



Article

Glucoselipid Biosurfactant Biosynthesis Operon of *Rouxiella badensis* DSM 100043^T: Screening, Identification, and Heterologous Expression in *Escherichia coli*

Andre Fahriz Perdana Harahap¹, Chantal Treinen^{1,2} , Leonardo Joaquim Van Zyl³ , Wesley Trevor Williams³ , Jürgen Conrad⁴, Jens Pfannstiel⁵, Iris Klaiber⁵, Jakob Grether¹, Eric Hiller¹ , Maliheh Vahidinasab¹ , Elvio Henrique Benatto Perino¹ , Lars Lilge¹ , Anita Burger³, Marla Trindade³ and Rudolf Hausmann^{1,*}

- ¹ Department of Bioprocess Engineering (150k), Institute of Food Science and Biotechnology, University of Hohenheim, Fruwirthstr. 12, 70599 Stuttgart, Germany; andrefahrizperdana.harahap@uni-hohenheim.de (A.F.P.H.); chantal.treinen@tum.de (C.T.); jakob.grether@uni-hohenheim.de (J.G.); eric.hiller@uni-hohenheim.de (E.H.); malihe.vahidinasab@uni-hohenheim.de (M.V.); eperino@uni-hohenheim.de (E.H.B.P.); lars.lilge@uni-hohenheim.de (L.L.)
 - ² Cellular Agriculture, TUM School of Life Sciences, Technical University of Munich, Gregor-Mendel-Str. 4, 85354 Freising, Germany
 - ³ Department of Biotechnology, Institute for Microbial Biotechnology and Metagenomics (IMBM), University of the Western Cape, Cape Town 7535, South Africa; vanzylly@gmail.com (L.J.V.Z.); wesleywt@gmail.com (W.T.W.); alburger@uwc.ac.za (A.B.); ituffin@uwc.ac.za (M.T.)
 - ⁴ Department of Organic Chemistry (130b), Institute of Chemistry, University of Hohenheim, Garbenstr. 30, 70599 Stuttgart, Germany; juergen.conrad@uni-hohenheim.de
 - ⁵ Core Facility Hohenheim, Mass Spectrometry Unit, University of Hohenheim, Ottilie-Zeller-Weg 2, 70599 Stuttgart, Germany; jens.pfannstiel@uni-hohenheim.de (J.P.); iris.klaiber@uni-hohenheim.de (I.K.)
- * Correspondence: rudolf.hausmann@uni-hohenheim.de

Abstract

Rouxiella badensis DSM 100043^T had been previously proven to produce a novel glucoselipid biosurfactant which has a very low critical micelle concentration (CMC) as well as very good stability against a wide range of pH, temperature, and salinity. In this study, we performed a function-based library screening from a *R. badensis* DSM 100043^T genome library to identify responsible genes for biosynthesis of this glucoselipid. The identified open reading frames (ORFs) were cloned into several constructs in *Escherichia coli* for gene permutation analysis and the individual products were analyzed using high-performance thin-layer chromatography (HPTLC). Products of interest from positive expression strains were purified and analyzed by liquid chromatography/electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) and nuclear magnetic resonance (NMR) for further structure elucidation. Function-based screening of 5400 clones led to the identification of an operon containing three ORFs encoding acetyltransferase GlcA (ORF1), acyltransferase GlcB (ORF2), and phosphatase/HAD GlcC (ORF3). *E. coli* pCAT2, with all three ORFs, resulted in the production of identical *R. badensis* DSM 100043^T glucoselipid with Glu-C_{10:0}-C_{12:1} as the main congener. ORF2-deletion strain *E. coli* pAFP1 primarily produced glucosemonolipids, with Glu-C_{10:0,3OH} and Glu-C_{12:0} as the major congeners, predominantly esterified at the C-2 position of the glucose moiety. Furthermore, fed-batch bioreactor cultivation of *E. coli* pCAT2 using glucose as the carbon source yielded a maximum glucoselipid titer of 2.34 g/L after 25 h of fermentation, which is 55-fold higher than that produced by batch cultivation of *R. badensis* DSM 100043^T in the previous study.

Keywords: biosurfactant; biosynthesis; *Escherichia coli*; glucoselipids; glycolipids; recombinant; *Rouxiella badensis*; screening



Academic Editor: Myung-Ji Seo

Received: 17 June 2025

Revised: 11 July 2025

Accepted: 11 July 2025

Published: 15 July 2025

Citation: Harahap, A.F.P.; Treinen, C.;

Zyl, L.J.V.; Williams, W.T.; Conrad, J.;

Pfannstiel, J.; Klaiber, I.; Grether, J.;

Hiller, E.; Vahidinasab, M.; et al.

Glucoselipid Biosurfactant

Biosynthesis Operon of *Rouxiella*

badensis DSM 100043^T: Screening,

Identification, and Heterologous

Expression in *Escherichia coli*.

Microorganisms **2025**, *13*, 1664.

<https://doi.org/10.3390/microorganisms13071664>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Glycolipids are a diverse group of bioactive molecules with potential applications in the biomedical, pharmaceutical, cosmetic, and surfactant fields [1,2]. The surfactant activity of glycolipids arises from their amphiphilic nature, as they possess both hydrophilic glycosyl groups and hydrophobic lipid residues. Microbial glycolipids are naturally produced as complex mixtures of congeners or homologues, which vary in the number of glycosyl units, the degree of acylation, the number of conjugated lipid chains, their lipid lengths, as well as the extent of unsaturation and substitution [3]. Despite their application potential, most glycolipids are commercially limited presumably due to their low yield and relatively high production costs [4]. Moreover, identification and functional characterization of the genes and/or biosynthetic gene clusters responsible for the bioproduction of most glycolipids remain largely underexploited, except for well-known glycolipids such as rhamnolipids, sophorolipids, and mannosylerythritol lipids [5–7]. Identification of complete biosynthesis pathways, genetic engineering and synthetic biology could offer promising solutions to the mentioned challenges in glycolipids production.

An earlier study conducted by Kügler et al. (2015) showed that *Rouxiella badensis* DSM 100043^T, a rod-shaped non-pathogenic Gram-negative enterobacter, produced glycolipid biosurfactant during bioreactor cultivation [8]. This was indicated by the formation of excessive foam and a significant decrease in surface tension over the batch time. The species *Rouxiella badensis* was proposed by Le Flèche-Matéos et al. (2017) who performed genome shotgun sequencing and assembly of strain *R. badensis* DSM 100043^T followed by phenotypic characterization [9]. A following study by Harahap et al. (2025) elucidated the chemical structure of this biosurfactant as a novel glycolipid with glucose as the carbohydrate moiety and both hydroxylated C_{12:1} and C_{10:0} fatty acids as its lipid moieties by means of NMR and LC-ESI/MS [10]. This glucoselipid biosurfactant exhibited excellent surface-active properties, shown by a low CMC of 5.69 mg/L with minimum surface tension of 24.59 mN/m as well as good stability under extreme conditions, making it highly promising for potential industrial applications in the future. The last discoveries of novel microbial glucoselipids were reported decades ago, including an anionic glucoselipid with a tetrameric oxyacyl side chain produced by *Alcanivorax borkumensis*, and the β-D-glucopyranosyl 3-(3'-hydroxytetradecanoyloxy) decanoate (Rubiwettin RG1) produced by *Serratia rubidaea* [11,12]. Since these publications, there has been a notable absence of further research on microbial glucoselipid discoveries.

This current study aimed to identify genes which might be involved in the biosynthesis of the novel glucoselipid biosurfactant in *R. badensis* DSM 100043^T. This was accomplished by functionally screening a genome fosmid library for biosurfactant activity. Subsequent Sanger sequencing of the positive clones was performed to reveal potential open reading frames (ORFs) for glucoselipid biosynthesis and their expression in *E. coli* was assessed [13]. The recombinantly produced glucoselipids were analyzed in comparison to the glucoselipids produced by *R. badensis* DSM 100043^T using mass spectrometry and NMR. This was intended to demonstrate the successful recombinant production of glucoselipids in *E. coli* and to gain preliminary insight into its biosynthesis pathway. Finally, *E. coli* was cultivated in a fed-batch bioreactor process to evaluate its glucoselipid production.

2. Materials and Methods

2.1. Chemicals and Bacterial Strains

All analytical-grade chemicals were primarily purchased from Carl Roth GmbH & Co. KG (Karlsruhe, Germany), unless stated otherwise. The bacterial strains and plasmids used in this study are listed in Table 1, while employed primers are provided in Table S1. *R. badensis* DSM 100043^T was grown in Luria–Bertani (LB) medium (10 g/L tryptone, 5 g/L yeast

extract, and 5 g/L NaCl) at 28 °C. *E. coli* EPI300 was cultivated at 37 °C in LB medium and agar (15 g/L) supplemented with 12.5 µg/mL chloramphenicol (Merck KGaA, Darmstadt, Germany) as selection marker for fosmid pCCERI and/or supplemented with 50 µg/mL kanamycin (Merck KGaA, Germany) as selection marker for pER1.3.50.2. *P. putida* MBD1 was grown at 30 °C in LB medium or agar supplemented with 30 µg/mL apramycin (Merck KGaA, Germany) for pCCERI fosmid selection. Exconjugants of *P. putida* MBD1 were grown on M9 minimal media containing 0.2% benzoate and 30 µg/mL apramycin [14]. All *E. coli* BL21(DE3) clones containing pET21a(+) derivatives were cultivated in LB medium or agar at 37 °C supplemented with 100 µg/mL ampicillin (GERBU Biotechnik GmbH, Heidelberg, Germany) and induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (GERBU Biotechnik GmbH, Heidelberg, Germany). *E. coli* K12 JM109 was grown at 37 °C in LB medium supplemented with 100 µg/mL ampicillin.

Table 1. List of strains and plasmids used in this study.

| Name | Genotype | Reference |
|---|--|---|
| Strains | | |
| <i>Rouxiella badensis</i> DSM 100043 ^T | Wild-type | [8] |
| <i>Escherichia coli</i> BL21(DE3) | <i>E. coli</i> BL21 F ⁻ <i>ompT hsdS</i> (rB – mB –) <i>dcm</i> + <i>Tetr gal λ</i> (DE3) <i>endA Hte</i> [<i>argU proL Camr</i>] [<i>argU ileY leuW Strep/Specr</i>] | Agilent Technologies (Santa Clara, CA, USA) |
| EPI300 | F ⁻ <i>mcrA Δ</i> (<i>mrr-hsdRMS-mcrBC</i>) <i>φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ</i> (<i>ara, leu</i>)7697 <i>galU galK λ⁻ rpsL nupG trfA tonA dhfr</i> | Epicentre (Illumina, San Diego, CA, USA) |
| K12 JM109 | | [15] |
| <i>Pseudomonas putida</i> MBD1 | <i>P. putida</i> KT2440 derivative; Kan ^r ; <i>φC31 attB site</i> ⁺ | [14] |
| 54F3 | <i>P. putida</i> MBD1 containing 54F3 fosmid for expression of lyso-ornithine lipid biosurfactant | [16] |
| Plasmids | | |
| pCCERI | pCC1FOS derivative, Cm ^r , Ap ^r , <i>φC31 integrase, attP</i> | [17] |
| pER1.3.50.2 | pRK2013 derivative with <i>trfA</i> gene deleted, Kan ^r | [17] |
| Fosmid 1.8.H6 | pCCFOS1 clone containing 2,399,010 bp to 2,429,311 bp of <i>R. badensis</i> (CP114060) | This study |
| pET21a(+) | Expression vector with a C-terminal His-tag, Amp ^r , <i>ori</i> , T7 promoter and terminator, MCS | Novagen (Merck KGaA, Darmstadt, Germany) |
| pCAT2 | pET21a(+) derivative; ORF 1; ORF2; ORF3 | This study |
| pCAT1 | pET21a(+) derivative; ORF 1; ORF2 | This study |
| pAFP1 | pET21a(+) derivative; ORF 1; ORF3 | This study |
| pAFP2 | pET21a(+) derivative; ORF 2; ORF3 | This study |
| pAFP3 | pET21a(+) derivative; ORF3 | This study |
| pAFP4 | pET21a(+) derivative; ORF1 | This study |
| pAFP5 | pET21a(+) derivative; ORF 2 | This study |

2.2. Genomic Library Screening

2.2.1. Preparation of Genomic DNA

Genomic DNA of *R. badensis* DSM 100043^T was extracted from the cell pellets obtained from 100 mL of 17 h fermentation cultures following protocol described by Wang et al. (1996) with modifications where a final concentration of 0.1 mg/mL proteinase K was added into the lysis buffer and pellets were incubated at 37 °C for 3 h [18]. An amount of 2 µL RNase A (50 U/mg) was added into the mixture after 1% SDS treatment and then incubated at 70 °C for 1 h. For extraction, an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added, mixed, and centrifuged at 13,000× g and 4 °C for 10 min. The supernatant

was collected and the step was repeated. An equal volume of chloroform:isoamyl alcohol (24:1) was added to the supernatant and centrifuged under the same conditions, again repeating the step. The final supernatant was mixed with 1/10 volume of 3 M sodium acetate and ice-cold 60% isopropanol, then stored at $-20\text{ }^{\circ}\text{C}$ overnight for precipitation. After centrifugation at $13,800\times g$ for 10 min, precipitated DNA was air-dried and dissolved in elution buffer (QIAprep[®] Spin Miniprep Kit, Hilden, Germany).

Genomic DNA Was further purified according to the protocol outlined by Liles et al. (2008) [19]. DNA was size-selected using a contour-clamped homogenous electric field (CHEF) gel apparatus (Bio-Rad CHEF-DR[®] III, Hercules, CA, USA) by electrophoresing the DNA into a low melting point agarose gel (1% *w/v*) using the manufacturer's recommended settings for separating fragments between 1 kb and 100 kb. Unstained DNA was excised following staining of the edge of the agarose gel with EtBr and the DNA recovered from the gel slice using agarase treatment (NEB, Ipswich, MA, USA). The recovered DNA was precipitated following standard protocols, washed using ice-cold 70% ethanol, air-dried, and suspended in an appropriate amount of TE buffer. DNA concentration was determined using a Qubit[™] 2.0 fluorometer (ThermoFisher Scientific Inc., Waltham, MA, USA). End-repairing of the DNA was performed using the CopyControl[™] Fosmid Library Production Kit (Epicentre[®], San Diego, CA, USA) following the manufacturer's protocol. End-repaired DNA was finally extracted twice to remove impurities [19].

