BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM HÀ NỘI

NGUYỄN MINH GIANG

KHAI THÁC DỮ LIỆU DNA ĐA HỆ GEN, BIỂU HIỆN VÀ NGHIÊN CỨU TÍNH CHẤT CỦA β-XYLOSIDASE TỪ VI SINH VẬT RUỘT MỐI *Coptotermes gestroi* Ở VIỆT NAM

LUẬN ÁN TIẾN SĨ SINH HỌC

Hà Nội, tháng 01 năm 2018

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM HÀ NỘI

NGUYỄN MINH GIANG

KHAI THÁC DỮ LIỆU DNA ĐA HỆ GEN, BIỂU HIỆN VÀ NGHIÊN CỨU TÍNH CHẤT CỦA β-XYLOSIDASE TỪ VI SINH VẬT RUỘT MỐI *Coptotermes gestroi* Ở VIỆT NAM

Chuyên ngành: DI TRUYỀN HỌC
Mã số: 62. 42. 01. 21

LUẬN ÁN TIẾN SĨ SINH HỌC

NGƯỜI HƯỚNG DẪN KHOA HỌC:
1. GS.TS. TRƯƠNG NAM HẢI
2. PGS.TS. ĐẶNG HỮU LANH

Hà Nội, tháng 01 năm 2018

LỜI CAM ĐOAN

Tôi xin cam đoan:

Đây là công trình nghiên cứu do chính tôi và các cộng sự tại Phòng Kỹ thuật di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam thực hiện. Các số liệu và kết quả trong luận án là trung thực, một phần đã được công bố trên các tạp chí chuyên ngành với sự đồng ý cho phép của đồng tác giả. Phần còn lại chưa được công bố trong bất kỳ công trình nào.

Hà Nội, ngày 9 tháng 9 năm 2017

Tác giả

Nguyễn Minh Giang

LỜI CẢM ƠN

Trong quá trình học tập và nghiên cứu, tôi đã nhận được sự chỉ bảo tận tình của GS.TS Trương Nam Hải, TS Đỗ Thị Huyền, Phòng Kỹ thuật Di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học Việt Nam và PGS.TS Đặng Hữu Lanh, Khoa Sinh học, Trường Đại học Sư phạm Hà Nội.

Trong quá trình thực nghiệm nghiên cứu, tôi đã nhận được sự giúp đỡ và tạo điều kiện của cán bộ Phòng Kỹ thuật Di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, thày cô giáo, bạn đồng nghiệp của Bộ môn Di truyền học, khoa Sinh học, Trường Đại học Sư phạm Hà Nội.

Tôi đã nhận được sự ủng hộ và giúp đỡ hết mình của bạn bè đồng nghiệp tại Trường Đại học Sư phạm Thành phố Hồ Chí Minh trong suốt bốn năm học tập và nghiên cứu.

Tôi sẽ không thể hoàn thành luận án này nếu thiếu gia đình nhỏ nơi mà chồng và hai con trai tôi đã dành cho tôi tình yêu, truyền cho tôi nghị lực, chia sẻ với tôi mọi khó khăn và tạo cho tôi cảm hứng sáng tạo trong cuộc sống.

Theo suốt hành trình học tập và nghiên cứu của tôi luôn có cha mẹ, anh chị em ở bên cạnh động viên, khuyến khích, dành cho tôi tình yêu vô điều kiện, giúp tôi đi đến thành công.

Tôi sẽ luôn ghi nhớ và cảm ơn tới thày cô, gia đình, bạn bè và đồng nghiệp đã giúp tôi hoàn thành công trình nghiên cứu này.

MỤC LỤC

LỜI CAM ĐOAN i

LỜI CẢM ƠN ii

MỤC LỤC iii

DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT vii

DANH MỤC BẢNG xi

DANH MỤC HÌNH xii

MỞ ĐẦU 1

1. Lý do chọn đề tài 1

2. Mục tiêu 3

2.1. Mục tiêu chung 3

2.2. Mục tiêu cụ thể 3

3. Nội dung nghiên cứu 3

4. Đối tượng 4

5. Phạm vi nghiên cứu 4

6. Ý nghĩa khoa học và thực tiễn của đề tài 4

7. Đóng góp mới của luận án 4

8. Nơi thực hiện đề tài luận án 5

Chương 1. TỔNG QUAN NGHIÊN CỨU 6

1.1. LIGNOCELLULOSE VÀ QUÁ TRÌNH CHUYỂN HÓA 6

1.1.1. Lignocellulose 6

1.1.2. Sự chuyển hóa lignocellulose 8

1.2. METAGENOMICS VÀ CÔNG CỤ TIN SINH HỌC KHAI THÁC DỮ LIỆU DNA ĐA HỆ GEN 10
1.2.1. Metagenomics 10
1.2.2. Một số công cụ tin sinh sử dụng để phân tích số liệu 13
1.2.3. Các nguồn dữ liệu 19
1.2.4. Mẫu dò DNA và ứng dụng 21
1.3. ENZYME β–xylosidase 22
1.3.1. Đặc điểm chung 22
1.3.2. Mô hình hoạt động 23
1.3.3. Cấu trúc không gian 24
1.3.4. Hoạt tính của β–xylosidase 25
1.3.5. Ứng dụng của β–xylosidase 26
1.3.6. Nguồn cung cấp β–xylosidase 26
1.4. KHU HỆ VI SINH VẬT VÀ ENZYME CHUYỂN HÓA LIGNOCELLULOSE 27
1.4.1. Một số khu hệ vi sinh vật chuyển hóa lignocellulose 27
1.4.2. Hệ vi sinh vật và enzyme thủy phân lignocellulose trong ruột mối 28
1.4.3. Tổng quan nghiên cứu về đa dạng vi sinh vật và enzyme chuyển hóa lignocellulose trong ruột mối *C. gestroi* ở Việt Nam 32
Chương 2. ĐỐI TƯỢNG, VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU 36
2.1. ĐỐI TƯỢNG VÀ VẬT LIỆU 36
2.1.1. Đối tượng 36
2.1.2. Hóa chất và thiết bị máy móc 37
2.2. PHƯƠNG PHÁP NGHIÊN CỨU 39
2.2.1. Phương pháp xây dựng mẫu dò 39

2.2.2. Các phương pháp xử lý số liệu bằng phần mềm tin sinh học 42

2.2.3. Các phương pháp vi sinh 45

2.2.4. Các phương pháp sinh học phân tử 45

Chương 3. KẾT QUẢ VÀ THẢO LUẬN 53

3.1. NGHIÊN CỨU ĐA DẠNG HỆ VI KHUẨN VÀ LIGNOCELLULASE THEO HỆ THỐNG PHÂN LOẠI CỦA CAZY TỪ DỮ LIỆU DNA ĐA HỆ GEN CỦA HỆ VI KHUẨN TRONG RUỘT MỐI *C. gestroi* 53

3.1.1. Nghiên cứu đa dạng hệ vi khuẩn sống trong ruột mối *C. gestroi* 53

3.1.2. Nghiên cứu đa dạng lignocellulase của vi sinh vật sống trong ruột mối *C. gestroi* 57

3.2. XÂY DỰNG PHƯƠNG PHÁP TÌM KIẾM GEN MÃ HÓA β–XYLOSIDASE TỪ DỮ LIỆU DNA ĐA HỆ GEN CỦA VI SINH VẬT TRONG RUỘT MỐI *C. gestroi* 62

3.2.1. Xác định các họ GH chứa β–xylosidase theo CAZY 62

3.2.2. Tìm kiếm các trình tự axit amin của β–xylosidase đã được nghiên cứu trong thực nghiệm 64

3.2.3. Nhóm các trình tự đã tìm kiếm được để xác định vùng tương đồng bằng ClustalW – PBIL 66

3.2.4. Xây dựng mẫu dò và giá trị tham chiếu 70

3.2.5. Khai thác trình tự gen mã hóa β–xylosidase bằng mẫu dò từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật ruột mối *C. gestroi* 73

3.2.6. Khảo sát cấu trúc bậc 3 của các β–xylosidase đã khai thác bằng mẫu dò ... 75

3.2.7. Dự đoán cấu trúc và chức năng của các gen mã hóa β–xylosidase bằng một số công cụ tin sinh học 79

3.2.8. Một số dự đoán chi tiết gen GL0112518 mã hóa β–xylosidase (*Xbx14*) 81

3.3. BIỂU HIỆN GEN *Xbx14* VÀ NGHIÊN CỨU TÍNH CHẤT CỦA β-XYLOSIDASE 85

3.3.1. Biểu hiện gen *Xbx14* 85

3.3.2. Tinh chế *Xbx14* 99

3.3.3. Nghiên cứu tính chất của *Xbx14* 104

KẾT LUẬN VÀ KIẾN NGHỊ 113

Kết luận 113

Kiến nghị 113

DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ 114

TÀI LIỆU THAM KHẢO 115

PHỤ LỤC 139

DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT

Tên viết tắt	Tên tiếng Anh	Tên tiếng Việt
APS	Ammonium persulfate	Ammonium persulfate
BGI	Beijing Genomics Institute	Viện nghiên cứu gen Bắc Kinh
BLAST	Basic Local Alignment Search Tool	Công cụ so sánh mức độ tương đồng về trình tự nucleotide/axit amin trực tuyến của NCBI
Bp	Base pair	Cặp base
BSA	Bovine serum albumin	Albumin huyết thanh bò
CAZY	Carbohydrate-Active enzymes	Cơ sở dữ liệu các họ protein chuyển hóa cacbohydrate
COG/ KOG	Clusters of Orthologous/ eukaryotic orthologous groups Group	Cơ sở dữ liệu protein của sinh vật nhân sơ và nhân chuẩn đơn bào/CSDL từ 7 hệ gen sinh vật nhân chuẩn: 3 loài động vật, 1 loài thực vật, *Arabidopsis thaliana*, 2 loài nấm và các ký sinh trùng nội bào
CSDL	-	Cơ sở dữ liệu
dNTP	2'-deoxyribonucleoside 5'-triphosphate	2'-deoxyribonucleoside 5'-triphosphate
DNA	Deoxyribonucleic acid	Axit deoxyribonucleic
EDTA	Ethylene diamine tetraacetic acid	Axit ethylene diamine tetraacetic
EBI	European Bioinformatics Institute	Viện nghiên cứu tin sinh học Châu Âu
EMBL	European Molecular Biology Laboratory	Phòng thí nghiệm phân tử sinh học Châu Âu

FGA	Functional gen arrays	Mảng gen chức năng
Expasy	Expert Protein Analysis System	Hệ thống phần mềm phân tích protein
eggNOG	Evolutionary genalogy of gens: Non-supervised Orthologous Groups	Cơ sở dữ liệu chứa các nhóm orthologous
GH	Glycoside hydrolase family	Họ enzyme thủy phân liên kết Glycoside
GO	Gen ontology	Bản thể gen
HiSpOD	High Specific Oligo Design	Thiết kế mẫu dò có độ đặc hiệu cao
His (H)	Histidin	Axit amin histidin
HTS	High Throughput Sequencing	Giải trình tự thông lượng cao
ID	Identification	Nhận biết
IPTG	Isopropy-β-D-thiogalactosidase	Chất cảm ứng Isopropy-β-D-thiogalactosidase
Kb	Kilo base	Đơn vị đo kích thước của axit nucleic
Kda	Kilodalton	Đơn vị đo trọng lượng của protein
KEGG	Kyoto Encyclopedia of Gens and Genomes	Cơ sở dữ liệu về hệ gen, enzyme và các sản phẩm sinh học
Km	Michaelis constant	Hằng số Michaelis
LB	Luria-Betani	Môi trường nuôi cấy LB
LBA/ LBC	Luria-Betani ampicillin/ Luria-Betani chloramphenicol	Môi trường nuôi cấy LB bổ sung ampicillin/chloramphenicol
MCS	Multi cloning site	Vùng đa nối
Mb	Megabyte	Megabyte
MEGAN	MEtaGenomic Analyser	Phần mềm phân tích trình tự đa hệ gen

		MEGAN
NCBI	National Center for Biotechnology Information	Trung tâm quốc gia về thông tin công nghệ sinh học (Mỹ)
NGS	Next Generation Sequencing	Giải trình tự gen thế hệ mới
NIH	National Institutes of Health	Viện Y tế quốc gia (Mỹ)
NR	Non-Redundant	Cơ sở dữ liệu chứa các trình tự non-redundant từ cơ sở dữ liệu NCBI
OD	Optimal density	Mật độ quang học
ORF	Open Reading Frame	Khung đọc mở
PBS	Phosphate-buffered saline	Đệm phosphate
PCR	Polymarase Chain Reaction	Phản ứng trùng hợp
PBD	Protein Data Bank	Ngân hàng protein
pI	isoelectrics point	Điểm đẳng điện
PHYRE	Protein Homology/analogY Recognition Engine	Protein tương đồng / tương tự
pNPX	4-nitrophenyl β-D-xylopyranoside	4-nitrophenyl β-D-xylopyranoside
Prim.cons	Prime consensus	Trình tự bảo tồn cao nhất
PSI-BLAST	Position-Specific Iterated BLAST	Công cụ so sánh mức độ tương đồng về trình tự nucleotide/axit amin trực tuyến của NCBI, sử dụng vị trí lặp đặc hiệu
RBS	Ribosome Binding Site	Vùng bám của Ribosome
RNA	Ribonucleic acid	Axit ribonucleic
SDS	Sodium dodecyl sulphate	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis	Điện di trên gel polyacrylamide biến tính
SIB	Swiss Institute of Bioinformatics	Viện nghiên cứu tin sinh học của Thụy Sỹ

SVM	Support Vector Machine	Vector hỗ trợ phân tích tự động
SwiProt	Swiss Protein	Dữ liệu protein của Thụy Sỹ chứa thông tin của protein tổng hợp từ các tài liệu đã phân tích và tính toán
TBI	Taiwan Bioinformatic Institute	Viện nghiên cứu tin sinh học Đài Loan
TEMED	N, N, N', N'-tetramethylethylenediamine	N, N, N', N'-tetramethylethylenediamine
TG	Termite gut	Ruột mối chứa các vi khuẩn với trùng roi
Tm	Temperature Melting	Nhiệt độ nóng chảy
w/v	Weight/volume	Khối lượng/thể tích
V_{max}	Maximum velocity	Vận tốc tối đa
Xbx14		Enzyme beta–xylosidase
UniProt	Universal Protein Resource	Nguồn protein phổ biến cung cấp các thông tin về trình tự và chức năng của protein

DANH MỤC BẢNG

Bảng 2.1. Thành phần và nồng độ của gel polyacrylamide 49

Bảng 3.1. Bảng so sánh các họ enzyme thủy phân lignocellulose (GH) của vi sinh vật trong ruột mối *C. gestroi* với một số loài mối khác…… 57

Bảng 3.2. Bảng thống kê số lượng ORF và họ GH của β–xylosidase trong ruột *C. gestroi*. 61

Bảng 3.3. Bảng các họ GH chứa β–xylosidase theo CAZY. 63

Bảng 3.4. Bảng tổng hợp dữ liệu đã được nghiên cứu chi tiết về β–xylosidase. 64

Bảng 3.5. Bảng so sánh độ tương đồng của mẫu dò với các trình tự thuộc GH43. 71

Bảng 3.6. Bảng tổng hợp ước đoán cấu trúc bậc ba của trình tự 4, 5, 9 bằng công cụ Swiss model. 72

Bảng 3.7. Bảng so sánh kết quả khai thác bằng mẫu dò với dự đoán của BGI. 73

Bảng 3.8. Bảng tổng hợp ước đoán cấu trúc bậc ba của β–xylosidase đã chọn bằng công cụ Swiss model. 75

Bảng 3.9. Bảng tổng hợp độ bao phủ và tương đồng với mẫu dò của các gen đã chọn mã hóa β–xylosidase. 80

Bảng 3.10. Bảng kết quả dự đoán pH và nhiệt độ hoạt động của các mã gen hoàn thiện mã hóa β–xylosidase. 81

DANH MỤC HÌNH

Hình 1.1. Cấu trúc của lignocellulose 6

Hình 1.2. Sơ đồ mô hình chuyển hóa sinh khối thành các sản phẩm 8

Hình 1.3. Vị trí tác dụng của một số hemicellulase trên mạch xylan 23

Hình 1.4. Mô hình glycoside hydrolase thủy phân liên kết glycoside 24

Hình 1.5. Cấu trúc không gian của enzyme thuộc GH43 ở *Cellvibrio japonicus* 24

Hình 1.6. Hoạt tính của β–xylosidase 25

Hình 1.7. Cấu tạo ruột mối 30

Hình 2.1. Các bước cơ bản trong nghiên cứu của luận án 39

Hình 2.2. Sơ đồ các bước cơ bản trong xây dựng mẫu dò 40

Hình 2.3. Đường chuẩn pNP được đo ở bước sóng 405 nm 51

Hình 2.4. Sự phụ thuộc của tốc độ phản ứng vào nồng độ cơ chất theo Lineweaver-Burk 52

Hình 3.1. Biểu đồ thống kê 20 loài vi khuẩn có ORF lớn nhất trong ruột *C. gestroi* ..54

Hình 3.2. Kết quả so sánh sự tương đồng của 11 trình tự mã hóa β–xylosidase và xây dựng mẫu dò cho β–xylosidase thuộc họ GH43 69

Hình 3.3. Trình tự mẫu dò của β–xylosidase thuộc họ GH43 70

Hình 3.4. Kết quả dự đoán tương đồng đặc hiệu trình tự và các gốc hoạt động của các trình tự được lựa chọn bằng mẫu dò GH43 75

Hình 3.5. Cấu trúc bậc ba của các β-xylosidase khuôn (template) 2exk.1.A (A), 1yrz.1.A (B), 1yi7.1.A (C), 3c7g.1.A (D) họ GH43 tương đồng với các trình tự được khai thác bằng mẫu dò sử dụng Swiss model 78

Hình 3.6. Kết quả dự đoán chức năng *Xbx14* bằng BLASTP 82

Hình 3.7. Mô hình cấu trúc không gian (A) dựa vào khuôn c1ylfC và trung tâm hoạt động (B) của enzyme *Xbx14* 83

Hình 3.8. Sơ đồ nguồn gốc mã gen GL0112518 84

Hình 3.9. Kết quả biến nạp vector pET22Xbx (A) và điện di đồ phân tích sản phẩm tách dòng gen *Xbx14* (B). 85

Hình 3.10. Sơ đồ vị trí cắt của enzyme hạn chế trên vector pET22Xbx (A), điện di sản phẩm PCR (B) và sản phẩm cắt pET22Xbx bằng *Nco*I và *Xho*I, *Hinc*II (C) 86

Hình 3.11. Điện di đồ phân tích sản phẩm protein đồng biểu hiện gen *Xbx14* với chaperone trong *E. coli Rosetta 1* trên gel SDS-PAGE 12,6% có SDS 89

Hình 3.12. Kết quả thử hoạt tính thô của *Xbx14* 89

Hình 3.13. Mật độ tế bào biểu hiện *E. coli Rosetta 1* (DE3) mang pET22Xbx cảm ứng IPTG ở các nhiệt độ 20°C, 25°C và 30°C 92

Hình 3.14. Điện di đồ phân tích sản phẩm biểu hiện *Xbx14* trong *E. coli Rosetta 1* sau 6 giờ cảm ứng 0,5 mM IPTG ở 20°C, 25°C, 30°C trên gel polyacrylamide 12,6% có SDS 92

Hình 3.15. Mật độ tế bào biểu hiện *E. coli Rosetta 1* (DE3) mang pET22Xbx cảm ứng IPTG ở các nồng độ từ 0,0 đến 1,5 mM 94

Hình 3.16. Điện di đồ phân tích protein biểu hiện *Xbx14* tổng số trong tế bào *E. coli Rosetta 1* cảm ứng 09 nồng độ IPTG trên gel polyacrylamide 12,6 % có SDS 94

Hình 3.17. Điện di đồ phân tích sản phẩm biểu hiện *Xbx14* tan trong tế bào *E. coli Rosetta 1* (DE3) cảm ứng 09 nồng độ IPTG trên gel polyacrylamide 12,6% có SDS . 95

Hình 3.18. Mật độ tế bào biểu hiện *E. coli Rosetta 1* (DE3) mang pET22Xbx cảm ứng ở các OD_{600} khác nhau 96

Hình 3.19. Điện di đồ phân tích sản phẩm biểu hiện của *Xbx14* trong tế bào *E. coli Rosetta 1* (DE3) cảm ứng IPTG tại các OD_{600} = 0,7; 1,0; 1,5 trên gel polyacrylamide 12,6% có SDS 96

Hình 3.20. Mật độ tế bào biểu hiện *E. coli Rosetta 1* (DE3) mang pET22Xbx thu mẫu sau 5, 8, 15 và 20 giờ cảm ứng 98

Hình 3.21. Điện di đồ phân tích sản phẩm biểu hiện của *Xbx14* trong tế bào *E. coli Rosetta 1* (DE3) thu mẫu sau 5, 8, 15 và 20 giờ cảm ứng IPTG trên gel polyacrylamide 12,6% có SDS 98

Hình 3.22. Điện di đồ phân tích kết quả tủa protein *Xbx14* biểu hiện trong tế bào *E. coli Rosetta 1* (DE3) bằng muối $(NH_4)_2SO_4$ theo phân đoạn 5% (A), 3% (B), 1% (C), 0,2% (D), 0,02% (E) và rửa tủa bằng đệm phosphate pH = 9 (F) 101

Hình 3.23. Kết quả kiểm tra độ sạch của *Xbx14* sau khi tinh chế bằng muối $(NH_4)_2SO_4$ và rửa bằng đệm phosphate (pH = 9) 103

Hình 3.24. Kết quả thử hoạt tính của *Xbx14* bằng cơ chất đặc hiệu pNPX 105

Hình 3.25. Hoạt tính *Xbx14* trên các cơ chất khác nhau 105

Hình 3.26. Ảnh hưởng của nhiệt độ đến hoạt tính *Xbx14* 107

Hình 3.27. Ảnh hưởng của pH đến hoạt tính của *Xbx14* ở 60^oC 107

Hình 3.28. Hoạt tính của *Xbx14* khi xử lý ở các nhiệt độ khác nhau 109

Hình 3.29. Ảnh hưởng một số ion kim loại và hóa chất lên hoạt tính của *Xbx14* 109

Hình 3.30. Sự phụ thuộc của tốc độ phản ứng của enzyme vào nồng độ cơ chất pNPX theo Linewever-Burk 112

MỞ ĐẦU

1. Lý do chọn đề tài

Lignocellulose là thành phần chính của sinh khối thực vật trong tự nhiên được thải ra với một khối lượng rất lớn. Ở nước ta, ước tính hàng năm, nguồn sinh khối này từ phế phẩm nông nghiệp như rơm, rạ, bã cây mía,… lên đến 50 triệu tấn. Đây chính là nguyên liệu của nhiều ngành công nghiệp để tạo ra các sản phẩm có giá trị như cồn sinh học, chất dinh dưỡng bổ sung cho người và vật nuôi,.... Hiện nay biện pháp xử lý các phế phẩm này chủ yếu là đốt gây lãng phí rất lớn, ảnh hưởng tiêu cực đến môi trường sống và sức khoẻ con người. Trong các phương pháp xử lý sinh khối thực vật, thì phương pháp sử dụng tác nhân vật lý và hóa học có nhiều nhược điểm như hiệu suất tạo đường đơn thấp, yêu cầu các thiết bị chuyên dụng phù hợp với giá thành cao, các dung môi vẫn tồn dư trong sản phẩm và gây ra các vấn đề về môi trường,…. Phương pháp xử lý bằng enzyme có nhiều ưu điểm vượt trội như chủ động điều khiển các phản ứng và sản phẩm của từng giai đoạn chuyển hóa, không tồn dư hóa chất nên thân thiện với môi trường,…. Vì vậy việc tìm kiếm và sản xuất nguồn enzyme thủy phân sinh khối thực vật, vẫn đang được quan tâm ở hầu hết các quốc gia trên thế giới trong đó có Việt Nam.

Trong các thành phần của lignocellulose thì hemicellulose có cấu trúc phức tạp và không đồng nhất, gồm nhiều loại phân tử sinh học khác nhau. Do đó để chuyển hóa triệt để hemicellulose cần nhiều nhóm enzyme như: xylanase, β-xylosidase, arabinofuranosidase, α-glucuronidase, acetyl xylan esterase, α-galactosidase, arabinanase, … Beta-xylosidase là enzyme có khả năng cắt ở vị trí đầu không khử trên mạch chính xylan và các phân tử đường đôi xylobiose thành đường đơn xylose, nên rất cần thiết trong hệ thống enzyme chuyển hóa hemicellulose.

Mối là loài có khả năng tiêu hóa lignocellulose rất hiệu quả do sự hỗ trợ của hàng loạt enzyme như cellulase, hemicellulase có nguồn gốc từ vi sinh vật sống trong ruột mối. Việc tìm kiếm các lignocellulase từ vi sinh vật sống trong ruột mối là một hướng được nhiều nghiên cứu quan tâm. Tuy nhiên hầu hết các vi sinh vật

trong ruột mối vẫn chưa nuôi cấy được, nên việc khai thác nguồn lignocellulase trước đây bị hạn chế. Hiện nay với sự ra đời của kỹ thuật Metagenomics cho phép giải trình tự toàn bộ hệ gen của quần xã sinh vật thu được trực tiếp từ mẫu môi trường. Kết quả giải trình DNA đa hệ gen tạo ra nguồn dữ liệu khổng lồ khó có thể xử lý bằng phương pháp thủ công, mà nhờ sự hỗ trợ của hàng loạt công cụ tin sinh học hiện đại đang liên tục xuất hiện, cập nhật và cải tiến. Đây chính là cơ sở thúc đẩy sự ra đời các dự án nghiên cứu về DNA đa hệ gen của hệ vi sinh vật có khả năng chuyển hóa lignocellulose, trong đó có các nghiên cứu về DNA đa hệ gen của vi sinh vật trong ruột mối.

Từ hỗ trợ tài chính của đề tài hợp tác Việt Nam - Nhật Bản "Phân lập gen mã hóa enzyme lignocellulolytic vi sinh vật trong ruột mối ở Việt Nam bằng kỹ thuật Metagenomics" giai đoạn 2012-2015, DNA đa hệ gen của vi sinh vật sống trong ruột mối *Coptotermes gestroi* được tách chiết và đọc trình tự. Sau khi xử lý số liệu đã xác định được 587 ORF mã hóa cho enzyme thủy phân lignocellulose. Cách tìm kiếm và lựa chọn gen mã hóa lignocellulase từ 587 ORF này đã sử dụng phương pháp thủ công như sau: Đầu tiên là chọn từng gen mã hóa enzyme thủy phân lignocellulose theo ước đoán ban đầu của công ty giải trình tự. Sau đó khảo sát vùng bảo thủ của protein và thiết lập sơ bộ cây phát sinh từng loại enzyme. Cuối cùng sẽ lựa chọn ORF theo tiêu chí (1) trình tự mới, (2) trung tâm hoạt tính rõ ràng, đặc hiệu và (3) trình tự đơn giản dễ biểu hiện. Tuy nhiên cách chọn gen này mất rất nhiều thời gian vì dữ liệu DNA đa hệ gen quá lớn. Hơn nữa, phần lớn gen đã được chọn rất khó biểu hiện thành công. Ví dụ, sau khi chọn được 8 gen để đưa vào biểu hiện thì 06 gen không biểu hiện hoặc biểu hiện không tan và chỉ có 02 gen biểu hiện, nhưng enzyme có hoạt tính yếu.

Những hạn chế của nghiên cứu trước đây đã đặt ra yêu cầu phải xây dựng được một phương pháp hiệu quả, để tìm kiếm nhanh được gen đích mã hóa đúng lignocellulase từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* và phải có hoạt tính đúng của enzyme mục tiêu sau khi biểu hiện. Ngoài ra, sự đa dạng hệ vi khuẩn và enzyme thủy phân lignocellulose trong ruột mối chưa được nghiên cứu. Đây chính là lý do để chúng tôi tiến hành đề tài: "**KHAI THÁC DỮ LIỆU DNA ĐA HỆ GEN, BIỂU HIỆN VÀ NGHIÊN CỨU TÍNH CHẤT CỦA**

β-XYLOSIDASE TỪ VI SINH VẬT RUỘT MỐI *Coptotermes gestroi* Ở VIỆT NAM".

2. Mục tiêu

2.1. Mục tiêu chung

Nghiên cứu được sự đa dạng hệ vi khuẩn và enzyme thủy phân lignocellulose trong ruột mối *C. gestroi* từ dữ liệu giải trình tự DNA đa hệ gen. Đồng thời xây dựng được một phương pháp hiệu quả để tìm kiếm nhanh được gen đích mã hóa đúng lignocellulase từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* và phải có hoạt tính đúng của enzyme mục tiêu sau khi biểu hiện.

2.2. Mục tiêu cụ thể

- Khai thác được dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* đã có để nghiên cứu đa dạng hệ vi khuẩn và enzyme thủy phân lignocellulose;

- Xây dựng được phương pháp mới và hiệu quả để tìm kiếm gen mục tiêu từ dữ liệu giải trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi*;

- Nghiên cứu được biểu hiện gen *Xbx14* và tính chất của enzyme tham gia thủy phân lignocellulose.

3. Nội dung nghiên cứu

- Nghiên cứu đa dạng hệ vi khuẩn và lignocellulase theo hệ thống phân loại của CAZY từ dữ liệu DNA đa hệ gen của hệ vi khuẩn trong ruột mối *C. gestroi* đã được thiết lập;

- Tìm kiếm các trình tự axit amin của β-xylosidase đã được nghiên cứu chi tiết về tính chất và lựa chọn công cụ tin sinh học đáng tin cậy để xây dựng mẫu dò. Sử dụng mẫu dò để tìm kiếm nhanh gen mã hóa β-xylosidase từ dữ liệu giải trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi*;

- Biểu hiện và nghiên cứu tính chất của β-xylosidase.

4. Đối tượng

Dữ liệu giải trình tự DNA đa hệ gen của vi sinh vật sống trong ruột mối *C. gestroi* của phòng Kỹ thuật Di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam; một số nguồn dữ liệu khai thác trên các CSDL và công cụ tin sinh học.

5. Phạm vi nghiên cứu

Nghiên cứu đặc điểm hệ vi khuẩn trong ruột mối *C. gestroi* và phương pháp xây dựng mẫu dò phục vụ cho việc khai thác và lựa chọn gen mong muốn từ dữ liệu đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* đã có. Thực nghiệm biểu hiện gen trong tế bào *E. coli Rosseta 1* và xác định các đặc điểm của β–xylosidase.

6. Ý nghĩa khoa học và thực tiễn của đề tài

6.1. Ý nghĩa khoa học

Đã phân tích được hệ vi khuẩn sống trong ruột mối *C. gestroi* ở Việt Nam và sự đa dạng các lignocellulase theo phân loại của CAZY;

Cung cấp một phương pháp mới và hiệu quả cho việc tìm kiếm gen mục tiêu từ dữ liệu giải trình tự DNA đa hệ gen bằng mẫu dò được xây dựng dựa trên các trình tự axit amin của β-xylosidase và công cụ tin sinh học;

Đã biểu hiện và xác định hoạt tính của β-xylosidase từ vi sinh vật tự do trong ruột mối *C. gestroi*.

6.2. Ý nghĩa thực tiễn

β-xylosidase là một enzyme mới, có hoạt tính tốt, hoạt động tối ưu trong môi trường kiềm và nhiệt độ cao. Đây là enzyme hứa hẹn mang lại hiệu quả cao khi ứng dụng trong thực tế thủy phân lignocellulose, đặc biệt kết hợp với phương pháp tiền xử lý sinh khối thực vật bằng kiềm và nhiệt.

7. Đóng góp mới của luận án

Đây là nghiên cứu đầu tiên về hệ vi khuẩn sống tự do và sự đa dạng enzyme thủy phân lignocellulose theo phân loại CAZY từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* ở Việt Nam;

Đã xây dựng và ứng dụng thành công mẫu dò dựa trên sự phân tích các trình tự axit amin thuộc nhóm β-xylosidase được nghiên cứu thực nghiệm. Sử dụng mẫu dò đã tìm kiếm được gen mã hóa cho β-xylosidase từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi*;

β–xylosidase là enzyme mới có nguồn gốc từ vi sinh vật trong ruột mối *C. gestroi* ở Việt Nam, đã được biểu hiện và tinh chế thành công, với hoạt tính tốt, hoạt động tối ưu trong môi trường kiềm và nhiệt độ cao.

8. Nơi thực hiện đề tài luận án

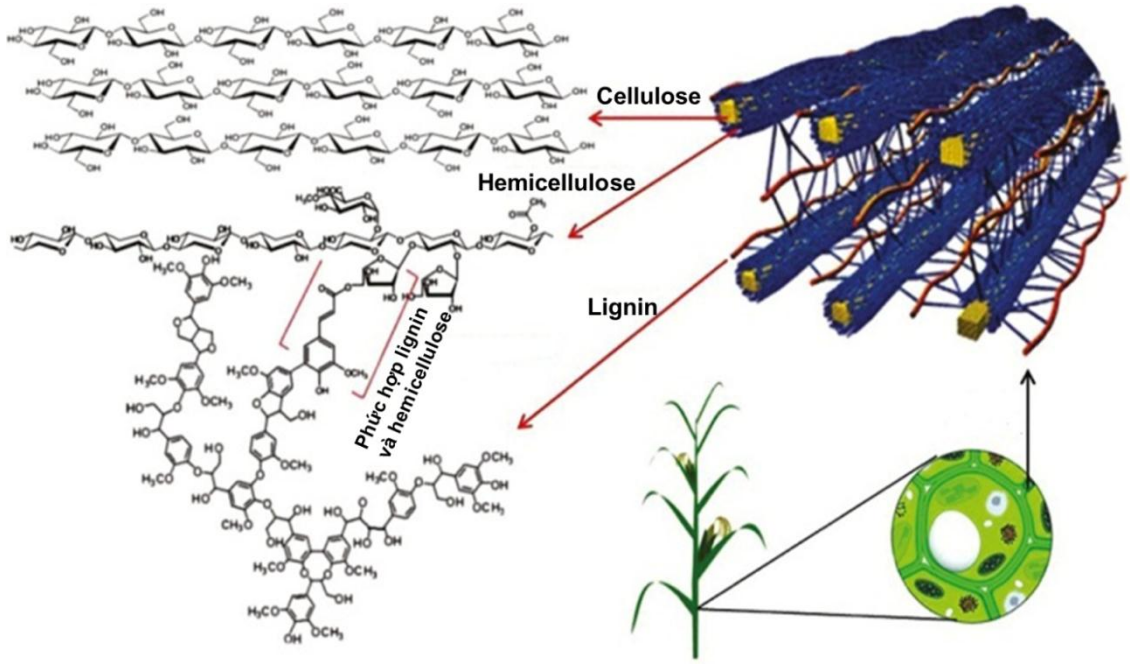
Luận án được nghiên cứu và thực nghiệm tại phòng Kỹ thuật di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam và Tổ Di truyền học, Khoa Sinh học, Trường Đại học Sư phạm Hà Nội từ tháng 12 năm 2013 đến tháng 6 năm 2017.

Chương 1. TỔNG QUAN NGHIÊN CỨU

1.1. LIGNOCELLULOSE VÀ QUÁ TRÌNH CHUYỂN HÓA

1.1.1. Lignocellulose

Lignocellulose là thành phần chính của sinh khối thực vật với cấu trúc bền vững. Trên thế giới ước tính thực vật tổng hợp lượng lignocellulose khoảng 2.10^{11} tấn/năm [142]. Đây là nguồn chất thải gây ra ô nhiễm môi trường nếu không được xử lý triệt để. Tuy nhiên, nguồn sinh khối được coi là rác thải này chính là nguyên liệu cho sản xuất nhiều sản phẩm có giá trị khác nhau như nhiên liệu sinh học, hóa chất, hay bổ sung dinh dưỡng cho thức ăn động vật ,…Ở Việt Nam nguồn sinh khối này đang được nông dân xử lý bằng cách đốt, gây lãng phí lớn và ảnh hưởng tiêu cực đến môi trường sống [170].



Hình 1.1. Cấu trúc của lignocellulose **[178]**

Lignocellulose được tạo thành từ ba polyme chính là cellulose (38-50%), hemicellulose (23-32%) và lignin (15-25%) [17] (Hình 1.1). Tỷ lệ các thành phần trong lignocellulose thay đổi tùy theo nguồn sinh khối như cỏ, ngũ cốc, gỗ mềm, gỗ cứng, phế phẩm nông nghiệp, rác thải nhà máy đường, nhà máy gỗ....

Lignin liên kết chặt chẽ với cellulose, hemicellulose và các thành phần khác, tạo nên độ bền cơ học ở thành tế bào thực vật, làm cho thành tế bào thực vật cứng, chắc và giòn. Đồng thời kết nối các tế bào với nhau tạo thành cấu trúc định hình ở thực vật. Thực vật càng già, lượng lignin tích tụ càng lớn. Lignin được cấu tạo từ 3 thành phần chính là p-coumaryl alcohol (H), coniferyl alcohol (G) và sinapyl alcohol (S). Tỷ lệ các thành phần không giống nhau tùy thuộc vào nguồn gốc sinh khối thực vật. Sự phân hủy lignin trong mô của thực vật thân gỗ lâu năm có thời gian dài nhất và đóng góp phần lớn chất mùn trong tự nhiên [63].

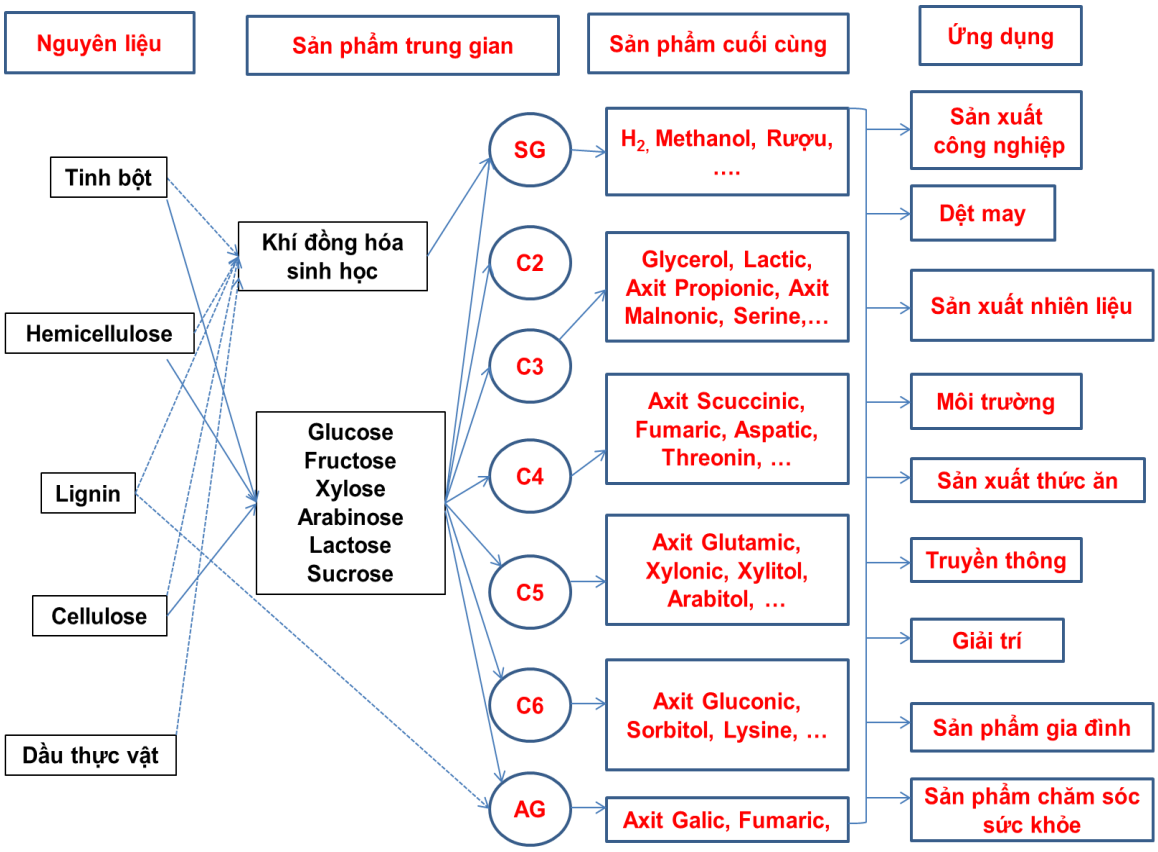
Thành phần chiếm tỷ lệ lớn nhất trong sinh khối thực vật chính là cellulose. Cellulose có cấu trúc cơ bản là các vi sợi, không phân nhánh xếp song song và liên kết với nhau bằng liên kết hydro trong thành tế bào [155]. Cellulose được hình thành từ các đơn vị glucose kết hợp với nhau qua liên kết β-1,4-glycoside, tạo thành cấu trúc tinh thể và cấu trúc vô định hình. Cấu trúc tinh thể rất vững chắc, bền với nhiệt và ổn định hơn so với cấu trúc vô định hình do tồn tại các liên kết hydro, lực van der Waal giữa các chuỗi polysaccharide và các vi sợi ở cạnh nhau [137]. Do đó trong quá trình tiền xử lý sinh khối phải phá vỡ được cấu trúc tinh thể của cellulose tạo điều kiện cho các enzyme có thể tiếp cận và xúc tác chuyển hóa hiệu quả. Cellulose có cấu trúc rất bền và khó bị thủy phân. Người và động vật không có enzyme phân giải cellulose (cellulase) nên không tiêu hóa được cellulose, vì vậy cellulose không có giá trị dinh dưỡng, nhưng có vai trò điều hòa hoạt động của hệ thống tiêu hóa.

Hemicellulose là một trong ba thành phần tạo nên sinh khối chính trong tự nhiên, liên kết chặt với cellulose và lignin. Đây là polysaccharide dị hợp có độ phân nhánh cao, thành phần chủ yếu là các phân tử đường 5 carbon (xylose, arabinose), đường 6 carbon (mannose, galactose, glucose) và một số axít. Hemicellulose được tạo ra từ nhiều chuỗi carbohydrate như xylan (đơn phân là xylose), xyloglucan (đơn phân là D-xylose và D-glucose), galactoglucomannan (đơn phân là D-glactose, D-glucose và D-mannose), glucomannan (đơn phân là D-glucose và D-mannose) và arabinogalactan (đơn phân là D-galactose và arabinose). Trong đó xylan là thành

phần nhiều nhất trong vật liệu gỗ cứng và rơm rạ; glucomannan lại chiếm chủ đạo trong nguyên vật liệu gỗ mềm [108]. Hemicellulose thường được phân loại theo thành phần cấu tạo khung cấu trúc chính như xylan, galactose, mananose,….

1.1.2. Sự chuyển hóa lignocellulose

Quá trình chuyển hóa lignocellulose tạo ra hàng loạt các sản phẩm trung gian, được ứng dụng trong nhiều lĩnh vực như công nghiệp, dệt may, môi trường, sản phẩm chăm sóc sức khỏe,.... đặc biệt trong sản xuất nhiên liệu sinh học (Hình 1.2).



***Hình 1.2. Sơ đồ mô hình chuyển hóa sinh khối thành các sản phẩm* [62]**

Ghi chú: C2, C3, C4, C5, C6, SG và AG là các sản phẩm chuyển hóa trung gian

Sự vững chắc trong cấu trúc của lignocellulose làm cho việc xử lý chúng một cách triệt để rất khó khăn và thường trải qua hai giai đoạn là: tiền xử lý để nới lỏng cấu trúc thành tế bào, phá vỡ liên kết của lignin và cấu trúc tinh thể của cellulose, sau đó sử dụng các enzyme để thủy phân các thành phần còn lại thành monome.

Trong thực tế sinh khối thực vật được tiền xử lý bằng phương pháp vật lý (như nghiền và xay sinh khối, sử dụng bức xạ electron và chiếu xạ vi sóng), phương pháp hóa học (như sử dụng kiềm, amoniac, axit, tác nhân oxi hóa và các dung môi hữu cơ) và phương pháp sinh học (sử dụng vi sinh vật hoặc enzyme để phân hủy trực tiếp). Trong các phương pháp trên phương pháp xử lý bằng vật lý và hóa học đều có những nhược điểm rất lớn như chưa phân hủy được hết lignin, hiệu suất tạo đường đơn thấp, cần có các thiết bị chuyên ngành phù hợp với giá thành quá cao hoặc sự tồn dư dung môi trong chế phẩm gây ức chế quá trình đường hóa và lên men tiếp theo, hoặc gây ra các vấn đề liên quan đến môi trường,... Phương pháp xử lý sinh học khắc phục được các nhược điểm trên, đặc biệt phương pháp sử dụng enzyme có nhiều ưu điểm như an toàn với môi trường, dễ dàng điều khiển và thu hồi sản phẩm theo từng giai đoạn xử lý,…. Do đó nhiều nghiên cứu đã tập trung tìm kiếm hệ thống enzyme giúp xử lý triệt để nguồn sinh khối thực vật khổng lồ được thải ra hàng năm. Hệ thống enzyme chuyển hóa lignocellulose gồm ba nhóm chính là ligninase, cellulase và hemicellulase.

Enzyme chuyển hóa lignin: Để phân hủy lignin cần nhiều enzyme khác nhau, trong đó nhóm peroxidase và oxidase đóng vai trò phá vỡ cấu trúc của polyme. Trong các peroxidase thì lignin peroxidase phân cắt sự trùng hợp của chuỗi polyme và mangan peroxidase giúp oxi hóa cả hợp chất phenolic và không phải là phenolic. Các thí nghiệm đã chứng minh mangan peroxidase đã phá vỡ cấu trúc của dimer, trimer, tetramer và phenolic của các oligo lignin. Bên cạnh các enzyme này thì laccase cũng là một enzyme oxi hóa một điện tử của phenol, vòng thơm amin và một số chất giàu điện tử khác. Ngoài ra laccase có thể oxi hóa các hợp chất không phải là phenolic với sự có mặt của một số chất phụ trợ khác nên cũng tham gia vào xử lý lignin [37].

Enzyme thủy phân cellulose: Cellulase là một phức hệ enzyme có tác dụng cắt liên kết β -1,4-glucoside trong cellulose để tạo thành đường đơn**.** Lớp enzyme này được tổng hợp chủ yếu từ vi sinh vật sinh trưởng trên môi trường sống có cellulose như nấm, vi khuẩn, động vật nguyên sinh và thực vật [7]. Để phân hủy hoàn toàn cellulose thành sản phẩm cuối cùng là glucose cần có sự kết hợp của ít nhất 3 loại

enzyme là: exoglucanase (EC 3.2.1.91), endoglucanase hay (EC 3.2.1.4) và β-glucosidase (EC 3.2.1.21) [178].

Enzyme thủy phân hemicellulose: Với cấu trúc không đồng nhất bao gồm rất nhiều thành phần khác nhau, do đó để phân giải hemicellulose cần nhiều nhóm enzyme như: endo-β-1,4-xylanase (E.C. 3.2.1.8), β-xylosidase (E.C.3.2.1.37), α-L-arabinofuranosidase (E.C.3.2.1.55), endo-α-1,5-arabinanase (E.C.2.1.99), α-glucuronidase (E.C.3.2.1.139), endo-β-1,4-mannanase (E.C.3.2.1.78), exo-β- 1,4-mannosidase (E.C.3.2.1.25), α-galactosidase (E.C.3.2.1.22), β-glucosidase (E.C.3.2.1.21), endo-galactanase (E.C.3.2.1.89), acetyl xylan esterase (E.C.3.1.1.72), acetyl mannan esterase (E.C.3.1.1.6), ferulic và *ρ*-cumaric axit esterases (E.C. 3.1.1.73),…[178].

1.2. METAGENOMICS VÀ CÔNG CỤ TIN SINH HỌC KHAI THÁC DỮ LIỆU DNA ĐA HỆ GEN

1.2.1. Metagenomics

1.2.1.1. Khái niệm

Thuật ngữ "Metagenomics" được sử dụng bởi Handelsman, Clardy, Goodman và một số tác giả khác, xuất bản lần đầu tiên vào năm 1998 [68]. Metagenomics là kỹ thuật phân tích đa hệ gen (DNA metagenome) của tất cả vi sinh vật thu nhận trực tiếp từ bất kỳ mẫu môi trường tự nhiên nào không thông qua nuôi cấy [69].

Theo ước tính Trái Đất có trên 10^{30} loài prokaryote đang sống [21], trong đó khoảng 95% - 99,9% vi sinh vật chưa thể nuôi cấy được trong phòng thí nghiệm [102]. Trong những năm gần đây, cùng với sự phát triển nhanh của công nghệ giải trình tự gen thế hệ mới, kỹ thuật Metagenomics cho phép giải mã khối lượng gen đồ sộ từ vi sinh vật thu được trong một mẫu môi trường, không cần thông qua nuôi cấy. Đây là nguồn dữ liệu di truyền vô cùng phong phú, cho phép khai thác tối đa nguồn gen quý.

1.2.1.2. Cách tiếp cận trong nghiên cứu Metagenomics

Metagenomics nghiên cứu DNA đa hệ gen của quần xã sinh vật thông qua ba bước gồm: 1) Tách chiết axit nucleic trong mẫu thu thập; 2) Thiết lập thư viện đa hệ

gen hoặc giải trình tự toàn bộ DNA đa hệ gen bằng máy giải trình tự thế hệ mới; 3) Sàng lọc gen dựa vào ngân hàng gen hoặc phân lập gen dựa vào số liệu giải trình tự gen.

Kỹ thuật Metagenomics có thể tiếp cận và nghiên cứu DNA đa hệ gen theo hai hướng: (1) Phân lập gen dựa trên việc thiết lập thư viện DNA đa hệ gen và (2) khai thác trình tự DNA và phân lập gen dựa trên dữ liệu giải trình tự trực tiếp DNA đa hệ gen. Trong đó, cách tiếp cận thứ nhất thường tốn rất nhiều thời gian và công sức, do phải sàng lọc một khối lượng lớn các dòng trong thư viện, yêu cầu số lượng dòng thư viện phải rất lớn và chất lượng thư viện phải cao [176]. Cách tiếp cận thứ hai đang ngày càng phổ biến và ứng dụng rộng rãi hơn nhờ kỹ thuật giải trình tự thông lượng cao (HTS) mẫu DNA đa hệ gen của quần xã vi sinh vật sống trong môi trường sinh thái nhất định. Với ưu điểm giải trình tự nhanh và chính xác, các hệ thống máy giải trình tự thế hệ mới (NGS) được sử dụng phổ biến như máy giải trình tự Roche 454 của 454 Life Sciences, HiSeq 2000 của Illumina và AB SoLid của Life Technologies [106]. Điểm khác biệt quan trọng giữa HTS và kỹ thuật giải trình tự Sanger truyền thống là thông lượng. Trong khi một Sanger điển hình tạo ra được 102 trình tự (với độ dài 600 đến 900 bp) thì HTS có thể sinh ra 10^6 đến 10^9 trình tự (với độ dài 100 đến 700 bp) cho một lần chạy [64]. Thiết bị giải trình tự gen hiện đại như hệ thống giải trình tự của Ilumina, một lần chạy trên bản 96 giếng, có thể cho ra khoảng 12-15 triệu trình tự và mỗi trình tự có độ dài trên dưới 100 bp [106].

1.2.1.3. Một số thành tựu ứng dụng Metagenomics

Sau gần 20 năm phát triển, kỹ thuật Metagenomics đã trở thành công cụ mạnh mẽ để phân tích đa dạng di truyền và loài, chức năng gen và protein, sự tương tác và tiến hóa của vi sinh vật trong nhiều loại môi trường khác nhau như: đất, nước, hệ tiêu hóa của người, dạ cỏ của động vật nhai lại,…. Trong một số năm trở lại đây khả năng giải trình tự nhanh chóng với chi phí thấp của hàng loạt máy thế hệ mới, đã làm bùng nổ cuộc cách mạng về số liệu di truyền [46]. Dữ liệu và siêu dữ liệu DNA đa hệ gen không chỉ dừng lại ở việc mô tả sự phát sinh loài hay một số đặc điểm của gen thông qua hệ thống di truyền 16S RNA, mà toàn bộ chức năng của gen, mối quan hệ giữa các gen trong một nhóm hay giữa các nhóm sinh vật đều được làm

sáng tỏ [127]. Các nghiên cứu đã sử dụng dữ liệu DNA đa hệ gen để thiết kế thí nghiệm thông lượng cao và thông lượng thấp, tập trung vào việc xác định vai trò của gen và vi sinh vật trong việc thành lập cộng đồng vi sinh vật động [192].

Metagenomics cũng được ứng dụng trong nhiều lĩnh vực của khoa học Trái Đất, khoa học sự sống, khoa học y sinh [180], sản xuất năng lượng, xử lý môi trường, công nghệ sinh học, nông nghiệp [69],…. Trong đó phải kể các ứng dụng vượt trội trong y học [38] như dự án của "Human Microbiom" bước đầu đã sử dụng trình tự DNA đa hệ gen của các cộng đồng vi khuẩn ở 15-18 vị trí khác nhau trên cơ thể 250 người để đánh giá sự thay đổi và mối quan hệ của chúng với sức khỏe của con người. Một nghiên cứu y tế khác của dự án MetaHit (Metagenomics of the Human Intestinal Tract) tiến hành ở 124 người từ Đan Mạch và Tây Ban Nha mắc bệnh đường ruột, thừa cân và cáu kỉnh đã công bố thông tin về sự đa dạng phát sinh loài vi khuẩn tiêu hóa. Nghiên cứu đã chứng minh rằng hai ngành vi khuẩn Bacteroidetes và Firmicutes chiếm hơn 90% của các loài được biết đến đang thống trị vi khuẩn đường ruột. Sử dụng tần số gen liên quan được tìm thấy trong ruột đã xác định 1.244 cụm gen của DNA đa hệ gen giữ vai trò vô cùng quan trọng cho sức khỏe của đường ruột [30]. Trong sản xuất năng lượng rất nhiều kết quả đã công bố về phân tích và so sánh DNA đa hệ gen giữa các vi sinh vật trong hệ thống lên men khí sinh học [167], trong đường tiêu hóa của động vật ăn cỏ như côn trùng, thú ăn cỏ [72]. Thế giới đã công bố khoảng 75 hệ gen của các loài vi sinh vật giữ vai trò nhất định trong quá trình sản xuất năng lượng sinh học [184] . Ở Việt Nam sử dụng kỹ thuật Metagenomics để phân lập gen đã được thực hiện trên một số đối tượng như gen mã hóa laccase của nấm Basidiomycetes trong mẫu đất rừng Nam Cát Tiên [89], gen mã hóa protease và amylase bền nhiệt từ thư viện DNA đa hệ gen của vi sinh vật suối nước nóng [194], gen mã hóa pectinase từ cộng đồng vi sinh vật đất [195], ….

