Integration of Biomass Conversion, Cell Factory Optimization and Bioprocess Design to Innovate the Valorization of Lignocellulosic Fibers

Driessen, Jasper

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Integration of Biomass Conversion, Cell Factory Optimization and Bioprocess Design to Innovate the Valorization of Lignocellulosic Fibers

PhD thesis

Jasper Leonard Simon Pieternel Driessen

The Novo Nordisk Center for Biosustainability

Technical University of Denmark

April 2022

Integration of Biomass Conversion, Cell Factory Optimization and Bioprocess Design to Innovate the Valorization of Lignocellulosic Fibers

PhD thesis was written by
Jasper Leonard Simon Pieternel Driessen

Supervised by
Professor Alex Toftgaard Nielsen
Senior Scientist Sheila Ingemann Jensen
Professor John Woodley

Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

PREFACE

This thesis is written as partial fulfilment of the requirements to obtain the degree of Doctor of Philosophy at the Technical University of Denmark. The work presented in this thesis was carried out between October 2018 and April 2022 at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark, Kongens Lyngby, Denmark, and during a research stay from October 2021 to December 2021 at the industrial partner International Flavors & Fragrances (IFF) Leiden, The Netherlands.

The work was supervised by Prof. Alex Toftgaard Nielsen, professor at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark and co-supervised by Sheila Ingemann Jensen Ph.D., senior scientist at the Technical University of Denmark and Prof. John Woodley, professor at the Department of Chemical and Biochemical Engineering of the Technical University of Denmark.

This project was funded by the Novo Nordisk Foundation within the Fermentation Based Biomanufacturing Initiative (FBM) framework, grant number NNF17SA0031362. The external stay was partially funded by the William Demant Foundation, Augustinus Foundation and Otto Mønsteds Foundation.

Jasper Driessen M.Sc. (WUR)

Kgs. Lyngby, April 2022

ASSESSMENT COMMITTEE

Senior researcher Group leader	Morten Nørholm[1]
Vice-President Global R&D	Casper Vroemen[2]
Professor Head of Division	Eva Nordberg Karlsson[3]

Affiliation:

[1]Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark
Kongens Lyngby, Denmark

[2] Health & Biosciences
International Flavors & Fragrances
Palo Alto, CA, USA

[3] Division of Biotechnology
Lund University
Lund, Sweden

ABSTRACT

If humankind wants to sustain itself on planet earth, a revolution in the way we produce our food, fuels, chemicals, and materials is indispensable. In this context, the replacement of fossil fuel-driven processes by fermentation based biomanufacturing using renewable biomass is one of the key technologies we have at our disposal to realise more sustainable practices. Second-generation biorefineries are likely to play a major role in future bio-manufacturing processes as they do not compete with food crops, but use cheap, widely available and inedible lignocellulosic biomass as feedstock. Despite the vast potential of second-generation biorefineries and years of dedicated research by academia and industry, the realisation of cost-effective commercial-scale second-generation biorefineries has been hampered by interactive technical, economic, and political challenges.

This thesis aims to contribute to overcoming some of the major hurdles associated with second-generation biorefineries, being the toxicity of biomass hydrolysates and the high cost of enzymes required in the process. The main work was done from the perspective of a case study, involving an on-site enzyme production process, using DDGS biomass as a fermentation feedstock. Regardless, the main findings in the specific context can have broader implications for the use of lignocellulose.

Lab-scale experiments were performed to compare and optimise different pretreatment setups. The experimental data was used to develop and compare data- and knowledge-driven computational models to aid the optimisation of pretreatment process parameters (Chapter 2). Next, the individual and combined effects of toxic compounds commonly released during biomass pretreatment were studied for *Bacillus subtilis* (Chapter 3). Subsequently, adaptive laboratory evolution of *B. subtilis* was used as a tool to obtain strains with increased tolerance towards biomass-associated, inhibitory compounds. Moreover, evolved strains not only showed increased tolerance, but were also capable of producing significant amounts of protein using biomass hydrolysate media (Chapter 4).

Finally, the cost-effectiveness of DDGS-based on-site enzyme production was explored by performing a techno-economic assessment (Chapter 5).

The presented results lead to an increased understanding of the toxicity of biomass hydrolysates and cost-effective enzyme production. As these are two major hurdles associated with the use of lignocellulose, this thesis contributes to the advancement of future second-generation biorefineries.

DANSK SAMMENFATNING

Hvis menneskeheden ønsker at bevare sig selv på jordkloden, er en revolution af den måde, vi producerer vores fødevarer, brændstof, kemikalier og materialer på, uundværlig. I denne sammenhæng er erstatningen af fossile processer med fermenteringsbaseret biofremstilling ved brug af vedvarende biomasse en af de nøgleteknologier, vi har til rådighed for at opnå en mere bæredygtig praksis. Anden generations bioraffinaderier vil sandsynligvis spille en stor rolle i fremtidige biofremstillingsprocesser, da de ikke konkurrerer med fødevareafgrøder, men bruger billig, vidt tilgængelig og uspiselig lignocelluloseholdig biomasse som råmateriale. På trods af det store potentiale i andengenerations-bioraffinaderier og mange års dedikeret forskning i den akademiske verden og industrien er realiseringen af omkostningseffektive andengenerations-bioraffinaderier i kommerciel skala blevet hæmmet af interaktive tekniske, økonomiske og politiske udfordringer.

Formålet med denne afhandling er at bidrage til at løse nogle af de største forhindringer forbundet med andengenerations bioraffinaderier, nemlig toksiciteten af biomassehydrolysater og de høje omkostninger til de enzymer, der kræves i processen. Hovedarbejdet blev udført i perspektivet af et casestudie, der involverede en on-site enzymproduktionsproces, ved brug af DDGS-biomasse som fermenteringsråmateriale. Resultaterne vil i den specifikke kontekst kunne have bredere implikationer for brugen af lignocellulose.

Eksperimenter i laboratorieskala blev udført for at sammenligne og optimere forskellige forbehandlingsopsætninger. De eksperimentelle data blev brugt til udvikling og sammenligning af data- og vidensdrevne beregningsmodeller for at hjælpe med optimering af forbehandlingsprocesparametre (Kapitel 2). Dernæst blev de individuelle og kombinerede virkninger af toksiske stoffer, der almindeligvis frigives under biomasseforbehandling, undersøgt for *Bacillus subtilis* (Kapitel 3). Efterfølgende blev adaptiv laboratorieevolution af *B. subtilis* brugt som et værktøj til at opnå stammer med en øget tolerance over for biomasse-associerede hæmmende stoffer. Desuden viste de udviklede stammer ikke kun øget tolerance, men var også i stand til at producere betydelige mængder protein ved hjælp af biomassehydrolysatmedier (Kapitel 4).

Endelig blev omkostningseffektiviteten af DDGS-baseret on-site enzymproduktion undersøgt ved at udføre en teknoøkonomisk vurdering (Kapitel 5).

De præsenterede resultater fører til en øget forståelse af toksiciteten af biomassehydrolysater og omkostningseffektiv enzymproduktion. Da disse er to store forhindringer forbundet med brugen af lignocellulose, bidrager denne afhandling til fremme af andengenerations bioraffinaderier i fremtiden.

NEDERLANDSE SAMENVATTING

Als de mensheid zichzelf staande wil houden op onze planeet, is er een drastische ommekeer nodig in de manier waarop we ons voedsel, onze brandstoffen, chemicaliën en materialen produceren.

In dit opzicht is de omzetting van biomassa tot deze broodnodige producten door microben een van de belangrijkste technologieën die we tot onze beschikking hebben om fossiele brandstoffen te vervangen en een duurzamere samenleving waar te maken.

Bioraffinage van lignocellulose zal een belangrijke rol spelen in toekomstige bioproductieprocessen, omdat het gebruik van deze oneetbare, plantaardige restproducten niet concurreert met onze voedselvoorziening. Ondanks de enorme kansen die bioraffinage van lignocellulose biedt en het jarenlange onderzoek dat is uitgevoerd door zowel academia als industrie, loopt de verwezenlijking van kosten-effectieve processen op commerciële schaal spaak. Dit komt door een samenspel van technische, economische en politieke struikelblokken.

Het doel van dit proefschrift is om bij te dragen aan de oplossing van enkele van deze problemen, namelijk de giftige stoffen die tijdens de voorbehandeling van biomassa vrij komen en de hoge kosten van enzymen die nodig zijn tijdens het omzettingsproces van lignocellulose. Voor het proefschrift werd er gebruik gemaakt van een casus, waarbij DDGS-biomassa (een bijproduct van de bioethanol-industrie) werd gebruikt als grondstof om enzymen lokaal op de bioraffinaderij te produceren. Ondanks het feit dat de belangrijkste bevindingen van het proefschrift betrekking hebben op deze casus, zijn de uitkomsten ook relevant voor het gebruik van lignocellulose in het algemeen.

Ten eerste werden experimenten uitgevoerd op laboratoriumschaal om verschillende voorbehandelingen van lignocellulose te vergelijken en te optimaliseren. De experimentele gegevens werden gebruikt voor de ontwikkeling van data- en kennisgedreven rekenmodellen om de variabelen van het voorbehandelingsproces te optimaliseren (Hoofdstuk 2). Vervolgens werden de individuele en gecombineerde toxische effecten van biomassa-gerelateerde stoffen op *Bacillus subtilis* bestudeerd (Hoofdstuk 3). Aansluitend werd adaptieve laboratoriumevolutie gebruikt als methode om stammen van *B. Subtilis* te verkrijgen met een verhoogde tolerantie voor deze toxische stoffen. Geëvolueerde stammen vertoonden niet alleen een verhoogde tolerantie, maar waren ook in staat om DDGS-biomassa te gebruiken voor de productie van significante hoeveelheden enzymen (Hoofdstuk 4).

Ten slotte werd de kosten-effectiviteit van het gebruik van DDGS-biomassa voor lokale enzymproductie onderzocht door middel van een technisch-economisch onderzoek (Hoofdstuk 5).

De bevindingen van dit proefschrift dragen bij aan een beter begrip van de toxiciteit van DDGS biomassahydrolysaten enerzijds, en hoe DDGS biomassa gebruikt kan worden als grondstof voor kosteneffectieve enzyme productie anderzijds. Aangezien beide een belangrijke hindernis vormen bij het gebruik van lignocellulose, draagt dit proefschrift bij aan de ontwikkeling van toekomstige bioraffinaderijen.

ACKNOWLEDGEMENTS

As everyone can tell from fascinating books they could not stop reading, movies that made them sit on the edge of their couch or music they fell spontaneously in love with: an intriguing story is not made of mere facts and figures. What makes a story worthwhile is the people in it. The quality of the former part of my story is for the assessment committee to judge. The latter component, however, is my privilege to write about.

First and foremost, I would like to express my deepest gratitude to my mentors during my PhD, Alex and Sheila. Calling you only supervisors doesn't suffice: besides scientific guidance, you offered a lot of room for personal growth and career advice. Taking me under your wings when I asked for help is something I will never forget. Alex, thank you for your contagious fascination with science and the world of biotech. Your positive energy and selfless way of looking at the person behind the student always ensured I left our meetings with new perspectives and belief in myself. Sheila, the way you help your students develop into highly skilled scientists, is truly admirable: believing in their capabilities to find their way and precisely knowing when to correct them. It's how you taught great molecular biologists in the past, and I am sure many will follow them in the future. Thanks for your patience, kindness and willingness to help.

I would also like to thank my co-supervisor, John Woodley, for the passionate introduction to the world of bioprocess design and supporting our ideas related to the FBM outreach mission. I really appreciate your enthusiasm for students' initiative and your efforts to boost interdepartmental research.

I want to thank all the new and old members of the BCF-group for the incredibly fun and caring environment we work in. Vivienne, thanks for borrowing your talented mind many times to me during the years, for the fun-factor: "foute hitjes in het lab", and of course, showing me the path North in the first place. Yixin, thanks a lot for your FBM-commitment and most of all, your excellent sense of humour; it was a treat every time. Philip, kerel, I appreciate your willingness to always help group members and openness for discussion a lot! Arrate, thanks for being the positive-minded, friendly, Basque spice of the lot and for all the little boomer moments we shared. Amalie, the McDrive-adventure was legendary. Ivan, many thanks for showing me the initial steps of molbio. Lucas, thanks for your bright ideas and help as an intern, taking over the responsibilities of the FBM-club, the bike advice, but foremost, putting friendship above all in times of trouble!

I would also express my gratitude to Celina and Zhijia for their help during my biomass experiments and Francesco for his help during his MSc thesis. Fernando, thanks for all the fun times we had together and the supreme Uruguayan BBQ; it has not been challenged since.

Next, I would like to thank all the PhD students at CfB for creating such an incredibly fun community of bright people to be part of. Thanks, Anja, (s)C(h)arolyn, Chris, Sophia, Brain, Matis and all others in the Beach Business for making sure friendships made at CfB are even improved outside of work.

Thank you, Morten, Rebecca and Susanne, for supporting all PhD students to the fullest. I am genuinely thankful for how you have supported me during the most challenging times.

Thanks to the FBM-community as a whole. Friederike, thank you so much for encouraging and supporting our crazy initiatives from the start! Thanks to the PhD students for creating a community where we can learn more from each other, than on our own. Thank you, Bjarke, Dina, Andreas and John, for challenging us to improve our outreach ideas and supporting them to the end.

Special thanks to Bram, Niko and Kate for their warm friendship. Beerke, although we already knew each other through and through, this academic adventure, with the high peaks and deep lows, strengthened our friendship even more. Thanks for always being the listening ear and support, the occasional moral mirror, and the joyful partner in crime. Thank you, Niko, for being the passionate, intelligent (secretly idealistic), firecracker you are. I think we did what FBM was made for, discussing science from different angles and creating synergy, while enjoying each other's company to the fullest. Thanks, both, for sparking the FBM outreach mission; I think it is something we can be proud of! Eeekaterina, thank you for being the intelligent, creative and positive-minded butterfly of science and all our adventures. I promise never to call you a colleague again.

Thanks to my housemates, the Heroes of Copenhagen, for being the little family while being abroad. Thanks, Ken, Joelle, and Maud for always listening to the tired PhD student coming home and making him laugh while enjoying many meals (and drinks) together. Max, from the moment I saw a Dutch name on the screen to the very last steps of the road, we shared the Danish journey and helped each other out. Thanks for all the mood-boosting runs, "being-done-with-life" couch potato sessions, bright inputs and support along the way. Juju, merci beaucoup for being the much-needed and caring friend. Thanks for inviting me to your home, celebrating your Pacs in Champagne

was one of the highlights of my PhD-time. After submitting this, the first thing I do is to open one of the French treats.

Many thanks to my industrial collaborators at IFF. Martijn, Bart, Marc, Sharief and Hans, thanks for your support of the project, for sharing your valuable input, and hosting me in Leiden during my external stay. Pauline and Casper, thank you for all your encouraging advice to follow the PhD-path during my time in Palo Alto. It was, in the end, the good choice indeed.

Bedankt ook, iedereen uit Wageningen, Limburg en de rest van Nederland voor alle steun die ik heb ontvangen tijdens het Deense avontuur. Lieve Nikki, jouw entree kwam pas redelijk aan het einde, maar heel erg bedankt voor de humoristische steun tijdens het wegbikken van de berg werk gedurende de laatste maanden. Abel, de uren in een donkere kroeg hebben plaats gemaakt voor de vele belletjes over en weer gedurende onze promotietijd. Zowel het lachen als het zanniken over alles blijkt hetzelfde te werken, bedankt daarvoor. Trompie, veel dingen, waaronder onze vriendschap, veranderen niet. Merci voor die laatste spellingscontrole.

Last but certainly not least, I would like to thank all of my family back home.

Zoonder jullie onvurwaardelijke liefde en steun waas ik noeit tot heer gekome. Hoe alder ik wear, hoe mier ik zeej daar ut werme nest woar ik vandan koom, extreem belangriek is. Deze basis maakt daat ik it beste oet meejzelf kaan hale en doarvur bin ik jullie intens dankbaar! Pap, bedankt vur ut brenge vaan rust en oow ovvertuugende en rotsvaste vertrouwen. Mam, bedankt vur oow zurgzaamheid en de fijnen thoes woar ik aaltied op trug kaan valle. Pieter en Willem, bedankt vur de broederliefde en de steun en hullup aas dur wer is waat veul water door de Maas hin moos.

PUBLICATIONS AND MANUSCRIPTS

I. Vollmer, N.I., Driessen, J.L.S.P., Yamakawa, C.K., Gernaey, K.V., Mussatto, S.I. and Sin, G., 2022. Model development for the optimization of operational conditions of the pretreatment of wheat straw. Chemical Engineering Journal, 430, p.133106. DOI: 10.1016/j.cej.2021.133106

II. Driessen, J.L.S.P. §, van der Maas, L. §, and Mussatto, S.I., 2021. Effects of Inhibitory Compounds Present in Lignocellulosic Biomass Hydrolysates on the Growth of Bacillus subtilis. Energies, 14(24), p.8419. DOI: 10.3390/en14248419

III. Driessen, J.L.S.P., Johnsen J., Mohamed E.T., Mussatto S.I., Feist A., Jensen S.I.*, Nielsen A.T.* 2022. Adaptive laboratory evolution of Bacillus subtilis to overcome toxicity of lignocellulosic DDGS biomass hydrolysate. *Prepared for submission*

IV. Driessen, J.L.S.P., Jensen S.I., Woodley J., Nielsen A.T. 2022. Techno-economic assessment of on-site enzyme production using DDGS-hydrolysate. *Prepared for submission*

§ These authors contributed equally.

Publications not included in this thesis

I. Dragone, G., Kerssemakers, A.A., Driessen, J.L.S.P., Yamakawa, C.K., Brumano, L.P. and Mussatto, S.I., 2020. Innovation and strategic orientations for the development of advanced biorefineries. Bioresource technology, 302, p.122847. DOI: 10.1016/j.biortech.2020.122847

PHD THESIS OUTLINE

The work performed in this thesis aims to overcome two hurdles of using lignocellulosic biomass: the toxicity of biomass hydrolysates and the high costs of enzymes in a biorefinery context. To do so, biomass conversion, cell factory optimisation, and bioprocess design integrate multiple aspects of the biorefinery framework.

This thesis is structured in 5 chapters. The introduction (Chapter 1) provides the current status of the biorefinery field, the main conclusions of this thesis and future perspectives. Subsequently, the main body of the PhD research is divided into four chapters (Chapters 2-5).

Chapter 1 introduces the role biorefineries play in a sustainable bio-based future. The current status and main challenges of second generation biorefineries are discussed. The toxicity of biomass hydrolysates and the high costs of enzymes are selected as two main hurdles to using lignocellulosic fibers. Both topics are being considered in a case-study framework. Distiller's Dried Grains with Solubles (DDGS) are used as a lignocellulosic feedstock for on-site enzyme production. In this light, the latest advances in bioprocess engineering and cell factory optimisation are discussed. Lastly, the main conclusions of the thesis and future perspectives of the field are provided.

Chapter 2 provides an example of how computational models can aid in optimising pretreatment. Lab-scale experiments of hydrothermal and dilute acid pretreatment were carried out. Subsequently, data-driven and knowledge-driven models were developed and compared.

Chapter 3 investigates the inhibitory effects of lignocellulose-derived compounds on *Bacillus subtilis*. As a standard chassis for enzyme production, a thorough characterisation of *Bacillus subtilis* is paramount. The considered compounds included sugar degradation products (furfural and 5-hydroxymethylfurfural), acetic acid, and seven phenolic compounds derived from lignin (benzoic acid, vanillin, vanillic acid, ferulic acid, p-coumaric acid, 4-hydroxybenzoic acid, and syringaldehyde).

Chapter 4 shows that adaptive laboratory evolution (ALE) is a successful method to obtain a tolerant *Bacillus subtilis* strain, able to use DDGS-derived hydrolysate as a fermentation medium. Next to growth, evolved strains also showed to be able to produce considerable amounts of enzymes. Whole-genome resequencing data revealed key mutations in global regulators and specific genes related to tolerance and cultivation conditions.

Chapter 5 assesses the economic effects of implementing an on-site enzyme production facility, using DDGS-biomass as a feedstock.

CONTENT

Chapter 1

INTRODUCTION, CONCLUDING REMARKS, AND FUTURE PERSPECTIVES

"I wish the world was twice as big –
and half of it was still unexplored."

Sir David Attenborough – Natural historian

Although many, including myself, would like to happily dream along with the wish of natural historian Sir David Attenborough, the truth is shocking; We have not only explored our normal-sized earth, but are also exploiting more of its resources than it can sustainably provide.

1.1 Towards a Carbon-Neutral Society

The Intergovernmental Panel on Climate Change (IPCC) of the United Nations (UN) provides a strong base of evidence that global warming is driven by increased anthropogenic greenhouse gas (GHG) emissions, predominantly caused by our insatiable consumption of fossil fuels [1,2]. Global warming is predicted to threaten human existence across the globe in numerous ways, varying from increases in hot extremes, marine heatwaves, and heavy precipitation; to reduced crop yields, loss of freshwater resources, and a rising sea level. As such, the IPCC stressed the need to reach net-zero CO_2 emissions globally by mid-century or sooner [3].

As around three-quarters of the global GHG emissions originate from combusting fossil fuels to provide energy, the International Energy Agency (IEA) identified the following key players to decarbonise the energy sector: electrification, renewable energy (solar, wind, hydropower, geothermal and bioenergy), hydrogen and hydrogen-based fuels and carbon capture, utilisation and storage, increased energy efficiency and behavioural changes [4]. Besides the use of fossil fuels as energy carriers, an increasing part of the worldwide oil (27 out of 173 EJ of oil) is used for petrochemicals [5,6]. Although the production does not directly contribute to higher GHG emissions, most petrochemically-derived products end up in landfills or are incinerated, leading to either pollution or indirect CO_2 emissions [7,8]. The use of fossil resources does not only lead to detrimental environmental effects, it is also finite, which makes transitioning to fossil fuel independence even more pressing. One of the ways to simultaneously reduce the dependence on oil and reduce GHG emissions is to leave all residual fossil carbon below the earth's surface and focus on the form of carbon already above ground: renewable biomass (**Figure 1**).

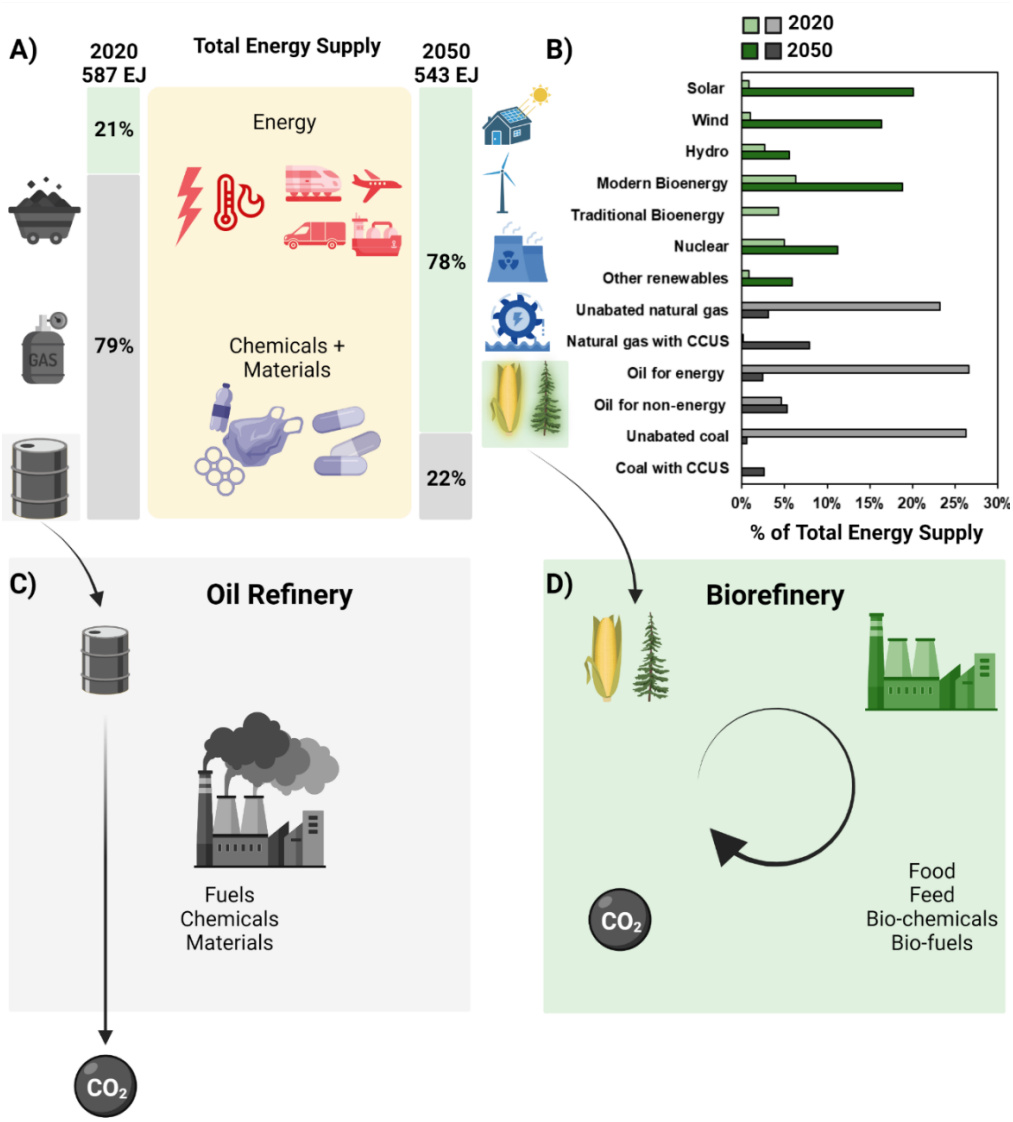


Figure 1 A) An overview of the total energy supply for 2020 and the envisioned energy supply according to the "A Roadmap for the Global Energy Sector" of the IEA to become carbon neutral in 2050 [4]. B) The contribution of each renewable and fossil-derived resource is given. C) Currently, linear production chains in oil refineries lead to increased GHG emissions. D) In contrast, biorefineries make use of renewable biomass to produce a variety of products (food, feed, biochemical and biofuels) in a circular, or even carbon negative, production process.

