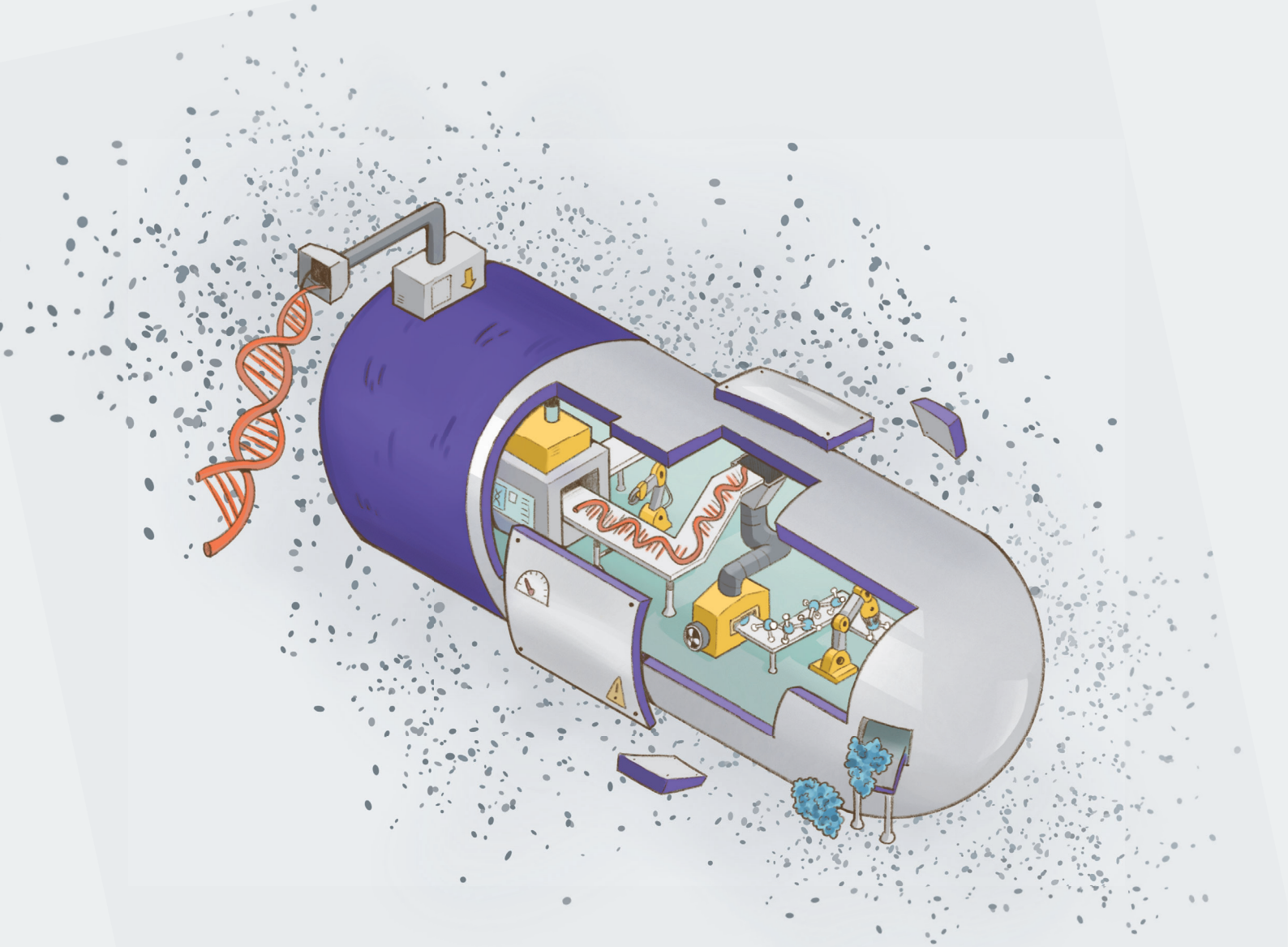
UNIVERSITAT DE BARCELONA

Recombinant production of enzymes for the paper industry

Jordi Ferrando Núñez

UNIVERSITAT DE BARCELONA
FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Recombinant production of enzymes for the paper industry



PhD Thesis

Jordi Ferrando Nuñez

PhD Program in Biotechnology

September 2025, Barcelona

UNIVERSITAT DE BARCELONA

Facultat de Farmàcia
i Ciències de l'Alimentació

Programa de doctorat en Biotecnologia

Departament de Biologia, Sanitat i Medi Ambient

Recombinant production of enzymes for the paper industry

Producció recombinant d'enzims per a la indústria paperera

Memòria presentada per Jordi Ferrando Núñez per optar al títol de doctor per la Universitat de Barcelona

Director i tutor

Dr. Pere Picart Faiget

Director

Dr. David Miñana i Galbis

Doctorand

Jordi Ferrando Núñez

Setembre 2025

La present tesi doctoral ha estat finançada pel Pla de Doctorats Industrials del Departament de Recerca i Universitats de la Generalitat de Catalunya, amb el suport de la Gestió d'Ajuts Universitaris de Recerca.

Acknowledgments

Fer una tesi doctoral és un procés llarg i exigent, que sovint posa a prova la paciència i la resistència, sobretot en aquells moments en què et preguntes si feia falta iniciar aquest camí. Per mi, sortir a córrer ha estat una manera de desconnectar. Córrer requereix constància i força de voluntat per no aturar-se fins a l'objectiu final, i sovint m'ha fet veure les similituds amb la tesi. Totes dues són trajectes llargs que demanen preparació, disciplina i la capacitat d'afrontar pujades i baixades, moments d'alegria i també de dubte o frustració. Tot i que al final la lluita és amb un mateix, el suport de la gent que t'acompanya és essencial: són els qui t'ajuden a avançar, que t'animen i amb qui comparteixes els moments viscuts, fins al punt d'esdevenir part de la feina feta. Per això vull dedicar aquestes primeres pàgines a agrair a totes les persones que m'han ajudat i amb qui he compartit aquesta experiència.

Primer de tot vull agrair al meu supervisor i director de tesi, Pere Picart, per donar-me no només l'oportunitat de realitzar aquesta tesi al teu grup, sinó també per acollir-me durant el màster i els mesos d'espera previs a la tesi. La teva actitud positiva en tot moment i l'habilitat de trobar la manera de fer riure la gent (5 minuts al dia), ha estat realment diferencial i de gran ajuda. Sempre que alguna cosa no funcionava del tot bé o no sortien les coses, sempre has trobat un segon punt de vista més optimista que m'ha ajudat a continuar endavant sense perdre la motivació. També vull agrair al meu codirector de tesi, David, per les teves revisions i aportacions dels articles i d'aquesta tesi. Sabia que l'article anava bé quan ja no trobaves cap detall a millorar. També vull fer menció a tots els membres del tribunal de la tesi i als revisors externs.

Vull donar les gràcies a tots els membres del departament, que al final ha estat l'ambient que m'ha permès dur a terme la tesi. En especial mencionar la Masha, que ha fet que la tesi es faci molt més amena. Tot i que al principi no deia massa cosa, la seva energia ha fet que s'acabi convertint en l'ànima del departament. Ens ha estat cuidant, alimentant-nos amb el banc d'aliments, regalant roses per Sant Jordi, i mil coses més. Xavi, Merche, Elena, Carme, Maria Jesús, Vicent, Ana, Maribel i Gisela, moltes gràcies a tots i totes. Durant la tesi també he compartit laboratori i temps amb l'Oriana, la Pilar i en Pepe, estudiants de màster que han donat vida al laboratori, que han fet una gran feina i que de tots he après alguna cosa. En especial agrair a l'Oriana, que a més va fer un treball excel·lent amb els carotenoides. Dono també les gràcies a la Mile, que a l'inici ens va deixar part del seu material i ha acabat sent de gran servei. Dana i Evelyn, gràcies també pel temps compartit al departament durant aquest camí.

També vull donar les gràcies a les tècniques del laboratori, la Maria Jesús i la Carme, per totes les vegades que m'han ajudat, i que ja el primer any en vaig perdre el compte. Segurament sense

la Maria Jesús encara estaria encallat intentant preparar alguna solució. Ha estat una sort tenir-te al laboratori, per qualsevol mena de problema o entrebanc sempre has estat per ajudar-me i donar una resposta. Moltes gràcies.

També vull expressar el meu agraïment a totes les persones d'entitats externes que han col·laborat en aquesta tesi. A l'Antoni Planas, de l'Institut Químic de Sarrià (IQS), moltes gràcies per obrir-me les portes del teu laboratori i per totes les valuoses aportacions en el darrer article de la tesi. A la meva amiga Clàudia Lliso, doctoranda de l'IQS, gràcies per la gran feina en la purificació d'enzims i, sobretot, per la teva constant disponibilitat per ajudar en qualsevol problema o dubte. També estic molt agraït a tot el grup de CELBIOTECH de la Universitat Politècnica de Catalunya per l'excel·lent acollida durant els mesos que hi vaig passar aplicant la cel·lulasa. En especial, vull agrair la supervisió i ajuda constant de la Blanca Roncero, així com els consells i el suport de diferents membres del professorat de CELBIOTECH, entre els quals voldria destacar l'Oriol Cusola, l'Úrsula Fillat, Cristina Valls i Teresa Vidal. Igualment, moltes gràcies als seus estudiants i doctorands, Geraldine i Eugenio, per rebre'm com un doctorand més del grup i fer-me sentir sempre tan còmode. Donar també les gràcies a Daniel R. Zeigler, exdirector del Bacillus Genetic Stock Center i reconegut expert en *Bacillus*, per haver proporcionat soques i eines genètiques de *Bacillus subtilis*, així com per la generosa i constant transferència del seu immens coneixement sobre aquesta soca, especialment durant el màster i l'inici de la tesi. Vull expressar el meu agraïment a l'Oriol, l'Albert i a tot el seu equip per rebre'm tan bé.

Els amics amb qui he compartit més temps durant aquests anys de tesi també han estat part essencial, sobretot l'Aleix, Pau, Marc, Bernat, Raúl i Pierluigi, que mai han deixat de preguntar com anava la tesi i que, amb tots els moments compartits, m'han ajudat a desconnectar i mantenir un equilibri. Entre tots ells, vull agrair especialment a l'Aleix Díaz, que amb el seu talent ha volgut col·laborar com a autor del dibuix de la portada d'aquesta tesi.

Finalment, vull agrair a la meva família, que sempre ha estat al meu costat, no només durant la tesi, sinó en tots els àmbits de la vida. No els dono les gràcies només pel suport en aquest camí, sinó també per tot el que han fet perquè avui sigui aquí. La meva mare, el meu pare i la meva germana han estat els exemples més clars que, malgrat les dificultats i els entrebancs, sempre cal tirar endavant. Aquesta és la lliçó més valuosa que m'han transmès. També vull agrair als meus tiets i tietes, cosins i cosines, que sempre s'han preocupat per mi i han estat disposats a donar-me un cop de mà quan ha calgut. Per acabar, vull mencionar les persones que ja no hi són, papa, avis i àvies, que sempre han estat una font de motivació per intentar fer que se sentin orgullosos.

Abstract

Enzymes have become indispensable tools in modern bioprocessing, offering sustainable and high-specificity alternatives to traditional chemical catalysts. In the pulp and paper industry, cellulases and amylases are increasingly used for reducing energy consumption, water use, and reliance on harsh chemicals. However, industrial implementation is often constrained by the need for catalytic activity under extreme conditions, such as elevated temperatures and alkaline pH. Successful industrial use thus requires both robust enzyme candidates and production systems capable of achieving high titers at low cost. *Bacillus subtilis* is a preferred microbial host in industrial platforms due to its GRAS (generally regarded as safe) or QPS (qualified presumption of safety) status and secretion capacity, yet its utility is often limited by bottlenecks in heterologous protein expression and secretion pathway, often leading to failed projects in both academic and industrial settings.

Within this context, this thesis presents a cohesive set of four publications that address these challenges through enzyme discovery and host strain engineering for developing advanced tools to enhance heterologous protein production in *B. subtilis*. The first paper reports on the identification and characterization of StachCel5, a novel GH5 endoglucanase mined from the genome of *Stachybotrys chartarum*, a well-known fungus producing thermostable and alkali-tolerant cellulases. Expressed in *Komagataella phaffii*, StachCel5 exhibited excellent thermostability at 50 °C, broad pH activity (pH 4–9), and strong specificity for β-1,4-glucans. Application in pulp refining demonstrated a 25% reduction in mechanical energy demand while improving fiber bonding and paper strength without compromising fiber integrity. These findings underline its potential for industrial use beyond paper and pulp industry, including in the biofuel, food, and detergent sectors.

Nonetheless, for industrial application, enzymes must be produced efficiently and at scale to be economically viable. The second paper of this thesis focuses on developing a modular, high-titer expression platform in *B. subtilis*. By employing the CRISPR-Cas9 system alongside promoter optimization, gene copy number modulation, and secretion enhancement, a stable and scalable AmyQ α-amylase production platform was established. A multi-copy strain was designed with up to six chromosomal *amyQ* copies under the dual promoter P_{amyQ}-P_{cry3A}, combined with overexpression of the chaperone PrsA and signal peptidase SppA. The engineered strain reached titers exceeding 1,400 U/mL in a 3 L bioreactor, demonstrating industrially relevant production capacity.

To overcome the time-intensive process of iterative genome editing, the third and fourth papers introduce a novel CRISPR-Cas9-based multiplex genome engineering platform for *B. subtilis*, featuring a colorimetric screening system based on the *crtMN* operon from *Staphylococcus aureus*. This operon confers a bright yellow pigmentation on the host, which is visibly lost upon successful gene replacement, enabling rapid, PCR-free selection of successful edited clones. Three open-access, patent-free *B. subtilis* KO7-S strains were constructed with one to three chromosomal *crtMN* insertions, along with a CRISPR-Cas9 single-plasmid toolkit (Paper 3). This system enables simultaneous, scarless integration of up to three gene copies in a single step, thus dramatically accelerating strain construction timelines. Building on this chassis, the fourth study demonstrates its use in carotenoid biosynthesis. By co-integrating multiple *crtMN* copies, enhancing precursor supply through *fpps* overexpression, and deleting *yisP*, a stable *B. subtilis* strain was created that produces C_{30} carotenoids at levels comparable to medium-copy plasmid systems, without the drawbacks of plasmid instability or antibiotic selection.

Together, these works define a robust pipeline for industrial biocatalyst development: from the discovery of robust enzymes suitable for process-relevant conditions, to the construction of stable and high-expression microbial platforms. This thesis contributes practical tools and strategies towards more sustainable, energy-efficient, and economically viable bioprocesses.

Keywords: CRISPR-Cas9; *Bacillus subtilis*; Pulp and Paper Industry; Microbial Cell Factories; Strain engineering; Amylase; Cellulase

Resum

Els enzims s'han convertit en eines indispensables en la indústria moderna, ja que ofereixen alternatives sostenibles i d'alta especificitat als catalitzadors químics tradicionals. A la indústria de la pasta i el paper, les cel·lulases i les amilases s'utilitzen cada cop més per a reduir el consum energètic, l'ús d'aigua i la dependència de productes químics agressius. Tanmateix, la seva aplicació a escala industrial sovint es veu limitada per la necessitat que mantinguin l'activitat catalítica en condicions extremes, com ara temperatures elevades o pH alcalí. Per tant, l'èxit en l'àmbit industrial requereix tant d'enzims robustos com de sistemes de producció capaços d'assolir alts rendiments a baix cost. *Bacillus subtilis* és un hoste habitualment escollit com a plataforma industrial gràcies al seu estatus GRAS (generalment reconegut com a segur) o QPS (presumpció qualificada de seguretat) i la seva capacitat de secreció. No obstant això, la seva utilitat es veu sovint limitada per colls d'ampolla en l'expressió i la secreció de proteïnes heteròlogues, fet que pot conduir a l'abandonament de projectes tant en l'àmbit acadèmic com en l'industrial.

En aquest context, aquesta tesi presenta un conjunt de quatre estudis que aborden aquests reptes mitjançant el descobriment d'enzims i l'enginyeria de soques hostes, amb l'objectiu de desenvolupar eines que permetin millorar la producció de proteïnes heteròlogues en *B. subtilis*. El primer article descriu la identificació i caracterització de StachCel5, una nova endoglucanasa GH5 descoberta al genoma de *Stachybotrys chartarum*, un fong conegut per produir cel·lulases termo-estables i tolerants a ambients alcalins. Expressada a *Komagataella phaffii*, StachCel5 va mostrar una gran estabilitat tèrmica a 50 °C, activitat en un ampli interval de pH (4–9) i alta especificitat per β-1,4-glucans. L'aplicació d'aquest enzim en el refinat de pasta va permetre reduir en un 25% el consum d'energia mecànica, alhora que millorava l'enllaç entre fibres i la resistència del paper, sense afectar la integritat de la fibra. Aquests resultats també posen de manifest el seu potencial d'aplicació més enllà de la indústria de la pasta i el paper, incloent-hi sectors com els biocombustibles, l'alimentació i els detergents.

Tot i això, per ser econòmicament viables, aquests enzims han de poder ser produïts de manera eficient i a gran escala. En aquest sentit, el segon article de la tesi se centra en el desenvolupament d'una plataforma d'expressió d'alt rendiment a *B. subtilis*. Mitjançant l'ús del sistema CRISPR-Cas9 combinat amb l'optimització de promotors, la modulació del nombre de còpies gèniques i la millora de les rutes de secreció, es va establir una plataforma estable i escalable per a la producció de l'α-amilasa AmyQ. D'aquesta manera, es va dissenyar una soca amb fins a sis còpies cromosòmiques del gen *amyQ* sota el promotor dual P_{amyQ}-P_{cry3A}, combinat

amb la sobreexpressió de la xaperona PrsA i la peptidasa senyal SppA. La soca obtinguda va assolir concentracions superiors a 1.400 U/mL en un bioreactor de 3 L, demostrant una capacitat de producció rellevant a nivell industrial.

Per a reduir el temps que requereix la integració genòmica iterativa, el tercer i quart article presenten una nova plataforma d'enginyeria genòmica basada en CRISPR-Cas9 per a *B. subtilis*, amb un sistema de cribratge colorimètric basat en l'operó *crtMN* de *Staphylococcus aureus*. Aquest operó confereix una pigmentació groga intensa a l'hoste, que desapareix quan la substitució gènica és exitosa, cosa que permet una selecció ràpida i sense PCR dels clons correctament editats. Es van construir tres soques *B. subtilis* KO7-S d'accés obert i lliure de patents, amb una, dues o tres insercions cromosòmiques de l'operó *crtMN*, juntament amb una eina CRISPR-Cas9 basada en un sol plasmidi (Article 3). Aquest sistema permet integrar fins a tres còpies gèniques de manera simultània, reduint notablement els temps de construcció de soques. A partir d'aquesta plataforma, el quart estudi n'aplica l'ús a la biosíntesi de carotenoides. Mitjançant la integració de múltiples còpies de *crtMN* i l'optimització del subministrament de precursors a través de la sobreexpressió de *fpps* i l'eliminació de *yisP*, es va aconseguir desenvolupar una soca estable de *B. subtilis* capaç de produir carotenoides C_{30} a nivells comparables als sistemes amb plasmidis de còpia mitjana, però sense els inconvenients de la inestabilitat plasmídica ni la necessitat d'antibiòtics.

En conjunt, aquests estudis defineixen un pla de treball robust per al desenvolupament de biocatalitzadors industrials: des del descobriment d'enzims robustos adaptats a condicions de procés industrial fins a la construcció de plataformes microbianes estables i d'elevada expressió. Finalment, aquesta tesi aporta eines i estratègies pràctiques per avançar cap a bioprocessos més sostenibles, eficients energèticament i viables econòmicament.

Paraules clau: CRISPR-Cas9; *Bacillus subtilis*; Indústria de la pasta i el paper; Fàbriques cel·lulars microbianes; Enginyeria de soques; Amilasa; Cel·lulasa

TABLE OF CONTENTS

Abbreviations and nomenclature

Cas CRISPR-associated

Cas9 Cas protein 9

CBH Cellobiohydrolase

CDD Conserved domain database

CMC Carboxymethyl cellulose

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

crRNA CRISPR RNA

CSM Counter-selectable markers

DR Direct repeats

DSB Double-stranded breaks

eDNA Environmental DNA

EG Endoglucanase

ER Endoplasmic reticulum

FBS Functional-based screening

FPP Farnesyl pyrophosphate

GAPDH Glyceraldehyde-3-phosphate dehydrogenase

GH5 Glycoside hydrolase family 5

GHG Greenhouse gas

GRAS Generally recognized as safe

gRNA RNA guide

GWP Global warming potential

HDR Homology-directed repair

IEA International Energy Agency

IPCC Intergovernmental Panel on Climate Change

NCBI National Center for Biotechnology Information

NHEJ Non-homologous end joining

PAM Protospacer-adjacent motif

PPI Pulp and paper industry

PTM Post-translational modification

QPS Qualified presumption of safety

SBS Sequence-based screening

Sec General secretion pathway

sgRNA Single guide RNA

SP Signal peptide

SSR Site-specific recombinase systems

Tat Twin-arginine translocation pathway

tracrRNA Trans-activating crRNA

α-MF α-mating factor

βG β-glucosidase

1 Introduction

1. Introduction

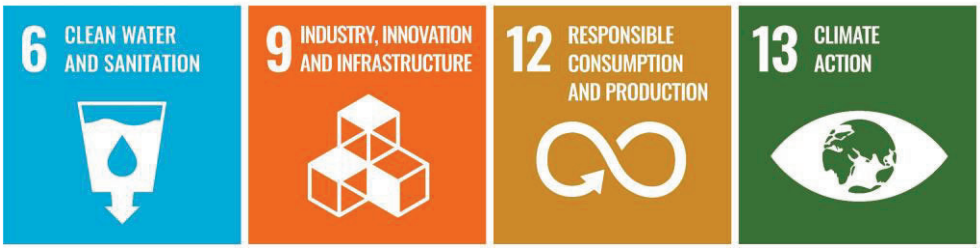
The Intergovernmental Panel on Climate Change (IPCC) of the United Nations has established that the rise in global temperatures is primarily caused by increased greenhouse gas (GHG) emissions resulting from human activities. Within this context, the global energy system is the bedrock of modern economies and societies. However, the production and consumption of energy accounts for 75% of GHG emissions, making it the primary driver of climate change, which threatens human existence worldwide through more frequent heatwaves, marine heat events, heavy rainfall, reduced crop yields, freshwater scarcity, and rising sea levels (International Energy Agency, 2024; Eyring *et al.*, 2021).

Industry, being the most energy-consuming and CO_2-emitting end-use sector, accounts for 39% of total final energy consumption and 45% of CO_2 emissions (IEA, 2024). Since 2000, industrial emissions have surged by around 70%, largely due to rising global demand for industrial goods. In 2022 alone, the sector emitted 9 gigatonnes (Gt) of CO_2. Although emissions experienced minor declines in 2020 and 2022, these reductions are far from sufficient to meet the International Energy Agency's (IEA) Net Zero Emissions by 2050 (NZE) target, which calls for a reduction to 7 Gt CO_2 by 2030 (IEA, 2021, 2023). Achieving this goal will require direct industrial emissions to drop by nearly 25% by the end of the decade, underscoring the need for urgent and accelerated action. Additionally, despite advancements in cleaner technologies, fossil fuels continue to dominate the energy mix in industry, posing a major challenge for decarbonization efforts (IEA, 2021).

The Pulp and Paper Industry (PPI) ranks among the five most energy-intensive industrial sectors and is widely recognized as a resource-intensive industry, requiring substantial amounts of energy, water and chemicals throughout its processes. Notably, this industry accounts for approximately 6% of global industrial energy consumption and contributes around 2% of direct industrial CO_2 emissions (Furszyfer Del Rio *et al.*, 2022). PPI is fundamental to daily life, producing essential materials such as tissues, coffee filters, and packaging. Global paper production already exceeds 400 million tons annually and is projected to reach 900 million tons by 2050, driven by the growing demand for bio-based packaging solutions and increased paperboard consumption, particularly in developing economies (Furszyfer Del Rio *et al.*, 2022). While the PPI has made notable progress in reducing emissions and adopting cleaner fuels, often cited as a reference within industrial decarbonization efforts (Bergquist & Söderholm, 2018), greater efforts are required to significantly cut emissions and align with the Net Zero Scenario outlined by the IPCC and IEA (IEA, 2021; Eyring *et al.*, 2021).

In this regard, the development of more sustainable industrial processes supports some of the United Nations Sustainable Development Goals (United Nations, 2015):

- **Goal 6**. Ensure availability and sustainable management of water and sanitation for all.
- **Goal 9.** Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation.
- **Goal 12.** Ensure sustainable consumption and production patterns.
- **Goal 13.** Take urgent action to combat climate change and its impacts.



1.1. Sustainable impact of Biotechnology in the Pulp and Paper Industry

The introduction of biotechnology into the PPI offers a promising pathway toward more sustainable and environmentally friendly production processes, providing additional advantages in developing cleaner technologies that yield high-quality final products. Enzymes, in particular, have emerged as transformative biocatalysts, offering a powerful alternative to conventional chemical treatments by enabling cleaner technologies, lowering energy consumption and reducing the reliance on harsh chemicals (Saira & Vauhkonen, 2024). One of the key advantages of enzymatic applications lies in their capacity to efficiently degrade lignocellulosic materials, thereby enhancing pulp extraction and pulp refining. This results in improved fiber separation and flexibility, while reducing water and energy consumption. Moreover, in bleaching-stages, lignin-degrading enzymes contribute to partial lignin removal, enhancing pulp brightness with fewer harmful by-products. Enzymatic treatments also show significant benefits in recycled paper processing, particularly in the deinking step. Enzymes help break down ink binders, adhesives and other sticky contaminants from recycled materials, thus reducing chemical usage and improving process stability (Yang *et al.*, 2023). Overall, enzymatic processes contribute to a more resource-efficient industry by lowering energy and chemicals consumption, improving drainage, enhancing paper qualities and increasing bleachability (Saira & Vauhkonen, 2024; Yang *et al.*, 2023). A summary of enzyme applications across key steps in paper production is illustrated in **Figure 1**.

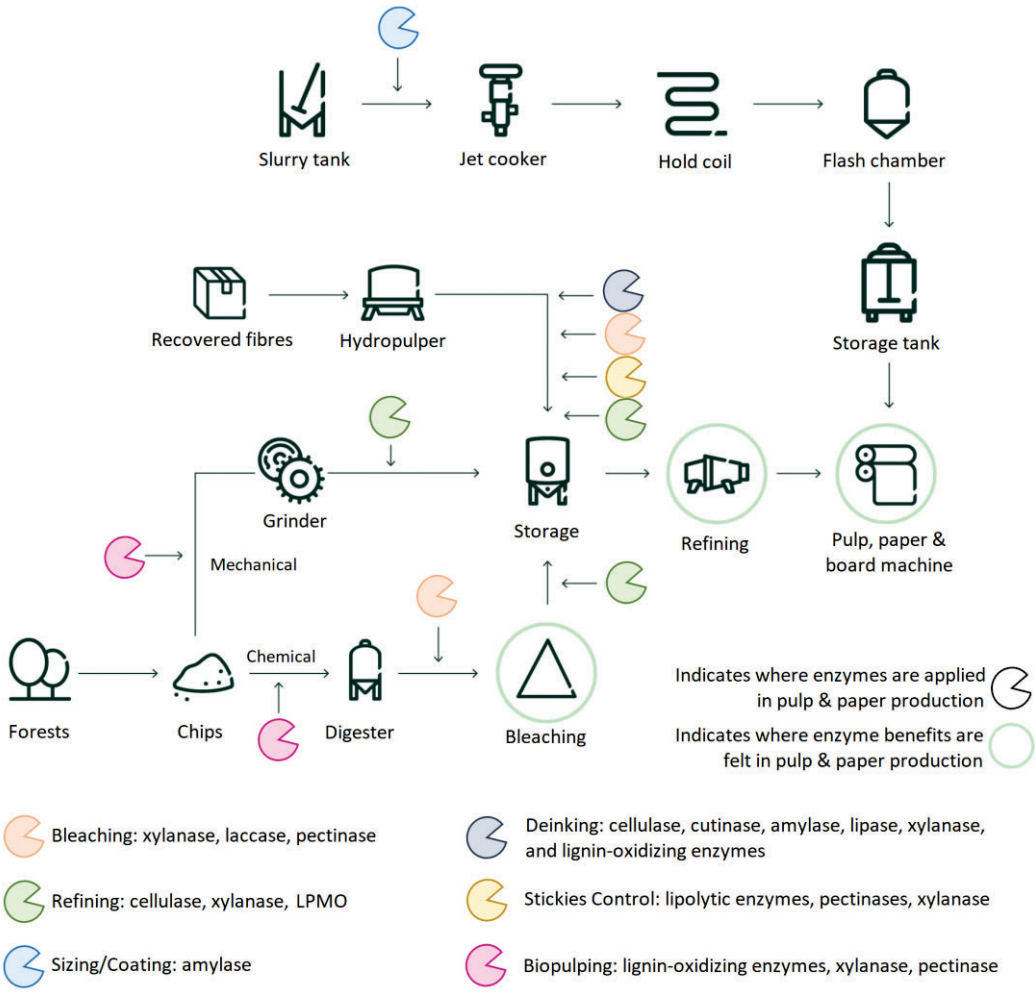


Figure 1. Diagram of enzyme applications in PPI. Arrows indicate specific stages where enzymatic treatments can be applied, including fiber refining, bleaching, and deinking in recycled paper processing. Enzyme interventions contribute to reduced chemical usage, improved fiber quality, and enhanced process efficiency. Adapted and modified from Novozymes A/S (2025).

Therefore, the integration of enzymes into industrial workflows can foster the pulp and paper sector's transition toward more sustainable practices by reducing pollutant generation and aligning with circular economy principles. This shift not only demonstrates the industry's commitment to sustainability but also contributes to addressing broader global environmental challenges. Additionally, the adoption of biotechnology can lead to significant cost reductions, as energy typically accounts for approximately 16% of total productions costs, and up to 30% in some regions (Furszyfer Del Rio *et al.*, 2022).

The industrial enzymes market is projected to grow from 7.9 billion USD in 2024 to 11.2 billion USD by 2029, at a compound annual growth rate (CAGR) of 7.2%. Specifically, the pulp and paper

enzymes market exceeded 170 million USD in 2022 and is expected to record over 5% CAGR between 2023 and 2032, driven by the abovementioned increasing demand for sustainable and eco-friendly products (Global Market Insights, 2023). Enzymes employed in the PPI encompass a diverse range of hydrolytic and oxidative enzymes, including lipases, pectinases, amylases, laccases, xylanases, cellulases, proteases, and esterases, as well as lignin-degrading enzymes and auxiliary enzymes such as lytic polysaccharide monooxygenases (LPMOs) (**Figure 1**) (Yang *et al.*, 2023). Among these, cellulase segment is dominating the paper and pulp market, largely due to their effectiveness in reducing energy consumption during the refining stage, a process that accounts for 15-27% of total energy consumption of a paper mill (Nagl *et al.*, 2023; Saira & Vauhkonen, 2024). On the other hand, amylases form the second largest enzyme segment and are expected to grow rapidly in the coming years, supported by the increasing demand for coatings with advantageous functional properties and the economic benefits of implementing on-site starch modification at production facilities (**Figure 2**) (Mondal *et al.*, 2022).

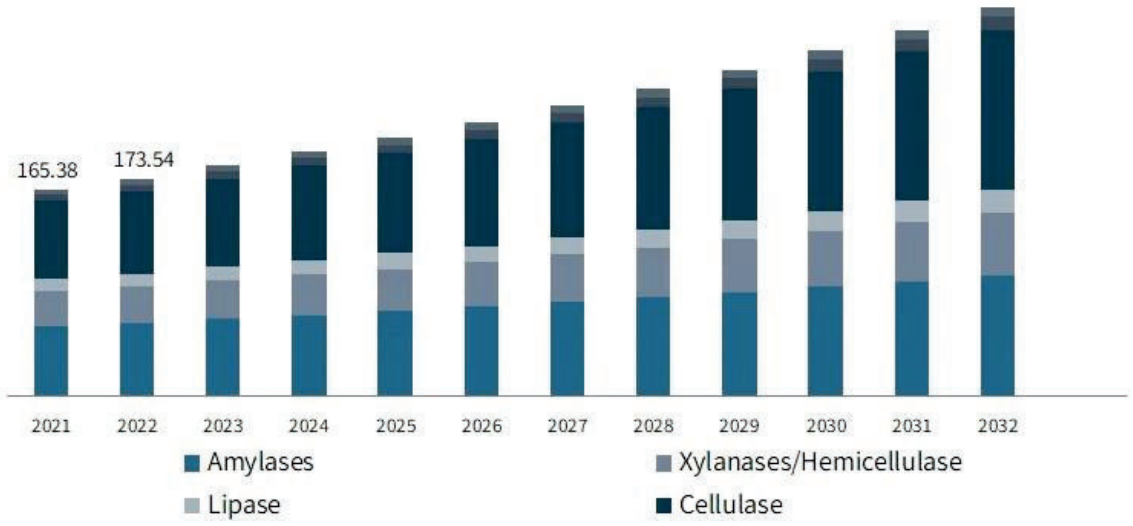


Figure 2. Projected market size of pulp and paper enzymes by product type from 2021 to 2032 (in USD million). Reprinted from *Pulp and paper enzymes market size, industry analysis, report 2032* (Global Market Insights, 2023).

Given their prominent roles in enhancing efficiency and sustainability in the PPI, this thesis focuses on the production of cellulases and amylases as key enzymatic solutions. However, the successful industrial implementation of enzymes requires not only the selection of suitable candidates for specific process conditions, but also their production at high titers. This often presents technical and economic challenges which, if not adequately addressed, may lead to

suboptimal performance or project failure. This work explores these constraints and provides solutions through the studies presented in the accompanying papers, which collectively contribute to the development of robust and scalable enzyme production strategies.

1.1.1. Cellulase application

Cellulose is the most abundant polysaccharide on Earth and the main component of cell wall plant biomass. It is a linear homopolymer of D-glucose units linked through β-1,4-glycosidic bonds. The hydroxyl groups on D-glucose units form strong intra- and intermolecular hydrogen bonds, along with Van der Waals interactions, leading to a tightly ordered crystalline cellulose, in contrast with the less organized amorphous regions within the cellulose matrix (Benalaya *et al.*, 2024). Complete hydrolysis of cellulose requires the cooperate action of several enzymes known as cellulases (**Figure 3**). First, endoglucanases (EGs) (EC 3.2.1.4) randomly cleave internal β-1,4 linkages within amorphous cellulose, releasing cello-oligosaccharides and exposing new chain ends for the action of cellobiohydrolases (CBH) (EC 3.2.1.91), which act as exo-glucanases. Second, CBHs progressively hydrolyze crystalline cellulose from both the reducing and non-reducing ends, thus releasing cellobiose as the main product. Finally, β-glucosidases (βG) (EC 3.2.1.21) completes the catalytic process converting cellobiose into glucose (Datta, 2024; Sutaoney *et al.*, 2024).

Mechanical refining is a critical unit operation in PPI that modifies the morphology and physicochemical structure of cellulose fibers to enhance their papermaking properties and improve the mechanical strength of the final product. Refining increases fiber flexibility and promotes pulp compaction, resulting in denser paper with lower bulk, opacity, and porosity (Motamedian *et al.*, 2019). During refining, fibers are subjected to random and repeated mechanical force (tensile, compressive, shear, and bending forces), which induces several structural irreversible changes: (i) external fibrillation, fibrils detach from the fiber surface but remain loosely attached, significantly increasing the available surface area for inter-fiber bonding; (ii) internal fibrillation, characterized by the disruption of hydrogen bonds between microfibrils, enhancing water uptake, fiber flexibility, and specific volume; (iii) generation of fines, which are small fiber fragments with high surface area that contribute to stronger fiber-fiber interactions; and (iv) fiber cutting and shortening, which enhances fiber swelling but, if excessive, can negatively impact fiber bonding and reduce the overall strength of the final paper product (Gharehkhani *et al.*, 2015; Motamedian *et al.*, 2019). The extent of these modifications depends on external factors such as fiber morphology, temperature, chemical environment, and

treatment conditions. These conditions, in turn, are influenced by equipment design and operating parameters, including pulp consistency and refining intensity (Bajpai, 2018).

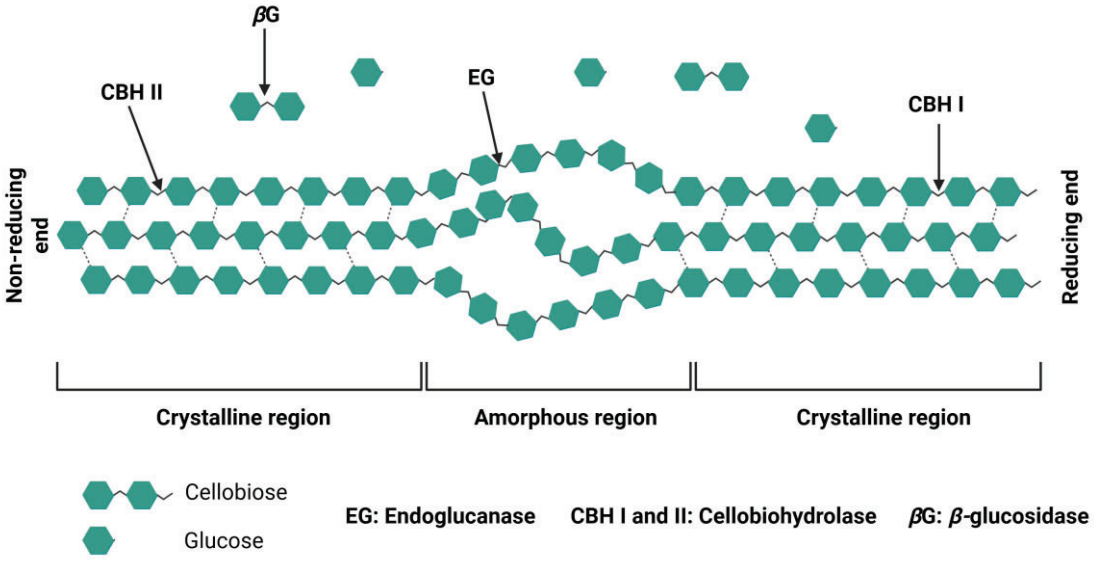


Figure 3. Schematic representation of cellulose hydrolysis by the synergistic action of cellulolytic enzymes. EGs randomly create nicks within the amorphous regions of cellulose, generating new ends. CBHs then bind to these exposed ends and degrade cellulose in a processive manner releasing cellobiose. Finally, the released cellobiose is converted into glucose by βGs, completing the saccharification process. Figure created with Biorender.

Mechanical refining accounts for up to 30-50% of the total energy consumed in papermaking, making it one of the most energy-intensive steps in the process, with energy demands ranging from 150 to 500 kWh per ton of paper (Kumar *et al.*, 2021; Saira & Vauhkonen, 2024). On the other hand, refining significantly impacts pulp drainage and the dewatering performance on the paper machine. Poor drainability can reduce machine runnability by increasing the time required for water removal, ultimately raising drying energy consumption and operational costs (Austin *et al.*, 2011). To address these challenges, enzymatic pre-treatment of pulp with cellulases prior to mechanical refining can improve energy efficiency. Studies have shown that cellulase application can reduce refining energy consumption by up to 40%, thereby contributing to more sustainable papermaking (Nagl *et al.*, 2023; Saira & Vauhkonen, 2024). Among cellulases, EGs are the most effective for this purpose, as they introduce weak points within cellulose fibers and enhance fiber fibrillation. This enzymatic action reduces the mechanical energy required to achieve comparable paper properties (Nagl *et al.*, 2021, 2023). Conversely, CBHs and βGs are generally not recommended for pulp treatment, as their hydrolytic activity on crystalline

cellulose can lead to excessive fiber degradation and reduced product quality (Nagl *et al.*, 2021; Wang *et al.*, 2017). Moreover, the specific demands of the PPI, such as high processing temperatures and alkaline pH, require enzymes with robust activity and stability under harsh conditions. Enzymes with low thermal tolerance, limited pH range, or poor catalytic efficiency may underperform or fail entirely in industrial settings. Therefore, identifying novel cellulases with enhanced thermostability, alkaline tolerance, and catalytic effectiveness is critical to optimizing their industrial application (Yang *et al.*, 2023).

In line with this objective, a key aim of the present thesis is the discovery of novel alkaline-tolerant and thermostable EGs suitable for refining. To support this goal, **Chapter 1** describes a database mining approach used to identify putative EGs with potential application in enzymatic refining.

1.1.2. Amylase application

Starch is one of the most abundant carbohydrate polymers in nature and represents the third most prevalent raw material in the paper, following cellulose fiber and mineral filler (Wang *et al.*, 2022). Most of the starch used in PPI is for surface sizing, a process that enhances surface strength, reduces linting, and improves the writing and printing quality of paper. Additionally, starch is widely employed as paper coating, where it acts as a binder for pigments and adhesives. These coatings contribute to the paper's whiteness, brightness, gloss, opacity and surface smoothness (Li *et al.*, 2019; Wang *et al.*, 2022).

Starch is a homo-polysaccharide mainly constituted by amylose and amylopectin. Amylose is a linear chain consisting of repetitive glucose units linked by α-1,4-glucosidic linkages, while amylopectin is highly branched, containing both α-1-4 linkages and α-1-6-glucosidic linkages in the branching points (Zhiguang *et al.*, 2025). Starch granules exhibit a semi-crystalline structure with varying degrees of crystallinity, primarily due to the organized radial arrangement of amylopectin, whereas the amorphous regions are predominantly composed of amylose. This crystalline organization renders starch insoluble in cold water, limiting its direct industrial use (Lacerda *et al.*, 2024). To address this limitation, starch must undergo gelatinization, a process in which starch granules are heated in water (typically above 50 °C). This disrupts the crystalline nature of their structure, leading to granule swelling, rupture, and a marked increase in viscosity, thus posing further challenges in surface sizing application (Lacerda *et al.*, 2024; Zhiguang *et al.*, 2025). To overcome the viscosity-related limitations of native starch, modification strategies are essential, particularly to reduce its molecular weight and, consequently, lower viscosity to levels suitable for surface sizing and paper coating. Among the commonly employed methods,

enzymatic breakdown and chemical treatments are the most prevalent. However, due to economic and environmental considerations, enzyme-catalysed hydrolysis is often preferred, especially in applications involving food-contact or medical-grade materials, where the use of milder and non-toxic processes is critical (Li *et al.*, 2019; Wang *et al.*, 2022).

Starch hydrolysis is catalyzed by amylases, which are of great significance for biotechnology and represent about 25% of the global enzyme market (Mondal *et al.*, 2022). Based on their mode of action, amylases are classified into four groups: i) α-amylase (EC 3.2.1.1) is typically a calcium-dependent metalloenzyme that randomly catalyses the endohydrolysis of internal α-1,4-glucosidic linkages in starch, resulting in a mixture of products composed of maltose (α-1,4-glucose dimers), maltotriose (α-1,4-glucose trimers), and branched oligosaccharides comprising 6–8 glucose units containing both α-1,4 and α-1,6 linkages. Notably, α-amylase cannot cleave terminal glucose residues or α-1,6-glycosidic bonds; ii) β-amylase (EC 3.2.1.2) is an exohydrolase enzyme that removes successive maltose units from the non-reducing end of starch polymers by cleaving α-1,4 linkages, but cannot bypass α-1,6-branch points; iii) γ-amylase (EC 3.2.1.3), commonly known as glucoamylase, hydrolyzes both α-1,4 and α-1,6 bonds, sequentially releasing glucose monomers from the non-reducing ends of amylose and amylopectin; and iv) debranching enzymes, such as pullulanases (EC 3.2.1.41), specifically hydrolyze α-1,6-glucosidic linkages, aiding in the complete depolymerization of branched starch structures (**Figure 4**) (Farooq *et al.*, 2021; Mondal *et al.*, 2022).

Among them, α-amylases are the main enzymes employed for producing enzyme-hydrolyzed starch in surface sizing applications, offering an environmentally friendlier alternative to chemical hydrolysis (Wang *et al.*, 2022). However, industrial starch processing requires heating to temperatures up to 80 °C to fully disrupt granular morphology and achieve gelatinization, thus highlighting the need for thermostable enzymes (Lacerda *et al.*, 2024). In this context, the α-amylase AmyQ from *Bacillus amyloliquefaciens* has been widely adopted for starch modification (Novozymes A/S, 2025; Wang *et al.,* 2023). Additionally, α-amylases can also be used as deinking agents in recycled paper processing, facilitating ink removal by hydrolyzing surface starch coatings and increasing cellulose fiber accessibility (Yang *et al.*, 2023). Given the potential role of α-amylases in the PPI, a major focus of this thesis is the optimization of AmyQ production in *Bacillus subtilis*. Specifically, **Chapters 2** and **3** explore strategies to enhance expression levels and secretion efficiency, alongside strain engineering strategies aimed at enabling cost-effective enzyme production.

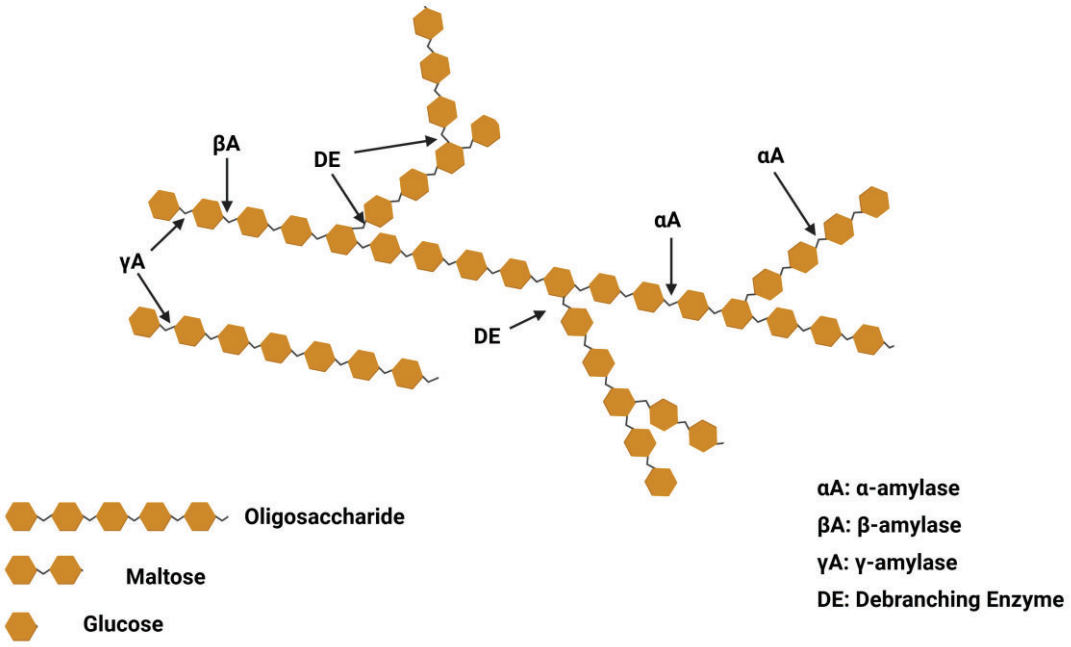


Figure 4. Scheme of starch enzymatic degradation via the synergistic action of α-amylase, β-amylase, γ-amylase and debranching enzymes. α-amylase randomly cleaves internal α-1,4-glycosidic bonds, generating shorter oligosaccharides. β-amylase acts from the non-reducing ends to release maltose units. γ-amylase hydrolyzes both α-1,4- and α-1,6-glycosidic bonds, yielding glucose units. Debranching enzyme (DE) acts on α-1,6 branch points, facilitating complete starch breakdown. Figure created with Biorender.

1.2. Challenges for enzyme implementation

Despite the well-recognized environmental and operational advantages of enzymes, their broader industrial adoption is often hindered by high production costs and technical limitations. These challenges are especially acute in cost-sensitive markets and developing economies, where enzyme prices can be prohibitive for small and medium-sized enterprises, hampering market penetration (Ferreira *et al.*, 2021; Pérez-Contreras *et al.*, 2024). The economics of enzyme production are commonly divided into capital expenditures (CAPEX) and operational expenditures (OPEX). CAPEX includes the initial investment for facility construction, with most costs deriving from equipment purchase and installation. Importantly, CAPEX is strongly influenced by the production rate and product titer, because lower productivities require larger equipment to meet the production goals. OPEX, on the other hand, comprises the recurring costs of plant operation; the largest contributors are feedstocks and ancillary chemicals (e.g., acids and bases), followed by utilities such as steam, cooling water, and aeration, as well as downstream-processing (de Lima *et al.*, 2022; Ferreira *et al.*, 2018; Khootama *et al.*, 2018).

Given the cost-sensitive nature of most industrial sectors, enzyme manufacturing must meet stringent cost-of-goods-sold (COGS) targets to remain competitive. Therefore, decreasing

production costs, evaluated through the titer, rate and yield metrics of the fermentation process, is essential to achieving economic viability (Konzock & Nielsen, 2024). Beyond cost, environmental impacts associated with enzyme production must also be addressed. This includes the entire life cycle, from raw material procurement to final product formulation. Significant inputs such as electricity, steam, feedstock ingredients, and water contribute heavily to both energy consumption and GHG emissions. Thus, a holistic view of the production process is crucial when the ultimate goal is to leverage enzyme technology to reduce global warming potential (GWP) and facilitate a transition from the traditional linear economy (make–use–dispose) to a circular model focused on resource efficiency (Hobusch *et al.*, 2024; Wowra *et al.*, 2023). For example, at industrial scale, Novozymes A/S reported that the fermentation step alone accounts for approximately 50% of the total environmental impact of enzyme manufacturing. Among all contributions, electricity usage and ingredient consumption were identified as the dominant sources of GWP (Nielsen, Oxenbøll, & Wenzel, 2007).

To address these multifaceted limitations, it is critical to optimize the cellular machinery responsible for gene expression, ensuring that engineered microbial strains achieve high enzyme titers (**Chapter 2**). In this context, modern genome editing technologies, especially CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeat-CRISPR associated protein 9), have become invaluable tools for rapidly and precisely enhancing the productivity of industrial host strains in a time-efficient manner (**Chapter 2** and **3**).

Finally, enzyme functionality must be tailored to specific industrial applications. For instance, in the paper industry, the discovery and engineering of novel enzymes with improved substrate specificity, stability, and catalytic efficiency is crucial for further enhancing its cost-effectiveness, and to boost its performance in processes such as bleaching, deinking, and fiber modification (**Chapter 1**).

1.3. Enzyme discovery through database mining

The PPI functions in a complex setting where enzyme performance is affected by factors like pH, temperature, and pulp composition, making it essential to find enzymes that can tolerate these conditions (Yang *et al.*, 2023). Microbial enzymes, which account for 85% of commercial enzymes, are preferred due to their higher stability, better yields, easier optimization, and cost-effectiveness compared to plant- or animal-based enzymes (Liu & Kokare, 2023; Singh *et al.*, 2019).

Screening natural microbial diversity through culture-dependent microbiological methods is a conventional approach for identifying novel biocatalysts with desirable properties. This typically involves enrichment and isolation of microorganisms under selective conditions, followed by taxonomic classification and enzyme identification using molecular or computational approaches (**Figure 5**) (Zhu *et al.*, 2022). However, these methods seriously limit the scope of finding new enzymes given the fact that less than 1% of environmental microbes are cultivable using standard laboratory techniques, thus excluding the vast majority of microbial diversity (Pham & Kim, 2012). Conversely, advances in next-generation sequencing (NGS) technologies and metagenomics methods have enabled direct access to environmental DNA (eDNA), allowing exploration of previously inaccessible microbial communities that may harbour valuable enzymatic functions (Nayfach *et al.*, 2021; Quince *et al.*, 2017).

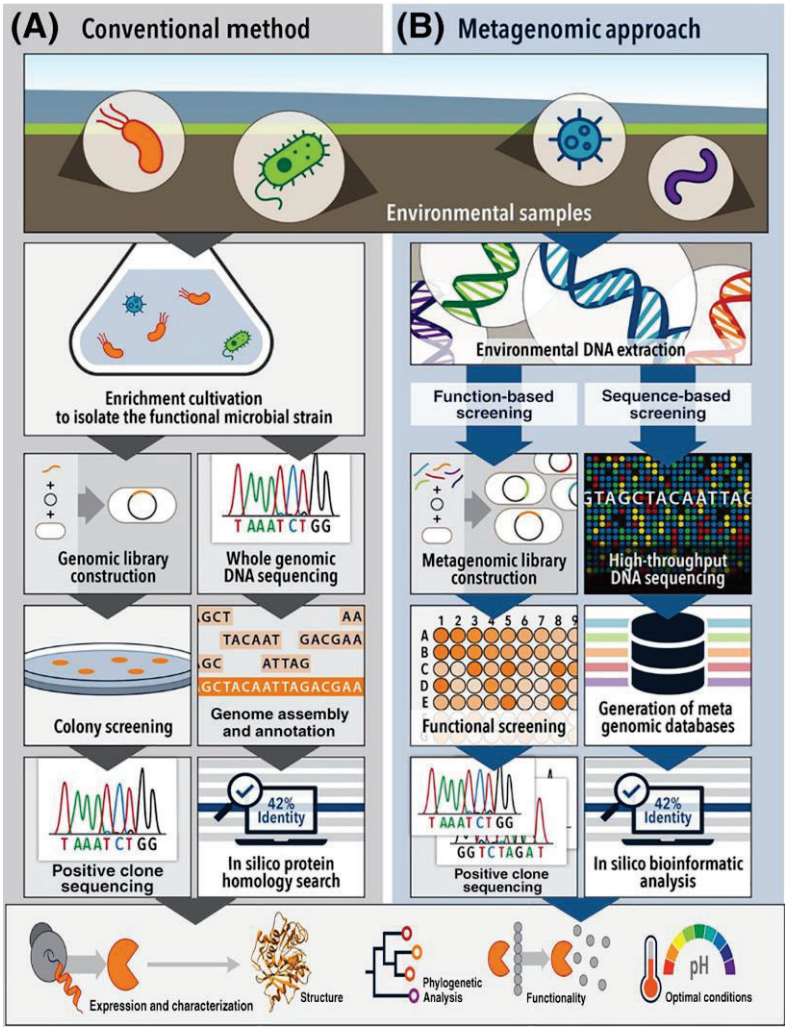


Figure 5. Schematic overview of conventional (A) and metagenomics-based (B) approaches for the discovery of novel enzymes. Conventional methods rely on cultivation, isolation, and screening of microorganisms, whereas metagenomics enables direct access to genetic material from environmental samples using functional- and sequence-based screening strategies. Adapted from Zhu *et al.*, 2022.

Generally, culture-independent strategies fall into two categories: i) functional-based screening (FBS), which involves the direct cloning of eDNA libraries into suitable expression systems for activity-based assays; and ii) sequence-based screening (SBS), which leverages bioinformatic tools to identify genes of interest based on sequence homology and functional annotation (**Figure 5**) (Ariaeenejad *et al.*, 2024). Contemporary enzyme discovery increasingly relies on *in-silico* analysis of sequences, including phylogenetic analyses, sequence similarity searches, genomic positional information, three-dimensional (3D) structural modelling, and machine learning predictions. These approaches enable rapid and cost-effective mining of metagenomic datasets for screening industrially relevant enzymes (Ariaeenejad *et al.*, 2024; Robinson *et al.*, 2021).

Despite the UniProt Knowledgebase cataloguing approximately 253 million protein sequences, fewer than 0.3% (around 500,000) have been manually curated, highlighting a substantial opportunity for novel enzyme discovery through database mining (Rhizobium, 2025). Major repositories supporting this effort include GenBank at the National Center for Biotechnology Information (NCBI), the European Molecular Biology Laboratory (EMBL) and the DNA Database of Japan (DDBJ) (Okido *et al.*, 2022; Sayers *et al.*, 2022; Thakur *et al.*, 2025). The database mining process can be broadly divided into two main steps: (i) identifying open reading frames (ORFs) annotated as putative enzymes; and (ii) screening, analysing and prioritizing candidates using bioinformatics tools (Kamble *et al.*, 2019). Sequence homology, conserved motifs, consensus patterns, and simple keyword-based searches are commonly used to retrieve relevant sequences. Candidates exhibiting ≤80% identity to known enzymes are typically prioritized, as they are more likely to represent novel biocatalysts while retaining essential catalytic features. Subsequent *in-silico* analysis typically encompass a myriad of complementary bioinformatics analyses to characterize physicochemical, phylogenetic and functional properties (Kamble *et al.*, 2019).

First, basic parameters such as molecular weight or theoretical pI can be calculated using tools like ExPASy ProtParam (Gasteiger *et al.*, 2003). Next, conserved domains and family assignments are identified via NCBI's Conserved Domain Database (CDD) and Pfam (Marchler-Bauer *et al.*, 2015). Homologous sequences can be retrieved with BLASTp, followed by multiple sequence alignment using MAFFT (Katoh *et al.*, 2002), MUSCLE (Edgar, 2004) or T-COFFEE (Notredame *et al.*, 2000) to highlight conserved residues and motifs critical for catalytic activity. Phylogenetic analysis using Molecular Evolutionary Genetics Analysis (MEGA) software place candidates within established enzyme families, providing insight into evolutionary relationships and functional predictions (Tamura *et al.*, 2021). Finally, 3D structural modelling using SWISS-Model

(Waterhouse *et al.*, 2018), Phyre2.2 (Powell *et al.*, 2025), or AlphaFold2 (Mirdita *et al.*, 2022) allows structural comparison with known enzyme. These models can be superimposed using PyMOL or Chimera to validate active-site residues and gain mechanistic insights (Rosignoli & Paiardini, 2022). This integrated computational pipeline enables the rapid prioritization of the most promising enzyme candidates for subsequent experimental validation and was therefore adopted in **Chapter 1**.

1.4. Recombinant protein production hosts

Microorganisms represent the dominant source of industrial enzymes, accounting for over 85% of the global enzyme market revenue. Their widespread adoption is largely attributed to key advantages including rapid growth, ease of cultivation and genetic manipulation, broad environmental availability, and simplified downstream processing (Liu & Kokare, 2023; Singh *et al.*, 2019). Ideally, a microbial production host should grow rapidly at moderate temperatures, produce high enzyme titers, and utilize inexpensive substrates efficiently. However, identifying a naturally occurring strain that meets all these requirements remains challenging. Many wild-type strains exhibit limitations such as slow grow rate, low enzyme productivities, or the generation of undesirable by-products that complicate downstream processing and compromise industrial scalability. As a result, strain improvement and host engineering are often necessary to optimize microbial systems for industrial enzyme production (Patel *et al.*, 2023).

Alternatively, a diverse array of microbial systems is employed for the heterologous expression of recombinant proteins, ranging from eukaryotic hosts such as yeast and filamentous fungi to prokaryotic organisms including Gram-positive and Gram-negative bacteria. The primary consideration in host selection is typically the origin of the target protein, whether prokaryotic or eukaryotic. Prokaryotic proteins are generally well-expressed in bacterial hosts, whereas eukaryotic proteins, particularly those requiring post-translational modifications (PTM) for proper folding, stability, or activity, are more effectively produced in yeast systems (Schütz *et al.*, 2023).

Noteworthy, a large fraction of industrial enzymes currently on the market are produced using hosts from the *Bacillus* genus, particularly *B. subtilis*. This prominence is mainly due to its genetic accessibility, excellent fermentation performance, and natural ability to secrete proteins into the extracellular medium, simplifying downstream processing (Herrmann *et al.*, 2024; Su *et al.*, 2020). Given its industrial relevance, *B. subtilis* has been selected as the main production host

in this thesis. **Chapters 2** and **3** are dedicated to investigating and optimizing its potential for efficient and high-yield recombinant enzyme production.

Nonetheless, several challenges such as limited PTMs, inadequate disulphide bond formation, poor secretion efficiency, protein misfolding or aggregation, and complex purification steps can hinder recombinant enzyme production in certain expression systems, particularly prokaryotic hosts (Schütz *et al.*, 2023). To bypass these limitations, the methylotrophic yeast *Komagataella phaffii* (formerly *Pichia pastoris*) has emerged as a preferred system for large-scale heterologous protein production. This yeast offers several advantages, including the ability to perform eukaryotic PTMs, efficient protein folding, and high-level secretion of recombinant proteins. Its well-developed genetic tools, combined with the capacity to grow to high cell densities, make this host an attractive and scalable platform for industrial enzyme manufacturing (Barone *et al.*, 2023; Duman-Özdamar & Binay, 2021). Consequently, *K. phaffii* was chosen in **Chapter 1** as the expression system for a functional fungal enzyme, enabling its subsequent biochemical characterization and application testing in the refining process.

1.4.1. *Bacillus subtilis*

B. subtilis is a low-G+C, Gram-positive, endospore-forming bacterium from the phylum *Bacillota* (formerly *Firmicutes*), predominantly found in soil and the plant rhizosphere (Earl *et al.*, 2008; Hohmann *et al.*, 2017). It has long been recognized as a model organism and is now widely used as a microbial workhorse in industrial biotechnology. In microbial fermentation, *B. subtilis* is employed for the large-scale production of heterologous proteins, predominantly enzymes, and value-added compounds such as carotenoids (Su *et al.*, 2020).

Several intrinsic features make *B. subtilis* particularly attractive for biotechnological applications: (i) its generally recognized as safe (GRAS) or qualified presumption of safety (QPS) status; (ii) rapid growth in low-cost media and the ability to reach high cell densities; (iii) efficient secretion of recombinant proteins into the culture medium, which simplifies downstream processing; (iv) lack of significant codon bias, allowing broad heterologous gene expression; and (v) a genetically well-characterized system that is easy to manipulate (Su *et al.*, 2020). Importantly, the high secretion capacity of *Bacillus* species and related genera has enabled the development of industrial strains capable of producing enzymes at gram-per-liter scales (Schallmey *et al.*, 2004; Su *et al.*, 2020). Additionally, *B. subtilis* stands out as one of the most well-studied and best-understood Gram-positive bacteria (Stülke *et al.*, 2023).

Within this species, strain 168 has emerged as a preferred chassis for recombinant protein production, synthetic biology, and microbial cell factory. It offers high transformation efficiency and a fully sequenced, well-annotated genome, making it particularly suitable for genetic engineering and pathway optimization (Zeigler *et al.*, 2008). Strain 168 is a domesticated derivative of the Marburg isolate (ATCC 6051), developed through mutagenesis to improve laboratory handling (**Figure 6**). While these modifications confer several benefits, such as reduced sporulation and enhanced transformability, they may also limit certain metabolic capabilities compared to wild-type strains (Chen *et al.*, 2023; Zeigler *et al.*, 2008). In fact, wild-type strains like ATCC 6051 and NCIB 3610 have shown higher protein production in some cases, although they retain undesirable traits for industrial use, such as excessive foaming, sporulation, high protease activity, and a ~1000-fold lower transformation efficiency than strain 168 (Zhang *et al.*, 2016).

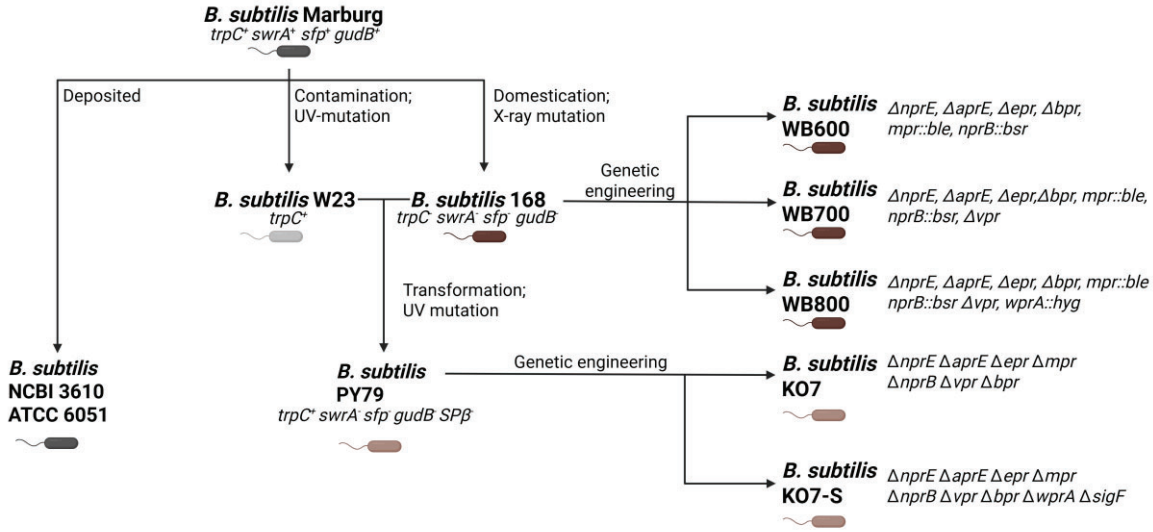


Figure 6. Genome heritage of some of the widely used *B. subtilis* strains in basic research and industrial applications. The figure illustrates the evolutionary relationships and genetic contributions from ancestral strains, highlighting the derivation of common laboratory and production strains from parental lineages. Genomic contributions are indicated by shading: black cells, original Marburg-like genomes; dark brown cells, 168-like genomes; light grey, W23 genome; light brown cells, 168-W23 hybrid genomes. Some key genetic differences include: mutations in *trpC* (indole-3-glycerol-phosphate synthase), *swrA* (activator of swarming motility), *sfp* (surfactin synthetase-activating enzyme), and *gudB* (glutamate dehydrogenase) genes; deletion of *sigF* (RNA polymerase forespore-specific sigma factor) and extracellular protease genes (*nprE, aprE, epr, mpr, nprB, vpr, bpr, wprA*); and curation of the *SPβ* prophage. Resistance markers include *ble* (bleomycin), *bsr* (blasticidin S), and *hyg* (hygromycin B). Modified from Chen *et al.*, 2023.

Several derivatives of 168, including WB600, WB700, and WB800, have been engineered by sequential deletion of extracellular protease genes to enhance recombinant protein stability in the culture medium (Wu *et* al., 2002; Zhang *et al.*, 2020). Other widely used strains, such as PY79, 3NA, and PS832, are hybrids incorporating genomic regions from the W23 strain (Zeigler *et al.*, 2008). Of particular interest is KO7-S, a non-sporulating derivative of PY79 lacking seven proteases and the sporulation-specific sigma factor *sigF* (Zeigler, unpublished data) (**Figure 6**).

Its domesticated background, high genetic accessibility, and well-documented performance in recombinant protein expression make this host a practical and effective chassis for industrial enzyme production. Building on this platform, the derivative strain KO7-S was employed for the expression of the model enzyme α-amylase AmyQ in this thesis (**Chapter 2** and **3**). KO7-S offers several other key advantages, including high transformability, the absence of extracellular proteases, inability to form spores, and patent-free status, all of which contribute to enhanced protein yield and process efficiency. However, selecting a suitable host is only the initial step. To meet the performance requirements of modern recombinant protein production, further strain engineering is essential. **Chapter 2** of this thesis focuses on optimizing key expression-related elements in KO7-S to improve productivity, stability, and scalability of the recombinant system.

1.4.1.1. Genetic manipulation of *Bacillus subtilis*

- **Integration vectors**

Classical chromosomal modification in *B. subtilis* typically involves the insertion of a selectable marker, such as an antibiotic resistance gene, via homologous recombination using integration vectors. These vectors, which remain essential from early molecular genetics to modern genomics, integrate into the chromosome through either single or double crossover events, depending on the number of homologous regions. This enables a range of genetic manipulations, including gene knockouts, gene-reporter fusions, and plasmid amplification (Arsov *et al.*, 2024; Dong & Zhang, 2014). However, the reliance on selectable markers restricts the number of feasible modifications due to limited available markers and regulatory concerns regarding antibiotic use, particularly in industrial settings. To address these limitations, marker recycling strategies have been developed, allowing the removal and reuse of selectable markers. These strategies primarily involve (i) counter-selectable markers (CSM), (ii) site-specific recombinase systems (SSR), and (iii) CRISPR-Cas9-based genome editing (Su *et al.*, 2020).

- **Counter-selectable markers systems**

CSM systems offer an efficient strategy for precise and scarless genome editing in *B. subtilis* by enabling selection for the loss rather than the presence of a genetic marker. These systems

typically involve integration of a construct containing a selectable marker (e.g., spectinomycin resistance) and a toxin or repressor gene, flanked by direct repeats (DRs). The construct is introduced at a target locus via homologous recombination under selection for the marker. Subsequent counter-selection eliminates cells retaining the cassette (Su *et al.*, 2020). In toxin-based systems, expression of a toxic gene, such as *upp* (Fabret *et al.*, 2002), *pyrF* (Suzuki *et al.*, 2012), and *mazF* (Zhang *et al.*, 2006), is conditionally induced, killing non-recombinant cells. Successful excision via recombination between DRs removes the cassette, preserving only the desired genomic alteration. Repressor-based systems use repressors like *xylR* (Jeong *et al.*, 2015), *blaI* (Brans *et al.*, 2004), *araR* (Liu *et al.*, 2008), or *lacI* (Zhang *et al.*, 2011) to suppress expression of a second selectable marker, which becomes active only upon cassette excision. While CSM strategies enable clean genomic modifications, their efficiency is often limited by leaky toxin/repressor expression and the need for prior strain engineering, making them labour-intensive and technically challenging (Dong & Zhang, 2014).

- **Site-specific recombinase systems**

SSR systems employ recombinases that mediate recombination between two site-specific recognition sites, enabling targeted DNA integration, deletion, or inversion (Su *et al.*, 2020). In *B. subtilis*, commonly used SSR systems include Cre/*loxP* from bacteriophage P1 (Yan *et al.*, 2008) and FLP/*FRT* from *Saccharomyces cerevisiae* (Chen *et al.*, 2010). These systems exhibit significantly higher recombination efficiency compared to endogenous mechanisms used in CSM, allowing efficient marker recycling and supporting multiple rounds of genome editing (Dong & Zhang, 2014). Nevertheless, SSR-based methods remain labor-intensive, often require multi-step protocols, and leave remnant sequences (scars) at the targeted site. As a more precise and versatile alternative, CRISPR-Cas systems, especially those using the Cas9 endonuclease, have been rapidly adapted for genome editing in *B. subtilis*, greatly expanding the molecular toolbox for high-resolution and scarless genetic manipulation (Altenbuchner, 2016).

- **CRISPR-Cas9 system**

CRISPR-Cas systems are adaptive, heritable immune mechanisms found in approximately 40% of bacteria and 90% of archaea, where they target and cleave invading genetic elements such as viruses and plasmids (Makarova *et al.*, 2020; Mojica *et al.*, 2005). These diverse systems can be broadly classified into two classes based on the configuration of their effector complexes: Class I employs multi-protein assemblies, whereas Class II relies on a single-component effector protein (Makarova *et al.*, 2020). Due to its simplicity and programmability, the Class II, type II CRISPR–Cas9 system from *Streptococcus pyogenes* has become a widely used platform for genome engineering (Doudna & Charpentier, 2014).

The CRISPR–Cas9 immune response involves three main stages: adaptation, expression and interference (**Figure 7**). During adaptation, a Cas protein complex recognizes a short protospacer-adjacent motif (PAM: NGG for class II type II) and cleaves a 20-nucleotide fragment of the invading DNA (known as protospacer) (Figure 7, step 1). This fragment is integrated into the CRISPR array (Figure 7, step 2). In the expression phase, the CRISPR array is transcribed into pre-crRNA, which base-pairs with trans-activating crRNA (tracrRNA), forming a dual-RNA structure (tracrRNA:crRNA), which is subsequently processed by Cas9 and RNase III to generate the mature guide RNA (gRNA) complex (Figure 7, steps 3 and 4). Finally, during interference, the gRNA guides Cas9 to a complementary DNA target flanked by the PAM sequence, enabling site-specific cleavage via a double-strand break (DSB) (Figure 7, steps 5 and 6) (Lander, 2016; Makarova *et al.*, 2020).

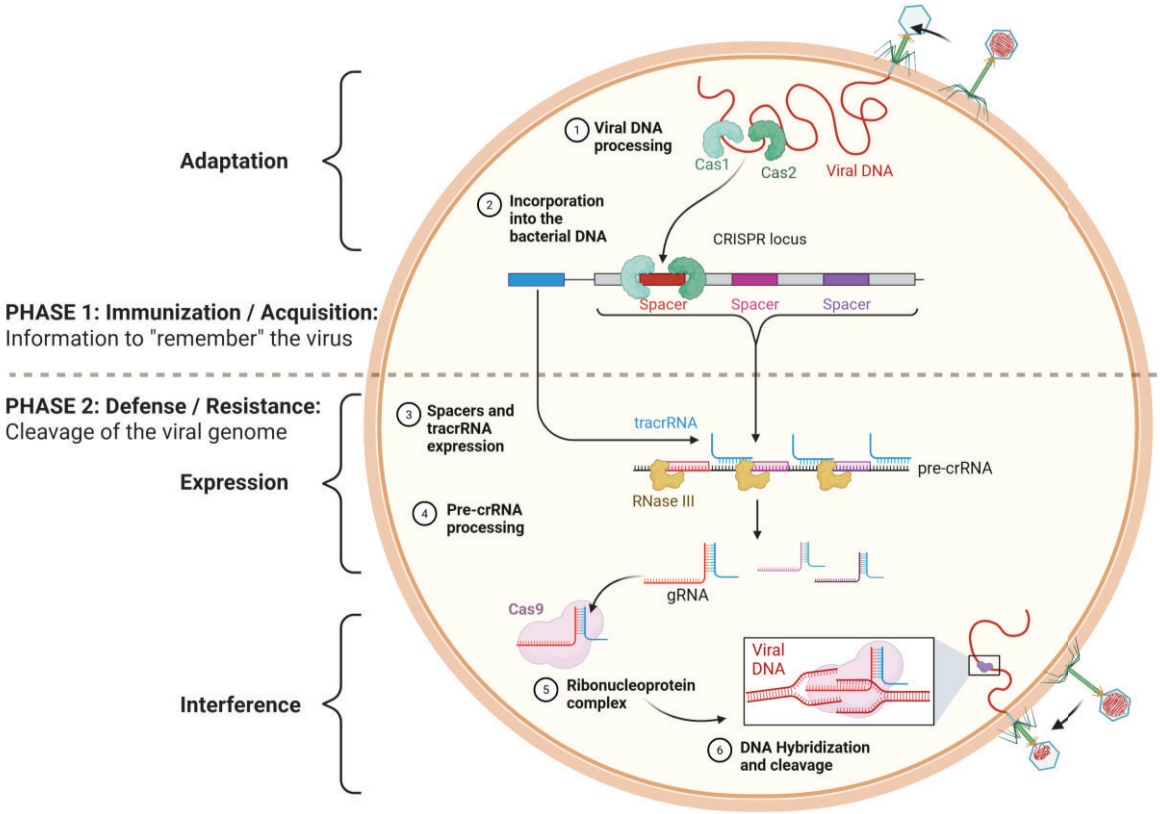


Figure 7. Schematic representation of the CRISPR-Cas9 adaptive immune system in *Streptococcus pyogenes*. Created by González Moraga, 2025.

A breakthrough in CRISPR-Cas9 genome editing was the development of a single-guide RNA (sgRNA) chimera, which combines the functions of crRNA and tracrRNA into a single molecule.

This innovation simplified the system by allowing target specificity to be reprogrammed through substitution of the 20-bp protospacer sequence at the 5' end of the sgRNA (Jinek *et al.*, 2012). Since DSBs represents one of the most precarious DNA lesions that can cause genomic instability and cell death, organisms have evolved two main repair pathways to preserve genome integrity: non-homologous end joining (NHEJ) and endogenous homology-directed repair (HDR). NHEJ is an error-prone process that often introduces random insertions or deletions (indels) at the repaired sites, while HDR enables precise, high-fidelity repair using donor DNA templates flanked by homology arms (**Figure 8**) (Song *et al.*, 2022). Importantly, due to the limited efficiency of NHEJ in *B. subtilis*, HDR has become the preferred approach for CRISPR-Cas9-mediated genome editing in this organism (Toymentseva & Altenbuchner, 2019).

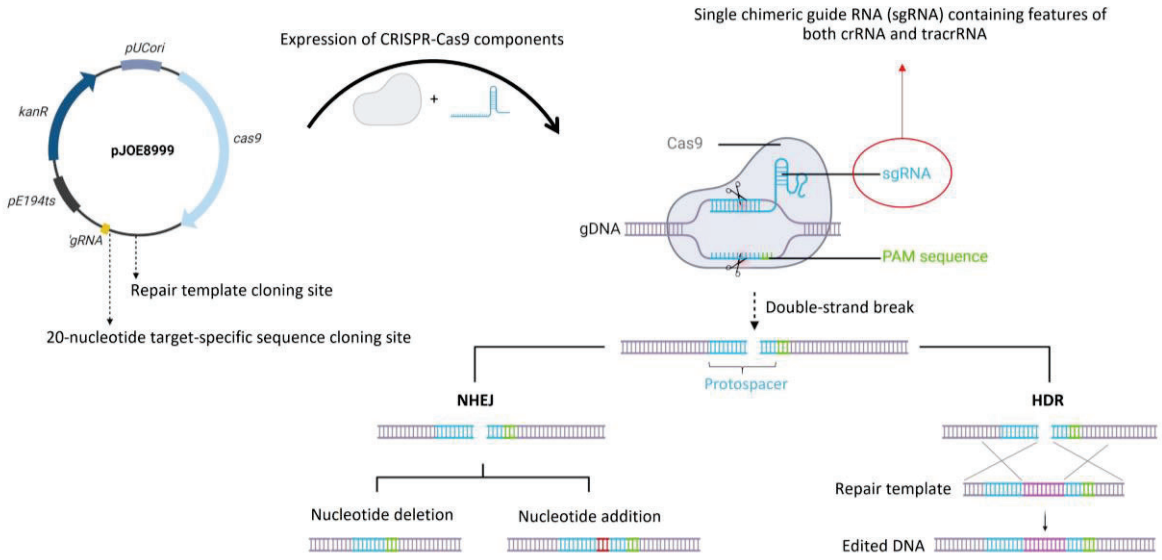


Figure 8. Schematic representation of the CRISPR-Cas9 vector pJOE8999, a single-plasmid system used in this thesis. Upon expression, Cas9 and sgRNA form a complex that specifically recognizes the target DNA sequence (protospacer). This is only possible if this sequence is followed by the Protospacer Adjacent Motif (PAM). Cas9 binding induces a dsDNA break, which is subsequently repaired by either non-homologous end joining (NHEJ) or homology-directed repair (HDR). Created with Biorender.

At present, three main CRISPR-Cas9-based genome editing strategies are widely used in *B. subtilis*: (i) The single-plasmid based system, in which Cas9, sgRNA, donor DNA, and other elements are assembled within a single vector; (ii) The two-plasmid-based system, where Cas9, sgRNA, and donor DNA are assembled on two separate plasmids, one expressing Cas9, and the other delivering the sgRNA and donor DNA template; (iii) The chromosomally integrated system, in which Cas9 is stably integrated into the host genome (Su *et al.*, 2020). In this thesis, the single-

plasmid based system was chosen due to its simplicity and ease of implementation (Altenbuchner, 2016), as depicted in **Figure 8**.

Recent advances in CRISPR-Cas9 technology have significantly enhanced the genetic engineering capabilities in *B. subtilis*. Notably, the development of rapid, all-in-one plasmid CRISPR-Cas9 systems, featuring self-curing plasmids, has accelerated iterative genome editing (Zou *et al.*, 2022). Additionally, Cas9 nickase variants have improved specificity and reduced cytotoxicity, allowing more precise iterative modifications (Liu *et al.*, 2019). The adoption of the CRISPR-Cpf1 (Cas12a) system further expanded the editing toolkit, owing to its intrinsic ability to process crRNA arrays, making it particularly well-suited for multiplex gene editing (Wu *et al.*, 2020). Despite these advances, most current systems still rely on complex vector assembly or sequential editing cycles. Efficiency tends to decline when multiple genomic loci are targeted simultaneously, resulting in prolonged screening and validation processes. Consequently, iterative genome modifications remain a daunting task, and continued refinement is needed to develop robust, user-friendly platforms capable of high-throughput, scarless, and simultaneous multilocus integration in *B. subtilis* in a timely fashion (**Chapter 3**).

1.4.1.2. Enhancing gene expression in *Bacillus subtilis*

The cellular abundance of a given protein is shaped by the integrated regulation of multiple molecular processes operating across distinct stages of gene expression. At the transcriptional level, DNA copy number and promoter strength primarily determine mRNA transcript levels. Post-transcriptionally, mRNA stability influences transcript longevity, thereby affecting the pool of templates available for translation. At the translational level, the ribosome binding site (RBS) sequence plays a central role, as it determines how frequently ribosomes initiate protein synthesis (Balakrishnan *et al.*, 2022; Nieuwkoop *et al.*, 2020; Salis *et al.*, 2009).

- **Promoter**

In bacterial systems such as *B. subtilis,* one of the key factors for achieving high-level expression of heterologous genes is the use of a strong and tightly regulated promoter. Importantly, RNA polymerases (RNAP) and associated sigma (σ) factors recognize the promoter and are recruited by the binding of regulatory proteins to specific sites within promoters to initiate transcription. In this regard, three types of promoter have been commonly used for the high-level expression of heterologous proteins in *B. subtilis*: constitutive promoters, inducible promoters and auto-inducible promoters (Schumann, 2007; Yu *et al.*, 2015).

Inducible promoters are widely used in *B. subtilis* for controlled gene expression, particularly when expressing toxic proteins. These systems respond to compounds such as sugars, e.g., sucrose (Liu & Du, 2012), maltose (Yue *et al.*, 2017) and xylose (Bhavsar *et al.*, 2001) or isopropyl-β-D-1-thiogalactopyranoside (IPTG) (Phan *et al.*, 2012). However, reliance on chemical inducers increases production costs and complicates large-scale fermentation, limiting industrial scalability (Yu *et al.*, 2015). To overcome these limitations, inducer-free alternatives have been explored, including promoters responsive to environmental cues such as pH (Atalla & Schumann, 2003), oxygen (Hoffmann *et al.*, 2021) or temperature (Huang *et al.*, 2013). Among these, constitutive and auto-inducible promoters are particularly attractive for industrial use due to their operational simplicity. Constitutive promoters enable continuous gene expression throughout fermentation without inducers, while auto-inducible promoters activate expression during specific growth phases (Yang *et al.*, 2017; Yu *et al.*, 2015). Promoters active from the late log to stationary phase are especially useful, enabling efficient, low-cost expression of heterologous and potentially toxic proteins (Yu *et al.*, 2015). Additionally, utilizing an inducer-independent promoter offers a simpler and more robust technical set up for the production line as there is no need for an additional feed line for inducer addition during fermentation. In this context, several strong, inducer-independent promoters have been widely applied in *B. subtilis*, including P_{aprE} (Chen *et al.*, 2015), P_{amyE} (Guan *et al.*, 2016), P_{hag} (Li *et al.*, 2022), P_{spoVG} (Li *et al.*, 2022), P_{43} (Liu *et al.*, 2019), P_{veg} (Yao *et al.*, 2023), P_{ylb} (Yu *et al.*, 2015). Notably, P_{spoVG} has shown superior performance, with 1.9-fold higher expression than the commonly used P_{43} (Liu *et al.*, 2018; Yang *et al.*, 2017; Yao *et al.*, 2023). Additionally, heterologous promoters such as P_{HpaII} from *Staphylococcus aureus* (Mu *et al.*, 2018) and P_{amyQ} from *B. amyloliquefaciens* (Kawabata *et al.*, 2012) have been utilized to further enhance expression levels in *B. subtilis*.

