ dữ liệu nhỏ thu được, số lượng trình tự trình tự có pI 5,0-8,0 chiếm số lượng lớn nhất với 113/131 trình tự chiếm tỉ lệ 86%. Nhóm pI \leq 5,0 có 12/131 trình tự (chiếm tỉ lệ 9%), còn lại là nhóm có giá trị 8,0 < pI < 10,0 với 6/131 chiếm tỉ lệ 5% số trình tự.

Đánh giá tương tự, trong số 673 trình tự beta-glucosidase trong bộ dữ liệu lớn, phần lớn các trình tự có số pI 5,0 < pI \leq 8,0 (chiếm 68% tổng số trình tự) thấp hơn so với tỷ lệ trình tự nhóm này trong bộ dữ liệu nhỏ. Số lượng trình tự có pI \leq 5,0 chiếm tỉ lệ 27% số lượng trình tự hoàn chỉnh và nhóm trình tự còn lại với giá trị 8,0 < pI \leq 10,0 chiếm tỉ lệ 5% tương đương với bộ dữ liệu nhỏ.

Ở dạ cỏ của động vật nhai lại nói chung, trong đó có dạ cỏ dê pH thường trong phạm vi 5,6-5,8 và phụ thuộc nhiều vào chế độ ăn do các sản phẩm phụ

được tạo ra trong quá trình tiêu hóa thức ăn. Dê được ăn thức ăn có chứa tannin sẽ có pH dạ cỏ cao hơn (khoảng 6,5) còn dê ăn chế độ thông thường có pH dao động dưới 6 [93]. Beta-Glucosidase trong dạ cỏ dê có chỉ số pI phân bố rộng, đặc biệt được nhìn thấy trong bộ dữ liệu lớn. Tuy nhiên, cũng có một phần đáng kể enzyme có pI gần với pH trong dạ cỏ dê. Nguồn enzyme này không chỉ là enzyme được tiết ra trong dạ cỏ để tiêu hóa cellulose mà chắc chắn có nhiều enzyme nội bào của các vi khuẩn cộng sinh giúp phân giải chất dinh dưỡng thành đường cho chính vi sinh vật sinh trưởng.



Hình 3.2: Phân bố pI của các trình tự beta-glucosidase trong hai bộ dữ liệu tương quan với độ dài trình tự.

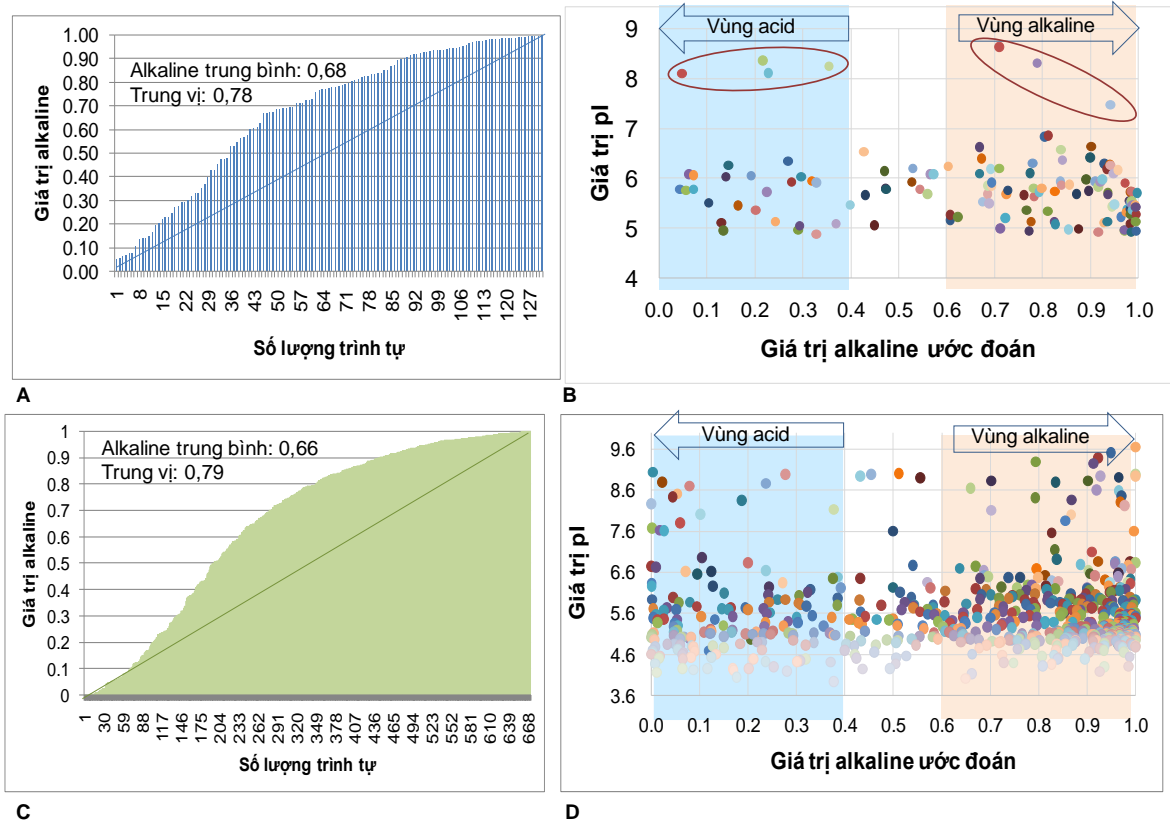
A: Đồ thị biểu diễn tỷ lệ các trình tự beta-glucosidase theo nhóm giá trị pI; B, C: Đồ thị biểu diễn mối tương quan giữa giá trị pI và độ dài của 131 trình tự beta-glucosidase trong bộ dữ liệu nhỏ và 673 trình tự beta-glucosidase trong bộ dữ liệu lớn

Từ kết quả hình 3.2 cho thấy nếu xét về mối tương quan giữa pI và độ dài của trình tự mã hóa enzyme thì đây là hai đại lượng độc lập, không phụ thuộc vào nhau. Do mức độ hạn chế của bộ dữ liệu nhỏ, chúng tôi không nhìn thấy xu hướng mối liên hệ phụ thuộc giữa pI và độ dài trình tự nhưng trong bộ dữ liệu lớn, có thể nhìn thấy vùng enzyme có pI lớn hơn 7,5 gồm các trình tự có kích thước ≤ 800 amino acid, kết quả thể hiện ở hình 3.2 C. Các trình tự lớn hơn 800 amino acid và các trình tự dài hơn thì có xu hướng có giá trị pI thấp hơn.

3.1.4. Ước đoán khoảng pH thích hợp cho hoạt động của beta-glucosidase

Phần mềm AcalPred ước đoán tương đối enzyme thuộc loại alkaline hoạt động ở vùng pH kiềm hay thuộc loại acid hoạt động ở vùng pH acid. Chỉ số alkaline càng gần về 1 xác suất chắc chắn enzyme hoạt động ở vùng pH kiềm càng cao và ngược lại chỉ số alkaline càng gần về 0 thì xác suất chắc chắn enzyme hoạt động ở vùng pH thấp càng cao. Dựa trên giá trị alkaline ước đoán của trình tự bằng phần mềm AcalPred cho thấy các enzyme beta-glucosidase trong dạ có 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 không chắc chắn hay vùng enzyme có thể hoạt động ở pH trung tính (lớn hơn 0,4 và nhỏ hơn 0,6) và vùng kiềm (từ 0,6 đến 1,0), thì 131 trình tự hoàn chỉnh thu được từ bộ dữ liệu nhỏ có 30/131 trình tự enzyme hoạt động ở vùng axit, 12/131 trình tự enzyme hoạt động ở vùng trung tính và 89/131 trình tự enzyme (chiếm 67,9%) hoạt động ở vùng pH kiềm. Nếu tính theo thống kê, chỉ số chắc chắn enzyme hoạt động tốt ở vùng pH kiềm thì trong bộ dữ liệu có 3 enzyme có chỉ số alkaline bằng 1, 11 enzyme có chỉ số alkaline lớn hơn hoặc bằng 0,99 đạt tỷ lệ 8,4%. Số enzyme có chỉ số alkaline nhỏ hơn 0,1 là 6 enzyme. Như vậy, tính tổng thể thì số lượng enzyme của bộ dữ liệu chắc chắn là enzyme hoạt động ở vùng pH kiềm hay pH axit là không cao. Kết quả này tương ứng với kết quả đánh giá giá trị trung bình và trung vị của chỉ số alkaline trong toàn bộ dữ liệu. Giá trị alkaline ước tính của trình tự trung bình đạt 0,68, giá trị trung vị ở ngưỡng 0,78 (Hình 3.3 A). Như vậy, phần lớn các beta-glucosidase là không ước đoán chắc chắn được có thể hoạt động được tốt ở vùng pH kiềm hay pH axit. Các trình tự

được ước đoán có giá trị alkaline đạt 0,8 thì mức độ chính xác của ước đoán mới đạt 80% và vẫn có 20% khả năng là enzyme không hoạt động tốt ở vùng pH kiềm.



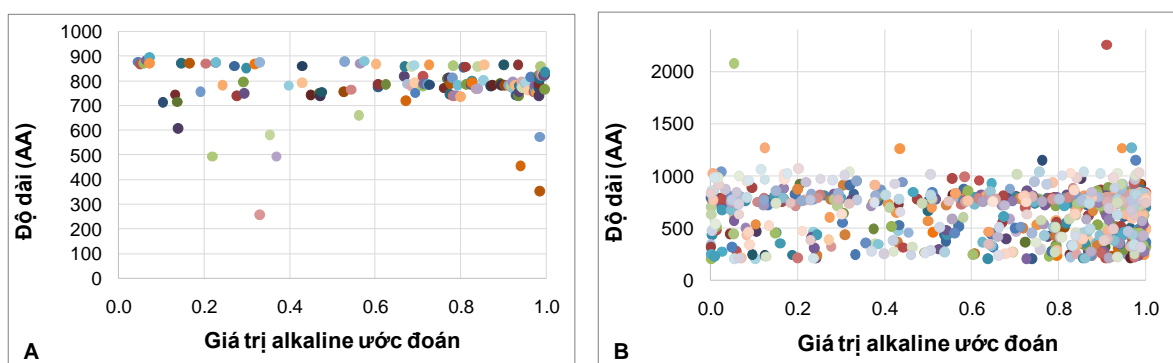
Hình 3.3: Phân bố giá trị alkaline của các trình tự beta-glucosidase trong hai bộ dữ liệu.

A, C: Đồ thị phân bố giá trị alkaline ước đoán của các trình tự beta-glucosidase từ bộ dữ liệu nhỏ và bộ dữ liệu lớn tương ứng; B,D: Phân bố giá trị alkaline của các trình tự tương ứng trong bộ dữ liệu nhỏ và bộ dữ liệu lớn trong mỗi quan hệ với chỉ số pI.

Từ kết quả hình 3.3 cho thấy với bộ dữ liệu lớn, trong số 673 trình tự hoàn thiện được đánh giá, nhóm trình tự có chỉ số alkaline thấp hơn hoặc bằng 0,4 có 168 trình tự, chiếm tỉ lệ 24,96%, nhóm trình tự có chỉ số alkaline lớn hơn 0,4 và nhỏ hơn 0,6 có 65/673 trình tự chiếm tỉ lệ 9,65% và nhóm trình tự có chỉ số alkaline lớn hơn 0,6 có 440 trình tự chiếm tỉ lệ 65,37%. Trong số đó, 45 trình tự có chỉ số alkaline lớn hơn hoặc bằng 0,99 và 2 trình tự có chỉ số alkaline bằng 1,0 (chiếm 6,7%). Các trình tự có chỉ số alkaline dưới 0,01 có 23 trình tự. Giá trị alkaline trung bình của bộ dữ liệu lớn là 0,66

và giá trị trung vị ở ngưỡng 0,78 (Hình 3.3 C). Giá trị trung bình và giá trị trung vị của chỉ số alkaline của các trình tự beta-glucosidase trong hai bộ dữ liệu là tương đồng nhau. Điều đó cho thấy số lượng trình tự chắc chắn hoạt động trong vùng pH kiềm hay pH axit là thấp.

Kết quả khảo sát mối liên quan giữa giá trị pI và chỉ số alkaline cho thấy, ở bộ dữ liệu nhỏ có 4 trình tự có pI cao và ước đoán hoạt động trong vùng pH thấp trong đó có 1 trình tự được xác định có chỉ số alkaline dưới 0,01. Đồng thời, có 3 trình tự có pI cao và có chỉ số alkaline cao, trong đó có 2 trình tự có chỉ số pI trên 8, chỉ số alkaline 0,7-0,8. Hai trình tự này rất có khả năng hoạt động tốt ở vùng pH trung tính hoặc hơi kiềm. Một trình tự có pI > 7 và chỉ số alkaline lớn hơn 0,9 sẽ có khả năng hoạt động tốt ở vùng pH kiềm. Tương tự như vậy, trong bộ dữ liệu lớn cũng có các trình tự có pI thấp (chủ yếu nằm trong khoảng 4,6-6) có giá trị alkaline tiệm cận về 0. Ba trình tự có chỉ số pI cao trên 8 cũng được ước đoán là enzyme acid với chỉ số alkaline xấp xỉ bằng 0 (Hình 3.3D). Bên cạnh đó cũng có số lượng lớn các trình tự có pI trong khoảng 4,6-6 và có giá trị alkaline tiệm cận về 1, ba trình tự có pI cao được ước đoán với giá trị alkaline bằng 1. Các enzyme này có khả năng hoạt động tốt ở pH alkaline. Thông thường, enzyme sẽ ở trạng thái tan trong môi trường có pH chênh lệch so với pI 2 đơn vị. pI được cho rằng không có liên quan tới pH tối ưu cho hoạt tính enzyme hoặc độ bền của enzyme trong môi trường pH nhưng giá trị pI có giá trị trong việc xác định enzyme ở trạng thái tan để có hoạt tính sinh học [94].



Hình 3.4: Mối quan hệ giữa giá trị alkaline ước đoán và độ dài trình tự.

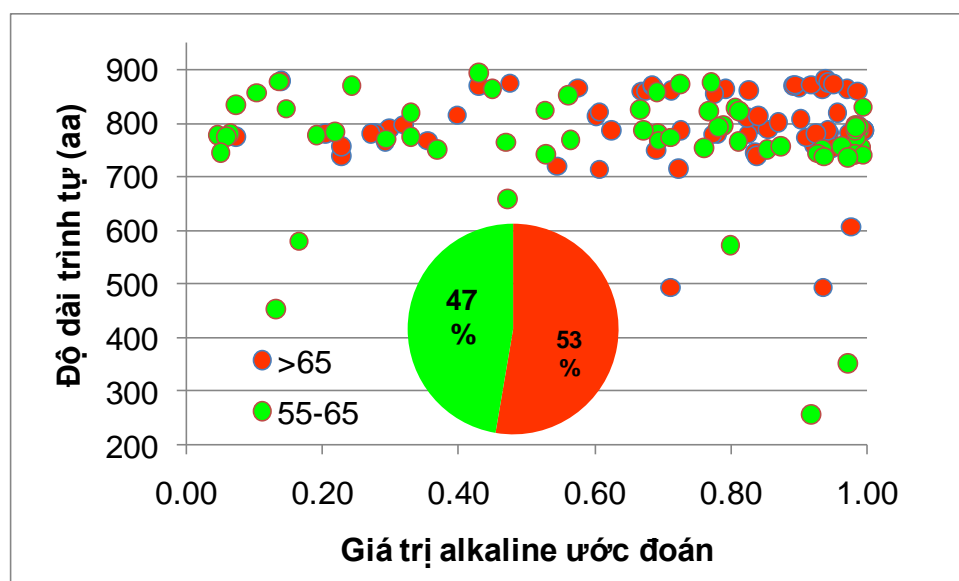
A: Bộ dữ liệu nhỏ; B: Bộ dữ liệu lớn

Kết quả khảo sát giá trị alkaline của các trình tự với độ dài trình tự chúng tôi cũng nhận thấy hai giá trị này là hai biến độc lập, không có liên quan đến nhau (Hình 3.4). Các trình tự được phân bố đều trong các vùng của đồ thị. Ở bộ dữ liệu lớn, hai trình tự có độ dài lớn nhất thì một trình tự có chỉ số alkaline ở vùng cao và 1 trình tự có chỉ số alkaline ở vùng thấp (Hình 3,4 B).

3.1.5. Ước đoán khả năng bền nhiệt của beta-glucosidase

Dựa trên phần mềm TBI (www.tbi.org.tw/tools), nhiệt độ nóng chảy T_m của các trình tự beta-glucosidase đã được ước đoán. Chỉ số nóng chảy của trình tự gợi ý cho khả năng bền nhiệt của enzyme, trong đó ở vùng T_m , enzyme bị biến tính và sẽ mất hoạt tính. Như vậy, chỉ số T_m càng cao thì khả năng enzyme bền nhiệt càng tốt. Chỉ số T_m không liên quan tới nhiệt độ tối ưu của enzyme nhưng gợi ý rằng nhiệt độ tối ưu cho hoạt động của enzyme phải thấp hơn chỉ số T_m .