1.2 The Biorefinery Concept

In contrast to fossil-driven oil refineries, biorefineries use renewable biomass as a feedstock. The IEA defines biorefining as "the sustainable synergetic processing of biomass into a spectrum of marketable food & feed ingredients, products (chemicals, materials), and energy (fuels, power, heat)[9]. Biomass conversion into products can be done by (thermo)chemical, (thermo)mechanical, and biochemical methods [10]. When using enzymes or microbes for the biological conversion of biomass, we enter the realm of industrial biotechnology [11]. In this thesis, the focus will be on the microbial conversion of biomass via fermentation. Microbial conversion is a convincing route to take as it involves relatively low energy input compared to conventional chemical reactions, the capability of producing a variety of organic molecules, high conversion selectivity and adaptability and only require modest infrastructure[12].

1.2.1 Sustainable Impact of Industrial Biotechnology

The UN developed the 17 Sustainable Development Goals (SDGs) as "a blueprint for governments, organisations, companies and citizens to achieve a better and more sustainable future for all" and address global challenges ranging from food insecurity to poverty to innovation and climate change [13]. Today, our increased ability to harness the capabilities of cells and enzymes for industrial biotechnology has positioned us to greatly contribute to the environmental, societal, and economic factors leading to sustainable development.

Environmental. Compared to its chemical, fossil–driven counterpart, bio-manufacturing uses less energy, lowers GHG emissions, and generates less waste and toxic by-products [14]. Intrinsically, truly sustainable production should lead to the decoupling of economic growth from environmental degradation [15].

Social. While making the industry more environmentally friendly, biorefineries will also lead to gains in social dimensions: creating jobs, developing new technology platforms as a basis for innovation, and conserving residual fossil resources for future generations for processes which cannot be replaced by bio-based alternatives [16].

Economic. Biorefineries can provide dynamic growth opportunities for different industries. While agricultural and forestry sectors are essential to provide the required feedstock, chemical industries have the resources to invest in research and development. They can bring new bio-based products to the market using existing marketing and sales capabilities [17,18]. In the last decade, global sales of industrial biorenewable chemicals (excluding biofuels) grew significantly from 82.94 billion US dollars in 2010 to 474.98 billion US dollars in 2020, representing 5.3% to 19.5% of the total chemical sales [19]. Significant examples from the past are the bio-based production of succinic acid, 1,4-butanediol by collaborative efforts of BASF and Genomatica, and 1,3-propanediol by Genencor[20–24]. In Europe, the USA, China, and other regions, governments acknowledge the sector's importance for the transition towards a circular bio-economy showing increasing support by tax incentives, mandatory use regulations, and R&D funding [25,26]. As a new technology, industrial biotechnology had the most significant influence on high-priced, low-volume pharmaceuticals. However, the greatest impact on future sustainability will result from targeting lower-priced chemicals, as these are produced on a much larger scale and therefore have a more significant impact on greenhouse gas emissions [27]. Moreover, biotechnology allows product diversification and tapping into new markets, as some chemicals or substances with innovative properties can only be produced biologically. Industrial enzymes, biofuels, fine chemicals, and platform intermediates are the main products on the market today, while biopolymers, bio-based bulk chemicals, and chemicals with innovative functionalities are predicted to make their debut in the coming years [28,29].

1.2.2 Different Generations of Biorefineries

Bio-based products require a renewable carbon and potentially a renewable energy source if the carbon does not provide enough energy. Biorefineries can be classified using different criteria. For example, the IEA developed a classification based on the platforms, products, feedstock and processes. However, in this thesis, biorefineries are classified according to different generations, i.e. the renewable feedstock used (Figure 2) [12]. Each generation entails its advantages, challenges, and commercialization status. First-generation biorefineries use an easily accessible and edible fraction of dedicated food crops (like sugarcane, soy or corn), while second-generation biorefineries use lignocellulose. Lignocellulose is often present as agricultural by-products, forest- and agro-industrial waste or dedicated non-edible crops. Next to hydrolysate fermentation, the gasification of lignocellulosic biomass to perform syngas fermentation is an alternative. Third generation and fourth generation biorefineries use algae and electricity-driven fermentation, respectively [12].

When considering the commercialization status, first-generation biorefineries are the mainstream practice for nearly all current bio-manufacturing practices. More than 1000 first-generation biorefineries have been constructed, primarily producing ethanol from corn in the USA or sugarcane in Brazil [12]. While economically viable production levels are reached due to the introduction of blending mandates, this practice has caused a lot of controversies as it uses crops, which can also be processed into food. Therefore, there is increasing pressure to switch to alternatives that are economically feasible and do not compromise other sustainability goals like zero hunger and global food security [13,30,31]. An example hereof is the new Directive on Renewable Energy (RED II) in the EU, which dictates conventional food-based biofuels will be limited to a maximum of 7% of the final consumption of energy in the road and rail transport sector in 2030 [32].

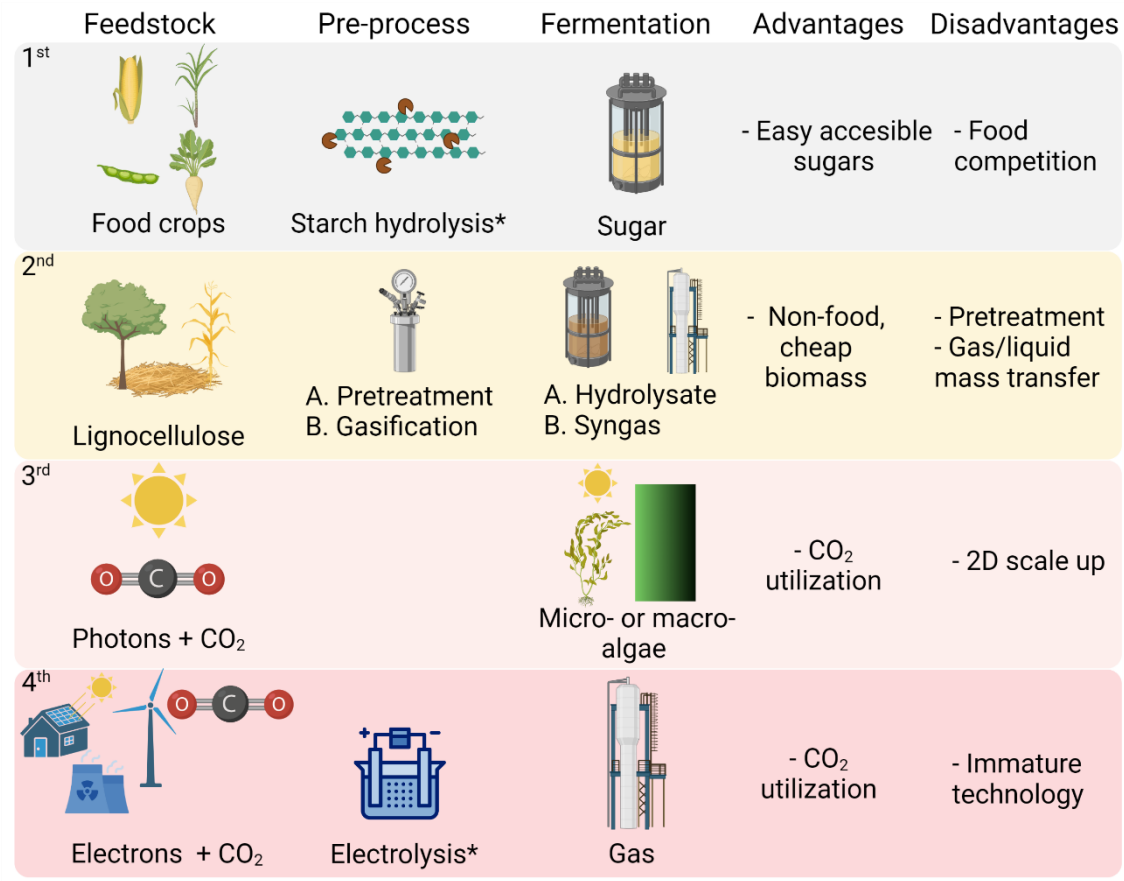


Figure 2 Different generations of biorefineries. *Not in each setup.

Considerable technical challenges still impede the commercialization of higher generation bioprocesses. Issues related to biomass pretreatment have slowed down the development of second-generation biorefineries. Nevertheless, the technological readiness level is up to the point that commercial-scale plants are operational [33,34]. Although gasification allows the use of all the biomass fractions, it involves extra costs to the process, [35] and the gas/liquid mass transfer can pose a challenge [36]. So far, the only realized commercial-scale syngas fermentation plant (Lanzatech) uses fossil-derived syngas [37]. Third generation algae fermentation often encounters problems related to scale up, as photosynthesis-driven cultivation requires to scale up in surface area and not volume. As such, third-generation is mainly shown at demonstration and pilots plants [38]. While having a lot of potentials, electricity-driven fourth-generation fermentation (directly or via H_2) is only in its conceptual stages and has the lowest technology readiness level [12,38,39].

It is likely that ultimately, the commercial manufacturing of bio-based products will be realized using different sources of renewable feedstock and renewable energy, translating into biorefineries of different generations. Nevertheless, as there is a short term need to replace fossil-derived practices to comply with sustainability goals with the biomass available on the planet, the use of lignocellulose biomass in second-generation biorefineries will have a crucial role to play [6,9].

1.2.3 Second-Generation Biorefineries

Lignocellulose is the most abundant biomass on the planet and includes residues and side streams from agriculture (sugarcane bagasse, corn stover), forestry (hard- and softwood), biorefineries (distillers grains), pulp mills, and dedicated energy crops. The main advantages of using these materials are that they are cheap and do not compete with the food supply or existing arable land, being more in line with sustainable development (global food security) [30,40–43].

Lignocellulose consists of polymeric structures: cellulose (40–50%), hemicellulose (25–30%) and lignin (15–20%) (Figure 3) [44]. Depending on the plant species, age, stage of growth, seasonal variation and crop handling, the fractions of these main constituents can vary significantly [45,46]. All three components of lignocellulose have a polymeric structure, consisting of characteristic monomers: Cellulose is homopolymer of glucose, forming long chains (over 10,000 molecules) by strong β-(l-4)-glycosidic bonds. Hemicellulose is a branched, shorter hetero-polymer (50–300 monosaccharide units), bound together by weaker α-1,4-glycosidic bonds and consisting mainly of pentoses (xylose, arabinose) and to a lesser extend hexoses (mannose, glucose, galactose). Acetyl-groups, galacturonic acid, or glucuronic acid can branch off the polymer chain. Lignin is a complex hydrophobic, cross-linked aromatic polymer, which protects plants against microbial degradation and is seen as the glue between cellulose and hemicellulose. In a simplified model of the lignin structure, three different building blocks can be differentiated, being p-coumaryl alcohol, sinapyl alcohol and coniferyl alcohol[47]. The ratio of the three building blocks in the heteropolymer varies between

different plant taxa, cell types, and individual cell wall layers [45,48]. The exact composition of each lignin molecule is more complex and still partly undiscovered.

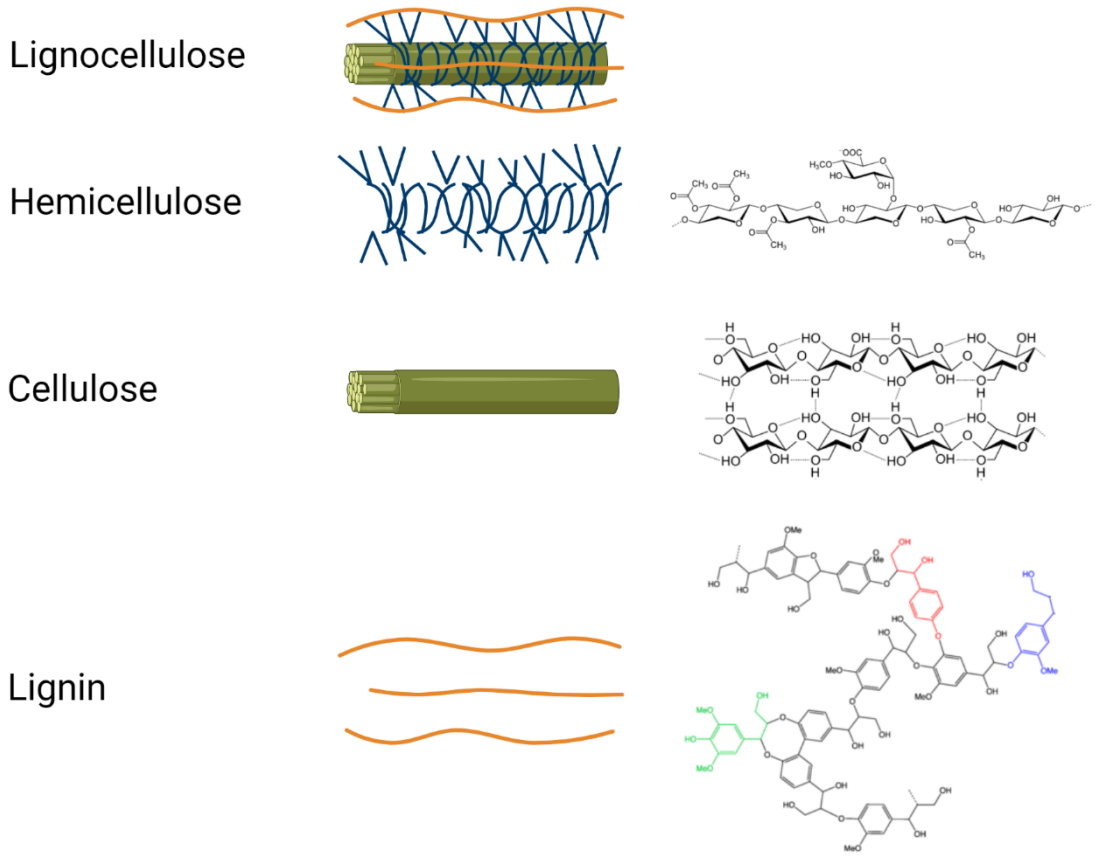


Figure 3 An example of typical lignocellulose composition and molecular structure of hemicellulose, cellulose and lignin. The depicted hemicellulose polymer is an example of an acetylated glucoronoxylan. Hydrogen bonds in the cellulose polymer are shown as dotted lines. The three different lignin monomers are distinguished by colour: p-coumaryl alcohol, sinapyl alcohol and coniferyl alcohol.

.

1.2.4 Biomass Pretreatment Methods

Due to its recalcitrant nature, the high amount of polymeric carbohydrates in lignocellulose are not freely accessible for enzymatic hydrolysis to monomeric sugars. As such, pretreatment is required to break the molecular bonds between the main constituents of lignocellulose by (a combination of) physical, chemical, and biological methods (Figure 4) [30].

During pretreatment, undesirable side reactions can lead to the formation of lignocellulose-derived compounds (often called inhibitors), which have inhibitory effects on microbial fermentation [49,50]. As microbial performance (titer, rates, yield) is key to a cost-competitive bioprocess, costly detoxification of lignocellulosic hydrolysate is often required [51,52]. The origin and composition of the feedstock and the harshness of the pretreatment dictate the concentration and distribution of the inhibitors. These can include furan aldehydes, such as furfural derived from xylose and 5-hydroxymethyl-furfural (5-HMF) derived from glucose, organic acids (mainly acetic, levulinic and formic acid), as well as lignin-derived (phenolic and aromatic) compounds [53–55].

During the enzymatic hydrolysis, a plethora of enzymes with varying specificities are needed to degrade all components of biomass efficiently [56]. There are three main cellulases which synergistically hydrolyse cellulose. These are endo-1,4-β-glucanases, cleaving the cellulose chain in the middle and reducing the degree of polymerisation, exo-1,4-β-glucanases, progressively cleaving off cellobiose from the cellulose chain ends, and β-glucosidases, hydrolysing the cellobiose unites to glucose monomers. As hemicellulose is more heterogeneous, it requires more different enzymes to hydrolyse the backbone and cleave off substituents [56].

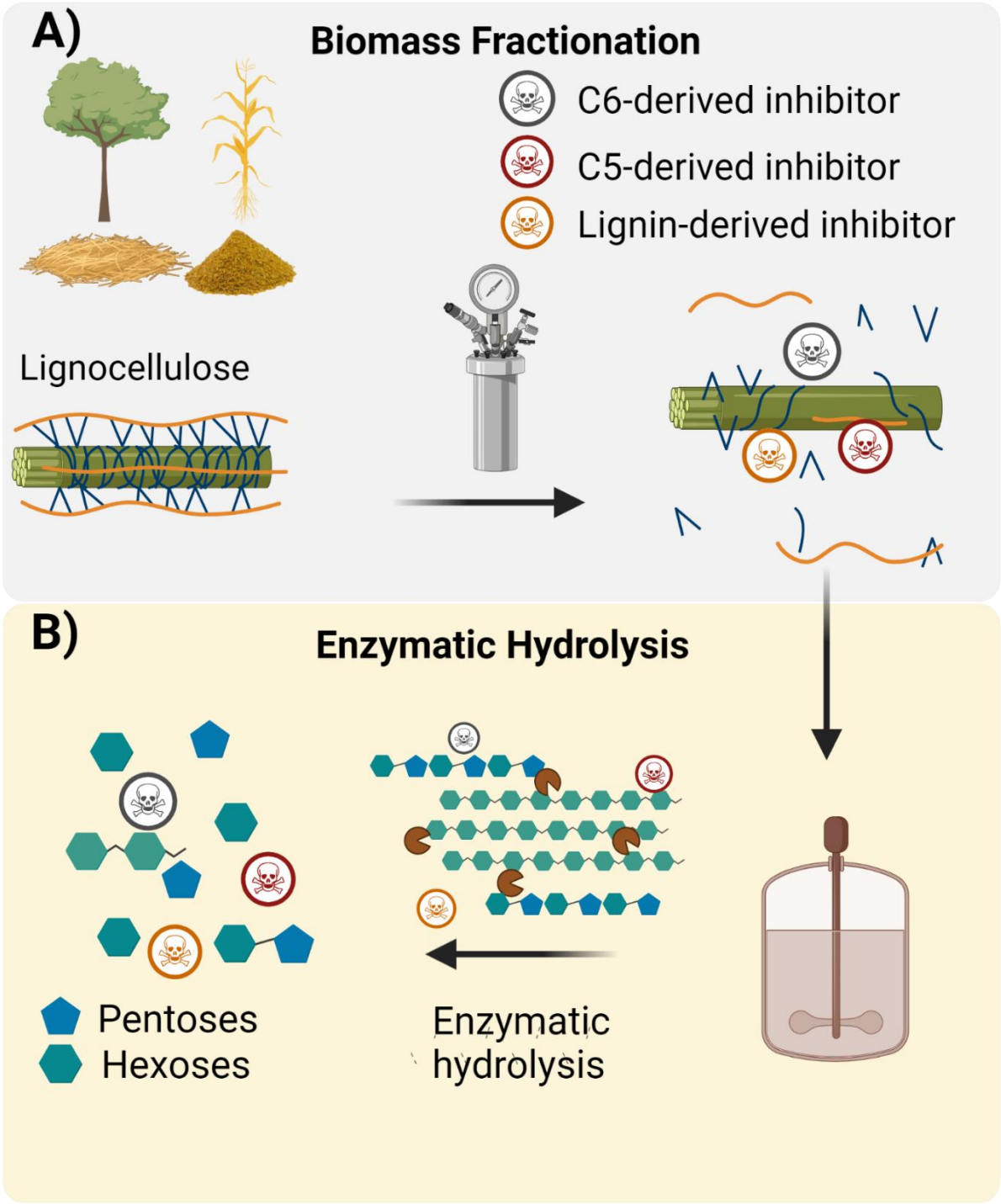


Figure 4 Overview of pretreatment (A) to make the biomass polymers more accessible for subsequent enzymatic hydrolysis (B).

Traditionally, pretreatment methods have been categorized according to their driving force being chemical, physical, physiochemical and biological. However, as the scope and focus of pretreatment changed to taking care of other constituents than cellulose, this classification has become somewhat outdated [48]. Hence, we distinct conventional and advanced pretreatment methods (Table 1).

Table 1 Common pretreatment methods, their mode of action, active agents, and targeted polymer fraction (CF: cellulose fraction, HF: hemicellulose fraction, LF: lignin fraction [48].

Pretreatment method	Mode of action	Active agent	Target polymer
Acid	Catalysis with H^+	H_2SO_4, H_3PO_4 and other strong acids	CF, HF
Dilute acid	Catalysis with H^+	H_2SO_4, H_3PO_4 and other strong acids	HF
Alkali	Catalysis with OH^-	NaOH, lime, Na_2CO_3 and strong bases alkaline compounds	LF
Hydrothermal	Catalysis with H^+	Auto-ionisation of H_2O under high pressure/temperature	HF
Steam explosion	Catalysis with H^+	Fibre separation by water expansion, catalyst may be added	HF
Ionic Liquids[57]	Fractionation of polymers	Large organic cation (e.g. imidazolium, phosphonium, pyrrolidinium, cholinium) and small inorganic anion	LF
Organosolv	Fractionation of polymers	Organic solvents e.g. ethanol, butanol, acetone. Catalyst may be added	LF
Biological	Degradation of material	Degradation by enzymes from xylophagous fungi	HF, CF, LF

Conventional pretreatment methods are thoroughly described in the literature and often involve acid, base, hot water or pressured steam [58,59]. The main advantages are that these methods are established, robust, and close to the concept of unit operation, which makes them easy to implement [60]. Conventional pretreatment methods are the most mature and are the backbone of current demo/industrial-scale facilities [61,62]. These methods often focus on high sugar conversion yields, which require harsh process conditions. Disadvantages of conventional methods include the need to use high temperatures and/or chemicals agents, the formation of inhibitory by-products due to harsh conditions, and issues related to recycling and waste handling of used chemicals [61].

In contrast, more advanced pretreatments were developed with the aim to recover more different biomass constituents, rather than only the cellulose fraction and overcome the limitations of current conventional methods. These methods often involve novel solvents (supercritical fluids and ionic liquids), deep eutectic solvents, inorganic salts or new technologies (such as popping, gamma-ray, electron beam radiation or combinations of pretreatment methods) [63]. Advanced methods comprise one or more of the following characteristics: being milder operation conditions, no formation of inhibitory compounds, low use of chemicals, high specificity towards lignin or efficient fractionation [64]. Despite the potential, further development of many advanced methods is hampered by health or sustainability concerns, complex equipment, poor performance or scalability issues. Moreover, the maturity of the more advanced pretreatment methods is low, as sufficient pilot/commercial scale data is lacking for well-funded quantitative techno-economic assessments [63].

Many efforts has been invested in studying the characteristics of both conventional and advanced pretreatment methods, including categorization, the effect of process conditions on yield and molecular reaction mechanisms leading to sugars and inhibitors [48,65–68]. Still, there are some crucial knowledge gaps. The enzymatic hydrolysis of cellulose is influenced by the general biomass composition and structure, the crystallinity of cellulose, degree of cellulose polymerization, accessible surface area, the protective layer of lignin and hemicellulose and the

degree of hemicellulose acetylation [66]. Different pretreatment methods have distinct effects on these features. Moreover, these features are altered simultaneously during the biomass pretreatment, which makes the establishment of clear-cut correlations between these factors and the biodegradability of the biomass during pretreatment difficult [66]. Also, the literature does not agree on the exact molecular mechanisms at hand, and the modelling of the formation of degradation products during pretreatment seems challenging [67].