2.2.2. Creation of Genomic Library

The genomic library was generated by using Copy Control[™] Fosmid Library Production Kit (Epicentre[®], San Diego, CA, USA) following the manufacturer's protocol. In this study, fosmid pCCERI was used instead of using the kit's supplied vector. The fosmid was linearized using restriction enzyme BstZ171 (NEB, Ipswich, MA, USA) followed by dephosphorylation using 1 U shrimp alkaline phosphatase, rSAP (NEB, Ipswich, MA, USA), following the manufacturer's protocols. The dephosphorylated fosmid was finally extracted twice to remove the impurities. The prepared genomic DNA was ligated with the pCCERI fosmid vector using T4 DNA ligase (NEB, Ipswich, MA, USA) following the manufacturer's instructions. Library packaging was performed using MaxPlax[™] Lambda Packaging Extract (Epicentre[®], San Diego, CA, USA). The phage-packed library was transfected into *E. coli* EPI300 containing conjugative helper plasmid pER1.3.50.2 to allow conjugation in other hosts. All molecular biology techniques including restriction endonuclease digestion, ligation, phosphorylation, and gel electrophoresis were performed according to standard protocols.

2.2.3. High-Throughput Conjugation and Screening

The phage-infected host cells were spread on Q-trays filled with 300 mL LB agar supplemented with 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol and 50 $\mu\text{g}/\text{mL}$ kanamycin. The colonies were picked automatically using a Genetix Qpix 2 XT robotic colony picker (Molecular Devices, Queensway, UK) and transferred into 96-well microtiter plates containing 200 μL fresh LB medium supplemented with 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol and 50 $\mu\text{g}/\text{mL}$ kanamycin. The microtiter plates were incubated at $37\text{ }^{\circ}\text{C}$ and 250 rpm for 17 h. The clones were stamped from the 96-well microtiter plates on Q-Trays containing 300 mL LB agar supplemented with 0.2% L-arabinose, 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol, and 50 $\mu\text{g}/\text{mL}$ kanamycin. An overnight culture of *P. putida* MBD1 was prepared in LB medium supplemented with 50 $\mu\text{g}/\text{mL}$ kanamycin and incubated at $30\text{ }^{\circ}\text{C}$ for 17 h. The overnight culture was diluted 25-fold in LB medium containing 10 mM MgCl_2 and incubated for 2 h at $30\text{ }^{\circ}\text{C}$ and 125 rpm. To inhibit restriction enzymes, the culture was incubated for 10 min at $42\text{ }^{\circ}\text{C}$. After that, 2 mL of the 2 h culture was added to 40 mL LB medium containing 10 mM MgCl_2 and

0.02% L-arabinose. Then, 100 μ L of the *P. putida* MBD1 culture was mixed with 10 μ L of the donor *E. coli* EPI 300 p.ER.1.3.50.2 culture containing the genomic library into a 96-well microtiter plate. The plates were incubated overnight at 30 °C without shaking and afterwards stamped on Q-Trays with M9-benzoate-apramycin agar which is selective for *P. putida* MBD1 [16]. The plates were incubated overnight at 30 °C and were afterwards transferred from the M9-benzoate-apramycin agar to fresh Q-trays containing LB agar supplemented with 30 μ g/mL apramycin. The screening was performed for both *E. coli* EPI 300 p.ER.1.3.50.2 and *P. putida* MBD1 plates. Clones were dotted on agar plates and cultured overnight at 30 °C. A mist of paraffin was sprayed on the colonies using an airbrush and biosurfactant activity was detected by the formation of halos around the biosurfactant producing colonies [16].

2.2.4. Bioinformatic Analysis of the Pathway Gene Products

Positive clones showing halos were analyzed via Sanger sequencing using primer pairs, pCCERI-FVD1—pCCERI-RVS1 and pCCERI-FVD2—pCCERI-RVS2 (Table S1), designed for the pCCERI fosmid. Sequencings were performed using ABI Prism 377 automated DNA sequencer (Central Analytical Facility, University of Stellenbosch, Stellenbosch, South Africa). The resulting sequences were aligned to the *R. badensis* DSM 100043^T genome (GCA_002093665) using the CLC genomics workbench v25.0 (Qiagen, Hilden, Germany) to reveal the open reading frames (ORFs) associated with the fosmid inserts.

AlphaFold (ColabFold; <https://github.com/sokrypton/ColabFold>; accessed on 24 May 2025) was used to predict the structure for the full length sequence of all characterized C1-phosphatases from the Huang et al. (2015) dataset and used together with the bona fide structure of the *Francisella tularensis* (3KD3) as input to Foldtree to construct a phylogeny tree [20,21]. The tree was then annotated using TVBOT [22]. Cagedat was used to compare the gene cluster with those on the Genbank database while clinker and TBtools-II v.2.155 were used to visualize the synteny and amino acid conservation between gene clusters [23,24].

2.2.5. Cloning of ORFs into the pET21a(+) System

Initially, primer pairs (Table S1) were designed to amplify two different fragments including removal of stop codon: fragment 1 included *glcA* and *glcB*, whereas fragment 2 included additional *glcC*. The removal of stop codon was meant to add a His-tag to the C-terminal of the protein, thus allowing the reading frame to extend and include the His-tag followed by subsequent stop codon and T7 terminator. Polymerase chain reaction (PCR) was performed using fosmid 1.8 H6 as template by Phusion[®] High-Fidelity DNA Polymerase (NEB, Ipswich, MA, USA) following the manufacturer's protocol. Both fragments were purified from the agarose gel using 0.5 U/ μ L agarase (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's protocol with slight changes. The amount of 100 mg molten agarose was digested with 1 U of agarase and the fragments were precipitated with sodium acetate. Centrifugation was performed at 11,000 *g* and 4 °C. The precipitated and washed fragments were dissolved in 20 μ L nuclease-free water. An initial cloning step to transfer the fragments into pJET1.2/blunt (CloneJET PCR Cloning Kit; ThermoFisher Scientific Inc., Waltham, MA, USA) was performed to facilitate cloning into pET21a. Following confirmation of cloning of the amplified fragments into pJET1.2/blunt, the fragments were liberated from these plasmids using NdeI and XhoI. Vector pET21a(+) was also linearized with NdeI and XhoI and then purified from agarose gel using NucleoSpin[®] Gel and PCR Clean-up (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany) following the manufacturer's protocol. Purified pET21a(+) was then ligated to fragments 1 or 2 and transformed into *E. coli* BL21(DE3). Constructs were confirmed

through restriction digestion and the fidelity of the amplified fragments established by Sanger sequencing by Eurofins Genomics (Ebersberg, Germany). Positive clones carrying pET21a(+) containing fragment 1 (glcAB) and fragment 2 (glcABC) will be referred to as *E. coli* pCAT1 and *E. coli* pCAT2, respectively.

2.3. Gene Permutation and Expression Analysis

The goal of gene permutations is typically to understand the variability and potential outcomes that can arise from different genetic combinations involving rearrangement or recombination of genes in various ways to study their effects, interactions, and glucoselipid biosynthesis. For this, five more constructs were created carrying different combinations of ORFs. For this, all primers (Table S1) and plasmid constructs were designed and analyzed using SnapGene software (version 7.1.1, GSL Biotech, San Diego, CA, USA). PCRs were carried out in a PCR thermal cycler peqSTAR XS (VWR™, Darmstadt, Germany) using Q5® High-Fidelity DNA polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany) according to standard protocols. PCRs for constructing pAFP4 and pAFP5 were conducted using PrimeSTAR® Max DNA Polymerase (Takara Bio Inc., Shiga, Japan) as exceptions. The Monarch® DNA Gel Extraction Kit and Monarch® PCR & DNA Cleanup Kit (New England Biolabs GmbH, Frankfurt am Main, Germany) were then used for plasmid fragments purification according to the manufacturer's protocol. The respective DNA fragments were joined via Gibson Assembly using the Gibson Assembly® Cloning Kit (New England Biolabs GmbH, Frankfurt am Main, Germany) [25]. The resulting constructs were transformed into chemically competent *E. coli* BL21(DE3) (New England Biolabs GmbH, Frankfurt am Main, Germany) following the manufacturer's protocol. The fidelity of all plasmid constructs was ensured by whole plasmid sequencing (Eurofins Genomics Germany GmbH). The plasmids were extracted using innuPREP Plasmid Mini Kit 2.0 by IST Innuscreen GmbH (Berlin, Germany) according to the manufacturer's instruction.

To evaluate the permutated gene(s) expression, shake-flask cultivations of all *E. coli* strains were performed in modified mineral salt media (MSM) supplemented with 10 g/L glucose and incubated at 37 °C and 160 rpm [10]. Induction with IPTG to a final concentration of 0.1 mM was performed when the OD₆₀₀ reached 1.0. Samples were collected after 6 h of cultivation and subjected to an emulsification assay and oil displacement test using olive oil [26,27]. The respective samples were extracted and qualitatively analyzed by HPTLC using thin-layer chromatography (TLC) silica plate 60 RP-18 F_{254S} (Merck KGaA, Germany) as well as p-anisaldehyde solution as staining agent [10].

2.4. Fed-Batch Bioreactor Cultivation

Fed-batch bioreactor cultivations were performed only with *E. coli* pCAT2 and pAFP1. The first preculture was carried out in a baffled shake flask with ampicillin-supplemented LB medium using an incubator shaker (New Brunswick™ Innova 44®R Eppendorf AG, Hamburg, Germany) at 37 °C and 120 rpm for 12 h. The second preculture was performed in chemically defined MSM as described by Riesenberger et al. (1991) with minor modification where additional (NH₄)₂SO₄ with final concentration of 5 g/L was added instead of thiamin HCl [28]. The second preculture, which was supplemented with 25 g/L glucose, was inoculated with the first preculture to reach an initial OD₆₀₀ of 0.2 and then cultured under the same conditions as the first preculture. The bioreactor cultivations were conducted using a 30 L stainless steel fermenter (ZETA GmbH, Lieboch, Austria) initially loaded with 10 L modified Riesenberger's medium for the batch phase. The batch medium was supplemented with 25 g/L glucose as a sole carbon source and ampicillin to prevent contamination. The bioreactor was then inoculated with previously prepared second preculture to reach an initial OD₆₀₀ of 0.2 and run at 37 °C, pH 7.0, and initial stirrer speed of 300 rpm. The pH

was controlled through addition of either 4 M H_3PO_4 or 20% (v/v) NH_3 solutions. The aeration rate was set to 2 L/min at the beginning and the minimum pO_2 level was always kept at 30% by adjusting the stirrer speed up to 900 rpm and the aeration up to 22 L/min. The end of the batch phase could be identified by typical second rise online measurement of pO_2 level as well as offline measurement of depleting glucose concentration [29]. For the fed-batch phase, 5 L glucose (50% w/v) feed solution containing 19.7 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ and 0.1 mM IPTG was pumped exponentially into the bioreactor with the initial feeding rate F_0 calculated based on the formula described by Hiller et al. (2024) [30]. The desired growth rate μ of 0.2 h^{-1} and maintenance coefficient m of $0.05 \text{ g}/(\text{g} \cdot \text{h})$ were considered into the formula. Antifoam (Struktol[®] SB590, Hamburg, Germany) was used by manual injection into the bioreactor to avoid excessive foam build-up.

Samples were collected at 2 h intervals from the bioreactor's sampling port and then OD_{600} was measured using a spectrophotometer (Biochrom WPA08000, Biochrom Ltd., Cambridge, UK). Samples were then centrifuged at $4816 \times g$ and $4 \text{ }^\circ\text{C}$ for 10 min (Multifuge X3R, Thermo Fisher Scientific, USA) to separate cells from the supernatant. The cell-free supernatant was used for glucose quantification using an enzymatic assay kit (R-Biopharm AG, Darmstadt, Germany) following the manufacturer's protocol. Prior to glucoselipid quantification, cell-free supernatant was extracted twice with ethyl acetate and then HPTLC-based (CAMAG AG, Muttenz, Switzerland) glucoselipid measurement supported with WinCATS Software 1.4.7 was performed according to Harahap et al. (2025) [10]. Each sample was applied in the length of 6 mm band on TLC silica plate 60 RP-18 F_{254S} (Merck KGaA, Germany) and developed with isopropyl acetate/methanol/acetic acid (100:10:1, $v/v/v$). The TLC plate was then derivatized using diphenylamine-aniline-phosphoric acid (DPA) solution and scanned at 620 nm for glucoselipid measurements. Purified glucoselipid by *R. badensis* DSM 100043^T was used as standard for this quantitative measurement. For cell dry weight (CDW) calculation, 40 mL fermentation broth was collected in triplicate and centrifuged as described previously to obtain the cell pellets. The pellets were washed with saline and dried overnight in an oven at $110 \text{ }^\circ\text{C}$. CDW was measured on scale and the mean correlation factor of 3.95 was applied.