Specifically, modifying the native P_{amyQ} promoter to match consensus -10 and -35 sequences resulted in a 4-fold increase in promoter activity (Cheng *et al.*, 2016; Widner *et al.*, 2000). An additional strategy to boost expression involves constructing tandem promoters, where multiple promoter elements are arranged sequentially to initiate transcription from several start sites (Liu *et al.*, 2018; Yao *et al.*, 2023). Notably, when the optimized P_{amyQ} was combined with the *cry3A* gene promoter (derived from *Bacillus thuringiensis*) to create the tandem promoter P_{amyQ}-P_{cry3A}, saturating levels of mRNA could be achieved from a single-copy gene on the chromosome (Widner *et al.*, 2000). Notably, the high activity of the *cry3A* promoter in *B. subtilis* has been attributed to the presence of a Shine-Dalgarno-like sequence (termed as STAB-SD), which contributes to significantly increase mRNA half-life (Widner *et al.*, 2000).

The work presented in **Chapter 2** utilizes a strong constitutive promoter system consisting of the consensus α-amylase promoter from *B. amyloliquefaciens* P_{amyQ} and the RNA-stability enhancing promoter from the IIIA crystal protein of *B thuringiensis* (P_{cry3A}). These two elements were arranged in a tandem to form the P_{amyQ}-P_{cry3A} promoter system, as originally described by Widner *et al.* (2000), which demonstrated high efficiency in driving protein expression and contributed to the development of a super-secreting *B. subtilis* strain. **Chapter 2** also includes a comparative analysis between P_{amyQ}-P_{cry3A} and the strong constitutive promoter P_{spoVG}. Furthermore, a novel triple tandem promoter (P_{spoVG}-P_{amyQ}-P_{cry3A}) was constructed and characterized, combining the strength of all three elements to enhance transcriptional output.

- **Gene copy number**

Once a suitable promoter has been chosen, maximizing enzyme secretion in *B. subtilis* typically requires increasing gene copy number (**Figure 9**). The higher the copy number, the greater the expression, until a threshold is reached, beyond which no further increase is observed (Watzlawick & Altenbuchner, 2019). Presumably, this is likely due to saturation of the post-transcriptional machinery, shifting the bottleneck downstream of transcription (Widner *et al.*, 2000). To achieve gene amplification, two main strategies are commonly employed: expression from autonomous plasmids, and chromosomal integration of multiple expression cassettes (Watzlawick & Altenbuchner, 2019). Replicative plasmids allow high expression levels, proportional to plasmid copy number (Schumann, 2007), but suffer from drawbacks such as plasmid instability and the need for antibiotic selection, both of which pose challenges for industrial use due to regulatory and environmental concerns.

In contrast, chromosomal integration is preferred for large-scale applications due to its genetic stability and antibiotic-free maintenance. While double-crossover integration provides stable insertions without antibiotic usage, it often results in low gene dosage unless multiple integration events are performed (Huang *et al.*, 2017; Yomantas *et al.*, 2011). Single-crossover vectors can increase gene dosage under antibiotic pressure, but the inserted copies are often unstable without continuous selection (Young, 1984). Recent advances in CRISPR/Cas9 genome editing have enabled the development of marker-free, plasmid-less *B. subtilis* strains through iterative chromosomal integration of multiple expression cassette copies. This approach combines genetic stability with high expression potential, eliminating the need of antibiotic resistance markers and offering a more environmentally sustainable approach (Watzlawick & Altenbuchner, 2019). Accordingly, this method was adopted in **Chapter 2** to maximize gene expression following promoter optimization. Nevertheless, because it relies on multiple rounds

of integration, this method remains tedious and time-consuming, highlighting the need for further refinement (an issue explored in **Chapter 3**).

1.4.1.4. Optimization of secretion machinery

Achieving high production titers in protein production process requires that all stages of gene expression function efficiently (**Figure 9**). Even with maximized transcription, protein yield can still be limited by downstream bottlenecks in the secretory pathway, often the next major challenge in high-level protein production. *B. subtilis* is known for its strong secretion capacity. However, heterologous proteins are usually secreted at lower levels than native proteins, making it crucial to address limitations within the secretion and folding machinery (Zhang *et al.*, 2020). *B. subtilis* utilizes at least four classical protein secretion pathways: the general secretion (Sec) pathway, the twin-arginine translocation (Tat) pathway, the pseudopilin export (Com) pathway, and the ATP-binding cassette (ABC) transporter system (Fu *et al.*, 2007). Among these, the general secretion (Sec) pathway is the main transport channel, responsible for translocating a broad range of proteins. The Tat pathway, though less commonly used, offers unique advantages due to its distinct translocation mechanism and components, making both Sec and Tat signal peptides (SPs) valuable tools for recombinant protein secretion (Zhang *et al.*, 2020). Despite structural differences, SPs for both pathways share a common architecture: a positively charged N-terminal region (N-region), a central hydrophobic domain (H-region) and a polar C-terminal region (C-region) containing a cleavage site for signal peptidase. Specifically, Tat-specific SPs are distinguished by a twin-arginine motif (RR) in the N-region, longer N-domains, and H-domains that lack helix-breaking residues (e.g. glycine or proline), which are common in Sec-dependent SPs (Freudl, 2018). These differences help direct proteins into the appropriate secretion machinery. Additionally, SPs from closely related species, such as *Bacillus licheniformis* and *B. amyloliquefaciens*, can also be effectively used to enhance secretion in *B. subtilis* (Fu *et al.*, 2007; Zhang *et al.*, 2020).

In *B. subtilis*, the common secretory pathways generally involve three functional stages: (i) presecretory protein synthesis, (ii) interaction with chaperone proteins and engagement with the translocase complex, and (iii) transmembrane translocation, SP cleavage, protein release and folding, and final passage through the cell wall to complete secretion (Chen *et al.*, 2024).

In the Sec pathway, protein translocation occurs in an unfolded state and can proceed via two modes: cotranslational and posttranslational export. In the cotranslational mode, precursor proteins are translocated concurrently with their synthesis, a process primarily used for membrane proteins (Freudl, 2018).

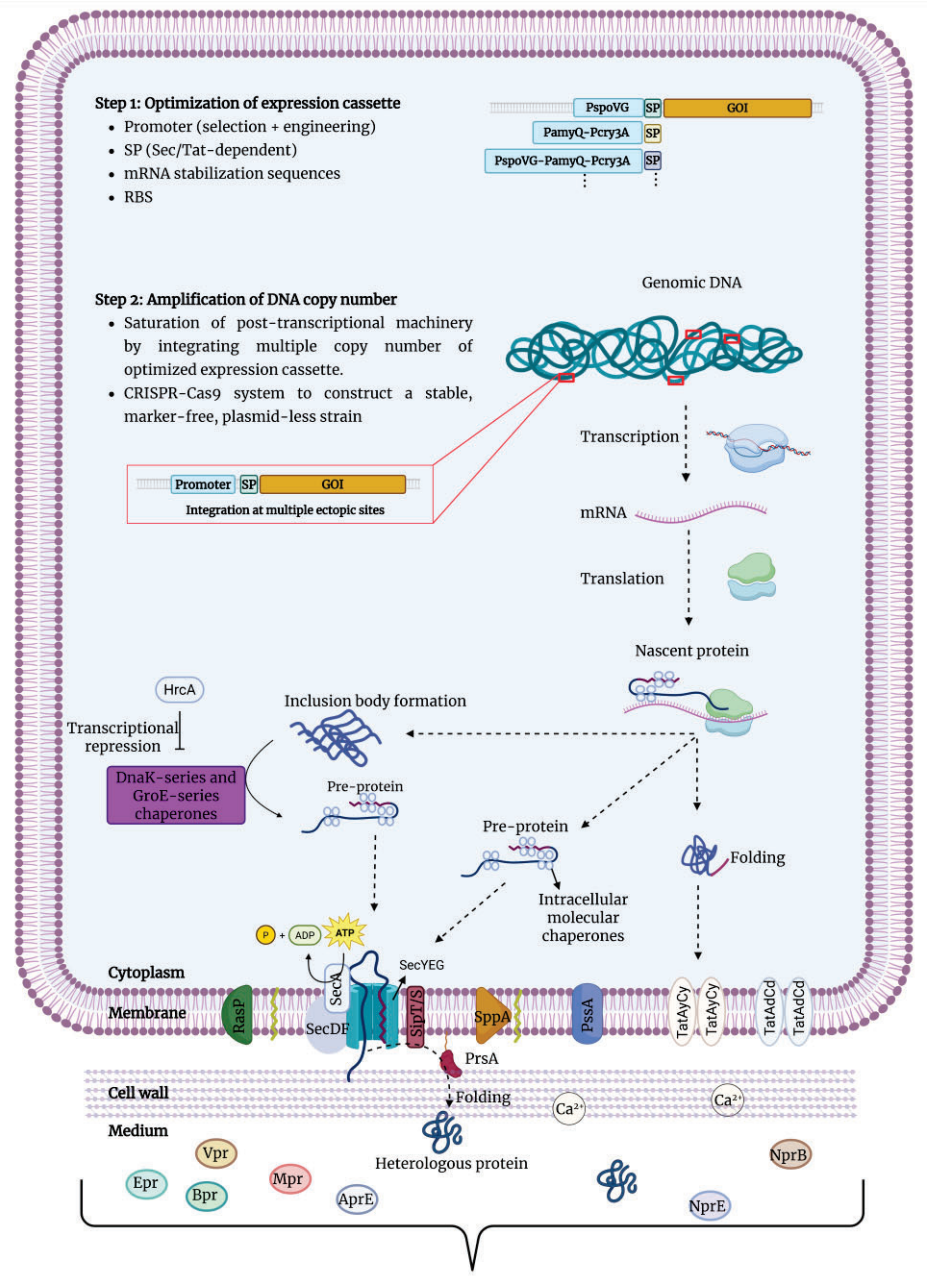


Figure 9. Framework for the systematic and comprehensive optimization of the *B. subtilis* expression system. The figure outlines key stages influencing recombinant protein yield, including expression cassette optimization, gene dosage strategies, and the major Tat- and Sec-dependent secretion pathways. Within the Sec-pathway, critical steps include preprotein translocation-competent state (DnaK- and GroE-series chaperones), targeting via signal peptides (SP), translocation (translocase subunits SecYEG, SecDF, and SecA), SP cleavage by signal peptidases (SipT, SipS), membrane clearance of SP remnants (signal peptide peptidases SppA and RasP), and extracytosolic folding assisted by the PrsA chaperone. The Tat-pathway is represented by TatAyCy and TatAdCd translocases. Additional engineering strategies include modulation of cell wall composition (e.g., phosphatidylserine synthase PssA) and deletion of extracellular proteases (Epr, Bpr, Vpr, Mpr, AprE, NprE, NprB). GOI = gene of interest. Created with BioRender.

In the posttranslational mode of the Sec pathway, precursor proteins are released from the ribosome in an unfolded, translocation-competent state. This state is maintained by general cytosolic chaperones such as the GroEL–GroES and DnaK–DnaJ–GrpE systems, which also prevent protein aggregation (Freudl, 2018). The unfolded preproteins are then delivered to the Sec translocase, where they are recognized by SecA, an ATP-dependent motor protein that translocates them stepwise through the SecYEG channel. The SecDF complex assists this process by exerting a proton motive force-driven pulling force from the periplasmic side of the membrane (Tsukazaki *et al.*, 2011). Efficient targeting to the translocase is critical, as preproteins may be poorly recognized or rejected if not properly directed. Therefore, the SP must both efficiently guide the protein to the translocase and be effectively cleaved by signal peptidases (Freudl, 2018).

Signal peptidase I, particularly SipS and SipT, mediates cleavage during the late stages of secretion. However, cleaved SP remnants remain embedded in the membrane and can interfere with protein transport. These remnants are subsequently degraded by specialized SP peptidases, including SppA and TepA, while RasP further facilitates membrane clearance by removing cleaved SPs and mislocalized precursor proteins (**Figure 9**) (Saito *et al.*, 2011; Yan & Wu, 2017; Zhang *et al.*, 2020).

After translocation across the cytoplasmic membrane, the extracytoplasmic folding factor PrsA plays a crucial role in assisting the proper folding of secreted proteins into their active conformations, thereby preventing aggregation and proteolytic degradation. PrsA functions as an efficient folding catalyst, particularly for exported amylases and other industrially relevant enzymes (Quesada-Ganuza *et al.*, 2019; Vitikainen *et al.*, 2001). Additionally, the efficiency of protein passage through the cell wall is influenced by its physical and chemical properties. High charge density and crosslinking in the peptidoglycan layer, primarily due to anionic polymers and lipoteichoic acids, attract cationic folding cofactors (e.g., metal ions) that can facilitate post-translocational folding (Harwood & Cranenburgh, 2008).

In contrast to the Sec pathway, the Tat pathway enables the secretion of fully folded proteins or multimeric enzyme complexes directly into the extracellular space (**Figure 9**). This capability may reduce proteolytic susceptibility and enhance enzyme yield (Zhang *et al.*, 2020). In *B. subtilis*, the Tat system comprises three TatA components (TatAc, TatAd, TatAy) and two TatC components (TatCd, TatCy), forming two distinct translocases: TatAdCd (induced under phosphate-limiting conditions) and TatAyCy (constitutively expressed and capable of broader substrate translocation) (Frain *et al.*, 2019; Zhang *et al.*, 2020). Notably, secretion pathway

specificity is dictated by the SP; thus, Sec-dependent proteins can be rerouted through the Tat pathway by replacing their native SP with a Tat-specific one, offering a versatile strategy to improve secretion performance (Liu *et al.*, 2014). In **Chapter 2**, the Tat pathway is evaluated as an alternative to the conventional Sec pathway to enhance secretion efficiency. Simultaneously, the highly efficient AmyQ SP was employed to direct protein export via the Sec pathway.

Despite significant advances in engineering *B. subtilis* as a microbial cell factory, critical bottlenecks along the protein secretion pathway, from transcription and translation to folding and export, remain insufficiently addressed. Many current strategies target isolated steps using plasmid-based systems, antibiotic selection, or costly inducers, which, while effective in laboratory settings, often lack scalability due to issues such as genetic instability, regulatory limitations, and elevated production costs. Consequently, a holistic strategy that systematically optimizes multiple layers of gene expression and secretion is essential for developing stable, high-performance strains suitable for industrial applications (**Chapter 2**).

Recent progress in genome editing, particularly through CRISPR-Cas9, combined with the deep functional understanding of *B. subtilis*, now enables the construction of plasmid-free, marker-free chassis strains with enhanced secretion capacity. These advances support the development of robust and scalable microbial platforms for industrial enzyme production. This thesis (**Chapters 2** and **3**) builds upon these innovations, aiming to contribute to the rational design of optimized *B. subtilis* strains tailored for high-yield, cost-effective biomanufacturing.

1.4.2. *Komagataella phaffii*

In 1954, Herman Phaff isolated a methylotrophic yeast from a black oak tree in Yosemite, later deposited as UCD-FST K-239 at UC Davis (Phaff, Miller, & Shifrine, 1956). Initially classified as *P. pastoris*, molecular taxonomy advances reclassified it as *Komagataella phaffii* in 2005 (Kurtzman, 2005). In the 1970s, Phillips Petroleum Company isolated a related strain, NRRL Y-11430, for methanol-based single-cell protein production, which was patented in 1980. This strain, likely a derivative from UCD-FST K-239, was later developed for recombinant protein expression by the Salk Institute and commercialized after its patent was acquired by Research Corporation Technologies in 1993 (Claes *et al.*, 2024). Widely used strains today, such as GS155 and X-33 from Invitrogen, trace their lineage to NRRL Y-11430. However, the majority of these strains, including those distributed by Invitrogen and the RCT Pichia expression system, are subjected to patents and material transfer restrictions that prohibit unlicensed commercial use, limiting their accessibility for cost-effective industrial applications (Claes *et al.*, 2024). To

overcome these barriers, license-free strains such as BG10, jointly developed by BioGrammatics and ATUM (formerly DNA2.0), offer a more accessible platform for commercial recombinant protein production (Juturu & Wu, 2018). Additionally, a genomic analysis recently confirmed that the Phaffi type strain UCD-FST K-239 is nearly identical to NRRL Y-11430. Based on this finding, an open-access *Pichia* chassis strain has been developed, enabling unrestricted, industry-standard recombinant protein production (Claes *et al.*, 2024).

K. phaffii is a methylotrophic yeast that has emerged as a leading platform for recombinant protein expression in both research and industrial settings (Moraes *et al.*, 2024). As a lower eukaryote, it combines advantageous traits of both prokaryote and eukaryote: the ease of cultivation, fast growth and highly scalable robust bioreactor processes of bacterial systems, along with the advanced protein secretion capabilities of eukaryotic cells (Barone *et al.*, 2023). Additional features of this yeast make it an ideal platform for recombinant protein production, including: (i) a well-characterized genetic background, complete genome sequence, and accessible molecular manipulation tools; (ii) proper protein folding and essential PTMs, including *O*- and *N*-linked glycosylation, and disulfide bond formation; (iii) robust secretory capacity with minimal endogenous protein secretion, facilitating high-yield expression and simplified downstream purification; (iv) its ability to grow to high cell densities in low-cost and renewable substrates; and (v) its GRAS or QPS status (Claes *et al.*, 2024; Ergün *et al.*, 2022; Karbalaei *et al.*, 2020; Lv & Cai, 2025). For all these reasons, *K. phaffii* BG10 was selected in **Chapter 1** for the heterologous expression of fungal enzymes.

1.4.2.1. *Komagataella phaffii* as a cell factory for recombinant proteins

K. phaffii is an attractive chassis to produce both high-value bioproducts (e.g., antibodies, hormones, vaccines) and low-value products (e.g., food and feed enzymes, processing aids). Successful recombinant protein expression in *K. phaffii* depends on the careful design of the expression system, including strain selection, a strong promoter to achieve high expression levels, gene dosage, codon optimization, and an effective secretory signal sequence. These elements are crucial for efficient protein expression and subsequent translocation and post-translational processing within the Endoplasmic Reticulum (ER) and Golgi apparatus (Juturu & Wu, 2018).

- **Promoter choice**

Promoter selection is critical in *K. phaffii* bioprocess development as it directly influences expression levels and dictates fermentation strategy (García-Ortega *et al.*, 2019). The two most widely used promoters are the strong methanol-inducible alcohol oxidase 1 promoter (P_{AOX1})

and constitutive glyceraldehyde 3-phosphate dehydrogenase promoter (P_{GAP}). Nonetheless, plenty of promoters are available in literature, which are also in continuous evolution (Çalik *et al.*, 2015).

K. phaffii metabolizes methanol as a carbon source via methanol utilization (MUT) pathway, which involves two alcohol oxidase *(AOX)* genes, *AOX1* and *AOX2*. *AOX1* encodes the major enzyme, which accounts up to 30% of the total soluble proteins in cells grown solely on methanol. Due to the high strength and regulation of its promoter, P_{AOX1} is currently the most widely used promoter in commercial protein production. Moreover, *AOX1* is repressed by most carbon sources other than methanol, ensuring high biomass accumulation before protein expression is induced (Barone *et al.*, 2023; Juturu & Wu, 2018).

However, methanol presents several drawbacks in industrial settings: (i) safety risks and expensive costs associated with storage and handling due to its flammability; (ii) its origin as a petrochemical, limiting alignment with circular economy principles; (iii) increased oxygen demand and heat generation during high-cell-density cultivation; (iv) formation of toxic by-products like formaldehyde and hydrogen peroxide; and (v) elevated risk of cell lysis and degradation of the product (Çalik *et al.*, 2015; García-Ortega *et al.*, 2019; Juturu & Wu, 2018).

As a result, there is growing interest in methanol-free expression systems. One such alternative is P_{GAP}, which drives the expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a highly expressed enzyme involved in glycolysis and gluconeogenesis. As a strong, constitutive promoter, P_{GAP} enables protein expression proportional to cell mass, making it ideal for producing non-toxic proteins that do not inhibit host cell growth (Çalik *et al.*, 2015). P_{GAP}-based systems typically use glucose or glycerol as carbon sources, though other substrates like fructose, sorbitol, mannitol, ethanol, and trehalose also support expression. While the carbon source can modulate expression levels, comparable yields have been reported with glucose and glycerol (García-Ortega *et al.*, 2019; Waterham *et al.*, 1997). Importantly, using renewable feedstock, such as glycerol from the biodiesel industry or corn steep liquor from the starch industry (Hahn-Hägerdal *et al.*, 2005; Jia Zheng *et al.*, 2012), further enhances the sustainability and economic feasibility of these systems (García-Ortega *et al.*, 2019). Overall, P_{GAP}-driven expression systems offer a methanol-free, scalable, and cost-effective alternative for large-scale recombinant protein production and, consequently, the enzyme expression strategy used in **Chapter 1** is based on P_{GAP}.

- **Integrative vectors and gene dosage**

The expression of any recombinant gene in *K. phaffii* has three phases: (i) Cloning the gene of interest into a suitable expression vector; (ii) insertion of the vector into the host genome; and (iii) screening of transformants to identify high expression strains (Karbalaei *et al.*, 2020). For stable expression, integrative vectors for P_{GAP}-driven expression systems are widely used. Among them, the pGAPZα vector series (Invitrogen, USA) are often employed for extracellular protein production. However, these vectors are protected by Intellectual Property (IP) rights, and their use in commercial manufacturing typically requires a license. Alternatively, IP-free expression systems such as those offered by ATUM.bio and the recently developed OPENPichia platform provide more accessible options for unrestricted use (Claes *et al.*, 2024; Juturu & Wu, 2018). After vector integration, gene dosage can be increased by selecting for clones that have incorporated multiple copies of the expression cassette. This is typically achieved by gradually increasing the concentration of the selective marker, allowing isolation of so-called "jackpot clone", which are the transformants with the highest expression levels (Naseem *et al.*, 2021).

- **Additional factors affecting protein expression in *K. phaffii***

An essential factor in maximizing heterologous protein expression in *K. phaffii* is codon optimization. Because codon usage can vary significantly between the host and foreign genes, the use of rare or unfavorable codons may lead to inefficient transcription or translation, and ultimately, low protein yields. Adapting the gene of interest to match the preferred codon usage of *K. phaffii* is therefore strongly recommended to enhance expression efficiency (Majeke *et al.*, 2020). Another key aspect is the secretory signal sequence**,** which directs the recombinant protein into the ER, the first of the yeast secretory pathway. Once inside the ER lumen, enzymes such as Kex2 and Ste13 usually cleave the SP (Juturu & Wu, 2018). Since the efficiency of translocation into the ER directly influences the overall secretion levels, numerous SPs have been developed for *K. phaffii* to support high-yield recombinant protein secretion. Nonetheless, the α-mating factor SP (α-MF) from *S. cerevisiae* is currently one of the most widely used SPs, and is commonly incorporated into commercial expression vectors (Lv & Cai, 2025). In **Chapter 1** of this thesis, *K. phaffii* was selected as the host for recombinant production of fungal enzymes due to the advantages discussed throughout this section. Expression was achieved using a P_{GAP}-driven system, along with codon optimization**,** gene copy number amplification**,** and the α-MF SP to maximize yield and secretion efficiency.

2 Aims and objectives

2. Aims and objectives

Enzymes have become transformative tools across a wide range of industrial applications. In the Pulp and Paper Industry, they are emerging as environmentally preferred alternatives to conventional chemical processes, offering the potential to reduce energy consumption, minimize water usage, and eliminate the need for harsh reagents. Despite these advantages, industrial implementation of enzymatic solutions remains limited by the need to identify process-compatible biocatalysts and enable their cost-effective, high-yield production at scale.

The work conducted in this thesis addresses these challenges through a two-pronged approach: (i) the discovery of novel enzymes adapted to the physicochemical conditions of the papermaking environmental via database mining and *in silico* screening, and (ii) the optimisation of enzyme expression in *B. subtilis,* a well-established industrial production host. To achieve this, the work includes the development of CRISPR-Cas9-based genome editing tools to streamline strain engineering, as well as a systematic strategy to enhance heterologous gene expression and protein secretion. The specific objectives of this thesis are as follows:

1- To identify novel EGs for application in the PPI:
   - 1.1 Use of database mining and *in silico* screening to rapidly prioritize enzyme candidates with potential relevance to papermaking conditions.
   - 1.2 Heterologous expression of the selected enzymes in *K. phaffii*, followed by purification and biochemical characterization of putative EGs.
   - 1.3 Performance assessment of promising enzyme candidates within the context of the industrial refining process used in papermaking.

2- To engineer a robust, super-secreting *B. subtilis* strain through systematic optimization of the secretory pathway:
   - 2.1 Construction of a stable, marker-free, plasmid-less strain using CRISPR-Cas9 mediated genome editing.
   - 2.2 Recombinant gene expression enhancement by optimizing promoter selection and increasing gene copy number.
   - 2.3 Identification and alleviation of rate-limiting steps in the protein secretion pathway to maximize extracellular enzyme yield.

3- To develop a novel CRISPR-Cas9-based genome editing strategy enabling one-step, simultaneous multi-gene integration in *B. subtilis* chromosome:
   - 3.1 Design and validation of a straightforward, colorimetric screening system for rapid identification of successful multi-locus insertions.

3.2 Assessment of the precision, efficiency, and genomic context-dependence of the CRISPR-Cas9 multiplex integration approach.

3.3 Leverage the colorimetric screening strains as a genetic platform for stable, chromosomal production of C_{30}-carotenoids in *B. subtilis*.

3 Reports

3. Reports

3.1 Report on the impact factor and participation in the publications

- Impact factor

The papers that are part of the doctoral thesis submitted by Jordi Ferrando Núñez have been published in international journals indexed in the Journal Citation Reports, as detailed below:

1. Ferrando, J., Pascual, C. L., Cusola, O., Roncero, M. B., Planas, A., & Picart, P. (2025). From mold to mill: StachCel5, a novel thermoalkaliphilic endoglucanase from *Stachybotrys chartarum* for pulp fiber biorefining. ***INTERNATIONAL JOURNAL OF BIOLOGICAL MACROMOLECULES***, *320*(Pt3), 145969. https://doi.org/10.1016/j.ijbiomac.2025.145969

This journal is included in the first quartile of the subject area: Biochemistry & Molecular Biology (29/319) and Polymer Science (6/94), with an impact factor of 8.5 for the year 2024.

2. Ferrando, J., Miñana-Galbis, D., & Picart, P. (2024). The Construction of an Environmentally Friendly Super-Secreting Strain of *Bacillus subtilis* through Systematic Modulation of Its Secretory Pathway Using the CRISPR-Cas9 System. ***INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES***, *25*(13), 6957. https://doi.org/10.3390/ijms25136957

This journal is included in the first quartile of the subject area: Biochemistry & Molecular Biology (72/319), with an impact factor of 4.9 for the year 2024 and a total of 3 citations (as of September 2025, Google Scholar).

3. Ferrando, J., Filluelo, O., Zeigler, D. R., & Picart, P. (2023). Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system. ***MICROBIAL CELL FACTORIES***, *22*(1), 21. https://doi.org/10.1186/s12934-023-02032-2

This journal is included in the first quartile of the subject area: Biotechnology & Applied Microbiology (41/174), with an impact factor of 4.3 for the year 2023 and a total of 18 citations (as of September 2025, Google Scholar).

4. Filluelo, O., Ferrando, J., & Picart, P. (2023). Metabolic engineering of *Bacillus subtilis* toward the efficient and stable production of C30-carotenoids. ***AMB EXPRESS***, *13*(1), 38. https://doi.org/10.1186/s13568-023-01542-x

This journal is included in the second quartile of the subject area: Biotechnology & Applied Microbiology (63/174), with an impact factor of 3.5 for the year 2023 and a total of 9 citations (as of September 2025, Google Scholar)."

- Participation

The doctoral candidate Jordi Ferrando Núñez has participated in the articles that are part of their doctoral thesis in the manner detailed below:

1. **Ferrando, J.**, Pascual, C. L., Cusola, O., Roncero, M. B., Planas, A., & Picart, P. (2025). From mold to mill: StachCel5, a novel thermoalkaliphilic endoglucanase from *Stachybotrys chartarum* for pulp fiber biorefining. *International Journal of Biological Macromolecules*, *320*(Pt3), 145969. https://doi.org/10.1016/j.ijbiomac.2025.145969

The PhD student carried out most of the experimental work, participated in the experimental design, and contributed to the writing of the original draft as well as the review and editing of the manuscript.

2. **Ferrando, J.**, Miñana-Galbis, D., & Picart, P. (2024). The Construction of an Environmentally Friendly Super-Secreting Strain of *Bacillus subtilis* through Systematic Modulation of Its Secretory Pathway Using the CRISPR-Cas9 System. *International Journal of Molecular Sciences*, *25*(13), 6957. https://doi.org/10.3390/ijms25136957

The PhD student carried out the experimental work, participated in the experimental design, and contributed to the writing of the original draft as well as the review and editing of the manuscript.

3. **Ferrando, J.**, Filluelo, O., Zeigler, D. R., & Picart, P. (2023). Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system. *Microbial Cell Factories*, *22*(1), 21. https://doi.org/10.1186/s12934-023-02032-2

The PhD student carried out most of the experimental work, participated in the experimental design, and contributed to the review and editing of the manuscript.

4. Filluelo, O., **Ferrando, J.**, & Picart, P. (2023). Metabolic engineering of *Bacillus subtilis* toward the efficient and stable production of C30-carotenoids. *AMB Express*, *13*(1), 38. https://doi.org/10.1186/s13568-023-01542-x

The PhD student carried out part of the experimental work and contributed to the review and editing of the manuscript.

None of the co-authors of the articles presented here have used the data described in these publications to prepare their doctoral theses.

Pere Picart

Firmado digitalmente por Pere Picart
Fecha: 2025.09.25 10:14:12 +02'00'

Dr. Pere Picart Faiget
Thesis director and supervisor

DAVID MIÑANA GALBIS - DNI 21651795D

Firmado digitalmente por DAVID MIÑANA GALBIS - DNI 21651795D
Fecha: 2025.09.23 17:01:53 +02'00'

Dr. David Miñana i Galbis
Thesis director

4 Publications

4. Publications

4.1. Chapter 1. Identification of new endoglucanases for their application in the Pulp and Paper Industry

4.1.1. Paper 1: From mold to mill: StachCel5, a novel thermoalkaliphilic endoglucanase from *Stachybotrys chartarum* for pulp fiber biorefining.

International Journal of Biological Macromolecules 320 (2025) 145969

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

From mold to mill: StachCel5, a novel thermoalkaliphilic endoglucanase from *Stachybotrys chartarum* for pulp fiber biorefining

Jordi Ferrando [a,*], Clàudia Lliso-Pascual [b], Oriol Cusola [c], M. Blanca Roncero [c], Antoni Planas [b,d], Pere Picart [a,*]

[a] Faculty of Pharmacy and Food Science Technology, Department of Biology, Healthcare and the Environment, Microbiology Section, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain
[b] Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, Via Augusta 390, 08017 Barcelona, Spain
[c] CELBIOTECH_Paper Engineering Research Group, Universitat Politècnica de Catalunya_BarcelonaTech, 08222 Terrassa, Spain
[d] Chemistry section, Royal Academy of Sciences and Arts of Barcelona, La Rambla 115, 08002 Barcelona, Spain

ARTICLE INFO

Keywords:
Stachybotrys
Genome mining
Cellulase
Biorefining
Pulp
Paper

ABSTRACT

StachCel5, a novel *endo*-β-1,4-glucanase from *Stachybotrys chartarum* IBT 7711, was identified through genome mining and exhibited 65 % sequence identity with known glycoside hydrolase family 5 (GH5) cellulases. The gene was codon-optimized, excluding its native signal peptide, and heterologously expressed in *Komagataella phaffii* under the control of constitutive glyceraldehye-3-phosphate dehydrogenase (GAP) promoter. The purified enzyme displayed a high specific activity of 287.6 U/mg on carboxymethyl cellulose (Na-CMC), retained over 87 % activity across pH 4–7, and maintained 77 % residual activity after 60 min at 50 °C. Structural modeling predicted a canonical $(\beta/\alpha)_8$ TIM-barrel fold, with Glu228 and Glu335 as conserved catalytic residues. Application trials on industrial hardwood and softwood pulps demonstrated that enzymatic pre-treatment with StachCel5 enhanced fiber fibrillation, reduced refining energy, and improved mechanical properties. In softwood pulp, the enzyme enabled equivalent tensile strength (67.8 Nm/g) as untreated controls with 25 % fewer refining revolutions (4500 vs 6000), reflecting significant energy savings. Improved tensile and burst indices, and preserved fiber integrity (zero-span tensile index) confirm the biorefining potential of StachCel5. These results establish StachCel5 as a thermostable and pH-tolerant biocatalyst suitable for enzymatic fiber modification in pulp and paper processing under industrially relevant conditions.

1. Introduction

Cellulose, the most abundant polysaccharide on Earth, constitutes a major component of plant biomass and is composed of β-1,4-linked D-glucose units forming crystalline and amorphous regions. Complete hydrolysis of cellulose requires the cooperate action of cellulases, namely endoglucanases, exoglucanases (cellobiohydrolases), and β-glucosidases [1]. These enzymes are pivotal in biomass valorization, contributing to circular bioeconomy efforts and sustainable industrial practices [2,3].

Cellulases have diverse applications across sectors, including food processing, animal feed, textiles, detergents, and notably, the pulp and paper industry [4,5]. In papermaking, mechanical refining modifies pulp fibers to enhance bonding and improve paper strength, yet it is highly energy-intensive, consuming 150–500 kWh/ton and contributing significantly to the carbon footprint of paper production [6,7]. Enzymatic pre-treatment with cellulases, particularly *endo*-β-1,4-glucanases (EC 3.2.1.4), has emerged as a promising strategy to reduce refining energy (up to 40 %) by enhancing fiber fibrillation and flexibility without over-degrading fiber integrity [7,8]. However, commercial deployment of cellulase-assisted refining faces several technical challenges. Many fungal endoglucanases display limited activity at neutral to alkaline pH and suffer from thermal instability under mill conditions, which typically operate at 40–60 °C and pH 5–8 [9,10]. Additionally, cellobiohydrolase and β-glucosidase activities are generally not considered beneficial for pulp modification, as their use may lead to excessive fiber degradation and compromise pulp strength, thus making enzyme selection and process tuning essential [11,12]. Therefore, identifying new cellulases with robust activity profiles, high substrate specificity, and compatibility with industrial conditions is critical for broader

* Corresponding authors.
E-mail addresses: jferranu7@alumnes.ub.edu (J. Ferrando), perepicart@ub.edu (P. Picart).

https://doi.org/10.1016/j.ijbiomac.2025.145969
Received 19 May 2025; Received in revised form 1 July 2025; Accepted 11 July 2025
Available online 12 July 2025

adoption.

Filamentous fungi remain a key source of industrial enzymes due to their natural ability to degrade lignocellulose and secrete diverse hydrolases [13]. Among them, *Stachybotrys chartarum* is a cellulolytic ascomycete frequently isolated from decaying cellulose-rich substrates, including paper waste and agricultural residues [14,15]. This species is known to produce thermostable and alkali-tolerant enzymes [16], yet its cellulolytic arsenal remains largely unexplored. To date, only a single lichenase from *S. chartarum* has been functionally characterized [17], despite the availability of complete genome sequences for multiple strains [18]. Recent advances in genome mining, protein modeling, and heterologous expression platforms now facilitate the discovery and engineering of novel fungal cellulases with customized properties [10,19].

In this study, we report the identification, recombinant production, and characterization of StachCel5, a novel GH5 *endo*-β-1,4-glucanase from *S. chartarum*. Using a genome mining approach, we selected a gene encoding a putative GH5 enzyme with a fungal cellulose-binding domain. The gene was codon-optimized and expressed in *Komagataella phaffii* for extracellular secretion. We performed structural modeling and comparative analyses, revealing conserved catalytic residues and stable TIM-barrel architecture. Biochemical assays showed that StachCel5 exhibits high specific activity, broad pH stability (4–9), and remarkable thermal resistance at 50 °C. We further validated its application in enzymatic pre-treatment of industrial softwood and hardwood pulps, demonstrating improved mechanical properties with reduced refining energy input. These results position StachCel5 as a promising candidate for incorporation into biocatalytic formulations in the pulp and paper, textile, and detergent industries. Its robust operational stability and substrate specificity address key limitations of current enzymatic tools, contributing to more sustainable and cost-effective industrial fiber modification processes.

2. Materials and methods

2.1. Strains, plasmid and media

Escherichia coli DH5α was used as the host strain for cloning and plasmid preparation. Chemically competent *E. coli* strains were prepared, and the transformations were carried out as described previously [20]. The methylotrophic yeast *K. phaffii* strain BG-10 and the expression vector pD915 were purchased from ATUM (Newark, California, USA). pD915 vector contains a *Saccharomyces cerevisiae*-derived α-pre-pro signal sequence for extracellular protein secretion, a constitutive glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter for gene expression, an alcohol oxidase 1 (AOX1) terminator, a zeocin resistance marker for selection, and a pUC origin of replication for propagation in *E. coli*. *E. coli* transformants were selected on low salt Luria–Bertani agar (1 % Tryptone, 0.5 % NaCl, 0.5 % Yeast extract and 1.5 % agar) plates supplemented with 25 μg/mL zeocin and incubated at 37 °C. For the selection of *K. phaffii* transformants, YPDS agar plates (1 % Yeast extract, 2 % Peptone, 2 % Dextrose, 1 % Sorbitol,1.5 % Agar) containing 300 μg/mL zeocin were used.

2.2. Gene identification and Bioinformatic analysis

Putative endoglucanase encoding sequences were retrieved from the publicly available genome sequence of *Stacybotrys chartarum* IBT 7711 (GenBank: KL647752.1) of the National Center for Biotechnology Information (NCBI), and those containing a cellulose binding domain (CBD), given its established role in enhancing enzyme-substrate interactions and improving catalytic efficiency [21] were selected and subjected to preliminary domain analysis to assess their functional classification.

Conserved domain analysis were performed using the NCBI Conserved Domain Database [22]. Signal peptide prediction was carried out using SignalP-6.0 online tool to evaluate the presence of N-terminal secretion signals [23]. The theoretical molecular weight (Mw) and isoelectric point (pI) of the deduced protein were estimated using the Expasy ProtParam tool [24]. Potential N-glycosylation sites were predicted using the NetNGlyc 1.0 server applying a threshold score of >0.5 to define likely glycosylation events [25]. A phylogenetic tree was constructed using the neighbor-joining method based on 40 top non-*Stachybotrys* BLASTp 'hits'. Tree reconstruction was carried out using the MEGA software (version 11), with bootstrapping of 1000 replicates [26]. Secondary structure prediction of cellulase candidate enzyme was performed using the Phyre2.2 online tool [27], while NetSurfP 3.0 was used to predict the relative surface accessibility of the amino acid residues [28]. Far-UV circular dichroism (CD) spectra of StachCel5 at 2.5 μM (20 mM phosphate buffer, pH 7 at 25 °C) was collected using Jasco J-810 spectro polarimeter (Jasco International Co., Japan) in the range of 195–250 nm using 1 mm quartz cuvette. Results have been expressed as mean residual ellipticity ($deg.cm^2.dmol^{-1}$). A total of 3 spectra were collected which were averaged and corrected by subtraction of the blank.

Tertiary structure was predicted exclusively by AlphaFold2 Colab [29,30] and the top-ranked structural model was selected for further analysis. Molecular graphics and visualization of the 3D structure were performed using the PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC. For the identification of conserved structural features, the amino acid sequence of StachCel5 was queried against the Protein Data Bank (PDB) using NCBI BLASTp database [31]. The top non-redundant sequences were aligned using the T-Coffee alignment program [32].

2.3. Recombinant protein expression and purification

The *S. chartarum* cellulase gene, excluding the putative signal peptide (first 16 amino acids), was codon-optimized for expression in *K. phaffii* and synthetized by GenScript (Piscataway, New Jersey, USA) (see supplementary materials, Fig. S1). The synthetic gene was cloned into the pD915 expression vector under the control of the constitutive GAP promoter and fused to the *S. cerevisiae* α-pre-pro signal peptide for extracellular secretion. Cloning was performed using *Sap*I (New England Biolabs, Ipswich, MA, USA) restriction sites. The resulting construct was transformed into *E. coli* DH5α for propagation, and its identity was verified by sequencing. Next, the plasmid was linearized using the unique *Avr*II (New England Biolabs, Ipswich, MA, USA) restriction site located within the GAP promoter and purified via sodium acetate-ethanol precipitation. Linearized plasmid was electroporated into *K. phaffii* BG-10 strain as previously described [33]. Transformants were selected on YPDS agar plates supplemented with 300 μg/mL zeocin and incubated at 30 °C for three days. Colonies were screened for carboxymethyl-cellulase (CMCase) activity in sterile 24-well plates containing 900 μL BMGY medium (1 % *w/v* yeast extract, 2 % w/v peptone, 100 mM pH 6.0 potassium phosphate buffer, 0.004 mg/L biotin, 1.34 % *v*/v yeast nitrogen base and 1 % v/v glycerol), supplemented with 300 μg/mL of zeocin. A glycerol stock of wild-type *K. phaffii* BG-10 was used as a negative control.

Expression of recombinant StachCel5 in cell-free supernatants was verified by 10 % sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). Gels were stained with Coomassie Brilliant Blue. Protein concentration was determined using the Bradford method employing the Bradford Protein Assay Kit (Bio-Rad; Hercules, California, USA). Bovine serum albumin (BSA) was used to generate a standard calibration curve for quantification.

Recombinant StachCel5 was purified from the culture supernatant of *K. phaffi* grown in 300 mL BMGY medium at 28 °C and 250 rpm for 72 h. The cell-free supernatant was first concentrated using a Vivaflow 50 tangential flow filtration system equipped with a 10 kDa molecular weight cut-off (MWCO) membrane (Sartorius Stedim Biotech GmbH, Göttingen, Germany). The buffer of the concentrated supernatant was then exchanged to the equilibration buffer (20 mM phosphate buffer,

pH 7.8) required for subsequent chromatography. The partially purified protein was loaded onto a HiTrap Q HP 1 mL anion exchange column connected to an ÄKTA start protein purification system (Cytiva Life Sciences, Marlborough, MA, USA). The column was pre-equilibrated with 20 mM phosphate buffer, pH 7.8, and the bound proteins were eluted using a linear pH gradient from pH 7.8 to 4.5, using 20 mM phosphate buffer pH 4.5. Fractions containing the target protein were identified by SDS–PAGE, pooled, and further concentrated using the same 10 kDa MWCO Vivaflow 50 device.

2.4. Cellulase activity assay and characterization

The hydrolytic activity of purified StachCel5 was assessed using the 3,5-dinitrosalicylic acid (DNS) method [34], which quantifies the release of reducing sugars from polysaccharide substrates. One unit (U) of cellulase activity was defined as the amount of enzyme that releases 1 μmol of reducing sugar per minute under the assay conditions. Standard assays for enzyme characterization (250 μL total volume) were performed using 1.5 % (*w/v*) sodium carboxymethyl cellulose (Na-CMC) as substrate, with an enzyme dosage of 2 μL (700 ng), and an incubation time of 2 min. Reactions were incubated at various temperatures and pH intervals, as detailed below, and subsequently stopped by the addition of 750 μL of DNS reagent. The mixture was then heated at 95 °C for 5 min for colour development, and absorbance was measured at 540 nm.

To determine the optimum pH, activity assays were performed at 65 °C using various 100 mM buffers across a pH range of 3.0–12.0: sodium citrate (pH 3.0–4.0), sodium acetate (pH 4.0–6.0), sodium phosphate (pH 6.0–7.0), Tris–HCl (pH 7.0–9.0), and sodium glycine (pH 9.0–12.0). pH stability was evaluated by pre-incubating the enzyme at 4 °C for 16 h in the respective buffers, followed by measuring residual activity under optimal conditions (100 mM sodium acetate buffer pH 5 at 65 °C). On the other hand, optimal temperature was determined by conducting the activity assay at temperatures ranging from 25 °C to 90 °C in 100 mM sodium acetate buffer (pH 5). Thermal stability was assessed by incubating the enzyme in the same buffer at 50 °C, 55 °C, and 60 °C for up to 1 h, with samples taken at different time intervals. Residual enzyme activity was then measured under optimal assay conditions (pH 5, 65 °C). Enzyme activity prior to incubation was taken as 100 % and used as the control.

To evaluate the influence of various effectors on StachCel5 activity, individual reactions were supplemented with 10 mM of different metal ions (Mg^{2+}, Ca^{2+}, Cu^{2+}, Co^{2+}, Ba^{2+}, Fe^{2+}, Zn^{2+}, Mn^{2+}, Ni^{2+}, Ag^{2+}, Pb^{2+}), the chelating agent EDTA, or the reducing agent β-mercaptoethanol. The effects of detergents—Tween 20, Triton X-100, Tween 80, and sodium dodecyl sulfate (SDS)—were also evaluated at a final concentration of 0.25 % (*v*/v). Each mixture was pre-incubated with the enzyme at room temperature for 60 min, after which residual enzymatic activity was determined using the DNS method under previously established optimal conditions. Enzymatic activity in the absence of any additive was defined as 100 % activity and used as the control.

Kinetic parameters of StachCel5, including the Michaelis-Menten constant (Km) and maximum reaction velocity (Vmax), were determined under optimal conditions by incubating varying concentrations of Na-CMC (6–21 mg/mL) for 3 min. The Km and Vmax values were derived from nonlinear regression fitting of the Michaelis-Menten equation using GraphPad Prism 10 for Windows. All assays were performed in triplicate, and the mean values were used for analysis.

2.5. Substrate specificity of StachCel5

Substrate specificity of StachCel5 was assessed by using 1.5 % (*w/v*) concentrations of various polysaccharides, including insoluble crystalline cellulose substrates (Avicel, filter paper), laminarin, xylan from oat spelts, potato starch, and Na-CMC (Merck KGaA, Germany), under standard reaction conditions. Additionally, the enzyme's activity on synthetic substrates *para*-nitrophenyl-β-D-cellobioside (*p*NPC) and *para*-nitrophenyl-β-D-glucopyranoside (*p*NPG) at a concentration of 2 mM was also evaluated. The reaction (250 μL total volume) was carried out at 50 °C for 5 min, followed by the addition of 750 μL of 1.0 M Na_2CO_3 to terminate the reaction. The amount of *para*-nitrophenol released was quantified spectrophotometrically by measuring absorbance at 405 nm. One unit of activity was defined as the amount of enzyme required to release 1 μmol of *para*-nitrophenol per minute.

2.6. Laboratory refining trials, sheet forming, and paper testing using softwood and hardwood pulp

The biorefining potential of endoglucanase StachCel5 was evaluated through its application to industrial dried, TCF (Totally Chlorine Free) bleached hardwood Kraft pulp from *Eucalyptus globulus*, supplied by ENCE (Pontevedra, Spain) and ECF (Elemental Chlorine Free) bleached softwood Kraft pulp, composed of 85 % Pine and 15 % Fir and Spruce, provided by Fiber Excellence (Saint-Gaudens, France). A concentrated crude extract from *K. phaffii* BG-10 expressing StachCel5 was used for this evaluation.

Enzyme treatments were conducted on 30 g (dry weight) of pulp at 10 % (*w/v*) pulp consistency in 50 mM sodium acetate buffer (pH 5). The enzyme was initially dissolved in the same buffer and then added to the pulp suspension to ensure homogeneous distribution. The enzymatic treatment was performed in BRAND® disposable polypropylene bags (Merck KGaA, Germany) placed in a water bath at 40 °C for 30 min, with an enzyme dosage of 0.5 CMCase units per gram of dry fiber. Samples were periodically removed from the water and kneaded for 10–15 s to ensure homogenous reaction. After the enzymatic treatment, the pulp was heated in a water bath at 90 °C for 30 min to terminate the reaction. The control treatment followed the same protocol but without enzyme addition.

Refining experiments on enzyme-treated and control pulp samples were conducted at 750, 1500, and 3000 revolutions for hardwood pulp, and 3000, 4500, and 6000 revolutions for softwood pulp, using a Mark VI PFI mill (Hamjern Maskin a.s., Hamar, Norway), in accordance with ISO 5264-2. Pulp properties were measured by drainage resistance (Schopper-Riegler number (°SR), ISO 5267-1) were measured. Handsheets with a grammage of 75 ± 3 g/m^2 were prepared using Rapid-Köthen equipment according to ISO 5269-2 and were tested mechanically according to the following standards: tensile index (ISO 1924-3), burst index (ISO 2758), tear index (ISO 1974) and zero-span tensile index (ISO 15361). Bendtsen air permeance (ISO 5636-3) and light scattering (ISO 9416) were also measured. Prior to testing, the handsheets were conditioned for 24 h at 23 °C and 50 % relative humidity according to ISO-187. A total of 8–9 paper samples were tested.

2.7. Scanning Electron microscopy (SEM) studies

SEM microscopy was used to analyze the surface morphology of paper sheets. Observations were conducted using a JEOL-6510 microscope under vacuum at an operating voltage of 10 kV. Samples were previously cut into small pieces and coated with carbon using a Cressington 108 carbon coater to obtain a conductive surface.

3. Results and discussion

3.1. Database mining for novel endoglucanases, phylogenetic analysis, and 3D structure

The publicly available GenBank non-redundant (nr) database for *Stachybotrys chartarum* IBT 7711 was screened to identify putative enzymes with endoglucanase activity containing a carbohydrate-binding domain (CBD). Five sequences were retrieved and subjected to preliminary domain analysis and BLASTp searches to determine their functional classification. Based on this analysis, the gene designated *stachCel5* (NCBI locus tag: S7711_03103) was identified as a putative

endoglucanase member of glycosyl hydrolase family 5 (GH5) and selected for further studies (see supplementary materials, Fig. S2).

According to NCBI annotation, the selected open reading frame (ORF) consists of 1263 bp and encodes a hypothetical protein of 402 amino acid residues (GenBank: KEY73803.1) (see supplementary materials, Fig. S3). Analysis performed through Conserved Domain Database (CDD) from NCBI confirmed that StachCel5 possess a fungal cellulose-binding module positioned between the residues Gly^{21} and Ile^{54} and a C-terminal catalytic domain belonging to the glycosyl hydrolase 5 family (residues Val^{104} to Gly^{371}) (see Supplementary material, Fig. S4), united by a peptide region rich in Thr and Ser residues (residues Asn^{55} to Gly^{103}). Additionally, a putative N-terminal signal peptide was predicted using the SignalP-6.0 server, with a probable cleavage site between residues 16 and 17. The estimated molecular mass and isoelectric point of StachCel5, excluding the signal peptide, were 42.8 KDa and 5.72, respectively.

Currently, GH5 is one of the largest and most diverse glycoside hydrolase families. While it primarily comprises *endo*-β-1,4-glucanase, many GH5 members exhibit broad substrate specificity, displaying additional activities such as mixed-linkage (β-1,3;1,4)-glucanase, xylanase, galactanase and xyloglucanase functions [35]. Phylogenetic analysis showed that StachCel5 clusters closely with GH5 endoglucanases from *Clonostachys* spp. and *Paramyrothecium foliicola*, (see Supplementary material, Fig. S5). Sequence alignment revealed 62–65 % identity and 99–100 % query coverage with these homologs based on the NCBI nr protein database. Additionally, StachCel5 was positioned near structurally characterized GH5 cellulases, including those from *Aspergillus niger*, *Thermoascus aurantiacus*, and *Penicillium verruculosum*, showing sequence identities of 60.8 %, 58.8 %, and 56.0 %, respectively, with approximately 75 % query coverage (Fig. 1).

Secondary structure prediction using Phyre2 indicated that StachCel5 consists of approximately 24 % α-helix and 13 % β-strand (see Supplementary Fig. S6a). This prediction was experimentally validated by circular dichroism (CD) spectroscopy, and the CD spectra were analyzed using BeStSel [36], confirming the presence of these secondary structural elements (see Supplementary Fig. S6b). The three-dimensional (3D) structure of the enzyme was modeled using Alphafold2, which has demonstrated exceptional accuracy in protein structure prediction [37]. The top ranked model was selected (see Supplementary material, Fig. S7) and subsequently visualized with PyMOL for further structural and functional analyses (Fig. 2a).

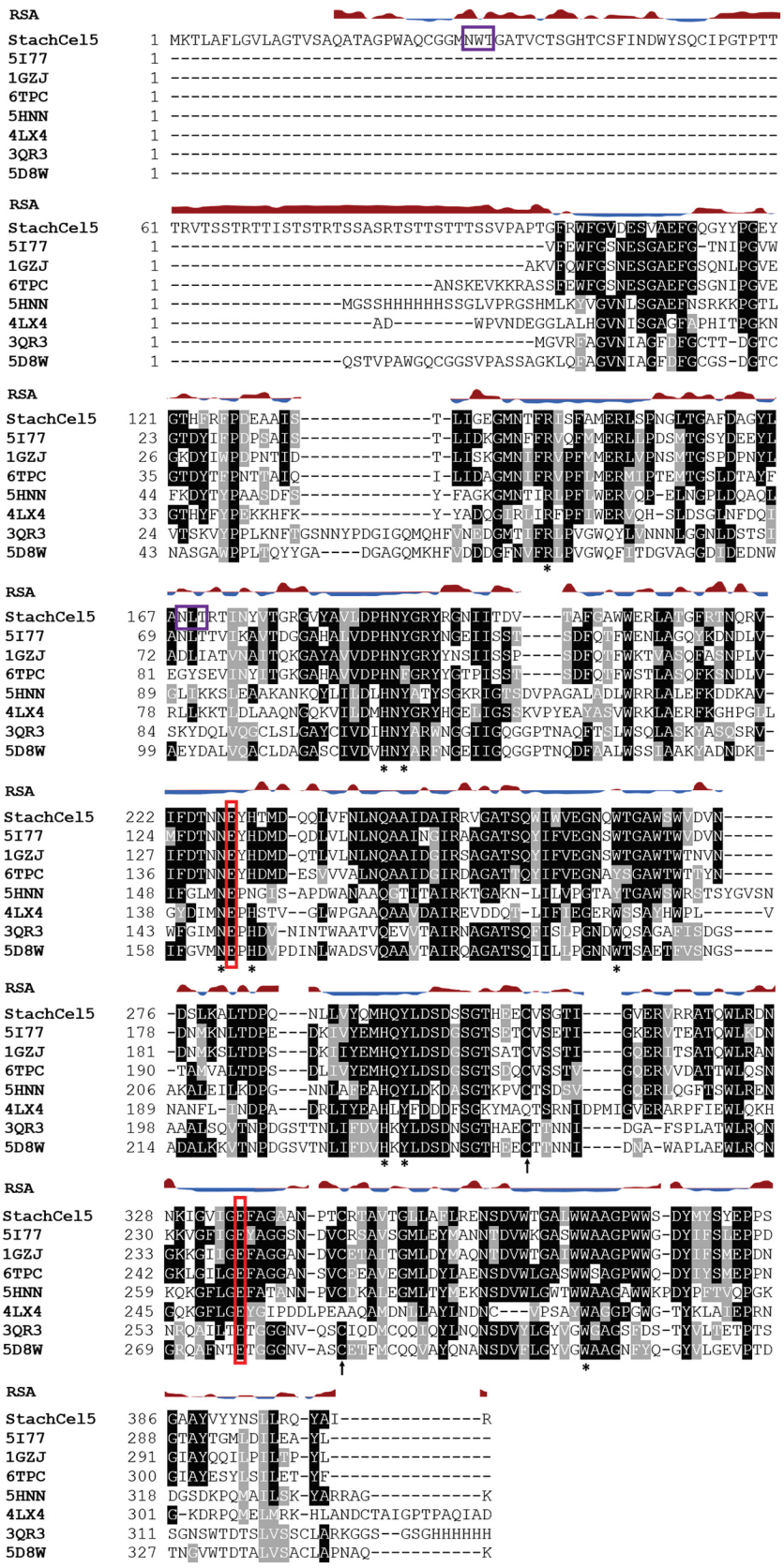
To identify the catalytic site of StachCel5 and gain insights into its molecular mechanism, its amino acid sequence was aligned with those of experimentally characterized GH5 family cellulases from *A. niger* [PDB ID: 5I77], *T. aurantiacus* [PDB ID: 1GZJ], and *P. verruculosum* [PDB ID: 6TPC]. The alignment revealed that residues Glu^{228} and Glu^{335} are evolutionarily conserved and likely represent the catalytic residues of StachCel5 (Figs. 1 and 2b). To further validate this hypothesis, the 3D structural model of StachCel5 was superimposed onto the GH5 endoglucanase from *A. niger* (PDB ID: 5I77) (Fig. 2c). The comparison revealed a high degree of structural similarity, particularly in the conserved $(\beta/\alpha)_8$-barrel fold architecture characteristic of GH5 enzymes, commonly known as the TIM-barrel, including alignment of the catalytic residues in both structures (Fig. 2c). This structural congruence is consistent with the sequence alignment results and supports the identification of Glu^{228} and Glu^{335} as the catalytic residues mediating glycosidic bond hydrolysis via a general acid–base catalysis mechanism, with Glu^{228} functioning as the proton donor and Glu^{335} serving as the nucleophile [38]. The distance between these residues plays a critical role in determining the catalytic mechanism: retaining enzymes typically exhibit a distance of approximately 5.5 Å, whereas inverting enzymes show about 10 Å between catalytic residues [39]. In the predicted 3D structure of StachCel5, the distance between Glu^{228} and Glu^{335} was measured at 4.9 Å (Fig. 2c), suggesting that the enzyme likely follows a retaining mechanism, as commonly observed among GH5 family members [38]. Furthermore, analysis of the predicted surface accessibility indicates that both catalytic residues are situated in a buried region within the protein core. Notably, sequence and structural analyses also reveal that other highly conserved residues are also predicted to be buried. This spatial arrangement is characteristic of many endoglucanases, where the catalytic residues reside in a well-defined cleft or pocket. Such partial burial serves to stabilize the electrostatic environment and ensures precise orientation of catalytic groups, which is essential for efficient substrate turnover. Although buried, these residues remain functionally accessible via a solvent-accessible channel that allows substrate entry and product release [40,41].

3.2. Heterologous expression and purification of the recombinant endoglucanase StachCel5

The yeast *K. phaffii* (commonly known as *Pichia pastoris*) is among the most widely used hosts for extracellular protein secretion, mainly due to its low secretion of endogenous proteins, which significantly simplifies downstream purification of the target protein [42]. Moreover, none of its native secreted proteins have been reported to display activity on lignocellulosic biomass, making *K. phaffii* an ideal host for the heterologous expression of fungal biomass-degrading enzymes [43]. The integrative vector pD915 and the *K. phaffii* BG-10 strain were employed as the heterologous expression system for StachCel5. The *stachCel5* gene from *S. chartarum* devoid of its signal peptide was synthesized with codon optimized for *K. phaffii*. This gene was cloned into the pD915 vector, generating an α-prepro signal sequence-*stachCel5* fusion construct under the control of the constitutive GAP promoter. The resulting plasmid was then transformed into *K. phaffii* BG-10 cells for recombinant expression. Recombinant clones were screened to identify the top-performing-secretory clone, exhibiting the highest carboxymethyl cellulase (CMCase) activity in culture supernatants (see section 2.3 for details). This screening aimed to isolate the so-called jackpot clone, characterized by a high copy number of the integrated expression cassette, which displayed the highest CMCase activity of 168 ± 13.1 U/mL (Fig. 3a). SDS-PAGE analysis of the cell-free culture supernatant revealed a prominent band with an estimated molecular weight of approximately 55 kDa (Fig. 3b). Although this observed molecular weight exceeds the theoretical molecular weight of 42.8 kDa predicted in silico, such discrepancy might be attributed to protein glycosylation, which can alter migration patterns in SDS-PAGE [44]. In particular, the presence of N-linked glycans, as well as potential hyperglycosylation, common for heterologous proteins expressed in *K. phaffii*, may explain the observed size difference. Supporting this, in silico analysis with the NetGlyc 1.0 Server, predicted two N-glycosylation sites at Asn^{30} and Asn^{168} in the StachCel5 sequence (Fig. 1). Additionally, *K. phaffii* is known to add O-linked glycans to serine and threonine residues, which may occur in the peptide linker connecting the fungal cellulose-binding domain (fCBD) and the catalytic domain of StachCel5 [44,45]. StachCel5 was purified to homogeneity from *K. phaffii* culture supernatants by ion exchange chromatography (Fig. 3c) and subsequently characterized.

3.3. Biochemical characterization of novel StachCel5

The activity of purified StachCel5 enzyme on carboxymethyl cellulose (CMC) was initially assessed across a range of pH values, with the enzyme retaining over 87 % of its activity at pH values between 4 and 7 and still maintained approximately 50 % of its maximum activity at alkaline pH 9 (Fig. 4a). At the optimal pH of 5, maximum activity of StachCel5 was observed at 65 °C, within a temperature range of 25 to 90 °C (Fig. 4b). Furthermore, StachCel5 exhibited stability over a broad pH range of pH 3 to pH 10, retaining 90–100 % activity. However, its activity significantly decreased at extreme pH values of pH 2 and pH 11, where only 49 % and 26 % of maximum activity were retained, respectively (Fig. 4c), thus indicating a strong stability of the StachCel5 enzyme. In thermal stability studies, StachCel5 demonstrated the greatest stability at 50 °C, retaining over 77 % residual activity after 60



(caption on next page)

Fig. 1. Multiple sequence alignment of GH5 family endoglucanases retrieved by BLASTp against the PDB database, highlighting conserved amino acid residues in StachCel5. Identical and similar residues are shaded in black and grey, respectively. Conserved catalytic glutamate residues, serving as the nucleophile and general acid/base in GH5 cellulases, are indicated by red boxes. Other key conserved active site residues are marked with asterisk (*). Position of cysteine residues are indicated with arrows. Predicted N-glycosylation sites in StachCel5 with a potential score > 0.5 are highlighted in purple boxes. Accession details and sequence identities to StachCel5 are as follows: 5I77 (*A. niger*, 60.8 %); 1GZJ (*T. aurantiacus*, 58.8 %); 6TPC (*P. verruculosum*, 56 %); 5HNN (*Xanthomonas axonopodis pv. citri*, 38.4 %); 4LX4 (*Stutzerimonas stutzeri* A1501, 31.4 %); 3QR3 (*Trichoderma reesei*, 31.5 %); 5D8W (*Ganoderma lucidum*, 34 %); and StachCel5 (this study). RSA: predicted Relative Surface Accessibility of StachCel5. Residues with RSA values >25 % are considered exposed (shown in red), while those with RSA ≤ 25 % are classified as buried (shown in blue).

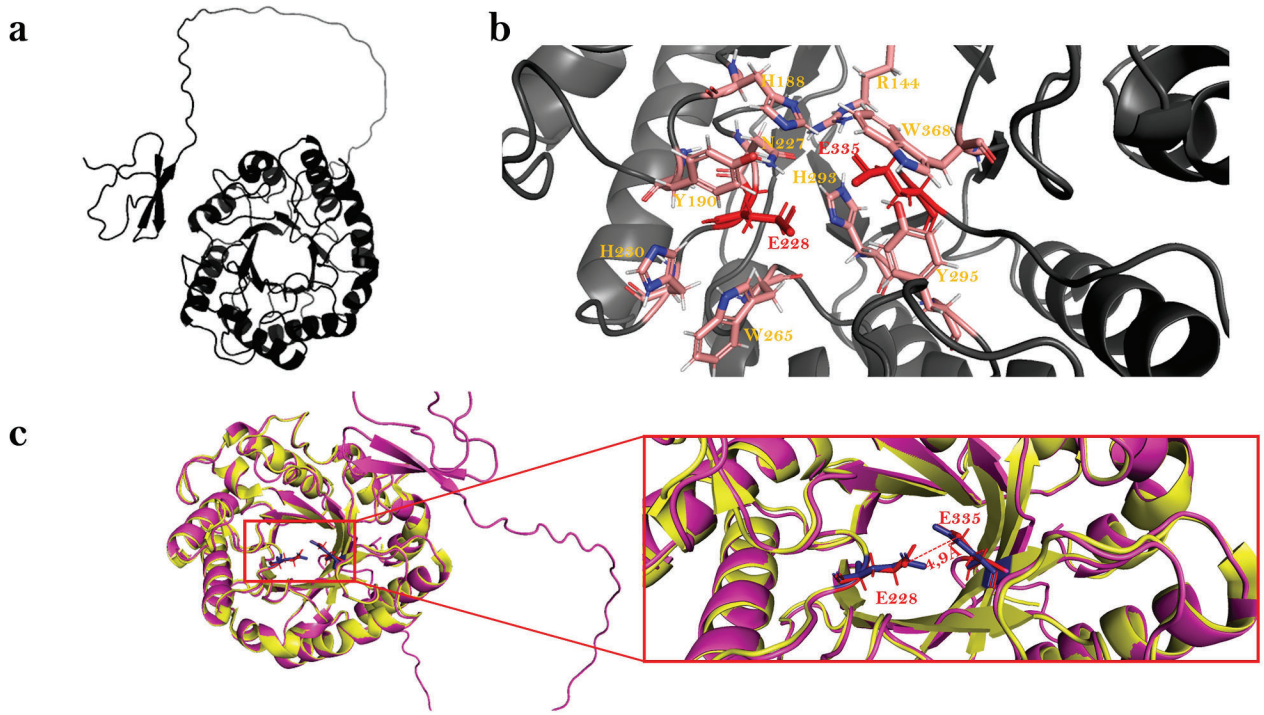


Fig. 2. Analysis of the predicted 3D structure of StachCel5. (a) Top ranked 3D structure model of StachCel5 predicted by Alphafold2 Colab. (b) Detailed view of the active site, highlighting highly conserved residues (R144, H188, Y190, N227, H230, W265, H293, Y295, W368) surrounding the catalytic glutamates E228 and E335 (shown as red sticks). (c) Superposition of StachCel5 (magenta) with the GH5 endoglucanase from *A. niger* (PDB ID: 5I77; yellow). The inset depicts the superimposed catalytic residues of StachCel5 (E228 and E335, red sticks) and endoglucanase 5I77 (blue sticks). The distance between the catalytic glutamates is indicated in red. All figures were generated using PyMOL.

min of incubation. At higher temperatures, the enzyme showed residual activities of about 60 % and 20 % after 15 min of incubation at 55 and 60 °C, respectively (Fig. 4d).

A comparative analysis of the enzymatic properties of StachCel5 with previously reported GH5 enzymes reveals both similarities and notable differences (Table 1). While most cellulases within the GH5 family exhibit optimal activity under acidic conditions (pH 3–5) and moderate temperatures (50 to 70 °C), StachCel5 demonstrates a broad pH activity range (pH 4–7), enhancing its applicability in biotechnological processes that operate under acidic to near-neutral conditions. The enzyme demonstrated high specific activity on CMC (287.6 ± 12.3 U/mg), positioning it among the most active GH5 family members, though still below the exceptionally active AnCel5A from *Aspergillus niger* (1726 U/mg) [46]. The combination of StachCel5's broad pH tolerance and elevated optimal temperature (65 °C) enhances its industrial relevance for applications demanding robust enzymatic performance under diverse operational conditions [4,9].

To further characterize StachCel5, the effects of various metal ions and chemical reagents on its enzymatic activity were evaluated as shown in Fig. 4e. Among the metal ions tested, most exhibited no significant influence on StachCel5 activity, with the exception of Mn^{2+} and Co^{2+}, which caused a slight improvement, whereas Zn^{2+} and Ag^{2+} decreased enzymatic activity by 21 % and 78 %, respectively. Notably, EDTA did not completely inhibit enzymatic activity, suggesting that StachCel5 is not dependent on metal ions for its catalytic function. Similarly, the addition of β-mercaptoethanol had no effect on enzymatic activity, indicating that disulfide bonds are not critical for maintaining the active conformation of StachCel5. Regarding the effects of detergents, no significant changes in activity were observed in the presence of Tween 20, Tween 80, or Triton X-100, whereas SDS completely abolished StachCel5 activity (Fig. 4e), an effect observed for other GH5 endoglucanases too [47].

A comparative analysis of kinetic parameters among GH5 endoglucanases from different organisms highlights the distinct catalytic performance of StachCel5 (Table 2). StachCel5 exhibited a moderate substrate affinity with a Km of 9.77 mg/mL and a high Vmax of 898.4 U/mg. Its turnover number (Kcat) reached 635.7 s^{-1}, placing it among the top-performing GH5 enzymes reported to date. Although *AnCel5A* from *Aspergillus niger* showed a higher catalytic efficiency (Kcat/Km = 227 mL/s/mg) [46], StachCel5 outperformed several fungal homologs such as MtEG5–1 from *Myceliophthora thermophila* (Kcat/Km = 13.9 mL/s/mg) [52] and GH5 from *Fomitopsis meliae* (Kcat/Km = 10.78 mL/s/mg) [50]. Notably, StachCel5 exhibited higher specific activity and Vmax compared to the widely used TrCel5A from *Trichoderma reesei*, a benchmark endoglucanase commonly incorporated into industrial biomass deconstruction enzyme cocktails [56]. The kinetic profile of StachCel5 indicates an efficient catalytic mechanism that balances substrate binding and turnover, positioning it as a promising candidate

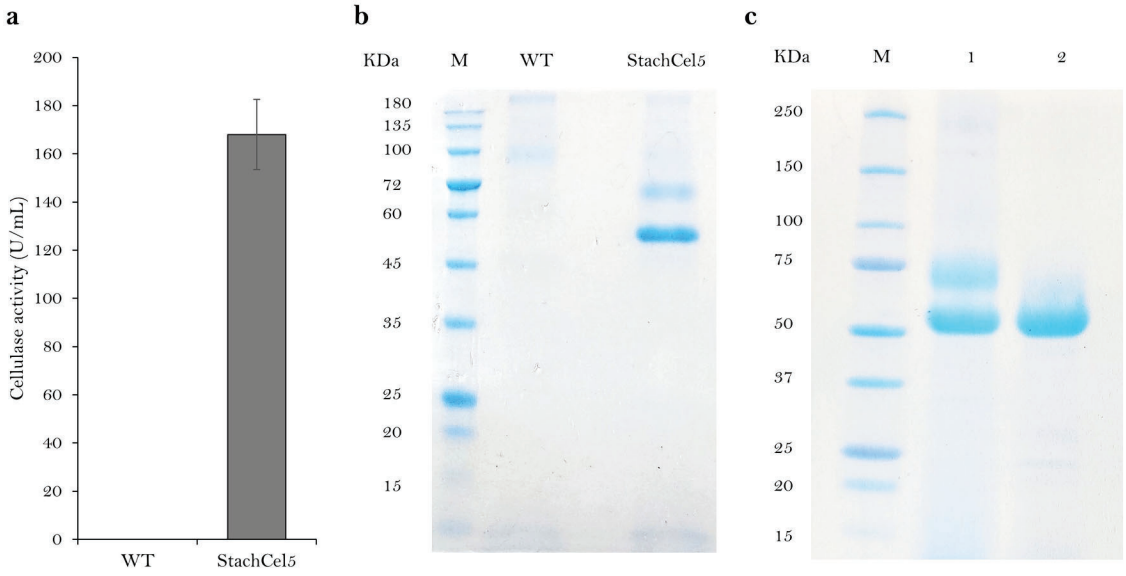


Fig. 3. Expression of recombinant endoglucanase StachCel5 in *K. phaffii* BG-10. (a) CMCase activity measured in culture supernatants of the wild-type *K. phaffii* BG-10 strain (WT) and the recombinant strain (StachCel5) expressing StachCel5. (b) SDS-PAGE analysis of culture supernatants from WT and recombinant *K. phaffii* BG-10 strains. (c) SDS-PAGE of purified StachCel5 from recombinant *K. phaffii* BG-10. Lane 1: culture supernatant before purification; Lane 2: purified StachCel5 fraction; Lane M: molecular weight marker.

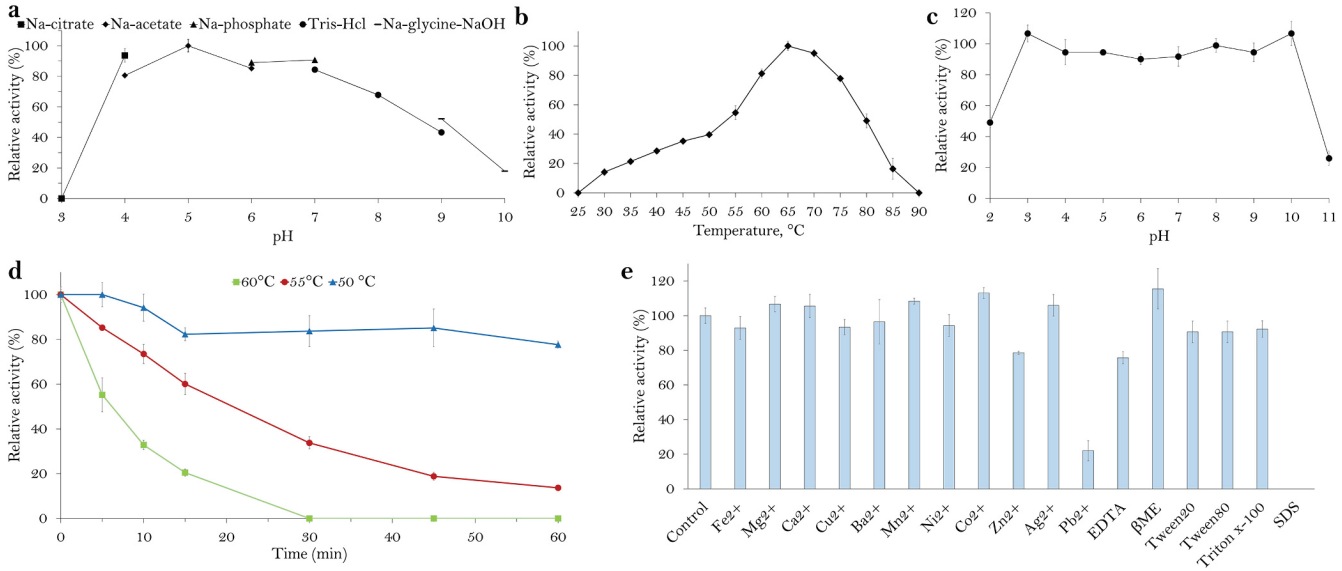


Fig. 4. Characterization of purified StachCel5 on 1.5 % carboxymethyl cellulose (CMC). (a) Effect of pH on enzyme activity at 65 °C (b) Effect of temperature on enzyme activity at pH 5.0 (c) pH stability of StachCel5, with 100 % residual activity corresponding to the activity at pH 5.0 before incubation (d) Thermostability assays of StachCel5, with 100 % residual activity corresponding to the activity before incubation (e) Effect of 10 mM metal ions, EDTA, β-Mercaptoethanol (βME) and detergents (0.25 %) on StachCel5 activity. The activity of StachCel5 in the absence of any solute was taken as 100 % (control). A relative activity of 100 % under the optimal conditions (pH 5 and 65 °C) corresponds to an absolute activity of 168.4 ± 4.2 U/mL. Results are expressed as the average of triplicate assays ± the standard error of the mean.

for biomass hydrolysis and other biotechnological applications.

3.4. Substrate specificity of novel StachCel5

Activity profiling of StachCel5 on various substrates revealed strong specificity for β-1,4-glucosidic linkages in CMC, displaying an activity of 287.6 ± 12.3 U/mg, and for mixed β-1,4 and β-1,3-glucosidic linkages in lichenan, with an activity of 480.4 ± 24.3 U/mg under optimal conditions. Family GH5 contains enzymes with diverse specificity, including cellulose *endo*-β-1,4-glucosidase (EC 3.2.1.4), endo-β-1,3/4-glucosidase (EC 3.2.1.6), glucan endo-β-1,3-glucosidase (EC 3.2.1.39) as well as other specificities (https://www.cazy.org/GH5_activity.html). StachCel5 shows higher activity on lichenan (with β-1,4 and β-1,3-glucosidic linkages) than with CM-cellulose (with all β-1,4 glucosidic linkages), but no activity on laminarin with β-1,3 glucosidic linkages. It indicates hydrolase activity on β-1,4-linkages with a specificity closer to endo-β-1,3/4-glucosidase (EC 3.2.1.6) as commonly found in the GH5 family [40,57].

In contrast, StachCel5 exhibited no hydrolytic activity toward xylan or starch. Moreover, no activity was detected against insoluble crystalline substrates such as Avicel and filter paper, or synthetic substrates *p*NPC and *p*NPG (data not shown). These findings emphasize the highly

Table 1
Properties comparison of StachCel5 with other fungal endoglucanases of family GH5. All properties are based on carboxymethyl cellulase (CMCase) activity. Not indicated (n.i.).

Enzyme	Organism	Optimum pH	Optimum TM (°C)	Specific CMC activity (U/mg)	References
StachCel5	*Stachybotrys chartarum*	4–7	65	287.6 ± 12.3	This study
Cel5A	*Thermoascus aurantiacus*	4	70	45	[48]
TrCel5A	*Trichoderma reesei*	5	55	220.57	[49]
AnCel5A	*Aspergillus niger*	4	70	1726 ± 19	[46]
GH5	*Fomitopsis meliae*	4.8	70	302.90	[50]
AgCMCase	*Aspergillus glaucus*	5	55	175.3 ± 5.1	[51]
*Mt*EG5–1	*Myceliophthora thermophila*	4–5	70	n.i.	[52]
GlCel5A	*Ganoderma lucidum*	3–4	60	104 ± 3	[53]
EGII	*Penicillium verruculosum*	4.5	70	50	[54]
EGL2	*Humicola grisea*	5	75	31.9	[55]

Table 2
Kinetic parameters of StachCel5 compared with other fungal GH5 endoglucanases. All values are based on carboxymethyl cellulase (CMCase) activity. n.i., not indicated.

Enzyme	Organism	Km (mg/mL)	Vmax (U/mg)	Kcat (s^{-1})	Kcat/Km (mL/s/mg)	References
StachCel5	*Stachybotrys chartarum*	9.77 ± 0.9	898.4 ± 15	635.7 ± 10.6	65.4 ± 7.1	This study
Cel5A	*Thermoascus aurantiacus*	37	82.6	n.i.	n.i.	[48]
TrCel5A	*Trichoderma reesei*	2.1	220.57	n.i.	n.i.	[49]
AnCel5A	*Aspergillus niger*	7.64 ± 0.4	2786 ± 79	1625 ± 46	227 ± 21	[46]
GH5	*Fomitopsis meliae*	12	556.58	129.41	10.78	[50]
AgCMCase	*Aspergillus glaucus*	7.22 mM	343.81 ± 2.77	n.i.	n.i.	[51]
MtEG5–1	*Myceliophthora thermophila*	6.11	91.74	84.96	13.9	[52]

specific endo-β-1,4-glucanase activity of StachCel5. Based on this data, StachCel5 can be classified within the GH5_5 subfamily, which predominantly comprises secreted bacterial and fungal enzymes exhibiting solely endo-β-1,4-glucanase activity (EC 3.2.1.4), with nearly half containing a carbohydrate-binding module [35].

In a nutshell, the novel endoglucanase StachCel5 demonstrates strong potential to advance biotechnological applications in demanding industrial processes, such as biofuel production, waste treatment, and lignocellulosic biomass bioconversion, where the performance of conventional cellulases is frequently compromised by harsh operational conditions [58].

3.5. Biotechnological application of StachCel5 in pulp biorefining

To evaluate the industrial applicability of StachCel5, its functional potential as a refining aid was examined under controlled laboratory conditions using both softwood and hardwood pulps. One of the key properties for assessing the outcome of the refining process is pulp drainability, characterized by the Schopper–Riegler number (°SR), which is mainly influenced by fiber external fibrillation and the presence of fine elements. As refining progresses, drainability typically decreases due to increased fiber bonding and the formation of a denser fiber network, thus retaining water more effectively and increasing the °SR number [59,60]. Treatment with StachCel5 significantly enhanced the refining degree for both softwood and hardwood pulps, as evidenced by a higher °SR in enzyme-treated pulps compared to the untreated control (Fig. 5). In softwood pulp, a notably higher °SR of 40 at 6000 revolutions was reached in comparison to 33.5 in control (Fig. 5a), while hardwood pulps showed a modest but meaningful increase from 29.3 to 31.8°SR at 3000 revolutions (Fig. 5b). These findings indicate that StachCel5 effectively promotes fiber modification, accelerating the refining process and potentially reducing energy consumption.