Frequently, reviews have been published summarizing existing pretreatment methods and updating the list with novel developments [48,62–66,69–71]. In short, three criteria should be fulfilled to result in an optimal pretreatment: a sharp separation of different fractions, high yields and minimal formation of inhibitory compounds. In addition, one would like to recycle the involved chemicals and minimize the required power input and costs [48]. Although these are general requirements, there is no standardized optimal pretreatment due to the huge variation observed for different lignocellulosic materials and the intrinsic trade-off between the optimization criteria. This requires to selecting on a case-to-case basis, taking the whole bioprocess (feedstock, cell factory, products etc.) and application into account when choosing for a specific pretreatment and employed conditions. [48,70,72]. As shown in a part of this thesis work, modelling can be an efficient tool to determine optimal conditions for hydrothermal and dilute acid pretreatment (**Chapter 2**) [73]. More computational decision tools, such as techno-economic assessments (**Chapter 5**) and life cycle assessments, are key to selecting the optimal pretreatment based on a holistic view of the entire bioprocess. However, the bottleneck for these tools is the lack of pilot-scale experimental data to provide accurate and quantitative conclusions. As such, it is likely these types of experiments will be a relevant topic of research in the coming years [63].

Steam explosion. In Chapter 4 and Chapter 5, a steam explosion pretreatment is applied to Dried Distiller's Grains with Solubles (DDGS). During this pretreatment method, steam is brought under high pressure, and upon the rapid release, it expands in the cell structure, causing it to disrupt. The addition of a dilute acid catalyst improves the hydrolysis of the hemicellulose fraction, the subsequent enzymatic hydrolysis of cellulose, and allows lower temperatures leading to less formation of inhibitors [66]. Steam explosion is used in this thesis as it is one of the most widely investigated methods in both lab-scale and different pilot plants and is considered a cost-effective method near commercialization [48,66,74].

1.3 Challenges of Second-Generation Biorefineries

The high availability of lignocellulose, the need to use non-food crops to replace petroleum-derived products, and the technological readiness level position second-generation biorefineries to significantly impact our current production patterns. In line with the sustainable development goals, locally available biomass can be a starting point for bio-based value chains that provide economical income and livelihood for people in both developed and developing countries [75,76].

Despite the tremendous amount of research performed on lab-scale second-generation biorefinery setups (varying feedstock sources, pretreatment methods, process conditions, production hosts, etc.), the current realisation of commercial-scale and economically viable production plants is lacking. Although the definitions for biorefineries vary between different institutional databases, it is clear that there only are a limited number of commercial-scale 2G bio-refineries, which exclusively produce biofuels [77]. While there are more than 1000 first generation biorefineries around the globe, only around 11 second-generation biorefineries have been built, all producing bioethanol [12,78]. Biofuelwatch, a non-profit organization situated in the UK, published an unsettling report that all 11 cellulosic bioethanol plants, have either failed and not operating anymore or are dealing with serious technical issues leading to lower production than projected [79]. This raises the question about why, despite the frequently published advantages and potentials of second-generation biorefineries, successful commercialisation is lagging. The implementation of new technologies is highly complex, and involved factors can be grouped as technical, economic, and political (**Figure 5**).

1.3.1 Technical Challenges

From a process technical perspective, the success of scale-up is highly dependent on the correct matching of feedstock to process design [80–82]. For 2G biorefineries, this requires a year-round feedstock availability at a competitive price. Therefore, it is crucial that second-generation biorefineries are near a reliable, long-term, cheap supply of biomass. Often lignocellulose derived from first-generation ethanol plants is used, like corn stover or sugarcane bagasse. The production, collection, transport, storage, and pre-processing of the enormous volumes of biomass, containing dirt and sand, has turned out to be exceptionally challenging in commercial-scale setups [33,78,81].

The technical immaturity of subsequent process steps, including pretreatment, enzymatic hydrolysis, fermentation, and downstream processing (DSP), has contributed significantly to the low success rate of commercial-scale 2G biorefineries [33]. Despite the many lab studies performed, optimal conditions found in the lab are hard to reproduce when applied in commercial-scale settings. This appears to be especially the case for the pretreatment step, where a quantitative comparison of economic and environmental performance between methods at industrial-scale is lacking [82].

1.3.2 Economic Challenges

From an economic perspective, the oil price, the product market, the high costs of second-generation biorefineries, and the lack of capital investments pose significant hurdles. The first contributing factor to the weak economic potential of biorefineries is the oil price. Widespread implementation of new oil extraction technologies led to an oversupply and thus lower prices of oil in recent years [83,84]. While the use of a finite fossil source is per definition unsustainable, in terms of economics, the oil price is of great importance for the competitiveness of bio-based fuels or products.

So far, commercial-scale 2G biorefineries have mainly focused on producing low priced ethanol leading to small operative margins. Although the IEA predicts that the market for biofuels will grow substantially in the coming years under the net-zero scenario, other options like electrification of transportation, may experience major innovations that make them (even more) preferable to biofuels [4]. The economic potential of 2G biorefineries could be increased by considering higher-value chemicals. As such, the US department of energy released a list of the 12 most promising chemicals in 2004, which has been updated recently by the IEA [85]. Besides diversifying the product portfolio, using different biomass fractions to produce value-added co-products next to the main product has often been mentioned to be a viable strategy to increase the economic robustness of a biorefinery considerably [28,86]. Nonetheless, the low oil price creates economic barriers for bio-based products to capture a share of the market [87].

Second, the previously mentioned technical challenges result in high capital and operational costs (CAPEX, OPEX). These are mainly related to the pretreatment and enzymatic hydrolysis required when using recalcitrant lignocellulosic biomass [33,82,83]. As such, capital costs of cellulosic biorefineries have shown to be up to five times higher compared to starch ethanol plants [83,88]. Additionally, the intrinsic low sugar concentrations associated with the use of lignocellulose lead to increased OPEX during downstream processing [61]. Applying harsher pretreatment conditions to achieve higher conversion yields lead to higher concentrations of inhibitory compounds and requires costly detoxification steps. Consequently, cellulose-derived sugars are currently estimated to be more costly than those of sugarcane or corn, according to Sanford, Chotani et al. [80]. Different strategies are being pursued to bring down the cost of lignocellulosic sugars, including improved enzyme activity, on-site enzyme production, development of tailor-made enzyme cocktails, improved production organisms, valorisation of lignin, and emerging consolidated bioprocessing [33,89–95].

The current fragile economics of second-generation biorefineries result in a risk profile which is not attractive for investors [33,78,83]. Venture-capital funds, normally willing to take risks for acceptable return of

investment, tend to regard 2G biorefineries as too capital intensive and technically complex[96]. More mainstream investors deem it too uncertain and seek more attractive alternatives [78].

From an industrial perspective, the choice to invest or even stop ongoing 2G biorefinery projects is based on their estimated profitability. Some of the companies, like Clariant, claim to have achieved a profitable process, while many others, like Dow-DuPont, Beta Renewables, BP, and Abengoa, have stopped their activities in the last years [33,34,79,82,97–99]. A potential strategy to maximize chances of success is to establish joint ventures and create synergy between smaller specialized companies with specialized technical solutions and bigger companies to complement the infrastructure and capital to scale up [78].

1.3.3 Political Challenges

Both global, national, and regional actors, such as the IEA, the Organisation for Economic Co-operation and Development (OECD), and national governments, promote the transition towards a bio-economy and the role of biorefinery development to combat climate change [100–102]. The realisation of biorefineries calls for public support as they are needed from a sustainability perspective but involve high capital costs and imply significant uncertainties [78,96].

To support biorefinery development, it is not common for governments to construct biorefineries themselves, but instead, implement policies to create conditions for industrial stakeholders to realize them. These often target the competitiveness of bio-based products compared to oil-derived products and can include direct subsidies, tax reliefs, and taxation on CO_2. Progressing the field of advanced biorefineries requires more than general mandates or taxation on CO_2, as these measures do not provide enough incentives to invest in commercial-scale plants [103]. Instead, clear industrialization policies are necessary to create a market and support technological development during early niche market phases. By doing so, first publicly supported commercial-scale plants allow learning-by-doing, formation of value chains and optimization of large-scale design options to go down the cost curve [83]. Tools to do so

can include public procurement and various types of price guarantees, combined with investment subsidies [103]. Due to the mentioned technical and economic challenges on top of the increasingly uncertain public policies in both Europe and North America, advanced biorefinery development (using lignocellulose) progressed slower than expected and many attempts for large-scale 2G plants have been abandoned [78,79,96,104].

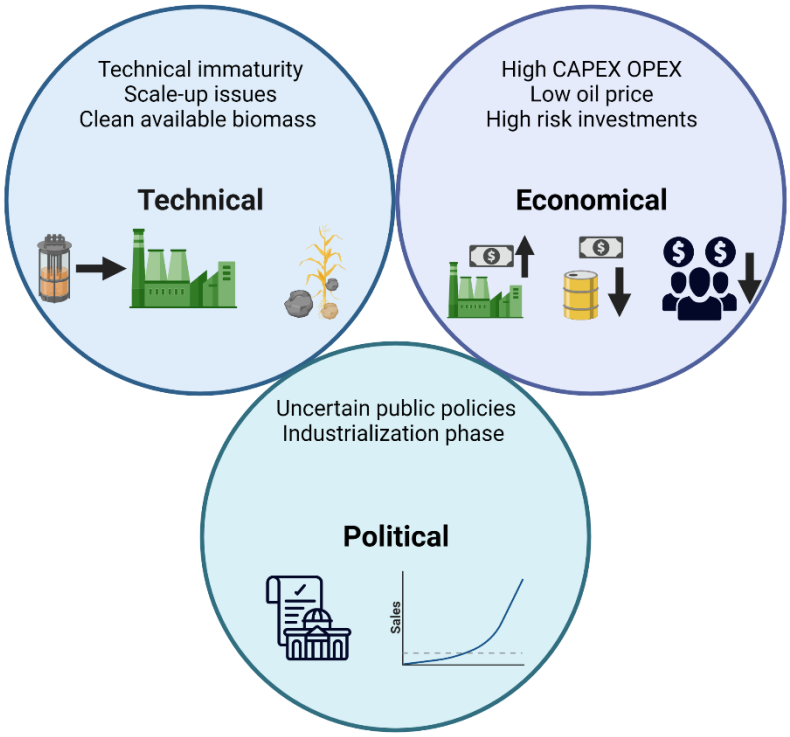


Figure 5 Main challenges encountered during commercialisation of second-generation biorefineries.

1.4 A Case Study: DDGS Valorization in a Dry-grind Bioethanol Plant

To increase the chance of success of a new technology to be commercially implemented, it is strategic to make use of a "bolt-on" configuration at a host facility, which is based on established technology. The rationale behind this approach is that the available feedstock and facilities are at a lower cost than would be the case for a stand-alone facility, lowering the risk of such an undertaking [83]. The main work of this thesis is done in a case study that involves the innovative valorisation of lignocellulosic DDGS via on-site enzyme production at an established dry-grind bioethanol facility.

1.4.1 Dry-grind Bioethanol Process

The majority of the current global bioethanol production (53%) is situated in the U.S., of which around 90% involves a dry mill process (**Figure 6A**). A simplified corn dry-milling process can be divided in five parts, being: milling of biomass, liquefaction, saccharification, fermentation, distillation, and recovery [105]. Here, the model made by the U.S. Department of Agriculture is used, but exact conditions per step can vary depending on the source used [105–108]. The process starts by milling the grains, after which thermostable amylases are added and the slurry is "cooked" at 110 °C for 15 minutes to hydrolyse the starch into oligosaccharides. The gelatinized mash is cooled down and further hydrolysed by glucoamylases at 61 °C for 5 hours to obtain a glucose-rich stream, which is subsequently fermented by yeast at 32 °C to produce ethanol. Ethanol is distilled and dehydrated, while the non-volatile part is separated into a liquid fraction (stillage) and a solid fraction (wet distiller's grain, WDG). Concentrated stillage is ultimately mixed with the WDG and dried to form distillers dried grain with solubles (DDGS) as a high nutritional by-product of the ethanol plant [108107].

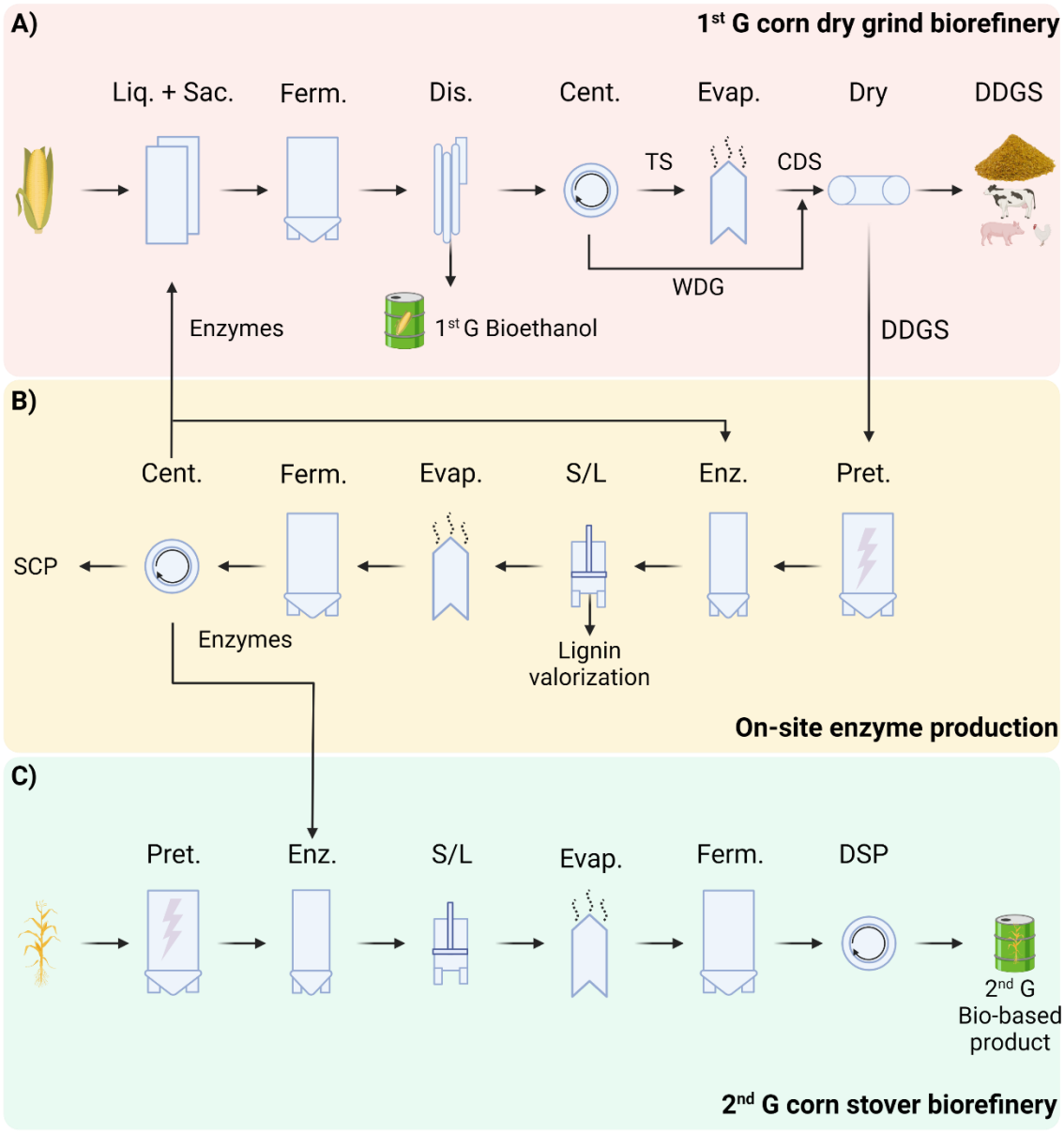


Figure 6 Schematic overview of A) current dry-grind bioethanol production process and production of DDGS as the major by-product. B) An on-site enzyme production facility as a strategy to valorise DDGS. C) An adjacent 2nd generation corn stover biorefinery which uses low-cost enzymes. Abbreviations: Liq. Liquefaction, Sac. Saccharification, Ferm. Fermentation, Cent. Centrifugation, TS. Thin stillage, WDG. Wet distiller's grains, CDS. Condensed Distiller's Solubles, Evap. Evaporation, Dry. Drying, DDGS. Distiller's Dried Grains with Solubles, Pret. Pretreatment, Enz. Enzymatic hydrolysis, S/L. Solid Liquid separation, SCP. Single Cell Protein.

1.4.2 The Potential of DDGS as Fermentation Feedstock

The selling of DDGS comprises 22% of the value output and is therefore of vital importance to the economic viability of the bioethanol industry [42,109]. Despite the recent decrease due to the COVID pandemic, the global bioethanol production (and thereby DDGS supply) is expected to grow to comply with sustainability pledges made around the globe [4,109]. Currently, DDGS is marketed as animal feed due to its high nutritional value as it is mainly composed of seed hull, germ, proteins, and oil. It is estimated that for every 100 kg of corn used, approximately 40.2 L ethanol, 32.3 kg of DDGS, and 32.3 kg of CO_2 are produced [42].

However, as DDGS can only account for 30% of the livestock feed, the animal feed market is expected to be saturated, endangering the economic viability of bioethanol plants [42,107]. Bearing these developments in mind, the need for alternative routes to add value to DDGS grows.

Recently, there has been an increasing interest in the potential of using DDGS as a substrate for microbial fermentation due to its rich nutritional composition. As such, DDGS can be an ideal starting point for the bio-manufacturing of a variety of products, including organic acids, biofuels or, in the context of this thesis, hydrolytic enzymes (**Figure 6B**) [107,110]. One of the main advantages of using DDGS as a feedstock for fermentation is that it allows the removal of the energy-intensive drying step of WDG to DDGS, which is a major cost factor in the bioethanol production process [105].

1.4.3 On-site Enzyme Production

On-site enzyme production is a strategy for generating value for the increasing supply of lignocellulosic DDGS. This bioprocess design would meet both needs of adding value to the increasing available DDGS by conversion into a high-value product and lower some of the risks and hurdles of using lignocellulose as a feedstock.

First of all, biomass availability is more reliable, as a clean DDGS stream would be readily accessible at the dry grind bioethanol plant itself. Overall risks related to technical immaturity of using lignocellulose are mitigated by making use of the "bolt-on" configuration as described above. Moreover, as enzymes are of higher value than biofuels and some bulk bio-based chemicals, higher economic margins are to be expected. In addition, using DDGS for on-site enzyme production should lead to lower enzyme production costs as a lower value lignocellulosic feedstock is used and formulation and transport costs of enzymes are omitted [83,89,111,112]. As enzymes are one of the biggest cost factors, lower enzyme production costs would contribute significantly to the economic viability of second-generation biorefineries [112] (**Figure 6C**).

The pretreatment of DDGS fiber (cellulose and hemicellulose) results in the release of both fermentable compounds (sugars) and unwanted formation of inhibitory compounds [110]. In **Chapter 3**, the inhibitory effects of lignocellulose-derived compounds on *Bacillus subtilis* are assessed. Subsequently, the tolerance of the enzyme-producing cell factory towards DDGS-hydrolysates is improved via adaptive laboratory evolution (**Chapter 4**). Lastly, a techno-economic assessment is performed to evaluate the economic effects of implementing an on-site enzyme production facility (**Chapter 5**).

As the integration of cell factory optimization and process engineering is essential for the successful scale-up of a new bioprocess [87], the advances of both fields and their relation to overcoming 2G-related challenges will be discussed in the following sections.

1.5 Advances in Bioprocess Engineering

From a bioprocess engineering perspective, the switch from first to second-generation biorefineries involves an essential extension of the process further upstream, being units involved with the biochemical conversion from lignocellulose to fermentable sugars (biomass pretreatment and enzymatic hydrolysis). Still, general principles in bioprocess engineering apply to both.

1.5.1 Bioprocess Design

The classical bioprocess. When considering the different stages of "classical" bioprocess development, it is obvious that it is a highly interdisciplinary endeavour [113]. Often, the process is initiated by the hands of highly skilled molecular biologists, which aim at optimizing the production of native metabolites or expression of heterologous biosynthetic pathways or proteins and improving the stability of the microbial cell factory. First, ideal culture conditions at lab-scale are defined. Subsequently, the scale-up process can assess whether conditions from the perspective of the microbe at a bigger scale are still meeting the requirements for efficient production. The initial stage is often still rather small (1-2L-bench top bioreactors). This stage is meant to provide calculated and measured parameters like mass-transfer coefficients, mixing time, gas holdup, specific oxygen consumption, power number, impeller shear rate, etc. Various modes of operation (batch, fed-batch or continuous fermentation), can be studied at this point to clarify which setup is most suited for the specific bioprocess. In the next phase, pilot-scale proportions (100-10,000 L) are used to reduce the risks of industrial implementations by providing a more representative version of the manufacturing process. Although the specifications of the bench-scale prototype (geometry, impeller design etc.) are often kept constant, increasing the size of the bioreactor have significant effects on the microbial activity. Full manufacturing scale (e.g. 20,000–2,000,000 L fermentations) requires even more

engineering input, involving auxiliary services facilities, continuous waste, and recycling streams and increased safety measures [114]. Besides fermentation; downstream processing or product recovery, is a crucial part in bioprocess development, as it represents a 20%-40% of the total processing cost for bulk chemicals [115]. Lastly, formulation, packaging, and marketing are the final steps of the bioprocess.

"Begin with the end in mind". Although the described bioprocess development starts with strain development and ends at marketing the product, it is a rather classical way of looking at bioprocess design. Instead, there is a clear shift to "begin with the end in mind". Instead of starting with the perspective of feedstock, lab conditions, and the current market, one shifts to having the product, large scale, and future markets as a perspective (**Figure 7**) [12,114]. Product specifications and required downstream processing schemes should be established at the start, as these have big influence on the upstream processes. Since bioprocess development takes time, one should establish price targets for the envisioned product and feedstock for future markets. Furthermore, adapting the microorganism to the constraints of industrial processes at the start of the process instead of vice versa can save a considerable amount of costs.

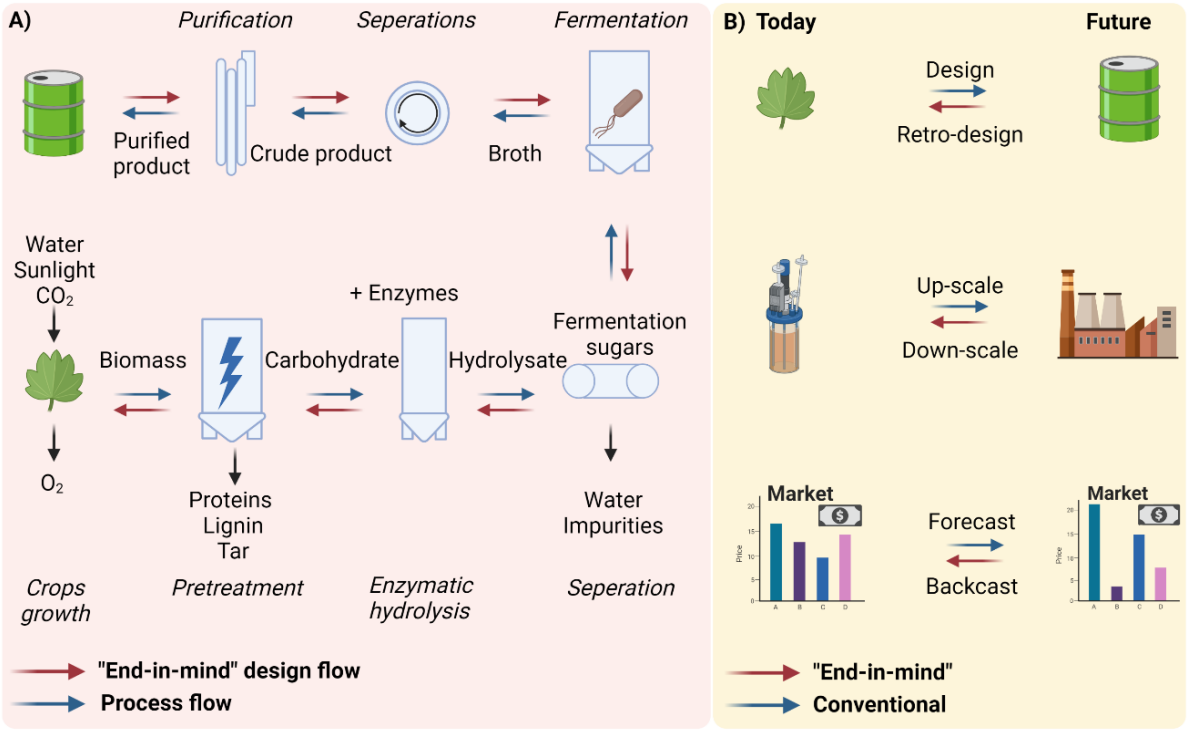


Figure 7 Bioprocess engineering with "Having the end in mind" A) A schematic overview of a bioprocess from the perspective of the final purified product, in which sugar hydrolysates are an intermediate platform. The design flow is the opposite of the process flow. B) Conventional versus "End-In-Mind" process design. Instead of the perspective of feedstock, lab conditions and current market situations, the focus is on the product, large-scale and future markets.

An example of starting with the end in mind is the global analysis of the future bio-based economy. As mentioned before, fossil fuels in the energy sector are expected to be mainly replaced by solar, wind, geothermal, and hydropower. However, in many scenarios, biomass will still play a major role [4,6,12]. In this regard, a key consideration is whether there will be enough renewable biomass available to drive the shift towards a bio-economy or whether this is impeded by global biomass supply limitations. A global analysis by Noorman and Heijnen shows that biomass actually could provide the necessary means to fulfil future carbon and energy demands, with lignocellulose being the main feedstock [6].

