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

TÓM TẮT LUẬN ÁN TIẾN SĨ SINH HỌC

Hà Nội, tháng 01 năm 2018

Công trình được hoàn thành tại: Tổ Di truyền học, khoa Sinh học, Trường ĐHSP Hà Nội và Phòng Kỹ thuật Di truyền, Viện Công nghệ Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học: **GS.TS. Trương Nam Hải**

PGS.TS. Đặng Hữu Lanh

Phản biện 1: GS.TS. Phan Hữu Tồn

Phản biện 2: GS.TS. Chu Hoàng Mậu

Phản biện 3: PGS.TS. Lê Huy Hàm

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án cấp Trường họp tại Trường Đại học Sư phạm Hà Nội vào hồi giờ 00 ngày tháng năm 2018

Có thể tìm hiểu luận án tại thư viện: Thư viện Quốc Gia, Hà Nội hoặc Thư viện Trường Đại học Sư phạm Hà Nội

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: Trình bày vai trò và thành phần cấu tạo của lignocellulose

1.1.2. Sự chuyển hóa lignocellulose: Mô tả hệ thống enzyme tham gia chuyển hóa lignocellulose

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

1.2.1. Metagenomics: Trình bày khái niệm chung về Metagenomics, cách tiếp cận trong nghiên cứu và ứng dụng.

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

Mô tả nguồn gốc, cơ sở xây dựng dữ liệu, cách sử dụng, độ tin cậy của một số công cụ tin sinh học sử dụng trong phân tích số liệu của luận án bao gồm: BLAST, EXPASY, Phyre 2, Swiss model, TBI, Alcapred và Phylogeny,…

1.2.3. Các nguồn dữ liệu: Mô tả chi tiết bao gồm nguồn gốc, lượng dữ liệu, cập nhật dữ liệu và độ tin cậy của hai nguồn dữ liệu sử dụng trong nghiên cứu gồm CAZY và NCBI.

1.2.4. Mẫu dò DNA và ứng dụng

Sơ lược chung về mẫu dò DNA và các ứng dụng, đặc biệt là các nghiên cứu về mẫu dò đã được sử dụng trong phân tích DNA đa hệ gen.

1.3. ENZYME β–xylosidase

Nêu tổng quan về enzyme β–xylosidase bao gồm tên gọi, mô hình hoạt động, cấu trúc không gian, hoạt tính, ứng dụng và nguồn cung cấp.

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

Tổng quan chung về các khu hệ vi sinh vật chuyển hóa lignocellulose trong tự nhiên.

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

Mô tả về đặc điểm sinh học của mối và sự đa dạng của vi sinh vật cộng sinh trong ruột mối – nguồn khai thác enzyme xử lý phân hủy lignocellulose.

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

Tóm tắt các kết quả đã nghiên cứu về mối, hệ vi sinh vật, enzyme thủy phân lignocellulose trong ruột của *C. gestroi* ở Việt Nam, cách chọn gen để biểu hiện từ dữ liệu DNA đa hệ gen và các vấn đề cần 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

Sử dụng dữ liệu trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* đã được chú thích bởi công ty giải trình tự BGI, bao gồm 125.432 ORF. Trong đó có 587 ORF mã hóa enzyme thủy phân lignocellulose làm nguồn dữ liệu để chọn trình tự gen mã số Termite 2_ GL0112518 mã hóa β–xylosidase.

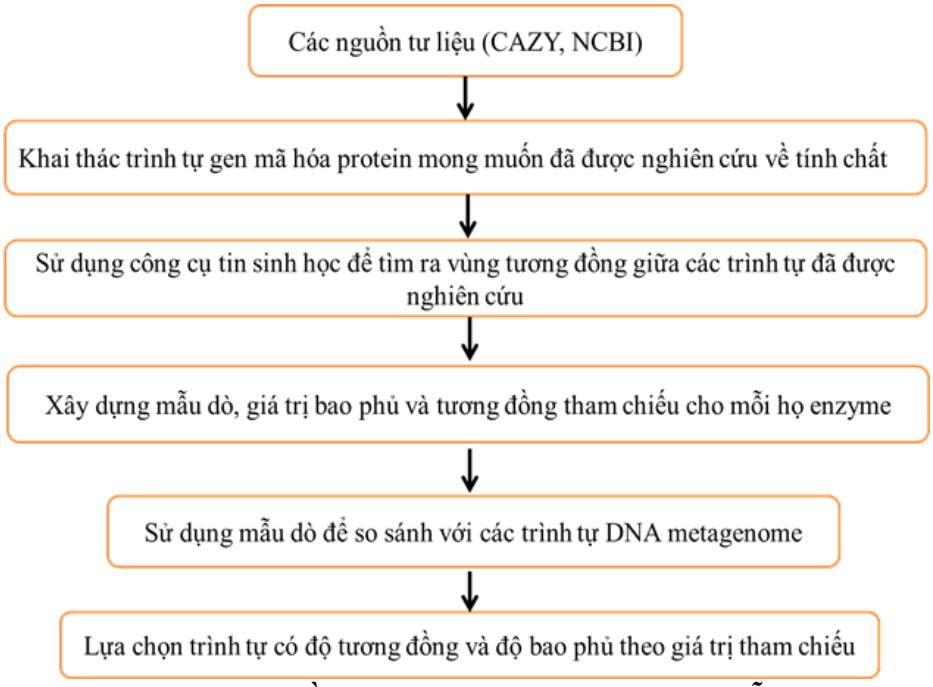
2.1.2. Hóa chất và thiết bị máy móc

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). Nghiên cứu sử dụng của máy móc của phòng Công nghệ trọng điểm, Viện Công nghệ Sinh học, Viện Hàn Lâm Khoa học và Công nghệ Việt Nam.

2.2. PHƯƠNG PHÁP NGHIÊN CỨU

2.2.1. Phương pháp xây dựng mẫu dò

Để xây dựng mẫu dò cho enzyme 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

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

Phân tích đa dạng các loài vi sinh vật điển hình trong ruột mối *C. gestroi*; Phân tích đa dạng ORF của enzyme thủy phân lignocellulose theo phân loại của CAZY; So sánh trình tự ORF của dữ liệu DNA đa hệ gen với cơ sở dữ liệu của NCBI bằng BLAST; Thiết kế mồi bằng phần mềm FastPCR; Kiểm tra vị trí của enzyme hạn chế trên gen quan tâm bằng phần mềm trực tuyến RestrictionMapper; Chuyển mã trình tự DNA sang trình tự axit amin bằng chương trình dịch mã EXPASY; Dự đoán cấu trúc của *Xbx14* bằng phần mềm Phyre2 và Swiss model; Dự đoán nguồn gốc của gen bằng công cụ BLAST-Explorer; Xác định độ sạch của protein *Xbx14* sau khi tinh chế bằng phần mềm Quantity One.

2.2.3. Các phương pháp vi sinh: Giữ chủng *E. coli;* cấy ria tạo khuẩn lạc đơn; nuôi cấy vi khuẩn.

2.2.4. Các phương pháp sinh học phân tử: Giải trình tự DNA bằng máy giải trình tự thế hệ mới HiSeq2000 của Illumina, biến nạp DNA plasmid vào vi khuẩn *E. coli* bằng sốc nhiệt**,** tách chiết DNA plasmid từ tế bào vi khuẩn *E.coli,* cắt kiểm tra gen, điện di DNA trên gel agarose, đồng biểu hiện gen.

2.2.5. Các phương pháp hóa sinh protein: Phương pháp điện di biến tính protein trên gel polyacrylamide – SDS, tinh chế protein bằng tủa phân đoạn, định lượng axit nucleic/protein bằng phương pháp đo mật độ quang, phương pháp xác định hoạt tính β–xylosidase, xác định ảnh hưởng của nhiệt độ và pH, xác định độ bền của enzyme, phương pháp xác định thông số động học của enzyme.

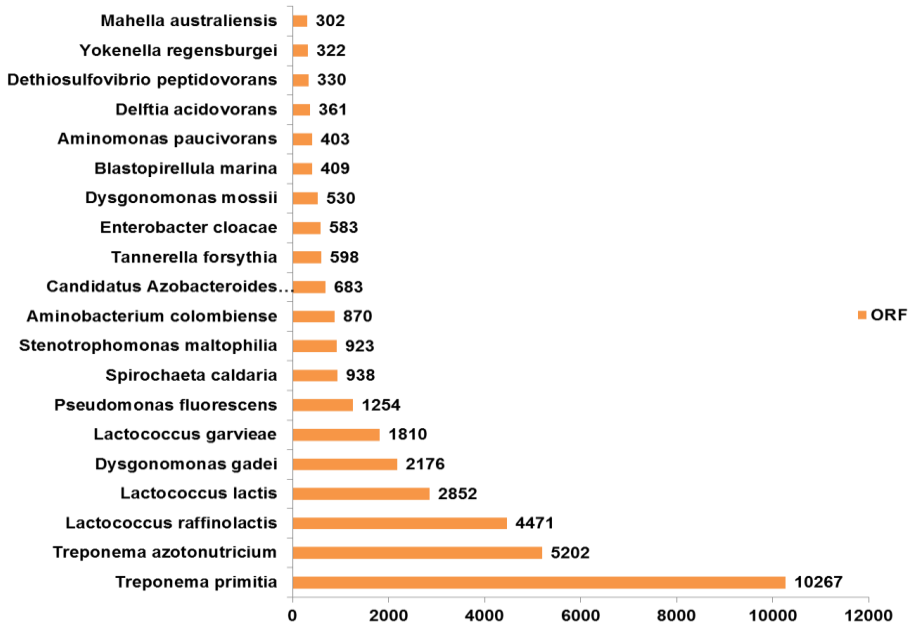
Chương 3. KẾT QUẢ VÀ THẢO LUẬN

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

Nghiên cứu trước đây đã sử dụng kỹ thuật Metagenomics giải trình tự đa hệ gen của vi sinh vật sống trong ruột mối *C. gestroi* ở Việt Nam. Trên cơ sở dữ liệu này chúng tôi sẽ tập trung vào phân tích vi sinh vật chiếm số lượng lớn nhất theo thống kê ban đầu đó 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. Dựa trên số liệu dự đoán về loài, chúng tôi tiến hành lọc và thống kê được 20 loài vi khuẩn có số lượng lớn hơn 300 ORF (Hình 3.1).



