

# NHÂN DÒNG VÀ PHÂN TÍCH TRÌNH TỰ GEN *glpF* MÃ HÓA PROTEIN XUYÊN MÀNG HỖ TRỢ VẬN CHUYỂN GLYCEROL Ở *Escherichia coli*

Phạm Thị Kiều Duyên<sup>1</sup>, Nguyễn Thị An Hòa<sup>1</sup>, Nguyễn Hà Mi<sup>1</sup>, Quyền Mỹ Linh<sup>1</sup>,  
Nguyễn Quang Huy<sup>1,2</sup>, Kiều Cẩm Nhung<sup>3</sup>, Chử Lương Luân<sup>2,3</sup>

<sup>1</sup>Khoa Sinh học, Trường Đại học Khoa học Tự Nhiên, Đại học Quốc gia Hà Nội

<sup>2</sup>Phòng Thí nghiệm Trọng điểm Công nghệ Enzym và Protein, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

<sup>3</sup>Bộ Khoa học và Công nghệ

## TÓM TẮT

Biodiesel là nguồn năng lượng tái tạo, hứa hẹn có thể thay thế cho nhiên liệu hóa thạch truyền thống bởi ưu điểm thân thiện với môi trường. Sự phát triển của công nghiệp biodiesel trên phạm vi toàn cầu đã dẫn đến dư thừa một trữ lượng lớn glycerol thô - sản phẩm phụ từ quá trình sản xuất. Điều này gây mất cân bằng giá thành glycerol thương mại và được xem là thách thức đối với tính bền vững của công nghiệp biodiesel. Từ đó, mục tiêu chuyển đổi glycerol thô thành sản phẩm có giá trị cao đã trở thành vấn đề cấp thiết ở nhiều quốc gia. *Escherichia coli* (*E. coli*) đã được chứng minh có thể sử dụng glycerol như một nguồn carbon thay thế thông qua sự kết hợp giữa glycerol kinase (GlpK) và glycerol 3-phosphate dehydrogenase (GlpD). Tuy nhiên, sự biểu hiện của gen aquaglyceroporin (*glpF*) là mấu chốt quan trọng, làm tăng tốc độ sinh trưởng và sự hình thành sinh khối. GlpF có cấu trúc bảo tồn, tính chọn lọc cao và được biết đến là protein vận chuyển duy nhất xúc tác cho quá trình khuếch tán glycerol qua màng trong của *E. coli*. Trong nghiên cứu này, gen *glpF* từ *E. coli* chủng DH5 alpha đã được nhân dòng vào vector pJET 1.2/ blunt và giải trình tự theo phương pháp Sanger. Trình tự gen đích được giải mã gồm 846 nucleotide và tương đồng 100% với trình tự gen tham chiếu. Kết quả so sánh trình tự amino acid suy diễn và phân tích cấu trúc protein bậc 3 cho kết quả tương đồng 100% với độ tin cậy là tuyệt đối. Kết quả nghiên cứu này cung cấp cơ sở khoa học, tạo nguyên liệu di truyền để phát triển chủng vi sinh vật tái tổ hợp giúp tăng cường biểu hiện gen *glpF* mã hóa cho protein xuyên màng hỗ trợ vận chuyển glycerol.

Từ khóa: *Escherichia coli*, glycerol, *glpF*, trình tự gen, nhân dòng.

## MỞ ĐẦU

Biodiesel là nhiên liệu có thể tái chế, được sản xuất từ dầu thực vật hoặc mỡ động vật với mục tiêu giảm thiểu ô nhiễm môi trường và thay thế cho nhiên liệu hóa thạch. Tuy nhiên, quá trình chế tạo dầu biodiesel dẫn đến dư thừa một lượng lớn glycerol - sản phẩm phụ của quá trình sản xuất. Theo tính toán, mỗi 9 kg dầu biodiesel được sản xuất từ quá trình este hóa sẽ tạo ra 1 kg glycerol thô (Zhou *et al.*, 2018). Mặc dù, từng được coi là loại "sản phẩm phụ" có giá trị, glycerol thô đang nhanh chóng trở thành một "sản phẩm thải" do chi phí xử lý cao và ảnh hưởng tới giá thành của glycerol tinh khiết. Theo nghiên cứu, sản lượng glycerol thô trên thị trường tăng nhanh và có liên quan trực tiếp tới sự phát triển của biodiesel. Mặt khác, glycerol thô chưa qua tinh chế không thể thải ra môi trường vì đốt trực tiếp sẽ tạo ra khí độc (Gupta & Kumar, 2012). Việc phát triển một quy trình bền vững với mục đích chuyển đổi glycerol thành sản phẩm có giá trị đang là vấn đề cấp thiết của các lĩnh vực kinh tế, năng lượng và công nghệ (Chiang *et al.*, 2020). Hiện nay, phương pháp chuyển đổi glycerol thường được tiếp cận bao gồm chuyển hóa hóa học và đặc biệt là lên men sử dụng vi sinh vật cải biến di truyền. Một số vi khuẩn có thể chuyển hóa glycerol thuộc chi *E. coli*, *Klebsiella*, *Enterobacter*, *Glucanobacter*, *Clostridium*, *Candida*, *Aspergillus*. Glycerol thô thường được chuyển đổi thành các sản phẩm có giá trị như: ethanol, propionic acid, biogas, citric acid, ethanol, acid docosaheptaenoic và một số acid béo omega-3 khác (Chilakamarry *et al.*, 2021). Với lợi ích về chi phí và bảo vệ môi trường, công nghệ DNA tái tổ hợp được xem là lĩnh vực triển vọng trong nghiên cứu chuyển hóa glycerol thô.