Other important properties used to assess the degree of refining in the paper handsheets include air permeance and the light scattering coefficient, both of which are expected to decrease as the refining degree increases due to enhanced fiber bonding and fiber flexibility, leading to a denser paper with a decreased opacity and porosity [61,62]. In line with this expectation, a slight but consistent decrease in both properties was observed in enzyme-treated paper handsheets compared to the untreated control. Specifically, air permeance values in enzyme-treated hardwood pulps declined by 2 % to 14 %, and by up to 20 % in softwood pulps. Likewise, light scattering coefficients were reduced by 2 % to 4 %

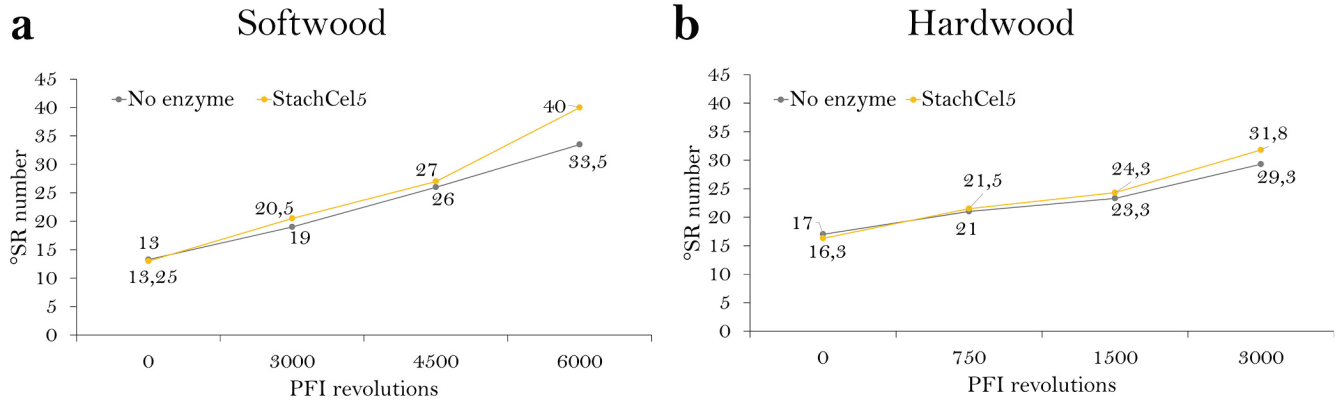


Fig. 5. Evolution of the degree of refining. Schopper-Riegler number (°SR) as a function of PFI refiner revolutions during laboratory refining of control (no enzyme) and StachCel5-treated pulps for (a) softwood and (b) hardwood. Enzyme treatments were applied at 0.5 CMCase units per gram of dry fiber for 30 min.

across different refining levels for both pulp types. These reductions provide evidence that the enzymatic treatment exerted a measurable effect on fiber morphology and sheet structure (see Supplementary material, Fig. S8). The results suggest that low-dose application of StachCel5 promotes external fibrillation by selectively cleaving cellulose chains at the fiber surface without inducing fiber damage or generating fines, thus improving inter-fiber bonding and paper strength properties [11,63]. Notably, StachCel5 enabled the attainment of a desirable refining degree with a reduced number of refiner revolutions, particularly in softwood pulp, thereby contributing to lower energy consumption during mechanical treatment.

3.6. Physico-mechanical properties of paper handsheets

To further evaluate the industrial applicability of StachCel5 in pulp refining, the balance between effective fiber modification and the

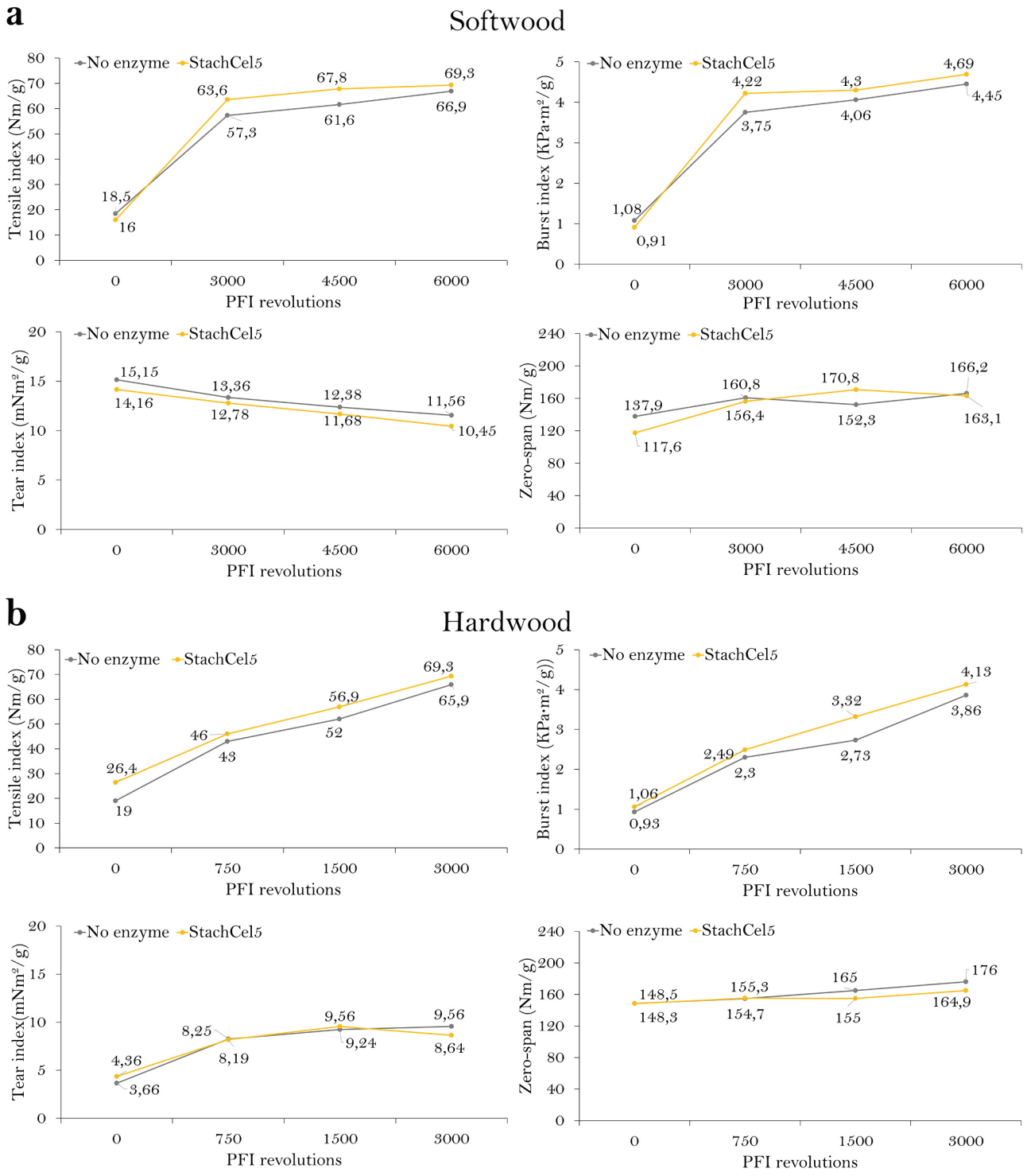


Fig. 6. Evolution of the physico-mechanical properties. Tensile, burst, tear and zero-span tensile indexes of handsheets obtained during the refining process for control (no enzyme) and StachCel5-treated pulps of (a) softwood and (b) hardwood. Enzyme treatments were applied at 0.5 CMCase units per gram of dry fiber for 30 min.

preservation of fiber integrity was examined. As an endoglucanase, StachCel5 catalyzes the hydrolysis of internal β-1,4-glycosidic linkages in cellulose, potentially facilitating fiber modification. However, if not carefully dosed, excessive hydrolysis may result in fiber over-degradation and increased fines content, ultimately impairing the physico-mechanical properties of the resulting handsheets [6,64]. To assess these outcomes, tensile strength index, burst index, tear index, and zero-span tensile index were measured in both untreated (control) and StachCel5-treated softwood and hardwood pulps, providing insight into the enzyme's effect on fiber and paper quality (Fig. 6).

The tensile index is crucial for assessing the refining process in the papermaking process, as it increases with the number of revolutions, primarily due to a denser paper structure and enhanced inter-fiber bonding through increased internal and external fibrillation [8,65]. Notably, softwood treated pulp followed a similar trend to the control but demonstrated a markedly enhanced refining efficiency, achieving the control's maximum tensile index (66.9 Nm/g at 6000 PFI revolutions) at only 4500 revolutions (67.8 Nm/g), thus demonstrating a ~ 25 % reduction in mechanical energy required (Fig. 6a). A similar trend was observed in hardwood pulp, though less pronounced, with tensile index values reaching 56.9 Nm/g at 1500 revolutions and 69.3 Nm/g at 3000 revolutions, compared to 52 Nm/g and a maximum of 65.9 Nm/g in the untreated counterpart, respectively (Fig. 6b). Tensile-related properties such as tensile energy absorption (T.E.A.), which is a measure of the paper's work to failure, and elongation exhibited parallel improvements (see Supplementary material, Fig. S9), supporting the enzyme's role as a refining aid. Notably, enzyme-treated handsheets demonstrated a superior T.E.A.–porosity profile relative to controls (see Supplementary material, Fig. S10), a key requirement for high-performance sack papers, which must resist mechanical stress during filling while maintaining adequate air resistance [66]. Furthermore, at a given drainage level (°SR), softwood pulps treated with StachCel5 showed higher tensile strength than untreated samples, indicating improved fiber development at comparable dewatering resistance (Fig. 7a). A similar, albeit less pronounced, effect was observed in hardwood pulp (Fig. 7b). These enhancements suggest potential energy savings in subsequent stages of papermaking, particularly during drying [67].

The burst index is a critical parameter for packaging-grade papers, as it reflects the material's ability to resist rupture under hydrostatic pressure, thereby improving durability during handling and transport [68]. In this study, StachCel5 treatment led to an increase in burst strength in handsheets produced from both softwood and hardwood pulps. For softwood pulp, the enzyme-treated sample reached a burst index of 4.69 kPa·m^2/g at 6000 PFI revolutions, compared to 4.45 kPa·m^2/g for the untreated control. This improvement mirrors the energy savings observed for tensile index, suggesting enhanced fiber bonding at reduced refining intensity. Hardwood pulp also showed notable gains, with the treated sample achieving a maximum burst index of 4.13 kPa·m^2/g at 3000 revolutions, relative to 3.86 kPa·m^2/g in the control (Fig. 6a). The enhancements in both tensile and burst indices are consistent with increased fiber bonding capacity, leading to the formation of a denser, more compact fibrous structure [6,65].

This structural compaction is further corroborated by scanning electron microscopy (SEM) of the handsheet surfaces (Fig. 8), which revealed a more uniformly bonded fiber network in enzyme-treated samples. These morphological improvements correspond with increased resistance to airflow and a reduced light scattering coefficient, indicating a tighter fiber matrix (Supplementary Fig. S8). Conversely, SEM images of unrefined pulps showed no significant differences in surface morphology, confirming that enzymatic action alone cannot replace mechanical refining (Supplementary Fig. S11). However, these findings demonstrate that StachCel5 facilitates energy-efficient fiber

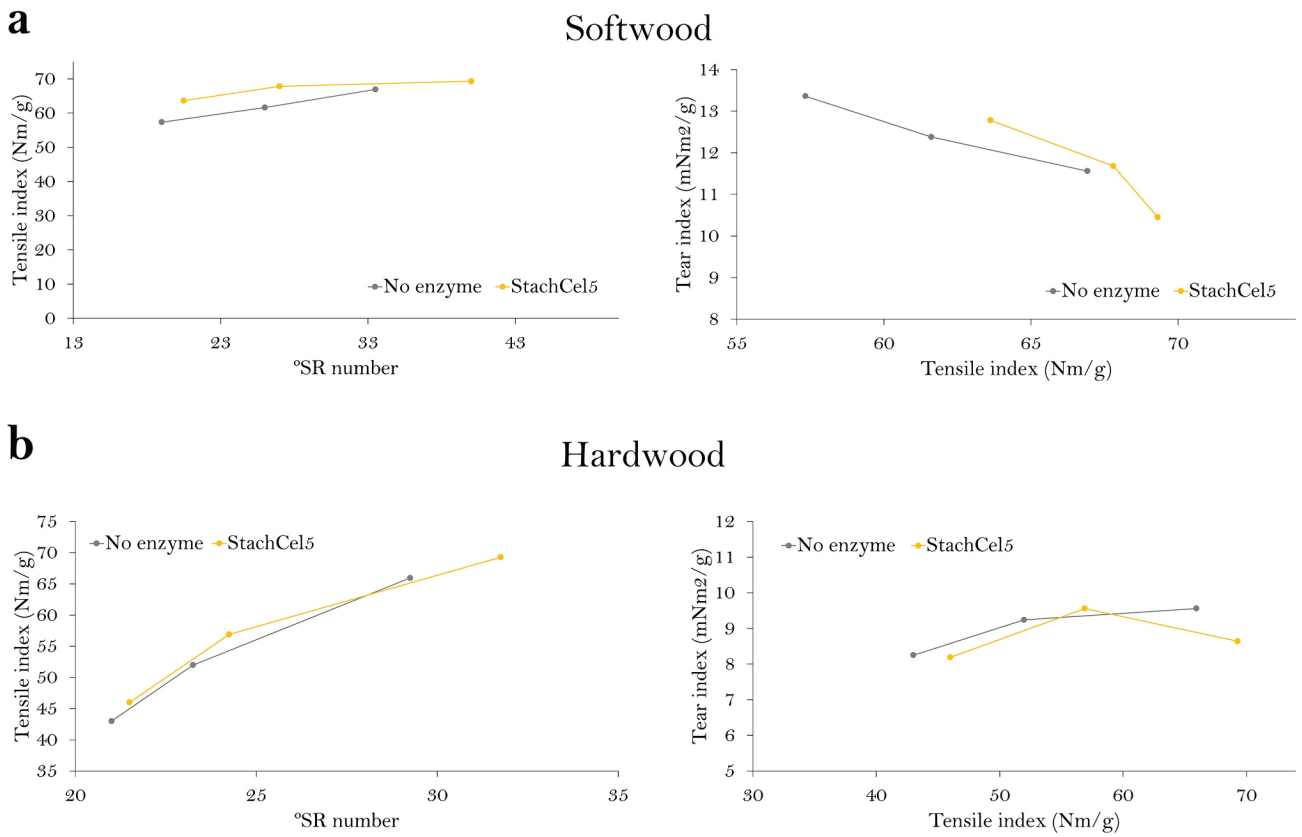


Fig. 7. Comparison of tensile index, °SR number and tear index. Tensile index as a function of pulp drainability (°SR) and tear index versus tensile index for handsheets produced from control (no enzyme) and StachCel5-treated pulps of (a) softwood and (b) hardwood. Increasing PFI mill revolutions are represented from left to right.

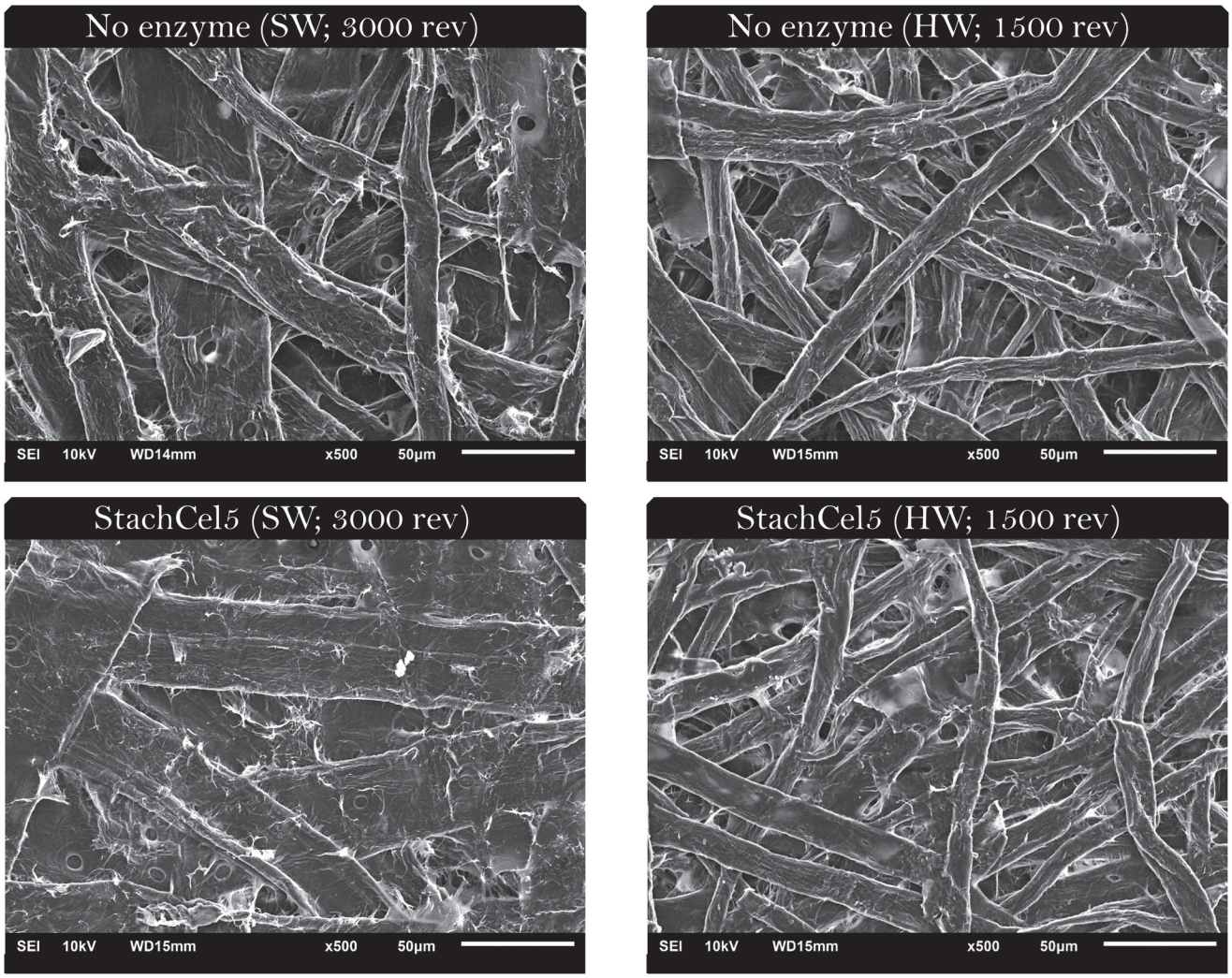


Fig. 8. Scanning electron microscopy (SEM) images of refined paper sheets produced from softwood (SW) and hardwood (HW) pulps. Handsheets were obtained after laboratory PFI refining at 3000 revolutions for SW and 1500 revolutions for HW.

development by reducing refining intensity without compromising key mechanical properties.

In contrast to tensile and burst indices, the tear index primarily reflects the intrinsic strength of individual fibers and typically declines with increased refining intensity, due to fiber shortening and degradation. This effect can be further amplified by cellulase activity, which may exacerbate fiber damage through additional hydrolysis [6,69,70]. To assess potential impacts on fiber integrity, we evaluated both the tear index and the zero-span tensile index of the resulting handsheets. In softwood pulp, enzymatic treatment produced only minor changes in the tear index, while the zero-span tensile index remained largely unchanged, indicating that intrinsic fiber strength was generally preserved (Fig. 6a). Notably, StachCel5 enabled higher tear index values at a given tensile index relative to the control, and at reduced PFI revolutions (Fig. 7a). The response in hardwood pulp differed, displaying a distinct performance profile (Fig. 7b). In this case, the tear and zero-span tensile indices remained like those of the control, indicating that StachCel5 facilitates an increase in tensile index at a given tear index, which also implies potential energy savings (Fig. 6b).

The results presented in this study suggest a fiber-type-dependent response to enzymatic treatment, with softwood pulps exhibiting greater potential for performance enhancement under the applied conditions. Such differential responses have been previously reported and are likely influenced by differences in substrate composition and fiber morphology, factors that play a critical role in determining enzymatic efficiency and outcome [8,11]. This may be partially attributed to the higher hemicellulose content typically found in hardwood compared to softwood, although it also depends on the organism source [71,72]. Since hemicellulose is known to inhibit cellulase activity by limiting its access to cellulose, its abundance could explain the reduced effect of enzymatic treatment on the refining degree in hardwood relative to softwood [73].

Overall, our findings are consistent with previous studies demonstrating the effectiveness of enzymatic pre-treatment in pulp biorefining. Comparable to commercial fungal cellulases (e.g., from *T. reesei*), StachCel5 enhanced both tensile and burst indices while enabling approximately 25 % reduction in refining energy for softwood pulp. Notably, at a low enzyme dosage (0.5 CMCase/g dried pulp, 30 min), StachCel5 achieved comparable or superior improvements in strength properties without compromising tear index, a common drawback of more aggressive cellulase treatments [6,8,11,71].

4. Conclusions

This study reports the discovery and characterization of StachCel5, a novel GH5 *endo*-β-1,4-glucanase from *S. chartarum*, identified via genome mining and heterologously expressed in *K. phaffii*. The enzyme exhibited broad pH stability (pH 4–8), thermostability at 50 °C, and high

specific activity on carboxymethyl cellulose, indicating strong suitability for industrial applications. Functionally, StachCel5 enhanced pulp refining efficiency, particularly in softwood pulp, by promoting fiber fibrillation and inter-fiber bonding while reducing mechanical energy input. Importantly, it preserved intrinsic fiber strength, maintaining or improving tear index at equivalent tensile strength levels. These outcomes suggest that StachCel5 improves fiber quality without inducing excessive fiber degradation, distinguishing it from other fungal endoglucanases. StachCel5 combines robust biochemical properties with favorable performance in pulp biorefining, offering a promising biocatalyst for energy-efficient, strength-preserving enzymatic treatment of lignocellulosic materials under industrially relevant conditions.

CRediT authorship contribution statement

Jordi Ferrando: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Conceptualization. **Clàudia Lliso-Pascual:** Writing – review & editing, Methodology, Investigation. **Oriol Cusola:** Resources, Conceptualization. **M. Blanca Roncero:** Writing – review & editing, Supervision, Resources, Conceptualization. **Antoni Planas:** Writing – review & editing, Supervision, Resources, Conceptualization. **Pere Picart:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Conceptualization.

Funding

This work was supported by the Pla de Doctorats Industrials del Departament de Recerca i Universitats de la Generalitat de Catalunya with the support of Gestio d'Ajuts Universitaris de Recerca, with grant number 2021 DI 77, awarded to J.F, and grant number 2022 DI 59, awarded to C.L.P.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationship that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2025.145969.

Data availability

Data will be made available on request.

References

[1] R. Datta, Enzymatic degradation of cellulose in soil: a review, Heliyon 10 (2024) e24022, https://doi.org/10.1016/j.heliyon.2024.e24022.
[2] M.G. Adsul, Cellulolytic enzymes recycling strategies for the economic conversion of lignocellulosic biomass to fuels, Process Biochem. 147 (2024) 62–74, https://doi.org/10.1016/j.procbio.2024.08.009.
[3] S. Pérez-Contreras, D.A. Avalos-de la Cruz, M.A. Lizardi-Jiménez, J.A. Herrera-Corredor, O. Baltazar-Bernal, R. Hernández-Martínez, Production of Ligninolytic and Cellulolytic Fungal Enzymes for Agro-Industrial Waste Valorization: Trends and Applicability, Catal 15 (2025) 30, 15 (2024) 30, https://doi.org/10.3390/catal15010030.
[4] U. Ejaz, M. Sohail, A. Ghanemi, Cellulases: from bioactivity to a variety of industrial applications, Biomimetics 6 (2021) 44, https://doi.org/10.3390/biomimetics6030044.
[5] P. Sutaoney, S.N. Rai, S. Sinha, R. Choudhary, A.K. Gupta, S.K. Singh, P. Banerjee, Current perspective in research and industrial applications of microbial cellulases, Int. J. Biol. Macromol. 264 (2024), https://doi.org/10.1016/j.ijbiomac.2024.130639.
[6] A. Kumar, A. Tazeb, C. Ram, Enzyme-assisted pulp refining: an energy saving approach, Phys. Sci. Rev. 6 (2021), https://doi.org/10.1515/psr-2019-0046.
[7] O. Saira, C. Vauhkonen, Cellulase-assisted refining in a paperboard mill: avoided emissions from energy savings. A case study of a Finnish paperboard mill, Cleaner Engineering and Technology 20 (2024) 100760, https://doi.org/10.1016/j.clet.2024.100760.
[8] M. Nagl, O. Haske-Cornelius, W. Bauer, G.S. Nyanhongo, G.M. Guebitz, Enhanced energy savings in enzymatic refining of hardwood and softwood pulp, Energy Sustain. Soc. 13 (2023) 1–15, https://doi.org/10.1186/s13705-023-00398-0/metrics.
[9] M. Yang, J. Li, S. Wang, F. Zhao, C. Zhang, C. Zhang, S. Han, Status and trends of enzyme cocktails for efficient and ecological production in the pulp and paper industry, J. Clean. Prod. 418 (2023) 138196, https://doi.org/10.1016/j.jclepro.2023.138196.
[10] S. Ghosh, L. Godoy, K.Y. Anchang, C.C. Achilonu, M. Gryzenhout, Fungal Cellulases: Current research and future challenges, in: A. Singh, M. Gupta, A. Passari (Eds.), Industrially Important Fungi for Sustainable Development: Volume 2: Bioprospecting for Biomolecules, Springer, Cham, 2021, pp. 263–298.
[11] M. Nagl, O. Haske-Cornelius, L. Skopek, A. Pellis, W. Bauer, G.S. Nyanhongo, G. Guebitz, Biorefining: the role of endoglucanases in refining of cellulose fibers, Cellulose 28 (2021) 7633–7650, https://doi.org/10.1007/s10570-021-04022-2.
[12] M. Wang, J. Du, D. Zhang, X. Li, J. Zhao, Modification of Different Pulps by Homologous Overexpression Alkali-Tolerant Endoglucanase in *Bacillus subtilis* Y106, Sci. Reports 71 (7) (2017) 1–10, https://doi.org/10.1038/s41598-017-03215-9 (2017).
[13] H. El-Gendi, A.K. Saleh, R. Badierah, E.M. Redwan, Y.A. El-Maradny, E.M. El-Fakharany, A comprehensive insight into fungal enzymes: structure, classification, and their role in mankind's challenges, J. Fungi 8 (2022), https://doi.org/10.3390/jof8010023.
[14] B.B. Jarvis, Stachybotrys chartarum: a fungus for our time, Phytochemistry 64 (2003) 53–60, https://doi.org/10.1016/s0031-9422(03)00275-9.
[15] S.R.M. Ibrahim, H. Choudhry, A.H. Asseri, M.A. Elfaky, S.G.A. Mohamed, G. A. Mohamed, Stachybotrys chartarum—A Hidden Treasure: Secondary Metabolites, Bioactivities, and Biotechnological Relevance, J. Fungi 8 (2022) 504, 2022. 504. 8, https://doi.org/10.3390/jof8050504.
[16] P. Picart, P. Diaz, F.I.J. Pastor, Stachybotrys atra BP-A produces alkali-resistant and thermostable cellulases, Antonie Van Leeuwenhoek 94 (2008) 307–316, https://doi.org/10.1007/s10482-008-9248-9.
[17] P. Picart, F. Goedegebuur, P. Díaz, F.I. Javier Pastor, Expression of novel β-glucanase Cel12A from Stachybotrys atra in bacterial and fungal hosts, Fungal Biol. 116 (2012) 443–451, https://doi.org/10.1016/j.funbio.2012.01.004.
[18] J. Semeiks, D. Borek, Z. Otwinowski, N.V. Grishin, Comparative genome sequencing reveals chemotype-specific gene clusters in the toxigenic black mold Stachybotrys, BMC Genomics 15 (2014), https://doi.org/10.1186/1471-2164-15-590.
[19] S. Ariaeenejad, J. Gharechahi, M. Foroozandeh Shahraki, F. Fallah Atanaki, J. L. Han, X.Z. Ding, F. Hildebrand, M. Bahram, K. Kavousi, G. Hosseini Salekdeh, Precision enzyme discovery through targeted mining of metagenomic data, Nat. Prod. Bioprospect. 141 (14) (2024) 1–17, https://doi.org/10.1007/S13659-023-00426-8.
[20] A.Y. Chang, V. Chau, J.A. Landas, Y. Pang, Preparation of calcium competent *Escherichia coli* and heat-shock transformation, JEMI Methods 1 (2017) 22–25.
[21] A. Bernardes, V.O.A. Pellegrini, F. Curtolo, C.M. Camilo, B.L. Mello, M.A. Johns, J. L. Scott, F.E.C. Guimaraes, I. Polikarpov, Carbohydrate binding modules enhance cellulose enzymatic hydrolysis by increasing access of cellulases to the substrate, Carbohydr. Polym. 211 (2019) 57–68, https://doi.org/10.1016/j.carbpol.2019.01.108.
[22] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R. C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J. S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI'S conserved domain database, Nucleic Acids Res. 43 (2015) D222–D226, https://doi.org/10.1093/nar/gku1221.
[23] F. Teufel, J.J. Almagro Armenteros, A.R. Johansen, M.H. Gíslason, S.I. Pihl, K. D. Tsirigos, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 6.0 predicts all five types of signal peptides using protein language models, Nat. Biotechnol. 40 (2022) 1023–1025, https://doi.org/10.1038/s41587-021-01156-3.
[24] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31 (2003) 3784–3788, https://doi.org/10.1093/nar/gkg563.
[25] R. Gupta, S. Brunak, Prediction of glycosylation across the human proteome and the correlation to protein function, Pac. Symp. Biocomput. (2002) 310–322, https://doi.org/10.1142/9789812799623_0029.
[26] K. Tamura, G. Stecher, S. Kumar, MEGA11: molecular evolutionary genetics analysis version 11, Mol. Biol. Evol. 38 (2021) 3022–3027, https://doi.org/10.1093/molbev/msab120.
[27] H.R. Powell, S.A. Islam, A. David, M.J.E. Sternberg, Phyre2.2: a community resource for template-based protein structure prediction, J. Mol. Biol. (2025) 168960, https://doi.org/10.1016/j.jmb.2025.168960.
[28] M.H. Høie, E.N. Kiehl, B. Petersen, M. Nielsen, O. Winther, H. Nielsen, J. Hallgren, P. Marcatili, NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning, Nucleic Acids Res. 50 (2022) W510–W515, https://doi.org/10.1093/nar/gkac439.
[29] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.

[30] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all, Nat. Methods 19 (2022) 679–682, https://doi.org/10.1038/s41592-022-01488-1.
[31] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T.L. Madden, NCBI BLAST: a better web interface, Nucleic Acids Res. 36 (2008) W5–W9, https://doi.org/10.1093/nar/gkn201.
[32] C. Notredame, D.G. Higgins, J. Heringa, T-coffee: a novel method for fast and accurate multiple sequence alignment, J. Mol. Biol. 302 (2000) 205–217, https://doi.org/10.1006/jmbi.2000.4042.
[33] J. Lin-Cereghino, W.W. Wong, S. Xiong, W. Giang, L.T. Luong, J. Vu, S.D. Johnson, G.P. Lin-Cereghino, Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris, Biotechniques 38 (2005) 44–48, https://doi.org/10.2144/05381bm04.
[34] I.P. Wood, A. Elliston, P. Ryden, I. Bancroft, I.N. Roberts, K.W. Waldron, Rapid quantification of reducing sugars in biomass hydrolysates: improving the speed and precision of the dinitrosalicylic acid assay, Biomass Bioenergy 44 (2012) 117–121, https://doi.org/10.1016/j.biombioe.2012.05.003.
[35] H. Aspeborg, P.M. Coutinho, Y. Wang, H. Brumer, B. Henrissat, Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5), BMC Evol. Biol. 12 (2012) 1–16, https://doi.org/10.1186/1471-2148-12-186.
[36] A. Micsonai, É. Moussong, F. Wien, E. Boros, H. Vadászi, N. Murvai, Y.H. Lee, T. Molnár, M. Réfrégiers, Y. Goto, Á. Tantos, J. Kardos, BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy, Nucleic Acids Res. 50 (2022) W90–W98, https://doi.org/10.1093/nar/gkac345.
[37] A.I. Matinja, N.H.A. Kamarudin, A.T.C. Leow, S.N. Oslan, M.S.M. Ali, Structural insights into cold-active lipase from Glaciozyma antarctica PI12: Alphafold2 prediction and molecular dynamics simulation, J. Mol. Evol. 92 (2024), https://doi.org/10.1007/s00239-024-10219-3.
[38] E. Drula, M.L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon, The carbohydrate-active enzyme database: functions and literature, Nucleic Acids Res. 50 (2022) D571–D577, https://doi.org/10.1093/nar/gkab1045.
[39] A. Planas, Glycoside Hydrolases: Mechanisms, Specificities, and Engineering, in: Glycoside Hydrolases, Academic Press, 2023, pp. 25–53.
[40] A.E. Naas, A.K. MacKenzie, B. Dalhus, V.G.H. Eijsink, P.B. Pope, Structural Features of a Bacteroidetes-Affiliated Cellulase Linked with a Polysaccharide Utilization Locus, Sci. Reports 51 (5) (2015) 1–12, https://doi.org/10.1038/srep11666 (2015).
[41] M. Lafond, G. Sulzenbacher, T. Freyd, B. Henrissat, J.G. Berrin, M.L. Garron, The quaternary structure of a glycoside hydrolase dictates specificity toward β-glucans, J. Biol. Chem. 291 (2016) 7183, https://doi.org/10.1074/JBC.M115.695999.
[42] G.D. Barone, A. Emmerstorfer-Augustin, A. Biundo, I. Pisano, P. Coccetti, V. Mapelli, A. Camattari, Industrial Production of Proteins with Pichia pastoris—Komagataella phaffii, Biomol 13 (2023) 441, 2023. 441. 13, https://doi.org/10.3390/biom13030441.
[43] A. Várnai, C. Tang, O. Bengtsson, A. Atterton, G. Mathiesen, V.G.H. Eijsink, Expression of endoglucanases in Pichia pastoris under control of the GAP promoter, Microb. Cell Factories 13 (2014) 57, https://doi.org/10.1186/1475-2859-13-57.
[44] A.V. Bernardi, D.K. Yonamine, S.A. Uyemura, T.M. Dinamarco, A thermostable *Aspergillus fumigatus* GH7 endoglucanase over-expressed in *Pichia pastoris* stimulates lignocellulosic biomass hydrolysis, Int. J. Mol. Sci. 20 (9) (2019) 2261, https://doi.org/10.3390/ijms20092261.
[45] B. Radoman, C. Grünwald-Gruber, B. Schmelzer, D. Zavec, B. Gasser, F. Altmann, D. Mattanovich, B. Radoman, B. Schmelzer, B. Gasser, F. Altmann, D. Mattanovich, D. Zavec, C. Grünwald-Gruber, The degree and length of O-glycosylation of recombinant proteins produced in Pichia pastoris depends on the nature of the protein and the process type, Biotechnol. J. 16 (2021) 2000266, https://doi.org/10.1002/biot.202000266.
[46] J. Zheng, H.-Q. Liu, X. Qin, K. Yang, J. Tian, X.-L. Wang, Y.-R. Wang, Y. Wang, B. Yao, H.-Y. Luo, H.-Q. Huang, Identification and mutation analysis of nonconserved residues on the TIM-barrel surface of GH5_5 Cellulases for catalytic efficiency and stability improvement, Appl. Environ. Microbiol. 88 (2022), https://doi.org/10.1128/aem.01046-22 e01046–22.
[47] J.J. Escuder-Rodríguez, M. González-Suarez, M.E. deCastro, A. Saavedra-Bouza, M. Becerra, M.I. González-Siso, Characterization of a novel thermophilic metagenomic GH5 endoglucanase heterologously expressed in *Escherichia coli* and *Saccharomyces cerevisiae*, Biotechnol. Biofuels Bioprod. 15 (2022) 1–13, https://doi.org/10.1186/s13068-022-02172-4.
[48] B.R. Dave, A.P. Sudhir, R.B. Subramanian, Purification and properties of an endoglucanase from Thermoascus aurantiacus, Biotechnol. Reports 6 (2015) 85–90, https://doi.org/10.1016/j.btre.2014.11.004.
[49] S. Samanta, A. Basu, U.C. Halder, S.K. Sen, Characterization of Trichoderma reesei endoglucanase ii expressed heterologously in Pichia pastoris for better biofinishing and biostoning, J. Microbiol. 50 (2012) 518–525, https://doi.org/10.1007/S12275-012-1207-5.
[50] A. Patel, A. Shah, Purification and characterization of novel, thermostable and non-processive GH5 family endoglucanase from Fomitopsis meliae CFA 2, Int. J. Biol. Macromol. 182 (2021) 1161–1169, https://doi.org/10.1016/j.ijbiomac.2021.04.110.
[51] Z. Li, X. Pei, Z. Zhang, Y. Wei, Y. Song, L. Chen, S. Liu, S.H. Zhang, The unique GH5 cellulase member in the extreme halotolerant fungus aspergillus glaucus CCHA is an endoglucanase with multiple tolerance to salt, alkali and heat: prospects for straw degradation applications, Extremophiles 22 (2018) 675–685, https://doi.org/10.1007/S00792-018-1028-5.
[52] W. Zhou, S. Tong, F.R. Amin, W. Chen, J. Cai, D. Li, Heterologous Expression and Biochemical Characterization of a Thermostable Endoglucanase (MtEG5–1) from Myceliophthora thermophila, Ferment 9 (2023) 462, 9 (2023) 462, https://doi.org/10.3390/fermentation9050462.
[53] G. Liu, Q. Li, N. Shang, J.W. Huang, T.P. Ko, W. Liu, Y. Zheng, X. Han, Y. Chen, C. C. Chen, J. Jin, R.T. Guo, Functional and structural analyses of a 1,4-β-endoglucanase from Ganoderma lucidum, Enzym. Microb. Technol. 86 (2016) 67–74, https://doi.org/10.1016/j.enzmictec.2016.01.013.
[54] A.S. Dotsenko, A.M. Rozhkova, A.V. Gusakov, Properties and N-glycosylation of recombinant endoglucanase II from Penicillium verruculosum, Mosc. Univ. Chem. Bull. 70 (2015) 283–286, https://doi.org/10.3103/S0027131415060024.
[55] S. Takashima, A. Nakamura, H. Masaki, T. Uozumi, Cloning, sequencing, and expression of a thermostable cellulase gene of humicola grisea, Biosci. Biotechnol. Biochem. 61 (1997) 245–250, https://doi.org/10.1271/bbb.61.245.
[56] A. Neis, L. da Silva Pinto, Glycosyl hydrolases family 5, subfamily 5: relevance and structural insights for designing improved biomass degrading cocktails, Int. J. Biol. Macromol. 193 (2021) 980–995, https://doi.org/10.1016/j.ijbiomac.2021.10.062.
[57] E.M. Glasgow, K.A. Vander Meulen, T.E. Takasuka, C.M. Bianchetti, L.F. Bergeman, S. Deutsch, B.G. Fox, Extent and origins of functional diversity in a subfamily of glycoside hydrolases, J. Mol. Biol. 431 (2019) 1217–1233, https://doi.org/10.1016/j.jmb.2019.01.024.
[58] N. Srivastava, M. Srivastava, P.K. Mishra, V.K. Gupta, G. Molina, S. Rodriguez-Couto, A. Manikanta, P.W. Ramteke, Applications of fungal cellulases in biofuel production: Advances and limitations 82 (2018) 2379–2386, https://doi.org/10.1016/j.rser.2017.08.074.
[59] S. Gharehkhani, E. Sadeghinezhad, S.N. Kazi, H. Yarmand, A. Badarudin, M. R. Safaei, M.N.M. Zubir, Basic effects of pulp refining on fiber properties—a review, Carbohydr. Polym. 115 (2015) 785–803, https://doi.org/10.1016/j.carbpol.2014.08.047.
[60] P. Przybysz, M. Dubowik, E. Małachowska, M. Kucner, M. Gajadhur, K. Przybysz, The effect of the refining intensity on the progress of internal fibrillation and shortening of cellulose fibers, BioResources 15 (2020) 1482–1499, https://doi.org/10.15376/biores.15.1.1482-1499.
[61] M. Krysztof, K. Olejnik, P. Kulpiński, A. Erdman, E. Sąsiadek, A comparative study of the effect of cellulose-based deep coating and pulp refining on the structural and mechanical properties of paper, BioResources 16 (2021) 5376–5389, https://doi.org/10.15376/biores.16.3.5376-5389.
[62] H. Ahadian, S. Ceccherini, E. Sharifi Zamani, J. Phiri, T. Maloney, Production of low-density and high-strength paperboards by controlled micro-nano fibrillation of fibers, J. Mater. Sci. 58 (2023) 17126–17137, https://doi.org/10.1007/S10853-023-09097-9.
[63] R. Sharma, G. Garnier, V. Haritos, A study of different actions of glucanases to modulate microfibrillated cellulose properties, Cellulose 29 (2022) 2323–2332, https://doi.org/10.1007/S10570-022-04451-7.
[64] S.G. Park, J.H. Tak, J.Y. Lee, K.S. Shin, S. Il Park, Evaluation of Cellulase effect on the refining process of softwood bleached Kraft pulp, BioResources 20 (2025) 1059–1068. https://ojs.bioresources.com/index.php/BRJ/article/view/24069.
[65] R.J. Kerekes, J.O. Heymer, J.D. McDonald, Refining pulp for tensile strength, Nord. Pulp Pap. Res. J. 36 (2021) 696–706, https://doi.org/10.1515/npprj-2021-0012.
[66] J. olson, B. allison, T. friesen, C. peters, Fiber fractionation for high-porosity sack Kraft paper, TAPPI J. 84 (2001) 66.
[67] N. Barrios, M. Gonzalez, R. Venditti, L. Pal, Synergistic cell-free enzyme cocktails for enhanced fiber matrix development: improving dewatering, strength, and decarbonization in the paper industry, Biotechnol. Biofuels Bioprod. 18 (2025) 1–23, https://doi.org/10.1186/S13068-025-02646-1.
[68] S. Bhardwaj, P. Kaur, N.K. Bhardwaj, Y.S. Negi, Surface application of different concentrations of chitosan on recycled paper and its impact on packaging properties, J. Coat. Technol. Res. 20 (2023) 1285–1298, https://doi.org/10.1007/S11998-022-00743-6.
[69] L. Cui, F. Meddeb-Mouelhi, F. Laframboise, M. Beauregard, Effect of commercial cellulases and refining on Kraft pulp properties: correlations between treatment impacts and enzymatic activity components, Carbohydr. Polym. 115 (2015) 193–199, https://doi.org/10.1016/j.carbpol.2014.08.076.
[70] T. Chen, Y. Xie, Q. Wei, X. (Alice) Wang, O. Hagman, O. Karlsson, J. Liu, Effect of refining on physical properties and paper strength of Pinus massoniana and China fir cellulose fibers, BioResources 11 (2016) 7839–7848, https://doi.org/10.15376/biores.11.3.7839-7848.
[71] O. Haske-Cornelius, A. Hartmann, F. Brunner, A. Pellis, W. Bauer, G.S. Nyanhongo, G.M. Guebitz, Effects of enzymes on the refining of different pulps, J. Biotechnol. 320 (2020) 1–10, https://doi.org/10.1016/j.jbiotec.2020.06.006.
[72] M. Nagl, O. Haske-Cornelius, L. Skopek, F. Bausch, A. Pellis, W. Bauer, G. S. Nyanhongo, G.M. Guebitz, Mechanistic investigation of the effect of endoglucanases related to pulp refining, Cellulose 29 (2022) 2579–2598, https://doi.org/10.1007/S10570-021-04386-5.
[73] S. Malgas, V.M. Kwanya Minghe, B.I. Pletschke, The effect of hemicellulose on the binding and activity of cellobiohydrolase I, Cel7A, from Trichoderma reesei to cellulose, Cellulose 27 (2020) 781–797, https://doi.org/10.1007/S10570-019-02848-5.

From Mold to Mill: StachCel5, a novel thermoalkaliphilic endoglucanase from *Stachybotrys chartarum* for pulp biorefining

Jordi Ferrando[1], Clàudia Lliso Pascual[2], Oriol Cusola[3], M. Blanca Roncero[3], Antoni Planas[2,4], and Pere Picart[1]

[1] Faculty of Pharmacy and Food Science Technology, Department of Biology, Healthcare and the Environment, Microbiology Section, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain

[2] Laboratory of Biochemistry, Institut Químic de Sarrià, University Ramon Llull, Via Augusta 390, 08017 Barcelona, Spain

[3]CELBIOTECH_Paper Engineering Research Group, Universitat Politècnica de Catalunya_BarcelonaTech, 08222 Terrassa, Spain

[4] Chemistry section, Royal Academy of Sciences and Arts of Barcelona, La Rambla 115, 08002 Barcelona, Spain

ATGAAGACTTTAGCATTTTTAGGAGTACTTGCCGGTACAGTTTCTGCTCAAGCAACAGCAGGACCTTGG
GCTCAGTGTGGAGGAATGAATTGGACTGGTGCCACAGTATGCACCTCTGGTCACACATGCTCTTTTATC
AACGATTGGTACTCCCAGTGCATACCAGGTACACCTACTACTACCAGAGTTACATCATCAACTCGAACC
ACCATCTCTACTTCCACCCGTACTTCTTCTGCTAGTAGAACCTCTACCACCTCTACCACTACATCCTCAGTC
CCTGCTCCTACAGGTTTCCGTTGGTTTGGAGTTGATGAGTCTGTTGCTGAATTTGGTCAAGGATACTATC
CTGGCGAATATGGAACTCATTTCCGTTTTCCTGATGAAGCTGCTATTAGTACACTGATTGGTGAAGGAA
TGAATACCTTCAGGATCAGTTTTGCTATGGAAAGGCTGTCTCCCAATGGCTTGACAGGAGCCTTTGATG
CTGGTTACTTGGCTAATTTAACACGAACGATCAACTACGTTACTGGACGTGGTGTTTATGCTGTTTTGGA
CCCTCATAACTATGGACGATATAGGGGTAACATTATTACAGATGTGACCGCATTTGGAGCATGGTGGG
AGAGGCTTGCAACAGGATTTAGGACCAATCAAAGAGTGATCTTCGATACTAATAACGAGTACCACACA
ATGGACCAACAGCTGGTTTTTAATCTTAATCAAGCAGCTATCGATGCAATCAGAAGAGTTGGCGCTACT
TCTCAATGGATTTGGGTAGAAGGAAATCAATGGACAGGAGCATGGTCTTGGGTCGATGTTAATGACTC
ACTGAAAGCATTGACCGATCCTCAGAACTTGCTTGTTTACCAGATGCATCAGTATTTGGATTCCGATTCC
TCAGGTACCCATGAGGAATGTGTTTCCGGTACTATTGGCGTTGAGAGAGTCAGAAGGGCCACTCAATG
GCTGAGAGATAACAACAAGATAGGAGTTATTGGAGAATTTGCTGGAGCAGCCAACCCTACATGTCGTA
CTGCTGTTACCGGTTTATTGGCTTTCCTTAGAGAAAATAGTGACGTTTGGACCGGTGCTTTATGGTGGG
CAGCTGGTCCTTGGTGGTCCGACTACATGTACTCTTACGAACCACCTTCCGGTGCTGCCTACGTCTACTA
TAACTCTCTATTGCGTCAGTATGCCATAAGATAA

Fig. S1 Codon-optimized *stachCel5* gene for expression in *Komagataella phaffii*. The predicted signal peptide is highlighted in yellow.

Hypothetical protein S7711_02738 [Stachybotrys chartarum IBT 7711] → GH1 family

Hypothetical protein S7711_03125 [Stachybotrys chartarum IBT 7711] → GH1 family

Hypothetical protein S7711_02093 [Stachybotrys chartarum IBT 7711] → GH7 family

Hypothetical protein S7711_03903 [Stachybotrys chartarum IBT 7711] → GH6 family

Hypothetical protein S7711_03103 [Stachybotrys chartarum IBT 7711] → GH5 family

Fig. S2 Putative cellulases from Stachybotrys chartarum IBT 7711 retrieved from the NCBI database. Accession numbers corresponds to predicted proteins containing glycosyl hydrolase catalytic domains and carbohydrate-binding domains (CBDs). Proteins S7711_02738 and S7711_03125 were excluded based on their glycosyl hydrolase families not being associated with endo-β-1,4-glucanase activity. BLASTp analysis indicated that S7711_02093 and S7711_03903 share high sequence similarity with exoglucanases, whereas S7711_03103 showed greater homology to known endoglucanases. Given the interest in identifying a novel endoglucanase, S7711_03103 was selected for further characterization in this study.

MKTLAFLGVLAGTVSAQATAGPWAQCGGMNWTGATVCTSGHTCSFINDWYSQCIPGTPTTTRVTSSTRT
TISTSTRTSSASRTSTTSTTTSSVPAPTGFRWFGVDESVAEFGQGYYPGEYGTHFRFPDEAAISTLIGEGMNTF
RISFAMERLSPNGLTGAFDAGYLANLTRTINYVTGRGVYAVLDPHNYGRYRGNIITDVTAFGAWWERLATG
FRTNQRVIFDTNNEYHTMDQQLVFNLNQAAIDAIRRVGATSQWIWVEGNQWTGAWSWVDVNDSLKAL
TDPQNLLVYQMHQYLDSDSSGTHEECVSGTIGVERVRRATQWLRDNNKIGVIGEFAGAANPTCRTAVTGL
LAFLRENSD VWTGALWWAAGPWWSDYMYSYEPPSGAAYVYYNSLLRQYAIR

Fig. S3 StachCel5 aminoacid sequence. Signal peptide is highlighted in yellow.

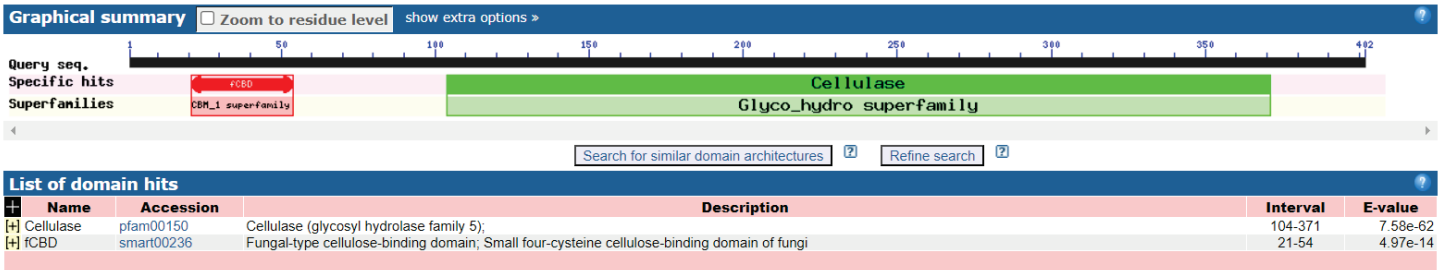


Fig. S4 Structural analysis of StachCel5. Schematic representation of conserved domains in the StachCel5 protein, highlighting the glycoside hydrolase family 5 (GH5) catalytic domain and the carbohydrate-binding domain (CBD).

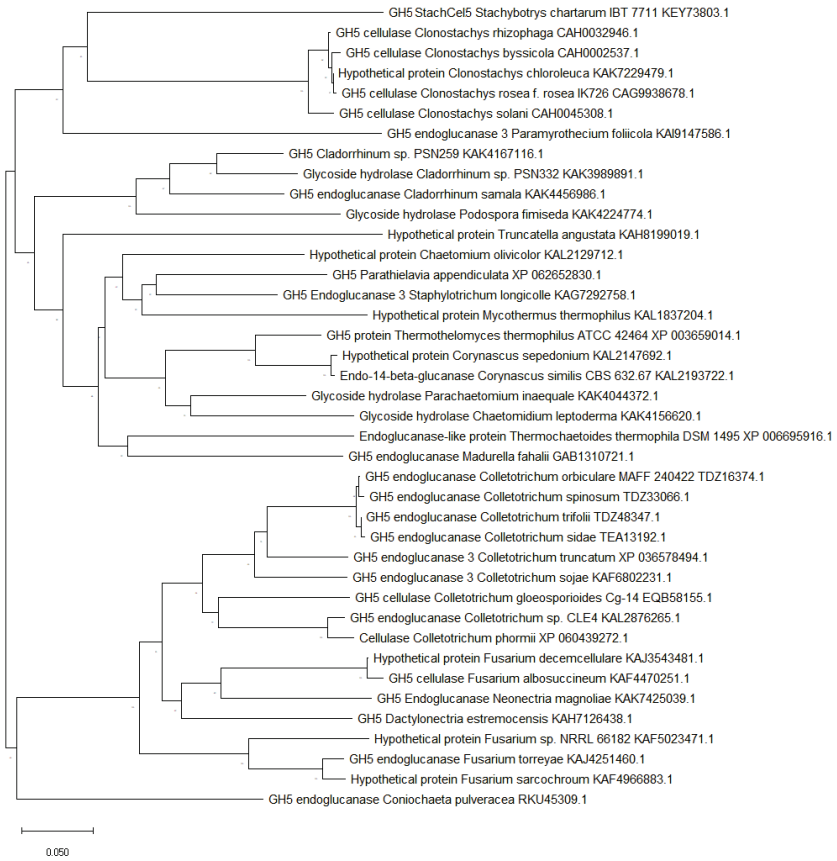


Fig. S5 Phylogenetic analysis of StachCel5. Phylogenetic tree showing the relationship between *StachCel5* and closely related GH5 proteins. The 40 non-*Stachybotrys* sequences were retrieved from the NCBI database as the top BLASTP hits against the StachCel5 protein sequence.

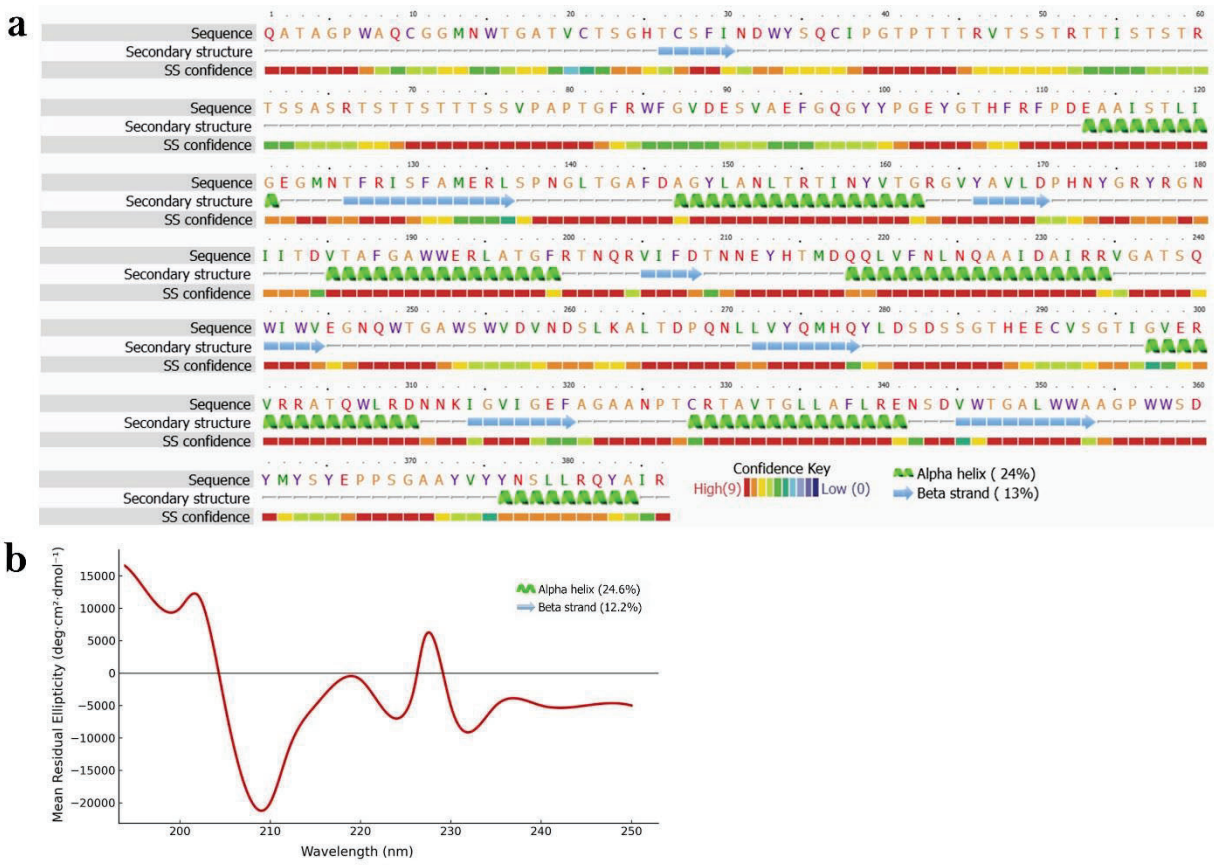


Fig. S6 Predicted secondary structure of StachCel5. (a) Secondary structure content predicted using Phyre2, showing the distribution of α-helices, β-sheets, and coil regions along the amino acid sequence. **(b)** Circular dichroism (CD) spectrum of StachCel5.

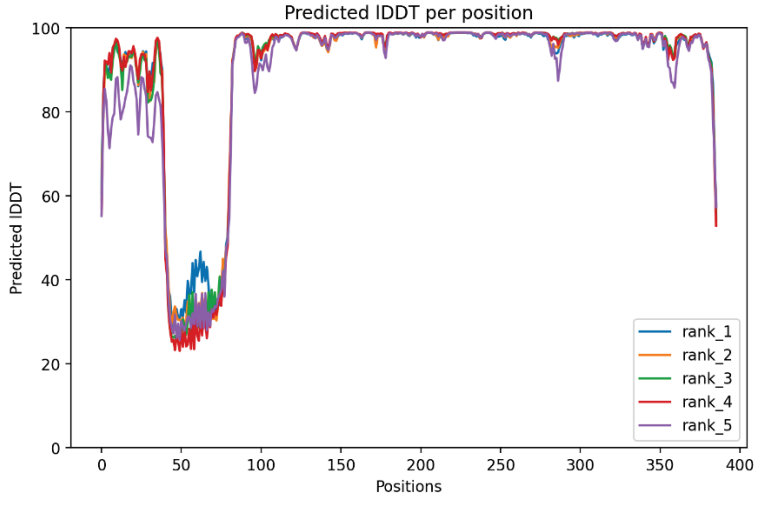


Fig. S7 Evaluation of the predicted 3D Alphafold2 models of the StachCel5 enzyme. Residues with a pLDDT score ≥ 90 indicate exceptionally high confidence, scores between 70 and 90 reflect high confidence, scores from 50 to 70 indicate low confidence, and scores below 50 correspond to very low confidence in the predicted structure.

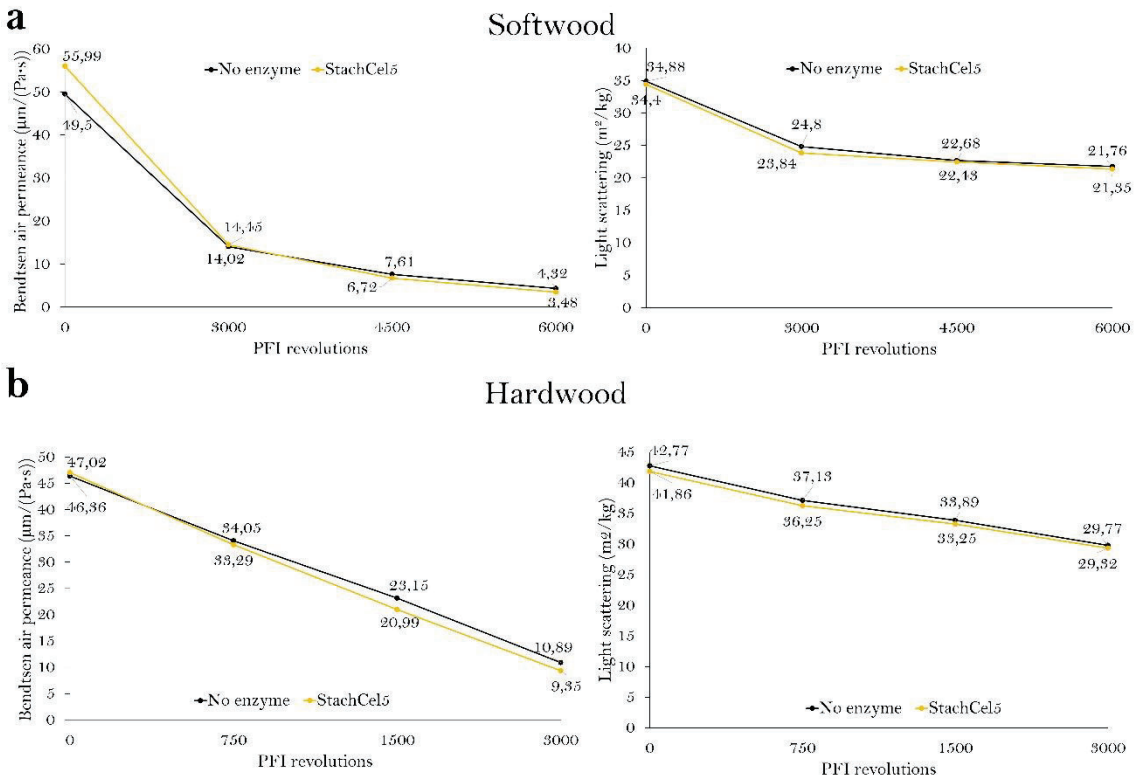


Fig. S8 Evolution of air permeability and light scattering coefficient of handsheets during refining. Measurements were taken for control (no enzyme) and *StachCel5* endoglucanase-treated samples from **(a)** softwood and **(b)** hardwood pulps.

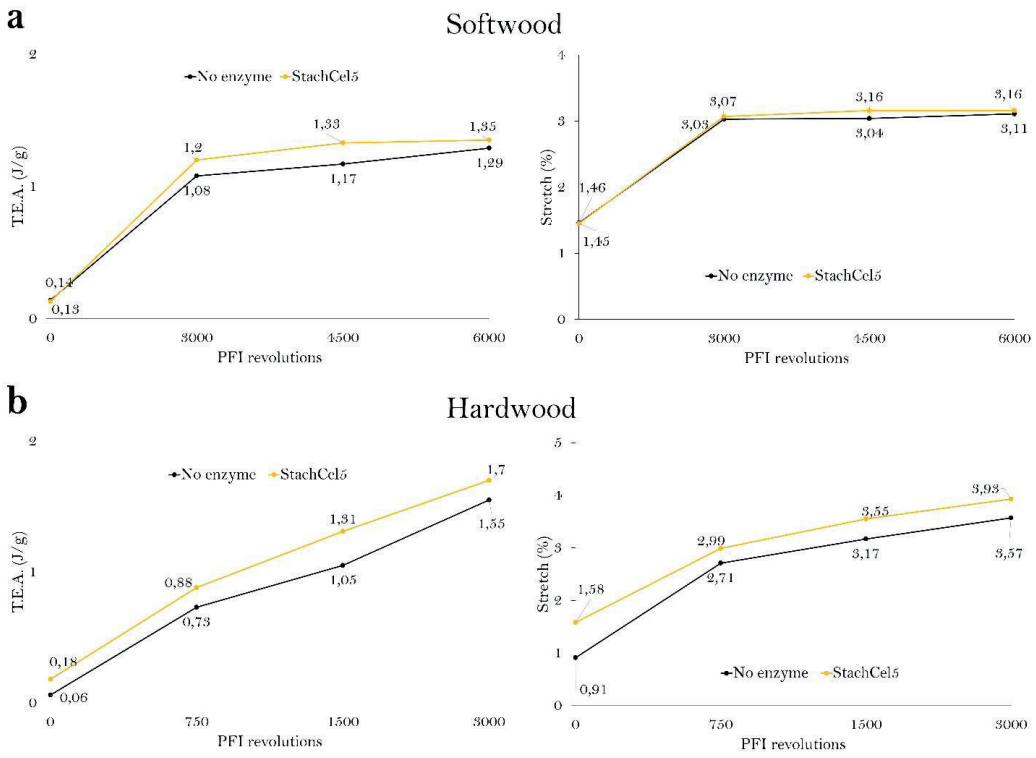


Fig. S9 Evolution of Tensile Energy Absorption (T.E.A.) and stretch (%) in handsheets during refining. Measurements compare control (no enzyme) and *StachCel5* endoglucanase-treated samples from **(a)** softwood and **(b)** hardwood pulps.

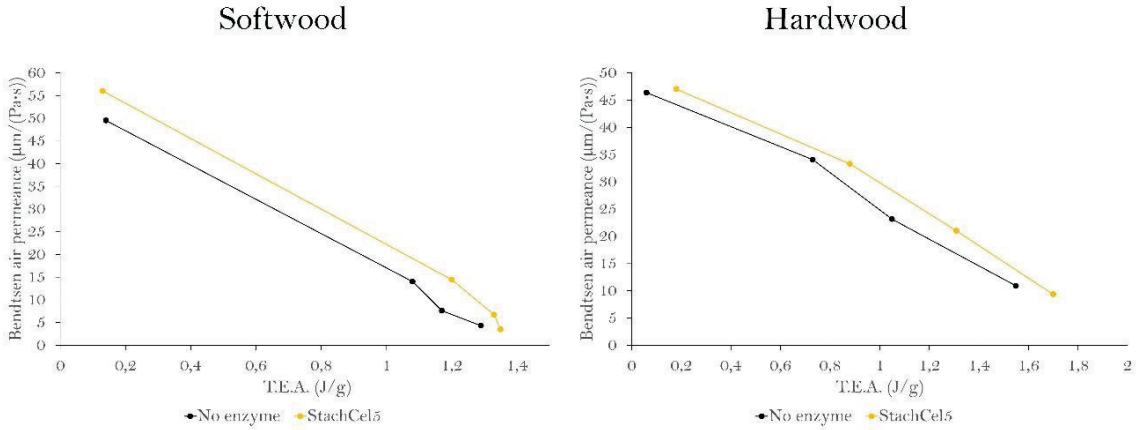


Fig. S10 Evolution of Tensile Energy Absorption (T.E.A.) and porosity performance of handsheets. Porosity was assessed by Bendtsen air permeance, with increasing PFI mill revolutions represented from left to right.

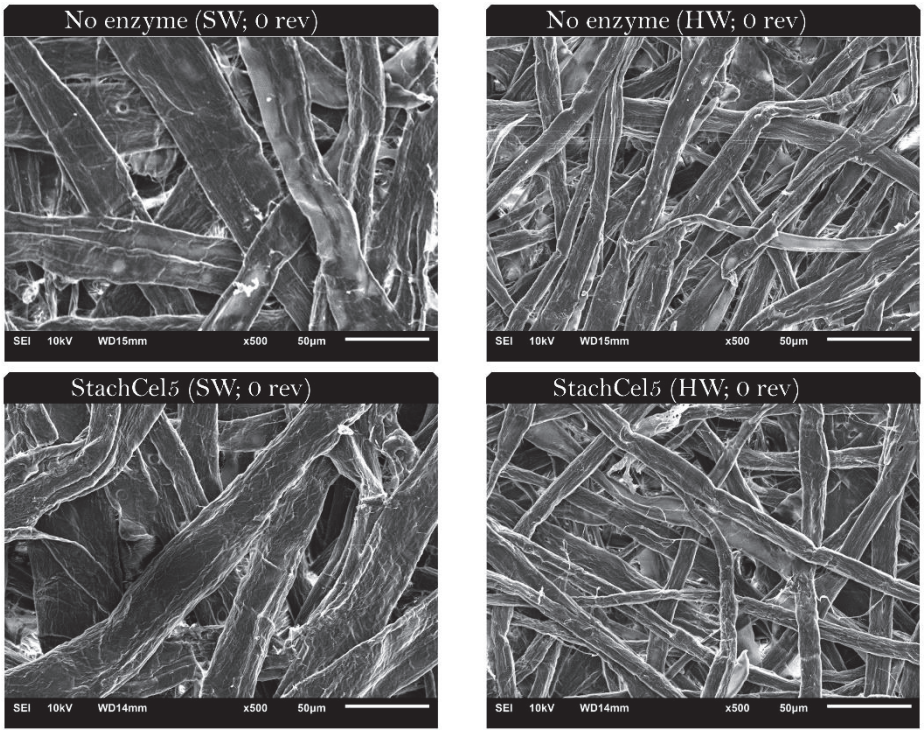


Fig. S11 Scanning-electron microscopy (SEM) images of unrefined handsheets from softwood (SW) and hardwood (HW) pulps prior to PFI refining.

4.2. Chapter 2 Construction of super-secreting *Bacillus subtilis* strains

4.2.1. Paper 2: The Construction of an Environmentally Friendly Super-Secreting Strain of *Bacillus subtilis* through Systematic Modulation of Its Secretory Pathway Using the CRISPR-Cas9 System

International Journal of
Molecular Sciences

MDPI

Article

The Construction of an Environmentally Friendly Super-Secreting Strain of *Bacillus subtilis* through Systematic Modulation of Its Secretory Pathway Using the CRISPR-Cas9 System

Jordi Ferrando, David Miñana-Galbis and Pere Picart *

Faculty of Pharmacy and Food Science Technology, Department of Biology, Healthcare and the Environment, Microbiology Section, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain; jferranu7@alumnes.ub.edu (J.F.); davidminyana@ub.edu (D.M.-G.)
* Correspondence: perepicart@ub.edu

Abstract: Achieving commercially significant yields of recombinant proteins in *Bacillus subtilis* requires the optimization of its protein production pathway, including transcription, translation, folding, and secretion. Therefore, in this study, our aim was to maximize the secretion of a reporter α-amylase by overcoming potential bottlenecks within the secretion process one by one, using a clustered regularly interspaced short palindromic repeat-Cas9 (CRISPR-Cas9) system. The strength of single and tandem promoters was evaluated by measuring the relative α-amylase activity of AmyQ integrated into the *B. subtilis* chromosome. Once a suitable promoter was selected, the expression levels of *amyQ* were upregulated through the iterative integration of up to six gene copies, thus boosting the α-amylase activity 20.9-fold in comparison with the strain harboring a single *amyQ* gene copy. Next, α-amylase secretion was further improved to a 26.4-fold increase through the overexpression of the extracellular chaperone PrsA and the signal peptide peptidase SppA. When the final expression strain was cultivated in a 3 L fermentor for 90 h, the AmyQ production was enhanced 57.9-fold. The proposed strategy allows for the development of robust marker-free plasmid-less super-secreting *B. subtilis* strains with industrial relevance.

Keywords: *Bacillus subtilis*; CRISPR-Cas9; heterologous protein; secretion pathway; bottleneck

Citation: Ferrando, J.; Miñana-Galbis, D.; Picart, P. The Construction of an Environmentally Friendly Super-Secreting Strain of *Bacillus subtilis* through Systematic Modulation of Its Secretory Pathway Using the CRISPR-Cas9 System. *Int. J. Mol. Sci.* **2024**, *25*, 6957. https://doi.org/10.3390/ijms25136957

Academic Editor: Dongsoo Yang

Received: 16 May 2024
Revised: 12 June 2024
Accepted: 18 June 2024
Published: 25 June 2024

1. Introduction

The Gram-positive bacterium *Bacillus subtilis* is widely used as a bacterial workhorse in microbial fermentations for the mass production of heterologous proteins. Several features of this bacteria are advantageous for biotechnological applications, including (i) its generally recognized as safe (GRAS) or qualified presumption of safety (QPS) status; (ii) rapid growth in usually inexpensive media, typically reaching high cell densities in large-scale fermentations; (iii) secretion of recombinant proteins into the culture medium, which simplifies its purification and eliminates the need for cell lysis; and (iv) ease of handling and manipulation, with genetically well-characterized expression systems. Therefore, the high secretion capacity of *Bacillus* species and their close relatives has enabled the development of industrial strains capable of producing enzymes at the scale of several grams per liter [1,2]. However, this capacity for the high-titer production of commercially relevant enzymes is limited by several bottlenecks within the protein production process, spanning from transcription and translation to folding and secretion, thus significantly reducing the overall yield of extracellular enzymes [3–5]. Briefly, the main bottlenecks within the secretory pathway of *B. subtilis* can be summarized as follows:

(A) At the transcriptional level, it is a pre-requisite to amplify the expression of the gene of interest, which is commonly achieved using strong or inducible promoters and

high-copy-number vectors [6,7]. However, the resulting strains harbor antibiotic resistance markers, which limits their industrial application due to legal restrictions and concerns related to antibiotic usage. Consequently, the construction of environmentally friendly multicopy strains is preferred [8–10]. (B) In the cytoplasm, the overexpressed proteins are prone to forming insoluble aggregates and inclusion bodies, rendering the protein inactive. By promoting the proper folding of proteins, which thus retain their biological activity, the overexpression of the GroE and DnaK series of intracellular chaperones has resulted in an enhanced expression of many enzymes [11–13]. (C) Within the membrane, the overexpression of exoproteins can cause jamming of the translocation machinery, thereby reducing the yield of secreted proteins. To overcome this barrier, several approaches have been adopted: (C.1) exploitation of the Tat pathway to reroute Sec-dependently secreted enzymes [14,15]; (C.2) overexpression of the translocon SecYEG [16,17]; (C.3) overexpression of the signal peptide processing enzymes SipS and SipT [18]; (C.4) overexpression of the signal peptide peptidase SppA, which cleaves remnant signal peptides left in the cell membrane after the action of SipS and SipT [19]; (C.5) overexpression of the signal peptide peptidase RasP, which prevents perturbation of the membrane and cell envelope [20]; and (C.6) cell surface engineering by modulating the expression of relevant cell surface enzymes, such as phosphatidylserine synthase (PssA) and cardiolipine synthase (ClsA), which have been shown to dramatically increase the overall presence of anionic membrane phospholipids, thus eliciting a higher secretion of enzymes [21]. (D) Once the membrane has been crossed and the extracytoplasmic compartment reached, overexpression of the major extracytoplasmic folding factor PrsA prevents proteolytic degradation while facilitating proper folding, thus enhancing the production of recombinant proteins [18,22]. A schematic representation of the aforementioned steps for the efficient expression of secreted proteins is depicted in Figure 1.

Intensive efforts have been made to surmount these limitations, focusing primarily on the engineering of expression vectors to allow for the overexpression, deletion, or genetic modification of individual steps in the overall production of specific extracellular recombinant proteins [23,24]. Although significant improvements have been reported, in many cases targeting specific genes, the use of plasmids, expensive inducers, and antibiotic selection markers hampers the application of the newly engineered strains to large-scale fermentation. Here, we outline for the first time a systematic analysis encompassing CRISPR-Cas9 genome editing modifications of most of the genes involved in the secretion of extracellular proteins, from transcription and translation to folding and secretion. As its expression is easy to measure, the α-amylase AmyQ was chosen as a readout to assess the effect of each genetic modification, with the ultimate goal of constructing a stable, environmentally friendly *B. subtilis* producer strain with practical application to industry.

Firstly, saturating levels of the specific *amyQ* message were reached by inserting six gene copies at ectopic sites within the *B. subtilis* chromosome under the control of a strong synthetic promoter. Secondly, we consecutively targeted various components of the post-transcriptional machinery, focusing on chaperones, folding processes, translocon systems, membrane stress factors, and metabolic loads, aiming to maximize the ability of *B. subtilis* to secrete the extracellular α-amylase AmyQ. Using this approach, we successfully debottlenecked the exoprotein secretion route at every level, resulting in a 26-fold increase in AmyQ expression. The ability of the newly constructed strain to produce the α-amylase AmyQ was verified using a 3 L fermentor, showing that the super-secreting engineered strain of *B. subtilis* holds commercial potential. We believe this work can provide a better understanding of the cellular mechanisms in the protein production pathway of this commonly used expression host and may offer some promising avenues for heterologous protein secretion in *B. subtilis*.

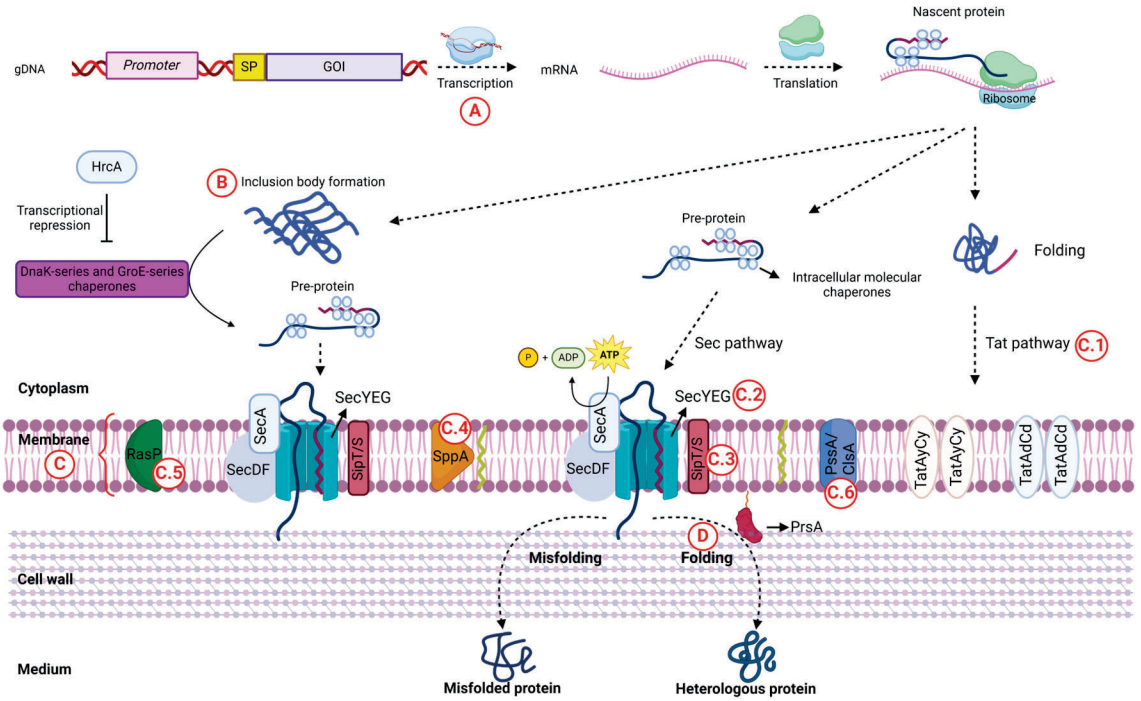


Figure 1. Schematic representation of strategies to overcome the main bottlenecks limiting the efficient expression of a secreted protein in *Bacillus subtilis*. (A) At the transcriptional level, maximum expression of the gene of interest (GOI) is achieved by using strong promoters, optimized signal peptides (SPs), and gene amplification. (B) In the cytoplasm, intracellular GroE and DnaK chaperones prevent protein aggregation. (C) Membrane modification strategies to enhance protein secretion include (C.1) exploiting the Tat pathway; (C.2) overexpressing the SecYEG translocon; (C.3–C.5) optimizing signal peptide processing of signal peptidases SipT and SipS and signal peptide peptidases SppA and RasP; and (C6) cell surface engineering by constructing PssA and ClsA mutants. Lastly, (D) overexpression of extracytoplasmic PrsA aids in proper protein folding and prevents degradation. Created with BioRender.com (accessed on 2 May 2024).

2. Results

2.1. The Effect of Single and Tandem Promoters on the Expression of Recombinant AmyQ α-Amylase in B. subtilis

The parental *B. subtilis* strain (BS1) inherently possesses a functional *amyE* gene, which encodes an extracellular amylase. Thus, we began by knocking out this gene to render the resulting BS2 strain inactive for amylase production (Figure S1a), thereby avoiding interference with the α-amylase screening method. Subsequently, BS2 was used as the initial strain for genetic manipulation. One of the most cost-effective and efficient methods to achieve high production of recombinant proteins is optimization of the promoter at the transcriptional level, as this element enables gene expression and regulation [10,25,26]. Maximizing the gene expression commonly involves using constitutive promoters. Accordingly, we initially selected the strong *spoVG* promoter to drive the expression of the α-amylase *amyQ* gene, as this widely used promoter is capable of delivering high-level gene expression in *B. subtilis* [27]. Thus, using the pJOE8999.1 plasmid as a backbone [28] and specific primers (Table S1), the pJOE3 vector was engineered for the in-frame replacement of the *spoVG* gene from strain BS2 with the *amyQ* gene, placing its expression under the control of the *spoVG* promoter (P_{spoVG}-*amyQ*) in strain BSQ1a (Figure 2a). The strength of this promoter was evaluated by culturing BSQ1a in a production medium at 37 °C and 220 rpm for 48 h, which resulted in α-amylase secretion into the medium of about 24.86 ± 1.53 U/mL (Figure 2b). This value served as a reference (control) for further analysis.

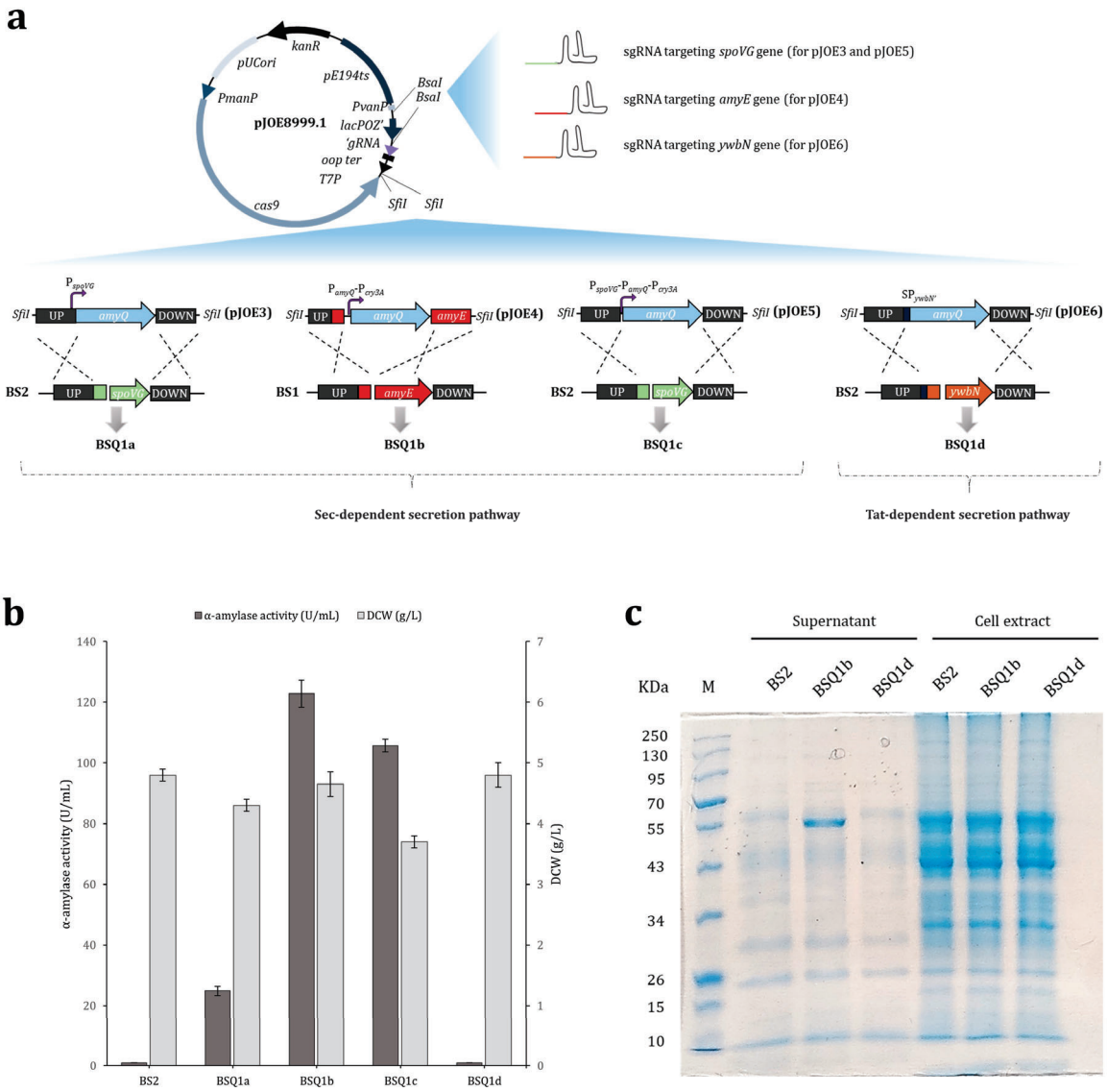


Figure 2. Construction, expression, and secretion of AmyQ α-amylase in strains BSQ1a, BSQ1b, BSQ1c, and BSQ1d. (**a**) Schematic representation of the BSQ1a-BSQ1d strain construction system using plasmids pJOE3-6. The 20 nt guide sequence targeting *spoVG* (for pJOE3 and pJOE5), *amyE* (for pJOE4), and *ywbN* (for pJOE6) was inserted into the *BsaI* sites of pJOE8999.1. The homologous repair template for each plasmid was inserted using the *SfiI* site. The BSQ1a strain harbors an *amyQ* copy at the *spoVG* locus under the control of the *spoVG* promoter (P_{spoVG}-*amyQ*). BSQ1b possesses an *amyQ* copy placed at the *amyE* locus under the control of double promoter P_{amyQ}-P_{cry3A}. In BsQ1c, the *amyQ* gene is inserted at the *spoVG* locus under the control of triple promoter P_{spoVG}-P_{amyQ}-P_{cry3A}. Integration of the *amyQ* gene containing the YwbN' signal peptide at the *ywbN* site yielded strain BSQ1d. In this strain, AmyQ secretion occurred through the Tat secretion pathway, whereas strains BSQ1a-c secreted AmyQ through the Sec pathway. (**b**) Extracellular AmyQ activity and dry cell weight (DCW) in engineered *Bacillus subtilis* BSQ1a-BSQ1d strains. BS2 was used as a control strain. The error bars represent the average ± standard deviation of three biological replicates. (**c**) SDS-PAGE analysis of supernatants and cell extracts derived from strains BS2, BSQ1b, and BSQ1d. The expected molecular weight of 55 kDa for AmyQ protein is indicated by the arrow. Lane M shows the molecular weight marker.