2.5. Structure Elucidation of Produced Glucoselipids

2.5.1. Glucoselipids Extraction and Purification

Prior to glucoselipid extraction, the bioreactor culture was first centrifuged as described previously and the cell-free supernatant was extracted twice with ethyl acetate according to Harahap et al. (2025) without addition of H_3PO_4 [10]. The organic phase was evaporated using a rotary vacuum evaporator (R-215, Büchi Labortechnik AG, Flawil, Switzerland) at 100 mbar and $40 \text{ }^\circ\text{C}$ to obtain the crude extract. Purification of glucoselipid produced by *E. coli* pCAT2 was performed in the same manner as that of glucoselipid produced by *R. badensis* DSM 100043^T as described previously by Harahap et al. (2025) using medium-pressure liquid chromatography (MPLC; SepacoreX50, Büchi, Flawil, Switzerland) and a reverse-phase C18 column (FlashPure EcoFlex, Büchi, Flawil, Switzerland) with the flowrate of the mobile phase set to 7.5 mL/min [10]. For purification of glucoselipid produced by *E. coli* pAFP1, minor modification on the gradient of acetonitrile (ACN)/water as mobile phases was carried out: 110 min 0–100% ACN and 10 min 100–100% ACN. The eluent was collected in glass tubes, each of which is referred to as a fraction. A total amount of 90 fractions were collected during MPLC purification process and pooled fractions containing glucoselipid, seen on the derivatized HPTLC plate, were then subjected to solvent evaporation using a rotary vacuum evaporator at 10 mbar and $40 \text{ }^\circ\text{C}$. The dried, purified glucoselipid samples were stored at $-20 \text{ }^\circ\text{C}$ for subsequent structural characterization using NMR spectroscopy and LC-ESI-MS/MS.

2.5.2. Nuclear Magnetic Resonance (NMR) Spectroscopy

The glycolipid samples were dissolved in 600 μL methanol- d_4 and transferred to a standard 5 mm NMR tube. 1D and 2D NMR-spectra were recorded on an Avance HD III 600 MHz spectrometer, equipped with a 5 mm BBO Prodigy cryo-probe (Bruker, Billerica, MA, USA). ^1H and ^{13}C chemical shifts were referenced to the residual solvent signal at $\delta_{\text{H/C}}$ 3.35 ppm/49.0 ppm. ^1H , ^{13}C , Heteronuclear Single-Quantum Coherence (HSQC), Heteronuclear Multiple Bond Correlation (HMBC), Correlation Spectroscopy (COSY), Total Correlation Spectroscopy (TOCSY), HSQCTOCSY and selective 1D-TOCSY spectra were recorded using standard Bruker pulse sequences at 298 K. The recorded NMR spectra were processed with Topspin 4.2.0 (copyright 2022, Bruker Biospin, Billerica, MA, USA) and SpinWorks 4.2.10 (Copyright 2019, K. Marat, University of Manitoba, CA, USA).

2.5.3. Mass Spectrometry Characterization

The LC-ESI/MS analysis of the glycolipid was performed on a 1290 UHPLC system (Agilent, Waldbronn, Germany) coupled to a Q-Exactive Plus Orbitrap mass spectrometer equipped with a heated electrospray ionization source (HESI, Thermo Fisher Scientific, Bremen, Germany) as previously described by Harahap et al. (2025) with modifications [10]. Glycolipids separations were achieved by an ACQUITY CSH C18 column (1.7 μm , 2.1 $\mu\text{m} \times 150$ mm, Waters, Eschborn, Germany). Gradient elution was carried out at a constant flow rate of 0.3 mL/min, with specific conditions for the glucoselipid samples from both *E. coli* pCAT2 and pAFP1 detailed in Table 2. The HESI source was operated in the positive and negative ion modes with a spray voltage of 4.0 kV in the positive ion mode and 3.5 kV in the negative ion mode. The ion transfer capillary temperature was set to 350 $^\circ\text{C}$ and the sweep gas and auxiliary pressure rates were set to 35 and 10, respectively. The S-lens RF level was set to 50%. The Q-Exactive Plus mass spectrometer was calibrated externally in the positive and negative ion modes using the manufacturer's calibration solutions (Pierce, Thermo Fisher Scientific, Bremen, Germany). Mass spectra were acquired at a resolution of 70,000 at m/z 200 using an Automatic Gain Control (AGC) target of 3.0×10^6 and a maximum ion injection time of 100 ms. Data-dependent MS/MS spectra in the mass range of 200 to 2000 m/z were generated for the five most abundant precursor ions with a resolution of 17,500 at m/z 200 using an AGC target of 1.0×10^6 and 100 ms maximum ion injection time. Xcalibur software version 4.3.73.11 and Compound Discoverer Software version 3.3 (both Thermo Fisher Scientific, San Jose, CA, USA) were used for data acquisition and data analysis. Identification and assignment of the glucoselipids were based on the precise m/z value of the precursor ions and manual inspection of the corresponding MS/MS spectra.

Table 2. LC-ESI/MS settings for analysis of purified glucoselipid samples.

| Parameters | <i>E. coli</i> pCAT2 Glucoselipids | <i>E. coli</i> pAFP1 Glucoselipids |
|--------------------|---|---|
| Column temperature | 40 $^\circ\text{C}$ | 50 $^\circ\text{C}$ |
| Injection volume | 0.37 μL | 3.0 μL |
| Mobile phase | A: 0.2% formic acid in water B: 0.2% formic acid in acetonitrile | A: 0.2% formic acid in water B: 0.2% formic acid in methanol |
| Gradient elution | 50–53% B from 0 to 15 min, 53–67% B from 15 to 17 min, 67–75% B from 17 to 25 min, 75–90% B from 25 to 30 min, 90–90% B (isocratic) from 30 to 35 min, 90–50% B from 35 to 36 min, 50–50% B from 36 to 37 min | 45–53% B from 0 to 5 min, 53–59% B from 5 to 10 min, 59–90% B from 10 to 20 min, 90–90% B (isocratic) from 20 to 25 min, 90–45% B from 25 to 26 min |

Table 2. Cont.

| Parameters | <i>E. coli</i> pCAT2 Glucoselipids | <i>E. coli</i> pAFP1 Glucoselipids |
|------------|------------------------------------|------------------------------------|
| Scan range | 200–1500 <i>m/z</i> | 100–1400 <i>m/z</i> |
| (N)CE * | 32 | 10 |

* Normalized collision energy.

3. Results

3.1. Identification of Genes Responsible for Glucoselipid Biosynthesis and Delineation of Their Functional Roles

To identify the genes responsible for glucoselipid synthesis in *R. badensis* DSM 100043^T, a genome fosmid library was constructed for screening in *P. putida*. Following conjugation to *P. putida*, a total of 5400 clones were screened resulting in four positive clones (Figure S1). The terminal ends of the fosmid insert sequences were mapped to the *R. badensis* DSM 100043^T genome (GCA_002093665) to delineate the genomic fragments captured in the respective clones which resulted in biosurfactant activity. This fragment represents the sequence from ~65,000 bp until ~101,000 bp on contig NZ_MRWE0100009.1. These clones revealed a shared operon consisting of three genes encoding ORFs designated as *glcA* (ORF1), *glcB* (ORF2), and *glcC* (ORF3) (Figure S2). The first ORF, encoding GlcA (WP_017492724), belongs to N-acetyltransferase (RimL; cI34333)/N-acyltransferase superfamilies (cI17182), specifically the GCN5-related N-acetyltransferases (GNAT) family (Table 3). The second ORF, encoding GlcB (WP_009635452), belongs to the 1-acyl-sn-glycerol-3-phosphate acyltransferase (Phospholipid synthase; PlsC; cI43057) and lysophospholipid acyltransferase (LPLAT; cI17185) superfamilies. The third ORF, encoding GlcC (WP_017492722), belongs to the haloacid dehalogenase-like hydrolase (HAD-like) and phosphoserine phosphatase (PSP) superfamilies (SerB; cI21460), specifically the C-1 type of HAD phosphatases. No closely related phosphatases appear to have been characterized, and the closest solved three-dimensional structure is that of a PSP of unknown function from *Francisella tularensis* (Q5NH99; 3KD3).

Table 3. The top two structural similarity hits using Foldseek for GlcA, GlcB and GlcC against AFDB50 and PDB100.

| Protein Description | Organism | Prob. | E Value | Seq. Ident % | Accession |
|--|------------------------------------|-------|------------------------|--------------|---------------------------|
| GlcA | | | | | |
| N-acetyltransferase domain-containing protein | <i>Pseudomonas fluorescens</i> | 1.00 | 8.92×10^{-23} | 40.4 | AF-A0A5E6RK03-F1-model_v4 |
| Crystal structure of a GNAT superfamily PA3944 acetyltransferase in complex with CoA | <i>Pseudomonas aeruginosa</i> PAO1 | 1.00 | 6.33×10^{-12} | 22 | 6EDD |
| GlcB | | | | | |
| Uncharacterized protein | <i>Rouxiella badensis</i> | 1.00 | 1.90×10^{-53} | 100 | AF-A0A1X0WHA5-F1-model_v4 |
| Crystal structure of the 1-acyl-sn-glycerophosphate (LPA) acyltransferase, PlsC | <i>Thermotoga maritima</i> MSB8 | 1.00 | 9.26×10^{-6} | 14.7 | 5KYM |

Table 3. Cont.

| Protein Description | Organism | Prob. | E Value | Seq. Ident % | Accession |
|--|---------------------------------------|-------|------------------------|--------------|-----------------------|
| GlcC | | | | | |
| Uncharacterized protein | <i>Serratia</i> sp. M24T3 | 1.00 | 1.18×10^{-39} | 77.8 | AF-I0QXG7-F1-model_v4 |
| Crystal structure of a phosphoserine phosphohydrolase-like protein | <i>Francisella tularensis</i> SCHU S4 | 1.00 | 7.70×10^{-13} | 23.4 | 3KD3 |

Although relatives of each individual ORF from the operon occur in disparate bacterial genomes, the conservation of the three-gene operon appears to be unique to certain species in the genera *Rouxiella* (25 genomes), *Pseudomonas* (46301 genomes), *Vibrio* (31462 genomes) and *Ottowia* (133 genomes) (Figure S3). The pathways in *Rouxiella* sp. are clearly unique, sharing low amino acid similarity and clustering away from those in other bacteria that contain a similar three-gene operon (Figure S4). Conservation of the three-gene operon in selected genomes suggests limited selective pressure to maintain the complete operon and perhaps the resultant product, whereas the presence of homologs of the individual genes in other genomes could suggest that related functions play a role in synthesis or modification of lipids produced by these organisms.

Analysis of the genomes of the six representative species of *Rouxiella* classified by the current Genome Taxonomy Database reveals that the pathway seems limited to *R. badensis*, *R. chamberiensis* and the unclassified *Rouxiella* sp. WC2420 (Figure 1). The genomic context within which these three genes find themselves is quite different when comparing the region between bacteria that contain the pathway (Figure S3). Expression of these genes in rhizosphere-associated bacteria as well as co-expression with genes potentially involved in pathogenesis (Phospholipase C, Type I secretion system) and plant growth promotion (auxin efflux carrier) suggests that this operon has been adapted for several different uses in these bacteria involved with host evasion/invasion or symbiosis (Figure S4). Biosurfactants are well-known to play a role in bacterial pathogenesis as well as plant growth promotion and in line with this, *R. badensis* was recently described as an emerging onion pathogen [31,32].

To delineate which genes are necessary for glucoselipid biosynthesis, we designed a series of constructs expressing different gene combinations of the pathway to assay glucoselipid production following expression in *E. coli* (Figure 2A). Initially, the growth behavior and product analyses among all expression strains were compared to *E. coli* BL21 (DE3) carrying an unmodified vector as control. All strains, including the control, exhibited similar growth behavior in shake flask cultivation, with the exception for *E. coli* pCAT1 (Figure S8). All strains, including the control, entered the stationary phase after 8 h of cultivation, reaching an average OD₆₀₀ of approximately 10. In contrast, *E. coli* pCAT1 only achieved an OD₆₀₀ of approximately 6. The reduced growth rate observed in *E. coli* pCAT1 indicates signs of stress, suggesting that the strain experienced a metabolic burden. It is possible that the combination of N-acetyltransferase (*glcA*) and acyltransferase (*glcB*) expression led to cellular stress due to the accumulation of new toxic intermediates or products. Additionally, the metabolites produced (e.g., free fatty acid) may have integrated into the cell membrane, potentially triggering membrane-associated stress responses [33,34]. Using emulsification and oil displacement assays, it was observed that the expression of the encoding genes in *E. coli* pCAT2 resulted in a positive phenotype (Figure 2B). Specifically, *E. coli* pCAT2 exhibited the highest values among all tested expression strains, with an emulsification unit of 138.9 EU/mL and an oil displacement diameter of 2.95 cm. In comparison, *E. coli* pAFP1 showed approximately half of these values, 68.8 EU/mL for emulsification assay and 1.45 cm for oil displacement, yet still higher than the other expression strains and the control. Since these two parameters are

indirect indicators of biosurfactant activity, a comparison with the HPTLC results in Figure 2C suggests that *E. coli* pAFP1 produced biosurfactant compounds at levels approximately half of those observed in *E. coli* pCAT2 [26,27]. Based on the p-anisaldehyde-stained RP18 TLC plate in Figure 2C, expression of all three genes (pCAT2) resulted in the production of a compound indicated by a prominent band at R_f 0.79, which matched the R_f value of the glucoselipid reference produced by *R. badensis* DSM 100043^T. Furthermore, expression of *glcA* and *glcC* (pAFP1) led to the production of several compounds, indicated by streaky bands between R_f 0.5 and 0.6, whereas expression of *glcB* and *glcC* (pAFP2) did not lead to the production of any glycolipids. Expression of both *glcA* and *glcB* (pCAT1) did not result in glycolipids synthesis, suggesting an absolute requirement for the phosphatase. Expressions of individual genes also showed no glycolipids synthesis.