*E. coli* mang plasmid thường được sử dụng trong lĩnh vực nghiên cứu tái tổ hợp với ưu điểm về thao tác di truyền và chi phí mở rộng sản xuất. Vi khuẩn này có khả năng sử dụng một số hợp chất làm nguồn carbon và glucose được ưu tiên sử dụng do ái lực cao với hệ thống phosphotransferase (PTS). Mặt khác, glucose cung cấp nhiều năng lượng cho tế bào thông qua việc khởi động quá trình đường phân dưới dạng glucose 6-phosphate và được tiêu thụ qua chu trình Krebs (Martínez-Gómez *et al.*, 2012). Tuy nhiên, *E. coli* có thể sinh trưởng trong môi trường giàu glycerol và sử dụng chúng như một nguồn carbon thay thế. Mặt khác, người ta còn chứng minh rằng glycerol có ảnh hưởng tích cực tới thời gian sống của tế bào và năng suất protein tái tổ hợp.

Glycerol có thể đi qua màng tế bào thông qua cơ chế khuếch tán thụ động không đặc hiệu, tích lũy ở tế bào chất và có thể di chuyển theo gradient nồng độ ra khỏi tế bào. Ngoài ra, glycerol cũng được xem là ví dụ điển hình cho

quá trình khuếch tán qua kênh vận chuyển trung gian ở *E. coli*. Tương tự như các aquaporin, GlpF là một kênh xuyên màng có vai trò thúc đẩy quá trình thẩm thấu glycerol. Sau khi qua màng tế bào, glycerol trong tế bào chất trải qua quá trình phosphoryl hóa cần ATP và được xúc tác bởi glycerol kinase (GlpK). Sản phẩm của quá trình được giữ lại trong tế bào dưới dạng glycerol-3-phosphate (G3P). Quá trình trình khử hydro diễn ra và dihydroxyacetone phosphate (DHAP) được tạo thành. G3P đóng vai trò là chất cảm ứng của operon *glpFK* và giúp *E. coli* có thể sử dụng glycerol hiệu quả. Mặt khác, G3P không phải chất nền cho GlpF nên được giữ lại bên trong tế bào mà không phụ thuộc gradient nồng độ (Sweet *et al.*, 1990). GlpF thuộc họ aquaporin (AQP), chứa chuỗi -Asn-Pro-Ala- bảo tồn, đã được xác định rõ về cấp độ di truyền và các đặc tính chức năng (Stroud *et al.*, 2003). GlpF gồm 1 xoắn dài xuyên màng với một vị trí uốn cong ở giữa và 2 chuỗi xoắn alpha xuyên màng một nửa, xiên chéo qua màng hình thành một phần của lỗ chọn lọc glycerol. GlpF có tính chọn lọc cao, không cho phép các hợp chất tích điện đi qua giúp duy trì gradient điện hóa trên màng. Trong một số nghiên cứu, *glpF* được xác định mã hóa cho protein xuyên màng GlpF và nằm chung operon với *glpK* mã hóa cho glycerol kinase (GlpK) (Sweet *et al.*, 1990). Hai protein GlpF và GlpK có quan hệ tương tác, sự xuất hiện của GlpF sẽ kích hoạt glycerol kinase (Voegelé *et al.*, 1993). Thêm vào đó, tốc độ vận chuyển glycerol qua GlpF nhanh hơn từ 100-1000 lần so với các phương thức vận chuyển khác (Heller *et al.*, 1980). Tại Việt Nam, đã có nghiên cứu tập trung vào tối ưu giá thành sản xuất biodiesel thay vì tận dụng nguồn sản phẩm có giá trị, sử dụng đối tượng: vi tảo biển, mỡ động vật. Tuy nhiên, những đối tượng nghiên cứu này còn giới hạn về tốc độ sinh trưởng, mùa vụ và khả năng mở rộng sản xuất. Chưa có nghiên cứu nào sử dụng công nghệ DNA tái tổ hợp trong nghiên cứu chuyển hóa trữ lượng glycerol thô dư thừa từ công nghiệp biodiesel.

Tóm lại, biểu hiện *glpF* được coi là mấu chốt quan trọng, làm tăng tốc độ sinh trưởng và sự hình thành sinh khối. Trong nghiên cứu này, chúng tôi tiến hành nhân dòng và phân tích trình tự *glpF*, bước đầu tạo nguyên liệu di truyền cho phát triển chủng tái tổ hợp mang biểu hiện gen quan tâm. Đây là một trong những nghiên cứu đầu tiên tại Việt Nam tiếp cận tới định hướng tận thu glycerol thô từ công nghiệp biodiesel sử dụng công nghệ tái tổ hợp.