Trong các dự án ứng dụng kỹ thuật Metagenomics để đánh giá đa dạng vi sinh vật và tìm kiếm gen mã hóa lignocellulose phải kể đến nhiều dự án giải trình tự DNA đa hệ gen của vi sinh vật sống trong ruột mối. Hàng loạt nghiên cứu đã đánh giá đa dạng vi sinh vật trong ruột của mối *R. speratus* [77], *N. ephratae* [182], *Microcerotermes* [76], *N. corniger* và *A. wheeleri* [70], *R. flavipes* [60], *C.*

formosanus [75] và khai thác gen mã hóa cho enzyme thủy phân lignocellulose từ hệ vi sinh vật ruột mối *N. ephratae* [182], *R. speratus* [171], *N. corniger* và *A. wheeleri* [70]. Kỹ thuật này cũng đã được sử dụng để phân tích DNA đa hệ gen của vi khuẩn sống tự do trong ruột mối *C. gestroi* ở Việt Nam [166].

1.2.2. Một số công cụ tin sinh sử dụng để phân tích số liệu

Việc phân tích và khai thác số liệu giải trình tự đa hệ gen có thể sử dụng rất nhiều phần mềm tin sinh học trực tuyến và khai thác miễn phí. Nghiên cứu này sử dụng một số công cụ tin sinh học để phân tích dữ liệu giải trình tự đa hệ gen của vi sinh vật ruột mối *C. gestroi* và tìm kiếm được gen mã hóa lignocellulase như sau:

1.2.2.1. BLAST (Basic Local Alignment Search Tool)

BLAST là một trong các công cụ nổi tiếng dùng trong sinh học tính toán, dựa trên giải thuật tìm kiếm những trình tự axit nucleic/protein tương đồng lưu trữ trên nhiều CSDL. Ba nguồn CSDL công cộng lớn nhất về trình tự DNA và protein được BLAST tham chiếu đến là NCBI, EMBL và DDBJ [205].

BLAST sử dụng giải thuật so sánh các chuỗi sinh học như chuỗi amino axit của protein hay chuỗi nucleotide của DNA. Khi được cung cấp một CSDL phù hợp, BLAST sẽ cho phép tìm kiếm chuỗi con giống với chuỗi có sẵn mà người dùng quan tâm. Để chạy BLAST cần đầu vào là chuỗi truy vấn (hay còn gọi là chuỗi đích) và một CSDL. BLAST sẽ tìm kiếm chuỗi con trong câu truy vấn giống với các chuỗi con trong CSDL. Thông thường, khi sử dụng, chuỗi truy vấn nhỏ hơn rất nhiều so với CSDL. Ý tưởng của BLAST dựa trên cơ sở xác suất những chuỗi bắt cặp trình tự thường sở hữu nhiều đoạn chuỗi con có tính tương đồng cao. Những chuỗi con này được mở rộng để tăng tính tương đồng trong quá trình tìm kiếm. Thuật toán của BLAST có 2 phần, một phần tìm kiếm và một phần đánh giá thống kê dựa trên kết quả tìm được.

BLASTall là tập hợp chương trình tìm kiếm được thiết kế để khám phá tất cả CSDL trình tự có sẵn, bất kể đó là protein hay DNA bao gồm: 1) BLASTp tìm kiếm tất cả trình tự protein tương đồng với trình tự protein cần phân tích trong CSDL protein; 2) BLASTn tìm kiếm tất cả trình tự nucleotide tương đồng với trình tự

DNA cần phân tích trong CSDL DNA; 3) TBLASTn tìm trình tự protein tương đồng trong CSDL DNA bằng cách dịch mỗi trình tự DNA ra tất cả 6 khung đọc mở; 4) BLASTx tìm trình tự nucleotide tương đồng trong CSDL protein bằng cách dịch trình tự nucleotide cần phân tích sang tất cả 6 khung đọc mở. Trong nghiên cứu DNA đa hệ gen, sau khi có được dự đoán về khung đọc mở cần quan tâm, sử dụng công cụ tìm kiếm trình tự axit amin tương đồng trong BLASTp có thể điều tra vùng bảo tồn của mỗi enzyme do ORF mã hóa [54].

1.2.2.2. EXPASY (Expert Protein Analysis System)

EXPASY là một cổng thông tin tài nguyên sinh học do Viện tin sinh học Thụy Sĩ thiết kế (SIB), cho phép mở rộng và tích hợp truy cập của công cụ phần mềm vào tài nguyên khoa học, CSDL và trong nhiều lĩnh vực của đời sống [196]. Các nhà khoa học có thể truy cập vào hàng loạt nguồn dữ liệu như protein, gen, phát sinh loài/ tiến hóa, hệ thống sinh học, di truyền học quần thể, transcriptomics,…. Ngoài trang web chính ở Thụy Sĩ, thì bảy chi nhánh khác được đặt ở hầu hết châu lục để phục vụ cho cộng đồng người dùng.

EXPASY bắt đầu hoạt động từ năm 1993 và hoạt động như một máy chủ proteomics để phân tích chuỗi protein. EXPASY cung cấp công cụ phân tích chuyên dụng giữa protein với protein và truy cập đến một loạt CSDL như SWISS-PROT, SWISS-2DPAGE,... Tất cả CSDL có sẵn trên EXPASY đều tham chiếu đến hầu hết CSDL trên toàn thế giới. Ví dụ SWISS-PROT tham chiếu chéo đến khoảng 50 CSDL khác nhau chuyên về protein và trình tự axit nucleic, mô hình cấu trúc 3D, phân loại sinh vật, thông tin di truyền, các vùng và dấu hiệu của họ protein/enzyme, sửa đổi sau dịch mã,... Công cụ phân tích có sẵn trong EXPASY đều thực hiện các nhiệm vụ cụ thể liên quan đến protein gồm: tìm kiếm tương tự, dự đoán thay đổi, dự đoán cấu trúc liên kết, phân tích cấu trúc bậc 1, 2, 3, 4,…. Dữ liệu của proteomics dựa trên nguồn thông tin về điện di và thí nghiệm quang phổ, do đó độ chính xác khá cao.

Trong những năm qua bộ sưu tập công cụ của EXPASY ra đời, hướng tới việc truy cập và hiển thị CSDL để phân tích chuỗi protein. EXPASY thiết lập hệ thống công cụ phân tích liên quan đến genome, protein, cấu trúc không gian, hệ thống sinh

học, tiến hóa, di truyền học quần thể, hóa dược, dịch mã,…. Để chạy các công cụ trong danh mục thiếp lập sẵn, người dùng chỉ cần nhập trình tự DNA hoặc axit amin vào, chọn chế độ phân tích và chờ kết quả sau từ vài phút đến vài giờ đồng hồ tùy vào từng loại công cụ. Trong nghiên cứu này một số công cụ của EXPASY được chúng tôi sử dụng để phục vụ cho mục đích dịch mã, dự đoán signal peptide, trọng lượng protein, điểm đẳng điện,… của protein.

1.2.2.3. Công cụ dự đoán cấu trúc không gian

Hai công cụ sử dụng để dự đoán cấu trúc không gian của protein trong nghiên cứu này là Phyre 2 và Swiss model.

Phyre 2 là công cụ dự đoán cấu trúc không gian dựa trên nguyên tắc và kỹ thuật mô hình tương đồng của chuỗi protein, dựa trên kết quả so sánh các vị trí axit amin bảo tồn. Công cụ này kế thừa từ mô hình 3D-PSSM của Lawrence Kelley và Bob MacCallum thuộc phòng thí nghiệm mô hình hóa sinh học phân tử tại Vương quốc Anh, máy chủ được giám sát bởi Michael Sternberg [197]. Cấu trúc bậc cao của protein được bảo tồn hơn trong quá trình tiến hóa so với trình tự chuỗi axit amin. Từ chuỗi protein quan tâm (mục tiêu) có thể được mô hình hóa với độ chính xác hợp lý dựa trên một chuỗi rất xa liên quan đến cấu trúc đã được biết đến về vị trí axit amin bảo tồn. Thông thường, trình tự axit amin của một chuỗi cấu trúc protein sẽ được xử lý bằng cách quét với một CSDL trình tự protein lớn để xuất ra kết quả. Nguồn dữ liệu sử dụng chính để so sánh khi tìm kiếm cấu trúc tương đồng của Phyre 2 chủ yếu là NCBI. Để dự đoán cấu trúc bậc cao của protein, người dùng sẽ nhập trình tự chuỗi axit amin và chờ từ 30 phút đến vài giờ (tùy thuộc vào chiều dài chuỗi, số lượng trình tự tương đồng, tần số lặp lại,…) công cụ sẽ đưa ra một dự báo cuối cùng. Thông tin tóm tắt của kết quả dự đoán sẽ chuyển đến email đăng ký. Bảng kết quả chính trong Phyre 2 cung cấp mức độ tin cậy của ước đoán, hình ảnh mô hình không gian, nguồn gốc hình ảnh, trung tâm hoạt động của enzyme,…. Đối với mỗi kết quả sẽ có một liên kết đưa người dùng đến nguồn gốc trình tự axit amin và cấu trúc không gian của CSDL đã sử dụng. Các tính năng xử lý hàng loạt cho phép người dùng gửi nhiều hơn một trình tự cần ước đoán. Theo mặc định, người sử dụng có giới hạn là 100 trình tự trong một lần nhập dữ liệu. Giới hạn này có thể

được nâng lên bằng cách liên lạc với quản trị, nhưng mất nhiều thời gian hơn. Trên cơ sở dự đoán cấu trúc không gian của phần mềm Phyre 2, nghiên cứu có thể sử dụng để xác định mối liên hệ giữa cấu trúc với chức năng của protein [8]. Kết quả này sẽ cung cấp thông tin cho việc lựa chọn gen biểu hiện trong vật chủ.

Swiss model được tạo ra vào năm 1986, tại Khoa Hóa sinh Y học của Đại học Geneva, Viện thông tin sinh học SIB Thụy Sĩ, sau đó kết hợp với thư viện dữ liệu EMBL [198]. Khoảng 85% trình tự protein CSDL này bắt nguồn từ việc dịch các trình tự mã hóa từ CSDL axit nucleic như EMBL-Bank/GenBank/DDBJ. Trong Swiss model có thể phân biệt hai lớp dữ liệu là dữ liệu cốt lõi và chú thích. Đối với mỗi lần nhập chuỗi, dữ liệu cốt lõi bao gồm thông tin về trình tự, nguồn trích dẫn của thư mục và phân loại (mô tả nguồn sinh học của protein). Dữ liệu chú thích bao gồm mô tả các mục như chức năng của protein, sửa đổi sau dịch mã (carbohydrate, phosphoryl hóa, acetyl hóa, GPI-anchor,…), vị trí và vùng bảo tồn (vùng liên kết canxi, kẽm, vị trí liên kết ATP,...), cấu trúc bậc hai (chuỗi alpha, beta,...), bậc cao (homodimer, heterotrimer,…), sự tương đồng với protein khác, bệnh liên quan đến sự thiếu hụt trong protein,… [10]. Để có được thông tin chú thích, Swiss model sử dụng nhận xét của chuyên gia bên ngoài và ấn phẩm báo cáo dữ liệu trình tự mới, so sánh với số liệu cũ để cập nhật định kỳ chú thích protein. Các dữ liệu được hợp nhất với nhau một cách khoa học, giảm thiểu sự dư thừa, đồng thời có thể tham chiếu đến hơn 100 CSDL khác nhau [10]. Swiss Protein ước đoán chi tiết được cấu trúc không gian của hầu hết trình tự protein dựa trên cả thông tin vùng bảo tồn và không bảo tồn, do đó cho phép dự đoán được mô hình cấu trúc protein đích. So sánh với kết quả dự đoán cấu trúc không gian protein của Phyre 2, thì kết quả dự đoán của Swiss Model sẽ mở rộng hơn. Để dự đoán kết quả, người dùng chỉ cần nhập một hoặc nhiều trình tự axit amin đích, công cụ sẽ tính toán để đưa ra nhiều thông tin liên quan đến cả hai lớp dữ liệu cốt lõi và chú thích. Đặc biệt cấu trúc không gian sẽ được mô tả một cách chi tiết, với các vùng chức năng đặc trưng của từng loại protein và cho phép người dùng thay đổi tham số đã mặc định.

1.2.2.4. Dự đoán khả năng chịu kiềm/axit

Tìm kiếm nguồn lignocellulase có hoạt tính tốt trong môi trường kiềm hoặc axit là mục tiêu của nhiều nghiên cứu ứng dụng. Vì vậy trong nghiên cứu tìm kiếm lignocellulase, chúng tôi cũng rất quan tâm đến mục tiêu này. Để dự đoán trước khả năng hoạt động tối ưu của enzyme được mã hóa bởi các ORF đã chọn, chúng tôi tiến hành tìm kiếm công cụ tin sinh học phù hợp. AcalPred là công cụ được cung cấp miễn phí bởi Trung tâm Tin sinh học, Đại học Khoa học Công nghệ và Điện tử, Trung Quốc, giúp dự đoán khả năng hoạt động của enzyme trong điều kiện kiềm hoặc axit [199]. Công cụ này dựa vào sự tổng hợp nhiều chỉ số khác nhau của protein đã nghiên cứu thực nghiệm bao gồm: thành phần axit amin, GO, nhóm axit amin được bảo tồn, giá trị điện tích,… Các chỉ số này sẽ là cơ sở để phần mềm thiết kế vector hỗ trợ SVM (Support Vector Machine) làm vector tham chiếu chung. Khi ta có một trình tự axit amin nhập vào, phần mềm sẽ tự động tính toán và cho ra kết quả cuối cùng về khả năng chịu kiềm hay axit của protein [104]. Kết quả cuối cùng phân loại thành hai nhóm dựa trên chỉ số phân tích. Nếu chỉ số phân tích càng gần với giá trị bằng 1, thì pH hoạt động tối ưu trong môi trường kiềm sẽ càng cao và ngược lại nếu chỉ số gần với 0 thì pH hoạt động tối ưu sẽ trong môi trường axit. Phần mềm này cho phép người dùng ước đoán cùng lúc hàng trăm trình tự axit amin khác nhau.

1.2.2.5. Phần mềm dự đoán khả năng chịu nhiệt

Việc ước đoán được khả năng chịu nhiệt của protein/enzyme cho phép tối ưu hóa hàng loạt quy trình công nghệ sinh học trong nhiều lĩnh vực như nông nghiệp, dược phẩm và môi trường [136]. 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, từ đó xây dựng vector hỗ trợ cho việc dự đoán đặc tính này [200]. Vector sử dụng chỉ số chịu nhiệt (Tm) đã thực nghiệm kiểm tra trên 75 bộ gen vi khuẩn hoàn chỉnh. Từ khả năng chịu nhiệt của gen được phân tích, sẽ tìm ra mối liên hệ với khả năng chịu nhiệt của chủng chủ sống trong môi trường có nhiệt độ khác nhau [94]. Để chạy chương trình, người dùng chỉ cần nhập trình tự axit amin cần xác định và chờ từ một đến hai phút, phần mềm sẽ tự động tính toán cho ra kết

quả. Kết quả dự đoán nhiệt độ hoạt động (Tm) ở ba mức độ: Nếu kết quả Tm lớn hơn 1 thì khả năng chịu nhiệt của protein là trên 65°C; nếu Tm nhỏ hơn 0 thì khả năng chịu nhiệt của protein là dưới 55°C và Tm trong giới hạn từ 0 đến 1 thì khả năng chịu nhiệt của protein là từ 55°C đến 65°C. Với phần mềm này người dùng có thể ước đoán cùng lúc khả năng chịu nhiệt của nhiều trình tự axit amin khác nhau.

1.2.2.6. Phần mềm dự đoán nguồn gốc của gen

Việc xây dựng cây phát sinh loài từ trình tự phân tử bằng công cụ tin sinh học giúp cho người sử dụng có thể xử lý số liệu nhanh và chính xác. Www.Phylogeny.fr là máy chủ phân tích sự phát sinh loài hoàn toàn tự động từ trình tự nucleotide [201]. Phần mềm này sử dụng dữ liệu của hàng loạt ngân hàng như NCBI, Swiss Protein, Uniprot, ... bắt đầu từ việc xác định tương đồng, sau đó dựng cây phát sinh suy diễn. Quá trình phân tích số liệu được liên kết với hàng loạt công cụ so sánh tương đồng khác, cuối cùng công cụ xử lý nào cho kết quả có điểm cao nhất sẽ được sử dụng để xây dựng cây phân loại. “Phylogeny analysis” cung cấp ba chế độ chính bao gồm: “One Click” là chế độ xử lý với tốc độ và tham số mặc định, phù hợp với hầu hết nghiên cứu. “Advance” cho phép người dùng tùy chỉnh các thông số của chương trình. Chế độ “A la Carte” linh hoạt hơn cho phép người dùng dựa trên phần mềm để thiết lập một cây phân loại với thông số mong muốn. Người dùng chỉ phải cung cấp chuỗi dữ liệu đích để phần mềm dựng cây phân loại của chính gen đó [42].

BLAST-Explorer là một công cụ nhỏ của hệ thống phân tích Phylogeny.fr, dựa trên sự kết hợp tìm kiếm tương đồng của BLAST với hàng loạt công cụ hỗ trợ cho việc phân biệt chính xác giữa các trình tự orthologous và paralogous. Để chạy BLAST-Explorer, người dùng chỉ cần nhập trình tự đích, chọn nguồn dữ liệu cần so sánh và gửi yêu cầu phân tích cây phát sinh. Kết quả được trả lại là một cây phân loại về mối quan hệ giữa trình tự gen cần phân tích với loài gần nhất và các loài có quan hệ tiến hóa xa hơn. Trong nghiên cứu này chúng tôi sẽ dùng công cụ BLAST-Explorer để dựng cây phân loại cho mã gen đã lựa chọn, phục vụ cho mục đích xác định nguồn gốc của gen.

1.2.3. Các nguồn dữ liệu

1.2.3.1. CAZY

CAZY (Carbohydrate-Active enZYmes) là CSDL được phát triển bởi nhóm Glycogenomics tại phòng nghiên cứu kiến trúc và chức năng của các đại phân tử sinh học (Architecture and Function of Biological Macromolecules: AFMB) ở Marseille của nước Pháp vào năm 1999. Đây là một hệ thống phân loại chứa CSDL về enzyme tham gia vào quá trình tổng hợp, trao đổi và vận chuyển carbohydrate [202]. Tính đến tháng 11/2013, CAZY chứa thông tin về gần 340.000 CAZYmes [107] và được chia thành 05 nhóm lớn khác nhau bao gồm: 94 họ glycosyl transferase (GH), 104 họ glycosyl transferase (GT), 22 họ polysaccharide lysase (PL), 16 họ carbohydrate esterase (CE) và 13 họ thực hiện hoạt động phụ trợ (AA). Hàng năm số lượng họ enzyme này tăng dần theo các công bố nghiên cứu mới về nhóm enzyme chuyển hóa carbohydrate. Trong các họ trên, họ thủy phân liên kết glycosyl (Hydrolases Glycosyl: GH) đã được giới thiệu bởi Bernard Henrissat vào năm 1991 [91]. Năm 2015 dữ liệu này chứa 135 họ GH và đến tháng 6 năm 2017 là 145 họ GH, với đặc điểm nhận biết khác nhau như: cấu trúc không gian, cơ chế xúc tác phản ứng, chất cho điện tử và proton [71]. Ngoài ra mỗi họ GH của CAZY sẽ thống kê các công bố về số lượng loài sinh vật chứa enzyme (bao gồm vi khuẩn, sinh vật nhân chuẩn, sinh vật chưa phân loại), cấu trúc và đặc điểm của enzyme. Nguồn dữ liệu của CAZY được cập nhật từ số liệu của Genbank. Cụ thể các bộ gen và họ enzyme mới được bổ sung ngay sau khi chúng xuất hiện trong phiên bản hàng ngày của GenBank và dựa trên công bố chi tiết hoạt động của ít nhất một gen mã hóa enzyme; Mặt khác các họ enzyme đã có sẵn đều được cập nhật thường xuyên thông tin liên quan (cấu tạo, hoạt tính, cơ chế hoạt động,… của enzyme). Nguồn dữ liệu này cho phép nhiều công cụ tin sinh học trực tuyến sử dụng để so sánh tương đồng và phân loại CSDL đích theo hệ thống phân loại của CAZY.

Hiện nay CAZY phát triển thêm công cụ CAZYchip, dựa trên thông tin DNA của 55.220 vi khuẩn khác nhau trong CSDL, để xác định những trình tự gen thuộc về từng họ cụ thể [2]. Trên cơ sở đó xây dựng được vùng trình tự đặc trưng cho từng họ GH, giúp so sánh và tìm kiếm nhanh các vị trí tương đồng của trình tự đích

với CSDL của CAZY. Do đó người dùng có thể nhập nguồn dữ liệu đích vào trong trong hệ thống trực tuyến của CAZY, chờ phân tích và nhận kết quả. Trong nghiên cứu này các ORF mã hóa lignocellulase từ DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* được so sánh với CSDL của CAZY để phân loại theo hệ thống GH.

1.2.3.2. The National Center for Biotechnology Information (NCBI)

NCBI là trung tâm quốc gia về thông tin công nghệ sinh học, tiến bộ khoa học và y tế của Hoa Kỳ [207]. NCBI cho phép truy cập để tìm kiếm thông tin liên quan đến y sinh học, di truyền và sử dụng dữ liệu theo hệ thống ứng dụng đã được thiết kế sẵn như thử nghiệm hóa sinh học, DNA và RNA, cấu trúc và vùng bảo tồn protein/gen, gen và biểu hiện gen, gen và thuốc, genome và bản đồ di truyền, phân loại, so sánh tương đồng,…. Hệ thống công cụ phân tích dữ liệu trên cho phép người dùng dễ dàng thao tác, sắp xếp, quan sát và đánh giá dữ liệu. NCBI thu thập dữ liệu nghiên cứu khoa học và y tế từ khắp nơi trên thế giới. Dữ liệu của NCBI bao gồm dữ liệu gốc do các nhà khoa học trên thế giới gửi đến và trình tự gen/axit amin do quá trình so sánh tham chiếu đến nguồn tài nguyên sẵn có của NCBI.

Sứ mệnh của NCBI là phát triển công nghệ thông tin mới hỗ trợ trong việc tìm hiểu về di truyền cơ bản, quá trình phân tử, kiểm soát sức khỏe và bệnh tật. Để thực hiện vai trò của mình, NCBI sử dụng phương pháp toán học nghiên cứu những vấn đề cơ bản của y sinh học ở cấp độ phân tử; Duy trì sự hợp tác chặt chẽ với một số viện, học viện, ngành công nghiệp và cơ quan chính phủ; Thúc đẩy truyền thông khoa học bằng cách tài trợ các cuộc họp, hội thảo và bài giảng liên quan; Hỗ trợ đào tạo nghiên cứu cơ bản và ứng dụng trong sinh học tính toán cho nghiên cứu sinh thông qua chương trình nghiên cứu. NCBI cũng là thành viên của cộng đồng khoa học quốc tế trong nghiên cứu tin học và đào tạo thông qua nhiều chương trình khoa học; Hỗ trợ và phối hợp với cộng đồng khoa học và y tế truy cập vào hàng loạt CSDL và phần mềm; Phát triển các tiêu chuẩn về CSDL cũng như danh pháp trong sinh học.

Nhiều dự án nghiên cứu cơ bản được thực hiện để tăng cường củng cố, công bố công khai CSDL và phần mềm ứng dụng của NCBI. Những nỗ lực hợp tác nghiên cứu, điều tra của NCBI với cộng đồng nghiên cứu bên ngoài, đã dẫn đến sự

phát triển của nhiều thuật toán tiên tiến (BLAST, PSI-BLAST, COG,…), nguồn lực cơ bản và phương pháp nghiên cứu mới.

1.2.4. Mẫu dò DNA và ứng dụng

1.2.4.1. Mẫu dò DNA

Trong sinh học phân tử, mẫu dò là một đoạn DNA hoặc RNA hay axit amin được sử dụng để phát hiện sự có mặt của trình tự nucleotide, protein tương đồng. Mẫu dò được thiết kế dựa trên những trình tự gen/axit amin có sẵn hoặc từ trình tự genome của sinh vật có quan hệ gần gũi với đối tượng nghiên cứu. Mục đích xây dựng mẫu dò để tìm kiếm gen chức năng được bảo tồn qua tiến hoá hay trình tự axit amin của protein là sản phẩm của gen. Thông qua việc so sánh tương đồng, từ đó tìm kiếm trình tự trọn vẹn của gen mã hoá cho protein trong genome,…[121].

So sánh tương đồng cũng được xem như công cụ hữu hiệu để khai thác gen từ nguồn dữ liệu trình tự DNA đa hệ gen [151]. Đồng thời có thể sử dụng phương pháp này nghiên cứu đặc tính và biểu hiện của gen [42]. Việc lựa chọn trình tự gen mong muốn được thực hiện bằng nhiều phần mềm dự đoán cấu trúc và chức năng của gen/protein. Một số phần mềm đã ra đời để tạo mẫu dò dựa trên thông tin về gen chức năng FGA (Functional Gen Arrays) như OligoArray [140], OligoPicker [181] và YODA [128]. Trong nghiên cứu phân tích dữ liệu DNA đa hệ gen, một số loại mẫu dò đã sử dụng công cụ tin sinh học thiết kế để tìm kiếm gen như: Hierarchical - mẫu dò phân cấp dựa trên gen chức năng [31]; PhylArray là mẫu dò tìm các vi khuẩn chưa từng được biết đến dựa trên sự thoái hóa và không thoái hóa cho mỗi taxon [118]; CommOligo thiết kế mẫu dò dựa trên việc sử dụng các trình tự tương đồng cao và thông tin của genome [103]; HiSpOD sử dụng trình tự nucleotide bảo tồn hoặc trình tự nucleotide riêng lẻ để tạo ra mẫu dò có độ đặc hiệu cao [50].

Nghiên của Zhou và cộng sự đã sử dụng công nghệ PathoChip kết hợp với PCR để phát hiện sự hiện diện tác nhân gây bệnh ung thư và một số bệnh khác ở người. PathoChip thiết kế mẫu dò dựa trên các trình tự DNA và RNA đặc trưng của mỗi nhóm sinh vật bao gồm 4200 tác nhân gây bệnh (virus, vi khuẩn, nấm, ký sinh

trùng) đã được công bố trong NCBI. Dữ liệu này được tập hợp thành một DNA đa hệ gen chứa 58 nhiễm sắc thể với độ dài 448,9 triệu bp. Công cụ này đã thiết kế khoảng 5,5 triệu mẫu dò từ DNA đa hệ gen trên dựa vào vị trí hoặc khu vực bảo tồn của bộ gen gây bệnh ở các loài [12].

1.2.4.2. Một số ứng dụng của mẫu dò DNA

Mẫu dò được sử dụng trong nhiều nghiên cứu khác nhau như nhận diện bản sao hoặc sản phẩm RNA của gen, chỉ ra loài sinh vật có quan hệ gần gũi với đối tượng nghiên cứu nhằm tìm kiếm gen chức năng được bảo tồn qua tiến hoá,…[123], kiểm tra mức độ biểu hiện của gen quan tâm ở các cơ quan [9], giai đoạn sinh trưởng [105], điều kiện môi trường [73]. Sử dụng mẫu dò để kiểm tra sinh vật biến đổi gen [74]; kiểm tra sự có mặt và biểu hiện của gen chuyển, gen của virus [4],…Trong thực tế việc sử dụng mẫu dò để tìm kiếm, sàng lọc trình tự gen mong muốn từ các ngân hàng DNA đa hệ gen đã được sử dụng [5].

Cách tiếp cận sử dụng mẫu dò để khai thác gen từ dữ liệu giải trình tự DNA đa hệ gen đã được Kushwaha và cộng sự tiến hành chọn lọc enzyme thủy phân mạch carbon từ vi sinh vật đất. Mẫu dò này được thiết kế dựa vào dữ liệu của NCBI và thuật toán tin sinh học của Metcap, cho phép tự động tạo ra mẫu dò thông lượng cao thay thế cho việc thiết kế mồi tìm kiếm trình tự gen đích. Mẫu dò tự động sử dụng nhiều thông số cho phép người dùng có thể thay đổi các tiêu chuẩn như: độ dài của mẫu dò, nhiệt độ nóng chảy và hàm lượng GC [97].

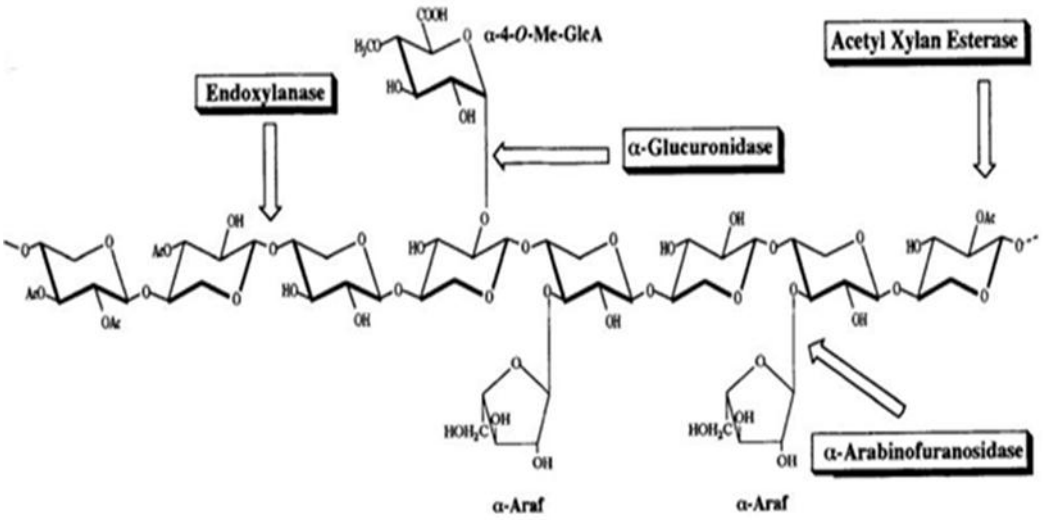
Như vậy hầu hết mẫu dò trong nghiên cứu DNA đa hệ gen đã công bố đều dựa trên CSDL sẵn có của các ngân hàng gen. Tất cả trình tự gen mã hóa cho protein đích được lựa chọn và nhập vào phần mềm một cách tự động, sau đó phần mềm sẽ sử dụng thuật toán so trùng để xây dựng nên mẫu dò. Tùy theo mục đích sử dụng mà người dùng có thể yêu cầu về độ dài của mẫu dò và thông số liên quan.

1.3. ENZYME β–xylosidase

1.3.1. Đặc điểm chung

Để phân giải hoàn toàn hemicellulose thành các đơn phân, đòi hỏi nhiều loại enzyme phân giải khác nhau do thành phần cấu trúc phức tạp. Tuy nhiên một số

enzyme đóng vai trò cắt các mạch chính và chuỗi nhánh của hemicellulose như endoxylanase, β-xylosidase, α-glucuronidase, α-arabinofuranosidase và esterase acetylxylan (Hình 1.3). Endo β-1,4 xylanase là một trong những enzyme phổ biến nhất trong quá trình thủy phân xylan. Enzyme này thủy phân liên kết 1,4 giữa các tiểu đơn vị xylose trong polyme của xylan để sản xuất oligosaccharide, từ đó β-xylosidase sẽ phân cắt thành xylose [57].

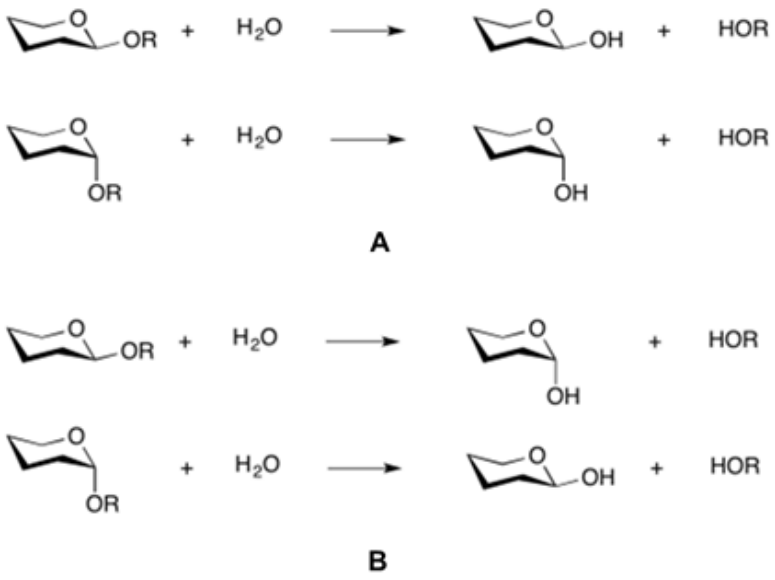


Hình 1.3. Vị trí tác dụng của một số hemicellulase trên mạch xylan

Xylan-1,4-β-xylosidase (EC 3.2.1.37) là một enzyme trong chuỗi chuyển hóa xylan ra sản phẩm cuối cùng. Ngoài tên hệ thống là xylan-1,4-β-xylosidase và tên thường gọi là β-xylosidase, enzyme này còn có một số tên gọi khác là xylobiase, exo-1,4-β-xylosidase, β-D-xylopyranosidase, exo-1,4-xylosidase, exo-1,4-β-D-xylosidase và 1,4-β-D-xylan xylohydrolase. Enzyme này có vai trò quan trọng phân giải hemicellulose nói riêng và sinh khối thực vật nói chung. Theo phân loại của CSDL CAZY thì β-xylosidase có cấu trúc khá đa dạng và được sắp xếp vào các họ GH1, 3, 30, 31, 39, 43, 51, 52, 54, 116 và 120.

1.3.2. Mô hình hoạt động

Trong hoạt động thủy phân của β–xylosidase thuộc các họ GH1, 3, 30, 31, 39, 51, 52, 54, 116 và 120 đều sử dụng cơ chế giữ nguyên cấu hình vòng, chỉ có họ GH43 sử dụng cơ chế nghịch đảo cấu hình vòng (Hình 1.4) [202].

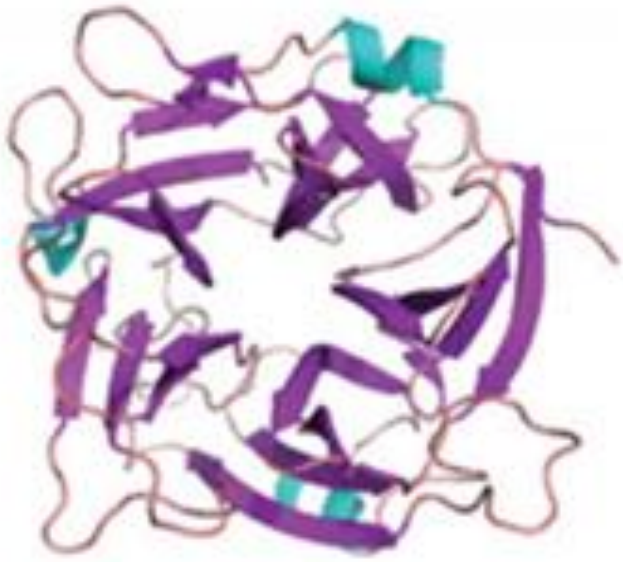


Hình 1.4. Mô hình glycoside hydrolase thủy phân liên kết glycoside

A. Cơ chế xúc tác giữ nguyên cấu hình vòng sau khi thủy phân

B. Cơ chế xúc tác nghịch đảo cấu hình vòng sau khi thủy phân

1.3.3. Cấu trúc không gian

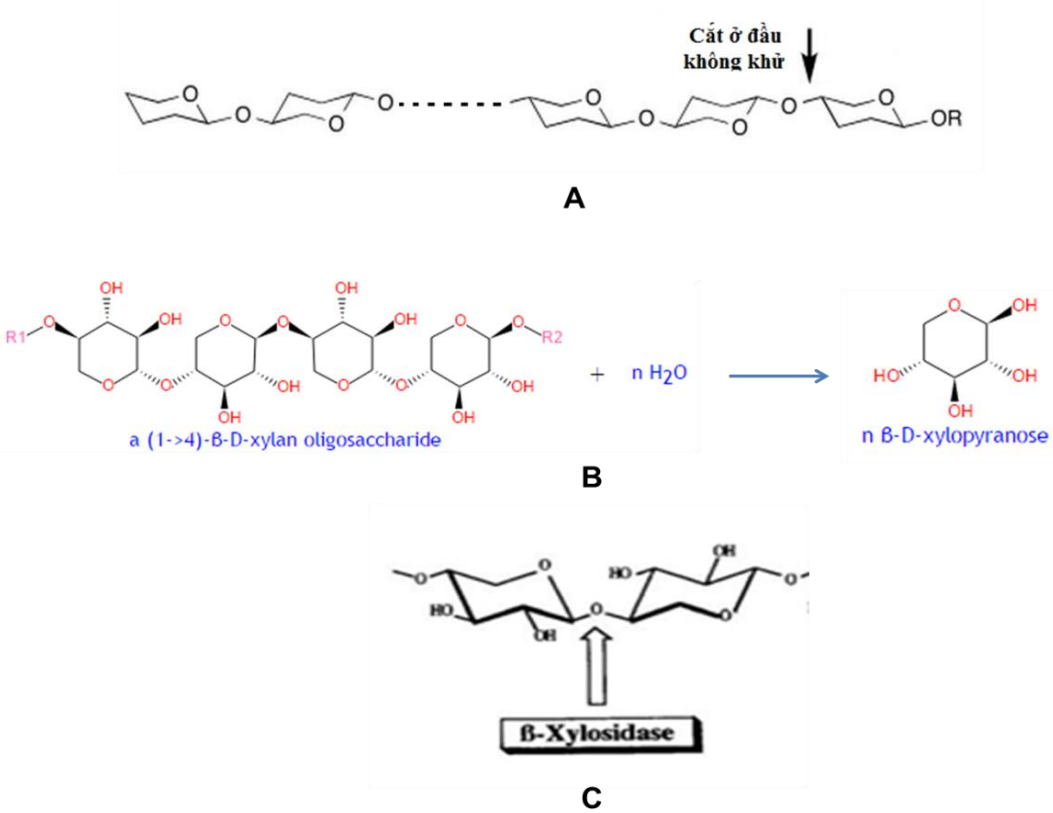


Hình 1.5. Cấu trúc không gian của enzyme thuộc GH43 ở Cellvibrio japonicus [13]

β–xylosidase có cấu trúc không gian khá đa dạng với một số mô hình như $(\beta/\alpha)_8$ (GH1, 30, 39, 51), $(\beta/\alpha)_8$ barrel (GH31) và 5-fold β-propeller (GH43). Đặc biệt enzyme thuộc họ GH43 có mô hình cấu trúc không gian gồm năm chuỗi gấp nếp β xếp thành cánh quạt tạo ra cấu trúc hình chữ V với các khe rãnh dài, xen kẽ

với chuỗi α. Trung tâm xúc tác là một túi sâu, được quyết định bởi cả hai axit amin là glutamate và aspartate (Hình 1.5). Hoạt động của hai axit amin gây ra hiện tượng đảo ngược vị trí chuỗi β và α (tức là đảo ngược cấu trúc lập thể của các nguyên tử carbon trong chất nền), nhưng không trở lại vị trí ban đầu, dẫn đến cấu hình trung tâm hoạt động của enzyme bị thay đổi. Đó chính là cơ sở giúp các enzyme trong họ GH43 có thể kết hợp với nhiều dạng cơ chất để thủy phân liên kết glycoside [13]. Các nghiên cứu khi sử dụng aryl-glycoside làm cơ chất cho thấy enzyme trong họ GH43 mang cả hai hoạt tính bao gồm alpha-L-arabinofuranosidase và beta-D-xylosidase.

1.3.4. Hoạt tính của β–xylosidase



Hình 1.6. Hoạt tính của β–xylosidase

A. *Vị trí cắt của β-xylosidase trên chuỗi xylan,* ***B.*** *β–xylosidase thủy phân (1,4)-β-D-xylan,* ***C.*** *Vị trí cắt của xylan β–xylosidase trên xylobiose*

β–xylosidase (E.C 3.2.1.37) là một enzyme có khả năng xúc tác phản ứng cắt liên kết β(1→4) glycoside ở đầu không khử của chuỗi xylan. Hoạt tính chủ yếu của

enzyme này là loại bỏ gốc D-xylose khỏi đầu không khử của (1,4)-β-D-xylan oligosaccharide (Hình 1.6A), nhờ đó chuỗi (1,4)-β-D-xylan oligosaccharide được thủy phân thành các tiểu phần β-D-xylanpyranose (Hình 1.6B). Cũng nhờ khả năng phân cắt liên kết β-1,4 này, β–xylosidase còn thủy phân xylobiose (phân tử đường đôi gồm 2 monome xylose liên kết với nhau bằng liên kết β-1,4) (Hình 1.6C). Một số hoạt tính exoglycosidase khác của enzyme này cũng được tìm thấy ở gan cừu [28].

1.3.5. Ứng dụng của β–xylosidase

β–xylosidase được coi là enzyme quan trọng phá vỡ liên kết glycoside của hemicellulose. Trong công nghiệp enzyme này được sử dụng để phục vụ cho việc sản xuất xylitol và giấy tái chế [86], giúp chuyển hóa sợi, giảm sử dụng chất tẩy trong sản xuất vải, bột giấy và giấy [160].

β–xylosidase kết hợp với endo - xylanase trong sản xuất bánh mì sẽ cải thiện quá trình nhào bột, tăng khả năng hấp thụ nước và độ xốp của bánh [160], giảm tính nhớt khi ủ bia [20], giúp tách nước và giải phóng hương thơm từ nho khi sản xuất rượu vang [96] và cải thiện khả năng tiêu hóa của vật nuôi khi bổ sung vào thức ăn [86].

Cùng với cellulase và các hemicellulase khác, β–xylosidase giúp phân hủy lignocellulose tạo ra các monosaccharide - nguyên liệu cung cấp cho sản xuất nhiên liệu sinh học [86]. Việc sản xuất nhiên liệu sinh học từ sinh khối thực vật đòi hỏi phải có hỗn hợp các cellulase và hemicellulase được xử lý qua từng bước khác nhau, trong đó β–xylosidase là một thành phần không thể thiếu [124].

1.3.6. Nguồn cung cấp β–xylosidase

β–xylosidase thường được tìm thấy ở sinh vật sử dụng sinh khối thực vật làm thức ăn, phần lớn trong số đó là vi khuẩn và nấm [18]. Nguồn cung cấp β–xylosidase khá đa dạng như từ thực vật rừng [34], vi sinh vật sống trong dạ cỏ của động vật nhai lại [78] và trong ruột mối [115],…. Trong số các nguồn khai thác trên, thì vi sinh vật trong ruột mối là một trong những đối tượng tiềm năng, nhờ vào khả năng phân giải hiệu quả lignocellulose.

Mối *C. gestroi* rất phổ biến ở Việt Nam cũng như một số quốc gia trên thế giới. Một số nghiên cứu đã chỉ ra có 16 họ GH của vi sinh vật sống trong ruột *C. gestroi* là GH2, 3, 5, 7, 10, 11, 16, 20, 26, 30, 42, 45, 47, 53, 77, 92 [61]. Vì vậy, chúng tôi sẽ tập trung nghiên cứu gen mã hóa β–xylosidase từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật ruột mối *C. gestroi* ở Việt Nam, để chọn lọc và biểu hiện, tinh chế thành công, cung cấp nguồn enzyme cho thực tiễn.

1.4. KHU HỆ VI SINH VẬT VÀ ENZYME CHUYỂN HÓA LIGNOCELLULOSE

1.4.1. Một số khu hệ vi sinh vật chuyển hóa lignocellulose

Tìm kiếm và khai thác được nguồn enzyme để xử lý sinh khối thực vật dư thừa sẽ ứng dụng hiệu quả trong thực tiễn như sản xuất nhiên liệu sinh học, thức ăn chăn nuôi, hạn chế ô nhiễm môi trường,…. Trong tự nhiên sự phân hủy lignocellulose chủ yếu là do vi sinh vật. Do đó mà quá trình phân hủy lignocellulose bằng các enzyme từ vi sinh vật có một ý nghĩa rất lớn.

Trong những năm gần đây, nhiều nghiên cứu tập trung vào việc tìm kiếm những enzyme có hoạt tính mạnh, giúp chuyển hóa nhanh lignocellulose thành đường theo hướng tiếp cận và sử dụng kỹ thuật Metagenomics. Hướng này cho phép khai thác được nhiều gen quý từ các vi sinh vật chưa nuôi cấy được vốn chiếm từ 99% đến 99,9% vi sinh vật có trong môi trường tự nhiên. Nhiều dự án nghiên cứu đã thực hiện theo hướng tiếp cận lập thư viện DNA đa hệ gen như: Từ thư viện DNA đa hệ gen của vi sinh vật trong dạ cỏ trâu đã phân lập được 13 dòng có hoạt tính endoglucanase và 14 gen cellulase [49]. Nhóm vi sinh vật phong phú nhất trong dạ cỏ là vi khuẩn, sau đó động vật nguyên sinh, cổ khuẩn và nấm. Vi khuẩn tồn tại trong dạ cỏ không giống nhau và phụ thuộc vào chế độ ăn của trâu [154]. Từ thư viện DNA đa hệ gen của vi sinh vật sống trong chất thải của nhà máy giấy, đã khai thác được bốn nhóm vi khuẩn phổ biến là *Spirochaetes, Proteobacteria, Bacteroidetes và Firmicutes*. Đồng thời đã phân lập được 2 dòng có hoạt tính endoglucanase, 3 dòng có hoạt tính exoglucanase và 2 dòng có hoạt tính β-glucosidase [187].

Tuy nhiên để hiểu rõ thành phần vi sinh vật, nhiều nghiên cứu đã sử dụng kỹ thuật Metagenomics theo hướng giải trình tự toàn bộ DNA đa hệ gen. Ví dụ kết quả phân tích trình tự DNA đa hệ gen với kích thước 674.959.683 bp của quần xã vi sinh vật kỵ khí phân hủy vụn gỗ Bạch Dương đã dự đoán được 893.380 gen. Trong đó số trình tự mã hóa hemicellulase chiếm tỷ lệ cao hơn so với cellulase. Các ngành vi khuẩn chiếm tỷ lệ cao nhất là *Firmicutes* (45,9%) và *Proteobacteria (*32,3%), tiếp theo là *Bacteroidetes* (9,9%) [185]. Nghiên cứu khu hệ vi sinh vật sống trong hai mẫu đất ủ phân hữu cơ ZC1 và ZC2 ở công viên vườn bách thú São Paulo (Brazil) có kích thước tương ứng là 836 Mbp và 842 Mb. Trong đó mẫu ZC1 có 15.623 gen và mẫu ZC2 có 76.175 gen được dự đoán. Vi khuẩn là loài chiếm ưu thế, sống trong mẫu ZC1 từ 84% đến 89% và trong mẫu ZC2 từ 93% đến 96%, còn lại là cổ khuẩn, virus và sinh vật nhân thực. Mẫu ZC1 và ZC2 đều có nhiều gen mã hóa enzyme tham gia thủy phân lignocellulose như: ZC1 có 112 trình tự mã hóa cellulase thuộc họ GH5 và 32 trình tự của vùng liên kết cellulose (CBD), 65 trình tự axit amin vùng dokerin (pfam00404), 36 trình tự vùng cohesin (pfam00963), 867 trình tự gen mã hóa cellulobiose phosphorylase. Mẫu ZC2 có 9 trình tự mã hóa cellulase, 6 trình tự axit amin vùng dokerin (pfam00404) và 8 trình tự vùng cohesin (pfam00963), 267 trình tự gen mã hóa cellulobiose phosphorylase. Đồng thời, ZC1 và ZC2 có 584 gen liên quan đến phân hủy pectin như pectate lyase và pectin methylesterase [110].

Trong số các khu hệ vi sinh vật chuyển hóa lignocellulose hiệu quả phải kể đến hệ vi sinh vật sống trong ruột mối. Nhiều nghiên cứu sử dụng kỹ thuật Metagenomics để tìm hiểu sự đa dạng của vi sinh vật và khai thác enzyme chuyển hóa lignocellulose từ cả mối bậc thấp và bậc cao đã công bố. Tổng quan về khu hệ vi sinh vật trong ruột mối sẽ được chi tiết trong nội dung tiếp theo.

1.4.2. Hệ vi sinh vật và enzyme thủy phân lignocellulose trong ruột mối

1.4.2.1. Đặc điểm sinh học của mối

Mối phân bố giữa 47° vĩ độ bắc và 47° vĩ độ nam và chiếm 68% bề mặt Trái Đất, với tổng số khoảng $2,4.10^{17}$ cá thể. Khu vực mối phân bố chính thuộc vùng

nhiệt đới và cận nhiệt đới [92]. Theo ước tính, hàng năm mối tiêu thụ khoảng 3 đến 7 tỷ tấn lignocellulose – với hiệu quả của quá trình tiêu hóa từ 79% đến 99% [172].

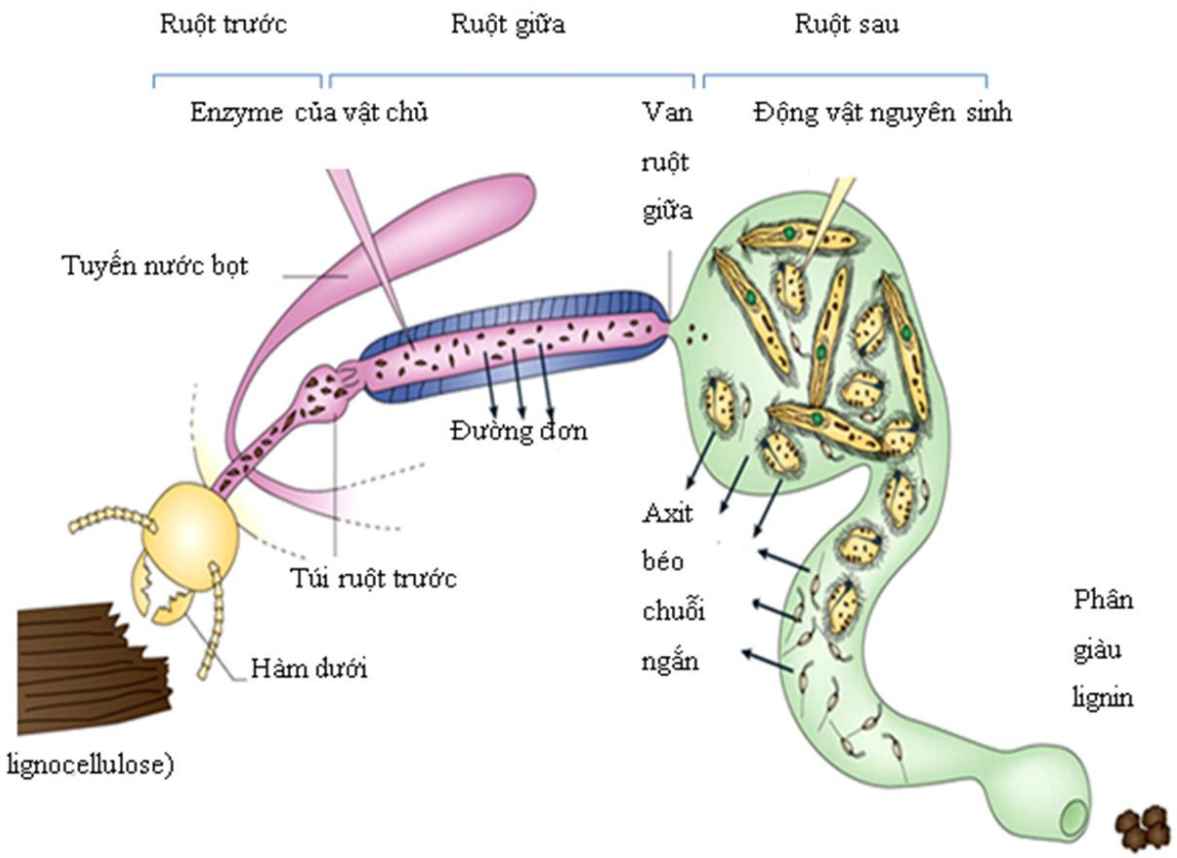
Ở Việt Nam, có 141 loài mối phân thành 4 họ: Kalotermitidae, Termopsidae, Rhinotermitidae, Termitidae. Trong đó ba loài *C. gestroi*, *Coptotermes emersoni*, *Coptotermes travians* thường xuyên được tìm thấy tại Hà Nội và loài *Coptotermes curvignathus* xuất hiện ở miền nam Việt Nam [125], [126], [174]. Trong các loài mối ở Việt Nam, *Coptotermes* là giống mối đáng chú ý nhất vì chúng là những loài mối ăn gỗ rất phổ biến, phá hoại nhiều công trình xây dựng của con người.

Mối sống thành tập đoàn, mang tính xã hội, chia thành đẳng cấp khác nhau bao gồm mối vua, mối chúa, mối lính, mối thợ và mối cánh. Chúng sống trong một tổ, được phân hoá giữ chức năng và nhiệm vụ khác nhau. Trong đó, mối thợ chịu trách nhiệm tìm kiếm thức ăn, tiêu hóa và nuôi dưỡng các thành viên còn lại trong tổ [36]. Mối bậc cao sử dụng nấm làm thức ăn, còn mối gỗ, mối đất và mối bậc thấp hoàn toàn chỉ sử dụng gỗ làm thức ăn [112]. Cho đến nay vẫn chưa được xác định chính xác sự tương tác giữa tế bào của ruột mối với vi sinh vật sống trong ruột mối để phân hủy lignocellulose.

Ruột mối có cấu trúc rất phức tạp, được chia ra làm ba bộ phận riêng biệt gồm ruột trước, ruột giữa và ruột sau (Hình 1.7). Ruột trước bao gồm dạ dày là nơi nghiền thức ăn và tuyến nước bọt tiết ra một số endoglucanase. Ruột giữa hình ống với môi trường hiếu khí – là nơi sản xuất endoglucanase ở mối bậc cao và diễn ra một vài sự thủy phân lignocellulose. Ruột sau có dạ cỏ - chính là một bể lên men kỵ khí. Đây là nơi cư trú của nhiều loài vi sinh vật, diễn ra phần lớn sự thủy phân cellulose và các quá trình lên men các sản phẩm từ sự thủy phân đó [152].