Scale up. In contrast to lab conditions, transport steps in large scale fermentations become rate limiting and are determined by phenomena like mass transfer, heat transfer, macromixing of limiting substrate in the bioreactor or the separation of the gas bubbles from the liquid [6]. The difference in fluid flow and reaction regime likely results in discrepancies in strain performance between lab and industrial-scale bioreactors. Downscaling the large-scale conditions in the lab allows for screening and selection of production hosts in more representative conditions. Here, cells are exposed to rapid fluctuations in concentrations of rate-limiting substrates, dissolved oxygen levels, pH, temperature or other relevant factors, to mimic conditions, which are perceived by the cells in large-scale reactors. To develop downscale simulators, a detailed analysis of the large-scale conditions is required first. To this end, computational tools, like computational fluid dynamics (CFD), play an increasing role in providing modelled data flow rates and concentration regimes of large-scale bioreactors. Using the output of these models, scale-down simulators and operating conditions can be designed with much more precision [116,117].

High throughput screening. Miniaturization of bioreactors could allow shortening bioprocess development by high-throughput screening of strain variants and process conditions. The advantages of micro-bioreactors are related to easier handling, smaller size, lower cost, and the potential for automation, while online cultivation data and controllability are maintained [118–120]. Nonetheless, fermentation is typically a bottleneck for the strain development process, which will be discussed in more detail in the next section.

1.5.2 The Ideal Full-scale Process

Four parameters can be used to evaluate the performance of the designed industrial bioprocess: titer, rates, yield (TRY), and downstream processing efficiency. An optimal bioprocess has the following characteristics [6,27]:

1) The product yield should be as close as possible to the thermodynamic maximum. This demands high conversion of the polymeric fractions to monosaccharides in the biomass pretreatment and enzymatic hydrolysis. For the subsequent fermentation, this entails that the available Gibbs free energy associated with the substrate is preserved in the product rather than in biomass, by-products or cell maintenance. As the cost of carbon often represents more than 50% of the total production costs, a high yield maximizes the gap between the cost of the substrate going into the process and the value of the product [121].

2) High volumetric productivity minimizes the capital investment needed for the process plant. Productivity is limited by transport rates, determined by industrial hardware, transfer area, and process characteristics.

3) A high product concentration prevents both high operational costs in downstream processing due to excessive water removal, and greater volumes of separation agents as high capital costs due to big fermentation volumes. For second-generation biorefineries, high product titers entail the use of high solid loadings during the enzymatic hydrolysis to produce high concentration sugar syrups. However, low amounts of free water during enzymatic hydrolysis lead to technical challenges, such as poor mixing, mass and heat transfer limitations, and reduced enzyme efficiency. Moreover, with increased solid loading, a drop in sugar yield is often observed [122]. Alternatively, concentration steps in the process are commonly applied to obtain higher sugar concentrations, but the concomitant increase of inhibitory compounds present in the biomass hydrolysate complicates this step.

4) The process should be scalable and have minimal maintenance requirements. This accounts for all stages of the process.

5) Microorganisms should be tailored to the process, involving non-filamentous phenotypes to prevent rheology issues, being thermos-tolerant to withstand high temperature for efficient cooling purposes and high kinetic rates, tolerance against products and oscillating environments. For the use of lignocellulosic hydrolysates, cell factories should be designed to tolerate the inhibitory compounds being formed during pretreatment and optimized to not only consume glucose but also co-utilize all other sugars present [123].

6) Integration of upstream, fermentation, and downstream processing can result in a more efficient overall process and lead to reduced costs. Examples of second-generation biorefineries are the combustion of lignin to provide heat and energy for downstream processing purposes or integrated pretreatment methods, which couple the pretreatment with at least one bioconversion step (enzymatic hydrolysis and fermentation) in the same reactor [63]. Consolidated bioprocessing is an example of integrated pretreatment, which allows enzyme production, enzymatic hydrolysis, and fermentation in one vessel [48,124].

1.5.3 Techno-economic Assessment of Bioprocesses

Techno-economic analysis (TEA) is an important tool for decision-making used in bioprocess design. By performing a TEA, the economic feasibility, bottlenecks and targets for process improvement of different design options can be identified during the initial stages of bioprocess development. As a result, both fruitless research efforts and wasted investments can be avoided [125,126].

The process of TEA can be described as follows: (1) selection of system and system boundary, (2) designing a process flow diagram, (3) analysis of mass and energy balances, (4) estimation of capital costs, and (5) estimation of operating costs [126].

It is common that economic trade-offs present themselves during the TEA for different design options. Optimization of the economic performance commonly yields several design solutions with equal economic performance. Subsequent, the selection of these designs is based on the most beneficial technical characteristics, such as safety, reliability, and ease of operating [125]. The limited availability of inputs, processing data, process equipment, and accessories associated with a designed bioprocess can form a major hindrance to performing a TEA. Data of closely relevant commercial systems is often used instead [126].

1.6 Advances in Microbial Cell Factory Development

Microbial cell factories (most often bacteria, yeast or fungi) are the working bees of industrial fermentation, being responsible for the biological conversion of renewable feedstock into higher-valued products. Although microbes are able to synthesize many chemical compounds, evolution has not selected them to overproduce human-desired products or be tailored to the envisioned bioprocess conditions. Therefore, cell factory engineering aims at optimizing the production of native metabolites, expression of heterologous biosynthetic pathways, or protein expression [127]. Next to production, the cell factory should be genetically stable and tolerant of the employed fermentation conditions. By doing so, scientists strive to improve the previously described performance indicators of industrial fermentation (yield, titer and productivity) to a level that results in a cost-competitive bioprocess.

Protein. Protein products include recombinant pharmaceutical proteins as well as industrial enzymes such as amylases, proteases, and lipases used for, e.g. laundry detergents, carbohydrate processing, and food processing. Host selection is based on the product characteristics: Pharmaceutical proteins often require post-translational modifications in eukaryotic cells, while other industrial enzymes can be produced in fast-growing prokaryotes.

Biochemical. While it is troublesome to chemically synthesize proteins, bio-based chemicals produced by fermentation have to compete with cheap, established, fossil-fuel driven chemical processes. Considerable effort has been invested to rank bio-based candidates based on their market prices and theoretical maximum yields to provide a better perspective on the economic potential of bio-based products compared to the petrochemical variant [128].

1.6.1 Model versus Non-model Organisms

Predominantly, model organisms, such as *Escherichia coli*, *Bacillus subtilis, Corynebacterium glutamicum* or *Saccharomyces cerevisiae,* are used as cell factories because of their well-characterized metabolism and physiology, the availability of abundant genetic tools and metabolic models, make them easy to work with. Moreover, these species have proved themselves in industrial-scale fermentations. However, the limited number of well-characterized species is not always the best choice to apply in an industrial setting, due to inflexible use of carbon sources or intolerance to the substrate, the product, the inhibitory compounds, the process conditions or oxidative stress. On the other hand, non-model microbes, like *P. thermoglucosidasius* [129], *M. thermoacetica* [130]or *K. Marxianus* [131], can have unique characteristics that are useful for industrial fermentation conditions or produce unique compounds, but lack established genetic toolboxes and metabolic models to take advantage of the riches biodiversity provides. Consequently, considerable strain engineering is focused on either improving tolerance, production or consumption of the model organisms in industrial settings or domesticating non-model organisms which already have the desired phenotype (Figure 8). In this thesis, I aimed at using the first approach using the model organism *B. subtilis*.

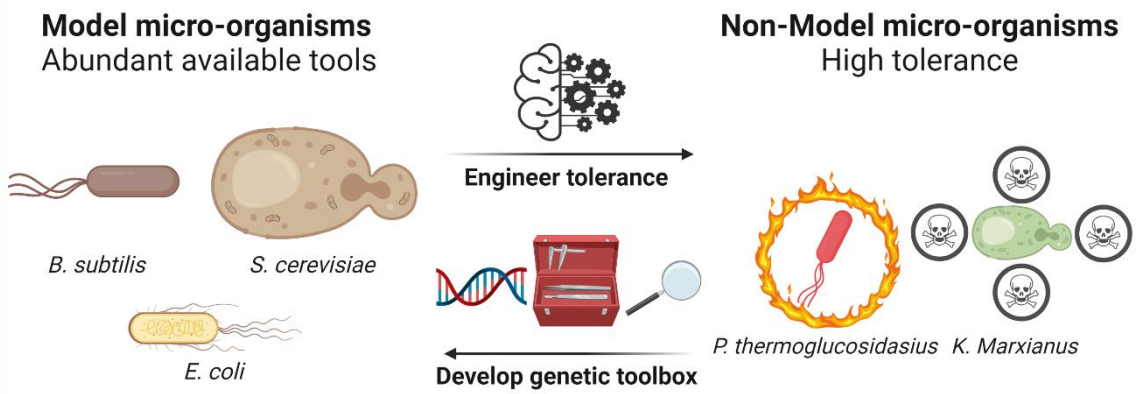


Figure 8 Strain engineering of Model organisms versus Non-Model organisms.

B. subtilis is a recognized cell factory for producing platform chemicals, biopolymers and enzymes in both industrial and academic environments. Its well-defined endogenous metabolism, distinct genetic background, generally recognized as safe (GRAS) status and well-established emerging genetic manipulation tools make this species a widespread candidate as a production chassis [132]. The available toolbox for *B. subtilis* includes computational models [133,134], genetic parts libraries [135–137], vector standards [138,139], Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) tools [140], and genome-scale deletion libraries [141].

1.6.2 Metabolic Engineering Using Synthetic Biology

Synthetic biology. The concept of synthetic biology is that any biological system can be seen as a combination of individual functional elements that can be modified or recombined to create new biological systems [142]. In regards to microbial cell factories; characterization and standardization of genetic elements (like promoters, terminators, Shine Dalgarno sequences, plasmid backbones or genes) has received a lot of attention to facilitate building new biological systems [143]. It is envisioned that future cell factory design and synthesis will combine a specific minimum amount of elements to efficiently produce a specific chemical or protein by using advanced modelling (incl. machine learning) and automation [144]. Whereas synthetic biology can be defined as a way to provide information and the components for biological systems, metabolic engineering applies the tools of synthetic biology to rewire the cell's metabolism to optimize the production of a desired compound [145].

Metabolic engineering. A more efficient biochemical production can be achieved by increasing pre-cursors and co-factor supply, up-regulating the expression of biochemical pathway enzymes, decreasing by-product formation, deleting product-degrading enzymes, removing product feedback inhibition and co-localizing pathway enzymes.

On the other hand, heterologous protein production can be improved by promoter and codon optimization, and improvement of secretion pathways [127]. Besides focussing on the production pathways, improved tolerance and co-utilization of the different substrates an organism can use can improve productivity.

In the perspective of using lignocellulose, metabolic engineering can be a useful tool to enhance the co-utilization of multiple substrates to increase product yields [146] or to degrade inhibitory compounds as to biologically detoxify biomass hydrolysates [147].

Although scientists are more and more equipped with advanced synthetic biology tools for genome editing and phenotypic characterization of cell factories, our recent knowledge about how metabolism is regulated is still limited. Therefore, developing cell factories for a commercial scale bioprocess is still a timely and costly affair, typically lasting for 6-8 years and requiring over 50 million USD [148]. Besides cell factory optimization, iterative prototyping is a widely adopted method in many engineering disciplines to efficiently explore the solution space by designing new experiments based on what is learned from data coming from previous designs. An example is the repeated optimization of CFD models by analysing the organism's response in scale-down experiments[149]. When applied in the field of synthetic biology and metabolic engineering, this process is commonly described as the design-build-test-learn cycle (DBTL-cycle, **Figure 9**). Considerable efforts have been invested in setting up so-called "bio-foundries" in both academic and industrial settings (IFF, Ginkgo, Zymergen), which adopt the DBTL-cycle in combination with rigorous automation and modelling to screen designs in a high-throughput manner and to speed up cell factory development. Issues to be addressed are related to the reliability and integration of hardware, the high capital investment needed to start biofoundries, the ethical

implication of using systems able to autonomous learn about biosystems [144,150,151].

Design. In the design phase, the blueprint for the cell factory, the product candidate and its synthesis is laid out, based on input from the previous learn phase, databases (KEGG [152], MetaCyc [153], BRENDA [154]) or theoretical hypothesis. The choice for the chassis organism is based on the inherent ability of the host to use the envisioned substrate and to express the product, tolerance to the product and the bioprocess conditions, strain physiology, and availability of genetic tools for the specific strain. In case metabolic pathways remain elusive, biosynthetic gene clusters or novel pathways can be identified by retrosynthetic methods using a variety of computational tools [155]. Standardization and characterization of biological parts (promoters, Shine-Dalgarno sequences, terminators etc.) facilitate the tuning of expression of the involved enzymes in the selected biosynthetic pathway. Genome-scale models allow *in silico* identification of beneficial changes for improved productivity of the target metabolite, like knockouts genes that prevent by-product formation, evaluation of metabolic fluxes, substrate consumption and biomass formation [156]. Moreover, artificial intelligence and computer-aided design will play an increasing role in assisting this phase.

Build. In the build phase, synthesis, assembly and insertion of the biological parts in the selected chassis strain should result in the designed cell factory. This phase includes constructing the metabolic pathway, deregulating the central carbon metabolism to rewire the flux to product and deleting pathways that might compete with product formation [148]. Decreasing costs of DNA synthesis (from oligonucleotides to whole genomes) and advances in genetic tool development (such as the development of CRISPR/CAS9 systems) accelerate this phase [157–159]. Although there are advances in characterization of biological elements, the behaviour of these elements is often context-dependent yielding it difficult to accurately program a metabolic pathway that leads to a desired level [148]. Here, the synthesis of large libraries can offer a way to screen millions of variants and select the best one [160]. Besides rational designs, cell factories can also be constructed by applying adaptive laboratory evolution, which will be discussed later [161].

Test. During the test phase, analytic tools provide data related to target molecules, transcripts, proteins, and metabolites of the designed system. Multi-omics analysis can provide a system-level view of the cell factory. For genomics and transcriptomics, next-generation sequencing is often used, whereas gas chromatography, liquid chromatography, mass spectrometry), Raman spectroscopy, and nuclear magnetic resonance are often used for proteomics and metabolomics analysis[162]. Depending on the output to be measured, the test phase can become the rate-limiting step of the strain engineering cycle as the runtime of some analysis are too long for the many variants created in the previous phase [155]. Although there have been advances within high-throughput multi-omics workflows, test capacity still poses a bottleneck in the cycles of strain engineering [144]. In contrast, developing a biosensor by coupling productivity to antibiotic resistance or fluorescence can yield specific data for many thousands of variants [163,164]. Consequently, high performing variants can be selected by fluorescence-activated cell sorting (FACS) (thousands of variants at the single-cell level) or microtiter plate screens.

Learn. The last phase of the cycle is concerned with the analysis of data to pinpoint flaws in the current biological system, which should be attended to in the next cycle. There is still a lot to gain in this phase, as data analysis is often not systematic, lacks statistical validation, relies on literature data, non-methodical observations and the researcher's intuition, who are responsible for the next round of optimization [148]. Output from the learn phase can improve metabolic models, the host organism or the production pathway. Also, learning how general metabolism is controlled can be particularly useful, as altering regulation can sometimes have more impact than simply overexpressing specific pathway enzymes [165]. On top of mechanistic models, artificial intelligence can contribute to redesign of biological systems by deducing patterns and trends using increasingly available data and computational power [166].

The degree to which the genotype-to-phenotype mapping of the cell factory is known, greatly influences the success of rational design used in the DBTL-cycle. However, as biological systems are often highly complex, our knowledge can be insufficient to perform such hypothesis-driven approaches. In contrast, evolution can be used as a driving force to naturally select mutants with increased fitness for strain optimization without the need to know the required genotype beforehand. As such, evolutionary engineering approaches can complement or replace entire steps of the DBTL cycle (**Figure 9**) [161].

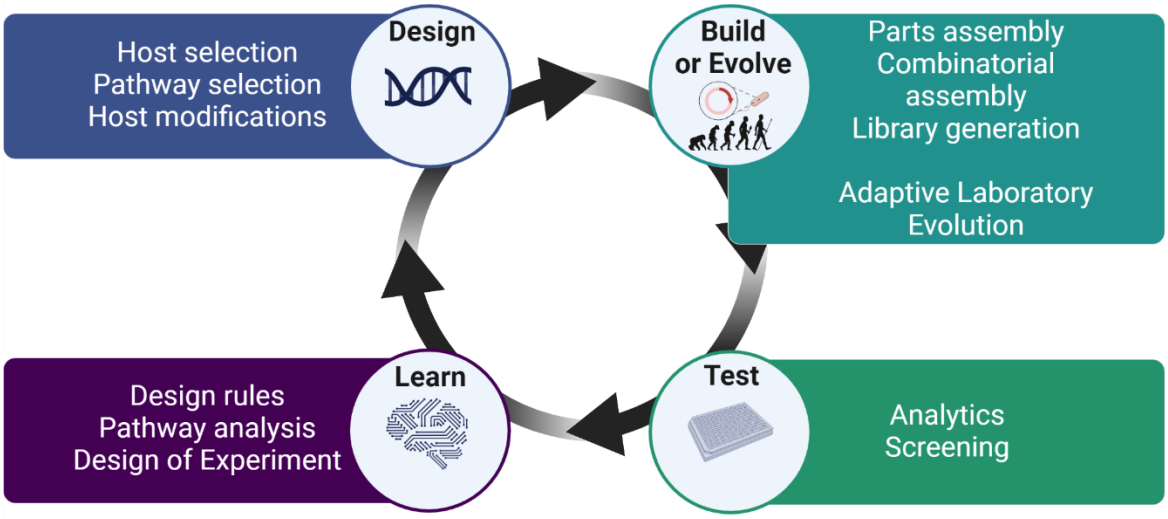


Figure 9 Schematic representation of the hypothesis-driven biological engineering by the DBTL cycle and the possibility of evolutionary engineering approaches to complement or replace entire steps of the DBTL cycle.

1.6.3 Evolutionary Engineering

As described earlier, cell factories should be tailored to the needs of the industrial process. This involves the introduction of biological functions, such as tolerance, genomic stability, co-consumption and high productivity, which are extremely complex in their regulation [167]. Applying evolutionary engineering can complement rational knowledge-based strain engineering (Figure 9), as it allows obtaining variants with the desired complex traits without knowledge of the genomic background. The increasing availability of low cost, high throughput DNA sequencing and the development of bioinformatics tools closes the genotype-phenotype knowledge gap, and accelerates the understanding of crucial intertwined regulatory networks [168].

Definitions. Evolutionary engineering involves the generation of spontaneous or some form of random or targeted mutagenesis in combination with *in vivo* selection for a desired phenotype [169]. Both directed evolution and adaptive laboratory evolution (ALE) are examples of evolutionary engineering and are sometimes used interchangeable, but have key differences. Directed evolution generally targets a particular gene for mutagenesis and screens for a desired trait, which is often independent of fitness effects. The potential of directed evolution has been recently acknowledged as the 2018 Nobel Prize for Chemistry was awarded to Frances Arnold, George P. Smith, and Gregory P. Winter for their work on creating enzymes with completely new properties [170]. In contrast, ALE experiments are carried out by prolonged culturing of cells in an environment with increasing selective pressure (for example a toxic compound). By time, variants which gained genome-wide mutations beneficial to the fitness will outcompete their ancestors and therefore be selected [161].

ALE is a popular method to overcome stress-induced growth inhibition by hydrolysate-associated inhibitory compounds, as it involves tuning the expression of multiple genes to alter complex physiological stress responses. As such, ALE has been used to yield tolerant strains for species like *Saccharomyces cerevisiae, Escherichia coli, Clostridium thermocellum and Corynebacterium glutamicum* [171–176].

Setups. The specific environment during the ALE experiment dictates what selective pressure is experienced by the cells and is thus critical for adaptation towards a desired phenotype. Currently, the most popular ALE methods are serial batch culturing and continuous chemostat cultures [177]. Serial batch culturing is popular for its simplicity, but fitness in serial batch culturing experiments can be different depending on the experimental conditions. When cells are kept in the exponential phase, fitness is directly related to growth rate, while if the culture undergoes different growth phases, the lag phase and survival during the stationary phase also come into play. For ALE experiments, it can be advantageous to keep cells in the exponential phase, however, this can be cumbersome when doing parallel batch experiments with changing growth rates. In contrast, a chemostat allows scientists to control the growth rate of the culture by setting the applied dilution rate. In addition, a bioreactor setup permits more control of environmental conditions such as pH and dissolved oxygen. However, it can be challenging to perform many ALE experiments in parallel, and undesired adaptations, like adhesive growth to the bioreactor wall, can occur [178].

Automation. In principle, ALE experiments are not extremely complicated to perform by hand, nor do they require sophisticated equipment. Nonetheless, automation can offer many advantages in terms of throughput, dynamic cell culturing, shorter timeframes and improved monitoring of the cultures, which cannot reasonably be performed manually [161]. The option to perform the same evolution experiment by many independent replicates results in much clearer outcomes as mutations which occur in multiple independent experiments are highly likely to be adaptive [179]. As such, an automated serial batch experiment combines the advantage of parallel experiments, while it keeps a constant selection pressure. In this thesis, automated adaptive laboratory evolution was employed to develop a *B. subtilis* strain with improved tolerance to inhibitors associated with lignocellulosic biomass (Chapter 4).

1.7 Concluding Remarks

The field of biotechnology offers great potential to minimize humanity's dependence on fossil fuels by revolutionizing the way we produce our chemicals and fuels. At first glance, the use of renewable biomass appears to be rather straightforward. However, nothing can be further from the truth. One major obstacle is that humanity does not have the luxury of merely using edible crops that are easy fermented to bulk fuels and chemicals, as it jeopardizes other aspects of sustainable development. To date, intertwined technical, political and economic hindrances have proven to be so complex that the realization of industrial-scale bio-based processes using non-edible crops is scarce.

This PhD thesis presents a small contribution to the development of second-generation biorefineries by studying some of the main challenges associated with the use of lignocellulosic biomass. The focus was to overcome the toxicity associated with lignocellulosic hydrolysates and explore ways to lower the production costs of enzymes that are required in the process. The main body of the thesis was done from the perspective of an on-site enzyme production process, using DDGS biomass as a fermentation feedstock. Regardless, the main findings in the specific context can have broader implications for the use of lignocellulose.

The first study of the thesis presented different models to optimize the operational conditions of small-scale pretreatment experiments using lignocellulosic biomass and showed the addition of small amounts of sulfuric acid significantly increased the performance of the pretreatment, i.e. high monomer yield and relatively low inhibitor formation (Chapter 2). In the second study of this thesis, compounds formed during pretreatment of lignocellulosic biomass were compared and showed to have clear independent and combined inhibitory effects on *B. subtilis* (Chapter 3). A tolerance adaptive laboratory evolution was applied to overcome this obstacle, and obtain *B. subtilis* strains with increased tolerance to inhibitory compounds present in DDGS-based hydrolysate (Chapter 4). Whole-genome resequencing data of independently evolved isolates revealed key mutations related to

tolerance and the employed cultivation conditions. Lastly, a techno-economic assessment showed that DDGS-based enzyme manufacturing was unfeasible under the current conditions (Chapter 5). Irrespective, the potential of DDGS-based on-site enzyme production was clear and conditions required to realize increased cost-effective production of enzymes were specified.

In conclusion, the presented outcomes lead to a better understanding of the toxicity of biomass hydrolysates and cost-effective enzyme production. As these are two major hurdles associated with using lignocellulose as a feedstock, this knowledge can be used to advance the development of second-generation biorefineries in the future.

1.8 Future Perspectives

"Man tries to make for himself in the fashion that suits him best a simplified and intelligible picture of the world; he then tries to some extent to substitute this cosmos of his for the world of experience, and thus to overcome it. This is what the painter, the poet, the speculative philosopher, and the natural scientists do, each in his own fashion."

- Albert Einstein, 'Principles of Research' 1934

As Einstein described so many years ago, I made my own simplified picture of the 'world of biotechnology'. Likewise, this cosmos was corrected many times during the PhD study, and I hope it will be corrected many times more in the time to come.

Without a doubt, the realization of second-biorefineries requires further efforts to solve the problems both inside and outside the scope of this thesis.

Although we showed that evolutionary engineering could increase the tolerance of cell factories to toxic compounds in biomass hydrolysates, this will not be sufficient by itself. As industrial implementation of lignocellulose require concentration steps to reach high sugar concentrations, the amounts of inhibitors increases concomitantly and will likely exceed the biological limits of cell factories. Therefore, the advancement of milder pretreatment processes will dictate the degree of detoxification required to keep the inhibitory compounds within the limits of biology. Innovative pretreatment methods employing milder process conditions are developed and well underway to be cost-competitive[180,181].