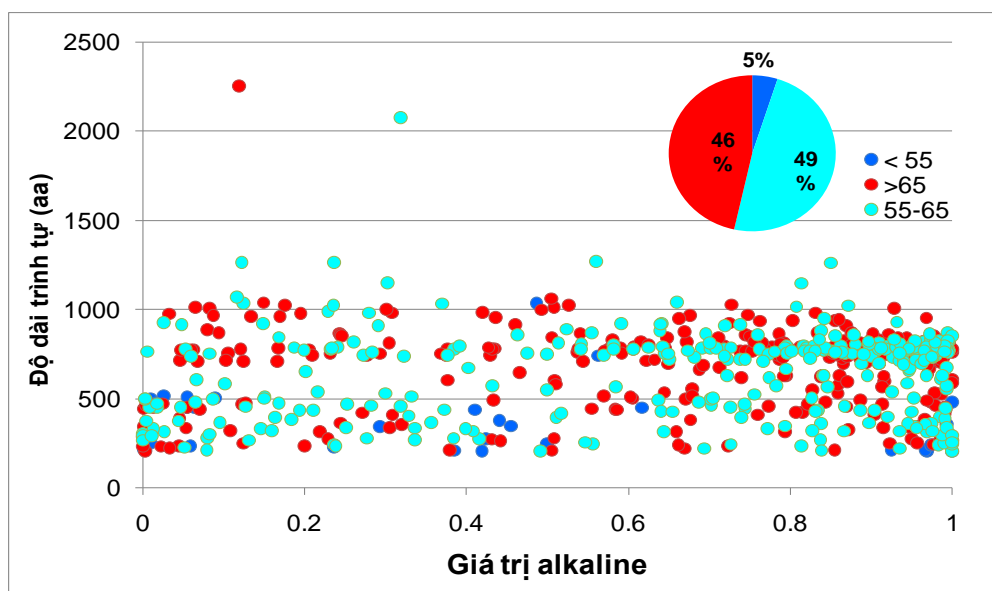
Nghiên cứu nhiệt độ T_m của các trình tự beta-glucosidase trong bộ dữ liệu nhỏ cho thấy 100% các trình tự có T_m từ 55°C trở lên và không có trình tự nào có T_m dưới 55°C. Trong đó, 53% số lượng enzyme (69 trình tự) có T_m trên 65°C, đồng nghĩa là 53% enzyme sẽ bị mất hoạt tính khi nhiệt độ tăng lên trên 65°C và 47% số enzyme còn lại bị mất hoạt tính ở nhiệt độ 55-65°C (Hình 3.5).



Hình 3.5: Phân bố nhiệt độ nóng chảy của các trình tự beta-glucosidase của bộ dữ liệu nhỏ trong mối tương quan với độ dài trình tự và giá trị alkaline.

Kết quả khảo sát mối liên hệ giữa độ dài trình tự với nhiệt độ nóng chảy của enzyme cho thấy, hai đại lượng này không có mối quan hệ với nhau (Hình 3.5). Nguyên nhân có thể do số lượng trình tự trong bộ dữ liệu nhỏ bị hạn chế và hầu hết các trình tự đều có kích thước lớn trên 700 amino acid, không có trình tự có kích thước trên 900 amino acid. Kết quả khảo sát mối liên hệ giữa nhiệt độ nóng chảy với giá trị alkaline của enzyme cũng cho thấy đây là hai biến độc lập, không có liên quan đến nhau. Các trình tự có T_m ở các vùng khác nhau nằm rải khắp các giá trị alkaline (Hình 3.5).

Với bộ dữ liệu lớn, giá trị T_m của 673 trình tự gồm đủ 3 nhóm: Nhóm có nhiệt độ nóng chảy cao $T_m > 65^\circ\text{C}$, nhóm có $T_m 55-65^\circ\text{C}$ và nhóm có T_m dưới 55°C với số lượng trình tự lần lượt theo ba nhóm là 329 (chiếm 49%), 313 (46%) và 55 trình tự (5%). So với bộ dữ liệu nhỏ, bộ dữ liệu này xuất hiện đáng kể số trình tự có nhiệt độ nóng chảy dưới 55°C và số lượng trình tự có $T_m 55-65^\circ\text{C}$ chiếm tỷ lệ cao nhất (Hình 3.6). Phân tích thống kê trong bộ dữ liệu lớn cũng nhận thấy điểm nóng chảy của enzyme không liên quan tới độ dài trình tự và không liên quan tới giá trị alkaline của các trình tự.



Hình 3.6: Phân bố nhiệt độ nóng chảy của các trình tự beta-glucosidase của bộ dữ liệu lớn trong mối tương quan với độ dài trình tự và giá trị alkaline.

3.2. NGHIÊN CỨU CẤU TRÚC MODULE CỦA BETA-GLUCOSIDASE TỪ VI KHUẨN TRONG DẠ CỎ DÊ

3.2.1. Khảo sát họ GH của beta-glucosidase

Cellulase nói chung và beta-glucosidase nói riêng tồn tại ở hai dạng là enzyme có cấu trúc module và enzyme không có cấu trúc module. Các enzyme không có cấu trúc module là phân tử enzyme chỉ có duy nhất vùng xúc tác, có hoạt tính xúc tác và không có các vùng phụ trợ khác. Các enzyme này thường phân cắt các cơ chất có cấu trúc đơn giản và các polysaccharide chuỗi ngắn. Enzyme có cấu trúc module là trong phân tử, ngoài vùng xúc tác ra enzyme còn chứa các vùng khác với các chức năng khác nhau có thể là có thêm vùng hoạt tính phụ hoặc các vùng hỗ trợ cho chức năng của enzyme. Các vùng này có cấu trúc bậc 3 tách biệt được và được gọi là các module. Các module phụ trợ thường giúp enzyme ổn định cấu trúc, ổn định vùng xúc tác enzyme-cơ chất, giúp làm tăng ái lực của enzyme với cơ chất,... Các module phụ trợ thường thấy trên các cellulose là CBM [95], FN3 [47], vùng X, nhân tố kết tập type C, Ig, ...

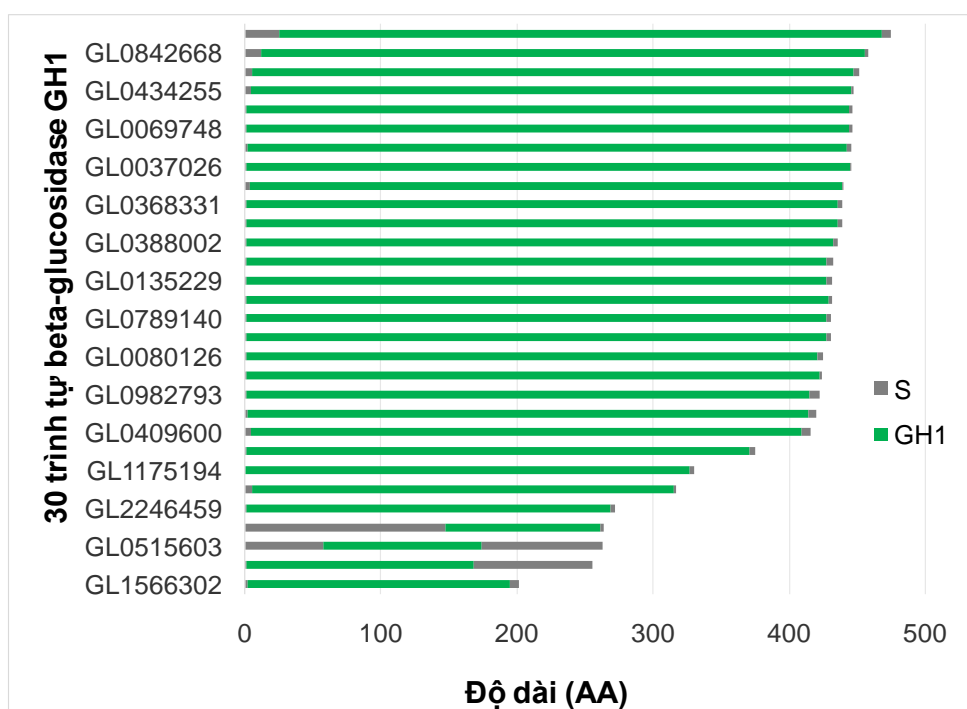
Trong bộ dữ liệu nhỏ, 131 trình tự beta-glucosidase đều được xác định thuộc họ GH3. Với số lượng trình tự beta-glucosidase lớn hơn (673 trình tự) thu được từ bộ dữ liệu lớn đã cung cấp cấu trúc module của beta-glucosidase của vi khuẩn trong dạ cỏ dê đa dạng hơn. Trong số các trình tự này, beta-glucosidase họ GH3 chiếm tỷ lệ cao nhất với 607 trình tự (chiếm 90,2%). Tiếp đến là trình tự thuộc các họ GH16, GH1 với 35 và 31 trình tự chiếm tương ứng 5,2% và 4,6% tổng số trình tự (Bảng 3.2). Beta-glucosidase từ vi khuẩn phổ biến thuộc họ GH3. Nghiên cứu beta-glucosidase của vi khuẩn từ rạn san hô cũng cho thấy enzyme chiếm ưu thế là enzyme thuộc họ GH3 và GH1 trong đó tỷ lệ GH1:GH3 là 1:1,2 [92].

Bảng 3.2: Số lượng các họ GH của beta-glucosidase của vi khuẩn trong dạ cỏ dê

Họ enzyme	Số lượng		
	Tổng số	Vùng xúc tác không đầy đủ	Vùng xúc tác hoàn thiện
GH3	607	149	458
GH16	35	0	35
GH1	31	1	30
Tổng	673	150	523

3.2.2. Cấu trúc module beta-glucosidase họ GH1

Cấu trúc module của toàn bộ 30 trình tự mã hóa enzyme GH1 trong bộ dữ liệu có vùng xúc tác hoàn thiện đã được phân tích. Từ kết quả hình 3.7 cho thấy, toàn bộ 30 trình tự đều không có module phụ trợ. Ngoài ra, từ bản đồ cấu trúc module cho thấy toàn bộ 30/30 trình tự đều không có tín hiệu tiết (SigP) trong cấu trúc module của trình tự điều này chứng minh các enzyme thuộc họ GH1 thu được từ bộ dữ liệu lớn đều là các enzyme nội bào. Độ dài của các enzyme dao động từ 202 amino acid (gen GL1566302) đến 476 amino acid (gen GL0842668).



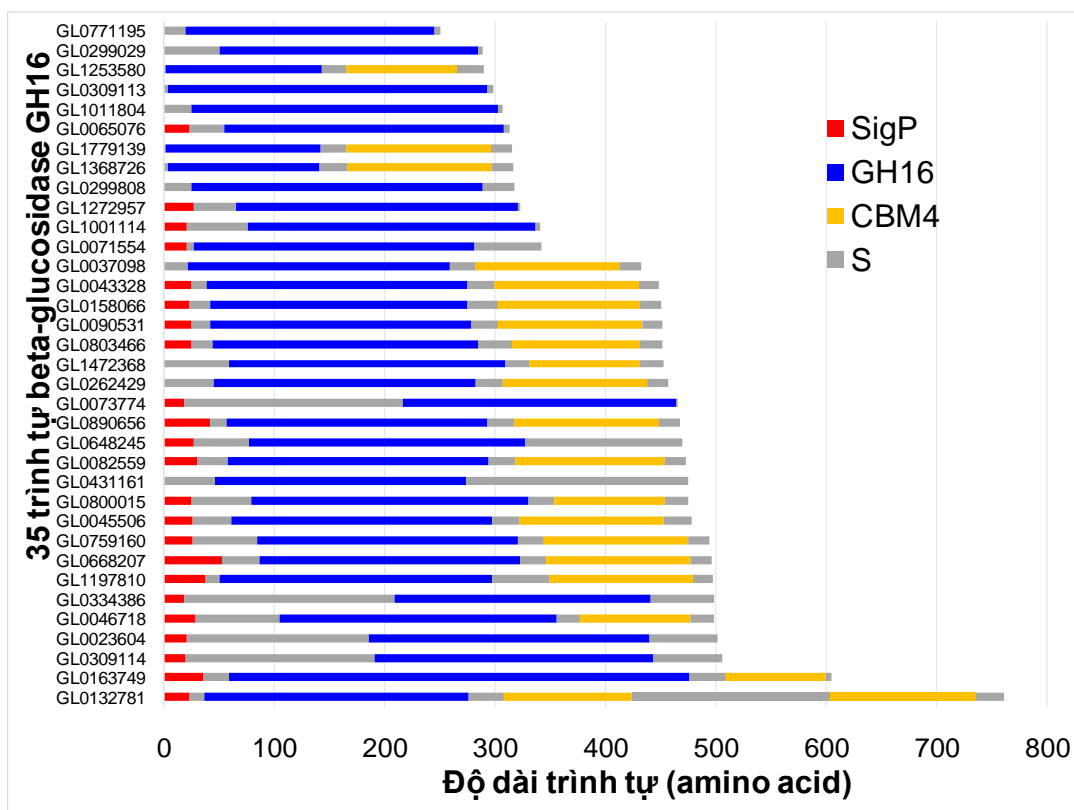
Hình 3.7: Cấu trúc module của 30 trình tự beta-glucosidase GH1.

S: vùng nội; GH1: glycoside hydrolase.

Thông thường các trình tự beta-glucosidase GH1 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. Các nghiên cứu trước đây cũng chỉ ra rằng, beta-glucosidase GH1 thường là các enzyme nội bào, không được tiết ra ngoài tế bào [96]. Vì là enzyme nội bào nên chúng thường phân cắt các có chất đơn giản, có cấu trúc ngắn để tạo glucose cung cấp nguồn đường cho vi khuẩn phát triển.

3.2.3. Cấu trúc module beta-glucosidase họ GH16

Từ kết quả hình 3.8 cho thấy tương tự như các beta-glucosidase GH1, beta-glucosidase GH16 cũng chỉ mang một vùng xúc tác thuộc họ GH16 và không có vùng xúc tác khác. Kích thước enzyme trong họ GH16 dao động trong khoảng rộng từ 250 amino acid (GL0771195) đến 761 amino acid (GL0132781). Chỉ có 2/35 trình tự có chiều dài lớn hơn 500 amino acid. Chiều dài này khá là tương đồng với nhóm trình tự nhóm GH1. Độ dài của trình tự phụ thuộc vào số lượng các module phụ trợ mà enzyme sở hữu.



Hình 3.8: Cấu trúc module của 35 trình tự beta-glucosidase GH16
 SigP: Vùng tín hiệu tiết, GH16: Glycoside hydrolase 16; CBM: Module liên kết với cơ chất carbohydrate.

Không giống như beta-glucosidase GH1, số lượng các trình tự mang vùng tín hiệu tiết chiếm tỷ lệ cao (23/35 trình tự có sự có mặt của tín hiệu tiết, chiếm 65,7% trình tự. Các trình tự mang tín hiệu tiết là dấu hiệu đặc trưng của enzyme được tiết ra khỏi màng trong của enzyme để ra khoang chu chất hoặc ra môi trường tùy theo vi khuẩn sinh enzyme là vi khuẩn Gram âm hay Gram dương. Ngoài vùng tín hiệu tiết ra, 20 trong số 35 enzyme GH16 có chứa module CBM4, chiếm 57,14%. Vùng này có độ dài 150 amino acid và được xác định là có nguồn gốc từ vi khuẩn. CBM4 được tìm thấy trong cấu trúc của các enzyme phân hủy lignocellulose, chúng có ái lực với nhiều dạng cơ chất như xylan, β -1,3-glucan, β -1,3-1,4-glucan, β -1,6-glucan và cellulose vùng vô định hình nhưng không có khả năng bám lên cellulose tinh thể. Các vùng CBM4 trong bộ dữ liệu luôn xuất hiện ở phía đầu C của enzyme, trong đó có một trình tự có mang 2 vùng CBM4 làm tăng kích thước enzyme lên trên 700 amino acid (Hình 3.8).

3.2.4. Cấu trúc module beta-glucosidase họ GH3

Beta-glucosidase GH3 là họ enzyme phong phú nhất trong cả hai bộ dữ liệu. Ở bộ dữ liệu nhỏ, 100% trình tự beta-glucosidase hoàn chỉnh thuộc họ GH3.

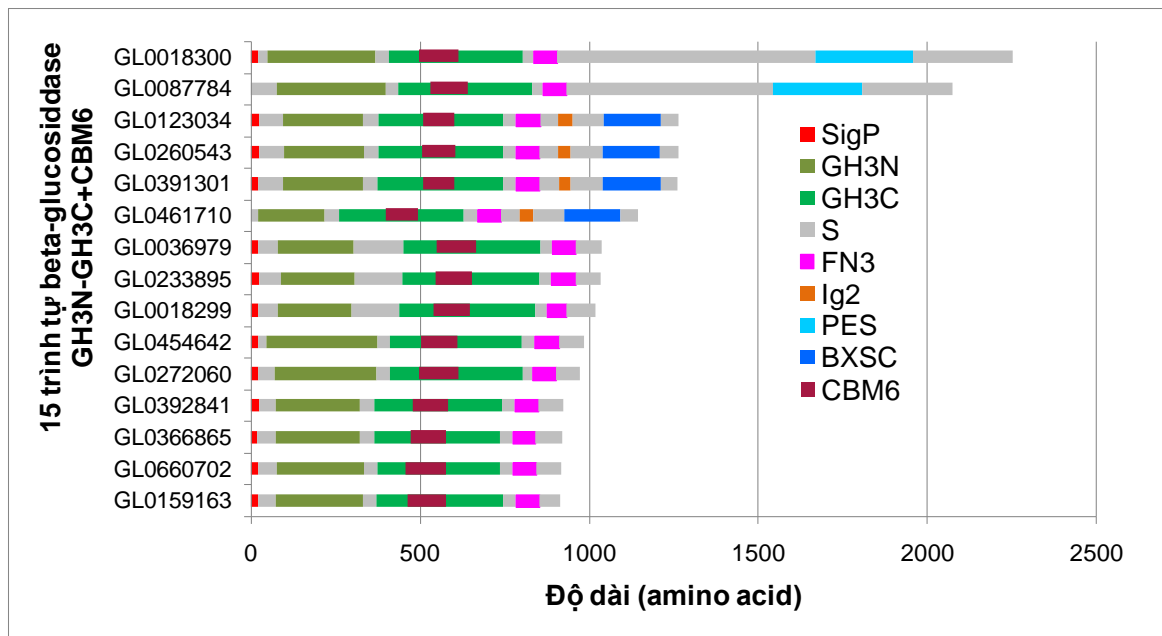
Bảng 3.3: Các dạng cấu trúc module beta-glucosidase GH3 của vi khuẩn trong dạ cỏ dê được khai thác từ bộ dữ liệu lớn

STT	Cấu trúc module	Số trình tự
1	GH3N-GH3C-FN3-CBM6	15
2	GH3N-GH3C-FN3 có PA14 nằm trên vùng GH3C	19
3	GH3N-GH3C-FN3-GH31	1
4	GH3N-GH3C-FN3-cellulase GH5	1
5	GH3N-FN3-GH3C	81
6	GH3N-GH3C	42
7	GH3N-GH3C-FN3	299
Tổng		458