Sự tiêu hóa lignocellulose của mối được thực hiện nhờ hệ enzyme được tiết ra từ bản thân mối và vi sinh vật sống trong ruột mối. Sau khi thức ăn được nghiền nhỏ ở miệng sẽ trộn với enzyme từ tuyến nước bọt và nghiền nhỏ ở túi ruột trước. Thức ăn qua ruột giữa sẽ tiếp tục được tiêu hóa một phần. Phần thức ăn còn lại sẽ bị thủy phân bởi enzyme của vi sinh vật sống ở ruột sau [131]. Với kích thước ruột nhỏ nên

thời gian vận chuyển của thức ăn qua đường ruột chỉ khoảng 24 giờ đã ảnh hưởng đáng kể đến thành phần của vi sinh vật trong ruột mối.



Hình 1.7. Cấu tạo ruột mối [24]

1.4.2.2. Hệ vi sinh vật sống tự do trong ruột mối

Dựa trên sự có mặt hay vắng mặt của động vật nguyên sinh trong ruột mối mà mối được chia làm hai nhóm là mối bậc thấp và mối bậc cao [1]. Động vật nguyên sinh chỉ có ở ruột mối bậc thấp, với số lượng từ 10^3 đến 10^5 cá thể, hơn 90% trong số đó chủ yếu tập trung ở ruột sau. Chúng còn là nơi sinh sống của rất nhiều loài vi sinh vật nhân sơ nội bào và ngoại bào [25].

Vi khuẩn chiếm trên 90% vi sinh vật, được tìm thấy trong cả mối bậc thấp và mối bậc cao. Trong đó, Spirochaetes là ngành đặc trưng của vi sinh vật ruột mối ăn gỗ, chiếm tỷ lệ cao (có thể hơn 50% tổng số prokaryote). Chúng có thể sống bám trên bề mặt tế bào động vật nguyên sinh hoặc sống tự do trong dịch ruột của ruột sau [82]. Thông thường những loài vi khuẩn có kích thước lớn sống tự do, còn

những loài vi khuẩn có kích thước nhỏ thường sống trong tế bào chất hoặc bám vào bề mặt ngoài của động vật nguyên sinh. Ngoài ra những loài sống cộng sinh bên trong động vật nguyên sinh thường chiếm một phần đáng kể của quần xã vi khuẩn trong ruột sau của mối [24]. Vi khuẩn đường ruột của hai loài mối bậc thấp phổ biến là *C. formosanus* và *R. speratus*, chủ yếu thuộc các ngành có khả năng chuyển hóa lignocellulose như: *Bacteroidetes, Spirochaetes, Fibrobacters, Elusimicrobia, Firmicutes* và *Proteobacteria*. Tỷ lệ loài vi khuẩn thuộc các ngành chiếm ưu thế của hai loài *C. formosanus* và *R. speratus* hoàn toàn khác nhau. Trong đó, Bacteroidetes chiếm tỷ lệ rất cao trong đường ruột của *C. formosanus*, còn Spirochaetes chiếm ưu thế trong đường ruột của *R. speratus* [24]. Thành phần quần xã vi khuẩn rất giống nhau giữa các loài mối trong cùng một chi, nhưng lại thay đổi mạnh giữa hai chi. Tuy nhiên quần xã vi khuẩn trong ruột mối mang tính đặc trưng cho từng loài mối [95].

1.4.2.3. Sự đa dạng enzyme tham gia vào chuyển hóa lignocellulose của mối

Các nghiên cứu trước đây đã chỉ ra trong ruột mối bậc thấp, động vật nguyên sinh là nguồn sản xuất enzyme phân hủy cellulose [138]. Trùng roi Parabasalia sống trong ruột mối có khả năng sản xuất exocellulobiohydrolase và endoglucanase cần thiết để phân hủy hiệu quả cellulose tinh thể. Gen mã hóa cellulase đã được phân lập từ trùng roi *Pseudotrichonympha grassii* sống trong ruột *C. formosanus* [55]. Hệ động vật nguyên sinh của mối *Coptotermes lacteus* có khả năng sản xuất β-1,4-glucosidase, endo-β-1,4-glucanase và exo-β-1,4-glucosidase. *Trichomitopsis termopsidis* sống trong ruột mối *Zootermopsis* có thể tiêu hóa được cả cellulose tinh thể lẫn xylan [130]. Ở ruột mối *Mastotermes darwiniensis, Hodotermopsis sjoestedti, Neotermes koshunensis, C. formosanus* và *Reticulitermes flavipes*, động vật nguyên sinh có khả năng tiết lignocellulase thuộc họ GH5 và GH7 [171].

Ở cả mối bậc thấp và bậc cao, lignocellulase được sản xuất chủ yếu từ hệ vi khuẩn sống trong ruột. Một số nghiên cứu đã phân lập được các gen đơn lẻ mã hóa enzyme thủy phân lignocellulose từ vi khuẩn như: Gen mã hóa cellulase kích thước 2,25 kb đã được phân lập từ *Enterobater loacase* trong ruột mối *Heterotermes indicola* [129]; Vi khuẩn *Dyella sp, Chryseobacterium sp* và *Bacillus* được phân lập

từ ruột *Reticulitermes speratus* sản xuất enzyme endo-beta-1,4-glucanase [29]; endocellulase được sản xuất từ hai loài vi khuẩn thuộc họ *Bacillus* sống trong ruột mối *C. formosanus* [169]. Ở mối bậc thấp như *R. flavipes*, có 18% cellulase, 31% hemicellulase, 57% chitinase và 13% alpha-carboxylhydrolase là có nguồn gốc từ vi khuẩn sống trong ruột sau [165].

1.4.3. Tổng quan nghiên cứu về đa dạng vi sinh vật và enzyme chuyển hóa lignocellulose trong ruột mối *C. gestroi* ở Việt Nam

1.4.3.1. Xác định đa dạng vi sinh vật trong ruột mối C. gestroi ở Việt Nam

Nghiên cứu trước đây của phòng Kỹ thuật Di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tiến hành thu thập sáu mẫu mối ở 6 địa điểm khác nhau (phường Văn Quán, quận Hà Đông; xã Tản Lĩnh, huyện Ba Vì; phố Bùi Xương Trạch, quận Thanh Xuân; phố Thái Hà, quận Đống Đa; phố Võng Thị, quận Tây Hồ và 1 địa điểm tại huyện Văn Lâm, tỉnh Hưng Yên), để giải trình tự DNA đa hệ gen trong dự án "Phân lập gen mã hóa lignocellulase của vi sinh vật sống trong ruột mối *C. gestroi* ở Việt Nam bằng kỹ thuật Metagenomics", giai đoạn 2012-2015 hợp tác giữa Việt Nam với Nhật Bản. Dựa vào trình tự DNA mã hóa rRNA 16S ti thể, sáu mẫu mối này đã được định danh đều thuộc loài *Coptotermes gestroi* [126].

DNA đa hệ gen của vi sinh vật sống trong ruột mối *C. gestroi* được tách chiết và đọc trình tự bằng máy giải trình tự HiSeq 2000 của Illumina. Sau đó, bằng phần mềm chuyên dụng SOAPdenovo và công cụ tin sinh học, máy đã sắp xếp các trình tự ngắn (read) đạt yêu cầu thành contig. Phần mềm MetagenAnnotator được sử dụng để dự đoán gen thuộc các contig. Trình tự gen tiếp tục được dự đoán theo 2 khía cạnh là chức năng sinh học và đơn vị phân loại dựa vào mức độ tương đồng của gen với CSDL tham khảo. Trên cơ sở bộ dữ liệu trình tự DNA đa hệ gen, độ đa dạng vi sinh vật trong ruột mối *C. gestroi* được đánh giá theo 2 cách: *Thứ nhất,* các trình tự đọc ngắn (read) chất lượng cao được so sánh trực tiếp với các trình tự của CSDL từ NCBI, RDP (Ribosomal Database Project) [33], hệ gen nấm và gen vi sinh vật đường ruột người. *Thứ hai,* tất cả các ORF được phần mềm MEGAN (MEtaGenomic ANalyser) [80] phân tích dựa vào CSDL NR (Non-redundant)

[120]. Kết quả cho thấy với kích thước của trình tự DNA đa hệ gen là 5,431.60 Mb, đã xác định được 125.431 ORF và dự đoán thuộc về 1.460 loài, 731 chi, 282 họ, 138 bộ, 65 lớp và 46 ngành của vi sinh vật. Trong đó, vi khuẩn chiếm 1.368 loài thuộc 628 chi, 217 họ, 97 bộ, 41 lớp và 22 ngành. Như vậy số ORF thuộc vi khuẩn là 80%, thuộc virus là 0,2%, cổ khuẩn là 0,42%, sinh vật nhân chuẩn là 0,58%, còn lại 18,8% là các ORF không dự đoán được [45]. Trong nghiên cứu này mới chỉ dừng lại phân tích sự đa dạng các loài vi sinh vật sống trong ruột mối *C. gestroi* ở mức thống kê chung dựa trên ước đoán chức năng của ORF, chưa đi sâu vào nghiên cứu vai trò của vi sinh vật chiếm số lượng lớn nhất là vi khuẩn trong chuyển hóa thức ăn. Nghiên cứu tiếp theo của chúng tôi sẽ khai thác và phân tích sâu về vai trò của vi khuẩn tham gia chuyển hóa thức ăn trong ruột mối *C. gestroi* ở Việt Nam.

1.4.3.2. Nghiên cứu về dự đoán chức năng gen và lựa chọn ORF để biểu hiện từ dữ liệu DNA đa hệ gen của sinh vật trong ruột C. gestroi

Sau khi giải trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* và dự đoán có 125.431 ORF, nghiên cứu đã tiếp tục dịch mã các ORF sang trình tự axit amin tương ứng. Dựa trên trình tự axit amin để so sánh với các CSDL eggNOG [134], COG [166] và KEGG [120] bằng BLASTP để dự đoán chức năng. Dựa vào CSDL eggNOG, trình tự axit amin tương ứng của ORF thể hiện chủ yếu chức năng của sinh vật nhân sơ. Dựa vào CSDL COG, 7.428 ORF được dự đoán liên quan đến quá trình chuyển hóa carbohydrate và xếp vào 210 nhóm COG. Dựa vào CSDL của KEGG, kết quả dự đoán cũng thể hiện số lượng ORF liên quan đến chuyển hóa carbohydrate chiếm 15.402 ORF và cao hơn rất nhiều so với kết quả dự đoán dựa vào CSDL COG. Kết quả dự đoán tổng hợp được 587 ORF mã hóa enzyme phân hủy lignocellulose thuộc về cellulase (316 ORF), hemicellulase (259 ORF), pectatelyase và pectinesterase (12 ORF), không có enzyme chuyển hóa lignin [45].

Trong 587 ORF mã hóa cho lignocellulase khai thác từ dữ liệu DNA đa hệ gen vi sinh vật trong ruột mối *C. gestroi,* chỉ có 99 ORF là chứa trọn vẹn gen (gen hoàn thiện, chứa cả hai đầu 3’ và 5’), còn lại 488 ORF là không hoàn thiện (mất một trong hai đầu hoặc mất cả hai đầu 3’ và 5’). Nhóm ORF hoàn thiện gồm có 55 ORF mã hóa cho 4 nhóm cellulase và 44 ORF mã hóa cho 6 nhóm hemicellulase. Trình

tự axit amin của cellulase và hemicellulase được mã hóa bởi ORF hoàn thiện đã được sử dụng để tìm kiếm các trình tự protein tương đồng trên ngân hàng gen NCBI bằng công cụ BLASTP. Qua đó xác định chức năng, nguồn gốc và đặc biệt là tính mới của gen (thông qua hệ số tương đồng tối đa). Kết quả tìm kiếm cho thấy, trong 99 ORF hoàn thiện mã hóa cho cellulase và hemicellulase có: 38 ORF tương đồng từ 86% đến 100%, 36 ORF tương đồng từ 60% đến 85% và 25 ORF có độ tương đồng thấp dưới 60% với các gen đã biết [45].

Nghiên cứu trước đây đã lựa chọn gen từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối theo 04 bước: (1) Trình tự các nucleotide trong ORF được chuyển sang trình tự axit amin trong chuỗi polipeptide và sử dụng để xác định vùng bảo thủ dựa trên BLASTP và BLAST PSI (lựa chọn các trình tự có vùng hoạt tính rõ ràng, đặc hiệu và đặc thù cho enzyme cần chọn). (2) Đánh giá so sánh tương đồng với các trình tự tương ứng trên ngân hàng gen (lựa chọn trình tự nằm trong nhóm enzyme đích có độ tin cậy cao và có tính mới dựa vào độ tương đồng khi so sánh). (3) Bước tiếp theo là truy nguyên nguồn gốc của gen (ưu tiên gen có nguồn gốc từ vi khuẩn). (4) Cuối cùng là lựa chọn trình tự đơn giản để dễ dàng biểu hiện gen trong *E. coli*. Các bước trên được thể hiện chi tiết trong luận án nghiên cứu "Nghiên cứu sàng lọc gen mã hóa enzyme tham gia thủy phân cellulose từ khu hệ vi sinh vật trong ruột mối bằng kỹ thuật Metagenomics" của nghiên cứu sinh Nguyễn Thị Thảo (2015) [193]. Theo cách làm trên, nghiên cứu sinh Thảo đã tiến hành chọn lọc và so sánh trình tự axit amin của 5 ORF hoàn thiện mã hóa endoglucanase với CSDL của NCBI bằng BlastP. Kết quả cho thấy cả 5 trình tự đều có tỷ lệ tương đồng từ 46% đến 88% so với các trình tự tương ứng trong CSDL của NCBI. Dựa vào tỷ lệ tương đồng trình tự nucleotide và trình tự axit amin, đã lựa chọn 2 ORF có mã số là GL0033071 và GL0130684 để nghiên cứu biểu hiện. Trình tự axit amin của 2 ORF này cũng được dự đoán chức năng bằng cách so sánh với CSDL khác như COG, Pfarm,… và dự đoán cấu trúc 3D bằng Phyre 2. Kết quả dự đoán cho thấy trình tự axit amin của GL0130684 có quan hệ rất gần gũi với endoglucanase của nhiều loài sinh vật và gần nhất với endoglucanase thuộc vi khuẩn *Yokenella regensburgei*. Đồng thời khi dự đoán bằng công cụ tin sinh học trình tự axit amin của GL0130684 cũng thể hiện hoạt tính endoglucanase rõ ràng. Tuy nhiên sau khi

tiến hành thực nghiệm biểu hiện trong *E. coli* thì hoạt tính endoglucanase lại khá thấp (11,793 U/mg). Ngoài ra, nhóm nghiên cứu cũng đã thử nghiệm biểu hiện 8 gen khác. Kết quả biểu hiện cho thấy: Protein không tan (02), hoặc tan ít (02), biểu hiện kém (02) và biểu hiện tan (02) và có hoạt tính yếu (Số liệu không công bố).

Những hạn chế của nghiên cứu trước đây đã đặt ra yêu cầu phải xây dựng được một phương pháp hiệu quả để tìm kiếm nhanh được gen đích mã hóa đúng lignocellulase từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* và phải có hoạt tính đúng của enzyme mục tiêu sau khi biểu hiện. Ngoài ra, sự đa dạng hệ vi khuẩn và enzyme thủy phân lignocellulose trong ruột mối cũng chưa được nghiên cứu. Những vấn đề tồn tại trên sẽ được chúng tôi sẽ giải quyết trong nghiên cứu này.

Chương 2. ĐỐI TƯỢNG, VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. ĐỐI TƯỢNG VÀ VẬT LIỆU

2.1.1. Đối tượng

Trong nghiên cứu chúng tôi sử dụng dữ liệu trình tự DNA đa hệ gen của vi sinh vật sống trong ruột mối *C. gestroi* đã được chú thích bởi BGI, bao gồm 125.432 ORF do phòng Kỹ thuật di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam cung cấp. Trong đó có 587 ORF mã hóa enzyme thủy phân lignocellulose làm nguồn dữ liệu để phân lập gen.

Các chủng vi sinh vật: chủng *E. coli* DH10B (F–*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L *nup*G λ–) của hãng Invitrogen (Mỹ) được sử dụng làm thể nhận trong thí nghiệm tách dòng gen.

Chủng *E. coli Rosetta 1* (F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^R)) là chủng thương mại của hãng Novagen được sử dụng làm chủng biểu hiện gen. Trong đó một số gen đã bị gây đột biến để thuận lợi cho thực nghiệm biểu hiện gen như: khuyết gen *omp*T tránh cho protein tái tổ hợp bị phá hủy ngay trong tế bào và giúp cải thiện thu hồi protein tái tổ hợp còn nguyên vẹn [66], khuyết *hsd*S không mã hóa protease có chức năng phân giải plasmid ngoại lai xâm nhập tế bào chủ làm tăng hiệu quả biến nạp vector tái tổ hợp, khuyết gen *gal* tránh việc tế bào sử dụng galactose làm nguồn carbon cho sinh trưởng [153], khuyết gen *dcm* giúp gen ngoại lai tránh bị methyl hóa, dẫn đến sai khác trong việc biểu hiện protein tái tổ hợp [19]. Mặt khác *E. coli Rosetta 1* chứa đoạn gen mã hóa cho T7- ARN polymease, có nguồn gốc từ bacteriophage T7 và đặt dưới sự điều khiển của promoter *lac*, được cảm ứng biểu hiện bởi IPTG. Ngoài ra *E. coli Rosetta 1* còn được thiết kế để tăng cường biểu hiện protein chứa các codon hiếm. Chúng chứa thêm gen mã hóa tRNA vận chuyển argU, argW, ileX, glyT, leuW, proL, metT, thrT, tyrU, thrU và bổ sung thêm 6 codon hiếm là AUA, AGG, AGA, CUA, CCC, và GGA.

Trình tự gen mã số Termite 2_ GL0112518 mã hóa β–xylosidase gồm 1077 nucleotide có được từ dữ liệu giải trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi (Phụ lục 1)*.

Vector pET22b(+) dùng làm vector biểu hiện gen *Xbx14* trong *E. coli Rosetta 1*. Hệ vector pET được Studier và cộng sự thiết kế dựa trên cơ sở promoter mạnh có nguồn gốc từ phage T7 và gen ngoại lai được biểu hiện dưới sự phiên mã của T7-RNA polymease [183]. Vector pET22b(+) có chiều dài 5493 bp, được sử dụng cho cả tách dòng và biểu hiện gen ngoại lai trong tế bào *E. coli*. Thành phần của vector pET22b(+) chứa đầy đủ yêu cầu phục vụ cho mục đích tách dòng và biểu hiện như: Trình tự khởi đầu sao chép (*ori*), gen chỉ thị chọn lọc (Amp^R) kháng Ampicillin, vùng đa nối MCS (Multiple Cloning Site) có chứa điểm cắt của một số enzyme hạn chế, promoter T7 kiểm soát phiên mã, gen quy định chuỗi tín hiệu tiết *pelB* trước đầu tận cùng N của protein có tác dụng định hướng protein tái tổ hợp chuyển ra khoang chu chất của tế bào, phía sau vị trí cắt đa điểm có một đoạn trình tự mã hóa cho 6 axit amin Histidine tích điện âm (đuôi His-tag) để tinh sạch protein tái tổ hợp bằng cột sắc kí ái lực, trình tự kiểm soát dịch mã như điểm bám của ribosome được bố trí thích hợp và codon khởi đầu ATG.

pG - KJE8 (TaKaRa) là vector biểu hiện chaperone, có kích thước 11,1 kb do hãng TaKaRa cung cấp. Đây là plasmid được thiết kế với mục đích biểu hiện nhiều loại chaperone khác nhau. Trong thành phần cấu tạo gồm gen mã hóa DnaK-DnaJ-GrpE dưới sự kiểm soát của promoter araB, được cảm ứng biểu hiện bởi L-arabinose và gen mã hóa GroES-GroEL dưới sự kiểm soát của promoter pzt – 1, được cảm ứng bởi Tetracyline. Trên vector có chứa gen kháng Chloramphenicol giúp cho việc kiểm soát sự tồn tại và chọn lọc plasmid.

Cặp mồi dùng để kiểm tra gen *Xbx14* từ DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* được đặt tổng hợp tại hãng Integrated DNA Technologies (Singapore).

GL0112518F-*Nco*I: A<u>CCATGG</u>ATGATAAAGTTACCAATCCGGTG

GL0112518R-*Xho*I: A<u>CTCGAG</u>ACGAGCAGCGTTCAG

2.1.2. Hóa chất và thiết bị máy móc

Các loại hóa chất: Hóa chất sử dụng trong nghiên cứu được đặt mua từ các công ty đạt tiêu chuẩn quốc tế như Bio-Lab (Mỹ), Fermentas (Mỹ), Sigma (Mỹ), Merck (Đức) bao gồm: SDS, Tris-HCl, EDTA, glycine, cao nấm men, bactor peptone, methanol, ethidium bromide, glycerol, ethanol, acetic acid, chloroform,

phenol, iso-amylalcohol, glucose, acrylamide/bis-acrylamide, hydrogen peroxide, imidazole, sodium carbonate, disodium hydrogen phosphate dodecahydrate, sodium dihydrogen phosphate dihydrate, sodium hydroxide, sodium chloride, potassium acetate, agar, dNTP, agarose, nickel (II) chloride, carboxymethylcellulose, xylan, 4-nitrophenyl β-D-glucopyranoside, sodium potassium tartrate, PEG 8000, 3,5-dinitrosalicylic axit và những hóa chất thông dụng khác dùng trong sinh học phân tử.

Các loại enzyme: *Taq* DNA polymease (Fermentas, Mỹ); các loại enzyme hạn chế (Bio-Lab và Fermentas, Mỹ); T4 DNA ligase (Fermentas, Mỹ).

Máy móc và thiết bị: Máy PCR (Applied Biosystems, Mỹ), máy ly tâm lạnh bé (Sorvall RC5B, Mỹ), máy ly tâm lạnh lớn (Sorvall RC5B, Mỹ), máy ổn nhiệt (Mỹ), máy điện di (Bio-Rad, Mỹ), máy biến tính, máy UV (Bio-Rad, Mỹ), máy lắc ổn nhiệt (New Jersey, Mỹ), cân phân tích (Precisa, Thụy Sĩ), cân điện tử (Precisa, Thụy Sĩ), máy đo pH (Metler, Thụy Sĩ), máy lắc ổn nhiệt, tủ nuôi cấy tế bào (Multitron, Đức), máy sắc kí FPLC (Amersham Pharmacia, Mỹ), máy hút chân không – speed Vac Sc 110ª (Savant, Mỹ), máy đo pH (Hana Instrument, Mỹ), máy NanoDrop (Implen, Đức), tủ lạnh sâu -80°C (Panasonic, Nhật), máy đọc ELISA ELx800 (BioTek, Mỹ), máy quang phổ UV 165°C (Shimadzu, Nhật Bản), máy giải trình tự thế hệ mới Hiseq 2000 (Illumina, Mỹ).

Môi trường nuôi cấy vi khuẩn E. coli: Luria-Bertani (LB) lỏng có 0,5% cao nấm men, 1% peptone và 0,1% NaCl. Môi trường LB bổ sung 100 mg/l Ampicillin (LBA), 20 mg/l Chloramphenicol (LBC).

Các dung dịch: Các dung dịch sử dụng trong tách chiết DNA plasmid từ *E. coli* gồm: Dung dịch I có thành phần gồm 50 mM glucose; 25 mM Tris-HCl, pH = 8,0 và 10 mM EDTA, pH = 8,0; Dung dịch II gồm có 0,2N sodium hydroxide và 1% SDS; Dung dịch III gồm 3M potassium acetate và 11,5% acetic axit. Dung dịch phenol: chloroform: isoamylalcohol tỷ lệ theo thể tích tương ứng 25: 24: 1.

Dung dịch sử dụng trong điện di DNA gồm dung dịch TAE đặc 50 lần (24,2 g Tris-base; 5,71 ml glacial acetic acid; 10 ml EDTA 0,5 M, pH = 8,0; bổ sung nước đến 100 ml). Dung dịch nhuộm gel ethidium bromide (EtBr) 0,5 µg/ml.

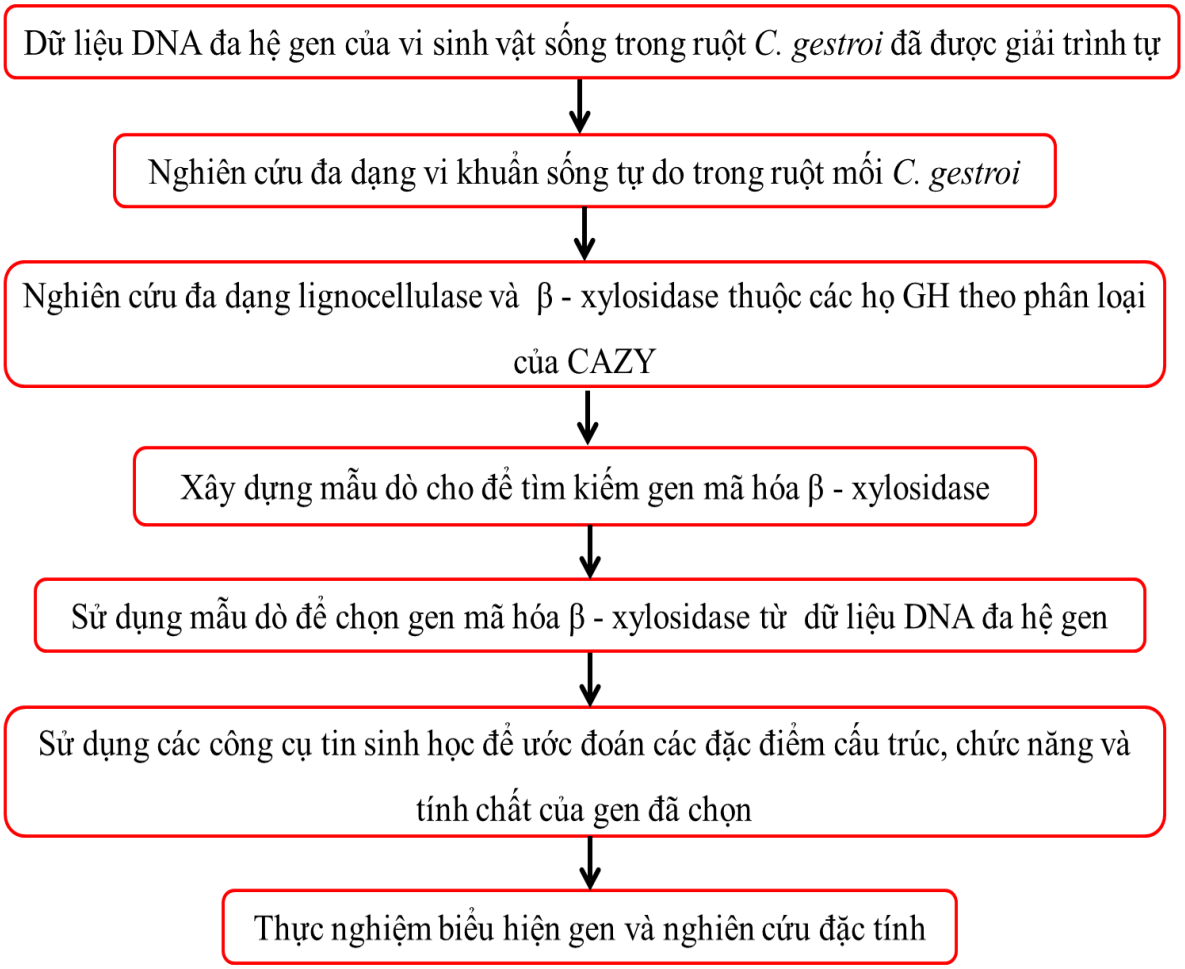
Dung dịch sử dụng trong điện di protein gồm đệm xử lý mẫu protein đặc 6 lần (7 ml Tris-HCl 1 M, pH = 6,8; 3 ml glycerol 100%; 1 g SDS; 0,6 ml 2-mercapto-

ethanol; 1,2 mg bromophenol). Đệm chạy điện di protein (0,05 M Tris; 0,192 M glycine; 0,1% SDS; pH = 8,4). Dung dịch nhuộm gel polyacrylamide kiểm tra protein (coomassie brilliant blue 0,1% w/v; methanol 30% v/v; acetic acid 10% v/v). Dung dịch tẩy nhuộm coomassie (methanol 40% v/v; acetic acid 10% v/v).

Dung dịch tinh chế protein gồm: muối $(NH_4)_2SO_4$, dung dịch PBS 10X (80 g NaCl; 2 g KCl; 2,4 g KH_2PO_4; 14,2 g Na_2HPO_4 bổ sung nước cho đến 1 lít) và một số dung môi hữu cơ khác, đệm phosphate (pH = 9).

2.2. PHƯƠNG PHÁP NGHIÊN CỨU

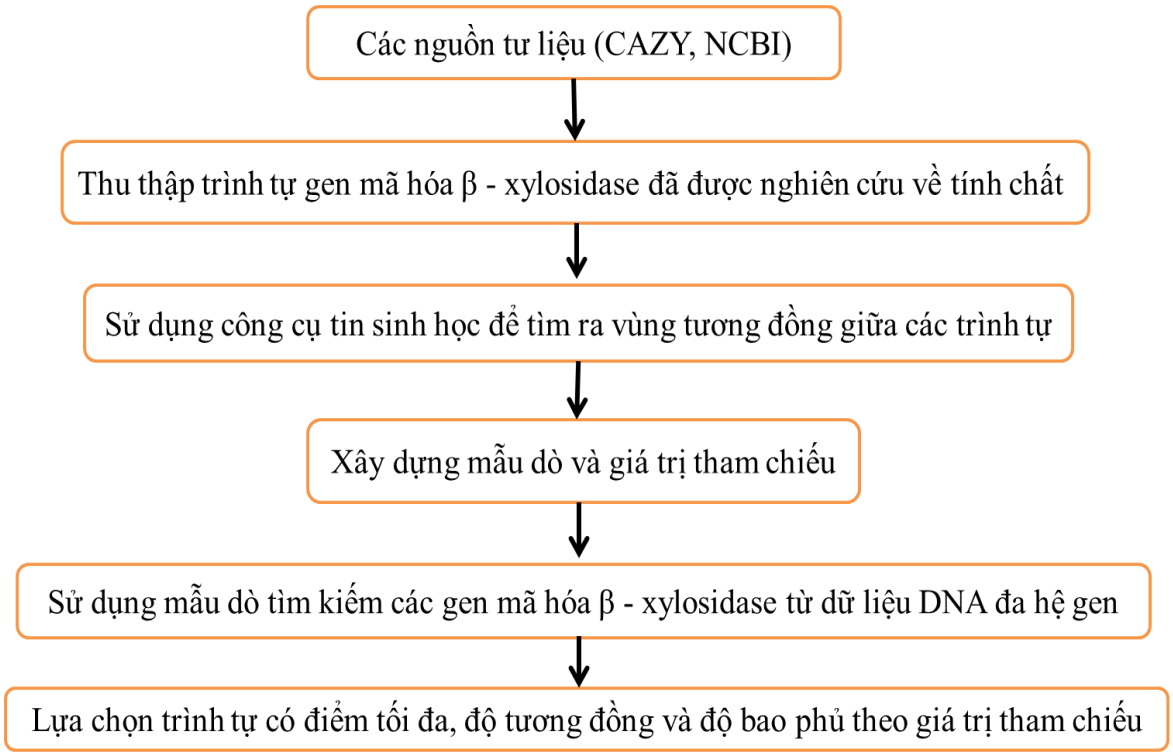
Đề tài được thực hiện theo các bước nghiên cứu cơ bản như sơ đồ hình 2.1:



Hình 2.1. Các bước cơ bản trong nghiên cứu của luận án

2.2.1. Phương pháp xây dựng mẫu dò

Để xây dựng mẫu dò cho β–xylosidase nghiên cứu thực hiện theo sơ đồ hình 2.2.



Hình 2.2. Sơ đồ các bước cơ bản trong xây dựng mẫu dò

2.2.1.1. Xác định các họ GH chứa β–xylosidase theo CAZY

Từ hệ thống phân loại của CAZY, nghiên cứu sẽ thống kê số lượng các họ GH chứa β–xylosidase. Sau đó sẽ so sánh sự khác nhau về cấu trúc không gian, axit amin cho và nhận proton trong quá trình hoạt động của enzyme thuộc các họ GH. Kết quả sẽ tổng hợp thành bảng phân loại để tiếp tục tìm hiểu thông tin cụ thể.

2.2.1.2. Tìm kiếm các trình tự axit amin của β–xylosidase đã được nghiên cứu đặc tính

Thu thập các trình tự axit amin của enzyme đã được nghiên cứu thực nghiệm thuộc mỗi họ GH để tìm ra sự tương đồng, từ đó xây dựng thành mẫu dò. Trong nghiên cứu này số liệu đã công bố về đặc điểm liên quan đến β–xylosidase như pH và nhiệt độ hoạt động tối ưu, cấu trúc không gian, trung tâm hoạt động của enzyme, trình tự gen/axit amin sẽ được tổng hợp lại thành bảng. Chỉ những nghiên cứu có thông tin chi tiết về đặc tính của enzyme mới được sử dụng cho nghiên cứu tiếp theo.

2.2.1.3. Nhóm các trình tự đã tìm kiếm được để tìm các vùng tương đồng bằng công cụ ClustalW - PBIL

ClustalW - PBIL là công cụ cho phép so sánh nhiều trình tự nucleotide hoặc axit amin [196]. Các trình tự axit amin của β-xylosidase sau khi được thu thập từ bước tìm kiếm sẽ được so sánh và tính toán mức độ tương đồng để cho ra một trình tự được cho là bảo tồn cao nhất (Ký hiệu *Prim.cons*) bằng công cụ ClustalW - PBIL. Kết quả so sánh bằng công cụ này chỉ ra vị trí axit amin giống nhau hoàn toàn hoặc một phần giữa các trình tự để người dùng dễ dàng quan sát và phân tích kết quả. Chúng tôi chọn tham số mặc định khi sử dụng công cụ ClustalW-PBIL trong nghiên cứu.

2.2.1.4. Xây dựng mẫu dò và xác định giá trị tham chiếu

Kết quả so sánh bằng công cụ ClustalW - PBIL sẽ xác định được vị trí bảo tồn cao và vị trí giống nhau giữa các trình tự. Trình tự mẫu dò bao gồm thành phần axit amin ở: (1) vị trí bảo tồn và giống nhau hoặc tương đối giống nhau; (2) vị trí không bảo tồn nhưng vẫn giống nhau trên 50% tổng số trình tự; (3) vị trí là số; (4) vị trí có từ trên hai loại axit amin giữa các trình tự bằng nhau.

Sau khi có trình tự axit amin của mẫu dò sẽ sử dụng để so sánh lại với chính các trình tự đã thu thập làm mẫu dò bằng BLASTP. Từ kết quả so sánh này sẽ xác định mức độ bao phủ và tương đồng của mẫu dò với chính các trình tự ban đầu sử dụng để thiết kế nó. Trên cơ sở đó xác định được giá trị tham chiếu về độ bao phủ, độ tương đồng và điểm tối đa khi sử dụng mẫu dò để lựa chọn gen mã hóa cho β–xylosidase từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi*.

2.2.1.5. Ước đoán cấu trúc bậc ba của trình tự mã hóa β–xylosidase khai thác

Cấu trúc bậc ba của phân tử được ước đoán dựa trên trình tự axit amin và so sánh với các protein đã được nghiên cứu về cấu trúc bằng công cụ Swiss model. Trong nghiên cứu này các trình tự thu thập để xây dựng mẫu dò sẽ được ước đoán cấu trúc không gian bằng dụng công cụ Swiss model, với tham số mặc định.

2.2.1.6 Tìm kiếm và lựa chọn gen mã hóa cho β–xylosidase từ số liệu DNA đa hệ gen của vi sinh vật trong ruột mối C. gestroi

Sau khi xây dựng được mẫu dò mã hóa cho β–xylosidase thuộc các họ GH, sẽ tiếp tục sử dụng công cụ BLASTP để so sánh mức độ tương đồng và bao phủ của mẫu dò, với trình tự axit amin từ dữ liệu giải trình tự từ vi sinh vật trong ruột mối *C. gestroi*. Trình tự được lựa chọn phải có giá trị tương đồng, độ bao phủ và điểm tối đa lớn hơn hoặc bằng giá trị tham chiếu.

2.2.2. Các phương pháp xử lý số liệu bằng phần mềm tin sinh học

2.2.2.1. Phân tích đa dạng các loài vi sinh vật điển hình trong ruột mối C. gestroi

Công cụ Excel của MS Word được sử dụng để lọc ORF của vi khuẩn, họ GH và lignocellulase từ 125.431 ORF từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật sống trong ruột mối *C. gestroi* đã có sẵn. Những loài vi sinh vật có số lượng bằng hoặc lớn hơn 300 ORF sẽ tiếp tục được tập trung nghiên cứu sâu hơn. Với mỗi loài sẽ tìm kiếm các thông tin liên quan đã được công bố, từ đó phân tích vai trò của chúng trong chuyển hóa lignocellulose.

2.2.2.2. Phân tích đa dạng ORF của lignocellulase theo phân loại của CAZY

Từ 587 ORF đã được ước đoán ban đầu, tiếp tục so sánh với CSDL của CAZY bằng chương trình chạy của CAZY trực tuyến theo tham số mặc định. Kết quả sẽ đưa ra bảng sắp xếp ORF vào các họ GH khác nhau. Sau khi có dữ liệu nhận được từ phân tích của CAZY, sẽ tiếp tục phân tích và so sánh với nghiên cứu tương tự ở các loài mối khác, để đánh giá sự đa dạng của lignocellulase.

2.2.2.3. So sánh trình tự ORF của dữ liệu trình tự DNA đa hệ gen có sẵn với CSDL của NCBI bằng công cụ BLAST

Với mục đích khai thác gen mã hóa enzyme mong muốn từ các ORF của dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối, BLAST trực tuyến của NCBI sẽ được sử dụng. Phần mềm này sẽ giúp so sánh mức độ tương đồng của ORF quan

tâm với trình tự trên ngân hàng gen. Quá trình thực hiện được tiến hành theo đường dẫn http://blast.ncbi.nlm.nih.gov/ [205].

Trong nghiên cứu này chúng tôi sử dụng công cụ BLASTP để so sánh sự tương đồng của trình tự axit amin với ngân hàng gen, theo tham số mặc định. Để chạy BLASTP, trình tự axit amin của β-xylosidase được định dạng FASTA, sẽ được nhập vào khung chính của công cụ. Sau đó chọn tham số mặc định của chương trình và nhấn lệnh BLAST, chờ một vài phút phần mềm sẽ trả lại kết quả. Thông thường kết quả có hai phần bao gồm: (1) phần hình chỉ ra vùng bảo tồn và hoạt tính của enzyme, (2) sơ đồ so sánh trình tự đích với các trình tự trong ngân hàng gen với chỉ số chi tiết về điểm tối đa, mức độ bao phủ, độ tương đồng và vật chủ chứa gen so sánh.

2.2.2.4. Thiết kế mồi bằng phần mềm FastPCR

Mồi đặc hiệu khuếch đại gen mã hóa *Xbx14* được thiết kế dựa vào trình tự đầu 5’ và 3’ của gen mong muốn có trong dữ liệu trình tự DNA đa hệ gen có sẵn. Sau đó, trình tự mồi được kiểm tra chất lượng và chỉnh sửa bằng phần mềm FastPCR, dựa trên nhiều thông số như nhiệt độ bắt cặp mồi, số nucleotide loại G và C,... theo tham số mặc định [203].

2.2.2.5. Kiểm tra vị trí của enzyme hạn chế trên gen quan tâm bằng phần mềm trực tuyến RestrictionMapper

Trình tự DNA của gen quan tâm được vào chương trình *RestrictionMapper* theo đường dẫn http://www.restrictionmapper.org/ [204]. Dựa trên việc đếm vị trí nucleotide trùng khớp với vị trí cắt của từng loại enzyme hạn chế, chương trình sẽ đưa ra kết quả số lượng và vị trí cắt của các enzyme hạn chế có trên.

2.2.2.6. Chuyển mã trình tự DNA sang trình tự axit amin bằng chương trình dịch mã EXPASY

Trình tự nucleotide của gen quan tâm được đưa vào chương trình EXPASY theo đường dẫn http://web.EXPASY.org/tools/translate/ [196]. Để chạy công cụ ”Translate tool” của EXPASY, trình tự nucleotide mã hóa β-xylosidase được định

dạng FASTA, sau đó nhập vào khung chính của công cụ. Tiếp theo sẽ chọn tham số mặc định của chương trình và nhấn lệnh ”Translate sequence”, chờ một vài phút. Chương trình sẽ đưa ra kết quả là trình tự axit amin tương ứng với 6 khung đọc khác nhau từ trình tự nucleotide ban đầu.

2.2.2.7. Dự đoán cấu trúc của protein bằng công cụ Phyre2 và Swiss model

Trình tự axit amin protein được xử lý trực tuyến bằng chương trình Phyre 2 theo đường dẫn http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index [197] và Swiss model theo đường dẫn https://swissmodel.expasy.org/interactive [198]. Bằng cách so sánh mức độ tương đồng của trình tự axit amin nghiên cứu với các trình tự có sẵn trên CSDL và đưa ra mô hình cấu trúc 3D thích hợp nhất theo tham số mặc định. Công cụ Swiss model sử dụng để dự đoán cấu trúc của các trình tự gen mã hóa enzyme đang tìm kiếm bằng mẫu dò. Công cụ Phyre 2 sử dụng để dự đoán cấu trúc của gen *Xbx14* mã hóa enzyme sau khi đã lựa chọn, để nghiên cứu trong thực nghiệm.

2.2.2.8. Dự đoán nguồn gốc của gen bằng công cụ BLAST-Explorer

Trình tự axit amin của *Xbx14* được xử lý trực tuyến bằng BLAST-Explorer của hệ thống phân tích Phylogeny.fr để dựng cây phân loại [201]. Sau đó chọn tham số mặc định của chương trình chờ một vài phút phần mềm sẽ trả lại kết quả dựng cây phân loại cho trình tự axit amin đích.

2.2.2.9. Xác định độ sạch của protein Xbx14 sau khi tinh chế bằng phần mềm Quantity One

Việc xác định khối lượng protein đích dựa vào băng protein chuẩn trên gel điện di bằng phần mềm Quantity One [206], là bước trung gian để có thể đánh giá độ sạch tương đối của protein. Để đánh giá độ sạch của *Xbx14*, mẫu protein sau khi tinh sạch được điện di trên gel polyacrylamide 12,6% có SDS cùng với thang chuẩn protein. Sau đó, gel được nhuộm bằng thuốc nhuộm comassie, rửa sạch và đưa lên máy scan để quét. Chế độ quét được cài đặt sao cho chất lượng ảnh đạt được tốt nhất. Sau đó, ảnh quét được đưa về chế độ đen trắng và chuyển vào phần mềm

Quantity One version 4.6. Bằng phần mềm này, lượng protein có trên giếng chạy được nhận biết và quét để định lượng tương đối protein tổng số trên giếng. Tuy nhiên, diện tích quét trên giếng nên là những vùng xác định thấy có băng protein. Bên cạnh đó, *Xbx14* trong giếng cũng được quét và định lượng. Độ sạch của *Xbx14* chính là tỷ lệ của khối lượng *Xbx14* trên khối lượng của protein ở mỗi đường chạy.

2.2.3. Các phương pháp vi sinh

2.2.3.1. Giữ chủng E. coli

Tế bào được nuôi cấy trong môi trường LB (bổ sung chất kháng sinh Ampicilin nồng độ cuối cùng là 100 µg/ml) đến khi OD_{600} đạt khoảng 2,0. Dịch tế bào tiếp tục được bổ sung glycerol vô trùng (nồng độ cuối là 15%) và chia thành các thể tích nhỏ trong ống eppendorf để giữ ở -80°C trong thời gian dài.

2.2.3.2. Cấy ria tạo khuẩn lạc đơn

Que cấy vô trùng được cho tiếp xúc với dịch nuôi cấy vi khuẩn *E. coli* và cấy ria trên đĩa petri chứa môi trường thạch. Đĩa thạch được ủ ở 37°C qua đêm, sau đó giữ ở 4°C và sử dụng trong vòng 14 ngày.

2.2.3.3. Nuôi cấy vi khuẩn

Chọn một khuẩn lạc đơn đường kính khoảng 2 mm trên đĩa thạch và nuôi cấy vào môi trường LB lỏng (bổ sung chất kháng sinh Ampicilin, Chloramphenicol có nồng độ cuối cùng lần lượt là 100 µg/ml và 20 µg/ml), nuôi cấy lắc 200 vòng/phút ở 37°C qua đêm.

2.2.4. Các phương pháp sinh học phân tử

2.2.4.1. Giải trình tự DNA bằng máy giải trình tự thế hệ mới HiSeq2000 của Illumina

Các deoxynucleoside triphosphate (dNTP) được sử dụng để giải trình tự đều gắn với nhân tố kết thúc đảo ngược đã được đánh dấu huỳnh quang. Mỗi loại nucleotide được đánh dấu huỳnh quang có màu sắc khác nhau. Chu kì giải trình tự bắt đầu khi dNTP thứ nhất được gắn bổ sung với khuôn DNA. dNTP này sẽ được

tia laser quét để xác định đỉnh màu sắc huỳnh quang tạo ra, đồng thời dưới tác động của laser, nhân tố kết thúc đảo ngược của dNTP đó sẽ bị phân hủy. Lúc này, dNTP thứ hai mới có thể tiếp tục gắn vào nucleotide thứ nhất và lại được xác định. Quá trình giải trình tự cứ lặp đi lặp lại như vậy cho đến khi sợi DNA bổ sung được tổng hợp xong. Trình tự nucleotide được xác định dựa trên các đỉnh màu sắc. Máy HiSeq 2000 giải trình tự đồng thời được 10 triệu cụm DNA.

2.2.4.2. Phương pháp biến nạp DNA plasmid vào vi khuẩn E. coli bằng sốc nhiệt

Quy trình tạo tế bào khả biến: cấy chuyển 1 khuẩn lạc *E. coli* DH10B vào 5 ml môi trường LB lỏng, lắc 200 vòng/phút ở 37°C qua đêm. Chuyển 0,5 ml dịch nuôi cấy chuyển sang 50 ml môi trường LB lỏng, tiếp tục cấy lắc 200 vòng/phút ở 37°C cho tới khi OD_{600} đạt 0,4 đến 0,6. Sau đó, dịch tế bào được đặt lên đá khoảng 1 giờ. Từ bước này trở đi, mẫu luôn luôn được tiến hành ở 4°C. Ly tâm thu tế bào ở tốc độ 4.000 vòng/phút trong 10 phút. Tế bào được hòa vào 50 ml $CaCl_2$ 100 mM (vô trùng, đã được làm lạnh) và ủ 30 phút. Ly tâm thu tế bào ở 4.000 vòng/phút trong 10 phút. Tế bào tiếp tục được hòa vào 25 ml $CaCl_2$ 100 mM và ủ 10 phút. Ly tâm 4.000 vòng/phút trong 10 phút để thu tế bào. Cuối cùng, tế bào được hòa vào 0,5 ml $CaCl_2$ 100 mM có bổ sung thêm 15% glycerol để bảo quản ở -80°C trong các ống eppendorf.

Quy trình biến nạp: Tế bào khả biến được lấy ra từ tủ lạnh -80°C được làm tan từ từ trên đá trong khoảng 30 phút. Sau đó, bổ sung vào mỗi ống tế bào khả biến 1 µg DNA, đảo nhẹ nhàng để DNA phân bố đều trong dịch tế bào khả biến. Ủ mẫu ở 4°C trong khoảng 30 phút. Mẫu được sốc nhiệt ở 42°C trong 1 phút 30 giây và ổn định mẫu ở 4°C trong 2 phút. Bổ sung khoảng 500 µl đến 600 µl LB lỏng, lắc hỗn hợp tế bào ở 37°C trong 45 phút đến 1 giờ. Hút 100 µl dịch tế bào và cấy trải trên đĩa môi trường LB đặc có bổ sung Ampicilin 100 µg/ml (Chloramphenicol 20 µg/ml) và ủ đĩa ở 37°C qua đêm.

2.2.4.3. Phương pháp tách chiết DNA plasmid từ tế bào vi khuẩn E. coli

Phương pháp này được thực hiện theo Sambrook và cộng sự (2001) [144] như sau: Cấy chuyển 1 khuẩn lạc riêng rẽ vào ống nghiệm chứa 5 ml môi trường LB có bổ sung Ampicillin (nồng độ cuối cùng là 100 µg/ml), nuôi cấy lắc 200 vòng/phút qua đêm ở 37°C. Hút 2 ml dịch nuôi cấy vào ống eppendorf, ly tâm 5000 vòng/phút trong 10 phút và thu tế bào. Tế bào được hòa vào 100 µl dung dịch I, trộn đều và để ở nhiệt độ phòng trong thời gian 5 phút. Bổ sung 200 µl dung dịch II, đảo đều nhẹ nhàng để tránh đứt gãy DNA, ủ hỗn hợp ở 4°C trong thời gian 10 phút. Tiếp tục bổ sung tiếp 150 µl dung dịch III, đảo đều nhẹ nhàng và nhanh. Tủa protein xuất hiện và có màu trắng đục và ủ trong thời gian 15 phút ở 4°C. Thêm 450 µl phenol/chloroform/isoamylalcohol, lắc nhẹ để các protein nhỏ hòa tan vào dịch có phenol. Ly tâm 13.000 vòng/phút trong 15 phút. Lúc này mẫu hình thành 3 lớp: lớp trên cùng có chứa DNA plasmid, lớp giữa là protein lớn không có khả năng hòa tan và DNA hệ gen, lớp dưới cùng là phenol/chloroform có chứa protein tạp. Dùng pipet hút nhẹ nhàng dung dịch lớp trên và chuyển sang ống eppendorf mới loại 1,5 ml, sau đó bổ sung 2 lần thể tích ethanol 100% và ủ mẫu ở -20°C trong 20 phút để tủa DNA. Ly tâm 13.000 vòng/phút trong 10 phút và thu tủa. Tủa DNA plasmid được rửa lại bằng ethanol 70% và ly tâm 13.000 vòng/phút trong 10 phút. Plasmid DNA được làm khô bằng máy hút chân không và hòa vào 20 đến 30 µl RNAase 1% để phân hủy hết RNA lẫn vào trong thời gian 30 phút ở 37°C. Plasmid DNA được giữ ở -20°C.

2.2.4.4. Phương pháp cắt kiểm tra gen

Phản ứng cắt kiểm tra plasmid tái tổ hợp Pet22Xbx, bằng các enzyme hạn chế, được tiến hành trong tổng thể tích là 10 µl gồm 3 µl DNA (10 µg/ml); 0,25 µl enzyme hạn chế (2 U/µl); 5,7 µl nước cất vô trùng. Hỗn hợp được trộn đều, ủ ở 37°C trong khoảng thời gian 120 phút. Sản phẩm của phản ứng cắt được kiểm tra bằng cách điện di trên gel agarose 0,8%.

2.2.4.5. Điện di DNA trên gel agarose

Mẫu DNA được tra vào giếng của gel agarose và điện di bằng dòng điện một chiều với hiệu điện thế 100V trong khoảng 30 đến 45 phút. Bản gel được lấy ra khỏi khuôn và ngâm vào dung dịch Ethidium bromide nồng độ 0,5 µg/ml trong khoảng 5 đến 10 phút rồi rửa lại bằng nước. Cuối cùng DNA được quan sát dưới ánh sáng tử ngoại với bước sóng 302 nm [144].

2.2.4.6. Phương pháp đồng biểu hiện gen

Khuẩn lạc mang gen *Xbx14* và chaperone được cấy chuyển vào 5 ml môi trường LB có 100 µg/ml Ampicilin và 20 µg/ml Chloramphenicol, lắc ở 37°C qua đêm với tốc độ 200 vòng/phút. Tế bào nuôi cấy qua đêm được chuyển sang môi trường mới sao cho OD_{600} khoảng 0,1. Môi trường nuôi cấy được bổ sung L–arabinose 500 mM và Tetracylin 5 µM, tiếp tục lắc 200 vòng/phút ở 37°C cho đến khi OD_{600} đạt 1 đến 1,6 thì tiến hành cảm ứng sinh tổng hợp protein ngoại lai bằng cách bổ sung IPTG. Dịch tiếp tục được nuôi và thu mẫu sau thời gian 15 giờ cảm ứng, ở nhiệt độ thích hợp. Các điều kiện ảnh hưởng đến sinh tổng hợp enzyme như nồng độ IPTG, nhiệt độ,... trong quá trình nuôi cấy chủng trong môi trường có chất cảm ứng đã được khảo sát. Tế bào sau khi biểu hiện được thu lại bằng ly tâm 6.000 vòng/phút trong 10 phút và thu tủa. Tủa tế bào được hòa lại bằng nước cất vô trùng đưa về cùng giá trị OD_{600} = 10 để thuận tiện cho việc so sánh kết quả. Protein được phân tích bằng phương pháp điện di biến tính trên gel polyacrylamide 12,6% có SDS [144].

Để đánh giá khả năng tan của protein tạo thành, tế bào biểu hiện đã hòa vào trong nước cất vô trùng (OD_{600} = 10), được phá vỡ bằng siêu âm với chu kỳ 10 giây siêu âm và 20 giây nghỉ trong tổng khoảng 30 đến 40 xung ở tần số 20 khz. Sau đó, dịch siêu âm được ly tâm 13.000 vòng/phút trong 10 phút để tách protein pha tan và pha không tan. Pha tan là phần dịch nổi được thu lại, còn phần cặn là pha không tan được hòa trở lại trong nước cất vô trùng với thể tích bằng thể tích phần dịch vừa hút ra và để điện di kiểm tra.

2.2.5. Các phương pháp hóa sinh protein

2.2.5.1. Điện di biến tính protein trên gel polyacrylamide - SDS

Phương pháp này được thực hiện theo Sambrook và cộng sự (2001) [144]. *Chuẩn bị gel*: gel polyacrylamide được chuẩn bị với thành phần và nồng độ thể hiện như Bảng 2.1.