## VẬT LIỆU VÀ PHƯƠNG PHÁP

### Chủng vi khuẩn và plasmid

*E. coli* chủng DH5 alpha (Invitrogen, Hoa Kỳ) được sử dụng cho mục đích tách chiết DNA tổng số và thiết kế các bước của thí nghiệm nhân dòng. Vector tạo dòng dùng sử dụng trong nghiên cứu là pJET 1.2/blunt (Invitrogen, Hoa Kỳ).

### Tách chiết genomic DNA

*E. coli* DH5 alpha được nuôi lắc tăng sinh trong 5 mL LB lỏng (160 vòng/phút, 37°C qua đêm). DNA tổng số của *E. coli* DH5 alpha được tách chiết bằng bộ sinh phẩm thương mại Genomic DNA Prep Kit (BioFact, Hàn Quốc). Sản phẩm tinh sạch được điện di trên gel agarose 1% có bổ sung thuốc nhuộm RedSafe™ Nucleic Acid Staining Solution 20000x (iNtRON Biotechnology, Hàn Quốc). Kết quả điện di được quan sát dưới tia tử ngoại.

### Thiết kế cặp mồi đặc hiệu khuếch đại gen *glpF*

Cặp mồi đặc hiệu khuếch đại gen *glpF* được chúng tôi nghiên cứu thiết kế dựa trên trình tự đã được công bố trên ngân hàng gen (GenBank: U00096.3). Bộ mồi được tổng hợp bằng phương pháp hóa học bởi Công ty TNHH MTV Sinh Hóa Phù Sa (PHUSA Biochem) và có trình tự: *glpF*-fw (5' – ATGAGTCAAACATCAACCTTG – 3'); *glpF*-Afl II- rv (5' –CTTAAGTTACAGCGAAGCTTTTTG – 3'). Phản ứng PCR khuếch đại gen *glpF* sử dụng bộ sinh phẩm Platinum™ SuperFi II PCR Master Mix (2X) (Invitrogen, Hoa Kỳ) chứa *Pfu* DNA polymerase biểu hiện hoạt tính exonuclease 3'-5' có khả năng đọc sửa và khuếch đại chính xác gen đích. Quá trình khuếch đại (PCR) được thực hiện bằng thiết bị luân nhiệt SimpliAmp™ Thermal Cycler (Applied Biosystem, Hoa Kỳ) theo chu trình: biến tính 95°C trong 5 phút; 35 chu kỳ với mỗi chu kỳ là 95°C trong 30 giây, 55°C trong 45 giây và 72°C trong 1 phút; cuối cùng giữ 72°C trong 10 phút. Sản phẩm PCR có dạng đầu bằng, kích thước theo tính toán là 846 bp. Kết quả khuếch đại được kiểm tra bằng điện di trên gel agarose TAE 1% ở điện áp 100V trong 25 phút. Tinh sạch sản phẩm PCR sử dụng Monarch DNA Gel Extraction Kit (NEB, Hoa Kỳ). Định lượng sản phẩm sau tinh sạch bằng máy đo quang phổ NanoDrop One (Thermo Fisher Scientific, Hoa Kỳ). Kiểm tra kích thước sản phẩm tinh sạch bằng kỹ thuật điện di trên gel agarose. Trong nghiên cứu này, công cụ ImageJ (Hoa Kỳ) được sử dụng với mục đích phân tích hình ảnh điện di.

### Phản ứng gắn *glpF* vào vector nhân dòng

Sau khi tinh chế, đoạn gen *glpF* được chèn trực tiếp vào vector pJET 1.2/blunt (Invitrogen, Hoa Kỳ) theo tỉ lệ: 1 µL vector pJET 1.2/blunt, 5 µL 2X reaction buffer, 1 µL T4 ADN ligase, 1,5 µL sản phẩm tinh sạch gen *glpF*, bổ sung thêm nuclease free water để đạt thể tích cuối cùng là 10 µL. Phản ứng nối được ủ ở nhiệt độ phòng (25°C) trong thời gian 2 giờ trước khi được sử dụng cho mục đích biến nạp.

### Biến nạp và PCR khuẩn lạc sàng lọc thể biến nạp

Toàn bộ sản phẩm gắn được trộn với 100 µL tế bào khả biến *E. coli* DH5 alpha. Hỗn hợp được giữ trên đá lạnh 15 phút, sau đó sốc nhiệt ở 42°C trong 1 phút và đặt lên đá lạnh trong 2 phút. Bổ sung 300 µL LB lỏng vào hỗn

hợp biến nạp, nuôi lắc tăng sinh ở 37°C, 160 rpm trong ít nhất 1 giờ. Cây trải 200 µL hỗn hợp biến nạp trên môi trường LB đặc có bổ sung 50 µg/mL ampicillin và nuôi qua đêm trong tủ ấm 37°C. Chọn 10 khuẩn lạc riêng rẽ trên đĩa thạch LB, hòa tan vào 40 µL nước cất khử trùng và dùng làm khuôn cho phản ứng PCR sử dụng cặp mồi của bộ kit nhân dòng (pJET fw/rv).