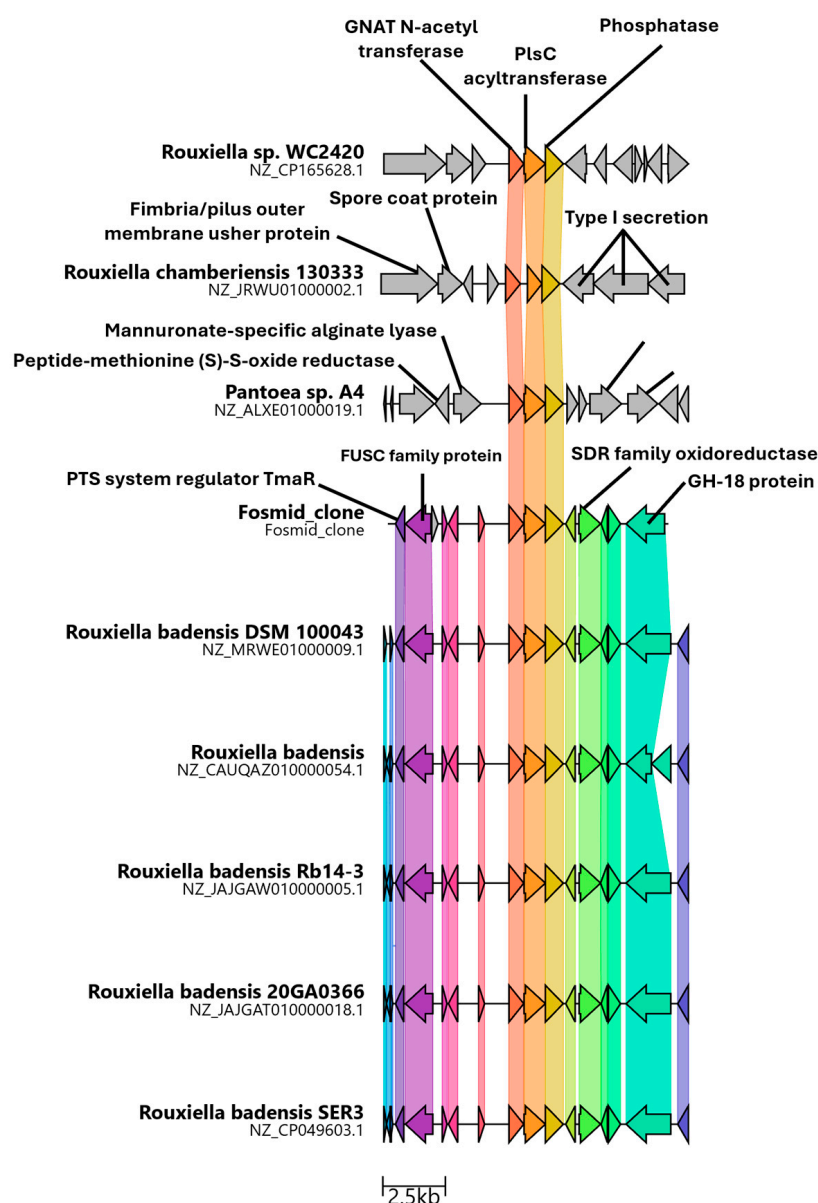


Figure 1. Comparison of the genomic regions encoding the glucoselipid ORFs found in *Rouxiella* sp. and closely related pathways.

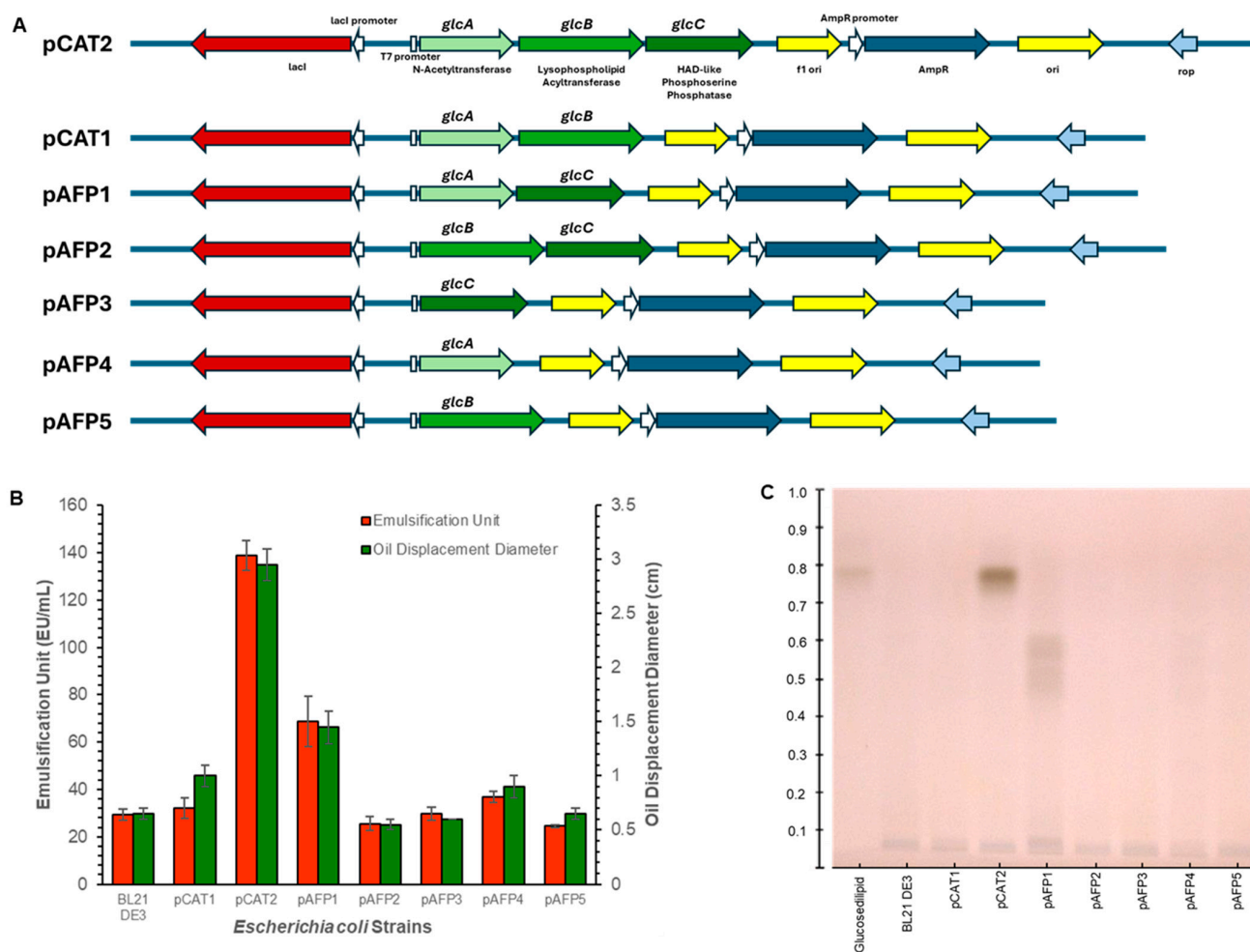


Figure 2. Effect of gene permutation on different expression strains' phenotypic characteristics: Schematic diagram of different plasmid constructs carrying different ORF(s) combinations, where ORF 1 is *glcA*, ORF2 is *glcB*, and ORF3 is *glcC* (A); Emulsification unit and oil displacement diameter of supernatant from all expression strains (B); and p-anisaldehyde-stained RP18 TLC plate of crude extracts showing prominent bands at R_f 0.79 and R_f 0.5–0.6 for *E. coli* pCAT2 and pAFP1, respectively (C). Please note in subfigure (C) that the glucosedi lipid on the first lane was obtained from *R. badensis* DSM 100043^T while all other lanes were sample from *E. coli* expression strains.

The source of the 3-hydroxy fatty acid donors for transfer to a glucose molecule resulting in the glycolipids are likely to come from the pool of acyl-carrier proteins, as intermediates of fatty acid synthesis, or as intermediates of fatty acid degradation in the form of long-chain acyl-CoA derivatives [35]. GlcA is clearly related to the GNAT family of acetyltransferases. The closest structural match to GlcA is 7KPS, an acetyltransferase from *Pseudomonas aeruginosa* with demonstrated ability to catalyze transfer of an acetyl group to polymyxin antibiotics (Figure S5A) [36]. Although the backbone is highly conserved, the acyl acceptor side of the active site geometry and surface charge differ substantially between GlcA and 7KPS (Figure S5B). It does display a greater overall positive charge as seen in 4KUA, known to perform O-acetylation of chloramphenicol. Although this enzyme family usually performs transfer of acetyl groups to ϵ -amino groups or primary amines on a large variety of biomolecules, rare examples have been described that either catalyze O-acetylation of hydroxyl groups [37–39]. Notably, a GNAT-related acetyltransferase found in *Lyso bacter enzymogenes* was found capable of O-acetylation of chloramphenicol using isobutyryl-CoA and isovaleryl-CoA as substrates [40]. Additionally, GNAT-related enzymes have been found to catalyze fatty acyl transfer of 3-hydroxy fatty acids to amine

groups on small molecules as in the case of phaeornamide and fatty acids as in the case of *Mycobacterium tuberculosis* Rv1347c [41,42]. All these references suggest that these enzymes can display substantial substrate flexibility; however, to our knowledge, no GNAT-related enzyme has been described performing O-acylation of hydroxyl groups with a fatty acid, likely required for synthesis of glucoselipids. The alignment of GlcA with several characterized GNAT-acetyltransferases, including examples of enzymes reported to perform O-acetylation, did not yield any information that could support the substrate selection for the enzyme, due to sequence divergence.

GlcB shows greatest structural similarity to 5KYM, a 1-acyl-sn-glycerol-3-phosphate (LPA) acyltransferase from *Thermotoga maritima* (Figure S6A) [43]. This enzyme class is characterized as having a narrow substrate range; however, some are known to catalyze acyl transfer to substrates other than LPA, such as ornithine lipid synthase A or sulfur-containing aminolipid synthase from *Ruegeria pomeroyi* [44–46]. Emerging evidence suggests a general role for these enzymes as responsible for the addition of a second fatty acyl chain to a variety of substrates. Comparison of the active site cleft geometry with that of 5KYM would suggest that GlcB is capable of accepting much bulkier substrates (Figure S6B). A hydrophobic two-helix motif present in both enzymes likely leads to it being membrane associated which would agree with a role for the biosurfactant as lipid membrane augments or if it's to be exported [43].

Several HAD-like sugar phosphatases capable of dephosphorylating sugar-6-phosphates have been described from *Pseudomonas fluorescens* (PFLU2693), *Thermophilus volcanium* (Q978Y6), *Bacteroides thetaiotaomicron* (Q8A2F3), *Eubacterium rectale* (D0VWU2), *Geobacillus kaustophilus* (Q5L139), *Saccharomyces cerevisiae* dog1 and dog2 (P38774, P38773), several enzymes from *E. coli* as well as an exceptional effort to capture the substrate preference for representative members for a significant portion of the superfamily [20,47–50]. This dataset includes the substrate range for the *F. tularensis* enzyme mentioned above which displays weak activity on glucose-6-phosphate (G6P) but has demonstrable activity on a wide range of phosphorylated substrates such as phosphoserine, phosphorylated sugar alcohols (mannitol and sorbitol), nucleotide monophosphate, glycerol phosphate, dihydroxyacetone phosphate and more. Substrate specificity prediction for phosphatases is notoriously difficult owing the wide variation in the structure of the “cap” domain likely responsible for substrate specificity [20]. To determine if GlcC could potentially use G6P as substrate we compared its putative structure to that of the characterized HAD C1-phosphatases. GlcC clustered with closely related enzymes from *Rouxiiella* sp., *Serratia* sp. and *Francisella* sp. (Figure S7). Although C1-phosphatases displaying activity on G6P from the Huang dataset are dominated by examples from the Pseudomonadota, these enzymes are not limited to this phylum and there appears to be little structural conservation that would enable prediction of activity on G6P as enzymes that do, and do not display activity on this substrate are evenly distributed throughout the tree.

The source of glucose that forms the scaffold for attachment of 3-hydroxy fatty acids is not immediately obvious. Any one of several activated glucose derivatives involved in glycolysis or cell wall synthesis (e.g., glucose-6-phosphate, glucose-1-phosphate, glucosamine-6-phosphate, N-acetylglucosamine-6-phosphate, UDP-glucose or UDP-N-acetylglucosamine) could be the source for the glucose moiety in the glucoselipids. However, complex nucleotide (e.g., UDP, TDP, or GDP)-activated sugars are generally less favorable as glucose donors for recent glucoselipids biosynthesis. Typically, in the case of glycolipids biosynthesis, only glycosyltransferases can catalyze the transfer of a carbohydrate moiety from these activated sugar forms to a lipid backbone such as rhamnosyltransferase RhIB in rhamnolipid biosynthesis, UDP-glucosyltransferase UGTA1 in

sophorolipid biosynthesis, and mannosyltransferase EMT1P in mannosylerythritol lipids (MELs) biosynthesis [51–53].

3.2. Structural Characterization of Produced Glucoselipids

A total of 384 mg and 263 mg of crude extracts were obtained after performing glucoselipids extraction on 500 mL of cell-free supernatant from the bioreactor cultivation of *E. coli* pCAT2 and pAFP1, respectively. The crude extracts were then subjected to individual purification step using MPLC and each fraction was visualized using HPTLC and subsequent staining. Fractions (46–51) containing glucosedilipids from *E. coli* pCAT2 began to elute at approximately 65% acetonitrile gradient while fractions (37–40) containing glucosemonolipids from *E. coli* pAFP1 eluted at approximately 40% acetonitrile gradient, suggesting more hydrophilic nature of glucosemonolipids. A final amount of 69 mg purified glucosedilipids in the form of white amorphous substance and 35 mg purified glucosemonolipids in the form of pale-yellow amorphous substance were obtained at the end of purification step. To gain a clearer understanding of the chemical structure of the synthesized glucoselipids, the purification procedures were repeated. The first purification was performed to collect and analyze specific fractions using LC-ESI-MS/MS, in order to study the variation in congeners and calculate their relative amounts. The process was repeated to isolate a targeted congener in a sufficient quantity for detailed structural analysis using NMR. For quality assurance regarding sample purity, fraction 48 (from *E. coli* pCAT2) was dried and subjected to NMR analysis, while fraction 37 through 40 (from *E. coli* pAFP1) were individually dried and analyzed using NMR.