Quy trình điện di: Mẫu được tra vào giếng và chạy điện di với cường độ dòng điện 10 mA cho mỗi bản gel cho đến khi mẫu qua hết lớp gel có nồng độ polyacrylamide thấp (5%). Sau đó, cường độ dòng điện được tăng lên 20 mA cho mỗi bản gel khi mẫu đến lớp gel có nồng độ polyacrylamide cao hơn (12,6%). Kết thúc điện di, gel được nhuộm Coomassie Brilliant Blue R250. Sau khi ủ với thuốc nhuộm, gel được rửa sạch bằng dung dịch tẩy nhuộm cho đến khi quan sát được rõ ràng các băng protein.

Bảng 2.1. Thành phần và nồng độ của gel polyacrylamide

Thành phần	Gel tách biến tính 12,5%	Gel cô biến tính 5%
dH_2O	0,55ml	0,45 ml
Tris-HCl pH 6,8	-	0,2 ml
Tris-HCl pH 8,8	1,125 ml	-
Glycerol 50%	0,9 ml	-
Acrylamid 30%	1,89 ml	0,14 ml
SDS 10%	45 µl	4 µl
APS 10%	30 µl	8 µl
TEMED	3 µl	1 µl
Tổng	4,543 ml	0,803 ml

2.2.5.2. Phương pháp tinh chế protein bằng tủa phân đoạn

Quy trình tinh chế: lượng protein tổng số biểu hiện trong *E. coli Rosetta 1* dưới dạng tan được thu lại và hòa vào trong nước. Sau đó bổ sung muối $(NH_4)_2SO_4$ để tủa theo phân đoạn từ nồng độ muối cao đến nồng độ muối thấp. Các phân đoạn được chia nhỏ dần để tìm ra nồng độ muối thấp nhất làm kết tủa hết protein mong

muốn. Các phân đoạn sau khi tủa được hòa lại vào trong nước và điện di kiểm tra trên gel polyacrylamide 12,6% có SDS. Phân đoạn sạch sẽ được giữ lại để kiểm tra hoạt tính và xác định thông số động học của enzyme.

2.2.5.3. Định lượng axit nucleic/protein bằng phương pháp đo mật độ quang

Để định lượng axit nucleic/protein, chúng tôi sử dụng phương pháp đo mật độ quang bằng máy đo NanoDrop. Máy hoạt động bằng cách kết hợp công nghệ sợi quang, tính năng căng thẳng bề mặt tự nhiên để giữ và đo mẫu giữa hai bệ quang mà không cần sử dụng cuvette hoặc đường ống. Hệ thống sử dụng các đường dẫn ngắn cho phép đo nồng độ axit nucleic, protein một cách chính xác với nhiều lựa chọn khác nhau. Khối lượng mẫu cần thiết cho phân tích quang phổ của máy rất ít nhưng vẫn thu được kết quả rất chính xác. Mẫu được đặt trực tiếp trên đầu của bề mặt phát hiện và một cột nước tạo thành một đường thẳng đứng giữa các đầu của sợi quang do sức căng thẳng bề mặt. Sau đó đèn flash xenon cung cấp nguồn ánh sáng và sợi quang phổ tuyến tính được sử dụng để phân tích ánh sáng đi qua mẫu [44]. Để xác định hàm lượng protein/axit nucleic, chỉ cần đưa từ 1 µl đến 2 µl mẫu lên thiết bị, chờ máy phân tích và trả kết quả phân tích sau từ 1s đến 2s. Mẫu sau khi đo xong được làm sạch một cách nhẹ nhàng bằng giấy làm sạch bề mặt quang học.

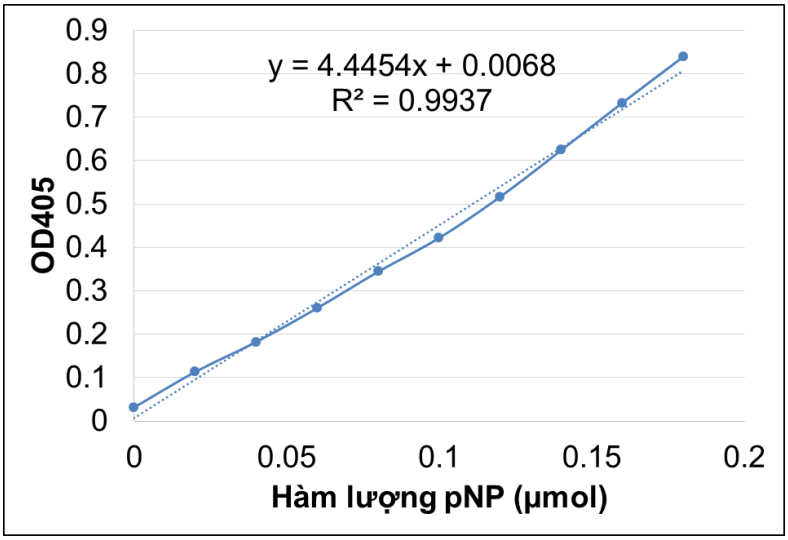
2.2.5.4. Phương pháp xác định hoạt tính β–xylosidase

Hoạt tính β–xylosidase được xác định dựa vào cơ chất được sử dụng phổ là 4-nitrophenyl β-D-xylopyranoside (pNPX). β–xylosidase thủy phân pNPX, giải phóng *p*-nitrophenol, nên hoạt tính của β–xylosidase được xác định bằng cách đo lượng *p*-nitrophenol được giải phóng ra ở bước sóng 405 nm. Một đơn vị hoạt tính β–xylosidase (IU) là lượng enzyme cần thiết để giải phóng 1 µmol p-nitrophenol trong 1 phút [40].

Thiết lập đường chuẩn nồng độ p-nitrophenol: Đồ thị chuẩn được thiết lập với thang chuẩn 10 nồng độ đường *p*-nitrophenol từ 0 mM đến 10 mM. *P*-nitrophenol được pha nồng độ 0,2 µmol/ml. Mỗi ống phản ứng có tổng thể tích là 1000 µl gồm *p*-nitrophenol và đệm phosphate, trong đó, thể tích *p*-nitrophenol giảm dần từ 1000

đến 0 µl, còn thể tích đệm phosphate tăng dần từ 0 đến 1000 µl. Tất cả các mẫu được đo OD ở bước sóng 405 nm bằng máy quang phổ UV 165^{0}C. Kết quả đo OD_{405} được sử dụng để thiết lập đường chuẩn biểu thị mối quan hệ giữa OD_{405} và nồng độ *p*-nitrophenol (Hình 2.3).

Hoạt tính *Xbx14* được xác định dựa theo phương pháp của Teng và cộng sự, 2011 [168]. Hỗn hợp phản ứng 50 µl gồm có 20 µl cơ chất pNPX (p-nitrophenyl – β- xylopyranoside) 10 mM, hòa trong đệm phosphate (pH = 7,0) được ủ với 30 µl dịch enzyme *Xbx14* ở 50°C trong 30 phút. Phản ứng được dừng lại khi bổ sung 200 µl Na_2CO_3 0,1M. Hoạt tính của enzyme được xác định thông qua việc đo độ hấp thụ pNP tạo ra ở bước sóng 405 nm. Lượng *p*-nitrophenol tạo ra được tính dựa vào đường chuẩn thể hiện mối quan hệ giữa OD_{405} và nồng độ *p*-nitrophenol.



Hình 2.3. Đường chuẩn pNP được đo ở bước sóng 405 nm

2.2.5.5. Xác định ảnh hưởng của nhiệt độ, pH và một số hóa chất

Phản ứng xác định ảnh hưởng của nhiệt độ và pH dựa theo phương pháp của Teng và cộng sự [168]. Ảnh hưởng của nhiệt độ trong thí nghiệm được tiến hành ở 20°C, 30°C, 40°C, 50°C, 60°C, 70°C, 80°C, và 90°C. Ảnh hưởng của pH đối với hoạt tính *Xbx14* được thay đổi theo đệm phosphate bao gồm: pH = 5, pH = 6, pH = 7, pH = 8, pH = 9, pH = 10, pH = 11 và pH = 12. Ảnh hưởng của một số hóa chất đến hoạt tính của enzyme được xác định theo phương pháp của Lee và cộng sự [100].

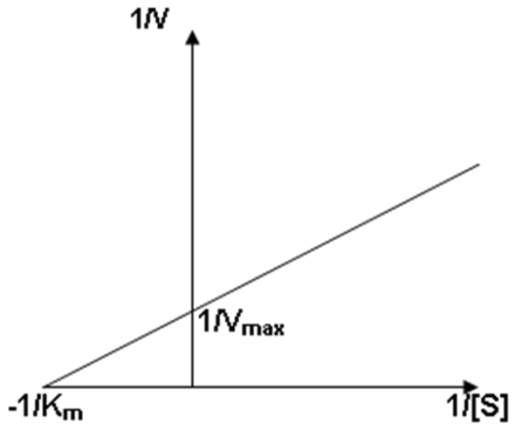
2.2.5.6. Xác định độ bền của enzyme

Xbx14 được ủ ở các nhiệt độ khác nhau: 20°C, 30°C, 40°C, 50°C, 60°C, 70°C, 80°C, 90°C trong khoảng thời gian là một giờ và hai giờ theo phương pháp của Pedersen và cộng sự (2007). Sau khi kết thúc thời gian ủ, mẫu được trộn trực tiếp với đệm và cơ chất. Phản ứng enzyme được thực hiện như phương pháp thử hoạt tính đã đề cập ở trên.

2.2.5.7. Xác định thông số động học của enzyme

Hoạt tính tính β–xylosidase được xác định ở các nồng độ cơ chất khác nhau: hỗn hợp phản ứng bao gồm 30 µl enzyme + 20 µl cơ chất pNPX pha trong đệm phosphat 1X (nồng độ từ 0 đến 20 mg/ml) được ủ ở 60°C trong 30 phút. Sau đó, bổ sung 200 µl Na_2CO_3 0,1M. Mẫu được đo OD ở bước sóng 405 nm. Lượng đường khử do *Xbx14* tạo ra trong các phản ứng được tính tương đương với lượng pNP theo đường chuẩn.

Để xác định giá trị K_m và V_{max}, đồ thị phụ thuộc của tốc độ phản ứng vào nồng độ cơ chất được thiết lập theo phương pháp của Linewever-Burk (Hình 2.4), trong đó V được coi là số micromole xylose (µM) được giải phóng ra trong 1 phút, 1/[S] được tính là 1/[pNPX], ở đó nồng độ cơ chất được tính là mM.



***Hình 2.4. Sự phụ thuộc của tốc độ phản ứng vào nồng độ cơ chất theo Lineweaver-Burk* [135]**

Chương 3. KẾT QUẢ VÀ THẢO LUẬN

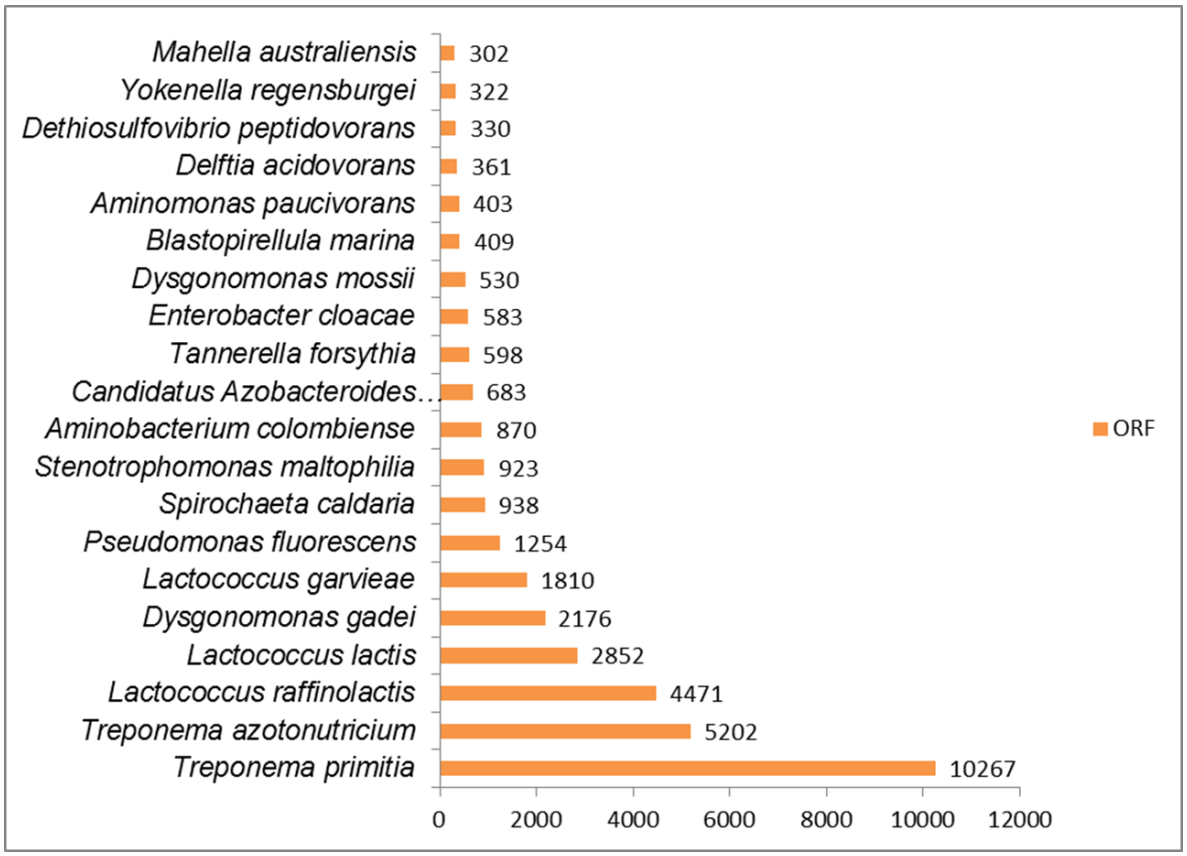
3.1. NGHIÊN CỨU ĐA DẠNG HỆ VI KHUẨN VÀ LIGNOCELLULASE THEO HỆ THỐNG PHÂN LOẠI CỦA CAZY TỪ DỮ LIỆU DNA ĐA HỆ GEN CỦA VI SINH VẬT TRONG RUỘT MỐI *C. gestroi*

3.1.1. Nghiên cứu đa dạng hệ vi khuẩn sống trong ruột mối *C. gestroi*

3.1.1.1. Thành phần các loài vi khuẩn có số lượng ORF lớn trong ruột mối C. gestroi

Trên thế giới đã có một số công bố về gen mã hóa lignocellulase của vi sinh vật ruột mối bậc thấp, nhưng hầu hết tập trung nghiên cứu sự đa dạng của động vật nguyên sinh và nguồn gen của chúng. Cho đến hiện nay các nghiên cứu đánh giá sâu về sự đa dạng của vi khuẩn cũng như sự đa dạng gen lignocellulase trong ruột mối bậc thấp còn hạn chế. Trên cơ sở dữ liệu giải trình tự DNA đa hệ gen của vi sinh vật sống trong ruột mối *C. gestroi* ở Việt Nam, chúng tôi căn cứ theo thống kê ban đầu và tập trung vào phân tích vi sinh vật có số lượng ORF lớn nhất chính là vi khuẩn. Đây là nhóm vi sinh vật tiết ra lignocellulase quyết định việc thủy phân hoàn toàn lignocellulose trong thức ăn của cả mối bậc thấp và mối bậc cao [24]. Mặt khác các enzyme có nguồn gốc từ vi khuẩn sống trong điều kiện khắc nhiệt như pH kiềm của ruột mối thường rất ổn định lên sẽ làm tăng khả năng thủy phân lignocellulose [133].

Dựa trên số liệu DNA đa hệ gen được so sánh với một số CSDL có uy tín như KEGG và eggNOG để dự đoán về loài, chúng tôi tiến hành lọc và thống kê được 20 loài có số lượng ORF lớn hơn 300 ORF. Tất cả các loài có số lượng ORF lớn đều thuộc về vi khuẩn (Hình 3.1). Vai trò của các loài vi khuẩn này sẽ được nghiên cứu chi tiết để tìm hiểu về mối quan hệ giữa chế độ ăn gỗ của vật chủ với hệ vi khuẩn cộng sinh trong ruột mối *C. gestroi* ở Việt Nam.



Hình 3.1. Biểu đồ thống kê 20 loài vi khuẩn có ORF lớn nhất trong ruột C. gestroi

3.1.1.2. Đặc điểm các loài vi khuẩn giúp C. gestroi chuyển hóa Nitơ

Ở Việt Nam mối *C. gestroi* là loài mối ăn gỗ, nên thức ăn rất nghèo Nitơ, vì vậy mối cần sự hỗ trợ của vi sinh vật sống trong ruột mối để bù đắp phần thiếu hụt chất dinh dưỡng thiết yếu chính là các axit amin. Khi phân tích vai trò của các loài vi khuẩn có số lượng ORF lớn cho thấy: Tham gia vào quá trình chuyển hóa Nitơ gồm có sáu loài là *Treponema primitia, Treponema azotaurricium, Aminobacterium colombiense, Aminomonas paucivorans, Candidatus Azobacteroides pseudotrichonymphae, Stenotrophomonas maltophilia*. Loài chiếm số lượng lớn nhất là *Treponema primitia* (10267 ORF) và *Treponema azotonutricium* (5202 ORF. Cho đến nay chúng tôi chưa tìm thấy công bố nào liên quan đến ba loài *Aminobacterium colombiense, Aminomonas paucivorans* và *Stenotrophomonas maltophilia* trong ruột của các loài mối khác mà chỉ có ở trong đường tiêu hóa của

C. gestroi. Theo nghiên cứu của Cleveland thì sự thiếu hụt nitơ nghiêm trọng trong chế độ ăn của mối có thể được bù đắp nhờ các vi khuẩn cố định nitơ trong ruột [32]. Sở dĩ mối phát triển mạnh trên gỗ do nhận được 30% - 50% từ con đường cố định nitơ nhờ vi khuẩn [43]. Như vậy chế độ ăn gỗ thiếu hụt Nitơ của *C. gestroi* được bù đắp lại bởi các loài vi khuẩn trên do chúng sản sinh ra enzyme ngoại bào tham gia vào quá trình chuyển hóa axit amin và giúp mối hấp thụ nguồn protein ít ỏi trong thức ăn [51]. Đồng thời cho phép các loài mối trong họ *Coptotermes*, trong đó có *C. gestroi* chỉ sử dụng gỗ làm thức ăn duy nhất vẫn có khả năng thống trị toàn cầu. Trong thực tế mối *C. gestroi* chỉ ăn gỗ nhưng được tìm thấy ở khắp nơi ở Việt Nam.

3.1.1.3. Đặc điểm các loài vi khuẩn giúp C. gestroi chuyển hóa lignocellulose

Sự tiêu hóa lignocellulose của mối được thực hiện nhờ enzyme được tiết ra từ bản thân mối và từ vi sinh vật sống trong ruột mối. Tuy nhiên các nghiên cứu đã chứng minh enzyme do vi sinh vật sống trong ruột mối quyết định hầu hết quá trình chuyển hóa thức ăn. Ở ruột sau của mối bậc thấp vi sinh vật chuyển hóa lignocellulose gồm vi khuẩn, nấm và động vật nguyên sinh sản xuất các cellulase (exoglucanase, endoglucanase, β-glucosidase) và hemicellulase (như xylanase, arabinosidase, mannosidase, arabinofuranosidase) có hoạt tính mạnh thủy phân hiệu quả cellulose và hemicellulose. Kết quả thống kê số loài có ORF lớn cho thấy tham gia vào quá trình chuyển hóa lignocellulose trong ruột *C. gestroi* có 16 loài vi khuẩn bao gồm: *Treponema primitia, Treponema azotaurricium, Lactococcus raffinolactis, Lactococcus lactis, Lactococcus garvieae, Spirochaeta caldaria, Enterobacter cloacae, Mahella australiensis, Tannerella forsythia, Delftia acidovorans, Dethiosulfovibrio peptidovoran, Blastopirellula marina, Pseudomonas fluorescens, Yokenella regensburgei, Dysgonomonas gadei* và *D. mossii.*

Theo một số nghiên cứu thì các loài vi khuẩn trên đóng vai trò cung cấp enzyme thủy phân lignocellulose như: cellulase được khai thác từ vi khuẩn *Pseudomonas fluorescens* trong ruột mối *Macrotermes subhyalinus* [11]; cellulase, hemicellulase và chitinase từ vi khuẩn *Spirochaeta caldaria* [51], trong ruột mối *Neotermes castaneus* [48]. Ba loài *Lactococcus raffinolactis, Lactococcus lactis,*

Lactococcus garvieae có số lượng ORF rất lớn ở ruột sau của mối ăn gỗ khô, liên quan đến quá trình lên men axit lactic [190]. Sự hiện diện của loài *Tannerella forsythia* sản xuất axit sialic, làm trung hòa độc tố sản sinh ra từ các loài vi sinh vật khác trong ruột mối, giúp chống lại mầm bệnh có trong nguồn thức ăn của mối *C. gestroi*. Đặc biệt loài *Treponema primitia và Treponema azotaurricium* vừa tham gia chuyển hóa nitơ, vừa tham gia chuyển hóa lignocellulose.

Trong số 16 loài có số lượng ORF lớn tham gia chuyển hóa lignocellulose trên, thì 5 loài chỉ xuất hiện trong ruột của mối *C. gestroi* ở Việt Nam bao gồm: *Mahella australiensis, Blastopirellula marina, Yokenella regensburgei (Koserella trabulsii), Delftia acidovorans, Dethiosulfovibrio peptidovorans.* Trong đó loài *Mahella australiensis* có khả năng lên men carbohydrate như cellobiose [159]. Loài *Blastopirellula marina* có khả năng sản xuất axit từ đường [58]. Loài *Delftia acidovorans* chứa enzyme thực hiện cả hai hoạt tính mono và diestease [88]. Loài *Dethiosulfovibrio peptidovorans* sản xuất các peptide và axit amin từ đường và axit béo [99]. *Yokenella regensburgei (Koserella trabulsii)* sản xuất cellulase trong quá trình phân hủy sinh khối thực vật [122]. Sự phong phú về số lượng ORF của 05 loài vi khuẩn tham gia phân giải lignocellulose trong ruột mối *C. gestroi* ở Việt Nam không có ở các loài mối khác có thể do sự khác biệt về nguồn thức ăn.

Tóm lại, dựa trên số lượng các ORF đã được ước đoán từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* ở Việt Nam, chúng tôi tìm thấy 20 loài vi khuẩn với số lượng ORF rất lớn, trong đó có 03 loài tham gia chuyển hóa Nitơ và 05 loài tham gia thủy phân lignocellulose không có ở các loài mối khác, giúp *C. gestroi* có thể tồn tại và phát triển trong điều kiện nguồn thức ăn giàu lignocellulose và thiếu hụt Nitơ. Như vậy kết quả nghiên cứu đa dạng các loài vi khuẩn có số ORF lớn trong ruột mối *C. gestroi* đã bổ sung thêm kết quả về hệ vi sinh vật sản xuất lignocellulase không chỉ có động vật nguyên sinh mà vi khuẩn cũng đóng góp phần lớn enzyme cần thiết cho mối tiêu hóa hiệu quả thức ăn.

3.1.2. Nghiên cứu đa dạng lignocellulase của vi sinh vật sống trong ruột mối *C. gestroi*

3.1.2.1. Đa dạng ORF mã hóa lignocellulase theo phân loại của CAZY

Theo ước đoán ban đầu của công ty giải trình tự BGI có 587 ORF mã hóa enzyme thủy phân lignocellulose từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi*. Để đánh giá sự đa dạng của các lignocellulase này, chúng tôi tiến hành sắp xếp chúng vào hệ thống GH theo phân loại của CAZY. Dựa vào kết quả phân tích của CAZY, sẽ dễ dàng so sánh với kết quả nghiên cứu tương tự từ các loài mối khác. Kết quả chi tiết được trình bày trong Bảng 3.1.

Bảng 3.1. Bảng so sánh các họ enzyme thủy phân lignocellulose (GH) của vi sinh vật trong ruột mối C. gestroi với một số loài mối khác

GH	*C. gestroi* (vi sinh vật)[1]	*Nasutitermes* (vi sinh vật)[2]	*R. speratus* (vi sinh vật)[3]	*R. flavipes* (vi sinh vật)[4]	*R. flavipes* (sinh vật nội sinh)[5]
GH1	187	22	–	–	3
GH2	36	23	–	1	1
GH3	205	69	1	10	2
GH4	116	14	–	–	–
GH5	136	56	3	11	2
GH6	21	–	–	–	–
GH7	20	–	.	35	10
GH8	13	5	9	1	1
GH9	16	9	–	–	2
GH10	39	46	.	1	1
GH11	52	14	.	10	5
GH12	7	–	–	–	–
GH13	10	48	–	–	10
GH16	3	1	–	3	3
GH17	2	–	–	–	–
GH18	1	17	–	5	8
GH20	–	15	–	3	4
GH23	–	52	–	–	–
GH25	–	1	–	–	–
GH26	6	15	.	5	1

GH27	24	4	–	–	1
GH28	2	6	–	–	–
GH30	2	–	–	1	3
GH31	6	26	–	–	–
GH32	16	–	–	–	–
GH35	2	3	–	–	–
GH36	44	5	–	–	–
GH37	1	–	–	–	1
GH38	–	11	–	–	2
GH39	50	3	–	–	–
GH42	30	24	–	1	–
GH43	60	16	.	–	1
GH44	6	6	–	–	–
GH45	–	4	2	4	1
GH47	–	–	–	1	–
GH51	50	18	–	–	–
GH52	–	3	–	–	–
GH53	3	12	–	1	1
GH55	1	–	–	–	–
GH57	–	17	–	–	–
GH58	–	1	–	–	–
GH62	1	–	.	–	–
GH65	–	6	–	–	–
GH67	20	10	–	–	–
GH68	1	–	–	–	–
GH70	–	–	–	–	1
GH74	–	7	–	–	–
GH76	–	–	–	–	1
GH77	1	14	–	1	–
GH78	1	–	–	–	–
GH79	18	–	–	–	–
GH82	1	–	–	–	–
GH85	–	–	–	–	1
GH86	2	–	–	–	–
GH87	4	–	–	–	–
GH88	–	9	–	–	–
GH91	–	1	–	–	–
GH92	–	2	–	1	–
GH93	1	–	–	–	–

GH94	4	68	–	–	–
GH95	–	12	–	–	–
GH97	7	–	–	–	–
GH98	–	1	–	–	–
GH103	–	3	–	–	–
GH106	–	2	–	–	–
GH109	–	3	–	–	–
GH113	3	–	–	–	–
GH115	3	–	–	–	–
GH116	1	–	–	–	–
GH117	40	–	–	–	–
GH119	2	–	–	–	–
GH126	3	–	–	–	–
GH128	3	–	–	–	–
GH130	10	–	–	–	–

[1]Do et al., 2014 (số liệu giải trình tự DNA đa hệ gen của vi sinh vật sống trong ruột mối *C. gestroi* bằng kỹ thuật Metagenomics giải trình tự toàn bộ đa hệ gen)

[2]Warnecke *et al*. [182] (số liệu giải trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *Nasutitermes* kỹ thuật Metagenomics giải trình tự toàn bộ đa hệ gen từ ngân hàng DNA đa hệ gen).

[3]Todaka *et al*. [171] (số liệu giải trình tự của vi sinh vật trong ruột mối *R. speratus* bằng phương pháp lập ngân hàng cDNA).

[4]Tartar *et al*. [165] (số liệu giải trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *R. flavipes* bằng phương pháp lập ngân hàng cDNA).

[5]Tartar *et al*. [165] (số liệu giải trình tự DNA đa hệ gen của vi sinh vật nội sinh trong ruột mối *R. flavipes* bằng phương pháp lập ngân hàng cDNA).

* GH sử dụng theo hệ thống phân loại của CAZY tại trang web: http://www.CAZY.org/

Dấu -: không có; dấu • không xác định số lượng gen

Kết quả phân tích của CAZY cho thấy 587 ORF từ vi sinh vật trong ruột mối *C. gestroi* được phân vào 52 họ GH khác nhau. So sánh với phương pháp lập thư viện cDNA, Todaka và cộng sự đã xác định 580 gen và dự đoán chỉ thuộc về 10 GH [171]. Warnecke và cộng sự đã sử dụng kỹ thuật Metagenomics, lập ngân hàng DNA đa hệ gen với khoảng 3 triệu trình tự đọc ngắn từ vi sinh vật trong ruột mối *Nasutitermes*. Kết quả đã xác định hơn 200 gen thủy phân cellulose và

hemicellulose nhưng cũng chỉ thuộc 45 họ GH khác nhau [146]. Như vậy so sánh với kết quả khi sử dụng kỹ thuật Metagenomics và giải trình tự toàn bộ DNA đa hệ gen cho thấy số họ GH của *C. gestroi* đa dạng hơn nhiều so với việc lập thư viện gen ở *R. speratus*, C. *formosanus, R. flavipes* hay thư viện DNA đa hệ gen của *Nasutitermes*.

Cũng từ kết quả này cho thấy trong số các họ GH dự đoán ở mối *C. gestroi* có thể có hoặc không có trong loài mối khác. Cụ thể có 22 GH (6, 12, 17, 32, 55, 62, 68, 78, 79, 82, 86, 87, 93, 97, 113, 115, 116, 117, 119, 126, 128 và 130) ở mối *C. gestroi* mà không có ở mối *Nasutitermes*; 24 GH (6, 7,12, 17, 30, 32, 37, 55, 62, 68, 78, 79, 82, 86, 87, 97, 113, 115, 116, 117, 119, 126, 128 và 130) của *C. gestroi* vắng mặt ở *Nasutitermes*. Tuy nhiên chỉ có 17 GH (20, 23, 28, 45, 52, 57, 58, 65, 74, 88, 91, 92, 98, 103, 106 và 109) có mặt trong *Nasutitermes* nhưng vắng mặt ở *C. gestroi*. Mặt khác số lượng ORF trong mỗi họ GH của *C. gestroi* hầu hết được dự đoán cao hơn ở *Nasutitermes* (Bảng 3.1).

Chúng tôi đặc biệt quan tâm đến các họ GH với số lượng ORF lớn chỉ xuất hiện trong ruột mối *C. gestroi* để xác định khả năng phân giải thành phần của lignocellulose, từ đó đưa ra hướng khai thác và chọn được gen mới trên nguồn dữ liệu DNA đa hệ gen đã có. Trong các họ GH theo phân loại của CAZY, họ GH117 tham gia vào thủy phân neoagarooligosaccharide do cắt liên kết alpha-1,3 tạo ra 3,6-anhydro-L-galactose và D-galactose; họ GH130 tham gia thủy phân mannooligosaccharide, mannosylglucose,.. là các chuỗi nhánh của hemicellulose. Hai họ enzyme này chỉ xuất hiện trong ruột mối *C. gestroi* với số lượng ORF khá lớn lần lượt là 40 (GH117) và 10 ORF (GH130). Điều này chứng tỏ, ruột mối *C. gestroi* có khả năng thủy phân mạch nhánh và dimer từ hemicellulose rất mạnh. Sự phong phú của hemicellulase từ dữ liệu DNA đa hệ gen của mối *C. gestroi* là cơ sở thuận lợi để chúng tôi tiếp tục nghiên cứu lựa chọn gen mã hóa nhóm enzyme này trong thực nghiệm.

3.1.2.2. Đa dạng ORF mã hóa β–xylosidase theo phân loại của CAZY

Từ số liệu phân tích về đa dạng enzyme chuyển hóa thành phần hemicellulose trong ruột mối *C. gestroi* như trên, chúng tôi quan tâm đến một trong những enzyme có ý nghĩa lớn trong quá trình thủy phân hemicellulose là β–xylosidase.

Hoạt động chính của β–xylosidase được tìm thấy trong ruột sau ở một số loài mối như *Pseudacanthotermes militaris* [15], *Neotermes castaneus* [48], *R. speratus* [157] và *R. santonensis* [164]. Trong ruột của mối *C. gestroi*, theo ước đoán có 46 ORF mã hóa β–xylosidase, trong đó 23 ORF được phân loại vào 13 loài, 23 ORF còn lại chưa được phân loại đến loài. Kiểm tra các ORF với hệ thống phân loại của CAZY cho thấy chúng được xếp vào 8 GH (GH1, 3, 5, 10, 39, 43, 51, 116) và chiếm ưu thế là GH43, 51, 116. Trong số các họ GH chứa β–xylosidase thì họ GH43 xuất hiện nhiều nhất ở 7 loài vi khuẩn và riêng loài *Treponema primitia* chứa ORF của cả 8 GH (Bảng 3.2). Bảng phân loại này giúp chúng tôi có cơ sở để xây dựng được phương pháp tìm kiếm gen mục tiêu và dễ dàng so sánh với kết quả chú giải gen từ công ty giải trình tự.

Bảng 3.2. Bảng thống kê số lượng ORF và họ GH của β–xylosidase trong ruột C. gestroi

Loài	Số ORF	GH
Bacteroides eggerthii	1	43
Clostridium hathewayi	1	51
Coprococcus eutactus	1	43
Dysgonomonas gadei	3	43, 51, 116
Enterobacter mori	1	43, 116
Lactococcus raffinolactis	2	43, 116
Leeuwenhoekiella blandensis	1	1
Mahella australiensis	1	3
Marvinbryantia formatexigens	2	1, 43, 116
Mitsuokella multacida	3	1, 51
Paenibacillus mucilaginosus	1	1
Treponema azotonutricium	1	51
Treponema primitia	5	1, 3, 5, 10, 39, 43, 51, 116

3.2. XÂY DỰNG PHƯƠNG PHÁP TÌM KIẾM GEN MÃ HÓA β–XYLOSIDASE TỪ DỮ LIỆU DNA ĐA HỆ GEN CỦA VI SINH VẬT TRONG RUỘT MỐI *C. gestroi*

Trong nghiên cứu này chúng tôi tiếp tục cải tiến và xây dựng một phương pháp mới để lựa chọn gen từ dữ liệu giải trình tự DNA đa hệ gen đã có, đó chính là phương pháp xây dựng mẫu dò. Phương pháp này dựa trên lý thuyết về các trình tự axit amin của một loại enzyme đều có vùng trình tự bảo thủ và gốc hoạt tính bảo tồn giống nhau. Đồng thời trình tự axit amin của cùng một loại enzyme chứa nhiều vị trí tương đồng. Như vậy từ các trình tự axit amin đã nghiên cứu về cùng một loại enzyme có thể tìm ra những vị trí bảo tồn và tương đồng nhau để xây dựng một mẫu dò chung. Khi đã có mẫu dò sẽ giúp lựa chọn nhanh trình tự gen từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* theo nguyên tắc so sánh tương đồng. Gen chọn phải mã hóa đúng enzyme đích và sau khi biểu hiện sẽ có hoạt tính tốt. Cụ thể các bước xây dựng mẫu dò sẽ được tiến hành để tìm kiếm gen mã hóa β–xylosidase từ DNA đa hệ gen của vi sinh vật ruột mối *C. gestroi* như sau:

3.2.1. Xác định các họ GH chứa β–xylosidase theo CAZY

Chúng tôi sử dụng dữ liệu phân loại của CAZY để xác định β–xylosidase thuộc về bao nhiêu họ GH. Trong mỗi họ GH này sẽ có thông tin về cấu trúc không gian, thành phần axit amin cho và nhận proton trong quá trình hoạt động của enzyme,... Đồng thời dựa vào sự bảo tồn cao về sự cuộn gấp trong cấu trúc không gian của protein, β–xylosidase sẽ được xếp vào từng nhóm lớn "clan". Kết quả chỉ ra enzyme này thuộc về bốn "clan" là GH-A, GH-D, GH-F, GH-O và được sắp xếp vào 11 họ GH1, 3, 30, 31, 39, 43, 51, 52, 54, 116, 120. Các thông tin chi tiết của từng họ được tổng hợp trong Bảng 3.3.

Theo kết quả này bốn họ GH1, GH30, GH39 và GH51 thuộc về "clan" GH-A giống nhau về cấu trúc không gian là $(\beta/\alpha)_8$. Họ GH31, GH41 thuộc về hai "clan" là GH-D, GH-F với mô hình cấu trúc không gian lần lượt là $(\beta/\alpha)_8$ barrel và 5-fold β-propeller. Họ GH52 và GH116 cùng thuộc về "clan" GH-O, nhưng vẫn chưa xác định được mô hình cấu trúc không gian. Các họ GH3, GH54 và GH120 chưa xác

định được cả "clan" và mô hình cấu trúc không gian. Tất cả các họ GH chứa β–xylosidase đều có chất cho điện tử và proton trong quá trình hoạt động của enzyme là glutamate (Glu) hoặc là Aspactate (Asp). Hai cơ chế phản ứng xúc tác thủy phân liên kết glycoside thường thấy nhất được Koshland đưa ra gồm có chế giữ nguyên và đảo ngược [93]. Kết quả tìm kiếm chỉ ra họ GH43 thuộc về cơ chế đảo ngược, còn lại đều thực hiện theo cơ chế giữ nguyên. Riêng GH54 mới chỉ xác định được cơ chế xúc tác mà chưa có thông số về clan, mô hình cấu trúc và chất cho/nhận điện tử.

Bảng 3.3. Bảng các họ GH chứa β–xylosidase theo CAZY

Mã E.C	**GH**	**Clan**	**Mô hình cấu trúc không gian**	**Chất cho điện tử xúc tác**	**Chất cho proton xúc tác**	**Cơ chế xúc tác**
3.2.1.37	1	GH-A	$(\beta/\alpha)_8$	Glu	Glu	Giữ nguyên
3.2.1.8	3			Asp	Glu	Giữ nguyên
3.2.1.8	30	GH-A	$(\beta/\alpha)_8$	Glu	Glu	Giữ nguyên
3.2.1.8	31	GH-D	$(\beta/\alpha)_8$ barrel	Asp	Asp	Giữ nguyên
3.2.1.8	39	GH-A	$(\beta/\alpha)_8$	Glu	Glu	Giữ nguyên
3.2.1.37	43	GH-F	5-fold β-propeller	Asp	Glu	Đảo ngược
3.2.1.37	51	GH-A	$(\beta/\alpha)_8$	Glu	Glu	Giữ nguyên
3.2.1.37	52	GH-O		Asp	Glu	Giữ nguyên
3.2.1.37	54					Giữ nguyên
3.2.1.37	116	GH-O		Asp	Glu	Giữ nguyên
3.2.1.37	120			Asp	Glu	Giữ nguyên

3.2.2. Tìm kiếm các trình tự axit amin của β–xylosidase đã được nghiên cứu trong thực nghiệm

Để xây dựng mẫu dò chúng tôi tiến hành tìm kiếm các trình tự axit amin hoặc gen mã hóa β–xylosidase đáp ứng được ba tiêu chí: (1) Trình tự gen mã hóa β–xylosidase đã được nghiên cứu thực nghiệm và phải có thông tin chi tiết về khả năng biểu hiện, nhiệt độ và pH hoạt động tối ưu, để đảm bảo trình tự này chắc chắn có hoạt tính đúng của β–xylosidase; (2) Chỉ chọn trình tự gen/axit amin từ vi khuẩn, để đảm bảo độ tương đồng và trùng khớp với vị trí axit amin bảo tồn; (3) Các trình tự không được khác nhau quá nhiều về độ dài, để sau khi xây dựng được mẫu dò sẽ tìm được giá trị tham chiếu tốt nhất về mức độ bao phủ và tương đồng của mẫu dò với trình tự đích cần tìm kiếm. Nguồn để tìm kiếm các trình tự gen/axit amin liên quan đến β–xylosidase rất đa dạng, nhưng chủ yếu từ CAZY và NCBI. Kết quả chi tiết được tổng hợp trong Bảng 3.4.

Bảng 3.4. Bảng tổng hợp dữ liệu đã được nghiên cứu chi tiết về β–xylosidase

Số thứ tự	Mã số trong GENBANK	Vi khuẩn	Số axit amin	pH hoạt động tối ưu	Nhiệt độ hoạt động tối ưu (^oC)	Nguồn thu thập số liệu
GH1						
1	CAD20872.1	*bacterium enrichment culture clone P11-6*	464	6	40	[116]
GH3						
1	CAD48309.1	*C. stercorarium*	715		50	[3]
GH30						
1	ABX45137.1.1	*Bifidobacterium breve*	448	6	45	[81]
GH43						
1	CAA29235.1.1	*Bacillus pumilus*	535	7	40	[186]
2	AAC97375.1	*Bacillus pumilus PLS*	535	6	45	[98]
3	AAC27699.1	*bacterium Bacillus sp. KK-1*	533		55	[191]

4	BAA02527.1	*Clostridium stercorarium*	473	7	65	[143]
5	BAC879411	*Clostridium stercorarium*	497	3,5	80	[162]
6	AFZ7887.1	*Enterobacter sp. enrichment culture clone nf1B6*	536	6	40	[27]
7	AAT9862.1	*Geobacillus stearothermophilus T-6 (XynB3)*	535	6,5	60	[148]
8	ABC750041	*Geobacillus thermoleovorans IT-08*	511	5	70	[175]
9	ADV16404.1	*Paenibacillus woosongensis*	477	6-7	30-45	[91]
10	AEF2882.1	*Thermobifida fusca TM51*	550	4,5	50	[59]
11	BAF982351	*Vibrio sp. XY-214*	535	7	36	[177]
GH52						
1	BAA74507.1	*Aeromonas caviae*	729	8,7	60	[163]
2	AGE344791	*Geobacillus stearothermophilus*	705	5,5	70	[79]
3	ABI49956.1	*Geobacillus stearothermophilus*	705	6,3	65	[22]
GH120						
1	ABM68042.1	*Thermoanaerobacterium saccharolyticum JW/SL-YS48*	636	6	65	[149]

Kết quả tìm kiếm được duy nhất 01 trình tự thuộc họ GH1, GH3, GH30 và GH120; 03 trình tự thuộc GH52; 11 trình tự thuộc GH43; các họ GH31, GH39, GH51, GH54, GH116 vẫn chưa tìm thấy công bố nào. Từ dữ liệu về số các trình tự mã hóa cho β–xylosidase đã được nghiên cứu tính chất sẽ tiếp tục so sánh để tìm ra vùng tương đồng.

3.2.3. Nhóm các trình tự đã tìm kiếm được để xác định vùng tương đồng bằng ClustalW – PBIL

```
Trình tự 1     --------------------KITNPVLKGFNPDPSICRAGEDYYMAVSTFEWFPGVQIY
Trình tự 2     --------------------KIINPVLKGFNPDPSICRVGEDYYMAVSTFEWFPGVQIY
Trình tự 3     --------------------KIINPVLKGFNPDPSICRVGEDYYIAVSTFEWFPGVQIH
Trình tự 7     -------------------SKIKNPILTGFHPDPSICRVGDDYYIAVSTFEWFPGVRIY
Trình tự 6     --------------------EITNPILTGFNPDPSLCRQGEDYYIATSTFEWFPGVRIY
Trình tự 11    ------------------TTTIQNPILKGFNPDPSIVRVGDDYYIATSTFEWFPGIQLH
Trình tự 10    ---TSPQVTSSPSREEPRAGTIRNPVLTGFYPDPSILRVGDDYYMATSTFEWYPGVTLH
Trình tự 8     --------------------EYSNPVIKGFYPDPSICRVGSDYYLVTSSFQYFPGVPIF
Trình tự 4     ----------------------RKQRFNPYLPSWEYIPDAEPYVFNGRVYIYGSHDRFN
Trình tự 9     ----------------------TKQGLNPYLPSWEYVPDGEPHVFNDRVYVYGSHDRFN
Trình tự 5     QPDNKHYKSAVRKWGDLGNGFYRNPVLNSDYSDPDVIRVGGDFYMVCSEFHYMGMPVLH
                                         :  :.   .. .    .  . :    : :     :
Prim.cons.    MQP2222222222222234MKI2NPVLKGFNPDPSICRVGEDYYIA2STFEWFPGV2I2
Trình tự 1     -HSKDLIHWRLAARPLQKTSQLDMKGNPDSGGV-WAPCLSYADGQFWLIYSDIKVVDG-
Trình tự 2     -HSKDLVHWRLAARPLQKTSQLDMKGNPDSGGV-WAPCLSYADGQFWLIYSDIKVVDG-
Trình tự 3     -HSKDLVNWRLIAHPLQRVSQLDMKGNPDSGGV-WAPCLSYSEGKFWLIYTDVKVVDG-
Trình tự 7     -HSKDLKNWRLVARPLNRLSQLNMIGNPDSGGV-WAPHLSYSDGKFWLIYTDVKVVEG-
Trình tự 6     -HSRDLKNWTLVSTPLDRVSMLDMKGNPDSGGI-WAPCLSYADGKFWLLYTDVKIVDS-
Trình tự 11    -HSRDLINWRLVGHALTRTSQLNMMGMDNSEGV-YAPALTYSDGTFWLCFSNVHSCRGG
Trình tự 10    -HSRDLVHWRPLGGALTETRLLDLAGRRDGAGV-WAPALSYRDGLFFLVFTNVASYSG-
Trình tự 8     -HSTNLINWNKIGYCLIRPSQLMLNNATNRSGI-FAPTLRYHEGIFYLITTNVTLKKN-
Trình tự 4     GHAFCLNDYVCWSAPVDDLSEWRYEG-VIYRK---TDDPLNPDGRMCLYAPDVTLGPD-
Trình tự 9     GHAFCLNDYVCWSAPVADLADWRYEG-VIYKK---TDDPLNPDGSMCLYAPDVTVGPD-
Trình tự 5     --SKDLVNWTIIGRVYDSLKHDPKYDNMEGYAKGSAPAIRYHNGRFYVYF-CTPDEPG-
                 :  * .:   .            .         :      :* : :         .
Prim.cons.     GHSKDLVNWRL22RPL2R2SQLDMKGNPDSGGVSWAPCLSY2DGKFWLIYTDVKVVDGG
Trình tự 1     PFKDGHNYLVTADAVDGEWS-DPVRLNSSG-DPSLFHDPSGKKYVLNMLWDHREKHHSF
Trình tự 2     PFKDGHNYLVTASEVDGDWS-EPIRLNSSG-DPSLFHDQSGKKYVLNMLWDHREKHHSF
Trình tự 3     AWKDCHNYLVTCETVDGDWG-EPIQLNSSG-DASLFHDVDGRKYLLNMLWDHRIGRHSF
Trình tự 7     QWKDGHNYLVTCDTIDGAWS-DPIYLNSSG-DPSLFHDEDGRKYLVNMYWDHRVDHHPF
Trình tự 6     PWKNGRNFLVTAPSIEGPWS-EPIPMGNGG-DPSLFHDDDGRKYYLYRPWGPRHHSNPH
Trình tự 11    NWMATPSYVVTADSIEGPWS-EPVPIGNYG-DPSLFHDDDGKKYMLNMIWGGRAKTNFF
Trình tự 10    NFWDAPNYVTTAPDITGPWS-DPVPLHSLG-DPSLFHDDDGRSWLLSTSMDWRPGRDAF
Trình tự 8     -------FIVMSEDLQGEWS-EPIWIDGWGGDPSLFFDNGKVYITGTND-NARGEE--
Trình tự 4     ----GRYYLYYVLDKVPVVS-VAVCDTPAG-KYEFYGYVR---YADGTRLGEREGD--
Trình tự 9     ----GRYYLYYVLDKVPIVS-VAVCDSPAG-EYEFYGYVR---YSDGTRLGERQGD--
Trình tự 5     ------LFMSTATDPAGPWSPLHEVVRVAGWDPCPFWDDG------NAYLGHSTVG-A
                     ::          .         * .   :              .
Prim.cons.     PWKDGHNYLVTADDIDGPWSPEP2PLNSSG2DPSLFHDDDG2KYLLNMLWGHREGHHSF
Trình tự 1    AGIALQEYSVSEKKLVGERKVIFKGTPIKLTEAPHLYYINDVYYLLTAEGGTRYEHAATI
Trình tự 2    AGIALQEYSVAEKKLIGRRKVIFKGTPIKLTEAPHLYHIGDNYYLLTAEGGTRYEHAATI
```

```
Trình tự 3    GGIVMQEFSVREKRLIHQPKIIFEGTDIKLTEAPHLYHIGDYYYLLTAEGGTRYEHAATI
Trình tự 7    YGIVLQEYSVEQKKLVGEPKIIFKGTDLRITEGPHLYKINGYYYLLTAEGGTRYNHAATI
Trình tự 6    NTIVMQEFDPQTGTLSPERKTLFTGTPLCYTEGAHLYRHAGWYYLMVAEXGTSYEHAVVV
Trình tự 11   GGIIMQEFDADEGKLVGAPKTVFEGTELGCTEGPQLLKKDDYYYLITAEGGTERNHAVTV
Trình tự 10   GGIVAQEFSVRDMKLVGEPVIIFTGTEAGVTEAPHIYKRDGWYYLVTAEGGTQWEHQVTV
Trình tự 8    LGIYQAEIDLKKGSIIGERKLIWKGTGGSYPEAPHLYKVNGWYYLLIAEGGTEYGHMVTV
Trình tự 4    ----WPQFDPAVLTE-GERTYLYTG----FCPKGDKSRKGAMATVLGPDMLTVVEEPVII
Trình tự 9    ----EPQFDPAVLTE-GEFTYLYTG----FCAIGDKSRKGAMATVLGRDMLTIVEEPVFV
Trình tự 5    GPIIIHKMSPDGTKLLDDGVIVYVG---KIAEGTKIYKRNGYYYLIIPEGGVKTG-WQTV
                    : .            :: *         .         ::  :  .       :
Prim.cons.    GGIV2QEFSVAEKKLVGERKIIF2GTP2KLTEAPHLYKI2GYYYLLTAEGGTRYEHAVTV
Trình tự 1     ARSSRIDGPYEVHPDNPILTAFHAPSHPLQKCGHASIVQTHTNEWYLAHLTGRPIH-SK
Trình tự 2     ARSSHIEGPYEVHPDNPILSAFHAPEHPLQKCGHASIVQTHTNEWYLAHLTGVPIQ-NK
Trình tự 3     ARSKHIEGPYEIHPDNPILTSWHEPRNPLQKCGHASIVETHTGEWYLAHLTGRPIH-AD
Trình tự 7     ARSTSLYGPYEVHPDNPLLTSWPYPRNPLQKAGHASIVHTHTDEWFLVHLTGRPLP-EG
Trình tự 6     LRAKTIDGPYELHPDVTMMTSWHLPENPLQKSGHGSLLQTHTGEWYMAYLTSRPLR-PG
Trình tự 11    -RSKHIWGPYEVHPENPILTSRFQEHAELSRAGHGFLVETQTGEWYMSHLCGRRIPNPE
Trình tự 10    ARSRSVTGPYEVDPAGPALTSRHVPEAPLQKAGHASMVETQHGEWYFAHLTGRPMP-P-
Trình tự 8     ARSKYPFGPFESCPFNPILTHRS--NHPLQAIGHADIVQYHDGSWWAVFHGTRPISYPP
Trình tự 4     VPSE----PYSRGSG—EGHEFFEAPSNPIRKKGDTYYFIYSSVVMHELCYATSKHPTKG
Trình tự 9     APSE----PYSKGSG--EGHEFFEAPSPIRKRGDTYYLIYSSVVMHELCYATSPFPTKG
Trình tự 5     LRSKSIYGPYERKVV-----LQTG-NTNINGPHQGALVELEGESWFMHFQDTGIL----
                 :     *:.                 :    .   .
Prim.cons.     ARSK2I3GPYEVHPDNPILTS2H2PE2PLQK2GHASIV2THTGEWYLAHLTGRPPTPPG
Trình tự 1     ESIFQQ-RGWCPLGRETAIQKLEWK-DGWPYVVGGKEGLLEVEAPA-MSVKEFSPTYHI
Trình tự 2     ESIFQQ-RGWCPLGRETAIQKLEWK-DGWPYVVGGKEGTLEVEAPK-IEEKDFAPTYHT
Trình tự 3     DSVIHQ-RGYCPLGRETAIQKLEWK-DGWPYVAGGKEGRVEVEAPR-ISETKFPSTYQE
Trình tự 7     QPLLEH-RGYCPLGRETAIQRLEWK-DGWPYVVGGNGPSLEIDGPS-VEEVSWEKDYDE
Trình tự 6     PLLASGGRGYCPLGRETGIARIEWR-DGWPYVEGGKHAQLTVKGPQ-VAEQPAAVQGSW
Trình tự 11    QFMPKYDNGFSILGRESALQKAHWQ-DDWPYIATGKTPVVEVEAPN-LPLHPWPES-PA
Trình tự 10    -------SGRCVLGRETALQKIEWSSDGWPRVRN-AEPLLEVPGPRGLAPHPWPQP-SE
Trình tự 8     ---------KHHLGRETCLAPIKWTDDGWPIIGYNGRIDIKMDAGYLPVKEKNIGDEII
Trình tự 4     ----FKYGGVIVSNCDLHIDSYKPAEKPMYYGGNNHGSIVEINGEWYIFYHRHTNG---
Trình tự 9     ----FTYQGVIVSNNDLHIDSYKPADKPMYYGGNNHGGAVEIQGQWYIFYHRHTNG---
Trình tự 5     -------------GRVCHLQPVTWV-DNWPLMGCDGEPVTVYRKPRAGKEYERTFP-QT
                             .    :       .
Prim.cons.     ES22QQ4RGYCPLGRETAIQKLEWKDDGWPYVGGGKE22LEVE2PRYI4EHP222TY3E
Trình tự 1     VDEFKDSSLNRHFQTLRIPFTDQIGSVTENPHHLRLYGQE-LTSKFTQAFVARRWQSFY
Trình tự 2     VDEFKESTLNRHFQTLRIPFTDQIGSLTEKPRHLRLFGRE-LTSKFTQAFVARRWQSFY
Trình tự 3     VDQFADRTLNMNFQTLRIPFTKELGSLTERPNHLRLYGRE-LTSTFTQAFVARRWQSLR
Trình tự 7     KDDFDGDTLNHHFQTLRIPLGEDIATLKARPGHLRLYGRE-LTSRFTQAFVARRWQHFH
Trình tự 6     RDDFDGSTLDPELQTLRIPFDDTLGSLTARPGYLRLYGND-LNSTFTQSTVARRWQHFI
Trình tự 11    RDEFIDPTLS-LISTLREPVSEKWLSLSERPGFLRLKGRH-LYSRYEQSMVARRFQAHN
Trình tự 10    TDHFDDPTPRPEWSTLRRPFDSSWVSLTERPGYLRIRGGQ-PAGLHEPSLVARRLQHRA
Trình tự 8     EDDFNSDIFSTDWNFIQNPRLEHYSS-KG RPSWLKMRGTEKTLNDINSPTFIRRQEHF
```

```
Trình tự 4     -TSFSRQGCMEKIKILEDGSIPQVEMTSCGSADEPLPGRG-YPAYIACNLFCGEESVYT
Trình tự 9     -TAFSRQGCMEPISFREDGTIPQVEMTSCGPNGGPLAGRG-YPAYLACNLFCKDEELYT
Trình tự 5     SDEFDGPELGLQWQWNHNPVNERWS-LSKRPGYLTLEAMY-ESLLKARNTLTQKLIGEK
                  *          .  .          .  .    : .            .
Prim.cons.     VDEFDD2TLN3HFQTLRIPFTEQ2GSLTERPGHLRLYGRESLTS5FTQAFVARRWQHF2
Trình tự 1     FEAETAVSFFPKNFQQAAGLVNYYNTENWTALQVTYDD-LGRILELSVCENLAFSQPLI
Trình tự 2     FEAETAVSFFPENFQQSAGLVNYYNTENWTALQVTYDE-LGRTLELSVCQNLAFSQPLA
Trình tự 3     FTAETAVDFNPETFQQAAGLVNYYNTENWTALQVTYDE-LGRILDVTICDNFTFSQPIK
Trình tự 7     FVAETKVSFRPTTFQQSAGLVNYYNTQNWTTLQITWHE-KGRILELMTCDHLVVDQPLR
Trình tự 6     FRAETRMQFSPVHFQQSAGLTCYYNSKNWSYCFVDYEE-QGRTIKVIQLDHNVPSWPLH
Trình tự 11    ATVETKLEFKPNTPYEMAGLCAYYARNGHYFLKMTAND-GERVLQVVGNINDVYGEYSN
Trình tự 10    CIFEACLEFKPEDFRQMAGITAYYNTRQWHYLRINRDD-GGVFAGVLTSDRGIIREVGR
Trình tự 8     NVS-TLLEFKPNQDNEEAGLTVYMNEKHHYEIAL-TKKNGRINVVLKKTVGDIQVVVN-
Trình tự 4     ---LTGAWMNN-QFPKITQDGKDGDEEPGYIANMKDSATAGFKYFDCKGIKSVKIKVR-
Trình tự 9     GFGASGAWMDS-RFPKITQDGKDGDEEMGYIANM-TDSATAGFKYFDCHGIRRMTIQV-
Trình tự 5     GTATTELNTENLKNQQRAGLAFLGGTQENWIG-VREGE---SYIKAVTAGIRYHGPEIE
                                : :         .      :
Prim.cons.     FTAETA22FKP2TFQQSAGLVNYYNTENW22LQVTYDET-GGR2L3V33CDNLV2SQPL
Trình tự 1     KK-IIIPDEIPYVYLKVTVQRE-TYTYSYSFDQQEWEKIDVPLESTHLSDDFIRGGG--
Trình tự 2     HK-IIIPDEVTYVYLKVTVRKE-TYKYSYSFDQKEWKEIDVPLESIHLSDDFIRGGG--
Trình tự 3     DK-IVIPRDIQYVYLRVNVEWE-TYYYSYSFNKKDWYKIEIPLESKKLSDDYVRGGG--
Trình tự 7     GREIVVPDDIEYVYLRVTVQAT-TYKYSYSFDGMNWIDLPVTFESYKLSDDYIKSRA--
Trình tự 6     EQPIPVPEQAESVWLRVDVDRL-VYRYSYSFDGETWHAVPVTYEAWKLSDDYIGGRG--
Trình tự 11    DV---VIGDADTVYMRLELKTQ-WYQYSYSLDGVDWYEIGPALNSTPLSDEGGPDIF--
Trình tự 10    RIS-V-TGWPKVFLRAEIDRN—LRFAV--SSDGSTWADMGVRLDMSILSDEYAEERFGN
Trình tự 8     YFS-------NTIIFSIQANPE-EYKFSFVDPNTGQTYLLGTGLTTLLSTEVAGGFT--
Trình tự 4     CR------------G--------VFEVKTSWNGEVLGKIPVEFSN-IWTEFSASIP---
Trình tự 9     CR------------G--------AFEIKTAWNGPVLGTIPVEFSN-VWKPYSTELV---
Trình tự 5     TP---------NVWFRAEIDLNGITQFYFSTDNENFIQLGGPCRL-----EAG------
                                                        :
Prim.cons.     2-SI2VPDDI2YVYLRV2VDRE-TYKYSYSFDGE5W3KIPVPLESTKLSD2YIRG2GGN
Trình tự 1     ----FFTGAFVGMQCQ-DTSGERLPADFKYFRYEETTE------
Trình tự 2     ----FFTGAFVGMQCQ-DTSGERLPADFHYFRYEETDE------
Trình tự 3     ----FFTGAFVGMQCQ-DTSGAHQHADFDYFMYKEL--------
Trình tự 7     ----AFTGAFVGMHCR-DGSGQNNYADFDYFLYKEL--------
Trình tự 6     ----FFTGAFVGLHCE-DISGDGCHADFDYFTYEPA--------
Trình tự 11    ----RFTGSFAALFVA-DITGQKRHADFDYFEYLEHSEQXENCE
Trình tự 10    PIMWGFTGAFLGLWAH-DMTGAGLPADFDFCTYRPQSPS-----
Trình tự 8     -------GVYFGLYATGNGKVCTAPAFFDWFKYIPEI-------
Trình tự 4     -----IPDGIHALYFTYRGSGSASLKSFTLCTD-----------
Trình tự 9     -----IPDGIQALYFTYTGMGSASLASFTLE-------------
Trình tự 5     ----FWKGARIALFSY-NTVMDGGRADFNWFRYEFE--------
                      .   . :              *
Prim.cons.     PIMWFFTGAFVGL2C2YD2SG5GLPADFDYF2YEE3SE2XENCE
```

Hình 3.2. Kết quả so sánh sự tương đồng của 11 trình tự mã hóa β–xylosidase và xây dựng mẫu dò cho β–xylosidase thuộc họ GH43

Ghi chú: Mức độ bảo thủ của các gốc axit amin được đánh dấu từ dấu "" đến dấu ":" dấu "." và không được đánh dấu. Các vị trí gạch chân trong trình tự Prim. cons sẽ được dùng làm mẫu dò. Trình tự 1,2 3,…. là các trình tự từ Bảng 3.4 thuộc họ GH43 từ số thứ tự 1 đến 11*

Để có thể xây dựng được mẫu dò, yêu cầu số lượng trình tự mã hóa β–xylosidase càng nhiều thì độ tin cậy của mẫu dò sẽ càng cao. Dựa trên số liệu thu thập trình tự axit amin thuộc về các họ GH khác nhau cho thấy, chỉ duy nhất một họ GH43 có số lượng trình tự đủ lớn (11 trình tự) để có thể so sánh tìm vùng tương đồng. Chúng tôi nhận thấy các enzyme thu thập được trong họ GH43 có nguồn gốc từ vi khuẩn, hoạt động tối ưu từ pH axit (pH = 3,5) đến trung tính (pH = 7) và nhiệt độ từ 30°C đến 80°C, độ dài không khác nhau nhiều từ 473 axit amin đến 550 axit amin (Bảng 3.4). Kết quả phân tích sau khi so sánh 11 trình tự thuộc GH43 bằng phần mềm ClustalW – PBIL như Hình 3.2.