Ở bộ dữ liệu lớn, enzyme thuộc họ GH3 là enzyme có độ đa dạng cao nhất. Beta-Glucosidase GH3 trong bộ dữ liệu được chia thành nhiều các nhóm trình tự cơ bản gồm nhóm có chứa vùng liên kết cơ chất CBM6 (15 trình tự); nhóm không chứa module CBM6 gồm có 592 trình tự trong đó có 41 trình tự chỉ có chứa vùng GH3N, 104 trình tự chỉ có vùng GH3C, 4 trình tự chỉ có

vùng FN3 nên số các trình tự có đủ hai vùng GH3N-GH3C còn lại là 443 trình tự. Các trình tự này (443 trình tự) được nhóm thành 6 loại cấu trúc module trong đó cấu trúc GH3N-GH3C-FN3 có tỷ lệ lớn nhất gồm 320 trình tự (chiếm 72,2 %), 42 trình tự chỉ có GH3N-GH3C (chiếm 9,4%), 81 trình tự có cấu trúc GH3C-FN3-GH3C (Bảng 3.3). Ở bộ dữ liệu nhỏ, trong số 131 trình tự beta-glucosidase GH3, 90,9% beta-glucosidase GH3 có chứa FN3 và chỉ có 9,1% không có FN3. Như vậy, tỷ lệ các trình tự có và không có vùng FN3 của hai bộ dữ liệu là tương đồng nhau. Điều này cho thấy FN3 là vùng module đặc thù của beta-glucosidase của vi khuẩn trong dạ cỏ dê.

a) Trình tự có vùng CBM6

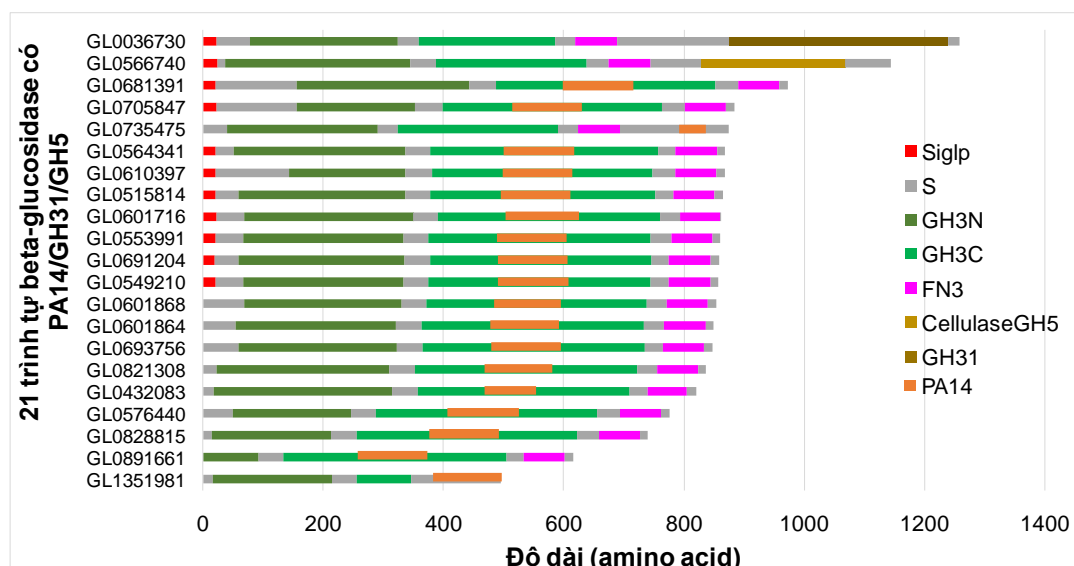


Hình 3.9: Cấu trúc module của 15 trình tự beta-glucosidase GH3 có CBM6
SigN: Tín hiệu tiết; *FN3*: Vùng giống fibronectin type III; *GH3N*, *GH3C*: Glycoside hydrolase 3 N, C; *S*: Vùng nối; *Ig*: Vùng giống immunoglobulin nhóm 2; *BXSC*: Vùng giống concavalin A tận cùng đầu C của beta xylosidase; *PES*: Pectinesterase; *CBM6*: Module liên kết carbohydrate (họ 6)

Từ 15 trình tự beta glucosidase thuộc nhóm GH3 được xác định có vùng xúc tác CMB6, các trình tự được lập bản đồ cấu trúc module như hình 3.9. Các trình tự thuộc nhóm này có chiều dài trình tự từ 914 amino acid (gen

GL0159163) đến 2253 amino acid (gen GL0018300). Các trình tự thuộc nhóm này có độ dài lớn hơn đáng kể so với các trình tự thuộc họ GH1 và GH16. Nguyên nhân là các enzyme họ GH3 thường có hai vùng GH3N (thường nằm ở tận cùng đầu N của enzyme, ngay sau vùng tín hiệu tiết) và vùng GH3C (thường nằm ở tận cùng đầu C), mỗi vùng có kích thước khoảng 350 amino acid, tạo hai cấu trúc TIM $(\alpha/\beta)_8$ và $(\alpha/\beta)_6$ kết hợp lại với nhau thành trung tâm hoạt động của enzyme. Trong số 15 trình tự, chỉ có 2 trình tự không có tín hiệu tiết. Điều đó cho thấy các BGL GH3 chứa CBM6 thường là enzyme tiết. Điều đặc biệt là CBM6 không nằm riêng rẽ ở một vùng của phân tử mà có mặt ngay trong trình tự của vùng GH3C (Hình 3.9). CBM6 gồm 120 gốc amino acid có vai trò làm tăng ái lực ở trung tâm xúc tác của enzyme với các cellulose vùng vô định hình. Chúng có thể liên kết với các cơ chất β -1,3-glucan, β -1,3-1,4-glucan, và β -1,4-glucan. Ngoài vùng CBM6 ra, nhóm enzyme này có 100% trình tự mang vùng FN3 liền kề phía đầu C của vùng GH3C. Cấu trúc FN3 đã được chứng minh giúp ổn định vùng trung tâm hoạt động của enzyme. Ngoài các vùng trên, 4/15 trình tự có thêm module BXSC và 2/15 trình tự có cấu trúc PES. Hai trình tự mang PES là hai trình tự có độ dài lớn nhất trong bộ dữ liệu. Các module PES có thể giúp enzyme có hoạt tính phụ là phân cắt pectin hoặc có chức năng làm tăng ái lực enzyme với cơ chất. Các cấu trúc concavalin A có khả năng kết dính cao, giúp gắn phân tử enzyme lên cơ chất [97]. Ngoài ra, trong số các enzyme này còn có 4 trình tự có thêm vùng Ig với chức năng chưa được nghiên cứu.

b) Trình tự beta-glucosidase GH3 có vùng PA14/GH31/GH5

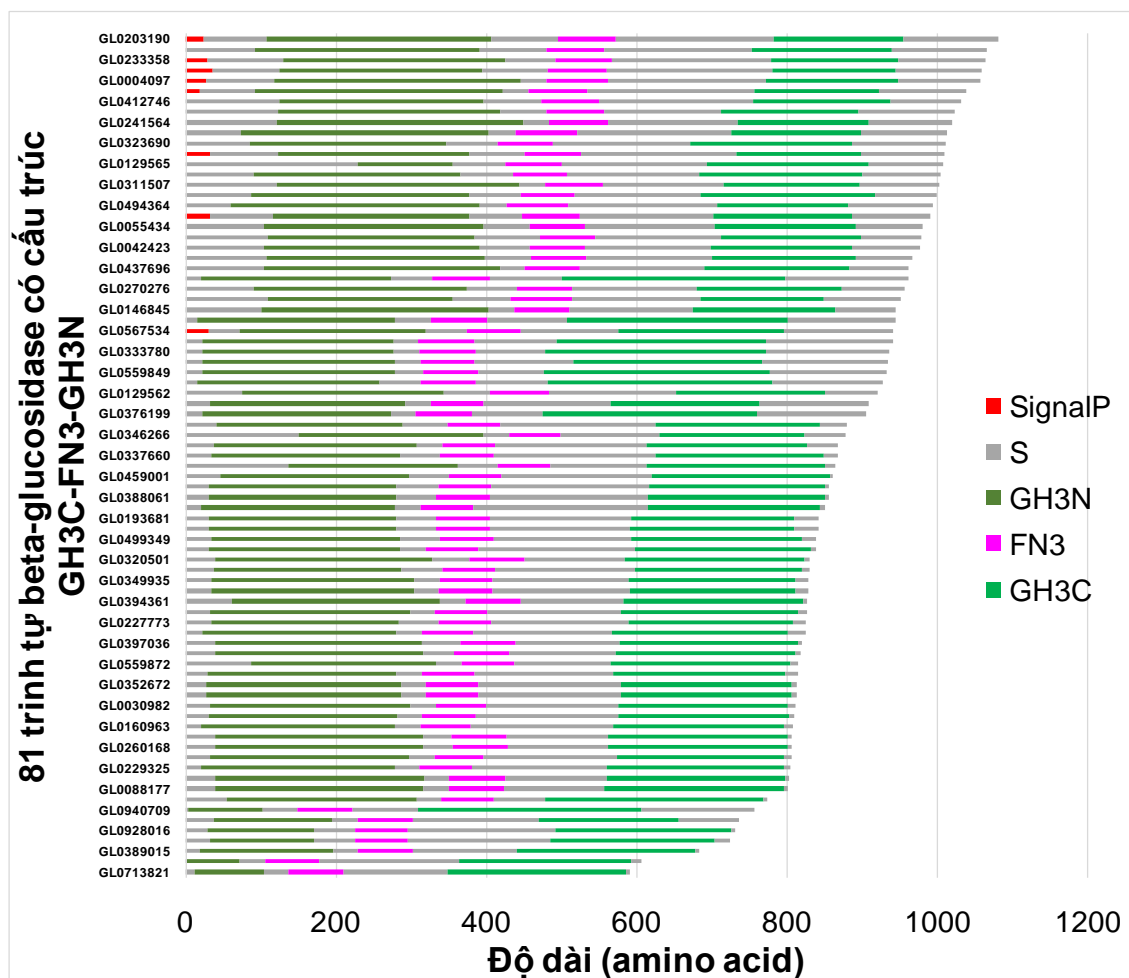


Hình 3.10: Cấu trúc module của 21 trình tự beta-glucosidase có PA14/GH31/GH5

SigN: Tín hiệu tiết; *FN3*: Vùng giống fibronectin type III; *GH3N*, *GH3C*: Glycoside hydrolase 3 N, C; *S*: Vùng nối; *GH31*: Glycoside hydrolase họ 31 có hoạt tính alpha-glucosidase; *Cellulase GH5*: Vùng xúc tác của cellulase họ GH5.

Tổng cộng 21 trình tự beta-glucosidase có thêm vùng phụ PA14/GH31/GH5 đã được đánh giá về cấu trúc module, trong đó 19 trình tự có vùng PA14, 1 trình tự có thêm vùng xúc tác của alpha-glucosidase GH31 và một trình tự có thêm vùng xúc tác của cellulase GH5. Các trình tự thuộc nhóm này có chiều dài từ 496 amino acid (gen GL1351981) đến 1,259 amino acid (gen GL0036730). Hai trình tự có hai vùng xúc tác GH5 và GH31 là hai trình tự có kích thước dài nhất và là hai enzyme được tiết ra môi trường do có chứa tín hiệu tiết (Hình 3.10). Tất cả các trình tự đều có vùng FN3 nằm sau vùng GH3C. Trong số 19 trình tự có thêm module PA14 thì có 9 trình tự có tín hiệu tiết. Tương tự như CBM6, các domain PA14 nằm ngay trong vùng GH3C. Cấu trúc PA14 này có vai trò quan trọng giúp tăng tính đặc hiệu cơ chất cho enzyme [98].

c) Trình tự có cấu trúc GH3C-FN3-GH3N



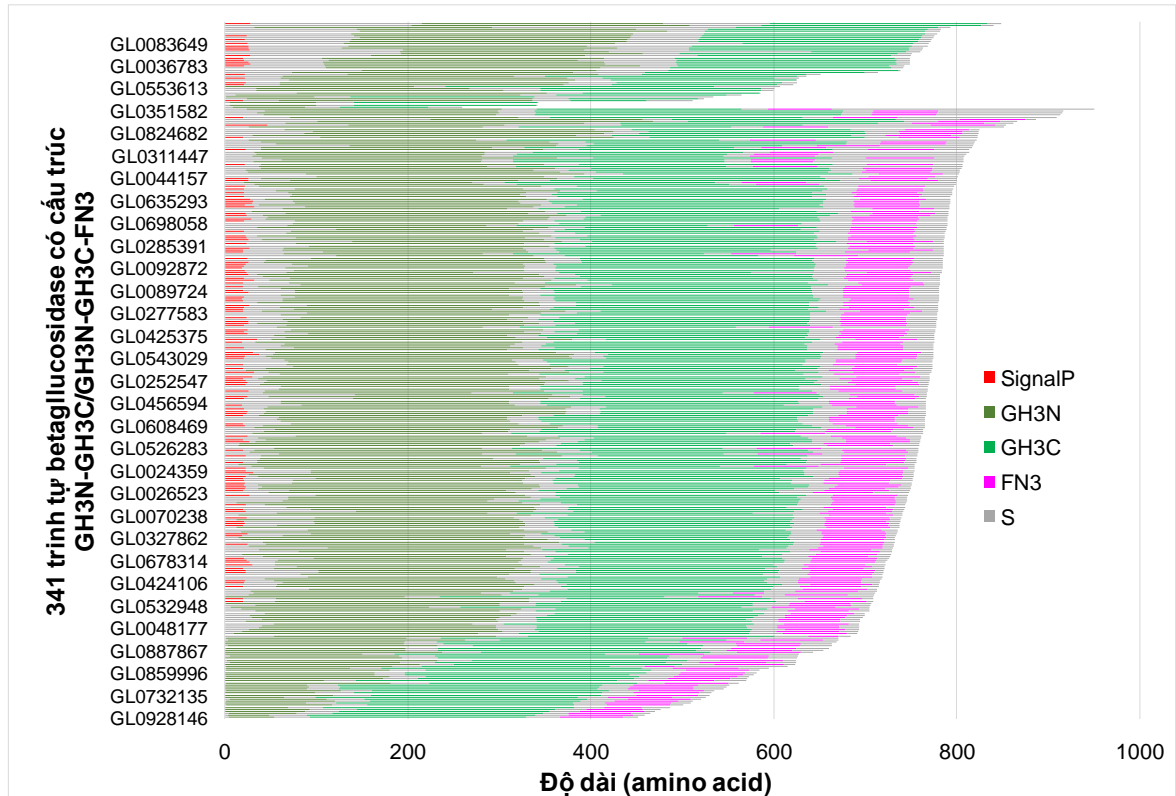
Hình 3.11: Cấu trúc module của 81 trình tự beta-glucosidase *GH3C-FN3 - GH3N*

SigN: Tín hiệu tiết; *FN3*: Vùng giống fibronectin type III; *GH3N*, *GH3C*: Glycoside hydrolase 3 N, C; *S*: Vùng nối.

Từ hình 3.11 cho thấy đây là nhóm enzyme đặc biệt mới được phát hiện trong bộ dữ liệu lớn này. Thông thường, các vùng GH3N thường ở phía tận cùng đầu N và vùng GH3C thường ở phía tận cùng đầu C, nhưng trong bộ dữ liệu này, vùng GH3C lại đứng ở tận cùng đầu N và ngược lại vùng GH3N lại ở tận cùng đầu C. Tổng số các trình tự có cấu trúc GH3N-FN3-GH3C trong bộ dữ liệu là 81 trình tự. Các trình tự thuộc nhóm này có chiều dài từ 592 amino acid (gen GL0713821) đến 1058 amino acid (gen GL0203190). Như vậy khoảng chiều dài của các trình tự GH3 nhóm này tương đồng với các enzyme cùng thuộc nhóm GH3 khác. Đánh giá chi tiết cấu trúc module cho thấy chỉ có 8/81 trình tự có vùng tín hiệu tiết. Như vậy, phần lớn các trình tự không có tín hiệu tiết và rất có thể enzyme không được tiết ra khỏi màng

trong của tế bào. Các vùng FN3 vẫn tuân theo quy luật là nằm sau vùng GH3C nên trong trường hợp này FN3 nằm xen giữa hai vùng xúc tác. Cấu trúc module của các enzyme này là khá đặc biệt nên một enzyme có tín hiệu tiết sẽ được nghiên cứu sâu hơn về dự đoán cấu trúc bậc ba.

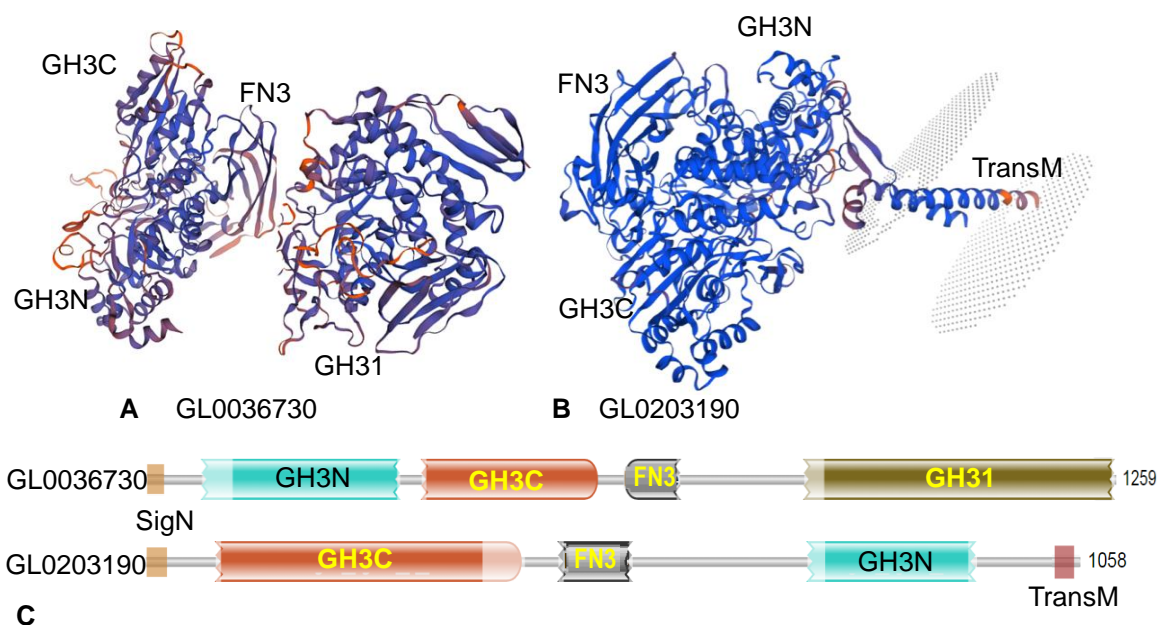
d) Trình tự có cấu trúc GH3N- GH3C- FN3



Hình 3.12: Cấu trúc module GH3N-GH3C/GH3N-GH3C-FN3 của 341 trình tự beta-glucosidase.