Evaluation of the 1D and 2D NMR spectra of fraction 48 and comparison with the ^1H and ^{13}C NMR data of glucosedilipid from *R. badensis* DSM 100043^T unambiguously established its identical chemical structure (Figure 3 and Figure S9) [10]. A distinct peak at 4.65 ppm in fraction 48 (*E. coli* pCAT2), which is entirely absent in glucosedilipid sample from *R. badensis* DSM 100043^T, is attributed to traces of water present in the sample. While analysis of 1D and 2D NMR spectra including COSY, TOCSY, selTOCSY, HSQCTOCSY, HSQC and HMBC of fraction 37 from *E. coli* pAFP1 revealed a mixture of three pairs of glucosemonolipid of α - and β -anomeric glucopyranosyl moieties substituted with one fatty acid each (Figure 4 and Figures S10–S15). The structure of the latter was unambiguously established as 3-hydroxydecanoic acid ($\text{C}_{10:0}$) by evaluation of the 2D NMR spectra. The acylation positions of the two major glucosemonolipid pairs were determined by HMBC of the respective sugar proton with the carboxyl C of the fatty acid. Thus, a HMBC correlation between 2- $\text{H}\alpha/\beta$ at δ 4.64/4.72 ppm and carboxyl carbon C-1' at δ 173.14/172.70 ppm indicated the C-2 monosubstituted glucopyranose as major compound 1 in fraction 37 whereas a HMBC correlation between 3- $\text{H}\alpha/\beta$ at δ 5.28/4.98 ppm and carboxyl carbon C-1' at δ 173.38/173.63 ppm showed the C-3 monosubstituted glucopyranose as second compound 2 in fraction 37. The C-6 acylation side of the third monosubstituted α - and β -glucopyranosyl pair 3 was only present in trace amounts and tentatively deduced by, e.g., the low-field-shifted methylene 6- $\text{H}_{\alpha/\beta}$ (β -anomer) at δ 4.46 and 4.23 ppm compared to methylene 6- $\text{H}_{\alpha/\beta}$ (β -anomer) at δ 3.91 and 3.71 ppm of compound 1 (Table S2). ^1H signal integration yielded a 59:28:13 ratio (1:2:3) of the three glucosemonolipid Glu- $\text{C}_{10:0}$ pairs 1–3. Although three additional MPLC fractions of *E. coli* pAFP1 were isolated (fractions 38–40), their purity was insufficient for full structural elucidation using NMR due to significant spectral overlaps. Therefore, MS-guided structural analysis was then performed to characterize all purified fractions instead.

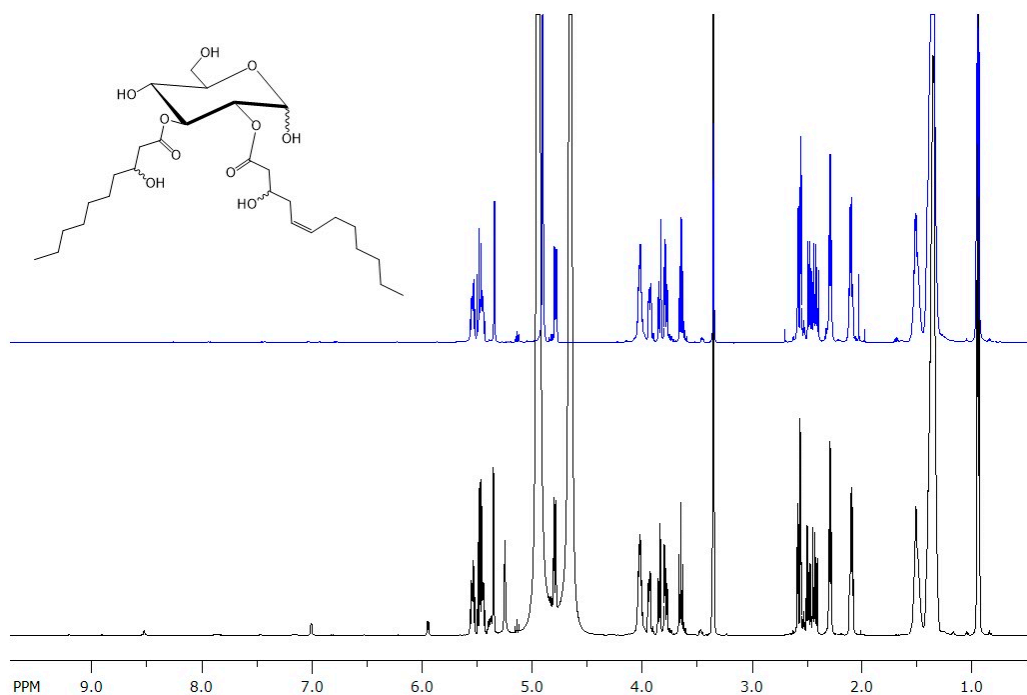


Figure 3. Comparison of ^1H NMR spectra of freshly dissolved glucosedi-lipid from *R. badensis* DSM 100043^T (blue) and *E. coli* pCAT2 (black) confirming the identical structure of both products as Glu-C_{10:0}-C_{12:1} [10]. Some impurities of unknown structure were detected between δ 5.80 and 8.80 ppm in the spectrum of glucosedi-lipid from fraction 48 (*E. coli* pCAT2).

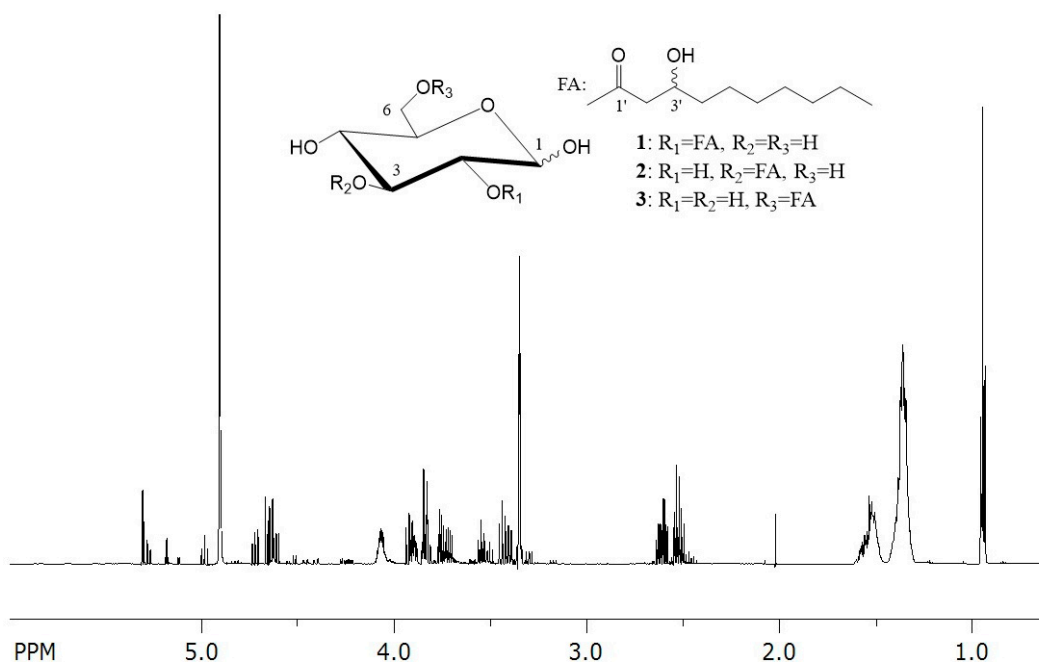


Figure 4. ^1H NMR spectrum of fraction 37 and the chemical structure of glucosemonolipids Glu-C_{10:0} 1–3 identified in fraction 37.

To further investigate the structure and congener composition of the produced glucoselipids, purified products from both strains were thoroughly examined using LC-ESI-MS/MS. The base peak chromatogram in the negative ion mode exhibited a prominent peak at RT 13.26 min for *E. coli* pCAT2's purified product (pooled fraction 46–51) as shown in Figure 5A. The mass spectrum at corresponding retention time showed a deprotonated molecular ion $[\text{M}-\text{H}]^-$ with a m/z of 545.3335 and formic acid adduct $[\text{M}+\text{FA}-\text{H}]^-$ with

a m/z of 591.3393. These findings enabled the determination of the molecular formula $C_{28}H_{49}O_{10}$ (error: 0.59 ppm). Furthermore, the mass spectrum exhibited two in-source fragment ions, which were likewise observed as prominent signals in the negative-ion-mode MS/MS spectrum of m/z 545.3335 (Figure 5B). These fragment ions corresponded to the fatty acid components 3-hydroxydecanoic acid ($C_{10:0}$) and 3-hydroxy-5-dodecenoic acid ($C_{12:1}$) with m/z of 187.1333 and 213.1492, respectively. Therefore, the compound was designated as glucosedi lipid Glu- $C_{10:0}$ - $C_{12:1}$ informed by similarity with MS data of *R. badensis* DSM 100043^T glucosedi lipid previously reported [10]. The acylation positions of $C_{10:0}$ and $C_{12:1}$ at C-3 and the C-2 position of the glucopyranosyl moiety were also unambiguously determined from ¹³C HMBC spectra of NMR results. Therefore, the structure of glucosedi lipid Glu- $C_{10:0}$ - $C_{12:1}$ elucidated by comprehensive NMR analysis was fully confirmed by the mass spectrometry results obtained from high-resolution LC-ESI-MS/MS. This findings suggest that the glucosedi lipid operon was successfully expressed heterologously by *E. coli* pCAT2.

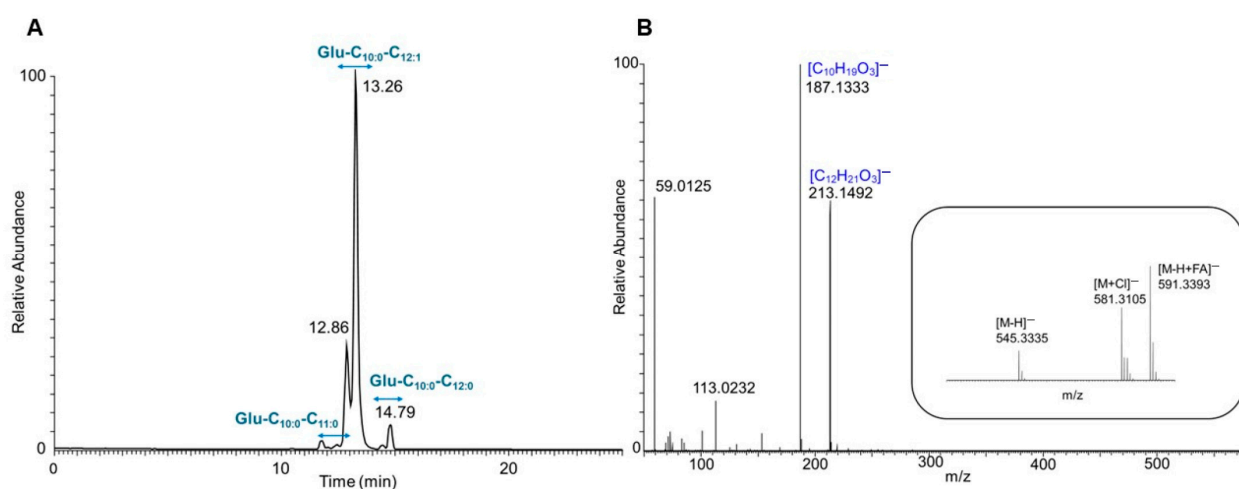


Figure 5. LC-ESI-MS/MS analysis of the purified glucosedi lipids produced by *E. coli* pCAT2 (pooled fraction 46–51). The total ion chromatogram (TIC) in the negative ion mode (A) shows diverse glucosedi lipid congeners including their isomers. The mass spectrum (B, inset) in the negative ion mode at retention time of 13.26 min shows Glu- $C_{10:0}$ - $C_{12:1}$ as the most abundant congener. The MS/MS spectrum of the protonated molecular ion m/z 545.3335 (B) exhibits a high degree of similarity to the MS/MS spectra of *R. badensis* DSM 100043^T glucosedi lipid [10].

Moreover, Figure 5A and Table 4 demonstrate that *E. coli* pCAT2 predominantly produced Glu- $C_{10:0}$ - $C_{12:1}$ (91.04%) while the minor glucosedi lipid congeners Glu- $C_{10:0}$ - $C_{11:0}$ (4.54%) and Glu- $C_{10:0}$ - $C_{12:0}$ (4.42%) were present in smaller amounts. The relative abundance of each congener was calculated based on the corresponding peak areas obtained from the extracted ion chromatogram (XIC) of the respective congener (Figure S16). Following a thorough analysis of the minor congeners, the mass spectrum at RT 14.79 min exhibited a deprotonated ion $[M-H]^-$ at m/z 547.3492, indicative of a molecular formula of $C_{28}H_{51}O_{10}$, as depicted in Figure S17. The MS/MS spectrum obtained from the fragmentation of the parent ion $[M-H]^-$ with m/z 547.3492 revealed two intense signals consistent with fatty acids, with m/z 187.1336 ($C_{10:0}$) and 215.1649, respectively. The m/z value of 215.1649 can be calculated as the difference of 2 (the addition of two hydrogen atoms) from the $C_{12:1}$ fatty acid. This results in the designation of 3-hydroxy-5-lauric acid ($C_{12:0}$). Subsequent analysis of the congener compound yielded the following structure: Glu- $C_{10:0}$ - $C_{12:0}$. An additional minor congener was detected based on the mass spectrum at RT 12.80 min where it exhibited a deprotonated ion $[M-H]^-$ with m/z 533.3331 and a predicted molecular formula of $C_{27}H_{49}O_{10}$ (error: 1.29 mmu) (Figure S18). MS/MS data from the

parent ion also exhibited two intense typical fatty acid signals with m/z 187.1333 ($C_{10:0}$) and 201.1491. The m/z value of 201.1491 can be explained as $\Delta m/z$ of -14 (reduction of one carbon chain $-CH_2-$) to the $C_{12:0}$ fatty acid. Accordingly, the fatty acid is designated as 3-hydroxy-5-undecanoic acid ($C_{11:0}$) and the congener compound was deduced as Glu- $C_{10:0}$ - $C_{11:0}$. Notably, all glucosedi lipid congeners produced by *E. coli* pCAT2 contained $C_{10:0}$ fatty acid paired with a second distinct fatty acid. In addition, the XICs from all glucosedi lipid congeners revealed the presence of both minor and major isomers in each congener (Figure S16). A comparison of the 1D selective TOCSYs NMR with those of the XICs of each congener indicated that the major and minor isomers were glucosedi lipids with α - and β -anomeric glucopyranosyl moieties, respectively [54].