### Tách chiết, tinh sạch và định lượng plasmid tái tổ hợp

Khuẩn lạc chứa DNA đoạn chèn đúng kích thước được nuôi riêng rẽ ở 37°C, qua đêm trong 5 mL môi trường LB lỏng chứa kháng sinh ampicillin. Sau đó, 5 mL dịch nuôi cấy được ly tâm 12000 vòng/ phút trong thời gian 2 phút để thu cặn tế bào. DNA plasmid được tinh sạch sử dụng kit Exprep Plasmid SV (Gene All, Hàn Quốc). Sản phẩm tinh sạch được định lượng bằng đo độ hấp thụ ánh sáng tử ngoại bước sóng A<sub>260</sub>.

### Kiểm tra plasmid tái tổ hợp bằng enzyme cắt giới hạn

Vector tái tổ hợp pJET-*glpF* được ủ với 2 enzyme giới hạn là *Bgl*II và *Afl*II (NEB, Hoa Kỳ). Phản ứng cắt được thực hiện theo tỉ lệ: 1 µL buffer 10X tương ứng, 1 µL DNA plasmid đã được tinh sạch, 1 µL enzyme giới hạn và bổ sung thêm nuclease free water để đạt thể tích cuối cùng là 10 µL. Ủ mẫu ở 37°C trong thời gian 45 phút trước khi điện di kiểm tra sản phẩm cắt.

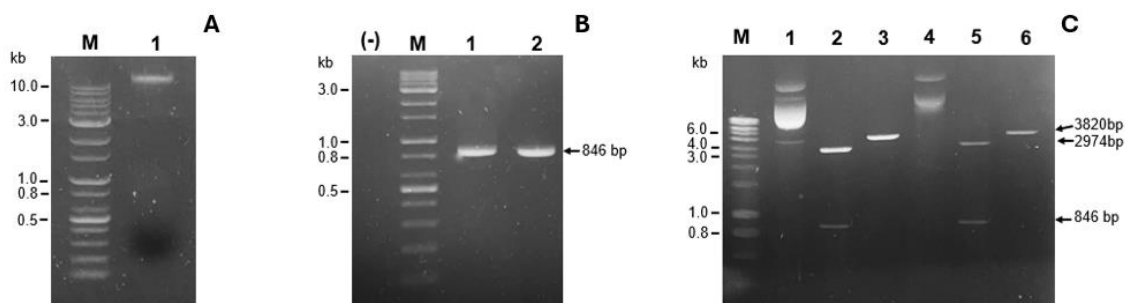
### Giải trình tự plasmid tái tổ hợp và phân tích bằng các công cụ tin sinh

Plasmid tái tổ hợp sau khi tinh sạch được dùng cho mục đích giải trình tự Sanger sử dụng cặp mồi đi kèm bộ kit nhân dòng pJET fw/rv. Giải trình tự được thực hiện tại công ty 1<sup>st</sup> BASE (Malaysia). Phân tích kết quả giải trình tự bằng phần mềm BioEdit. Công cụ SnapGene được sử dụng để đưa ra trình tự amino acid suy diễn. Sử dụng CLUSTAL OMEGA cho mục đích so sánh trình tự gen và trình tự amino acid với trình tự tham chiếu. Phân tích cấu trúc không gian ba chiều của protein được thực hiện qua phần mềm Protein Homology/analogy Recognition Engine V 2.0. Prankweb cũng được dùng để xác định trung tâm hoạt động dựa vào cấu trúc 3D của protein GlpF quan tâm. Xác định cấu trúc xoắn xuyên màng dự đoán sử dụng công cụ TMHMM 2.0. Bên cạnh đó, điểm đẳng điện (pI) và khối lượng phân tử (Mw) được tính toán bằng ExPASy pI Calculator.

## KẾT QUẢ VÀ THẢO LUẬN

### Kết quả PCR khuếch đại gen *glpF* và định lượng sản phẩm tinh sạch

Chúng tôi nghiên cứu tiến hành nhân bản đoạn gen mã hóa GlpF sử dụng cặp mồi được thiết kế đặc hiệu với khuôn là DNA tổng số của *E. coli* DH5 alpha (Hình 1A, giếng 1). Kết quả điện di cho thấy, band đích sáng rõ, kích thước 846 bp phù hợp với tính toán ban đầu (Hình 1B, giếng 1), không có band phụ, đối chứng âm không xuất hiện band. Điều này chứng tỏ cặp mồi do nhóm nghiên cứu thiết kế có độ đặc hiệu cao. Sau khi tinh chế, kết quả đo nanodrop chứng tỏ sản phẩm tinh sạch gen *glpF* có độ sạch tốt và nồng độ tinh sạch đạt yêu cầu (Bảng 1, STT 1). Mặt khác, trên gel agarose TAE 1% xuất hiện band đích sáng rõ, kích thước theo tính toán khoảng 846 bp (Image J, Hoa Kỳ) phù hợp với giả thiết, không có band phụ (Hình 1B, giếng 2). Từ đó kết luận, thiết kế thành công cặp mồi khuếch đại gen *glpF* từ genome *E. coli* DH5 alpha và đoạn gen đích sau khi tinh chế hoàn toàn phù hợp làm nguyên liệu cho các thí nghiệm tiếp theo.