Các trình tự thuộc họ GH3 còn lại với 341 trình tự có hai dạng cấu trúc module là GH3N-GH3C và GH3N-GH3C-FN3, đã được đánh giá. Từ kết quả hình 3.12 cho thấy chiều dài của nhóm trình tự này từ 452 amino acid (gen GL0928146) đến 917 amino acid (gen GL0351582). Đánh giá cấu trúc module cho thấy trong số 341 trình tự này có 42 trình tự có cấu trúc GH3N-GH3C và 299 trình tự cấu trúc GH3N-GH3C-FN3. Sự có mặt của tín hiệu tiết (SigP) được xác định trên 140/341 trình tự chiếm 41,05% số trình tự nhóm này. Những enzyme mang tín hiệu này được xếp vào nhóm những enzyme ngoại bào.

e) Khảo sát cấu trúc bậc 3 của hai trình tự điển hình



Hình 3.13: Cấu trúc bậc ba và module của *GL0036730* (*GH3N-GH3C-FN3-GH31*) và *GL0203190* (*GH3C-FN3-GH3N*)

A, B: Cấu trúc bậc ba; *C:* Cấu trúc module; *FN3:* Vùng giống fibronectin 3; *TransM:* Vùng chuyển màng; *SigN:* Tín hiệu tiết

Các beta-glucosidase của vi khuẩn trong dạ cỏ dê chủ yếu thuộc họ GH3 và trong họ này cấu trúc điển hình của các trình tự là có thêm vùng phụ trợ FN3. Tuy nhiên, kết quả khảo sát cấu trúc module cho thấy FN3 luôn luôn nằm phía đầu C của vùng GH3C. Vì vậy, trong nhiều trình tự cấu trúc module thường là GH3N-GH3C-FN3 [99]. Ngược lại, có trình tự có vùng GH3C lại đứng ở tận cùng phía đầu N, ngay liền kề tín hiệu tiết của enzyme tiết, tiếp theo là FN3 và cuối cùng mới đến GH3N. Vì vậy, để chắc chắn hơn, chúng tôi đã khảo sát cấu trúc điển hình của hai trình tự là *GL0036730* (*GH3N-GH3C-FN3-GH31*) và *GL0203190* (*GH3C-FN3-GH3N*).

Kết quả từ hình 3.13 cho thấy, các vùng GH3N là vùng TIM barrel có cấu trúc $(\alpha/\beta)_8$ và vùng GH3C có cấu trúc $(\alpha/\beta)_6$ cùng góp phần tạo thành trung tâm hoạt động của enzyme, trong khi đó vùng FN3 tạo thành từ các phiến beta là vùng tương đối độc lập, vươn ra bên ngoài phân tử. Vùng FN3 thường nằm đối diện với trung tâm xúc tác và các nghiên cứu chỉ ra nó có vai trò làm ổn định vị trí xúc tác [100]. Các nghiên cứu trước chỉ ra rằng vùng GH3N mang gốc aspartate có vai trò là ái nhân, vùng GH3C mang gốc glutamate là gốc cho/nhận điện tử [100]. Các gốc này trực tiếp tấn công cơ chất tạo các liên kết cộng hóa trị với cơ chất, hình thành mạng lưới truyền

điện tử để phân giải cơ chất. Cả hai enzyme đánh giá này đều là hai enzyme tiết vì có vùng tận cùng đầu N là vùng tín hiệu tiết. Tín hiệu tiết này giúp enzyme được tiết ra khỏi màng trong của tế bào. Enzyme GL0036730 không có vị trí chuyển màng. Kết quả khảo sát cho thấy đây là enzyme được tiết hoàn toàn ra ngoài. Ngược lại, enzyme GL0203190 lại có vùng chuyển màng nằm ở phía GH3N và được gắn chặt vào màng trong của tế bào (Hình 3.13 C) giúp enzyme neo vào màng tế bào để phân giải cơ chất thành đường, cung cấp năng lượng cho tế bào. Điều này cũng nói lên, các enzyme trong tự nhiên cũng rất đa dạng, hoàn hảo, có thể thay đổi các trật tự module nhưng vẫn đảm bảo cấu trúc và chức năng cho enzyme.

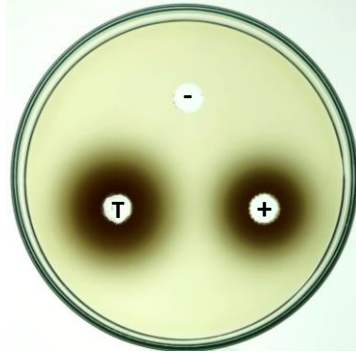
3.3. NGHIÊN CỨU CÁC ĐẶC ĐIỂM CỦA ENZYME ĐƯỢC ƯỚC ĐOÁN VÀ TRÊN THỰC NGHIỆM

Để kiểm tra sơ bộ độ tin cậy của kết quả ước đoán đặc điểm của enzyme dựa trên các công cụ tin sinh học và trên thực nghiệm, một enzyme beta-glucosidase mã hóa từ gen mã số GL0036730 được ước đoán có hoạt tính nghiêng về kiềm nhẹ với chỉ số alkaline 0,851 và dự đoán có T_m nằm trong khoảng 55-65°C (Phụ lục 2).

Phòng kỹ thuật di truyền đã tiến hành biểu hiện enzyme trong tế bào *Escherichia coli*, tinh chế enzyme để nghiên cứu đặc điểm. Trong khuôn khổ luận văn, chúng tôi đã sử dụng enzyme tinh chế để đánh giá về hoạt tính của enzyme trên cơ chất đặc hiệu esculin, độ bền nhiệt và khoảng pH hoạt động của enzyme trên thực nghiệm.

3.3.1. Đánh giá hoạt tính beta-glucosidase

Mẫu protein pha tan sau khi siêu âm tế bào ở $OD_{600\text{ nm}} = 10$ được nhỏ lên trên đĩa có esculin. Kết quả cho thấy, xung quanh giếng có chứa protein pha tan chuyển màu nâu thẫm, khác biệt so với mẫu đối chứng, chứng tỏ beta-glucosidase trong dịch pha tan đã phân cắt esculin thành glucose và esculetin (Hình 3.14). Kết quả này có thể khẳng định rằng enzyme được xác định đúng là enzyme có mang hoạt tính beta-glucosidase.

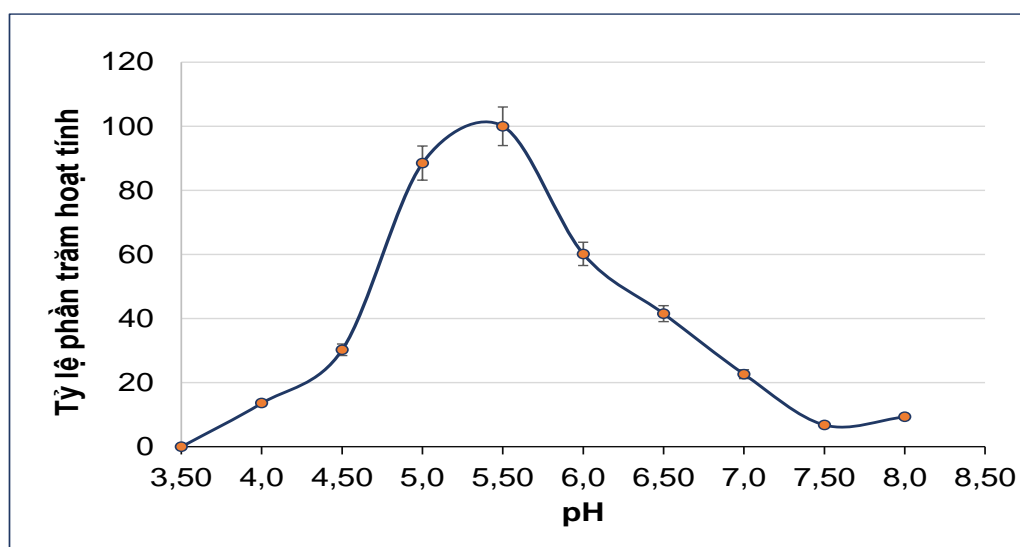


Hình 3.14: Định tính hoạt tính enzyme tinh sạch trên cơ chất esculin

*T: Protein beta-glucosidase tinh sạch;
(+): cellulase (Sigma, 0,05 U); (-): PBS*

3.3.2. Ảnh hưởng của pH tới hoạt tính beta-glucosidase

Sự thay đổi của pH sẽ dẫn đến sự ion hóa các nguyên tử và phân tử amino acid, thay đổi hình dạng và cấu trúc của protein, do đó làm tăng hay giảm chức năng của protein. Mỗi enzyme đều có khoảng pH hoạt động tối ưu nhất định. Việc thay đổi pH ở ngoài khoảng tối ưu sẽ làm giảm hoạt tính của enzyme. Enzyme beta-glucosidase GL0036730 được ước đoán là hoạt động tốt ở pH trung tính đến kiềm nhẹ. Kết quả khảo sát hoạt độ của enzyme ở các pH khác nhau 3,5 - 8 (Hình 3.15) cho thấy enzyme hoạt động tốt nhất ở pH 5-5,5. Hoạt độ enzyme cao nhất pH 5,5 đạt 17 U/mg. Hoạt độ enzyme giảm nhanh ở vùng pH axit. Ở pH 4,5 enzyme chỉ giữ được 30,25% hoạt tính. Enzyme có khả năng hoạt động tốt hơn ở vùng pH trung tính. Trong đó ở pH 6,0, enzyme giữ được 60,14% hoạt tính; ở pH 7,0 enzyme giữ được 22,6% hoạt tính nhưng khi tăng lên 7,5 enzyme chỉ còn giữ được 6,8% hoạt tính. Như vậy, so với ước đoán, các enzyme này hoạt động ở vùng pH trung tính nghiêng axit và không đạt được độ kiềm như mong muốn.



Hình 3.15: Ảnh hưởng của pH tới hoạt tính enzyme beta-glucosidase trên trình tự GL0036730

Tuy nhiên, do enzyme có hai vùng chức năng. Vùng GH31 có hoạt tính phân cắt các liên kết alpha-glucoside nên có hoạt tính alpha-glucosidase. Sự tồn tại của hai vùng hoạt tính có khả năng gây ảnh hưởng tới ước đoán tính chất của enzyme. Chúng tôi đã tách riêng phần GH3 được ước đoán có hoạt tính beta-glucosidase và kiểm tra lại xem enzyme thuộc dạng alkaline hay acid thì điểm số alkaline của enzyme giảm xuống còn 0,75. Với chỉ số alkaline này thì việc ước đoán enzyme thuộc loại alkaline chỉ đúng 75%. Đối với enzyme nghiên cứu, enzyme hoạt động trong vùng pH ngả axit nhẹ và bị ảnh hưởng bởi vùng xúc tác GH31. Để kết luận được chính xác ảnh hưởng của trình tự enzyme hay tỷ lệ sai số của phần mềm thì cần phải có các thực nghiệm kiểm chứng trên số mẫu lớn.

Bảng 3.4: Giá trị pH và pI của một số enzyme đã được nghiên cứu [101]

Enzyme	pI	pH tối ưu	Khoảng pH mà enzyme bền
Xylan endo-1,3-beta-xylosidase từ <i>Vibrio sp.</i>	3.6	7	5.0-8.0
Acetylxyylan esterase từ <i>Bacillus pumilus</i>	4.8	8	8.0-9.5
Beta-glucosidase I từ <i>Penicillium aurantiogriseu</i>	3.6	6	5.0-7.5
Glycine C-acetyltransferase từ <i>Bos</i>	5.2	7.5	6.5-8.0

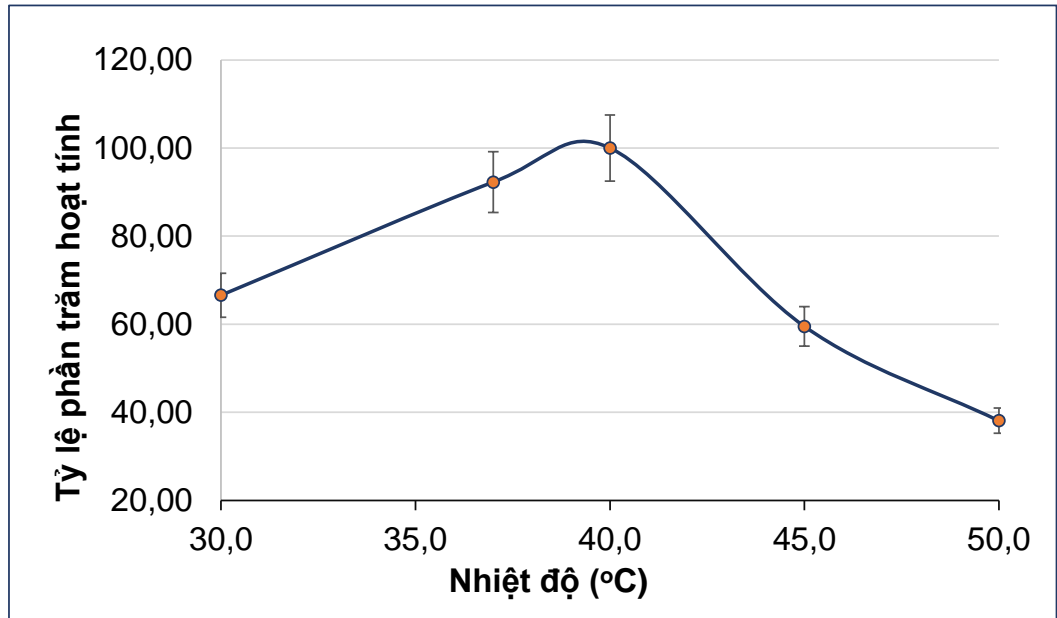
<i>taurus</i>			
Cellulase CH43 từ <i>Bacillus sp.</i>	5.4	6.5	6-10
Alpha-N-arabinofuranosidase từ <i>Aspergillus niger</i>	3.5	4	4 - 7
Beta-glucosidase từ <i>Termitomyces clypeatus</i>	4.5	5	2 -10
Glucan endo-1,6-beta-glucosidase từ <i>Acremonium sp.</i>	4.9	5	4.5-9.0
Mannan 1,4-mannobiosidase từ <i>Aeromonas hydrophil</i>	5.9	6	5.0-8.5
Alpha-N-arabinofuranosidase từ <i>Butyrivibrio fibrisolvens</i>	6	6	4.0-8.0
Glucan 1,3-beta-glucosidase từ <i>Acremonium</i>	5.1	5	4.5 - 10
Glutamate methylesterase từ <i>Rattus norve</i>	4.45	4	2-10
3-Phytase từ <i>Klebsiella terrigena</i>	5.5	5	3.0 -9.0
Alpha-galactosidase từ <i>Trichoderma reesei</i>	5.2	4	4 - 8
<i>Laccase từ Lentinula edodes</i>	6.9	3	3 -7

Vùng tô màu chỉ các enzyme có giá trị pH tối ưu nằm sát với giá trị pI.

Enzyme được dự đoán có pI là 5,61 và hoạt tính enzyme đạt tối ưu ở pH 5,5 gần sát với khoảng pI làm cho nhóm nghiên cứu đã có nghi ngờ vì thông thường theo lý thuyết enzyme sẽ bị tủa ở pH này nên có thể dẫn đến không còn hoạt tính. Trong trường hợp này, ảnh hưởng của pH đã được thử nghiệm lại nhưng vẫn cho kết quả tương tự. Qua nghiên cứu tài liệu, tác giả Talley và Alexoy (2010) cũng đã nghiên cứu *in silico* đặc điểm của nhiều enzyme và kiểm tra lại bằng thực nghiệm thì thấy rằng có nhiều enzyme có pH tối ưu gần hoặc trùng với giá trị pI như nghiên cứu này (Bảng 3.4) [101]. Việc tìm hiểu

nguyên nhân để giải thích cho hiện tượng trên sẽ cần có những nghiên cứu sâu hơn.

3.3.3. Ảnh hưởng của nhiệt độ tới hoạt tính beta-glucosidase



Hình 3.16: Ảnh hưởng của nhiệt độ lên hoạt tính enzyme beta-glucosidase trên trình tự GL0036730

Enzyme được ước đoán có điểm nóng chảy ở nhiệt độ 55-65°C nên enzyme có khả năng hoạt động tốt ở nhiệt độ dưới 55°C. Kết quả kiểm tra từ thực nghiệm cho thấy, enzyme hoạt động tốt nhất ở 40°C với hoạt tính đạt 19 U/mg. Enzyme bị giảm nhanh hoạt tính khi tăng nhiệt độ. Cụ thể, ở nhiệt độ 45°C, enzyme chỉ duy trì được 60% hoạt tính và khi nhiệt độ tăng lên 50°C thì enzyme chỉ duy trì được 3% hoạt tính (Hình 3.16). Điều này chứng tỏ, khi tăng nhiệt độ lên 55°C, enzyme sẽ bị biến tính.