Dựa vào mức độ bảo tồn các vị trí axit amin trong chuỗi polipeptide, khi so sánh 11 trình tự cho thấy: Có 7 vị trí axit amin bảo tồn cao nhất, hoàn toàn giống nhau (màu đỏ), 32 vị trí bảo tồn ở mức trung bình, giống nhau ở đa số các trình tự (màu xanh lục) và 30 vị trí bảo tồn thấp hơn, giống nhau ở một số trình tự (màu xanh lam). Ở các vị trí axit amin còn lại chỉ so sánh sự giống nhau ở cùng một số vị trí có: 71 vị trí axit amin giống nhau từ 8 đến 10 trình tự (màu cam), 102 vị trí giống nhau từ 6 đến 7 trình tự (màu hồng), ở các vị trí khác loại axit amin nào chiếm tỷ lệ lớn hơn sẽ được giữ lại (màu đen). Trong trường hợp số lượng giữa các loại axit amin bằng nhau thì trình tự cuối cùng sẽ để số - đại diện cho số lượng loại axit amin lặp lại nhiều nhất (Ví dụ số 2, nghĩa là giữa 11 trình tự có 02 loại axit amin có số lượng lớn nhất bằng nhau). Những vị trí axit amin chỉ có duy nhất ở ít hơn 50% trình tự sẽ được đánh dấu là màu tím (Hình 3.2). Kết quả này cho thấy trình tự các axit amin của β–xylosidase thuộc họ GH43 ở vi khuẩn bảo tồn rất cao. Tính bảo tồn

cao của axit amin trong các trình tự β–xylosidase cho thấy việc lựa chọn trình tự chỉ từ vi khuẩn đảm bảo cho số liệu mẫu dò đáng tin cậy.

3.2.4. Xây dựng mẫu dò và giá trị tham chiếu

Tất cả 11 trình tự axit amin của β–xylosidase thuộc họ GH43 rất điển hình về chiều dài, vị trí axit amin bảo tồn và trung tâm hoạt động. Do đó chúng tôi sẽ lựa chọn cả những vị trí axit amin không giống nhau hoàn toàn ở 11 trình tự, để đảm bảo được độ dài và việc hình thành cấu trúc không gian chính xác.

Mẫu dò này sẽ được so sánh lại với 11 trình tự để xác định điểm tối đa, mức độ bao phủ và tương đồng bằng BLASTP. Trên cơ sở đó xác định giá trị tham chiếu cho mẫu dò khi sử dụng để lựa chọn gen mã hóa cho β–xylosidase từ số liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi*. Căn cứ vào kết quả so sánh giữa 11 trình tự thuộc họ GH43 chúng tôi sẽ lựa chọn các vị trí được bảo tồn (màu đỏ, màu xanh lục, màu xanh lam), vị trí đa số axit amin giống nhau (màu cam và màu hồng), vị trí chiếm tỷ lệ lớn hơn (màu đen) và vị trí có loại axit amin bằng nhau (số) để đảm bảo độ dài cho mẫu dò. Những vị trí axit amin có mặt ít hơn 50% trình tự (màu tím) sẽ loại bỏ.

```
I2NPVLKGFNPDPSICRVGEDYYIA2STFEWFPGV2I2HSKDLVNWRL22RPL2R2SQLDM
KGNPDSGGVWAPCLSY2DGKFWLIYTDVKVVDGPWKDGHNYLVTADDIDGPWSPEP2PLNS
SGFDPSLFHDDDG2KYLLNMLWGHREGHHSFGGIV2QEFSVAEKKLVGERKIIF2GTP2KL
TEAPHLYKI2GYYYLLTAEGGTRYEHAVTVARSK2I3GPYEVHPDNPILTS2H2PE2PLQK
2GHASIV2THTGEWYLAHLTGRPIPTPPRGYCPLGRETAIQKLEWKDGWPYVGGGKE22LE
VE2PRYI4EHP222TY3EVDEFDD2TLN3HFQTLRIPFTEQ2GSLTERPGHLRLYGRESLT
S5FTQAFVARRWQHF2FTAETA22FKP2TFQQSAGLVNYYNTENW22LQVTYDETGGR2L3
V33CDNLV2SQPL3DDI2YVYLRV2VDREYKYSYSFDGE5W3KIPVPLESTKLSD2YIRG2
GGFFTGAFVGL2C2YD2SG5GLPADFDYF2YEE3
```

Hình 3.3. Trình tự mẫu dò của β–xylosidase thuộc họ GH43

Kết quả phân tích, mẫu dò của họ GH43 được xây dựng bao gồm có 464 axit amin. Trong đó chứa toàn bộ 7 axit amin màu đỏ, hoàn toàn giống nhau bảo tồn cao nhất; 32 vị trí màu xanh lục, giống nhau bảo tồn trung bình; có 30 vị trí màu xanh

lam, bảo tồn ở mức độ thấp hơn; 71 vị trí axit amin màu cam, giống nhau từ 8 đến 10 trình tự, 102 vị trí màu hồng, giống nhau từ 6 đến 7 trình tự, 165 vị trí màu đen và 57 vị trí là số (Hình 3.2). Trình tự mẫu dò của họ GH43 như hình 3.3.

Chúng tôi đã so sánh mức độ tương đồng giữa mẫu dò với từng trình tự đã sử dụng bằng BLASTP để tìm giá trị tham chiếu cho việc sử dụng mẫu dò trong khai thác gen. Kết quả cho thấy các trình tự mã hóa β–xylosidase GH43 phải có điểm tối đa trên 17, độ bao phủ và độ tương đồng tối thiểu 60% và 30% so với mẫu dò (Bảng 3.5). Tuy nhiên, kết quả so sánh cũng cho thấy mẫu dò này phù hợp nhất với trình tự số 1, 2, 3, 6, 7, 8, 10, 11 (chỉ số điểm tối đa trên 200, độ tương đồng trên 37% và độ bao phủ trên 99%) và không đại diện tốt cho trình tự số 4, 5, 9 vì chỉ số điểm tối đa và độ tương đồng thấp.

Bảng 3.5. Bảng so sánh độ tương đồng của mẫu dò với các trình tự thuộc GH43

Trình tự	**Điểm tối đa**	**Tổng điểm**	**Độ bao phủ** (%)	**Giá trị E**	**Độ tương đồng** (%)
Trình tự 2	608	608	100	0	72
Trình tự 1	603	603	99	0	71
Trình tự 3	596	596	99	0	71
Trình tự 7	547	547	99	0	66
Trình tự 6	444	444	99	5E-155	56
Trình tự 10	358	377	98	9E-122	51
Trình tự 11	351	351	99	4E-119	49
Trình tự 8	231	248	99	2E-73	37
Trình tự 5	101	134	83	3E-27	30
Trình tự 4	25,8	76,2	61	0,003	33
Trình tự 9	17,3	113	60	1,2	46

Tìm hiểu sâu hơn về trình tự 4, 5 và 9 cho thấy: trình tự 4 và 9 đều mã hóa enzyme hai chức năng β-xylosidase/α-arabinofuranosidase với hoạt tính của α -

arabinofuranosidase mạnh hơn [91] và trình tự 5 mã hóa cho β–xylosidase [161]. Do đó chúng tôi sử dụng công cụ Swiss model dự đoán cấu trúc không gian để kiểm chứng thêm mức độ chính xác hoạt tính của trình tự axit amin 4, 5 và 9. Kết quả cho thấy cấu trúc không gian của cả 03 trình tự này đều tương đồng cao với cấu trúc β-xylosidase/α-arabinofuranosidase (Bảng 3.6). Kết quả này có thể giả định rằng trình tự 4, 5 và 9 sử dụng làm mẫu dò có điểm tối đa thấp có thể là do chúng thực hiện đồng thời 02 hoạt tính, nhưng hoạt tính alpha-arabinofuranosidase hoạt động mạnh hơn so với hoạt tính của β-xylosidase.

Bảng 3.6. Bảng tổng hợp ước đoán cấu trúc bậc ba của trình tự 4, 5, 9 bằng công cụ Swiss model

Trình tự	**Khuôn**	**Hoạt tính**	**Độ bao phủ**	**Độ tương đồng**	**Phương pháp**	**Cấu trúc**	**Phối tử**
Trình tự 4	4nov.1.A	Xylosidase/arabinofuranosidase Xsa43E	0,6	30,28	X-ray, 1,3Å	Mono-mer	1 x CA
Trình tự 5	3c2u.1.A	Xylosidase/arabinofuranosidase	0,94	27,08	X-ray, 1,3Å	Homotetramer	4 x B3P
Trình tự 9	5glk.1.A	beta-xylosidase/ arabinofuranosidase	0,56	36,84	X-ray, 1,7Å	Mono-mer	Không có

Để khai thác tối đa các trình tự gen mã hóa enzyme có hoạt tính của β–xylosidase, khi sử dụng mẫu dò sẽ chọn những trình tự có điểm tối đa, độ tương đồng và độ bao phủ tối thiểu lần lượt là 17, 30% và 60%. Tuy nhiên để đảm bảo cho khả năng gen đã chọn biểu hiện thành công, chúng tôi đề xuất nên chọn các trình tự có điểm tối đa, độ tương đồng và độ bao phủ tối thiểu lần lượt là 200, 37% và 99% với mẫu dò.

Tóm lại dữ liệu về trình tự gen/axit amin được tìm kiếm chủ yếu từ NCBI và CAZY thỏa mãn các điều kiện: có nguồn gốc từ vi khuẩn, đã được nghiên cứu thực

nghiệm về đặc điểm hoạt động của enzyme, không quá khác nhau về chiều dài sẽ được thu thập để làm mẫu dò cho từng họ GH. Trong số 11 họ GH chứa β–xylosidase, chỉ họ GH43 thu thập số trình tự đủ lớn để xây dựng mẫu dò có độ dài là 464 axit amin, với giá trị tham chiếu về điểm tối đa, độ tương đồng và độ bao phủ của mẫu dò với gen lựa chọn tối thiểu là 17, 30% và 60%. Tuy nhiên nên lựa chọn trình tự gen có điểm tối đa, độ tương đồng và độ bao phủ tối thiểu tốt nhất là 200, 37%, 99% để đưa vào thực nghiệm.

3.2.5. Khai thác trình tự gen mã hóa β–xylosidase bằng mẫu dò từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật ruột mối *C. gestroi*

Theo kết quả của công ty giải trình BGI dựa vào CSDL của KEGG đã chú giải 46 trình tự từ DNA đa hệ gen của vi sinh vật trong ruột mối có hoạt tính β–xylosidase thuộc họ GH43. Chúng tôi sẽ sử dụng mẫu dò để kiểm tra lại độ chính xác của ước đoán từ BGI, bằng cách so sánh tương đồng với tất cả dữ liệu giải trình tự DNA đa hệ gen đã có. Kết quả cho thấy, nếu sử dụng điểm tối đa trên 17, độ bao phủ và độ tương đồng tối thiểu 60% và 30% thì chúng tôi lựa chọn được 25 trình tự, trong đó có 20 trình tự trùng với dự đoán BGI. Tuy nhiên vẫn có 05 trình tự sử dụng mẫu dò tìm kiếm được lại không nằm trong dự đoán của BGI. Nếu sử dụng điểm tối đa trên 200, độ tương đồng tối thiểu là 37% và độ bao phủ thấp nhất là 99% để chọn gen đưa vào thực nghiệm, thì chỉ duy nhất 01 trình tự đạt yêu cầu về các chỉ số tham chiếu (Bảng 3.7).

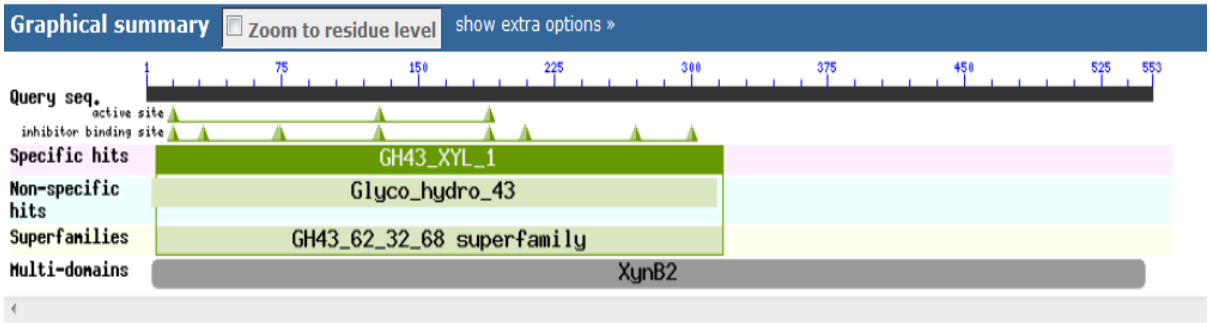
Bảng 3.7. Bảng so sánh kết quả khai thác bằng mẫu dò với dự đoán của BGI

Kết quả dự đoán của BGI	**Kết quả khai thác bằng mẫu dò**					
	Mã gen	**Điểm tối đa**	**Tổng điểm**	**Độ bao phủ**	**Giá trị E**	**Độ tương đồng (%)**
GL0104795	GL0104795	382	382	99	1e-130	48
GL0020896	GL0020896	285	285	62	3e-96	54
GL0117445	GL0117445	285	285	62	3e-96	54

GL0076106	GL0076106	243	259	85	2e-77	38
GL0006405	GL0006405	227	271	68	7e-75	57
GL0044986	GL0044986	225	256	87	7e-72	37
GL0077826	GL0077826	216	265	67	9e-70	48
GL0112518	GL0112518	216	265	67	9e-70	56
GL0001262	GL0001262	214	266	82	8e-67	35
GL0095948	GL0095948	213	213	63	3e-53	39
GL0015489	GL0015489	201	201	67	2e-49	42
GL0088906	GL0088906	154	193	62	2e-48	53
GL0016592	GL0016592	130	160	64	8e-40	49
GL0021333	GL0021333	99,8	117	77	2e-27	30
GL0090776	GL0090776	54,7	119	63	4e-12	37
GL0083296	GL0083296	43,9	58,1	67	4e-09	35
GL0091901	GL0091901	40,4	54,3	64	3e-08	35
GL0079004	GL0079004	37,0	85,1	65	5e-07	36
GL0050001	GL0050001	35,8	112	67	8e-07	36
GL0106540	GL0106540	34,7	72,0	66	2e-06	35
	GL0119754	93,2	156	61	3e-26	38
	GL0039878	55,8	103	69	2e-13	36
	GL0122352	27,3	102	69	4e-04	32
	GL0068837	24,6	115	66	0,002	35
	GL0042431	22,7	84,3	61	0,011	31

Chúng tôi tiếp tục khảo sát lại vùng bảo thủ bằng BLASTP tất cả các trình tự dự đoán mã hóa β–xylosidase của BGI và mẫu dò dự đoán cho thấy: Tất cả 25 trình

tự được lựa chọn bằng mẫu dò đều chứa những vùng đặc thù cho β–xylosidase (specific hit) và có vị trí hoạt động (active sites) (Hình 3.4). Trong khi đó rất nhiều trình tự gen mã hóa β–xylosidase do BGI dự đoán không mã hóa cho enzyme này. Điều này chứng tỏ việc xây dựng mẫu dò để tìm kiếm và lựa chọn được ít trình tự gen hơn so với dự đoán của BGI, nhưng chính xác hơn.



Hình 3.4. Kết quả dự đoán tương đồng đặc hiệu trình tự và các gốc hoạt động của các trình tự được lựa chọn bằng mẫu dò GH43

Chú thích: GH43_XYL: họ glycosyl 43 có khả năng phân hủy cấu trúc beta-D-xyloside; XynB2: enzyme beta-xylosidase

3.2.6. Khảo sát cấu trúc bậc 3 của các β–xylosidase đã khai thác bằng mẫu dò

Cấu trúc bậc ba của phân tử protein cho phép ước đoán cụ thể hơn về trung tâm hoạt động, cấu hình phân tử và phối tử liên quan đến hoạt tính sinh học của enzyme. Để xác định tính chính xác của các trình tự đã lựa chọn bằng mẫu dò một lần nữa, chúng tôi tiếp tục kiểm tra lại cấu trúc không gian bằng công cụ Swiss model. Kết quả cho thấy đa số trình tự được lựa chọn bằng mẫu dò đều được ước đoán có cấu trúc tương đồng cao với β–xylosidase. Tuy nhiên một số trình tự có giá trị điểm tối đa thấp dưới 100, thì có hoạt tính khác β–xylosidase như xylanase và arabinofuranosidase (Bảng 3.8).

Bảng 3.8. Bảng tổng hợp ước đoán cấu trúc bậc ba của β–xylosidase đã chọn bằng công cụ Swiss model

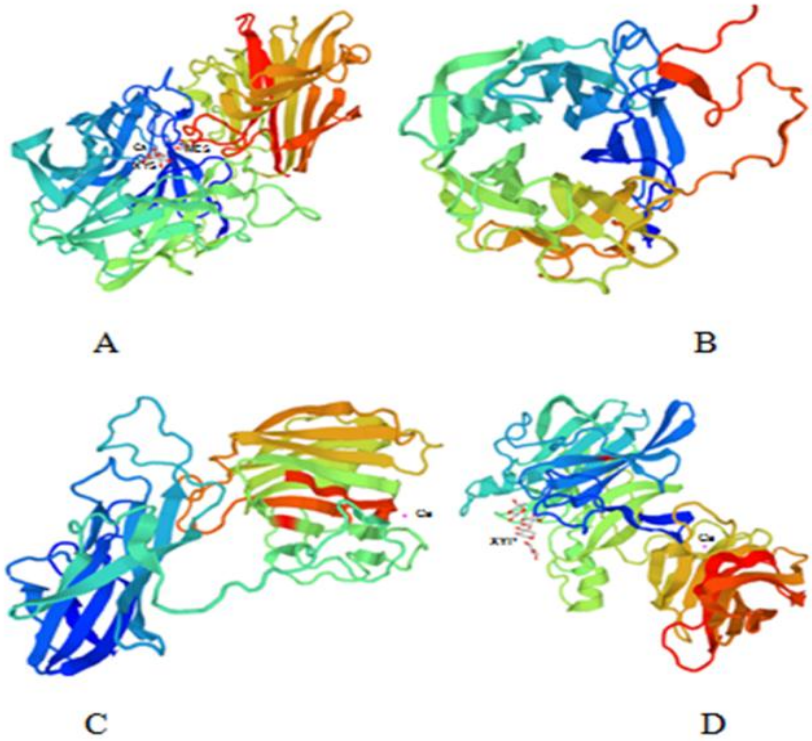
Mã gen	Điểm tối đa	Khuôn	Hoạt tính	Độ bao phủ	Độ tương đồng	Phương pháp	Cấu trúc	Phối tử

GL0104795	382	3c2u.1.A	Xylosidase/A-rabinosidase	0,95	54,77	X-ray, 1.3Å	Homo-tetramer	4 x B3P
GL0020896	285	3c2u.1.A	beta-D-xylosidase	0,93	52,11	X-ray, 1.3Å		4 x B3P
GL0117445	285	2exj.1.A	beta-D-xylosidase	0,96	55,34	X-ray, 2.2Å		4 x XYS - XYS, 4 x CA, 3 x MES
GL0076106	243	2exk.1.A	beta-D-xylosidase	0,94	37,55	X-ray, 2.2Å		4 x XYS-XYS, 4 x CA, 3 x MES
GL0006405	227	2exk.1.A	beta-D-xylosidase	0,98	59,53	X-ray, 2.2Å		4 x XYS-XYS, 4 x CA, 3 x MES
GL0044986	225	2exj.1.A	beta-D-xylosidase	0,94	40,29	X-ray, 2.2Å		4 x XYS-XYS, 4 x CA, 3 x MES
GL0077826	216	1yrz.1.A	beta-D-xylosidase	0,96	57,85	X-ray, 2.0Å	Monomer	Không
GL0112518	216	1yrz.1.A	beta-D-xylosidase	0,93	58,86	X-ray, 2.0Å	Monomer	Không
GL0001262	214	2exk.1.A	beta-D-xylosidase	0,94	36,73	X-ray, 2.2Å	Homotetramer	4 x XYS-XYS, 4 x CA, 3 x MES
GL0095948	173	1yi7.1.A	beta-D-xylosidase	0,93	46,52	X-ray, 1.9Å	Homotetramer	4 x CA
GL0015489	161	2exh.1.A	beta-D-xylosidase	0,98	46,56	X-ray, 1.9Å	Homotetramer	4 x CA, 3 x MES
GL0088906	154	1yrz.1.A	beta-D-xylosidase	0,96	64,1	X-ray, 2.0Å	Monomer	Không
GL0016592	130	1yi7.1.A	beta-D-xylosidase	0,98	48,9	X-ray, 1.9Å	Homotetramer	4 x CA

GL0021333	99,8	1yrz.1.A	beta-D-xylosidase	0,96	31,65	X-ray, 2.0Å	Monomer	Không
GL0090776	54,7	2exj.1.A	Endo-1,4-beta-xylanase	0,42	24,22	X-ray, 2.2Å	Homotetramer	4 x XYS-XYS, 4 x CA, 3 x MES
GL0083296	43,9	3c7g.1.A	Endo-1,4-beta-xylanase	0,93	33,88	X-ray, 2.0Å	Monomer	4 x XYP, 1 x CA
GL0091901	40,4	3c7g.1.A	Endo-1,4-beta-xylanase	0,91	31,94	X-ray, 2.0Å	Monomer	4 x XYP, 1 x CA
GL0079004	37	1yrz.1.A	beta-1,4-xylosidase	0,94	18,62	X-ray, 2.0Å	Monomer	Không
GL0050001	35,8	3qee.1.A	β-xylosidase/ α-L-arabinofuranosidase	0,99	56,51	X-ray, 1.6Å	Monomer	1 x CA
GL0106540	34,7	4nov.1.A	β-xylosidase/α-arabinofuranosidase	0,93	55,74	X-ray, 1.3Å	Monomer	1 x CA
GL0119754	93,2	2exh.1.A	beta-1,4-xylosidase	0,81	42,77	X-ray, 1.9Å	Homotetramer	4 x CA, 3 x MES
GL0039878	55,8	1yrz.1.A	beta-1,4-xylosidase	0,94	31,09	X-ray, 2.0Å	Monomer	Không
GL0122352	27,3	3kst.1.A	Endo-1,4-beta-xylanase	0,9	36	X-ray, 1.7Å	Monomer	1 x CA
GL0068837	24,6	2exk.1.A	Endo-1,4-beta-xylanase	0,89	25,68	X-ray, 2.2Å	Homotetramer	4 x XYS-XYS, 4 x CA, 3 x MES
GL0042431	22,7	3c7f.1.A	Endo-1,4-beta-xylanase	0,82	32,08	X-ray, 1.5Å	Monomer	1 x CA

Chú thích: B3P:2-[3-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino] propylamino]-2-(hydroxymethyl)propane-1,3-diol; XYS: xylopyranose; CA: Ca^{++}; MES: 2-morpholin-4-ium-4-ylethanesulfonate

Xem xét dự đoán cấu trúc không gian của trình tự có điểm tối đa cao trên 100 cho thấy chúng đều tương đồng với beta-D-xylosidase và thường có vị trí liên kết với ion Ca^{++} nằm ở vùng liên kết giữa các phân tử polyme để tạo dạng tetramer. Đồng thời các trình tự này còn có vùng liên kết với xylopyranose đặc thù trên phân tử cơ chất của β–xylosidase, chứng tỏ chúng có hoạt tính phân cắt xylose thành đường (Hình 3.5). Như vậy hoàn toàn phù hợp giữa ước đoán cấu trúc và chức năng của phân tử. Từ kết quả trên cho thấy, chỉ số điểm tối đa là rất quan trọng để xem xét tính xác thực của gen và protein khi sử dụng các công cụ tin sinh học để ước đoán. Kết quả này một lần nữa giúp chúng tôi khẳng định việc chọn chỉ số điểm tối đa trên 200 là giá trị tham chiếu khi sử dụng mẫu dò làm tăng độ chính xác khi chọn gen đưa vào thực nghiệm.



Hình 3.5. Cấu trúc bậc ba của các β-xylosidase khuôn (template) 2exk.1.A (A), 1yrz.1.A (B), 1yi7.1.A (C), 3c7g.1.A (D) họ GH43 tương đồng với các trình tự được khai thác bằng mẫu dò sử dụng Swiss model

B3P: 2-[3-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]propylamino]-2-(hydroxymethyl)propane-1,3-diol; XYS: xylopyranose; CA: Ca^{++}; MES: 2-morpholin-4-ium-4-ylethanesulfonate

Theo kết quả Bảng 3.7 nếu dựa trên điểm tối đa trên 200, độ tương đồng tối thiểu là 37% và độ bao phủ thấp nhất là 99% để chọn gen đưa vào thực nghiệm, thì chỉ duy nhất 01 trình tự đạt yêu cầu về các chỉ số tham chiếu. Do đó để có nhiều sự lựa chọn gen hơn khi đưa vào thực nghiệm, chúng tôi vẫn sử dụng mẫu dò của họ GH43 với chỉ số tham chiếu về độ bao phủ là 60%, độ tương đồng là 30% và nhưng điểm tối đa phải trên 200. Như vậy với các chỉ số này, chúng tôi lọc được 11 mã gen từ dữ liệu trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* bằng mẫu dò là: GL0104795, GL0020896, GL0117445, GL0076106, GL0006405, GL0044986, GL0077826, GL0112518, GL0001262, GL0095948, GL0015489.

3.2.7. Dự đoán cấu trúc và chức năng của các gen mã hóa β–xylosidase bằng một số công cụ tin sinh học

Việc lựa chọn gen đưa vào thực nghiệm sẽ ưu tiên các trình tự được ước đoán ban đầu là ORF hoàn thiện. Trong số 11 mã gen, có 04 mã gen là hoàn thiện gồm GL0104795, GL0076106, GL0001262 và GL0112518 có mức độ bao phủ và tương đồng với mẫu dò như Bảng 3.9.

Nếu theo tiêu chí tốt nhất để chọn gen đưa vào thực nghiệm là điểm tối đa trên 200, độ tương đồng và độ bao phủ là 37% và 99%, thì trong 04 mã gen hoàn thiện chỉ có mã gen GL0104795 thỏa mãn các tiêu chuẩn trên. Tuy nhiên, ngoài việc chọn được gen mã hóa β–xylosidase biểu hiện thành công trong thực nghiệm, chúng tôi còn quan tâm đến khả năng enzyme hoạt động tối ưu trong môi trường kiềm và nhiệt độ cao. Do đó 04 gen hoàn thiện sẽ tiếp tục được chọn lọc bằng các phần mềm dự đoán pH và nhiệt độ hoạt động tối ưu bằng công cụ Alcapred [199] và TBI [200].

Bảng 3.9. Bảng tổng hợp độ bao phủ và tương đồng với mẫu dò của các gen đã chọn mã hóa β–xylosidase

STT	Mã gen	Mức độ bao phủ (%)	Mức độ tương đồng (%)	Gen hoàn thiện
1	GL0104795	99	48	X
2	GL0020896	62	54	
3	GL0117445	62	54	
4	GL0076106	85	38	X
5	GL0006405	68	57	
6	GL0044986	87	37	
7	GL0077826	67	48	
8	GL0112518	67	56	X
9	GL0001262	82	35	X
10	GL0095948	63	39	
11	GL0015489	67	42	

Một số nghiên cứu đã xác định được ở mối bậc cao có sự thay đổi pH từ trung tính ở ruột trước và ruột giữa, đến ruột sau pH tăng lên đến kiềm tính và có thể đạt đến giá trị 12 [87]. Đây được cho là điều kiện khá lý tưởng để nới lỏng và phá vỡ các liên kết của lignin trong thành phần của lignocellulose. Đồng thời với điều kiện này, các vi khuẩn cũng có khả năng sản xuất enzyme hoạt động tốt trong môi trường kiềm. Dựa trên cơ sở các nghiên cứu đã công bố về pH trong ruột mối, chúng tôi dự đoán có thể tìm được lignocellulase hoạt động tốt trong môi trường kiềm.

Kết quả sử dụng công cụ TBI dự đoán cho thấy cả 4 gen đều mã hóa enzyme hoạt động tối ưu ở nhiệt độ từ 55℃ đến 65℃. Khi dự đoán bằng công cụ Alcapred cho thấy mã gen GL0104795, GL0001262, GL0076106 và GL0112518 với chỉ số dự đoán lần lượt là 0,217171; 0,509051; 0,871792 và 0,984522 (Bảng 3.10). Theo kết quả dự đoán của công cụ Alcapred nếu chỉ số dự đoán càng gần với giá trị bằng 1, thì pH hoạt động tối ưu trong môi trường kiềm sẽ càng cao và ngược lại càng gần với 0 thì pH hoạt động tối ưu sẽ trong môi trường axit. Như vậy mã gen GL0112518

được dự đoán với kết quả chịu kiềm cao nhất sẽ được ưu tiên lựa chọn để đưa vào thực nghiệm.

Bảng 3.10. Bảng kết quả dự đoán pH và nhiệt độ hoạt động của các mã gen hoàn thiện mã hóa β–xylosidase

STT	Mã gen	Số axit amin	pH hoạt động	Nhiệt độ hoạt động
1	GL0104795	561	0,217171	55°C - 65°C
2	GL0076106	553	0,871792	55°C - 65°C
3	GL0001262	555	0,509051	55°C - 65°C
4	GL0112518	359	0,984522	55°C - 65°C

Tóm lại, sử dụng mẫu dò của β–xylosidase của họ GH43, với độ dài là 446 axit amin, giá trị tham chiếu về mức độ tương đồng, độ bao phủ và điểm tối đa lần lượt là trên 60%, 30% và 200, chúng tôi lựa chọn được 11 mã gen mã hóa β–xylosidase từ dữ liệu giải trình tự DNA đa hệ gen. Trong đó chỉ có mã gen GL0112518 là hoàn thiện, được dự đoán có khả năng hoạt động tối ưu ở nhiệt độ cao và pH kiềm.

3.2.8. Một số dự đoán chi tiết gen GL0112518 mã hóa β–xylosidase (*Xbx14*)

Sau khi chọn được mã gen GL0112528 có chiều dài là 1077 bp mã hóa β-xylosidase bằng mẫu dò với các dự đoán về nhiệt độ và pH hoạt động tối ưu của enzyme như trên. Chúng tôi sẽ tìm hiểu thông tin chi tiết liên quan cấu trúc, chức năng và nguồn gốc của gen để có định hướng trong nghiên cứu thực nghiệm.

3.2.8.1. *Kết quả dự đoán chức năng của Xbx14 bằng BLASTP*

Các ORF hoàn thiện mã hóa cho lignocellulase có độ tương đồng càng thấp thì khả năng chúng là gen mới và mã hóa cho những enzyme mới với những đặc tính sinh lý hóa mới càng cao. Do đó, những ORF mã hóa cho enzyme có độ tương đồng với các protein đã biết dưới 86%, nhất là dưới 60% được đặc biệt quan tâm. Kết quả dự đoán chức năng bằng BLASTP cho thấy mã gen này mã hóa cho *Xbx14* thuộc họ GH43, chứa những vùng đặc thù và vị trí hoạt động đúng như dự đoán ban đầu khi lựa chọn bằng mẫu dò (Hình 3.6). Họ enzyme này chứa nhiều loại enzyme thủy

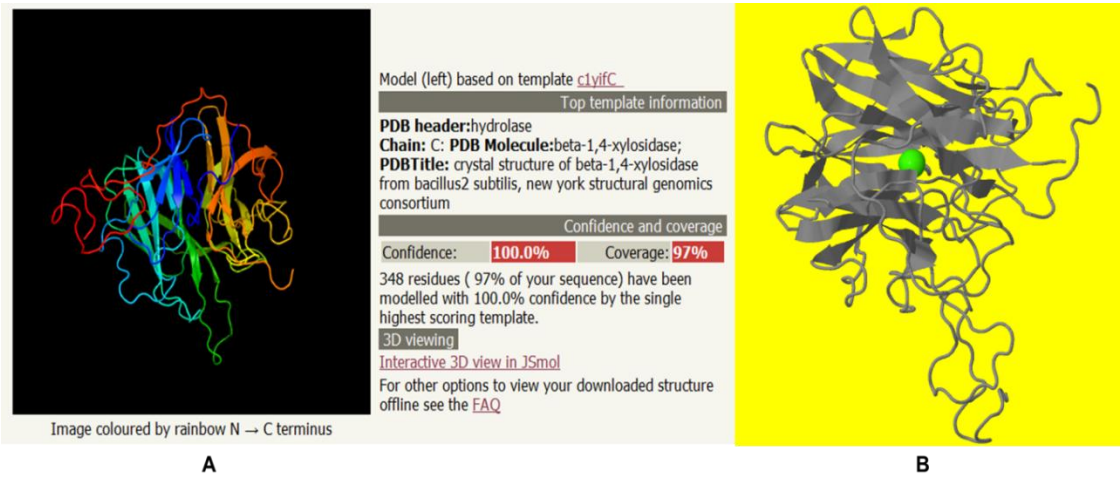
phân lignocellulose khác nhau như β-xylosidase (EC 3.2.1.37); α-L-arabinofuranosidase (EC 3.2.1.55),…. Dựa trên đánh giá tương đồng cho thấy mã gen này có độ bao phủ axit amin cao với xylan 1,4 – beta-xylosidase của nhiều loài vi khuẩn như: *Paenibacillus sp. NFR01* (ID: SET10311.1), *Thalassospira sp. KO164* (ID: OCK06431.1), *Fictibacillus solisalsi* (ID: SDN26714.1), *Bacillus daliensis* (ID: SDO23417.1), *Treponema primitia* (ID: WP_010255887.1), *Treponema succinifaciens* (Mã: WP_013700334.1), *Treponema sp. C6A8* (ID: WP_027728697.1),…. Tuy nhiên độ tương đồng của axit amin so với các trình tự trong NCBI cao nhất chỉ có 68%. Do đó nghiên cứu này có thể sẽ tìm ra một gen mã hóa *Xbx14* mới so với các trình tự gen đã được công bố.

Hình 3.6. Kết quả dự đoán chức năng Xbx14 bằng BLASTP

3.2.8.2. Kết quả dự đoán cấu trúc không gian của Xbx14 bằng PHYRE 2 (Protein Homology/analogY Recognition Engine)

Phân tích cấu trúc bậc hai *Xbx14* được mã hóa bởi mã gen lựa chọn có 41% là các chuỗi gấp nếp β, 1% là chuỗi α và 62% trình tự axit amin không xác định

được (Hình 3.7A). Mô hình cấu trúc không gian cho thấy các chuỗi axit amin của *Xbx14* cuộn gấp lại dạng hình cầu, có độ tương đồng và bao phủ cao nhất là 97% so với cấu trúc beta-xylosidase của *Bacillus 2 subtillis* (Hình 3.7B). Protein này có độ dài là 533 axit amin và trọng lượng phân tử là 61,41 Kda, được biểu hiện trong tế bào *E. coli* BL21(DE3). Trung tâm hoạt động xúc tác thủy phân của cả hai enzyme là beta - xylosidase và alpha-L- arabinanase [83]. Cấu trúc *Xbx14* cũng tương đồng cao (trên 50%) với beta-xylosidase của vi khuẩn *Bacillus halodurans c-125* [179], *Geobacillus stearothermophilus* và *Clostridium acetobutylicum* [26].

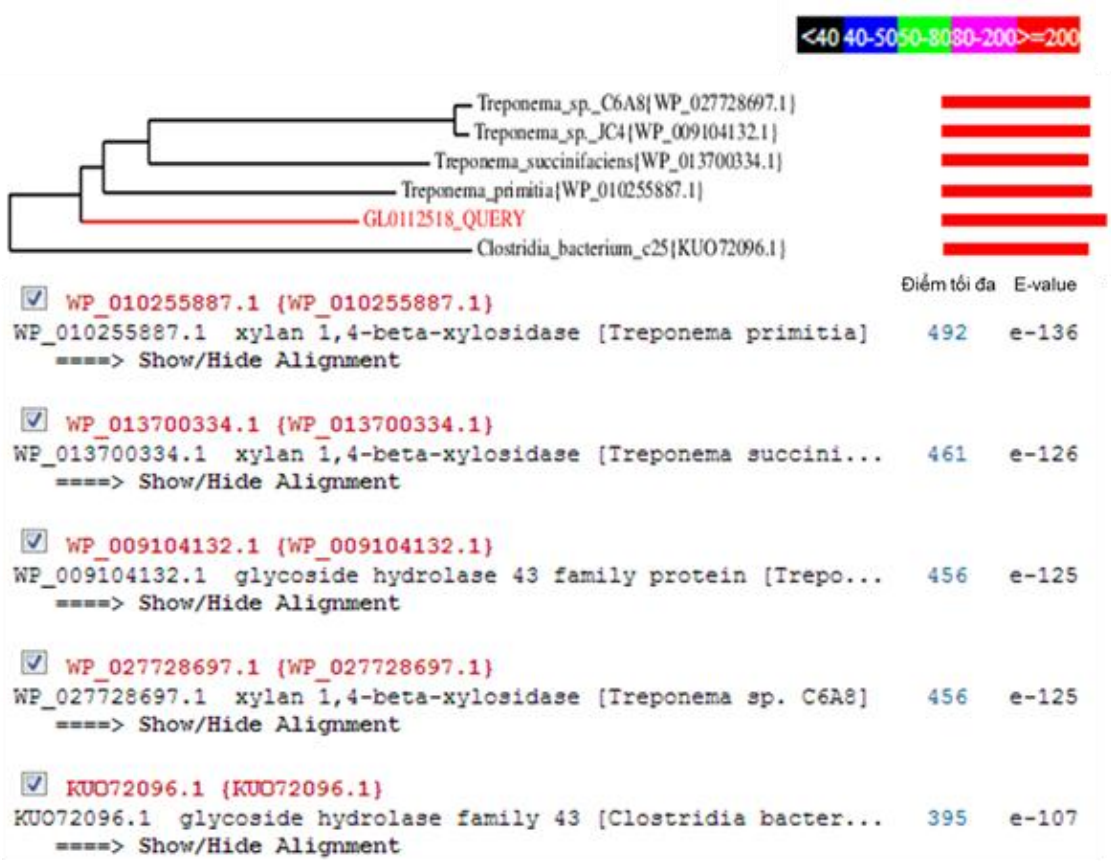


Hình 3.7. Mô hình cấu trúc không gian (A) dựa vào khuôn c1ylfC và trung tâm hoạt động (B) của enzyme Xbx14

Nghiên cứu chi tiết về cơ chế hoạt động của enzyme này từ vi khuẩn *Geobacillus stearothermophilus T-6* (XynB3) cho thấy: Mỗi tiểu đơn vị của XynB3 gồm năm chuỗi beta tạo thành một cái túi, được đóng kín tại đầu N bằng một chuỗi beta khác. Đây chính là vùng xúc tác của enzyme. Cấu trúc này có tác dụng giới hạn độ dài cơ chất và enzyme phân cắt cơ chất lần lượt từ ngoài vào trong. Axit amin Asp ở vị trí 128 có vai trò điều chỉnh pK và giữ cho trung tâm hoạt động định hướng chính xác ở bề mặt. Mặt khác Asp128 giữ vai trò liên kết cơ chất với trung tâm hoạt động thông qua vị trí 2-O của glycon. Đây là lý do giải thích vai trò quan trọng của Asp128 trong sự ổn định quá trình chuyển đổi trạng thái và là một bất

biến trong trung tâm hoạt động [26]. Dựa trên dự đoán cấu trúc không gian cho thấy cả hai đầu N và C của protein đều không bị cuộn vào trong cấu trúc mà nằm lộ ra ngoài. Với đặc điểm này sẽ giúp cho việc tinh sạch protein tái tổ hợp khi gắn đuôi His – tag sẽ thuận lợi.

3.2.8.3. Kiểm tra nguồn gốc của mã gen GL0112518 bằng công cụ Blast-Explorer



Hình 3.8. Sơ đồ nguồn gốc mã gen GL0112518

Sử dụng công cụ Blast-Explorer của Phylogeny.fr để kiểm tra nguồn gốc của mã gen GL0112518. Kết quả cho thấy mã gen này nằm trong nhánh phát sinh của vi khuẩn *Treponema primitia* (Hình 3.8). Đây chính là nhóm vi khuẩn có số lượng ORF lớn nhất trong ruột mối *C. gestroi* (Hình 3.1).

Tóm lại sử dụng BLASTP, chúng tôi dự đoán mã gen GL0112518 có chiều dài là 1077 bp, mã hóa đúng β–xylosidase. Enzyme này được dự đoán có cấu trúc không gian dạng hình cầu và được hình thành chủ yếu từ các chuỗi beta bằng công

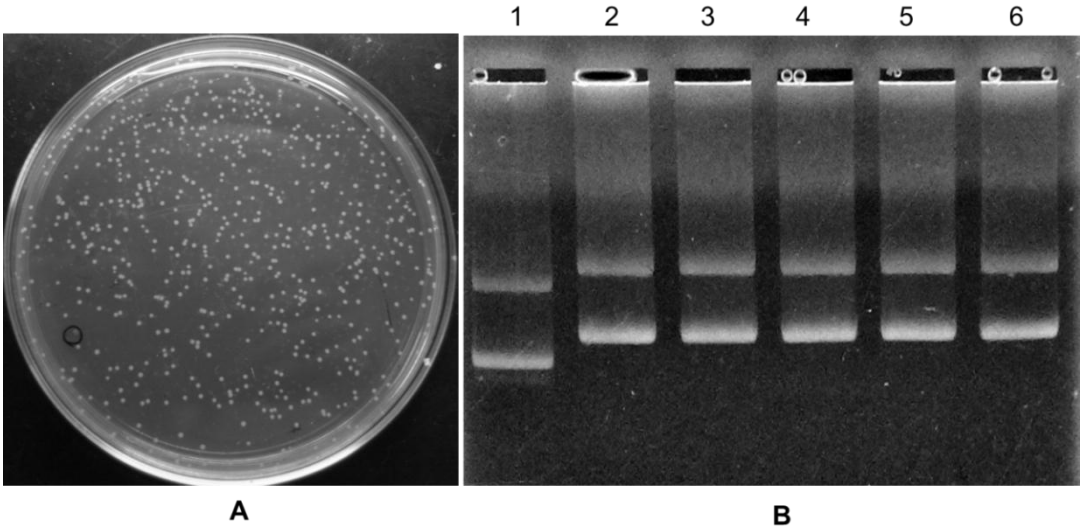
cụ Phyre2. Sử dụng công cụ kiểm tra nguồn gốc Blast-Explorer cho thấy, gen chọn tương đồng cao nhất với trình tự gen từ vi khuẩn *Treponema primitia* – vi khuẩn có số lượng ORF lớn nhất trong ruột mối *C. gestroi*.

3.3. BIỂU HIỆN GEN *Xbx14* VÀ NGHIÊN CỨU TÍNH CHẤT CỦA β-XYLOSIDASE

3.3.1. Biểu hiện gen *Xbx14*

3.3.1.1. Kiểm tra gen Xbx14 trong vector pET22b(+)

Mã gen GL0112518 được chọn bằng mẫu dò được đặt tổng hợp và gắn sẵn vào vector biểu hiện pET22b(+). Sau đó được nhân dòng trong tế bào *E. coli* DH10B, kiểm tra bằng enzyme hạn chế, phản ứng PCR và giải trình tự.

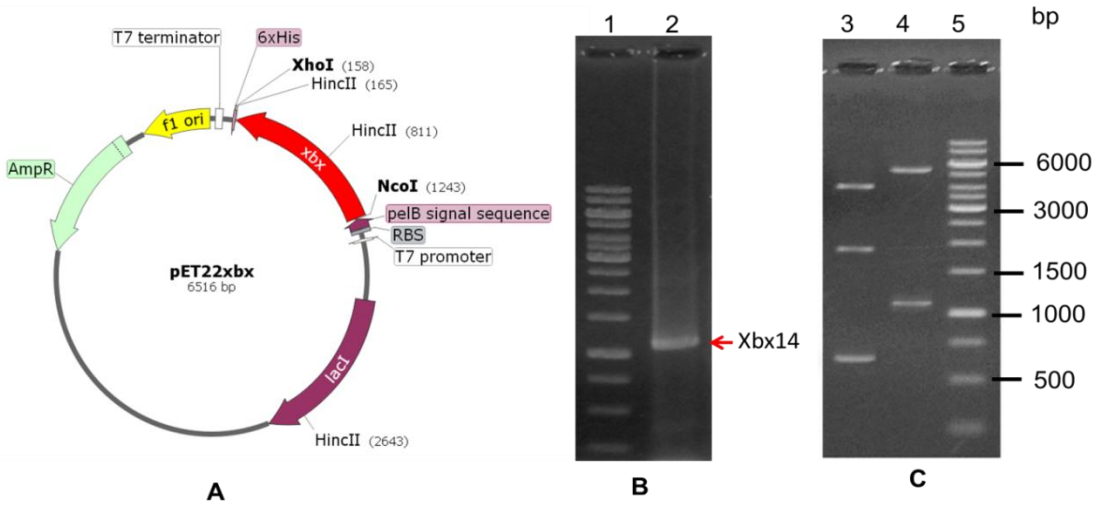


Hình 3.9. Kết quả biến nạp vector* pET22Xbx *(A) và điện di đồ phân tích sản phẩm tách dòng gen Xbx14 (B). *Đường chạy 1: Plasmid* pET22b(+)*; đường chạy 2-6: Plasmid* pET22Xbx

Để thuận lợi cho quá trình phân lập, ghép nối gen vào vector biểu hiện, chúng tôi đã khảo sát trình tự enzyme hạn chế có trong gen bằng phần mềm trực tuyến Restriction mapper3. Kết quả cho thấy gen *Xbx14* không chứa trình tự cắt của enzyme hạn chế *Nco*I và *Xho*I. Mặt khác trong vùng đa nối của vector biểu hiện pET22b(+) có vị trí cắt của hai enzyme này. Do đó hai đầu mồi xuôi và mồi ngược được thêm trình tự nhận biết của enzyme *Nco*I và *Xho*I phục vụ cho việc kiểm tra

gen. Kết quả biến nạp vector pET22Xbx vào tế bào DH10B và tách plasmid để kiểm tra như Hình 3.9.

Sản phẩm biến nạp gen và trải trên đĩa LBA cho thấy các khuẩn lạc mọc tròn, đều và đặc trưng của vi khuẩn (Hình 3.9A). Phân tích bằng điện di sản phẩm tách plasmid 05 khuẩn lạc trên đĩa biến nạp pET22Xbx cho thấy, tất cả đều chạy cao hơn so với vector đối chứng không mang gen (Hình 3.9B). Kết quả này chứng tỏ vector pET22b(+) có đoạn DNA chèn vào vùng đa nối, nên kích thước lớn hơn và sẽ chạy chậm hơn trên hình ảnh.