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

Tham gia vào quá trình chuyển hóa Nitơ gồm có sáu loài là *Treponema primitia, Treponema azotaurricium, Candidatus Azobacteroides pseudotrichonymphae, Aminobacterium colombiense, Aminomonas paucivorans, Stenotrophomonas maltophilia.* Trong đó ba loài *Aminobacterium colombiense, Aminomonas paucivorans và Stenotrophomonas maltophilia* chỉ có ở *C. gestroi*.

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

Vi khuẩn sống trong ruột *C. gestroi* có 16 loài với số lượng ORF lớn tham gia vào quá trình chuyển hóa lignocellulose, 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* (Hình 3.1). Trong số 16 loài điển hình tham gia chuyển hóa lignocellulose trên, thì 5 loài chỉ xuất hiện trong ruột của mối *C. gestroi* bao gồm*: Mahella australiensis, Blastopirellula marina, Yokenella regensburgei (Koserella trabulsii), Delftia acidovorans, Dethiosulfovibrio peptidovorans.*

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 góp phần cung cấp các 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 enzyme này, chúng tôi tiến hành sắp xếp chúng vào hệ thống GH trong bảng 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. So sánh các họ enzyme thủy phân lignocellulose (GH) của vi sinh vật cộng sinh trong ruột mối C. gestroi với một số loài mối khác

GH	*C. gestroi* (sinh vật cộng sinh)[1]	*Nasutitermes* (sinh vật cộng sinh)[2]	*R. speratus* (sinh vật cộng sinh)[3]	*R. flavipes* (sinh vật cộng sinh)[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	–	–	–
GH10	–	3	–	–	–
GH10	–	2	–	–	–
GH10	–	3	–	–	–
GH11	3	–	–	–	–
GH11	3	–	–	–	–
GH11	1	–	–	–	–
GH11	40	–	–	–	–
GH11	2	–	–	–	–
GH12	3	–	–	–	–
GH12	3	–	–	–	–
GH13	10	–	–	–	–

[1]Do et al., 2014 (số liệu giải trình tự DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi* bằng phương pháp Metagenomics

[2]Warnecke *et al*. (số liệu giải trình tự DNA đa hệ gen của vi sinh vật cộng sinh trong ruột mối *Nasutitermes* bằng phương pháp Metagenomics bằng việc lập ngân hàng DNA đa hệ gen)

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

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

[5]Tartar *et al*. (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 phân loại theo dữ liệu của CAZY tại trang web: http://www.CAZY.org/
Dấu -: không có; dấu • không xác định số lượng gen

Từ kết quả so sánh cho thấy sự đa dạng hơn về cả số họ GH và số lượng ORF thuộc về từng họ GH trong ruột của *C. gestroi*. Chúng tôi đặc biệt quan tâm đến các họ GH chỉ xuất hiện trong ruột mối *C. gestroi* để xác định khả năng phân giải các thành phần của lignocellulose, từ đó đưa ra hướng khai thác và chọn được gen mới từ dữ liệu DNA đa hệ gen đã có sẵn. Trong các họ GH theo phân loại của CAZY, họ GH117 và GH130 tham gia thủy phân 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 các dimer có nguồn gốc từ hemicellulose rất hiệu quả. Sự phong phú của các hemicellulase từ dữ liệu DNA đa hệ gen của mối *C. gestroi* là cơ sở thuận lợi để lựa chọn gen mã hóa enzyme phân hủy hemicellulose trong nghiên cứu thực nghiệm.

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

Trong nghiên cứu này chúng tôi quan tâm đến nhóm hemicellulase và một trong những enzyme có ý nghĩa lớn trong quá trình thủy phân hemicellulose là β–xylosidase. Theo ước đoán của công ty giải trình tự BGI, thì trong ruột của mối *C. gestroi*, có 48 ORF mã hóa β–xylosidase, trong đó 23 ORF được phân loại vào 13 loài, số còn lại chưa được loại đến loài. Các ORF được xếp vào 8 GH khác nhau và chiếm ưu thế là các GH43, 51, 116. Trong số các họ GH chứa β–xylosidase thì họ GH43 xuất hiện nhiều nhất ở 7 loài vi khuẩn khác nhau. Đặc biệt loài *Treponema primitia* chứa ORF của cả 8 GH (Bảng 3.2).

Bảng 3.2. Đa dạng ORF và GH của β–xylosidase trong ruột mối C. gestroi

Loài	Số ORF	GH	Loài	Số ORF	GH
Bacteroides eggerthii	1	43	*Lactococcus raffinolactis*	2	43, 116
Clostridium hathewayi	1	51	*Leeuwenhoekiella blandensis*	1	1
Coprococcus eutactus	1	43	*Marvinbryantia formatexigens*	2	1, 43, 116
Paenibacillus mucilaginosus	1	1	*Treponema primitia*	5	1, 3, 5, 10, 39, 43, 51, 116

Enterobacter mori	1	43,	*Treponema azotonutricium*	1	51
Mahella australiensis	1	3	*Dysgonomonas gadei*	3	43, 51, 116
Mitsuokella multacida	3	1, 51			

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*

Chúng tôi tiến hành xây dựng phương pháp mới để chọn gen từ dữ liệu giải trình tự DNA đa hệ gen đã có bằng mẫu dò. Phương pháp xây dựng mẫu dò được dựa trên lý thuyết về trình tự axit amin của một loại enzyme đều có vùng trình tự bảo thủ và gốc hoạt tính bảo tồn giống nhau. Khi đã có mẫu dò có thể tìm nhanh trình tự gen đích từ dữ liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi.* Gen chọn sẽ 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 dữ liệu trình tự 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

Nghiên cứu sử dụng bảng phân loại của CAZY để xác định β–xylosidase thuộc về từng họ GH, với đặc điểm chi tiết về cấu trúc không gian, thành phần axit amin cho và nhận proton trong quá trình hoạt động của enzyme. Dựa trên sự bảo tồn cao về sự cuộn gấp trong cấu trúc không gian của protein, sẽ được xếp vào các 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.

Bảng 3.3. 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 thực nghiệm

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

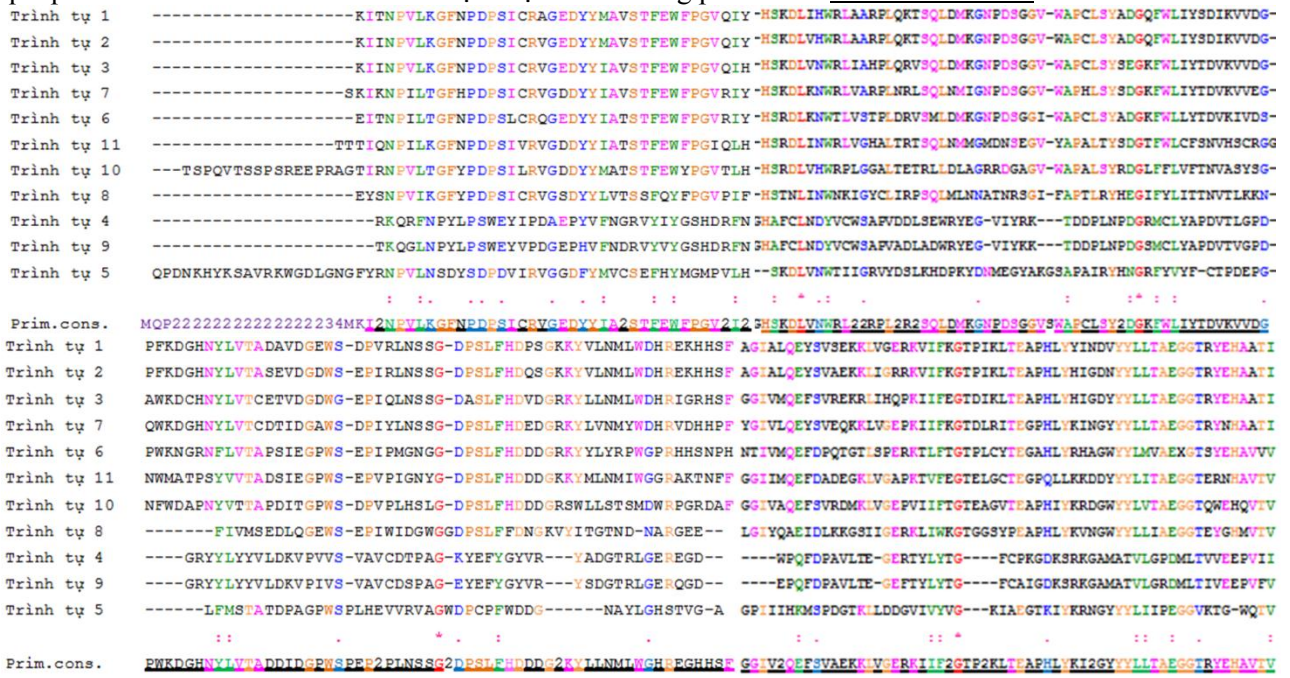
Số thứ tự	**Mã số trong GENBANK**	**Vi khuẩn**	**Số axit amin**	**pH hoạt động tối ưu**	**Nhiệt độ hoạt động tối ưu (^oC)**
		GH1			
1	CAD20872.1	*bacterium enrichment culture clone P11-6*	464	6	40
		GH3			
1	CAD48309.1	*C. stercorarium*	715		50
		GH30			
1	ABX45137.1.1	*Bifidobacterium breve*	448	6	45
		GH43			
1	CAA29235.1.1	*Bacillus pumilus*	535	7	40
2	AAC97375.1	*Bacillus pumilus PLS*	535	6	45
3	AAC27699.1	*bacterium Bacillus sp. KK-1*	533		55
4	BAA02527.1	*Clostridium stercorarium*	473	7	65
5	BAC879411	*Clostridium stercorarium*	497	3,5	80
6	AFZ7887.1	*Enterobacter sp. enrichment culture clone*	536	6	40