Hình 1. Kết quả điện di

Giếng M: Thang chuẩn DNA (A) Sản phẩm tách DNA tổng số, giếng 1: DNA tổng số của *E. coli* chủng DH5 alpha; (B) Sản phẩm khuếch đại và tinh sạch gen *glpF*, giếng (-): đối chứng âm, giếng 1: sản phẩm PCR *glpF* từ hệ gen của *E. coli* chủng DH5-α sử dụng cặp mồi *glpF*- fw/*Afl* II- rv, giếng 2: sản phẩm tinh sạch gen *glpF*; (C) Sản phẩm cắt pJET-*glpF* bằng enzyme cắt giới hạn, giếng 1,4: Không cắt, giếng 2,5: cắt với enzyme *Bgl*II, giếng 3,6: cắt với enzyme *Afl*II.

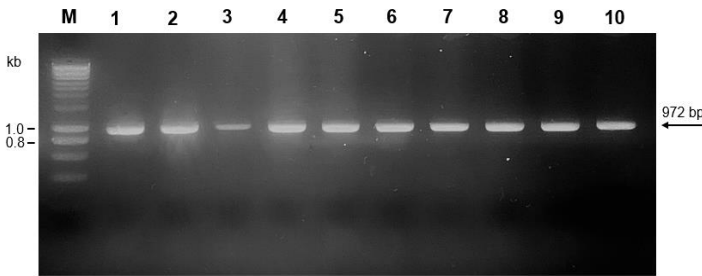
### Kết quả nhân dòng gen mã hóa GlpF bằng vector pJET 1.2/ blunt

Sản phẩm PCR có dạng đầu bằng nên gen *glpF* nên sau tinh chế có thể được nối trực tiếp vào vector nhân dòng pJET 1.2/ blunt theo tỉ lệ đã được mô tả trong phần phương pháp. Sau đó, vector tái tổ hợp được biến nạp vào tế bào khả biến *E. coli* chủng DH5 alpha. Tại vị trí đa điểm trên vector nhân dòng có chứa gen gây độc *Eco47IR* bị gián đoạn bởi đoạn chèn, trình khởi động *Plac* được sửa đổi cho phép biểu hiện gen *Eco47IR* ở mức độ vừa

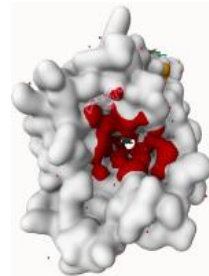
phải, đủ để chọn lọc plasmid. Vì vậy, chỉ những vi khuẩn mang DNA tái tổ hợp mới có thể mọc thành khuẩn lạc. Do đó, hoàn toàn có thể bỏ qua bước sàng lọc xanh/ trắng tốn kém. Kết quả biến nạp cho thấy, khuẩn lạc trên đĩa LB - ampicillin xuất hiện tròn, rõ, rời rạc và không có khuẩn lạc vệ tinh, phù hợp để thực hiện sàng lọc bằng PCR khuẩn lạc (Hình 2A). Tuy nhiên, *gfpF* có sẵn trong hệ gen của vi khuẩn *E. coli* DH5 alpha nên không thể sử dụng cặp mồi do chúng tôi thiết kế để kiểm tra khuẩn lạc. Mặt khác, cặp mồi pJET fw/rv bám vào hai đầu vector nhân dòng và bị gián đoạn bởi vùng MCS nên thích hợp cho mục đích kiểm tra đoạn chèn. Kết quả sàng lọc khuẩn lạc bằng phương pháp PCR cho thấy cả 10/10 khuẩn lạc được chọn đều xuất hiện band đích (972bp) (Hình 2B, giếng 1-10). Sự chênh lệch kích thước đến từ việc chúng tôi đã nghiên cứu sử dụng cặp mồi của vector nhân dòng, phù hợp với giả thiết ban đầu.