Hình 3.10. Sơ đồ vị trí cắt của enzyme hạn chế trên vector* pET22Xbx *(A), điện di sản phẩm PCR (B) và sản phẩm cắt* pET22Xbx *bằng Nco*I *và Xho*I*, Hinc*II *(C)

Đường chạy 1, 5: Thang DNA chuẩn (Fermentas), đường chạy 2: Sản phẩm PCR gen Xbx14, đường chạy 3: cắt pET22Xbx *Hinc*II*, đường chạy 4: cắt* pET22Xbx *bằng Nco*I *và Xho*I

Nghiên cứu tiến hành kiểm tra gen *Xbx14* trong vector pET22Xbx bằng phản ứng PCR, cắt kiểm tra bằng enzyme hạn chế và giải trình tự gen. Chúng tôi chọn một dòng (đường chạy 3 của hình 3.9B) trong 05 dòng plasmid chạy cao hơn đối chứng để làm khuôn cho PCR và cắt kiểm tra bằng enzyme hạn chế. Sơ đồ enzyme cắt hạn chế trên vector pET22Xbx được kiểm tra bằng phần mềm trực tuyến tại trang web http://www.restrictionmapper.org/. Kết quả cho thấy enzyme *Xho*I, *Nco*I chỉ có 1 điểm cắt trên vector, còn *Hinc*II có 3 điểm cắt trên vector (Hình 3.10A). Như vậy khi được xử lý bằng cặp enzyme cắt hạn chế *Xho*I và *Nco*I, vector sẽ được

cắt thành 2 đoạn, một đoạn có kích thước khoảng 1,1 kb tương đương với kích thước của gen *Xbx14* và đoạn thứ hai ở vị trí khoảng 5,5 kb tương đương với kích thước của vector pET22b(+). Khi xử lý với enzyme cắt hạn chế *Hinc*II sẽ thu được 3 đoạn có kích thước khoảng 4 kb, 1,8 kb và 0,6 kb.

Kết quả PCR bằng cặp mồi đặc hiệu cho duy nhất một băng sáng, sắc nét và tương đương với kích thước của gen *Xbx14* (Hình 3.10B). Khi sử dụng enzyme cắt riêng rẽ *Hinc*II cho ra ba băng với kích thước lần lượt là 4 kb, 1.8 kb và 0.6 kb, cắt đồng thời bằng *Nco*I và *Xho*I cho ra hai băng sáng rõ với kích thước khoảng 1,1 kb và 5,5 kb đúng như tính toán lý thuyết (Hình 3.10C). Kết quả cắt dòng plasmid số 3 bằng các enzyme hạn chế trên hoàn toàn chính xác với tính toán lý thuyết, chứng tỏ gen *Xbx14* đã được gắn vào vector pET22b(+).

Tuy nhiên để kiểm tra chính xác trình tự của gen *Xbx14* trong vector pET22Xbx, chúng tôi tiến hành giải trình tự gen *Xbx14* trong plamid dòng số 3 bằng cặp mồi đặc hiệu. Kết quả sau khi giải trình tự cho thấy trình tự gen *Xbx14* gồm 1093 nucleotide, trong đó có 16 nucleotide là trình tự thêm vào trong mồi. Phần còn lại gồm 1077 bp tương đồng 100% so với trình tự gen mã số Termite 2_ GL0112518. Như vậy có thể khẳng định trình tự gen mã hóa *Xbx14* đã được gắn vào vector pET22b(+) đúng như trình tự gốc ban đầu *(Phụ lục 2).*

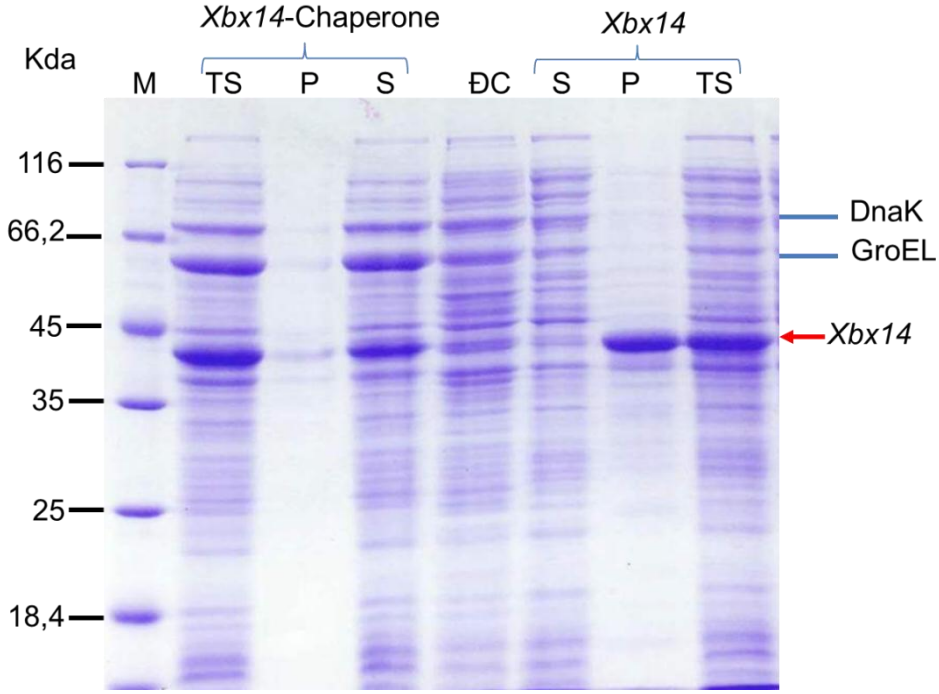
3.3.1.2. Đồng biểu hiện gen mã hóa Xbx14 với chaperone trong E. coli Rosetta 1

Mỗi gen ngoại lai có ảnh hưởng khác nhau tới sự sinh trưởng, trao đổi chất của chủng chủ. Thông thường những sản phẩm của gen có lợi hoặc phù hợp tốt với chủng chủ sẽ có xu thế biểu hiện tốt. Ngược lại, sản phẩm của gen gây độc và có ảnh hưởng không tốt tới sự sinh trưởng cũng như trao đổi chất của tế bào thì không được biểu hiện, hoặc biểu hiện yếu, hoặc ở dạng không hoạt động [139]. Do đó, không có một quy luật chung nào cho phép biểu hiện tối ưu nhiều loại protein khác nhau trong một hệ biểu hiện. Vì vậy, để thu được lượng lớn protein ngoại lai *Xbx14* ở dạng tan, chúng tôi tiến hành khảo sát các yếu tố ảnh hưởng đến quá trình biểu hiện, để tìm ra điều kiện tối ưu nhất.

Vector tái tổ hợp pET22Xbx được biến nạp vào chủng biểu hiện *E. coli Rosetta 1* để khảo sát biểu hiện. Sau khi nuôi cấy ở 37°C, OD_{600} đạt từ 0,6 đến 0,8,

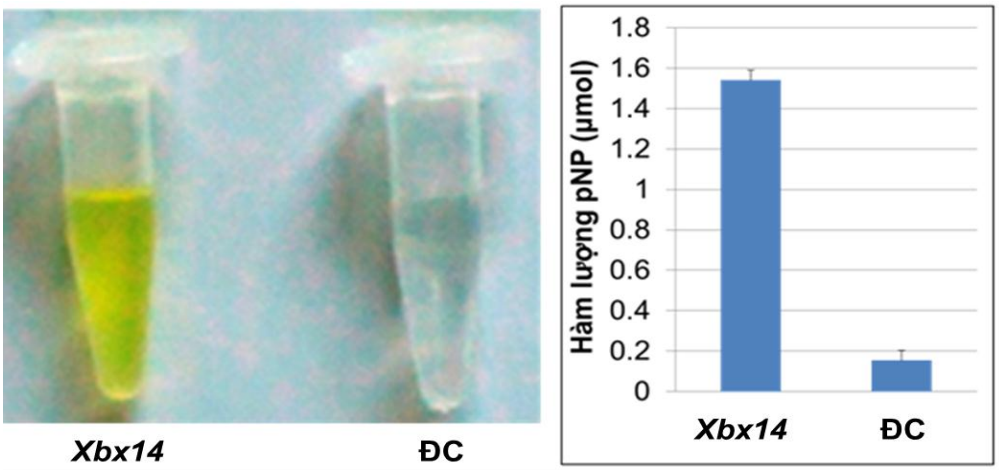
mẫu được chia thành hai phần, một phần cảm ứng IPTG và một phần không cảm ứng làm đối chứng. Kết quả protein *Xbx14* biểu hiện rất tốt nhưng không tan. Sau đó chúng tôi tiếp tục tối ưu các điều kiện biểu hiện như nhiệt độ, nồng độ chất cảm ứng, thời điểm cảm ứng, thời gian thu mẫu,… nhưng không cải thiện được khả năng tan của *Xbx14*. Một trong những nguyên nhân có thể do protein tái tổ hợp *Xbx14* được sản xuất trong *E. coli* không tạo ra cấu trúc không gian chính xác. Trong một số trường hợp, sự biểu hiện các phân tử chaperone tạo thuận lợi cho sản xuất protein và quá trình cuộn gấp cấu trúc đúng của protein. Vì vậy nghiên cứu này tiếp tục tiến hành đồng biểu hiện gen *Xbx14* trong vector pET22Xbx với các chaperone GroEL (60 kDa), DnaK (70 kDa) và DnaJ (40 kDa) trong vector pG - KJE8. Phức hệ DnaK-DnaJ-GrpE tham gia vào các hoạt động khác nhau như ngăn ngừa sự kết tập protein mới tổng hợp, giúp cho quá trình cuộn gấp đúng khi hình thành cấu trúc không gian, sắp xếp lại các vị trí bị đứt quãng và phân tổ protein, do đó tăng cường độ hòa tan của protein tái tổ hợp. Chaperone GroES-GroEL là protein thiết yếu cho sự tăng trưởng của vi khuẩn ở mọi nhiệt độ, phân hủy nhanh protein kết tập và hỗ trợ sản xuất protein có kích thước lớn [109]. Ngoài ra các chaperone này còn tham gia vào quá trình chuyển vị protein qua màng, chế biến RNA ribosome,….

Để tạo được chủng *E. coli Rosetta 1* đồng biểu hiện protein *Xbx14* và chaperone, đầu tiên chúng tôi biến nạp plasmid pG - KJE8 vào tế bào *E. coli Rosetta 1*. Tế bào sau khi biến nạp được chọn lọc trên môi trường LB có Chloramphenicol nồng độ 20 µg/ml để tạo chủng *E. coli Rosetta 1* mang pG - KJE8. Chủng tái tổ hợp này sẽ được sử dụng làm tế bào khả biến. Sau đó tiếp tục biến nạp pET22Xbx vào và chọn lọc trên môi trường LBAC chứa 2 loại kháng sinh là Ampicillin nồng độ 100 µg/ml và Chloramphenicol với nồng độ 20 µg/ml. Mẫu tế bào sau khi đồng biểu hiện được thu lại và kiểm tra kết quả. Điện di đồ cho thấy khi tiến hành đồng biểu hiện với chaperone thì cả DnaK, DnaJ, GroEL và *Xbx14* đều xuất hiện ở dạng tan với băng to, đậm và đúng kích thước lý thuyết. Mặt khác protein *Xbx14* không xuất hiện ở dòng đối chứng không cảm ứng IPTG. So sánh với sản phẩm biểu hiện gen *Xbx14* không có chaperone, thì protein *Xbx14* hầu hết nằm ở pha tủa (Hình 3.11). Điều này chứng tỏ đồng biểu hiện với chaperone đã giúp cho *Xbx14* đã biểu hiện ở dạng tan.



Hình 3.11. Điện di đồ phân tích sản phẩm protein đồng biểu hiện gen Xbx14 với chaperone trong E. coli Rosetta 1 trên gel SDS-PAGE 12,6% có SDS

M: Thang protein chuẩn (Fermentas), ĐC: Protein tổng số không cảm ứng IPTG, Xbx14: protein Xbx14 biểu hiện không có chaperone, Xbx14-Chaperone: protein Xbx14 đồng biểu hiện với chaperone pG-KJE8, TS: protein ở dạng tổng số, S: protein ở dạng tan, P: protein ở dạng tủa



Hình 3.12. Kết quả thử hoạt tính thô của Xbx14

Kết quả này phù hợp với hầu hết nghiên cứu đồng biểu hiện enzyme chuyển hóa sinh khối thực vật với chaperone như biểu hiện beta-galactosidase tái tổ hợp

cùng với hệ thống GroESL ở 30°C tan rất tốt, với phức hệ chaperone KJE protein tái tổ hợp hòa tan ở mọi nhiệt độ thử [119]. Đồng biểu hiện chaperone với xylanase đã làm tăng độ tan trong nước từ 33% đến 40% [84]. Kết quả này được giải thích là do trong tế bào chất, mỗi loại chaperone có thể giúp cuộn gấp chính xác một số loại protein xác định nên làm tăng khả năng tan của protein [139]. Đồng thời sự có mặt của chaperone cũng giúp tăng sinh khối của vi khuẩn *E. coli* ở nhiệt độ thấp [139].

Sau khi đồng biểu hiện gen *Xbx14* ở dạng tan, chúng tôi tiến hành kiểm tra hoạt tính thô bằng cơ chất đặc hiệu pNPX của β-xylosidase theo phương pháp của Teng và cộng sự. Kết quả kiểm tra cho thấy ống đối chứng (ĐC) không có màu vàng đặc trưng của pNP do sự xúc tác chuyển hóa của β-xylosidase như ống có protein *Xbx14* tái tổ hợp (Hình 3.12). Như vậy sản phẩm đồng biểu hiện *Xbx14* với chaperone giúp protein tái tổ hợp tan và có hoạt tính. Do đó chúng tôi sẽ chọn đồng biểu hiện gen *Xbx14* với chaperone để phục vụ cho các nghiên cứu tiếp theo.

3.3.1.3. Tối ưu điều kiện biểu hiện gen Xbx14

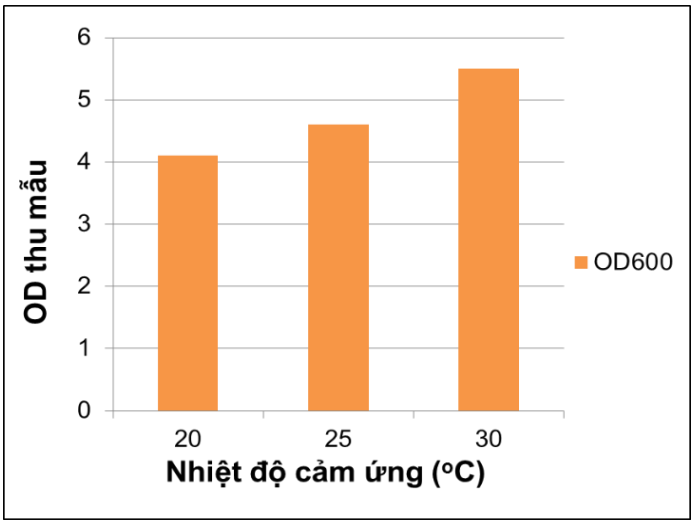
a. Nghiên cứu ảnh hưởng của nhiệt độ đến biểu hiện gen Xbx14 trong tế bào E. coli Rosetta 1

Nhiệt độ nuôi cấy là yếu tố hàng đầu ảnh hưởng đến hiệu suất và chất lượng protein. Thông thường tế bào *E. coli* sinh trưởng tốt nhất ở 37°C và quá trình sinh tổng hợp protein cũng diễn ra rất mạnh. Vì vậy khi biểu hiện gen ngoại lai ở nhiệt độ này, chủng chủ sẽ ồ ạt sản xuất một lượng lớn protein. Điều này cũng đồng nghĩa là một lượng lớn protease sẽ được tổng hợp và sẽ phân cắt protein ngoại lai. Mặt khác khi sinh trưởng ở tốc độ cao, các bản sao plasmid trong tế bào giảm cũng sẽ làm giảm chất lượng của sản phẩm protein tạo ra. Do đó nhiệt độ tối ưu cho biểu hiện trong tế bào *E. coli* thường thấp hơn 37°C sẽ cải thiện độ hòa tan của protein tái tổ hợp [158]. Ở nhiệt độ thấp thời gian sinh trưởng và tổng hợp protein chậm lại, giúp cho protein cuộn xoắn đúng cấu trúc nên giảm sự kết tập protein [90]. Hơn nữa hầu hết các protease hoạt động kém tích cực ở nhiệt độ thấp, nên làm giảm khả năng phân hủy protein ngoại lai. Việc lựa chọn nhiệt độ thích hợp đóng vai trò quan trọng trong việc thu sinh khối tế bào và sản lượng protein tái tổ hợp dạng tan.

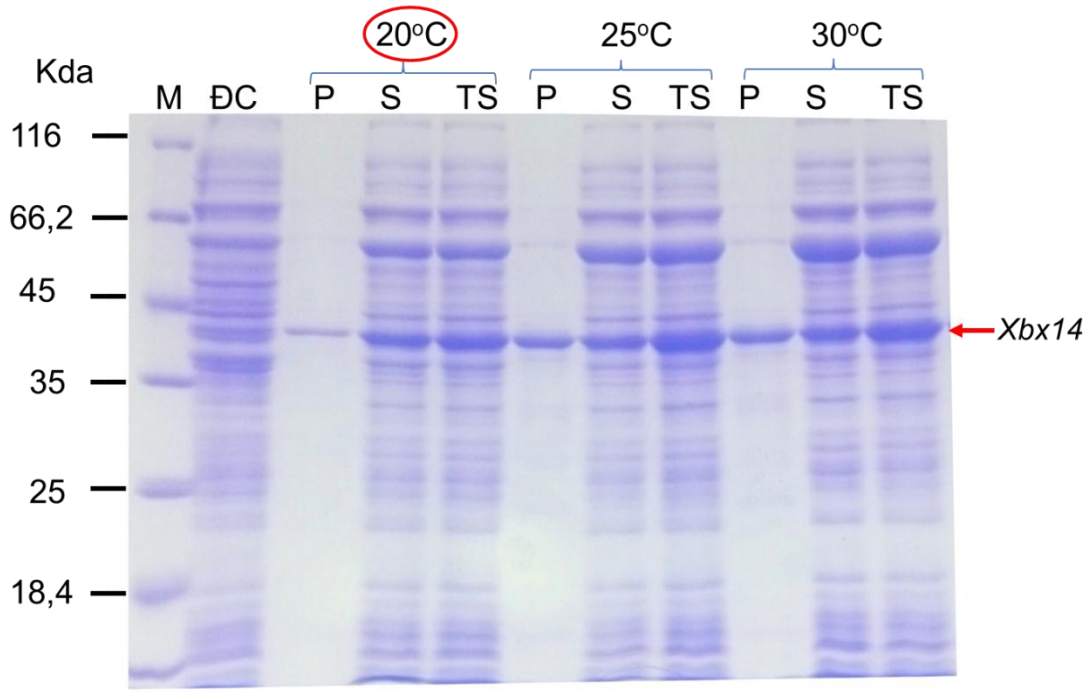
Kỹ thuật hạn chế protein tái tổ hợp không tan là nuôi cấy ở nhiệt độ thấp [147] và đã chứng minh thông qua cải thiện độ tan của một số protein như interferon α-2 của người, subtilisin E, chuỗi ricin A, bacterial luciferase, Fab fragments, β-lactamase, lipoxygenase L-2 của lúa gạo, lypoxygenase L-1 của đậu nành, kanamycin nuclotidyltransferase và glycogen phosphorylase của thỏ [158],…. Mặt khác khi protein mục tiêu tăng lên, thì các protein nội sinh của vi khuẩn giảm, sẽ giúp cho quá trình tinh chế protein dễ dàng và hiệu quả hơn [145].

Trong nghiên cứu chúng tôi lựa chọn ba nhiệt độ thấp hơn nhiệt độ sinh trưởng của *E. coli* là 20°C, 25°C và 30°C để kiểm tra khả năng biểu hiện của *Xbx14*. Mặt khác theo khuyến cáo của nhà cung cấp vector pG-KJE8 thì nhiệt độ thích hợp để các chaperone biểu hiện ở dạng tan là dưới 25°C. Chủng biểu hiện *E. coli Rosetta 1* mang hai vector pG-KJE8 và Pet22Xbx được nuôi sau khi cảm ứng IPTG sẽ tiếp tục nuôi lắc trong 5 giờ ở ba nhiệt độ trên. Mẫu tế bào thu được sẽ đo OD_{600} để so sánh sinh khối thu được cuối cùng.

Kết quả cho thấy mật độ tế bào thu được sau 5 giờ cảm ứng cao nhất ở 30°C và thấp nhất ở 20°C (Hình 3.13). Như vậy ở nhiệt độ càng cao thì sinh khối tế bào thu được càng lớn là hoàn toàn phù hợp với quy luật sinh trưởng của *E. coli*. Tế bào thu được sẽ tiếp tục phá vỡ bằng siêu âm để kiểm tra khả năng tan của protein tái tổ hợp. Kết quả điện di đồ cho thấy protein *Xbx14* được tạo thành có cả dạng tan và dạng không tan ở cả 20°C, 25°C và 30°C (Hình 3.14). Mặc dù lượng protein tái tổ hợp tổng số *Xbx14* được tạo ra ở 30°C là cao nhất, nhưng lượng protein *Xbx14* dạng tan ở 20°C là nhiều hơn. Mặt khác theo khuyến cáo của hãng cung cấp vector *pG-KJE8,* nhiệt độ để biểu hiện chaperone không nên vượt quá 25°C. Như vậy để lượng chaperone tạo ra ở dạng tan một cách ổn định trong tế bào *E. coli Rosetta 1,* giúp cho việc hình thành *Xbx14* ở dạng tan tốt hơn, chúng tôi sẽ ưu tiên chọn nhiệt độ thấp hơn hoặc bằng 25°C. Căn cứ vào kết quả lượng protein *Xbx14* tan ở các nhiệt độ trên, chúng tôi sẽ lựa chọn nhiệt độ 20°C để tiếp tục biểu hiện trong thí nghiệm tiếp theo.



Hình 3.13. Mật độ tế bào biểu hiện E. coli Rosetta 1 (DE3) mang* pET22Xbx *cảm ứng IPTG ở các nhiệt độ 20°C, 25°C và 30°C



Hình 3.14. Điện di đồ phân tích sản phẩm biểu hiện Xbx14 trong E. coli Rosetta 1 sau 6 giờ cảm ứng 0,5 mM IPTG ở 20°C, 25°C, 30°C trên gel polyacrylamide 12,6% có SDS

Đường chạy ĐC: đối chứng âm, protein tổng số của dòng tế bào biểu hiện mang vector pET22Xbx *không cảm ứng IPTG, đường chạy TS: protein tổng số biểu hiện trong tế bào mang* pET22Xbx, *đường chạy S: protein dạng tan biểu hiện trong*

tế bào mang pET22Xbx, *đường chạy P: protein dạng không tan biểu hiện trong tế bào mang* pET22Xbx, *đường chạy M: protein chuẩn unstained (Thermo scientific)*

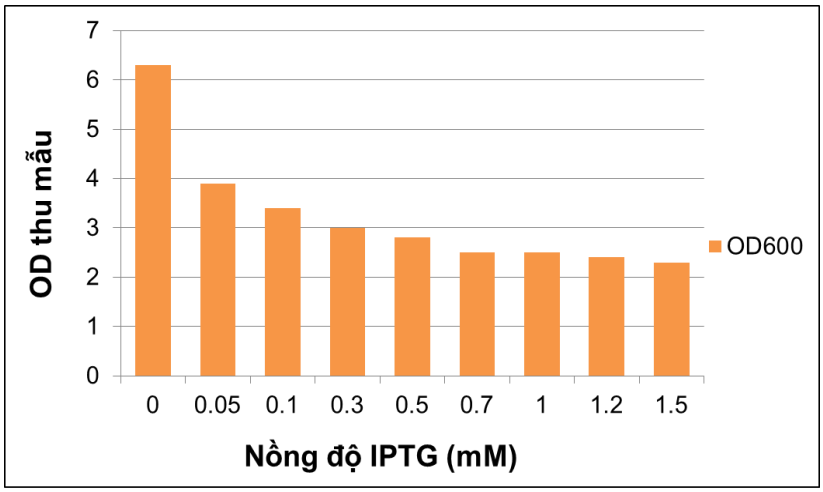
b. Nghiên cứu ảnh hưởng của nồng độ IPTG đến hiệu quả biểu hiện Xbx14

IPTG là chất cảm ứng tổng hợp protein *Xbx14* dưới sự điều khiển của promoter T7 trong vector pET22Xbx. Trong môi trường không có chất cảm ứng, gen *LacI* nằm trên hệ gen vi khuẩn được phiên mã tổng hợp protein *LacI* - chất ức chế có ái lực cao với operon *lac* trên vector biểu hiện. Khi được cảm ứng bằng IPTG, *LacI* không còn ái lực với operator *lac*, sẽ rời khỏi phức hợp operator trên DNA của chủng chủ và plasmid. Khi đó promoter *Lac* hoạt động, khởi động sinh tổng hợp T7 RNA polymease và chính enzyme này sẽ bám vào vị trí T7 promoter để khởi đầu quá trình sinh tổng hợp protein ngoại lai. Nồng độ IPTG cảm ứng ảnh hưởng đến hiệu quả biểu hiện gen ngoại lai. IPTG được sử dụng như một yếu tố trung hòa chất ức chế, do đó không cần bổ sung một lượng quá nhiều trong quá trình cảm ứng. Ngoài ra IPTG còn là chất gây độc cho tế bào, ở nồng độ cao sẽ ức chế sự sinh trưởng của vi khuẩn [53]. Mặt khác giá thành của IPTG khá cao, nên việc xác định nồng độ IPTG phù hợp là rất cần thiết cho quá trình sản xuất protein tái tổ hợp ở quy mô lớn.

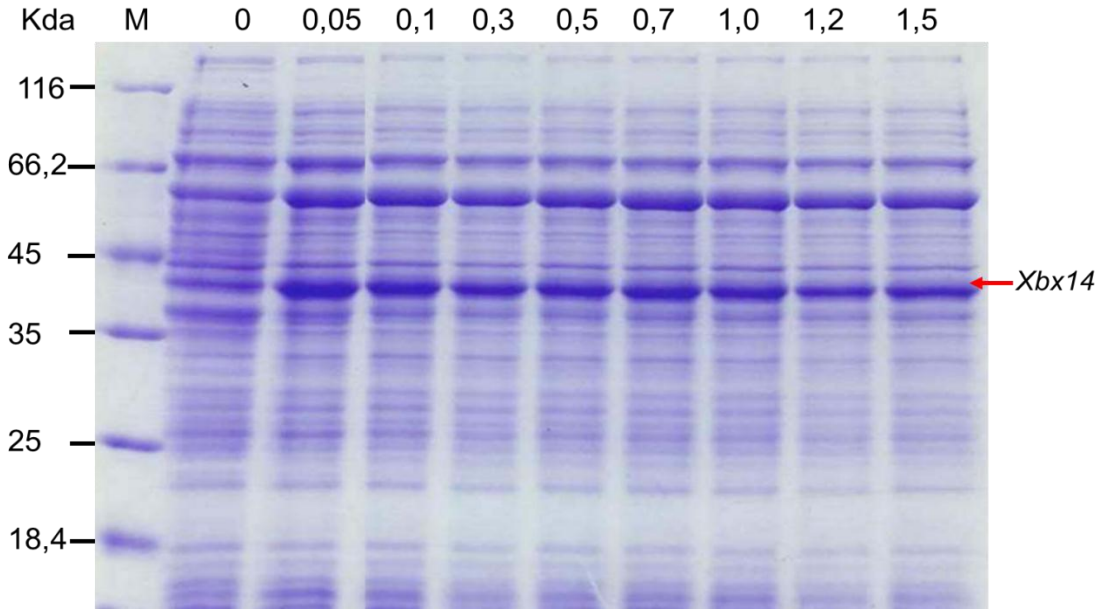
Để kiểm tra ảnh hưởng của nồng độ IPTG cảm ứng đến hiệu quả biểu hiện gen *Xbx14* trong tế bào *E. coli Rosetta 1* ở 20^oC, chúng tôi đã chọn 9 nồng độ IPTG cảm ứng là 0 mM; 0,05 mM; 0,1 mM; 0,3 mM; 0,5 mM; 0,7mM; 1mM; 1,2 mM và 1,5 mM. Mẫu tế bào sau khi nuôi đến OD_{600} = 1 được chia thành 9 phần bằng nhau và bổ sung IPTG theo dải nồng độ trên. Tất cả mẫu đã cảm ứng được nuôi lắc tiếp ở 20^oC trong thời gian 5 giờ. Sau đó thu tế bào và đo OD_{600} để so sánh sinh khối cuối cùng.

Kết quả cho thấy nồng độ IPTG càng tăng từ 0,0 mM đến 1,5 mM thì sinh khối tế bào thu được giảm dần (Hình 3.15). Đặc biệt ngay sau khi cảm ứng IPTG sinh khối tế bào với nồng độ IPTG rất thấp là 0,05 mM, sinh khối tế bào đã giảm gần một nửa. Điều đó chứng tỏ sự biểu hiện của gen đã làm ảnh hưởng tới quá trình trao đổi chất của chủng chủ. Hiệu quả biểu hiện protein *Xbx14* phụ thuộc đồng thời cả lượng tế bào thu được và lượng protein tái tổ hợp ở dạng tan. Do đó mẫu tế bào

cảm ứng với nồng độ IPTG khác nhau sẽ được phá vỡ bằng siêu âm để kiểm tra mức độ biểu hiện và dạng tan của *Xbx14*. Kết quả cảm ứng biểu hiện *Xbx14* tổng số ở nồng độ IPTG khác nhau được kiểm tra bằng điện di trên gel polyacrylamide 12,6% có SDS (hình 3.16).



Hình 3.15. Mật độ tế bào biểu hiện E. coli Rosetta 1 (DE3) mang* pET22Xbx *cảm ứng IPTG ở các nồng độ từ 0,0 đến 1,5 mM

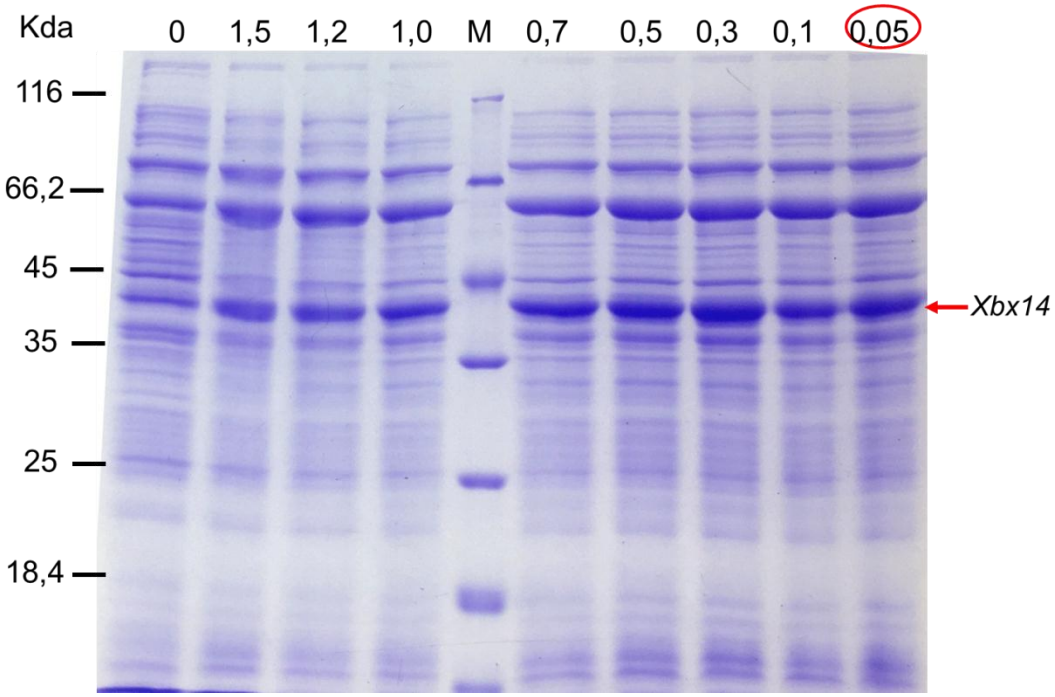


Hình 3.16. Điện di đồ phân tích protein biểu hiện Xbx14 tổng số trong tế bào E. coli Rosetta 1 cảm ứng 09 nồng độ IPTG trên gel polyacrylamide 12,6 % có SDS

Đường chạy 0,0 – 1,5: là protein tổng số biểu hiện khi tế bào được cảm ứng nồng độ IPTG từ 0,0 đến 1,5 mM, đường chạy M: protein chuẩn (Thermo scientific)

Từ điện di đồ kiểm tra protein tổng số cho thấy ở nồng độ IPTG là 0,05 mM lượng protein tái tổ hợp *Xbx14* thu được nhiều nhất. Khi nồng độ IPTG cảm ứng tăng dần thì lượng *Xbx14* tổng số không thay đổi nhiều và giảm xuống ở nồng độ IPTG cao là 1,5 mM. Kết quả này có thể do khi cảm ứng nồng độ IPTG cao tế bào thích nghi kém nên sinh trưởng chậm và kèm theo đó lượng protein tạo ra cũng giảm xuống [173].

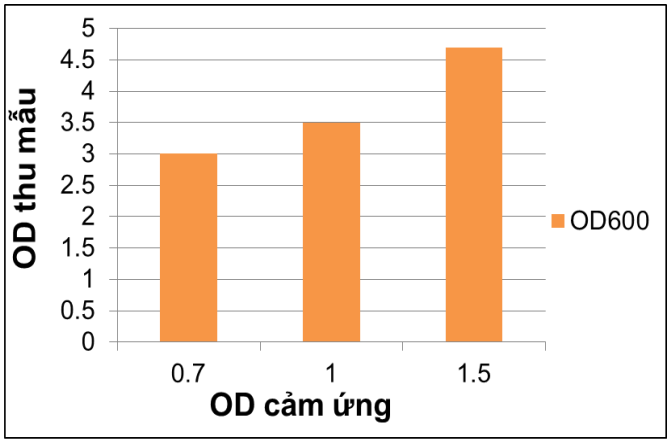
Lượng *Xbx14* dạng tan ở các nồng độ IPTG cảm ứng khác nhau cũng tỷ lệ thuận với lượng protein *Xbx14* tổng số (Hình 3.16, Hình 3.17). Lượng protein tan giảm mạnh chủ yếu là do lượng tế bào trong mẫu ít hơn. Do đó để thu được sinh khối tế bào và lượng protein *Xbx14* nhiều nhất, cũng như tỷ lệ protein dạng tan cao nhất thì nồng độ IPTG cảm ứng thích hợp là 0,05 mM.



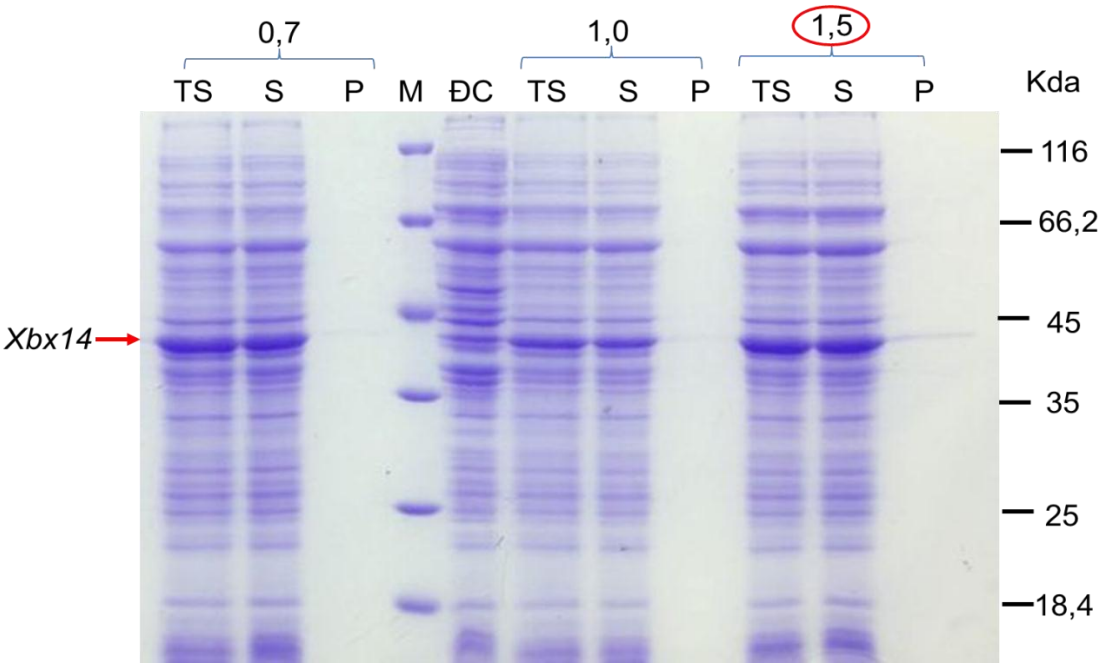
Hình 3.17. Điện di đồ phân tích sản phẩm biểu hiện Xbx14 tan trong tế bào E. coli Rosetta 1 (DE3) cảm ứng 09 nồng độ IPTG trên gel polyacrylamide 12,6% có SDS

Các đường chạy 0,0 – 1,5: là protein tổng số biểu hiện khi tế bào được cảm ứng nồng độ IPTG từ 0,0 đến 1,5 mM, đường chạy M: protein chuẩn (Thermo scientific)

c. Nghiên cứu ảnh hưởng của thời điểm cảm ứng đến hiệu quả biểu hiện protein Xbx14



Hình 3.18. Mật độ tế bào biểu hiện E. coli Rosetta 1 (DE3) mang* pET22Xbx *cảm ứng ở các OD_{600} khác nhau



Hình 3.19. Điện di đồ phân tích sản phẩm biểu hiện của Xbx14 trong tế bào E. coli Rosetta 1 (DE3) cảm ứng IPTG tại các OD_{600} = 0,7; 1,0; 1,5 trên gel polyacrylamide 12,6% có SDS

Đường chạy ĐC: đối chứng âm, protein tổng số của dòng tế bào biểu hiện không cảm ứng IPTG, đường chạy TS: protein tổng số biểu hiện trong tế bào mang

pET22Xbx, *đường chạy S: protein dạng tan, đường chạy P: protein dạng không tan, đường chạy M: protein chuẩn unstained (Thermo scientific)*

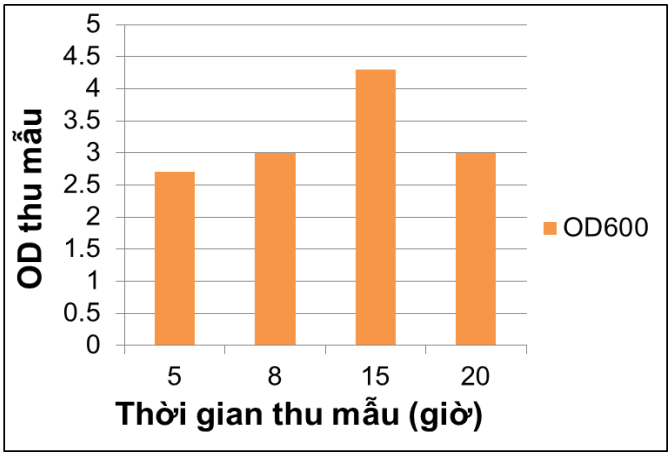
Giai đoạn sinh trưởng của tế bào ảnh hưởng lớn đến sự tổng hợp và khả năng hoạt động của protein tái tổ hợp. Mỗi loại protein tái tổ hợp biểu hiện trong *E. coli* sẽ có thời điểm thích hợp để cảm ứng riêng [156]. Ví dụ khi nghiên cứu biểu hiện gen mã hóa beta-defensin-4 của người cho thấy thời điểm cảm ứng thích hợp ở OD_{600} trong khoảng 0,8 – 1 [188], của cystein protease từ *Porphyromonas gingivalis* là 0,6 [39]. Trong nghiên cứu này chúng tôi lựa chọn cảm ứng IPTG ở mật độ tế bào ở 3 giá trị khác nhau bao gồm OD_{600} = 0,7; 1,0 và 1,5. Mật độ tế bào thu được sau 5 giờ cảm ứng được đo OD để kiểm tra. Kết quả chỉ ra khi cảm ứng IPTG ở OD_{600} là 1,5 thu được lượng sinh khối tế bào lớn nhất (Hình 3.18). Mật độ tế bào cảm ứng này cũng khá cao so với các nghiên cứu trước đó về OD cảm ứng gen mã hóa beta-xylosidase nằm trong giới hạn từ 0,6 đến 1,0. Tuy nhiên kết quả này có thể do thí nghiệm của chúng tôi tiến hành đồng biểu hiện với các chaperone, nên lượng chaperone tạo ra trước càng nhiều thì khả năng protein *Xbx14* được bảo vệ và hình thành cấu trúc đúng cũng sẽ tăng lên.

Mẫu tế bào thu được tiếp tục kiểm tra lượng protein *Xbx14* tổng số và ở dạng tan. Kết quả kiểm tra bằng điện di trên gel polyacrylamide 12,6% có SDS cho thấy, khi cảm ứng tại thời điểm OD_{600} là 0,7 và 1,5 thu được lượng protein *Xbx14* tổng số và tan tương đương nhau (Hình 3.19). Tuy nhiên cảm ứng ở OD_{600} = 1,5 thu được sinh khối tế bào lớn hơn. Do đó chúng tôi sẽ chọn cảm ứng ở OD_{600} = 1,5 cho các thí nghiệm tiếp theo.

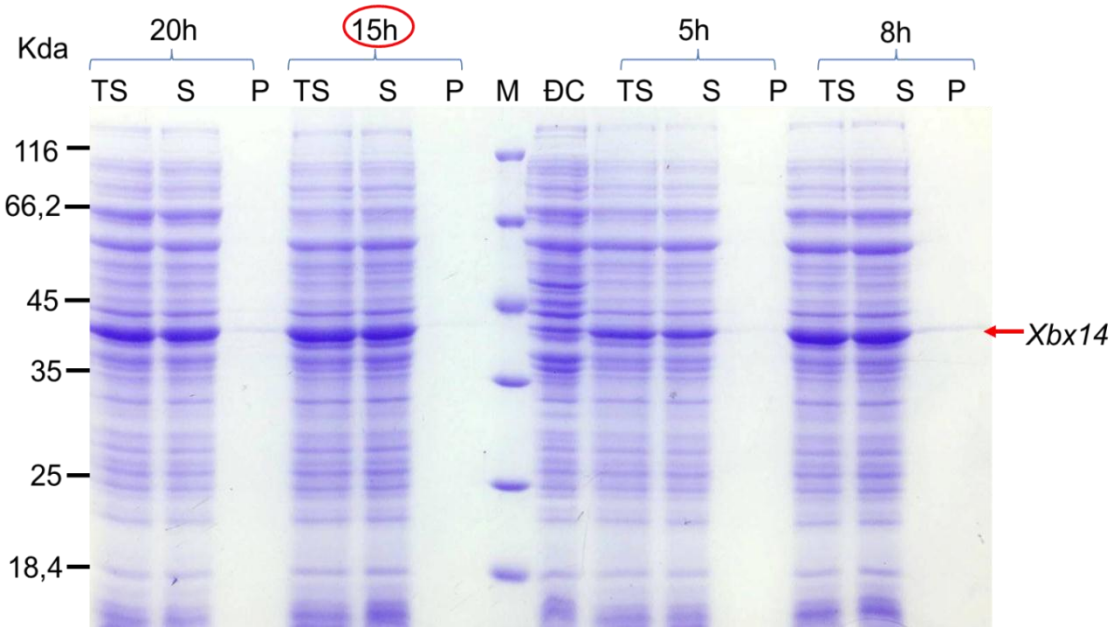
d. Nghiên cứu ảnh hưởng thời gian thu mẫu đến hiệu quả biểu hiện protein Xbx14

Trong điều kiện nuôi cấy thích hợp, tế bào sinh tổng hợp protein mục tiêu ngay sau khi được cảm ứng IPTG. Do đó mức độ tích lũy sản phẩm protein ngoại lai theo lý thuyết sẽ tăng lên và đạt cực đại ở một thời điểm nhất định. Tuy nhiên khi thời gian nuôi cấy kéo dài, nguồn dinh dưỡng cạn dần và tế bào chết có thể sản sinh ra sản phẩm không mong muốn như các protease sẽ phân giải protein tái tổ hợp dẫn đến giảm sản lượng. Vì vậy chúng tôi sẽ nghiên cứu để xác định ảnh hưởng của

thời gian thu mẫu đến hiệu quả biểu hiện gen *Xbx14*. Tế bào sau khi cảm ứng 0,05 mM IPTG sẽ được nuôi tiếp ở 20°C, mẫu sẽ được thu lại để kiểm tra sau 5 giờ, 8 giờ, 15 giờ và 20 giờ cảm ứng.



Hình 3.20. Mật độ tế bào biểu hiện E. coli Rosetta 1 (DE3) mang* pET22Xbx *thu mẫu sau 5, 8, 15 và 20 giờ cảm ứng



Hình 3.21. Điện di đồ phân tích sản phẩm biểu hiện của Xbx14 trong tế bào E. coli Rosetta 1 (DE3) thu mẫu sau 5, 8, 15 và 20 giờ cảm ứng IPTG trên gel polyacrylamide 12,6% có SDS

Đường chạy ĐC: đối chứng âm, protein tổng số của dòng tế bào biểu hiện không cảm ứng IPTG, đường chạy TS: protein tổng số biểu hiện trong tế bào mang

pET22Xbx, *đường chạy S: protein dạng tan, đường chạy P: protein dạng không tan, đường chạy M: protein chuẩn unstained (Thermo scientific)*

Kết quả đo OD_{600} ở thời điểm thu mẫu khác nhau cho thấy mật độ tế bào tăng dần từ 5 giờ và đạt cao nhất tại thời điểm 15 giờ, sau đó giảm tại thời điểm 20 giờ (Hình 3.20). Mẫu tế bào thu được tiếp tục được phá bằng siêu âm để kiểm tra mức độ biểu hiện và dạng tan của *Xbx14*. Điện di đồ cho thấy tại thời điểm thu mẫu sau 5 giờ cảm ứng lượng protein *Xbx14* tổng số và tan thu được ít nhất do mật độ tế bào thấp. Tại thời điểm thu mẫu sau 8 giờ, 15 giờ mật độ tế bào tiếp tục tăng và lượng *Xbx14* biểu hiện tổng số và tan là tương đương nhau. Tuy nhiên tại thời điểm thu mẫu sau 20 giờ cảm ứng, mặc dù mật độ tế bào đã giảm, do môi trường thiếu chất dinh dưỡng, nhưng lượng protein *Xbx14* tổng số và tan vẫn ổn định (Hình 3.21). Kết quả này có thể giải thích, là do sự có mặt của các chaperone đã bảo vệ enzyme *Xbx14* tránh khỏi sự phân cắt protease giải phóng từ các tế bào bị chết. Như vậy đồng biểu hiện *Xbx14* với chaperone thu được sản lượng protein tái tổ hợp ổn định. Kết hợp với số liệu mật độ tế bào khi thu mẫu, chúng tôi lựa chọn thời điểm thu mẫu là sau 15 giờ cảm ứng ở 20^oC.

Tóm lại các điều kiện tối ưu đồng biểu hiện *Xbx14* với các chaperone trong tế bào *E. coli Rosetta 1* bao gồm: nhiệt độ biểu hiện 20^oC, nồng độ IPTG cảm ứng 0,05 mM, OD_{600} cảm ứng 1,5 và thu mẫu sau 15 giờ cảm ứng.

3.3.2. Tinh chế *Xbx14*

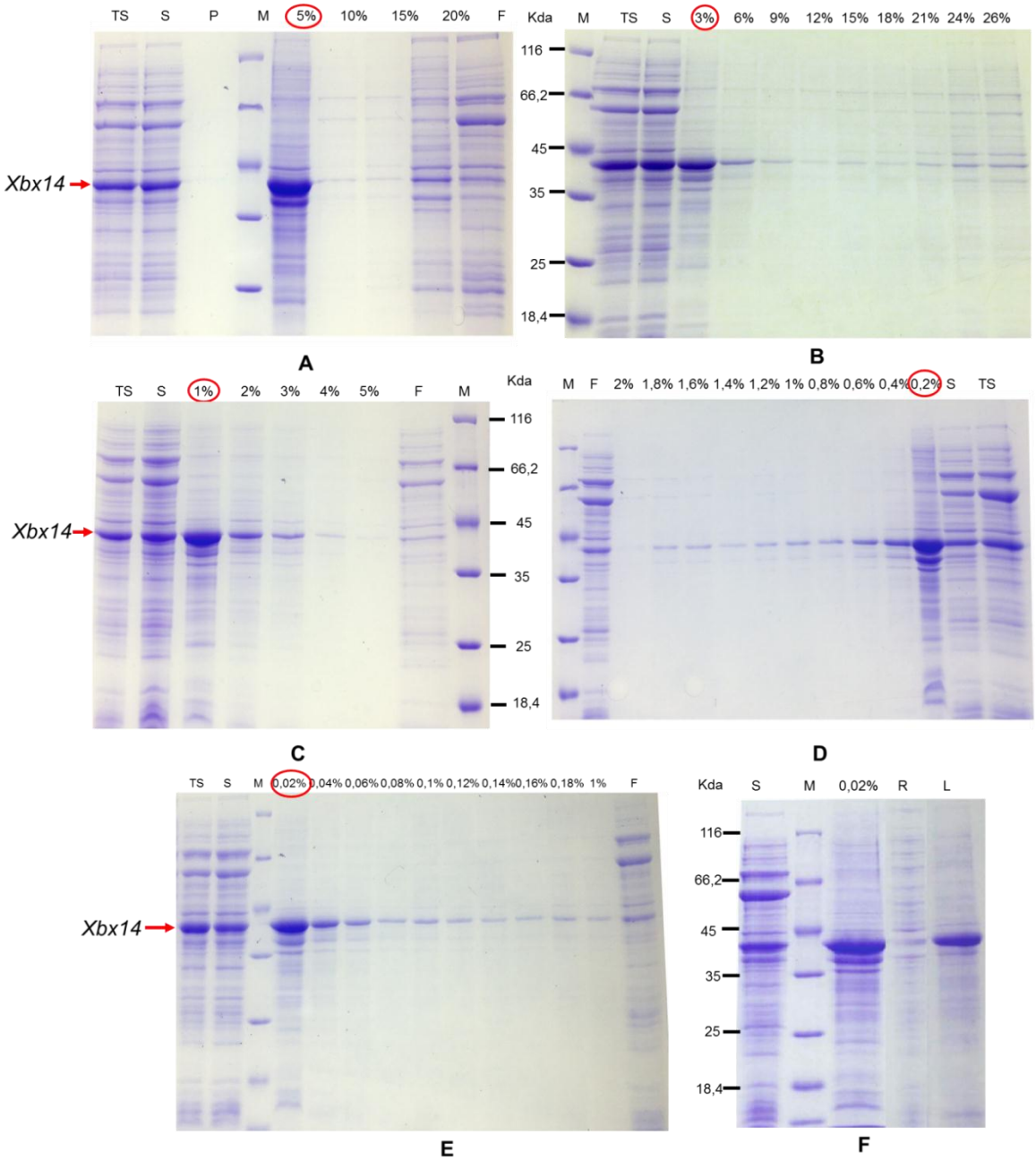
Theo thiết kế ban đầu gen GL0112518 được gắn vào vector biểu hiện pET22b(+) tại vị trí đa nối, phía sau vị trí cắt đa điểm có một đoạn trình tự mã hóa cho 6 axit amin Histidine tích điện âm được gọi là đuôi His-tag. Đuôi này giúp protein tái tổ hợp được giữ lại khi chúng đi qua cột sắc kí ái lực, do lực hút tĩnh điện giữa phần điện tích âm của đuôi His-tag với điện tích dương của các cation có trong cột tinh sạch. Các protein này sẽ được tách khỏi cột khi dịch đệm có pH thích hợp chảy qua.

Trong các thí nghiệm tối ưu trên, chúng tôi hòa tủa tế bào sau khi thu mẫu vào trong nước, siêu âm để phá tế bào cho thấy *Xbx14* tan hoàn toàn trong nước. Muốn sử dụng cột sắc ký ái lực để tinh sạch, thì protein *Xbx14* tan sau khi phá tế bào phải được hòa bằng chính đệm tinh chế. Do đó mẫu protein *Xbx14* tan trong nước sẽ được bổ sung muối trong đệm tinh chế trước khi đưa lên cột. Tuy nhiên sau khi bổ sung muối protein *Xbx14* bị tủa hoàn toàn. Đây là lý do chúng tôi không sử dụng được cột sắc ký ái lực, mà sử dụng muối $(NH_4)_2SO_4$ để tinh chế bằng cách tủa phân đoạn.

3.3.2.1. Tinh chế Xbx14 bằng muối $(NH4)_2SO4$

Cơ sở khoa học để tinh chế protein bằng tủa muối dựa trên độ hòa tan của protein (đặc biệt là các protein hình cầu) bị ảnh hưởng khi bổ sung nồng độ muối thấp (<0,15 M). Khi muối được thêm vào dung dịch, sức căng bề mặt nước tăng lên, dẫn đến sự tương tác giữa protein và nước tăng lên. Protein đáp ứng với tình trạng này bằng cách giảm thiểu tiếp xúc thông qua việc thu nhỏ cấu trúc làm giảm diện tích bề mặt, sau đó dẫn đến kết tủa [184]. Các yếu tố ảnh hưởng đến nồng độ muối làm protein kết tủa bao gồm: số lượng và vị trí của các nhóm phân cực, trọng lượng phân tử của protein, pH của dung dịch và nhiệt độ [184].

Muối $(NH_4)_2SO_4$, được sử dụng để tủa protein vì có khả năng vừa làm trung hòa điện (do các ion tác động tương hỗ với nhóm tích điện trái dấu), vừa loại bỏ lớp vỏ hydrat của phân tử protein. Mặt khác muối $(NH_4)_2SO_4$ không làm biến tính protein và dễ dàng hòa tan tủa protein bằng các đệm khác nhau. Mỗi loại protein bị kết tủa ở nồng độ muối không giống nhau, nên dễ dàng phân tách từng loại protein ra khỏi hỗn hợp [184]. Trong nghiên cứu này chúng tôi sử dụng muối $(NH_4)_2SO_4$ để tủa theo phân đoạn, lần lượt từ phân đoạn có nồng độ cao đến nồng độ thấp dần *(Phụ lục 3)*. Nồng độ muối $(NH_4)_2SO_4$ sử dụng thấp nhất có khả năng tủa được nhiều nhất *Xbx14* và chứa ít các protein không mong muốn sẽ được sử dụng để tinh chế lượng lớn.