KẾT LUẬN VÀ KIẾN NGHỊ

KẾT LUẬN

1. Từ hai bộ dữ liệu giải trình tự DNA metagenome của vi khuẩn trong dạ cỏ dê có kích thước 8,46 Gb và 48,66 Gb, tổng số 804 trình tự beta-glucosidase hoàn thiện có kích thước 200 đến 2253 amino acid đã được khai thác. Kích thước beta-glucosidase phổ biến trong khoảng 700-900 amino acid. Các trình tự có pI nằm trong khoảng từ 3,9 đến 9,67, tập trung chủ yếu ở khoảng pI 5-8. Mặc dù số lượng trình tự có chỉ số alkaline từ 0,6 đến 1 chiếm tỷ lệ lớn nhưng số lượng enzyme có chỉ số alkaline từ 0,99-1 chỉ chiếm 8,4%, số enzyme có chỉ số alkaline dưới 0,1 chiếm ~5%. Trong cả hai bộ dữ liệu, số lượng trình tự được ước đoán có nhiệt độ nóng chảy trên 65°C chiếm khoảng 50%.
2. Toàn bộ trình tự mã hóa enzyme beta-glucosidase trong bộ dữ liệu nhỏ với 131 trình tự hoàn thiện đều thuộc họ GH3. Trong tổng số 673 trình tự beta-glucosidase hoàn thiện của bộ dữ liệu lớn có 90,2% trình tự thuộc họ GH3, 5,2% trình tự thuộc họ GH16 và 4,6% trình tự thuộc họ GH1. Khoảng trên 90% enzyme thuộc họ GH3 có chứa module FN3. Đây được coi là đặc trưng cho các enzyme beta-glucosidase thu từ dạ cỏ dê. Trong bộ dữ liệu lớn còn tìm thấy cấu trúc module khá đặc biệt có vùng GH3C nằm ở tận cùng đầu N, vùng GH3N nằm ở đầu C và vùng FN3 luôn đứng sau vùng GH3C. Kết quả ước đoán cấu trúc bậc 3 của hai trình tự enzyme điển hình *GL0036730* và *GL0203190* đã được đánh giá và phân tích cho thấy sự bảo tồn về cấu trúc bậc 3 của enzyme mặc dù có sự khác nhau về sắp xếp vị trí trình tự module.
3. Enzyme *GL0036730* có hai vùng xúc tác GH3, GH31 được ước đoán có chỉ số alkaline 0,851, Tm 55-65°C và pI 5,61 đã được kiểm chứng lại nhiệt độ và pH hoạt động của enzyme. Trên thực nghiệm, enzyme beta-glucosidase tái tổ hợp thể hiện hoạt tính tốt nhất ở 40°C và pH 5,5. Khi loại bỏ vùng GH31, giá trị alkaline của enzyme bị giảm xuống 0,75.

KIẾN NGHỊ

Số lượng beta-glucosidase thu được từ hai bộ dữ liệu là rất lớn nhưng việc nghiên cứu biểu hiện, kiểm chứng lại tính chất enzyme từ hai bộ dữ liệu còn rất hạn chế. Vì vậy, cần có nghiên cứu biểu hiện các enzyme có cấu trúc, đặc điểm khác nhau để kiểm chứng lại kết quả ước đoán tính chất enzyme từ các công cụ tin sinh học để hướng tới lựa chọn *in silico* được chính xác các enzyme có đặc tính tốt cho ứng dụng vào thực tiễn.

BÀI BÁO

Đỗ Thị Huyền, **Nguyễn Minh Chiến**, Trương Nam Hải (2024) 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. Hội nghị Công nghệ sinh học toàn quốc 2024. 32-37

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] Tran T., Son J., Awais M., Ko J., Yang D., Jung S., 2023, β -Glucosidase and its application in bioconversion of ginsenosides in *Panax ginseng*, *Bioengineering (Basel)*, 10(4), pp. 484.
- [2] Ma L., Liu X., Guo L., 2024, Discovery of plant chemical defence mediated by a two-component system involving β -glucosidase in *Panax* species, *Nat Commun*, 15(1).
- [3] Seeberger P., Werz D., 2007, Synthesis and medical applications of oligosaccharides, *Nature*, 446(7139), pp. 1046–1051.
- [4] Rocha R., Mariano D., Almeida D., 2023, Thermostabilizing mechanisms of canonical single amino acid substitutions at a GH1 β -glucosidase probed by multiple MD and computational approaches, *Proteins*, 91(2), pp. 218–236.
- [5] Godse R., Awane H., Tripathi J., Kulkarni R., 2021, Unconventional β -glucosidases: A promising biocatalyst for industrial biotechnology, *Appl Biochem Biotechnol*, 193(9), pp. 2993–3016.
- [6] Ahmed A., Hassan Nasim F., Batool K., Bibi A., 2017, Microbial β -glucosidase: Sources, production and applications, *Journal of Applied & Environmental Microbiology*, 5(1).
- [7] Huyen D., Thao N., Thanh M., 2014, Mining biomass-degrading genes through Illumina-based *de novo* sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam, *J Biosci Bioeng*, 118(6), pp. 665–671.
- [8] Huyen D., Ngoc L., Khoa., 2018, Metagenomic insights into lignocellulose-degrading genes through Illumina based *de novo* sequencing of the microbiome in vietnamese native goats rumen, *The Journal of general and applied microbiology*, 64(3), pp. 108–116.
- [9] Singh G., Verma A., Kumar V., 2016, Catalytic properties, functional attributes and industrial applications of β -glucosidases, *3 Biotech*, 6(1), pp. 3.
- [10] Erkanli M., El-Halabi K., Kim J., 2024, Exploring the diversity of β -glucosidase: Classification, catalytic mechanism, molecular characteristics, kinetic models, and applications, *Enzyme and Microbial Technology*, 173, p. 110363.

- [11] Zhang R., Cao C., Bi J., Li Y., 2022, Fungal cellulases: protein engineering and post-translational modifications, *Appl Microbiol Biotechnol*, 106(1), pp. 1–24.
- [12] Tribolo S., Berrin J., Kroon P., Czjzek M., Juge N., 2007, The crystal structure of human cytosolic beta-glucosidase unravels the substrate aglycone specificity of a family 1 glycoside hydrolase, *J Mol Biol*, 370(5), pp. 964–975.
- [13] Singhanian R., Patel A., Sukumaran R., Larroche C., Pandey A., 2013, Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production, *Bioresour Technol*, 127, pp. 500–507.
- [14] Godse R., Bawane H., Tripathi J., Kulkarni R., 2021, Unconventional β -Glucosidases: A Promising Biocatalyst for Industrial Biotechnology, *Appl Biochem Biotechnol*, 193(9), pp. 2993–3016.
- [15] Treebupachatsakul T., Nakazawa N., Shinbo H., 2016, Heterologously expressed *Aspergillus aculeatus* β -glucosidase in *Saccharomyces cerevisiae* is a cost-effective alternative to commercial supplementation of β -glucosidase in industrial ethanol production using *Trichoderma reesei* cellulases, *J Biosci Bioeng*, 121(1), pp. 27–35.
- [16] Sørensen A., Lübeck M., Ahring B., 2013, Fungal Beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials, *Biomolecules*, 3(3), pp. 612–631.
- [17] Kabera N., Semana E., Mussa R., He X., 2014, Plant Secondary Metabolites: Biosynthesis, Classification, Function and Pharmacological Properties, *Journal of Pharmacy and Pharmacology*, 2, pp. 377-392.
- [18] Karimi E., Oskoueian E., Hendra R., Oskoueian A., Jaafar Z., 2012, Phenolic compounds characterization and biological activities of *Citrus aurantium* bloom, *Molecules*, 17(2), pp. 1203-1218 .
- [19] Servili M., Sordini B., Esposto S., 2013, Biological Activities of Phenolic Compounds of Extra Virgin Olive Oil, *Antioxidants (Basel)*, 3(1), pp. 1–23.
- [20] Setchell K., Brown N., Brasher W., 2002, Evidence for lack of absorption of soy isoflavone glycosides in humans, supporting the crucial role of intestinal metabolism for bioavailability, *Am J Clin Nutr*, 76(2), pp. 447–453.

- [21] Day A., Dupont M., Ridley S., 1998, Deglycosylation of flavonoid and isoflavonoid glycosides by human small intestine and liver beta-glucosidase activity, *FEBS Lett*, 436(1), pp. 71–75.
- [22] Maicas S., Mateo J., 2005, Hydrolysis of terpenyl glycosides in grape juice and other fruit juices: a review, *Appl Microbiol Biotechnol*, 67(3) pp. 322–335.
- [23] Gunata Y., Bayonove C., Baumes R., Cordonnier R., 1985, The aroma of grapes I. Extraction and determination of free and glycosidically bound fractions of some grape aroma components, *Journal of Chromatography A*, 331, pp. 83–90.
- [24] Krammer G., Winterhalter P., Schwab M., Schreier P., 1991, Glycosidically bound aroma compounds in the fruits of *Prunus* species: apricot (*P. armeniaca*, L.), peach (*P. persica*, L.), yellow plum (*P. domestica*, L. ssp. *syriaca*), *J. Agric. Food Chem.*, 39(4), pp. 778–781.
- [25] Sakho M., Chassagne D., Crouzet J., 1997, African Mango Glycosidically Bound Volatile Compounds, *J. Agric. Food Chem.*, 45(3), pp. 883–888.
- [26] Roscher R., Herderich M., Steffen J., Schreier P., Schwab P., 1996, 2,5-Dimethyl-4-hydroxy-3[2H]-furanone 6'O-malonyl-beta-D-glucopyranoside in strawberry fruits, *Phytochemistry*, 43(1), pp. 155–159.
- [27] Mateo J., Jiménez M., 2000, Monoterpenes in grape juice and wines, *J Chromatogr A*, 881(1–2), pp. 557–567.
- [28] Gunata Z., Bitteur S., Brilouet J., Bayonove C., Cordonier R., Sequential enzymic hydrolysis of potentially aromatic glycosides from grape, *Carbohydr. res*, 184, pp. 139–149.
- [29] Baffi M., Martin N., Tobal T., 2013, Purification and characterization of an ethanol-tolerant β -glucosidase from *Sporidiobolus pararoseus* and its potential for hydrolysis of wine aroma precursors, *Appl Biochem Biotechnol*, 171(7), pp. 1681–1691.
- [30] Michlmayr H., Wurbs B., Kulke K., 2010, A β -glucosidase from *Oenococcus oeni* ATCC BAA-1163 with potential for aroma release in wine: Cloning and expression in *E. coli*, *World J Microbiol Biotechnol*, 26(7), pp. 1281–1289.

- [31] Michlmayr H., Schümann C., Silva N., Kulbe K., Hierro A., 2010, Isolation and basic characterization of a beta-glucosidase from a strain of *Lactobacillus brevis* isolated from a malolactic starter culture, *J Appl Microbiol*, 108(2), pp. 550–559.
- [32] Srisongkram T., 2022, Basic principles and applications of animal cell culture in pharmaceutical sciences.
- [33] Atta-Mensah S., 1990, Detoxification of cassava during gari preparation, *Journal of Food*.
- [34] Ghanbarzadeh B., Ataefard M., Etezad S., Mahdavi S., 2024, Optical and Printing Properties of Deinked Office Waste Printed Paper, *Progress in Color, Colorants and Coatings*, 17(1), pp. 75–84.
- [35] Park N., Cha J., Kim D., Park C., 2007, Enzymatic characterization and substrate specificity of thermostable beta-glycosidase from hyperthermophilic archaea, *Sulfolobus shibatae*, expressed in *E. coli*, *J Microbiol Biotechnol*, 17(3), pp. 454–460.
- [36] Bruins M., Strubel M., Lieshout J., Janssen A., Boom R., 2003, Oligosaccharide synthesis by the hyperthermostable β -glucosidase from *Pyrococcus furiosus*: kinetics and modelling, *Enzyme and Microbial Technology*, 33(1), pp. 3–11 .
- [37] Otto R., Bornscheuer U., Syldatk C., Schmid R., 1998, Synthesis of aromatic n-alkyl-glucoside esters in a coupled β -glucosidase and lipase reaction, *Biotechnology Letters*, 20(4), pp. 437–440.
- [38] Charoenwattanasatien R., Pengthaisong S., Breen M., 2016, Bacterial β -glucosidase reveals the structural and functional basis of genetic defects in human glucocerebrosidase 2 (GBA2), *ACS Chem. Biol.*, 11(7), pp. 1891–1900.
- [39] Ouyang B., Wang G., Zhang N., Zuo J., Huang Y., Zhao X., 2023, Recent advances in β -glucosidase sequence and structure engineering: A brief review, *Molecules*, 28(13).
- [40] Deflandre B., Jadot C., Herman R., 2022, Structure and function of BcpE2, the most promiscuous GH3-family glucose scavenging beta-glucosidase, *mBio*, vol. 13, no. 4, pp. e00935-22.
- [41] Jeng W., Wang N., Lin M., 2011, Structural and functional analysis of three β -glucosidases from bacterium *Clostridium cellulovorans*, fungus

- Trichoderma reesei* and termite *Neotermes koshunensis*, *Journal of Structural Biology*, 173(1), pp. 46–56.
- [42] Chida N., Sato T., 2012, 2.8 Chiral Pool Synthesis: Chiral Pool Syntheses Starting from Carbohydrates, *Elsevier*, pp. 207–239.
- [43] Dale M., Kopfler W., Chait I., Byers L., 1986, Beta.-Glucosidase: substrate, solvent, and viscosity variation as probes of the rate-limiting steps, *ACS Publications*.
- [44] Davies G., Henrissat B., 1995, Structures and mechanisms of glycosyl hydrolases, *Structure*, 3(9), pp. 853–859.
- [45] Guillén D., Sánchez S., Rodríguez R., 2009, Carbohydrate-binding domains: Multiplicity of biological roles, *Applied microbiology and biotechnology*, 85, pp. 1241–1249.
- [46] Hansen C., 1992, Fibronectin type III-like sequences and a new domain type in prokaryotic depolymerases with insoluble substrates, *FEBS Journal*, 305(2), pp. 91–96.
- [47] Perrakis A., Ouzounis C., Wilson K., 1997, Evolution of immunoglobulin-like modules in chitinases: their structural flexibility and functional implications, *Folding and Design*, 2(5), pp. 291–294.
- [48] Guillén D., Sánchez S., Rodríguez-Sanoja R., 2010, Carbohydrate-binding domains: multiplicity of biological roles, *Appl. Microbiol. Biotechnol.*, 85(5), pp. 1241–1249.
- [49] Campbell I., Spitzfaden C., 1994, Building proteins with fibronectin type III modules, *Structure*, 2(5), pp. 333–337.
- [50] Little E., Bork P., Doolittle R., 1994, Tracing the spread of fibronectin type III domains in bacterial glycohydrolases, *J Mol Evol*, 39(6), pp. 631–643.
- [51] Kataeva I., Seidel R., Shah A., West L., Li X., Ljungdahl L., 2002, The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface, *Microbiol*, 68(9), pp. 4292–4300.
- [52] Kataeva I., Uversky V., Brever M., 2004, Interactions between immunoglobulin-like and catalytic modules in *Clostridium thermocellum* cellulosomal cellobiohydrolase CbhA, *Protein .*, 17(11), pp. 759–769.

- [53] Isorna P., Polaina J., Latorre-García L., Cañada F., González B., Sanz-Aparicio, 2007., Crystal structures of *Paenibacillus polymyxa* beta-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases, *J Mol Biol*, 371(5), pp. 1204–1218.
- [54] Sørensen A., Lübeck M., Ahring B., 2013, Fungal Beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials, *Biomolecules*, 3(3), pp. 612–631.
- [55] Li Y., Jiang C., Wan X., Zhang Z., Li D., 2005, Purification and partial characterization of beta-glucosidase from fresh leaves of tea plants, *Acta Biochim Biophys Sin (Shanghai)*, 37(6), pp. 363–370.
- [56] Ait N., Creuzet N., Cattaneo J., 1982, Properties of β -Glucosidase Purified from *Clostridium thermocellum*, *Microbiology*, 128(3), pp. 569–577.
- [57] Kengen S., Luesink E., Stams A., Zehnder A., 1993, Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*, *Biochem*, 213(1), pp. 305–312.
- [58] Spiridonov N., Wilson D., 2001, Cloning and biochemical characterization of BglC, a beta-glucosidase from the cellulolytic actinomycete *Thermobifida fusca*, *Curr Microbiol*, 42(4), pp. 295–301.
- [59] Hong M., Kim Y., Park C., Lee J., Kim Y., Oh D., 2009, Characterization of a recombinant beta-glucosidase from the thermophilic bacterium *Caldicellulosiruptor saccharolyticus*, *J Biosci Bioeng*, 108(1), pp. 36–40.
- [60] An D., Cui C., Lee H., 2010, Identification and characterization of a novel *Terrabacter ginsenosidimutans* sp. nov. beta-glucosidase that transforms ginsenoside Rb1 into the rare gypenosides XVII and LXXV, *Appl Environ Microbiol*, 76(17), pp. 5827–5836.
- [61] Naz S., Ikram N., Rajoka M., Sadaf S., Akhtar M., 2010, Enhanced production and characterization of a β -glucosidase from *Bacillus halodurans* expressed in *Escherichia coli*, *Biochemistry Moscow*, 75(4), pp. 513–518.