When considering the costs of enzymes required for second-generation biorefineries, integrated enzyme production using DDGS biomass offers significant potential cost reductions. To validate the techno-economic assessment performed, more research is required to optimize the protein extraction of DDGS-biomass, and the conversion of DDGS-hydrolysate towards enzymes. Moreover, an additional life-cycle

assessment of DDGS-based on-site enzyme production is crucial to provide a well-founded claim of more sustainable practice based on quantitative data. As the change in production location, setup and feedstock entail considerable technical and financial risks, strong collaboration between biorefineries and enzyme-producing companies is essential. Enzyme-producing companies have decades of experience in both R&D and first-hand industrial practice. Using highly engineered production hosts, optimized fermentation processes, and relevant data on production costs is crucial to model the entailed risk and benefits.

From a more holistic perspective, the advancement of cell factory development will significantly support the realization of second-generation biorefineries. Optimization of the production chassis for the use of second-generation hydrolysates includes the co-utilization of different substrates and improved tolerance to hydrolysate-based media in industrial conditions. Decreased cost for strain development and increased speed of the DBTL-cycle is essential to driving this field. The build step will benefit from further development of high-throughput DNA assembly techniques and decreased DNA synthesis and sequencing costs. Moreover, high-throughput screening methods and even more standardization of biological parts will improve the test and design steps. Lastly, the introduction of artificial intelligence and improved metabolic models can increasingly support learn and design steps. The emerging biofoundries will most likely have an important role in the rapid development of strains tailored to different biorefinery setups.

Looking beyond the cell factory, the interdisciplinary character of bioprocess development makes it paramount that scientists of different disciplines collaborate at each step of the journey. This includes, but is not limited to molecular biology, process engineering, and computational sciences. Being “king of your own hill” can tailor advantages in an academic setting. However, the realization of a bioprocess is doomed to fail when transparency, collaboration, trust and synergy between researchers of different fields are not prioritized. In the end, none of us is smarter than all of us.

The implementation of industrial-scale second-generation biorefineries should be technically feasible, as there are multiple examples. However, the past also shows that the development of 2G biorefineries will hamper without considerable public support and clear governmental policies to make initial projects economically attractive for industry stakeholders to participate. New production processes need time to "learn by doing", form value chains, optimize large-scale design options to lower costs and become competitive.

Ultimately, the geographically dispersed biomass should be used as efficient as possible in a cascading way that ensures food, feed and the necessary carbon source for chemicals, while less is used for energy. Currently, first-generation biorefineries produce not only fuels but also an increasing number of bio-based chemicals. In the longer term, commercial manufacturing of bio-based products will likely be realized using edible crops, lignocellulose and a combination of CO_2 renewable energy, i.e. biorefineries of different generations. Even solutions beyond the realm of biotechnology, like carbon capture and storage or sustainable chemical processes using renewable energy (power-to-x), will be used to produce chemicals. Assessment of standardized economic and sustainability metrics associated with these solutions will be decisive for their implementation.

As public support is essential to drive the initial stages of sustainable development, communication about the benefits and urgency of bio-based production patterns deserve more attention. This is not only essential to realize an increased awareness of the climate crisis at hand, but also offers society perspective and solutions to rally behind. In the end, science can present great technologies to sustain ourselves on the planet in the future, but it takes societal and political commitment to realize them.

1.9 References

1. Field, C. B., Barros, V. R., Mastrandrea, M. D., Mach, K. J., Abdrabo, M.-K., Adger, N., Anokhin, Y. A., Anisimov, O. A., Arent, D. J., Barnett, J. *Summary for policymakers. In Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. Contribution ofWorkingGroup II to the Fifth Assessment Report ofthe Intergovernmental Panel on Climate Change. Cambridge*. *Josef Settele* (Alistair Woodward, 2014).
2. Intergovernmental Panel on Climate Change. *Climate Change 2014 Mitigation of Climate Change*. *Climate Change 2014 Mitigation of Climate Change* (2014). doi:10.1017/cbo9781107415416
3. Allen, M. *et al.* Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels. *Aromar Revi*
4. IEA. Net Zero by 2050: A Roadmap for the Global Energy Sector. *Int. Energy Agency* 224 (2021).
5. BP. BP Energy Outlook – 2018 edition.
6. Noorman, H. J. & Heijnen, J. J. Biochemical engineering's grand adventure. *Chem. Eng. Sci.* **170**, 677–693 (2017).
7. Roddy, D. J. Biomass in a petrochemical world. *Interface Focus* **3**, (2013).
8. Serpell, O. & Paren, B. Balancing Act: Can Petrochemicals Be Both Emissions Free and Zero-Waste? (2021).
9. Geoff Bell, A. *et al.* IEA BIOENERGY Task42 BIOREFINING Sustainable and synergetic processing of biomass into marketable food & feed ingredients, products (chemicals, materials) and energy (fuels, power, heat) The Netherlands (coordinator) Electronic copies. (2014).
10. European Commission. *EU Biorefinery Outlook to 2030*. (2021). doi:10.2777/103465
11. Vaishnav, P. & Demain, A. L. Industrial Biotechnology (Overview). *Encycl. Microbiol.* 665–680 (2019). doi:10.1016/B978-0-12-809633-8.13064-X
12. Straathof, A. J. J. *et al.* Grand Research Challenges for Sustainable Industrial Biotechnology. *Trends Biotechnol.* **37**, 1042–1050 (2019).
13. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. Available at: https://sdgs.un.org/2030agenda. (Accessed: 3rd February 2022)
14. Industrialization of biology: A roadmap to accelerate the advanced manufacturing of chemicals. *Ind. Biol. A Roadmap to Accel. Adv. Manuf. Chem.* 1–150 (2015). doi:10.17226/19001
15. Sanyé-Mengual, E., Secchi, M., Corrado, S., Beylot, A. & Sala, S. Assessing the

decoupling of economic growth from environmental impacts in the European Union: A consumption-based approach. *J. Clean. Prod.* **236**, 117535 (2019).

16. International Renewable Energy Agency - IRENA. *Global Energy Transformation: The REmap transition pathway (Background report to 2019 Edition)*. (2019).

17. Festel, G. Industrial biotechnology: Market size, company types, business models, and growth strategies. *https://home.liebertpub.com/ind* **6**, 88–94 (2010).

18. Nieuwenhuizen, P. J. & Lyon, D. Anticipating opportunities in industrial biotechnology: Sizing the market and growth scenarios. *J. Commer. Biotechnol.* **17**, 159–164 (2011).

19. Festel, G. Economic Aspects of Industrial Biotechnology. *Adv. Biochem. Eng. Biotechnol.* **173**, 53–74 (2020).

20. Nakamura, C. E. & Whited, G. M. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **14**, 454–459 (2003).

21. Burgard, A., Burk, M. J., Osterhout, R., Van Dien, S. & Yim, H. Development of a commercial scale process for production of 1,4-butanediol from sugar. *Curr. Opin. Biotechnol.* **42**, 118–125 (2016).

22. Yim, H. *et al.* Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat. Chem. Biol. 2011 77* **7**, 445–452 (2011).

23. Choi, S. *et al.* Highly selective production of succinic acid by metabolically engineered Mannheimia succiniciproducens and its efficient purification. *Biotechnol. Bioeng.* **113**, 2168–2177 (2016).

24. Clarke, L. & Kitney, R. Developing synthetic biology for industrial biotechnology applications. *Biochem. Soc. Trans.* **48**, 113 (2020).

25. JRC Publications Repository - Consequences, Opportunities and Challenges of Modern Biotechnology for Europe. Available at: https://publications.jrc.ec.europa.eu/repository/handle/JRC36951. (Accessed: 23rd January 2022)

26. Okun, D. T. & Rogowsky, R. A. Industrial Biotechnology: Development and Adoption by the U.S. Chemical and Biofuel Industries.

27. Woodley, J. M. Towards the sustainable production of bulk-chemicals using biotechnology. *N. Biotechnol.* **59**, 59–64 (2020).

28. Werpy, T. & Petersen, G. Top Value Added Chemicals from Biomass Volume I. *Us Nrel* Medium: ED; Size: 76 pp. pages (2004). doi:10.2172/15008859

29. Sherpa Group & European Commission. KET – INDUSTRIAL BIOTECHNOLOGY. (2011).

30. Naik, S. N., Goud, V. V, Rout, P. K. & Dalai, A. K. Production of first and second

generation biofuels: A comprehensive review. doi:10.1016/j.rser.2009.10.003

31. Hazell, P. & Pachauri, R. K. BIOENERGY AND AGRICULTURE: PROMISES AND CHALLENGES. (2020).

32. Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources. Available at: https://eur-lex.europa.eu/eli/dir/2018/2001/oj. (Accessed: 6th February 2022)

33. Chandel, A. K., Garlapati, V. K., Singh, A. K., Antunes, F. A. F. & da Silva, S. S. The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization. *Bioresour. Technol.* **264**, 370–381 (2018).

34. Clariant. Welcome at Clariant in Romania. Available at: https://www.clariant.com/en/Company/Contacts-and-Locations/Key-Sites/Romania. (Accessed: 9th February 2022)

35. Redl, S. *et al.* Thermodynamics and economic feasibility of acetone production from syngas using the thermophilic production host Moorella thermoacetica. *Biotechnol. Biofuels* **10**, 1–17 (2017).

36. Heijstra, B. D., Leang, C. & Juminaga, A. Gas fermentation: cellular engineering possibilities and scale up. *Microb. Cell Fact.* **16**, 60 (2017).

37. World's First Commercial Waste Gas to Ethanol Plant Starts Up | LanzaTech. Available at: https://www.lanzatech.com/2018/06/08/worlds-first-commercial-waste-gas-ethanol-plant-starts/. (Accessed: 3rd February 2022)

38. Brennan, L. & Owende, P. Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products. *Renew. Sustain. Energy Rev.* **14**, 557–577 (2010).

39. Woo, H. M. Solar-to-chemical and solar-to-fuel production from CO2 by metabolically engineered microorganisms. *Curr. Opin. Biotechnol.* **45**, 1–7 (2017).

40. Popp, J., Lakner, Z., Harangi-Rákos, M. & Fári, M. The effect of bioenergy expansion: Food, energy, and environment. *Renew. Sustain. Energy Rev.* **32**, 559–578 (2014).

41. Deng, Y. Y., Koper, M., Haigh, M. & Dornburg, V. Country-level assessment of long-term global bioenergy potential. *Biomass and Bioenergy* **74**, 253–267 (2015).

42. Chatzifragkou, A. & Charalampopoulos, D. Distiller's dried grains with solubles (DDGS) and intermediate products as starting materials in biorefinery strategies. in *Sustainable Recovery and Reutilization of Cereal Processing By-Products* (2018). doi:10.1016/B978-0-08-102162-0.00003-4

43. Zabed, H., Sahu, J. N., Suely, A., Boyce, A. N. & Faruq, G. Bioethanol production

from renewable sources: Current perspectives and technological progress. *Renew. Sustain. Energy Rev.* **71**, 475–501 (2017).

44. Menon, V. & Rao, M. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. *Prog. Energy Combust. Sci.* **38**, 522–550 (2012).

45. van der Pol, E. C., Bakker, R. R., Baets, P. & Eggink, G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. *Appl. Microbiol. Biotechnol.* **98**, 9579–9593 (2014).

46. Bhatia, S. K. *et al.* Recent developments in pretreatment technologies on lignocellulosic biomass: Effect of key parameters, technological improvements, and challenges. *Bioresour. Technol.* **300**, (2020).

47. Vanholme, R., Demedts, B., Morreel, K., Ralph, J. & Boerjan, W. Lignin Biosynthesis and Structure. *Plant Physiol.* **153**, 895–905 (2010).

48. Galbe, M. & Wallberg, O. Pretreatment for biorefineries: a review of common methods for efficient utilisation of lignocellulosic materials. *Biotechnol. Biofuels 2019 121* **12**, 1–26 (2019).

49. Pereira, J. P. C., Verheijen, P. J. T. & Straathof, A. J. J. Growth inhibition of S. cerevisiae, B. subtilis, and E. coli by lignocellulosic and fermentation products. *Appl. Microbiol. Biotechnol.* **100**, 9069–9080 (2016).

50. Palmqvist, E. & Hahn-Hägerdal, B. Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms of inhibition. *Bioresour. Technol.* **74**, 25–33 (2000).

51. Jönsson, L. J. & Martín, C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour. Technol.* **199**, 103–112 (2016).

52. Gupta, R., Hemansi, Gautam, S., Shukla, R. & Kuhad, R. C. Study of charcoal detoxification of acid hydrolysate from corncob and its fermentation to xylitol. *J. Environ. Chem. Eng.* **5**, 4573–4582 (2017).

53. van der Pol, E. C., Bakker, R. R., Baets, P. & Eggink, G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. *Applied Microbiology and Biotechnology* **98**, 9579–9593 (2014).

54. Van Der Maas, L. The effect of inhibitory compounds in biomass hydrolysate on Bacillus subtilis. 1–17 (2020).

55. Monlau, F. *et al.* Do furanic and phenolic compounds of lignocellulosic and algae biomass hydrolyzate inhibit anaerobic mixed cultures? A comprehensive review. *Biotechnol. Adv.* **32**, 934–951 (2014).

56. Van Dyk, J. S. & Pletschke, B. I. A review of lignocellulose bioconversion using

enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy. *Biotechnol. Adv.* **30**, 1458–1480 (2012).

57. Costa, S. P. F., Azevedo, A. M. O., Pinto, P. C. A. G. & Saraiva, M. L. M. F. S. Environmental Impact of Ionic Liquids: Recent Advances in (Eco)toxicology and (Bio)degradability. *ChemSusChem* **10**, 2321–2347 (2017).

58. Huang, R., Su, R., Qi, W. & He, Z. Bioconversion of Lignocellulose into Bioethanol: Process Intensification and Mechanism Research. *BioEnergy Res. 2011 44* **4**, 225–245 (2011).

59. Ruiz, H. A. *et al.* Engineering aspects of hydrothermal pretreatment: From batch to continuous operation, scale-up and pilot reactor under biorefinery concept. *Bioresour. Technol.* **299**, 122685 (2020).

60. Perez-Cantu, L. *et al.* Comparison of pretreatment methods for rye straw in the second generation biorefinery: Effect on cellulose, hemicellulose and lignin recovery. *Bioresour. Technol.* **142**, 428–435 (2013).

61. Alberts, G. *et al.* IRENA INNOVATION OUTLOOK ADVANCED LIQUID BIOFUELS.

62. Kumar, B., Bhardwaj, N., Agrawal, K., Chaturvedi, V. & Verma, P. Current perspective on pretreatment technologies using lignocellulosic biomass: An emerging biorefinery concept. *Fuel Process. Technol.* **199**, 106244 (2020).

63. Duque, A., Álvarez, C., Doménech, P., Manzanares, P. & Moreno, A. D. Advanced Bioethanol Production: From Novel Raw Materials to Integrated Biorefineries. *Process. 2021, Vol. 9, Page 206* **9**, 206 (2021).

64. Naresh Kumar, M., Ravikumar, R., Thenmozhi, S., Ranjith Kumar, M. & Kirupa Shankar, M. Choice of Pretreatment Technology for Sustainable Production of Bioethanol from Lignocellulosic Biomass: Bottle Necks and Recommendations. *Waste and Biomass Valorization* **10**, 1693–1709 (2019).

65. Bhatia, S. K. *et al.* Recent developments in pretreatment technologies on lignocellulosic biomass: Effect of key parameters, technological improvements, and challenges. *Bioresour. Technol.* **300**, 122724 (2020).

66. Zheng, Y., Shi, J., Tu, M. & Cheng, Y. S. Principles and Development of Lignocellulosic Biomass Pretreatment for Biofuels. *Adv. Bioenergy* **2**, 1–68 (2017).

67. Rasmussen, H., Sørensen, H. R. & Meyer, A. S. Formation of degradation compounds from lignocellulosic biomass in the biorefinery: sugar reaction mechanisms. *Carbohydr. Res.* **385**, 45–57 (2014).

68. Pedersen, M. & Meyer, A. S. Lignocellulose pretreatment severity – relating pH to biomatrix opening. *N. Biotechnol.* **27**, 739–750 (2010).

69. Amore, A., Ciesielski, P. N., Lin, C. Y., Salvachuá, D. & Nogué, V. S. I.

Development of Lignocellulosic Biorefinery Technologies: Recent Advances and Current Challenges. *Australian Journal of Chemistry* **69**, 1201–1218 (2016).

70. Sankaran, R. *et al.* Recent advances in the pretreatment of microalgal and lignocellulosic biomass: A comprehensive review. *Bioresour. Technol.* **298**, 122476 (2020).

71. Moreno, A. D. & Olsson, L. Pretreatment of lignocellulosic Feedstocks. *Extrem. Enzym. Process. Lignocellul. Feed. to Bioenergy* 31–52 (2017). doi:10.1007/978-3-319-54684-1_3

72. Singh, S. Designing tailored microbial and enzymatic response in ionic liquids for lignocellulosic biorefineries. *Biophys. Rev.* **10**, 911–913 (2018).

73. Vollmer, N. I. *et al.* Model development for the optimization of operational conditions of the pretreatment of wheat straw. *Chem. Eng. J.* **430**, 133106 (2022).

74. Galbe, M. & Zacchi, G. Pretreatment of Lignocellulosic Materials for Efficient Bioethanol Production. *Adv. Biochem. Eng. Biotechnol.* **108**, 41–65 (2007).

75. Sultana, A. & Kumar, A. Optimal configuration and combination of multiple lignocellulosic biomass feedstocks delivery to a biorefinery. *Bioresour. Technol.* **102**, 9947–9956 (2011).

76. Nanda, S., Azargohar, R., Dalai, A. K. & Kozinski, J. A. An assessment on the sustainability of lignocellulosic biomass for biorefining. *Renew. Sustain. Energy Rev.* **50**, 925–941 (2015).

77. Vollmer, N. I. Conceptual Process Design in Biomanufacturing. (2022).

78. Alfano, S., Berruti, F., Denis, N. & Santagostino, A. The future of second-generation biomass. 1–5 (2016). Available at: https://www.mckinsey.com/business-functions/sustainability/our-insights/the-future-of-second-generation-biomass. (Accessed: 19th January 2022)

79. Ernsting, A. & Smolker, R. Dead End Road: The false promises of cellulosic biofuels. *Bioenergy Int.* 1–44 (2018).

80. Sanford, K., Chotani, G., Danielson, N. & Zahn, J. A. Scaling up of renewable chemicals. *Curr. Opin. Biotechnol.* **38**, 112–122 (2016).

81. Okolie, J. A., Mukherjee, A., Sonil Nanda, |, Dalai, A. K. & Kozinski, J. A. Next-generation biofuels and platform biochemicals from lignocellulosic biomass. (2021). doi:10.1002/er.6697

82. Dale, B. Time to Rethink Cellulosic Biofuels? *Biofuels, Bioprod. Biorefining* **12**, 5–7 (2018).

83. Lynd, L. R. *et al.* Cellulosic ethanol: status and innovation. *Curr. Opin. Biotechnol.* **45**, 202–211 (2017).

84. Crude oil prices. Available at: https://ourworldindata.org/grapher/crude-oil-prices. (Accessed: 7th February 2022)

85. de Jong, E. *et al.* Bio-Based Chemicals A 2020 Update Bio-Based Chemicals With input from: (pdf version) Published by IEA Bioenergy. (2020).

86. Jarboe, L. R. Progress and challenges for microbial fermentation processes within the biorefinery context. *A-Z Biorefinery* 447–471 (2022). doi:10.1016/B978-0-12-819248-1.00019-1

87. Sanford, K., Chotani, G., Danielson, N. & Zahn, J. A. Scaling up of renewable chemicals. *Curr. Opin. Biotechnol.* **38**, 112–122 (2016).

88. Wright, M. M. & Brown, R. C. Comparative economics of biorefineries based on the biochemical and thermochemical platforms. *Biofuels, Bioprod. Biorefining* **1**, 49–56 (2007).

89. Dragone, G. *et al.* Innovation and strategic orientations for the development of advanced biorefineries. *Bioresource Technology* **302**, (2020).

90. Lynd, L. R., Wyman, C. E. & Gerngross, T. U. Biocommodity engineering. *Biotechnol. Prog.* **15**, 777–793 (1999).

91. Sanford, K., Chotani, G., Danielson, N. & Zahn, J. A. Scaling up of renewable chemicals. *Curr. Opin. Biotechnol.* **38**, 112–122 (2016).

92. Nielsen, J., Larsson, C., van Maris, A. & Pronk, J. Metabolic engineering of yeast for production of fuels and chemicals. *Curr. Opin. Biotechnol.* **24**, 398–404 (2013).

93. Ragauskas, A. J. *et al.* Lignin valorization: Improving lignin processing in the biorefinery. *Science (80-.).* **344**, (2014).

94. Zhang, R. *et al.* Lignin valorization meets synthetic biology. *Engineering in Life Sciences* **19**, 463–470 (2019).

95. Bommarius, A. S., Sohn, M., Kang, Y., Lee, J. H. & Realff, M. J. Protein engineering of cellulases. *Curr. Opin. Biotechnol.* **29**, 139–145 (2014).

96. Hellsmark, H. & Söderholm, P. Innovation policies for advanced biorefi nery development: key considerations and lessons from Sweden. (2016). doi:10.1002/bbb.1732

97. Biofuelsdigest.com. Beta Renewables, Biochemtex ink deal for commercial-scale cellulosic biofuels project in Slovakia : Biofuels Digest. Available at: https://www.biofuelsdigest.com/bdigest/2014/10/06/beta-renewables-biochemtex-ink-deal-for-commercial-scale-cellulosic-biofuels-project-in-slovakia/. (Accessed: 9th February 2022)

98. Biofuelsdigest.com. Beta Renewables in cellulosic ethanol crisis, as Grupo M&G parent files for restructuring : Biofuels Digest. Available at: https://www.biofuelsdigest.com/bdigest/2017/10/30/beta-renewables-in-cellulosic-ethanol-crisis-as-grupo-mg-parent-files-for-restructuring/.

(Accessed: 9th February 2022)

99. Biofuelsdigest.com. DowDuPont to exit cellulosic biofuels business : Biofuels Digest. Available at: https://www.biofuelsdigest.com/bdigest/2017/11/02/breaking-news-dowdupont-to-exit-cellulosic-ethanol-business/. (Accessed: 9th February 2022)

100. IEA Bioenergy - Task 42. Biorefineries: adding value to the sustainable utilisation of biomass.

101. De Besi, M. & McCormick, K. Towards a Bioeconomy in Europe: National, Regional and Industrial Strategies. *Sustain. 2015, Vol. 7, Pages 10461-10478* **7**, 10461–10478 (2015).

102. Staffas, L., Gustavsson, M. & McCormick, K. Strategies and Policies for the Bioeconomy and Bio-Based Economy: An Analysis of Official National Approaches. *Sustain. 2013, Vol. 5, Pages 2751-2769* **5**, 2751–2769 (2013).

103. Hellsmark, H. & Söderholm, P. Innovation policies for advanced biorefinery development: key considerations and lessons from Sweden. *Biofuels, Bioprod. Biorefining* **11**, 28–40 (2017).

104. Huenteler, J., Diaz Anadon, L., Lee, H. & Santen, N. Commercializing Second-Generation Biofuels Scaling Up Sustainable Supply Chains and the Role of Public Policy rapporteur's report Energy Technology Innovation Policy.

105. Kwiatkowski, J. R., McAloon, A. J., Taylor, F. & Johnston, D. B. Modeling the process and costs of fuel ethanol production by the corn dry-grind process. *Ind. Crops Prod.* **23**, 288–296 (2006).

106. Perkis, D., Tyner, W. & Dale, R. Economic analysis of a modified dry grind ethanol process with recycle of pretreated and enzymatically hydrolyzed distillers' grains. *Bioresour. Technol.* **99**, 5243–5249 (2008).

107. Chatzifragkou, A. *et al.* Biorefinery strategies for upgrading Distillers' Dried Grains with Solubles (DDGS). *Process Biochemistry* (2015). doi:10.1016/j.procbio.2015.09.005

108. Vohra, M., Manwar, J., Manmode, R., Padgilwar, S. & Patil, S. Bioethanol production: Feedstock and current technologies. *J. Environ. Chem. Eng.* **2**, 573–584 (2014).

109. Schwarck, R. *et al.* RFA ETHANOL INDUSTRY OUTLOOK 2021. (2021).

110. Iram, A., Cekmecelioglu, D. & Demirci, A. Distillers' dried grains with solubles (DDGS) and its potential as fermentation feedstock. *Applied Microbiology and Biotechnology* **104**, 6115–6128 (2020).

111. Humbird, D. *et al.* Process design and economics for conversion of lignocellulosic biomass to ethanol. *NREL Tech. Rep. NREL/TP-5100-51400* **303**, 275–3000 (2011).

112. Ellilä, S. *et al.* Development of a low-cost cellulase production process using Trichoderma reesei for Brazilian biorefineries. *Biotechnol. Biofuels* **10**, 30 (2017).

113. Doran, P. M. Bioprocess Development: An Interdisciplinary Challenge. *Bioprocess Eng. Princ.* 3–8 (1995). doi:10.1016/b978-012220855-3/50001-8

114. Crater, J. S. & Lievense, J. C. Scale-up of industrial microbial processes. *FEMS Microbiol. Lett.* **365**, 138 (2018).