Next, we explored the possibility of using two or more tandem promoters to further improve the expression levels of the *amyQ* gene, as this strategy has been shown to be effective [29–31]. Thus, the dual tandem promoter P_{amyQ}-P_{cry3A}, composed of the engi-

neered *amyQ* promoter and the *cry3A* promoter (Figure S2), and the triple tandem promoter P_{spoVG}-P_{amyQ}-P_{cry3A}were constructed using specific primers (Table S1) and inserted into the backbone vector pJOE8999.1 to yield plasmids pJOE4 and pJOE5, respectively. The two constructs were subsequently inserted into the *B. subtilis* chromosome to obtain the recombinant strains BSQ1b (P_{amyQ}-P_{cry3A}-*amyQ*) and BSQ1c (P_{spoVG}-P_{amyQ}-P_{cry3A}-*amyQ*). The identity of each recombinant strain was confirmed by PCR and Sanger sequencing (Figure S1b–d). A comprehensive scheme depicting the construction of each AmyQ-producing strain is provided in Figure 2a. The strength of the tandem promoters to drive *amyQ* expression was evaluated by culturing both strains in production media, and the levels of secreted α-amylase were compared with the control. Additionally, the dry cell weight (DCW) of each strain was calculated (Figure 2b). The maximum value of α-amylase activity was 122.8 ± 4.4 U/mL in strain BSQ1b (the P_{amyQ}-P_{cry3A} promoter), which was 4.9-fold higher compared to BSQ1a, whereas the α-amylase activity in the BSQ1c strain (the P_{spoVG}-P_{amyQ}-P_{cry3A}promoter) was about 1.2-fold lower than in BSQ1b. Notably, while the DCW of strains BSQ1a and BSQ1b was similar (about 4.3 and 4.7 g/L, respectively), that of BSQ1c was markedly lower (3.7 g/L). Conclusively, the dual promoter P_{amyQ}-P_{cry3A} led to the most pronounced expression level of *amyQ* among the promoters investigated here, and consequently, strain BSQ1b was selected for further studies.

2.2. Secretory Expression of AmyQ via the Non-Classical (Tat) Secretion Pathway

Previous studies have shown that overexpression of secreted proteins in *B. subtilis* can cause jamming at the cell membrane due to a shortage of Sec pathway components [11,32]. However, *B. subtilis* also possesses the twin-arginine translocation (Tat) system, which facilitates the transport of fully folded proteins across membranes [33–35]. Therefore, seeking to harness the Tat pathway as a complementary secretion route, the widely used twin-arginine signal peptide of *B. subtilis* YwbN [14,36,37] was engineered (YwbN′), as described by Yang and coworkers [38], and placed in frame with the *amyQ* gene (strain BSQ1d) (Figure S1e), with the aim of redirecting AmyQ amylase secretion to the Tat pathway. A scheme of the proposed alternative secretion route for AmyQ and the construction of strain BSQ1d is depicted in Figure 2a. Surprisingly, the α-amylase activity assays showed no activity in the culture supernatant of the recombinant BSQ1d strain (Figure 2b). SDS-PAGE analysis revealed a prominent band with the expected molecular mass (55-kDa) for AmyQ in the supernatants of strain BSQ1b, which secretes AmyQ through the Sec pathway, but this band was found in neither the supernatant nor the cell extract of strain BSQ1d (Figure 2c). These results suggest that $SP_{ywbN'}$ cannot direct the extracellular secretion of AmyQ using the Tat pathway, and the use of this route was therefore discarded for further studies.

2.3. Maximizing amyQ Expression by Inserting Multiple amyQ Gene Copies into the B. subtilis Chromosome

To ascertain whether *amyQ* expression could be achieved independently of the integration site, multiple copies of the *amyQ* gene were placed under the control of the previously selected double promoter P_{amyQ}-P_{cry3A} and were independently inserted into the BS2 chromosome, one by one. Each expression cassette (P_{amyQ}-P_{cry3A}-*amyQ*), hereafter named *amyQ*_Ec, was inserted at ectopic sites within the BS2 chromosome, which were chosen based on their potential to be deleted without affecting strain growth [39,40]. The sets of vectors, integration sites, and primers are provided in the Materials and Methods section, and the resulting strains were named BSQ1e (*amyQ*_Ec at *pksG*), BSQ1f (*amyQ*_Ec at *ppsE*), BSQ1g (*amyQ*_Ec at *cotB*), BSQ1h (*amyQ*_Ec at *ylbP*), and BSQ1i (*amyQ*_Ec at *veg*) (Figure 3a). Enzymatic assays and DCW analysis showed that each of the strains harboring one copy of *amyQ*_Ec secreted similar levels of α-amylase to the medium, with activity values of 115.5–122.8 U/mL. This indicates that the integration of *amyQ*_Ec at the selected ectopic sites did not affect its expression. Similarly, comparable DCW values were obtained

for all strains (Figure 3b). Overall, we concluded that *amyQ*_Ec was uniformly expressed regardless of the integration site.

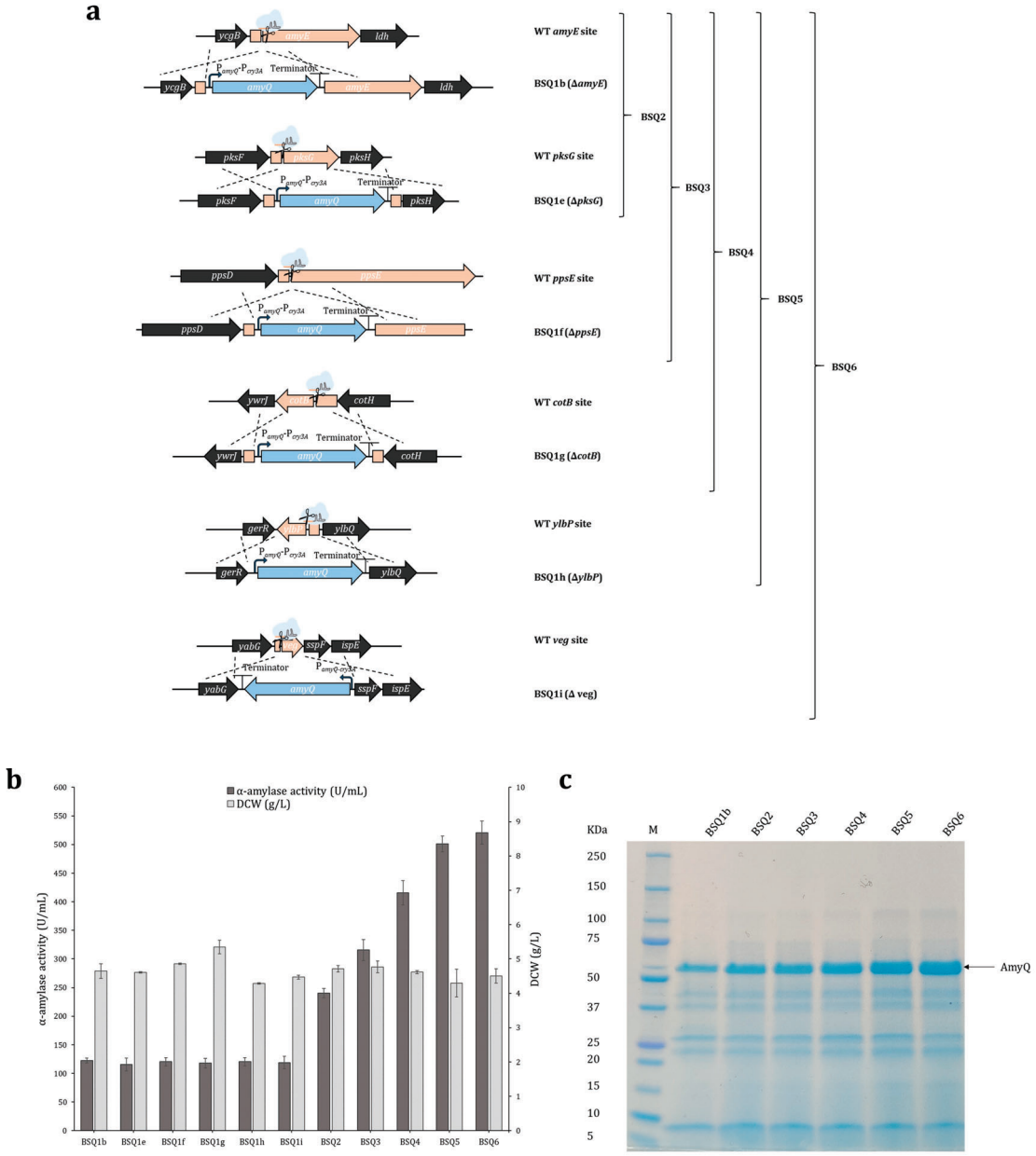


Figure 3. Construction and analysis of *Bacillus subtilis* strains with multiple *amyQ* expression cassette (*amyQ*_Ec) integrations. (**a**) Schematic representation of the construction system for strains BSQ1b, BSQ1e-i, and BSQ2-6. Individual integration of *amyQ*_Ec at the *amyE*, *pksG*, *ppsE*, *cotB*, *ylbP*, and *veg* loci yielded BSQ1b, BSQ1e, BSQ1f, BSQ1g, BSQ1h, and BSQ1i, respectively. Iterative integration of *amyQ*_Ec yielded strains BSQ2 (Δ*amyE*, Δ*pksG*), BSQ3 (Δ*amyE*, Δ*pksG*, Δ*ppsE*), BSQ4 (Δ*amyE*, Δ*pksG*, Δ*ppsE*, Δ*cotB*), BSQ5 (Δ*amyE*, Δ*pksG*, Δ*ppsE*, Δ*cotB*, Δ*ylbP*), and BSQ6 (Δ*amyE*, Δ*pksG*, Δ*ppsE*, Δ*cotB*, Δ*ylbP*, Δ*veg*) containing two to six *amyQ* copies, respectively. (**b**) Extracellular AmyQ activity and dry cell weight (DCW) in engineered *B. subtilis* strains. The error bars represent the average ± standard deviation of three biological replicates. (**c**) SDS-PAGE showing the protein supernatants from BSQ1b (1), BSQ2 (2), BSQ3 (3), BSQ4 (4), BSQ5 (5), and BSQ6 (6). Supernatants were prepared from cells incubated at 37 °C and 220 rpm for 48 h, cleared by centrifugation, and 15 µL of cleared supernatant proteins was loaded onto the SDS-PAGE gel in each lane. The estimated molecular weight for AmyQ protein of 55 kDa is indicated by an arrow. Lane M corresponds to the molecular weight marker.

The first dedicated step in the protein secretion route is the transcription of a particular gene into mRNA. To achieve maximum expression, one of the most adopted strategies is to amplify the copy number of the gene of interest [8,26,41]. The higher the copy number, the greater the expression, until the point is reached where increasing the gene copy number no longer results in increased expression [9]. In this context, iterative genome editing was performed using successive double, triple, quadruple, quintuple, and sextuple *amyQ*_Ec integration to yield strains BSQ2 (*amyQ*_Ec at *amyE* and *pksG*), BSQ3 (a*myQ*_Ec at *amyE*, *pksG*, and *ppsE*), BSQ4 (*amyQ*_Ec at *amyE*, *pksG*, *ppsE*, and *cotB*), BSQ5 (*amyQ*_Ec at *amyE*, *pksG*, *ppsE*, *cotB*, and *ylbP*), and BSQ6 (*amyQ*_Ec at *amyE*, *pksG*, *ppsE*, *cotB*, *ylbP*, and *veg*). The identity of each recombinant strain was successfully confirmed by PCR and Sanger sequencing (Figure S1f–j). A comprehensive diagram of the multiple integration sites of *amyQ*_Ec in each strain is provided in Figure 3a.

α-amylase assays revealed a continuous increase in the levels of enzymatic activity until the copy number of *amyQ*_Ec reached the value of six. The α-amylase secretion by strain BSQ2 was almost two-fold higher compared to BSQ1b (122.8 ± 4.4 U/mL), reaching a value of 240.4 ± 8.0 U/mL (two-fold; $p < 0.05$). Strains BSQ3, BSQ4, and BSQ5 secreted higher levels of α-amylase compared to strain BSQ2, albeit the rates of increase were lower, reaching values of 315.7 ± 18.4 U/mL (2.6-fold; $p < 0.05$), 415.7 ± 21.1 U/mL (3.4-fold; $p < 0.05$), and 501.2 ± 13.5 U/mL (4.1-fold; $p < 0.05$), respectively. Finally, strain BSQ6 exhibited about 520.6 ± 20.2 U/mL of α-amylase activity (4.2-fold; $p < 0.05$), which was similar to strain BSQ5 (no significant differences; $p > 0.05$), suggesting the levels of the *amyQ*-specific message had saturated the post-transcriptional machinery of this strain (Figure 3b). SDS–PAGE analysis of the supernatants from the various strains (Figure 3c) showed an increasingly prominent band of 55 kDa, which corresponds to the molecular weight of AmyQ. Overall, the levels of α-amylase activity were notably increased, being 20.9-fold higher in BSQ6 than the values observed in the initial BSQ1a strain (Table 1).

Table 1. Cumulative effect on AmyQ α-amylase activity of successive modifications with significant positive effects. Dry cell weight (DCW) values of the engineered strains are also shown. Experiments were carried out in triplicate, and data are presented as mean values ± standard deviation.

Strain	Amylase Activity (U/mL)	Cumulative Increase [a]		DCW (g/L)
		Total Fold Change	*p*-Value [b]	
BSQ1a	24.9 ± 1.5	1		4.3 ± 0.1
BSQ1b	122.8 ± 4.4	4.9	<0.001	4.7 ± 0.2
BSQ2	240.4 ± 8.0	9.7	<0.001	4.7 ± 0.1
BSQ3	315.7 ± 18.4	12.7	0.003	4.8 ± 0.2
BSQ4	415.7 ± 21.1	16.7	0.003	4.6 ± 0.1
BSQ5	501.2 ± 13.5	20.2	0.004	4.3 ± 0.4
BSQ6	520.6 ± 20.2	20.9	0.3	4.5 ± 0.2
BSQ6_7	579.0 ± 25.4	23.3	0.04	4.3 ± 0.3
BSQ6_9	604.3 ± 13.5	24.3	0.004 [c]	4.3 ± 0.2
BSQ6_11	656.8 ± 24.3	26.4	0.03	4.4 ± 0.01
BSQ6_11_F [d]	1439.2 ± 92.7	57.9	<0.001	22.4 ± 0.7

[a] The increase by one significant edition after another in the total fold change. [b] *p*-value from Student's *t*-test of each strain compared with the precedent strain. [c] *p*-value from Student's *t*-test compared with BSQ6. [d] BSQ6_11 values in a 3 L fermentor at 90 h.

2.4. Enhancing AmyQ Secretion in the BSQ6 Strain through Genome Modification of the Sec Pathway and Other Potential Post-Transcriptional Bottlenecks

Previous studies have indicated that saturating levels of mRNA expression may potentially lead to the saturation of the secretory machinery, thus hindering effective protein secretion [26]. In this scenario, it has been shown that the overexpression of some components of the secretory machinery can facilitate the secretion and proper folding of the protein of interest [3,11,42]. Therefore, to comprehensively circumvent the potential bottlenecks lying downstream of transcription, we adopted an approach based on targeting

several post-transcriptional constraints and evaluated the effect of these mutations in strain BSQ6. Hence, a set of strains harboring genome modifications within the Sec pathway secretion route were constructed using the CRISPR-Cas9 system. Firstly, a truncation or deletion of the *hag*, *pssA*, *yusX*, or *hrcA* genes was independently performed in BSQ6 to obtain strains BSQ6_1, BSQ6_2, BSQ6_3, and BSQ6_4, respectively. Secondly, the insertion of an extra copy of the signal peptidase *sipT* gene at the *thrC* site, placed under the control of the dual promoter P_{amyQ}-P_{cry3A}, led to the construction of strain BSQ6_5. Thirdly, the in-frame replacement of the *spoVG*, *yqeZ*, *sigX*, and *pel* genes with *rasP*, *sppA*, the operon *secYEG*, and *prsA* allowed for the overexpression of these genes in strains BSQ6_6, BSQ6_7, BSQ6_8, and BSQ6_9, respectively. Finally, strain BSQ6_10 was designed to contain a functionally active GudB protein. To this end, a 9 bp direct repeat, which produced a cryptic product in strains derived from *B. subtilis* 168, was deleted within the wild-type *gudB* gene sequence [43]. Figure 4a provides a comprehensive schematic representation of the multiple modifications in strain BSQ6. The identity of each recombinant strain was verified by PCR and Sanger sequencing (Figures S1k–r and S3).

The ability of the BSQ6 strain mutants to secrete α-amylase into the media and their DCWs were individually evaluated using shake-flask cultures (Figure 4b). Only two of the ten mutants analyzed, BSQ6_7 (harboring an extra copy of the *sppA* gene) and BSQ6_9 (harboring an extra copy of *prsA*), exhibited significantly enhanced *amyQ* expression, with 1.1- and 1.2-fold higher AmyQ secretion compared to strain BSQ6 (both $p < 0.05$), respectively (Table 1). Instead, BSQ6_6 (harboring an extra copy of the *rasP* gene) and BSQ6_8 (harboring an extra copy of the operon *secYEG*) showed reduced levels of α-amylase activity, both being 1.2-fold lower than the parental strain. Notably, strain BSQ6_2 (with a truncated copy of the *pssA* gene) showed a prominent growth defect, and therefore the levels of α-amylase secreted into the media were dramatically reduced. The rest of the mutants displayed values of α-amylase activity similar to the control BSQ6 strain (all $p > 0.05$) (Figure 4b).

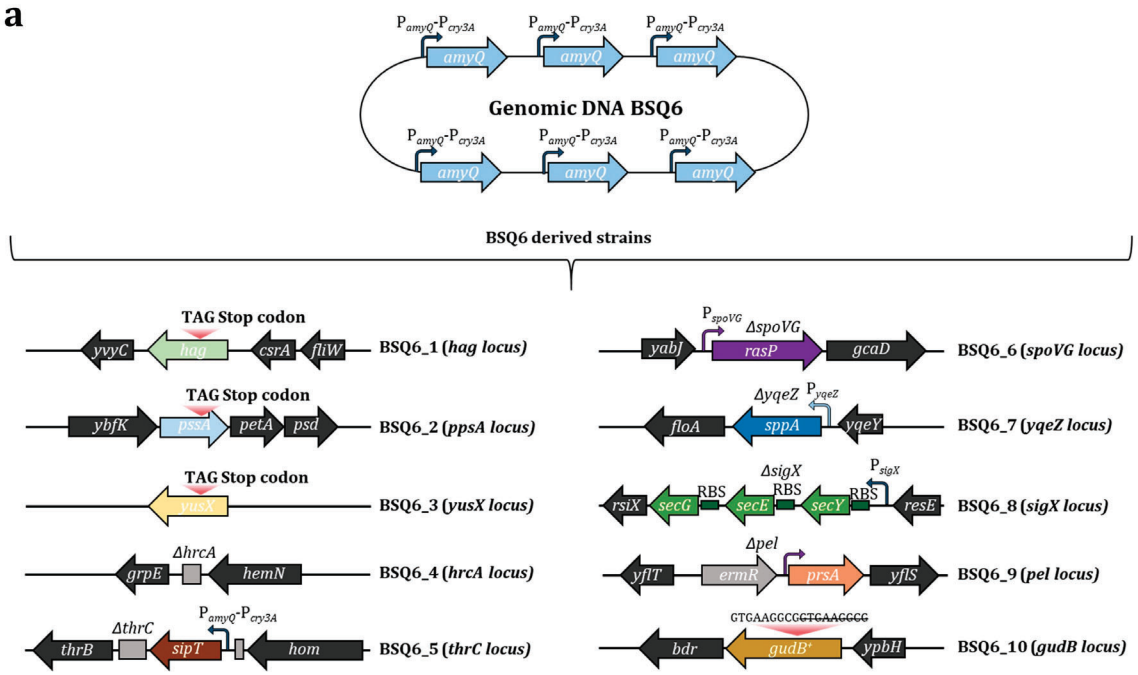


Figure 4. *Cont.*

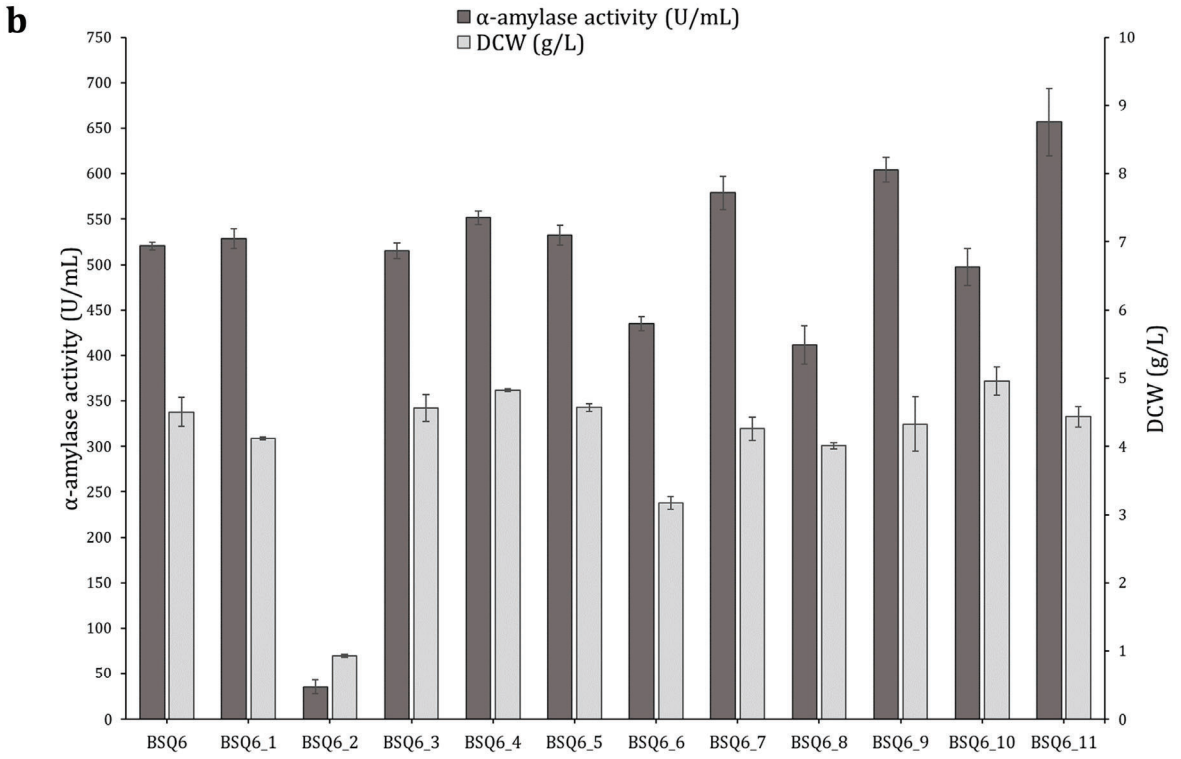


Figure 4. Expression and secretion of AmyQ α-amylase by *Bacillus subtilis* BSQ6 variants. (**a**) Depiction of the multiple BSQ6 mutant strains. Strains BSQ6_1 to BSQ1_3 possess a truncated copy of the *hag*, *pssA*, and *yusX* genes, respectively, achieved by adding a premature stop codon. Strain BSQ6_4 has a clean deletion of the *hrcA* gene. Strains BSQ6_5 to BSQ6_9 were constructed to contain an extra copy of the *sipT*, *rasP*, *sppA*, *secYEG*, or *prsA* genes, respectively, under the control of strong promoters. A functional *gudB* gene was restored in strain BSQ6_10 by deleting 9 bp. Strain BSQ6_11 with combinational overexpression of *sppA* and *prsA*. RBS (ribosome binding site). (**b**) Extracellular AmyQ activity and dry cell weight (DCW) in engineered *B. subtilis* strains. The error bars represent the average ± standard deviation of three biological replicates. BSQ6 was used as a control.

Subsequently, to further develop our *B. subtilis* strain as a high expressor of the AmyQ α-amylase, we investigated whether the double mutant BSQ6_11, carrying both beneficial mutations (overexpression of *sppA* and *prsA*), could enhance AmyQ production. As detailed in Table 1, BSQ6_11 displayed outstanding secretion of the AmyQ α-amylase, with a remarkable 26.4-fold increase compared to the initial strain BSQ1a.

2.5. Scale-Up of α-Amylase Production in a 3 L Fermentor

The expression efficiency of BSQ6_11 was further explored in a 3 L fermentor. The fermentor was inoculated with 4% (*v*/*v*) of freshly cultured BSQ6_11 grown in production medium at 37 °C for 18 h. To maintain cell growth and α-amylase production, we chose a fed-batch strategy with pulse feeding of highly concentrated sucrose (250 g/L and 580 g/L) and soy peptone (250 g/L), which were fed intermittently in two pulses of 0.12 L in response to an increased DO signal. During the growth phase (Figure 5), the maximum biomass in the fermentor reached a DCW of 27.8 g/L at 78 h. The activity of α-amylase in the medium was continuously increased and reached a maximum of 1439.2 ± 92.7 U/mL at 90 h, with a productivity of 16 U/mL h. This was 2.2-fold greater than the α-amylase activity in the fermentation supernatant of the same BSQ6_11 strain in the shake-flask cultures and corresponded to a remarkable 57.9-fold higher AmyQ secretion compared to the parental strain BSQ1a (Table 1). The high activity of α-amylase indicated that the engineered strain BSQ6_11 was a suitable host for the industrial production of the AmyQ α-amylase.

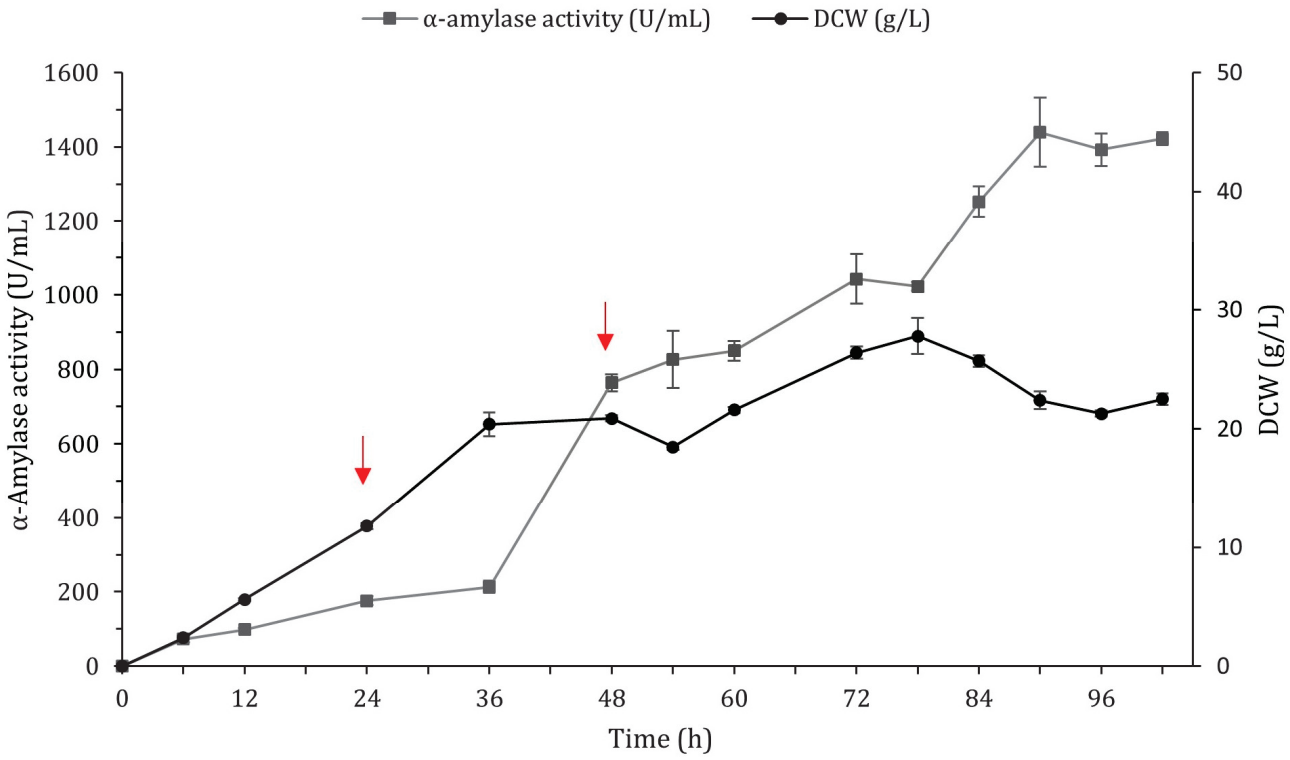


Figure 5. Production of α-amylase AmyQ in recombinant strain BSQ6_11 in 3 L fermentor. AmyQ production and DCW were monitored as a function of time. Solid square: α-amylase activity in medium. Solid circle: DCW. Red arrows indicate pulse feedings at 24 and 48 h. Error bars indicate the standard deviation from the mean of the three experimental data replicates.

3. Discussion

Despite the extensive knowledge available on the secretion of enzymes by *B. subtilis*, we believe there is scope for improving its capacity to overproduce commercially significant enzymes. The *B. subtilis* secretory pathway can be divided into three stages: transcription, translocation and folding, and secretion [44]. Along this pathway, each step is a potential bottleneck for high-level production, and these restrictions should be identified if the yields of heterologous proteins are to be improved. Here, we used the CRISPR-Cas9 system to systematically tackle, one by one, the different steps throughout the secretion route of *B. subtilis*. By employing this approach, it was possible to pinpoint targets whose modification could enhance the capacity of the secretion machinery, thus maximizing the secretion of the reporter AmyQ α-amylase in *B. subtilis*, as well as constructing an environmentally friendly super-secreting *B. subtilis* strain with practical application to industry.

To enhance or appropriately adjust the gene expression levels in *B. subtilis*, it is essential to study the promoters that regulate transcription levels, as they constitute one of the most important elements for facilitating the high production of recombinant proteins [45,46]. Thus, we evaluated the capacity of single, double, and triple promoters to dictate the expression of the AmyQ α-amylase and found the expression strength of the double promoter (P_{amyQ}-P_{cry3A}) to be the greatest (Figure 2b). This outcome was probably due to the presence of mRNA stabilization sequences, which have been shown to increase the expression of industrial enzymes [47,48]. As strength is not necessarily correlated with the number of copies of the promoter [45], the weakness of P_{spoVG}-P_{amyQ}-P_{cry3A} might have been caused by aligning more than two promoters [49]. Additionally, differences in origin can influence the cooperativity of the tandem promoter.

Once a suitable promoter has been chosen, achieving the maximum secretion of a particular enzyme in *B. subtilis* almost always requires the amplification of the gene copy number. This is accomplished by either using high-copy-number replicative plasmids or

inserting multiple copies of the desired gene into the chromosome [8,9]. Hence, environmentally friendly, plasmid-less, marker-free multicopy-*amyQ* strains were constructed by sequentially inserting up to six *amyQ* gene copies under the control of the P_{amyQ}-P_{cry3A} promoter at ectopic sites within the *B. subtilis* chromosome. As shown in Table 1, the higher the copy number of the *amyQ*_Ec cassette, the higher the *amyQ* gene expression, until at the sixth copy, the levels of α-amylase activity were similar to those achieved with five copies. Comparably to our results, the group of Altenbuchner showed that the integration of five copies of the β-glucosidase *ganA* into the chromosome of *B. subtilis* was required to achieve the maximum expression [9], whereas in the case of the protease *aprL*, only one copy was sufficient [26]. This discrepancy might stem from a lower translation initiation rate owing to different gene sequences downstream of the start codon [50], indicating that the number of copies necessary to achieve the maximum expression is variable and depends on each specific gene. In any case, at this point, we could conclude that the levels of *amyQ*-specific mRNA had saturated the post-transcriptional machinery, and the bottleneck for expression had shifted downstream of transcription.

To elucidate which other factors might impede the AmyQ production process, we targeted essential genes involved in the post-transcriptional stage and non-essential Sec pathway components, aiming to improve the secretion of recombinant AmyQ in *B. subtilis*. After transcription, the first step toward the successful secretion of the newly formed pre-protein in *B. subtilis* requires the action of intracellular chaperones that prevent inappropriate folding or aggregation of the pre-protein within the cell [51]. Previous reports show that inactivation of the repressor HrcA enables the overexpression of intracellular chaperones [52], and the secretion of the AmyS α-amylase has been significantly enhanced using this approach [13]. However, contrary to expectations, *hrcA* deletion exerted no significant effect on AmyQ production and therefore cannot be considered a bottleneck in our *B. subtilis* α-amylase production system.

B. subtilis possesses a well-developed and highly efficient Sec pathway that is commonly exploited for the production of industrial enzymes. However, an alternative system, the Tat pathway, exists to facilitate the transport of proteins that fold too tightly or rapidly in the cytosol and compatibility with the Sec pathway [53]. The nature of the signal peptide directs the nascent protein to the Sec or Tat translocase, which is efficiently cleaved by signal peptidases [54] prior to the export of the protein to the extracytoplasmic compartment. Here, we used the signal peptide of the typically Tat-secreted YwbN protein [34,55] to evaluate whether this route was compatible with AmyQ α-amylase secretion, which we hoped would be enhanced via both routes simultaneously. Although successful secretion of alkaline α-amylase has been achieved using an engineered Tat-dependent YwbN signal peptide [38], its application in our study resulted in no α-amylase activity being detected in the extracellular media (Figure 2b). Our findings are in agreement with previous reports pointing to Tat incompatibility with AmyL α-amylase secretion, probably due to the inability of the protein to fold in the cytoplasm, a prerequisite for Tat secretion [14].

High gene expression levels may result in the saturation of the Sec translocon capacity, which may be due to a shortage of translocons or the unavailability of signal peptidases [18,21]. Therefore, the effect of both artificial *secYEG* operon overexpression and signal peptidase *sipT* overexpression on AmyQ secretion was evaluated. While the extra copy of the *sipT* gene resulted in a slight 1.02-fold improvement in AmyQ production, the overexpression of *secYEG* unexpectedly diminished α-amylase activity (Figure 4). Although increasing the number of translocons [16] and the overexpression of SipT [18] can be accompanied by a concomitant increase in the amount of exoenzymes, in other cases, these modifications have had a negligible or even a reducing effect [11], which indicates they have an inconsistent impact on amylase secretion. In the present study, we verified that the translocon SecYEG and the signal peptidase SipT have different effects on the secretion of different proteins.

Bottlenecks in the late stages of protein secretion in *B. subtilis* include the extracellular chaperone prsA [18] and the signal peptide peptidases SppA and RasP [20,56]. Gratifyingly,

PrsA overexpression significantly enhanced AmyQ production, with the values of secreted α-amylase exceeding those of BSQ6 by 16%. This result is in line with those of previous studies that have highlighted PrsA as a primary bottleneck in α-amylase secretion, as it facilitates the folding process of mature proteins into a stable and active conformation [11,18,22]. Additionally, the insertion of an extra copy of the *sppA* gene under the control of the strong promoter *yqeZ* [57] significantly increased the α-amylase activity by 11% compared to BSQ6 (Figure 4). This is in agreement with a previous study, which found that the overexpression of *sppA* enhanced the production of α-amylase in *Bacillus licheniformis*, suggesting that this peptidase is necessary for the efficient processing of cleaved signal peptides and to maintain the proper secretion of mature proteins across the membrane under conditions of hyper-secretion [19]. Intriguingly, the overexpression of RasP severely impaired the secretion of AmyQ, in contrast with previous research in which its overexpression markedly improved the secretion of the AmyAc α-amylase [20]. Our hypothesis is that excessive RasP may have some unknown harmful effects on the physiological characteristics of *B. subtilis*, resulting in a significant drop in α-amylase production.

Acquiring a systematic and thorough understanding of Sec pathway components and non-associated Sec pathway proteins that might facilitate target protein secretion could be the key to significantly increasing the yield of recombinant proteins. Indeed, besides optimizing the Sec-related components, this study aimed to further engineer the *B. subtilis* host as a chassis to improve the production of the AmyQ α-amylase. To achieve this, firstly, the *B. subtilis* cell surface was modified by knocking out the *pssA* gene. Previous reports have suggested that the deletion of this gene results in the increased presence of anionic membrane phospholipids in *B. subtilis*, leading to a higher amount of α-amylase being released into the medium [21]. Conversely, *pssA* deletion caused severe growth arrest, indicating that this gene is vital for maintaining the viability of *Bacillus* cells under the conditions of high *amyQ* gene expression. Secondly, reduced expression of the oligoendopeptidase YusX or YusZ has been associated with high secretion levels of enzymes in *B. subtilis* [58]. Nonetheless, the ability to secrete AmyQ in Δ*yusX* mutants remained stable, indicating a negligible effect.

Lastly, we wanted to ascertain whether knocking out specific genes to increase the growth rate of *B. subtilis* would result in higher productivities, as previously reported [39]. To investigate this hypothesis, two strategies were adopted. On the one hand, we examined the effect of deleting the non-essential *hag* gene on the growth rate of BSQ6. The rationale for this approach was that the metabolic energy wasted on expressing the subunit flagellin protein (*hag* gene), one of the most highly expressed genes in the *B. subtilis* genome [59], could be redirected toward the synthesis of α-amylase protein and essential genes, resulting in an increased growth rate, as described in previous studies [60]. However, no significant differences in the growth rate or α-amylase production were observed in BSQ6_1 lacking the *hag* gene. This is not surprising considering that metabolism self-regulation by the cell can often minimize the performance of knock-out strains [61]. On the other hand, we wanted to re-establish the activity of the *gudB* gene, encoding glutamate dehydrogenase (GDH), a gene truncated in the strain *B. subtilis* 168 and its derivatives. As our production medium contained soy peptone, which is rich in amino acids of the glutamate family, we speculated that restoring GudB activity would allow for more efficient utilization of nitrogen sources and therefore confer a growth advantage compared to cells lacking a functional enzyme [43]. However, no significant differences between the strains were observed. As *B. subtilis* carries an additional GDH enzyme encoded by the *rocG* gene, our hypothesis is that although soy peptone is rich in glutamate, the final concentration used in this study may not have been high enough for an additional GDH enzyme to be advantageous [62].

In this work, two pulses of fresh nutrients (sucrose and peptone) proved to be optimal for enhancing AmyQ production in a 3 L fermentor. In pulse fed-batch fermentation, the substrate concentration is kept within certain limits by the pulses to meet the requirements of *B. subtilis* metabolism, thus avoiding starvation and directing the energy obtained from

the substrates towards both maintaining cellular metabolism and producing the α-amylase AmyQ. Accordingly, this strategy resulted in a 2.2-fold enhancement of α-amylase activity (1439 U/mL) in comparison to in the shake flasks (657 U/mL). Remarkably, the level of secreted AmyQ reached at the fermentor stage was higher than the value of 1100 U/mL for AmyQ α-amylase activity achieved in previous studies by using the high-copy pKHT10 plasmid in *B. subtilis* [63,64].

This is the first report that describes an improvement in α-amylase extracellular production levels in *B. subtilis* by merely using the CRISPR-Cas9 system, rendering an industrial strain devoid of plasmids and antibiotic selection markers and bypassing the need for expensive inducers. Although the impact of gene modifications within the secretory pathway of heterologous proteins might be variable, probably depending on each specific gene, we consider the strategy presented in this work to obtain the maximum secretion levels from multiple copy gene insertion, along with combinational Sec pathway analysis, a promising approach that will facilitate the construction of robust, ecologically safe, industrial strains of *B. subtilis* in forthcoming years.

4. Materials and Methods

4.1. Bacterial Strains, Plasmids, and Culture Conditions

All the bacterial strains and plasmids used in this study are listed in Tables 2 and 3, respectively. *Escherichia coli* DH5α served as the host for cloning and plasmid preparation. Chemically competent *E. coli* strains were prepared and the transformation protocol performed as described previously [65]. *B. subtilis* BS1, which is an asporogenous strain (Δ*sigF*) with reduced lysis (Δ*lytC*) and deficient in seven extracellular proteases (Δ*nprE*, Δ*aprE*, Δ*epr*, Δ*mpr*, Δ*nprB*, Δ*vpr*, Δ*bpr*), was used as a host for AmyQ expression. *B. subtilis* was transformed according to the method of Yasbin et al. [66], using plasmid DNA isolated from the *rec*$^+$ strain *E. coli* Turbo (New England Biolabs, Ipswich, MA, USA). The plasmid pJOE8999.1 [28], an *E. coli*/*B. subtilis* shuttle vector harboring the CRISPR-Cas9 system, was used to edit the *B. subtilis* genome. Transformants of *E. coli* and *B. subtilis* were selected on Luria–Bertani (LB) agar plates at 37 °C, supplemented with appropriate antibiotics. To select plasmids in *E. coli*, kanamycin was used at a final concentration of 30 μg/mL, while for *B. subtilis*, kanamycin was used at a final concentration of 6 μg/mL, as well as erythromycin and lincomycin at 2 μg/mL and 25 μg/mL, respectively. All the strains were incubated under shaking conditions at 200 rpm. All the experiments were repeated at least three times, and mean values were used for comparisons.

Table 2. Strains used in this study.

Strain	Characteristics	Reference
E. coli DH5α	*fhuA2 lac(del)U169 phoA glnV44 Φ80′ lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Laboratory stock
E. coli NEB® turbo	*F′ proA + B + lacIq ΔlacZM15/fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)TetS endA1 thi-1 Δ(hsdS-mcrB)5*	Laboratory stock
B. subtilis strains		
BS0	Δ*nprE*, Δ*aprE*, Δ*epr*, Δ*mpr*, Δ*nprB*, Δ*vpr*, Δ*bpr*, Δ*sigF*	BGSC
BS1	BS0 derivative, Δ*lytC*	Laboratory stock
BS2	BS1 derivative, Δ*amyE*	This work
BSQ1a	BS2 derivative, *amyQ* (P_{spoVG}) knock-in mutant (Δ*spoVG*)	This work
BSQ1b	BS1 derivative, *amyQ* (P_{amyQ}-P_{cry3A}) knock-in mutant (Δ*amyE*)	This work
BSQ1c	BS2 derivative, *amyQ* (P_{spoVG}-P_{amyQ}-P_{cry3A}) knock-in mutant (Δ*spoVG*)	This work

Table 2. *Cont.*

Strain	Characteristics	Reference
BSQ1d	BS2 derivative, *amyQ* ($SP_{ywbN'}$) knock-in mutant (Δ*ywbN*)	This work
BSQ1e	BS2 derivative, *amyQ* (P_{amyQ}-P_{cry3A}) knock-in mutant (Δ*pksG*)	This work
BSQ1f	BS2 derivative, *amyQ* (P_{amyQ}-P_{cry3A}) knock-in mutant (Δ*ppsE*)	This work
BSQ1g	BS2 derivative, *amyQ* (P_{amyQ}-P_{cry3A}) knock-in mutant (Δ*cotB*)	This work
BSQ1h	BS2 derivative, *amyQ* (P_{amyQ}-P_{cry3A}) knock-in mutant (Δ*ylbP*)	This work
BSQ1i	BS2 derivative, *amyQ* (P_{amyQ}-P_{cry3A}) knock-in mutant (Δ*veg*)	This work
BSQ2	BSQ1b derivative, *amyQ* double knock-in mutant (Δ*amyE*, Δ*pksG*)	This work
BSQ3	BSQ2 derivative, *amyQ* triple knock-in mutant (Δ*amyE*, Δ*pksG*,Δ*ppsE*)	This work
BSQ4	BSQ3 derivative, *amyQ* quadruple knock-in mutant (Δ*amyE*, Δ*pksG*,Δ*ppsE*, Δ*cotB*)	This work
BSQ5	BSQ4 derivative, *amyQ* quintuple knock-in mutant (Δ*amyE*, Δ*pksG*, Δ*ppsE*, Δ*cotB*, Δ*ylbP*)	This work
BSQ6	BSQ4 derivative, *amyQ* sextuple knock-in mutant (Δ*amyE*, Δ*pksG*, Δ*ppsE*, Δ*cotB*, Δ*ylbP*, Δ*veg*)	This work
BSQ6_1	BSQ6 derivative, Δ*hag*	This work
BSQ6_2	BSQ6 derivative, Δ*pssA*	This work
BSQ6_3	BSQ6 derivative, Δ*yusX*	This work
BSQ6_4	BSQ6 derivative, Δ*hrcA*	This work
BSQ6_5	BSQ6 derivative, *sipT* knock-in mutant (Δ*thrC*)	This work
BSQ6_6	BSQ6 derivative, *rasP* knock-in mutant (Δ*spoVG*)	This work
BSQ6_7	BSQ6 derivative, *sppA* knock-in mutant (Δ*yqeZ*)	This work
BSQ6_8	BSQ6 derivative, *secYEG* knock-in mutant (Δ*sigX*)	This work
BSQ6_9	BSQ6 derivative, *prsA* knock-in mutant (Δ*pel*; erm^r)	This work
BSQ6_10	BSQ6 derivative, restored *gudB* gene	This work
BSQ6_11	BSQ6 derivative, *sppA* and *prsA* knock-in mutant (Δ*yqeZ*, Δ*pel*, erm^r)	This work

BGSC: Bacillus Genetic Stock Center.

4.2. DNA Manipulations

Standard molecular techniques were carried out following the standard methods [67]. The genes encoding engineered *prsA* (Figure S4) and the α-amylase *amyQ* gene harboring its native signal peptide and the dual promoter P_{amyQ}-P_{cry3A} (Figure S2) were synthesized by NZYtech (Lisboa, Portugal). Chromosomal DNA was extracted from *B. subtilis* strains using the NZY Tissue gDNA Isolation kit). Plasmid DNA was isolated from *E. coli* using the NZYSpeedy Miniprep kit. Digested DNA fragments from agarose gel and amplified DNAs in PCRs were isolated using the NZYGelpure kit. All kits and enzymes were purchased from NZYtech (Lisboa, Portugal). Gibson assembly was performed according to the manufacturer's instructions (Invitrogen, Waltham, MA, USA). All the DNA constructs were sequenced by Macrogen (Seoul, Republic of Korea).

4.3. Construction of the Integration Vectors for amyQ Overexpression and Sec Pathway Modulation

All the integration vectors and primers used in this study are listed in Tables 2 and S1, respectively. In all cases, the integration vectors were constructed using the pJOE8999.1 plasmid as the parental plasmid and required two consecutive steps: (i) cloning specific sgRNA and (ii) cloning a specific editing template.

4.4. Cloning of sgRNA

The design of sgRNA for gene editing of *B. subtilis* was accomplished using the sgRNA Designer tool, provided by the Broad Institute [68]. For sgRNA construction targeting each desired gene, two complementary oligonucleotides were ordered (Macrogen, Seoul, Republic of Korea) with the respective overhangs, annealed, and cloned into the vector pJOE8999.1. In brief, both complementary oligonucleotides were mixed at a final concentration of 10 µM in annealing buffer (10× stock containing 100 mM Tris-HCl pH 7.5, 1 M NaCl, and 1 mM ethylenediaminetetraacetic acid (EDTA) (pH 8)), kept at 98 °C for 5 min, and slowly cooled to room temperature. Then, the annealed oligonucleotides were treated with polynucleotide kinase to phosphorylate the 5′ ends, according to the manufacturer's instructions (Invitrogen, Waltham, MA, USA), and ligated with the *BsaI*-cleaved and dephosphorylated plasmid pJOE8999.1 to incorporate specific target sequences into the vector.

4.5. Cloning of the Editing Templates

In a second step, to construct the editing templates, two homologous arms of a similar length and the desired template to be inserted were separately amplified and then fused together by splicing with overlap extension PCR (SOEing-PCR, Table S2). The resulting PCR product was digested with *SfiI* and ligated into the *SfiI*-digested pJOE8999.1 vector to obtain each editing plasmid.

For the construction of the knock-out plasmids pJOE2, pJOE12, pJOE13, pJOE14, pJOE15, and pJOE20, the corresponding sgRNA and homologous repair template were inserted into the plasmid pJOE8999.1 using the following primers: TS2F/TS2R (sgRNA targeting *amyE*), P2_1F/P2_1R and P2_2F/P2_2R (template for *amyE* deletion); TS12F/TS12R (sgRNA targeting *hag*), P12_1F/P12_1R and P12_2F/P12_2R (template for *hag* deletion); TS13F/TS13R (sgRNA targeting *pssA*), P13_1F/P13_1R and P13_2F/P13_2R (template for *pssA* deletion); TS14F/TS14R (sgRNA targeting *yusX*), P14_1F/P14_1R and P14_2F/P14_2R (template for *yusX* deletion); TS15F/TS15R (sgRNA targeting *hrcA*), P15_1F/P15_1R and P15_2F/P15_2R (template for *hrcA* deletion); and TS20F/TS20R (sgRNA targeting *gudB*), P20_1F/P20_1R and P20_2F/P20_2R (template for *gudB* restoration). In the pJOE12, pJOE13, and pJOE14 editing plasmids, the repair template was designed to remove 6 bp of the native sequence to insert a *XhoI* restriction site and 5 bp of a random sequence, causing gene frameshift mutation and consequently a loss of function. The removal of 9 pb of cryptic *gudB* using pJOE20 restored *gudB* function instead. However, pJOE2 and pJOE15 were designed to produce a partial deletion of the corresponding gene, leaving only a few amino acids intact.

In an analogous manner, the knock-in plasmids pJOE3, pJOE4, pJOE5, pJOE6, pJOE7, pJOE8, pJOE9, pJOE10, pJOE11, pJOE16, pJOE17, pJOE18, and pJOE19 were constructed using the following primers: TS3F/TS3R (sgRNA targeting *spoVG*), P3_1F/P3_1R, P3_2F/P3_2R, and P3_3F/P3_3R (template for P_{spoVG}-*amyQ* integration), P5_1F/P5_1R, P5_2F/P5_2R, and P5_3F/P5_3R (template for P_{spoVG}-P_{amyQ}-P_{cry3A}-*amyQ* integration), P17_1F/P17_1R, P17_2F/P17_2R, and P17_3F/P17_3R (template for *rasP* integration); TS4F/TS4R (sgRNA targeting *amyE*), P4_1F/P4_1R, P4_2F/P4_2R, and P4_3F/P4_3R (template for P_{amyQ}-P_{cry3A}-*amyQ* integration); TS6F/TS6R (sgRNA targeting *ywbN*), P6_1F/P6_1R, P6_2F/P6_2R, and P6_3F/P6_3R (template for $SP_{ywbN'}$-*amyQ* integration); TS7F/TS7R (sgRNA targeting *pksG*), P7_1F/P7_1R, P7_2F/P7_2R, and P7_3F/P7_3R (template for P_{amyQ}-P_{cry3A}-*amyQ* integration); TS8F/TS8R (sgRNA targeting *ppsE*), P8_1F/P8_1R, P8_2F/P8_2R, and P8_3F/P8_3R (template for P_{amyQ}-P_{cry3A}-*amyQ* integration); TS9F/TS9R (sgRNA targeting *cotB*), P9_1F/P9_1R, P9_2F/P9_2R, and P9_3F/P9_3R (template for P_{amyQ}-P_{cry3A}-*amyQ* integration); TS10F/TS10R (sgRNA targeting *ylbP*), P10_1F/P10_1R, P10_2F/P10_2R, and P10_3F/P10_3R (template for P_{amyQ}-P_{cry3A}-*amyQ* integration); TS11F/TS11R (sgRNA targeting *veg*), P11_1F/P11_1R, P11_2F/P11_2R, and P11_3F/P11_3R (template for P_{amyQ}-P_{cry3A}-*amyQ* integration); TS16F/TS16R (sgRNA targeting *thrC*), P16_1F/P16_1R, P16_2F/P16_2R, P16_3F/P16_3R, and P16_4F/P16_4R (template for *sipT* integration); TS18F/TS18R (sgRNA

targeting *yqeZ*), P18_1F/P18_1R, P18_2F/P18_2R, and P18_3F/P18_3R (template for *sppA* integration); and TS19F/TS19R (sgRNA targeting *sigX*), P19_1F/P19_1R, P19_2F/P19_2R, P19_3F/P19_3R, P19_4F/P19_4R, and P19_5F/P19_5R (template for *secYEG* overexpression). The latter template was constructed using Gibson assembly following the manufacturer's instructions.

Due to cloning issues during *prsA* template construction, a different approach based on Gibson assembly was adopted for *prsA* insertion at the *pel* locus. Four different fragments were amplified by PCR, corresponding to the (i) 5′ *pel* homologous region (primers AF/AR); (ii) the erythromycin resistance gene from the pBS2EXylR plasmid (BF/BR primers); (iii) synthetic *prsA* (primers CF/CR); and (iv) the 3′ *pel* homologous region (primers DF/DR). The amplicons were designed to allow Gibson assembly to occur following the manufacturer's instructions (Thermo Fischer Scientific, Waltham, MA, USA). The resulting reaction mixture was used as a template to amplify the 8 kb fragment using the primers AF/DR, which served as donor DNA for *B. subtilis* transformation [69].

Table 3. Plasmids used in this study.

Plasmid	Characteristics	Reference
pBS2EXylRPxylA	Plasmid containing the xylose-inducible promoter/xylose repressor system	[70]
pJOE8999.1	P_{manP}-*cas9*, pUC, pE194^{ts}, kan^{r}	[28]
pJOE2	*amyE* gene knock-out plasmid derived from pJOE8999.1	This work
pJOE3	*amyQ* gene (P_{spoVG}) knock-in plasmid derived from pJOE8999.1. Integration at the *spoVG* locus.	This work
pJOE4	*amyQ* gene (P_{amyQ}-P_{cry3A}) knock-in plasmid derived from pJOE8999.1. Integration at the *amyE* locus.	This work
pJOE5	*amyQ* gene (P_{spoVG}-P_{amyQ}-P_{cry3A}) knock-in plasmid derived from pJOE8999.1. Integration at the *spoVG* locus.	This work
pJOE6	*amyQ* gene ($SP_{ywbN'}$) knock-in plasmid derived from pJOE8999.1. Integration at the *ywbN* locus.	This work
pJOE7	*amyQ* gene (P_{amyQ}-P_{cry3A}) knock-in plasmid derived from pJOE8999.1. Integration at the *pksG* locus.	This work
pJOE8	*amyQ* gene (P_{amyQ}-P_{cry3A}) knock-in plasmid derived from pJOE8999.1. Integration at the *ppsE* locus.	This work
pJOE9	*amyQ* gene (P_{amyQ}-P_{cry3A}) knock-in plasmid derived from pJOE8999.1. Integration at the *cotB* locus.	This work
pJOE10	*amyQ* gene (P_{amyQ}-P_{cry3A}) knock-in plasmid derived from pJOE8999.1. Integration at the *ylbP* locus.	This work
pJOE11	*amyQ* gene (P_{amyQ}-P_{cry3A}) knock-in plasmid derived from pJOE8999.1. Integration at the *veg* locus.	This work
pJOE12	*hag* gene knock-out plasmid derived from pJOE8999.1	This work
pJOE13	*pssA* gene knock-out plasmid derived from pJOE8999.1	This work
pJOE14	*yusX* gene knock-out plasmid derived from pJOE8999.1	This work
pJOE15	*hrcA* gene knock-out plasmid derived from pJOE8999.1	This work
pJOE16	*sipT* gene knock-in plasmid derived from pJOE8999.1. Integration at the *thrC* locus.	This work
pJOE17	*rasP* gene knock-in plasmid derived from pJOE8999.1. Integration at the *spoVG* locus.	This work
pJOE18	*sppA* gene knock-in plasmid derived from pJOE8999.1. Integration at the *yqeZ* locus.	This work

Table 3. *Cont.*

Plasmid	Characteristics	Reference
pJOE19	*secYEG* artificial operon knock-in plasmid derived from pJOE8999.1. Integration at the *sigX* locus.	This work
pJOE20	*gudB* gene restoration plasmid derived from pJOE8999.1	This work

4.6. Plasmid Curing and Genome Edition Modifications

To cure the plasmid from recombinant strains, edited colonies were passaged three times on LB agar plates (without any antibiotics) at 50 °C for 24 h. Plasmid curing yielded the optimum results when the cells were streaked for single colonies at each passage. Colonies cured of the editing plasmid were confirmed by streaking them onto LB agar plates containing kanamycin or no antibiotics; colonies cured of plasmids fail to grow at 37 °C. The identity of each genome modification in the mutant strains was verified by colony PCR using the relevant primers and Sanger sequencing (Table S1 and Figure S1).

4.7. Quantification of α-Amylase Activity in Shake Flasks

Overnight cultures of recombinant *B. subtilis* strains in production medium (12 g/L sucrose, 18 g/L peptone, 2 g/L $(NH_4)_2SO_4$, 18.3 g/L $K_2HPO_4 \cdot 3H_2O$, 6 g/L KH_2PO_4, 1 g/L Na^+citrate, 0.2 g/L $MgSO_4 \cdot 7H_2O$, 1 g/L $CaCl_2$, 0.12 g/L $FeSO_4 \cdot 7H_2O$, 30 mg/L $MnSO_4 \cdot H_2O$, 12 mg/L $CuSO_4 \cdot H_2O$, and 12 mg/L $ZnCl_2$) were diluted to 0.1 OD_{600} in 25 mL of production media and grown at 37 °C and 220 rpm for 48 h. The culture supernatant, which was used as the crude enzyme solution, was obtained at various intervals during cultivation by centrifugation at 8000× *g* for 20 min at 4 °C. The enzyme activity was calculated by incubating 1 µL of the supernatant with 250 µL of 0.75% (*w*/*v*) soluble starch solution (prepared in Tris-HCl buffer at a pH of 6.5) as a substrate for 1 min at 80 °C. The reaction was terminated by adding 0.75 mL of 3,5-dinitrosalicylic acid (DNS) reagent, followed by incubating the mixture in a boiling water bath for 5 min. Absorbance was read at 540 nm and compared to a standard curve. One unit (IU) of amylase activity was defined as the amount of enzyme that liberated 1 µmol of maltose from soluble starch per minute under the assay conditions.

4.8. SDS-PAGE Analysis

After the growth of the cells for 48 h at 37 °C and 220 rpm in production media, 8 mL of the cells was harvested by centrifugation (12,000× *g*, 10 min, 4 °C) to obtain supernatant (crude enzyme solution) and a cell pellet, which was then resuspended in 800 µL of 0.05 M Tris-HCl with a pH of 7. The crude cell extract was prepared using an ultrasonic homogenizer (Bandelin Sonoplus HD 3100, Berlin, Germany) for 1 min at 70% amplitude using the following cycle: work 1 s, stop 1 s. After centrifugation, the supernatant of the lysate and the supernatant of the cell cultures were analyzed using sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) with a 10% separation gel. Proteins were visualized with Coomassie Brilliant Blue.

4.9. Dry Cell Weight

The dry cell weight (DCW) of the parental strain BS1 was determined by centrifuging 10 mL of BS1 culture broth at 12,000× *g* for 15 min at 4 °C. The resulting pellet was washed three times in 0.9% (*w*/*v*) NaCl solution before drying it in an oven at 105 °C to obtain a constant weight. OD_{600} was monitored using a Genesys 30 visible spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and used to convert the optical density at 600 nm (OD_{600}) into the DCW. According to the formula, 1 OD_{600} was equivalent to 0.352 g/L. This formula was used to calculate the DCW of each recombinant strain.

4.10. Three-Liter Fermentor Experiments

Seed culture was started by inoculating 50 mL of the production medium into a 500 mL shake flask with a 20 µL sample of frozen glycerol stock (stored at −80 °C). The resulting cultures were incubated at 37 °C and 220 rpm for 18 h. In a 3 L fermentor (Sartorius AG, Göttingen, Germany), the fermentation medium (30 g/L sucrose, 30 g/L peptone, 2 g/L $(NH_4)_2SO_4$, 18.3 g/L $K_2HPO_4{\cdot}3H_2O$, 6 g/L KH_2PO_4, 1 g/L Na^+citrate, 0.2 g/L $MgSO_4{\cdot}7H_2O$, 1 g/L $CaCl_2$, 0.12 g/L $FeSO_4{\cdot}7H_2O$, 30 mg/L $MnSO_4{\cdot}H_2O$, 12 mg/L $CuSO_4{\cdot}H_2O$, and 12 mg/L $ZnCl_2$) was inoculated with seed cultures at a ratio of 4% (*v*/*v*). Fed-batch fermentation with pulse feeding was started as a 0.9 L batch and was carried out at 37 °C, maintaining the pH at 7 by using 4 M H_3PO_4 and 2 M NaOH and the dissolved oxygen (DO) at approximately 40% by cascading the agitation speed between the ranges of 700–1000 rpm and injecting air mixed with pure oxygen. Two pulse feeds of 0.12 L at 24 h and 48 h were added to the fermentor. The first pulse contained 250 g/L sucrose and 250 g/L soy peptone, along with other nutrients, whereas in the second pulse, the sucrose was increased to 580 g/L and was fed in when the dissolved oxygen concentration in the reactor increased abruptly, indicating complete consumption of the carbon source. At defined time intervals, the medium was sampled. The culture supernatant was obtained by centrifugation at 12,000× *g* for 15 min at 4 °C and used as the crude enzyme solution for further analysis.

4.11. Statistical Analysis

All the samples were analyzed in triplicate, and the data are presented as the mean ± standard deviation for each sample point. All the data were collected to analyze the variance at $p < 0.05$, and a t test was applied to compare the mean values using the R-jamovi statistical software (version 2.5).

5. Conclusions

Selecting a suitable promoter along with increasing the gene copy number proved to be the most important determinants to achieve the highest levels of *amyQ* expression. However, once the levels of the gene-specific message had saturated the post-transcriptional machinery, we found that the deficiency of the PrsA lipoprotein and the accumulation of signal peptides in the cytoplasmatic membrane constituted the most critical rate-limiting steps in our AmyQ protein production system, achieving a stunning 57.9-fold increase in enzyme activity and the production of AmyQ compared to the control at the fermentor stage. This high-level production provides a basis for enhanced industrial production of the α-amylase AmyQ. We believe this approach could be also valuable in the expression of other enzymes in *B. subtilis*.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ijms25136957/s1.

Author Contributions: Conceptualization, P.P. and D.M.-G.; methodology, J.F. and P.P.; validation, J.F.; investigation, J.F.; resources, P.P. and D.M.-G.; writing—original draft preparation, J.F. and P.P.; writing—review and editing, J.F., P.P. and D.M.-G.; supervision, P.P. and D.M.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Pla de Doctorats Industrials del Departament de Recerca i Universitats de la Generalitat de Catalunya with the support of Gestio d'Ajuts Universitaris de Recerca, with grant number 2021 DI 77, awarded to J.F.N.

Data Availability Statement: All the data supporting the conclusions of this study are included within the article and its Supplementary Materials.

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

References

1. Schallmey, M.; Singh, A.; Ward, O.P. Developments in the use of *Bacillus* species for industrial production. *Can. J. Microbiol.* **2004**, *50*, 1–17. [CrossRef] [PubMed]
2. van Dijl, J.M.; Hecker, M. *Bacillus subtilis*: From soil bacterium to super-secreting cell factory. *Microb. Cell Fact.* **2013**, *12*, 3. [CrossRef] [PubMed]
3. Li, W.; Zhou, X.; Lu, P. Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res. Microb.* **2004**, *155*, 605–610. [CrossRef] [PubMed]
4. Yan, S.; Wu, G. Bottleneck in secretion of α-amylase in *Bacillus subtilis*. *Microb. Cell Fact.* **2017**, *16*, 124. [CrossRef] [PubMed]
5. Zhao, L.; Ye, B.; Zhang, Q.; Cheng, D.; Zhou, C.; Cheng, S.; Yan, X. Construction of second generation protease-deficient hosts of *Bacillus subtilis* for secretion of foreign proteins. *Biotechnol. Bioeng.* **2019**, *116*, 2052–2060. [CrossRef] [PubMed]
6. Schumann, W. Production of recombinant proteins in *Bacillus subtilis*. *Adv. Appl. Microbiol.* **2007**, *62*, 137–189. [PubMed]
7. Wang, Y.; Shi, Y.; Hu, L.; Du, G.; Chen, J.; Kang, Z. Synthetic and systems biotechnology engineering strong and stress-responsive promoters in *Bacillus subtilis* by interlocking sigma factor binding motifs. *Synth. Syst. Biotechnol.* **2019**, *4*, 197–203. [CrossRef] [PubMed]
8. Ferrando, J.; Filluelo, O.; Zeigler, D.R.; Picart, P. Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: Development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system. *Microb. Cell Fact.* **2023**, *22*, 21. [CrossRef]
9. Watzlawick, H.; Altenbuchner, J. Multiple integration of the gene ganA into the *Bacillus subtilis* chromosome for enhanced β-galactosidase production using the CRISPR/Cas9 system. *AMB Express* **2019**, *9*, 158. [CrossRef]
10. Zhou, C.; Ye, B.; Cheng, S.; Zhao, L.; Liu, Y.; Jiang, J.; Yan, X. Promoter engineering enables overproduction of foreign proteins from a single copy expression cassette in *Bacillus subtilis*. *Microb. Cell Fact.* **2019**, *18*, 111. [CrossRef]
11. Chen, J.; Fu, G.; Gai, Y.; Zheng, P.; Zhang, D.; Wen, J. Combinatorial Sec pathway analysis for improved heterologous protein secretion in *Bacillus subtilis*: Identification of bottlenecks by systematic gene overexpression. *Microb. Cell Fact.* **2015**, *14*, 92. [CrossRef] [PubMed]
12. Wu, S.C.; Ye, R.; Wu, X.C.; Ng, S.C.; Wong, S.L. Enhanced secretory production of a single-chain antibody fragment from *Bacillus subtilis* by coproduction of molecular chaperones. *J. Bacteriol.* **1998**, *180*, 2830–2835. [CrossRef] [PubMed]
13. Yao, D.; Su, L.; Li, N.; Wu, J. Enhanced extracellular expression of *Bacillus stearothermophilus* α-amylase in *Bacillus subtilis* through signal peptide optimization, chaperone overexpression and α-amylase mutant selection. *Microb. Cell Fact.* **2019**, *18*, 69. [CrossRef] [PubMed]
14. Kolkman, M.A.B.; Van Der Ploeg, R.; Bertels, M.; Van Dijk, M.; Van Der Laan, J.; Van Dijl, J.M.; Ferrari, E. The twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct either Tat- or Sec-dependent secretion of different cargo proteins: Secretion of active subtilisin via the *B. subtilis* Tat pathway. *Appl. Environ. Microbiol.* **2008**, *74*, 7507–7513. [CrossRef] [PubMed]
15. Ren, G.H.; Cao, L.C.; Kong, W.; Wang, Z.J.; Liu, Y.H. Efficient secretion of the β-galactosidase Bgal1-3 via both Tat-dependent and Tat-independent pathways in *Bacillus subtilis*. *J. Agric. Food Chem.* **2016**, *64*, 5708–5716. [CrossRef] [PubMed]
16. Mulder, K.C.L.; Bandola, J.; Schumann, W. Construction of an artificial *secYEG* operon allowing high level secretion of α-amylase. *Protein Expr. Purif.* **2013**, *89*, 92–96. [CrossRef] [PubMed]
17. Oswald, J.; Njenga, R.; Natriashvili, A.; Sarmah, P.; Koch, H.G. The Dynamic SecYEG Translocon. *Front. Mol. Biosci.* **2021**, *8*, 664241. [CrossRef]
18. Vitikainen, M.; Pummi, T.; Airaksinen, U.; Wahlström, E.; Wu, H.; Sarvas, M.; Kontinen, V.P. Quantitation of the capacity of the secretion apparatus and requirement for PrsA in growth and secretion of α-amylase in *Bacillus subtilis*. *J. Bacteriol.* **2001**, *183*, 1881–1890. [CrossRef]
19. Cai, D.; Wang, H.; He, P.; Zhu, C.; Wang, Q.; Wei, X.; Nomura, C.T.; Chen, S. A novel strategy to improve protein secretion via overexpression of the SppA signal peptide peptidase in *Bacillus licheniformis*. *Microb. Cell Fact.* **2017**, *16*, 70. [CrossRef]
20. Neef, J.; Bongiorni, C.; Goosens, V.J.; Schmidt, B.; van Dijl, J.M. Intramembrane protease RasP boosts protein production in *Bacillus*. *Microb. Cell Fact.* **2017**, *16*, 57. [CrossRef]
21. Cao, H.; Heel, A.J.; Ahmed, H.; Mols, M.; Kuipers, O.P. Cell surface engineering of *Bacillus subtilis* improves production yields of heterologously expressed α-amylases. *Microb. Cell Fact.* **2017**, *16*, 56. [CrossRef] [PubMed]
22. Quesada-Ganuza, A.; Antelo-Varela, M.; Mouritzen, J.C.; Bartel, J.; Becher, D.; Gjermansen, M.; Hallin, P.F.; Appel, K.F.; Kilstrup, M.; Rasmussen, M.D.; et al. Identification and optimization of PrsA in *Bacillus subtilis* for improved yield of amylase. *Microb. Cell Fact.* **2019**, *18*, 158. [CrossRef]
23. Vojnovic, S.; Aleksic, I.; Ilic-Tomic, T.; Stevanovic, M.; Nikodinovic-Runic, J. *Bacillus* and *Streptomyces* spp. as hosts for production of industrially relevant enzymes. *Appl. Microbiol. Biotechnol.* **2024**, *108*, 185. [CrossRef] [PubMed]
24. Zhang, K.; Su, L.; Wu, J. Recent Advances in Recombinant Protein Production by *Bacillus subtilis*. *Annu. Rev. Food Sci. Technol.* **2020**, *11*, 295–318. [CrossRef] [PubMed]
25. Miao, C.C.; Han, L.L.; Lu, Y.B.; Feng, H. Construction of a high-expression system in *Bacillus* through transcriptomic profiling and promoter engineering. *Microorganisms* **2020**, *8*, 1030. [CrossRef]
26. Widner, B.; Thomas, M.; Sternberg, D.; Lammon, D.; Behr, R.; Sloma, A. Development of marker-free strains of *Bacillus subtilis* capable of secreting high levels of industrial enzymes. *J. Ind. Microbiol. Biotechnol.* **2000**, *25*, 204–212. [CrossRef]

27. Li, Y.; Wu, Y.; Liu, Y.; Li, J.; Du, G.; Lv, X.; Liu, L. A genetic toolkit for efficient production of secretory protein in *Bacillus subtilis*. *Bioresour. Technol.* **2022**, *363*, 127885. [CrossRef]
28. Altenbuchner, J. Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **2016**, *82*, 5421–5427. [CrossRef]
29. Gan, T.; Fang, J.; Wang, Y.; Liu, K.; Sang, Y.; Chen, H.; Lu, Y.; Zhu, L.; Chen, X. Promoter engineering for efficient production of sucrose phosphorylase in *Bacillus subtilis* and its application in enzymatic synthesis of 2-O-α-D-glucopyranosyl-L-ascorbic acid. *Enzym. Microb. Technol.* **2023**, *169*, 110267. [CrossRef]
30. Guan, C.; Cui, W.; Cheng, J.; Liu, R.; Liu, Z.; Zhou, L.; Zhou, Z. Construction of a highly active secretory expression system via an engineered dual promoter and a highly efficient signal peptide in *Bacillus subtilis*. *New Biotechnol.* **2016**, *33*, 372–379. [CrossRef]
31. Jun, J.S.; Jeong, H.E.; Hong, K.W. Exploring and engineering novel strong promoters for high-level protein expression in *Bacillus subtilis* DB104 through transcriptome analysis. *Microorganisms* **2023**, *11*, 2929. [CrossRef] [PubMed]
32. Ma, R.J.; Wang, Y.H.; Liu, L.; Bai, L.L.; Ban, R. Production enhancement of the extracellular lipase LipA in *Bacillus subtilis*: Effects of expression system and Sec pathway components. *Protein Expr. Purif.* **2018**, *142*, 81–87. [CrossRef] [PubMed]
33. Frain, K.M.; Robinson, C.; Maarten Van Dijl, J. Transport of folded proteins by the Tat System. *Protein J.* **2019**, *38*, 377–388. [CrossRef] [PubMed]
34. Jongbloed, J.D.H.; Grieger, U.; Antelmann, H.; Hecker, M.; Nijland, R.; Bron, S.; Van Dijl, J.M. Two minimal Tat translocases in *Bacillus*. *Mol. Microbiol.* **2004**, *54*, 1319–1325. [CrossRef] [PubMed]
35. Palmer, T.; Berks, B.C. The twin-arginine translocation (Tat) protein export pathway. *Nat. Rev. Microbiol.* **2012**, *10*, 483–496. [CrossRef] [PubMed]
36. Liu, R.; Zuo, Z.; Xu, Y.; Song, C.; Jiang, H.; Qiao, C.; Xu, P.; Zhou, Q.; Yang, C. Twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct Tat-dependent secretion of methyl parathion hydrolase. *J. Agric. Food Chem.* **2014**, *62*, 2913–2918. [CrossRef] [PubMed]
37. Pan, X.; Yang, Y.; Liu, X.; Li, D.; Li, J.; Guo, X.; Zhou, Z. Secretory expression of a heterologous protein, Aiio-AIO6BS, in *Bacillus subtilis* via a non-classical secretion pathway. *Biochem. Biophys. Res. Commun.* **2016**, *478*, 881–886. [CrossRef] [PubMed]
38. Yang, H.; Ma, Y.; Zhao, Y.; Shen, W.; Chen, X. Systematic engineering of transport and transcription to boost alkaline α-amylase production in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 2973–2985. [CrossRef] [PubMed]
39. Morimoto, T.; Kadoya, R.; Endo, K.; Tohata, M.; Sawada, K.; Liu, S.; Ozawa, T.; Kodama, T.; Kakeshita, H.; Kageyama, Y.; et al. Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res.* **2008**, *15*, 73–81. [CrossRef]
40. Zhu, B.; Stülke, J. SubtiWiki in 2018: From genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res.* **2018**, *46*, D743–D748. [CrossRef]
41. Young, M. Gene amplification in *Bacillus subtilis*. *Microbiology* **1984**, *130*, 1613–1621. [CrossRef]
42. Neef, J.; Bongiorni, C.; Schmidt, B.; Goosens, V.J.; van Dijl, J.M. Relative contributions of non-essential Sec pathway components and cell envelope-associated proteases to high-level enzyme secretion by *Bacillus subtilis*. *Microb. Cell Fact.* **2020**, *19*, 6298–6305. [CrossRef] [PubMed]
43. Belitsky, B.R.; Sonenshein, A.L. Role and regulation of *Bacillus subtilis* glutamate dehydrogenase genes. *J. Bacteriol.* **1998**, *180*, 6298–6305. [CrossRef]
44. Simonen, M.; Palva, I. Protein secretion in *Bacillus* species. *Microbiol. Rev.* **1993**, *57*, 109. [CrossRef] [PubMed]
45. Li, M.; Wang, J.; Geng, Y.; Li, Y.; Wang, Q.; Liang, Q.; Qi, Q. A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*. *Microb. Cell Fact.* **2012**, *11*, 19. [CrossRef]
46. Yang, S.; Du, G.; Chen, J.; Kang, Z. Characterization and application of endogenous phase-dependent promoters in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 4151–4161. [CrossRef] [PubMed]
47. Lee, S.J.; Pan, J.G.; Park, S.H.; Choi, S.K. Development of a stationary phase-specific autoinducible expression system in *Bacillus subtilis*. *J. Biotechnol.* **2010**, *149*, 16–20. [CrossRef]
48. Mathy, N.; Bé, L.; Pellegrini, O.; Daou, R.; Wen, T.; Condon, C.N. 5′-to-3′ exoribonuclease activity in bacteria: Role of RNase J1 in rRNA maturation and 5′ stability of mRNA. *Cell* **2007**, *129*, 681–692. [CrossRef]
49. Liu, X.; Wang, H.; Wang, B.; Pan, L. Efficient production of extracellular pullulanase in *Bacillus subtilis* ATCC6051 using the host strain construction and promoter optimization expression system. *Microb. Cell Fact.* **2018**, *17*, 163. [CrossRef]
50. Salis, H.M.; Mirsky, E.A.; Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. biotechnol.* **2009**, *27*, 946–950. [CrossRef]
51. Tjalsma, H.; Bolhuis, A.; Jongbloed, J.D.H.; Bron, S.; van Dijl, J.M. Signal peptide-dependent protein transport in *Bacillus subtilis*: A genome-based survey of the secretome. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 515–547. [CrossRef]
52. Yuan, G.; Wong, S.L. Isolation and characterization of *Bacillus subtilis* groE regulatory mutants: Evidence for orf39 in the dnaK operon as a repressor gene in regulating the expression of both groE and dnaK. *J. Bacteriol.* **1995**, *177*, 6462. [CrossRef] [PubMed]
53. Barnett, J.P.; Van Der Ploeg, R.; Eijlander, R.T.; Nenninger, A.; Mendel, S.; Rozeboom, R.; Kuipers, O.P.; Van Dijl, J.M.; Robinson, C. The twin-arginine translocation (Tat) systems from *Bacillus subtilis* display a conserved mode of complex organization and similar substrate recognition requirements. *FEBS J.* **2009**, *276*, 232–243. [CrossRef] [PubMed]
54. Westers, L.; Westers, H.; Quax, W.J. *Bacillus subtilis* as cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta–Mol. Cell Res.* **2004**, *1694*, 299–310. [CrossRef] [PubMed]

55. Fu, L.L.; Xu, Z.R.; Li, W.F.; Shuai, J.B.; Lu, P.; Hu, C.X. Protein secretion pathways in *Bacillus subtilis*: Implication for optimization of heterologous protein secretion. *Biotechnol. Adv.* **2007**, *25*, 1–12.
56. Bolhuis, A.; Matzen, A.; Hyyryläinen, H.L.; Kontinen, V.P.; Meima, R.; Chapuis, J.; Venema, G.; Bron, S.; Freudl, R.; Van Dijl, J.M. Signal peptide peptidase- and ClpP-like proteins of *Bacillus subtilis* required for efficient translocation and processing of secretory proteins. *J. Biol. Chem.* **1999**, *274*, 24585–24592. [CrossRef]
57. Yu, X.; Xu, J.; Liu, X.; Chu, X.; Wang, P.; Tian, J.; Wu, N.; Fan, Y. Identification of a highly efficient stationary phase promoter in *Bacillus subtilis*. *Sci. Rep.* **2015**, *5*, 18405. [CrossRef]
58. Nielsen, A.K.; Rasmussen, M.D. Mutated Prokaryotic Cells with High Secretion-Levels. U.S. Patent No. 7,951,556, 31 May 2011.
59. Buescher, J.M.; Liebermeister, W.; Jules, M.; Uhr, M.; Muntel, J.; Botella, E.; Hessling, B.; Kleijn, R.J.; Le Chat, L.; Lecointe, F.; et al. Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. *Science* **2012**, *335*, 1099–1103. [CrossRef] [PubMed]
60. Liu, Y.; Su, A.; Tian, R.; Li, J.; Liu, L.; Du, G. Developing rapid growing *Bacillus subtilis* for improved biochemical and recombinant protein production. *Metab. Eng. Commun.* **2020**, *11*, e00141. [CrossRef]
61. Colletti, P.F.; Goyal, Y.; Varman, A.M.; Feng, X.; Wu, B.; Tang, Y.J. Evaluating factors that influence microbial synthesis yields by linear regression with numerical and ordinal variables. *Biotechnol. Bioeng.* **2011**, *108*, 893–901. [CrossRef]
62. Gunka, K.; Stannek, L.; Care, R.A.; Commichau, F.M. Selection-driven accumulation of suppressor mutants in *Bacillus subtilis*: The apparent high mutation frequency of the cryptic gudB gene and the rapid clonal expansion of gudB+ suppressors are due to growth under selection. *PLoS ONE* **2013**, *8*, e00141. [CrossRef] [PubMed]
63. Palva, I. Molecular cloning of α-amylase gene from *Bacillus amyloliquefaciens* and its expression in *B. subtilis*. *Gene* **1982**, *19*, 81–87. [CrossRef] [PubMed]
64. Lulko, A.T.; Veening, J.W.; Buist, G.; Smits, W.K.; Blom, E.J.; Beekman, A.C.; Bron, S.; Kuipers, O.P. Production and secretion stress caused by overexpression of heterologous alpha-amylase leads to inhibition of sporulation and a prolonged motile phase in *Bacillus subtilis*. *Appl. Environ. Microbiol.* **2007**, *73*, 5354–5362. [CrossRef] [PubMed]
65. Chang, A.Y.; Chau, V.W.; Landas, J.A.; Pang, Y. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. *JEMI Methods* **2017**, *1*, 22–25.
66. Yasbin, R.E.; Wilson, G.A.; Young, F.E. Transformation and transfection in lysogenic strains of *Bacillus subtilis*: Evidence for selective induction of prophage in competent cells. *J. Bacteriol.* **1975**, *121*, 296–304. [CrossRef] [PubMed]
67. Russell, D.W.; Sambrook, J. Molecular cloning: A laboratory manual. *Cold Spring Harb. Lab.* **2001**, *1*, 112.
68. Doench, J.G.; Fusi, N.; Sullender, M.; Hegde, M.; Vaimberg, E.W.; Donovan, K.F.; Smith, I.; Tothova, Z.; Wilen, C.; Orchard, R.; et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **2016**, *34*, 184–191. [CrossRef]
69. Wu, G.; Drufva, E.; Wu, K. Fast genome editing in *Bacillus subtilis*. *Eng. Life Sci.* **2019**, *19*, 471–477. [CrossRef]
70. Popp, P.F.; Dotzler, M.; Radeck, J.; Bartels, J.; Mascher, T. The Bacillus BioBrick Box 2.0: Expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. *Sci. Rep.* **2017**, *7*, 15058. [CrossRef]