- [62] Del Pozo M., Arroso L., Martinez J., 2012, Microbial β -glucosidases from cow rumen metagenome enhance the saccharification of lignocellulose in combination with commercial cellulase cocktail, *Biotechnology for Biofuels*, 5, pp.73.
- [63] Uchiyama T., Miyazaki K., Yaoi K., 2013, Characterization of a novel β -glucosidase from a compost microbial metagenome with strong transglycosylation activity, *J Biol Chem*, 288(25), pp. 18325–18334.
- [64] Tiwari P., Misra B., Sangwan N., 2013, Beta-glucosidases from the fungus *Trichoderma*: an efficient cellulase machinery in biotechnological applications, *BioMed Research International*, 2013, pp. 203735.
- [65] Zang X., Liu M., Fan Y., Xu J., Xu X., Li H., 2018, The structural and functional contributions of β -glucosidase-producing microbial communities to cellulose degradation in composting, *Biotechnol Biofuels*, 11, pp. 51.
- [66] Kaushal G., Rai A., Singh S., 2021, A novel β -glucosidase from a hot-spring metagenome shows elevated thermal stability and tolerance to glucose and ethanol, *Enzyme and Microbial Technology*, 145, pp. 109764.
- [67] Ariaeenejad S., Rarhban M., Kavousi K., 2020, A novel high glucose-tolerant β -glucosidase: Targeted computational approach for metagenomic screening, *Front Bioeng Biotechnol*, 8, pp. 813.
- [68] Crespim E., Zanphorlin L., Souza F., 2016, A novel cold-adapted and glucose-tolerant GH1 β -glucosidase from *Exiguobacterium antarcticum* B7, *International Journal of Biological Macromolecules*, 82, pp. 375–380.
- [69] Chan C., Sin L., Chan K., 2016, Characterization of a glucose-tolerant β -glucosidase from *Anoxybacillus* sp. DT3-1, *Biotechnology for Biofuels*, 9(1), pp. 174.
- [70] Sinha S., Datta S., 2016, β -Glucosidase from the hyperthermophilic archaeon *Thermococcus* sp. is a salt-tolerant enzyme that is stabilized by its reaction product glucose, *Appl Microbiol Biotechnol*, 100(19), pp. 8399–8409.

- [71] Tiwari R., Singh P., Singh S., Nain P., Nain L., Shukla P., 2017, Bioprospecting of novel thermostable β -glucosidase from *Bacillus subtilis* RA10 and its application in biomass hydrolysis, *Biotechnology for Biofuels*, 10(1), pp. 246.
- [72] Singh B., 2020, Exploring microbial diversity for biotechnology: the way forward, *Trends Biotechnol*, 28(3), pp. 111–116.
- [73] Ferrer M., Beloqui A., Timmis K., Golyshin, 2009 P., Metagenomics for mining new genetic resources of microbial communities, *J Mol Microbiol Biotechnol*, 16(1–2), pp. 109–123.
- [74] Handelsman J., 2004, Metagenomics: Application of Genomics to Uncultured Microorganisms, *Microbiol Mol Biol Rev*, 68(4), pp. 669–685.
- [75] Kim S., Lee C ., Kim M., 2007, Screening and characterization of an enzyme with beta-glucosidase activity from environmental DNA, *J Microbiol Biotechnol*, 17(6), pp. 905–912.
- [76] Feng Y., Duan C., Liu L., Tang J., Feng J., 2009, Properties of a metagenome-derived beta-glucosidase from the contents of rabbit cecum *Biosci Biotechnol Biochem*, 73(7), pp. 1470–1473.
- [77] Coci M., Odermatt N., Salcher M., Pernthaler J., Corno G., 2015, Ecology and Distribution of Thaumarchaea in the Deep Hypolimnion of Lake Maggiore, *Archaea*, 2015, pp. 590434.
- [78] Tatusov R., Ferodova N., Jackon J., 2003, The COG database: an updated version includes eukaryotes, *BMC Bioinformatics*, 4(1), pp. 41.
- [79] Guo H., Feng Y., Mo X., Duan C., Tang J., Feng J., 2008, Cloning and expression of a beta-glucosidase gene umcel3G from metagenome of buffalo rumen and characterization of the translated product, *Sheng Wu Gong Cheng Xue Bao*, 24(2), pp. 232–238.
- [80] Giang M., Huyen Đ., Hai T., 2016, Using bioinformatic tools in exploited gene encoding enzyme to decompose lignocellulose from metagenome of free - living bacteria in the gut of the lower termite *Coptotermes gestroi*, *Vietnam Journal of Biotechnology*, 14(1).
- [81] Fan G., Li Q., Zuo Y., 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 Biochemistry*, 48(7), pp. 1048–1053.

- [82] Lin H., Chen W., Ding H., 2013, AcalPred: A Sequence-Based Tool for Discriminating between Acidic and Alkaline Enzymes, *PLoS ONE*, 8(10).
- [83] Ebrahimi M., Lakizadeh A., Agha-Golzadeh P., Ebrahimie E., 2011, Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes, *PLoS One*, 6(8).
- [84] Dyson Z., Seviour R., Tucci J., Petrovski S., 2016, Genome sequences of *Pseudomonas oryzae* phage POR1 and *Pseudomonas aeruginosa* phage PAE1, *Genome Announc.*, 4(3), pp. e01515-15.
- [85] Bitencourt T., Macedo C., Fanco C., 2016, Transcription profile of *Trichophyton rubrum* conidia grown on keratin reveals the induction of an adhesin-like protein gene with a tandem repeat pattern, *BMC Genomics*, 17.
- [86] Jiang S., Cai D., Sun Y., Teng Y., 2016, Isolation and characterization of putative functional long terminal repeat retrotransposons in the *Pyrus* genome, *Mob DNA*, 7.
- [87] Kroj T., Chanclud E., Michel-Romiti C., Grand X., Morel J., 2016, Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread, *New Phytol*, 210(2), pp. 618–626.
- [88] Ariyaratna H., Francki M., 2016, Phylogenetic relationships and protein modelling revealed two distinct subfamilies of group II HKT genes between crop and model grasses, *Genome*, pp. 1–9.
- [89] Thakur R., Shankar J., 2016, In silico analysis revealed high-risk single nucleotide polymorphisms in human pentraxin-3 gene and their impact on innate immune response against microbial pathogens, *Front Microbiol*, 7.
- [90] Veena V., Poornima P., Parvatham R., Kalalselvi K., 2011, Isolation and characterization of beta-glucosidase producing bacteria from different sources, *African Journal of Biotechnology*, 10, pp. 14907–14912

- [91] Su H., Xiao H., Yu K., 2021, High diversity of β -glucosidase-producing Bacteria and their genes associated with Scleractinian corals, *Int J Mol Sci*, 22(7), pp. 3523.
- [92] Bhattacharya A., Pletschke B., 2014, Review of the enzymatic machinery of *Halothermothrix orenii* with special reference to industrial applications, *Enzyme Microb Technol*, 55, pp. 159–169.
- [93] Rabee A., Sallam A., Sarba E., 2024, Rumen fermentation and microbiota in Shami goats fed on condensed tannins or herbal mixture, *BMC Vet Res*, 20, pp. 35.
- [94] Shaw K., Grimsley G., Yakovlev G., Makarov A., Pace C., 2001, The effect of net charge on the solubility, activity, and stability of ribonuclease Sa, *Protein Sci*, 10(6), pp. 1206–1215.
- [95] Sidar A., Albuquerque E., Voshol G., Ram A., Vijgenboom E., Punt P., 2020, Carbohydrate Binding Modules: Diversity of Domain Architecture in Amylases and Cellulases From Filamentous Microorganisms, *Front Bioeng Biotechnol*, 8, pp. 871.
- [96] Mohamad Sobri M., Abd-Aziz S., Abu Bakar F., Ramli N., 2020, In-silico characterization of glycosyl hydrolase family 1 β -glucosidase from *Trichoderma asperellum* UPM1, *Int J Mol Sci*, 21(11), pp. 4035.
- [97] Ansari S., Husain Q., 2011, Immobilization of β galactosidase on concanavalin A-layered calcium alginate-cellulose beads and its application in lactose hydrolysis in continuous spiral bed reactors, *Polish Journal of Chemical Technology*, 13(4), pp. 15–20.
- [98] Yoshida E., Hidaka E., Fufhinobu S., 2010, Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 β -glucosidase from *Kluyveromyces marxianus*, *Biochem J*, 431(1), pp. 39–49.
- [99] Mohsin I., Poudel N., Li D., Papageorgiou A., 2019, Crystal structure of a GH3 β -glucosidase from the thermophilic fungus *Chaetomium thermophilum*, *Int J Mol Sci*, 20(23), pp. 5962.
- [100] Karkehabadi S., Hansson H., Kim S., 2018, Structural studies of a glycoside hydrolase family 3 β -glucosidase from the model fungus *Neurospora crassa*, *Acta Crystallogr F Struct Biol Commun*, 74(12), pp. 787–796.

- [101] Talley K., Alexov E., 2010, On the pH-optimum of activity and stability of proteins, *Proteins*, 78(12), pp. 2699–2706.

PHỤ LỤC

Bảng phụ lục 1: *Kết quả dự đoán một số tính đặc điển của ORF mã hóa từ bộ dữ liệu nhỏ*

STT	Mã gen	Tm	aa	Akaline Score	Acid Score
1	[denovogenes]_2771	>65	862	0,783	0,217
2	[denovogenes]_2827	>65	859	0,669	0,331
3	[denovogenes]_4170	55-65	785	0,625	0,375
4	[denovogenes]_4750	55-65	755	0,971	0,029
5	[denovogenes]_4840	55-65	751	0,977	0,023
6	[denovogenes]_4109	>65	787	0,675	0,325
7	[denovogenes]_2748	>65	864	0,934	0,066
8	[denovogenes]_9726	>65	606	0,139	0,861
9	[denovogenes]_2821	>65	859	0,686	0,314
10	[denovogenes]_3282	55-65	829	0,987	0,013
11	[denovogenes]_4764	55-65	754	0,475	0,525
12	[denovogenes]_2747	>65	864	0,901	0,099
13	[denovogenes]_2574	>65	879	0,575	0,425
14	[denovogenes]_2639	55-65	873	0,227	0,773
15	[denovogenes]_2703	>65	868	0,602	0,398
16	[denovogenes]_2809	>65	859	0,840	0,160
17	[denovogenes]_2859	55-65	857	0,987	0,013
18	[denovogenes]_2942	55-65	851	0,298	0,702
19	[denovogenes]_4394	55-65	773	0,945	0,055
20	[denovogenes]_5067	55-65	741	0,930	0,070
21	[denovogenes]_11003	55-65	580	0,354	0,646
22	[denovogenes]_2632	>65	874	0,227	0,773
23	[denovogenes]_2709	>65	867	0,319	0,681
24	[denovogenes]_2731	>65	865	0,855	0,145
25	[denovogenes]_3206	55-65	834	0,997	0,003
26	[denovogenes]_3327	55-65	826	0,996	0,004
27	[denovogenes]_4524	55-65	766	0,996	0,004
28	[denovogenes]_5096	>65	739	0,277	0,723

29	[denovogenes]_3628	>65	812	0,781	0,219
30	[denovogenes]_3894	>65	798	0,826	0,174
31	[denovogenes]_3914	>65	796	0,292	0,708
32	[denovogenes]_3916	55-65	796	0,922	0,078
33	[denovogenes]_3979	55-65	793	0,950	0,050
34	[denovogenes]_4005	55-65	792	0,827	0,173
35	[denovogenes]_4040	>65	790	0,893	0,107
36	[denovogenes]_4221	>65	782	0,793	0,207
37	[denovogenes]_4252	55-65	781	0,895	0,105
38	[denovogenes]_4279	>65	779	0,685	0,315
39	[denovogenes]_4309	55-65	778	0,918	0,082
40	[denovogenes]_4506	55-65	767	0,836	0,164
41	[denovogenes]_4509	>65	767	0,838	0,162
42	[denovogenes]_4563	55-65	764	0,544	0,456
43	[denovogenes]_4719	>65	756	0,936	0,064
44	[denovogenes]_21483	55-65	454	0,940	0,060
45	[denovogenes]_3920	>65	796	0,943	0,057
46	[denovogenes]_3960	55-65	794	0,952	0,048
47	[denovogenes]_3999	55-65	792	0,429	0,571
48	[denovogenes]_4081	>65	788	0,713	0,287
49	[denovogenes]_4103	>65	787	0,776	0,224
50	[denovogenes]_4110	>65	787	0,691	0,309
51	[denovogenes]_4125	55-65	786	0,724	0,276
52	[denovogenes]_4133	55-65	786	0,607	0,393
53	[denovogenes]_4148	>65	786	0,824	0,176
54	[denovogenes]_4178	55-65	784	0,727	0,273
55	[denovogenes]_4222	>65	782	0,911	0,089
56	[denovogenes]_4240	55-65	781	0,903	0,097
57	[denovogenes]_4266	>65	780	0,398	0,602
58	[denovogenes]_4273	>65	779	0,871	0,129
59	[denovogenes]_4317	55-65	778	0,712	0,288
60	[denovogenes]_4363	55-65	775	0,943	0,057
61	[denovogenes]_4378	55-65	774	0,607	0,393

62	[denovogenes]_4456	55-65	769	0,841	0,159
63	[denovogenes]_4532	>65	766	0,926	0,074
64	[denovogenes]_4616	>65	760	0,957	0,043
65	[denovogenes]_4763	>65	754	0,976	0,024
66	[denovogenes]_2805	>65	860	0,271	0,729
67	[denovogenes]_2671	>65	870	0,204	0,796
68	[denovogenes]_4974	55-65	745	0,774	0,226
69	[denovogenes]_2426	55-65	894	0,072	0,928
70	[denovogenes]_3311	55-65	827	0,990	0,010
71	[denovogenes]_3366	55-65	824	0,993	0,007
72	[denovogenes]_5084	55-65	740	0,935	0,065
73	[denovogenes]_3382	55-65	823	0,995	0,005
74	[denovogenes]_4441	55-65	770	0,761	0,239
75	[denovogenes]_2746	>65	864	0,726	0,274
76	[denovogenes]_2761	55-65	863	0,691	0,309
77	[denovogenes]_7571	55-65	659	0,562	0,438
78	[denovogenes]_11451	55-65	572	0,985	0,015
79	[denovogenes]_2596	55-65	877	0,529	0,471
80	[denovogenes]_2656	>65	871	0,165	0,835
81	[denovogenes]_2658	>65	871	0,073	0,928
82	[denovogenes]_2662	>65	871	0,147	0,853
83	[denovogenes]_2878	55-65	856	0,812	0,188
84	[denovogenes]_3396	55-65	822	0,983	0,017
85	[denovogenes]_3406	55-65	822	0,990	0,010
86	[denovogenes]_39321	55-65	352	0,985	0,015
87	[denovogenes]_4008	55-65	792	0,693	0,307
88	[denovogenes]_4712	55-65	756	0,191	0,809
89	[denovogenes]_4837	55-65	751	0,694	0,306
90	[denovogenes]_4879	55-65	750	0,470	0,530
91	[denovogenes]_5006	>65	744	0,132	0,868
92	[denovogenes]_5059	55-65	741	0,789	0,211
93	[denovogenes]_5080	55-65	740	0,783	0,217
94	[denovogenes]_5120	>65	738	0,983	0,017

95	[denovogenes]_5628	>65	719	0,672	0,328
96	[denovogenes]_17198	>65	493	0,219	0,781
97	[denovogenes]_2535	>65	883	0,064	0,936
98	[denovogenes]_2610	>65	876	0,046	0,954
99	[denovogenes]_2623	55-65	875	0,330	0,670
100	[denovogenes]_2648	>65	872	0,058	0,942
101	[denovogenes]_2679	>65	870	0,565	0,435
102	[denovogenes]_2683	55-65	869	0,050	0,950
103	[denovogenes]_2796	>65	860	0,430	0,570
104	[denovogenes]_2914	>65	854	0,806	0,194
105	[denovogenes]_3486	55-65	819	0,668	0,332
106	[denovogenes]_4396	55-65	773	0,982	0,018
107	[denovogenes]_4718	55-65	756	0,527	0,473
108	[denovogenes]_4901	>65	749	0,293	0,707
109	[denovogenes]_5018	55-65	743	0,449	0,551
110	[denovogenes]_5153	55-65	737	0,472	0,528
111	[denovogenes]_5214	55-65	735	0,800	0,200
112	[denovogenes]_5758	>65	714	0,137	0,864
113	[denovogenes]_5791	>65	713	0,103	0,897
114	[denovogenes]_3678	>65	809	0,768	0,232
115	[denovogenes]_4134	>65	786	0,813	0,187
116	[denovogenes]_4402	>65	773	0,973	0,027
117	[denovogenes]_3708	>65	807	0,985	0,015
118	[denovogenes]_3591	>65	814	0,964	0,036
119	[denovogenes]_3834	>65	801	0,854	0,146
120	[denovogenes]_17221	>65	493	0,369	0,631
121	[denovogenes]_4104	>65	787	0,983	0,017
122	[denovogenes]_3450	>65	820	0,995	0,005
123	[denovogenes]_3618	>65	812	0,772	0,228
124	[denovogenes]_4257	>65	781	0,243	0,757
125	[denovogenes]_68027	55-65	257	0,330	0,670
126	[denovogenes]_3474	>65	819	0,712	0,288
127	[denovogenes]_4224	>65	782	0,874	0,126

128	[denovogenes]_4251	>65	781	0,927	0,073
129	[denovogenes]_4269	>65	780	0,937	0,063
130	[denovogenes]_4277	>65	779	0,973	0,027
131	[denovogenes]_4353	>65	775	0,918	0,082

Bảng phụ lục 2: Dự đoán, tính chất đặc điểm của ORF hoàn chỉnh thu từ bộ dữ liệu lớn

STT	Mã gen	Tm	aa	Akaline score	Acid score
1	>GL0519266	<55	200	1,000	0,000
2	>GL1566302	<55	202	0,420	0,580
3	>GL1697874	<55	202	0,969	0,031
4	>GL2616464	55-65	202	0,492	0,508
5	>GL2655146	55-65	203	1,000	0,000
6	>GL0549687	>65	204	0,003	0,997
7	>GL1535347	<55	206	0,004	0,996
8	>GL1754634	<55	206	0,385	0,615
9	>GL2447691	55-65	206	0,839	0,161
10	>GL2597643	>65	206	0,506	0,494
11	>GL1201108	<55	207	0,967	0,033
12	>GL2364104	<55	207	0,926	0,074
13	>GL1864084	>65	208	0,855	0,145
14	>GL0277841	55-65	209	0,080	0,920
15	>GL1354621	>65	209	0,380	0,620
16	>GL1042507	<55	211	0,001	0,999
17	>GL1661557	55-65	217	0,936	0,064
18	>GL0473750	55-65	218	0,694	0,306
19	>GL0490756	>65	218	0,670	0,330
20	>GL1846626	>65	221	0,033	0,967
21	>GL1526434	55-65	224	0,052	0,948
22	>GL0755083	<55	226	0,237	0,763
23	>GL0720331	>65	228	0,023	0,977
24	>GL1559271	>65	229	0,000	1,000
25	>GL2551123	>65	230	0,200	0,800
26	>GL0188734	55-65	231	0,239	0,761
27	>GL1013789	55-65	231	0,814	0,186
28	>GL1128079	<55	231	0,059	0,941
29	>GL1917898	>65	232	0,045	0,955
30	>GL2629756	>65	233	0,723	0,277