**Kết quả kiểm tra plasmid tái tổ hợp bằng enzyme cắt giới hạn**

Trước hết, 2/10 khuẩn lạc đã được chọn lọc từ phản ứng PCR được tinh sạch plasmid theo quy trình đã mô tả trong phần phương pháp. Kết quả kiểm tra hấp thụ ánh sáng tử ngoại bước sóng A<sub>260</sub> lần lượt là 210 ng/ μL và 160 ng/ μL, hoàn toàn phù hợp cho các thí nghiệm tiếp theo. Hai enzyme giới hạn được sử dụng để kiểm tra plasmid tái tổ hợp là *Bgl*II và *Afl*III với mục đích kiểm tra kích thước plasmid, vị trí và kích thước đoạn chèn. Kết quả điện di đồ sản phẩm cắt với enzyme giới hạn đã được thể hiện (Hình 1C), kích thước band đích tiếp tục được tính toán sử dụng công cụ Image J. Trong đó, enzyme cắt mở vòng *Afl*III chứa 1 điểm cắt trên vector pJET 1.2/ blunt tại vị trí 750/751 bp nên khi cắt sẽ tạo ra 1 band đích kích thước khoảng 3820 bp (Hình 1C, giếng 3,6). Mặt khác, vector nhân dòng mang 2 vị trí nhận biết enzyme *Bgl*II tại vị trí (337/338 bp và 383/384 bp) kế bên MCS nên được sử dụng để kiểm tra kích thước đoạn chèn. Do đó, khi cắt với *Bgl*II sẽ cho 2 band đích có kích thước lần lượt khoảng 2974 bp (vector) và 846 bp (đoạn chèn) (Hình 1C, giếng 2,5). Bên cạnh đó, kết quả điện di vector tái tổ hợp không cắt với enzyme xuất hiện band kích thước lớn, có thể do giải thích do plasmid tồn tại ở trạng thái siêu xoắn, một dạng lý tưởng cho mục đích giải trình tự (Hình 1C, giếng 1,4). Bước đầu kết luận gen *gfpF* đã được chèn chính xác vào vị trí MCS và được biến nạp thành công vào tế bào khả biến.



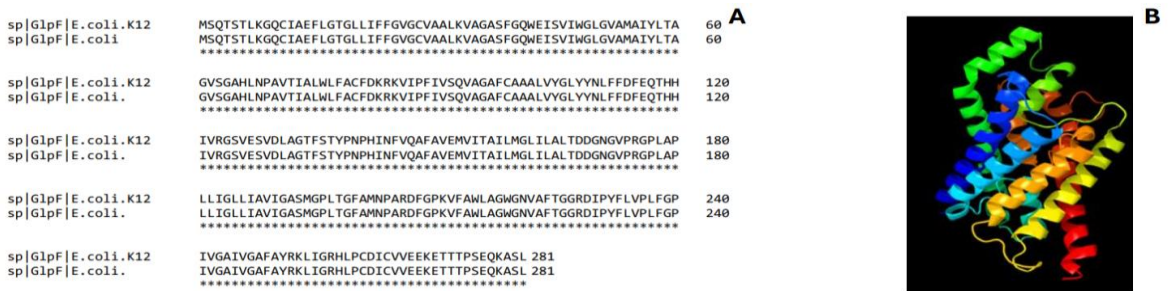
**Hình 2. Kết quả PCR khuẩn lạc sàng lọc thể biến nạp**  
M: Thang chuẩn DNA, 1-10: khuẩn lạc 1-10.



**Hình 3. Trung tâm hoạt động của GfpF**  
(Vùng màu đỏ thể hiện trung tâm hoạt động).

**Kết quả giải trình tự gen *gfpF* và phân tích sử dụng các công cụ tin sinh**

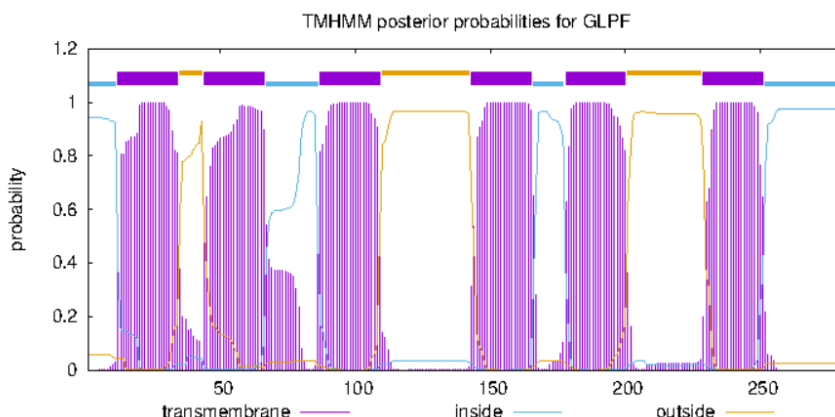
Chúng tôi tiếp tục nghiên cứu giải trình tự nucleotide để khẳng định vector tái tổ hợp pJET-*gfpF* đã mang chính xác đoạn gen quan tâm. Kết quả so sánh trình tự bằng công cụ tin sinh CLUSTAL OMEGA cho thấy gen *gfpF* có kích thước 846bp và tương đồng 100% với trình tự gen *gfpF* đã công bố trên ngân hàng gen (GenBank: AAA23886.1). Các nghiên cứu trước đó thường sử dụng chủng gốc *E. coli* K12 MG1655, *Corynebacterium glutamicum*, *Pseudomonas aeruginosa* hoặc các chủng vi khuẩn khác phục vụ cho mục đích tách nguồn gen. Mặc dù *E. coli* DH5 alpha là một dạng dẫn xuất của chủng gốc *E. coli* K12 MG1655 và gen *gfpF* có tính bảo tồn cao. Theo dự đoán, gen *gfpF* có độ tương đồng khoảng 99% về trình tự nucleotide giữa 2 chủng vi khuẩn này, sự sai khác có thể do các biến thể nhỏ hoặc các đột biến điểm trên gen. Kết quả nghiên cứu này của chúng tôi cho thấy gen *gfpF* từ *E. coli* DH5 alpha được kiểm tra trình tự trong nghiên cứu tương đồng tuyệt đối với gen này ở chủng *E. coli* K12 MG1655, chứng tỏ tính bảo tồn tiến hóa cao của gen.