Construction of an environmentally friendly super-secreting strain of *Bacillus subtilis* by systematic modulation of its secretory pathway using the CRISPR-Cas9 system

Jordi Ferrando, David Miñana-Galbis and Pere Picart

SUPPLEMENTARY MATERIAL

Additional file 1

Figure S1. PCR verification of *B. subtilis* mutants

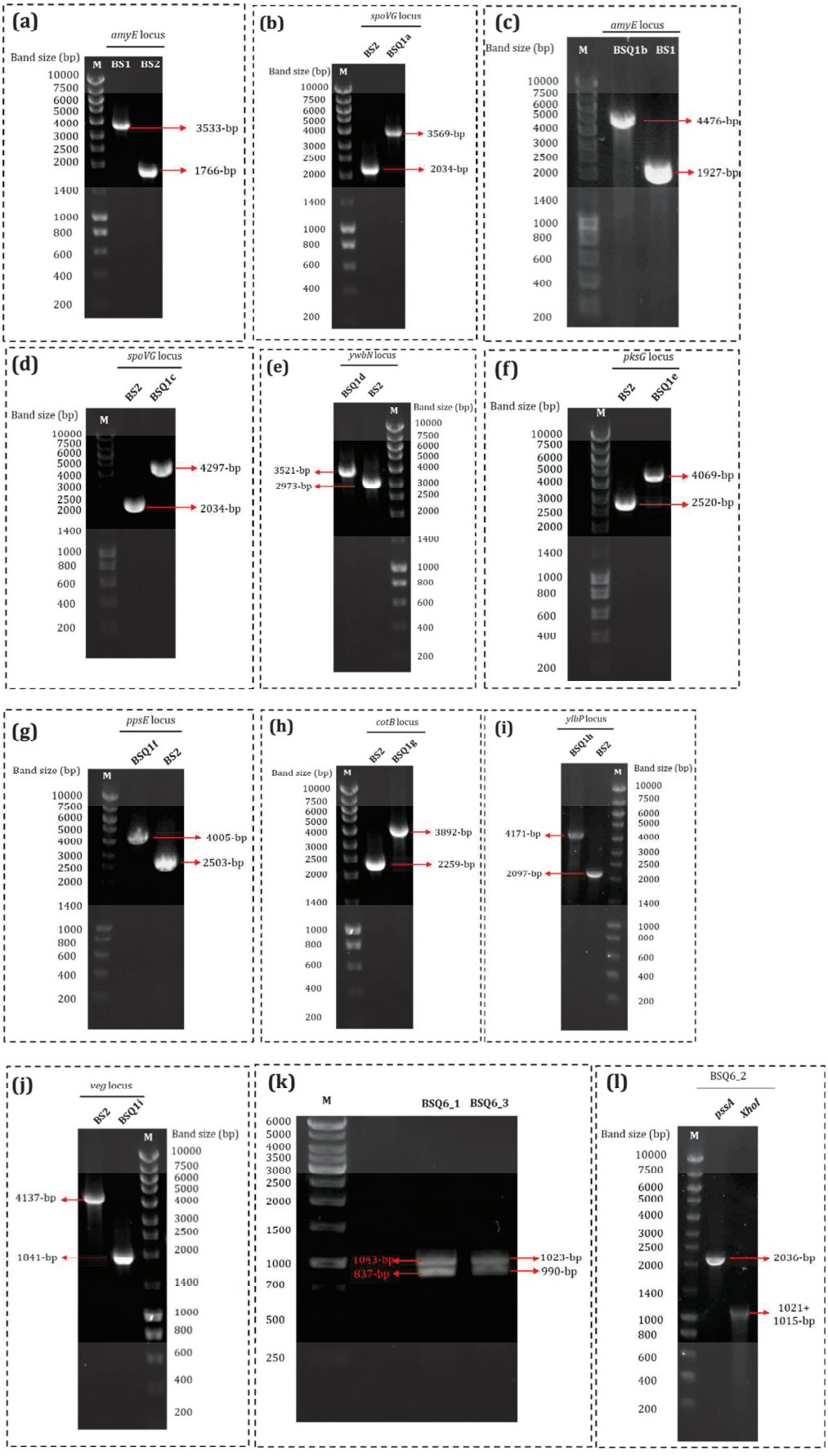
Figure S2. *amyQ* gene and dual promoter (P_{amyQ}-P_{cry3A}) sequence with codon optimization for *B. subtilis*

Figure S3. Sanger sequencing results of *gudB* gene in BSQ6_10 strain

Figure S4. Sequence of the synthetic *prsA* gene used in this study

Table S1. Primers designed in this study

Table S2. Splicing with overlap extension PCR (SOEing-PCR) program


(a)
amyE locus
Band size (bp)
M BS1 BS2
10000
7500
6000
5000
4000
3000
2500
2000
1400
1000
800
600
400
200
3533-bp
1766-bp
(b)
spoVG locus
BS2 BSQ1a
Band size (bp)
M
3569-bp
2034-bp
(c)
amyE locus
Band size (bp)
M BSQ1b BS1
4476-bp
1927-bp
(d)
spoVG locus
BS2 BSQ1c
Band size (bp)
M
4297-bp
2034-bp
(e)
ywbN locus
BSQ1d BS2
Band size (bp)
M
3521-bp
2973-bp
(f)
pksG locus
BS2 BSQ1e
Band size (bp)
M
4069-bp
2520-bp
(g)
ppsE locus
BSQ1f BS2
Band size (bp)
M
4005-bp
2503-bp
(h)
cotB locus
BS2 BSQ1g
Band size (bp)
M
3892-bp
2259-bp
(i)
ylbP locus
BSQ1h BS2
Band size (bp)
M
4171-bp
2097-bp
(j)
veg locus
BS2 BSQ1i
Band size (bp)
M
4137-bp
1041-bp
(k)
M
BSQ6_1 BSQ6_3
6000
5000
4000
3500
3000
2500
2000
1500
1000
700
500
250
1043-bp
837-bp
1023-bp
990-bp
(l)
BSQ6_2
pssA XhoI
Band size (bp)
M
2036-bp
1021+
1015-bp

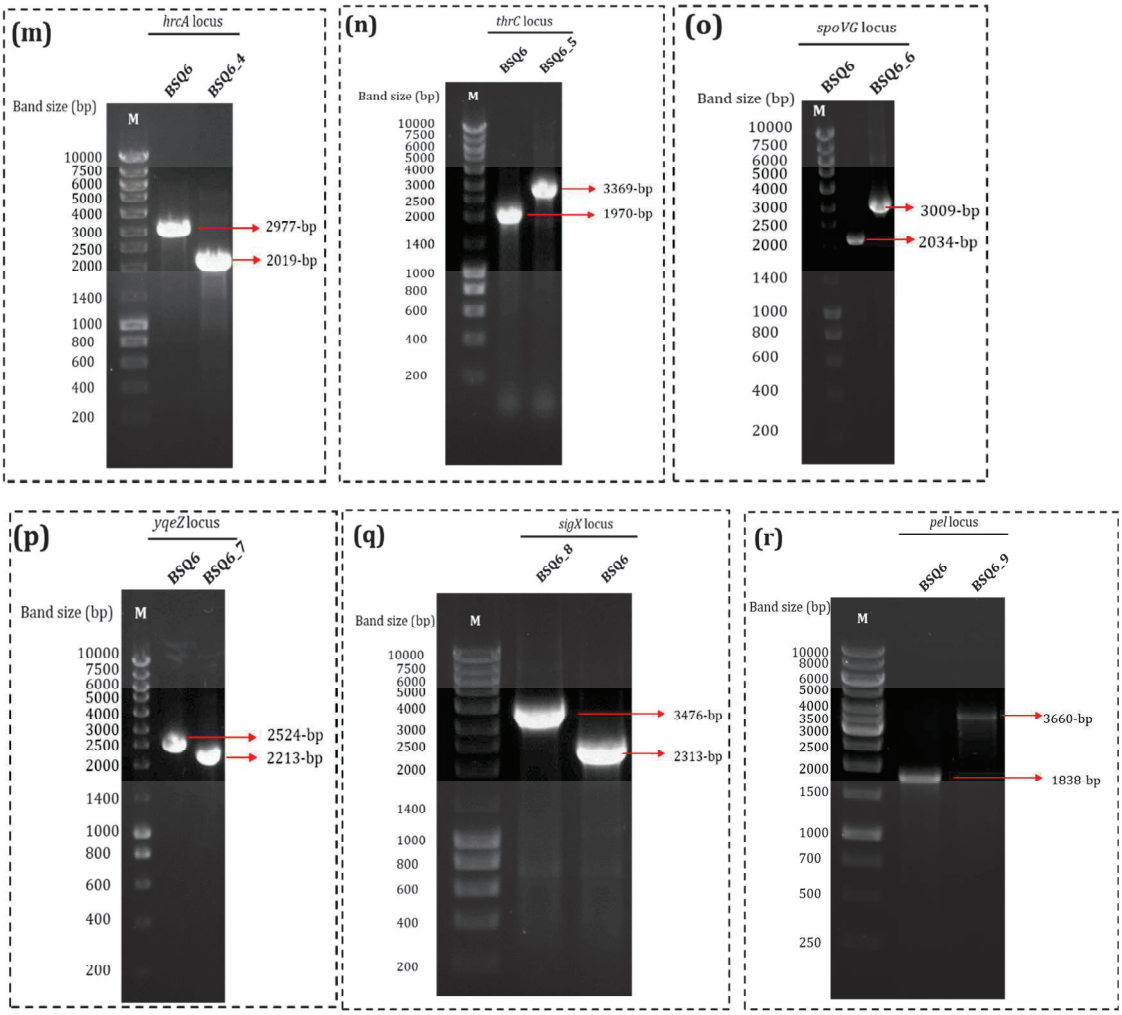


Figure S1. PCR verification of *B. subtilis* mutants. **(a)** *amyE* deletion in BS2 strain. BS1 strain was used as control; **(b)** P_{spoVG}-*amyQ* integration in BSQ1a. BS2 strain was used as control; **(c)** P_{amyQ}-P_{cry3A}-*amyQ* (*amyQ*_Ec) integration in BSQ1b strain. BS1 strain was used as control; **(d)** P_{spoVG}-P_{amyQ}-P_{cry3A}-*amyQ* integration in BSQ1c strain. BS2 strain was used as control; **(e)** *amyQ* gene with $SP_{ywbN'}$ integration in BSQ1d. BS2 strain was used as control; **(f)** *amyQ*_Ec integration in BSQ1e. BS2 strain was used as control; **(g)** *amyQ*_Ec integration in BSQ1f. BS2 strain was used as control; **(h)** *amyQ*_Ec integration in BSQ1g. BS2 strain was used as control; **(i)** *amyQ*_Ec integration in BSQ1h. BS2 strain was used as control; **(j)** *amyQ*_Ec integration in BSQ1i. BS2 strain was used as control; **(k)** insertion of Stop codon within *hag* and *yusX* genes in BSQ6_1 and BSQ6_3, respectively. PCR product digested with *XhoI*; **(l)** insertion of Stop codon within *pssA* gene in BSQ1_2. pssA, PCR product. XhoI, PCR product after digestion with *XhoI*; **(m)** *hrcA* deletion in BSQ6_4. BSQ6 strain was used as control; **(n)** *sipT* integration in BSQ6_5. BSQ6 strain was used as control; **(o)** *rasP* integration in BSQ6_6. BSQ6 strain was used as control; **(p)** *sppA* integration in BSQ6_7. BSQ6 strain was used as control; **(q)** *secYEG* operon integration in BSQ6_8. BSQ6 strain was used as control; **(r)** *prsA* integration in BSQ6_9. BSQ6 strain was used as control. M corresponds to the molecular weight marker. For knock-out strains, the following corresponding verification primer pairs were used: P1_1F/P1_2R (for *lytC*), P2_1F/P2_2R (for *amyE*), P12_1F/P12_2R (for *hag*), P13_1F/P13_2R (for *pssA*), P14_1F/P14_2R (for *yusX*) and P15_1F/P15_2R (for *hrcA*); whereas for knock-in strains the following primers were used: P5_1F/P5_3R (for *spoVG* site), P4_1F/P4_3R (for *amyE* site), P6_1F/P6_3R (for *ywbN* site), P7_1F/P7_3R (for *pksG* site), P8_1F/P8_3R (for *ppsE* site), P9_1F/P9_3R (for *cotB* site), P10_1F/P10_3R (for *ylbP* site), P11_1F/P11_3R (for *veg* site), P16_1F/P16_3R (for *thrC* site), P18_1F/P18_3R (for *yqeZ* site), P19_1F/P19_5R (for *sigX* site), P20_1F/P20_2R (for *gudB* site) and pelF/pelR (for *pel* site).

TGCTGTCCAGACTGTCCGCTGTGTAAAAATAAGGAATAAAGGGGGGTTGACATTATTTTACTGATATGTATAATATAA
TTTGTATAAGAAAATGTCGAAACGTAAGATGAAACCTTAGATAAAAGTGCTTTTTTTGTTGCAATTGAAGAATTATTA
ATGTTAAGCTTAATTAAAGATAATATCTTTGAATTGTAACGCCCCTCAAAAGTAAGAACTACAAAAAAAGAATACGTT
ATATAGAAATATGTTTGAACCTTCTTCAGATTACAAATATATTCGGACGGACTCTACCTCAAATGCTTATCTAACTAT
AGAATGACATACAAGCACAACCTTGAAAATTTGAAAATATAACTACCAATGAACTTGTTCATGTGAATTATCGCTGTA
TTTAATTTTCTCAATTCAATATATAATATGCCAATACATTGTTACAAGTAGAAATTAAGACACCCTTGATAGCCTTAC
TATACCTAACATGATGTAGTATTAAATGAATATGTAAATATATTTATGATAAGAAGCGACTTATTTATAATCATTACA
TATTTTTCTATTGGAATGATTAAGATTCCAATAGAATAGTGTATAAATTATTTATCTTGAAAGGAGGGATGCCTAAAA
ACGAAGAACATTAAAAACATATATTTGCACCGTCTAATGGATTTATGAAAAATCATTTTATCAGTTTGAAAATTATGT
ATTATGATAAGAAAGGGAGGACAAAC**ATGATTCAAAAACGAAAGCGGACAGTTTCGTTCAGACTTGTGCTTATGTGCA**
CGCTGTTATTTGTCAGTTTGCCGATTACAAAAACATCAGCCGTTAACGGCACACTTATGCAATACTTCGAATGGTACA
CACCTAACGATGGCCAACATTGGAAACGTCTTCAAAACGATGCTGAACATCTTTCTGATATCGGCATCACAGCTGTTT
GGATACCTCCTGCTTACAAAGGCCTTTCTCAATCTGATAACGGCTACGGCCCTTACGATCTTTACGATCTTGGCGAAT
TCCAACAAAAAGGCACAGTTCGTACAAAATACGGCACAAAATCTGAACTTCAAGATGCTATCGGCTCTCTTCATTCTC
GTAACGTTCAAGTTTACGGCGATGTTGTTCTTAACCATAAAGCTGGCGCTGATGCTACAGAAGATGTTACAGCTGTTG
AAGTTAACCCTGCTAACCGTAACCAAGAAACATCTGAAGAATACCAAATCAAAGCTTGGACAGATTTCCGTTTCCCTG
GCCGTGGCAACACATACTCTGATTTCAAATGGCATTGGTACCATTTCGATGGCGCTGATTGGGATGAATCTCGTAAAA
TCTCTCGTATCTTCAAATTCCGTGGCGAAGGCAAAGCTTGGGATTGGGAAGTTTCTTCTGAAAACGGCAACTACGATT
ACCTTATGTACGTTGATGTTGATTACGATCATCCTGATGTTGTTGCTGAAACAAAAAAATGGGGCATCTGGTACGCTA
ACGAACTTTCTCTTGATGGCTTCCGTATCGATGCTGCTAAACATATCAAATTCTCTTTCCTTCGTGATTGGGTTCAAG
CTGTTCGTCAAGCTACAGGCAAAGAAATGTTCACAGTTGCTGAATACTGGCAAAACGATGCTGGCAAACTTGAAAACT
ACCTTAACAAAACATCTTTCAACCAATCTGTTTTCGATGTTCCTCTTCATTTCAACCTTCAAGCTGCTTCTTCTCAAG
GCGGCGGCTACGATATGCGTCGTCTTCTTGATGGCACAGTTGTTTCTCGTCATCCTGAAAAAGCTGTTACATTCGTTG
AAAACCATGATACACAACCTGGCCAATCTCTTGAATCTACAGTTCAAACATGGTTCAAACCTCTTGCTTACGCTTTCA
TCCTTACACGTGAATCTGGCTACCCTCAAGTTTTCTACGGCGATATGTACGGCACAAAAGGCACATCTCCTCGTGAAA
TCCCTTCTCTTAAAGATTCTATCGAACCTATCCTTAAAGCTCGTAAAGAATACGCTTACGGCCCTCAACATGATTACA
TCGATCATCCTGATGTTATCGGCTGGACACGTGAAGGCGATTCTTCTGCTGCTAAATCTGGCCTTGCTGCTCTTATCA
CAGATGGCCCTGGCGGCTCTAAACGTATGTACGCTGGCCTTAAAAACGCTGGCGAAACATGGTACGATATCACAGGCA
ACCGTTCTGATACAGTTAAAATCGGCTCTGATGGCTGGGGCGAATTCCATGTTAACGATGGCTCTGTTTCTATCTACG
TTCAAAAATAA*GGTAATAAAAAAACACCTCCAAGCTGAGTGCGGGTATCAGCTTGGAGGTGCGTTTATTTTTTCAGCC*
GTATGACAAGGTCGGCATCAGGTGTGACAAATACGGTATGCTGGCTGTCATAGGTGACAAATCCGGGTTTTGCGCCGT
TTGGCTTTTTCACATGTCTGATTTTTGTATAATCAACAGGCACGGAGCCGGAATCTTTCGCCTTGGAAAAATAAGCGG
CGATCGTAGCTGCTTCCAATATGGATTGTTCATCGGGATCGCTGCTTTTAATCACAACGTG

Figure S2. *amyQ* gene and dual promoter (P_{amyQ}-P_{cry3A}) sequence with codon optimization for *B. subtilis*. The signal peptide SP_{amyQ} is marked in bold. Promoter and terminator are underlined.

Cryptic *gudB* gene in *B. subtilis* BS1 strain due to a duplication of 9 bp:

ATGGCAGCCGATCGAAACACCGGTCATACAGAAGAGGACAAACTTGATGTATTAAAATCAACCCAAACCGTAATACATAAGGCT
CTGGAAAAATTGGGATATCCCGAAGAGGTATACGAATTGTTAAAAGAGCCGATGAGATTATTAACGGTAAAAATACCTGTTCGT
ATGGACGACGGTTCAGTAAAGATTTTCACAGGATATCGTGCGCAGCACAATGACTCTGTCGGTCCAACGAAAGGCGGGATACG
TTTTCACCCGAACGTAACAGAAAAAGAG**GTGAAGGCGGTGAAGGCG**CTTTCAATTTGGATGAGTTTAAAATGCGGCATAATTG
ATCTTCCATATGGCGGTGGTAAAGGCGGAATTGTTTGTGATCCAAGGGATATGTCGTTTAGAGAGCTGGAGCGTCTGAGCAGA
GGGTATGTCAGAGCGATCAGCCAAATTGTCGGCCCGACAAAAGACGTGCCGGCACCGGATGTATTTACAAACTCACAAATCAT
GGCTTGGATGATGGATGAGTATTCAAGAATTGATGAATTTAATTCGCCTGGATTTATTACAGGCAAACCGCTTGTGCTTGGCGG
ATCTCACGGGAGAGAATCTGCGACAGCAAAAGGTGTTACCATCTGTATTAAAGAAGCGGCTAAGAAGAGAGGCATCGATATTA
AAGGTGCGCGTGTCGTTGTCCAAGGCTTCGGAAACGCGGGAAGCTATTTGGCAAAATTTATGCATGATGCGGGGGCAAAAGTT
GTCGGCATCTCAGATGCGTATGGCGGACTTTATGATCCGGAAGGCCTTGATATCGATTATTTACTCGACCGACGCGACAGCTTC
GGTACCGTAACAAAGCTTTTCAACGATACCATTACCAACCAAGAGCTGCTGGAGCTGGATTGTGATATTCTCGTTCCTGCTGCGA
TTGAAAATCAAATTACAGAAGAAAATGCCCATAATATCCGGGCTAAAATTGTCGTTGAAGCAGCGAACGGACCAACAACGCTTG
AAGGAACAAAAATTCTTTCAGACCGGGACATTCTGCTTGTACCAGACGTGCTGGCAAGTGCCGGTGGCGTAACAGTTTCTTATT
TTGAATGGGTTCAGAATAACCAAGGCTTCTACTGGAGTGAAGAAGAGGTAGAAGAAAAATTAGAAAAAATGATGGTCAAATCA
TTTAACAATATTTACGAAATGGCTAACAACCGAAGAATTGACATGAGGCTCGCTGCATATATGGTCGGCGTTCGCAAAATGGCT
GAAGCTTCGCGTTTTAGAGGCTGGATATAA

Sequencing results of restored *gudB* gene in BSQ6_10 strain with removed 9-bp direct repeat:

ATGGCAGCCGATCGAAACACCGGTCATACAGAAGAGGACAAACTTGATGTATTAAAATCAACCCAAACCGTAATACATAAGGCT
CTGGAAAAATTGGGATATCCCGAAGAGGTATACGAATTGTTAAAAGAGCCGATGAGATTATTAACGGTAAAAATACCTGTTCGT
ATGGACGACGGTTCAGTAAAGATTTTCACAGGATATCGTGCGCAGCACAATGACTCTGTCGGTCCAACGAAAGGCGGGATACG
TTTTCACCCGAACGTAACAGAAAAAGAG**GTGAAGGCG**CTTTCAATTTGGATGAGTTTAAAATGCGGCATAATTGATCTTCCATA
TGGCGGTGGTAAAGGCGGAATTGTTTGTGATCCAAGGGATATGTCGTTTAGAGAGCTGGAGCGTCTGAGCAGAGGGTATGTCA
GAGCGATCAGCCAAATTGTCGGCCCGACAAAAGACGTGCCGGCACCGGATGTATTTACAAACTCACAAATCATGGCTTGGATG
ATGGATGAGTATTCAAGAATTGATGAATTTAATTCGCCTGGATTTATTACAGGCAAACCGCTTGTGCTTGGCGGATCTCACGGG
AGAGAATCTGCGACAGCAAAAGGTGTTACCATCTGTATTAAAGAAGCGGCTAAGAAGAGAGGCATCGATATTAAAGGTGCGCG
TGTCGTTGTCCAAGGCTTCGGAAACGCGGGAAGCTATTTGGCAAAATTTATGCATGATGCGGGGGCAAAAGTTGTCGGCATCTC
AGATGCGTATGGCGGACTTTATGATCCGGAAGGCCTTGATATCGATTATTTACTCGACCGACGCGACAGCTTCGGTACCGTAAC
AAAGCTTTTCAACGATACCATTACCAACCAAGAGCTGCTGGAGCTGGATTGTGATATTCTCGTTCCTGCTGCGATTGAAAATCAA
ATTACAGAAGAAAATGCCCATAATATCCGGGCTAAAATTGTCGTTGAAGCAGCGAACGGACCAACAACGCTTGAAGGAACAAA
AATTCTTTCAGACCGGGACATTCTGCTTGTACCAGACGTGCTGGCAAGTGCCGGTGGCGTAACAGTTTCTTATTTTGAATGGGTT
CAGAATAACCAAGGCTTCTACTGGAGTGAAGAAGAGGTAGAAGAAAAATTAGAAAAAATGATGGTCAAATCATTTAACAATAT
TTACGAAATGGCTAACAACCGAAGAATTGACATGAGGCTCGCTGCATATATGGTCGGCGTTCGCAAAATGGCTGAAGCTTCGC
GTTTTAGAGGCTGGATATAA

Figure S3. Sanger sequencing results of *gudB* gene in BSQ6_10 strain. The 9-bp repeat (GTGAAGGCG) is marked in bold.

AATAGCCAACATTGACATTATTTTACTGATATGTATAATATAATTTGTATAAGAAAATGAGAGGGAGAGG
AAACATGAAGAAGATTGCAATTGCGGCGATTACAGCGACAAGCGTGCTGGCTCTCAGCGCATGCTCTGGC
GGCGATTCTGAAGTTGTTGCTGAAACAAAAGCTGGCAACATCACAAAAGAAGATCTTTACCAAACACTTA
AAGATAACGCTGGCGCTGATGCTCTTAACATGCTTGTTCAAAAAAAAGTTCTTGATGATAAATACGATGT
TTCTGATAAAGAAATCGATAAAAAACTTAACGAATACAAAAAATCTATGGGCGATCAACTTGATCAACTT
ATCAAACAAAAAGGCGAAGATTTCGTTAAAGAACAAATCAAATACGAACTTCTTATGAAAAAAGCTGCTA
AAGATAACATCAAAGTTACAGATGATGATGTTAAAGAATACTACGATGGCCTTAAAGGCAAAATCCATCT
TTCTCATATCCTTGTTAAAGAAAAAAAAACAGCTGAAGAAGTTGAAAAAAAACTTAAAAAAGGCGAAAAA
TTCGAAGATCTTGCTAAAGAATACTCTACAGATGGCACAGCTGAAAAAGGCGGCGATCTTGGCTGGGTTG
GCAAAGATGATAACATGGATAAAGATTTCGTTAAAGCTGCTTTCGCTCTTAAAACAGGCGAAATCTCTGG
CCCTGTTAAATCTCAATTCGGCTACCATATCATCAAAAAAGATGAAGAACGTGGCAAATACGAAGATATG
AAAAAAGAACTTAAAAAAGAAGTTGAAGAACAAAAACTTAACGATCAAACAGAACTTCAATCTGTTATCG
ATAAACTTGTTAAAGATGCTGATCTTAAAGTTAAAGATAAAGAACTTAAAAAACAAGTTGATCAACGTCA
AGCTCAAACATCTTCTTCTTCTTAA*CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCT*
TTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCC
TTTCTGCGTTTATA

Figure S4. Sequence of synthetic *prsA* gene used in this study. Promoter is underlined and terminator is in italics

Name	Sequence	Purpose	PCR Product
TS1F	5' – TACGAAAAATGTAATTATTGTAGG	**Target sequence for *lytC* gene deletion**	
TS1R	5' – AAACCCTACAATAATTACATTTTT		
		PCR of homology template for *lytC* gene deletion	**2101**
P1_1F	5' – AAGGCCAACGAGGCCGAACTACTGTACCTGATACAAC	Upstream *lytC* gene homologous arm	1108
P1_1R	5' – GGTGTTCCCATGCCTCGAGAATAATTACATTTTTAGTCTGCATC		
P1_2F	5' – CTCGAGGCATGGGAACACCTGCTGTTTCTTC	Downstream *lytC* gene homologous arm	993
P1_2R	5' – AAGGCCTTATTGGCCTCTCTCGTTCCAAGATTAGCC		
TS2F	5' – TACGACAGCACCGTCGATCAAAAG	**Target sequence for *amyE* gene deletion**	
TS2R	5' – AACCTTTTGATCGACGGTGCTGT		
		PCR of homology template for *amyE* gene deletion	**1766**
P2_1F	5' – AAGGCCAACGAGGCCGAGTTATTCATTGCAGAAGCGCA	Upstream *amyE* gene homologous arm	890
P2_1R	5' – GTTGTATCCCCTGCCAGAACCAAATGAAAC		
P2_2F	5' – GTTCTGGCAGGGGATACAACCAACGCAAAAGTG	Downstream *amyE* gene homologous arm	876
P2_2R	5' –AAGGCCTTATTGGCCGTGGCTCCAAACAGGAAGC		
TS3F	5' – TACGAAATCAGACATGCTCGGACG	**Target sequence for *spoVG* replacement**	
TS3R	5' – AAACCGTCCGAGCATGTCTGATTT		
		PCR of homology template for *spoVG* gene replacement for *amyQ gene* (single promoter)	**2894**
P3_1F	5' – AAGGCCAACGAGGCCCGGGCCTTATTCACAAGGG	Upstream *spoVG* gene homologous arm	530
P3_1R	5' – CTTTCGTTTTTGAATCATAGTAGTTCACCACCTTTTCCC		
P3_2F	5' – GTGAACTACTATGATTCAAAAACGAAAGCGGAC	*amyQ* gene amplification	1829
P3_2R	5' – GGTTATTTTCACGTTGTGATTAAAAGCAGCG		
P3_3F	5' – GCTTTTAATCACAACGTGAAAATAACCAAAAAGCAAGGACTG	Downstream *spoVG* gene homologous arm	535
P3_3R	5' – AAGGCCTTATTGGCCTTACTGTTCCATCGTCTCTGCTG		
		PCR of homology template for *amyQ* gene integration at *amyE* site	**4476**
P4_1F	5' – AAGGCCAACGAGGCCGAGTTATTCATTGCAGAAGCGCA	Upstream *amyE* gene homologous arm	951
P4_1R	5' – GCGGACAGTCTGGACAGCAGATCGACGGTGCTGTAAGCT		
P4_2F	5' – AGCTTACAGCACCGTCGATCTGCTGTCCAGACTGTCCGC	*amyQ* gene amplification	2557
P4_2R	5' – CCATGCATGAAGAATGGTTCCTCACGTTGTGATTAAAAGCAGCGAT		
P4_3F	5' – ATCGCTGCTTTTAATCACAACGTGAGGAACCATTCTTCATGCATGG	Downstream *amyE* gene homologous arm	968
P4_3R	5' – AAGGCCTTATTGGCCCTTCAAATAAAGCACTCCCGC		
		PCR of homology template for *spoVG* gene replacement for *amyQ gene* (triple promoter)	**4297**
P5_1F	5' – AAGGCCAACGAGGCCCTTGCGAAAAGAAGCATGAAAAC	Upstream *spoVG* gene homologous arm	864
P5_1R	5' – CCCTATATAAAAGCATTAGTGTATC		
P5_2F	5' – ACACTAATGCTTTTATATAGGGTGCTGTCCAGACTGTCCG	*amyQ* gene amplification	2557
P5_2R	5' – GGTTATTTTCACGTTGTGATTAAAAGCAGCG		
P5_3F	5' – GCTTTTAATCACAACGTGAAAATAACCAAAAAGCAAGGACTG	Downstream *spoVG* gene homologous arm	876
P5_3R	5' – AAGGCCTTATTGGCCATTAACTCCGAGCGTTTCTTGG		
TS6F	5' – TACGTTTAAAATGGGGAGCGATGG	**Target sequence for *ywbN* replacement**	

TS6R	5′ – AAACCCATCGCTCCCCATTTTAAA		
		PCR of homology template for *ywbN* gene replacement for *amyQ gene* (Tat pathway)	**3521**
P6_1F	5′ – AAGGCCAACGAGGCCCTCAGTGGAAAAAGGAGACC		
P6_1R	5′ – GATTTTGAATTGTCCCGTCTGTGAATTTGTTTTGGCTTTTTTCTCTGTTTATCGCTCATGATGTTACAAAACT	Upstream *ywbN* gene homologous arm	1062
P6_2F	5′ – CACAGACGGGACAATTCAAAATCGGGAGCGATGGCAGGGGCAGCCGTTGCGGTTAACGGCACACTTATGCAAT	*amyQ* gene amplification	1736
P6_2R	5′ – GCATGCGTATAAAGCAGAACCCACGTTGTGATTAAAAGCAGC		
P6_3F	5′ – GGTTCTGCTTTATACGCATGC		
P6_3R	5′ – AAGGCCTTATTGGCCGTTTCCAGTGCTGATTTGAATG	Downstream *ywbN* gene homologous arm	723

Table S1. Primers designed in this study.

Name	Sequence	Purpose	PCR Product
TS7F	5′ – TACGGAGGTGTTTCAAATAGATCC	**Target sequence of *pksG* gene for *amyQ* gene integration**	
TS7R	5′ – AAACGGATCTATTTGAAACACCTC		
		PCR of homology template for *amyQ* gene integration at *pksG* site	**4069**
P7_1F	5′ – AAGGCCAACGAGGCCCTTGGTGCTTTAATGGATCTTTC	Upstream *pksG* gene homologous arm	
P7_1R	5′ – CAGAGCAAGTGATCAGCAATTC		928
P7_2F	5′ – GAATTGCTGATCACTTGCTCTGTGCTGTCCAGACTGTCCG	*amyQ* gene amplification	
P7_2R	5′ – CGTAGGAAAAACAGCCAATTCGCACGTTGTGATTAAAAGCAGCG		2557
P7_3F	5′ – CGAATTGGCTGTTTTTCCTACG	Downstream *pksG* gene homologous arm	
P7_3R	5′ – AAGGCCTTATTGGCCCGCAAACGCAATAAATGCTTTC		884
TS8F	5′ – TACGACCGTCAGAAATCCCAGGCG	**Target sequence of *ppsE* gene for *amyQ* gene integration**	
TS8R	5′ – AAACCGCCTGGGATTTCTGACGGT		
		PCR of homology template for *amyQ* gene integration at *ppsE* site	**4005**
P8_1F	5′ – AAGGCCAACGAGGCCGATTAAACTGAACGCAGACTTGC	Upstream *ppsE* gene homologous arm	
P8_1R	5' – CTAGAAGAGCGACTGCTCAACATAC		686
P8_2F	5' – GTTGAGCAGTCGCTCTTCTAGTGCTGTCCAGACTGTCCG	*amyQ* gene amplification	
P8_2R	5' – CAGGAGCAACGACAAGATTAAAGCACGTTGTGATTAAAAGCAGCG		2557
P8_3F	5' – CTTTAATCTTGTCGTTGCTCCTG	Downstream *ppsE* gene homologous arm	
P8_3R	5′ – AAGGCCTTATTGGCCGACGTGTAATTGCTGTTTCTGAC		762
TS9F	5′ – TACGCACTATAGTATAATGGCCGT	**Target sequence of *cotB* gene for *amyQ* gene integration**	
TS9R	5′ – AAACACGGCCATTATACTATAGTG		
		PCR of homology template for *amyQ* gene integration at *cotB* site	**3892**
P9_1F	5' – AAGGCCAACGAGGCCCACAAGAGGAACTTTGGAAGG	Upstream *cotB* gene homologous arm	
P9_1R	5' – CATCAAGATCACCAGGCTATTC		670
P9_2F	5' – GAATAGCCTGGTGATCTTGATGTGCTGTCCAGACTGTCCG	*amyQ* gene amplification	
P9_2R	5' – GTTACTGTATATCGTGGAGGTCCACGTTGTGATTAAAAGCAGCG		2557
P9_3F	5' – GACCTCCACGATATACAGTAAC	Downstream *cotB* gene homologous arm	
P9_3R	5' – AAGGCCTTATTGGCCGCTGGTATTGTATTCACCTCG		665
TS10F	5' –TACGGCTTATCAACTATAAAACGC	**Target sequence of *ylbP* gene for *amyQ* gene integration**	
TS10R	5' –AAACGCGTTTTATAGTTGATAAGC		
		PCR of homology template for *amyQ* gene integration at *ylbP* site	**4171**
P10_1F	5' – AAGGCCAACGAGGCCGATAAGCTGGCGTTGTCAG	Upstream *ylbP* gene homologous arm	
P10_1R	5' – ACAAATCTCCCCCTTTGTTG		780
P10_2F	5' – CAACAAAGGGGGAGATTTGTCACGTTGTGATTAAAAGCAGC	*amyQ* gene amplification	
P10_2R	5' – AATAGGCCGGTTGTTTTGATTTTGCTGTCCAGACTGTCCG		2557
P10_3F	5' – AAATCAAAACAACCGGCCTATT	Downstream *ylbP* gene homologous arm	
P10_3R	5' – AAGGCCTTATTGGCCAAGCATGCGGATCATACAAC		834
TS11F	5' –TACGCTGACGTTAAAAGCAAACGG	**Target sequence of *veg* gene for *amyQ* gene integration**	
TS11R	5' –AAACCCGTTTGCTTTTAACGTCAG		
		PCR of homology template for *amyQ* gene integration at *veg* site	**4137**
P11_1F	5' – AAGGCCAACGAGGCCGCCAATATCAGCATCAGGAG	Upstream *veg* gene homologous arm	
P11_1R	5' – TGCATCCACCTCACTACATTT		797
P11_2F	5' – AAATGTAGTGAGGTGGATGCACACGTTGTGATTAAAAGCAGC	*amyQ* gene amplification	
P11_2R	5' – CAAAAGGTTCACTGCCCGTTATGCTGTCCAGACTGTCCG		2557
P11_3F	5' – TAACGGGCAGTGAACCTTTTG	Downstream *veg* gene homologous arm	
P11_3R	5' –AAGGCCTTATTGGCCAGCCACCGGAATTACCTTC		783

Name	Sequence	Purpose	PCR Product
TS12F	5′ – TACGAGGTCTTCGCATCAACCGTG	**Target sequence for *hag* gene deletion**	
TS12R	5′ – AAACCACGGTTGATGCGAAGACCT		
		PCR of homology template for *hag* gene deletion	**1880**
P12_1F	5′ – AAGGCCAACGAGGCCGCGGGATTCCAGGCTTTTTAG	Upstream *hag* gene homologous arm	837
P12_1R	5′ – CATCTCCCCTAAGATCTCGAGTTGATGCGAAGACCTGAAGAAAG		
P12_2F	5′ – CTCGAGATCTTAGGGGAGATGACGCAGCAGGT	Downstream *hag* gene homologous arm	1043
P12_2R	5′ – AAGGCCTTATTGGCCGTTCAATTGATCATCCCCTATGC		
TS13F	5′ – TACGACGATAGGAAACTTCATTTG	**Target sequence for *pssA* gene deletion**	
TS13R	5′ – AAACCAAATGAAGTTTCCTATCGT		
		PCR of homology template for *pssA* gene deletion	**2036**
P13_1F	5′ – AAGGCCAACGAGGCCCTATGTTTCTACTCG	Upstream *pssA* gene homologous arm	1021
P13_1R	5′ – AGCAATCCCTATCGCTCGAGGAAGTTTCCT		
P13_2F	5′ – CTCGAGCGATAGGGATTGCTGGCGATTCAT	Downstream *pssA* gene homologous arm	1015
P13_2R	5′ – AAGGCCTTATTGGCCGATTCGTCAGCTCTA		
TS14F	5′ – TACGCTTGAAACCTATAAAGCGCG	**Target sequence for *yusX* gene deletion**	
TS14R	5′ – AAACCGCGCTTTATAGGTTTCAAG		
		PCR of homology template for *yusX* gene deletion	**2013**
P14_1F	5′ – AAGGCCAACGAGGCCTTCCGCTTTTTCAGGCTGTC	Upstream *yusX* gene homologous arm	990
P14_1R	5′ – TCCCAGCCCAGTCCTCGAGTTTATAGGTTTCAAGCCGGAATC		
P14_2F	5′ – CTCGAGGACTGGGCTGGGAGAATGTCCTGA	Downstream *yusX* gene homologous arm	1023
P14_2R	5′ – AAGGCCTTATTGGCCACGCACAGCCAGCTTGATG		
TS15F	5′ – TACGTATTAAATCGGCACAGCCGG	**Target sequence for *hrcA* gene deletion**	
TS15R	5′ – AAACCCGGCTGTGCCGAATTTAATA		
		PCR of homology template for *hrcA* gene deletion	**2019**
P15_1F	5′ – AAGGCCAACGAGGCCCGTGCTAAAGCCCTCAAGC	Upstream *hrcA* gene homologous arm	1005
P15_1R	5′ – GTCCGAAGTCACCAGCAGCTGACGATTTGTTAAC		
P15_2F	5′ – GCTGCTGGTGACTTCGGACTTGTCAAAAG	Downstream *hrcA* gene homologous arm	1014
P15_2R	5′ – AAGGCCTTATTGGCCCTTTCAGCGTATGATTTAAGGTG		
TS16F	5′ – TACGACGGGATCATTTAAAGATCG	**Target sequence of *thrC* gene for *sipT* gene integration**	
TS16R	5′ – AAACCGATCTTTAAATGATCCCGT		
		PCR of homology template for *thrC* gene replacement for *spiT* gene	**3369**
P16_1F	5′ – AAGGCCAACGAGGCCGATGGGAATTGTGAACGGCAC	Upstream *thrC* gene homologous arm	1017
P16_1R	5′ – GCGGACAGTCTGGACAGCATTTAAATGATCCCGTAGGATTGACG		
P16_2F	5′ – CGTCAATCCTACGGGATCATTTAAATGCTGTCCAGACTGTCCGC	$P_{amyQ\text{-}cry3a}$ amplification	728
P16_2R	5′ – GTATTCGTATTTTTTTCCTCGGTCATGTTTGTCCTCCCTTTCTTATCAT		
P16_3F	5′ – ATGATAAGAAAGGGAGGACAAACATGACCGAGGAAAAAAATACGAATAC	*sipT* gene amplification	582
P16_3R	5′ – TGCCACAGCCATAACCATTCCTTGGCTATTATCGGTTCAGTGTTCA		
P16_4F	5′ – TGAACACTGAACCGATAATAGCCAAGGAATGGTTATGGCTGTGGCA	Downstream *thrC* gene homologous arm	1042
P16_4R	5′ – AAGGCCTTATTGGCCTGAATCAAAGCCGGGGCCTA		
		Gibson assembly primers for *prsA* gene integration at *pel* site	**8109**
AF	5′ – GCTCCTACTGAAAGATTCAGC	Upstream *pel* gene homologous arm	2702
AR	5′ – GCCATGTCACTATTGCTTCAGGTCCAAATCATACTCCGGATC		
BF	5′ – GATCCGGAGTATGATTTGGACCTGAAGCAATAGTGACATGGC	*erythromycin* resistance gene amplification from pBS2EXxylRPxylA	1721
BR	5′ – GTAAAATAATGTCAATGTTGGCTATTCAGGATGGCCTTCTGCTTAG		
CF	5′ – CTAAGCAGAAGGCCATCCTGAATAGCCAACATTGACATTATTTTAC	Synthetic *prsA* gene amplification	1064
CR	5′ – GATTTCACATTAGCAGAAGCATCTATAAACGCAGAAAGGCCCAC		
DF	5′ – GTGGGCCTTTCTGCGTTTATAGATGCTTCTGCTAATGTGAAATC	Downstream *pel* gene homologous arm	2622
DR	5′ – CACTTGATCGTTGACGCTTAC		
pelF	5′ – GTCTTAGCTCACGACGATGA	PCR for *prsA* integration verification	
pelR	5′ – TCCCCTTTTCCATTCTACTCCT		

Name	Sequence	Purpose	PCR Product
		PCR of homology template for *spoVG* gene replacement for *rasP gene*	**3009**
P17_1F	5′ – AAGGCCAACGAGGCCCTTGCGAAAAGAAGCATGAAAAC	Upstream *spoVG* gene homologous arm	864
P17_1R	5′ – CGCTATAACTGTATTCACGAACATAGTAGTTCACCACCTTTTCCC		
P17_2F	5′ – GGGAAAAGGTGGTGAACTACTATGTTCGTGAATACAGTTATAGCG	*rasP* gene amplification	1269
P17_2R	5′ – CAGTCCTTGCTTTTTGGTTATTTTTTACAAAAACAGCCGCTGGATAT		
P17_3F	5′ – ATATCCAGCGGCTGTTTTTGTAAAAAATAACCAAAAAGCAAGGACTG	Downstream *spoVG* gene homologous arm	876
P17_3R	5′ – AAGGCCTTATTGGCCATTAACTCCGAGCGTTTCTTGG		
TS18F	5′ – TACGTAACGCAGACAAGGCGATTG	**Target sequence for *yqeZ* replacement**	
TS18R	5′ – AAACCAATCGCCTTGTCTGCGTTA		
		PCR of homology template for *yqeZ* gene replacement for *sppA gene*	**2213**
P18_1F	5′ – AAGGCCAACGAGGCCGGTGGATTTATGAGTCTTCTTGAGC	Upstream *yqeZ* gene homologous arm	586
P18_1R	5′ – GGGCGGTGTATCCCTCCTTC		
P18_2F	5′ – GAAGGAGGGATACACCGCCCATGAATGCAAAAAGATGGATTGC	*sppA* gene amplification	1008
P18_2R	5′ – CATATAACTTCTCCTCGTTTCTATTCTACTTCGCATAGAGATACATC		
P18_3F	5′ – AATAGAAACGAGGAGAAGTTATATG	Downstream *yqeZ* gene homologous arm	619
P18_3R	5′ – AAGGCCTTATTGGCCCATGTCAGGGTTTTCAAGCAC		
TS19F	5′ – TACGTACATTTGACTGGGATACAC	**Target sequence for *sigX* replacement**	
TS19R	5′ – AAACGTGTATCCCAGTCAAATGTA		
		PCR of homology template for *sigX* gene replacement for *artificial secYEG operon*	**3476**
P19_1F	5′ – AAGGCCAACGAGGCCAGTCATGAGCTGAGAACAC	Upstream *sigX* gene homologous arm	856
P19_1R	5′ – GTTGGAGATTGTTTTAAACAATTGAAACCCCTCCGTTCAC		
P19_2F	5′ – GTGAACGGAGGGGTTTCAATTGTTTAAAACAATCTCCAACTTTATGC	*secY* gene amplification	1296
P19_2R	5′ – GTAAAAGACCTCCACAATTTCTAGTTTTTCATAAATCCACGGTAG		
P19_3F	5′ – CGTGGATTTATGAAAAACTAGAAATTGTGGAGGTCTTTTACATGC	*secE* gene amplification	200
P19_3R	5′ – CACCTCCAGACTCACTTATTCAACTATTAAACGAATTAATTGAG		
P19_4F	5′ – TTAATTCGTTTAATAGTTGAATAAGTGAGTCTGGAGGTGTATGG	*secG* gene amplification	252
P19_4R	5′ – CTGAGGCGAACGATGGTCTTCTATAGGATATAAGCAAGCGCAATC		
P19_5F	5′ – GCTTGCTTATATCCTATAGAAGACCATCGTTCGCCTCAG	Downstream *sigX* gene homologous arm	872
P19_5R	5′ – AAGGCCTTATTGGCCAATCATCAACTTCTGACTCC		
TS20F	5′ – TACGAAAAGAGGTGAAGGCGGTGA	**Target sequence for *gudB* replacement**	
TS20R	5′ – AAACTCACCGCCTTCACCTCTTTT		
		PCR of homology template for *gudB* repair	**1203**
P20_1F	5′ – AAGGCCAACGAGGCCGAAGATCTAGGGTCACATACG	Upstream *gudB* gene homologous arm	596
P20_1R	5′ – CCAAATTGAAAGCGCCTTCACCTCTTTTTCTG		
P20_2F	5′ – GAGGTGAAGGCGCTTTCAATTTGG	Downstream *gudB* gene homologous arm	607
P20_2R	5′ – AAGGCCTTATTGGCCGAATATCACAATCCAGCTCCAG		

Table S2. Splicing with overlap extension PCR (SOEing-PCR) program.

OVERLAP EXTENSION PCR	TEMPERATURE	TIME
STEP 1		
Initial Denaturation	95°C	5 minutes
15 cycles	95°C	30 seconds
	X °C *	1:30 minutes
	72°C	X minutes **
Final Extension	72°C	5 minutes
STEP 2		
Initial Denaturation	95°C	5 minutes
30 cycles	95°C	30 seconds
	X °C *	30 seconds
	72°C	X minutes **
Final Extension	72°C	5 minutes
Hold	4 °C	

** Depending on the melting temperature (Tm) of the primer

** Depending on the length of the fragment to be amplified

4.3. Chapter 3 Expanding the CRISPR-Cas9 toolbox in *Bacillus subtilis*

4.3.1. Paper 3: Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system

4.3.2. Paper 4: Metabolic engineering of *Bacillus subtilis* toward the efficient and stable production of C30-carotenoids.

Ferrando *et al. Microbial Cell Factories* (2023) 22:21
https://doi.org/10.1186/s12934-023-02032-2

Microbial Cell Factories

RESEARCH

Open Access

Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system

Jordi Ferrando[1†], Oriana Filluelo[1†], Daniel R. Zeigler[2] and Pere Picart[1*]

Abstract

Background Despite recent advances in genetic engineering tools for effectively regulating and manipulating genes, efficient simultaneous multigene insertion methods have not been established in *Bacillus subtilis*. To date, multilocus integration systems in *B. subtilis*, which is one of the main industrial enzyme producers and a GRAS (generally regarded as safe) microbial host, rely on iterative rounds of plasmid construction for sequential insertions of genes into the *B. subtilis* chromosome, which is tedious and time consuming.

Results In this study, we present development and proof-of-concept of a novel CRISPR-Cas9-based genome-editing strategy for the colorimetric detection of one-step multiple gene insertion in *B. subtilis*. First, up to three copies of the *crtMN* operon from *Staphylococcus aureus*, encoding a yellow pigment, were incorporated at three ectopic sites within the *B. subtilis* chromosome, rendering engineered strains able to form yellow colonies. Second, a single CRISPR-Cas9-based plasmid carrying a highly specific single guide RNA (sgRNA) targeting *crtMN* operon and a changeable editing template was constructed to facilitate simultaneous insertion of multiple gene-copies through homology-directed repair (HDR). Upon transformation of engineered strains with engineered plasmids, strains harboring up to three gene copies integrated into the chromosome formed white colonies because of the removal of the *crtMN* operon, clearly distinguishable from yellow colonies harboring undesired genetic modifications. As a result, construction of a plasmid-less, marker-free, high-expression stable producer *B. subtilis* strain can be completed in only seven days, demonstrating the potential that the implementation of this technology may bring for biotechnology purposes.

Conclusions The novel technology expands the genome-editing toolset for *B. subtilis* and means a substantial improvement over current methodology, offering new application possibilities that we envision should significantly boost the development of *B. subtilis* as a chassis in the field of synthetic biology.

[†]Jordi Ferrando and Oriana Filluelo have contributed equally to this work

*Correspondence:
Pere Picart
perepicart@ub.edu
Full list of author information is available at the end of the article

Keywords *B. subtilis*, CRISPR-Cas9, Colorimetric screening, Genome engineering, Multigene insertion

Background

Bacillus is a genus of gram-positive, rod-shaped bacteria, which due to their wide distribution, safety in work, ease of cultivation, and susceptibility to genetic transformation, have been widely used to produce heterologous proteins [1, 2]. *B. subtilis* and related *Bacillus* strains are the dominant enzyme-producing microorganisms in applied and industrial microbiology owing to their ability to secrete enzymes at very high levels [3–5], in addition to their extensive use to produce drug precursors, platform compounds, biofuels and biopolymers [6, 7]. To achieve maximum expression of a particular gene in *Bacillus*, it is highly desirable to amplify the copy number of the gene of interest [8]. To this end, a traditional approach involves the introduction of replicative plasmids, where the level of gene expression is dictated by the copy number of the plasmid in the cells [9]. However, the use of antibiotic resistance markers limits their use in industrial applications due to both the genetic instability of many recombinant plasmids in the absence of selection and the restrictions and concerns derived from the massive abuse of antibiotics, promoting the emergence of bacterial resistance.

Protein expression attributed to a single copy of the integrated gene usually results in lower yields of product compared to the use of high copy-number vectors. To circumvent this limitation, the use of a single-crossover integrative vector, which creates direct repeats of the target fragment upon its insertion into the chromosome, offers the possibility of amplifying the integrated plasmid by growing cultures in increasing concentrations of the selective antibiotic and, therefore, amplifying gene dosage in the chromosome [10]. However, although some reports indicate that gene amplification is stable under non-selective conditions [11, 12], other authors show that integrated plasmid copies are gradually lost in the absence of selection [13–15], resulting in unstable strains not suitable for industrial applications. On the other hand, gene-replacement strategies based on double-crossover recombination integrative plasmids allow introduction of mutations that are stable in the absence of ongoing selection. Nevertheless, this approach has the disadvantage that resulting strains have low gene dosage unless multiple rounds of gene insertion are performed [16–18], until reaching expression levels comparable to that of cells carrying multiple copies of a recombinant plasmid. Therefore, the construction of environmentally friendly, marker-free industrial strains of *B. subtilis* with multicopy genes is limited by the availability of selection markers, involving labor-intensive methods of introducing recycling markers. Such methods for optimal marker recycling in *B. subtilis* have been developed based on: (i) counter-selectable markers such as *mazF*, *blaI*, *ysbC* and *uppC* [19–23]; (ii) site-specific recombinase systems such as Cre/LoxP and FLP/FRT [24, 25]; and (iii) the λ-Red phage mediated single-stranded DNA recombination system [26], thus allowing the removal of the selectable marker once the desired chromosome modification is performed in order to reuse it in further rounds of modification. Nevertheless, these methods are still time consuming, laborious, and quite inefficient.

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeat-CRISPR associated protein) systems, especially with nuclease Cas9, were rapidly adapted for genome editing in *B. subtilis*, thus facilitating the introduction of gene mutations, deletions, and insertions [27–30]. Typically, a single plasmid containing the Cas9 endonuclease is targeted to a specific site by a 20-nucleotide sgRNA also present in the vector. By means of homologous recombination with a plasmid-cloned editing template, this system enables genome editing and cell survival. The introduction of the CRISPR-Cas9 system simultaneously removed the need for selection markers in genome editing and radically increased editing efficiencies, becoming one of the most powerful tools for genome engineering in *B. subtilis* [31–35]. However, to the best of our knowledge, no editing efficiencies have been reported for simultaneous integration of multiple genes in *B. subtilis*, possibly due to the low integration efficiencies achieved during the process.

For the first time, the present study aimed to develop both a CRISPR-Cas9-mediated genomic multigene insertion method in *B. subtilis* and a colorimetric high-throughput screening method for identification of multicopy clones. To this end, the *crtMN* operon encoding yellow pigment from *S. aureus* was first integrated into the *B. subtilis* chromosome at three ectopic sites, thus obtaining a yellowish-pigmented strain. Using a single CRISPR-Cas9-based plasmid harboring a unique high-efficiency target site and changeable editing template, we developed a white/yellow colorimetric screening for white colonies because of the removal of the *crtMN* operon, which if still present leads to the formation of yellow colonies in *B. subtilis*. Thus, an easy, fast, and suitable method to identify white clones for genomic

double- (DGI) and triple-gene integration (TGI) in a single step was established. By using this technology, we were able to construct plasmid-less, marker-free stable producer strains harboring up to three copies of an α-amylase gene inserted into the *B. subtilis* genome in only one week. Additionally, we expanded this technology for the fine-tuning of gene expression by switching a constitutive promoter for a xylose-inducible promoter based on the xylose-repressor system. This system extends the repertoire of molecular toolboxes for genetic manipulations and biotechnological endeavors by enabling simultaneous integration of multiple gene copies, which we anticipate will enhance the development of *B. subtilis* as a platform to produce important enzymes and other commodities.

Results

Genetic manipulation of colony color in *B. subtilis*

In *S. aureus*, the yellow C30 carotenoid 4,4′-diaponeurosporene is synthesized by the 2385-bp *crtMN* operon [36]. It has been reported that plasmid-mediated recombinant expression of this operon in *B. subtilis* results in a colony color change from white to yellow [37]. In this study, we wanted to ascertain whether a strain harboring the *crtMN* operon inserted into its chromosome would also produce yellow colonies, as a preliminary step to implement our colorimetric screening method. For this purpose, the pJOE8999 plasmid, which has been widely used for CRISPR-Cas9-mediated engineering in *B. subtilis* [27], was used to exchange the *spoVG* gene of recipient *B. subtilis* strain (BsMN0) with the *crtMN* operon, setting the expression of the encoded yellow C30 carotenoid under the control of the strong *spoVG* promoter (P_{spoVG}), and flanked by rho-independent transcriptional terminator (T_{spoVG}) downstream of the same gene [38]. The employed editing vector, denoted as pJOE891, contained a sgRNA targeting the *spoVG* gene and a homology repair expression cassette (MN_Ec) composed of the 530-bp upstream flanking genomic region of *spoVG* followed by the *crtMN* operon and the 535-bp downstream flanking genomic region of *spoVG* (Fig. 1). Upon transformation of BsMN0 with pJOE891, we obtained the BsMN1 strain, which satisfactorily produced the yellow pigment, thus turning the *B. subtilis* colonies from white to yellow in color. Next, a set of 2 vectors was constructed, namely pJOE892 and pJOE893. The first one was engineered to replace the extracellular amylase gene of *B. subtilis* (*amyE*) with MN_Ec, rendering the BsMN2 strain inactive for amylase production. The latter was constructed to facilitate the insertion of the MN_Ec into the already inactivated extracellular protease *aprE* gene to obtain BsMN3 strain. Noticeably, both constructs contained the 530 bp upstream arm and the 535-bp downstream arm flanking genomic region of *spoVG*. Next, iterative genome editing was performed by a successive double- and triple-MN_Ec integration to yield strains BsMN4 and BsMN5, harboring two and three copies of MN_Ec integrated at specific sites, respectively. A comprehensive scheme for the construction of each yellow pigment-producing strain is depicted in Fig. 1a. Using specific primers, the identity of each recombinant strain was demonstrated by PCR verification and Sanger sequencing, showing that MN_Ec was successfully inserted into the *B. subtilis* chromosome (Fig. 1b). Irrespective of the *crtMN* operon copy number, all engineered strains formed yellow-pigmented colonies with a highly similar appearance on LB-agar plates, corroborating the successful expression of the yellow C30-carotenoid (Fig. 1c).

Establishment of a white/yellow colorimetric screening method

The basis of the CRISPR-Cas9-mediated white/yellow colorimetric screening method proposed in this work is the generation of double-strand breaks (DSBs) at *crtMN* operon target sites in yellow-pigmented strains and their repair through HDR, thus enabling selection for white colonies because of the removal of the *crtMN* operon, which if still present will lead to the formation of yellow colonies. However, it is crucial that DSBs generated by this system are exclusively repaired through HDR because the NHEJ (non-homologous end joining) system may lead to unintended rearrangements at the target region [39], which may consequently interfere with the screening system. Thus, to explore the

(See figure on next page.)

Fig. 1 Schematic representation of BsMN1-BsMN5 strains construction system. **a** Integration of MN_Ec at *spoVG* locus using plasmid pJOE891, yielding BsMN1. MN_Ec integration at *amyE* locus using plasmid pJOE892, to yield BsMN2. MN_Ec integration at *aprE* locus using plasmid pJOE893, yielding BsMN3. BsMN4 strain contains two MN_Ec copies integrated at *spoVG* and *amyE* locus sites, whereas BsMN5 contains three MN_Ec copies integrated at *spoVG*, *amyE* and *aprE* locus sites. DSB: double-strand breaks. **b** Confirmation of the *spoVG*, *amyE* and *aprE* genes replacement for the MN_Ec in BsMN5 strain. Lanes 1, 3 and 5 corresponds to amplification bands of 4134-bp, 4883-bp and 5124-bp using primers P17F/P17R, P18F/P18R and P19F/P19R to verify MN_Ec integrations at *spoVG*, *amyE* and *aprE* locus sites with gDNA from BsMN5 as template, respectively. Lanes 2, 4 and 6 corresponds to amplifications bands of 2043-bp, 3413-bp and 1830-bp using the same primers with gDNA from BsMN0 as control. M corresponds to the molecular marker weight. **c** LB-agar plates showing white/yellow colonies of strains BsMN0-BsMN5

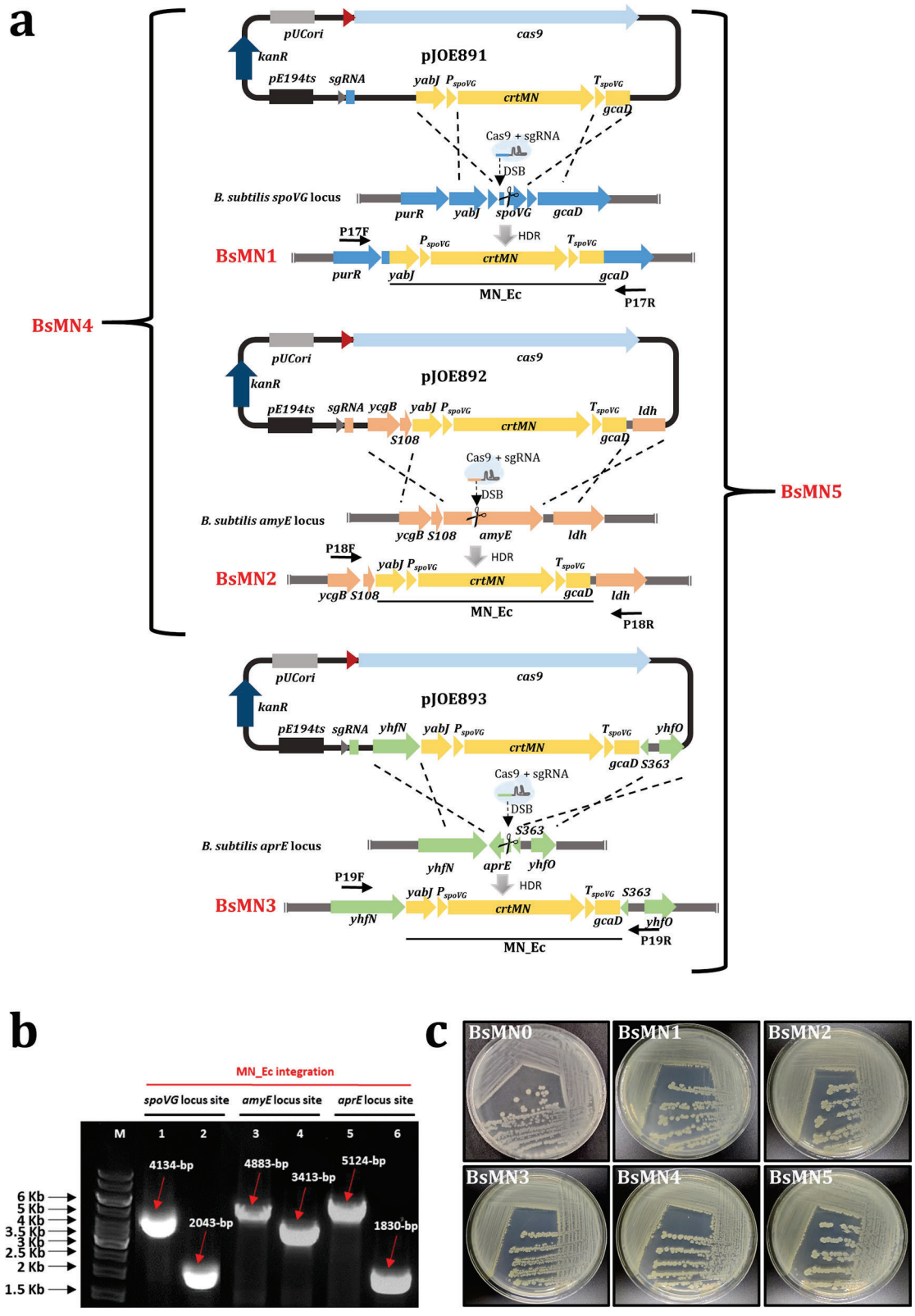


Fig. 1 (See legend on previous page.)

ability of the unwanted NHEJ system to repair DSBs produced at the *crtMN* operon site, a high-scoring sgRNA sequence targeting the *crtMN* operon (5′-ACC AGAAGATCAAAGAAAAGCGG-3′) was identified using the online sgRNA Designer tool provided by the Broad Institute [40]. This target site was determined to be unique following BLASTN analysis against the *B. subtilis* chromosome, and the closest homolog had 5 mismatches, which thereby prevents off-target effects. Plasmid pJOE8999 was engineered to contain this target sequence yielding pJOE894, which was used to transform BsMN1-BsMN3 strains. The lack of colonies obtained upon transformation (Table 1) demonstrated that the introduction of a DSB at the *crtMN* operon is lethal for the cells unless we provide an editing template, ensuring that NHEJ system will not disturb the colorimetric screening method.

Development, evaluation, and validation of novel CRISPR-Cas9 system for simultaneous insertion of multiple gene-copies

To demonstrate the functionality of the CRSIPR-Cas9-based white/yellow screening method, we first constructed plasmid pJOE895. This plasmid was engineered to contain a homology repair expression cassette (Q_Ec) composed of the 530-bp homologous region upstream of *spoVG,* followed by a synthetic 1829-bp DNA fragment coding for amylase gene

Table 1 Summary of efficiency results and CFU for SGI, DGI and TGI using plasmids pJOE894, pJOE895 and pJOE896. Experiments were carried out in triplicates and data are presented as mean values ± standard deviation

Gene integration number	Plasmid	CFU/μg plasmid[a]	No. of Kan^r white clones (% white clones)	% Kan^s white clones (total no. of colonies tested)[b]	Efficiency for Kan^s white clones (%)[c]	Purpose
SGI (*spoVG* locus)	pJOE894[d]	0	0	0	0	NHEJ functionality in *B. subtilis*
	pJOE895[e]	150 ± 21	147 ± 18 (98.2 ± 1.4)	95 (40)	100 (8/8)	*ΔcrtMN* (Q_Ec) knock-in at *spoVG* site
	pJOE896[f]	53 ± 7	49 ± 5 (92.7 ± 2.7)	90 (40)	100 (8/8)	*ΔcrtMN* (Qxyl_Ec) knock-in at *spoVG* site
SGI (*amyE* locus)	pJOE894	0	0	0	0	NHEJ functionality in *B. subtilis*
	pJOE895	147 ± 32	143 ± 30 (97.5 ± 0.6)	92.5 (40)	100 (8/8)	*ΔcrtMN* (Q_Ec) knock-in at *amyE* site
	pJOE896	42 ± 4	38 ± 2 (90 ± 3.8)	87.5 (40)	100 (8/8)	*ΔcrtMN* (Qxyl_Ec) knock-in at *amyE* site
SGI (*aprE* locus)	pJOE894	0	0	0	0	NHEJ functionality in *B. subtilis*
	pJOE895	139 ± 34	137 ± 33 (98.6 ± 0.4)	97.5 (40)	100 (8/8)	*ΔcrtMN* (Q_Ec) knock-in at *aprE* site
	pJOE896	61 ± 11	55 ± 8 (90.6 ± 5.9)	95 (40)	100 (8/8)	*ΔcrtMN* (Qxyl_Ec) knock-in at *aprE* site
DGI (*spoVG* + *amyE*)	pJOE895	138 ± 22	95 ± 15 (68.8 ± 2.8)	92.5 (40)	100 (18/18)	*ΔcrtMN* (Q_Ec) double knock-in at *spoVG* and *amyE* locus site
	pJOE896	38 ± 9	4 ± 3 (10 ± 8.8)	100 (12)	100 (12/12)	*ΔcrtMN* (Qxyl_Ec) triple knock-in at *spoVG* and *amyE* locus site
TGI (*spoVG* + *amyE* + *aprE*)	pJOE895	120 ± 19	6 ± 2 (4.9 ± 1.2)	94.4 (18)	100 (17/17)	*ΔcrtMN* (Q_Ec) triple knock-in at *spoVG*, *amyE* and *aprE* locus site
	pJOE896	35 ± 6	0	0	0	*ΔcrtMN* (Qxyl_Ec) triple knock-in at *spoVG*, *amyE* and *aprE* locus site

[a] Total number of transformants

[b] Clones were tested after three passages at 50 ºC

[c] Clones that showed PCR products out of total number of clones analyzed by colony PCR is depicted in parenthesis

[d] pJOE894 contains no repair template

[e] pJOE895 contains Q_Ec (2894-bp) as a repair template

[f] pJOE896 contains Qxyl_Ec (4476-bp) as a repair template

amyQ from *Bacillus amyloliquefaciens* and the 535-bp homologous region downstream of *spoVG*. We used an α-amylase as a secreted reporter system due to easy quantification of its extracellular production in production media. The rationale behind our procedure is that transformation of yellow-pigmented strains (BsMN1-BsMN5) with pJOE895 will allow the selective cleavage of the chromosome at specific sites with an integrated *crtMN* operon. After chromosome cleavage, homologous recombination with editing template will restore chromosome integrity and replace the *crtMN* operon with Q_Ec, thus leading to the formation of white colonies in successful edits (BsQ1–BsQ5). Conversely, unsuccessful genome edits will keep the *crtMN* operon intact, and the bacterial colonies will thus remain yellow. A scheme of the molecular mechanism proposed for the colorimetric detection of multiple *crtMN* operon replacement for *amyQ* gene is depicted in Fig. 2.

Here, as a proof of principle, we tested the possibility of one-, two- and three-copy *amyQ* gene integration into the *B. subtilis* chromosome in a single step. For this application, BsMN1-BsMN5 strains were transformed with pJOE895 and resulting transformants were selected on LB plates supplemented with kanamycin (to select for the plasmid) at 37 ºC for 24 h. After incubation, the resultant colonies showed a filamentous aspect with irregular borders and without significant differences in their colony color (Fig. 3a–e). These filamentous colonies were immediately streaked on new LB plates with kanamycin and mannose and incubated at 37 ºC for an additional 24 h. The following day, we could readily distinguish between yellow and white bacterial clones in resulting strains (BsQ1–BsQ5) (Fig. 3f–j). Noticeably, some filamentous colonies derived from double- (DGI) and triple-*amyQ* gene integration (TGI) were unable to grow in the new plate. We then patched randomly selected white clones derived from each single *amyQ*-gene integration (SGI), DGI and TGI onto LB plates without antibiotics at 50 ºC to facilitate plasmid curing, which was achieved for most of the white clones tested after three passages at 50 ºC (Table 1). The plasmid-cured cells were checked by colony-PCR using the outer primers from the specific insertion sites to confirm the successful integrations. All white clones tested contained the desired replacements, as confirmed by checking the size of PCR fragments spanning the desired integration and by DNA sequencing (Fig. 3k). Overall, this meant it took around one week to construct a high-copy *B. subtilis* strain harboring three *amyQ* gene copies ready for characterization or another round of genetic manipulation (see Fig. 4). These results validate our procedure to discern between unedited yellow colonies from edited white colonies. Furthermore, the power and convenience of the high-throughput colorimetric screening method was duly proved, demonstrated by its ability for the rapid in vitro screening of large number of transformants and a dramatic decrease in the rate of false positives, allowing the identification of multigene insertion *B. subtilis* strains with 100% positivity among all clones tested (Table 1).

This procedure was also employed to monitor integration efficiency by simply counting colony color upon transformation. Therefore, integration efficiencies were systematically calculated as the ratio of the number of white colonies (correctly edited transformants) to the total number of transformants (white colonies plus yellow colonies). The simultaneous integration efficiency achieved for TGI was very low (4.9% ± 0.8%), significantly lower than efficiency achieved for DGI (68.8 ± 2.8%). In contrast, very high efficiencies that ranged from 97.5 to 98.6% for single *amyQ* gene integration (SGI) were achieved (Table 1). Moreover, CFUs were counted to analyze cell growth with no significant differences observed in the total number of transformants obtained from SGI, DGI, and TGI (Table 1). Gratifyingly, not only was the goal of inserting two- and three-gene copies into the *B. subtilis* chromosome in a single step accomplished for the first time by this novel methodology, but also the simple and fast identification of positive clones with desired genetic modifications.

Expanding the genetic toolbox for promoter switching

A key aspect of the novel editing strategy presented here is that a unique sgRNA along with changeable editing template would be sufficient for simultaneous targeting and insertion of multiple gene copies in *B. subtilis*. Hence, the flexibility of editing template construction opens the possibility of using this system to insert any gene under the control of desired promoters, which are indispensable control elements to accurately regulate expression of target proteins [41]. To begin exploring this application, we aimed to switch the P_{spoVG} controlling *amyQ* gene expression for the xylose-inducible promoter xylose-repressor system (P_{xyl}). For this purpose, an engineered editing template (Qxyl_Ec) composed of 536-bp homologous upstream arm, followed by the 3405-bp *amyQxyl* gene (*amyQ* gene under the control of P_{xyl}) and 535-bp homologous downstream arm, was constructed and inserted into the plasmid pJOE894 to yield pJOE896. We placed a stop codon at position + 6 (relative to *crtM* gene start codon) in the homologous upstream arm to prevent expression from the *spoVG* promoter. Upon transformation of yellow-pigmented strains with pJOE896, we were again able to recover white colonies from SGI and DGI integration experiments (Fig. 5a, b), although with very low efficiency for the double integrations (10% ± 8.8%;

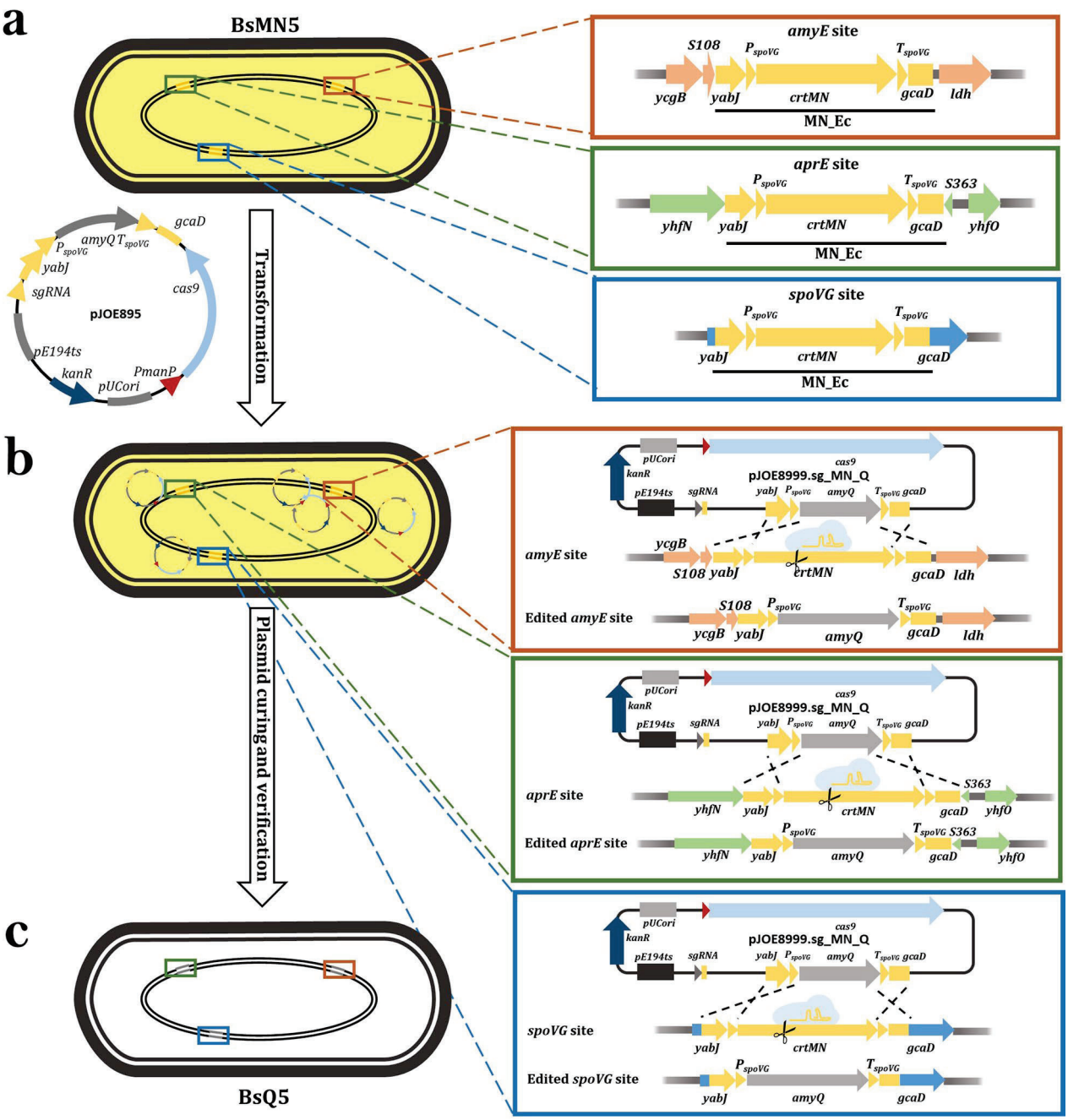


Fig. 2 Outline of the proposed molecular mechanism of *crtMN* operon replacement for *amyQ* gene. **a** First step is to transform yellowish-pigmented BsMN5 strain with pJOE895 editing plasmid. Insets show MN_Ec integrated at *spoVG*, *amyE*, and *aprE* locus sites. **b** Upon transformation, the expression of Cas9 is guided by a specific sgRNA targeting multiple *crtMN* operon, thus generating DSBs at *spoVG*, *amyE* and *aprE* locus sites within BsMN5 chromosome. Insets depict homologous recombination with editing template provided by the plasmid, resulting in the replacement of three *crtMN* operon with three *amyQ* gene-copies, in a single step. Successful edited colonies (BsQ5) will appear as white clones, in contrast to unedited colonies that will remain yellow. **c** The final step is to verify the loss of the plasmid and the identity of the genome modification

Table 1). In contrast, no white colonies for TGI were observed in three independent experiments (Fig. 5c, Table 1). After curing of the plasmid, all recovered white clones tested contained the desired Qxyl_Ec insertion, as verified by PCR amplification and Sanger sequencing (Fig. 5d), thus confirming the successful insertion of two Qxyl_Ec copies into the *B. subtilis* chromosome. Finally, we observed a marked decrease in the number of

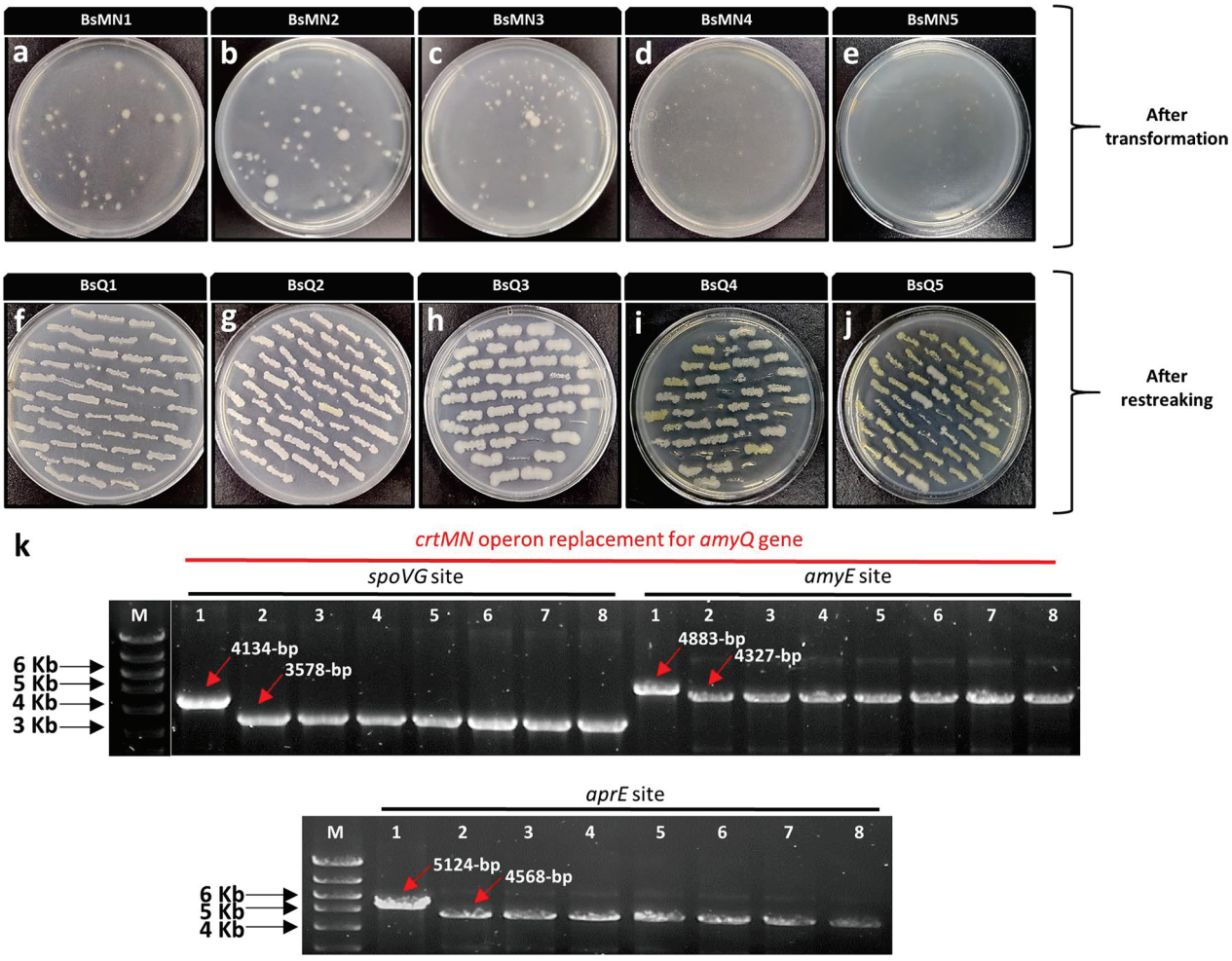


Fig. 3 Transformation of pJOE895 plasmid and BsMN1-BsMN5-targeting plasmid. **a**–**e** Resulting colonies after transformation in LB plates with kanamycin. **f**–**j** Resulting clones after restreaking obtained transformants in LB plates with kanamycin and mannose. **k** PCR verification of randomly selected white clones derived from BsQ5 strain (as shown in panel j). The three *crtMN* operon replacements for *amyQ* gene were confirmed by visualization of correctly sized PCR products using P17F/P17R primers (*spoVG* site), P18F/P18R (*amyE* site) and P19F/P19R (*aprE* site), corresponding to lanes 2 to 8, which showed a reduction in size relative to control product with same primers (lane 1) at different locus sites. M corresponds to the molecular weight marker

CFUs (35 to 61 CFU/µg DNA) in comparison to previous experiments with pJOE895 (120 to 150 CFU/µg DNA, Table 1).