31	>GL2058787	>65	236	0,662	0,338
32	>GL1266659	<55	238	0,974	0,026
33	>GL2345077	55-65	238	0,984	0,016
34	>GL0537551	>65	240	0,977	0,023
35	>GL2246079	55-65	241	0,727	0,273
36	>GL2361367	55-65	241	0,237	0,763
37	>GL1777315	<55	242	0,500	0,500
38	>GL1347953	55-65	243	1,000	0,000
39	>GL1587250	55-65	244	0,556	0,444
40	>GL0692408	>65	245	0,124	0,876
41	>GL0696711	>65	246	0,923	0,077
42	>GL1620222	55-65	249	0,990	0,010
43	>GL0654245	>65	250	0,957	0,043
44	>GL0771195	<55	250	0,010	0,990
45	>GL1765931	55-65	251	1,000	0,000
46	>GL2169064	55-65	251	0,547	0,453
47	>GL0229597	<55	256	0,996	0,004
48	>GL1606574	>65	259	0,986	0,014
49	>GL0864617	>65	261	0,442	0,558
50	>GL1477159	>65	263	1,000	0,000
51	>GL0515603	55-65	264	1,000	0,000
52	>GL0712621	55-65	264	0,001	0,999
53	>GL1435421	55-65	264	0,131	0,869
54	>GL1860948	55-65	265	0,336	0,664
55	>GL0307424	55-65	267	0,838	0,162
56	>GL1748212	>65	268	0,432	0,568
57	>GL2074522	>65	270	0,229	0,771
58	>GL2197469	>65	270	0,951	0,049
59	>GL1585343	<55	273	0,424	0,576
60	>GL2246459	55-65	273	0,416	0,584
61	>GL1798865	>65	274	0,508	0,492
62	>GL1863986	55-65	274	0,277	0,723
63	>GL1647112	55-65	275	0,384	0,616

64	>GL2205698	55-65	277	0,080	0,920
65	>GL2065540	.65	282	0,002	0,998
66	>GL1261069	>65	283	0,000	1,000
67	>GL0299029	55-65	288	0,001	0,999
68	>GL1253580	55-65	289	0,011	0,989
69	>GL1547440	55-65	291	1,000	0,000
70	>GL1690952	55-65	292	0,989	0,011
71	>GL0309113	55-65	298	0,083	0,917
72	>GL1467459	<55	300	0,997	0,003
73	>GL1541080	55-65	300	0,000	1,000
74	>GL1011804	<55	306	0,004	0,996
75	>GL1043823	55-65	310	0,969	0,031
76	>GL1616487	>65	311	0,660	0,340
77	>GL0194899	>65	312	0,219	0,781
78	>GL0065076	55-65	313	0,027	0,973
79	>GL0504684	55-65	313	0,645	0,355
80	>GL1908453	>65	313	0,794	0,206
81	>GL0509157	55-65	314	0,409	0,591
82	>GL1494468	55-65	314	0,976	0,024
83	>GL1532195	55-65	314	0,867	0,133
84	>GL1779139	55-65	315	0,936	0,064
85	>GL1368726	55-65	316	0,827	0,173
86	>GL0299808	55-65	317	0,160	0,840
87	>GL0517657	>65	318	0,109	0,891
88	>GL1272957	>65	322	0,872	0,128
89	>GL1394839	55-65	327	0,967	0,033
90	>GL0451117	55-65	328	0,399	0,601
91	>GL1082369	55-65	328	0,772	0,228
92	>GL1323716	55-65	328	0,013	0,987
93	>GL1925261	55-65	330	0,146	0,854
94	>GL1068851	55-65	331	0,336	0,664
95	>GL1175194	>65	331	0,054	0,946
96	>GL1819884	55-65	331	0,957	0,043

97	>GL0891279	55-65	333	0,252	0,748
98	>GL0722561	>65	335	0,949	0,051
99	>GL0320311	>65	336	0,305	0,695
100	>GL0701392	>65	336	0,915	0,085
101	>GL1733172	55-65	336	0,993	0,007
102	>GL1001114	>65	340	0,001	0,999
103	>GL0071554	<55	342	0,293	0,707
104	>GL1565779	>65	342	0,959	0,041
105	>GL1397096	<55	343	0,455	0,545
106	>GL1447068	55-65	343	0,954	0,046
107	>GL1591200	55-65	344	0,794	0,206
108	>GL1498740	>65	350	0,321	0,679
109	>GL0687424	55-65	354	0,967	0,033
110	>GL1763829	<55	354	0,994	0,006
111	>GL0790401	>65	355	0,966	0,034
112	>GL0994579	55-65	357	0,839	0,161
113	>GL1652427	55-65	357	0,975	0,025
114	>GL0928350	55-65	359	0,946	0,054
115	>GL1393176	>65	361	0,245	0,755
116	>GL1733374	55-65	361	0,898	0,102
117	>GL0637780	55-65	362	0,357	0,643
118	>GL1668519	55-65	363	0,096	0,904
119	>GL1723798	55-65	371	0,044	0,956
120	>GL0652595	55-65	372	0,005	0,995
121	>GL0648969	<55	376	0,441	0,559
122	>GL0952173	>65	377	0,677	0,323
123	>GL1682374	55-65	379	0,184	0,816
124	>GL1543535	>65	381	0,967	0,033
125	>GL1390113	55-65	385	0,992	0,008
126	>GL1094445	>65	388	0,046	0,954
127	>GL0916088	55-65	392	0,740	0,260
128	>GL1503024	55-65	392	0,511	0,489
129	>GL1644445	55-65	393	0,164	0,836

130	>GL0660797	55-65	394	0,921	0,079
131	>GL1551395	55-65	394	0,985	0,015
132	>GL0689447	55-65	401	0,329	0,671
133	>GL0986010	>65	406	0,309	0,691
134	>GL1371709	>65	407	0,760	0,240
135	>GL0692689	>65	408	0,904	0,096
136	>GL1350846	>65	412	0,992	0,008
137	>GL1427142	55-65	414	0,518	0,482
138	>GL0409600	>65	417	0,272	0,728
139	>GL0676724	>65	421	0,816	0,184
140	>GL0847998	>65	421	0,808	0,192
141	>GL0555399	55-65	423	0,655	0,345
142	>GL0982793	55-65	424	0,950	0,050
143	>GL1309745	55-65	425	0,832	0,168
144	>GL0080126	55-65	426	0,641	0,359
145	>GL0037098	55-65	432	0,211	0,789
146	>GL0558956	55-65	432	0,905	0,095
147	>GL0789140	55-65	432	0,887	0,113
148	>GL0135229	55-65	433	0,194	0,806
149	>GL0237065	55-65	433	0,834	0,166
150	>GL0827597	<55	434	0,411	0,589
151	>GL0880940	>65	434	0,586	0,414
152	>GL1381292	55-65	435	0,373	0,627
153	>GL1390908	>65	436	0,992	0,008
154	>GL0388002	>65	437	0,071	0,929
155	>GL1040434	>65	439	0,584	0,416
156	>GL0044913	>65	440	0,001	0,999
157	>GL0368331	>65	440	0,555	0,445
158	>GL0166767	55-65	441	0,745	0,255
159	>GL0901780	55-65	442	0,992	0,008
160	>GL0037026	>65	447	0,014	0,986
161	>GL0854333	<55	447	0,617	0,383
162	>GL0043328	<55	448	0,064	0,936

163	>GL0069748	>65	448	0,055	0,945
164	>GL0395899	55-65	448	0,047	0,953
165	>GL0434255	55-65	449	0,726	0,274
166	>GL0158066	55-65	450	0,018	0,982
167	>GL0090531	55-65	451	0,008	0,992
168	>GL0803466	<55	451	0,054	0,946
169	>GL0871800	>65	452	0,870	0,130
170	>GL0928146	>65	452	0,991	0,009
171	>GL1472368	>65	452	0,009	0,991
172	>GL0466815	55-65	453	0,127	0,873
173	>GL1426961	>65	455	0,773	0,227
174	>GL0262429	55-65	456	0,319	0,681
175	>GL0936818	55-65	456	0,867	0,133
176	>GL0741913	55-65	458	0,283	0,717
177	>GL0842668	>65	459	0,004	0,996
178	>GL1251050	>65	459	0,980	0,020
179	>GL0923269	55-65	462	0,703	0,297
180	>GL0073774	>65	465	0,123	0,877
181	>GL1356242	55-65	465	0,744	0,256
182	>GL0537552	>65	466	0,914	0,086
183	>GL0890656	55-65	467	0,253	0,747
184	>GL1311059	>65	467	0,823	0,177
185	>GL0648245	55-65	469	0,018	0,982
186	>GL1038633	>65	470	0,024	0,976
187	>GL0082559	55-65	472	0,169	0,831
188	>GL0431161	>65	474	0,127	0,873
189	>GL0800015	<55	474	0,017	0,983
190	>GL0907284	>65	476	0,975	0,025
191	>GL1114278	<55	476	1,000	0,000
192	>GL0045506	55-65	478	0,065	0,935
193	>GL0631197	>65	478	0,844	0,156
194	>GL0741912	55-65	479	0,687	0,313
195	>GL1305452	>65	487	0,970	0,030

196	>GL1260344	>65	489	0,433	0,567
197	>GL1374953	>65	489	0,887	0,113
198	>GL1184407	55-65	491	0,637	0,363
199	>GL0759160	55-65	494	0,087	0,913
200	>GL0557997	55-65	496	0,702	0,298
201	>GL0668207	55-65	496	0,004	0,996
202	>GL1351981	>65	496	0,671	0,329
203	>GL0942712	>65	497	0,606	0,394
204	>GL1197810	55-65	497	0,150	0,850
205	>GL0046718	55-65	498	0,003	0,997
206	>GL0334386	<55	498	0,010	0,990
207	>GL0023604	<55	501	0,089	0,911
208	>GL0511958	55-65	501	0,953	0,047
209	>GL0706834	55-65	501	0,827	0,173
210	>GL1186057	55-65	501	0,705	0,295
211	>GL0309114	<55	505	0,055	0,945
212	>GL0719344	>65	507	0,603	0,397
213	>GL0234738	55-65	509	0,151	0,849
214	>GL0710620	55-65	509	0,333	0,667
215	>GL0631577	>65	511	0,679	0,321
216	>GL0752533	>65	511	0,570	0,430
217	>GL1208122	<55	513	0,026	0,974
218	>GL1090879	55-65	521	0,757	0,243
219	>GL0712622	>65	524	0,988	0,012
220	>GL0535182	55-65	526	0,300	0,700
221	>GL0765380	>65	526	0,859	0,141
222	>GL0812685	>65	527	0,982	0,018
223	>GL0732135	>65	530	0,979	0,021
224	>GL0724817	>65	532	0,645	0,355
225	>GL1048170	55-65	534	0,926	0,074
226	>GL0918869	55-65	535	0,217	0,783
227	>GL0773938	>65	545	0,827	0,173
228	>GL0506284	55-65	548	0,500	0,500

229	>GL0394439	>65	551	0,679	0,321
230	>GL1204243	55-65	553	0,994	0,006
231	>GL0849679	>65	562	0,914	0,086
232	>GL0842209	55-65	564	0,585	0,415
233	>GL0841753	55-65	565	0,846	0,154
234	>GL1002088	55-65	568	0,784	0,216
235	>GL0406889	55-65	570	0,432	0,568
236	>GL0711549	55-65	571	0,994	0,006
237	>GL1061378	>65	573	0,859	0,141
238	>GL0944918	>65	578	0,511	0,489
239	>GL0859996	55-65	581	0,102	0,898
240	>GL0922393	55-65	584	0,945	0,055
241	>GL0780037	>65	585	1,000	0,000
242	>GL0713821	>65	592	0,870	0,130
243	>GL0385480	>65	595	0,917	0,083
244	>GL0769550	>65	600	0,377	0,623
245	>GL0553613	>65	601	0,509	0,491
246	>GL0163749	55-65	604	0,067	0,933
247	>GL1036629	55-65	606	0,986	0,014
248	>GL1043965	>65	607	1,000	0,000
249	>GL0734093	55-65	615	0,774	0,226
250	>GL0891661	>65	616	0,740	0,260
251	>GL0445939	>65	621	0,838	0,162
252	>GL0771136	>65	623	0,916	0,084
253	>GL0522263	>65	624	0,793	0,207
254	>GL0553171	55-65	625	0,969	0,031
255	>GL0879194	>65	625	0,840	0,160
256	>GL1090758	>65	625	0,861	0,139
257	>GL0037781	55-65	626	0,896	0,104
258	>GL0388580	>65	626	0,795	0,205
259	>GL0395652	55-65	626	0,989	0,011
260	>GL0914216	55-65	627	0,841	0,159
261	>GL0762440	55-65	628	0,842	0,158

262	>GL0778232	55-65	628	0,723	0,277
263	>GL0297237	>65	636	0,970	0,030
264	>GL0596420	>65	643	0,466	0,534
265	>GL0123984	55-65	651	0,202	0,798
266	>GL0887867	55-65	654	0,947	0,053
267	>GL0814024	>65	660	0,688	0,312
268	>GL0539782	>65	664	0,972	0,028
269	>GL0703166	55-65	664	0,795	0,205
270	>GL0907357	55-65	668	0,948	0,052
271	>GL0975607	55-65	669	0,403	0,597
272	>GL0767758	>65	671	0,713	0,287
273	>GL0918382	55-65	671	0,993	0,007
274	>GL0389015	>65	684	0,693	0,307
275	>GL0701640	>65	684	0,970	0,030
276	>GL0686099	55-65	691	0,940	0,060
277	>GL0619664	>65	692	0,650	0,350
278	>GL0048177	55-65	693	0,963	0,037
279	>GL0206064	55-65	693	0,888	0,112
280	>GL0536626	>65	693	0,866	0,134
281	>GL0650486	>65	693	0,910	0,090
282	>GL0702665	>65	694	0,976	0,024
283	>GL0927080	55-65	694	0,910	0,090
284	>GL0806428	>65	698	0,894	0,106
285	>GL0673718	>65	700	0,944	0,056
286	>GL0897161	>65	700	0,986	0,014
287	>GL0496508	>65	705	0,068	0,932
288	>GL0515495	55-65	705	0,767	0,233
289	>GL0922134	>65	705	0,544	0,456
290	>GL0532948	>65	706	0,125	0,875
291	>GL0737456	>65	707	0,166	0,834
292	>GL0522030	55-65	709	0,935	0,065
293	>GL0598335	>65	709	0,622	0,378
294	>GL0901505	>65	709	0,890	0,110

295	>GL0638034	55-65	710	0,940	0,060
296	>GL0562319	55-65	711	0,966	0,034
297	>GL0152472	>65	713	0,968	0,032
298	>GL0826775	>65	713	0,103	0,897
299	>GL0530204	>65	714	0,047	0,953
300	>GL0505231	>65	715	0,633	0,367
301	>GL0706681	>65	715	0,987	0,013
302	>GL0424106	55-65	716	0,956	0,044
303	>GL0675892	>65	716	0,942	0,058
304	>GL0598966	>65	718	0,787	0,213
305	>GL0317758	55-65	719	0,745	0,255
306	>GL0400817	55-65	719	0,723	0,277
307	>GL0545616	55-65	720	0,651	0,349
308	>GL0559570	55-65	720	0,739	0,261
309	>GL0022225	55-65	721	0,964	0,036
310	>GL0885366	>65	721	0,957	0,043
311	>GL0853264	>65	722	0,887	0,113
312	>GL0926687	>65	722	0,865	0,135
313	>GL0678314	55-65	724	0,825	0,175
314	>GL0833645	>65	725	0,935	0,065
315	>GL0112527	>65	726	0,850	0,150
316	>GL0001633	>65	727	0,965	0,035
317	>GL0001460	>65	728	0,990	0,010
318	>GL0052557	>65	729	0,791	0,209
319	>GL0617030	55-65	729	0,976	0,024
320	>GL0751240	55-65	730	0,682	0,318
321	>GL0611376	>65	731	0,978	0,022
322	>GL0675014	55-65	731	0,927	0,073
323	>GL0928016	>65	732	0,871	0,129
324	>GL0039691	55-65	733	0,975	0,025
325	>GL0327862	>65	733	0,972	0,028
326	>GL0797055	>65	733	0,892	0,108
327	>GL0737821	55-65	734	0,901	0,099

328	>GL0564957	>65	735	0,972	0,028
329	>GL0291920	55-65	736	0,719	0,281
330	>GL0513466	55-65	736	0,323	0,677
331	>GL0954694	55-65	737	0,060	0,940
332	>GL0073007	>65	738	0,913	0,087
333	>GL0552229	55-65	738	0,980	0,020
334	>GL0756970	>65	738	0,983	0,017
335	>GL0064999	>65	739	0,277	0,723
336	>GL0631566	<55	739	0,563	0,437
337	>GL0743072	>65	739	0,974	0,026
338	>GL0818215	55-65	740	0,570	0,430
339	>GL0828815	>65	740	0,375	0,625
340	>GL0070238	>65	741	0,786	0,214
341	>GL0407733	>65	741	0,430	0,570
342	>GL0091490	>65	742	0,210	0,790
343	>GL0786844	>65	742	0,057	0,943
344	>GL0036783	55-65	743	0,273	0,727
345	>GL0519267	55-65	743	0,900	0,100
346	>GL0952332	55-65	743	0,377	0,623
347	>GL0653607	55-65	744	0,975	0,025
348	>GL0730021	>65	744	0,054	0,946
349	>GL0096660	55-65	745	0,774	0,226
350	>GL0207272	>65	746	0,920	0,080
351	>GL0341835	>65	746	0,858	0,142
352	>GL0567574	55-65	746	0,891	0,109
353	>GL0395788	55-65	747	0,500	0,500
354	>GL0026523	>65	748	0,944	0,056
355	>GL0175879	55-65	749	0,082	0,918
356	>GL0338225	>65	749	0,232	0,768
357	>GL0489002	>65	749	0,293	0,707
358	>GL0608187	>65	749	0,902	0,098
359	>GL0928587	>65	749	0,369	0,631
360	>GL0097473	>65	750	0,962	0,038