		nf1B6			
7	AAT9862.1	*Geobacillus stearothermophilus T-6 (XynB3)*	535	6,5	60
8	ABC750041	*Geobacillus thermoleovorans IT-08*	511	5	70
9	ADV16404.1	*Paenibacillus woosongensis*	477	6-7	30-45
10	AEF2882.1	*Thermobifida fusca TM51*	550	4,5	50
11	BAF982351	*Vibrio sp. XY-214*	535	7	36
GH52					
1	BAA74507.1	*Aeromonas caviae*	729	8,7	60
2	AGE344791	*Geobacillus stearothermophilus*	705	5,5	70
3	ABI49956.1	*Geobacillus stearothermophilus*	705	6,3	65
GH120					
1	ABM68042.1	*Thermoanaerobacterium saccharolyticum JW/SL-YS48*	636	6	65

Chúng tôi tiến hành tìm kiếm các trình tự axit amin hoặc gen mã hóa β–xylosidase đáp ứng được ba tiêu chí: (1) Các 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 các trình tự gen/axit amin từ vi khuẩn, để đảm bảo độ tương đồng và trùng khớp các 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 các trình tự đích cần tìm. Nguồn để tìm kiếm các trình tự 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.

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

ClustalW – PBIL là công cụ cho phép so sánh nhiều trình tự axit amin để xác định mức độ bảo tồn của các vị trí axit amin và sự tương đồng giữa các trình tự. Kết quả so sánh các trình tự của công cụ này sẽ đưa ra một trình tự bảo tồn cao nhất (Ký hiệu *Prim. cons*). Sau khi nhập dữ liệu ClustalW-PBIL sẽ tính toán và trả kết quả sau vài phút tùy vào số lượng trình tự cần so sánh. 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 (Bảng 3.4), 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 với nhau. Chúng tôi nhận thấy các enzyme thu thập được trong họ GH43 hoạt động tối ưu ở pH axit (pH = 3,5) đến trung tính (pH = 7) và nhiệt độ từ 30°C đến 80°C, có nguồn gốc từ vi khuẩn, 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.



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*

Dựa vào mức độ bảo tồn các vị trí axit amin, 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 vị trí. Kết quả so sánh 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 có ít hơn 50% trình tự sẽ được đánh dấu là màu tím. Kết quả này cho thấy trình tự axit amin của β–xylosidase thuộc họ GH43 ở vi khuẩn bảo tồn rất cao, đảm bảo cho số liệu mẫu dò đáng tin cậy.

3.2.4. Xây dựng mẫu dò và giá trị tham chiếu

```
I2NPVLKGFNPDPSICRVGEDYYIA2STFEWFPGV2I2HSKDLVNWRL22RPL2R2SQ
LDMKGNPDSGGVWAPCLSY2DGKFWLIYTDVKVVDGPWKDGHNYLVTADDIDGPWSPE
P2PLNSSGFDPSLFHDDDG2KYLLNMLWGHREGHHSFGGIV2QEFSVAEKKLVGERKI
IF2GTP2KLTEAPHLYKI2GYYYLLTAEGGTRYEHAVTVARSK2I3GPYEVHPDNPIL
TS2H2PE2PLQK2GHASIV2THTGEWYLAHLTGRPIPTPPRGYCPLGRETAIQKLEWK
DGWPYVGGGKE22LEVE2PRYI4EHP222TY3EVDEFDD2TLN3HFQTLRIPFTEQ2G
SLTERPGHLRLYGRESLTS5FTQAFVARRWQHF2FTAETA22FKP2TFQQSAGLVNYY
NTENW22LQVTYDETGGR2L3V33CDNLV2SQPL3DDI2YVYLRV2VDREYKYSYSFD
GE5W3KIPVPLESTKLSD2YIRG2GGFFTGAFVGL2C2YD2SG5GLPADFDYF2YEE3
```

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

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 các 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 để xây dựng mẫu dò. Những vị trí khác (màu tím) sẽ loại bỏ. Mẫu dò này sẽ được so sánh lại với 11 trình tự để xác định mức độ bao phủ và tương đồng bằng BLASTP.

Mẫu dò của họ GH43 được xây dựng bao gồm có 464 axit amin. Trong đó chứa toàn bộ 7 axit amin màu đỏ, 32 vị trí màu xanh lục, 30 vị trí màu xanh lam, 71 vị trí axit amin màu cam, 102 vị trí màu hồng, 165 vị trí màu đen và 57 vị trí là số (Hình 3.2). Trình tự mẫu dò thu được cho họ GH43 như hình 3.3.

Kết quả so sánh mức độ tương đồng giữa mẫu dò với từng trình tự 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 các 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 các trình tự số 4, 5, 9 vì chỉ số điểm tối đa và độ tương đồng thấp. 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 và trình tự 5 mã hóa cho β–xylosidase. Do đó để kiểm chứng chính xác hoạt tính của trình tự axit amin 4, 5 và 9 một lần nữa chúng tôi sử dụng công cụ Swiss model. Kết quả cho thấy cấu trúc không gian của cả 03 trình tự này đều tương đồng cao với cấu trúc của 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 chức năng, nhưng chức năng alpha-arabinofuranosidase hoạt động mạnh hơn, còn với trình tự 5 nghiên cứu chưa thử hoạt tính alpha-arabinofuranosidase.

Bảng 3.5. So sánh mức độ tương đồng của mẫu dò với các trình tự thuộc họ 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

Bảng 3.6. Ước đoán cấu trúc bậc ba của trình tự 4, 5, 9 bằng 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, 3Å	monomer	1 x CA
Trình tự 5	3c2u.1.A	Xylosidase/arabinofuranosidase	0,94	27,08	X-ray, 3Å	homo-tetramer	4 x B3P
Trình tự 9	5glk.1.A	β-xylosidase/α-arabinofuranosidase	0,56	36,84	X-ray, 1.7Å	monomer	None

Để khai thác tối đa các trình tự gen mã hóa enzyme có hoạt tính của β–xylosidase, khi sử dụng mẫu dò các 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 chắc chắn trình tự gen lựa chọn khi đưa vào thực nghiệm thành công, 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%.

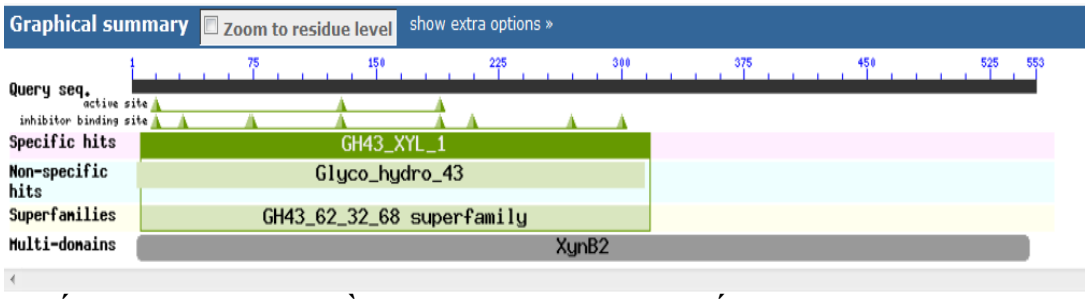
3.2.5. Khai thác trình tự gen mã hóa β–xylosidase bằng mẫu dò từ số liệu DNA đa hệ gen của vi sinh vật trong ruột mối *C. gestroi*

Công ty BGI khi giải trình tự dựa vào CSDL của KEGG đã chú giải 46 trình tự có hoạt tính β–xylosidase thuộc họ GH43. Khi sử dụng mẫu dò, nếu sử dụng chỉ số 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 (Bảng 3.7).