Monitoring amylase AmyQ secretion on liquid cultures

The effect of multiple Q_Ec and Qxyl_Ec copies inserted into the *B. subtilis* chromosome in the release of extracellular amylases was investigated by culturing recombinant strains in amylase production media and measuring the values of α-amylase activity secreted to the media. Strain BsQ2, which had a truncated copy of the *amyE* gene in the chromosome, rendering the strain inactive for amylase production, was used to evaluate the effect of one Q_Ec insertion and compared to BsQ4 and BsQ5 with two and three copies, respectively. Enzymatic assays showed that the higher the copy number of Q_Ec, the greater the values of α-amylase activity secreted to the media, achieving a value of 21.9 ± 2.1 U/ml in BsQ2, which was doubled in BsQ4 (40.2 ± 5.9 U/ml) and almost tripled in BsQ5 (53 ± 2.7 U/ml) (Fig. 6a). Although BsQxyl2 and BsQxyl4 showed low levels of α-amylase activity in production media (4.9 ± 0.9 U/ml and 7.0 ± 1.3 U/ml, respectively), seemingly resulting from basal promoter activity, α-amylase activity was highly induced with xylose: 24.7 ± 2.2 U/ml and 42.5 ± 4.8 U/ml, respectively (Fig. 6a), demonstrating the tight regulation of *amyQ* gene expression achieved through this system. Overall, these results

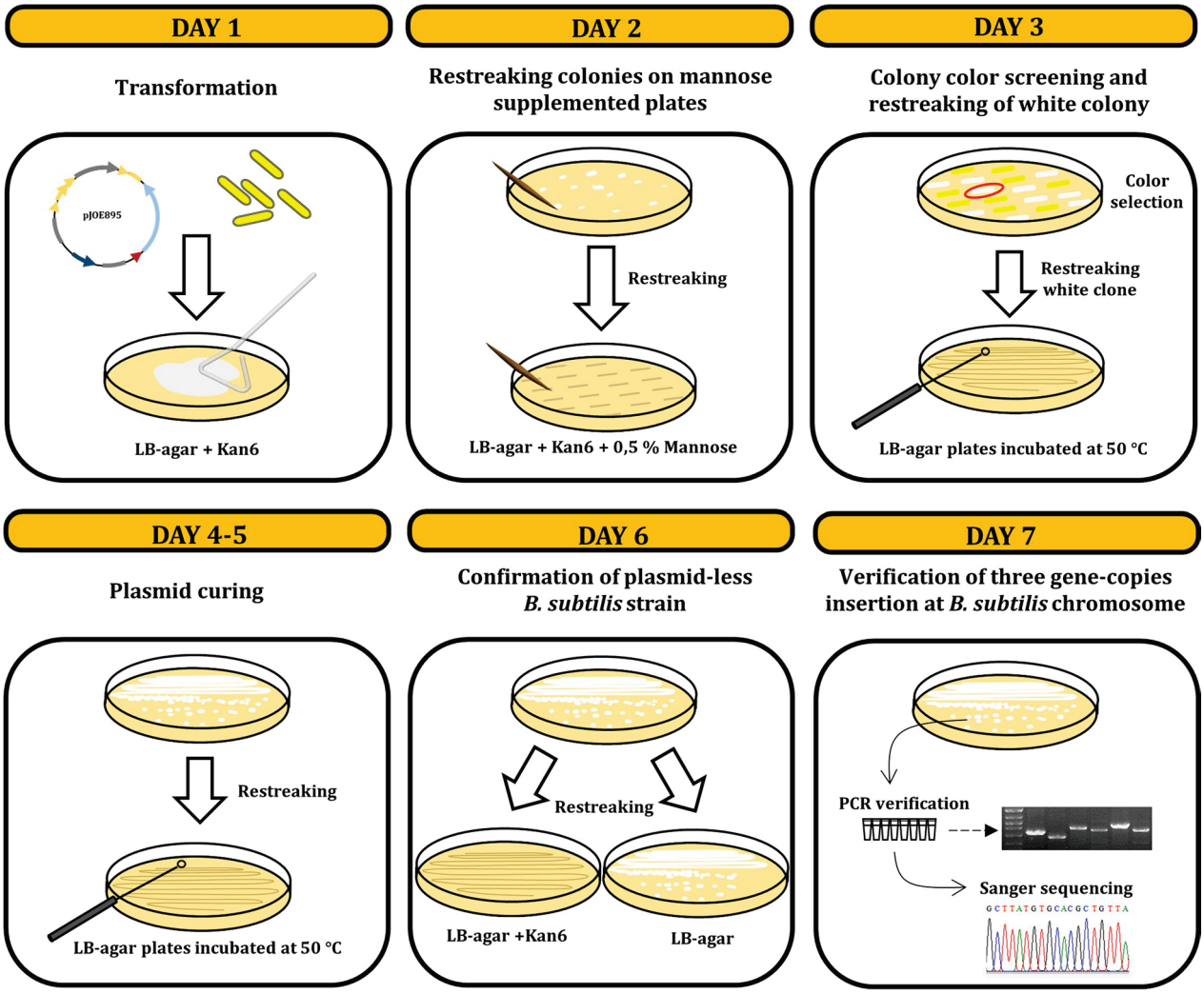


Fig. 4 Schematic diagram of the white/yellow colorimetric screening method for the one-step multiple gene insertion detection in *B. subtilis* using the CRISPR-Cas9 system. Day 1: Transformation of *B. subtilis* with a single plasmid carrying a specific sgRNA targeting *crtMN* operon and desired editing template. Day 2: Transformants were spread on a LB plate containing kanamycin and mannose. Day 3: Resulting white clones were restreaked on LB-agar plates and incubated at 50 ºC. Day 4–5: White clones were cured from the plasmid after three passages at 50 ºC. Day 6: Colonies were replicated on LB plates with and without kanamycin. Day 7: Antibiotic sensitive colonies were then subjected to PCR verification and Sanger sequencing to verify their identities

confirm the dependence of extracellular amylase activity on *amyQ* gene copy-number and promoter type, thus highlighting the importance of multicopy strains for maximizing gene expression. Noticeably, we demonstrated that engineering the editing template could allow gene of interest expression with high precision.

The stability of strains BsQ5 and BsQxyl4 in the production of amylase AmyQ without antibiotic selection was tested. Overnight cultures of both strains in LB were diluted 1:1000 in the same medium. The cells were grown in shaking flasks at 37 °C to stationary phase and diluted again 1000-fold. This was repeated five times and in the last transfer, when the stationary phase was reached, strains were cultured in production media and α-amylase activity was determined. Figure 6b shows that both strains produced similar levels of α-amylase activity for at least 100 generations (every round of growth to stationary phase corresponds to about ten generations without antibiotic supplementation), demonstrating that this technology is a valuable tool for constructing stable producer *B. subtilis* strains.

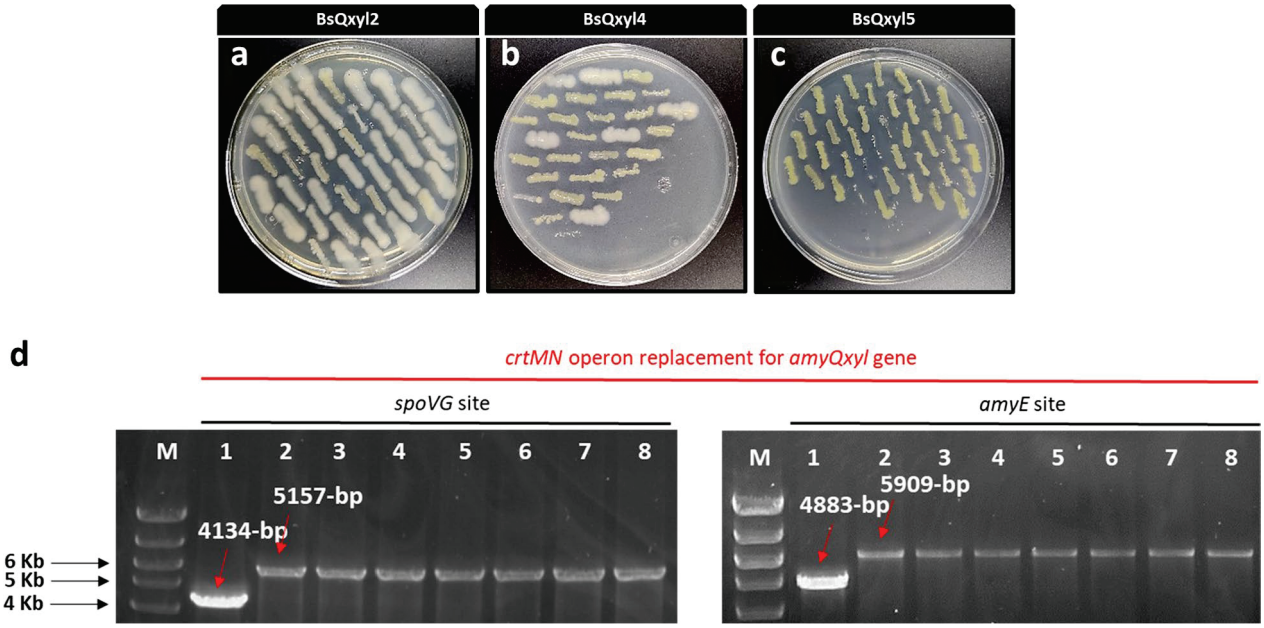


Fig. 5 Transformation of BsMN2, BsMN4 and BsMN5 with pJOE896 plasmid. **a**–**c** Resulting white and yellow clones after transformation and restreaking of the obtained colonies in LB plates with kanamycin and mannose. **d** PCR verification of randomly selected white clones derived from BsQxyl4 strain (as shown in panel b). The two *crtMN* operon replacements for *amyQxyl* gene were confirmed by visualization of correctly sized PCR products using P17F/P17R primers (*spoVG* site) and P18F/P18R (*amyE* site) (lanes 2 to 8), which showed an increase in size relative to control product with same primers (lane 1) at different locus sites. M corresponds to the molecular weight marker

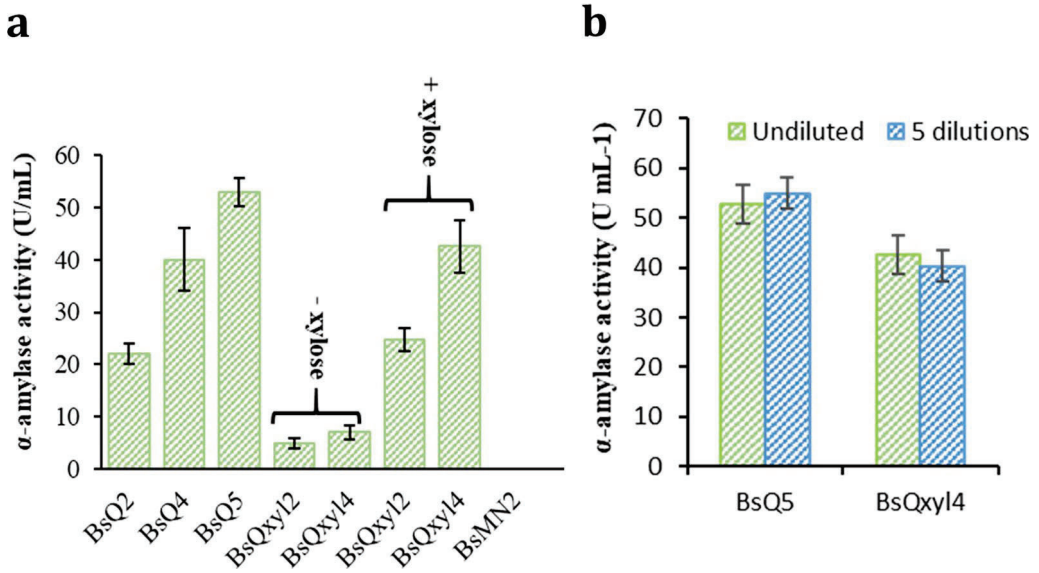


Fig. 6 Detection of α-amylase activity in engineered B. subtilis strains. **a** α-amylase activity derived from supernatants of strains: BsQ2, BsQ4, BsQ5, BsQxyl2 and BsQxyl4. BsMN2 was used as a control strain. Experiments without (-xylose) and with xylose (+xylose) are indicated. **b** α-amylase activity in supernatants from BsQ5 and BsQxyl4 and from the same strains diluted 1000-fold and grown to stationary phase repeatedly 5 times without antibiotics in LB media. The error bars represent the average ± standard deviation of three biological replicates

Discussion

Various high efficiency, simultaneous multilocus integration methods based on two-plasmid CRISPR-Cas9 systems have been implemented in *E. coli* [42, 43], yeasts [44, 45], and fungi [46, 47], making them suitable host cells to be used as a chassis in the field of synthetic biology. However, although the emergence of the CRISPR/Cas9 system has led to numerous new applications and developments in *B. subtilis* [29, 30, 48–51], efficient methods for simultaneous gene insertion in *B. subtilis* have not yet been reported. The increase of gene copy number and its integration into multiple neutral sites is the most promising way to achieve high yield and stable productivity, as a preliminary step for industrial strain development [16–18, 52, 53]. However, although this strategy is a valuable approach for the construction of multicopy *B. subtilis* strains, it relies on iterative rounds of plasmid construction for sequential insertions of gene copies into the *B. subtilis* chromosome, which is tedious and time consuming. This prompted us to investigate new methods to facilitate multigene insertions in a single step, therefore bypassing the big efforts needed for multistep construction of plasmids and time-consuming curing of the plasmids in iterative genome editing.

The single-plasmid-based CRISPR-Cas9 system established in this study is preferred over genome engineering methods based on two-plasmid systems because of both the higher plasmid transformation efficiencies achieved and a marked decrease in the burden of the host strain [54]. This technology boasts two unique features that facilitate the simultaneous insertion of multiple gene copies into the *B. subtilis* chromosome. First, an invariable sgRNA was selected to reduce the off-target effects of endonuclease Cas9 and to ensure high specificity for the *crtMN* operon. Second, a changeable editing template containing the gene of interest flanked with homologous arms was demonstrated to be sufficient to replace the *crtMN* operon with the desired gene and allow white/yellow screening for successful insertions.

Although HDR efficiency using a single-plasmid system is greatly dependent on the length of the homologous repair template [51], our results corroborated that 530-bp homology arms were enough to allow multicopy integration with high efficiency. Additionally, we observed that the natural NHEJ system present in *B. subtilis* is too weak to repair DSBs in *B. subtilis*, which is in accordance with previous reports [51, 55]. Hence, as a proof of principle, the novel genome-editing method has been proven to be efficient for the simultaneous insertion of three 2894-bp, and two 4476-bp fragments containing Q_Ec and Qxyl_Ec, respectively, with Qxyl_Ec being the largest fragment reported to be integrated into the *B. subtilis* genome using the CRISPR-Cas9 system, so far.

The simultaneous insertion of three Qxyl_Ec copies into the *B. subtilis* genome was unsuccessful. We propose that the reason for failure might be associated with the size of the editing template. Qxyl_Ec is 4476-bp, which may be too large to efficiently initiate the double exchange, thus significantly lowering the probability for the insertion of more than one gene at a time. Therefore, for further practical application using this system, the size of the editing template should be carefully considered for expression cassettes longer than 3-kb.

Despite the very low efficiencies achieved for DGI and TGI, the colorimetric screening method adopted in this work represents a good solution to tackle the issue of low editing efficiencies achieved in multiplex genome editing, allowing the straightforward identification of positive clones with the desired genetic changes. Remarkably, this procedure resulted in saving time and cost and allowed the rapid and convenient construction of high-level expression of multicopy genes encoding secreted α-amylase in *B. subtilis* in only seven days.

Multilocus integration experiments resulted in the isolation of recombinant colonies showing a filamentous aspect, forming colonies with irregular borders and different sizes. The presence of this colony type may be related to the SOS response in *B. subtilis*. We suspect that the bacteria were edited during growth, thus causing some damage to the *B. subtilis* DNA which, as previously reported, triggers a physiological response called the SOS response [56], thus inhibiting cell division and causing bacteria to appear filamentous [33, 57–59]. Although most of the colonies grew normally after being subcultured on mannose plates, the growth of a few colonies was impaired, which irremediably led to cell death. We suspect that this phenomenon might be due to their inability to repair damaged DNA, which is supported by a similar observation recently reported by the group of Guo [33].

The versatility and wide range of applicability of the novel technology is considerable and was demonstrated in this study by engineering editing template to allow the replacement of constitutive P_{spoVG} for xylose-inducible P_{xyl}. As one of the main difficulties in genetic engineering and synthetic biology is how to control the expression of a certain protein at a given level, our system provides an efficient, and facile approach for achieving desired production goals by selecting a suitable promoter element to accurately regulate expression of target proteins. For each application, the only requisite is to engineer the editing template to contain the desired cargo (gene of interest with or without modifications) flanked by 530-bp homologous arms to enable its multiple insertion under the control of desired promoter. In the case that higher levels of protein

production are required, the use of dual and multiple tandem promoters has been shown to enhance productivity in *B. subtilis* [60–62], which can be also achieved by rational design of the editing template to contain the synthetic promoter. Conversely, if temporal control of gene expression is needed, the use of inducible promoters is of great interest, especially for the fine induction of difficult to express or toxic proteins [63]. Conclusively, editing template engineering strategies will allow fine-tuning of multicopy gene expression, which was successfully evaluated by means of a model protein, amylase AmyQ from *B. amyloliquefaciens* in this study. While the strategy presented here has been adopted for *B. subtilis*, we hope that this methodology will be applicable not only to other *Bacillus* species but to other relevant bacterial strains used in industrial applications, such as *Pseudomonas putida* and *Corynebacterium glutamicum*, among others [64]. Strains and plasmids developed in this work will be publicly available to the scientific community through the Bacillus Genetic Stock Center (https://bgsc.org/).

Conclusions

Efficient multigene integration methods have not been developed in the model gram-positive *B. subtilis*, which is considered one of the dominant bacterial workhorses in microbial fermentation. For the first time, a simple, rapid, and convenient approach to facilitate the simultaneous insertion of up to three gene copies in *B. subtilis* was established through the development of a high-throughput colorimetric screening method combined with the use of a single CRISPR-Cas9-based plasmid carrying a unique high-efficiency target site and a changeable editing template with 530-bp homologous arms. Additionally, we demonstrated that the system could be adapted to modulate the promoter controlling the expression of the gene of interest by rational design of the editing template. This novel technology can be potentially used as a routine method for the construction of marker-free, plasmid-less, high-expression stable producer strains in a timely fashion, which provides promising prognoses for future developments of *B. subtilis* as a microbial cell factory in industrial settings.

Methods

Strains, media and growth conditions

Escherichia coli strain DH5α was used as the host strain for routine molecular cloning and plasmid construction operations. For transformation of *B. subtilis*, plasmid DNA was isolated from the rec^+ strain *E. coli* Turbo (New England Biolabs). *B. subtilis* strain BsMN0, an asporogenous strain with seven protease genes inactivated, was purchased from the Bacillus Genetic Stock Center (BGSC, Ohio) and served as the recipient strain for the genome editing experiments. Chemically competent *E. coli* cells and transformation protocol were performed as described previously [65]. *B. subtilis* strains were transformed according to the method previously described [66]. The strains involved in this study are listed in Table 2. Strains were propagated in Lysogeny Broth (LB) medium (10 g/L tryptone, 5 g/L yeast

Table 2 Strains used in this study

Strain	Genotype or description	Source/Reference
E. coli DH5α	*fhuA2 lac(del)U169 phoA glnV44 Φ80′ lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Laboratory stock
E. coli NEB® turbo	*F′ proA + B + lacIq ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)*TetS *endA1 thi-1 Δ(hsdS-mcrB)5*	Laboratory stock
B. subtilis strains		
BsMN0	*ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr ΔsigF*	Laboratory stock
BsMN1	BsMN0 with *ΔspoVG* (MN_Ec) knock-in mutant	This study
BsMN2	BsMN0 with *ΔamyE* (MN_Ec) knock-in mutant	This study
BsMN3	BsMN0 with *ΔaprE* (MN_Ec) knock-in mutant	This study
BsMN4	BsMN0 with *ΔspoVG ΔamyE* (MN_Ec) double knock-in mutant	This study
BsMN5	BsMN0 with *ΔspoVG ΔamyE ΔaprE* (MN_Ec) triple knock-in mutant	This study
BsQ1	BsMN1 with ΔMN_Ec (Q_Ec) knock-in mutant	This study
BsQ2	BsMN2 with ΔMN_Ec (Q_Ec) knock-in mutant	This study
BsQ3	BsMN3 with *ΔMN_Ec* (Q_Ec) knock-in mutant	This study
BsQ4	BsMN4 with ΔMN_Ec (Q_Ec) double knock-in mutant	This study
BsQ5	BsMN5 with ΔMN_Ec (Q_Ec) triple knock-in mutant	This study
BsQxyl2	BsMN2 with ΔMN_Ec (Qxyl_Ec) knock-in mutant	This study
BsQxyl4	BsMN4 with ΔMN_Ec (Qxyl_Ec) double knock-in mutant	This study

extract, and 10 g/L NaCl) and on LB agar (15 g/L agar) plates at 37 °C. The plasmid used for the expression of the CRISPR–Cas9 constructs is based on vector pJOE8999 [27]. Kanamycin was used for screening in *E. coli* and *B. subtilis* at final concentrations of 30 µg/mL and 6 µg/mL, respectively. To induce the CRISPR-Cas9 system in *B. subtilis* cells, 0.5% D-mannose was added.

Plasmid construction

The plasmids and primers used in this study are listed in Additional file 2: Tables S1 and S2, respectively. All plasmids construction were performed using the pJOE8999 plasmid as the parental plasmid and required two consecutive steps: (i) cloning specific sgRNA; and (ii) cloning specific editing template.

Cloning of sgRNA

For sgRNA construction targeting each specific gene, two complementary oligonucleotides were ordered (Macrogen, Korea) with the respective overhangs, annealed and cloned into the vector pJOE8999. In brief, both complementary oligonucleotides were mixed at a final concentration of 10 µM in annealing buffer (10 × stock contained 100 mM Tris-HCl pH 7.5, 1 M NaCl and 1 mM Ethylenediaminetetraacetic acid (EDTA) (pH 8), kept at 98 °C for 5 min and slowly cooled to room temperature. Then, the annealed oligonucleotides were treated with polynucleotide kinase to phosphorylate the 5' ends, according to manufacturer's instructions (Invitrogen), and ligated to *BsaI* cleaved and dephosphorylated plasmid pJOE8999 to incorporate specific target sequences to the vector.

Cloning of editing template

In a second step, to construct editing templates, two homologous arms of similar length and the desired cargo to be inserted were separately amplified and were then fused together by splicing with overlap extension PCR (SOEing-PCR, Additional file 2: Table S3). Regarding the desired cargo, the *amyQ* gene was synthesized by Nzytech (Portugal) with codon optimization for *B. subtilis* (Additional file 1: Fig. S1), and cloned into pBS2EXxylRPxylA vector by using restriction sites *XbaI* and *PstI*, resulting in the plasmid pBS2EXxylRPxylA_Q. This plasmid was used as a template for *amyQ* and *amyQxyl* gene amplifications, whereas pHY_crtMN vector provided by the Maeda's group [37], was used as a template for *crtMN* operon amplification. SOEing-PCR products of the expected size along with pJOE8999 plasmids harboring specific target sequence were digested with *SfiI* and ligated with T4 DNA-ligase (Nzytech) to incorporate the desired cargo to the corresponding editing plasmid. All plasmids were verified by Sanger sequencing. A more detailed explanation for the construction of each editing plasmid is given in the supplementary material (Additional file 1: Fig. S2–S6).

Construction of multicopy strains and plasmid curing

After transformation of *B. subtilis* strains with editing plasmids, resulting colonies obtained on LB agar plates with kanamycin were streaked into LB agar plates supplemented with kanamycin and 0.5% of D-mannose and incubated at 37 °C for 24 h. After incubation, white colonies were selected among yellow colonies and passaged three times on LB agar plates (without any antibiotics) at 50 °C for 24 h to cure the plasmid. Plasmid curing achieved best results when cells are streaked for single colonies at each passage. The colonies cured of editing plasmid were confirmed by streaking them onto LB agar plates containing kanamycin or no antibiotics; colonies cured of plasmid fail to grow at 37 °C. The presence of multiple gene copies was verified by colony PCR using relevant primers and Sanger sequencing (Additional file 2: Table S3).

Quantification of α-amylase activity in liquid cultures

Overnight cultures of recombinant *B. subtilis* strains in production medium (12 g/L sucrose, 18 g/L peptone, 2 g/L $(NH_4)_2SO_4$, 18.3 g/L $K_2HPO_4{\cdot}3H_2O$, 6 g/L KH_2PO_4, 1 g/L Na^+ citrate, 0.2 g/L $MgSO_4{\cdot}7H_2O$, 0.12 g/L $FeSO_4{\cdot}7H_2O$, 30 mg/L $MnSO_4{\cdot}H_2O$, 12 mg/L $CuSO_4{\cdot}H_2O$ and 12 mg/L $ZnCl_2$) were diluted to 0.1 OD_{600} in 25 mL of production media and were grown at 37 °C and 220 rpm for 2 days. In the case of xylose-inducing experiments, 1% xylose was added at OD_{600} between 0.6 and 0.8 and strains were then cultured for additional 48 h at same conditions. After 48 h, the culture supernatants were obtained by centrifugation at 8000 g for 20 min at 4 °C and were used as crude α-amylase samples. The standard assay mixture contained 1.5% soluble starch in a final volume of 0.25 ml of 50 mM Tris–HCl buffer at pH 6.5. The mixture was assayed with 2 µl of the α-amylase solution and incubated at 80 °C for 2 min. Next, the reaction was stopped by the addition of 0.75 mL of 3,5- dinitrosalicylic acid (DNS) reagent then heated for 5 min in boiling water bath and cooled on ice. The absorbance was read at 540 nm and compared to a standard calibration curve for maltose (Additional file 1: Fig. S7). One unit of amylase activity was defined as the amount of enzyme which liberated 1 µmol of maltose from soluble starch per minute under the assay conditions [67].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12934-023-02032-2.

Additional file 1: Figure S1. *amyQ* gene sequence with codon optimization for *B. subtilis*. **Figure S2.** Construction system of vector pJOE891. **Figure S3.** Construction system of vector pJOE892. **Figure S4.** Construction system of vector pJOE893. **Figure S5.** Construction system of vector pJOE894 and pJOE895. **Figure S6.** Construction system of vector pJOE896. **Figure S7.** Standard calibration curve for maltose.

Additional file 2: Table S1. Plasmids used in this study. **Table S2.** Primers designed in this study. **Table S3.** Splicing with overlap extension PCR (SOEing-PCR) program.

Acknowledgements
The authors thank Dr. Isamu Maeda for providing us with the plasmid pHY_crtMN.

Author contributions
Conceptualization, PP and DRZ; methodology, JF and OF; validation, JF and OF; investigation, JF and OF; resources, PP; writing-original draft preparation, PP; writing-review and editing, PP and DRZ; supervision, PP and DRZ. All authors read and approved the final manuscript.

Funding
This work was supported by the Pla de Doctorats Industrials del Departament de Recerca i Universitats de la Generalitat de Catalunya with the support of Gestió d'Ajuts Universitaris de Recerca for grant number 2021 DI 77 awarded to J.F.N.

Availability of data and materials
All data supporting the conclusions of this study are included within the article and its additional files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no financial and non-financial interests.

Author details
[1]Microbiology Section, Department of Biology, Healthcare and Environment, Faculty of Pharmacy and Food Sciences, Universitat de Barcelona, Barcelona, Catalonia, Spain. [2]Subtilogic LLC, Columbus, OH, USA.

Received: 4 January 2023 Accepted: 25 January 2023
Published online: 31 January 2023

References

1. Earl AM, Losick R, Kolter R. Ecology and genomics of *Bacillus subtilis*. Trends Microbiol. 2008;16:269–75.
2. Schallmey M, Singh A, Ward OP. Developments in the use of *Bacillus* species for industrial production. Can J Microbiol. 2004;50:1–17.
3. Outtrup H, Jorgensen ST. The importance of *Bacillus* species in the production of industrial enzymes. In: Applications and systematics of *Bacillus* and relatives. Oxford, UK: Blackwell Science Ltd.; 2008. p. 206–18.
4. Dijl JMV, Hecker M. *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. Microb Cell Fact. 2013;12(1):3–8.
5. Liu DY, Mao ZT, Guo JX, Wei LY, Ma HW, Tang YJ, et al. Construction, model-based analysis, and characterization of a promoter library for fine-tuned gene expression in *Bacillus subtilis*. ACS Synth Biol. 2018;7(7):1785–97.
6. Danilova I, Sharipova M. The practical potential of *Bacilli* and their enzymes for industrial production. Front Microbiol. 2020;11:1782–8.
7. Su Y, Liu C, Fang H, Zhang D. *Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine. Microb Cell Fact. 2020;19(1):173–84.
8. Young M. Gene amplification in *Bacillus subtilis*. J Gen Microbiol. 1984;130:1613–21.
9. Schumann W. Production of recombinant proteins in *Bacillus subtilis*. Adv Appl Microbiol. 2007;62:137–89.
10. Piggot PJ, Curtiss CAM. Analysis of the regulation of gene expression during *Bacillus subitlis* sporulation by manipulation of the copy number of *spo-lacZ* fusions. J Bacteriol. 1987;1987169:1260–6.
11. Vazquez-Cruz C, Ochoa-Sanchez J, Olmedo-Alvarez G. Pulse-field gel-electrophoretic analysis of the amplification and copy-number stability of an integrational plasmid in *Bacillus subtilis*. Appl Microbiol Biotechnol. 1996;46(1):55–60.
12. Jannière L, Niaudet B, Pierre E, Ehrlich SD. Stable gene amplification in the chromosome of *Bacillus subtilis*. Gene. 1985;40(1):47–55.
13. Leenhouts KJ, Kok J, Venema G. Stability of integrated plasmids in the chromosome of *Lactococcus lactis*. Appl Environ Microbiol. 1990;56(9):2726–35.
14. van der Laan JC, Gerritse G, Mulleners LJ, van der Hoek RA, Quax WJ. Cloning, characterization, and multiple chromosomal integration of a *Bacillus* alkaline protease gene. Appl Environ Microbiol. 1991;57:901–9.
15. Petit MA, Joliff G, Mesas JM, Klier A, Rapoport G, Ehrlich SD. Hypersecretion of a cellulase from clostridium thermocellum in *Bacillus subtilis* by induction of chromosomal DNA amplification. Biotechnology (NY). 1990;8:559–63.
16. Yomantas YA, Abalakina EG, Golubeva LI, Gorbacheva LY, Mashko SV. Overproduction of *Bacillus amyloliquefaciens* extracellular glutamyl-endopeptidase as a result of ectopic multi-copy insertion of an efficiently expressed mpr gene into the *Bacillus subtilis* chromosome. Microb Cell Fact. 2011;10:64–73.
17. Huang K, Zhang T, Jiang B, Yan X, Mu W, Miao M. Overproduction of *Rummeliibacillus pycnus* arginase with multi-copy insertion of the *argR-pyc* cassette into the Bacillus subtilis chromosome. Appl Microbiol Biotechnol. 2017;101:6039–48.
18. Wang J-J, Rojanatavorn K, Shih JCH. Increased production of *Bacillus* keratinase by chromosomal integration of multiple copies of the *kerA* gene. Biotechnol Bioeng. 2004;87:459–64.
19. Zhang XZ, Yan X, Cui ZL, Hong Q, Li SP. *mazF*, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. Nucleic Acids Res. 2006;34(9): e71.
20. Brans A, Filee P, Chevigne A, Claessens A, Joris B. New integrative method to generate *Bacillus subtilis* recombinant strains free of selection markers. Appl Environ Microbiol. 2004;70(12):7241–50.
21. Defoor E, Kryger MB, Martinussen J. The orotate transporter encoded by *oroP* from *Lactococcus lactis* is required for orotate utilization and has utility as a food-grade selectable marker. Microbiology. 2007;153:3645–59.
22. Fabret C, Ehrlich SD, Noirot P. A new mutation delivery system for genome-scale approaches in *Bacillus subtilis*. Mol Microbiol. 2002;46(1):25–36.
23. Shi T, Wang G, Wang Z, Fu J, Chen T, Zhao X. Establishment of a markerless mutation delivery system in *Bacillus subtilis* stimulated by a double-strand break in the chromosome. PLoS ONE. 2013;8(11):e81370.
24. Yan X, Yu HJ, Hong Q, Li SP. Cre/lox system and PCR-based genome engineering in *Bacillus subtilis*. Appl Environ Microbiol. 2008;74(17):5556–62.
25. Chen PT, Jeifu S, Chao YP, Ho T, Yu SM. Construction of chromosomally located T7 expression system for production of heterologous secreted proteins in *Bacillus subtilis*. J Agric Food Chem. 2010;58(9):5392–9.
26. Wang Y, Weng J, Waseem R, Yin X, Zhang R, Shen Q. *Bacillus subtilis* genome editing using ssDNA with short homology regions. Nucleic Acids Res. 2012;40(12):e91.
27. Altenbuchner J. Editing of the *Bacillus subtilis* genome by the CRISPRCas9 system. Appl Environ Microbiol. 2016;82:5421–7.
28. Burby PE, Simmons LA. CRISPR/Cas9 editing of the bacillus subtilis genome. Bio Protoc. 2017;7(8):e2272.

29. So Y, Park SY, Park EH, Park SH, Kim EJ, Pan JG, et al. A highly efficient CRISPR-Cas9-mediated large genomic deletion in *Bacillus subtilis*. Front Microbiol. 2017;8:1167–78.
30. Westbrook AW, Moo-Young M, Chou CP. Development of a CRISPR-Cas9 tool kit for comprehensive engineering of *Bacillus subtilis*. Appl Environ Microbiol. 2016;82:4876–95.
31. Lu Z, Yang S, Yuan X, Shi Y, Ouyang L, Jiang S, et al. CRISPRassisted multidimensional regulation for fine-tuning gene expression in *Bacillus subtilis*. Nucleic Acids Res. 2019;47(7):e40.
32. Price MA, Cruz R, Baxter S, Escalettes F, Rosser SJ. CRISPR-Cas9 In Situ engineering of subtilisin E in *Bacillus subtilis*. PLoS ONE. 2019;14(1):e0210121.
33. Tian J, Xing B, Li M, Xu C, Huo YX, Guo S. Efficient large-scale and scarless genome engineering enables the construction and screening of *Bacillus subtilis* biofuel overproducers. Int J Mol Sci. 2022;23(9):4853–71.
34. Liu D, Huang C, Guo J, Zhang P, Chen T, Wang Z, et al. Development and characterization of a CRISPR/Cas9n-based multiplex genome editing system for *Bacillus subtilis*. Biotechnol Biofuels. 2019;12:197–213.
35. Wu Y, Liu Y, Lv X, Li J, Du G, Liu L. CAMERS-B: CRISPR/Cpf1 assisted multiple-genes editing and regulation system for *Bacillus subtilis*. Biotechnol Bioeng. 2020;117(6):1817–25.
36. Pelz A, Wieland KP, Putzbach K, Hentschel P, Albert K, Götz F. Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*. J Biol Chem. 2005;280:32493–8.
37. Yoshida K, Ueda S, Maeda I. Carotenoid production in *Bacillus subtilis* achieved by metabolic engineering. Biotechnol Lett. 2009;1:1789–93.
38. Banner CD, Moran CP Jr, Losick R. Deletion analysis of a complex promoter for a developmentally regulated gene from *Bacillus subtilis*. J Mol Biol. 1983;168(2):351–65.
39. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, et al. Identification of a DNA nonhomologous end-joining complex in bacteria. Science. 2002;297(5587):1686–9.
40. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34:184–91.
41. Fu G, Yue J, Li D, Li Y, Lee SY, Zhang D. An operator-based expression toolkit for *Bacillus subtilis* enables fine-tuning of gene expression and biosynthetic pathway regulation. Proc Natl Acad Sci. 2022;119(11):e2119980119.
42. Zhu X, Zhao D, Qiu H, Fan F, Man S, Bi C, et al. The CRISPR/Cas9-facilitated multiplex pathway optimization (CFPO) technique and its application to improve the *Escherichia coli* xylose utilization pathway. Metab Eng. 2017;43:37–45.
43. Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. Appl Environ Microbiol. 2015;81(7):2506–14.
44. Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, et al. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR–Cas. Cell Syst. 2015;1:88–96.
45. Wang L, Deng A, Zhang Y, Liu S, Liang Y, Bai H, et al. Efficient CRISPR–Cas9 mediated multiplex genome editing in yeasts. Biotechnol Biofuels. 2018;11:277–92.
46. Katayama T, Nakamura H, Zhang Y, Pascal A, Fujii W, Maruyama JI. Forced recycling of an AMA1-based genome-editing plasmid allows for efficient multiple gene deletion/integration in the industrial filamentous fungus *Aspergillus oryzae*. Appl Environ Microbiol. 2019;85(3):e01896.
47. Pohl C, Kiel J, Driessen AJM, Bovenberg RAL, Nygard Y. CRISPR/Cas9 based genome editing of *Penicillium chrysogenum*. ACS Synth Biol. 2016;5:754–64.
48. Song Y, He S, Abdallah II, Jopkiewicz A, Setroikromo R, van Merkerk R, et al. Engineering of multiple modules to improve amorphadiene production in *Bacillus Subtilis* using CRISPR-Cas9. J Agric Food Chem. 2021;69:4785–94.
49. Wu G, Drufva E, Wu K. Fast genome editing in *Bacillus subtilis*. Eng Life Sci. 2019;19(6):471–7.
50. Sachla AJ, Alfonso AJ, Helmann JD. A simplified method for CRISPR-Cas9 engineering of *Bacillus subtilis*. Microbiol Spectr. 2021;9(2):e0075421.
51. Toymentseva AA, Altenbuchner J. New CRISPR-Cas9 vectors for genetic modifications of *Bacillus* species. FEMS Microbiol Lett. 2019;366(1):1–8.
52. Watzlawick H, Altenbuchner J. Multiple integration of the gene *ganA* into the *Bacillus subtilis* chromosome for enhanced β-galactosidase production using the CRISPR/Cas9 system. AMB Express. 2019;9(1):158–68.
53. Widner B, Thomas M, Sternberg D, Lammon D, Behr R, Sloma A. Development of marker-free strains of *Bacillus subtilis* capable of secreting high levels of industrial enzymes. J Ind Microbiol Biotech. 2000;25:204–12.
54. Hong KQ, Liu DY, Chen T, Wang Z-W. Recent advances in Crispr/Cas9 mediated genome editing in *Bacillus subtilis*. World J Microbiol Biotechnol. 2018;34:153–61.
55. Zheng X, Li SY, Zhao GP, Wang J. An efficient system for deletion of large DNA fragments in *Escherichia coli* Via introduction of both Cas9 and the non-homologous end joining system from *Mycobacterium smegmatis*. Biochem Biophys Res Commun. 2017;485:768–74.
56. Au N, Kuester-Schoeck E, Mandava V, Bothwell LE, Canny SP, Chachu K, Colavito SA, Fuller SN, Groban ES, Hensley LA, O'Brien TC, Shah A, Tierney JT, Tomm LL, O'Gara TM, Goranov AI, Grossman AD, Lovett CM. Genetic composition of the *Bacillus subtilis* SOS system. J Bacteriol. 2005;187(22):7655–66.
57. Kawai Y, Moriya S, Ogasawara N. Identification of a protein, Ynea, responsible for cell division suppression during the SOS response in *Bacillus subtilis*. Mol Microbiol. 2003;47:1113–22.
58. Claessen D, Emmins R, Hamoen LW, Daniel RA, Errington J, Edwards DH. Control of the cell elongation-division cycle by shuttling of Pbp1 protein in *Bacillus subtilis*. Mol Microbiol. 2008;68:1029–46.
59. Dion MF, Kapoor M, Sun Y, Wilson S, Ryan J, Vigouroux A, et al. *Bacillus Subtilis* cell diameter is determined by the opposing actions of two distinct cell wall synthetic systems. Nat Microbiol. 2019;4:1294–305.
60. Song Y, Nikoloff JM, Fu G, Chen J, Li Q, Xie N, et al. Promoter screening from *Bacillus subtilis* in various conditions hunting for synthetic biology and industrial applications. PLoS ONE. 2016;11(7):e0158447.
61. Liu X, Wang H, Wang B, Pan L. Efficient production of extracellular pullulanase in *Bacillus subtilis* ATCC6051 using the host strain construction and promoter optimization expression system. Microb Cell Fact. 2018;17(1):163–74.
62. Liu Z, Zheng W, Ge C, Cui W, Zhou L, Zhou Z. High-level extracellular production of recombinant nattokinase in *Bacillus subtilis* WB800 by multiple tandem promoters. BMC Microbiol. 2019;19(1):89–102.
63. Saccardo P, Corchero JL, Ferrer-Miralles N. Tools to cope with difficult-to-express proteins. Appl Microbiol Biotechnol. 2016;100:4347–55.
64. Regassa H, Bose D, Mukherjee A. Review of microorganisms and their enzymatic products for industrial bioprocesses. Ind Biotechnol. 2021;17:214–26.
65. Chang AY, Chau VWY, Landas JA, Pang Y. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. JEMI methods. 2017;1:22–5.
66. Yasbin RE, Wilson GA, Young FE. Transformation and transfection in lysogenic strains of *Bacillus subtilis*: evidence for selective induction of prophage in competent cells. J Bacteriol. 1975;121:296–304.
67. Yang H, Liu L, Li J, Du G, Chen J. Heterologous expression, biochemical characterization, and overproduction of alkaline α-amylase from *Bacillus alcalophilus* in *Bacillus subtilis*. Microb Cell Fact. 2011;10:77–85.

Publisher's Note

Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system

Jordi Ferrando, Oriana Filluelo, Daniel R Zeigler and Pere Picart

SUPPLEMENTARY MATERIAL

Additional file 1

Figure S1: *amyQ* gene sequence with codon optimization for *B. subtilis*

Figure S2: Construction system of vector pJOE891

Figure S3: Construction system of vector pJOE892

Figure S4: Construction system of vector pJOE893

Figure S5: Construction system of vector pJOE894 and pJOE895

Figure S6: Construction system of vector pJOE896

Figure S7: Standard calibration curve for maltose

TAAGAAAG**GGAGG**ACAAACATGATTCAAAAACGAAAGCGGACAGTTTCGTTCAGACTTGTGCTTATGTGCACGCTGTTATTTG
TCAGTTTGCCGATTACAAAAACATCAGCCGTTAACGGCACACTTATGCAATACTTCGAATGGTACACACCTAACGATGGCCAA
CATTGGAAACGTCTTCAAAACGATGCTGAACATCTTTCTGATATCGGCATCACAGCTGTTTGGATACCTCCTGCTTACAAAGG
CCTTTCTCAATCTGATAACGGCTACGGCCCTTACGATCTTTACGATCTTGGCGAATTCCAACAAAAAGGCACAGTTCGTACAA
AATACGGCACAAAATCTGAACTTCAAGATGCTATCGGCTCTCTTCATTCTCGTAACGTTCAAGTTTACGGCGATGTTGTTCTT
AACCATAAAGCTGGCGCTGATGCTACAGAAGATGTTACAGCTGTTGAAGTTAACCCTGCTAACCGTAACCAAGAAACATCTGA
AGAATACCAAATCAAAGCTTGGACAGATTTCCGTTTCCCTGGCCGTGGCAACACATACTCTGATTTCAAATGGCATTGGTACC
ATTTCGATGGCGCTGATTGGGATGAATCTCGTAAAATCTCTCGTATCTTCAAATTCCGTGGCGAAGGCAAAGCTTGGGATTGG
GAAGTTTCTTCTGAAAACGGCAACTACGATTACCTTATGTACGTTGATGTTGATTACGATCATCCTGATGTTGTTGCTGAAAC
AAAAAAATGGGGCATCTGGTACGCTAACGAACTTTCTCTTGATGGCTTCCGTATCGATGCTGCTAAACATATCAAATTCTCTT
TCCTTCGTGATTGGGTTCAAGCTGTTCGTCAAGCTACAGGCAAAGAAATGTTCACAGTTGCTGAATACTGGCAAAACGATGCT
GGCAAACTTGAAAACTACCTTAACAAAACATCTTTCAACCAATCTGTTTTCGATGTTCCTCTTCATTTCAACCTTCAAGCTGC
TTCTTCTCAAGGCGGCGGCTACGATATGCGTCGTCTTCTTGATGGCACAGTTGTTTCTCGTCATCCTGAAAAAGCTGTTACAT
TCGTTGAAAACCATGATACACAACCTGGCCAATCTCTTGAATCTACAGTTCAAACATGGTTCAAACCTCTTGCTTACGCTTTC
ATCCTTACACGTGAATCTGGCTACCCTCAAGTTTTCTACGGCGATATGTACGGCACAAAAGGCACATCTCCTCGTGAAATCCC
TTCTCTTAAAGATTCTATCGAACCTATCCTTAAAGCTCGTAAAGAATACGCTTACGGCCCTCAACATGATTACATCGATCATC
CTGATGTTATCGGCTGGACACGTGAAGGCGATTCTTCTGCTGCTAAATCTGGCCTTGCTGCTCTTATCACAGATGGCCCTGGC
GGCTCTAAACGTATGTACGCTGGCCTTAAAAACGCTGGCGAAACATGGTACGATATCACAGGCAACCGTTCTGATACAGTTAA
AATCGGCTCTGATGGCTGGGGCGAATTCCATGTTAACGATGGCTCTGTTTCTATCTACGTTCAAAAATAAGGTAATAAAAAAA
CACCTCCAAGCTGAGTGCGGGTATCAGCTTGGAGGTGCGTTTATTTTTTCAGCCGTATGACAAGGTCGGCATCAGGTGTGACA
AATACGGTATGCTGGCTGTCATAGGTGACAAATCCGGGTTTTGCGCCGTTTGGCTTTTTCACATGTCTGATTTTTGTATAATC
AACAGGCACGGAGCCGGAATCTTTCGCCTTGGAAAAATAAGCGGCGATCGTAGCTGCTTCCAATATGGATTGTTCATCGGGAT
CGCTGCTTTTAATCACAACGTGA

Figure S1. *amyQ* gene sequence with codon optimization for *B. subtilis*. The ribosomal binding site is marked in bold and the signal peptide is underline.

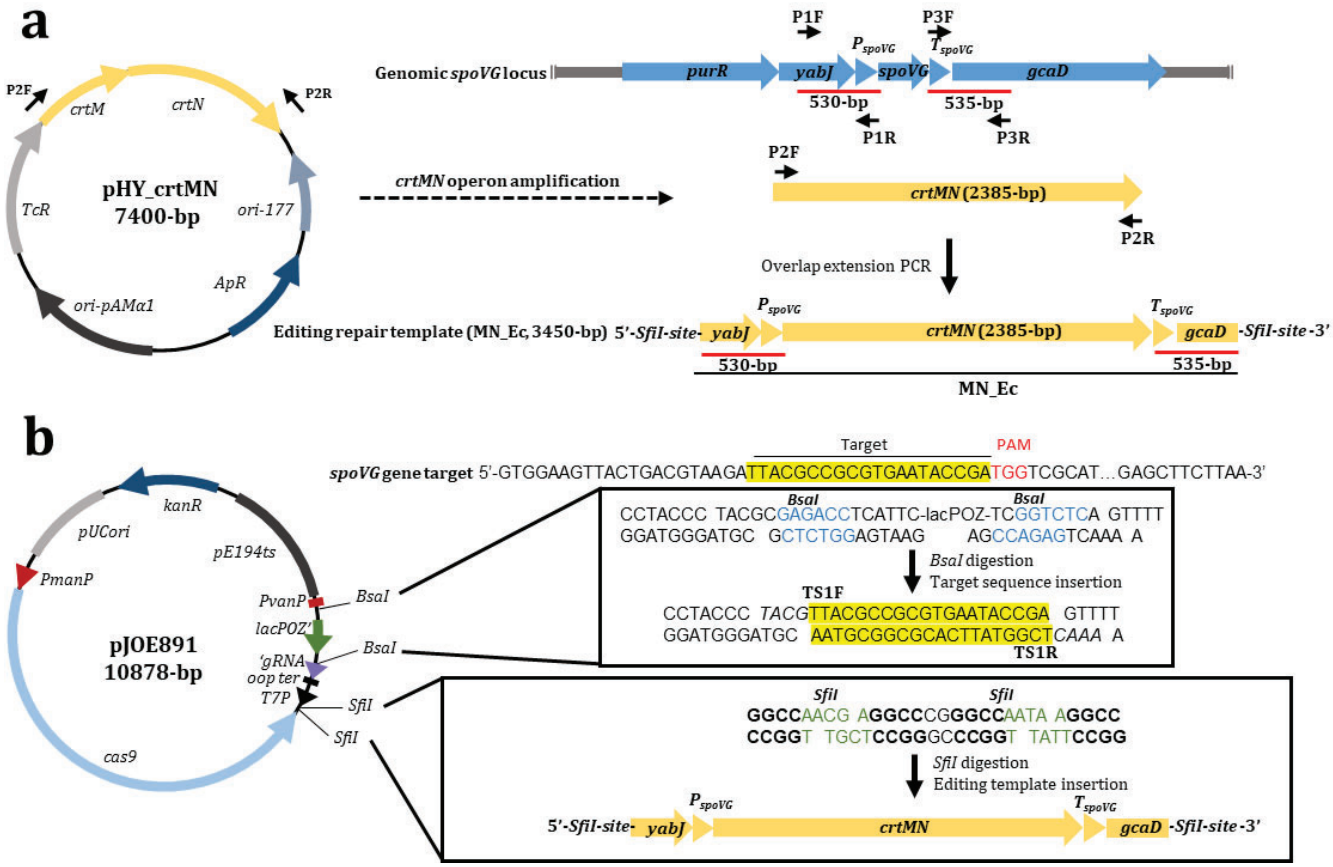


Figure S2. Construction system of vector pJOE891 (**a**) Diagram showing the editing repair template construction (MN_Ec) design consisting of *crtMN* operon flanked by *spoVG* homologous arms. (**b**) Physical map of pJOE891 vector containing pUCori, *E. coli* replication origin; pE194ts, *B. subtilis* temperature-sensitive replication origin; *kanR*, kanamycin resistance marker; *cas9*, Cas9 encoding gene; P_{manP}, mannose-inducible promoter of Cas9 encoding gene; 'gRNA, binding scaffold for Cas9; P_{vanP}, semisynthetic promoter of the sgRNA; *lacPOZ'*, *lacZ* α fragment; oop ter, λ *oop* terminator; T7P, T7 promoter. Insets show the insertion of *spoVG* gene target sequence (TS1F/TS1R primers) and editing repair template (3450-bp) in the *BsaI* and *SfiI* sites of the pJOE891 vector, respectively.

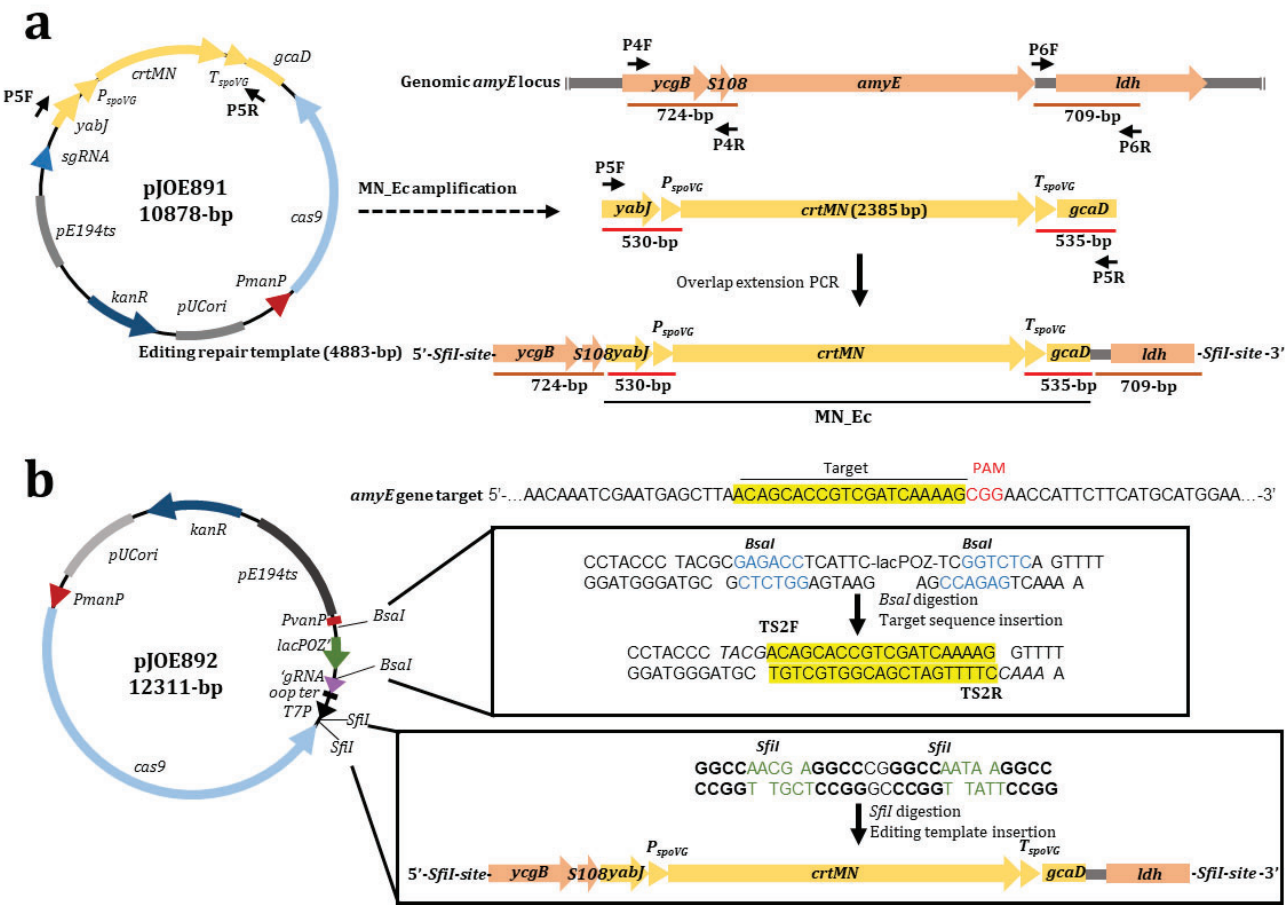


Figure S3. Construction system of vector pJOE892. (**a**) Diagram showing the editing repair template construction design consisting of MN_Ec flanked by *amyE* homologous arms. (**b**) Physical map of pJOE892 vector. Insets show the insertion of *amyE* gene target sequence (TS2F/TS2R primers) and editing repair template (4883-bp) in the *BsaI* and *SfiI* sites of the pJOE892 vector, respectively.

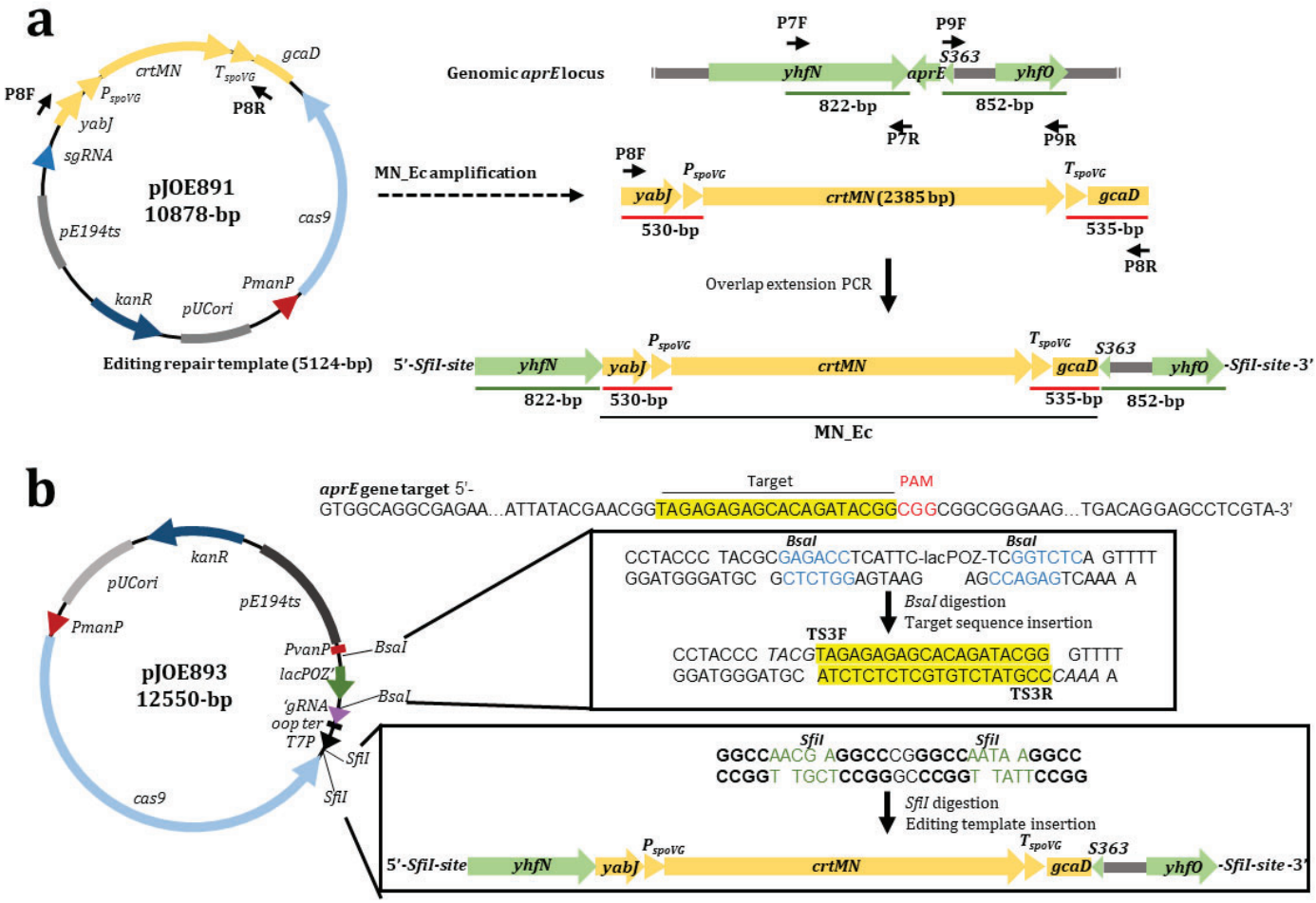


Figure S4. Construction system of vector pJOE893. (**a**) Diagram showing the editing repair template construction consisting of MN_Ec flanked by *aprE* homologous arms. (**b**) Physical map of pJOE893 vector. Insets show the insertion of *aprE* gene target sequence (TS3F/TS3R primers) and editing repair template (5124-bp) in the *BsaI* and *SfiI* sites of the pJOE893 vector, respectively.

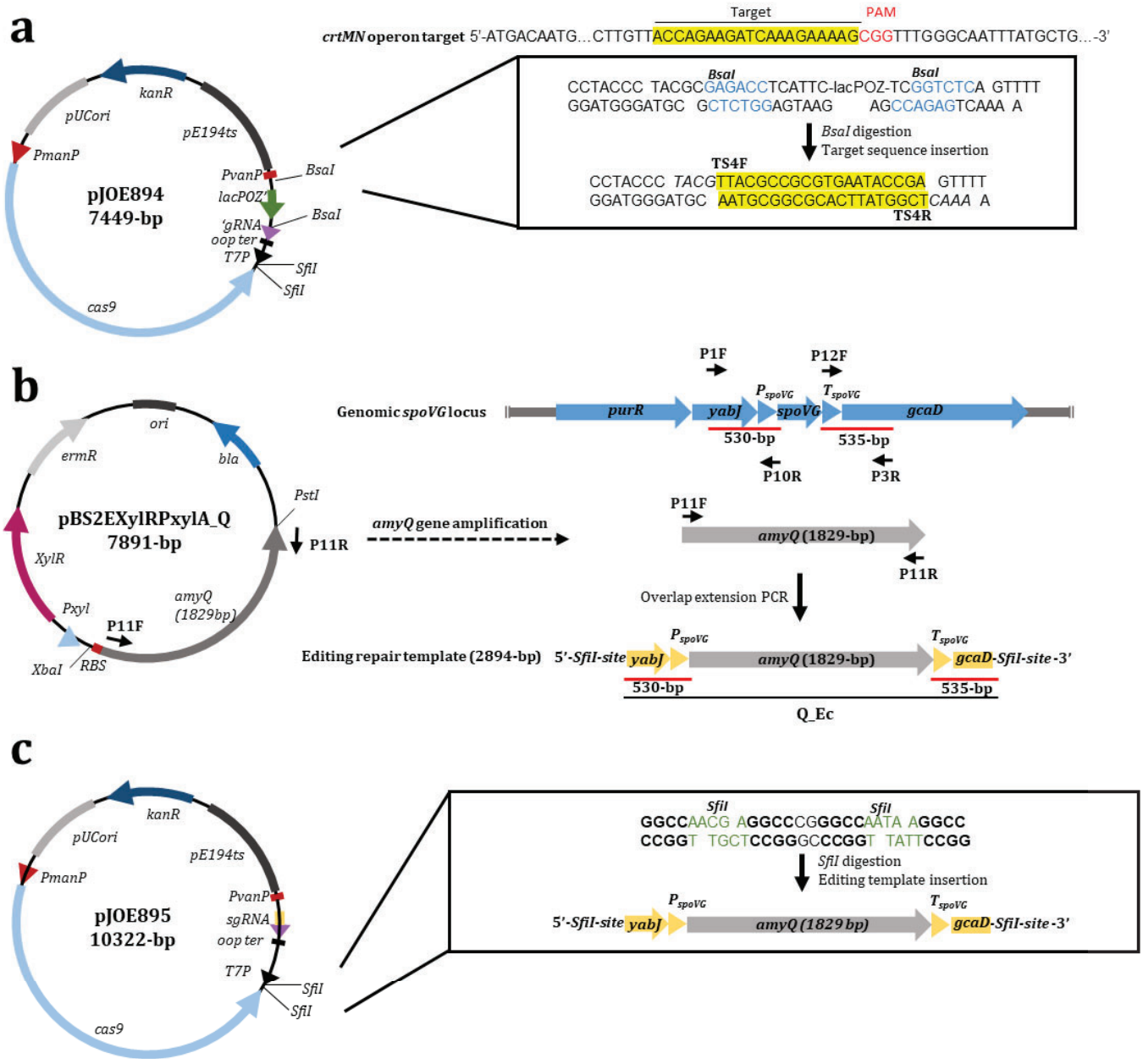


Figure S5. Construction system of vector pJOE894. (**a**) Construction system of pJOE894 plasmid. Inset show the insertion of *crtMN* operon target sequence (TS4F/TS4R primers) in the *BsaI* site of the pJOE8999 vector. (**b**) Diagram showing the editing repair template (Q_Ec) construction design consisting of *amyQ* gene (1829-bp) flanked by *spoVG* homologous arms (**c**) Physical map of pJOE895. Inset show the insertion of Q_Ec in the *SfiI* site of the pJOE895 vector.

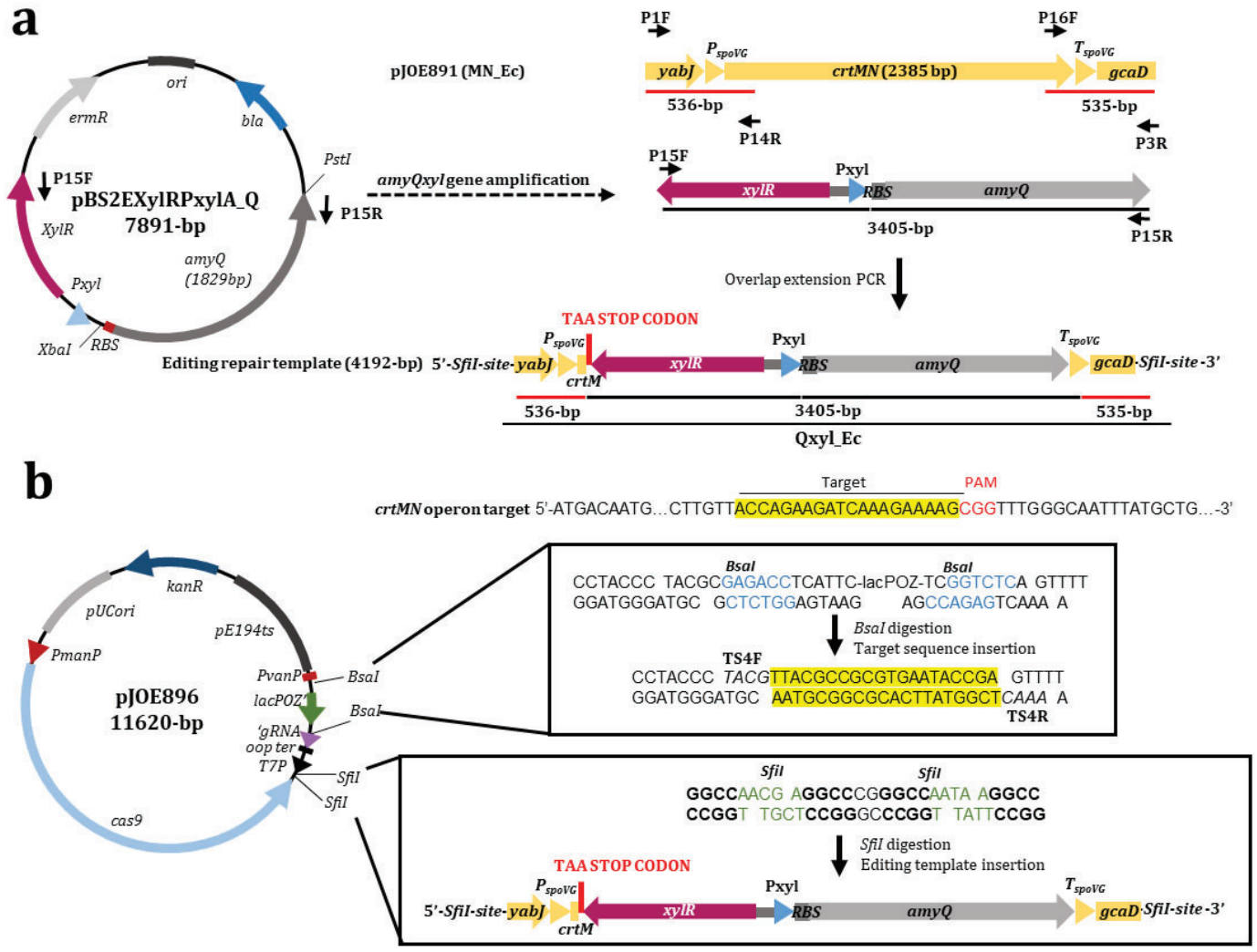


Figure S6. Construction system of vector pJOE896. (**a**) Diagram showing the editing repair template construction (Qxyl_Ec) design consisting of *amyQxyl* gene (3405-bp) flanked by *spoVG* homologous arms. (**b**) Physical map of pJOE896 vector. Insets show the insertion of *crtMN* operon target sequence (TS4F/TS4R primers) and Qxyl_Ec in the *BsaI* and *SfiI* sites of the pJOE896, respectively.

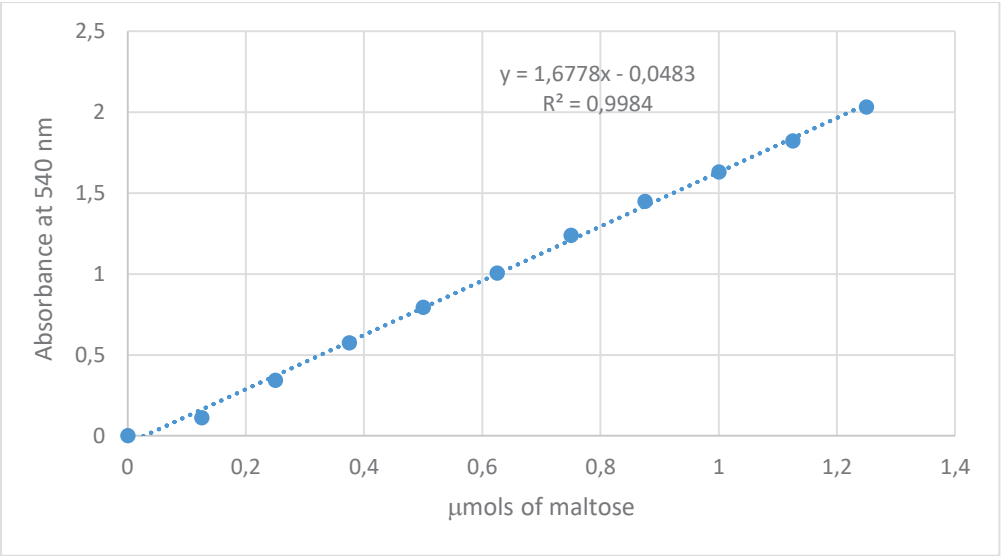


Figure S7. Standard calibration curve for the determination of maltose released in the α-amylase assay.

Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system

Jordi Ferrando, Oriana Filluelo, Daniel R Zeigler and Pere Picart

SUPPLEMENTARY MATERIAL.

Additional file 2

Table S1: Plasmids used in this study

Table S2: Primers designed in this study

Table S3: Splicing with overlap extension PCR (SOEing-PCR) program

Table S1. Plasmids used in this study.

Plasmid	Description	Source/Reference
pHY_crtMN	Plasmid pHY300PLK containing the *crtMN* operon from *Staphylococcus aureus*	(1)
pJOE8999	$P_{manP\text{-}cas9}$, pUC, pE194^{ts}, *kan*^{r}	(2)
pBS2EXxylRPxylA	Plasmid containing the xylose-inducing promoter xylose-repressor system	(3)
pBS2EXxylRPxylA_Q	Plasmid pBS2EXxylRPxylA containing the *amyQ* gene	This study
pJOE891	Plasmid used to replace the *spoVG* locus for MN_Ec	This study
pJOE892	Plasmid used to replace the *amyE* locus for MN_Ec	This study
pJOE893	Plasmid used to replace the *aprE* locus for MN_Ec	This study
pJOE894	Plasmid used to confirm that the NHEJ system is not active in *B. subtilis*	This study
pJOE895	Plasmid used for the simultaneous insertion of multiple Q_Ec copies	This study
pJOE896	Plasmid used for the simultaneous insertion of multiple Qxyl_Ec copies	This study

Table S2. Primers designed in this study.

Name	Sequence	Purpose	PCR Product
TS1F	5' – TACGTTACGCCGCGTGAATACCGA	**Target sequence for *spoVG* gene replacement**	
TS1R	5' – AAACTCGGTATTCACGCGGCGTAA		
		PCR of homology template for *spoVG* replacement for *crtMN* operon (MN_Ec)	**3450-bp**
P1F	5' – AAGGCCAACGAGGCCCGGGCCTTATTCACAAGGG	upstream *spoVG* gene homologous arm	530-bp
P1R	5' – AAATTCATATCCATCATTGTCATAGTAGTTCACCACCTTTTC		
P2F	5' – GAAAAGGTGGTGAACTACTATGACAATGATGGATATGAATTT	*crtMN* operon amplification	2385-bp
P2R	5' – TCCTTGCTTTTTGGTTATTTTTTATACGCCCCGCTCAAT		
P3F	5' – ATTGAGCGGGGCGTATAAAAAAATAACCAAAAAGCAAGGA	downstream *spoVG* gene homologous arm	535-bp
P3R	5' – AAGGCCTTATTGGCCTTACTGTTCCATCGTCTCTGCTG		
TS2F	5' – TACGACAGCACCGTCGATCAAAAG	**Target sequence for amyE gene replacement**	
TS2R	5' – AAACCTTTTGATCGACGGTGCTGT		
		PCR of homology template for *amyE* replacement for *crtMN* operon	**4883-bp**
P4F	5' – AAGGCCAACGAGGCCGAGTTATTCATTGCAGAAGCGCA	upstream *amyE* gene homologous arm	724-bp
P4R	5' – ACCGATGTGAAGACTGGAGAA		
P5F	5' – TTCTCCAGTCTTCACATCGGTTCGGGCCTTATTCACAAGGG	Expression cassete containing *crtMN* operon amplification (MN_Ec)	3450-bp
P5R	5' – TTACTGTTCCATCGTCTCTGCTG		
P6F	5' – CAGCAGAGACGATGGAACAGTAAGCGGTTCTCTTCCCCATTGA	downstream *amyE* gene homologous arm	709-bp
P6R	5' – AAGGCCTTATTGGCCTACGTTTTGAGGCGCTGCG		
TS3F	5' – TACGTAGAGAGAGCACAGATACGG	**Target sequence for *aprE* gene replacement**	
TS3R	5' – AAACCCGTATCTGTGCTCTCTCTA		
		PCR of homology template for *aprE* replacement for *crtMN* operon	**5124-bp**
P7F	5' – AAGGCCAACGAGGCCGCTCATACAAGCTTCTTGCC	upstream *aprE* gene homologous arm	822-bp
P7R	5' – CACTCTTTACCCTCTCCTTTTAA		
P8F	5' – TTAAAAGGAGAGGGTAAAGAGTGTCGGGCCTTATTCACAAGGG	Expression cassete containing *crtMN* operon amplification (MN_Ec)	3450-bp
P8R	5' – TTACTGTTCCATCGTCTCTGCTG		
P9F	5' – CAGCAGAGACGATGGAACAGTAATAGTAAAAAGAAGCAGGTTCCT	downstream *aprE* gene homologous arm	852-bp
P9R	5' – AAGGCCTTATTGGCCCTTTACGCTTTGCGTTCTCG		

Table S2. Primers designed in this study (continued).

Name	Sequence	Purpose	PCR Product
TS4F	5' – TACGACCAGAAGATCAAAGAAAAG	**Target sequence for *crtMN* operon replacement**	
TS4R	5' – AAACCTTTTCTTTGATCTTCTGGT		
		PCR of homology template for crtMN *operon* replacement for amyQ *gene* (Q_Ec)	**2894-bp**
P1F	5' – AAGGCCAACGAGGCCCGGGCCTTATTCACAAGGG	upstream *spoVG* gene homologous arm	530-bp
P10R	5' – CTTTCGTTTTTGAATCATAGTAGTTCACCACCTTTTCCC		
P11F	5' – GTGAACTACTATGATTCAAAAACGAAAGCGGAC	*amyQ* gene amplification	1829-bp
P11R	5' – GGTTATTTTCACGTTGTGATTAAAAGCAGCG		
P12F	5' – GCTTTTAATCACAACGTGAAAATAACCAAAAAGCAAGGACTG	downstream *spoVG* gene homologous arm	535-bp
P3R	5' – AAGGCCTTATTGGCCTTACTGTTCCATCGTCTCTGCTG		
		Cloning *the* amyQ gene in plasmid pBS2EXxylRPxylA *using* XbaI-PstI sites	
P13F	5' – GGTCTAGAGATAAGAAAGGGAGGACAAACA	*amyQ* gene with ribosome-binding site amplification	1850-bp
P13R	5' – TACTGCAGCACGTTGTGATTAAAAGCAGCG		
		PCR of homology template for crtMN *operon* replacement for amyQ-xylR gene (Qxyl_Ec)	**4670-bp**
P1F	5' – AAGGCCAACGAGGCCCGGGCCTTATTCACAAGGG	upstream *spoVG* site homologous arm	630-bp
P14R	5' – TTATTGATCTTCTGGTAACAAGTCA		
P15F	5' – TGACTTGTTACCAGAAGATCAATAATGGCCGACTTTAGATATTTTCGT	*amyQxyl* gene amplification	3405-bp
P15R	5' – TGGATGCGTACTTGCACCTGCCACGTTGTGATTAAAAGCAGCG		
P16F	5' – GCAGGTGCAAGTACGCATCCA	downstream *spoVG* site homologous arm	635-bp
P3R	5' – AAGGCCTTATTGGCCTTACTGTTCCATCGTCTCTGCTG		
		Validation primers used in the genome editing experiments	
P17F	5' – TTCAGGCTCCTCAAACCGCA	PCR verification with outside primers for gene insertion at *spoVG* site	4134-bp (*crtMN* operon)
P17R	5' – TTACCAGTCTGGTAAGCGGC		3578-bp (*amyQ* gene)
			5154-bp (*amyQxyl* gene)
P18F	5' – GAGTTATTCATTGCAGAAGCGCA	PCR verification with outside primers for gene insertion at *amyE* site	4883-bp (*crtMN* operon)
P18R	5' – TACGTTTTGAGGCGCTGCG		4327-bp (*amyQ* gene)
			6098-bp (*amyQxyl* gene)
P19F	5' – GCTCATACAAGCTTCTTGCC	PCR verification with outside primers for gene insertion at *aprE* site	5124-bp (*crtMN* operon)
P19R	5' – CTTTACGCTTTGCGTTCTCG		4568-bp (*amyQ* gene)
			6339-bp (*amyQxyl* gene)

Table S3. Splicing with overlap extension PCR (SOEing-PCR) program.

OVERLAP EXTENSION PCR	TEMPERATURE	TIME
STEP 1		
Initial Denaturation	95°C	5 minutes
15 cycles	95°C	30 seconds
	X °C *	1:30 minutes
	72°C	X minutes **
Final Extension	72°C	5 minutes
STEP 2		
Initial Denaturation	95°C	5 minutes
30 cycles	95°C	30 seconds
	X °C *	30 seconds
	72°C	X minutes **
Final Extension	72°C	5 minutes
Hold	4 °C	

** Depending on the melting temperature (Tm) of the primer

** Depending on the length of the fragment to be amplified

References

1. Yoshida K, Ueda S, Maeda I. Carotenoid production in *Bacillus subtilis* achieved by metabolic engineering. Biotechnol Lett. 1, 1789-1793 (2009).
2. Altenbuchner J. Editing of the *Bacillus subtilis* Genome by the CRISPRCas9 System. Appl Environ Microbiol. 82, 5421–5427 (2016).
3. Popp PF, Dotzler M, Radeck J, Bartels J, Mascher T. The *Bacillus* BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. Sci Rep. 7, 15058-15070 (2017).

4.3.2. Paper 4: Metabolic engineering of *Bacillus subtilis* toward the efficient and stable production of C30-carotenoids.

Filluelo *et al. AMB Express* (2023) 13:38
https://doi.org/10.1186/s13568-023-01542-x

AMB Express

ORIGINAL ARTICLE Open Access

Metabolic engineering of *Bacillus subtilis* toward the efficient and stable production of C_{30}-carotenoids

Oriana Filluelo[1†], Jordi Ferrando[1†] and Pere Picart[1*]

Abstract

Commercial carotenoid production is dominated by chemical synthesis and plant extraction, both of which are unsustainable and can be detrimental to the environment. A promising alternative for the mass production of carotenoids from both an ecological and commercial perspective is microbial synthesis. To date, C_{30} carotenoid production in *Bacillus subtilis* has been achieved using plasmid systems for the overexpression of biosynthetic enzymes. In the present study, we employed a clustered regularly interspaced short palindromic repeat-Cas9 (CRISPR-Cas9) system to develop an efficient, safe, and stable C_{30} carotenoid-producing *B. subtilis* strain, devoid of plasmids and antibiotic selection markers. To this end, the expression levels of *crtM* (dehydrosqualene synthase) and *crtN* (dehydrosqualene desaturase) genes from *Staphylococcus aureus* were upregulated by the insertion of three gene copies into the chromosome of *B. subtilis*. Subsequently, the supply of the C_{30} carotenoid precursor farnesyl diphosphate (FPP), which is the substrate for CrtMN enzymes, was enhanced by expressing chromosomally integrated *Bacillus megaterium*-derived farnesyl diphosphate synthase (FPPS), a key enzyme in the FPP pathway, and abolishing the expression of farnesyl diphosphate phosphatase (YisP), an enzyme responsible for the undesired conversion of FPP to farnesol. The consecutive combination of these features resulted in a stepwise increased production of C_{30} carotenoids. For the first time, a *B. subtilis* strain that can endogenously produce C_{30} carotenoids has been constructed, which we anticipate will serve as a chassis for further metabolic engineering and fermentation optimization aimed at developing a commercial scale bioproduction process.

Key points

- Overexpression of chromosomally integrated *crtMN* genes improved C_{30} carotenoid production
- Overexpression of FPPS and branch pathway attenuation further enhanced C_{30} carotenoid yield
- A stable plasmid-less, marker-less C_{30} carotenoid-producing *B. subtilis* strain was constructed

Keywords *B. subtilis, C_{30} carotenoids, CRISPR-Cas9, Metabolic engineering*

[†]Oriana Filluelo and Jordi Ferrando have contributed equally to this work.

*Correspondence:
Pere Picart
perepicart@ub.edu

[1]Faculty of Pharmacy and Food Science Technology, Department of Biology, Healthcare and the Environment, Microbiology Section, University of Barcelona, Avinguda Joan XXIII, 27-31, Barcelona 08028, Spain

Introduction

Terpenoids (also known as isoprenoids) constitute one of the largest and structurally most diverse groups of natural products with diverse biological functions (Zhang and Hong 2020). An economically important class of terpenoids are the carotenoids, which are ubiquitous lipid-soluble pigments responsible for the red, yellow, and orange colors of plants, algae, fungi, and bacteria (Cardoso et al. 2017). Although commercial carotenoid production is dominated by chemical synthesis and plant extraction, these processes are not sustainable or ecological. Carotenoids are chemically synthesized under harsh conditions, generating byproducts and hazardous waste, whereas sourcing carotenoids from plant extracts is generally dependent on the seasons and geographic areas, which cannot always be standardized (Siziya et al. 2022). Therefore, microbial production is emerging as one of the most promising safe and environmentally friendly options to satisfy the fast-growing demands for carotenoids (Siziya et al. 2022).