Table 4. Summary of the glucoseli pids congeners produced by strains *E. coli* pCAT2 and *E. coli* pAFP1 as identified by means of LC-ESI-MS/MS analysis.

| Strain | Glucoseli pid | Molecular Formula | RT * [min] | m/z [M-H] ⁻ | Fatty Acid's m/z [M-H] ⁻ | Relative Abundance (%) | |
|----------------------|------------------------------|----------------------|--------------|------------------------------|---------------------------------------|------------------------------|-------|
| <i>E. coli</i> pCAT2 | Glu- $C_{10:0}$ - $C_{11:0}$ | $C_{27}H_{50}O_{10}$ | 12.3 | 533.333 | $C_{11}H_{21}O_3^-$ (201) | $C_{10}H_{19}O_3^-$ (187) | 4.54 |
| | | | 12.8 | | | | |
| | Glu- $C_{10:0}$ - $C_{12:1}$ | $C_{28}H_{50}O_{10}$ | 12.8 13.2 | 545.333 | $C_{12}H_{21}O_3^-$ (213) | $C_{10}H_{19}O_3^-$ (187) | 91.04 |
| <i>E. coli</i> pAFP1 | Glu- $C_{10:0}$ | $C_{16}H_{30}O_8$ | 6.3 | 349.186 | $C_{10}H_{19}O_3^-$ (187) | - | 33.95 |
| | | | 6.7 | | | | |
| | | | 7.1 | | | | |
| Glu- $C_{11:0}$ | $C_{17}H_{32}O_8$ | 11.0 | 363.202 | $C_{11}H_{21}O_3^-$ (201) | - | 3.63 | |
| | | 11.8 | | | | | |
| Glu- $C_{12:0}$ | $C_{18}H_{34}O_8$ | 18.0 | 377.218 | $C_{12}H_{23}O_3^-$ (215) | - | 4.07 | |
| | | 18.4 | | | | | |
| Glu- $C_{12:1}$ | $C_{18}H_{32}O_8$ | 12.3 | 375.202 | $C_{12}H_{21}O_3^-$ (213) | - | 41.44 | |
| | | 13.1 | | | | | |
| Glu- $C_{14:0}$ | $C_{20}H_{38}O_8$ | 23.0 | 405.249 | $C_{14}H_{27}O_3^-$ (243) | - | 16.91 | |

* Retention times are provided for each identified isomer.

LC-ESI-MS/MS analysis of *E. coli* pAFP1's purified product (pooled fraction 37–40) demonstrates that it synthesizes glucosemonolipid congeners, with Glu- $C_{10:0}$ (33.95%) and Glu- $C_{12:1}$ (41.44%) being the two most prevalent congeners, which comprise the primary components of the glucosedi lipid Glu- $C_{10:0}$ - $C_{12:1}$ (Figure 6A). As demonstrated in Table 4, the analysis identified additional minor congeners, including Glu- $C_{12:0}$ (4.07%) and Glu- $C_{11:0}$ (3.63%), which have also been observed to contribute to the formation of minor glucosedi lipid congeners. Interestingly, an additional congener Glu- $C_{14:0}$ which had not been previously detected was also identified with a relatively high abundance of 16.91%. The relative abundance of each congener was calculated based on the corresponding peak areas obtained from the XIC of the respective glucosemonolipid congener (Figure S19). As illustrated in Figure 6A, the base peak chromatogram in the negative ion mode demonstrates that the glucosemonolipid Glu- $C_{10:0}$ elutes within the RT interval of approximately 6.3–7.1 min as three isomeric forms. This finding is consistent with the NMR results, which revealed the presence of glucosemonolipid Glu- C_{10} in three distinct compounds differing in their acylation positions. The mass spectrum at RT 6.34 min exhibited the deprotonated molecular ion [M-H]⁻ with m/z 349.1869 and a predicted molecular formula of $C_{16}H_{29}O_8$ (calcd: 349.1868; error: 0.03 ppm) (Figure 6B). The elimination of the

glucosyl moiety (162 Da) results in the production of a remnant of m/z 187.1332, which was also identified in the mass spectrum as a prominent signal, commonly associated with 3-hydroxydecanoic acid ($C_{10:0}$). The mass spectra (ESI full MS and MS²) of the other glucosemonolipid congeners produced by *E. coli* pAFP1, along with their corresponding fragmentation patterns, are presented in Figures S20–S23.

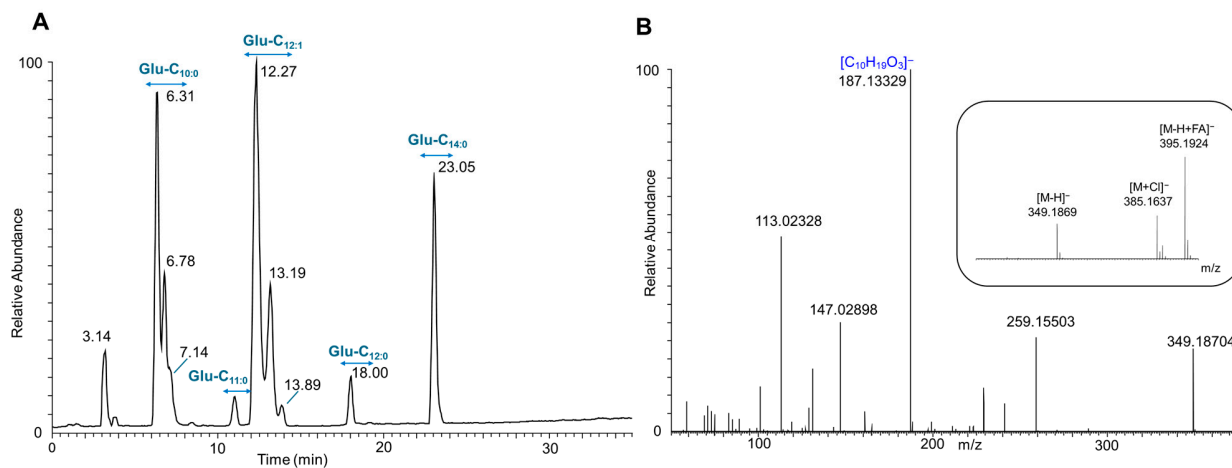


Figure 6. LC-ESI-MS/MS analysis of the purified glucosemonolipids produced by *E. coli* pAFP1 (pooled fraction 37–40). The total ion chromatogram (TIC) in the negative ion mode (A) shows diverse glucosemonolipid congeners including their isomers. The mass spectrum (B, inset) in the negative ion mode at retention time of 6.31 min exhibited a deprotonated molecular ion with a m/z 349.1869. The MS/MS spectrum of the deprotonated molecular ion m/z 349.1869 (B) revealed fragmentation patterns of Glu- $C_{10:0}$ which is one of the most abundant glucosemonolipid congeners. This observation is consistent with the results obtained by nuclear magnetic resonance (NMR) analysis.

Notably, one of the minor glucoselipid congeners produced by the *E. coli* pCAT2 and pAFP1 contained an odd-chain hydroxylated $C_{11:0}$ fatty acid. This finding is of particular interest, as *E. coli* typically synthesizes even-chain fatty acids, and the occurrence of an odd-chain fatty acids is rare [55]. It is possible that this observation can be explained by several factors, including metabolic alterations resulting from the heterologous gene expression or the formation of artifacts during the downstream processing and sample preparation steps.

In view of the previously outlined structure of glucosedilipid and the subsequent validation in the present study, a salient question emerges: by what mechanism does the activity of these three identified genes culminate in the biosynthesis of glucosedilipid [10]? The expression of distinct gene combinations has revealed that the phosphatase GlcC is indispensable for the synthesis of compounds, indicating that, within the context of *Escherichia coli*, no other gene can adequately substitute for this function. Alternatively, if such a substitution were possible, the resulting compound production would be below our detection limit. This finding further substantiates the sequential mechanism of the acetyltransferase GlcA, which functions by initially catalyzing acylation at the C-2 hydroxyl group of the glucose scaffold with a 3-hydroxy fatty acid. Subsequent to this initial acylation, GlcA facilitates the addition of another 3-hydroxy fatty acid to the hydroxyl group at the C-3 position by GlcB, thereby culminating in the completion of glucosedilipid synthesis. This finding indicates that the acyltransferase GlcB recognizes the glucosemonolipids as a substrate and is unable to catalyze direct acylation of the activated glucose scaffold. Nevertheless, the elucidation of the complete biosynthetic pathway of the recently identified glucosedilipid would be a particularly intriguing direction for future research.

3.3. Fed-Batch Bioreactor Cultivation of *E. coli* pCAT2 for Glucosedilipid Production

Fed-batch cultivations of *E. coli* pCAT2 were carried out in a pilot-scale bioreactor system to compare recombinant glucosedilipid production performance with the previously described bioreactor cultivation using the wild-type producer *R. badensis* DSM 100043^T [10]. As can be seen in Figure 7, initial glucose concentration of 25 g/L was completely consumed after 10 h cultivation during the batch phase yielding an average CDW of 8.10 ± 0.25 g/L and relatively low glucosedilipid concentration of 0.21 ± 0.03 g/L. Subsequently, the feeding phase was started with the exponential pumping of feed solution and IPTG injection, maintaining a fixed growth rate of 0.2 h^{-1} . The feeding phase lasted for 15 h, thus giving a total cultivation time of 25 h.

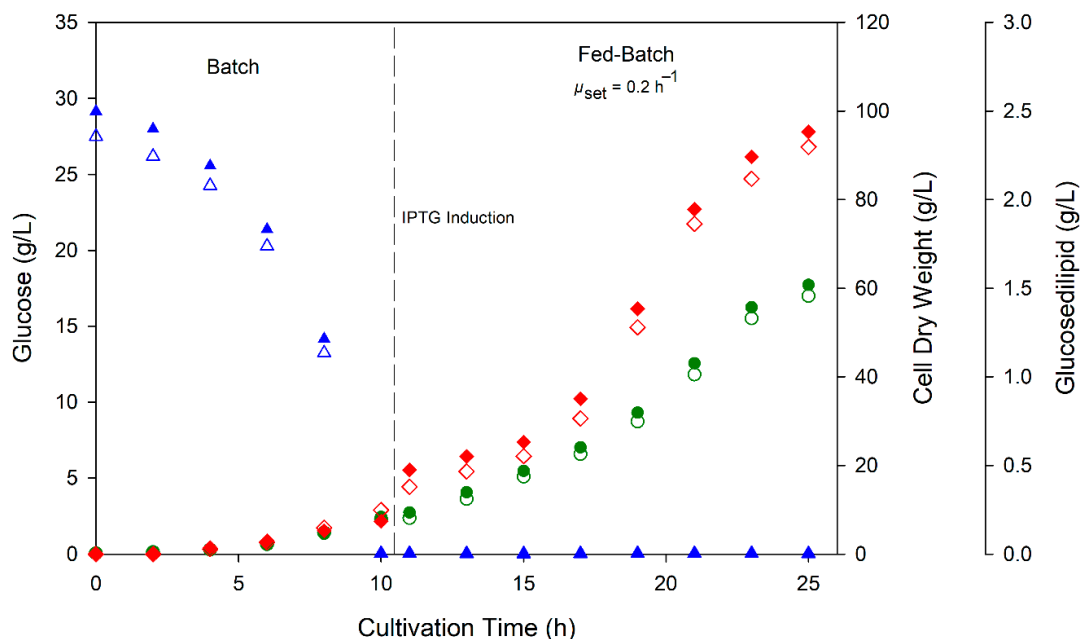


Figure 7. Time course of fed-batch bioreactor cultivation for the recombinant production of glucosedilipid by *E. coli* pCAT2 showing CDW (green circle), glucose concentration (blue triangle) and glucosedilipid concentration (red diamond) from biological duplicate experiments.

With fed-batch process utilizing glucose as the sole carbon source, high cell density was achieved with maximum cell dry weight of 59.5 ± 1.3 g/L. There was no glucose accumulation detected during the feeding phase at a set growth rate of 0.2 h^{-1} . As can be seen in Figure 7, glucosedilipid production followed the trend of biomass formation where 2.34 ± 0.04 g/L glucosedilipid concentration was reached at the end of cultivation. The glucosedilipid titer achieved in this study is 55-fold higher than that produced by the wild-type producer *R. badensis* DSM 100043^T in batch bioreactor cultivation as previously described by Harahap et al., 2025 [10]. The stable product accumulation across the cultivation suggests that the glucosedilipid is not rapidly consumed or broken down by *E. coli*. This stability enhances further downstream processing predictability and simplifies product recovery.