Bảng 3.7. So sánh số lượng trình tự 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



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*

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, cho thấy chúng 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.

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

Để xác định mức độ chính xác của các trình tự đã lựa chọn bằng mẫu dò, chúng tôi tiếp tục kiểm tra lại cấu trúc không gian bằng công cụ Swiss model. Kết quả cho thấy đa số các 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 (Max score) thấp dưới 100, thì hoạt tính khác β–xylosidase như hoạt tính arabinofuranosidase và xylanase (Bảng 3.8). Kết quả này cho thấy, chỉ số điểm tối đa là rất quan trọng để xem xét tính xác thực của gen, 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 làm giá trị tham chiếu khi sử dụng mẫu dò sẽ giúp cho độ chính xác cao hơn khi chọn gen đưa vào thực nghiệm.

Bảng 3.8. Ước đoán cấu trúc bậc ba của các trình tự bằng 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/ Arabinosidase	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	homo-tetramer	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	homo-tetramer	4 x CA
GL0015489	161	2exh.1.A	beta-D-xylosidase	0,98	46,56	X-ray, 1.9Å	homo-tetramer	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	homo-tetramer	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	homo-tetramer	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	xylan 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	xylan beta-1,4-xylosidase	0,81	42,77	X-ray, 1.9Å	homo-tetramer	4 x CA, 3 x MES
GL0039878	55,8	1yrz.1.A	xylan 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	homo-tetramer	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

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% được kiến nghị ở trên 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 để thực nghiệm sẽ ưu tiên các trình tự từ kết quả được lựa chọn bằng mẫu dò và 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.

Bảng 3.9. Độ bao phủ và tương đồng với mẫu dò của các gen 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	

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 đặc biệt quan tâm đến khả năng enzyme hoạt động tối ưu trong môi trường kiềm và nhiệt độ cao. Do đó sử dụng công cụ Alcapred và TBI để tiếp tục chọn lọc 04 gen hoàn thiện bằng các dự đoán pH và nhiệt độ hoạt động tối ưu.

Bảng 3.10. Kết quả dự đoán pH và nhiệt độ hoạt động của các mã gen β–xylosidase

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

Sử dụng phần mềm TBI dự đoán nhiệt độ hoạt động tối ưu của enzyme cho thấy: cả 4 enzyme đều có khả năng hoạt động tối ưu ở nhiệt độ từ 55℃ đến 65℃. Khi dự đoán pH hoạt động tối ưu bằng phần mềm Alcapred, kết quả cho thấy các 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 phần mềm này chỉ số 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.

Tóm lại, sử dụng mẫu dò của enzyme 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 tối thiểu lần lượt là 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 có sẵn. 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.

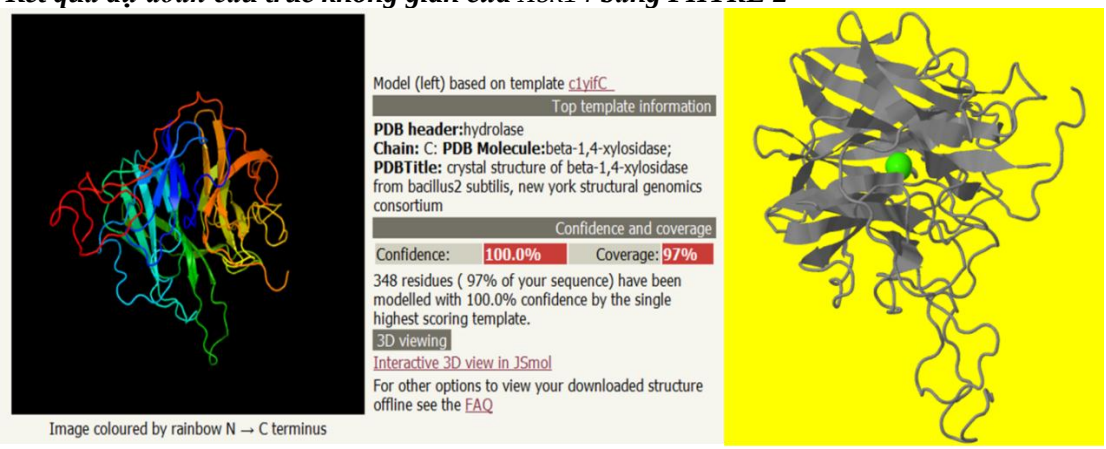
3.2.8. Một số dự đoán chi tiết gen GL0112518 mã hóa β–xylosidase (*Xbx14*)

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

Mã gen GL0112528 có chiều dài là 1077 bp. Kết quả dự đoán bằng BLASTP cho thấy mã gen này mã hóa cho *Xbx14* thuộc họ GH43, đúng như dự đoán ban đầu khi lựa chọn bằng mẫu dò (Hình 3.6).

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

3.2.8.2. Kết quả dự đoán cấu trúc không gian của Xbx14 bằng PHYRE 2



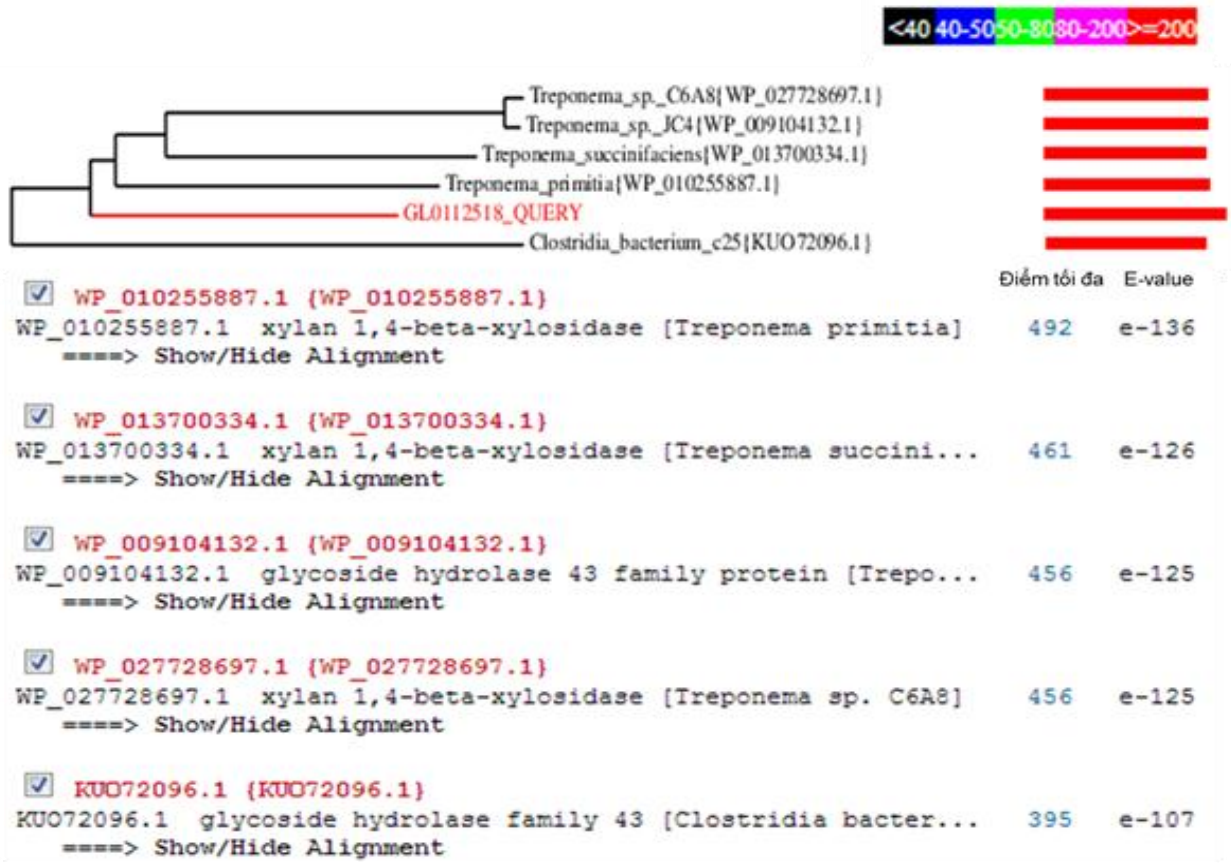
***Hình 3.7. Mô hình cấu trúc không gian (A) dựa vào khuôn c1ylfC và trung tâm hoạt động (B) của enzyme** Xbx14*

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).

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

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 các vi khuẩn *Treponema primitia* (Hình 3.8). Đây chính là nhóm vi khuẩn chiếm số lượng ORF lớn nhất, đặc trưng 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. Sử dụng công cụ kiểm tra nguồn gốc cho thấy, mã gen GL0112518 tương đồng cao nhất với trình tự gen từ vi khuẩn *Treponema primitia* – vi khuẩn đặc trưng trong ruột mối *C. gestroi*.