**Hình 4. Kết quả so sánh trình tự amino acid và xây dựng cấu trúc không gian 3 chiều dự đoán GfpF từ *E. coli* DH5 alpha**  
(A) Kết quả so sánh trình tự amino acid (Dấu sao thể hiện trình tự tương đồng) (B) Mô hình cấu trúc không gian 3 chiều GfpF trên *E. coli* chủng DH5 alpha (Cấu trúc xoắn α được thể hiện bằng hình lát mỏng dạng cuộn).

Từ trình tự nucleotide, chúng tôi sử dụng phần mềm SnapGene để đưa ra trình tự amino acid suy diễn và tiếp tục dùng CLUSTAL OMEGA cho mục đích so sánh trình tự. Kết quả cho thấy, chiều dài protein dự đoán là 281 amino acid, tương đồng 100% với trình tự tham chiếu trên RCSB PDB (Mã: 1FX8) (Hình 4A). Điểm đẳng điện (pI) theo tính toán bằng công cụ ExPASy pI Calculator là 6,18 và khối lượng phân tử là 29,78 kDa, phù hợp với nghiên cứu đã được công bố (Weissenborn et al., 1992). Bên cạnh đó, GlpF hoạt động tốt ở môi trường từ trung tính đến kiềm thấp, pH được xác định trong khoảng 6,5 đến 7,5 và hoạt động ở nhiệt độ phát triển điển hình của vi khuẩn (37°C).

GlpF là protein màng có cấu trúc quyết định chức năng, vì vậy mô phỏng cấu trúc 3D và dự đoán chuỗi xuyên màng là cần thiết. Chúng tôi sử dụng phần mềm Protein Homology/analogy Recognition Engine V 2.0 để dự đoán cấu trúc không gian 3 chiều đặc trưng của protein GlpF (Hình 4B). Kết quả cho thấy, protein quan tâm có cấu trúc tương đồng với protein hỗ trợ vận chuyển glycerol ở *E. coli* (GlpF) với độ tin cậy 100%. Cụ thể, cấu trúc không gian 3 chiều dự đoán của protein GlpF đặc trưng với cấu trúc 6 chuỗi xoắn alpha (lần lượt từ TM1 đến TM6), đầu N và đầu C nằm trong tế bào chất. Các chuỗi xoắn này tạo thành kênh hẹp chọn lọc ở trung tâm protein. Kênh này được lót bằng các gốc axit amin phân cực và tích điện, tạo điều kiện tương tác với glycerol. Kết quả phân tích bằng TMHMM 2.0 cho thấy: 6 chuỗi xoắn alpha (TM1 đến TM6) xuyên màng dự đoán (Hình 5), phù hợp với mô tả về cấu trúc đặc trưng của protein quan tâm. Thêm vào đó, GlpF là một protein thuộc họ aquaporin nên thường không có peptide tín hiệu. GlpF được tổng hợp với các miền xuyên màng đã có sẵn, vì vậy không cần peptide tín hiệu để tích hợp vào màng. Thực tế, chưa có nghiên cứu nào trước đây báo cáo về peptide tín hiệu của GlpF vì chức năng định hướng vào màng được coi là một quá trình sinh tổng hợp thông thường. Thay vào đó, cấu trúc miền xuyên màng của protein này là đặc điểm quyết định chức năng và thường được quan tâm nghiên cứu.



**Hình 5. Kết quả dự đoán cấu trúc xuyên màng của GlpF**  
(màu tím thể hiện trạng thái xuyên màng dự đoán).

Khi sử dụng Prankweb, 40 amino acid được xác định liên quan tới trung tâm hoạt động của protein GlpF dựa vào cấu trúc không gian ba chiều. Không giống như các enzyme có vị trí “active site” cụ thể, trung tâm hoạt động của GlpF nằm ở kênh trung tâm, giúp vận chuyển và tương tác với glycerol (Hình 3). Các gốc amino acid chính được xác định có vai trò quan trọng trong việc quyết định tính đặc hiệu và tạo điều kiện thuận lợi cho quá trình vận chuyển glycerol bao gồm: Arginine (ARG 206) - tương tác với các nhóm hydroxyl của glycerol; Asparagine (ASN 68, ASN 203) - hình thành liên kết với glycerol, loại trừ phân tử lớn hơn; Serine (SER 63, SER 136) - duy trì cấu hình của kênh. Có thể nói, chức năng của GlpF nằm ở việc tăng quá trình vận chuyển glycerol qua màng tế bào thay vì xúc tác phản ứng sinh hóa với một vị trí trung tâm hoạt động như với các loại enzyme truyền thống.