4. Conclusions

Through the function-based screening of the *R. badensis* DSM 100043^T genome library, we identified an operon containing three genes responsible for the biosynthesis of glucosedilipid confirmed by mass spectrometry and NMR-based structure elucidation. The genes were identified to encode N-acetyltransferase GlcA (ORF1), acyltransferase GlcB (ORF2), and phosphatase/HAD GlcC (ORF3). Heterologous expression of this operon

(pCAT2) in *E. coli* resulted in the production of identical *R. badensis* DSM 100043^T glucosediacyl lipid Glu-C_{10:0}-C_{12:1} as its major congener. In contrast, deletion of acyltransferase-encoding gene *glcB* (ORF2) in *E. coli* pAFP1 led to the production of glucosemonolipid, particularly Glu-C_{10:0} and Glu-C_{12:0} as its primary congeners, with a predominant acylation site at the C-2 position of the glucose moiety. This finding suggests that GlcA functions sequentially by first catalyzing the acylation, preferably at the C-2 hydroxyl group of the glucose scaffold, using a 3-hydroxy fatty acid as the acyl donor. These findings also indicate that the acyltransferase GlcB likely catalyzed the second acylation of 3-hydroxy fatty acid molecule to the glucose moiety, thereby providing a basis for subsequent enzyme characterization studies. Furthermore, the fed-batch high cell density fermentation of *E. coli* pCAT2 utilizing glucose as the sole carbon source resulted in 2.34 g/L glucosediacyl lipid titer after 25 h fermentation, which was 55-fold greater than the previously reported titer in the batch bioreactor fermentation of *R. badensis* DSM 100043^T using glycerol as the carbon source. The results presented lay the foundation for future studies aimed at elucidating the biosynthesis pathway and enabling the production of glucoselipids for biotechnological applications as novel biosurfactants.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms13071664/s1>, Table S1: Primers used in this study.; Table S2: ¹H and ¹³C NMR chemical shifts, spin multiplicity, *J* coupling constants of Compound 1-3 of glucosemonolipid of Fraction 37 (*E. coli* pAFP1) in methanol-*d*₄ at 600 (¹H) and 150 MHz (¹³C).; Figure S1: Halo formation of the clones 1.15 G9 and 1.8 H6 shown from four positive clones.; Figure S2: Open reading frames of sequenced *R. badensis* region. ORFs of interest are marked in green: N-acetyltransferase, lysophospholipid acyltransferase, phosphoserine phosphatase/ HAD.; Figure S3: Extended comparison of genomic regions encoding glucoselipid pathway related genes among different phyla.; Figure S4: Comparison of amino acid similarity among glucoselipid synthesis pathway ORFs in bacteria that have retained all three ORFs. The cladogram represents hierarchical clustering where the similarities were calculated using Euclidean distance as the measure of dissimilarity. Clustering was performed using the complete linkage method.; Figure S5: A) Cartoon representation of the structural alignment of GlcA (green) with 7KPS (cyan) with RMSD of 0.944. The structure of Coenzyme-A is shown as stick model in the binding pocket based on the 7KPS structure. B) Comparison of the surface geometry and charge distribution on the acyl acceptor side of GlcA in comparison with 7KPS and 4KUA, known to catalyze O-acetylation of chloramphenicol.; Figure S6: A) Cartoon representation of the structural alignment of GlcB (cyan) with 5KYM (purple) with RMSD of 4.7. B) Comparison of the surface geometry and charge distribution of the catalytic cleft of GlcB in comparison with 5KYM known to catalyze acylation of 1 acyl-glycerol-3-phosphate.; Figure S7: Structure-based phylogenetic relationship between all C1-phosphatase members that have been assayed for substrate range and select members from other phosphatase families (Q819K1, 1K7H, dog1 and dog2). The outer ring and legend indicate the position and color scheme assigned to different phyla. The inner ring indicates those proteins that display activity on G6P (green squares), those with no activity on G6P (brown squares) and the phosphatase in the pathway under investigation (red square). The data is based on the publication by Huang et al. 2015 [20]; Figure S8: Growth curve of shake flask culture of various recombinant *E. coli* constructs containing different permuted ORFs with *E. coli* BL21 DE3 empty vector as control strain. IPTG inductions were done at *t* = 1 h. All plots represent the mean values obtained from duplicate experiments.; Figure S9: Comparison of ¹³C NMR spectra of the *E. coli* pCAT2's glucosediacyl lipid (black) and glucosediacyl lipid from *R. badensis* DSM 100043^T (blue) in methanol-*d*₄ at 150 MHz confirming identical structure. Additional small NMR signals in the upper arise from the α/β anomeric equilibrium in solution.; Figure S10: ¹³C NMR spectrum of glucosemonolipid Glu-C_{10:0} (fraction 37) from *E. coli* pAFP1 in methanol-*d*₄ at 150 MHz.; Figure S11: Expansion of ¹H NMR spectrum of glucosemonolipid Glu-C_{10:0} (fraction 37) from *E. coli* pAFP1 and selective 1D TOCSY spectra displaying the individual glucopyranosyl ¹H spinsystems of the glucose monolipids 1-3 identified in fraction 37 in methanol-*d*₄ at 600 MHz. Trace 1: α -anomer

of **1**, trace 2: α -anomer of **2**, trace 3: α -anomer of **3**, trace 4: β -anomer of **2**, trace 5: β -anomer of **1**; Figure S12: 2D COSY spectrum of glucosemonolipid Glu-C_{10:0} (fraction 37) from *E. coli* pAFP1 in methanol-*d*₄ at 600 MHz.; Figure S13: Expansion of 2D TOCSY spectrum of glucosemonolipid Glu-C_{10:0} (fraction 37) from *E. coli* pAFP1 displaying the glucopyranosyl ¹H spinsystems of the glucose monolipids **1–3** identified in fraction 37 in methanol-*d*₄ at 600 MHz.; Figure S14: 2D gHSQC spectrum of glucosemonolipid Glu-C_{10:0} (fraction 37) from *E. coli* pAFP1 in methanol-*d*₄ at 600 MHz.; Figure S15: 2D gHMBC spectrum of glucosemonolipid Glu-C_{10:0} (fraction 37) from *E. coli* pAFP1 in methanol-*d*₄ at 600 MHz.; Figure S16: LC-ESI-MS/MS analysis in negative ion mode of the purified glucosedilipids produced by *E. coli* pCAT2 (pooled fraction 46-51). From top to bottom: total ion chromatogram (TIC) showing overall glucosedilipid congeners; extracted ion chromatogram (XIC) of Glu-C_{10:0}-C_{11:0} isomers at RT 12.38 and 12.80 min; XIC of Glu-C_{10:0}-C_{12:1} isomers at RT 12.86 and 13.3 min; XIC of Glu-C_{10:0}-C_{12:0} at RT 14.42 and 14.75 min; and photodiode array (PDA) total scan between 190–400 nm.; Figure S17: Mass spectra (MS² and ESI full MS (inset)) glucosedilipid Glu-C_{10:0}-C_{12:0} as one minor congener produced by *E. coli* pCAT2.; Figure S18: Mass spectra (MS² and ESI full MS (inset)) glucosedilipid Glu-C_{10:0}-C_{11:0} as one minor congener produced by *E. coli* pCAT2.; Figure S19: LC-ESI-MS/MS analysis in negative ion mode of the purified glucosemonolipids produced by *E. coli* pAFP1 (pooled fraction 37-40) From top to bottom: total ion chromatogram (TIC) showing overall glucosemonolipid congeners; extracted ion chromatogram (XIC) of Glu-C_{10:0} isomers with *m/z* [M-H]⁻ of 349.186; XIC of Glu-C_{11:0} isomers with *m/z* [M-H]⁻ of 363.202; XIC of Glu-C_{12:1} with *m/z* [M-H]⁻ of 375.202; XIC of Glu-C_{12:0} with *m/z* [M-H]⁻ of 377.218; and XIC of Glu-C_{14:0} with *m/z* [M-H]⁻ of 405.249.; Figure S20: Mass spectra (MS² and ESI full MS (inset)) glucosemonolipid Glu-C_{11:0} as one minor congener produced by *E. coli* pAFP1.; Figure S21: Mass spectra (MS² and ESI full MS (inset)) glucosemonolipid Glu-C_{12:1} as one minor congener produced by *E. coli* pAFP1.; Figure S22: Mass spectra (MS² and ESI full MS (inset)) glucosemonolipid Glu-C_{12:0} as one minor congener produced by *E. coli* pAFP1.; Figure S23: Mass spectra (MS² and ESI full MS (inset)) glucosemonolipid Glu-C_{14:0} as one minor congener produced by *E. coli* pAFP1.

Author Contributions: Conceptualization, R.H., A.B., M.T. and L.J.V.Z. Supervision, R.H. and M.T. Funding acquisition, A.B., R.H. and M.T. Project administration, A.B., E.H.B.P., M.V. and L.L. Methodology, A.F.P.H., C.T., L.L. and J.G. Resources, M.V., E.H. and W.T.W. Investigation, A.F.P.H., C.T., I.K. and W.T.W. Data curation and software, L.J.V.Z., E.H., J.G. and E.H.B.P. Formal analysis, A.F.P.H., C.T. and J.C. Validation, J.C., J.P. and I.K. Writing—original draft, A.F.P.H., J.C., J.P. and L.J.V.Z. Writing—review and editing, R.H., A.F.P.H., M.T., C.T. and L.J.V.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The molecular biology experiments and research stay were funded under the Wissenschaftlich-Technologischen Zusammenarbeit mit Südafrika (No. 01DG17018) project by the German Federal Ministry of Education and Research and the South African National Research Foundation (UID105876). The first author, A.F.P.H., received a Research Grant for Doctoral Programmes in Germany (No. 57552340) from the German Academic Exchange Service (DAAD). The 600 MHz NMR spectrometer was co-funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project number 317898569).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors acknowledge the outstanding technical assistance provided by Eike Grunwaldt, at Bioprocess Engineering Department (Universität Hohenheim, Germany), on bioreactor experiments and Mario Wolf, at Organic Chemistry Department (Universität Hohenheim, Germany) on all NMR experiments. The authors would also like to express their gratitude to Misri Gozan (Universitas Indonesia, Indonesia) as the doctoral mentor of first author, A.F.P.H.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-------------------|---|
| ACN | Acetonitrile |
| CDW | Cell dry weight |
| CMC | Critical micelle concentration |
| COSY | Correlation Spectroscopy |
| DNA | Deoxyribonucleic Acid |
| DPA | Diphenylamine-aniline-phosphoric acid |
| HMBC | Heteronuclear Multiple Bond Correlation |
| HPTLC | High-performance thin-layer chromatography |
| HSQC | Heteronuclear Single-Quantum Coherence |
| IPTG | Isopropyl β -D-1-thiogalactopyranoside |
| LB | Luria–Bertani |
| LC-ESI/MS | Liquid chromatography/electrospray ionization mass spectrometry |
| MSM | Mineral salt medium |
| NMR | Nuclear magnetic resonance |
| OD ₆₀₀ | Optical density at 600 nm |
| ORF | Open reading frame |
| PCR | Polymerase chain reaction |
| R_f | Retardation factor |
| RT | Retention time |
| TOCSY | Total Correlation Spectroscopy |
| TIC | Total ion chromatogram |
| TLC | Thin-layer chromatography |
| XIC | Extracted ion chromatogram |

References

1. Varvaresou, A.; Iakovou, K. Biosurfactants in cosmetics and biopharmaceuticals. *Lett. Appl. Microbiol.* **2015**, *61*, 214–223. [[CrossRef](#)] [[PubMed](#)]
2. Lukic, M.; Pantelic, I.; Savic, S. An overview of novel surfactants for formulation of cosmetics with certain emphasis on acidic active substances. *Tenside Surfactants Deterg.* **2016**, *53*, 7–19. [[CrossRef](#)]
3. Otzen, D.E. Biosurfactants and surfactants interacting with membranes and proteins: Same but different? *Acta BBA—Biomembr.* **2017**, *1859*, 639–649. [[CrossRef](#)] [[PubMed](#)]
4. Abdel-Mawgoud, A.M.; Stephanopoulos, G. Simple glycolipids of microbes: Chemistry, biological activity and metabolic engineering. *Synth. Syst. Biotechnol.* **2018**, *3*, 3–19. [[CrossRef](#)] [[PubMed](#)]
5. Kulakovskaya, E.; Kulakovskaya, T. *Extracellular Glycolipids of Yeasts: Biodiversity, Biochemistry, and Prospects*; Elsevier: Amsterdam, The Netherlands, 2014. [[CrossRef](#)]
6. Van Bogaert, I.N.A.; Zhang, J.; Soetaert, W. Microbial synthesis of sophorolipids. *Process. Biochem.* **2011**, *46*, 821–833. [[CrossRef](#)]
7. Wittgens, A.; Tiso, T.; Arndt, T.T.; Wenk, P.; Hemmerich, J.; Müller, C.; Wichmann, R.; Küpper, B.; Zwick, M.; Wilhelm, S.; et al. Growth independent rhamnolipid production from glucose using the non-pathogenic *Pseudomonas putida* KT2440. *Microb. Cell Fact.* **2011**, *10*, 80. [[CrossRef](#)] [[PubMed](#)]
8. Kügler, J.H.; Muhle-Goll, C.; Hansen, S.H.; Völp, A.R.; Kirschhöfer, F.; Köhl, B.; Brenner-Weiss, G.; Luy, B.; Syltatk, C.; Hausmann, R. Glycolipids produced by *Rouxiella* sp. DSM 100043 and isolation of the biosurfactants via foam-fractionation. *AMB Express* **2015**, *5*, 82. [[CrossRef](#)] [[PubMed](#)]
9. Le Flèche-Matéos, A.; Kugler, J.H.; Hansen, S.H.; Syltatk, C.; Hausmann, R.; Lomprez, F.; Vandenberg, M.; Manuguerra, J.-C.; Grimont, P.A.D. *Rouxiella badensis* sp. nov. and *Rouxiella silvae* sp. nov. isolated from peat bog soil and emendation description of the genus *Rouxiella*. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 1255–1259. [[CrossRef](#)] [[PubMed](#)]
10. Harahap, A.F.P.; Conrad, J.; Wolf, M.; Pfannstiel, J.; Klaiber, I.; Grether, J.; Hiller, E.; Vahidinasab, M.; Salminen, H.; Treinen, C.; et al. Structure Elucidation and Characterization of Novel Glycolipid Biosurfactant Produced by *Rouxiella badensis* DSM 100043^T. *Molecules* **2025**, *30*, 1798. [[CrossRef](#)] [[PubMed](#)]
11. Abraham, W.R.; Meyer, H.; Yakimov, M. Novel glycine containing glucolipids from the alkane using bacterium *Alcanivorax borkumensis*. *Biochim. Biophys. Acta—Lipids Lipid Metab.* **1998**, *1393*, 57–62. [[CrossRef](#)] [[PubMed](#)]