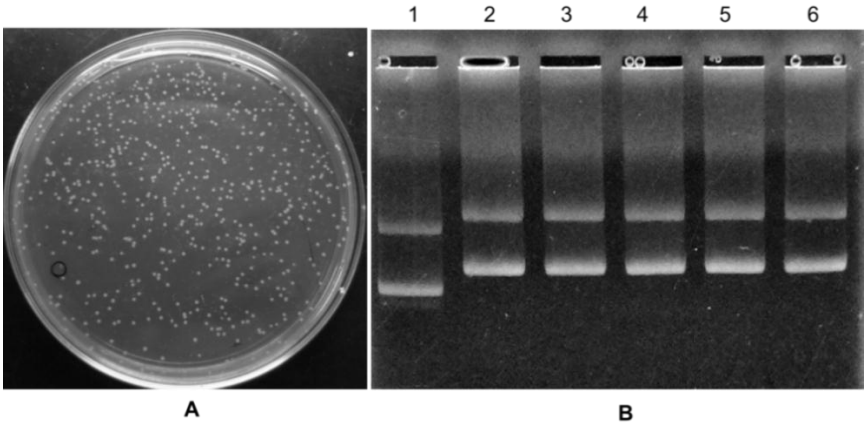
Hình 3.8. Sơ đồ nguồn gốc mã gen GL0112518

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(+)

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.

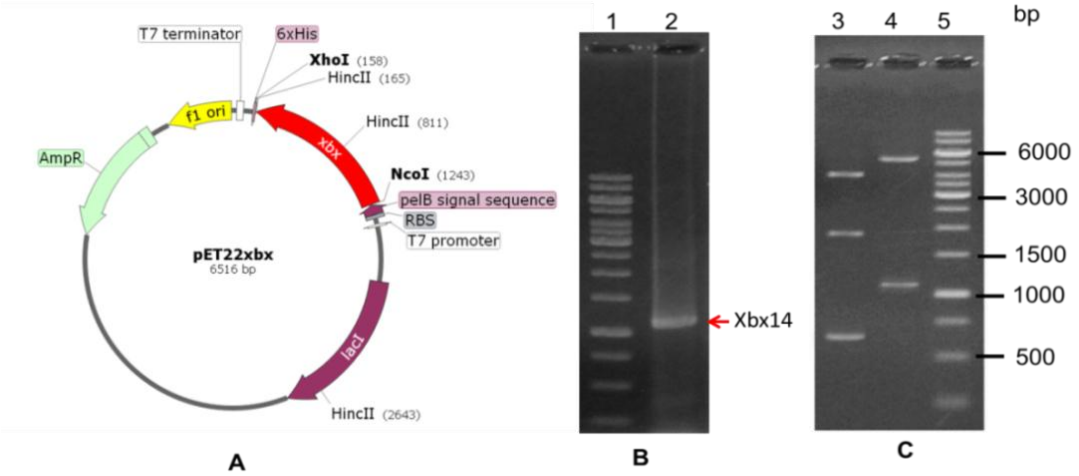


Hình 3.9. Kết quả biến nạp vector Pet22Xbx ***(A) và điện di đồ phân tích sản phẩm tách dòng gen*** *Xbx14* ***(B). Đường chạy 1: Plasmid Pet22b(+); Đường chạy 2-6: Plasmid*** Pet22Xbx

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.

Sơ đồ các enzyme cắt hạn chế trên vector pET22Xbx cho thấy enzyme *Xho*I, *Nco*I chỉ có 1 điểm cắt, còn *Hinc*II có 3 điểm cắt trên vector (Hình 3.10A). 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 HincII cho ra ba băng với kích thước lần lượt là 4 kb, 1.8 kb và 0.6 kb và cắt đồng thời bằng *Nco*I, *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(+).



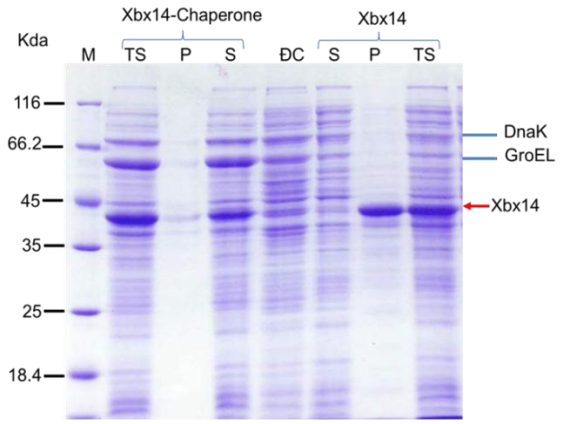
Hình 3.10. Sơ đồ vị trí cắt của enzyme cắt hạn chế HincII, XhoI và NcoI trên vector Pet22Xbx ***(A), điện di sản phẩm PCR (B) và sản phẩm cắt*** Pet22Xbx ***bằng NcoI và XhoI, HincII (C)***

Đường chạy 1, 5: Thang DNA chuẩn (Fermentas), đường chạy 2: Sản phẩm PCR gen Xbx14, đường chạy 3: cắt Pet22Xbx *HincII, đường chạy 4:* Pet22Xbx *bằng NcoI và XhoI*

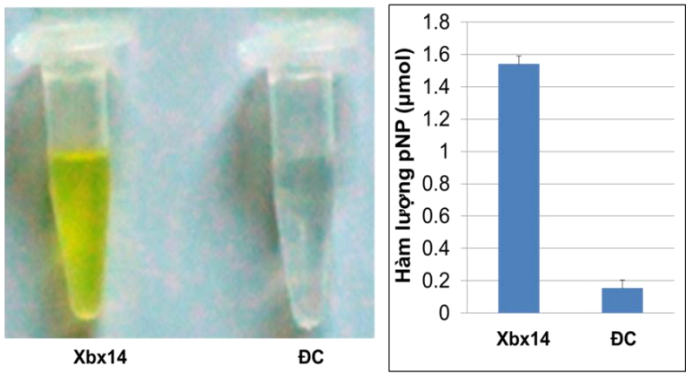
Kết quả sau khi giải trình tự bằng mồi xuôi và mồi ngược đặc hiệu đã được xử lý cho thấy trình tự gen *Xbx14* đã được gắn vào vector tách dòng pET22b(+) đúng như trình tự gốc ban đầu (Phụ lục 1).

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

Vector tái tổ hợp pET22Xbx được biến nạp vào chủng biểu hiện *E.coli Rosetta 1* để khảo sát biểu hiện. Kết quả protein *Xbx14* biểu hiện rất tốt, nhưng không tan. Vì vậy nghiên cứu này tiếp tục tiến hành đồng biểu hiện gen *Xbx14* trong vector pET22Xbx và gen tổng hợp các chaperone GroEL (60 kDa), DnaK (70 kDa) và DnaJ (40 kDa) trong vector pG - KJE8. Kết quả đồng biểu hiện với chaperone sau 5 giờ cảm ứng, được phân tích bằng điện di trên gel SDS-PAGE 12,6% (Hình 3.11). 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. Khi so sánh với sản phẩm biểu hiện gen *Xbx14* không có chaperone, thì protein *Xbx14* hầu hết nằm ở pha tủa. Điều này chứng tỏ đồng biểu hiện với chaperone đã giúp cho *Xbx14* đã biểu hiện tốt và ở dạng tan. Đồng thời kết quả thử hoạt tính thô 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).



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

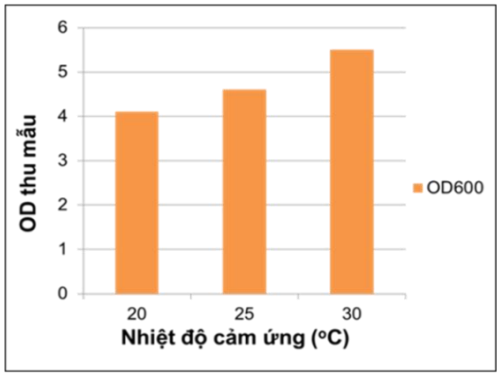


Hình 3.12. Kết quả thử hoạt tính thô của Xbx14

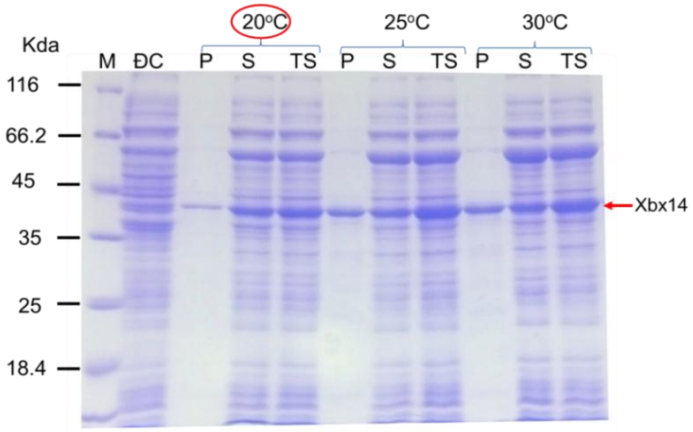
M: Thang protein chuẩn (Fermentas); ĐC: Protein tổng số không cảm ứng IPTG; Xbx14: protein Xbx14 biểu hiện không có chaperone; Xbx14-Chaperone: protein Xbx14 đồng biểu hiện với chaperone pG-KJE8; TS: protein ở dạng tổng số; S: protein ở dạng tan; P: protein ở dạng tủa

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



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

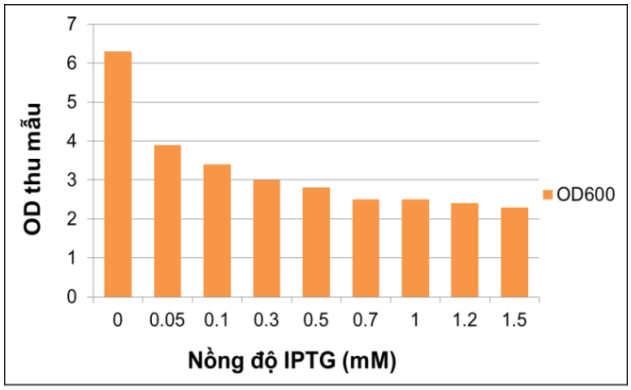


Hình 3.14. Điện di đồ phân tích biểu hiện Xbx14 trong chủng E. coli Rosetta 1 sau 6 giờ cảm ứng 0,5 mM IPTG ở 20°C, 25°C, 30°C trên gel PAGE-SDS 12,6%

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)*

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. Kết quả thu tế bào 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). 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, 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. Kết hợp khuyến cáo của hãng cung cấp vector *pG-KJE8* và kết quả lượng protein tái tổ hợp tan ở các nhiệt độ trên, chúng tôi sẽ lựa chọn nhiệt độ 20°C để biểu hiện gen *Xbx14* trong các thí nghiệm tiếp theo.