Hình 3.22. Điện di đồ phân tích kết quả tủa protein Xbx14 biểu hiện trong tế bào E. coli Rosetta 1 (DE3) bằng muối $(NH_4)_2SO_4$ theo phân đoạn 5% (A), 3% (B), 1% (C), 0,2% (D), 0,02% (E) và rửa tủa bằng đệm phosphate pH = 9 (F)

Đường chạy TS: protein tổng số biểu hiện trong tế bào mang pET22Xbx, đường chạy S: protein dạng tan, đường chạy 5%, 10%,... tương ứng nồng độ muối $(NH_4)_2SO_4$ sử dụng để tủa Xbx14, đường chạy F: dịch cuối cùng sau khi tủa protein tan, đường chạy R: dịch sau khi rửa tủa protein Xbx14 bằng đệm phosphate

(pH=9), đường chạy L: Xbx14 sau khi rửa bằng đệm phosphate, đường chạy M: protein chuẩn unstained (Thermo scientific)

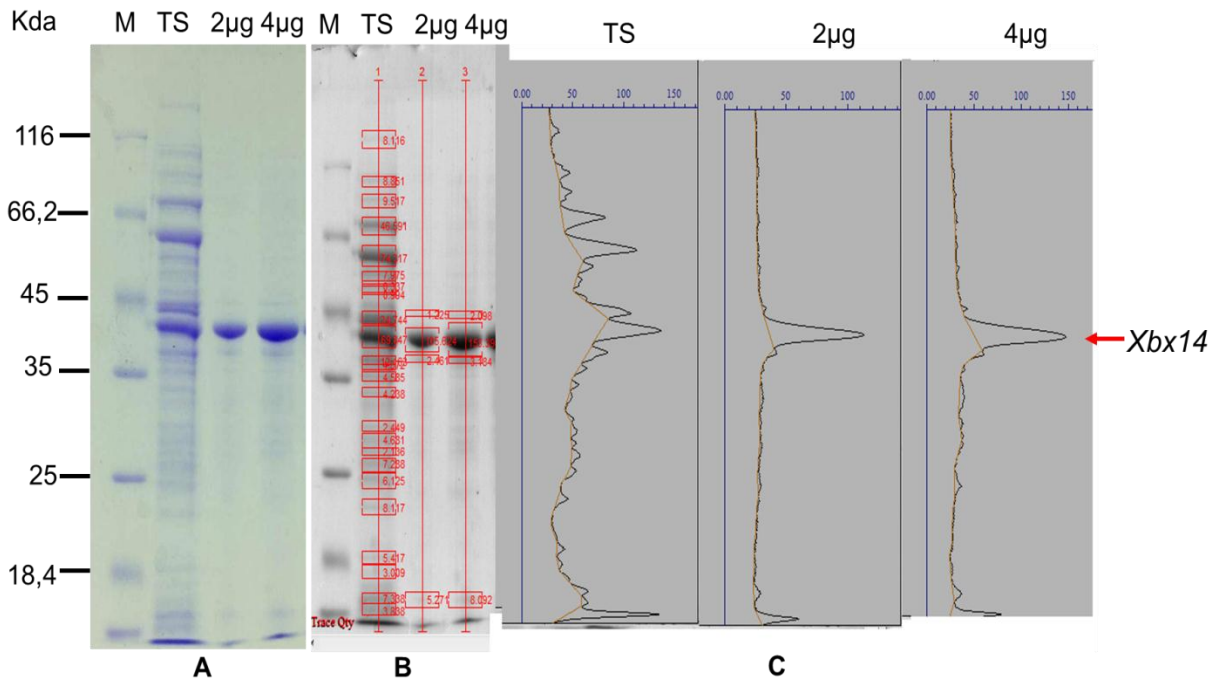
Sau khi bổ sung muối $(NH_4)_2SO_4$ vào dung dịch phá tế bào trong nước cho thấy *Xbx14* bị tủa theo từng phân đoạn từ nồng độ muối cao, đến nồng độ muối rất thấp. Trong hình 3.22A, *Xbx14* bị tủa ở nồng độ 5% muối $(NH_4)_2SO_4$, và tiếp tục bị tủa khi sử dụng các nồng độ thấp hơn là 3% (Hình 3.22B), 1% (Hình 3.22C), 0,2% (Hình 3.22D) và 0,02% (Hình 3.22E). Càng ở nồng độ muối thấp protein không mong muốn bị tủa cùng *Xbx14* ít dần. Hầu như các loại protein có kích thước lớn hơn *Xbx14* không bị tủa, do đó mà lượng protein tạp còn rất ít. Protein *Xbx14* bị tủa ở nồng độ 0,02% $(NH_4)_2SO_4$ rất thấp so với nghiên cứu tủa β-xylosidase từ *Aspergillus tamarii Kita* là 60% [56], *Emericella nidulans* là 80% [113] và *Pseudozyma hubeinsis* là 90% [117]. *Xbx14* bị tủa ở nồng độ muối thấp có thể do cấu trúc không gian có dạng hình cầu [52], đúng như dự đoán ban đầu (Hình 3.8). Mặt khác với nồng độ muối thấp thì sản phẩm protein tinh sạch không cần phải qua giai đoạn loại muối vì ít ảnh hưởng đến hoạt tính của enzyme.

Để tiếp tục làm sạch các protein không mong muốn, nghiên cứu dựa trên điểm đẳng điện của protein *Xbx14* và khả năng hòa tan trong đệm với pH khác nhau. Protein có xu hướng kết tụ và kết tủa tại pI của chúng bởi vì không có điện đẩy tĩnh điện giữ chúng tách nhau ra. Protein có điểm đẳng điện khác nhau do các chuỗi axit amin khác nhau (số lượng tương đối của nhóm anion và cation). Về lý thuyết các loại protein khác nhau có thể tách ra bằng cách điều chỉnh độ pH của dung dịch. Nghĩa là khi pH được điều chỉnh đến đúng pI của một protein, nó sẽ kết tủa và tách ra khỏi hỗn hợp các protein khác trong dung dịch. Protein *Xbx14* có điểm đẳng điện là 7,59 và sau khi tiến hành phá tế bào cho thấy *Xbx14* bị tủa gần như hoàn toàn trong đệm phosphate. Vì vậy *Xbx14* sau khi được tủa bằng 0,02% $(NH_4)_2SO_4$ sẽ tiếp tục được rửa bằng đệm phosphate với pH khác nhau. Sau đó tiến hành ly tâm thu tủa và dịch tan để kiểm tra. Kết quả chỉ ra, hầu hết protein không mong muốn đã bị hòa tan, còn *Xbx14* vẫn tủa trong đệm phosphate pH = 9. Như vậy việc sử dụng đệm phosphate pH = 9 để rửa tủa, tiếp tục giúp tinh sạch *Xbx14* (Hình 3.22F) theo đúng

lý thuyết về điểm đẳng điện của protein. So sánh với phương pháp sử dụng cột sắc ký ái lực thì phương pháp tinh sạch sử dụng muối $(NH_4)_2SO_4$ và đệm phosphate pH = 9 này đơn giản và hiệu quả hơn. Đặc biệt có thể dễ dàng sử dụng phương pháp này tinh chế được lượng lớn protein Xbx14 trong một thời gian ngắn.

3.3.2.2. Kiểm tra độ sạch của Xbx14 sau tinh chế bằng phần mềm Quantity one

Quan sát bằng mắt thường hình ảnh điện di sản phẩm sau khi tinh chế và rửa bằng đệm phosphate pH = 9 cho thấy xuất hiện băng protein *Xbx14* to và đậm, kèm theo một số băng protein mờ nhạt khác. Bằng phần mềm Quantity One, độ sạch tương đối của protein *Xbx14* sau tinh chế được đánh giá ở mức định lượng. Mẫu được tính toán cho mỗi đường chạy là 2 µg và 4 µg protein (Hình 3.23A). Một vị trí nền (BGR) có màu sắc và độ sáng giống với nền ở vị trí băng protein *Xbx14* trên ảnh quét bản gel được lựa chọn. Mỗi băng điện di được chỉ ra bởi một đỉnh đường cong tương ứng trên đồ thị (Hình 3.23B và Hình 3.23C). Mỗi đỉnh trên đường cong được phần mềm tự động tính khối lượng protein tương ứng.



Hình 3.23. Kết quả kiểm tra độ sạch của Xbx14 sau khi tinh chế bằng muối $(NH_4)_2SO_4$ và rửa bằng đệm phosphate (pH = 9)

Độ sạch của protein *Xbx14* tinh chế được thể hiện thông qua tỷ lệ giữa khối lượng protein *Xbx14* và tổng khối lượng protein trên cùng đường chạy. Kết quả phân tích hình ảnh điện di được quét và tính toán lượng protein *Xbx14* so với protein tổng số chiếm 19,8%. Tuy nhiên sau khi tinh chế thì *Xbx14* chiếm tỷ lệ 92,7%. Kết quả tinh chế *Xbx14* bằng tủa muối này, thu được độ sạch cao hơn rất nhiều so với kết quả tinh chế β- xylosidase từ *Pseudozyma hubeinsis* chỉ có độ sạch là 53,12% [117].

Sử dụng muối $(NH_4)_2SO_4$ ở nồng độ 0,02% để tủa phân đoạn và rửa tủa *Xbx14* bằng đệm phosphate (pH = 9) đã thu *Xbx14* với độ sạch 92,7% đủ điều kiện để xác định hoạt tính của enzyme. Phương pháp tinh sạch này khá đơn giản để thu được lượng lớn protein *Xbx14* khi ứng dụng trong công nghiệp. Kết quả *Xbx14* bị tủa ở nồng độ muối rất thấp cũng cho phép dự đoán mô hình cấu trúc không gian của *Xbx14* có dạng hình cầu, theo đúng dự đoán của phần mềm Phyre 2.

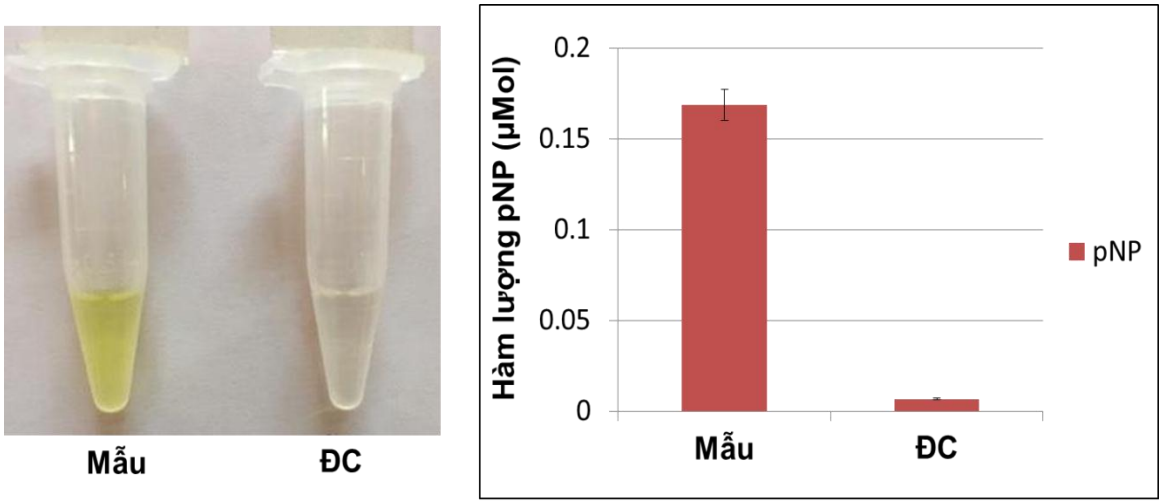
3.3.3. Nghiên cứu tính chất của *Xbx14*

3.3.3.1. Xác định hoạt tính của Xbx14

Hoạt tính của *Xbx14* được xác định bằng cơ chất pNPX 10 mM pha trong đệm phosphate (pH = 7) theo phương pháp của Teng và cộng sự [168]. Phản ứng được trộn giữa 30 µl enzyme (0,253 mg/ml) và 20 µl cơ chất pNPX 10 mM ủ ở 50°C trong 1,5 giờ. Ở mẫu đối chứng enzyme và cơ chất để riêng rẽ, ủ cùng điều kiện với mẫu phản ứng. Phản ứng được dừng lại bằng cách bổ sung 200 µl Na_2CO_3 0,1 mM và đo OD_{405} để xác định lượng pNP tạo ra theo đường chuẩn.

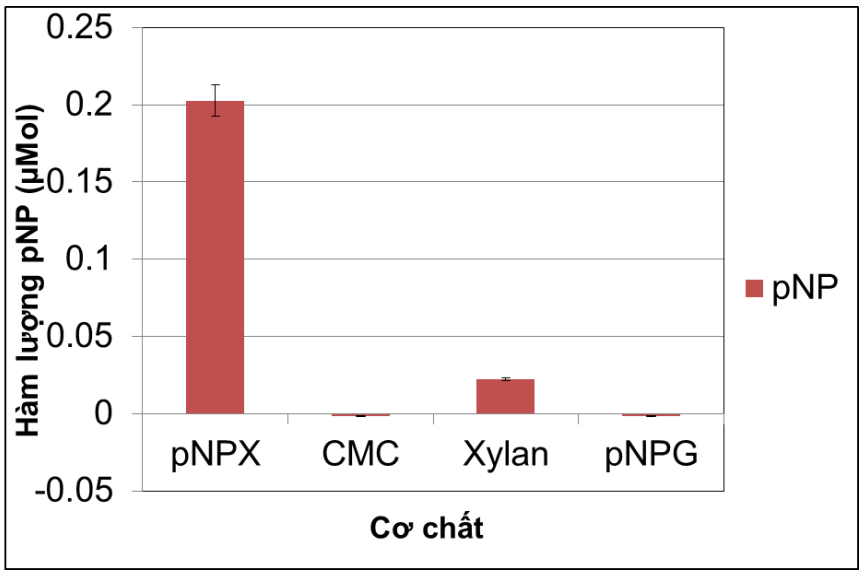
Sau khi ủ 1,5 giờ bằng mắt thường quan sát cho thấy ở mẫu phản ứng xuất hiện màu vàng đặc trưng của pNP, còn mẫu đối chứng (ĐC) không xuất hiện màu vàng. Hoạt tính của enzyme tiếp tục được xác định chính xác bằng phương pháp đo Elisa ở bước sóng 405 nm. Kết quả đo được tính toán dựa theo đường chuẩn cho thấy sau 1,5 giờ phản ứng, tạo ra lượng pNP là 0,1692 µMol (Hình 3.24). Như vậy protein *Xbx14* sau khi tinh chế vẫn giữ nguyên được hoạt tính xúc tác của β-

xylosidase như mẫu thử enzyme thô ban đầu. Kết quả này khẳng định việc tinh chế bằng muối $(NH_4)_2SO_4$ và rửa tủa bằng đệm phosphate pH = 9 hoàn toàn giữ nguyên được hoạt tính enzyme và đạt hiệu quả cao trong nghiên cứu này.



Hình 3.24. Kết quả thử hoạt tính của Xbx14 bằng cơ chất đặc hiệu pNPX

3.3.3.2. Xác định tính đặc hiệu cơ chất của Xbx14



Hình 3.25. Hoạt tính Xbx14 trên các cơ chất khác nhau

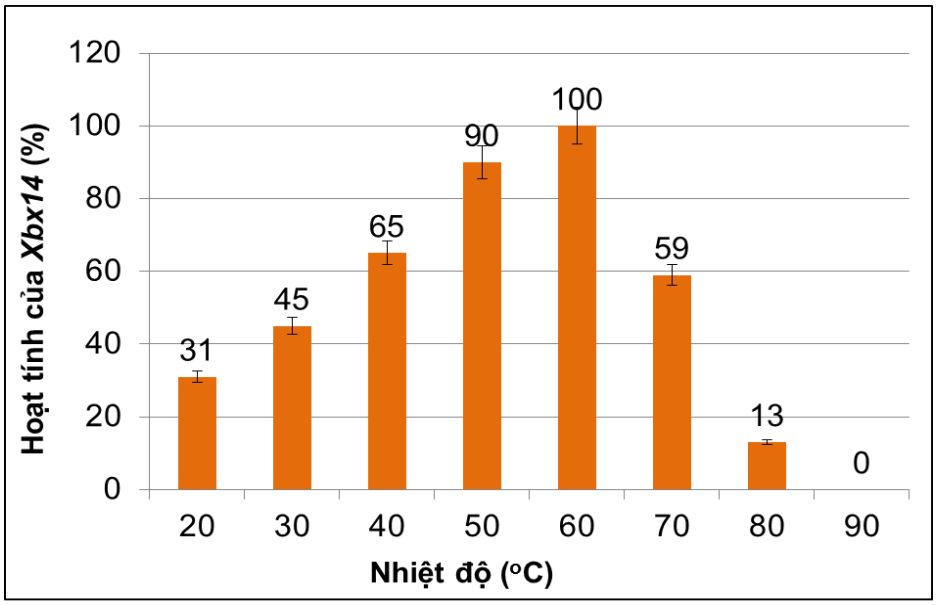
Một số cơ chất phổ biến là CMC, *p*-nitrophenyl-β-glucoside (*p*NPG), 4-nitrophenyl β- D-xylopyranoside (pNPX) và xylan được sử dụng để kiểm tra tính đặc hiệu của *Xbx14*. Đây là những cơ chất đặc hiệu của các cellulase và

hemicellulase. Trong đó, pNPX là cơ chất đặc hiệu dạng tan của β–xylosidase, xylan là cơ chất đặc hiệu của xylanase, CMC là cơ chất đặc hiệu của endoglucanase và *p*NPG là cơ chất đặc hiệu của β-glucosidase.

Kết quả kiểm tra cho thấy, *Xbx14* thể hiện hoạt tính cao nhất trên cơ chất pNPX, có hoạt tính yếu với cơ chất xylan và không phân cắt cơ chất CMC và pNPG (Hình 3.25). Điều này chứng tỏ *Xbx14* có hoạt tính đặc hiệu của β–xylosidase phân cắt đường đôi xylose và phân tử đường ở đầu của mạch xylan. Như vậy có thể khẳng định *Xbx14* mang đúng hoạt tính của β–xylosidase.

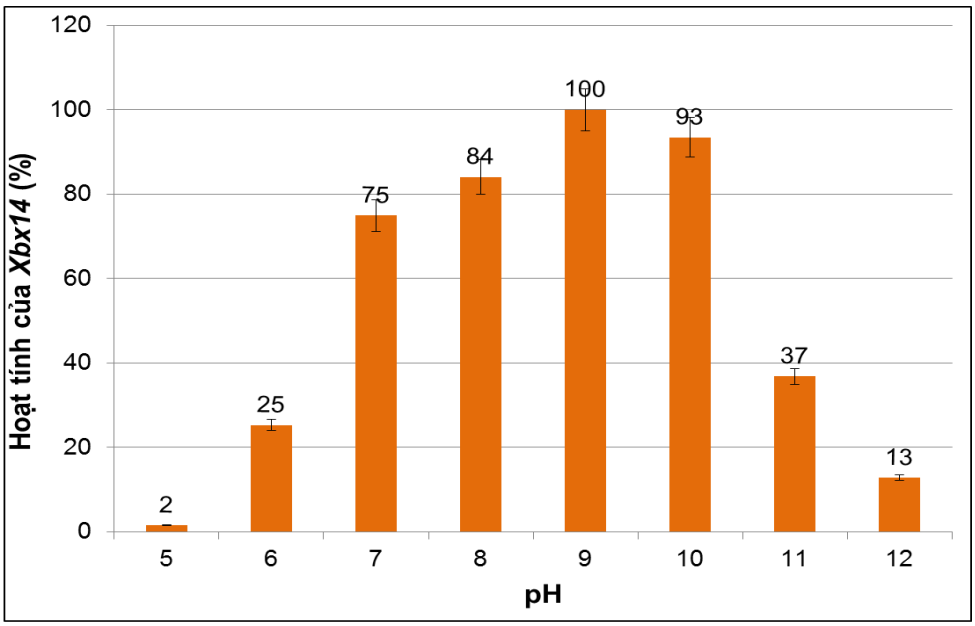
3.3.3.3. Nghiên cứu ảnh hưởng của nhiệt độ đến hoạt tính của Xbx14

Nghiên cứu ảnh hưởng của nhiệt độ đến hoạt tính của enzyme *Xbx14* theo một dải liên tục là 20°C, 30°C, 40°C, 50°C, 60°C, 70°C, 80°C và 90°C. Mẫu được ủ ở các nhiệt độ trên trong 1,5 giờ cho thấy hoạt tính β–xylosidase tăng dần từ 20°C (31%) đến 60°C (100%) và giảm dần khi nhiệt độ tăng đến 80°C và giảm mạnh ở nhiệt độ 90°C (0%) (Hình 3.26). Như vậy, nhiệt độ thích hợp nhất cho hoạt động của *Xbx14* là 60°C. Nhiệt độ hoạt động tối ưu này hoàn toàn đúng với ước đoán ban đầu bằng công cụ TBI đã sử dụng khi chọn gen (từ 55°C đến 65°C). Nhiệt độ hoạt động tối ưu này thấp hơn so với β–xylosidase *ở Geobacillus thermoleovorans IT-08* được phân lập từ suối nước nóng là 70°C [175], tương đương với hoạt động của β–xylosidase ở *Geobacillus stearothermophilus T-6 (XynB3)* [148],... nhưng cao hơn hẳn so với nhiệt độ hoạt động tối ưu của β–xylosidase ở đa số vi khuẩn như *Selenomonas ruminantium GA192* [35], *Vibrio sp. XY-214* [177], *Bacillus pumilus* [186],…. Theo một số nghiên cứu về enzyme trong ruột mối đã chỉ ra nhiệt độ hoạt động tối ưu thường dao động từ 30°C đến 40°C, phù hợp với môi trường ruột mối [16]. Tuy nhiên enzyme *Xbx14* có khả năng hoạt động tối ưu ở 60°C cao hơn hẳn môi trường ruột mối thông thường. Đây là enzyme có khả năng chịu nhiệt rất tốt và có thể là một enzyme hoàn toàn mới có nguồn gốc từ vi khuẩn *Treponema primitia* so với các nghiên cứu đã công bố.



Hình 3.26. Ảnh hưởng của nhiệt độ đến hoạt tính Xbx14

3.3.3.4. Nghiên cứu ảnh hưởng của pH đến hoạt tính của Xbx14



Hình 3.27. Ảnh hưởng của pH đến hoạt tính của Xbx14 ở 60°C

Theo ước đoán bằng phần mềm Alcapred, *Xbx14* là enzyme có khả năng hoạt động tốt ở vùng pH kiềm. Do đó trong nghiên cứu này chúng tôi chọn dải pH = 5, 6, 7, 8, 9, 10, 11, 12 để thử hoạt tính của *Xbx14*. Sau khi kiểm tra hoạt tính chỉ ra, *Xbx14* không có hoạt tính ở pH = 5, hoạt tính tăng dần từ pH = 6 (25%) đến pH = 9 và đạt giá trị cao nhất ở pH = 9 (100%). Sau đó hoạt tính giảm nhẹ ở pH = 10 còn

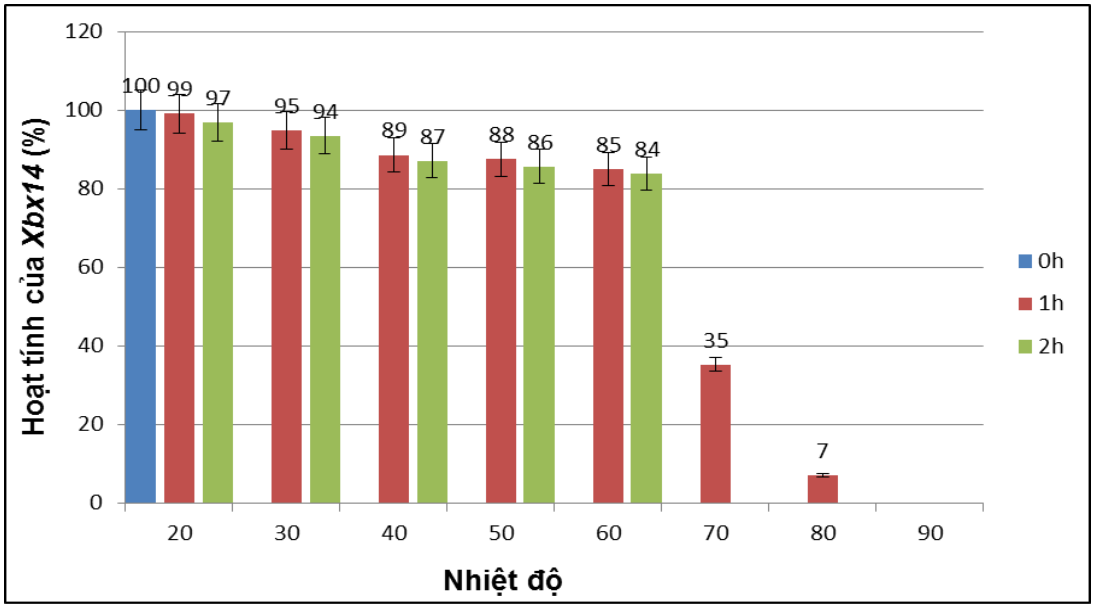
93% và giảm sâu ở pH = 12 chỉ còn 13% (Hình 3.27). Như vậy pH hoạt động tối ưu của *Xbx14* cho thấy enzyme này hoạt động tối ưu trong môi trường kiềm tính, đúng với dự đoán ban đầu khi sử dụng công cụ Alcapred. Kết quả này phù hợp với một số nghiên cứu về pH trong ruột của mối cho rằng từ ruột trước đến ruột giữa của mối bậc cao có giá trị pH trung tính, nhưng phần đầu tiên của ruột sau có pH kiềm tính đạt đến giá trị 12 như ở mối *R. flavipes, Nasutitermes lujae* [23]. Đây được cho là điều kiện khá lý tưởng để nới lỏng và phá vỡ các liên kết của lignin trong thành phần của lignocellulose. Đồng thời điều kiện này cũng làm tăng hiệu quả tiêu hóa thức ăn trong ruột mối *C. gestroi*. Từ hoạt động của *Xbx14* cho phép dự đoán môi trường đoạn đầu ruột sau của mối *C. gestroi* là môi trường kiềm tương tự như các loài mối khác [87].

So sánh với các nghiên cứu về β–xylosidase thuộc họ GH43 từ vi khuẩn và nấm cho thấy sự không đồng nhất về pH hoạt động tối ưu. Ở vi khuẩn dải pH hoạt động tối ưu thay đổi từ trung tính (*Bacillus spp.* [14]) đến axit (*Selenomonas ruminantium GA192* (pH = 5,3) [35]). Chỉ có β-xylosidase được phân lập từ thư viện DNA đa hệ gen của vi sinh vật trong mẫu ủ rơm rạ có pH hoạt động tối ưu từ 8 đến 10 [111]. Trong khi đó hầu như các β–xylosidase từ nấm có pH hoạt động tối ưu nằm trong giá trị axit như *Penicillium sclerotiorum* (pH = 2,5); *Aspergillus brasiliensis, A. niger* (pH = 5) [132],... Như vậy *Xbx14* hoạt động tối ưu ở pH = 9 là enzyme mới so với các công bố từ vi khuẩn trong ruột mối *C. gestroi*. Enzyme này sẽ ứng dụng rất hiệu quả khi kết hợp với điều kiện tiền xử lý sinh khối thực vật bằng dung dịch kiềm. Đây cũng chính là enzyme mà chúng tôi đặt mục tiêu tìm kiếm trong quá trình chọn gen biểu hiện. Kết quả này cũng chứng minh việc sử dụng công cụ Alcapred ban đầu để dự đoán pH hoạt động tối ưu của enzyme *Xbx14* là hoàn toàn chính xác.

3.3.3.5. Nghiên cứu độ bền nhiệt của Xbx14

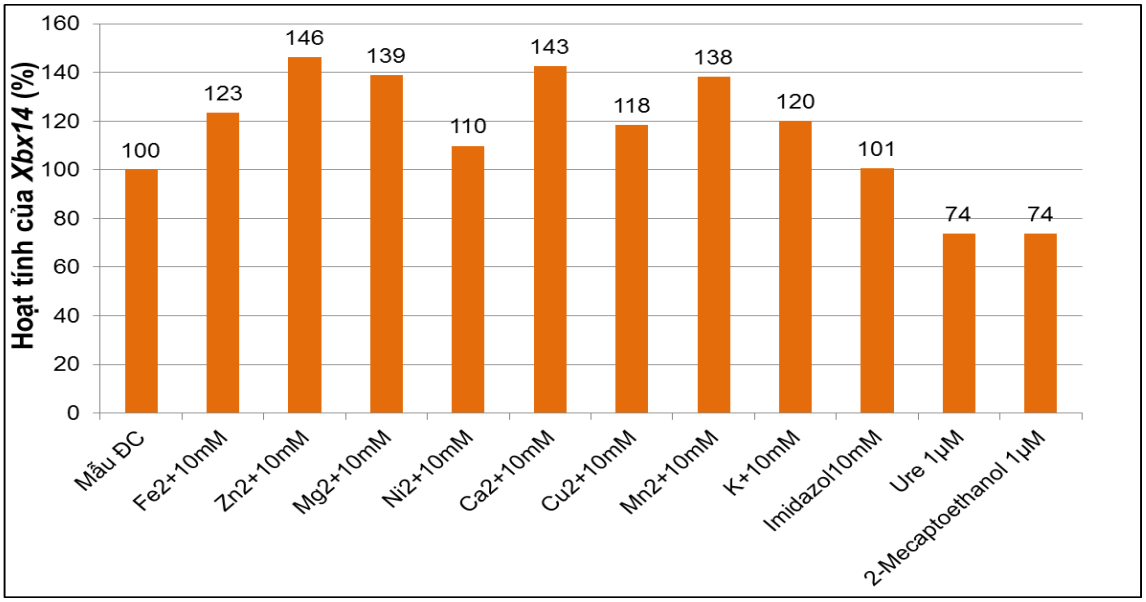
Xbx14 được xử lý ở nhiệt độ 20°C, 30°C, 40°C, 50°C, 60°C, 70°C, 80°C và 90°C trong các khoảng thời gian 1 giờ và 2 giờ trước khi ủ với cơ chất, để kiểm tra độ bền nhiệt. Ở nhiệt độ lớn hơn 70°C, *Xbx14* mất hoàn toàn hoạt tính dù là thời gian xử lý ngắn nhất (1 giờ). Ở nhiệt độ từ 20°C đến 60°C, hoạt tính enzyme gần như ổn định sau khi xử lý nhiệt từ 1 đến 2 giờ. Như vậy enzyme *Xbx14* khá bền

nhiệt trong khoảng nhiệt độ 20°C đến 60°C (Hình 3.28). Độ bền nhiệt ở 60°C của *Xbx14* tương tự như β–xylosidase có nguồn gốc từ nấm *Aspergillus brasiliensis*, *A. niger* [132] và từ vi sinh vật trong mẫu ủ phân hữu cơ [47]. Sự bền nhiệt của *Xbx14* ở nhiệt độ hoạt động tối ưu 60°C tiếp tục làm tăng hiệu quả sử dụng enzyme trong công nghiệp nếu tiền xử lý sinh khối thực vật bằng nhiệt.



Hình 3.28. Hoạt tính của Xbx14 khi xử lý ở các nhiệt độ khác nhau

3.3.3.6. Nghiên cứu ảnh hưởng của một số ion kim loại và hóa chất lên hoạt tính của Xbx14



Hình 3.29. Ảnh hưởng một số ion kim loại và hóa chất lên hoạt tính của Xbx14

Thông thường, các ion kim loại có ảnh hưởng đến hoạt động của *Xbx14* theo hướng làm tăng hoặc giảm hoạt tính. Nhiều nghiên cứu cho thấy, trong khoảng nồng độ từ 1 mM đến 10 mM, hoạt tính của β–xylosidase có thể tăng dần hoặc giảm dần theo sự tăng nồng độ ion. Để đánh giá ảnh hưởng của ion kim loại lên hoạt tính của *Xbx14*, chúng tôi đã tiến hành theo phương pháp được mô tả bởi Lee và cộng sự [100]. Nồng độ của các ion kim loại sử dụng trong phản ứng là 10 mM, nồng độ của một số chất khử như 2-mecaptoethanol, urea là 1μM và của imidazol là 10 mM. Trong nghiên cứu này, chúng tôi lựa chọn 9 ion kim loại phổ biến (K^+, Mn^{2+}, Cu^{2+}, Ca^{2+}, Ni^{2+}, Mg^{2+}, Zn^{2+}, Fe^{2+}), imidazole, Urea và 2 - mercaptoethanol để đánh giá ảnh hưởng của chúng đến hoạt tính của *Xbx14*. Tất cả được tiến hành ở điều kiện hoạt động tối ưu của *Xbx14* là 60°C và pH = 9.

Kết quả thí nghiệm cho thấy các ion kim loại đều làm tăng hoạt tính của enzyme *Xbx14* từ 110% đến 146% (Hình 3.29). Kết quả này tương tự với nghiên cứu ảnh hưởng của các ion kim loại hóa trị hai bao gồm Ca^{2+}, Co^{2+}, Fe^{2+}, Mn^{2+} và Ni^{2+} làm tăng hoạt tính của β–xylosidase, do chúng đã tham gia vào quá trình chuyển từ trạng thái dimer sang tetramer trong họ GH43 [85]. Tuy nhiên trong một nghiên cứu khác chỉ ra ion Cu^{2+}, Hg^{2+}, Mn^{2+}, Pb^{2+}, Zn^{2+} lại ức chế hơn 90% hoạt tính của β–xylosidase của *Vibrio* sp. XY-214 [177], hoặc ion Ba^{2+}, Ca^{2+}, Fe^{2+}, Mg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+} và Na^+ không ảnh hưởng đến hoạt tính β–xylosidase của vi khuẩn *C. stercorarium* [161]. Hai ion Zn^{2+} và Ca^{2+} làm tăng hoạt tính của *Xbx14* lên nhiều nhất, có thể do Ca^{2+} là ion nằm ở vùng liên kết giữa các phân tử polyme để tạo dạng tetramer, nên giúp làm ổn định cấu trúc của *Xbx14* [67]. Do đó làm tăng hiệu quả tương tác và phân cắt giữa *Xbx14* với cơ chất. Kết quả của chúng tôi tương tự như nghiên cứu ảnh hưởng của ion Ca^{2+} đã làm tăng hoạt tính của β–xylosidase từ *Humicola grisea var. thermoidea* [6], *T. thermophilus* [67]. Nếu như các ion kim loại nặng Zn^{2+}, Cu^{2+}, Mg^{2+} làm tăng hoạt tính của *Xbx14* thì một số nghiên cứu khác chỉ ra các ion này ức chế hoạt động của β–xylosidase [114].

Kiểm tra ảnh hưởng của Imidazol, Urê và 2-Mecaptoethanol cho thấy ở nồng độ 10 mM imidazol không ảnh hưởng đến hoạt tính của *Xbx14*, tuy nhiên ở nồng độ

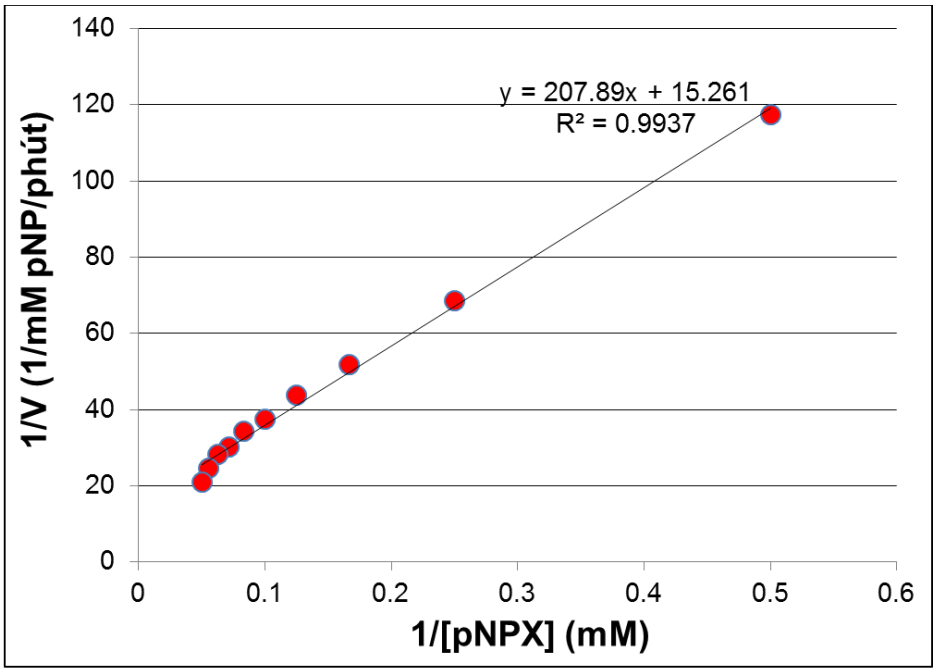
1 µM chất khử 2- mercaptoethanol và 1 µM urea làm giảm đáng kể đến hoạt tính của *Xbx14* (Hình 3.29). Kết quả nghiên cứu ảnh hưởng của Imidazol đến hoạt tính của *Xbx14* tương tự như nghiên cứu hoạt động của beta-xylosidase RS223BX thuộc họ GH43. Nguyên nhân là do imidazole là phân tử trung gian trong liên kết của Ca^{2+} với cơ chất và chính imidazole cung cấp một cơ chế để ổn định, giúp kích hoạt hoạt động của enzyme [41]. Như vậy imidazol là thành phần giúp ổn định cơ chế xúc tác của *Xbx14* nên không làm giảm hoạt tính của *Xbx14*. Tuy nhiên 2-mercaptoethanol là chất phá hủy các liên kết disunfit và tượng tự như Urê làm biến tính protein, nên hoạt tính của *Xbx14* bị giảm. Như vậy ảnh hưởng của một số ion kim loại đến hoạt tính của *Xbx14* có một số sai khác so với các nghiên cứu đã công bố. Kết quả này sẽ là cơ sở để hạn chế các loại hóa chất làm giảm hoạt tính của enzyme hoặc sử dụng ion thích hợp làm tăng hoạt tính của *Xbx14* khi cần thiết trong phản ứng thủy phân sinh khối thực vật.

3.3.3.7. Nghiên cứu đặc điểm động học của Xbx14

Hằng số Km của *Xbx14* được xác định bởi dải nồng độ cơ chất pNPX từ 2,0 đến 20 mg/ml trong đệm phosphate pH = 9. Thông số động học được xác định từ đồ thị Lineweaver-Burke. Kết quả cho thấy, với lượng protein *Xbx14* là 0,253 mg/ml, phương trình phụ thuộc giữa tốc độ phản ứng với nồng độ cơ chất đã tuân theo hàm y = ax+b với độ tin cậy cao R^2 = 0,9937 (Hình 3.30). Dựa trên phương trình, Km và Vmax của *Xbx14* tương ứng là 3,18 mM và 6,5 µmol/phút. Do đó, hoạt tính riêng của *Xbx14* đạt được là 24,54 U/mg.

Một số nghiên cứu công bố Km của β–xylosidase thuộc họ GH43 ở vi khuẩn nằm trong khoảng 0,038 mM đến 28 mM. Ví dụ Km của β–xylosidase từ *Vibrio* sp. XY-214 là 0,244 mM [189]; *Clostridium acetobutylicum* là 3,7 mM [101]; *Thermoanaerobacterium* là 28 mM [150]. Như vậy, so với β–xylosidase tạo ra từ vi khuẩn khác nhau thì các giá trị giá trị Km của *Xbx14* là 3,18 mM là khá tốt so với các nghiên cứu đã công bố. Giá trị Vmax của β–xylosidase thuộc họ GH43 ở các loài vi khuẩn cũng biến thiên khá rộng. Ví dụ như Vmax của *Paenibacillus woosongensis* là 1,4 µmol/phút [91]; *Bacillus pumilus* 1,23 µmol/phút [186];

Clostridium stercorarium là 15 µmol/phút [162],… Với *Xbx14*, Vmax là 6,5 µmol/phút, nằm trong giá trị trung gian dải biến thiên tốc độ phản ứng của β–xylosidase từ vi khuẩn.



Hình 3.30. Sự phụ thuộc của tốc độ phản ứng của enzyme vào nồng độ cơ chất pNPX theo Linewever-Burk

Các nghiên cứu về hoạt tính đặc hiệu của β–xylosidase từ nấm có tần số biến thiên từ rất thấp (2 U/mg) đến rất cao (50 U/mg) ở *Fusarium sp.* [141]. Sự biến thiên này cũng tương tự như ở vi khuẩn, có thể từ 1,25 U/mg ở *Caulobacter crescentus* [65] đến 17,5 U/mg ở *Clostridium acetobutylicum* [101], …. Như vậy *Xbx14* có hoạt tính riêng là 24,54 U/mg cao hơn hẳn hoạt tính của đa số β–xylosidase có nguồn gốc từ nấm và vi khuẩn đã công bố.

Tóm lại kết quả nghiên cứu hoạt tính của *Xbx14* cho thấy đây là một enzyme mới, có khả năng hoạt động tối ưu ở nhiệt độ cao (60°C), môi trường kiềm (pH = 9) và bền nhiệt, với hoạt tính riêng tốt là 24,54 U/mg. Enzyme này hứa hẹn sẽ mang lại nhiều lợi ích kinh tế khi kết hợp với phương pháp tiền xử lý sinh khối thực vật bằng kiềm và nhiệt trong thực tiễn.

KẾT LUẬN VÀ KIẾN NGHỊ

Kết luận

1. Nghiên cứu đã khai thác được 20 loài vi khuẩn có số lượng ORF lớn phù hợp với chế độ ăn gỗ giàu lignocellulose và nghèo Nitơ của mối *C. gestroi* từ dữ liệu trình tự DNA đa hệ gen. Trong đó có 03 loài tham gia chuyển hóa Nitơ và 05 loài tham gia thủy phân lignocellulose không có ở các loài mối khác. Đồng thời xác định các lignocellulase của vi sinh vật ruột mối *C. gestroi* thuộc về 52 họ GH khác nhau theo phân loại của CAZY.
2. Xây dựng được phương pháp tìm kiếm gen từ dữ liệu trình tự DNA đa hệ gen bằng mẫu dò dựa vào các trình tự axit amin đã nghiên cứu tính chất và sự hỗ trợ của các công cụ tin sinh học, theo định hướng ứng dụng về khả năng chịu nhiệt và chịu kiềm của enzyme. Sử dụng mẫu dò thuộc họ GH43 có độ dài là 464 axit amin đã tìm kiếm được gen GL0112518 mã hóa β–xylosidase từ dữ liệu DNA đa hệ gen của vi khuẩn ruột mối *C. gestroi;*
3. Biểu hiện thành công gen *Xbx14* trong tế bào *E. coli Rosetta 1* có hoạt tính β–xylosidase đúng như dự đoán ban đầu, với các thông số động học theo phương trình Lineweaver-Burk là Km = 3,18 mM, Vmax = 6,5 μmol/phút và hoạt tính riêng đạt được là 24,54 U/mg. Đây là một enzyme mới, có hoạt tính tốt với khả năng hoạt động tối ưu trong điều kiện nhiệt độ cao và pH kiềm.

Kiến nghị

Thử nghiệm sử dụng enzyme *Xbx14* kết hợp với các cellulase và hemicellulase khác để làm hỗn hợp enzyme thủy phân hoàn toàn lignocellulose.

DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ

1. Nguyễn Minh Giang, Đỗ Thị Huyền, Trương Nam Hải (2015). Sử dụng công cụ tin sinh trong nghiên cứu metagenomic - hướng nghiên cứu và ứng dụng mới trong sinh học. Tạp chí khoa học Trường ĐHSP TP.HCM, số tháng 2 năm 2015.

2. Nguyễn Minh Giang, Đỗ Thị Huyền, Trương Nam Hải (2016). Sử dụng một số công cụ tin sinh khai thác gen mã hóa enzyme thủy phân lignocellulose từ dữ liệu metagenome của vi sinh vật trong ruột mối *Coptotermes gestroi*. Tạp chí Công nghệ Sinh học 14(1): 39-47.

3. Nguyen Minh Giang, Do Thi Huyen, Truong Nam Hai (2016). In silicon mining for alkaline enzymes from metagenomic DNA data of gut microbes of the lower termite *Coptotermes gestroi* in Viet Nam. Tạp chí Sinh học 38(3): 374-383. DOI: 10.15625/0866-7160/v38n3.7811.

4. Nguyễn Minh Giang, Đỗ Thị Huyền, Phùng Thu Nguyệt, Trương Nam Hải (2017). Xây dựng probe để khai thác và chọn gen mã hóa xylan 1-4 beta xylosidase từ dữ liệu giải trình tự đa hệ gen. Tạp Chí Công Nghệ Sinh Học. *Chấp nhận đăng*.

5. Nguyễn Minh Giang, Nguyễn Minh Thu, Nguyễn Thị Duyên, Đỗ Thị Huyền, Hồ Thị Thương, Phạm Bích Ngọc, Chu Hoàng Hà, Trương Nam Hải (2017). Thiết kế vector biểu hiện pCB301-*Xbxs14*-ELP và biểu hiện gen Xbxs14 mã hóa Xylan 1,4-beta xylosidase từ vi sinh vật ruột mối *Coptotermes gestroi* trong cây thuốc lá *Nicotiana benthamiana. Tạp chí Sinh học.* Tạp chí Sinh học, 39(2): 239-248. DOI: 10.15625/0866-7160/v39n2.9843.

6. Nguyễn Minh Giang, Phạm Hải Như, Nguyễn Minh Thu, Đỗ Thị Huyền, Đinh Nho Thái, Trương Nam Hải (2017). Biểu hiện gen *Xbxs14* mã hóa Xylan 1,4-beta xylosidase có nguồn gốc từ vi sinh vật ruột mối *Coptotermes gestroi* trong tế bào *Escherichia coli* Rosetta (DE3). Tạp Chí Công Nghệ Sinh Học 15(3):1-7.

TÀI LIỆU THAM KHẢO

1. Abe, T., Bignell, D.E., and Higashi, M. (2000) Termites: evolution, sociality, symbioses, ecology. Kluwer Academic Publishers, Dordrecht; Boston.

2. Abot, A., Arnal, G., Auer, L., Lazuka, A., Labourdette, D., Lamarre, S., et al. (2016) *CAZyChip: dynamic assessment of exploration of glycoside hydrolases in microbial ecosystems*. BMC Genomics. 17 pp. 671.

3. Adelsberger, H., Hertel, C., Glawischnig, E., Zverlov, V.V., and Schwarz, W.H. (2004) *Enzyme system of Clostridium stercorarium for hydrolysis of arabinoxylan: reconstitution of the in vivo system from recombinant enzymes*. Microbiology. 150 (7), pp. 2257–2266.

4. Agrawal, A., Tripp, R.A., Anderson, L.J., and Nie, S. (2005) *Real-Time Detection of Virus Particles and Viral Protein Expression with Two-Color Nanoparticle Probes*. Journal of Virology. 79 (13), pp. 8625–8628.

5. Akama, T., Kawashima, A., Tanigawa, K., Hayashi, M., Ishido, Y., Luo, Y., et al. (2013) *Comprehensive Analysis of Prokaryotes in Environmental Water Using DNA Microarray Analysis and Whole Genome Amplification*. Pathogens. 2 (4), pp. 591–605.

6. Almeida, E.M. de, Polizeli, M. de L.T.M., Terenzi, H.F., and Jorge, J.A. (1995) *Purification and biochemical characterization of beta-xylosidase from Humicola grisea var. thermoidea*. FEMS Microbiology Letters.

7. Anwar, Z., Gulfraz, M., and Irshad, M. (2014) *Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: A brief review*. Journal of Radiation Research and Applied Sciences. 7 (2), pp. 163–173.

8. Ariyarathna, H.A.C.K. and Francki, M.G. (2016) *Phylogenetic relationships and protein modelling revealed two distinct subfamilies of group II HKT genes between crop and model grasses*. Genome. 59 (7), pp. 509–517.

9. Bagger, F.O., Rapin, N., Theilgaard-Mönch, K., Kaczkowski, B., Jendholm, J., Winther, O., et al. (2012) *HemaExplorer: a Web server for easy and fast visualization of gene expression in normal and malignant hematopoiesis*. Blood. 119 (26), pp. 6394–6395.

10. Bairoch, A. and Apweiler, R. (2000) *The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000*. Nucleic Acids Research. 28 (1), pp. 45–48.

11. Bakare, M.K., Adewale, I.O., Ajayi, A., and Shonukan, O.O. (2005) *Purification and characterization of cellulase from the wild-type and two improved mutants of Pseudomonas fluorescens*. ResearchGate. 4 (9), pp. 898–904.

12. Baldwin, D.A., Feldman, M., Alwine, J.C., and Robertson, E.S. (2014) *Metagenomic Assay for Identification of Microbial Pathogens in Tumor Tissues*. mBio. 5 (5), pp. e01714-14.

13. Bandyopadhyay, A. (2015) *Bioprocessing Technologies in Biorefinery for Sustainable Production of Fuels, Chemicals, and Polymes, Edited by: Shang-Tian Yang, Hesham El-Ensashy & Nuttha Thongchul, 2nd Edition,*. CLEAN – Soil, Air, Water. 43 (11), pp. 1556–1557.

14. Banka, A.L., Albayrak Guralp, S., and Gulari, E. (2014) *Secretory Expression and Characterization of Two Hemicellulases, Xylanase, and β-Xylosidase, Isolated from Bacillus Subtilis M015*. Applied Biochemistry and Biotechnology. 174 (8), pp. 2702–2710.

15. Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., et al. (2013) *Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics*. Biotechnology for Biofuels. 6 (1), pp. 78.

16. Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., et al. (2013) *Mining for hemicellulases in the fungus-growing termite*

Pseudacanthotermes militaris using functional metagenomics. Biotechnology for Biofuels. 6 pp. 78.

17. Beguin, P. (1990) *Molecular Biology of Cellulose Degradation*. Annual Review of Microbiology. 44 (1), pp. 219–248.

18. Benassi, V.M., de Lucas, R.C., Jorge, J.A., and Polizeli, M. de L.T. de M. (2015) *Screening of thermotolerant and thermophilic fungi aiming β-xylosidase and arabinanase production*. Brazilian Journal of Microbiology. 45 (4), pp. 1459–1467.

19. Bhagwat, A.S., Sohail, A., and Roberts, R.J. (1986) *Cloning and characterization of the dcm locus of Escherichia coli K-12*. Journal of Bacteriology. 166 (3), pp. 751–755.

20. Biely, P. (1985) *Microbial xylanolytic systems*. Trends in Biotechnology. 3 (11), pp. 286–290.

21. Bomar, L., Maltz, M., Colston, S., and Graf, J. (2011) *Directed Culturing of Microorganisms Using Metatranscriptomics*. mBio. 2 (2), pp. e00012-11.

22. Bravman, T., Zolotnitsky, G., Shulami, S., Belakhov, V., Solomon, D., Baasov, T., et al. (2001) *Stereochemistry of family 52 glycosyl hydrolases: a beta-xylosidase from Bacillus stearothermophilus T-6 is a retaining enzyme*. FEBS Letters. 495 (1–2), pp. 39–43.

23. Brune, A., Emerson, D., and Breznak, J.A. (1995) *The Termite Gut Microflora as an Oxygen Sink: Microelectrode Determination of Oxygen and pH Gradients in Guts of Lower and Higher Termites.* Applied and Environmental Microbiology. 61 (7), pp. 2681–2687.

24. Brune, A. (2014) *Symbiotic digestion of lignocellulose in termite guts*. Nature Reviews Microbiology. 12 (3), pp. 168–180.

25. Brune, A. (2012) Microbial Symbioses in the Digestive Tract of Lower Termites. in: E. Rosenberg, U. Gophna (Eds.), Benef. Microorg. Multicell. Life Forms, Springer Berlin Heidelberg, pp. 3–25.

26. Brüx, C., Ben-David, A., Shallom-Shezifi, D., Leon, M., Niefind, K., Shoham, G., et al. (2006) *The structure of an inverting GH43 beta-xylosidase from Geobacillus stearothermophilus with its substrate reveals the role of the three catalytic residues*. Journal of Molecular Biology. 359 (1), pp. 97–109.

27. Campos, E., Negro Alvarez, M.J., Sabarís di Lorenzo, G., Gonzalez, S., Rorig, M., Talia, P., et al. (2014) *Purification and characterization of a GH43 β-xylosidase from Enterobacter sp. identified and cloned from forest soil bacteria*. Microbiological Research. 169 (2–3), pp. 213–220.

28. Chinchetru, M.A., Cabezas, J.A., and Calvo, P. (1989) *Purification and characterization of a broad specificity beta-glucosidase from sheep liver*. The International Journal of Biochemistry. 21 (5), pp. 469–476.

29. Cho, M.-J., Kim, Y.-H., Shin, K., Kim, Y.-K., Kim, Y.-S., and Kim, T.-J. (2010) *Symbiotic adaptation of bacteria in the gut of Reticulitermes speratus: low endo-beta-1,4-glucanase activity*. Biochemical and Biophysical Research Communications. 395 (3), pp. 432–435.

30. Chroneos, Z.C. (2010) *Metagenomics: Theory, methods, and applications*. Human Genomics. 4 (4), pp. 282–283.

31. Chung, W.-H., Rhee, S.-K., Wan, X.-F., Bae, J.-W., Quan, Z.-X., and Park, Y.-H. (2005) *Design of long oligonucleotide probes for functional gene detection in a microbial community*. Bioinformatics (Oxford, England). 21 (22), pp. 4092–4100.

32. Cleveland, L.R. (1926) *Symbiosis Among Animals with Special Reference to Termites and Their Intestinal Flagellates*. The Quarterly Review of Biology. 1 (1), pp. 51–60.

33. Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., et al. (2009) *The Ribosomal Database Project: improved alignments and new tools for rRNA analysis*. Nucleic Acids Research. 37 (Database issue), pp. D141-145.

34. Corrêa, J.M., Christi, D., Torre, C.L.D., Henn, C., Conceição-Silva, J.L. da, Kadowaki, M.K., et al. (2016) *High levels of β-xylosidase in Thermomyces lanuginosus: potential use for saccharification*. Brazilian Journal of Microbiology. 47 (3), pp. 680.

35. Corrêa, J.M., Christi, D., Torre, C.L.D., Henn, C., Conceição-Silva, J.L. da, Kadowaki, M.K., et al. (2016) *High levels of β-xylosidase in Thermomyces lanuginosus: potential use for saccharification*. Brazilian Journal of Microbiology. 47 (3), pp. 680.

36. Costa-Leonardo, A.M. and Haifig, I. (2014) Termite Communication During Different Behavioral Activities. in: Biocommunication Anim., Springer, Dordrecht, pp. 161–190.

37. Cullen, D. and Kersten, P.J. (1996) Enzymology and Molecular Biology of Lignin Degradation. in: Biochem. Mol. Biol., Springer, Berlin, Heidelberg, pp. 295–312.