## KẾT LUẬN

Nghiên cứu đã khuếch đại và nhân dòng thành công gen *glpF* trên đối tượng *E. coli* DH5 alpha. Gen *glpF* có chiều dài 846 bp mã hóa cho protein xuyên màng (GlpF) gồm 281 amino acid. Kết quả phân tích trình tự nucleotide và trình tự amino acid có độ tương đồng là 100% so với trình tự tham chiếu đã được công bố trước đó trên ngân hàng gen. Cấu trúc không gian ba chiều của protein dự đoán và các chuỗi xoắn xuyên màng đã được mô phỏng hoàn toàn phù hợp với cấu trúc đặc trưng của GlpF. Các gốc amino acid chịu trách nhiệm chính cũng đã được đề cập. Đây là những kết quả bước đầu trong việc tạo vật liệu di truyền cho các nghiên cứu biểu hiện và đánh giá vai trò của protein GlpF trong chuyển hóa glycerol ở *E. coli* cải biến di truyền.

## TÀI LIỆU THAM KHẢO

Chiang CJ, Ho YJ, Hu MC, Chao YP (2020). Rewiring of glycerol metabolism in *Escherichia coli* for effective production of recombinant proteins. *Biotechnol Biofuels*, 13, 205.

- Chilakamarry CR, Sakinah A MM, Zularisam AW, Pandey A (2021). Glycerol waste to value added products and its potential applications. *Syst Microbiol Biomanufacturing*, 1(4): 378-396.
- Gupta M, Kumar N (2012). Scope and opportunities of using glycerol as an energy source. *Renew Sustain Energy Rev*, 16(7): 4551-4556.
- Heller KB, Lin EC, Wilson TH (1980). Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J Bacteriol*, 144(1): 274-278.
- Martínez-Gómez K, Flores N, Castañeda HM, Martínez-Batallar G, Hernández-Chávez G, Ramírez OT, Gosset G, Encarnación S, Bolívar F (2012). New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microb Cell Fact*, 11(1): 46.
- Stroud RM, Nollert P, Miercke L (2003). The glycerol facilitator GlpF its aquaporin family of channels, and their selectivity. *Adv Protein Chem*, 63: 291-316.
- Sweet G, Gandor C, Voegelé R, Wittekindt N, Beuerle J, Truniger V, Lin EC, Boos W (1990). Glycerol facilitator of *Escherichia coli*: Cloning of *glpF* and identification of the *glpF* product. *J Bacteriol*, 172(1): 424-430.
- Voegelé RT, Sweet GD, Boos W (1993). Glycerol kinase of *Escherichia coli* is activated by interaction with the glycerol facilitator. *J Bacteriol*, 175(4): 1087-1094.
- Weissenborn DL, Wittekindt N, Larson TJ (1992). Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. *J Bio Chem* 267(9): 6122-6131.

## MOLECULAR CLONING AND SEQUENCE ANALYSIS OF *glpF* GENE ENCODING THE GLYCEROL FACILITATOR IN *Escherichia coli*

Pham Thi Kieu Duyen<sup>1</sup>, Nguyen Thi An Hoa<sup>1</sup>, Nguyen Ha Mi<sup>1</sup>, Quyen My Linh<sup>1</sup>,  
Nguyen Quang Huy<sup>1,2</sup>, Kieu Cam Nhung<sup>3</sup>, Luan Luong Chu<sup>2</sup>

<sup>1</sup>Faculty of Biology, VNU University of Science, Vietnam National University, Hanoi

<sup>2</sup>National Key Laboratory of Enzyme and Protein Technology, VNU University of Science, Vietnam National University, Hanoi

<sup>3</sup>Ministry of Science and Technology

### SUMMARY

Biodiesel, a renewable energy source, promises to replace conventional fossil fuels for its environmental benefits. The global expansion of the biodiesel industry has led to an excessive redundancy of crude glycerol - a major byproduct of the production process. This issue has imbalanced commercial glycerol's price and posed a significant challenge to the sustainability of the biodiesel industry. Therefore, the conversion of crude glycerol into high-value products has become an urgent issue in many countries. *Escherichia coli* (*E. coli*) has been demonstrated to utilize glycerol as a carbon source by the combined action of glycerol kinase (GlpK) and glycerol 3-phosphate dehydrogenase (GlpD). However, the expression of aquaglyceroporin gene (*glpF*) plays a crucial role in enhancing growth and biomass formation of bacteria. GlpF exhibits a conservative structure, high selectivity, and is known as the sole protein transport catalyzing glycerol diffusion across the *E. coli* inner membrane. In this study, the *glpF* gene from *E. coli* DH5 $\alpha$  was successfully cloned in the pJET 1.2/blunt vector and sequenced by the Sanger method. The decoded sequence of target gene consisted of 846 nucleotides and was 100% homologous to the reference gene sequence. Furthermore, the comparison of the deduced amino acid sequence and the analysis of the protein's three-dimensional structure also revealed 100% similarity with absolute confidence. These findings of the study would provide scientific foundation, generate genetic resources for developing recombinant microorganisms which enhance the expression of the *glpF* gene encoding a transmembrane protein that facilitates glycerol transport.

**Keywords:** *Escherichia coli*, glycerol, *glpF*, molecular cloning, sequence analysis.

\* Author for correspondence: Tel: 0985464929; Email: luancl@vnu.edu.vn