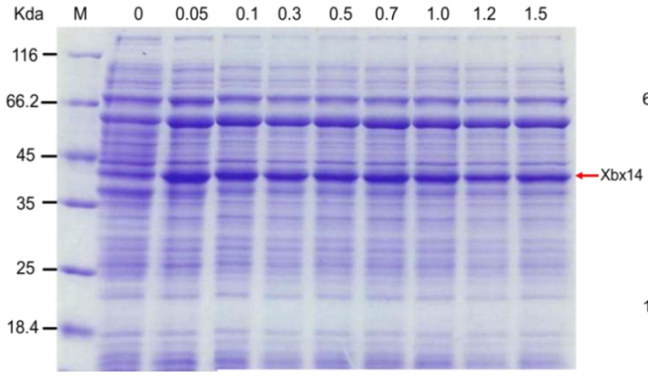
b. Nghiên cứu ảnh hưởng của nồng độ IPTG đến hiệu quả cảm ứng



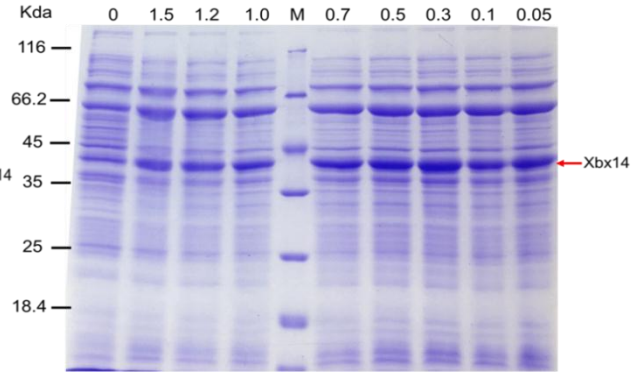
Hình 3.15. Mật độ tế bào biểu hiện E. coli Rosetta 1 (DE3) mang Pet22Xbx ***cảm ứng IPTG ở các nồng độ từ 0,0 đến 1,5 mM***

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°C, 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. Tất cả các mẫu đã cảm ứng được nuôi ở 20°C trong thời gian 5 giờ.

Mẫu tế bào cảm ứng ở mỗi nồng độ IPTG thu được, sẽ 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). Kết quả cảm ứng biểu hiện *Xbx14* tổng số ở các nồng độ IPTG khác nhau được kiểm tra bằng điện di trên gel polyacrylamide 12,6% có SDS (Hình 3.16). Từ điện di đồ kiểm tra protein *Xbx14* tổng số cho thấy: ở nồng độ IPTG là 0,05 mM lượng protein tái tổ hợp *Xbx14* thu được nhiều nhất.



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

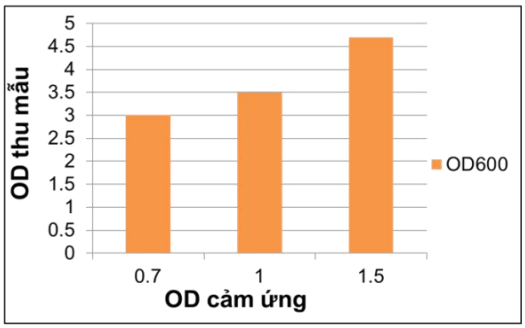


Hình 3.17. Điện di đồ phân tích protein Xbx14 tan biểu hiện trong tế bào E. coli Rosetta 1 (DE3) cảm ứng 09 nồng độ IPTG trên gel PAGE-SDS 12,6%

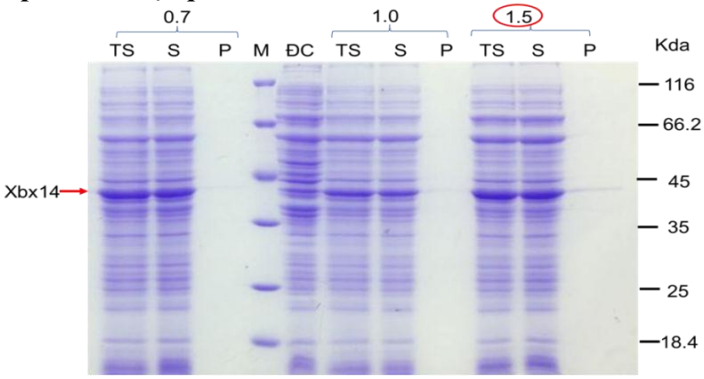
Các đường chạy 0,0 – 1,5: là protein tổng số/tan biểu hiện khi tế bào được cảm ứng nồng độ IPTG từ 0,0 đến 1,5 mM; đường chạy M: protein chuẩn (Thermo scientific)

Kiểm tra lượng *Xbx14* dạng tan cho thấy ở các nồng độ IPTG cảm ứng khác nhau thì lượng protein tái tổ hợp tan cũng tỷ lệ thuận với lượng protein *Xbx14* tổng số (Hình 3.16, Hình 3.17). 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 nhất là 0,05 mM.

c. 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

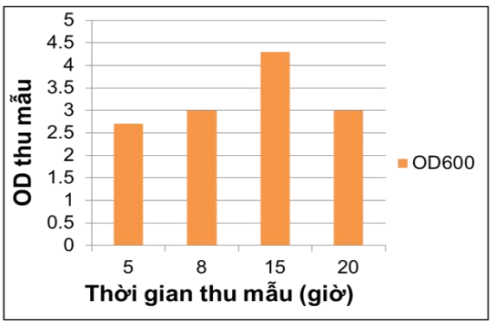


Hình 3.19. Điện di đồ protein Xbx14 biểu hiện trong tế bào E. coli Rosetta 1 (DE3) cảm ứng IPTG tại các OD_{600} = 0,7; 1,0; 1,5 trên gel PAGE-SDS 12,6%

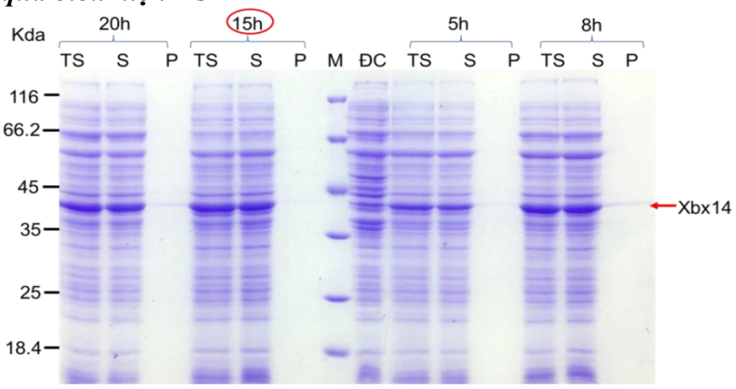
Đường chạy ĐC: đối chứng âm, protein tổng số của dòng tế bào biểu hiện không cảm ứng IPTG; đường chạy TS: protein tổng số biểu hiện trong tế bào mang Pet22Xbx*; đường chạy S: protein dạng tan; đường chạy P: protein dạng không tan; đường chạy M: protein chuẩn unstained (Thermo scientific)*

Mẫu tế bào được nuôi ở 25°C để cảm ứng biểu hiện chaperone và 20°C cảm ứng biểu hiện *Xbx14*. 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). 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. Tuy nhiên cảm ứng ở OD_{600} = 1,5 thu được sinh khối tế bào lớn hơn (Hình 3.19). Do đó chúng tôi sẽ chọn cảm ứng ở OD_{600} = 1,5 cho các thí nghiệm tiếp theo.

d. Ảnh hưởng thời gian thu mẫu đến hiệu quả biểu hiện Xbx14



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 protein Xbx14 biểu hiện trong tế bào E. coli Rosetta 1 (DE3) thu mẫu sau 5, 8, 15 và 20 giờ cảm ứng IPTG trên gel PAGE-SDS 12,6%

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).*

Tế bào sau khi cảm ứng 0,05 mM IPTG ở OD_{600} là 1,5 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. Kết quả đo OD_{600} ở các 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 biểu hiện sau khi thu, tiếp tục được phá bằng siêu âm để kiểm tra mức độ biểu hiện và dạng tan của *Xbx14*. Kết quả cho thấy lượng protein *Xbx14* tổng số và tan thu được tại thời điểm thu mẫu sau 5 giờ cảm ứng ít nhất, sau 8 giờ và 15 giờ tăng lên và 20 giờ 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 hợp với số liệu mật độ tế bào thu mẫu cuối cùng, chúng tôi lựa chọn thời điểm thu mẫu là sau 15 giờ cảm ứng, ở 20°C.