38. Culligan, E.P., Sleator, R.D., Marchesi, J.R., and Hill, C. (2014) *Metagenomics and novel gene discovery: promise and potential for novel therapeutics*. Virulence. 5 (3), pp. 399–412.

39. Curtis, M.A., Aduse-Opoku, J., and Rangarajan, M. (2001) *Cysteine proteases of Porphyromonas gingivalis*. Critical Reviews in Oral Biology and Medicine: An Official Publication of the American Association of Oral Biologists. 12 (3), pp. 192–216.

40. Dashtban, M., Maki, M., Leung, K.T., Mao, C., and Qin, W. (2010) *Cellulase activities in biomass conversion: measurement methods and comparison*. Critical Reviews in Biotechnology. 30 (4), pp. 302–309.

41. Db, J., Jd, B., K, W., Cc, L., Vj, C., I, D., et al. (2015) *X-ray Crystal Structure of Divalent Metal-Activated beta-xylosidase, RS223BX*. Applied Biochemistry and Biotechnology. 177 (3), pp. 637–648.

42. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008) *Phylogeny.fr: robust phylogenetic analysis for the non-specialist*. Nucleic Acids Research. 36 (Web Server issue), pp. W465-469.

43. Desai, M.S. and Brune, A. (2012) *Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites*. The ISME Journal. 6 (7), pp. 1302–1313.

44. Desjardins, P., Hansen, J.B., and Allen, M. (2009) *Microvolume Protein Concentration Determination using the NanoDrop 2000c Spectrophotometer*. Journal of Visualized Experiments : JoVE. (33),.

45. Do, T.H., Nguyen, T.T., Nguyen, T.N., Le, Q.G., Nguyen, C., Kimura, K., et al. (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*. Journal of Bioscience and Bioengineering. 118 (6), pp. 665–671.

46. Dolan, M.F. (2001) *Speciation of termite gut protists: the role of bacterial symbionts*. International Microbiology: The Official Journal of the Spanish Society for Microbiology. 4 (4), pp. 203–208.

47. Dougherty, M.J., D'haeseleer, P., Hazen, T.C., Simmons, B.A., Adams, P.D., and Hadi, M.Z. (n.d.) (2012) *Glycoside Hydrolases from a targeted Compost Metagenome, activity-screening and functional characterization*. BMC Biotechnology. 12.

48. Dröge, S., Fröhlich, J., Radek, R., and König, H. (2006) *Spirochaeta coccoides sp. nov., a Novel Coccoid Spirochete from the Hindgut of the Termite Neotermes castaneus*. Applied and Environmental Microbiology. 72 (1), pp. 392–397.

49. Duan, C.-J., Xian, L., Zhao, G.-C., Feng, Y., Pang, H., Bai, X.-L., et al. (2009) *Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens*. Journal of Applied Microbiology. 107 (1), pp. 245–256.

50. Dugat-Bony, E., Missaoui, M., Peyretaillade, E., Biderre-Petit, C., Bouzid, O., Gouinaud, C., et al. (2011) *HiSpOD: probe design for functional DNA microarrays*. Bioinformatics (Oxford, England). 27 (5), pp. 641–648.

51. Dunne, C., Crowley, J.J., Moënne-Loccoz, Y., Dowling, D.N., Bruijn, s, and O'Gara, F. (1997) *Biological control of Pythium ultimum by Stenotrophomonas maltophilia W81 is mediated by an extracellular proteolytic activity*. Microbiology. 143 (12), pp. 3921–3931.

52. Duong-Ly, K.C. and Gabelli, S.B. (2014) *Salting out of proteins using ammonium sulfate precipitation*. Methods in Enzymology. 541 pp. 85–94.

53. Dvorak, P., Chrast, L., Nikel, P.I., Fedr, R., Soucek, K., Sedlackova, M., et al. (2015) *Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway*. Microbial Cell Factories. 14 pp. 201.

54. Dyson, Z.A., Seviour, R.J., Tucci, J., and Petrovski, S. (2016) *Genome Sequences of Pseudomonas oryzihabitans Phage POR1 and Pseudomonas aeruginosa Phage PAE1*. Genome Announcements. 4 (3), pp. e01515-15.

55. Ebert, A. and Brune, A. (1997) *Hydrogen Concentration Profiles at the Oxic-Anoxic Interface: a Microsensor Study of the Hindgut of the Wood-Feeding Lower Termite Reticulitermes flavipes (Kollar).* Applied and Environmental Microbiology. 63 (10), pp. 4039–4046.

56. El-Gindy, A.A., Saad, R.R., and Fawzi, E.M. (2015) *Purification of β-xylosidase from Aspergillus tamarii using ground oats and a possible application on the fermented hydrolysate by Pichia stipitis*. Annals of Microbiology. 65 (2), pp. 965–974.

57. Ellis, J.T. and Magnuson, T.S. (2012) *Thermostable and Alkalistable Xylanases Produced by the Thermophilic Bacterium Anoxybacillus flavithermus TWXYL3*. ISRN Microbiology. 2012 pp. 517524.

58. Elshahed, M.S., Youssef, N.H., Luo, Q., Najar, F.Z., Roe, B.A., Sisk, T.M., et al. (2007) *Phylogenetic and Metabolic Diversity of Planctomycetes from Anaerobic, Sulfide- and Sulfur-Rich Zodletone Spring, Oklahoma*. Applied and Environmental Microbiology. 73 (15), pp. 4707–4716.

59. Fekete, C.A. and Kiss, L. (2012) *Purification and characterization of a recombinant β-D-xylosidase from Thermobifida fusca TM51*. The Protein Journal. 31 (8), pp. 641–650.

60. Fisher, M.L. (2006) *Comparison of Subterranean Termite (Rhinotermitidae: Reticulitermes) Gut Bacterial Diversity Within and Between Colonies and to Other Termite Species Using Molecular Techniques (ARDRA and 16S rRNA Gene Sequencing)*. ResearchGate.

61. Franco Cairo, J.P.L., Leonardo, F.C., Alvarez, T.M., Ribeiro, D.A., Büchli, F., Costa-Leonardo, A.M., et al. (2011) *Functional characterization and target discovery of glycoside hydrolases from the digestome of the lower termite Coptotermes gestroi*. Biotechnology for Biofuels. 4 pp. 50.

62. GenuGreen (2012) *Bio-based products over Petroleum-based products*. Denmark Business-Focused Energy Study Tour.

63. Ghaffar, S.H. and Fan, M. (2013) *Structural analysis for lignin characteristics in biomass straw*. Biomass and Bioenergy. 57 pp. 264–279.

64. Glenn, T.C. (2011) *Field guide to next-generation DNA sequencers*. Molecular Ecology Resources. 11 (5), pp. 759–769.

65. Graciano, L., Corrêa, J.M., Gandra, R.F., Seixas, F.A.V., Kadowaki, M.K., Sampaio, S.C., et al. (2012) *The cloning, expression, purification, characterization and modeled structure of Caulobacter crescentus β-*

Xylosidase I. World Journal of Microbiology and Biotechnology. 28 (9), pp. 2879–2888.

66. Grodberg, J. and Dunn, J.J. (1988) *ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymease during purification*. Journal of Bacteriology. 170 (3), pp. 1245–1253.

67. Guerfali, M., Gargouri, A., and Belghith, H. (2008) *Talaromyces thermophilus beta-D-xylosidase: purification, characterization and xylobiose synthesis*. Applied Biochemistry and Biotechnology. 150 (3), pp. 267–279.

68. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) *Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products*. Chemistry & Biology. 5 (10), pp. R245-249.

69. Handelsman, J. (2004) *Metagenomics: application of genomics to uncultured microorganisms*. Microbiology and Molecular Biology Reviews: MMBR. 68 (4), pp. 669–685.

70. He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H., et al. (2013) *Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites*. PloS One. 8 (4), pp. e61126.

71. Henrissat, B. and Davies, G. (1997) *Structural and sequence-based classification of glycoside hydrolases*. Current Opinion in Structural Biology. 7 (5), pp. 637–644.

72. Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., et al. (2011) *Metagenomic discovery of biomass-degrading genes and genomes from cow rumen*. Science (New York, N.Y.). 331 (6016), pp. 463–467.

73. Holben, W.E., Jansson, J.K., Chelm, B.K., and Tiedje, J.M. (1988) *DNA Probe Method for the Detection of Specific Microorganisms in the Soil*

Bacterial Community. Applied and Environmental Microbiology. 54 (3), pp. 703–711.

74. Holst-Jensen, A., Bertheau, Y., de Loose, M., Grohmann, L., Hamels, S., Hougs, L., et al. (2012) *Detecting un-authorized genetically modified organisms (GMOs) and derived materials*. Biotechnology Advances. 30 (6), pp. 1318–1335.

75. Hongoh, Y. (2010) *Diversity and genomes of uncultured microbial symbionts in the termite gut*. Bioscience, Biotechnology, and Biochemistry. 74 (6), pp. 1145–1151.

76. Hongoh, Y., Deevong, P., Hattori, S., Inoue, T., Noda, S., Noparatnaraporn, N., et al. (2006) *Phylogenetic diversity, localization, and cell morphologies of members of the candidate phylum TG3 and a subphylum in the phylum Fibrobacteres, recently discovered bacterial groups dominant in termite guts*. Applied and Environmental Microbiology. 72 (10), pp. 6780–6788.

77. Hongoh, Y., Ohkuma, M., and Kudo, T. (2003) *Molecular analysis of bacterial microbiota in the gut of the termite Reticulitermes speratus (Isoptera; Rhinotermitidae)*. FEMS Microbiology Ecology. 44 (2), pp. 231–242.

78. Howard, B.H., Jones, G., and Purdom, M.R. (1960) *The pentosanases of some rumen bacteria*. Biochemical Journal. 74 (1), pp. 173–180.

79. Huang, Z., Liu, X., Zhang, S., and Liu, Z. (2014) *GH52 xylosidase from Geobacillus stearothermophilus: characterization and introduction of xylanase activity by site‐directed mutagenesis of Tyr509*. Journal of Industrial Microbiology & Biotechnology. 41 (1), pp. 65–74.

80. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) *MEGAN analysis of metagenomic data*. Genome Research. 17 (3), pp. 377–386.

81. Hyun, Y.-J., Kim, B., and Kim, D.-H. (2012) *Cloning and characterization of ginsenoside Ra1-hydrolyzing beta-D-xylosidase from Bifidobacterium breve K-110*. Journal of Microbiology and Biotechnology. 22 (4), pp. 535–540.

82. Iida, null, Ohkuma, null, Ohtoko, null, and Kudo, null (2000) *Symbiotic spirochetes in the termite hindgut: phylogenetic identification of ectosymbiotic spirochetes of oxymonad protists*. FEMS Microbiology Ecology. 34 (1), pp. 17–26.

83. Information, N.C. for B., Pike, U.S.N.L. of M.R., BethesdaMD, and 20894USA (n.d.) (2005) *1YIF: Crystal Structure Of Beta-1,4-Xylosidase From Bacillus Subtilis, New York Structural Genomics Consortium - NCBI structure*.

84. Jhamb, K. and Sahoo, D.K. (2012) *Production of soluble recombinant proteins in Escherichia coli: Effects of process conditions and chaperone co-expression on cell growth and production of xylanase*. Bioresource Technology. 123 pp. 135–143.

85. Jordan, D.B., Braker, J.D., Wagschal, K., Lee, C.C., Chan, V.J., Dubrovska, I., et al. (2015) *X-ray Crystal Structure of Divalent Metal-Activated β-xylosidase, RS223BX*. Applied Biochemistry and Biotechnology. 177 (3), pp. 637–648.

86. Jordan, D.B. and Wagschal, K. (2010) *Properties and applications of microbial beta-D-xylosidases featuring the catalytically efficient enzyme from Selenomonas ruminantium*. Applied Microbiology and Biotechnology. 86 (6), pp. 1647–1658.

87. Kappler, A. and Brune, A. (1999) *Influence of gut alkalinity and oxygen status on mobilization and size-class distribution of humic acids in the hindgut of soil-feeding termites*. Applied Soil Ecology. 13 (3), pp. 219–229.

88. Khan, S., Sistla, S., Dhodapkar, R., and Parija, S.C. (2012) *Fatal Delftia acidovorans infection in an immunocompetent patient with empyema*. Asian Pacific Journal of Tropical Biomedicine. 2 (11), pp. 923–924.

89. Khánh H.Q. and Ngọc N.B. (2014) *Ứng dụng kỹ thuật metagenomics để nhận diện gene mã hóa laccase của nấm Basidiomycetes trong mẫu đất rừng Nam Cát Tiên*. Tạp chí Phát triển Khoa học và Công nghệ. 16 (3T), pp. 60–74.

90. Khow, O. and Suntrarachun, S. (2012) *Strategies for production of active eukaryotic proteins in bacterial expression system*. Asian Pacific Journal of Tropical Biomedicine. 2 (2), pp. 159–162.

91. Kim, Y.A. and Yoon, K.-H. (2010) *Characterization of a Paenibacillus woosongensis beta-Xylosidase/alpha-Arabinofuranosidase produced by recombinant Escherichia coli*. Journal of Microbiology and Biotechnology. 20 (12), pp. 1711–1716.

92. König, H. (2006) *Bacillus species in the intestine of termites and other soil invertebrates*. Journal of Applied Microbiology. 101 (3), pp. 620–627.

93. Koshland, D.E. (1953) *Stereochemistry and the Mechanism of Enzymatic Reactions*. Biological Reviews. 28 (4), pp. 416–436.

94. Ku, T., Lu, P., Chan, C., Wang, T., Lai, S., Lyu, P., et al. (2009) *Predicting melting temperature directly from protein sequences*. Computational Biology and Chemistry. 33 (6), pp. 445–450.

95. Kudo, T. (2009) *Termite-microbe symbiotic system and its efficient degradation of lignocellulose*. Bioscience, Biotechnology, and Biochemistry. 73 (12), pp. 2561–2567.

96. Kundu, A. and Ray, R.R. (2013) *Production of intracellular β-xylosidase from the submerged fermentation of citrus wastes by Penicillium janthinellum MTCC 10889*. 3 Biotech. 3 (3), pp. 241–246.

97. Kushwaha, S.K., Manoharan, L., Meerupati, T., Hedlund, K., and Ahrén, D. (2015) *MetCap: a bioinformatics probe design pipeline for large-scale targeted metagenomics*. BMC Bioinformatics. 16 pp. 65.

98. La Grange, D.C., Pretorius, I.S., and van Zyl, W.H. (1997) *Cloning of the Bacillus pumilus beta-xylosidase gene (xynB) and its expression in Saccharomyces cerevisiae*. Applied Microbiology and Biotechnology. 47 (3), pp. 262–266.

99. Labutti, K., Mayilraj, S., Clum, A., Lucas, S., Glavina Del Rio, T., Nolan, M., et al. (2010) *Permanent draft genome sequence of Dethiosulfovibrio peptidovorans type strain (SEBR 4207)*. Standards in Genomic Sciences. 3 (1), pp. 85–92.

100. Lee, C.C., Braker, J.D., Grigorescu, A.A., Wagschal, K., and Jordan, D.B. (2013) *Divalent metal activation of a GH43 β-xylosidase.* Enzyme and Microbial Technology. 52 (2), pp. 84–90.

101. Lee, S.F. and Forsberg, C.W. (1987) *Isolation and Some Properties of a β-d-Xylosidase from Clostridium acetobutylicum ATCC 824*. Applied and Environmental Microbiology. 53 (4), pp. 651–654.

102. Lelie, D. van der, Taghavi, S., McCorkle, S.M., Li, L.-L., Malfatti, S.A., Monteleone, D., et al. (2012) *The Metagenome of an Anaerobic Microbial Community Decomposing Poplar Wood Chips*. PLOS ONE. 7 (5), pp. e36740.

103. Li, X., He, Z., and Zhou, J. (2005) *Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation*. Nucleic Acids Research. 33 (19), pp. 6114–6123.

104. Lin, H., Chen, W., and Ding, H. (2013) *AcalPred: a sequence-based tool for discriminating between acidic and alkaline enzymes*. PloS One. 8 (10), pp. e75726.

105. Lin, R., Glazebrook, J., Katagiri, F., Orf, J.H., and Gibson, S.I. (2015) *Identification of differentially expressed genes between developing seeds of different soybean cultivars*. Genomics Data. 6 pp. 92–98.

106. Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012) *Comparison of Next-Generation Sequencing Systems*. Journal of Biomedicine and Biotechnology. 2012.

107. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) *The carbohydrate-active enzymes database (CAZy) in 2013*. Nucleic Acids Research. 42 (Database issue), pp. D490-495.

108. Maki, M., Leung, K.T., and Qin, W. (2009) *The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass*. International Journal of Biological Sciences. 5 (5), pp. 500–516.

109. Martínez-Alonso, M., García-Fruitós, E., Ferrer-Miralles, N., Rinas, U., and Villaverde, A. (2010) *Side effects of chaperone gene co-expression in recombinant protein production*. Microbial Cell Factories. 9 pp. 64.

110. Martins, L.F., Antunes, L.P., Pascon, R.C., Oliveira, J.C.F. de, Digiampietri, L.A., Barbosa, D., et al. (2013) *Metagenomic Analysis of a Tropical Composting Operation at the São Paulo Zoo Park Reveals Diversity of Biomass Degradation Functions and Organisms*. PLOS ONE. 8 (4), pp. e61928.

111. Maruthamuthu, M. and van Elsas, J.D. (2017) *Molecular cloning, expression, and characterization of four novel thermo-alkaliphilic enzymes retrieved from a metagenomic library*. Biotechnology for Biofuels. 10 pp. 142.

112. Mathew, G.M., Ju, Y.-M., Lai, C.-Y., Mathew, D.C., and Huang, C.C. (2012) *Microbial community analysis in the termite gut and fungus comb of Odontotermes formosanus: the implication of Bacillus as mutualists*. FEMS Microbiology Ecology. 79 (2), pp. 504–517.

113. Matsuo, M. and Yasui, T. (1984) *Purification and Some Properties of β-Xylosidase from Emericella nidulans*. Agricultural and Biological Chemistry. 48 (7), pp. 1853–1860.

114. Matsuzawa, T., Kaneko, S., and Yaoi, K. (2015) *Screening, identification, and characterization of a GH43 family β-xylosidase/α-arabinofuranosidase from a compost microbial metagenome*. Applied Microbiology and Biotechnology. 99 (21), pp. 8943–8954.

115. Mattéotti, C., Haubruge, E., Thonart, P., Francis, F., De Pauw, E., Portetelle, D., et al. (2011) *Characterization of a new β-glucosidase/β-xylosidase from the*

gut microbiota of the termite (Reticulitermes santonensis). FEMS Microbiology Letters. 314 (2), pp. 147–157.

116. Mattéotti, C., Haubruge, E., Thonart, P., Francis, F., De Pauw, E., Portetelle, D., et al. (2011) *Characterization of a new β-glucosidase/β-xylosidase from the gut microbiota of the termite (Reticulitermes santonensis)*. FEMS Microbiology Letters. 314 (2), pp. 147–157.

117. Mhetras, N., Liddell, S., and Gokhale, D. (2016) *Purification and characterization of an extracellular β-xylosidase from Pseudozyma hubeiensis NCIM 3574 (PhXyl), an unexplored yeast*. AMB Express. 6 pp. 73.

118. Militon, C., Rimour, S., Missaoui, M., Biderre, C., Barra, V., Hill, D., et al. (2007) *PhylArray: phylogenetic probe design algorithm for microarray*. Bioinformatics (Oxford, England). 23 (19), pp. 2550–2557.

119. Missiakas, D. and Raina, S. (1997) *Protein folding in the bacterial periplasm*. Journal of Bacteriology. 179 (8), pp. 2465–2471.

120. Mitra, S., Rupek, P., Richter, D.C., Urich, T., Gilbert, J.A., Meyer, F., et al. (2011) *Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG*. BMC Bioinformatics. 12 (1), pp. 1–8.

121. Mitsuhashi, M., Cooper, A., Ogura, M., Shinagawa, T., Yano, K., and Hosokawa, T. (1994) *Oligonucleotide probe design--a new approach*. Nature. 367 (6465), pp. 759–761.

122. Mofeed, N.M.M. (2012) *In silico identification of potential biomass and cell wall degrading enzymes in the microbial community of the Red Sea Atlantis-II brine pool using metagenomic approach*.

123. Muro, M.A. de (2005) Probe Design, Production, and Applications. in: Med. Biomethods Handb., Humana Press, pp. 13–23.

124. Mustafa, G., Kousar, S., Rajoka, M.I., and Jamil, A. (2016) *Molecular cloning and comparative sequence analysis of fungal β-Xylosidases*. AMB Express. 6 (1), pp. 30.

125. Nguyen, D.-K., Nguyen, T.-V., Trinh, V.-H., Nguyen, V.-Q., Le, V.-T., Nguyen, T.-H., et al. (2007) Fauna of Vietnam. Termite: Isoptera. The Science and Technology of Ha Noi, Vietnam, Ha Noi.

126. Nguyen, T., Do, T., Duong, T., Le, Q., Dao, T., Nguyen, T., et al. (2014) *Identification of Vietnamese Coptotermes pest species based on the sequencing of two regions of 16S rRNA gene*. Bulletin of Insectology. 67 (1), pp. 131–136.

127. Nimchua, T., Thongaram, T., Uengwetwanit, T., Pongpattanakitshote, S., and Eurwilaichitr, L. (2012) *Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes*. Journal of Microbiology and Biotechnology. 22 (4), pp. 462–469.

128. Nordberg, E.K. (2005) *YODA: selecting signature oligonucleotides*. Bioinformatics (Oxford, England). 21 (8), pp. 1365–1370.

129. Nurizzo, D., Nagy, T., Gilbert, H.J., and Davies, G.J. (2002) *The structural basis for catalysis and specificity of the Pseudomonas cellulosa alpha-glucuronidase, GlcA67A*. Structure (London, England: 1993). 10 (4), pp. 547–556.

130. Odelson, D.A. and Breznak, J.A. (1985) *Nutrition and Growth Characteristics of Trichomitopsis termopsidis, a Cellulolytic Protozoan from Termites*. Applied and Environmental Microbiology. 49 (3), pp. 614–621.

131. Ohkuma, M. (2008) *Symbioses of flagellates and prokaryotes in the gut of lower termites*. Trends in Microbiology. 16 (7), pp. 345–352.

132. Pedersen, M., Lauritzen, H.K., Frisvad, J.C., and Meyer, A.S. (2007) *Identification of thermostable beta-xylosidase activities produced by Aspergillus brasiliensis and Aspergillus niger*. Biotechnology Letters. 29 (5), pp. 743–748.

133. Pinheiro, G.L., Correa, R.F., Cunha, R.S., Cardoso, A.M., Chaia, C., Clementino, M.M., et al. (2015) *Isolation of aerobic cultivable cellulolytic bacteria from different regions of the gastrointestinal tract of giant land snail Achatina fulica*. Frontiers in Microbiology. 6.

134. Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., et al. (2012) *eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges*. Nucleic Acids Research. 40 (Database issue), pp. D284-289.

135. Price, N.C. (1985) *The determination of Km values from lineweaver-burk plots*. Biochemical Education. 13 (2), pp. 81–81.

136. Pucci, F., Bourgeas, R., and Rooman, M. (2016) *Predicting protein thermal stability changes upon point mutations using statistical potentials: Introducing HoTMuSiC*. Scientific Reports. 6 pp. 23257.

137. Quiroz-Castañeda, R.E. and Folch-Mallol, J.L. (2013) *Hydrolysis of Biomass Mediated by Cellulases for the Production of Sugars*.

138. Radek, R. (n.d.) (1999) *Flagellates, bacteria, and fungi associated with termites: Diversity and function in nutrition - A review*. Ecotropica. 5 pp. 183–196.

139. Rosano, G.L. and Ceccarelli, E.A. (2014) *Recombinant protein expression in Escherichia coli: advances and challenges*. Frontiers in Microbiology. 5.

140. Rouillard, J.-M., Herbert, C.J., and Zuker, M. (2002) *OligoArray: genome-scale oligonucleotide design for microarrays*. Bioinformatics (Oxford, England). 18 (3), pp. 486–487.

141. Saha, B.C. (2003) *Purification and properties of an extracellular beta-xylosidase from a newly isolated Fusarium proliferatum*. Bioresource Technology. 90 (1), pp. 33–38.

142. Saini, J.K., Saini, R., and Tewari, L. (2015) *Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments*. 3 Biotech. 5 (4), pp. 337–353.

143. Sakka, K., Yoshikawa, K., Kojima, Y., Karita, S., Ohmiya, K., and Shimada, K. (1993) *Nucleotide sequence of the Clostridium stercorarium xylA gene encoding a bifunctional protein with beta-D-xylosidase and alpha-L-arabinofuranosidase activities, and properties of the translated product*. Bioscience, Biotechnology, and Biochemistry. 57 (2), pp. 268–272.

144. Sambrook, J. and Russell, D. (2001) Molecular cloning: a laboratory manual. in: Cold Spring Harbor Laboratory Press, New Yorkp. I (5.4-5.17).

145. San-Miguel, T., Pérez-Bermúdez, P., and Gavidia, I. (2013) *Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature*. SpringerPlus. 2.

146. Scharf, M.E. and Tartar, A. (2008) *Termite digestomes as sources for novel lignocellulases*. Biofuels, Bioproducts and Biorefining. 2 (6), pp. 540–552.

147. Schein, C.H. (1989) *Production of Soluble Recombinant Proteins in Bacteria*. Nature Biotechnology. 7 (11), pp. 1141–1149.

148. Shallom, D., Leon, M., Bravman, T., Ben-David, A., Zaide, G., Belakhov, V., et al. (2005) *Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from Geobacillus stearothermophilus T-6*. Biochemistry. 44 (1), pp. 387–397.

149. Shao, W., Xue, Y., Wu, A., Kataeva, I., Pei, J., Wu, H., et al. (2011) *Characterization of a novel beta-xylosidase, XylC, from Thermoanaerobacterium saccharolyticum JW/SL-YS485*. Applied and Environmental Microbiology. 77 (3), pp. 719–726.

150. Shao, W., Xue, Y., Wu, A., Kataeva, I., Pei, J., Wu, H., et al. (2011) *Characterization of a Novel β-Xylosidase, XylC, from Thermoanaerobacterium*

saccharolyticum JW/SL-YS485. Applied and Environmental Microbiology. 77 (3), pp. 719–726.

151. Sharpton, T.J. (2014) *An introduction to the analysis of shotgun metagenomic data*. Frontiers in Plant Science. 5.

152. Shi, J., Chinn, M.S., and Sharma-Shivappa, R.R. (2008) *Microbial pretreatment of cotton stalks by solid state cultivation of Phanerochaete chrysosporium*. Bioresource Technology. 99 (14), pp. 6556–6564.

153. Sim, M., Seok, H.-S., and Kim, J. (2013) *A Next-generation Sequence Clustering Method for E. Coli through Proteomics-genomics Data Mapping*. Procedia Computer Science. 23 pp. 96–101.

154. Singh, K.M., Reddy, B., Patel, A.K., Panchasara, H., Parmar, N., Patel, A.B., et al. (2014) *Metagenomic analysis of buffalo rumen microbiome: Effect of roughage diet on Dormancy and Sporulation genes*. Meta Gene. 2 pp. 252–268.

155. Siqueira, G., Bras, J., and Dufresne, A. (2010) *Cellulosic Bionanocomposites: A Review of Preparation, Properties and Applications*. Polymes. 2 (4), pp. 728–765.

156. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q. (2009) *Practical protocols for production of very high yields of recombinant proteins using Escherichia coli*. Protein Science : A Publication of the Protein Society. 18 (5), pp. 936–948.

157. Slaytor, M., Sugimoto, A., Azuma, J.-I., Murashima, K., and Inoue, T. (1997) *Cellulose and Xylan Utilisation in the Lower Termite Reticulitermes speratus*. Journal of Insect Physiology. 43 (3), pp. 235–242.

158. Sørensen, H.P. and Mortensen, K.K. (2005) *Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli*. Microbial Cell Factories. 4 pp. 1.

159. Sun, L., Liu, T., Müller, B., and Schnürer, A. (2016) *The microbial community structure in industrial biogas plants influences the degradation rate of straw and cellulose in batch tests*. Biotechnology for Biofuels. 9.

160. Sunna, A. and Antranikian, G. (1997) *Xylanolytic enzymes from fungi and bacteria*. Critical Reviews in Biotechnology. 17 (1), pp. 39–67.

161. SURYANI, KIMURA, T., SAKKA, K., and OHMIYA, K. (2004) *Sequencing and Expression of the Gene Encoding the Clostridium stercorarium β-Xylosidase Xyl43B in Escherichia coli*. Bioscience, Biotechnology, and Biochemistry. 68 (3), pp. 609–614.

162. Suryani, null, Kimura, T., Sakka, K., and Ohmiya, K. (2004) *Sequencing and expression of the gene encoding the Clostridium stercorarium beta-xylosidase Xyl43B in Escherichia coli*. Bioscience, Biotechnology, and Biochemistry. 68 (3), pp. 609–614.

163. Suzuki, T., Kitagawa, E., Sakakibara, F., Ibata, K., Usui, K., and Kawai, K. (2001) *Cloning, expression, and characterization of a family 52 beta-xylosidase gene (xysB) of a multiple-xylanase-producing bacterium, Aeromonas caviae ME-1*. Bioscience, Biotechnology, and Biochemistry. 65 (3), pp. 487–494.

164. Tarayre, C., Bauwens, J., Brasseur, C., Mattéotti, C., Millet, C., Guiot, P.A., et al. (2015) *Isolation and cultivation of xylanolytic and cellulolytic Sarocladium kiliense and Trichoderma virens from the gut of the termite Reticulitermes santonensis*. Environmental Science and Pollution Research International. 22 (6), pp. 4369–4382.

165. Tartar, A., Wheeler, M.M., Zhou, X., Coy, M.R., Boucias, D.G., and Scharf, M.E. (2009) *Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite Reticulitermes flavipes*. Biotechnology for Biofuels. 2 pp. 25.

166. Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) *The COG database: a tool for genome-scale analysis of protein functions and evolution*. Nucleic Acids Research. 28 (1), pp. 33–36.

167. Techtmann, S.M. and Hazen, T.C. (2016) *Metagenomic applications in environmental monitoring and bioremediation*. Journal of Industrial Microbiology & Biotechnology. 43 (10), pp. 1345–1354.

168. Teng, C., Jia, H., Yan, Q., Zhou, P., and Jiang, Z. (2011) *High-level expression of extracellular secretion of a β-xylosidase gene from Paecilomyces thermophila in Escherichia coli*. Bioresource Technology. 102 (2), pp. 1822–1830.

169. Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., et al. (2014) *Molecular traces of alternative social organization in a termite genome*. Nature Communications. 5 pp. 3636.

170. Tipayarom, D., Thi, N., and Oanh, N.T. (2007) *Effects from Open Rice Straw Burning Emission on Air Quality in the Bangkok Metropolitan Region*. ScienceAsia. 33 pp. 339–345.

171. Todaka, N., Moriya, S., Saita, K., Hondo, T., Kiuchi, I., Takasu, H., et al. (2007) *Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of Reticulitermes speratus*. FEMS Microbiology Ecology. 59 (3), pp. 592–599.

172. Tokuda, G., Tsuboi, Y., Kihara, K., Saitou, S., Moriya, S., Lo, N., et al. (2014) *Metabolomic profiling of 13C-labelled cellulose digestion in a lower termite: insights into gut symbiont function*. Proc. R. Soc. B. 281 (1789), pp. 20140990.

173. Tolia, N.H. and Joshua-Tor, L. (2006) *Strategies for protein coexpression in Escherichia coli*. Nature Methods. 3 (1), pp. 55–64.

174. Trinh, V.-H., Tran, T.-H., and Nguyen, T.-H. (2010) Diversity of termite species in Vietnam. in: Singapore.

175. Tripuspaningsih, N.N., Suwanto, A., Suhartono, M.T., Achmadi, S.S., Yogiara, and Kimura, T. (2008) *Cloning, Sequencing, and Characterization of the Xylan Degrading Enzymes from Geobacillus thermoleovurans IT-08*. ResearchGate. 9 (2), pp. 177–187.

176. Uchiyama, T. and Miyazaki, K. (2009) *Functional metagenomics for enzyme discovery: challenges to efficient screening*. Current Opinion in Biotechnology. 20 (6), pp. 616–622.

177. Umemoto, Y., Onishi, R., and Araki, T. (2008) *Cloning of a Novel Gene Encoding β-1,3-Xylosidase from a Marine Bacterium, Vibrio sp. Strain XY-214, and Characterization of the Gene Product*. Applied and Environmental Microbiology. 74 (1), pp. 305–308.

178. Volynets, B., Ein-Mozaffari, F., and Dahman, Y. (2016) *Biomass processing into ethanol: pretreatment, enzymatic hydrolysis, fermentation, rheology, and mixing*. Green Processing and Synthesis. 6 (1), pp. 1–22.

179. Wagschal, K., Jordan, D.B., and Braker, J.D. (2012) *Catalytic properties of β-d-xylosidase XylBH43 from Bacillus halodurans C-125 and mutant XylBH43-W147G*. Process Biochemistry.

180. Wang, J., Qi, J., Zhao, H., He, S., Zhang, Y., Wei, S., et al. (2013) *Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease*. Scientific Reports. 3 pp. 1843.

181. Wang, X. and Seed, B. (2003) *Selection of oligonucleotide probes for protein coding sequences*. Bioinformatics (Oxford, England). 19 (7), pp. 796–802.

182. Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., et al. (2007) *Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite*. Nature. 450 (7169), pp. 560–565.

183. William Studier, F., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) *[6] Use of T7 RNA polymease to direct expression of cloned genes*. Methods in Enzymology. 185 pp. 60–89.

184. Wingfield, P.T. (2001) *Protein Precipitation Using Ammonium Sulfate*. Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.]. APPENDIX 3 pp. Appendix-3F.

185. Xie, L., Zhang, L., Zhong, Y., Liu, N., Long, Y., Wang, S., et al. (2012) *Profiling the metatranscriptome of the protistan community in Coptotermes formosanus with emphasis on the lignocellulolytic system*. Genomics. 99 (4), pp. 246–255.

186. Xu, W.Z., Shima, Y., Negoro, S., and Urabe, I. (1991) *Sequence and properties of beta-xylosidase from Bacillus pumilus IPO. Contradiction of the previous nucleotide sequence*. European Journal of Biochemistry / FEBS. 202 (3), pp. 1197–1203.

187. Xu, Y.-Q., Duan, C.-J., Zhou, Q.-N., Tang, J.-L., and Feng, J.-X. (2006) *[Cloning and identification of cellulase genes from uncultured microorganisms in pulp sediments from paper mill effluent]*. Wei Sheng Wu Xue Bao = Acta Microbiologica Sinica. 46 (5), pp. 783–788.

188. Xu, Z., Zhong, Z., Huang, L., Peng, L., Wang, F., and Cen, P. (2006) *High-level production of bioactive human beta-defensin-4 in Escherichia coli by soluble fusion expression*. Applied Microbiology and Biotechnology. 72 (3), pp. 471–479.

189. Y, U., R, O., and T, A. (2008) *Cloning of a novel gene encoding beta-1,3-xylosidase from a marine bacterium, Vibrio sp. strain XY-214, and characterization of the gene product., Cloning of a Novel Gene Encoding β-1,3-Xylosidase from a Marine Bacterium, Vibrio sp. Strain XY-214, and Characterization of the Gene Product*. Applied and Environmental Microbiology, Applied and Environmental Microbiology. 74, 74 (1, 1), pp. 305, 305–308.

190. Yan Yang, S., Zheng, Y., Huang, Z., Min Wang, X., and Yang, H. (2016) *Lactococcus nasutitermitis sp. nov. isolated from a termite gut*. International Journal of Systematic and Evolutionary Microbiology. 66 (1), pp. 518–522.

191. Yoon, K.H., Yun, H.N., and Jung, K.H. (1998) *Molecular cloning of a Bacillus sp. KK-1 xylanase gene and characterization of the gene product*. Biochemistry and Molecular Biology International. 45 (2), pp. 337–347.

192. Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S.G., and Alvarez-Cohen, L. (2015) *High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats*. mBio. 6 (1), pp. e02288-14.

193. Nguyễn Thị Thảo (2015). "Nghiên cứu *gen* mã hóa enzyme tham gia thủy *phân* cellulose *từ* khu hệ vi khuẩn ruột mối bằng kỹ thuật *Metagenomics"*. Luận án tiến sỹ Sinh học.

194. Trần Đình Mấn (2013). "Nghiên cứu sàng lọc các enzyme bền nhiệt từ các khu hệ vi sinh vật của các nguồn nước nóng bằng kỹ thuật Metagenomics". Đề tài VAST năm 2013, mã số VAST03.01/12-13.

195. Đặng Duy Đức (2015). "Cloning gen mã hóa pectinase từ cộng đồng vi sinh vật không thông qua nuôi cấy". Khóa luận tốt nghiệp.

196. http://www.EXPASY.org

197. http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index.

198. http://swissmodel.expasy.org/.

199. http://lin.uestc.edu.cn/server/AcalPred

200. http://www.tbi.org.tw/tools

201. http://www.phylogeny.fr

202. http://www.CAZY.org

203. http://primerdigital.com/fastpcr

204. http://www.restrictionmapper.org

205. http://blast.ncbi.nlm.nih.gov

206. http://www.bio-rad.com/en-ch/product/quantity-one-1-d-analysis-software

207. www.ncbi.nlm.nih.gov

PHỤ LỤC

Phụ lục 1: Trình tự nucleotide của mã gen Termite-2_GL0112518 (Xbx14)

ATGGATAAAGTTACCAATCCGGTGCTTACCGGTTTTCACGCCGACCCCTCGATTGTGAG
AGTCGGGGATTATTTTTACATCGCTAATTCCACTTTTGAATGGTATCCAGGCGTGGAAC
TGCACCGTTCAAAAAACTTGGCGAATTGGGAATCGCTGCCTTCGCCGCTGGGCGAACG
GCGGCTTCTGGACATGGAAGGCGCGCGCGCGTCCTGCGGCATCTGGGCGCCCTGCTTG
AGCTACGCCGACGGGCTTTTCTGGCTCATCTATACCAACGTGCGCACCTGGAACGCGG
GGCCGTGGAAGGACTGCCCCAACTACCTGACAACCGCCAAGTCAATCGAAGGCCCGT
GGTCAGACCCCGTGTTCCTCAACTGCTCAGGCTTTGACCCCTCGCTTTTTCATGACGAT
GACGGCAGAAAGTGGCTGGTCAACATGGAGTGGGACTACCGCAAGCCGGGCGACCCC
GAAGGCCCGCAGTTTTCGGGCATACTGATTCAGGAATACAGCCCCGCGGAAAAAAGG
CTCGCGGGGCCGGTTCGCAAGATTTTCACGGGTTCCCCGATAGCCTGCGTGGAAGGCC
CGCACGTCTACAAGCGGGACGGCTGGTACTACCTGCTCACCGCCGAGGGCGGCACGGT
GTATAACCACGCGGCGACCCTTGCCCGCTCCCGCGCGTTGGAAGGGCCTTACGAGATT
CACCCGCAGAACCCGCTTATCAGTTCGCGGGGGAAGCCGGAACTGCGCCTGCAAAAA
GCGGGGCACGCGAGCTGGTGCGAGACCGCCGACGGCAGAACCTACCTGGCCTTCTTGT
GCGGGAGGCCGCTGCCCGGCACGCAAAACTGCCCGCTGGGGCGGGAGACTTCGATAG
CCGAGCTGGTCTGGCACGAGGGGTGGCCGTATGTCAAAGGCGAAGACGGAAACAGGC
AGAATTTCCCCGCGGACACTTTCGAGCCTCCCGTGAAAATTGCCGCCCCGGCGCGAAA
GAGGCGCGGGCGCTCTATCAGTTTGACGGCCCCGCCATCCACGGCGACTTCAAGACTC
TCCGCGTTCCCGCCGACCCTGAACGCTGCTCGTTGA.

Phụ lục 2: Kết quả giải trình tự gen Xbx14

```
Trình tự gốc ---------------------------ACCATGGATGATAAAGTTACCAATCCGGTGGGTTAGGC
Mồi xuôi     ---------------------------ACCATGGATGATAAAGTTACCAATCCGGTGGGTTAGGC
Mồi ngược    ------------------------------------------------------------------
                                                 ***************************
Nu bảo tồn   ----------------------------ACCATGGATGATAAAGTTACCAATCCGGTGGGTTAGGC
Trình tự gốc ATGGATAAAGTTACCAATCCGGTGCTTACCGGTTTTCACGCCGACCCCTCGATTGTGAGA
Mồi xuôi     ATGGATAAAGTTACCAATCCGGTGCTTACCGGTTTTCACGCCGACCCCTCGATTGTGAGA
Mồi ngược    ------------------------------------------------------------------
             ******************************************************************
Nu bảo tồn   ATGGATAAAGTTACCAATCCGGTGCTTACCGGTTTTCACGCCGACCCCTCGATTGTGAGA
Trình tự gốc GTCGGGGATTATTTTTACATCGCTAATTCCACTTTTGAATGGTATCCAGGCGTGGAACTG
Mồi xuôi     GTCGGGGATTATTTTTACATCGCTAATTCCACTTTTGAATGGTATCCAGGCGTGGAACTG
Mồi ngược    ------------------------------------------------------------------
             ******************************************************************
Nu bảo tồn   GTCGGGGATTATTTTTACATCGCTAATTCCACTTTTGAATGGTATCCAGGCGTGGAACTG
Trình tự gốc CACCGTTCAAAAAACTTGGCGAATTGGGAATCGCTGCCTTCGCCGCTGGGCGAACGGCGG
Mồi xuôi     CACCGTTCAAAAAACTTGGCGAATTGGGAATCGCTGCCTTCGCCGCTGGGCGAACGGCGG
Mồi ngược    CACCGTTCAAAAAACTTGGCGAATTGGGAATCGCTGCCTTCGCCGCTGGGCGAACGGCGG
```

```
             ************************************************************
Nu bảo tồn   CACCGTTCAAAAAACTTGGCGAATTGGGAATCGCTGCCTTCGCCGCTGGGCGAACGGCGG
Trình tự gốc CTTCTGGACATGGAAGGCGCGCGCGCGTCCTGCGGCATCTGGGCGCCCTGCTTGAGCTAC
Mồi xuôi     CTTCTGGACATGGAAGGCGCGCGCGCGTCCTGCGGCATCTGGGCGCCCTGCTTGAGCTAC
Mồi ngược    CTTCTGGACATGGAAGGCGCGCGCGCGTCCTGCGGCATCTGGGCGCCCTGCTTGAGCTAC
             ************************************************************
Nu bảo tồn   CTTCTGGACATGGAAGGCGCGCGCGCGTCCTGCGGCATCTGGGCGCCCTGCTTGAGCTAC
Trình tự gốc GCCGACGGGCTTTTCTGGCTCATCTATACCAACGTGCGCACCTGGAACGCGGGGCCGTGG
Mồi xuôi     GCCGACGGGCTTTTCTGGCTCATCTATACCAACGTGCGCACCTGGAACGCGGGGCCGTGG
Mồi ngược    GCCGACGGGCTTTTCTGGCTCATCTATACCAACGTGCGCACCTGGAACGCGGGGCCGTGG
             ************************************************************
Nu bảo tồn   GCCGACGGGCTTTTCTGGCTCATCTATACCAACGTGCGCACCTGGAACGCGGGGCCGTGG
Trình tự gốc AAGGACTGCCCCAACTACCTGACAACCGCCAAGTCAATCGAAGGCCCGTGGTCAGACCCC
Mồi xuôi     AAGGACTGCCCCAACTACCTGACAACCGCCAAGTCAATCGAAGGCCCGTGGTCAGACCCC
Mồi ngược    AAGGACTGCCCCAACTACCTGACAACCGCCAAGTCAATCGAAGGCCCGTGGTCAGACCCC
             ************************************************************
Nu bảo tồn   AAGGACTGCCCCAACTACCTGACAACCGCCAAGTCAATCGAAGGCCCGTGGTCAGACCCC
Trình tự gốc GTGTTCCTCAACTGCTCAGGCTTTGACCCCTCGCTTTTTCATGACGATGACGGCAGAAAG
Mồi xuôi     GTGTTCCTCAACTGCTCAGGCTTTGACCCCTCGCTTTTTCATGACGATGACGGCAGAAAG
Mồi ngược    GTGTTCCTCAACTGCTCAGGCTTTGACCCCTCGCTTTTTCATGACGATGACGGCAGAAAG
             ************************************************************
Nu bảo tồn   GTGTTCCTCAACTGCTCAGGCTTTGACCCCTCGCTTTTTCATGACGATGACGGCAGAAAG
Trình tự gốc TGGCTGGTCAACATGGAGTGGGACTACCGCAAGCCGGGCGACCCCGAAGGCCCGCAGTTT
Mồi xuôi     TGGCTGGTCAACATGGAGTGGGACTACCGCAAGCCGGGCGACCCCGAAGGCCCGCAGTTT
Mồi ngược    TGGCTGGTCAACATGGAGTGGGACTACCGCAAGCCGGGCGACCCCGAAGGCCCGCAGTTT
             ************************************************************
Nu bảo tồn   TGGCTGGTCAACATGGAGTGGGACTACCGCAAGCCGGGCGACCCCGAAGGCCCGCAGTTT
Trình tự gốc TCGGGCATACTGATTCAGGAATACAGCCCCGCGGAAAAAAGGCTCGCGGGGCCGGTTCGC
Mồi xuôi     TCGGGCATACTGATTCAGGAATACAGCCCCGCGGAAAAAAGGCTCGCGGGGCCGGTTCGC
Mồi ngược    TCGGGCATACTGATTCAGGAATACAGCCCCGCGGAAAAAAGGCTCGCGGGGCCGGTTCGC
             ************************************************************
Nu bảo tồn   TCGGGCATACTGATTCAGGAATACAGCCCCGCGGAAAAAAGGCTCGCGGGGCCGGTTCGC
Trình tự gốc AAGATTTTCACGGGTTCCCCGATAGCCTGCGTGGAAGGCCCGCACGTCTACAAGCGGGAC
Mồi xuôi     AAGATTTTCACGGGTTCCCCGATAGCCTGCGTGGAAGGCCCGCACGTCTACAAGCGGGAC
Mồi ngược    AAGATTTTCACGGGTTCCCCGATAGCCTGCGTGGAAGGCCCGCACGTCTACAAGCGGGAC
             ************************************************************
Nu bảo tồn   AAGATTTTCACGGGTTCCCCGATAGCCTGCGTGGAAGGCCCGCACGTCTACAAGCGGGAC
Trình tự gốc GGCTGGTACTACCTGCTCACCGCCGAGGGCGGCACGGTGTATAACCACGCGGCGACCCTT
Mồi xuôi     GGCTGGTACTACCTGCTCACCGCCGAGGGCGGCACGGTGTATAACCACGCGGCGACCCTT
Mồi ngược    GGCTGGTACTACCTGCTCACCGCCGAGGGCGGCACGGTGTATAACCACGCGGCGACCCTT
             ************************************************************
Nu bảo tồn   GGCTGGTACTACCTGCTCACCGCCGAGGGCGGCACGGTGTATAACCACGCGGCGACCCTT
Trình tự gốc GCCCGCTCCCGCGCGTTGGAAGGGCCTTACGAGATTCACCCGCAGAACCCGCTTATCAGT
Mồi xuôi     GCCCGCTCCCGCGCGTTGGAAGGGCCTTACGAGATTCACCCGCAGAACCCGCTTATCAGT
Mồi ngược    GCCCGCTCCCGCGCGTTGGAAGGGCCTTACGAGATTCACCCGCAGAACCCGCTTATCAGT
```

```
             ************************************************************
Nu bảo tồn   GCCCGCTCCCGCGCGTTGGAAGGGCCTTACGAGATTCACCCGCAGAACCCGCTTATCAGT
Trình tự gốc TCGCGGGGGAAGCCGGAACTGCGCCTGCAAAAAGCGGGGCACGCGAGCTGGTGCGAGACC
Mồi xuôi     TCGCGGGGGAAGCCGGAACTGCGCCTGCAAAAAGCGGGGCACGCGAGCTGGTGCGAGACC
Mồi ngược    TCGCGGGGGAAGCCGGAACTGCGCCTGCAAAAAGCGGGGCACGCGAGCTGGTGCGAGACC
             ************************************************************
Nu bảo tồn   TCGCGGGGGAAGCCGGAACTGCGCCTGCAAAAAGCGGGGCACGCGAGCTGGTGCGAGACC
Trình tự gốc GCCGACGGCAGAACCTACCTGGCCTTCTTGTGCGGGAGGCCGCTGCCCGGCACGCAAAAC
Mồi xuôi     GCCGACGGCAGAACCTACCTGGCCTTCTTGTGCGGGAGGCCGCTGCCCGGCACGCAAAAC
Mồi ngược    GCCGACGGCAGAACCTACCTGGCCTTCTTGTGCGGGAGGCCGCTGCCCGGCACGCAAAAC
             ************************************************************
Nu bảo tồn   GCCGACGGCAGAACCTACCTGGCCTTCTTGTGCGGGAGGCCGCTGCCCGGCACGCAAAAC
Trình tự gốc TGCCCGCTGGGGCGGGAGACTTCGATAGCCGAGCTGGTCTGGCACGAGGGGTGGCCGTAT
Mồi xuôi     TGCCCGCTGGGGCGGGAGACTTCGATAGCCGAGCTGGTCTGGCACGAGGGGTGGCCGTAT
Mồi ngược    TGCCCGCTGGGGCGGGAGACTTCGATAGCCGAGCTGGTCTGGCACGAGGGGTGGCCGTAT
             ************************************************************
Nu bảo tồn   TGCCCGCTGGGGCGGGAGACTTCGATAGCCGAGCTGGTCTGGCACGAGGGGTGGCCGTAT
Trình tự gốc GTCAAAGGCGAAGACGGAAACAGGCAGAATTTCCCCGCGGACACTTTCGAGCCTCCCGTG
Mồi xuôi     GTCAAAGGCGAAGACGGAAACAGGCAGAATTTCCCCGCGGACACTTTCGAGCCTCCCGTG
Mồi ngược    GTCAAAGGCGAAGACGGAAACAGGCAGAATTTCCCCGCGGACACTTTCGAGCCTCCCGTG
             ************************************************************
Nu bảo tồn   GTCAAAGGCGAAGACGGAAACAGGCAGAATTTCCCCGCGGACACTTTCGAGCCTCCCGTG
Trình tự gốc AAAATTGCCGCCCCGGCGCGAAAGAGGCGCGGGCGCTCTATCAGTTTGACGGCCCCGCCA
Mồi xuôi     AAAATTGCCGCCCCGGCGCGAAAGAGGCGCGGGCGCTCTATCAGTTTGACGGCCCCGCCA
Mồi ngược    AAAATTGCCGCCCCGGCGCGAAAGAGGCGCGGGCGCTCTATCAGTTTGACGGCCCCGCCA
             ************************************************************
Nu bảo tồn   AAAATTGCCGCCCCGGCGCGAAAGAGGCGCGGGCGCTCTATCAGTTTGACGGCCCCGCCA
Trình tự gốc TCCACGGCGACTTCAAGACTCTCCGCGTTCCCGCCGACCCTGAACGCTGCTCGTTGACGG
Mồi xuôi     ------------------------------------------------------------
Mồi ngược    TCCACGGCGACTTCAAGACTCTCCGCGTTCCCGCCGACCCTGAACGCTGCTCGTTGACGG
              ***********************************************************
Nu bảo tồn   TCCACGGCGACTTCAAGACTCTCCGCGTTCCCGCCGACCCTGAACGCTGCTCGTTGACGG
Trình tự gốc ATGACTTGCGACGAGCAGAGCTCA---------------------------------------
Mồi xuôi     ------------------------------------------------------------
Mồi ngược    ATGACTTGCGACGAGCAGAGCTCA ----------------------------------
             ************************

Nu bảo tồn   ATGACTTGCGACGAGCAGAGCTCA----------------------------------
```

Phụ lục 3: Công thức pha muối và lượng $(NH4)_2SO_4$ bão hòa bổ sung để tủa Xbx14 ở các nồng độ khác nhau

1. Công thức pha muối $(NH4)_2SO_4$ bão hòa ở các nhiệt độ khác nhau

Nhiệt độ (°C)	**0**	**10**	**20**	**30**	**40**	**50**	**60**
Khối lượng $(NH4)_2SO_4$ (g)	70,6	73	75,4	78,1	81,2	84,3	87,4

2. Lượng $(NH4)_2SO_4$ bão hòa bổ sung để tủa *Xbx14* ở các nồng độ khác nhau

Dịch ban đầu (ml)	**Bổ sung amoni sulfat bão hòa (ml)**	**Nồng độ bão hòa (%)**	**Nồng độ (%)**
10	0	0	0
10,31	0,31	3,00679	2,345296
10,64	0,33	6,015038	4,691729
10,99	0,35	9,008189	7,026388
11,37	0,38	12,04925	9,398417
11,81	0,44	15,32599	11,95428
12,24	0,43	18,30065	14,27451
12,67	0,43	21,0734	16,43725
13,1	0,43	23,66412	18,45802
13,6	0,5	26,47059	20,64706

Dịch ban đầu (ml)	**Bổ sung amoni sulfat bão hòa (ml)**	**Nồng độ bão hòa (%)**	**Nồng độ (%)**
10	0	0	0
10,02	0,02	0,199601	0,155689
10,041	0,021	0,408326	0,318494
10,061	0,02	0,606302	0,472915
10,081	0,02	0,803492	0,626724
10,102	0,021	1,009701	0,787567
10,122	0,02	1,205295	0,94013
10,142	0,02	1,400118	1,092092
10,163	0,021	1,603857	1,251009
10,184	0,021	1,806756	1,409269
10,205	0,021	2,008819	1,566879

Dịch ban đầu (ml)	Bổ sung amoni sulfat bão hòa (ml)	Nồng độ bão hòa (%)	Nồng độ (%)
10	0	0	0
10,0021	0,0021	0,020996	0,016377
10,0041	0,002	0,040983	0,031967
10,0061	0,002	0,060963	0,047551
10,0081	0,002	0,080934	0,063129
10,0101	0,002	0,100898	0,078701
10,0121	0,002	0,120854	0,094266
10,0141	0,002	0,140801	0,109825
10,0161	0,002	0,160741	0,125378
10,0181	0,002	0,180673	0,140925
10,0201	0,002	0,200597	0,156466
10,0221	0,002	0,220513	0,172