115. Straathof, A. J. J. The Proportion of Downstream Costs in Fermentative Production Processes. *Compr. Biotechnol. Second Ed.* **2**, 811–814 (2011).

116. Haringa, C. *et al.* Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. *Eng. Life Sci.* **16**, 652–663 (2016).

117. Wang, G., Haringa, C., Noorman, H., Chu, J. & Zhuang, Y. Developing a Computational Framework To Advance Bioprocess Scale-Up. *Trends Biotechnol.* **38**, 846–856 (2020).

118. Schäpper, D., Alam, M. N. H. Z., Szita, N., Eliasson Lantz, A. & Gernaey, K. V. Application of microbioreactors in fermentation process development: A review. *Anal. Bioanal. Chem.* **395**, 679–695 (2009).

119. Long, Q. *et al.* The development and application of high throughput cultivation technology in bioprocess development. *J. Biotechnol.* **192**, 323–338 (2014).

120. Lye, G. J., Ayazi-Shamlou, P., Baganz, F., Dalby, P. A. & Woodley, J. M. Accelerated design of bioconversion processes using automated microscale processing techniques. *Trends Biotechnol.* **21**, 29–37 (2003).

121. Chotani, G., Peres, C., Schuler, A. & Moslemy, P. Bioprocessing Technologies. *Bioprocess. Renew. Resour. to Commod. Bioprod.* **9781118175**, 133–166 (2014).

122. Da Silva, A. S. A. *et al.* Constraints and advances in high-solids enzymatic hydrolysis of lignocellulosic biomass: a critical review. *Biotechnol. Biofuels 2020 131* **13**, 1–28 (2020).

123. Li, X. *et al.* Highly efficient hemicellulose utilization for acetoin production by an engineered Bacillus subtilis. *J. Chem. Technol. Biotechnol.* **93**, 3428–3435 (2018).

124. Olson, D. G., McBride, J. E., Joe Shaw, A. & Lynd, L. R. Recent progress in consolidated bioprocessing. *Curr. Opin. Biotechnol.* **23**, 396–405 (2012).

125. Towler, G. & Sinnott, R. *Economic evaluation of projects*. *Chemical Engineering Design* (2022). doi:10.1016/b978-0-12-821179-3.00009-1

126. Shah, A., Baral, N. R. & Manandhar, A. *Technoeconomic Analysis and Life Cycle Assessment of Bioenergy Systems*. *Advances in Bioenergy* **1**, (Elsevier, 2016).

127. Davy, A. M., Kildegaard, H. F. & Andersen, M. R. Cell Factory Engineering. *Cell Syst.* **4**, 262–275 (2017).

128. Straathof, A. J. J. & Bampouli, A. Potential of commodity chemicals to become bio-based according to maximum yields and petrochemical prices. *Biofuels, Bioprod. Biorefining* **11**, 798–810 (2017).

129. Mol, V. *et al.* Genome-scale metabolic modeling of P. thermoglucosidasius NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism. *Metab. Eng.* **65**, 123–134 (2021).

130. Redl, S. *et al.* Thermodynamics and economic feasibility of acetone production from syngas using the thermophilic production host Moorella thermoacetica. *Biotechnol. Biofuels* **10**, (2017).

131. Lacerda, M. P., Oh, E. J. & Eckert, C. The Model System Saccharomyces cerevisiae Versus Emerging Non-Model Yeasts for the Production of Biofuels. *Life 2020, Vol. 10, Page 299* **10**, 299 (2020).

132. Gu, Y. *et al.* Advances and prospects of Bacillus subtilis cellular factories: From rational design to industrial applications. *Metabolic Engineering* **50**, 109–121 (2018).

133. Kocabaş, P., Çalık, P., Çalık, G. & Özdamar, T. H. Analyses of extracellular protein production in Bacillus subtilis – I: Genome-scale metabolic model reconstruction based on updated gene-enzyme-reaction data. *Biochem. Eng. J.* **127**, 229–241 (2017).

134. Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28791–28799 (2007).

135. Guiziou, S. *et al.* A part toolbox to tune genetic expression in Bacillus subtilis. *Nucleic Acids Res.* **44**, 7495–7508 (2016).

136. Castillo-Hair, S. M., Fujita, M., Igoshin, O. A. & Tabor, J. J. An Engineered B. subtilis Inducible Promoter System with over 10 000-Fold Dynamic Range. *ACS Synth. Biol.* **8**, 1673–1678 (2019).

137. Liu, D. *et al.* Construction, Model-Based Analysis, and Characterization of a Promoter Library for Fine-Tuned Gene Expression in Bacillus subtilis. *ACS Synth. Biol.* **7**, 1785–1797 (2018).

138. Falkenberg, K. B. *et al.* The ProUSER2.0 Toolbox: Genetic Parts and Highly Customizable Plasmids for Synthetic Biology in Bacillus subtilis. *ACS Synth. Biol.* acssynbio.1c00130 (2021). doi:10.1021/ACSSYNBIO.1C00130

139. Radeck, J., Meyer, D., Lautenschläger, N. & Mascher, T. Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for Bacillus subtilis. *Sci. Reports 2017 71* **7**, 1–13 (2017).

140. Altenbuchner, J. Editing of the Bacillus subtilis genome by the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **82**, 5421–5427 (2016).

141. Koo, B. M. *et al.* Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Syst.* **4**, 291-305.e7 (2017).

142. de Lorenzo, V. & Danchin, A. Synthetic biology: discovering new worlds and new words. *EMBO Rep.* **9**, 822–827 (2008).

143. Galdzicki, M. *et al.* The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol. 2014 326* **32**, 545–550 (2014).

144. Chao, R., Mishra, S., Si, T. & Zhao, H. Engineering biological systems using automated biofoundries. *Metab. Eng.* **42**, 98–108 (2017).

145. García-Granados, R., Lerma-Escalera, J. A. & Morones-Ramírez, J. R. Metabolic engineering and synthetic biology: Synergies, future, and challenges. *Front. Bioeng. Biotechnol.* **7**, 36 (2019).

146. Zhang, B. *et al.* Production of acetoin through simultaneous utilization of glucose, xylose, and arabinose by engineered Bacillus subtilis. *PLoS One* **11**, 1–14 (2016).

147. Chandel, A. K., da Silva, S. S. & Singh, O. V. Detoxification of Lignocellulose Hydrolysates: Biochemical and Metabolic Engineering Toward White Biotechnology. *Bioenergy Res.* **6**, 388–401 (2013).

148. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).

149. Wang, G. *et al.* Coupled metabolic-hydrodynamic modeling enabling rational scale-up of industrial bioprocesses. *Biotechnol. Bioeng.* **117**, 844–867 (2020).

150. Hillson, N. *et al.* Building a global alliance of biofoundries. *Nat. Commun. 2019 101* **10**, 1–4 (2019).

151. Holowko, M. B., Frow, E. K., Reid, J. C., Rourke, M. & Vickers, C. E. Building a biofoundry. *Synth. Biol.* **6**, (2021).

152. Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci.* **31**, 47–53 (2022).

153. MetaCyc: Metabolic Pathways From all Domains of Life. Available at: https://metacyc.org/. (Accessed: 30th January 2022)

154. Enzyme Database - BRENDA. Available at: https://www.brenda-enzymes.org/. (Accessed: 30th January 2022)

155. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* **3**, 135 (2015).

156. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol. 2019 201* **20**,

1–18 (2019).

157. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).

158. Luo, Z. *et al.* Whole genome engineering by synthesis. *Sci. China. Life Sci.* **61**, 1515–1527 (2018).

159. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**, 816–821 (2012).

160. Zeng, W., Guo, L., Xu, S., Chen, J. & Zhou, J. High-Throughput Screening Technology in Industrial Biotechnology. *Trends Biotechnol.* **38**, 888–906 (2020).

161. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56**, 1–16 (2019).

162. Misra, B. B., Langefeld, C., Olivier, M. & Cox, L. A. Integrated Omics: Tools, Advances, and Future Approaches. *J. Mol. Endocrinol.* **62**, R21–R45 (2018).

163. D'Ambrosio, V. & Jensen, M. K. Lighting up yeast cell factories by transcription factor-based biosensors. *FEMS Yeast Res.* **17**, (2017).

164. Zhang, J., Jensen, M. K. & Keasling, J. D. Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol.* **28**, 1–8 (2015).

165. Ostergaard, S., Olsson, L., Johnston, M. & Nielsen, J. Increasing galactose consumption by Saccharomyces cerevisiae through metabolic engineering of the GAL gene regulatory network. *Nat. Biotechnol.* **18**, 1283–1286 (2000).

166. Lawson, C. E. *et al.* Machine learning for metabolic engineering: A review. *Metab. Eng.* **63**, 34–60 (2021).

167. Yang, J., Kim, B., Kim, G. Y., Jung, G. Y. & Seo, S. W. Synthetic biology for evolutionary engineering: From perturbation of genotype to acquisition of desired phenotype. *Biotechnol. Biofuels* **12**, 1–14 (2019).

168. Shendure, J. *et al.* DNA sequencing at 40: past, present and future. *Nat. 2017 5507676* **550**, 345–353 (2017).

169. Shepelin, D., Hansen, A. S. L., Lennen, R., Luo, H. & Herrgård, M. J. Selecting the Best: Evolutionary Engineering of Chemical Production in Microbes. *Genes (Basel).* **9**, (2018).

170. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chemie Int. Ed.* **57**, 4143–4148 (2018).

171. Almario, M. P., Reyes, L. H. & Kao, K. C. Evolutionary engineering of *Saccharomyces cerevisiae* for enhanced tolerance to hydrolysates of lignocellulosic biomass. *Biotechnol. Bioeng.* **110**, 2616–2623 (2013).

172. Koppram, R., Albers, E. & Olsson, L. Evolutionary engineering strategies to

enhance tolerance of xylose utilizing recombinant yeast to inhibitors derived from spruce biomass. *Biotechnol. Biofuels* **5**, 32 (2012).

173. Linville, J. L. *et al.* Industrial Robustness: Understanding the Mechanism of Tolerance for the Populus Hydrolysate-Tolerant Mutant Strain of Clostridium thermocellum. *PLoS One* **8**, e78829 (2013).

174. Qin, D. *et al.* An auto-inducible Escherichia coli strain obtained by adaptive laboratory evolution for fatty acid synthesis from ionic liquid-treated bamboo hydrolysate. *Bioresour. Technol.* **221**, 375–384 (2016).

175. Wallace-Salinas, V. & Gorwa-Grauslund, M. F. Adaptive evolution of an industrial strain of Saccharomyces cerevisiae for combined tolerance to inhibitors and temperature. *Biotechnol. Biofuels* **6**, 151 (2013).

176. Wang, X., Khushk, I., Xiao, Y., Gao, Q. & Bao, J. Tolerance improvement of Corynebacterium glutamicum on lignocellulose derived inhibitors by adaptive evolution. *Appl. Microbiol. Biotechnol.* **102**, 377–388 (2018).

177. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution - principles and applications for biotechnology. *Microb. Cell Fact.* **12**, 1–17 (2013).

178. Rao, V. S. H. & Rao Sekhara, P. R. Global stability in chemostat models involving time delays and wall growth. *Nonlinear Anal. Real World Appl.* **5**, 141–158 (2004).

179. Bailey, S. F., Rodrigue, N., Kassen, R. & Barlow, M. The Effect of Selection Environment on the Probability of Parallel Evolution. doi:10.1093/molbev/msv033

180. BBI, U. Successful scale-up of the FABIOLA™ process in a lignocellulose biorefinery pilot plant – UNRAVEL BBI. Available at: http://unravel-bbi.eu/successful-scale-up-of-the-fabiola-process-in-a-lignocellulose-biorefinery-pilot-plant/. (Accessed: 24th April 2022)

181. Smit, A. & Huijgen, W. Effective fractionation of lignocellulose in herbaceous biomass and hardwood using a mild acetone organosolv process. *Green Chem.* **19**, 5505–5514 (2017).

Figures created with BioRender.com.

Chapter 2

MODEL DEVELOPMENT FOR THE OPTIMIZATION OF OPERATIONAL CONDITIONS OF THE PRETREATMENT OF WHEAT STRAW

Nikolaus I. Vollmer[1], Jasper L. S. P. Driessen[2], Celina K. Yamakawa[3], Krist V. Gernaey[1], Solange I. Mussatto[3], Gürkan Sin[1]*

[1] Process and Systems Engineering Research Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 228A, 2800 Kongens Lyngby, Denmark

[2] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800 Kongens Lyngby, Denmark

[3] Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 223, 2800 Kongens Lyngby, Denmark

*** Corresponding author**
Gürkan Sin
Email: gsi@kt.dtu.dk

This chapter comprises a published manuscript.
DOI: 10.1016/j.cej.2021.133106

Abstract

The underlying study presents models for the optimization of operational conditions of the pretreatment of what straw. Experiments for hydrothermal and dilute acid pretreatment are performed and analyzed. The highest xylose monomer yield obtained for dilute acid pretreatment is $Y_{Xyl} = 98$ % Y_Xyl=98 % at a temperature of $T = 195$℃, a reaction time of $t = 18\ min$ and a dilute acid concentration of $C_{acid} = 1.25\ wt$ %. The data is used to fit a response surface model (RSM), a Gaussian process regression model (GPR), and a mechanistic model based on thermodynamic principles and first-order rate equations. Each model is used in an optimization problem to predict the optimal operational conditions that maximize the xylose yield. The conditions found by the mechanistic model ($T = 191.6$ ℃, $t = 18\ min$, $C_{acid} = 1.13\ wt.\%$) with $C_{Xyl,mech} = 23.47\ wt.\%$ and the GPR ($T = 195$ ℃, $t = 18\ min$, $C_{acid} = 1.25\ wt.\%$) with $C_{Xyl,GPR} = 23.23\ wt.\%$ are in agreement and stand out compared to the RSM metamodeling approach ($T = 182.4$ ℃, $t = 26.2\ min$, $C_{acid} = 1.25\ wt.\%$), which yields $CC_{Xyl,RSM} = 25.72\ wt.\%$. Considering the scenario of uncertainty in the feedstock composition, the optimization under this uncertainty with the mechanistic model yields slightly different conditions ($T = 182.6$℃, $t = 18\ min$, $C_{acid} = 0.84\ wt.\%$) and $C_{Xyl,mech,uc} = 20.88\ wt.\%$. Given the underlying phenomena in the biomass pretreatment, all models have shortcomings; however, the mechanistic model is validated best overall and is thus recommended for further engineering purposes as, e.g., the conceptual process design of biorefineries.

2.1 Introduction

Biorefineries as a replacement for traditional chemical process routes play an essential role in developing more sustainable production patterns demanded by the 2030 Sustainable Development Agenda of the United Nations [1,2]. Nevertheless, despite harboring this immense potential, there are still significant knowledge gaps in industrial biotechnology regarding the use of optimal cell factories in fermentation processes and along the value chain. The latter mainly refers to questions about the optimal use of feedstocks for a set of various products, as well as questions about process intensification and integration, which remain unanswered [3–6].

There is a multitude of explanations for this fact, but foremost the fragile economic potential due to different cost-drivers has impaired a breakthrough of this concept for the time being [7,8]. The fundamental problem is a lack of facilitation in practical applications of the conceptual design of a biorefinery. Such conceptual approaches can describe the inherent complexity of bioprocesses under the use of non-conventional feedstocks to find an optimal and cost-efficient setup. However, such conceptual approaches are seldom employed despite many approaches being published [3,9–12]. From a technical perspective, many problems remain unsolved: lignocellulosic biomass as a sugar source for fermentation processes requires an extensive pretreatment process of the biomass to reduce the recalcitrance of the feedstock and to release both the sugars from the hemicellulosic fraction and the cellulosic fraction respectively. There has been steady progress in developing new pretreatment methods or improving existing methods further, but despite the abundant number of methods, selecting a suitable pretreatment is far from straightforward [13,14]. Furthermore, all pretreatment methods are associated with high capital and operational expenditures due to the extreme process conditions, which remains a prominent issue regarding the economics of a biorefinery [15].

From a holistic perspective, the biomass pretreatment both has a critical influence on the overall process economics and also a crucial role in the actual design process for a biorefinery, as critical decisions on the recovery of the fractions in the lignocellulosic biomass and subsequently the possible product sets for the biorefinery are made [16]. Independent from the potentially viable products, the optimal pretreatment unit fulfills three criteria: 1) a precise split between the fractions in the biomass, 2) high respective yields for the monomers in the fractions, and 3) a low formation of inhibitors in the pretreatment process. Depending on which products are supposed to be produced, the applied method can vary, but the requirements stay the same [13].

For a biorefinery with a viable economic potential, the utilization of the feedstock is the key factor. Hence, for its conceptual design, computational alternatives provide the potential to yield conceptually feasible solutions. The crucial factor in this is the availability of models that accurately describe the underlying unit operations to design the process [9].

This work focuses on the development of a biomass pretreatment model for this precise purpose. As a case study, the biotechnological production of xylitol in a biorefinery is chosen: in 2004, the US Department of Energy declared xylitol as one of the top 12 chemicals to be produced in a biorefinery [17]. Xylitol can be produced from xylose via fermentation in a suitable cell factory. Xylose is the main constituent of the hemicellulosic fraction in lignocellulosic biomass. These conditions make it an ideal product for the production in a biorefinery; however, research on xylitol production mainly focuses either on the pretreatment, or the fermentation, or the downstream processing, and a holistic perspective on the whole process is lacking [18–21].

In this study's scope, a suitable feedstock and pretreatment method for the given case are selected for the introduced criteria. Based on this, experiments for the selected feedstock and pretreatment methods are designed and performed. After analyzing the results of these experiments, different model candidates are calibrated to the data and subsequently validated, including an assessment of their robustness. All models are then taken to optimize the pretreatment conditions towards a selected objective for the case study. Lastly, these results are

compared, and the most suitable model candidate for the conceptual process design of a biorefinery is selected.

The remainder of this study is structured as follows: in Section 2, biomass pretreatment methods and feedstocks, in general, are introduced, and also the criteria applicable for selecting a method and a suitable feedstock in process design. Section 3 describes the experimental procedure of feedstock analysis, the design, performance, and analysis of the pretreatment experiments. Section 4 introduces both the data-driven models and the knowledge-driven models, which are employed in this study. For the knowledge-driven model, it is explained in detail how the parameters of the model are identified and estimated, as well as the employed procedure for the robustness assessment via uncertainty and sensitivity analysis. The procedure of validating all models is explained, as well as the optimization with all models. Section 5 includes the results of the feedstock analysis, both pretreatment experiments, the calibration and validation of data-driven models, the calibration, validation, and the robustness assessment of the knowledge-driven model, as well as the optimization study. Lastly, in Section 6, the study's primary outcomes are summarized, and instigations for further work are given.

2.2 Material and Methods

2.2.1 Feedstock Composition

For the performed experiments, wheat straw of the variety *Triticum aestivum* from a field in Freerslev sogn (Faxe, Denmark) of the harvest of fall 2018 was taken as feedstock. The wheat was harvested and dried. The chaff was separated from the wheat; the straw was dried again and then milled and ground. The resulting particle size after milling is in the range between 18 and 40 mm.

First, the respective feedstock is analyzed regarding its composition, particularly the hemicellulosic, the cellulosic, and the lignin fraction. This is performed with extractive-free biomass as described by the National Renewable Energy Laboratory (NREL) [22]. The extractives have been removed as described by the NREL [23]. Glucose, xylose, arabinose, acetic acid, formic acid, furfural, and 5-HMF were quantified by high-performance liquid chromatography (HPLC), using a Dionex Ultimate 3000 UHPLC+ Focused System (Dionex Softron GmbH, Germany) with a Bio-Rad Aminex column HPX-87H ($300 mm \times 7.8\, mm$) at 60 °C, and 5.0 mM sulfuric acid as mobile phase at a flow rate of 0.6 mL/min. Sugars and acids were detected using a Shodex RI-101 refractive index detector, whereas 5-HMF and furfural were detected using ultraviolet measurements at 254 nm.

2.2.2 Pretreatment Experiments

For each point in the design of experiments, a batch experiment is performed. The used batch vessel is a non-stirred pressure vessel (Parr Series 4760, 600 mL, Parr Instrument Company, Moline, IL). In preparation for the experiments, the milled biomass's moisture content and its dry mass are determined with an automated moisture analyzer (MB 163-M, VWR International) as a first step. For the hydrothermal pretreatment, the corresponding amount of water is weighed to meet

the set solid-to-liquid ratio. For the dilute acid pretreatment, the sulfuric acid and its water content are determined according to the experiment's acid concentration. The amount of water is weighed respectively. Then, the acid is mixed with the water. Subsequently, the biomass and the liquid are added alternatingly in small portions into the reaction vessel until it is filled. Lastly, the mixture is stirred thoroughly by hand to achieve an equal distribution of the liquid.

As soon as this is achieved, the reactor is closed and sealed, and put into a heating bath that is set to the desired temperature for the experiment. After a transition period of $t = 5\,min$ when the reaction vessel's temperature rises to the silicone bath's temperature level, the time measurement is started. After the desired residence time, the reaction vessel is taken out of the heating bath and cooled rapidly to inhibit further degradation reactions. After approximately 30 min, the reactor vessel is opened, and the liquid phase is separated from the solid phase through a simple sieve. The solid residue is put in a mechanical press to extract the maximum amount of free liquid phase, and hereafter its moisture content is determined again. The volume of the total amount of liquid hydrolysate is determined as well as its pH value. Lastly, the hydrolysate is filtered once with a vacuum filter in order to remove solid residues. The hydrolysate samples are also analyzed by HPLC, as depicted in Section 2.2.1.

2.2.3 Post-hydrolysis Experiments

The hydrothermal pretreatment with its neutral pH range proves to yield a significant amount of oligomeric sugars, indicating slower depolymerization reactions [13]. Therefore, the hydrolysate is commonly subjected to a post-hydrolysis step to increase the degradation of the oligomers into sugar monomers. This post-hydrolysis is performed as described by the NREL [24]. Hereafter, the samples are again analyzed by HPLC to determine the concentrations of the created monomers.

2.2.4 Response Surface Methodology

RSM as statistical analysis is described by Box and Wilson as follows: The actual response surface is a second-order polynomial fitted to experimental data to predict optimal conditions for the given set of factors D of the design of experiments. The general form of the polynomial is the following:

$$y = \alpha + \beta \cdot b + \sum_{i \in D} \gamma_i \cdot x_i + \sum_{i \in D} \sum_{j \geq i \in D} \delta_{ji} \cdot x_i \cdot x_j, \qquad (2\text{-}1)$$

In which x_i denote the factors, y the response and α, β, γ and δ the coefficients of the polynomial [25]. The RSM model with a second-order polynomial simplifies the underlying system significantly, but its use for statistical analysis of experiments remains high [26]. In this study, the RSM model in the *rsm* library in R is used [27]. After calibrating the model by fitting its parameters to experimental data, the validation of the model is crucial to confirm both the goodness of fit of the estimation and its predictive capability for the system it was calibrated for [28]. In the case of the RSM model, the model is validated by an analysis of variance (ANOVA). Furthermore, common validation metrics which quantitatively express both the goodness of fit and the predictive capacities are the coefficient of determination R^2 and the root mean square error $RMSE$, which are both used in this study. A detailed description is provided with the supplementary material.

For the hydrothermal pretreatment, the factors for the design of experiments are chosen to be the reaction temperature T in the interval $T = [160,195]$ °C and the reaction time in the interval $t = [20,60]\ min$. The central composite design (CCD) is chosen with $\alpha = 0.714$, two cube center points, and one star center point, which yields eleven experimental points. For the dilute acid pretreatment, the factors for the design of experiments are chosen to be the reaction temperature T in the interval $T = [160,195]$ °C, the reaction time t in the interval $t = [20,60]\ min$ and the acid concentration C_{ac} in the interval $C_{ac} = [0.8,1.3]\ wt\ \%$. The CCD is chosen to be circumscribed with $\alpha = 1.69$, two cube center points, and one star center point, which yields seventeen experimental points. The response variable y for all is the

xylose concentration. A detailed description of the setup of the design of experiments can be found in the supplementary material. The scripts for the design of experiments are provided through a GitHub repository [29].

2.2.5 Gaussian Process Regression

A GPR model's predictive capacities derive from the eponymous stochastic process: the prediction of interpolated values is governed by prior covariances of training data points and described by specific kernel functions, whose parameters are fitted to these points [30]. This is expressed as follows:

$$y = \mu(x) + \sigma^2 \cdot \mathcal{Z}(x, \omega), \qquad \mu(x) = \rho \cdot \beta(x). \tag{2-2}$$

Here, the predicted output y is described by $\mu(x)$ as the mean value of the stochastic process with the input x and σ^2 as its variance; $\mathcal{Z}(x, \omega)$ denotes a zero mean unit variance stochastic process with the mentioned kernel function ω. For the mean value, ρ relates to parameters which are fitted based on the training data and $\beta(x)$ describes a set of basis functions. Regarding the employable basis and kernel functions, the reader is referred to the book by Rasmussen [31]. In this study, the *fitrgp* function of the Statistics and Machine Learning Toolbox in MATLAB is used for fitting the GPR model. Analogously to the RSM model, the model input x consists of the three design factors T, t and C_{ac}, and the model output y respectively corresponds to the measured concentrations. The model validation is performed by cross-validation equally by calculating the coefficient of determination, and the root mean square error for the model, as described in Section 2.2.4. Further explanation on the validation is given in the supplementary material.