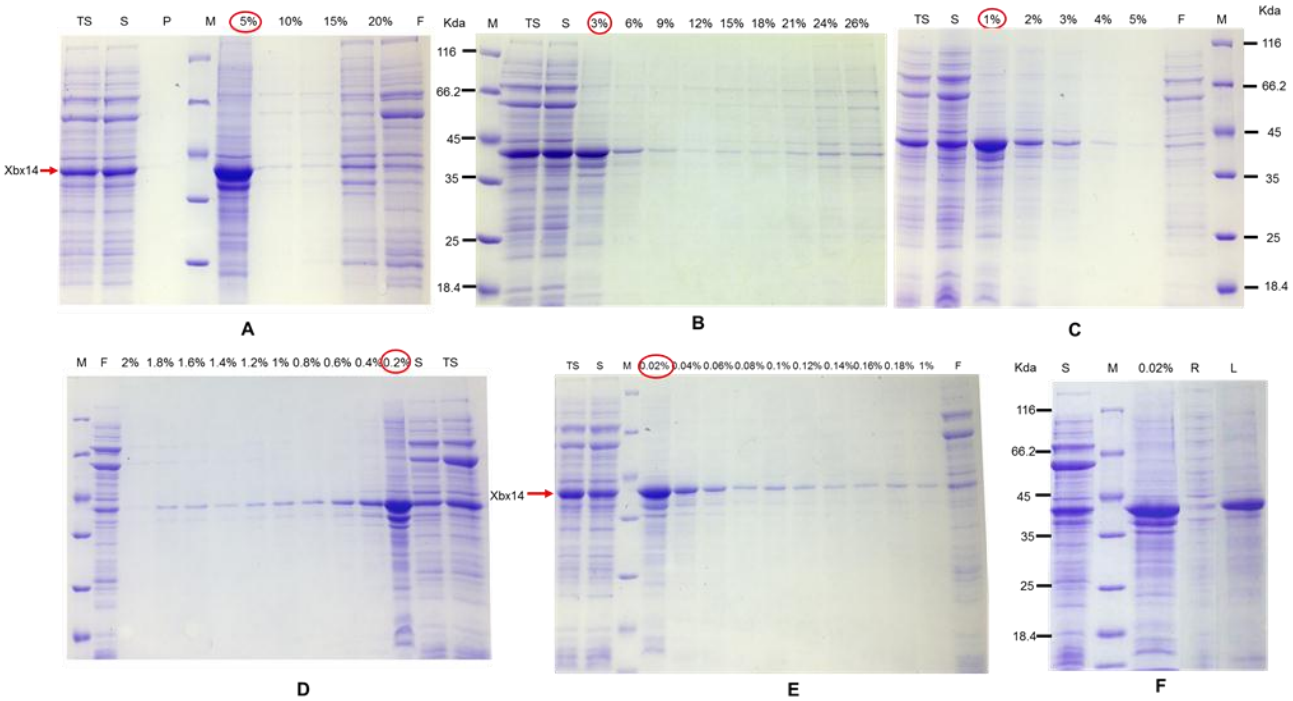
Như vậy các điều kiện tối ưu đồng biểu hiện *Xbx14* với các chaperone trong tế bào *E. coli Rosetta 1* bao gồm: nhiệt độ biểu hiện 20°C, nồng độ IPTG cảm ứng 0,05 mM, OD_{600} cảm ứng 1,5 và thu mẫu sau 15 giờ cảm ứng.

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 có một đoạn trình tự mã hóa cho 6 axit amin Histidine tích điện âm để sử dụng trong tinh chế bằng cột sắc kí ái lực. Tuy nhiên mẫu protein *Xbx14* chỉ tan trong nước và bị tủa hoàn toàn ở nồng độ muối rất thấp. 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.

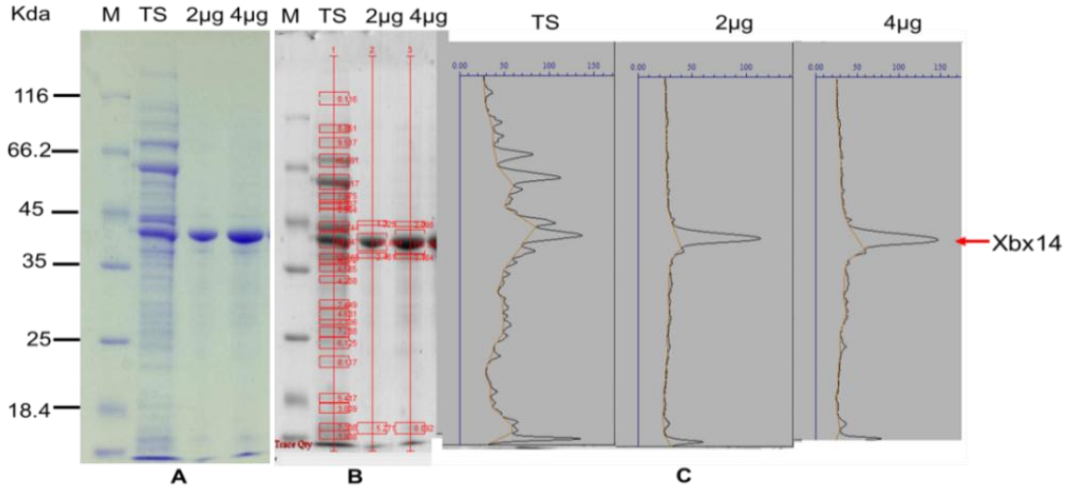
3.3.2.1. Tinh chế Xbx14 bằng muối $(NH4)_2SO4$

Kết quả cho thấy *Xbx14* bị tủa theo từng phân đoạn từ nồng độ muối cao, đến nồng độ muối rất thấp. Hình 3.22A cho thấy *Xbx14* bị tủa ở nồng độ muối 5% $(NH_4)_2SO_4$, tiếp tục bị tủa khi sử dụng các nồng độ muối $(NH_4)_2SO_4$ 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 nhiễm còn rất ít. Để tiếp tục làm sạch các protein tạp nhiễm, chúng tôi dựa trên điểm đẳng điện của protein *Xbx14* và khả năng hòa tan trong đệm có pH khác nhau. Kết quả chỉ ra, hầu hết các 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).



Hình 3.22. Điện di đồ 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 đường chạy đệm phosphate; M: protein chuẩn unstained (Thermo scientific)*

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



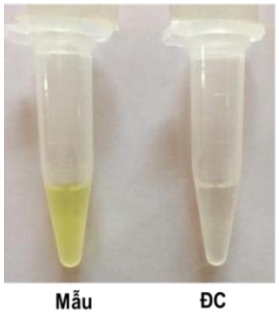
Hình 3.23. Kết quả kiểm tra độ sạch của Xbx14 sau khi tinh chế bằng muối $(NH_4)_2SO_4$ và rửa bằng đệm phosphate (pH = 9)

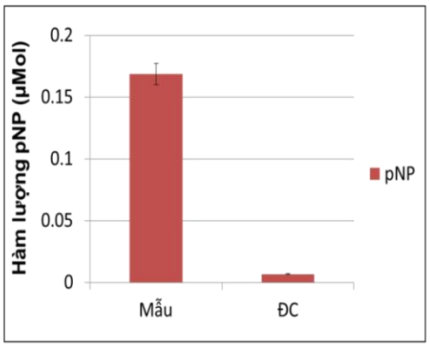
Bằng phần mềm Quantity One, độ sạch tương đối của protein *Xbx14* đã tinh chế được đánh giá ở mức định lượng. Mẫu được đo hàm lượng và tính toán cho mỗi đường chạy là 2 µg và 4 µg protein (Hình 3.23A). Một vị trí nền (BGR) có màu sắc và độ sáng giống với nền ở vị trí băng protein *Xbx14* trên ảnh quét 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 3.23C). Kết quả phân tích hình ảnh điện di được quét và được tính toán của protein *Xbx14* so với với protein tổng số chiếm 19,8%. Tuy nhiên sau khi tinh chế thì *Xbx14* chiếm tỷ lệ 92,7%.

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. 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 sẽ được xác định chính xác bằng phương pháp đo Elisa ở bước sóng 405 nm. Kết quả đo được tính toán theo đường chuẩn cho thấy, sau 2 giờ phản ứng 30 µl Xbx14 (0,253 mg/ml) thủy phân 20 µl pNPX 10mM tạo ra lượng pNP là 0,1692 µMol (Hình 3.24).