B. subtilis is generally recognized as safe (GRAS), has a high growth rate, and is easy to genetically manipulate and cultivate, with a wide substrate range (Earl, 2008; Schallmey et al., 2004). In addition, it is one of the highest producer of isoprene (the smallest terpenoid) among eubacteria, thus constituting an ideal microbial host for use as a terpenoid cell factory (Kuzma et al. 1995; Wagner et al. 2000; Julsing et al. 2007; Moser and Pichler 2019; Guan et al. 2015). This bacterium is able to initiate terpenoid biosynthesis from simple carbon sources through the methylerythritol 4-phosphate (MEP) pathway, a route with eight enzymatic reactions leading to the synthesis of isopentenyl diphosphate (IPP; C5) and dimethylallyl diphosphate (DMAPP; C5), the universal precursors of all terpenoids (Guan et al. 2015). The consecutive condensation of IPP and DMAPP is catalyzed by prenyl diphosphate synthase (*IspA*) to produce starting precursors for the synthesis of different classes of terpenoids: geranyl diphosphate (GPP; C10), a monoterpenoid precursor; farnesyl diphosphate (FPP; C15) for the production of sesquiterpenoids, triterpenoids and C_{30}-carotenoids, and geranylgeranyl diphosphate (GGPP; C20), the precursor of diterpenoids and carotenoids (Moser and Pichler 2019). Most carotenoids contain a 40-carbon backbone (C_{40} carotenoids), including β-carotene, lycopene and astaxanthin, whereas those with 30-carbon backbones (C_{30} carotenoids), such as 4,4'-diaponeurosporene (DNP) and 4,4'- diapolycopene (DLP), are synthesized by a limited group of bacteria, including *Staphylococcus aureus* (Marshall and Wilmoth 1981), and *Heliobacteria spp*. (Takaichi et al. 1997). Genes responsible for C_{30} carotenoid biosynthesis in *S. aureus* have been characterized (Pelz et al. 2005; Wieland et al. 1994).) The first dedicated enzyme in the C_{30} carotenoid synthetic pathway is CrtM (dehydrosqualene synthase), which catalyzes the head-to-head condensation of two molecules of FPP to dehydrosqualene. The enzyme CrtN (dehydrosqualene desaturase) then converts dehydrosqualene to the yellow C_{30} carotenoid, DNP, a relatively unstable compound that can suffer further oxidation by CrtMN to yield DLP. The action of these two enzymes probably constitutes the most common route of C_{30} carotenoid biosynthesis in bacteria. Notably, these yellow pigments have attracted interest from the pharmaceutical industry owing to their powerful antioxidant activities (Yoshida et al. 2009), as well as their role as immunomodulators, significantly enhancing the immune system (Jing et al. 2017, 2019; Liu et al. 2016, 2017). Consequently, microbial cell engineering approaches aimed at improving C_{30} carotenoid yields are required to achieve industrial-scale production.

To date, the metabolic engineering of *B. subtilis* toward enhanced C_{30} carotenoid production has focused on using two-plasmid systems comprising pHY_crtMN (Yoshida et al. 2009), mediating *crtMN* gene overexpression under tetracycline selection, and xylose-inducible pHCMC04G (Xue et al. 2015), mediating stable overexpression of all MEP pathway enzymes under non-selection conditions (Abdallah et al. 2020). However, two-plasmid systems may impose a metabolic burden on the host cells, leading to lower growth rates and increased productivity costs (Wu et al. 2016). Another drawback is the high-cost of the inducer compounds and, more importantly, the requirement for antibiotic usage, which is restricted by governmental regulations and can thus hinder the establishment of a commercially viable industry. On the other hand, very little work has been done to explore the effects of modulating *crtMN* gene expression and other competing branch pathways (which can limit FPP availability) on C_{30} carotenoid production, leaving room for improvement. In this study, we initially compared the expression levels of plasmid-based and chromosomally integrated *crtMN* genes, and then implemented CRISPR-Cas9-based metabolic engineering strategies to achieve an efficient C_{30} carotenoid-producing strain of *B. subtilis*, a bacterium that naturally produces yellow pigments (Fig. 1). Thus, with the aim of increasing the supply of the carotenoid precursor FPP, we planned (i) to introduce a chromosomally integrated copy of FPPS (farnesyl diphosphate synthase), and (ii) to abolish the activity of a competing branch pathway that uses FPP. With this approach, it was envisaged that we could construct a stable and efficient C_{30} carotenoid-producing *B. subtilis* strain that was plasmid- and marker-free, an attribute of paramount importance for its potential development into a commercially viable bioprocess.

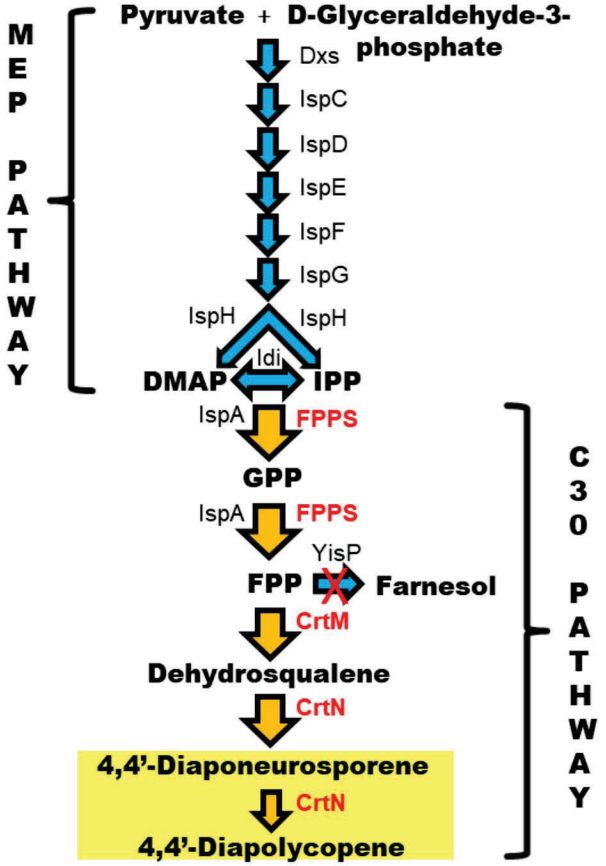


Fig. 1 Metabolic pathways associated with terpenoid biosynthesis in *B. subtilis* and engineering strategies for the production of yellow C_{30} carotenoids 4,4′-diaponeurosporene and 4,4′-diapolycopene (C_{30} pathway). Foreign genes are marked in red. Yellow arrows outlined in black indicate the reactions reinforced by chromosomic overexpression of the *fpps* gene (farnesyl diphosphate synthase) from *B. megaterium* DSM 319, *crtM* (squalene desaturase) and *crtN* (dehydrosqualene desaturase or diapophytoene desaturase) genes from *S. aureus*, and deletion of the *yisP* (farnesyl diphosphate phosphatase) gene, yielding the C_{30} carotenoid pigments 4,4′-diaponeurosporene and 4,4′-diapolycopene. Enzymes in the MEP (Methylerythritol 4-phosphate) pathway: 1-deoxy-D-xylulose-5-phosphate synthase (*Dxs*); 1-deoxy-D-xylulose-5-phosphate reductoisomerase or 2-C-methyl-D-erythritol 4-phosphate synthase (*Dxr*, also known as *IspC*); 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (*IspD*); 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase (*IspE*); 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*IspF*); (E)-4- hydroxy-3-methylbut-2-enyl-diphosphate synthase (*IspG*); 4-hydroxy-3- methylbut-2-enyl diphosphate reductase (*IspH*); and isopentenyl-diphosphate delta-isomerase (*Idi*). Geranyltransferase (*IspA*, also known as *YqiD*) refers to the *B. subtilis* gene responsible for the supply of GPP (geranyl diphosphate) and FPP (farnesyl diphosphate) precursors

Materials and methods

Bacterial strains

The *E. coli* NEB® turbo strain (New England Biolabs) was used as the host strain for routine molecular cloning and plasmid construction operations, and *B. subtilis* KO7-S (Bacillus Genetic Stock Center), an asporogenous strain with seven inactivated protease genes, was used as a host strain for C_{30} carotenoid production. DNA isolation and manipulations were carried out using standard protocols. The bacterial strains employed in this research are listed in Table 1.

Table 1 Bacterial strains and plasmids used in this study

Strain	Genotype or description	Source/Reference
E. coli NEB® turbo	*F′ proA + B + laclq ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)*TetS	Laboratory stock
B. megaterium DSM 319	Source of *fpps* gene	DSM
B. subtilis 168 KO7-S	*ΔnprE ΔaprE Δepr Δmpr ΔnprB Δvpr Δbpr ΔsigF*	BGSC
BsMN0	*B. subtilis* KO7-S strain harboring the plasmid pHY_crtMN	This study
BsMN1	Contains one *crtMN* gene-copy integrated into the genome	Ferrando et al., 2023
BsMN2	Contains two *crtMN* gene-copies integrated into the genome	Ferrando et al., 2023
BsMN3	Contains three *crtMN* gene-copies integrated into the genome	Ferrando et al., 2023
BsMN4	*B. subtilis* KO7-S strain harboring the plasmid pBS0E_crtMN	This study
BsMN5	BsMN3 strain with an *fpps* gene from *B. megaterium* replacing the *sigX* gene	This study
BsMN6	BsMN5 strain with a truncated copy of the *yisP* gene	This study
Plasmid	**Description**	**Source/Reference**
pHY_crtMN	Plasmid pHY300PLK containing the *crtMN* operon from *Staphylococcus aureus*	Yoshida et al. 2009
pBS0E	Plasmid containing the xylose-inducing promoter xylose-repressor system	Popp et al. 2017
pJOE8999	PmanP-cas9, pUC, pE194ts, *kanr*	Altenbuchner 2016
pJOE8999_VG_MN	Plasmid used to replace the *spoVG* locus for the *crtMN genes*	Ferrando et al., 2023
pBS0E_crtMN	Plasmid containing *crtMN* genes under the control of xylose-inducing promoter	This study
pJOE8999.sg_sigX_fpps	Plasmid used to replace the *sigX* locus for the *fpps* gene	This study
pJOE8999.sg_ΔyisP	Plasmid used to delete the *yisP* gene	This study

Medium and culture conditions

E. coli strains were cultured in Luria-Bertani (LB) medium at 37 °C, while *B. subtilis* KO7-S strains were grown in Tryptic Soy Broth (TSB) (17 g/l tryptone, 3 g/l

soytone, 2.5 g/l dextrose, 5.0 g/l NaCl, 2.5 g/l K_2HPO_4) or *Bacillus subtilis* 1 (BS1) medium, typically used in industrial fermentation (Wenzel et al. 2011). The BS1 medium contained standard salts (in g/l: 2 (NH4)2SO4; 18.3 K2HPO4·3H2O; 6 KH2PO4; 1 Na+-citrate; 0.2 MgSO4·7H2O), trace metals (in mg/l: 120 FeSO4·7H2O; 30 MnSO4·H2O; 12 CuSO4·5H2O; 12 ZnCl2) and was supplemented with 12 g sucrose/l and 18 g soybean meal/l (Sigma Aldrich). All strains were incubated at 37 ºC on a rotatory shaker at 200 rpm. When necessary, the growth media were supplemented with antibiotics at the following concentrations: 30 µg/ml kanamycin for *E. coli*, and 6 µg/ml kanamycin or 10 µg/ml tetracycline or 2 µg/ml erythromycin for *B. subtilis*. To induce the CRISPR-Cas9 system in *B. subtilis* cells, 0.5% D-mannose was added.

Plasmid construction and primers

The plasmids used in this study are listed in Table 1 and the primers in Table S1. For the insertion of *crtMN* genes into the pBS0E vector (Popp et al. 2017), the pHY_crtMN plasmid (Yoshida et al. 2009) was used as a template to amplify *crtMN* genes using primers P1F/P1R. The resulting DNA amplicon was treated with *Eco*RI and *Spe*I and cloned into the replicative plasmid pBS0E for the construction of the xylose-inducible pBS0E_crtMN vector. CRISPR-Cas9-mediated genome editing in *B. subtilis* was performed using the pJOE8999 vector as the parental plasmid, according to a previously described method (Altenbuchner 2016).

Chromosomal integration of the *fpps* gene

To generate the *sigX* gene replacement by the *fpps* gene, oligonucleotides for 20 pb gRNA (TS1F and TS1R) were synthesized and ligated to *Bsa*I-digested pJOE8999. *sigX*-targeting gRNA containing pJOE8999 was named pJOE8999.g_sigX. A repair template for *fpps* integration into the *sigX* gene was constructed in vitro by overlap extension PCR of three fragments as follows: the 800-bp upstream flanking genomic region of *sigX* (P2F/P2R primers) followed by the *fpps* gene (P3F/P3R primers) and the 800-bp downstream flanking genomic region of *sigX* (P4F/P4R primers). Homologous arms were amplified using the *B. subtilis* KO7-S chromosome as a template, while the *fpps* gene was amplified using genomic DNA from *B. megaterium* DSM 319. The fused fragment was digested with *Sfi*I and then ligated into pJOE8999.g_sigX, which had also been digested with *Sfi*I to obtain the editing plasmid pJOE8999.g_sigX_fpps, used for FPPS overexpression.

Deletion of the *yisP* gene

To generate the *yisP* knockout mutant, a procedure similar to the one described above was performed. Primers TS2F and TS2R targeting the *yisP* gene were synthesized and ligated to the vector, thus obtaining plasmid pJOE8999.g_yisP. A 1.6 kb repair template, containing the 800-bp upstream region and 800-bp downstream region of the *yisP* gene, was PCR-amplified using the *B. subtilis* KO7-S genome as a template. Primer sets P5F/P5R and P6F/P6R were used to amplify each fragment and fused together by overlapping PCR. The repair template was further digested with *Sfi*I for ligation with pJOE8999.g_yisP to obtain the editing plasmid pJOE8999.sg_ΔyisP, which was used to delete the *yisP* gene.

Transformation and plasmid curing

The well-established plasmids (1 µg) were then transformed to *B. subtilis* KO7-S according to the standard methods described by Yasbin and coworkers (Yasbin et al. 1975). For the CRISPR-Cas9-induced genome editing, the resulting transformants were passaged three times on LB agar plates (without any antibiotics) at 50 °C for 24 h to cure the plasmid. The colonies were confirmed as cured of the editing plasmid by streaking them onto LB agar plates containing kanamycin or no antibiotics; plasmid cured colonies fail to grow at 37 °C. To confirm whether the desired insertion or deletion in the genome of *B. subtilis* had been performed, a colony PCR was conducted to amplify the target fragments from the bacterial chromosome and validated by further Sanger sequencing.

Extraction of carotenoids from *B. subtilis*

Carotenoids were extracted from the engineered *B. subtilis* cells according to the literature (Xue et al. 2015) with some modifications. Briefly, recombinant strains were inoculated in 50 ml TSB at an optical density (OD_{600}) of 0.05 and cultured for 24 h at 37 ºC (250 rpm). In the case of xylose-inducing experiments, 1% xylose was added at an OD_{600} of 0.6, and strains were then cultured for an additional 24 h in the same conditions. Samples were collected by centrifugation at 8000 *g* for 15 min and washed with 1 ml TE buffer (10 mM Tris/HCl, 1 mM EDTA, pH 8.0). The cells were resuspended in 500 µl TE buffer. To extract the carotenoids, cell suspensions were lysed with 25 µl of 100 mg/ml lysozyme, followed by incubation for 15 min at 37 °C. The cell lysate was then transferred into a glass tube, covered in aluminum foil to avoid light exposure, and centrifuged for 20 min at 2100 g. The supernatant was removed, and 1 ml acetone was added to the pellets. These were vortexed for 4 min, heated for 2 min in a water-bath at 55ºC, and then vortexed again for 2 min. After centrifugation at 2300 g for 15 min, the supernatants were collected and transferred to a new glass tube. The acetone extraction was repeated four times. Next, the acetone extracts were evaporated, and the remaining carotenoids were dissolved in 100 µl acetone and collected in HPLC vials, prior to their analysis

Table 2 Comparison of dry cell weight, titer and yield of C_{30} carotenoids produced by engineered *B. subtilis* strains and relative increase compared to the control strain

Strain	DCW (g/L culture)	Titer Carotenoids (mg/L culture)	Yield Carotenoids (mg/g DCW)[a]	Relative increase[b]
BsMN0	1.36 ± 0.09	1.01 ± 0.08	0.74 ± 0.07	1
BsMN1	1.23 ± 0.13	2.40 ± 0.13	1.95 ± 0.12	2.64
BsMN2	1.35 ± 0.05	2.96 ± 0.07	2.19 ± 0.08	2.96
BsMN3	1.43 ± 0.06	3.30 ± 0.11	2.31 ± 0.16	3.12
BsMN4	1.87 ± 0.16	4.22 ± 0.23	2.26 ± 0.32	3.05
BsMN5	1.32 ± 0.09	4.49 ± 0.19	3.39 ± 0.33	4.58
BsMN6	1.47 ± 0.08	6.51 ± 0.12	4.42 ± 0.19	5.97
BsMN6[c]	2.98 ± 0.14	9.11 ± 0.36	3.20 ± 0.24	NA[d]

[a] The total amount of carotenoids was measured in triplicate (±standard deviation)

[b] The relative increase is calculated as the amount of carotenoids produced in the engineered *B. subtilis* strain divided by the amount of carotenoids produced in the control strain (BsMNO) harboring the pHYCrtMN vector

[c] Strain cultured in BS1 medium

[d] Not applicable

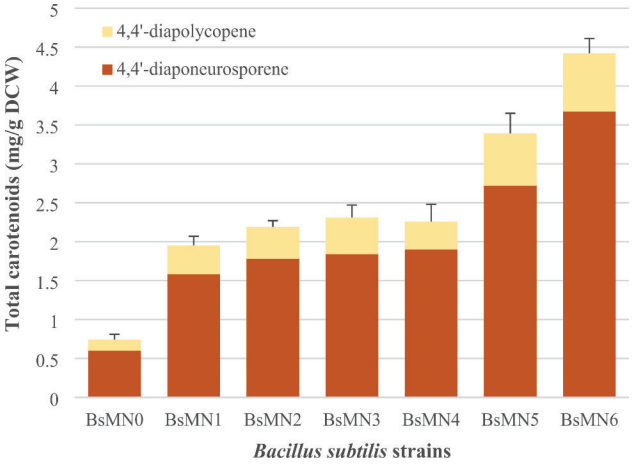


Fig. 2 Quantitative analysis of C_{30} carotenoids produced by engineered *B. subtilis* strains. Samples were analyzed by HPLC after C_{30} carotenoid extraction with acetone. Quantification of each C_{30} carotenoid was performed comparing peak areas with the standard reference curve, and then normalized to the dry cell mass of each culture. The amount of DNP is indicated in orange and the amount of DLP in yellow. The experiments were performed in triplicate

using an HPLC system. Cell dry weight was determined by pelleting and drying a fraction of the culture.

HPLC analysis of carotenoids

Carotenoid extracts were analyzed with a Shimadzu HPLC system equipped with a Gemini® NX-C18 column (5 µm, 110 Å, 250×4.60 mm) and a UV/VIS detector at 25 °C. The mobile phase consisted of acetonitrile and water (85:15%) at a flow rate of 2 ml/min. DNP and DLP were identified from their absorption spectra and quantified by comparing their peak areas using an standard calibration curve prepared with known amounts of β-carotene (quantified by absorbance), then multiplying by the molar extinction coefficient (ε) of β-carotene (138,900 M−1 cm−1 at 450 nm) (Britton et al., 2004), and dividing by the ε value for the carotenoid in question (147,000 M−1 cm−1 at 440 nm for DNP, 185,000 M−1 cm−1 at 470 nm for DLP) (Furubayashi et al. 2014). Production weights of carotenoids were then normalized to the dry cell weight (DCW) of each culture.

Results

Dependence on the *crtMN* gene copy number in C_{30} carotenoid production

A set of plasmid-less, marker-free *B. subtilis* strains harboring one (BsMN1), two (BsMN2) or three copies (BsMN3) of *crtMN* genes in their chromosomes under the control of the constitutive *spoVG* promoter were previously constructed by our research group, but not characterized (Ferrando et al. 2023). Therefore, to investigate the effect of multiple *crtMN* gene copy expression on the intracellular accumulation of C_{30} carotenoids, cells of a stationary overnight culture in TSB were diluted to an OD_{600} of 0.05 in TSB and grown in shake flasks at 220 rpm and 37 °C for 24 h. Then, samples were taken to quantify both the DCW and the total amounts of DNP and DLP by HPLC. The latter were calculated as mg/g DCW to allow comparison between the strains. The parental *B. subtilis* strain (BsMN0) containing only the pHY_crtMN plasmid was used as a control.

After 24 h of growth, all engineered *B. subtilis* strains had an OD_{600} of 7–8, with DCW values of 1.23–1.47 g/L, showing a slight increase in DCW as the *crtMN* gene copy number increased (Table 2). HPLC chromatogram analysis revealed two major peaks at 450 nm, which eluted at 2.4 and 2.8 min, with absorption spectra for each peak identical to those of DLP and DNP, respectively (Fig. S1) (Takaichi 2000; Takaichi et al. 1997). As the two peaks were present in the chromatograms of all samples, both compounds were calculated individually as well as together as total carotenoids, with the results provided in Fig. 2; Table 2. Surprisingly, the BsMN1 strain harboring a single copy of *crtMN* genes produced a titer of 2.40±0.13 mg/L carotenoids with a yield of 1.95±0.12 mg/g DCW, which was already more than a 2-fold increase in total carotenoid production compared to strain BsMN0 containing the pHY_crtMN plasmid (0.74±0.12 mg/g DCW). We observed that DCW and carotenoid yield slightly increased with increasing *crtMN* copy number and the highest titer of 3.30±0.11 mg/L carotenoids was achieved in BsMN3, with a yield of 2.31±0.16 mg/g DCW, which constituted a 3.12-fold increase in carotenoid production compared to BsMN0 (Fig. 2 and Table 2). The yield obtained in BsMN0 was comparable with previously reported values (Xue et al.,

2015; Abdallah et al. 2020), which demonstrates the feasibility and robustness of the comparative studies.

The low carotenoid yield obtained in BsMN0 suggested that *crtMN* genes are poorly expressed through the pHY_crtMN plasmid. To test this hypothesis, we cloned the *crtMN* genes in the xylose-inducible medium copy number pBS0E plasmid (Popp et al. 2017), which is particularly useful for overcoming bottlenecks in protein overproduction generated by limited expression of targeted genes (Toymentseva et al. 2012). The *B. subtilis* strain bearing the pBS0E_crtMN plasmid (BsMN4) showed a higher cell growth compared to BsMN0 – BsMN3 strains, with a DCW of 1.87 g/L, probably due to the addition of an extra carbon source (D-xylose inducer) to the media. As expected, BsMN4 exhibited a notable increase in carotenoid yield (3.05-fold) compared to BsMN0, demonstrating a higher expression of *crtMN* genes through this plasmid (Table 2). More importantly, the yield obtained for strain BsMN4 was similar to that of BsMN3, indicating that plasmid-bearing and multicopy strains had a comparable performance.

Optimization of the C_{30} carotenoid biosynthetic pathway

In the C_{30} carotenoid metabolic pathway in *B. subtilis*, farnesyl diphosphate synthase (IspA) converts the universal terpenoid precursors DMAPP and IPP to FPP, which is the substrate for CrtMN enzymes in C_{30} carotenoid biosynthesis (Fig. 1). In order to further improve the production of C_{30} carotenoids, we aimed to increase the FPP supply, as studies report that enhanced FPP availability drives metabolic flux toward their synthesis (Xue et al. 2015; Abdallah et al. 2020; Song et al. 2021). This has been achieved previously by introducing either an extra copy of *ispA* to release the theoretical bottleneck within the metabolic pathway or an improved variant of the enzyme with enhanced catalytic properties (Zhao et al. 2013). In the present study, farnesyl diphosphate synthase (encoded by the *fpps* gene) from *B. megaterium* DSM 319, which is an active highly specific enzyme exclusively yielding FPP (Hartz et al., 2018), was overexpressed to enhance the FPP pool. To this end, plasmid pJOE8999.sigX_fpps was constructed for the replacement of the *sigX* gene of BsMN3 (codifying for sigma factor SigX) with the *fpps* gene, setting the expression of the encoded FPPS under the control of a strong *sigX* promoter (Song et al. 2016), and strain BsMN5 was generated (Fig. 3a). The insertion of the *fpps* gene in cured transformant cells was confirmed by diagnostic PCR (Fig. 3d) and further Sanger sequencing. Fermentation studies revealed a remarkable 46.8% increase in the production of C_{30} carotenoids compared with BsMN3 (Fig. 2and Table 2). Additionally, BsMN5 grew at a similar rate to the parental strain BsMN3, indicating that the overexpression of FPPS did not affect cell growth in TSB medium. Based on these results, we surmised that heterologous expression of FPPS in *B. subtilis* is beneficial for the construction of a high-yielding C_{30} carotenoid-producing strain.

Branch pathway engineering to increase C_{30} carotenoid production

To provide enough FPP for C_{30} carotenoid biosynthesis, it is crucial to attenuate branch pathways that use this precursor as the starting material. In the biosynthesis of farnesol lipids, each FPP molecule is converted to farnesol by the action of farnesyl diphosphate phosphatase (YisP) (Fig. 1); therefore, this branch pathway was selected as a candidate for engineering. Plasmid pJOE8999_ΔyisP was constructed to knock out a 770-bp fragment of *yisP* in strain BsMN5 and inactivate the function of YisP, thus blocking the synthesis of farnesol in the newly generated strain BsMN6 (Fig. 3b and c). Disruption of the *yisP* gene in resulting transformants was confirmed by PCR amplification, as previously (Fig. 3e), and further verified by sequencing. The positive clone was cured from the plasmid and subjected to fermentation for 24 h to measure the production of DLP and DNP. Again, BsMN6 growth was similar to the parental strain BsMN5, indicating that *yisP* disruption in BsMN6 did not affect cell growth. However, C_{30} carotenoid production in strain BsMN6 was significantly enhanced, being 130.4% relative to BsMN5 after fermentation (Fig. 2; Table 2). Overall, combining the simultaneous overexpression of farnesyl diphosphate synthase, dehydrosqualene synthase, and dehydrosqualene desaturase encoded by *fpps*, *crtM* and *crtN*, respectively, and the disruption of the *yisP* gene positively affected C_{30} carotenoid production in strain BsMN6, which was up to 6-fold higher compared to the control strain BsMN0 (Fig. 2; Table 2).

Stability of BsMN6 in C_{30} carotenoid production and its cultivation in industrial fermentation medium

The stability of C_{30} carotenoid production in strain BsMN6 without antibiotic selection was tested. An overnight culture of BsMN6 in TSB was diluted 1:1000 in the same medium. The cells were grown in shake flasks at 37 °C to the stationary phase and diluted again 1000-fold. This was repeated five times and in the last transfer, when the stationary phase was reached, the strain was cultured again in TSB and the C_{30} carotenoid yield was determined. As shown in Fig. 4a, BsMN6 produced similar levels of C_{30} carotenoids for at least 50 generations (every round of growth to stationary phase corresponds to about ten generations without antibiotic supplementation, calculated by dividing the length of the exponential growth phase (about 300 min) by the doubling time of BsMN6 (approximately 30 min) in TSB medium),

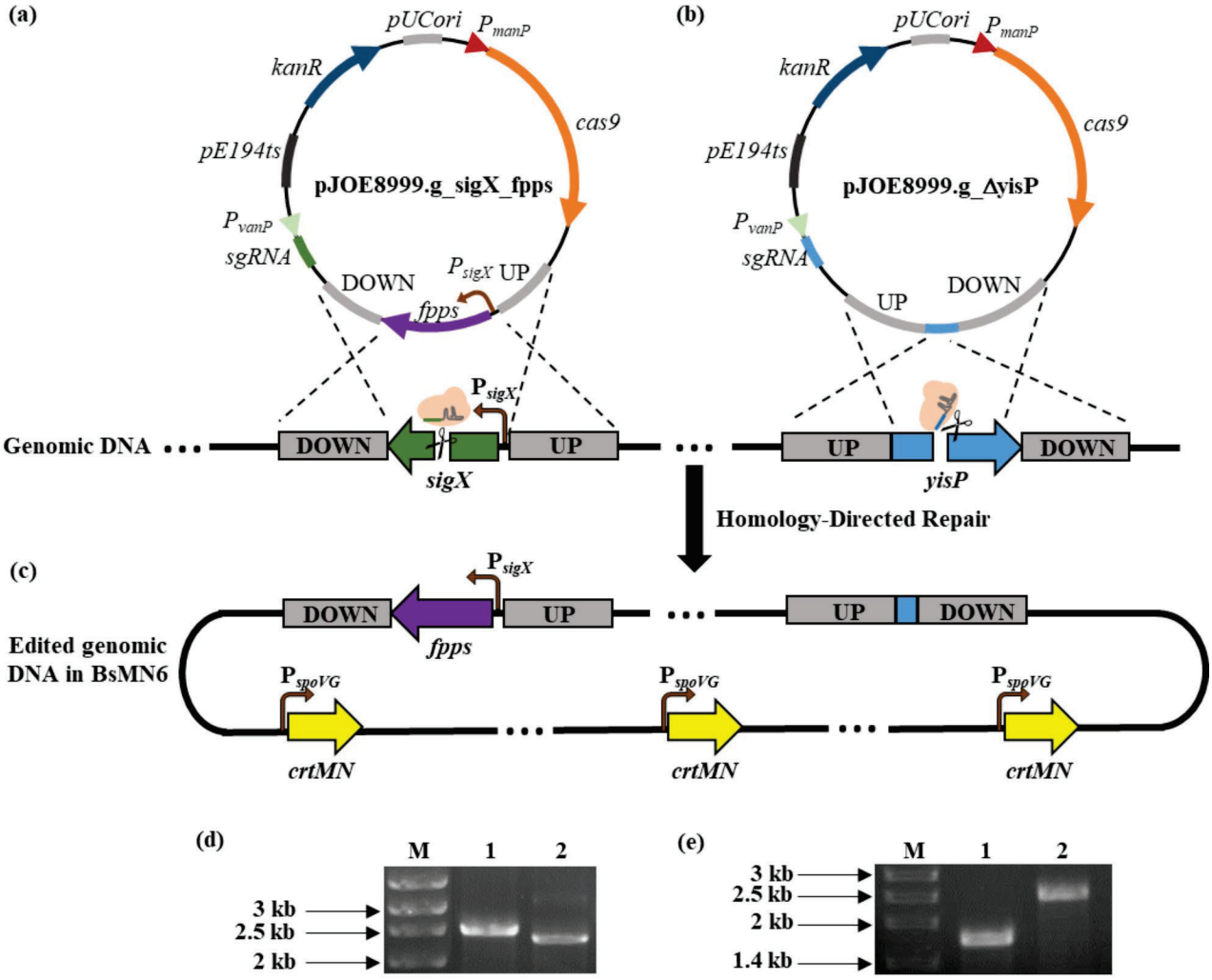


Fig. 3 Engineering of the genome-integrated farnesyl diphosphate synthase (FPPS) and disruption of farnesyl diphosphate phosphatase (YisP) in *B. subtilis*. (**a**) pJOE8999.g_sigX_fpps was designed to allow the replacement of the *sigX* gene from *B. subtilis* by the *fpps* gene from *B. megaterium* under the control of a strong promoter P_{sigX}. (**b**) pJOE8999.g_ΔyisP was constructed for the disruption of the *yisP* gene from *B. subtilis* (**c**) Upon transformation, the resulting *B. subtilis* strain harboring both genomic modifications along with three gene-copies of the *crtMN* genes under the control of the constitutive promoter P_{sigX} was designated as BsMN6. (**d**) Confirmation of the *sigX* gene replacement by *fpps* in the BsMN5 strain. Lane 1 corresponds to an amplification band of 2.5 kb using primers P2F/P4R to verify *fpps* integration at the *sigX* locus site in BsMN5. Lane 2 corresponds to an amplification band of 2.25 kb using the same primers in recipient strain BsMN3. M corresponds to the molecular marker weight. (**e**) Confirmation of the *yisP* gene disruption in strain BsMN6. Lane 1 corresponds to an amplification band of 1.75 kb using primers P5F/P6R to verify *yisP* deletion in BsMN6. Lane 2 corresponds to an amplification band of 2.5 kb using the same primers in recipient strain BsMN5. M corresponds to the molecular marker weight

demonstrating that BsMN6 achieved a high yield of C_{30} carotenoids with stable productivity.

To date, recombinant production of C_{30} carotenoids in *B. subtilis* has been exclusively tested by culturing engineered strains in TSB medium at the shake flask level (Yoshida et al. 2009; Xue et al. 2015; Abdallah et al. 2020). However, TSB is a nutritious medium designed to support the growth of a wide variety of microorganisms, and inappropriate for *B. subtilis* fermentation on an industrial scale due to its high cost. We therefore decided to investigate the capacity of strain BsMN6 to accumulate C_{30} carotenoids in BS1, a commonly used industrial bacterial feed (Wenzel et al. 2011). To this end, BsMN6 was cultured for 24 h in TSB and BS1 media before analyzing DCW and C_{30} carotenoid production. As shown in Fig. 4b and c; Table 2, BsMN6 was able to double the cell biomass concentration when grown in BS1 medium (2.98±0.14 g/L culture) compared to the same strain growing in TSB medium (1.47±0.08 g/L culture). Although the yield of C_{30} carotenoids obtained in TSB (4.42±0.19 mg/g DCW) was higher compared to BS1 medium (3.20±0.24 mg/g DCW), the titer of C_{30} carotenoids obtained in the latter was 40% higher than the titer obtained in TSB, reaching a value of 9.11±0.36 mg/L C_{30}

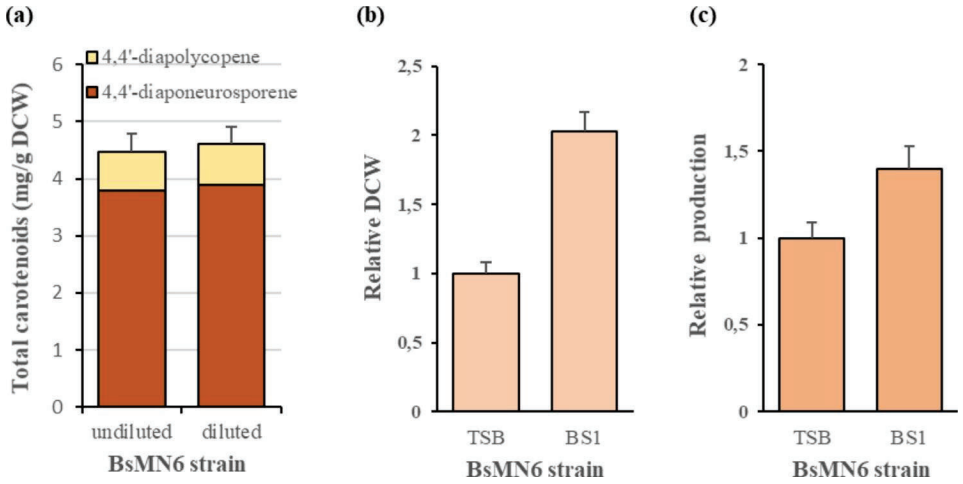


Fig. 4 Stability and C_{30}carotenoid production in strain BsMN6. (**a**) C_{30} carotenoid production in the BsMN6 strain diluted 1000-fold and grown to the stationary phase, repeated 5 times, without antibiotics in TSB media. (**b**) Relative DCW and (**c**) relative titers of C_{30} carotenoids produced by strain BsMN6 cultured in TB and BS1 media, after 24 h of fermentation. The error bars represent the average ± standard deviation of three biological replicates

carotenoids. This indicates that BS1 medium can stimulate cell growth and had a significantly positive effect on the C_{30} carotenoid titer in comparison with TSB.

Discussion

The market demand for carotenoids is continuing to grow due to their antioxidant, anti-inflammatory, and anticancer properties. In particular, the biotechnological production of carotenoids to replace artificial pigments is rapidly gaining interest, despite technological, economic, and legislative limitations. *E. coli* and *B. subtilis* strains have been engineered to accumulate C_{30} carotenoids utilizing suitable expression vectors for relevant *crtMN* genes, the overexpression of MEP pathway enzymes, and the concomitant use of antibiotic drugs and plasmids. However, the current trend in industrial bioprocesses is to circumvent the use of antibiotic selection markers by developing marker-free production systems due to concerns derived from the massive overuse of antibiotics. In many areas of biotechnology, restrictions on antibiotic usage have been imposed by regulatory authorities (Mingon et al., 2015). In the present work, we constructed a plasmid-less, marker-free strain of *B. subtilis*, a bacterium that can naturally produce C_{30} carotenoids in the absence of any inducer or antibiotic compound. Optimization steps involving *crtMN* gene dosage and an enhanced supply of the precursor FPP were carried out using the CRISPR-Cas9 system, resulting in the generation of an efficient, safe, and stable C_{30} carotenoid-producing *B. subtilis* strain.

Reliance on the use of plasmids and antibiotic selection markers constitutes a major limiting factor for the implementation of an optimal *B. subtilis* chassis able to execute the functions needed for efficient C_{30} carotenoid production. To bypass this limitation, an interesting option is to maintain the cloned genes by genome integration, thus ensuring high stability in the absence of antibiotic selection pressure. Nevertheless, the main drawback of this approach is that the resulting strains have a low gene dosage unless multiple gene copies are integrated into the genome (Yomantas et al. 2011; Huang et al. 2017; Wang et al. 2004), until reaching expression levels comparable to those of cells carrying multiple copies of a recombinant plasmid. Our study clearly shows that the low copy number pHY_crtMN plasmid (5–15 per cell), a derivative of pHY_300PLK (Ishiwa and Shibahara 1985), is an unfavorable vector for maximizing *crtMN* gene expression. We hypothesize that the reason for the low expression achieved is that *crtMN* genes are the second and third genes transcribed from the promoter of the tetracycline resistant gene (Isamu Maeda personal communication). Within an operon, the expression of a gene at the first position is expected to be higher compared to the gene at the second position, which in turn should be more expressed than a gene at the third position (Lim et al. 2011). In contrast, C_{30} carotenoid production in cells carrying multiple copies of the xylose-inducible medium copy number pBS0E_crtMN plasmid (15–25 per cell) was significantly improved; more importantly, its performance was comparable to the plasmid-less strain harboring three *crtMN* gene copies in the chromosome. Presumably, when these conditions occur, increasing the copy number no longer enhances expression levels (Widner et al. 2000) and the potential bottlenecks in C_{30} carotenoid production rely on the expression of other rate-limiting enzymes in the biosynthetic pathway. Notably, the insertion of three *crtMN* gene copies into the *B. subtilis* chromosome debottlenecked an unexplored rate-limiting step in the C_{30} carotenoid biosynthetic route and at the same time alleviated the need for antibiotic

selection for plasmid maintenance. Moreover, its stability and potential ecological safety suggests that the engineered *B. subtilis* strain has great promise as an efficient C_{30} carotenoid cell factory with practical application in industrial settings (García-Moyano et al. 2020; Su et al. 2020).

To further improve the *B. subtilis* carotenoid production capacity, we focused on modulating some of the well-recognized regulatory elements that tightly control the metabolic flux to C_{30} carotenoid biosynthesis from the universal precursors DMAPP and IPP. Specifically, our aim was to enhance the FPP pool and also ameliorate its consumption by removing the competing pathway yielding farnesol. The first attempt to overexpress the *fpps* gene from *B. megaterium* resulted in a significant improvement (1.46-fold) of C_{30} carotenoid production. This result is in accordance with a previous study that achieved 1.36-fold higher carotenoid yields by introducing an extra copy of the homologous *fpps* gene from *B. subtilis* (*ispA*) (Xue et al. 2015). The additional expression of the *fpps* gene from *Saccharomyces cerevisiae* also increased the supply of the precursor FPP (Song et al. 2021). We therefore conclude that the heterologous expression of FPPS from *B. megaterium* increased C30 carotenoid biosynthesis in *B. subtilis*, similarly to the values obtained when an extra copy of the native IspA was overexpressed (Xue et al. 2015). It has also been reported that attenuation of a competing FPP-consuming pathway toward C55 heptaprenyl diphosphate contributed to a 1.15-fold increase in terpenoid synthesis (Song et al. 2021). Accordingly, we assumed that abolishing non-essential expression of *yisP*, the only phosphatase that catalyzes the conversion of FPP to farnesol, would also lead to less FPP consumption in this competing pathway, and the resulting extra FPP could be used by CrtMN enzymes to increase C_{30} carotenoid yield. In the Δ*yisP* mutant, known to exhibit no FPP phosphatase activity (Feng et al. 2014), excess FPP was distributed to increase the carotenoid yield in the engineered strain 1.39-fold (Fig. 2; Table 2). Thus, for the first time, the role of *yisP* knockout in an increased accumulation of C_{30} carotenoids in *B. subtilis* was demonstrated.

Cell engineering techniques have been previously used to improve C_{30} carotenoid productivity in *E. coli* and *B. subtilis*. *E. coli* strains were engineered to accumulate C_{30} carotenoids, with production levels ranging from 0.5 mg/ gDCW to 10.8 mg/L (Chae et al. 2010; Kim et al. 2010, 2022; Takemura et al. 2021). *B. subtilis* has also been engineered using two-plasmid systems comprising pHY_crtMN (Yoshida et al. 2009), mediating *crtMN* gene overexpression, and xylose-inducible pHCMC04G (Xue et al. 2015), mediating stable overexpression of all MEP pathway enzymes. In total, the yield of C_{30} carotenoids achieved was 21 mg/g DCW, the highest production in *B. subtilis* reported to date (Abdallah et al. 2020). In the present study, the combination of chromosomal overexpression of farnesyl diphosphate synthase, dehydrosqualene synthase and dehydrosqualene desaturase encoded by *fpps*, *crtM* and *crtN*, respectively, with the simultaneous disruption of the *yisP* gene, resulted in a titer of 9.11 mg/L C_{30} carotenoids, and a yield of 4.42 mg/g DCW. Although the C_{30} carotenoid accumulation is similar to that achieved in *E. coli* strains and lower (4.7-fold) than in *B. subtilis* overexpressing the eight enzymes of the MEP pathway, it should be noted that we only focused on improving the last three steps downstream of the MEP pathway. Consequently, one could expect that combining both strategies would serve to obtain a superior productive strain. Additionally, we demonstrated that routinely used industrial bacterial feed (antibiotic- and xylose-inducer-free) may provide a cost-effective bioprocess for the industrial production of C_{30} carotenoids. In a nutshell, taking advantage of its inherent capacity to synthesize C_{30} carotenoids, we have developed a plasmid-less, marker-free, *B. subtilis* strain that can serve as a stepping stone for further genetic engineering and fermentation process optimization targeted at a sustainable and efficient production of C_{30} carotenoids.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13568-023-01542-x.

Supplementary Material 1

Acknowledgements
We thank Dr. Isamu Maeda for providing us with the plasmid pHY_crtMN.

Author contribution
PP designed research. OF and JF conducted experiments. PP analyzed the data. OF, JF and PP wrote the manuscript. All authors read and approved the manuscript.

Funding information
This work was supported by the *Pla de Doctorats Industrials del Departament de Recerca i Universitats de la Generalitat de Catalunya* and *Gestió d' Ajuts Universitaris de Recerca* for grant number 2021 DI 77.

Data availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest
The authors declare no financial or commercial conflict of interest.

Ethical statement
This article does not describe any studies with human participants or animals performed by any of the authors.

Received: 5 February 2023 / Accepted: 5 April 2023
Published online: 29 April 2023

References

Abdallah II, Xue D, Pramastya H, van Merkerk R, Setroikromo R, Quax WJ (2020) A regulated synthetic operon facilitates stable overexpression of multigene terpenoid pathway in *Bacillus subtilis*. J Ind Microbiol Biotechnol 47(2):243–249. https://doi.org/10.1007/s10295-019-02257-4

Altenbuchner J (2016) Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 System. Appl Environ Microbiol 15(17):5421–5427. https://doi.org/10.1128/AEM.01453-16

Britton G, Liaaen-Jensen S, Pfander H (2004) Carotenoids Handbook. Birkhäuser Verlag

Cardoso LAC, Karp SG, Vendruscolo F, Kanno KYF, Zoz LIC, Carvalho JC (2017) Biotechnological Production of Carotenoids and their applications in Food and Pharmaceutical Products. Carotenoids. InTech. doi:https://doi.org/10.5772/67725

Chae HS, Kim K-H, Kim SC, Lee PC (2010) Strain-dependent carotenoid productions in metabolically engineered *Escherichia coli.* Appl Biochem Biotechnol 162(8):2333–2344. https://doi.org/10.1007/s12010-010-9006-0

Earl AM, Losick R, Kolter R (2008) Ecology and genomics of *Bacillus subtilis*. Trends Microbiol 16(6):269–275. https://doi.org/10.1016/j.tim.2008.03.004

Feng X, Hu Y, Zheng Y, Zhu W, Li K, Huang CH, Ko TP, Ren F, Chan HC, Nega M, Bogue S, López D, Kolter R, Götz F, Guo RT, Oldfield E (2014) Structural and functional analysis of *Bacillus subtilis* YisP reveals a role of its product in biofilm production. Chem Biol 21(11):1557–1563. https://doi.org/10.1016/j.chembiol.2014.08.018

Ferrando J, Filluelo O, Zeigler DR, Picart P (2023) Barriers to simultaneous multilocus integration in *Bacillus subtilis* tumble down: development of a straightforward screening method for the colorimetric detection of one-step multiple gene insertion using the CRISPR-Cas9 system. Microb Cell Fact 22:21. https://doi.org/10.1186/s12934-023-02032-2

Furubayashi M, Li L, Katabami A, Saito K, Umeno D (2014) Construction of carotenoid biosynthetic pathways using squalene synthase. FEBS Lett 588(3):436–442. https://doi.org/10.1016/j.febslet.2013.12.003

García-Moyano A, Larsen Ø, Gaykawad S, Christakou E, Boccadoro C, Puntervoll P, Bjerga GEK (2020) Fragment Exchange plasmid tools for CRISPR/Cas9-Mediated gene integration and protease production in *Bacillus subtilis*. Appl Environ Microbiol 87(1):e02090–e02020. https://doi.org/10.1128/AEM.02090-20

Guan Z, Xue D, Abdallah II, Dijkshoorn L, Setroikromo R, Lv G, Quax WJ (2015) Metabolic engineering of *Bacillus subtilis* for terpenoid production. Appl Microbiol Biotechnol 99(22):9395–9406. https://doi.org/10.1007/s00253-015-6950-1

Hartz P, Milhim M, Trenkamp S, Bernhardt R, Hannemann F (2018) Characterization and engineering of a carotenoid biosynthesis operon from *Bacillus megaterium*. Metab Eng 49:47–58. https://doi.org/10.1016/j.ymben.2018.07.017

Huang K, Zhang T, Jiang B, Yan X, Mu W, Miao M (2017) Overproduction of Rummeliibacillus pycnus arginase with multi-copy insertion of the argr-pyc 23 cassette into the *Bacillus subtilis* chromosome. Appl Microbiol Biotechnol 101:6039–6048. https://doi.org/10.1007/s00253-017-8355-9

Ishiwa H, Shibahara H (1985) New shuttle vectors for *Escherichia coli* and *Bacillus subtilis*. Jpn J Genet 60:235–243. https://doi.org/10.1016/0378-1119(84)90161-6

Jing Y, Liu H, Xu W, Yang Q (2017) Amelioration of the DSS-induced colitis in mice bypretreatment with 4,4′-diaponeurosporene-producing *Bacillus subtilis*. Exp Ther Med 14(6):6069–6073. https://doi.org/10.3892/etm.2017.5282

Jing Y, Liu H, Xu W, Yang Q (2019) 4,4′-Diaponeurosporene-producing *Bacillus subtilis* promotes the development of the mucosal immune system of the piglet gut. Anat Rec (Hoboken) 302:1800–1807. https://doi.org/10.1002/ar.24102

Julsing MK, Rijpkema M, Woerdenbag HJ, Quax WJ, Kayser O (2007) Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis*. Appl Microbiol Biotechnol 75(6):1377–1384. https://doi.org/10.1007/s00253-007-0953-5

Kim J, Kong MK, Lee SY, Lee PC (2010) Carbon sources-dependent carotenoid production in metabolically engineered *Escherichia coli*. World J Microbiol Biotechnol 26(12):2231–2239. https://doi.org/10.1007/s11274-010-0408-5

Kim M, Jung DH, Hwang CY, Siziya IN, Park YS, Seo MJ (2022) 4,4′-Diaponeurosporene production as C_{30} carotenoid with antioxidant activity in recombinant *Escherichia coli*. Appl Biochem Biotechnol Sep 6. https://doi.org/10.1007/s12010-022-04147-5

Kuzma J, Nemecek-Marshall M, Pollock WH, Fall R (1995) Bacteria produce the volatile hydrocarbon isoprene. Curr Microbiol 30(2):97–103. https://doi.org/10.1007/BF00294190

Lim HN, Lee Y, Hussein R (2011) Fundamental relationship between operon organization and gene expression. Proc Natl Acad Sci U S A 108(26):10626–10631. https://doi.org/10.1073/pnas.1105692108

Liu H, Xu W, Chang X, Qin T, Yin Y, Yang Q (2016) 4,4′-diaponeurosporene, a C_{30} carotenoid, effectively activates dendritic cells via CD36 and NF-kappaB signaling in a ROS independent manner. Oncotarget 7(27):40978–40991. https://doi.org/10.18632/oncotarget.9800

Liu H, Xu W, Yu Q, Yang Q (2017) 4,4′-Diaponeurosporene-producing *Bacillus subtilis* increased mouse resistance against *Salmonella typhimurium* infection in a CD36-Dependent manner. Front Immunol 8:483. https://doi.org/10.3389/fimmu.2017.00483

Marshall JH, GJ Wilmoth (1981) Pigments of *Staphylococcus aureus*, a series of triterpenoid carotenoids. J Bacteriol 147:900–913. https://doi.org/10.1128/jb.147.3.900-913.1981

Mignon C, Sodoyer R, Werle B (2015) Antibiotic-free selection in biotherapeutics: now and forever. Pathogens 4(2):157–181. https://doi.org/10.3390/pathogens4020157

Moser S, Pichler H (2019) Identifying and engineering the ideal microbial terpenoid production host. Appl Microbiol Biotechnol 103(14):5501–5516. https://doi.org/10.1007/s00253-019-09892-y

Pelz A, Wieland KP, Putzbach K, Hentschel P, Albert K, Götz F (2005) Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*. J Biol Chem 280:32493–32498. https://doi.org/10.1074/jbc.M505070200

Popp PF, Dotzler M, Radeck J, Bartels J, Mascher T (2017) The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. Sci Rep 7(1):15058. https://doi.org/10.1038/s41598-017-15107-z

Schallmey M, Singh A, Ward OP (2004) Developments in the use of *Bacillus* species for industrial production. Can J Microbiol 50:1–17. https://doi.org/10.1139/w03-076

Siziya IN, Hwang CY, Seo MJ (2022) Antioxidant potential and capacity of microorganism-sourced C_{30} Carotenoids-A review. Antioxid (Basel) 11(10):1963. https://doi.org/10.3390/antiox11101963

Song Y, Nikoloff JM, Fu G, Chen J, Li Q, Xie N, Zheng P, Sun J, Zhang D (2016) Promoter screening from *Bacillus subtilis* in various Conditions Hunting for Synthetic Biology and Industrial Applications. PLoS ONE 11(7):e0158447. https://doi.org/10.1371/journal.pone.0158447

Song Y, He S, Abdallah II, Jopkiewicz A, Setroikromo R, van Merkerk R, Tepper PG, Quax WJ (2021) Engineering of multiple modules to improve Amorphadiene production in *Bacillus subtilis* using CRISPR-Cas9. J Agric Food Chem 69(16):4785–4794. https://doi.org/10.1021/acs.jafc.1c00498

Su Y, Liu C, Fang H, Zhang D (2020) *Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine. Microb Cell Fact 19(1):173. https://doi.org/10.1186/s12934-020-01436-8

Takaichi S (2000) Characterization of carotenes in a combination of a C(18) HPLC column with isocratic elution and absorption spectra with a photodiode-array detector. Photosynth Res 65(1):93. https://doi.org/10.1023/A:1006445503030

Takaichi S, Inoue K, Akaike M, Kobayashi M, Oh-oka H, Madigan MY (1997) The major carotenoid in all known species of *heliobacteria* is the C_{30} carotenoid 4,4′-diaponeurosporene, not neurosporene. Arch Microbiol 168:277–281. https://doi.org/10.1007/s002030050499

Takemura M, Takagi C, Aikawa M, Araki K, Choi SK, Itaya M, Shindo K, Misawa N (2021) Heterologous production of novel and rare C_{30}-carotenoids using *Planococcus* carotenoid biosynthesis genes. Microb Cell Fact 20(1):194. https://doi.org/10.1186/s12934-021-01683-3

Toymentseva AA, Schrecke K, Sharipova MR, Mascher T (2012) The LIKE system, a novel protein expression toolbox for *Bacillus subtilis* based on the *lial* promoter. Microb. Cell Fact 11:143. doi: https://doi.org/10.1186/1475-2859-11-143

Wagner WP, Helmig D, Fall R (2000) Isoprene biosynthesis in *Bacillus subtilis* via the methylerythritol phosphate pathway. J Nat Prod 63:37–40. https://doi.org/10.1021/np990286p

Wang J-J, Rojanatavorn K, Shih JCH (2004) Increased production of *Bacillus* keratinase by chromosomal integration of multiple copies of the *kerA* gene. Biotechnol Bioeng 87:459–464. https://doi.org/10.1002/bit.20145

Wenzel M, Müller A, Siemann-Herzberg M, Altenbuchner J (2011) Self-inducible *Bacillus subtilis* expression system for reliable and inexpensive protein production by high-cell-density fermentation. Appl Environ Microbiol 77(18):6419–6425. https://doi.org/10.1128/AEM.05219-11

Widner B, Thomas M, Sternberg D, Lammon D, Behr R, Sloma A (2000) Development of marker-free strains of *Bacillus subtilis* capable of secreting high levels of industrial enzymes. J Ind Microbiol Biotech 25:204–212. https://doi.org/10.1038/sj.jim.7000051

Wieland B, Feil C, Gloria-Maercker E, Thumm G, Lechner M, Bravo JM, Poralla K, Goëtz F (1994) Genetic and biochemical analyses of the biosynthesis of the

yellow carotenoid 4,40 -diaponeurosporene of *Staphylococcus aureus*. J Bacteriol 176:7719–7726. https://doi.org/10.1016/S0021-9258(19)39076-3

Wu G, Yan Q, Jones JA, Tang YJ, Fong SS, Koffas MAG (2016) Metabolic burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. Trends Biotechnol 34(8):652–664. https://doi.org/10.1016/j.tibtech.2016.02.010

Xue D, Abdallah II, de Haan IE, Sibbald MJ, Quax WJ (2015) Enhanced C_{30} carotenoid production in *Bacillus subtilis* by systematic overexpression of MEP pathway genes. Appl Microbiol Biotechnol 99(14):5907–5915. https://doi.org/10.1007/s00253-015-6531-3

Yasbin RE, Wilson GA, Young FE (1975) Transformation and transfection in lysogenic strains of *Bacillus subtilis*: evidence for selective induction of prophage in competent cells. J Bacteriol 121:296–304. https://doi.org/10.1128/jb.121.1.296-304.1975

Yomantas YA, Abalakina EG, Golubeva LI, Gorbacheva LY, Mashko SV (2011) Overproduction of *Bacillus amyloliquefaciens* extracellular glutamylendopeptidase as a result of ectopic multi-copy insertion of an efficiently expressed *mpr* gene into the *Bacillus subtilis* chromosome. Microb Cell Fact 10:64–73. https://doi.org/10.1186/1475-2859-10-64

Yoshida K, Ueda S, Maeda I (2009) Carotenoid production in *Bacillus subtilis* achieved by metabolic engineering. Biotechnol Lett 31(11):1789–1793. https://doi.org/10.1007/s10529-009-0082-6

Zhang C, Hong K (2020) Production of terpenoids by Synthetic Biology Approaches. Front Bioeng Biotechnol 8:347. https://doi.org/10.3389/fbioe.2020.00347

Zhao J, Li Q, Sun T, Zhu X, Xu H, Tang J, Zhang X, Ma Y (2013) Engineering central metabolic modules of *Escherichia coli* for improving beta-carotene production. Metab Eng 17:42–50. https://doi.org/10.1016/j.ymben.2013.02.002

Publisher's Note

Metabolic engineering of *Bacillus subtilis* toward the efficient and stable production of C_{30}-carotenoids

Oriana Filluelo, Jordi Ferrando and Pere Picart

SUPPLEMENTARY MATERIAL

Figure S1: HPLC chromatogram and absorption spectra of C_{30} carotenoids

Table S1: Primers designed in this study

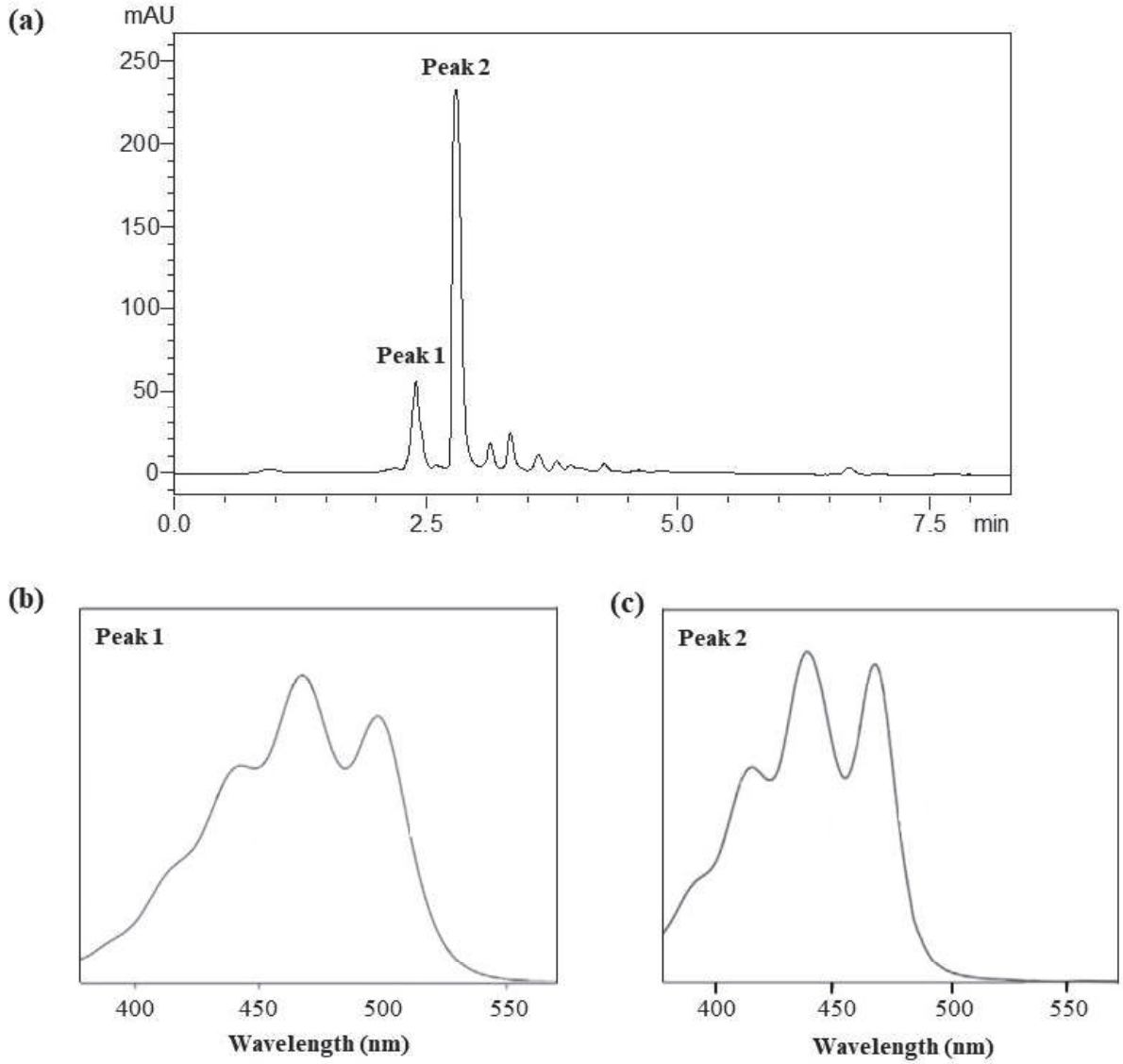


Figure S1. Analysis of C_{30} carotenoids produced by engineered *B. subtilis* strains. a) Chromatogram of C_{30} carotenoids separated by HPLC. The quantity of each C_{30} carotenoid was determined by comparing HPLC profiles peak areas to that of a β-carotene standard curve. The two major peaks correspond to

4,4'-diapolycopene (peak 1) and 4,4'-diaponeurosporene (peak 2) at 450 nm. b) UV/Vis absorption spectra for compound corresponding to individual peak 1 (4,4'-diapolycopene) and c) UV/Vis absorption spectra for compound corresponding to individual peak 2 (4,4'-diaponeurosporene).

Table S1. Primers designed in this study*.

Name	Sequence	Purpose
		PCR for *crtMN* genes cloning into pBS0E
P1F	5′ – CCGGAATTCTAAGGAGGAAAAACATATGACAATGATGGATATGAATTT	PCR product containing *crtMN* genes flanked by *EcoRI* and *SpeI*
P1R	5′ – GGACTAGTTTATACGCCCCGCTCAAT	
TS1F	5′ – TACGTACATTTGACTGGGATACAC	**Target sequence for *sigX* gene replacement**
TS1R	5′ – AAACGTGTATCCCAGTCAAATGTA	
		PCR of homology template for *sigX* replacement for *fpps*
P2F	5′ – AA**GGCCAACGAGGCC**AGTCATGAGCTGAGAACAC	upstream *sigX* gene homologous arm
P2R	5′ – CATTTGAAACCCCTCCGTT	
P3F	5′ – AACGGAGGGGTTTCAAATGGTGAATGAAATGAATTTAAAAGAG	*fpps* gene amplification
P3R	5′ – TGAGGCGAACGATGGTCTTTAACTATCACGTTTTGCTATATA	
P4F	5′ – AGACCATCGTTCGCCTCA	downstream *sigX* gene homologous arm
P4R	5′ – AA**GGCCTTATTGGCC**AATCATCAACTTCTGACTCC	
TS2F	5′ – TACGCGGTAAACAACATCTATGGA	**Target sequence for *YisP* gene deletion**
TS2R	5′ – AAACTCCATAGATGTTGTTTACCG	
		PCR of homology template for *YisP gene* deletion
P5F	5′ – AA**GGCCAACGAGGCC**AGAAGGATAGATGAGCAGGG	upstream *yisP* gene homologous arm
P5R	5′ – CTCATTGCTGGTAAGCTTCTTT	
P6F	5′ – AAAGAAGCTTACCAGCAATGAGAAAACACCCGGCCTTGATTGA	downstream *yisP* gene homologous arm
P6R	5′ – AA**GGCCTTATTGGCC**TGGCCATGTCCACATCTATC	

* *Sfi*I restriction sites are indicated in **bold**. *Eco*RI restriction site is underlined and S*pe*I restriction site is doble underlined

5 General discussion

5. General discussion

Biocatalysis has emerged as a key driver of the bio-based economy, with enzymes increasingly integrated into industrial processes due to their environmental and economic advantages over traditional chemical catalysts (Molina-Espeja *et al.*, 2023). However, in the PPI, enzymes must retain activity under harsh conditions, including high temperatures, alkaline pH, and prolonged processing times, which often hinder their performance (Yang *et al.*, 2023). Therefore, the discovery and development of robust novel enzymes with enhanced stability and catalytic efficiency under such conditions is critical to advancing greener, safer, and more sustainable industrial operations (Molina-Espeja *et al.*, 2023). This thesis presents four interconnected papers that collectively advance biocatalysis by bridging enzyme discovery with host strain engineering. It begins with the identification of industrially robust enzymes suited to demanding process conditions and progresses toward the development of optimized microbial platforms for their efficient, high-level production (**Figure 10**). By integrating targeted enzyme selection with rational host design, these contributions provide a practical and scalable framework for enabling more sustainable, energy-efficient, and economically viable bioprocesses in industrial biotechnology.

In **Chapter 1** of this thesis, a novel EG candidate was identified from the publicly available genome of *Stachybotrys chartarum* IBT 7711, an ascomycete commonly associated with cellulosic substrates (Ibrahim *et al.*, 2022). This organism is known to produce a cellulolytic system with reported alkali resistance and thermostability, traits highly desirable for biocatalysts used in PPI applications (Picart *et al.*, 2008). Among five putative cellulases retrieved from genomic data, the gene designated *stachCel5* (NCBI locus tag: S7711_03103) was selected based on its modular architecture, which is typical of efficient fungal cellulases. Domain analysis revealed a configuration consisting of an N-terminal cellulose-binding domain (CBD) and a C-terminal GH5 catalytic domain, an arrangement frequently associated with high endo-β-1,4-glucanase activity and substrate affinity (Neis & da Silva Pinto, 2021). These structural features support its potential as a promising biocatalyst for PPI processes requiring enzymatic resilience in alkaline and thermophilic environments (M. Yang *et al.*, 2023).

Additionally, the predicted N-terminal SP in StachCel5 supports its secretion, consistent with a functional role in extracellular polysaccharide degradation. Phylogenetic analysis placed StachCel5 in close proximity to EGs from *Clonostachys* and *Paramyrothecium* species, indicating potential evolutionary conservation of function. Sequence identity of 60-65% with these homologs, along with strong BLASTp alignment coverage with experimentally characterized GH5

EGs from *Aspergillus niger*, *Thermoascus aurantiacus*, and *Penicillium verruculosum* reinforced its classification as a member of GH5. Structural modeling with AlphaFold2 and superposition onto the crystal structure of the *A. niger* GH5 EG (PDB ID: 5I77) confirmed conservation of the canonical $(\beta/\alpha)_8$ TIM-barrel fold and the spatial arrangement of the catalytic residues. Sequence alignment identified Glu^{228} and Glu^{335} as the putative catalytic residues, aligning with known catalytic motifs of GH5 enzymes (Drula *et al.*, 2022). Notably, the inter-residue distance of 4.9 Å between the catalytic glutamates suggests that StachCel5 likely operates via a retaining mechanism, consistent with the catalytic behavior of most GH5 EGs (Planas, 2023).

Building on the selection of StachCel5 as a promising EG candidate, *K. phaffii* was employed as the heterologous expression system in **Paper 1**. While prokaryotic hosts such as *Escherichia coli* and *B. subtilis* are widely used for recombinant protein production due to their fast growth rates, well-characterized genetics, and cost-effective cultivation, they present notable limitations for the expression of complex eukaryotic proteins, particularly fungal enzymes (Schütz *et al.*, 2023). In the case of *B. subtilis*, which is the primary expression host optimized throughout this thesis (**Chapters 2** and **3**), its secretion capacity and GRAS or QPS status make it a valuable industrial platform. However, it lacks the PTM machinery (such as glycosylation and disulfide bond formation) required for the proper folding, stability, and activity of many eukaryotic enzymes. *E. coli* shares similar constraints and often forms inclusion bodies when expressing complex proteins, requiring additional refolding steps that can compromise activity and yield. In contrast, eukaryotic hosts such as yeasts and filamentous fungi offer more suitable expression environments for fungal enzymes due to their ability to perform essential PTMs and facilitate correct protein folding and secretion. *K. phaffii*, in particular, combines the benefits of high cell density growth, strong promoters, and efficient secretion pathways, while maintaining relatively low production costs compared to higher eukaryotes (Schütz *et al.*, 2023; K. Zhang *et al.*, 2020). Its ability to produce active, properly folded fungal enzymes renders it a strategic choice for the initial biochemical characterization of StachCel5. Therefore, *K. phaffii* was used in **Paper 1** to enable functional expression and robust characterization of StachCel5 under conditions that better reflect its native folding requirements. In contrast**, Papers 2, 3,** and **4** focused on adapting *B. subtilis* to express more scalable, prokaryote-compatible enzyme variants suitable for industrial deployment.

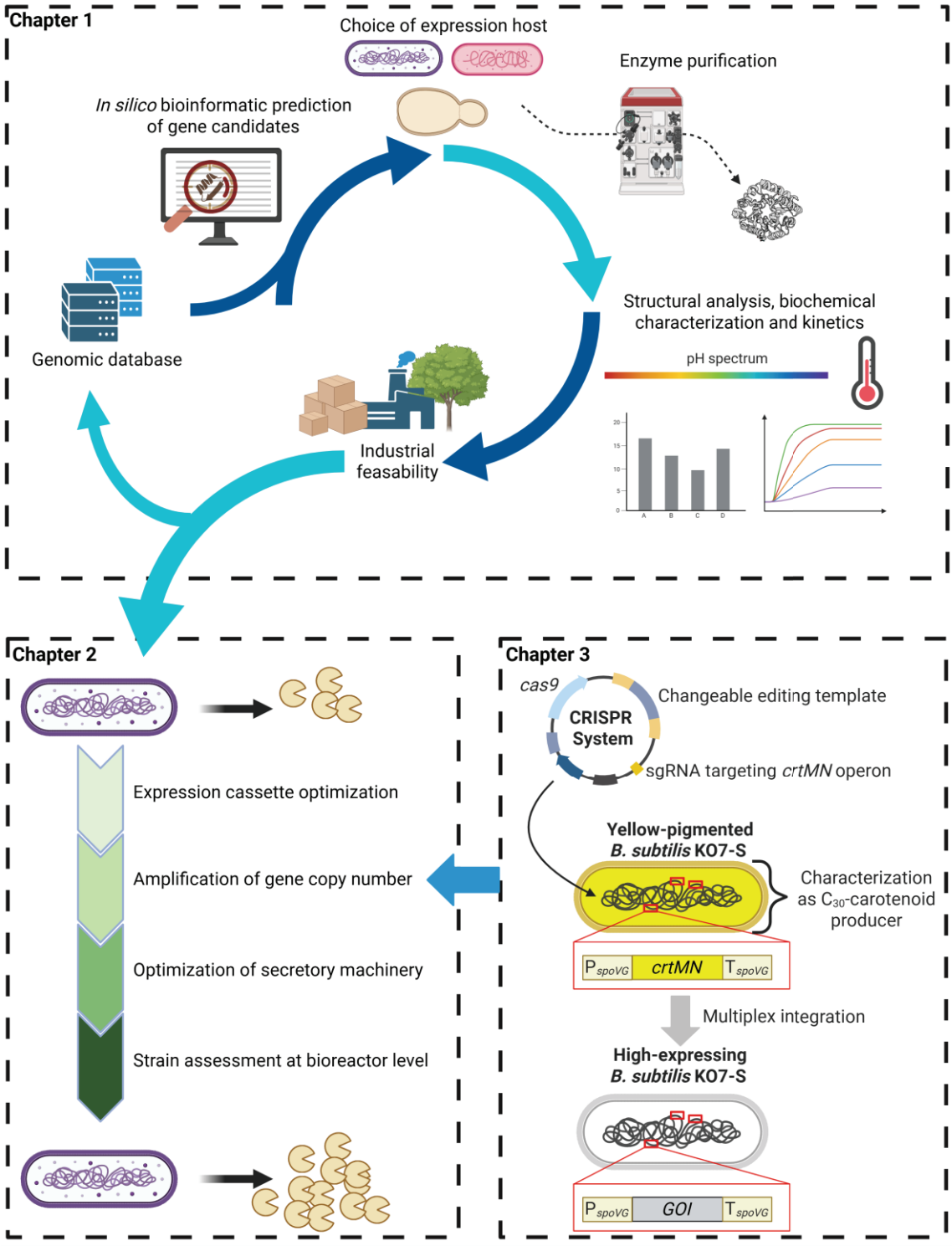


Figure 10. Overview of the thesis workflow, integrating enzyme discovery, host strain engineering, and new genome editing tools to enable robust and scalable production of industrially relevant enzymes. A novel endoglucanase, StachCel5, was identified through database mining and selected for its high stability and performance in pulp refining (Chapter 1). To enable high-level enzyme production, a modular *B. subtilis* platform was developed via promoter tuning, gene amplification, and secretion enhancement (Chapter 2). A CRISPR-Cas9-based multiplex genome editing system with colorimetric screening was further introduced for rapid, simultaneous gene integration. This colorimetric system also enabled stable production of C_{30} carotenoids, demonstrating its flexibility (Chapter 3). Created with Biorender.

The *stachCel5* gene was codon-optimized for *K. phaffii* and engineered to replace its native SP with the yeast α-factor secretion signal, a commonly used strategy to enhance extracellular secretion in this host (Lv & Cai, 2025). Expression was driven by the constitutive GAP promoter, enabling continuous enzyme production without induction. This approach was effective, as StachCel5 was successfully secreted into the culture medium. Although the theoretical molecular weight of StachCel5 is predicted to be 42.8 kDa, SDS-PAGE analysis of the culture supernatant revealed a prominent band at approximately 55 kDa. This shift is consistent with post-translational glycosylation, a frequent modification in *K. phaffii*-expressed proteins (Bernardi *et al.*, 2019). Supporting this, bioinformatic analysis predicted two putative N-glycosylation sites at Asn^{30} and Asn^{168}.

Although the GH5 family are known for their catalytic versatility, including endo-β-1,4-glucanase, mixed-linkage β-1,3/1,4 glucanase, xylanase, and galactanase activities (Aspeborg *et al.*, 2012), StachCel5 exhibited a remarkably narrow substrate range, with high specificity for β-1,4-glucosidic linkages. StachCel5 demonstrated robust activity, outperforming several fungal homologs except for AnCel5A from *A. niger* (**Table 1**). Notably, no detectable activity was observed on laminarin, a polysaccharide composed solely of β-1,3-glucosidic linkages. This lack of activity implies that StachCel5 acts specifically at β-1,4 linkages, and the enhanced performance on lichenan relative to carboxymethyl cellulose (CMC) may reflect increased substrate accessibility or a favorable conformation due to the mixed-linkage structure. Moreover, StachCel5 showed no activity on a broad range of alternative polysaccharides and model substrates, including xylan, starch, Avicel, filter paper, and synthetic substrates *para*-nitrophenyl-β-D-cellobioside (*p*NPC) and *para*-nitrophenyl-β-D-glucopyranoside (*p*NPG). The absence of activity against these substrates underscores the enzyme's substrate specificity. Collectively, these results support the classification of StachCel5 as a highly specific endo-β-1,4-glucanase targeting amorphous cellulose, and its assignment to the GH5_5 subfamily, which primarily comprises secreted fungal and bacterial enzymes with this activity (EC 3.2.1.4) (Aspeborg *et al.*, 2012; Neis & da Silva Pinto, 2021).

Biochemical characterization revealed that StachCel5 is a robust enzyme, retaining over 87% of its maximal CMCase activity between pH 4 and 7, and approximately 50% activity at alkaline pH 9. This broad pH tolerance surpasses that of many fungal GH5 EGs, enhancing its potential for diverse industrial processes (**Table 1**) (Ferrando *et al.*, 2025; Rodrigues & Odaneth, 2021). The enzyme exhibited optimal activity at pH 5 and 65 °C, consistent with closely related fungal homologues. Kinetic analysis further emphasized its industrial relevance. Although StachCel5 exhibited a moderate substrate affinity (K_m = 9.77 mg/mL) and high maximal velocity (V_{max} =

898.4 U/mg), its turnover number (K_{cat} = 635.7 s^{-1}) and catalytic efficiency (K_{cat}/K_m = 65.4 mL/s/mg) rank it among the more efficient fungal GH5 EGs (**Table 1**). While not as catalytically potent as the highly active AnCel5A from *A. niger* (Zheng *et al.*, 2022), StachCel5 outperformed several fungal homologs, including TrCel5A from *Trichoderma reesei*, a widely used benchmark EG in industrial enzyme cocktails (Neis & da Silva Pinto, 2021), thus highlighting its competitiveness for biomass deconstruction. Further characterization showed that StachCel5 retained activity in the presence of various metal ions and chemical agents. The enzyme remained active with EDTA, indicating metal-independent catalysis, and was unaffected by β-mercaptoethanol, suggesting disulfide bonds are not essential for its activity. However, activity was completely inhibited by SDS, a behavior consistent with other fungal GH5 EGs (Escuder-Rodríguez *et al.*, 2022).

Table 1. Optimal properties comparison of StachCel5 with other fungal endoglucanases of family GH5. All properties are based on carboxymethyl cellulase (CMCase) activity. Not indicated (n.i.). Modified from Ferrando *et al*., 2025.

Enzyme	pH	TM (°C)	Specific activity (U/mg)	K_m (mg/mL)	V_{max} (U/mg)	K_{cat} (s^{-1})	K_{cat}/K_m (ml/s/mg)	References
StachCel5	4-7	65	287.6	9.77	898.4	635.7	65.4	This study
Cel5A	4	70	45	37	82.6	n.i.	n.i.	(Dave *et al.*, 2015)
TrCel5A	5	55	220.6	2.1	220.6	n.i.	n.i.	(Samanta *et al.*, 2012)
AnCel5A	4	70	1,726	7.6	2786	1,625	227	(Jie Zheng *et al.*, 2022)
GH5	4.8	70	302.9	12	556.6	129.4	10.8	(A. Patel & Shah, 2021)
AgCMCase	5	55	175.3	7.2 mM	343.8	n.i.	n.i.	(Z. Li *et al.*, 2018)
*Mt*EG5-1	5	70	n.i.	6.1	91.7	85	13.9	(Zhou *et al.*, 2023)

Given its broad pH and temperature tolerance, high specific activity on cellulose, and resistance to metal ions and chemical agents, StachCel5 emerges as a highly promising EG for pulp modification, where selective fiber treatment is critical for improving paper properties and reducing energy consumption. Beyond its relevance to the PPI, these robust features position StachCel5 as a versatile biocatalyst for a wide range of industrial applications, including food and beverage processing, bioethanol production from lignocellulosic waste, and the development of detergents and textile formulations (Rodrigues & Odaneth, 2021). Accordingly, the final part of this first chapter evaluated its potential as a refining aid in pulp processing. When applied to

pulp refining, StachCel5 significantly improved refining efficiency, particularly in softwood pulps. Treated pulp samples exhibited higher drainability resistance (°SR), while handsheets produced from these pulps showed reduced air permeability and lower light scattering, key indicators of enhanced fiber bonding and sheet density (Ahadian *et al.*, 2023; Gharehkhani *et al.*, 2015; Krysztof *et al.*, 2021). These findings were corroborated by scanning electron microscopy (SEM), which revealed a denser and more uniformly bonded fiber network following enzymatic treatment.

Given that paper must often perform under mechanical stress, assessing its physico-mechanical properties is essential to determine its end-use suitability. Among these, tensile and burst indices are key indicators of paper strength. The tensile index reflects the maximum force a paper can withstand when stretched, while the burst index indicate how much pressure a paper can handle before rupturing (Bhardwaj *et al.*, 2023; Kerekes *et al.*, 2021). Across all refining intensities, StachCel5-treated pulps consistently outperformed untreated controls in both parameters. Notably, equivalent mechanical strength was achieved with fewer PFI revolutions, demonstrating that enzyme pretreatment enabled substantial energy savings. For example, a tensile index of 66.9 Nm/g achieved at 6,000 revolutions in untreated pulp was matched at only 4,500 revolutions (67.8 Nm/g) in enzyme-treated pulp, representing a 25% reduction in refining energy. Importantly, this improved performance was not accompanied by fiber degradation. This was evidenced by the stable tear index, which is strongly influenced by fiber length, and the zero-span tensile index, a direct indicator of intrinsic fiber strength (Zeng *et al.*, 2012). These findings indicate that the low enzyme dosage applied in **Paper 1** selectively modified fiber surfaces without causing excessive hydrolysis, fiber shortening, or fines generation, which could otherwise impair the physico-mechanical properties of the handsheets (Park *et al.*, 2025). Interestingly, the refining improvement was more pronounced in softwood than hardwood pulps, likely due to the higher hemicellulose content in hardwood, which may limit enzyme accessabiliy to cellulose (Haske-Cornelius *et al.*, 2020; Malgas *et al.*, 2020; Nagl *et al.*, 2022). Overall, the novel GH5 endo-β-1,4-glucanase StachCel5, identified via genome mining from *S. chartarum*, demonstrated broad pH stability (pH 4–8), thermal stability at 50 °C, and high specific activity on CMC. Beyond its robust biochemical profile, StachCel5 functioned as an effective and energy-efficient refining aid, enhancing fiber bonding and mechanical strength while reducing mechanical input, thus offering a sustainable alternative for pulp and paper processing.

Once a suitable enzyme is identified for industrial application, the next critical step toward reducing manufacturing costs and meeting growing global demand is the optimization of the

production strain (**Figure 10**). High protein titers are essential, as they enable smaller fermentation volumes, reduce downstream processing burdens, and ultimately enhances the overall economic viability of the process (Abbate *et al.*, 2023; Pothakos *et al.*, 2018). Given that a substantial proportion of industrial enzymes are produced using *Bacillus* species, particularly *B. subtilis* (Su *et al.*, 2020), **Chapter 2** focused on engineering a super-secreting *B. subtilis* KO7-S strain suitable for the industrial production of AmyQ, a well-characterized α-amylase widely used in the PPI. Achieving commercially viable yields in *B. subtilis* requires systematic optimization across multiple stages of protein production, including transcription, translation, folding, and secretion. In line with this, **Chapter 2** employed a rational, stepwise genome engineering strategy using CRISPR-Cas9 to construct an environmentally friendly, plasmid-free production chassis. Unlike plasmid-based systems or traditional marker-recycling methods, the chromosomal integration strategy employed here offers several advantages for industrial application, including improved genetic stability, elimination of antibiotic resistance markers and the use of expensive inducers, and a more streamlined strain development process.