12. Matsuyama, T.; Kaneda, K.; Ishizuka, I.; Toida, T.; Yano, I. Surface-active novel glycolipid and linked 3-hydroxy fatty acids produced by *Serratia rubidaea*. *J. Bacteriol.* **1990**, *172*, 3015–3022. [[CrossRef](#)] [[PubMed](#)]
13. Handelsman, J.; Liles, M.; Mann, D.; Riesenfeld, C.; Goodman, R.M. Cloning the metagenome: Culture-independent access to the diversity and functions of the uncultivated microbial world. *Methods Microbiol.* **2002**, *33*, 241–255. [[CrossRef](#)]
14. Martinez, A.; Kolvek, S.J.; Yip, C.L.T.; Hopke, J.; Brown, K.A.; MacNeil, I.A.; Osburne, M.S. Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. *Appl. Environ. Microbiol.* **2004**, *70*, 2452–2463. [[CrossRef](#)] [[PubMed](#)]
15. Yanisch-Perron, C.; Vieira, J.; Messing, J. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **1985**, *33*, 103–119. [[CrossRef](#)] [[PubMed](#)]
16. Williams, W.; Kunorozva, L.; Klaiber, I.; Henkel, M.; Pfannstiel, J.; Van Zyl, L.J.; Hausmann, R.; Burger, A.; Trindade, M. Novel metagenome-derived ornithine lipids identified by functional screening for biosurfactants. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 4429–4441. [[CrossRef](#)] [[PubMed](#)]
17. O’Gara, F. Novel or improved expression and screening systems for high—Throughput discovery of new bioactive compounds from marine metagenomic libraries. In *Marine Microbial Biodiversity, Bioinformatics, and Biotechnology*; European Union’s Seventh Framework Programme: Brussels, Belgium, 2014. Available online: <https://api.semanticscholar.org/CorpusID:221376119> (accessed on 22 May 2025).
18. Wang, Y.; Zhang, Z.; Ruan, J. A proposal to transfer *Microbispora bispora* (Lechevalier 1965) to a new genus, *Thermobispora* gen. nov., as *Thermobispora bispora* comb. nov. *Int. J. Syst. Bacteriol.* **1996**, *46*, 933–938. [[CrossRef](#)] [[PubMed](#)]
19. Liles, M.R.; Williamson, L.L.; Rodbumrer, J.; Torsvik, V.; Goodman, R.M.; Handelsman, J. Recovery, Purification, and Cloning of High-Molecular-Weight DNA from Soil Microorganisms. *Appl. Environ. Microbiol.* **2008**, *74*, 3302–3305. [[CrossRef](#)] [[PubMed](#)]
20. Huang, H.; Pandya, C.; Liu, C.; Al-Obaidi, N.F.; Wang, M.; Zheng, L.; Keating, S.T.; Aono, M.; Love, J.D.; Evans, B.; et al. Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1974–E1983. [[CrossRef](#)] [[PubMed](#)]
21. Moi, D.; Bernard, C.; Steinegger, M.; Nevers, Y.; Langleib, M.; Dessimoz, C. Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-positive bacteria and their viruses. *bioRxiv* **2023**. [[CrossRef](#)]
22. Xie, J.; Chen, Y.; Cai, G.; Cai, R.; Hu, Z.; Wang, H. Tree Visualization by One Table (tvBOT): A web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* **2013**, *51*, W587–W592. [[CrossRef](#)] [[PubMed](#)]
23. Gilchrist, C.L.M.; Booth, T.J.; Van Wersch, B.; Van Grieken, L.; Medema, M.H.; Chooi, Y.H. cblaster: A remote search tool for rapid identification and visualization of homologous gene clusters. *Bioinform. Adv.* **2021**, *1*, vbab016. [[CrossRef](#)] [[PubMed](#)]
24. Chen, C.; Wu, Y.; Li, J.; Wang, X.; Zeng, Z.; Xu, J.; Liu, Y.; Feng, J.; Chen, H.; He, Y.; et al. TBtools-II: A ‘one for all, all for one’ bioinformatics platform for biological big-data mining. *Mol. Plant* **2023**, *16*, 1733–1742. [[CrossRef](#)] [[PubMed](#)]
25. Gibson, D.G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **2011**, *498*, 349–361. [[CrossRef](#)] [[PubMed](#)]
26. Patil, J.R.; Chopade, B.A. Studies on bioemulsifier production by *Acinetobacter* strains isolated from healthy human skin. *J. Appl. Microbiol.* **2001**, *91*, 290–298. [[CrossRef](#)] [[PubMed](#)]
27. Morikawa, M.; Hirata, Y.; Imanaka, T. A study on the structure-function relationship of lipopeptide biosurfactants. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2000**, *1488*, 211–218. [[CrossRef](#)] [[PubMed](#)]
28. Riesenberg, D.; Schulz, V.; Knorre, W.; Pohl, H.-D.; Korz, D.; Sanders, E.; Roß, A.; Deckwer, W.-D. High cell density cultivation of *Escherichia coli* at controlled specific growth rate. *J. Biotechnol.* **1991**, *20*, 17–27. [[CrossRef](#)] [[PubMed](#)]
29. Henkel, M.; Zwick, M.; Beuker, J.; Willenbacher, J.; Baumann, S.; Oswald, F.; Neumann, A.; Siemann-Herzberg, M.; Syldatk, C.; Hausmann, R. Teaching bioprocess engineering to undergraduates: Multidisciplinary hands-on training in a one-week practical course. *Biochem. Mol. Biol. Educ.* **2015**, *43*, 189–202. [[CrossRef](#)] [[PubMed](#)]
30. Hiller, E.; Off, M.; Hermann, A.; Vahidinasab, M.; Perino, E.H.B.; Lilge, L.; Hausmann, R. The influence of growth rate-controlling feeding strategy on the surfactin production in *Bacillus subtilis* bioreactor processes. *Microb. Cell Fact.* **2024**, *23*, 260. [[CrossRef](#)] [[PubMed](#)]
31. Sharma, J.; Sundar, D.; Srivastava, P. Biosurfactants: Potential Agents for Controlling Cellular Communication, Motility, and Antagonism. *Front. Mol. Biosci.* **2021**, *8*, 727070. [[CrossRef](#)] [[PubMed](#)]
32. Zhao, M.; Tyson, C.; Gitaitis, R.; Kvitko, B.; Dutta, B. *Rouxiiella badensis*, a new bacterial pathogen of onion causing bulb rot. *Front. Microbiol.* **2022**, *13*, 1054813. [[CrossRef](#)] [[PubMed](#)]
33. Snoeck, S.; Guidi, C.; De Mey, M. ‘Metabolic Burden’ Explained: Stress Symptoms and Its Related Responses Induced by (Over)Expression of (Heterologous) Proteins in *Escherichia coli*; BioMed Central Ltd.: London, UK, 2024. [[CrossRef](#)]
34. Lennen, R.M.; Kruziki, M.A.; Kumar, K.; Zinkel, R.A.; Burnum, K.E.; Lipton, M.S.; Hoover, S.W.; Ranatunga, D.R.; Wittkopp, T.M.; Marner, W.D.; et al. Membrane stresses induced by overproduction of free fatty acids in *Escherichia coli*. *Appl. Environ. Microbiol.* **2011**, *77*, 8114–8128. [[CrossRef](#)] [[PubMed](#)]
35. Fujita, Y.; Matsuoka, H.; Hirooka, K. Regulation of fatty acid metabolism in bacteria. *Mol. Microbiol.* **2007**, *66*, 829–839. [[CrossRef](#)] [[PubMed](#)]

36. Baumgartner, J.T.; Mohammad, T.S.H.; Czub, M.P.; Majorek, K.A.; Arolli, X.; Variot, C.; Anonick, M.; Minor, W.; Ballicora, M.A.; Becker, D.P.; et al. Gcn5-Related N-Acetyltransferases (GNATs) with a Catalytic Serine Residue Can Play Ping-Pong Too. *Front. Mol. Biosci.* **2021**, *8*, 646046. [[CrossRef](#)] [[PubMed](#)]
37. Asensio, T.; Dian, C.; Boyer, J.B.; Rivière, F.; Meinel, T.; Giglione, C. A Continuous Assay Set to Screen and Characterize Novel Protein N-Acetyltransferases Unveils Rice General Control Non-repressible 5-Related N-Acetyltransferase2 Activity. *Front. Plant Sci.* **2022**, *13*, 832144. [[CrossRef](#)] [[PubMed](#)]
38. Daigle, D.M.; Hughes, D.W.; Wright, G.D. Prodigious substrate specificity of AAC(6′)-APH(2′′), an aminoglycoside antibiotic resistance determinant in enterococci and staphylococci. *Chem. Biol.* **1999**, *6*, 99–110. [[CrossRef](#)] [[PubMed](#)]
39. Majorek, K.A.; Kuhn, M.L.; Chruszcz, M.; Anderson, W.F.; Minor, W. Structural, functional, and inhibition studies of a Gcn5-related N-acetyltransferase (GNAT) superfamily protein PA4794: A new C-terminal lysine protein acetyltransferase from *Pseudomonas aeruginosa*. *J. Biol. Chem.* **2013**, *288*, 30223–30235. [[CrossRef](#)] [[PubMed](#)]
40. Zhang, W.; Huffman, J.; Li, S.; Shen, Y.; Du, L. Unusual acylation of chloramphenicol in *Lysobacter enzymogenes*, a biocontrol agent with intrinsic resistance to multiple antibiotics. *BMC Biotechnol.* **2017**, *17*, 59. [[CrossRef](#)] [[PubMed](#)]
41. Hubrich, F.; Bösch, N.M.; Chepkirui, C.; Morinaka, B.I.; Rust, M.; Gugger, M.; Robinson, S.L.; Vagstad, A.L.; Piel, J. Ribosomally derived lipopeptides containing distinct fatty acyl moieties. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2113120119. [[CrossRef](#)] [[PubMed](#)]
42. Frankel, B.A.; Blanchard, J.S. Mechanistic analysis of *Mycobacterium tuberculosis* Rv1347c, a lysine N ϵ -acyltransferase involved in mycobactin biosynthesis. *Arch. Biochem. Biophys.* **2008**, *477*, 259–266. [[CrossRef](#)] [[PubMed](#)]
43. Robertson, R.M.; Yao, J.; Gajewski, S.; Kumar, G.; Martin, E.W.; O Rock, C.; White, S.W. A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer. *Nat. Struct. Mol. Biol.* **2017**, *24*, 666–671. [[CrossRef](#)] [[PubMed](#)]
44. Smith, A.F.; Silvano, E.; Päuker, O.; Guillonneau, R.; Quareshy, M.; Murphy, A.; A Mausz, M.; Stirrup, R.; Rihtman, B.; Aguilo-Ferretjans, M.; et al. A novel class of sulfur-containing aminolipids widespread in marine roseobacters. *ISME J.* **2021**, *15*, 2440–2453. [[CrossRef](#)] [[PubMed](#)]
45. Vasilopoulos, G.; Heflik, L.; Czolkoss, S.; Heinrichs, F.; Kleetz, J.; Yesilyurt, C.; Tischler, D.; Westhoff, P.; Exterkate, M.; Aktas, M.; et al. Characterization of multiple lysophosphatidic acid acyltransferases in the plant pathogen *Xanthomonas campestris*. *FEBS J.* **2024**, *291*, 705–721. [[CrossRef](#)] [[PubMed](#)]
46. Aygun-Sunar, S.; Bilaloglu, R.; Goldfine, H.; Daldal, F. *Rhodobacter capsulatus* OlsA is a bifunctional enzyme active in both ornithine lipid and phosphatidic acid biosynthesis. *J. Bacteriol.* **2007**, *189*, 8564–8574. [[CrossRef](#)] [[PubMed](#)]
47. Maleki, S.; Hrudikova, R.; Zotchev, S.B.; Ertesvåg, H. Identification of a new phosphatase enzyme potentially involved in the sugar phosphate stress response in *Pseudomonas fluorescens*. *Appl. Environ. Microbiol.* **2017**, *83*, e02361-16. [[CrossRef](#)] [[PubMed](#)]
48. Kankanamge, L.S.P.; Ruffner, L.A.; Touch, M.M.; Pina, M.; Beuning, P.J.; Ondrechen, M.J. Functional annotation of haloacid dehalogenase superfamily structural genomics proteins. *Biochem. J.* **2023**, *480*, 1553–1569. [[CrossRef](#)] [[PubMed](#)]
49. Kuznetsova, E.; Nocek, B.; Brown, G.; Makarova, K.S.; Flick, R.; Wolf, Y.I.; Khusnutdinova, A.; Evdokimova, E.; Jin, K.; Tan, K.; et al. Functional diversity of haloacid dehalogenase superfamily phosphatases from *Saccharomyces cerevisiae*: Biochemical, structural, and evolutionary insights. *J. Biol. Chem.* **2015**, *290*, 18678–18698. [[CrossRef](#)] [[PubMed](#)]
50. Kuznetsova, E.; Proudfoot, M.; Gonzalez, C.F.; Brown, G.; Omelchenko, M.V.; Borozan, I.; Carmel, L.; Wolf, Y.I.; Mori, H.; Savchenko, A.V.; et al. Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J. Biol. Chem.* **2006**, *281*, 36149–36161. [[CrossRef](#)] [[PubMed](#)]
51. Bazire, A.; Dufour, A. The *Pseudomonas aeruginosa* rhlG and rhlAB genes are inversely regulated and RhlG is not required for rhamnolipid synthesis. *BMC Microbiol.* **2014**, *14*, 160. [[CrossRef](#)] [[PubMed](#)]
52. Saerens, K.M.J.; Van Bogaert, I.N.A.; Soetaert, W. Characterization of sophorolipid biosynthetic enzymes from *Starmerella bombicola*. *FEMS Yeast Res.* **2015**, *15*, fov075. [[CrossRef](#)] [[PubMed](#)]
53. Wada, K.; Koike, H.; Fujii, T.; Morita, T. Targeted transcriptomic study of the implication of central metabolic pathways in mannosylerythritol lipids biosynthesis in *Pseudozyma antarctica* T-34. *PLoS ONE* **2020**, *15*, e0227295. [[CrossRef](#)] [[PubMed](#)]
54. Lopes, J.F.; Gaspar, E.M.S.M. Simultaneous chromatographic separation of enantiomers, anomers and structural isomers of some biologically relevant monosaccharides. *J. Chromatogr. A* **2008**, *1188*, 34–42. [[CrossRef](#)] [[PubMed](#)]
55. Royce, L.A.; Liu, P.; Stebbins, M.J.; Hanson, B.C.; Jarboe, L.R. The damaging effects of short chain fatty acids on *Escherichia coli* membranes. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 8317–8327. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.