2.2.6 Mechanistic Model

The backbone of most published models for the pretreatment of lignocellulosic biomass are mass and energy balances; however, different types of reaction equations are employed. Mostly, these are pseudo-first or second-order reaction kinetics [32–34]. Generally, it is rather difficult to develop a wholesome model, as some reaction mechanisms and some components are unknown and vary highly between different feedstocks and pretreatment methods [35]. Furthermore, measuring these components in experimental setups is not straightforward, as subsequent reactions occur relatively fast at a particular stage. Hence, mechanistic models for biomass pretreatment commonly only capture one specific pretreatment method in a certain range of operational conditions for a selected number of components, depending on the provided experimental data and the respectively estimated parameters.

For the model in this study, we choose an equivalent approach based on mass balances and reaction kinetics. The considered set of components J in this model is listed in Table 1:

Table 1: List of considered components in the pretreatment model.

Number	Component	Shorthand symbol in model
1.	Xylan	(Xyn)
2.	Xylose	(Xyl)
3.	Cellulose	(Cel)
4.	Glucose	(Glu)
5.	Arabinan	(Arn)
6.	Arabinose	(Ara)
7.	Acetyl Groups	(Act)
8.	Acetic Acid	(Aac)
9.	Furfural	(Fur)
10.	5-Hydroxymethylfurfural (5-HMF)	(Hmf)
11.	Formic acid	(Fac)
12.	Further Degradation Products	(Deg)

The set of occurring reactions I which are considered in the model are illustrated in the following Figure 1:

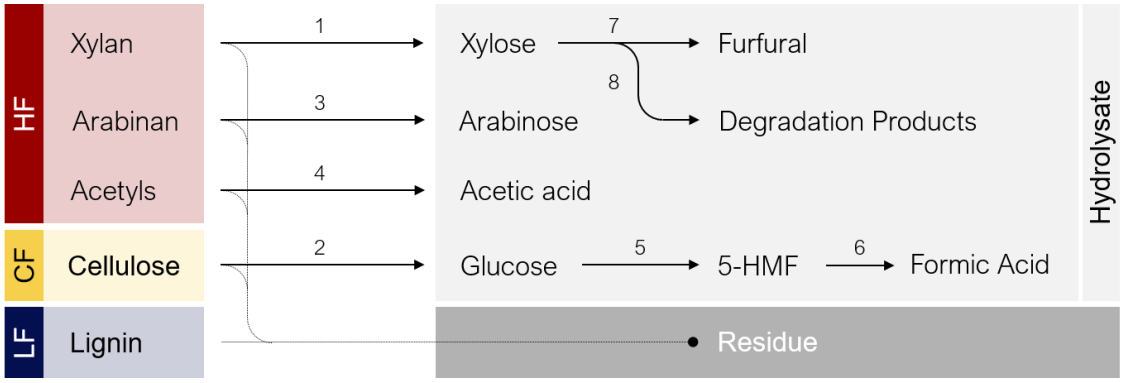


Figure 1: Illustration of the considered reactions occurring during the pretreatment, describing the transfer of components of the HF, CF, and LF into hydrolysate and residue

Hence, the number of reactions is $|I| = 8$. The reaction equations, including their stoichiometry, are provided in the supplementary material. Each reaction $i \in I$ occurs at a reaction rate r_i which can be calculated with the rate constant k_i of the reaction and the concentrations of the participating reactants C_j in $wt\%$ or $g/100g$ of dry biomass for $j \in J_i \subset J$ with J_i as a subset of the set of all reactants J.

$$r_i = k_i \cdot \prod_{j \epsilon J_i} C_j \quad \forall\, i \in I. \tag{2-3}$$

The rate constant k_i for every reaction i can be determined by the Arrhenius law as follows:

$$k_i = A_i \cdot exp\left(-\frac{E_{A,i}}{\tilde{R} \cdot T}\right) \cdot C_{acid}^{n_i} \quad \forall\, i \in I, \tag{2-4}$$

with A_i as the frequency factor in min^{-1}, $E_{A,i}$ as the activation energy in $kJ \cdot mol^{-1}$, $\tilde{R}$ as the universal gas constant in $kJ \cdot mol^{-1} \cdot K^{-1}$, T as the temperature of the reaction in K, C_{acid} as the concentration of supplied acid in case of dilute acid pretreatment in $wt\%$ or $g/100g$ of dry biomass and n_i as reaction order exponent for the participation of acid in each respective reaction.

For each component C_j a component balance can be formulated:

$$\frac{dC_j}{dt} = \sum_{i \in I_J} s_i \cdot r_i \quad \forall\, j \in J, \tag{2-5}$$

with $I_J \subset I$ as the subset of reactions which involve component j and s_i as the stoichiometric factor for the reaction r_i. All stoichiometric factors can be summarized in the stoichiometric matrix S with the dimensions $|I| \times |J|$:

$$S = \begin{pmatrix} -1 & 1.136 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1.111 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1.136 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1.364 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{2-6}$$

The values for the reaction towards xylose, glucose arabinose, and acetic acid take into account the anhydrous form in which the individual molecule is present in the biomass's polymeric structure [24]. All equations as indicated are implemented in MATLAB and solved subsequently. The differential equations are solved with the *ode15s* solver. The model validation is performed by a train & test single-split validation. Also, the coefficient of determination and the root mean square error for the model are used as described in Section 2.2.4. Further explanation regarding the validation is provided through the supplementary material.

2.2.7 Parameter Estimation & Identifiability Analysis

2.2.7.1 Parameter Estimation

The approach in this section is based on maximum-likelihood estimation described by Sin and Gernaey (2016) [36]. A detailed summary can be found in the supplementary material. The values which are determined by the maximum-likelihood estimation are the mean estimate $\hat{\theta}_i$ of the parameter i, the standard deviation σ_i and the lower and upper bound of the 95 % confidence interval of the estimate l_i and u_i.

2.2.7.2 Identifiability Analysis

Depending on the modelled system, a mechanistic model can comprise a very high number of parameters that are supposed to be estimated. Depending on the model structure and the amount of available experimental data, this can lead to high standard deviations for the estimated values. Hence, it is paramount to identify a subset of parameters $\theta_k \in \theta$ which is significant for the model output and uniquely estimable by parameter estimation from a distinct data set [37]. The described identifiability analysis methodology is based on a local sensitivity analysis known as one-factor-at-a-time (OAT) method as described by Brun et al. (2002) [38]. The analysis quantifies the significance of each parameter – expressed by the value δ^{msqr} – and describes the collinearity of each possible combination of datasets by a collinearity index γ. A detailed explanation regarding the calculation is given in the supplementary material.

2.2.8 Uncertainty and Sensitivity Analysis

2.2.8.1 Monte Carlo-Based Uncertainty Analysis

Due to several error sources introduced in the model, e.g., measurement errors in experimental data, there is a need to quantify the uncertainty in the model output [37]. Several ways of performing this assessment exist, of which a Monte Carlo method is employed here. According to Sin et al. (2009), Monte Carlo-based uncertainty analysis is practically performed in four steps: 1) definition of input uncertainty, in this case, the uncertainty from measurement errors in the parameter estimation, 2) sampling, in this study Latin Hypercube Sampling, 3) Monte Carlo simulations, and 4) result analysis. A detailed description of the procedure can be found in the supplementary material. The entire procedure is implemented in MATLAB.

2.2.8.2 Variance-Based Sensitivity Analysis

After performing an uncertainty analysis, which describes the model output uncertainty with uncertain input, the complementary analysis is a sensitivity analysis, aiming to apportion the output uncertainty on the different inputs [37]. For this, commonly, the first-order sensitivity index S_i and the total sensitivity index S_{Ti} are calculated for each variable. They describe respectively the sensitivity of the model output to the respective parameter solitarily, hence the first order, and the sensitivity of the model output to the parameter including all interactions with other parameter, hence the total sensitivity [39,40]. In particular, in this study, a global sensitivity analysis based on the work of Saltelli et al. (2010) is implemented: the numerical calculation of both sensitivity indices is performed with Sobol sequence sampling: As a first step, two sampling matrices A and B are generated by Sobol sampling. Subsequently, two mixed matrices A_B^i and B_A^i are generated, in which column i from the one matrix is replaced by the same column of the respective other matrix, and all other columns are maintained. Then, the model outputs are calculated for all four sampling matrices, and the respective sensitivity measures can be calculated. The first-order sensitivity index is calculated as:

$$S_i = V(y) - \frac{1}{2N}\sum_{j=1}^{N}\left(y_B(j) - y_{ABi}(j)\right)^2, \tag{2-7}$$

and the total sensitivity index as:

$$S_i = \frac{1}{2N}\sum_{j=1}^{N}\left(y_A(j) - y_{ABi}(j)\right)^2, \tag{2-8}$$

according to the referred methods [41]. A detailed description of the backgrounds of the method can be found in the supplementary material. The *easyGSA* toolbox in MATLAB is used for the analysis [39].

2.2.9 Optimization

2.2.9.1 Optimization Excluding Uncertainty

To maximize the monomer yield by optimizing the operative conditions of the biomass pretreatment, all presented data and knowledge-driven models can be employed. The optimization problem, which needs to be solved for this is independent of the model and can be formulated as follows:

$$\max y = f(x)$$
$$x \in X \subset \mathbb{R}^n \tag{2-9}$$

The optimization problem is defined as a constrained nonlinear program (NLP). Dependent on each model, different optimization techniques are applicable; they are described in detail for each model in the supplementary material. For the optimization with the RSM model, the *rsm* package in R is chosen [27]. For the optimization with the GPR model, the *ga* solver in MATLAB is chosen. Finally, for the optimization with the mechanistic model, the *fmincon* solver with a sequential quadratic programming algorithm and a multi-start setup with the *MultiStart* function in MATLAB is chosen.

2.2.9.2 Optimization Including Uncertainty

As models predict results with inherent uncertainties, ideally, these should be considered when optimizing. Furthermore, the assumptions taken before building a model can significantly impact the prediction of the model output. Therefore, taking these into account while optimizing the operational conditions can potentially find a more robust optimum under more robust conditions. A more detailed introduction is given in the supplementary material. For this, the *MOSKopt* solver in MATLAB is used [42].

2.3 Results

2.3.1 Feedstock Analysis

The weight percentages per total dry weight of biomass of the composition of the wheat straw were determined according to the description in Section 2.2.1 and are listed in the following Table 2:

Table 2: Composition analysis of the fiber fraction of wheat straw.

Component	This study	Vassilev et al., 2012 [43]
Cellulose	40.7 wt. % ($\sigma = 0.0003$)	44.5 wt. %
Hemicellulose	33.6 wt. % ($\sigma = 0.0025$)	33.2 wt. %
Lignin (soluble & insoluble)	24.9 wt. % ($\sigma = 0.0041$)	22.3 wt. %

Vassilev et al. (2012) reviewed nine different research papers that indicate the composition of wheat straw and present the mean value [43], which is used for a comparison here. Thus, it becomes evident that the wheat straw composition in this study is in good agreement with the literature values; only the present amount of cellulose is slightly lower than the reference value.

Regarding the composition of the hemicellulosic fraction, also several values are reported in the literature. Hence, it is assumed that an average value of 80% is attributed to xylan, 10% to arabinan, and 10% to acetyl groups for straw [43,44]. This assumption will be revisited and also taken into account in the optimization under uncertainty in subsection 2.3.8.

2.3.2 Hydrothermal Pretreatment

For the experiments, a solid-to-liquid ratio of 1: 10 (w/w) is chosen with a dry mass of feedstock of 30 g. The experiments are performed as described in Section 2.2.2, and the results are listed in Table 3.

Table 3: HPLC analysis of the autothermal pretreatment experiments.

No.	Temperature	Time	Xyl	Glu	Ara	Aac	Hmf	Fac	Fur
	°C	Min	wt%	wt%	wt%	wt%	wt%	wt%	wt%
1	165	40	1.62	0.56	0.15	0.61	0.00	0.13	0.00
2	177.5	54	1.67	0.46	0.54	1.31	0.01	0.23	0.02
3	177.5	40	1.70	0.54	0.41	0.92	0.01	0.18	0.01
4	177.5	26	1.62	0.57	0.11	0.47	0.00	0.10	0.00
5	190	40	1.69	0.39	0.67	1.33	0.01	0.29	0.03
6	160	20	1.59	0.57	0.07	0.39	0.00	0.00	0.00
7	160	60	1.66	0.58	0.86	0.75	0.00	0.15	0.00
8	177.5	40	1.54	0.49	0.32	0.85	0.01	0.17	0.01
9	177.5	40	1.65	0.53	0.39	0.92	0.00	0.18	0.01
10	195	60	1.66	0.29	0.84	2.95	0.05	0.87	0.54
11	195	20	1.55	0.54	0.14	0.54	0.00	0.12	0.00

The yield of xylose Y_{Xyl} (w/w) in the hydrothermal pretreatment lies at around $Y_{Xyl} = 5 - 6\,\%$. It becomes evident that the selected pretreatment method combined with the feedstock and the experimental condition is not a suitable combination for the given

criteria in Section 2.2.1. The most obvious reason for this is possibly the high recalcitrance of the feedstock. Due to the mild conditions of the hydrothermal pretreatment, the polymeric chains are not entirely broken down into monomers, but many oligomeric sugars are released from the biomass and are not depolymerized furtherly [45,46].

For a further analysis of the hydrolysate, a post-hydrolysis is performed as described in Section 2.2.3. The detailed results are listed in Table 4.

Table 4: Results from the HPLC analysis of the posthydrolyzed autothermal pretreatment experiments.

No.	Temperature	Time	Xyl	Glu	Ara	Aac	Hmf	Fac	Fur
	°C	Min	wt%	wt%	wt%	wt%	wt%	wt%	wt%
1	165	40	1.82	1.12	0.34	0.43	0.14	0.18	0.04
2	177.5	54	3.95	1.42	1.45	1.63	0.10	0.27	0.17
3	177.5	40	1.60	1.50	1.12	0.96	0.11	0.26	0.12
4	177.5	26	2.18	1.31	0.31	0.46	0.18	0.22	0.04
5	190	40	5.84	1.59	1.76	1.87	0.09	0.31	0.36
6	160	20	2.13	1.23	0.24	0.35	0.16	0.18	0.03
7	160	60	1.97	1.09	0.50	0.53	0.11	0.20	0.05
8	177.5	40	1.47	1.37	1.24	1.66	0.12	1.70	0.11
9	177.5	40	1.53	1.39	1.13	0.91	0.11	0.23	0.12
10	195	60	15.29	1.63	1.70	3.62	0.12	0.66	1.16
11	195	20	2.34	1.36	0.43	0.54	0.14	0.25	0.06

As a result, the only notable increase in the concentration of xylose monomers occurred for the experiment with the highest temperature and the longest reaction time; the yield is $Y_{Xyl} = 65.7$ %. Other studies also report a maximum hemicellulosic sugar recovery for post-hydrolyzed fractions of hydrothermally pretreated wheat straw of around 60 % to 70 % [45,46]. Considering that such reaction conditions are

unfavorable from an economic perspective, the additional post-hydrolysis step does not significantly improve the feasibility of the hydrothermal pretreatment in this setup.

2.3.3 Dilute Acid Pretreatment

For the experiments, a solid-to-liquid ratio of 1: 10 (w/w) is chosen with a dry mass of feedstock of 30 g. The chosen acid is sulfuric acid and the acid concentration C_{ac} concerning the solid mass is chosen to be in the interval $C_{ac} = [0.8,1.3]$ %. The experiments are performed as described in Section 2.2.2, and the results are listed in **Table 5**.

For the sugar monomer yield, it becomes apparent that the dilute acid pretreatment releases a significantly higher amount of xylose monomers compared to the results of the hydrothermal pretreatment in Section 2.2.3. Especially points in the octants with a higher temperature than the center point, longer reaction times than the central point, and higher and lower acid concentrations than the central point – points 3, 4, 8, 9, 10, 13, 14, and 15 – show very high monomer yields. Taking into account the stoichiometric factor, an average hydrolysate volume of $V = 250\,mL$, as well as the amount of xylan in hemicellulose, the highest yield is obtained for point ten with $Y_{Xyl} =$ 98 %. Similar high yields have been reported for a combined dilute-acid and steam explosion pretreatment of wheat straw, for the pretreatment of wheat straw with a subsequent enzymatic hydrolysis step, or for lower absolute monomer concentrations. [47–49]. The degradation of the cellulosic fraction is comparatively small. As the furfural production occurs by the degradation of xylose, the amount of furfural for these conditions is low. On the other hand, the amount of acetic acid for these conditions is high, indicating an equally high yield for this reaction. In conclusion, the dilute acid pretreatment for combining the given feedstock, pretreatment method, and experimental conditions seems advantageous and a good option for the study.

Table 5: Results from the HPLC analysis of the dilute acid pretreatment experiments.

No.	T ℃	Time Min	Acid wt. %	Xyl wt. %	Glu wt. %	Ara wt. %	Aac wt. %	Hmf wt. %	Fac wt. %	Fur wt. %	Xyl yield %
1	160	10	0.8	1.47	0.58	1.18	0.38	0.00	0.00	0.00	6.21
2	160	26	1.7	17.84	1.80	2.25	2.82	0.17	0.00	0.28	75.38
3	173	18	1.25	20.28	1.63	2.66	2.85	0.12	0.00	0.09	85.69
4	186	26	1.7	19.48	0.40	2.63	4.11	0.22	0.31	1.99	82.35
5	186	10	0.8	4.61	0.75	2.15	0.88	0.03	0.00	0.01	19.48
6	160	10	1.7	3.15	0.51	1.50	0.81	0.01	0.00	0.00	13.31
7	160	26	0.8	12.79	1.06	2.37	1.65	0.07	0.00	0.03	54.07
8	173	18	1.25	21.04	1.84	2.64	2.95	0.15	0.00	0.13	88.92
9	186	26	0.8	23.27	2.48	2.72	3.57	0.28	0.24	0.56	98.34
10	195	18	1.25	23.29	2.27	2.72	3.63	0.23	0.19	0.34	98.42
11	151	18	1.25	1.44	0.61	1.68	1.49	0.01	0.00	0.00	6.07
12	173	4.55	1.25	1.44	0.56	0.96	0.31	0.00	0.00	0.00	6.10
13	173	18	1.25	20.45	1.72	2.55	2.95	0.14	0.22	0.12	86.43
14	173	31.45	1.25	21.30	2.56	2.62	3.71	0.26	0.28	1.01	90.02
15	173	18	2.0	21.03	2.12	2.56	4.62	0.20	0.23	0.28	88.89
16	173	18	0.5	13.80	1.24	1.70	2.04	0.10	0.13	0.10	58.32
17	186	10	1.7	1.66	0.67	1.57	0.38	0.01	0.00	0.00	7.01

2.3.4 Model Calibration of Mechanistic Model

All model calibration and validation will only be performed based on the dilute acid pretreatment. Furthermore, published studies indicate that the prediction of xylose with a rate constant, as shown in equation (2-4), might impair predictions over the large design space considered in this study [32,50]. Preliminary analyses with the presented experimental data confirm this. Hence, the mechanistic model will only be fitted to a spherical sector of the design space described earlier in Section 2.3.3, as the resulting xylose concentrations were the highest over the whole design space. The reasoning behind this is the potential application in a biorefinery context, where a maximum amount of monomers, as one of the defined criteria, is required. Hence, the data points for estimation are points 3, 8, 10, 13, 14, 15, and 16, as listed in **Table 5**. Data points 4 and 9 are used for the model validation. The design space including these points is illustrated in **Figure 2**.

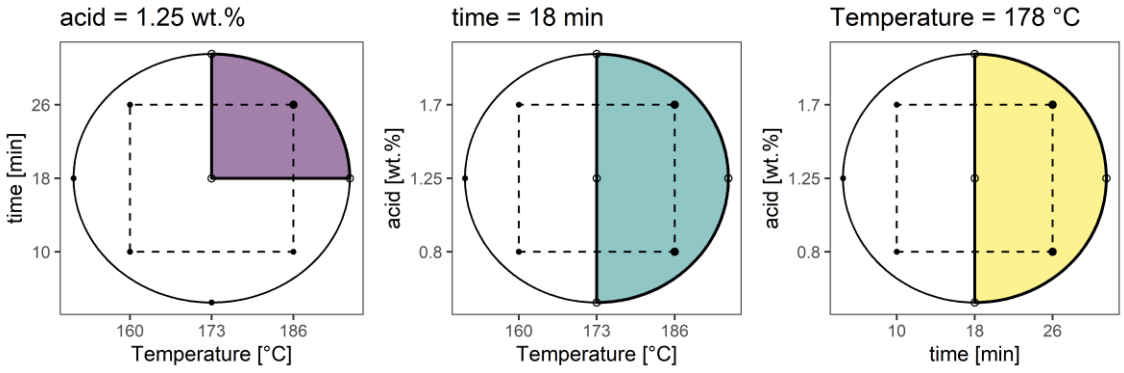


Figure 2: Design space of the experiments for all 17 operational conditions(• not estimated, ○ training data set, ● testing data set).

The scripts for the identifiability analysis and the parameter estimation are provided through a GitHub repository [29].

2.3.4.1 Identifiability Analysis

The Identifiability analysis for the kinetic model is performed as described in Section 2.2.7.2. The values which are determined by the maximum-likelihood estimation are the mean estimate $\hat{\theta}_i$ of the parameter i, the standard deviation σ_i and the upper and lower bound of the 95 % confidence interval of the estimate l_i and u_i.

The used values for the parameters are obtained by a preliminary run of the parameter estimation as described in Section 2.2.7.1 with arbitrary initial values: all frequency factors were set to $A_i = 1 \cdot 10^{10}\ s^{-1}$, all activation energies were set to $E_i = 100\ kJ \cdot mol^{-1}$ and all reaction order exponents were set to $n_i = 1$. The values for the initial parameter values are provided in the supplementary material. The resulting values for δ_{msqr} are illustrated in **Figure 3**.

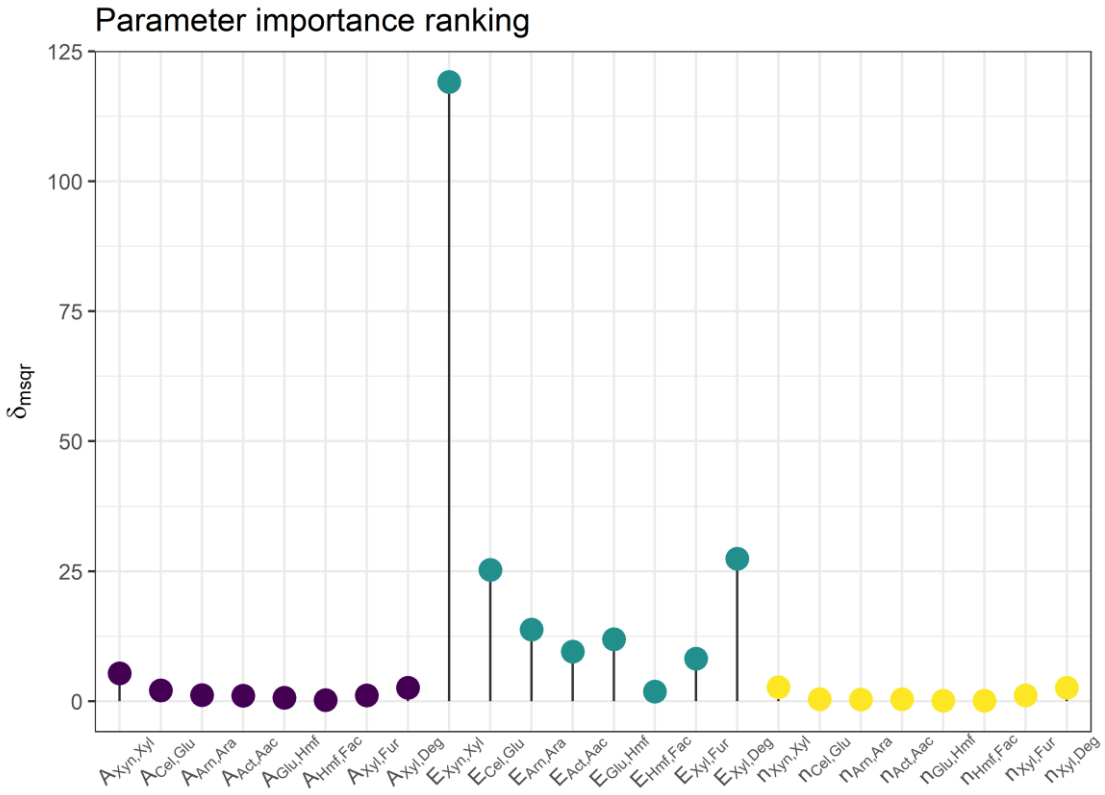


Figure 3: Values for the parameter significance calculated as δ_{msqr} for all parameter values.