A critical initial step in enhancing recombinant protein production is optimizing the expression cassette, particularly the promoter, which governs the timing and level of gene transcription (Nieuwkoop *et al.*, 2020). In this context, single, dual, and triple promoter configurations were evaluated to drive *amyQ* expression. Among them, the dual promoter P_{amyQ}-P_{cry3A} yielded the highest α-amylase activity, reaching 122.8 ± 4.4 U/mL, a 4.9-fold improvement over the strong constitutive P_{spoVG} promoter (**Table 2**). This enhancement likely reflects the synergistic effect of σ^A-dependent consensus sequences in P_{amyQ} and mRNA-stabilizing elements within P_{cry3A}, both known to boost transcription and mRNA stability in *B. subtilis* (Widner *et al.*, 2000). Notably, the use of strong constitutive promoters to drive consistent expression levels eliminated the need for costly inducers, facilitating scalability and compatibility with continuous bioprocessing. In parallel, the secretion route was also evaluated, as the choice of SP determines whether the enzyme is exported via the Sec or Tat pathway (Zhang *et al.*, 2020). While the Sec pathway is the predominant route for industrial enzymes, and was in deep analyzed in **Paper 2**, the potential of Tat-dependent secretion by fusing AmyQ to a Tat-specific SP was also investigated. However, no extracellular α-amylase activity was detected, indicating that AmyQ is incompatible with Tat-mediated export. This result is consistent with previous findings for other α-amylases and suggests that these enzymes are unable to fold appropriately in the cytoplasm to meet the folding requirements of the Tat system (Kolkman *et al.*, 2008).

Following expression cassette optimization, the next logical strategy to maximize gene expression was to increase gene copy number. Traditionally, this is achieved through high-copy-

number plasmids or sequential chromosomal integrations. However, these classical approaches are constrained by the limited availability of selection markers, reliance on marker-recycling protocols, and often result in genetic scars that complicate industrial approval (Su *et al.*, 2020). To overcome these limitations, **Paper 2** employed the CRISPR-Cas9 genome editing system to construct environmentally friendly, plasmid-free, markerless *B. subtilis* strains with up to six chromosomal copies of the *amyQ* gene, each controlled by the strong dual promoter P_{amyQ}-P_{cry3A}.

Ectopic integration sites were carefully selected based on their non-essential nature, ensuring cell viability and minimizing unintended physiological effects. Several of these loci have previously been shown to be dispensable, and their targeted deletion can, in some cases, enhance production performance (Aguilar-Suárez *et al.*, 2024; Xu *et al.*, 2023). Moreover, the integration strategy enabled simultaneous removal of native amylases (*amyE*), secondary metabolite clusters (*pksG*, *ppsE*), and genes involved in biofilm formation (*veg*) or sporulation (*cotB*), thereby reducing metabolic burden and further optimizing the strain for industrial enzyme production. Interestingly, no significant differences in growth or enzyme yield were observed across the various genomic integration sites, indicating functional equivalence under the tested conditions. As expected, α-amylase production increased with gene copy number, plateauing beyond five copies (**Table 2**), likely due to limitations in transcriptional or translational capacity. This saturation effect mirrors findings for other recombinant proteins, where optimal gene dosage is context-dependent and influenced by expression cassette architecture, codon usage, and translational dynamics (Watzlawick & Altenbuchner, 2019; Widner *et al.*, 2000).

Table 2. Effect on AmyQ α-amylase activity of successive modifications with significant positive effects in *B. subtilis* KO7-S. Modified from Ferrando *et al.,* 2024.

Strain	Modification	Amylase activity at 48 h. (U/mL)	Total fold change
BSQ1a	P_{spoVG}-*amyQ*	24.9 ± 1.5	1
BSQ1b	P_{amyQ}-P_{cry3A}-*amyQ*	122.8 ± 4.4	4.9
BSQ2	2(P_{amyQ}-P_{cry3A}-*amyQ*)	240.4 ± 8.0	9.7
BSQ3	3(P_{amyQ}-P_{cry3A}-*amyQ*)	315.7 ± 18.4	12.7
BSQ4	4(P_{amyQ}-P_{cry3A}-*amyQ*)	415.7 ± 21.1	16.7
BSQ5	5(P_{amyQ}-P_{cry3A}-*amyQ*)	501.2 ± 13.5	20.2
BSQ6	6(P_{amyQ}-P_{cry3A}-*amyQ*)	520.6 ± 20.2	20.9
BSQ6_11	6(P_{amyQ}-P_{cry3A}-*amyQ*) + PrsA + SppA	656.8 ± 24.3	26.4
BSQ6_11*	6(P_{amyQ}-P_{cry3A}-*amyQ*) + PrsA + SppA	1439.2 ± 92.7	57.9

*Values in a 3 L bioreactor at 90 h.

These results highlight that while increasing gene dosage can significantly boost protein yields, careful design of regulatory elements within the expression cassette is equally critical to reduce the time-consuming process of iterative gene integration. In support of this, a range of modular toolkits, such as BioBrick Box 1.0 (Radeck *et al.*, 2013) and 2.0 (Popp *et al.*, 2017), ProUSER 2.0 (Falkenberg *et al.*, 2021), and the more recently introduced SubtiToolKit (Caro-Astorga *et al.*, 2025), have been developed to enable rapid screening and optimization of expression constructs in *B. subtilis*, streamlining the development of robust industrial production strains

As a final step in the systematic strain engineering strategy, **Paper 2** addressed post-translational bottlenecks that often limit efficient protein secretion in *B. subtilis*. Specifically, the overexpression of the chaperone PrsA and the SP peptidase SppA significantly improved extracellular α-amylase yields. PrsA, a well-characterized extracellular chaperone, facilitates the proper folding of secreted proteins in the extracytoplasmic space, while SppA prevents the accumulation of uncleaved SPs, a common issue in overexpressing strains (Cai *et al.*, 2017; Quesada-Ganuza *et al.*, 2019). These targeted modifications led to a final shake flask yield of 657 U/mL, corresponding to a 26.4-fold increase relative to the baseline single-copy P_{spoVG}-driven strain (**Table 2**).

To confirm industrial scalability, the engineered strain was cultivated in a 3 L stirred-tank fermenter, where it achieved 1439.2 U/mL of α-amylase activity with a productivity of 16 U/mL·h (**Table 2**). Beyond high productivity, the use of a marker-free, genome-integrated system also confers enhanced genetic stability, eliminates the need for antibiotic selection, and improves regulatory compliance, key advantages for industrial enzyme production.

In summary, the work presented in **Chapter 2** outlines a rational and modular strain optimization pipeline for *B. subtilis*, leveraging promoter engineering, gene dosage tuning, and post-translational enhancement to construct a genetically stable, high-yield production platform. This approach not only sheds light on key expression bottlenecks limiting recombinant protein production in *B. subtilis* but also highlights the utility of CRISPR-Cas9-based genome editing as a versatile and efficient tool for accelerating strain development in industrial biotechnology. Nonetheless, a notable limitation of the strategy described in **Chapter 2** was the inherently iterative nature of multigene integration. Despite the use of CRISPR-Cas9 to streamline the editing process, sequential chromosomal insertion of multiple gene copies still required multiple rounds of plasmid-construction, transformation, selection, and validation. As such, the approach remains time-consuming and labor-intensive, emphasizing the need for more efficient multiplex genome editing platforms to further accelerate industrial strain development.

Multiplex genome editing in *B. subtilis* using CRISPR systems has seen remarkable progress, particularly for introducing multiple point mutations, as shown in Cas9n- and Cpf1-based strategies (Liu *et al.*, 2019; Wu *et al.*, 2020). However, when it comes to simultaneous gene deletions or insertions, these systems fall short of industrial needs, with editing efficiencies dropping sharply or failing altogether beyond two loci. For instance, Cpf1-based approaches achieved knockouts at only two loci simultaneously, while insertions were limited to a single chromosomal site, severely restricting its utility for complex strain engineering (Wu *et al.*, 2020). Compounding these technical barriers is the dependency on dual-plasmid systems, which adds layers of complexity to cloning and transformation workflows, ultimately hindering the pace of strain construction.

These shortcomings underscore a critical technological gap in current genome engineering toolkits: the lack of a simple, robust, and scalable method for multi-site gene integration in *B. subtilis*. To overcome this bottleneck, **Chapter 3** introduces a streamlined, one-step platform for multiplexed chromosomal integration of target genes at distinct chromosomal loci. This approach represents a major advancement in the genetic engineering of *B. subtilis*, offering a practical solution to accelerate the construction of next-generation production strains tailored for industrial biotechnology.

Given the inherently low editing efficiencies associated with multiplex gene insertions, **Paper 3** prioritized the development of a rapid, visual and efficient screening system to identify successful multi-locus integrants, eliminating the need for labor-intensive, low-throughput PCR-based validation. As the number of targeted loci increases, conventional screening methods become increasingly burdensome and impractical, posing a major bottleneck for scalable genome engineering. To address this, the study implemented, for the first time, a colorimetric selection system based on the *crtM* and *crtN* carotenoid biosynthesis genes from *Staphylococcus aureus*, which confer a vivid yellow pigmentation when expressed in *B. subtilis* (Yoshida *et al.*, 2009). This visually distinct phenotype offers a robust, substrate-free, and equipment-independent readout for identifying edited colonies. Unlike conventional screening systems such as X-gal/LacZ, chromoproteins, or fluorescent proteins, which often rely on expensive substrates, specialized detection equipment, or might suffer from toxicity and metabolic burden at high expression levels, the *crtMN* system is biologically well-tolerated, and compatible with high-throughput applications. Additionally, carotenoid production has been reported to enhance oxidative stress resistance in *B. subtilis*, providing a potential side benefit under CRISPR-associated stress conditions (Yoshida *et al.*, 2009). By enabling simple and scalable visual screening, this system greatly accelerates the identification of successful multi-locus integrants,

paving the way for more efficient implementation of complex genome editing strategies in *B. subtilis*.

The core innovation of the method developed in **Paper 3** lies in using *crtMN* insertions as removable colorimetric markers for CRISPR-Cas9-mediated genome editing. In this system, successful gene integration is visually confirmed by the loss of yellow pigmentation, as the incoming DNA construct replaces the *crtMN* operon at the target locus. Edited colonies thus appear white, enabling rapid, substrate-free, and equipment-independent identification of correct integrants. This simple yet powerful visual readout eliminates the need for PCR-based screening, greatly streamlining multiplex genome engineering. To implement this method, three patent-free *B. subtilis* KO7-S strains were constructed, each harboring one, two, or three chromosomal copies of *crtMN*, integrated at well-characterized loci: *spoVG*, *aprE*, and *amyE*. These insertions were achieved by targeting a synthetic expression cassette (MN_Ec), comprising the *crtMN* operon flanked by 530 base pairs (bp) upstream and 535 bp downstream homology arms of the *spoVG* gene. The operon was expressed under the strong endogenous P_{spoVG} promoter and terminated by a rho-independent T_{spoVG} terminator, resulting in stable and vivid yellow pigmentation across strains. This design enables precise and modular replacement of *crtMN* with desired constructs at multiple loci, making it a versatile platform for one-step multiplex editing in *B. subtilis*.

Building on the *crtMN*-tagged strains, **Paper 3** introduced a streamlined, single-plasmid CRISPR-Cas9 platform to enable simultaneous multilocus genome editing in *B. subtilis*. This system leverages a single, invariant sgRNA designed to specifically target the *crtMN* sequence with high precision and minimal off-target activity. By using the same *crtMN*-targeting sgRNA across all loci, the approach simplifies vector design and ensures consistency in editing. The editing template, tailored to the desired gene insertion, is flanked by standarized *spoVG* homology arms. These arms match those used in the *crtMN* cassettes previously integrated at the *spoVG*, *aprE*, and *amyE* loci, thereby enabling HDR at any or all of these sites. Upon Cas9-induced DSBs at the *crtMN* loci, HDR facilitates the seamless replacement of *crtMN* with the new construct. Successful integration is instantly visualized by a loss of yellow pigmentation, as colonies lacking *crtMN* appear white, thus being a clear, substrate-free readout that eliminates PCR screening. This method offers a robust, scalable, and visually traceable system for high-efficiency multiplex editing, greatly accelerating strain development workflows.

To validate the system's performance, the *amyQ* gene was used as a model for testing single-, double- and triple-locus integrations. The initial editing template (Q_Ec) enabled replacement

of *crtMN* with *amyQ* under the control of constitutive P_{spoVG}. Single-gene integrations (SGI) exhibited very high efficiency (97.5-98.6%), while double-gene integrations (DGI) achieved moderate efficiency (68.8%). In contrast, triple-gene integration (TGI) displayed a marked decline in efficiency (4.9% ± 0.8%), highlighting the critical utility of the colorimetric screening system for rapidly identifying correctly edited clones. Notably, all white colonies tested carried the expected genetic replacement, confirming the system's accuracy and reliability.

The versatility of the platform was further demonstrated by modifying the expression cassette. A new editing template (QXyl_Ec), placing *amyQ* under the control of the inducible P_{xyl} promoter, was designated to evaluate the system's compatibility with larger constructs. This 4,476 bp insert was successfully integrated at one and two loci, although double insertion efficiency dropped to ~10%, and no triple integration insertions were observed. This decline was attributed to the increased insert size, compared to *Q_Ec* (2,894 bp), highlighting the importance of template size as a limiting factor in multiplex integration (Huang *et al.*, 2022). Overall, these results demonstrate that while the system is robust and flexible, careful optimization of template architecture is essential to maximize editing efficiency across multiple loci. A notable advantage of this genome editing strategy lies in its modular and reusable design: only the editing template needs to be adapted for each new application, while the same sgRNA targeting the *crtMN* sequence can be consistently reused, greatly simplifying the workflow. Moreover, the single-plasmid-based CRISPR-Cas9 system established in **Paper 3** offers clear benefits over dual-plasmid approaches, including reduced transformation and vector construction complexity. Although HDR efficiency is influenced by the length of the homologous arms (Tian *et al.*, 2022), our findings confirm that the 530-bp of flanking sequences are sufficient to support efficient multicopy integration. Moreover, we observed that the natural NHEJ system present in *B. subtilis* is too weak to repair DSBs, which is in accordance with previous reports (Toymentseva & Altenbuchner, 2019). Despite a reduction in editing efficiency for multiple insertions, the colorimetric screening strategy developed in this thesis provides a powerful, cost-effective, and user-friendly solution to bypass these limitations. The distinct loss of yellow pigmentation upon successful editing enables rapid and reliable identification of positive clones, accelerating strain development workflows**.**

Remarkably, this system enabled the construction of *B. subtilis* strains expressing secreted α-amylase from multiple chromosomal loci in just seven days, demonstrating its speed and efficiency. The system's broad applicability was further validated through promoter swapping, allowing customized control of gene expression. Importantly, both the CRISPR-Cas9 plasmid and the engineered *crtMN*-tagged *B. subtilis* KO7-S strains are publicly available and free from

intellectual property restrictions, promoting open-access adoption. Combined with the KO7-S host's favorable industrial traits (lack of extracellular proteases and non-sporulating phenotype), this toolbox provides a robust and scalable platform for developing high-performance protein secretion strains tailored for biotechnological applications.

Although carotenoid production was not an initial focus of this thesis, the robust yellow pigmentation observed in the engineered *B. subtilis* strains from **Paper 3** sparked further investigation into their potential for carotenoid biosynthesis. This distinct and stable phenotype not only confirmed the functionality and reliability of the colorimetric screening system but also highlighted *B. subtilis* as a promising chassis for the sustained and efficient production of carotenoids. The unexpected strength and consistency of pigment expression underscored the strain's compatibility with heterologous isoprenoid pathways, thus revealing an untapped potential for metabolic engineering. Based on this observation, **Paper 4** pivoted toward developing *B. subtilis* as a cell factory for C_{30}-carotenoid biosynthesis, laying the groundwork for expanding its role beyond protein secretion into the production of high-value pigments and other terpenoid-derived compounds.

Given the rising demand for sustainable carotenoids due to their antioxidant properties and high industrial value (Siziya *et al.*, 2022), **Paper 4** addressed key limitations in this microbial production system by focusing on gene dosage optimization, metabolic flux redirection, and regulatory compatibility. While *B. subtilis* naturally synthesizes farnesyl pyrophosphate (FPP), a central precursor in carotenoid biosynthesis, it does not natively produce C_{30} carotenoids. These compounds are typically synthesized by bacteria such as *Staphylococcus aureus* through the *crtM* and *crtN* genes, which encode enzymes converting FPP into 4,4'-diaponeurosporene and subsequently into 4,4'-diapolycopene, the characteristic yellow C_{30} carotenoids (Yoshida *et al.*, 2009). To reconstitute this pathway in *B. subtilis*, up to three chromosomal copies of *crtMN* were integrated at the *spoVG*, *aprE*, and *amyE* loci, under the control of the strong native *spoVG* promoter as established in **Paper 3**. This scarless, multi-copy chromosomal integration strategy not only circumvented the instability and metabolic burden of plasmid-based expression systems, but also effectively resolved gene dosage limitations. Remarkably, the resulting strain achieved carotenoid titers comparable to those obtained with a medium-copy-number plasmid (pBS0EXylRP from Popp *et al.*, 2017, ~15–25 copies per cell), demonstrating that stable chromosomal expression can rival plasmid-based production yields. To further boost precursor availability, the *fpps* gene from *Bacillus megaterium*, encoding farnesyl diphosphate synthase (Hartz *et al.*, 2018), was genomically integrated, enhancing FPP synthesis. In parallel, deletion of *yisP,* a gene diverting FPP toward farnesol lipid biosynthesis (Feng *et al.*, 2014), redirected

metabolic flux toward carotenoid production. Together, these interventions led to a 6-fold increase in carotenoid yield compared to a strain harboring a single *crtMN* copy.

Overall, **Chapter 3** showcases a streamlined and modular CRISPR-Cas9 platform for multiplex genome editing in *B. subtilis*, combining precise multi-locus insertion with a robust colorimetric screening method based on the *crtMN* operon. Beyond its primary utility in genome engineering, this study demonstrates the broader applicability of *crtMN* as both a selectable marker and a functional biosynthetic module. These findings highlight *B. subtilis* as a versatile chassis for the stable and efficient production of industrially relevant C_{30} carotenoids, expanding its biotechnological potential beyond protein secretion.

Ultimately, this thesis provides a blueprint for accelerating the construction of high-performance *B. subtilis* strains and broadens the biotechnological scope of this GRAS/QPS organism beyond its traditional roles, paving the way toward its use in next-generation biofactories for sustainable bioproduction. Collectively, by combining advanced enzyme discovery with scalable, precise microbial engineering tools, this thesis lays the groundwork for more efficient, sustainable, and innovative solutions in industrial biotechnology.

Looking ahead, the tools and strategies established in this thesis provide a foundation for multiple avenues of **future research and potential applications**:

- The discovery and characterization of StachCel5 exemplify the untapped potential of underexplored microbial genomes for industrial enzyme development. While its efficacy in pulp refining has been demonstrated in this thesis, its favorable properties point to broader applications in biomass valorization processes, including biofuel production, textile treatment, agricultural waste conversion, and nanocellulose production. Moreover, StachCel5 represents a strong candidate for further protein engineering, particularly with the advent of AI-driven design tools (Singh *et al.,* 2025; Wang *et al.,* 2025), which could be employed to improve its thermostability, catalytic efficiency, or substrate specificity to better align with industrial demands.
- On the microbial engineering front, the stepwise genome editing strategy applied to construct high-producing *B. subtilis* strains offers a modular, plasmid-free, and environmentally friendly alternative to traditional approaches. To further enhance strain productivity, this framework could be integrated with transcriptomic analyses and ribosome profiling to identify rate-limiting steps in protein synthesis and secretion (Wang *et al.,* 2024). In parallel, CRISPR interference (CRISPRi) represents a powerful high-throughput tool for

systematically screening genes that indirectly influence protein production, thereby uncovering novel targets for strain optimization (Zhang *et al.,* 2024).

- The multilocus genome editing system developed in this thesis may serve as the bedrock for future refinements. One promising direction involves the use of Cas nickase to promote HDR via single-strand breaks, thereby enhancing editing efficiency while reducing cytotoxicity. Coupled with strategies that inhibit nick re-ligation and employ heterologous recombination systems, editing efficiency could be further improved (Liu *et al.*, 2019; Wu *et al.*, 2020). Moreover, boosting the natural competence of *B. subtilis* through overexpression of the competence regulators *comK* and *comS* may increase transformation efficiency and co-editing success (Deng *et al.,* 2021). The visual screening platform could likewise be refined by incorporating shorter chromogenic or fluorescent markers, streamlining high-throughput editing workflows as recently demonstrated in this strain (Caro-Astorga *et al.*, 2025).

6 Conclusions

6. Conclusions

- A novel GH5 endo-β-1,4-glucanase (StachCel5) is identified from *Stachybotrys chartarum* through genome mining and bioinformatic analysis, demonstrating the effectiveness of *in silico* discovery strategies for industrial enzymes from underexplored fungi.
- Heterologous expression in *Komagataella phaffii* enables efficient secretion and production of active StachCel5, validating this yeast as a robust host for fungal enzyme expression.
- Structural analysis reveals that StachCel5 follows a retaining catalytic mechanism, with conserved glutamate residues (Glu^{228} and Glu^{335}) embedded in a TIM-barrel fold typical of GH5 endoglucanases.
- StachCel5 exhibits high specific activity on carboxymethyl cellulose, broad pH stability (4-9), and notable thermostability at 50 °C, supporting its suitability for industrial conditions.
- The enzyme shows strict specificity for β-1,4-glucan linkages, with no activity on xylan, laminarin, starch, crystalline cellulose, or synthetic analogs, confirming its potential as a precise tool for targeted cellulose modification.
- StachCel5 shows resistance to inhibition by most metal ions and common reagents such as EDTA and β-mercaptoethanol, while being completely inhibited by SDS, consistent with related GH5 enzymes.
- Kinetic characterization (K_m = 9.77 mg/mL, V_{max} = 898.4 U/mg, K_{cat} = 635.7 s^{-1}, K_{cat}/K_m = 65.4 mL/s/mg) places StachCel5 among the top-performing fungal GH5 endoglucanase reported to date, combining high turnover rates with favorable substrate affinity.
- In refining trials, StachCel5 improved pulp fibrillation and inter-fiber bonding in softwood pulp, reduced mechanical energy input, and enhanced paper tensile and burst strength while preserving fiber integrity, thus highlighting strong biorefining potential.
- A stepwise CRISPR-Cas9 genome engineering framework has been established to construct stable, high-expression, and marker-free *Bacillus subtilis* strains for industrial enzyme production.
- Among tested configurations, the dual promoter P_{amyQ}-P_{cry3A} has yielded the highest expression in *B. subtilis*.

- Iterative chromosomal integration of up to six copies of the *amyQ* gene has resulted in a 20.9-fold increase in α-amylase production, with expression plateauing beyond five copies.
- Analysis of rate-limiting steps has identified secretion bottlenecks, including limited PrsA chaperone availability and signal peptide accumulation, providing clear targets for further strain development.
- The Tat secretion pathway is incompatible with AmyQ export under the tested conditions.
- The final engineered *B. subtilis* strain, BSQ6_11, has achieved a final 58-fold increase in AmyQ production in a 3L bioreactor, demonstrating the industrial potential of the CRISPR-Cas9 strain engineering approach.
- A single-plasmid CRISPR-Cas9 system for *B. subtilis* has been developed to enable simultaneous multilocus integration of up to three gene copies in a single step, coupled with a novel *crtMN*-based colorimetric selection system that allows rapid visual screening of edited clones without PCR.
- Single and double gene insertions have shown high editing efficiencies (>90% and ~69%, respectively), while triple insertions have remained feasible though less efficient (~5%).
- Editing efficiency has been inversely correlated with the size of the editing template, highlighting the need for optimization in large-fragment genome integration.
- Homology arms of 530-bp have proven sufficient for high-efficiency homologous recombination in multiplex editing applications.
- The CRISPR-Cas9 platform has been validated through successful, accurate multi-locus integration of *amyQ*, enabling rapid development of stable, high-expression strains.
- Beyond genome editing, the *crtMN* pigmentation system has been repurposed for stable chromosomal production of C_{30} carotenoids, achieving titers comparable to medium-copy plasmid systems.
- Integration of *fpps* and deletion of the competing *yisP* gene have enhanced precursor flux and led to a final 6-fold increase in carotenoid production, showcasing the engineered strain's value for metabolic engineering.

7 References

7. References

Abbate, E., Andrion, J., Apel, A., Biggs, M., Chaves, J., Cheung, K., Ciesla, A., Clark-ElSayed, A., Clay, M., Contridas, R., Fox, R., Hein, G., Held, D., Horwitz, A., Jenkins, S., Kalbarczyk, K., Krishnamurthy, N., Mirsiaghi, M., Noon, K., … Yerramsetty, K. (2023). Optimizing the strain engineering process for industrial-scale production of bio-based molecules. *Journal of Industrial Microbiology & Biotechnology*, *50*(1), kuad025. https://doi.org/10.1093/jimb/kuad025

Aguilar Suárez, R., Kohlstedt, M., Öktem, A., Neef, J., Wu, Y., Ikeda, K., Yoshida, K., Altenbuchner, J., Wittmann, C., & van Dijl, J. M. (2024). Metabolic profile of the genome-reduced *Bacillus subtilis* strain IIG-Bs-27-39: An attractive chassis for recombinant protein production. *ACS Synthetic Biology*, *13*(7), 2199–2214. https://doi.org/10.1021/acssynbio.4c00254

Ahadian, H., Ceccherini, S., Sharifi Zamani, E., Phiri, J., & Maloney, T. (2023). Production of low-density and high-strength paperboards by controlled micro-nano fibrillation of fibers. *Journal of Materials Science*, *58*(44), 17126–17137. https://doi.org/10.1007/s10853-023-09097-9

Altenbuchner, J. (2016). Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system. *Applied and Environmental Microbiology*, *82*(17), 5421–5427. https://doi.org/10.1128/aem.01453-16

Ariaeenejad, S., Gharechahi, J., Foroozandeh Shahraki, M., Fallah Atanaki, F., Han, J.-L., Ding, X.-Z., Hildebrand, F., Bahram, M., Kavousi, K., & Hosseini Salekdeh, G. (2024). Precision enzyme discovery through targeted mining of metagenomic data. *Natural Products and Bioprospecting*, *14*(1), 1–17. https://doi.org/10.1007/s13659-023-00426-8

Arsov, A., Armenova, N., Gergov, E., Petrov, K., & Petrova, P. (2024). Cloning Systems in *Bacillus*: Bioengineering of metabolic pathways for valuable recombinant products. *Fermentation*, *10*(1), 50. https://doi.org/10.3390/fermentation10010050

Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., & Henrissat, B. (2012). Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology*, *12*(1), 186. https://doi.org/10.1186/1471-2148-12-186

Atalla, A., & Schumann, W. (2003). The pst operon of *Bacillus subtilis* is specifically induced by alkali stress. *Journal of Bacteriology*, *185*(16), 5019–5022. https://doi.org/10.1128/jb.185.16.5019-5022.2003

Austin, P. C., Mack, J., McEwan, M., Afshar, P., Brown, M., & Maciejowski, J. (2011). Improved energy efficiency in paper making through reducing dryer steam consumption using advanced process control. In *Proceedings of the Paper Conference and Trade Show*.

Bajpai, P. (2018). Refining and pulp characterization. In C. J. Biermann (Ed.), *Biermann's handbook of pulp and paper: Volume 2: Paper and board making* (pp. 1–34). Elsevier. https://doi.org/10.1016/b978-0-12-814238-7.00001-5

Balakrishnan, R., Mori, M., Segota, I., Zhang, Z., Aebersold, R., Ludwig, C., & Hwa, T. (2022). Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria. *Science*, *378*(6624), eabk2066. https://doi.org/10.1126/science.abk2066

Barone, G. D., Emmerstorfer-Augustin, A., Biundo, A., Pisano, I., Coccetti, P., Mapelli, V., & Camattari, A. (2023). Industrial production of proteins with *Pichia pastoris—Komagataella phaffii*. *Biomolecules*, *13*(3), 441. https://doi.org/10.3390/biom13030441

Benalaya, I., Alves, G., Lopes, J., & Silva, L. R. (2024). A review of natural polysaccharides: sources, characteristics, properties, food, and pharmaceutical applications. *International Journal of Molecular Sciences*, *25*(2), 1322. https://doi.org/10.3390/ijms25021322

Bergquist, A. K., & Söderholm, K. (2018). The greening of the pulp and paper industry: Sweden in comparative perspective. In T. Kuhlberg & A. J. McIvor (Eds.), *Technological transformation in the global pulp and paper industry 1800–2018: Comparative perspectives* (pp. 65–87). Springer International Publishing. https://doi.org/10.1007/978-3-319-94962-8

Bernardi, A. V., Yonamine, D. K., Uyemura, S. A., & Dinamarco, T. M. (2019). A Thermostable *Aspergillus fumigatus* GH7 endoglucanase over-expressed in *Pichia pastoris* stimulates lignocellulosic biomass hydrolysis. *International Journal of Molecular Sciences*, *20*(9), 2261. https://doi.org/10.3390/ijms20092261

Bhardwaj, S., Kaur, P., Bhardwaj, N. K., & Negi, Y. S. (2023). Surface application of different concentrations of chitosan on recycled paper and its impact on packaging properties. *Journal of Coatings Technology and Research*, *20*(4), 1285–1298. https://doi.org/10.1007/s11998-022-00743-6

Bhavsar, A. P., Zhao, X., & Brown, E. D. (2001). Development and characterization of a xylose-dependent system for expression of cloned genes in *Bacillus subtilis*: Conditional complementation of a teichoic acid mutant. *Applied and Environmental Microbiology*, *67*(1), 403–410. https://doi.org/10.1128/aem.67.1.403-410.2001

Brans, A., Filée, P., Chevigné, A., Claessens, A., & Joris, B. (2004). New integrative method to generate *Bacillus subtilis* recombinant strains free of selection markers. *Applied and Environmental Microbiology*, *70*(12), 7241-7250. https://doi.org/10.1128/aem.70.12.7241-7250.2004

Cai, D., Wang, H., He, P., Zhu, C., Wang, Q., Wei, X., Nomura, C. T., & Chen, S. (2017). A novel strategy to improve protein secretion via overexpression of the SppA signal peptide peptidase in *Bacillus licheniformis*. *Microbial Cell Factories*, *16*(1), 70. https://doi.org/10.1186/s12934-017-0688-7

Çalik, P., Ata, Ö., Güneş, H., Massahi, A., Boy, E., Keskin, A., Öztürk, S., Zerze, G. H., & Özdamar, T. H. (2015). Recombinant protein production in *Pichia pastoris* under glyceraldehyde-3-phosphate dehydrogenase promoter: From carbon source metabolism to bioreactor operation parameters. *Biochemical Engineering Journal*, *95*, 20–36. https://doi.org/10.1016/j.bej.2014.12.003

Caro-Astorga, J., Rogan, M., Malcı, K., Ming, H., Debenedictis, E., James, P., & Ellis, T. (2025). SubtiToolKit: a bioengineering kit for *Bacillus subtilis* and Gram-positive bacteria. *Trends in Biotechnology*, *43*(6), 1446–1469. https://doi.org/10.1016/j.tibtech.2025.02.004

Chen, J., Gai, Y., Fu, G., Zhou, W., Zhang, D., & Wen, J. (2015). Enhanced extracellular production of α-amylase in *Bacillus subtilis* by optimization of regulatory elements and overexpression of PrsA lipoprotein. *Biotechnology Letters*, *37*(4), 899–906. https://doi.org/10.1007/s10529-014-1755-3

Chen, P. T., Shaw, J. F., Chao, Y. P., David Ho, T. H., & Yu, S. M. (2010). Construction of chromosomally located T7 expression system for production of heterologous secreted proteins in *Bacillus subtilis*. *Journal of Agricultural and Food Chemistry*, *58*(9), 5392–5399. https://doi.org/10.1021/jf100445a

Chen, T., Brul, S., & Hugenholtz, J. (2023). Exploring the potential of *Bacillus subtilis* as cell factory for food ingredients and special chemicals. *Microbial Cell Factories*, *22*(1), 1–12. https://doi.org/10.1186/s12934-023-02208-w

Chen, Y., Li, M., Yan, M., Chen, Y., Saeed, M., Ni, Z., Fang, Z., & Chen, H. (2024). *Bacillus subtilis*: Current and future modification strategies as a protein secreting factory. *World Journal of Microbiology and Biotechnology*, *40*(6), 195. https://doi.org/10.1007/s11274-024-03997-x

Cheng, J., Guan, C., Cui, W., Zhou, L., Liu, Z., Li, W., & Zhou, Z. (2016). Enhancement of a high efficient autoinducible expression system in *Bacillus subtilis* by promoter engineering. *Protein Expression and Purification*, *127*, 81–87. https://doi.org/10.1016/j.pep.2016.07.008

Claes, K., Van Herpe, D., Vanluchene, R., Roels, C., Van Moer, B., Wyseure, E., Vandewalle, K., Eeckhaut, H., Yilmaz, S., Vanmarcke, S., Çıtak, E., Fijalkowska, D., Grootaert, H., Lonigro, C., Meuris, L., Michielsen, G., Naessens, J., van Schie, L., De Rycke, R., … Callewaert, N. (2024). OPENPichia: Licence-free *Komagataella phaffii* chassis strains and toolkit for protein expression. *Nature Microbiology*, *9*(3), 864–876. https://doi.org/10.1038/s41564-023-01574-w

Datta, R. (2024). Enzymatic degradation of cellulose in soil: A review. *Heliyon*, *10*(1), e24022. https://doi.org/10.1016/j.heliyon.2024.e24022

Dave, B. R., Sudhir, A. P., & Subramanian, R. B. (2015). Purification and properties of an endoglucanase from *Thermoascus aurantiacus*. *Biotechnology Reports*, *6*, 85–90. https://doi.org/10.1016/j.btre.2014.11.004

de Lima, E. A., Mandelli, F., Kolling, D., Matsusato Souza, J., de Oliveira Filho, C. A., Ribeiro da Silva, M., Lobo de Mesquita Sampaio, I., Lopes Junqueira, T., Ferreira Chagas, M., Teodoro, J. C., de Morais, E. R., & Murakami, M. T. (2022). Development of an economically competitive *Trichoderma*-based platform for enzyme production: Bioprocess optimization, pilot plant scale-up, techno-economic analysis and life cycle assessment. *Bioresource Technology*, *364*, 128019. https://doi.org/10.1016/j.biortech.2022.128019

Deng, A., Sun, Z., Wang, T., Cui, D., Li, L., Liu, S., Huang, F., & Wen, T. (2021). Simultaneous multiplex genome engineering via accelerated natural transformation in *Bacillus subtilis*. *Frontiers in Microbiology, 12*, 714449. https://doi.org/10.3389/fmicb.2021.714449

Dong, H., & Zhang, D. (2014). Current development in genetic engineering strategies of *Bacillus* species. *Microbial Cell Factories*, *13*(1), 1–11. https://doi.org/10.1186/1475-2859-13-63

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096. https://doi.org/10.1126/science.1258096

Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., & Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research*, *50*(D1), D571–D577. https://doi.org/10.1093/nar/gkab1045

Duman-Özdamar, Z. E., & Binay, B. (2021). Production of industrial enzymes via *Pichia pastoris* as a cell factory in bioreactor: Current status and future aspects. *Protein Journal*, *40*(3), 367-376. https://doi.org/10.1007/s10930-021-09968-7

Earl, A. M., Losick, R., & Kolter, R. (2008). Ecology and genomics of *Bacillus subtilis*. *Trends in microbiology*, *16*(6), 269-275. https://doi.org/10.1016/j.tim.2008.03.004

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797. https://doi.org/10.1093/nar/gkh340

Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., & Sun, Y. (2021). Human influence on the climate system. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 423–552). Cambridge University Press. https://doi.org/10.1017/9781009157896.005

Ergün, B. G., Laçın, K., Çaloğlu, B., & Binay, B. (2022). Second generation *Pichia pastoris* strain and bioprocess designs. *Biotechnology for Biofuels and Bioproducts*, *15*(1), 1–19. https://doi.org/10.1186/s13068-022-02234-7

Escuder-Rodríguez, J. J., González-Suarez, M., deCastro, M. E., Saavedra-Bouza, A., Becerra, M., & González-Siso, M. I. (2022). Characterization of a novel thermophilic metagenomic GH5 endoglucanase heterologously expressed in *Escherichia coli* and *Saccharomyces cerevisiae*. *Biotechnology for Biofuels and Bioproducts*, *15*(1), 1–13. https://doi.org/10.1186/s13068-022-02172-4

Fabret, C., Ehrlich, S. D., & Noirot, P. (2002). A new mutation delivery system for genome-scale approaches in *Bacillus subtilis*. *Molecular Microbiology*, *46*(1), 25–36. https://doi.org/10.1046/j.1365-2958.2002.03140.x

Falkenberg, K. B., Mol, V., de la Maza Larrea, A. S., Pogrebnyakov, I., Nørholm, M. H. H., Nielsen, A. T., & Jensen, S. I. (2021). The ProUSER2.0 toolbox: Genetic parts and highly customizable plasmids for synthetic biology in *Bacillus subtilis*. *ACS Synthetic Biology*, *10*(12), 3278–3289.

https://doi.org/10.1021/acssynbio.1c00130

Farooq, M. A., Ali, S., Hassan, A., Tahir, H. M., Mumtaz, S., & Mumtaz, S. (2021). Biosynthesis and industrial applications of α-amylase: A review. *Archives of Microbiology*, *203*(4), 1281-1292. https://doi.org/10.1007/s00203-020-02128-y

Feng, X., Hu, Y., Zheng, Y., Zhu, W., Li, K., Huang, C. H., Ko, T. P., Ren, F., Chan, H. C., Nega, M., Bogue, S., López, D., Kolter, R., Götz, F., Guo, R. T., & Oldfield, E. (2014). Structural and functional analysis of *Bacillus subtilis* YisP reveals a role of its product in biofilm production. *Chemistry and Biology*, *21*(11), 1557–1563. https://doi.org/10.1016/j.chembiol.2014.08.018

Ferrando, J., Lliso-Pascual, C., Cusola, O., Roncero, M. B., Planas, A., & Picart, P. (2025). From mold to mill: StachCel5, a novel thermoalkaliphilic endoglucanase from *Stachybotrys chartarum* for pulp fiber biorefining. *International Journal of Biological Macromolecules, 320(3), 145969.* https://doi.org/10.1016/j.ijbiomac.2025.145969

Ferrando, J., Miñana-Galbis, D., & Picart, P. (2024). The Construction of an environmentally friendly super-secreting strain of *Bacillus subtilis* through systematic modulation of its secretory pathway using the CRISPR-Cas9 system. *International Journal of Molecular Sciences*, *25*(13), 6957. https://doi.org/10.3390/ijms25136957/s1

Ferreira, R. D. G., Azzoni, A. R., & Freitas, S. (2018). Techno-economic analysis of the industrial production of a low-cost enzyme using *E. coli*: The case of recombinant β-glucosidase. *Biotechnology for Biofuels*, *11*(1), 81. https://doi.org/10.1186/s13068-018-1077-0

Ferreira, R. G., Azzoni, A. R., & Freitas, S. (2021). On the production cost of lignocellulose-degrading enzymes. *Biofuels, Bioproducts and Biorefining*, *15*(1), 85–99. https://doi.org/10.1002/bbb.2142

Frain, K. M., van Dijl, J. M., & Robinson, C. (2019). The Twin-arginine pathway for protein secretion. *EcoSal Plus*, *8*(2), 10-1128.. https://doi.org/10.1128/ecosalplus.esp-0040-2018

Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems. *Microbial Cell Factories*, *17*(1), 52. https://doi.org/10.1186/s12934-018-0901-3

Fu, L. L., Xu, Z. R., Li, W. F., Shuai, J. B., Lu, P., & Hu, C. X. (2007). Protein secretion pathways in *Bacillus subtilis*: Implication for optimization of heterologous protein secretion. *Biotechnology Advances*, *25*(1), 1–12. https://doi.org/10.1016/j.biotechadv.2006.08.002

Furszyfer Del Rio, D. D., Sovacool, B. K., Griffiths, S., Bazilian, M., Kim, J., Foley, A. M., & Rooney, D. (2022). Decarbonizing the pulp and paper industry: A critical and systematic review of sociotechnical developments and policy options. *Renewable and Sustainable Energy Reviews*, *167*, 112706. https://doi.org/10.1016/j.rser.2022.112706

García-Ortega, X., Cámara, E., Ferrer, P., Albiol, J., Montesinos-Seguí, J. L., & Valero, F. (2019). Rational development of bioprocess engineering strategies for recombinant protein production in *Pichia pastoris* (*Komagataella phaffii*) using the methanol-free GAP promoter. Where do we stand? *New Biotechnology*, *53*, 24-34. https://doi.org/10.1016/j.nbt.2019.06.002

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784-3788. https://doi.org/10.1093/nar/gkg563

Gharehkhani, S., Sadeghinezhad, E., Kazi, S. N., Yarmand, H., Badarudin, A., Safaei, M. R., & Zubir, M. N. M. (2015). Basic effects of pulp refining on fiber properties—A review. *Carbohydrate Polymers*, *115*, 785-803. https://doi.org/10.1016/j.carbpol.2014.08.047

Global Market Insights. (2023). *Pulp and paper enzymes market size, industry analysis, report 2032*. https://www.gminsights.com/industry-analysis/pulp-and-paper-enzymes-market

González Moraga, S. (2025). *CRISPR-Cas9 adaptive immune system of Streptococcus pyogenes against bacteriophages* [BioRender template]. https://app.biorender.com/biorender-templates/details/t-5f2aa9452c88bf00b025a42d-crispr-cas9-adaptive-immune-system-of-streptococcus-pyogenes

Guan, C., Cui, W., Cheng, J., Liu, R., Liu, Z., Zhou, L., & Zhou, Z. (2016). Construction of a highly active secretory expression system via an engineered dual promoter and a highly efficient signal peptide in *Bacillus subtilis*. *New Biotechnology*, *33*(3), 372-379. https://doi.org/10.1016/j.nbt.2016.01.005

Hahn-Hägerdal, B., Karhumaa, K., Larsson, C. U., Gorwa-Grauslund, M., Görgens, J., & van Zyl, W. H. (2005). Role of cultivation media in the development of yeast strains for large scale industrial use. *Microbial Cell Factories*, *4*(1), 31. https://doi.org/10.1186/1475-2859-4-31

Hartz, P., Milhim, M., Trenkamp, S., Bernhardt, R., & Hannemann, F. (2018). Characterization and engineering of a carotenoid biosynthesis operon from *Bacillus megaterium*. *Metabolic Engineering*, *49*, 47-58. https://doi.org/10.1016/j.ymben.2018.07.017

Harwood, C. R., & Cranenburgh, R. (2008). *Bacillus* protein secretion: an unfolding story. *Trends in Microbiology*, *16*(2), 73-79. https://doi.org/10.1016/j.tim.2007.12.001

Haske-Cornelius, O., Hartmann, A., Brunner, F., Pellis, A., Bauer, W., Nyanhongo, G. S., & Guebitz, G. M. (2020). Effects of enzymes on the refining of different pulps. *Journal of Biotechnology*, *320*, 1-10. https://doi.org/10.1016/j.jbiotec.2020.06.006

Herrmann, L. W., Letti, L. A. J., Penha, R. de O., Soccol, V. T., Rodrigues, C., & Soccol, C. R. (2024). *Bacillus* genus industrial applications and innovation: First steps towards a circular bioeconomy. *Biotechnology Advances*, *70*, 108300. https://doi.org/10.1016/j.biotechadv.2023.108300

Hobusch, M., Kırtel, O., Meramo, S., Sukumara, S., & Hededam Welner, D. (2024). A life cycle assessment of early-stage enzyme manufacturing simulations from sustainable feedstocks. *Bioresource Technology*, *400*, 130653. https://doi.org/10.1016/j.biortech.2024.130653

Hoffmann, M., Braig, A., Fernandez Cano Luna, D. S., Rief, K., Becker, P., Treinen, C., Klausmann, P., Morabbi Heravi, K., Henkel, M., Lilge, L., & Hausmann, R. (2021). Evaluation of an oxygen-dependent self-inducible surfactin synthesis in *B. subtilis* by substitution of native promoter PsrfA by anaerobically active PnarG and PnasD. *AMB Express*, *11*(1), 57. https://doi.org/10.1186/s13568-021-01218-4

Hohmann, H.-P., van Dijl, J. M., Krishnappa, L., & Prágai, Z. (2017). Host organisms: *Bacillus subtilis*. In C. Wittmann & J. C. Liao (Eds.), *Industrial biotechnology: Microorganisms* (Chapter 7). Wiley-VCH. https://doi.org/10.1002/9783527807796.ch7

Huang, K., Zhang, T., Jiang, B., Yan, X., Mu, W., & Miao, M. (2017). Overproduction of *Rummeliibacillus pycnus* arginase with multi-copy insertion of the arg $^{R.pyc}$ cassette into the *Bacillus subtilis* chromosome. *Applied Microbiology and Biotechnology*, *101*(15), 6039-6048. https://doi.org/10.1007/s00253-017-8355-9

Huang, S., Xue, Y., Zhou, C., & Ma, Y. (2022). An efficient CRISPR/Cas9-based genome editing system for alkaliphilic *Bacillus* sp. N16-5 and application in engineering xylose utilization for d-lactic acid production. *Microbial Biotechnology, 15*(11), 2730–2743. https://doi.org/10.1111/1751-7915.14131

Huang, W. Z., Wang, J. J., Chen, H. J., Chen, J. T., & Shaw, G. C. (2013). The heat-inducible essential response regulator WalR positively regulates transcription of *sigI*, *mreBH* and *lytE* in *Bacillus subtilis* under heat stress. *Research in Microbiology*, *164*(10), 998-1008. https://doi.org/10.1016/j.resmic.2013.10.003

Ibrahim, S. R. M., Choudhry, H., Asseri, A. H., Elfaky, M. A., Mohamed, S. G. A., & Mohamed, G. A. (2022). *Stachybotrys chartarum*—A hidden treasure: secondary metabolites, bioactivities, and biotechnological relevance. *Journal of Fungi*, *8*(5), 504. https://doi.org/10.3390/jof8050504

International Energy Agency. (2021). *Net zero by 2050*. https://www.iea.org/reports/net-zero-by-2050

International Energy Agency. (2023). *Direct CO2 emissions from industry in the Net Zero Scenario, 2000-2030*. IEA. https://www.iea.org/data-and-statistics/charts/direct-co2-emissions-from-industry-in-the-net-zero-scenario-2000-2030-2

International Energy Agency. (2024). *Global energy and climate model*. https://www.iea.org/reports/global-energy-and-climate-model

Jeong, D. E., Park, S. H., Pan, J. G., Kim, E. J., & Choi, S. K. (2015). Genome engineering using a synthetic gene circuit in *Bacillus subtilis*. *Nucleic Acids Research*, *43*(6), e42. https://doi.org/10.1093/nar/gku1380

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816-821. https://doi.org/10.1126/science.1225829

Juturu, V., & Wu, J. C. (2018). Heterologous protein expression in *Pichia pastoris*: Latest research progress and applications. *ChemBioChem*, *19*(1), 7-21. https://doi.org/10.1002/cbic.201700460

Kamble, A., Srinivasan, S., & Singh, H. (2019). *In-silico* bioprospecting: Finding better enzymes. *Molecular Biotechnology*, *61*(1), 53-59. https://doi.org/10.1007/s12033-018-0132-1

Karbalaei, M., Rezaee, S. A., & Farsiani, H. (2020). *Pichia pastoris*: A highly successful expression system for optimal synthesis of heterologous proteins. *Journal of Cellular Physiology*, *235*(9), 5867-5881. https://doi.org/10.1002/jcp.29583

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059-3066. https://doi.org/10.1093/nar/gkf436

Kawabata, Y., Kimura, K., & Funane, K. (2012). Extracellular production of cycloisomaltooligosaccharide glucanotransferase and cyclodextran by a protease-deficient *Bacillus subtilis* host-vector system. *Applied Microbiology and Biotechnology*, *93*(5), 1877–1884. https://doi.org/10.1007/s00253-011-3671-y

Kerekes, R. J., Heymer, J. O., & McDonald, J. D. (2021). Refining pulp for tensile strength. *Nordic Pulp and Paper Research Journal*, *36*(4), 696-706. https://doi.org/10.1515/npprj-2021-0012

Khootama, A., Putri, D. N., & Hermansyah, H. (2018). Techno-economic analysis of lipase enzyme production from *Aspergillus niger* using agro-industrial waste by solid state fermentation. *Energy Procedia*, *153*, 143-148. https://doi.org/10.1016/j.egypro.2018.10.054

Kolkman, M. A. B., Van Der Ploeg, R., Bertels, M., Van Dijk, M., Van Der Laan, J., Van Dijl, J. M., & Ferrari, E. (2008). The Twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct either Tat- or Sec-dependent secretion of different cargo proteins: Secretion of active subtilisin via the *B. subtilis* Tat pathway. *Applied and Environmental Microbiology*, *74*(24), 7507-7513. https://doi.org/10.1128/aem.01401-08

Konzock, O., & Nielsen, J. (2024). TRYing to evaluate production costs in microbial biotechnology. *Trends in Biotechnology*, *42*(11), 1339-1347. https://doi.org/10.1016/j.tibtech.2024.04.007

Krysztof, M., Olejnik, K., Kulpiński, P., Erdman, A., & Sąsiadek, E. (2021). A Comparative study of the effect of cellulose-based deep coating and pulp refining on the structural and mechanical properties of paper. *BioResources*, *16*(3), 5376-5389. https://doi.org/10.15376/biores.16.3.5376-5389

Kumar, A., Tazeb, A., & Ram, C. (2021). Enzyme-assisted pulp refining: An energy saving approach. *Physical Sciences Reviews*, *6*(2), 20190046. https://doi.org/10.1515/psr-2019-0046

Kurtzman, C. P. (2005). Description of *Komagataella phaffii* sp. nov. and the transfer of *Pichia pseudopastoris* to the methylotrophic yeast genus *Komagataella*. *International journal of systematic and evolutionary microbiology*, *55*(2), 973-976. https://doi.org/10.1099/ijs.0.63491-0

Lacerda, L. D., da Silveira, N. P., Bondam, A. F., & Hoffmann, J. F. (2024). Starch gelatinization behavior: The impact of granular structure. *Starch/Staerke*, *76*(7–8), 2300143. https://doi.org/10.1002/star.202300143

Lander, E. S. (2016). The Heroes of CRISPR. *Cell, 164*(1), 18–28. https://doi.org/10.1016/j.cell.2015.12.041

Li, H., Qi, Y., Zhao, Y., Chi, J., & Cheng, S. (2019). Starch and its derivatives for paper coatings: A review. *Progress in Organic Coatings*, *135*, 213-227. https://doi.org/10.1016/j.porgcoat.2019.05.015

Li, Y., Wu, Y., Liu, Y., Li, J., Du, G., Lv, X., & Liu, L. (2022). A genetic toolkit for efficient production of secretory protein in *Bacillus subtilis*. *Bioresource Technology*, *363*, 127885. https://doi.org/10.1016/j.biortech.2022.127885

Li, Z., Pei, X., Zhang, Z., Wei, Y., Song, Y., Chen, L., Liu, S., & Zhang, S. H. (2018). The unique GH5 cellulase member in the extreme halotolerant fungus *Aspergillus glaucus* CCHA is an endoglucanase with multiple tolerance to salt, alkali and heat: prospects for straw degradation applications. *Extremophiles*, *22*(4), 675–685. https://doi.org/10.1007/s00792-018-1028-5

Liu, D., Huang, C., Guo, J., Zhang, P., Chen, T., Wang, Z., & Zhao, X. (2019). Development and characterization of a CRISPR/Cas9n-based multiplex genome editing system for *Bacillus subtilis*. *Biotechnology for Biofuels*, *12*(1), 1–17. https://doi.org/10.1186/s13068-019-1537-1

Liu, H., Wang, S., Song, L., Yuan, H., Liu, K., Meng, W., & Wang, T. (2019). Trehalose production using recombinant trehalose synthase in *Bacillus subtilis* by integrating fermentation and biocatalysis. *Journal of Agricultural and Food Chemistry*, *67*(33), 9314-9324. https://doi.org/10.1021/acs.jafc.9b03402

Liu, R., Zuo, Z., Xu, Y., Song, C., Jiang, H., Qiao, C., Xu, P., Zhou, Q., & Yang, C. (2014). Twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct Tat-dependent secretion of methyl parathion hydrolase. *Journal of Agricultural and Food Chemistry*, *62*(13), 2913-2918. https://doi.org/10.1021/jf405694n

Liu, S., Endo, K., Ara, K., Ozaki, K., & Ogasawara, N. (2008). Introduction of marker-free deletions in *Bacillus subtilis* using the AraR repressor and the *ara* promoter. *Microbiology*, *154*(9), 2562-2570. https://doi.org/10.1099/mic.0.2008/016881-0

Liu, S. L., & Du, K. (2012). Enhanced expression of an endoglucanase in *Bacillus subtilis* by using the sucrose-inducible *sacB* promoter and improved properties of the recombinant enzyme. *Protein Expression and Purification*, *83*(2), 164-168. https://doi.org/10.1016/j.pep.2012.03.015

Liu, X., & Kokare, C. (2023). Microbial enzymes of use in industry. In G. Brahmachari (Ed.), *Biotechnology of microbial enzymes* (2nd ed., pp. 405-444). Academic Press. https://doi.org/10.1016/b978-0-12-803725-6.00011-x

Liu, X., Wang, H., Wang, B., & Pan, L. (2018). Efficient production of extracellular pullulanase in *Bacillus subtilis* ATCC6051 using the host strain construction and promoter optimization expression system. *Microbial Cell Factories*, *17*(1), 163. https://doi.org/10.1186/s12934-018-1011-y

Lv, W., & Cai, M. (2025). Advancing recombinant protein expression in *Komagataella phaffii*: opportunities and challenges. *FEMS Yeast Research*, *25*, foaf010. https://doi.org/10.1093/femsyr/foaf010

Majeke, B. M., García-Aparicio, M., Biko, O. D., Viljoen-Bloom, M., van Zyl, W. H., & Görgens, J. F. (2020). Synergistic codon optimization and bioreactor cultivation toward enhanced secretion of fungal lignin peroxidase in *Pichia pastoris*: Enzymatic valorization of technical (industrial) lignins. *Enzyme and Microbial Technology, 139,* 109593. https://doi.org/10.1016/j.enzmictec.2020.109593

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67-83. https://doi.org/10.1038/s41579-019-0299-x

Malgas, S., Kwanya Minghe, V. M., & Pletschke, B. I. (2020). The effect of hemicellulose on the binding and activity of cellobiohydrolase I, Cel7A, from *Trichoderma reesei* to cellulose. *Cellulose*, *27*(2), 781-797. https://doi.org/10.1007/s10570-019-02848-5

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z.,

Yamashita, R. A., Zhang, D., Zheng, C., & Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(D1), D222-D226. https://doi.org/10.1093/nar/gku1221

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, *19*(6), 679–682. https://doi.org/10.1038/s41592-022-01488-1

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, *60*(2), 174-182. https://doi.org/10.1007/s00239-004-0046-3

Molina-Espeja, P., Sanz-Aparicio, J., Golyshin, P. N., Robles-Martín, A., Guallar, V., Beltrametti, F., Müller, M., Yakimov, M. M., Modregger, J., Van Logchem, M., Corvini, P., Shahgaldian, P., Degering, C., Wieland, S., Timm, A., De Carvalho, C. C., Re, I., Daniotti, S., Thies, S., … Ferrer, M. (2023). Enzymes for consumer products to achieve climate neutrality. *Oxford Open Climate Change*, *3*(1), 3. https://doi.org/10.1093/oxfclm/kgad003

Mondal, S., Mondal, K., Halder, S. K., Thakur, N., & Mondal, K. C. (2022). Microbial amylase: Old but still at the forefront of all major industrial enzymes. *Biocatalysis and Agricultural Biotechnology*, *45*, 102509. https://doi.org/10.1016/j.bcab.2022.102509

Moraes, L. M. P. de, Marques, H. F., Reis, V. C. B., Coelho, C. M., Leitão, M. de C., Galdino, A. S., Porto de Souza, T. P., Piva, L. C., Perez, A. L. A., Trichez, D., de Almeida, J. R. M., De Marco, J. L., & Torres, F. A. G. (2024). Applications of the methylotrophic yeast *Komagataella phaffii* in the context of modern biotechnology. *Journal of Fungi, 10*(6), 411. https://doi.org/10.3390/jof10060411

Motamedian, H. R., Halilovic, A. E., & Kulachenko, A. (2019). Mechanisms of strength and stiffness improvement of paper after PFI refining with a focus on the effect of fines. *Cellulose*, *26*(6), 4099-4124. https://doi.org/10.1007/s10570-019-02349-5

Mu, D., Lu, J., Qiao, M., Kuipers, O. P., Zhu, J., Li, X., Yang, P., Zhao, Y., Luo, S., Wu, X., Jiang, S., & Zheng, Z. (2018). Heterologous signal peptides-directing secretion of *Streptomyces mobaraensis* transglutaminase by *Bacillus subtilis*. *Applied Microbiology and Biotechnology*, *102*(13), 5533-5543. https://doi.org/10.1007/s00253-018-9000-y

Nagl, M., Haske-Cornelius, O., Bauer, W., Nyanhongo, G. S., & Guebitz, G. M. (2023). Enhanced energy savings in enzymatic refining of hardwood and softwood pulp. *Energy, Sustainability and Society*, *13*(1), 19. https://doi.org/10.1186/s13705-023-00398-0

Nagl, M., Haske-Cornelius, O., Skopek, L., Bausch, F., Pellis, A., Bauer, W., Nyanhongo, G. S., & Guebitz, G. M. (2022). Mechanistic investigation of the effect of endoglucanases related to pulp refining. *Cellulose*, *29*(4), 2579-2598. https://doi.org/10.1007/s10570-021-04386-5

Nagl, M., Haske-Cornelius, O., Skopek, L., Pellis, A., Bauer, W., Nyanhongo, G. S., & Guebitz, G. (2021). Biorefining: The role of endoglucanases in refining of cellulose fibers. *Cellulose*, *28*(12), 7633-7650. https://doi.org/10.1007/s10570-021-04022-2

Naseem, M. U., Tajti, G., Gaspar, A., Szanto, T. G., Borrego, J., & Panyi, G. (2021). Optimization of *Pichia pastoris* expression system for high-level production of margatoxin. *Frontiers in Pharmacology*, *12*, 733610. https://doi.org/10.3389/fphar.2021.733610

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I. M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T. B. K., Nielsen, T., Kirton, E., Faria, J. P., Edirisinghe, J. N., Henry, C. S., … Eloe-Fadrosh, E. A. (2021). A genomic catalog of Earth's microbiomes. *Nature Biotechnology*, *39*(4), 499-509. https://doi.org/10.1038/s41587-020-0718-6

Neis, A., & da Silva Pinto, L. (2021). Glycosyl hydrolases family 5, subfamily 5: Relevance and structural insights for designing improved biomass degrading cocktails. *International Journal of Biological Macromolecules*, *193*(Pt A), 980-995. https://doi.org/10.1016/j.ijbiomac.2021.10.062

Nielsen, P. H., Oxenbøll, K. M., & Wenzel, H. (2007). Cradle-to-gate environmental assessment of enzyme

products produced industrially in Denmark by Novozymes A/S. *International Journal of Life Cycle Assessment*, *12*(6), 432-438. https://doi.org/10.1065/lca2006.08.265.1

Nieuwkoop, T., Finger-Bou, M., van der Oost, J., & Claassens, N. J. (2020). The ongoing quest to crack the genetic code for protein production. *Molecular Cell*, *80*(2), 193-209. https://doi.org/10.1016/j.molcel.2020.09.014

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205-217. https://doi.org/10.1006/jmbi.2000.4042

Novozymes A/S. (2025). *Pulp and paper biosolutions*. Novonesis. Retrieved June 2025, from https://www.novonesis.com/en/biosolutions/industrial-bioprocessing/pulp-and-paper

Novozymes A/S. (2025). *BAN-480 LS* [Product page]. Novonesis. Retrieved August 2025, from https://www.novonesis.com/en/biosolutions/industrial-bioprocessing/starch/specialties/ban-480-ls

Okido, T., Kodama, Y., Mashima, J., Kosuge, T., Fujisawa, T., & Ogasawara, O. (2022). DNA Data Bank of Japan (DDBJ) update report 2021. *Nucleic Acids Research*, *50*(D1), D102-D105. https://doi.org/10.1093/nar/gkab995

Park, S. G., Tak, J. H., Lee, J. Y., Shin, K. S., & Park, S. Il. (2025). Evaluation of cellulase effect on the refining process of softwood bleached kraft pulp. *BioResources*, *20*(1), 1059-1068. https://ojs.bioresources.com/index.php/brj/article/view/24069

Patel, A. K., Dong, C.-D., Chen, C.-W., Pandey, A., & Singhania, R. R. (2023). Production, purification, and application of microbial enzymes. In G. Brahmachari (Ed.), *Biotechnology of microbial enzymes* (2nd ed., pp. 25–57). Academic Press. https://doi.org/10.1016/b978-0-443-19059-9.00019-0

Patel, A., & Shah, A. (2021). Purification and characterization of novel, thermostable and non-processive GH5 family endoglucanase from *Fomitopsis meliae* CFA 2. *International Journal of Biological Macromolecules*, *182*, 1161-1169. https://doi.org/10.1016/j.ijbiomac.2021.04.110

Pérez-Contreras, S., Avalos-de la Cruz, D. A., Lizardi-Jiménez, M. A., Herrera-Corredor, J. A., Baltazar-Bernal, O., & Hernández-Martínez, R. (2024). Production of ligninolytic and cellulolytic fungal enzymes for agro-industrial waste valorization: Trends and applicability. *Catalysts*, *15*(1), 30. https://doi.org/10.3390/catal15010030

Phaff, H. J., Miller, M. W., & Shifrine, M. (1956). The taxonomy of yeasts isolated from *Drosophila* in the Yosemite region of California. *Antonie van Leeuwenhoek*, *22*(1), 145-161. https://doi.org/10.1007/bf02538322

Pham, V. H. T., & Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends in Biotechnology*, *30*(9), 475-484. https://doi.org/10.1016/j.tibtech.2012.05.007

Phan, T. T. P., Nguyen, H. D., & Schumann, W. (2012). Development of a strong intracellular expression system for *Bacillus subtilis* by optimizing promoter elements. *Journal of Biotechnology*, *157*(1), 167-172. https://doi.org/10.1016/j.jbiotec.2011.10.006

Picart, P., Diaz, P., & Pastor, F. I. J. (2008). *Stachybotrys atra* BP-A produces alkali-resistant and thermostable cellulases. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *94*(2), 307-316. https://doi.org/10.1007/s10482-008-9248-9,

Planas, A. (2023). Glycoside hydrolases: Mechanisms, specificities, and engineering. In A. Goyal & K. Sharma (Eds.), *Foundations and frontiers in enzymology: Glycoside hydrolases* (pp. 25–53). Academic Press. https://doi.org/10.1016/b978-0-323-91805-3.00011-3

Popp, P. F., Dotzler, M., Radeck, J., Bartels, J., & Mascher, T. (2017). The *Bacillus* BioBrick Box 2.0: Expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. *Scientific Reports, 7*(1), 15058. https://doi.org/10.1038/s41598-017-15107-z

Pothakos, V., Debeer, N., Debonne, I., Rodriguez, A., Starr, J. N., & Anderson, T. (2018). Fermentation titer optimization and impact on energy and water consumption during downstream processing. *Chemical Engineering & Technology*, *41*(12), 2358. https://doi.org/10.1002/ceat.201800279

Powell, H. R., Islam, S. A., David, A., & Sternberg, M. J. E. (2025). Phyre2.2: A community resource for template-based protein structure prediction. *Journal of Molecular Biology*, 168960. https://doi.org/10.1016/j.jmb.2025.168960

Quesada-Ganuza, A., Antelo-Varela, M., Mouritzen, J. C., Bartel, J., Becher, D., Gjermansen, M., Hallin, P. F., Appel, K. F., Kilstrup, M., Rasmussen, M. D., & Nielsen, A. K. (2019). Identification and optimization of PrsA in *Bacillus subtilis* for improved yield of amylase. *Microbial Cell Factories*, *18*(1), 158. https://doi.org/10.1186/s12934-019-1203-0

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology, 35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Dürr, F., Emenegger, J., Kelterborn, S., Sauer, C., Fritz, G., Gebhard, S., & Mascher, T. (2013). The *Bacillus* BioBrick Box: Generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *Journal of Biological Engineering*, *7*(1), 29. https://doi.org/10.1186/1754-1611-7-29

Rhizobium, G. E. (2025). UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Research*, *53*(D1), 13–14. https://doi.org/10.1093/nar/gkae1010

Robinson, S. L., Piel, J., & Sunagawa, S. (2021). A roadmap for metagenomic enzyme discovery. *Natural Product Reports*, *38*(11), 1994-2023. https://doi.org/10.1039/d1np00006c

Rodrigues, V. J., & Odaneth, A. A. (2021). Industrial application of cellulases. *Current Status and Future Scope of Microbial Cellulases*, 189-209. https://doi.org/10.1016/b978-0-12-821882-2.00007-7

Rosignoli, S., & Paiardini, A. (2022). Boosting the full potential of PyMOL with structural biology plugins. *Biomolecules*, *12*(12), 1764. https://doi.org/10.3390/biom12121764

Saira, O., & Vauhkonen, C. (2024). Cellulase-assisted refining in a paperboard mill: Avoided emissions from energy savings. A case study of a Finnish paperboard mill. *Cleaner Engineering and Technology*, *20*, 100760. https://doi.org/10.1016/j.clet.2024.100760

Saito, A., Hizukuri, Y., Matsuo, E. I., Chiba, S., Mori, H., Nishimura, O., Ito, K., & Akiyama, Y. (2011). Post-liberation cleavage of signal peptides is catalyzed by the site-2 protease (S2P) in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13740–13745. https://doi.org/10.1073/pnas.1108376108

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, *27*(10), 946-950. https://doi.org/10.1038/nbt.1568

Samanta, S., Basu, A., Halder, U. C., & Sen, S. K. (2012). Characterization of *Trichoderma reesei* endoglucanase ii expressed heterologously in *Pichia pastoris* for better biofinishing and biostoning. *Journal of Microbiology*, *50*(3), 518-525. https://doi.org/10.1007/s12275-012-1207-5

Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Schoch, C. L., Sherry, S. T., & Karsch-Mizrachi, I. (2022). GenBank. *Nucleic Acids Research*, *50*(D1), D161–D164. https://doi.org/10.1093/nar/gkab1135

Schallmey, M., Singh, A., & Ward, O. P. (2004). Developments in the use of *Bacillus* species for industrial production. *Canadian journal of microbiology*, *50*(1), 1-17. https://doi.org/10.1139/w03-076

Schumann, W. (2007). Production of recombinant proteins in *Bacillus subtilis*. *Advances in Applied Microbiology*, *62*, 137-189. https://doi.org/10.1016/s0065-2164(07)62006-1

Schütz, A., Bernhard, F., Berrow, N., Buyel, J. F., Ferreira-da-Silva, F., Haustraete, J., van den Heuvel, J., Hoffmann, J. E., de Marco, A., Peleg, Y., Suppmann, S., Unger, T., Vanhoucke, M., Witt, S., & Remans, K. (2023). A concise guide to choosing suitable gene expression systems for recombinant protein

production. *STAR Protocols*, *4*(4), 102572. https://doi.org/10.1016/j.xpro.2023.102572

Singh, N., Lane, S., Yu, T., Lu, J., Ramos, A., Cui, H., & Zhao, H. (2025). A generalized platform for artificial intelligence-powered autonomous enzyme engineering. *Nature Communications, 16*(1), 5648. https://doi.org/10.1038/s41467-025-61209-y

Singh, R. S., Singh, T., & Pandey, A. (2019). Microbial enzymes—An overview. In R. S. Singh, R. R. Singhania, A. Pandey, & C. Larroche (Eds.), *Advances in enzyme technology* (pp. 1–40). Elsevier. https://doi.org/10.1016/b978-0-444-64114-4.00001-7

Siziya, I. N., Hwang, C. Y., & Seo, M. J. (2022). Antioxidant potential and capacity of microorganism-sourced C_{30} carotenoids—A Review. *Antioxidants, 11*(10), 1963. https://doi.org/10.3390/antiox11101963

Song, Y., He, S., Jopkiewicz, A., Setroikromo, R., van Merkerk, R., & Quax, W. J. (2022). Development and application of CRISPR-based genetic tools in *Bacillus* species and *Bacillus* phages. *Journal of Applied Microbiology*, *133*(4), 2280-2298. https://doi.org/10.1111/jam.15704

Stülke, J., Grüppen, A., Bramkamp, M., & Pelzer, S. (2023). *Bacillus subtilis*, a Swiss army knife in science and biotechnology. *Journal of Bacteriology*, *205(5),* e00102-23. https://doi.org/10.1128/jb.00102-23

Su, Y., Liu, C., Fang, H., & Zhang, D. (2020). *Bacillus subtilis*: A universal cell factory for industry, agriculture, biomaterials and medicine. *Microbial Cell Factories*, *19*(1), 173. https://doi.org/10.1186/s12934-020-01436-8

Sutaoney, P., Rai, S. N., Sinha, S., Choudhary, R., Gupta, A. K., Singh, S. K., & Banerjee, P. (2024). Current perspective in research and industrial applications of microbial cellulases. *International Journal of Biological Macromolecules*, *264*(Pt 1), 130639. https://doi.org/10.1016/j.ijbiomac.2024.130639

Suzuki, H., Murakami, A., & Yoshida, K. I. (2012). Counterselection system for *Geobacillus kaustophilus* HTA426 through: Disruption of *pyrF* and *pyrR*. *Applied and Environmental Microbiology*, *78*(20), 7376-7383. https://doi.org/10.1128/aem.01669-12

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, *38*(7), 3022-3027. https://doi.org/10.1093/molbev/msab120

Thakur, M., Brooksbank, C., Finn, R. D., Firth, H. V., Foreman, J., Freeberg, M., Gurwitz, K. T., Harrison, M., Hulcoop, D., Hunt, S. E., Leach, A. R., Levchenko, M., Marques, D., McDonagh, E. M., Mithani, A., Parkinson, H., Perez-Riverol, Y., Perova, Z., Sarkans, U., … McEntyre, J. (2025). EMBL's European Bioinformatics Institute (EMBL-EBI) in 2024. *Nucleic Acids Research*, *53*(D1), D10–D19. https://doi.org/10.1093/nar/gkae1089

Tian, J., Xing, B., Li, M., Xu, C., Huo, Y. X., & Guo, S. (2022). Efficient large-scale and scarless genome engineering enables the construction and screening of *Bacillus subtilis* biofuel overproducers. *International Journal of Molecular Sciences*, *23*(9), 4853. https://doi.org/10.3390/ijms23094853/s1

Toymentseva, A. A., & Altenbuchner, J. (2019). New CRISPR-Cas9 vectors for genetic modifications of *Bacillus* species. *FEMS Microbiology Letters*, *366*(1), fny284. https://doi.org/10.1093/femsle/fny284

Tsukazaki, T., Mori, H., Echizen, Y., Ishitani, R., Fukai, S., Tanaka, T., Perederina, A., Vassylyev, D. G., Kohno, T., Maturana, A. D., Ito, K., & Nureki, O. (2011). Structure and function of a membrane component SecDF that enhances protein export. *Nature*, *474*(7350), 235-238. https://doi.org/10.1038/nature09980

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. https://sdgs.un.org/2030agenda

Vitikainen, M., Pummi, T., Airaksinen, U., Wahlström, E., Wu, H., Sarvas, M., & Kontinen, V. P. (2001). Quantitation of the capacity of the secretion apparatus and requirement for PrsA in growth and secretion of α-amylase in *Bacillus subtilis*. *Journal of Bacteriology*, *183*(6), 1881-1890. https://doi.org/10.1128/jb.183.6.1881-1890.2001

Wang, B., Kes, M. B., van Saparoea, A. C. V. d. B., Dugar, G., Luirink, J., & Hamoen, L. W. (2024). Inactivation of the conserved protease LonA increases production of xylanase and amylase in *Bacillus subtilis*. *Microbial Cell Factories, 23*(1), 335. https://doi.org/10.1186/s12934-024-02616-6

Wang, M., Du, J., Zhang, D., Li, X., & Zhao, J. (2017). Modification of different pulps by homologous overexpression alkali-tolerant endoglucanase in *Bacillus subtilis* Y106. *Scientific Reports*, *7*(1), 3321. https://doi.org/10.1038/s41598-017-03215-9

Wang, T., Wang, F., Ma, R., & Tian, Y. (2022). Enzymatically modified starch for paper surface sizing: Enzymes with different action modes and sites. *Carbohydrate Polymers*, *291*, 119636. https://doi.org/10.1016/j.carbpol.2022.119636

Wang, Y., Dong, J., Jin, Z., & Bai, Y. (2023). Analysis of the action pattern of sequential α-amylases from *B. stearothermophilus* and *B. amyloliquefaciens* on highly concentrated soluble starch. *Carbohydrate Polymers, 320,* 121190. https://doi.org/10.1016/j.carbpol.2023.121190

Wang, Z., Xie, D., Wu, D., Luo, X., Wang, S., Li, Y., Yang, Y., Li, W., & Zheng, L. (2025). Robust enzyme discovery and engineering with deep learning using CataPro. *Nature Communications, 16*, 2736. https://doi.org/10.1038/s41467-025-58038-4

Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V., & Cregg, J. M. (1997). Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene*, *186*(1), 37-44. https://doi.org/10.1016/s0378-1119(96)00675-0

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296-W303. https://doi.org/10.1093/nar/gky427

Watzlawick, H., & Altenbuchner, J. (2019). Multiple integration of the gene ganA into the *Bacillus subtilis* chromosome for enhanced β-galactosidase production using the CRISPR/Cas9 system. *AMB Express*, *9*(1), 158. https://doi.org/10.1186/s13568-019-0884-4

Widner, B., Thomas, M., Sternberg, D., Lammon, D., Behr, R., & Sloma, A. (2000). Development of marker-free strains of *Bacillus subtilis* capable of secreting high levels of industrial enzymes. *Journal of Industrial Microbiology and Biotechnology*, *25*(4), 204-212. https://doi.org/10.1038/sj.jim.7000051

Wowra, K., Hegel, E., Scharf, A., Grünberger, A., & Rosenthal, K. (2023). Estimating environmental impacts of early-stage bioprocesses. *Trends in Biotechnology*, *41*(9), 1199-1212. https://doi.org/10.1016/j.tibtech.2023.03.011

Wu, S. C., Yeung, J. C., Duan, Y., Ye, R., Szarka, S. J., Habibi, H. R., & Wong, S. L. (2002). Functional production and characterization of a fibrin-specific single-chain antibody fragment from *Bacillus subtilis*: Effects of molecular chaperones and a wall-bound protease on antibody fragment production. *Applied and Environmental Microbiology*, *68*(7), 3261-3269. https://doi.org/10.1128/AEM.68.7.3261-3269.2002

Wu, Y., Liu, Y., Lv, X., Li, J., Du, G., & Liu, L. (2020). CAMERS-B: CRISPR/Cpf1 assisted multiple-genes editing and regulation system for *Bacillus subtilis*. *Biotechnology and Bioengineering*, *117*(6), 1817-1825. https://doi.org/10.1002/bit.27322

Xu, X., Meier, F., Blount, B. A., Pretorius, I. S., Ellis, T., Paulsen, I. T., & Williams, T. C. (2023). Trimming the genomic fat: Minimising and re-functionalising genomes using synthetic biology. *Nature Communications*, *14*(1), 1984. https://doi.org/10.1038/s41467-023-37748-7

Yan, S., & Wu, G. (2017). Bottleneck in secretion of α-amylase in *Bacillus subtilis*. *Microbial Cell Factories*, *16*(1), 124. https://doi.org/10.1186/s12934-017-0738-1

Yan, X., Yu, H. J., Hong, Q., & Li, S. P. (2008). Cre/lox system and PCR-based genome engineering in *Bacillus subtilis*. *Applied and Environmental Microbiology*, *74*(17), 5556-5562. https://doi.org/10.1128/aem.01156-08

Yang, M., Li, J., Wang, S., Zhao, F., Zhang, C., Zhang, C., & Han, S. (2023). Status and trends of enzyme cocktails for efficient and ecological production in the pulp and paper industry. *Journal of Cleaner Production*, *418*, 138196. https://doi.org/10.1016/j.jclepro.2023.138196

Yang, S., Du, G., Chen, J., & Kang, Z. (2017). Characterization and application of endogenous phase-dependent promoters in *Bacillus subtilis*. *Applied Microbiology and Biotechnology*, *101*(10), 4151-4161. https://doi.org/10.1007/s00253-017-8142-7

Yao, D., Han, X., Gao, H., Wang, B., Fang, Z., Li, H., Fang, W., & Xiao, Y. (2023). Enhanced extracellular production of raw starch-degrading α-amylase in *Bacillus subtilis* through expression regulatory element modification and fermentation optimization. *Microbial Cell Factories*, *22*(1), 118. https://doi.org/10.1186/s12934-023-02116-z

Yomantas, Y. A. V., Abalakina, E. G., Golubeva, L. I., Gorbacheva, L. Y., & Mashko, S. V. (2011). Overproduction of *Bacillus amyloliquefaciens* extracellular glutamyl-endopeptidase as a result of ectopic multi-copy insertion of an efficiently-expressed mpr gene into the *Bacillus subtilis* chromosome. *Microbial Cell Factories*, *10*(1), 64. https://doi.org/10.1186/1475-2859-10-64

Yoshida, K., Ueda, S., & Maeda, I. (2009). Carotenoid production in *Bacillus subtilis* achieved by metabolic engineering. *Biotechnology Letters*, *31*(11), 1789-1793. https://doi.org/10.1007/s10529-009-0082-6

Young, M. (1984). Gene amplification in *Bacillus subtilis*. *Journal of General Microbiology*, *130*(7), 1613-1621. https://doi.org/10.1099/00221287-130-7-1613

Yu, X., Xu, J., Liu, X., Chu, X., Wang, P., Tian, J., Wu, N., & Fan, Y. (2015). Identification of a highly efficient stationary phase promoter in *Bacillus subtilis*. *Scientific Reports*, *5*(1), 1-9. https://doi.org/10.1038/srep18405

Yue, J., Fu, G., Zhang, D., & Wen, J. (2017). A new maltose-inducible high-performance heterologous expression system in *Bacillus subtilis*. *Biotechnology Letters*, *39*(8), 1237–1244. https://doi.org/10.1007/s10529-017-2357-7

Zeigler, D. R., Prágai, Z., Rodriguez, S., Chevreux, B., Muffler, A., Albert, T., Bai, R., Wyss, M., & Perkins, J. B. (2008). The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *Journal of Bacteriology*, *190*(21), 6983-6995. https://doi.org/10.1128/jb.00722-08

Zeng, X., Retulainen, E., Heinemann, S., & Fu, S. (2012). Fibre deformations induced by different mechanical treatments and their effect on zero-span strength. *Nordic Pulp and Paper Research Journal*, *27*(2), 335-342. https://doi.org/10.3183/npprj-2012-27-02-p335-342

Zhang, C., Zhang, X., Yao, Z., Lu, Y., Lu, F., & Lu, Z. (2011). A new method for multiple gene inactivations in *Bacillus subtilis* 168, producing a strain free of selectable markers. *Canadian Journal of Microbiology*, *57*(5), 427-436. https://doi.org/10.1139/W11-035

Zhang, K., Duan, X., & Wu, J. (2016). Multigene disruption in undomesticated *Bacillus subtilis* ATCC 6051a using the CRISPR/Cas9 system. *Scientific Reports*, *6*(1), 27943. https://doi.org/10.1038/srep27943

Zhang, K., Su, L., & Wu, J. (2020). Recent advances in recombinant protein production by *Bacillus subtilis. Annual review of food science and technology, 11*(1), 295-318. https://doi.org/10.1146/annurev-food-032519

Zhang, K., Luo, H., Zhu, X., Liu, W., Yu, X., Tao, W., Lin, H., Hou, M., & Wu, J. (2024). Construction of *Bacillus subtilis* chassis strain with enhanced α-amylase expression capability based on CRISPRi screening. *International Journal of Biological Macromolecules, 283*(Pt 1), 137497. https://doi.org/10.1016/j.ijbiomac.2024.137497

Zhang, X. Z., Yan, X., Cui, Z. L., Hong, Q., & Li, S. P. (2006). *mazF*, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. *Nucleic Acids Research*, *34*(9), e71-e71. https://doi.org/10.1093/nar/gkl358

Zheng, Jia, Zhao, W., Guo, N., Lin, F., Tian, J., Wu, L., & Zhou, H. (2012). Development of an industrial

medium and a novel fed-batch strategy for high-level expression of recombinant β-mananase by *Pichia pastoris*. *Bioresource Technology*, *118*, 257-264. https://doi.org/10.1016/j.biortech.2012.05.065

Zheng, Jie, Liu, H.-Q., Qin, X., Yang, K., Tian, J., Wang, X.-L., Wang, Y.-R., Wang, Y., Yao, B., Luo, H.-Y., & Huang, H.-Q. (2022). Identification and mutation analysis of nonconserved residues on the tim-barrel surface of GH5_5 cellulases for catalytic efficiency and stability improvement. *Applied and Environmental Microbiology*, *88*(17), e01046-22. https://doi.org/10.1128/aem.01046-22

Zhiguang, C., Haixia, Z., Min, C., Fayong, G., & Jing, L. (2025). The fine structure of starch: A review. *Npj Science of Food*, *9*(1), 50. https://doi.org/10.1038/s41538-025-00414-x

Zhou, W., Tong, S., Amin, F. R., Chen, W., Cai, J., & Li, D. (2023). Heterologous expression and biochemical characterization of a thermostable endoglucanase (MtEG5-1) from *Myceliophthora thermophila*. *Fermentation*, *9*(5), 462. https://doi.org/10.3390/fermentation9050462

Zhu, B., Wang, D., & Wei, N. (2022). Enzyme discovery and engineering for sustainable plastic recycling. *Trends in Biotechnology*, *40*(1), 22-37. https://doi.org/10.1016/j.tibtech.2021.02.008

Zou, Y., Qiu, L., Xie, A., Han, W., Zhang, S., Li, J., Zhao, S., Li, Y., Liang, Y., & Hu, Y. (2022). Development and application of a rapid all-in-one plasmid CRISPR-Cas9 system for iterative genome editing in *Bacillus subtilis*. *Microbial Cell Factories*, *21*(1), 173. https://doi.org/10.1186/s12934-022-01896-0

UNIVERSITAT DE BARCELONA

Sustainable Biotechnology and Bioremediation Group

DOCTORATS INDUSTRIALS
RECERCA ESTRATÈGICA · COL·LABORATIVA · APLICADA

Generalitat de Catalunya
Departament de Recerca i Universitats