361	>GL0429316	55-65	750	0,649	0,351
362	>GL0658852	55-65	750	0,877	0,123
363	>GL0201852	55-65	751	0,966	0,034
364	>GL0293362	55-65	751	0,977	0,023
365	>GL0744351	55-65	751	0,645	0,355
366	>GL0213163	55-65	752	0,868	0,132
367	>GL0393087	>65	752	0,105	0,895
368	>GL0075078	>65	753	0,592	0,408
369	>GL0113329	55-65	753	0,796	0,204
370	>GL0515482	>65	753	0,974	0,026
371	>GL0024359	55-65	754	0,475	0,525
372	>GL0034702	>65	754	0,854	0,146
373	>GL0316302	55-65	754	0,936	0,064
374	>GL0446967	>65	754	0,971	0,029
375	>GL0003249	55-65	756	0,996	0,004
376	>GL0215843	>65	756	0,906	0,094
377	>GL0452183	55-65	756	0,941	0,059
378	>GL0284153	55-65	757	0,856	0,144
379	>GL0338611	>65	757	0,955	0,045
380	>GL0940709	>65	757	0,946	0,054
381	>GL0303480	>65	758	0,992	0,008
382	>GL0526283	>65	758	0,537	0,463
383	>GL0568771	>65	758	1,000	0,000
384	>GL0283870	>65	759	0,964	0,036
385	>GL0663988	55-65	759	0,864	0,136
386	>GL0140000	>65	760	0,914	0,086
387	>GL0256060	55-65	760	0,284	0,716
388	>GL0299299	>65	760	0,957	0,043
389	>GL0132781	55-65	761	0,781	0,219
390	>GL0582776	>65	761	0,937	0,063
391	>GL0655062	55-65	761	0,991	0,009
392	>GL0325559	55-65	762	0,006	0,994
393	>GL0053129	55-65	763	0,819	0,181

394	>GL0230128	55-65	763	0,802	0,198
395	>GL0514365	55-65	763	0,831	0,169
396	>GL0632411	55-65	763	0,824	0,176
397	>GL0141073	>65	765	0,571	0,429
398	>GL0273927	55-65	765	0,992	0,008
399	>GL0340400	55-65	765	0,906	0,094
400	>GL0423274	>65	765	0,738	0,262
401	>GL0608469	>65	765	0,966	0,034
402	>GL0843842	>65	765	0,753	0,247
403	>GL0026605	55-65	766	0,996	0,004
404	>GL0109862	>65	766	0,968	0,032
405	>GL0127563	>65	766	0,926	0,074
406	>GL0184156	>65	766	0,746	0,254
407	>GL0187597	55-65	766	0,989	0,011
408	>GL0382741	>65	766	0,988	0,012
409	>GL0456594	55-65	766	0,852	0,148
410	>GL0798219	55-65	766	0,674	0,326
411	>GL0832646	>65	766	0,993	0,007
412	>GL0099893	55-65	767	0,966	0,034
413	>GL0588393	>65	767	0,980	0,020
414	>GL0068310	>65	768	0,960	0,040
415	>GL0173810	>65	769	0,947	0,053
416	>GL0253283	>65	769	0,060	0,940
417	>GL0363614	55-65	769	0,841	0,159
418	>GL0659259	55-65	769	0,945	0,055
419	>GL0083649	>65	770	0,048	0,952
420	>GL0168347	55-65	770	0,984	0,016
421	>GL0204059	>65	770	0,875	0,125
422	>GL0106310	55-65	771	0,199	0,801
423	>GL0220407	>65	771	0,842	0,158
424	>GL0252547	55-65	771	0,994	0,006
425	>GL0365823	>65	771	0,935	0,065
426	>GL0114000	>65	772	0,064	0,936

427	>GL0039564	>65	773	0,206	0,794
428	>GL0216268	>65	773	1,000	0,000
429	>GL0497418	>65	773	0,973	0,027
430	>GL0047454	55-65	774	0,542	0,458
431	>GL0060441	>65	774	0,794	0,206
432	>GL0544800	55-65	774	0,607	0,393
433	>GL0657330	55-65	774	0,985	0,015
434	>GL0722658	>65	774	0,890	0,110
435	>GL0771530	55-65	774	0,952	0,048
436	>GL0093099	55-65	775	0,719	0,281
437	>GL0153524	55-65	775	0,918	0,082
438	>GL0215060	55-65	775	0,965	0,035
439	>GL0236348	>65	775	0,121	0,879
440	>GL0255037	55-65	775	0,583	0,417
441	>GL0332371	>65	775	0,993	0,007
442	>GL0429464	>65	775	0,989	0,011
443	>GL0512198	55-65	775	0,993	0,007
444	>GL0543029	>65	775	0,985	0,015
445	>GL0549825	55-65	775	0,947	0,053
446	>GL0142622	>65	776	1,000	0,000
447	>GL0220406	55-65	776	0,053	0,947
448	>GL0337071	55-65	776	0,386	0,614
449	>GL0425375	55-65	776	0,762	0,238
450	>GL0576440	>65	776	0,433	0,567
451	>GL0582322	>65	776	0,607	0,393
452	>GL0678094	55-65	776	0,978	0,022
453	>GL0071396	55-65	777	0,910	0,090
454	>GL0136774	55-65	777	0,232	0,768
455	>GL0096647	55-65	778	0,702	0,298
456	>GL0257977	55-65	778	0,947	0,053
457	>GL0261387	>65	778	0,538	0,462
458	>GL0579409	>65	778	0,377	0,623
459	>GL0030699	>65	779	0,435	0,565

460	>GL0141582	>65	779	0,990	0,010
461	>GL0277583	>65	779	0,756	0,244
462	>GL0426099	>65	779	0,576	0,424
463	>GL0709068	>65	779	0,799	0,201
464	>GL0134099	55-65	780	0,692	0,308
465	>GL0239484	>65	780	0,915	0,085
466	>GL0463690	>65	780	0,646	0,354
467	>GL0547964	>65	780	0,768	0,232
468	>GL0577287	55-65	780	0,996	0,004
469	>GL0089724	55-65	781	0,820	0,180
470	>GL0097023	55-65	781	0,709	0,291
471	>GL0137500	>65	781	0,869	0,131
472	>GL0281020	55-65	781	0,844	0,156
473	>GL0411370	>65	781	0,927	0,073
474	>GL0521530	55-65	781	0,903	0,097
475	>GL0650224	55-65	781	0,957	0,043
476	>GL0700692	55-65	781	0,794	0,206
477	>GL0086540	>65	782	0,793	0,207
478	>GL0099721	55-65	782	0,913	0,087
479	>GL0167272	>65	782	0,874	0,126
480	>GL0409824	>65	782	0,924	0,076
481	>GL0733035	>65	782	0,909	0,091
482	>GL0770843	>65	782	0,168	0,832
483	>GL0891722	55-65	782	0,991	0,009
484	>GL0411288	55-65	783	0,234	0,766
485	>GL0104388	55-65	784	0,727	0,273
486	>GL0847351	>65	784	0,590	0,410
487	>GL0092872	55-65	785	0,931	0,069
488	>GL0106816	55-65	785	0,188	0,812
489	>GL0120838	55-65	785	0,954	0,046
490	>GL0160578	>65	785	0,935	0,065
491	>GL0182620	>65	785	0,428	0,572
492	>GL0708017	55-65	785	0,625	0,375

493	>GL0023528	>65	786	0,983	0,017
494	>GL0047457	>65	786	0,796	0,204
495	>GL0055977	55-65	786	0,965	0,035
496	>GL0224783	>65	786	0,607	0,393
497	>GL0259440	55-65	786	0,242	0,758
498	>GL0270274	>65	786	0,790	0,210
499	>GL0285391	>65	786	0,813	0,187
500	>GL0687348	55-65	786	0,969	0,031
501	>GL0824285	>65	786	0,890	0,110
502	>GL0051962	55-65	787	0,968	0,032
503	>GL0487163	>65	787	0,983	0,017
504	>GL0619839	>65	787	0,833	0,167
505	>GL0800302	55-65	788	0,932	0,068
506	>GL0182221	>65	789	0,855	0,145
507	>GL0253287	>65	789	0,914	0,086
508	>GL0042421	>65	790	0,871	0,129
509	>GL0186086	>65	790	0,778	0,222
510	>GL0218658	55-65	790	0,643	0,357
511	>GL0224918	>65	790	0,987	0,013
512	>GL0698058	>65	790	0,893	0,107
513	>GL0128614	>65	791	0,863	0,137
514	>GL0605517	55-65	791	0,982	0,018
515	>GL0627622	55-65	791	0,796	0,204
516	>GL0030811	55-65	792	0,863	0,137
517	>GL0063380	>65	792	0,978	0,022
518	>GL0273453	>65	792	0,980	0,020
519	>GL0027942	55-65	793	0,950	0,050
520	>GL0062700	>65	793	0,835	0,165
521	>GL0428107	>65	793	0,718	0,282
522	>GL0504892	>65	793	0,990	0,010
523	>GL0635293	55-65	793	0,392	0,608
524	>GL0117941	55-65	795	0,830	0,170
525	>GL0269791	55-65	795	0,817	0,183

526	>GL0612948	55-65	795	0,991	0,009
527	>GL0477897	55-65	796	0,922	0,078
528	>GL0089214	>65	797	0,993	0,007
529	>GL0404686	>65	799	0,626	0,374
530	>GL0563286	55-65	799	0,582	0,418
531	>GL0044157	>65	800	0,968	0,032
532	>GL0683781	>65	800	0,902	0,098
533	>GL0071300	>65	801	0,877	0,123
534	>GL0088177	55-65	802	0,898	0,102
535	>GL0191639	55-65	802	0,881	0,119
536	>GL0691075	>65	803	0,917	0,083
537	>GL0586869	55-65	804	0,978	0,022
538	>GL0808633	>65	804	0,788	0,212
539	>GL0229325	>65	805	0,950	0,050
540	>GL0343990	55-65	806	0,542	0,458
541	>GL0499098	>65	806	0,915	0,085
542	>GL0109676	>65	807	0,970	0,030
543	>GL0260168	>65	807	0,979	0,021
544	>GL0290983	>65	807	0,985	0,015
545	>GL0785644	55-65	807	0,936	0,064
546	>GL0033216	>65	808	0,887	0,113
547	>GL0160963	>65	808	0,766	0,234
548	>GL0311508	>65	808	0,761	0,239
549	>GL0409589	>65	808	0,305	0,695
550	>GL0311447	55-65	809	0,514	0,486
551	>GL0476095	55-65	810	0,700	0,300
552	>GL0055435	>65	811	0,793	0,207
553	>GL0030982	55-65	812	0,910	0,090
554	>GL0352458	>65	813	0,975	0,025
555	>GL0352672	55-65	813	0,980	0,020
556	>GL0048970	>65	814	0,756	0,244
557	>GL0301606	>65	814	0,964	0,036
558	>GL0125444	>65	816	0,997	0,003

559	>GL0559872	55-65	816	0,261	0,739
560	>GL0271670	>65	817	0,673	0,327
561	>GL0020327	>65	818	0,616	0,384
562	>GL0559018	55-65	818	0,966	0,034
563	>GL0373558	55-65	819	0,897	0,103
564	>GL0177166	>65	820	0,867	0,133
565	>GL0397036	55-65	821	0,955	0,045
566	>GL0432083	>65	821	0,967	0,033
567	>GL0726261	>65	822	0,934	0,066
568	>GL0514179	>65	823	0,995	0,005
569	>GL0650433	55-65	823	0,996	0,004
570	>GL0590974	55-65	824	0,993	0,007
571	>GL0824682	55-65	824	0,995	0,005
572	>GL0274898	55-65	825	0,984	0,016
573	>GL0227773	>65	826	0,985	0,015
574	>GL0263654	55-65	826	0,974	0,026
575	>GL0394361	>65	827	0,581	0,419
576	>GL0572267	55-65	827	0,974	0,026
577	>GL0349935	>65	829	0,909	0,091
578	>GL0375480	>65	829	0,900	0,100
579	>GL0320501	55-65	830	0,836	0,164
580	>GL0368628	55-65	830	0,929	0,071
581	>GL0821308	>65	837	0,711	0,289
582	>GL0293981	>65	839	0,945	0,055
583	>GL0499349	>65	840	0,648	0,352
584	>GL0710997	55-65	841	0,168	0,832
585	>GL0193681	>65	842	0,506	0,494
586	>GL0367947	>65	842	0,934	0,066
587	>GL0693756	>65	847	0,750	0,250
588	>GL0416264	>65	849	0,246	0,754
589	>GL0601864	55-65	849	0,856	0,144
590	>GL0217892	55-65	851	1,000	0,000
591	>GL0168813	>65	852	0,977	0,023

592	>GL0601868	>65	854	0,244	0,756
593	>GL0619037	55-65	854	0,879	0,121
594	>GL0388061	55-65	856	0,761	0,239
595	>GL0510502	55-65	857	0,464	0,536
596	>GL0549210	>65	857	0,746	0,254
597	>GL0691204	>65	858	0,923	0,077
598	>GL0553991	>65	861	0,787	0,213
599	>GL0459001	55-65	862	0,973	0,027
600	>GL0561446	>65	862	0,903	0,097
601	>GL0601716	55-65	862	0,696	0,304
602	>GL0426313	>65	865	0,243	0,757
603	>GL0515814	>65	865	0,541	0,459
604	>GL0631348	55-65	866	0,995	0,005
605	>GL0610397	>65	868	0,094	0,906
606	>GL0234409	>65	869	0,876	0,124
607	>GL0337660	55-65	869	0,555	0,445
608	>GL0564341	>65	869	0,834	0,166
609	>GL0735475	>65	874	0,670	0,330
610	>GL0346266	55-65	878	0,640	0,360
611	>GL0209413	55-65	880	0,839	0,161
612	>GL0705847	>65	884	0,080	0,920
613	>GL0790368	55-65	887	0,524	0,476
614	>GL0376199	55-65	906	0,291	0,709
615	>GL0048340	55-65	909	0,720	0,280
616	>GL0121362	>65	909	0,869	0,131
617	>GL0567534	>65	912	0,460	0,540
618	>GL0159163	55-65	914	0,049	0,951
619	>GL0326938	55-65	914	0,739	0,261
620	>GL0569660	55-65	916	0,640	0,360
621	>GL0351582	55-65	917	0,641	0,359
622	>GL0660702	55-65	918	0,592	0,408
623	>GL0366865	55-65	920	0,149	0,851
624	>GL0129562	>65	921	0,725	0,275

625	>GL0392841	55-65	925	0,026	0,974
626	>GL0215958	55-65	928	0,932	0,068
627	>GL0559849	>65	933	0,763	0,237
628	>GL0203999	>65	935	0,803	0,197
629	>GL0333780	>65	936	0,856	0,144
630	>GL0166255	>65	942	0,861	0,139
631	>GL0063477	55-65	946	0,843	0,157
632	>GL0146845	>65	946	0,663	0,337
633	>GL0586240	>65	951	0,969	0,031
634	>GL0266450	>65	952	0,436	0,564
635	>GL0270276	>65	958	0,170	0,830
636	>GL0321409	>65	959	0,134	0,866
637	>GL0058591	>65	963	0,676	0,324
638	>GL0437696	>65	963	0,088	0,912
639	>GL0334404	>65	967	0,748	0,252
640	>GL0681391	>65	973	0,033	0,967
641	>GL0272060	>65	974	0,196	0,804
642	>GL0042423	55-65	978	0,280	0,720
643	>GL0093389	>65	979	0,308	0,692
644	>GL0111993	>65	979	0,832	0,168
645	>GL0055434	>65	982	0,421	0,579
646	>GL0454642	55-65	987	0,229	0,771
647	>GL0494364	>65	995	0,493	0,507
648	>GL0271671	>65	1000	0,301	0,699
649	>GL0311507	>65	1003	0,083	0,917
650	>GL0026884	>65	1005	0,929	0,071
651	>GL0129565	>65	1009	0,065	0,935
652	>GL0323690	>65	1012	0,509	0,491
653	>GL0146986	55-65	1014	0,808	0,192
654	>GL0018299	55-65	1019	0,872	0,128
655	>GL0241564	>65	1020	0,176	0,824
656	>GL0184556	>65	1021	0,527	0,473
657	>GL0238182	55-65	1023	0,236	0,764

658	>GL0429315	>65	1024	0,728	0,272
659	>GL0004097	55-65	1030	0,370	0,630
660	>GL0412746	55-65	1033	0,124	0,876
661	>GL0233895	<55	1035	0,487	0,513
662	>GL0233358	>65	1036	0,150	0,850
663	>GL0036979	55-65	1037	0,660	0,340
664	>GL0203190	>65	1058	0,505	0,495
665	>GL0585941	55-65	1067	0,116	0,884
666	>GL0566740	55-65	1144	0,814	0,186
667	>GL0461710	55-65	1146	0,303	0,697
668	>GL0036730	55-65	1259	0,851	0,149
669	>GL0391301	55-65	1262	0,236	0,764
670	>GL0260543	55-65	1264	0,123	0,877
671	>GL0123034	55-65	1266	0,561	0,439
672	>GL0087784	55-65	2075	0,319	0,681
673	>GL0018300	>65	2253	0,119	0,881