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

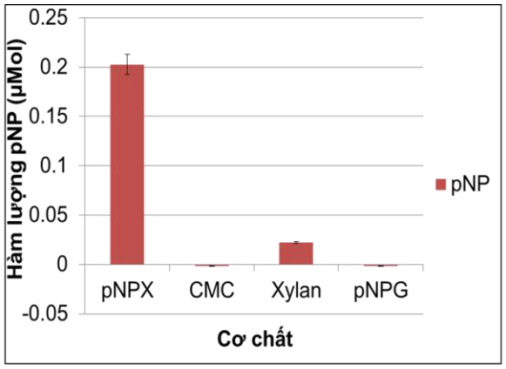


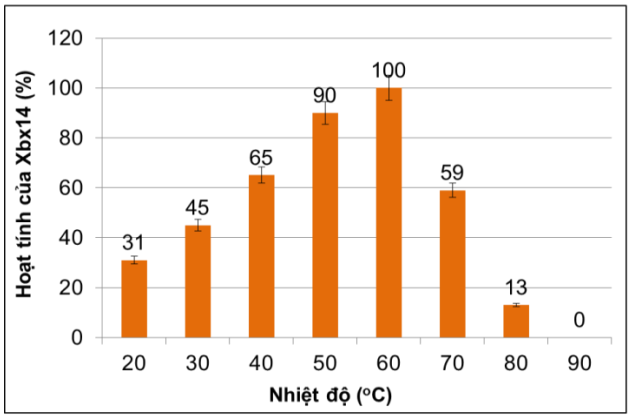
Fig 3.25. Kết quả xác định hoạt tính của Xbx14 trên các cơ chất khác nhau

3.3.3.2. Xác định tính đặc hiệu cơ chất của Xbx14

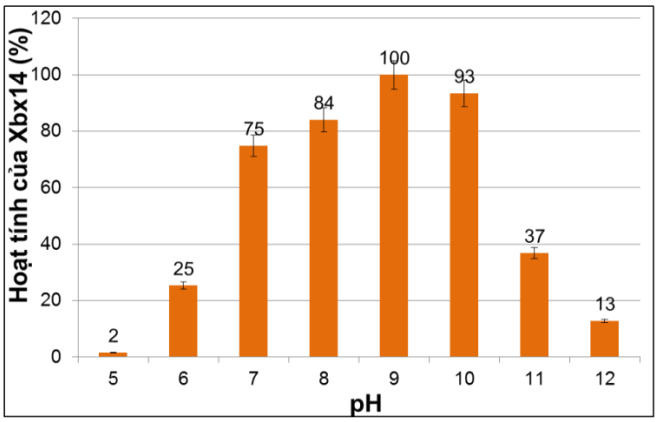
Một số cơ chất phổ biến là CMC, *p*-nitrophenyl-β-glucoside (*p*NPG), 4-nitrophenyl β- D-xylopyranoside (pNPX) và xylan được sử dụng để kiểm tra tính đặc hiệu của *Xbx14*. Trong đó, pNPX là cơ chất đặc hiệu dạng tan của *Xbx14*, CMC là cơ chất đặc hiệu của endoglucanase, xylan là cơ chất đặc hiệu của xylanase 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 enzyme β–xylosidase phân cắt các đường đôi xylose và phân cắt phân tử đường ở các đầu của mạch xylan.

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

Tiếp tục nghiên cứu hoạt động tối ưu của enzyme *Xbx14* theo một dải nhiệt độ là 20°C, 30°C, 40°C, 50°C, 60°C, 70°C, 80°C và 90°C. Kết quả cho thấy hoạt tính β–xylosidase tăng dần từ 20°C đến 60°C và giảm dần khi nhiệt độ tăng đến 80°C và giảm mạnh ở nhiệt độ 90°C. Như vậy, nhiệt độ thích hợp nhất cho hoạt động của *Xbx14* là 60°C (Hình 3.26), cao hơn hầu hết các nghiên cứu đã công bố về nhiệt độ hoạt động của β–xylosidase và đúng với dự đoán của công cụ TBI. Đây hứa hẹn sẽ là một enzyme mới, có khả năng chịu nhiệt rất tốt.



Hình 3.26. Ảnh hưởng nhiệt độ đến hoạt tính Xbx14



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

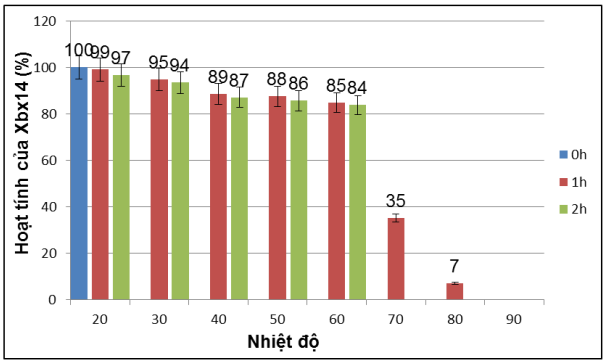
3.3.3.4. Nghiên cứu ảnh hưởng của pH đến hoạt tính của Xbx14

Theo ước đoán bằng phần mềm Alcapred, *Xbx14* là enzyme có khả năng hoạt động tốt ở vùng pH kiềm. Kết quả kiểm tra hoạt tính cho thấy, enzyme không có hoạt tính ở pH = 5, hoạt tính tăng dần từ pH = 6 đến pH = 9 và đạt giá trị cao nhất ở pH = 9. Sau đó hoạt tính giảm nhẹ ở pH = 10 và giảm sâu ở pH = 12 (Hình

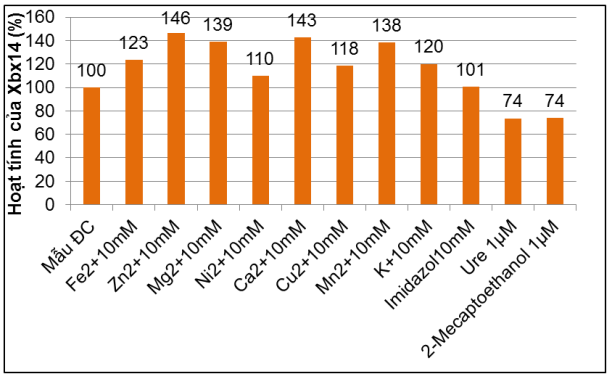
3.27). Như vậy pH hoạt động tối ưu của *Xbx14* thích hợp trong môi trường kiềm tính, đúng với dự đoán ban đầu khi sử dụng công cụ Alcapred. Enzyme này có tiềm năng ứng dụng rất hiệu quả khi kết hợp với các điều kiện xử lý sinh khối thực vật bằng dung dịch kiềm. Đây cũng chính là enzyme mà chúng tôi đang muốn tìm kiếm trong nghiên cứu này.

3.3.3.5. Nghiên cứu độ bền nhiệt của Xbx14

Xbx14 được xử lý ở nhiệt độ từ 20°C, 30°C, 40°C, 50°C, 60°C, 70°C, 80°C và 90°C trong các khoảng thời gian khác nhau 1 giờ và 2 giờ trước khi ủ với cơ chất, để kiểm tra độ bền nhiệt. Kết quả cho thấy ở nhiệt độ lớn hơn 70°C, *Xbx14* mất hoàn toàn hoạt tính dù là thời gian xử lý ngắn nhất (1 giờ). Ở các 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).



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

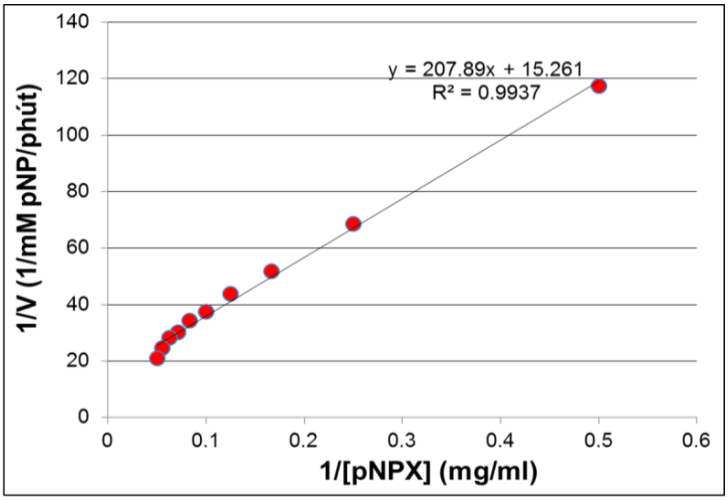


Hình 3.29. Ảnh hưởng một số ion kim loại và hóa chất lên hoạt tính của Xbx14

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

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 hoặc Shi và cộng sự. Kết quả chỉ ra ở nồng độ 10 mM, các ion kim loại đều làm tăng hoạt tính của enzyme *Xbx14*. Imidazol với nồng độ 10 mM không ảnh hưởng đến hoạt tính của *Xbx14*, tuy nhiên ở nồng độ 1 µM 2- mercaptoethanol và 1 µM urea làm giảm đáng kể đến hoạt tính của *Xbx14* (Hình 3.29).

3.3.3.7. Nghiên cứu đặc điểm động học của Xbx14



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

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. So với hoạt tính riêng của các β-xylosidase đã được nghiên cứu từ nấm và vi khuẩn, thì hoạt tính của *Xbx14* khá cao.

Tóm lại 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 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* 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.

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 đa hệ gen 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.