The figure clearly illustrates that the most significant parameters are all activation energies, whereas both the frequency factors and the reaction order exponents are of minor significance. This is in agreement with similar investigated results, e.g., in Prunescu et al. (2015) [34]. In consequence, the activation energies are selected as the parameters to be estimated.

2.3.4.2 Parameter Estimation

The mechanistic model's parameter estimation is performed as described in Section 2.2.7.1 for the eight parameters selected in Section 2.3.4.1, namely the activation energies. The estimated parameters' initial values are the results from the initial estimation, as performed in Section 2.2.7.1. They are listed in the supplementary material. With these set values, the estimated values for the activation energies, their standard deviations, and upper and lower bounds for the confidence intervals result are given in **Table 6**:

Table 6: Values of the estimated parameters of the pretreatment model.

Parameter	$\widehat{\boldsymbol{\theta}}_i$	$\boldsymbol{\sigma}_i$	$\boldsymbol{l}_i$	$\boldsymbol{u}_i$
$E_{Xyn,Xyl}$	121.29	0.1048	121.08	121.5
$E_{Cel,Glu}$	39.246	0.2618	38.718	39.775
$E_{Arn,Ara}$	56.917	0.3761	56.158	57.677
$E_{Act,Aac}$	38.66	0.2469	38.161	39.159
$E_{Glu,HMF}$	66.425	1.2743	63.851	68.998
$E_{HMF,Fac}$	38.036	2.5472	32.891	43.18
$E_{Xyl,Fur}$	25.907	0.7316	24.429	27.384
$E_{Xyl,Deg}$	37.374	0.2053	36.959	37.789

Also, the following correlation matrix results from the parameter estimation:

$$CORR = \begin{pmatrix} 1 & 0 & 0.02 & 0 & 0 & 0 & 0.01 & 0.56 \\ 0 & 1 & 0 & 0 & 0.69 & 0.01 & 0 & 0 \\ 0.02 & 0 & 1 & 0 & 0 & 0 & 0.57 & 0.28 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.69 & 0 & 0 & 1 & 0.01 & 0 & 0 \\ 0 & 0.01 & 0 & 0 & 0.01 & 1 & 0 & 0 \\ 0.01 & 0 & 0.57 & 0 & 0 & 0 & 1 & 0.51 \\ 0.56 & 0 & 0.28 & 0 & 0 & 0 & 0.51 & 1 \end{pmatrix}$$

The resulting standard deviations of the parameter estimation are low and thus indicate a very good fit. This is equally reflected in the correlation matrix, where only a few estimated parameters have a significant but not critical covariance. In comparison with other reported values, the found values for the activation energies in this study are in good agreement with findings in other studies [34,51,52]. Overall, it can be concluded that the parameter estimation in combination with the identifiability analysis leads to an excellent estimation result that can be further investigated.

2.3.5 Model Validation

With the given experimental data points for the dilute acid pretreatment, the RSM, the GPR, and the mechanistic model are fitted. For the GPR model. the used basis function β is constant, and the used kernel function ω is an exponential function. For the mechanistic model, the split between training and testing data set is described in Section 2.3.4. The scripts for all model validations are provided through a GitHub repository [29].

The metrics for each model are calculated as indicated in Section 2.2.4. This comprises the analysis of variance (ANOVA) for the RSM model, a k-fold cross-validation for the GPR with $k = 5$ and a single-split

validation with training and testing data set for the mechanistic model. The validation metrics for all models for training and testing data sets (if applicable) are indicated in Table 7:

Table 7: Validation metrics of the RSM, the GPR, and the mechanistic model.

	RSM	GPR	Mechanistic Model
R^2_{train}	0.914	0.999	0.988
R^2_{test}	–	0.141	0.305
$RMSE_{train}$	2.537	0.969	0.842
$RMSE_{test}$	–	1.277	0.682

The ANOVA returns $F = 6.403$ on 10 and 6 degrees of freedom, in connection with $p = 0.0169 < 0.05$. Hence, the null hypothesis is rejected, and the model is considered validated. In conclusion, the RSM model combined with a CCD appears to be a viable combination to build a data-driven model from experimental data in this study's scope. The GPR training data set metrics indicate an excellent fit, whereas the testing data set $RMSE$ increases significantly, indicating a possibly inferior predictive capacity. For the mechanistic model, both metrics for both data sets show a very good fit with even lower $RMSE$ for the testing set than the GPR; hence, the model also validates well.

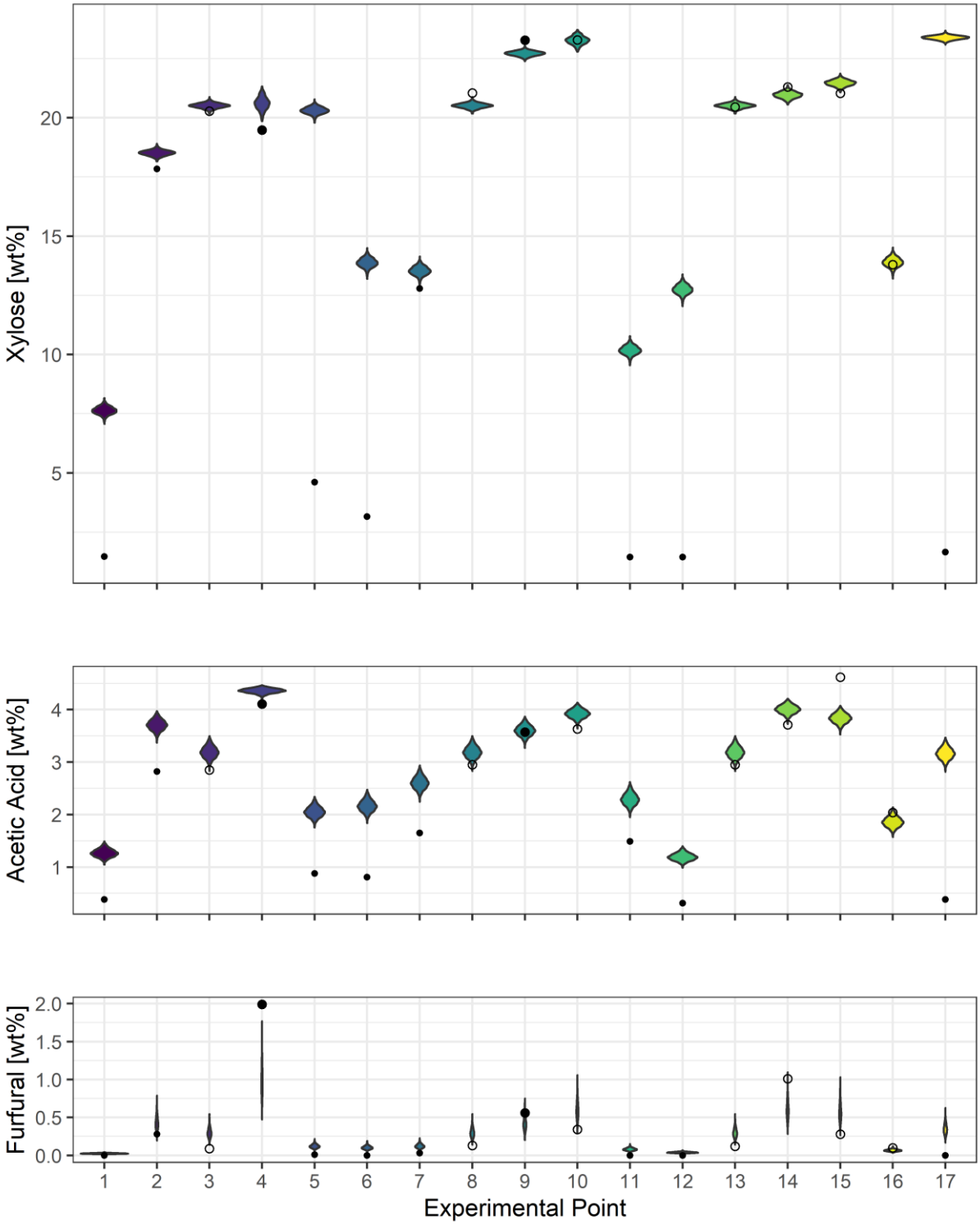
2.3.6 Uncertainty and Sensitivity Analysis

2.3.6.1 Uncertainty Analysis

The uncertainty analyses for all model outputs of the mechanistic model were performed as described in Section 2.2.8.1. The considered uncertainty here is the one deriving from the estimation of the model parameters. The errors are assumed to be normally distributed according to their standard deviation, as measurement errors from the experimental data as error source are assumed to have random character. With the correlation matrix and the standard deviation of the parameters, a multivariate normal random distribution is generated for $N = 1000$ Monte Carlo samples. The Monte Carlo procedure is performed N times for all 17 operational conditions. The scripts for the uncertainty analysis are provided through a GitHub repository [29]. The results for the uncertainty in the predictions of xylose, acetic acid, and furfural are illustrated in the following violin plot in **Figure 4**. The violin plots for the other components are provided with the supplementary material.

For the uncertainty in the xylose prediction, it becomes apparent that the uncertainty in the prediction is generally very low and that the model predicts closely the experimental results from both the training and the testing data set. For data points that were excluded from the estimation, however, the model overpredicts these values significantly. As already mentioned in the introduction of Section 2.5.5, this is expected and reflects the fact that based on this expected behavior, only a subset of all available data points was chosen for the estimation as described in subsection 2.3.4.2.

Figure 4: Violin plots for the results of the Monte Carlo-based uncertainty analysis for N=1000 samples for all 17 operational conditions (• not estimated, ○ training data set, ● testing data set) for the concentrations of xylose, acetic acid, and furfural.



For the uncertainty in the xylose prediction, it becomes apparent that the uncertainty in the prediction is generally very low and that the model predicts closely the experimental results from both the training and the testing data set. For data points that were excluded from the estimation, however, the model overpredicts these values significantly. As already mentioned in the introduction of Section 2.5.5, this is expected and reflects the fact that based on this expected behavior, only a subset of all available data points was chosen for the estimation as described in subsection 3.4.2.

Similar behavior for the uncertainty in the acetic acid prediction is displayed, however at a smaller scale as the concentrations are generally lower: the data points which were not used for estimation are slightly overpredicted, and all but one data point in the training and testing data set are predicted within the range of the model uncertainty. For point 15, the model underpredicts the acetic acid concentration. The used acid concentration for this data point is $C_{ac} = 2.0\ wt\%$, which possibly explains a higher release of acetyls from the polymeric structure than predicted by the model as the reaction might be more than proportionally dependent on the acid concentration.

The uncertainty in the furfural prediction also happens at a generally low level; in contrast to the xylose and acetic acid predictions, the model does mostly not overpredict the values for data points, even for points that have not been estimated. Solely for point 15, the model underpredicts the furfural concentration. The operational conditions for this data point with $T = 186$ °C, $t = 26\ min$ and $C_{ac} = 2.0\ wt\%$ are all close to the upper bound of their range. Hence, these conditions are rather severe for the pretreatment, potentially leading to increased xylose degradation and a high furfural concentration. In conclusion, the mechanistic model shows low uncertainty in predicting all considered component concentrations.

2.3.6.2 Sensitivity Analysis

Similar to the uncertainty analyses, the sensitivity analyses were performed for all mechanistic model outputs for all three operational parameters. The considered ranges for the variables are $T = [173,195]$ °C, $t = [18,30]\ min$ and $C_{ac} = [0.5,2.0]\ wt$ %. All sensitivity analyses were performed with $N = 8192$ Sobol samples. The scripts for the sensitivity analysis are provided through a GitHub repository [29]. The resulting first-order and total sensitivity indices for all three operational variables for all output concentrations are presented in the heatmap in Figure 5.

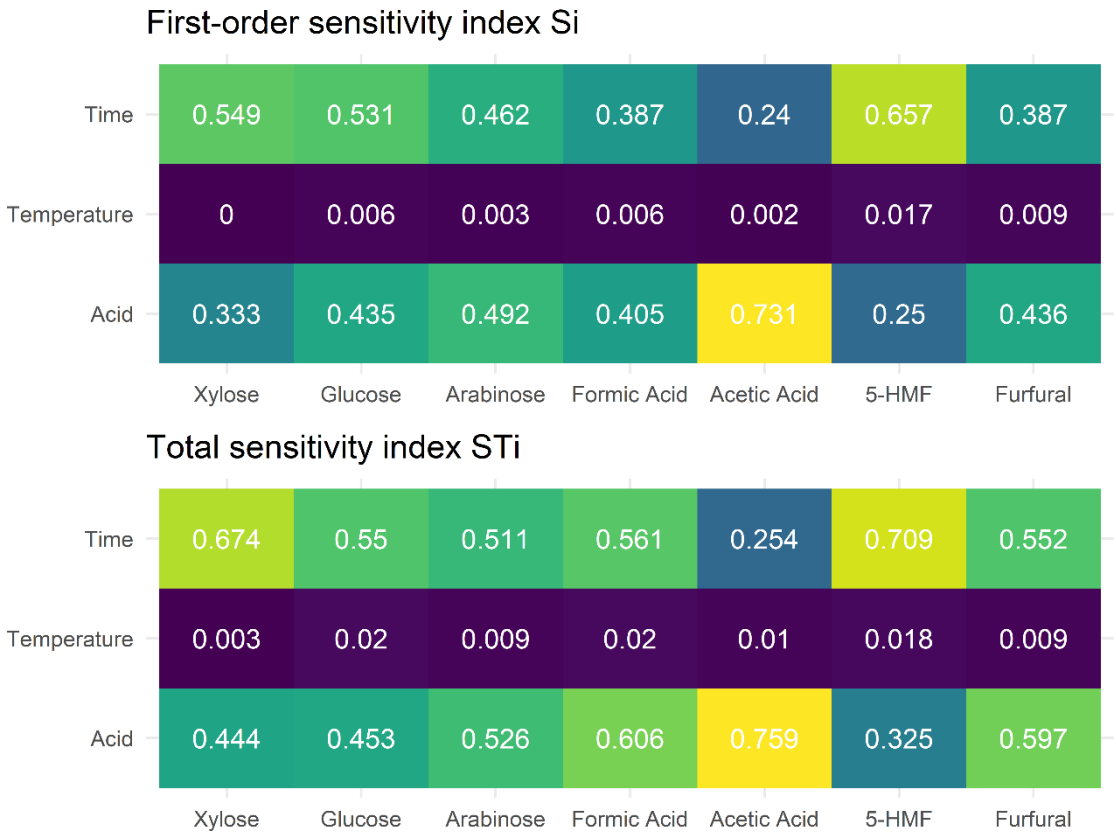


Figure 5: Heat map for the values of the sensitivity analysis of all operational parameters for all model outputs of the mechanistic model.

Reviewing the results for the first-order sensitivity indices in detail, a tendency for moderate sensitivities in the time and acid concentration becomes apparent. Equally, a generally low sensitivity for the temperature is seen. Reciprocally, the results for the total sensitivity indices illustrate that only the influence of the temperature in the given

interval would be negligible. Referring back to the uncertainty analysis results in subsection 2.5.6.1, the conclusions regarding the underprediction of acetic acid for high acid concentrations agree with a high first-order sensitivity index for the acid concentration on the output of acetic acid. Similarly, the furfural output is susceptible to the acid concentration due to a higher release of xylose monomers; hence the higher the acid concentration, the higher the furfural output. Combined with a high temperature and time, this can yield very high furfural outputs, whereas a maximal acid concentration with moderate temperature and time yields moderate furfural outputs. The importance of all three parameters in different absolute orders is also seen in other studies [32,50].

2.3.7 Optimization of Operational Conditions

As the last part of the presented study, the operational conditions of the dilute acid pretreatment are optimized with all developed models. The scripts for all optimization setups are provided through a GitHub repository [29]. The considered objective is the output concentration of xylose, which is to be maximized. The objective function, as stated in (2-9), is the respective model evaluation function. The considered bounds for the input variables as stated in (2-9) are $T = [173,195]$ °C, $t = [18,30]\ min$ and $C_{ac} = [0.5,2.0]\ wt$. However, a further constraint is included that the optimal point can only lie within the design space of the CCD for which the design of experiments is valid. This is expressed with the following equation:

$$\left|\frac{T \cdot °C^{-1} - 173}{13}\right| + \left|\frac{t \cdot min^{-1} - 18}{8}\right| + \left|\frac{C_{ac} - 1.25}{0.45}\right| \leq \alpha, \quad (3\text{-}1)$$

With the value of $\alpha = 1.69$ for the CCD of the dilute acid pretreatment experiments.

2.3.7.1 Results

The optimized variables, as well as the value for the objective function, are listed in Table 8:

Table 8: Results for the optimized operational variables and the value for the objective function of all models.

Variable	Unit	RSM	GPR	Mechanistic Model
T	°C	182.4	195	191.6
t	min	26.2	18	18
C_{ac}	$wt\%$	1.25	1.25	1.13
$\boldsymbol{C_{Xyl}}$	$\boldsymbol{wt\%}$	**25.72**	**23.23**	**23.47**

For the RSM, the point does not lie in the vicinity of already measured data points but rather within the design space. A single experiment for these conditions is performed to verify these conditions, as described in Section 2.2.2. The resulting concentration of xylose is $C_{Xyl} = 20.97\ wt\%$ and hence a deviation from the predicted value by 18.5 %. The RSM model hence overpredicted the maximum concentration significantly.

For the GPR, although the predicted concentration is notably lower than the one predicted with the RSM model, the experimental value for this point is $C_{Xyl} = 23.29\ wt\%$, which is an excellent prediction. It is to mention that the point is part of the training data set of the GPR, namely point 10.

The predicted xylose concentration is even higher for the mechanistic model than the one predicted by the GPR model, while the difference within both the selected conditions and the output concentrations is marginal. This point is not verified experimentally due to the close vicinity of this point to point 10 of the experimental design and the minimal difference in the output concentrations within the uncertainty range of the experimental values. Nonetheless, given the sensitivity analysis results in subsection 2.3.6.2, a slightly reduced acid concentration compared to point 10 is conclusive as the most sensitive

operational condition for the xylose concentration is the acid concentration.

2.3.7.2 Comparison

After calibrating all models to experimental data and calibrating them respectively, conclusions regarding their comparability and performance shall be drawn in this section. Figure 6 shows contour plots for all three introduced models, with the output being the xylose concentration for respectively two of the three operational conditions, while the third operational condition is fixed at its mean value.

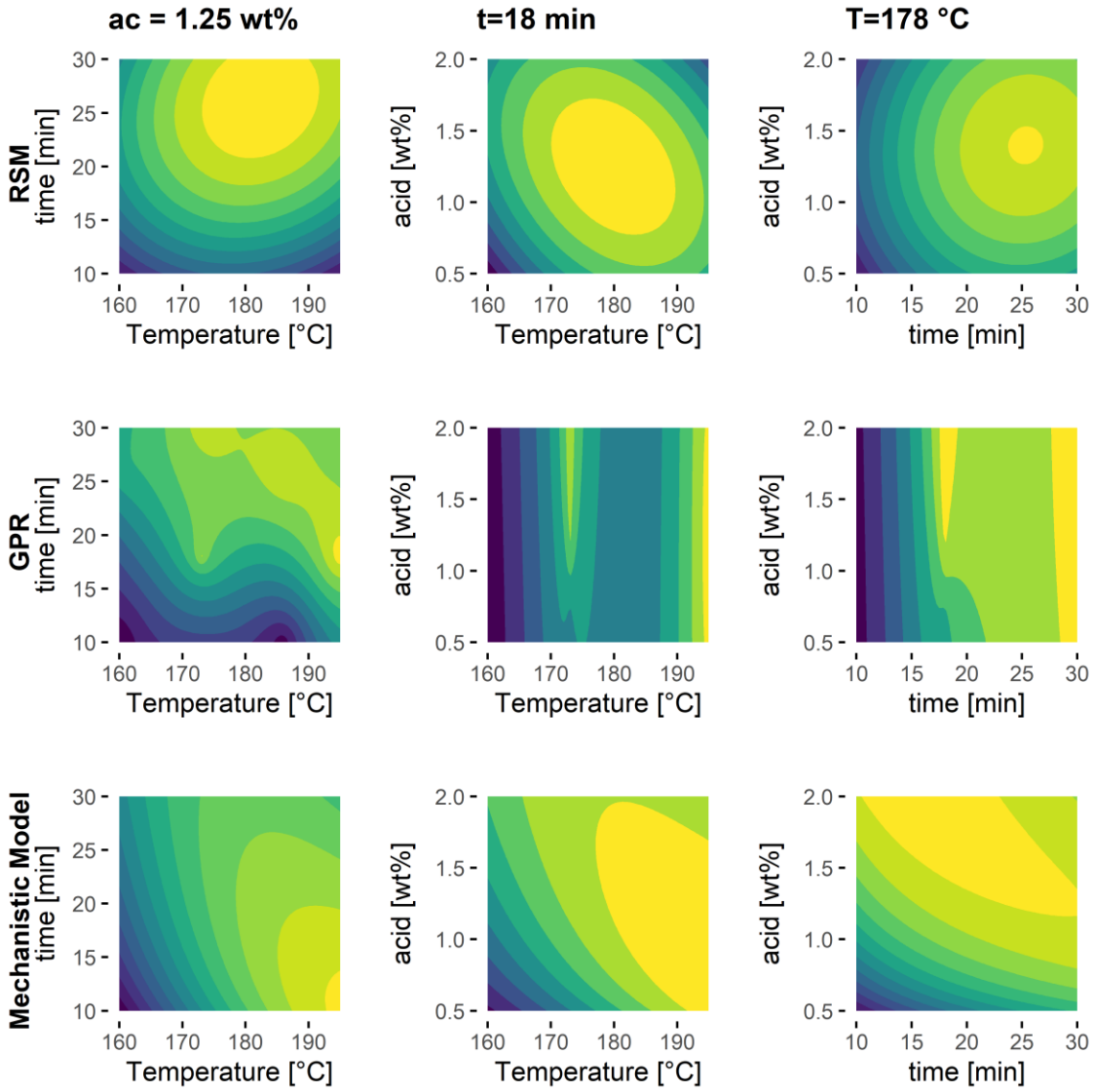


Figure 6. Contour plots of the xylose concentration for two out of three operational conditions, with the third fixed at its mean value.

For the RSM model, the contour of the xylose output is characterized by the underlying second-order polynomial, illustrating the global optimum for a maximum xylose output as a single point within the design space. For the GPR model, the contour plot illustrates the general behavior, which accounts for GPR models: The prediction in close vicinity of known points is excellent, whereas the uncertainty in prediction significantly increases for areas that do not lie in the vicinity to known data points. This also relates to the fact that the validation metrics – particularly the RMSE – for the testing data set were the highest of all models. This leads to the fact that several local maxima can be seen in the contours for the training data points. Lastly, the contour plots of the mechanistic model resemble the kinetic behavior of the first-order rate expressions, which predicts the global optimum towards higher temperatures, average times, and lower acid concentrations. Lower xylose concentrations are, however, overpredicted.

As a concluding remark, it should be stated that all models involve a certain level of heuristics and approximations since the exact mechanisms of all occurring reactions in biomass pretreatment with the inherently high number of components is infeasible to describe. Hence, using a particular model always indicates a trade-off between different characteristics, and there is no best one-fits-all candidate modeling approach. To this end, it is noted that data sets resulting from typical designs of experiments usually are analyzed by RSM, where our study highlights shortcomings of this methodology [26,53]. In point of fact, to determine optimal conditions, predicting unseen data with a small error is of vital importance. Therefore, a diligent approach is the benchmark of different modeling approaches against each other regarding the prediction of unseen or test data sets. Since the mechanistic model is validated best in this regard, it is used for the optimization under uncertainty and verifies the assumptions taken regarding the feedstock.

2.3.8 Optimization of Operational Conditions under Uncertainty

The purpose of the optimization step is to use the model to rigorously explore the design space for any conditions that can improve process yield. Therefore, the model is optimized under uncertainty to assess the effects of both the model uncertainties and the assumption regarding the feedstock. The uncertainty in the parameters is expressed through the covariance matrix and used for sampling. For the assumption in the feedstock composition, it is assumed that the original amount of xylan in hemicellulose of 80% is uniformly distributed in an interval of [70,80] %, and the arabinan and acetyl amounts are uniformly distributed in an interval of [10,15] % as a worst-case scenario. The solver is performing $N = 250$ Monte Carlo simulations for each solver iteration and initially performs 25 simulations and subsequently performs 75 iterations. The chosen infill criterion is *mcFEI,* and the corresponding infill solver is a particle swarm solver. The scripts for the optimization under uncertainty are provided through a GitHub repository [29].

Additionally, two further constraints for the concentration of acetic acid and furfural are introduced in order to assure that the predicted optimum does not yield higher concentrations of both components compared to the optimum, which was found without considering the uncertainty in the feedstock:

$$C_{Aac} \leq 3.705\ wt\%, \tag{3-2}$$

$$C_{Fur} \leq 0.5289\ wt\%. \tag{3-3}$$

When hedging against the uncertainty with the mean value of the predictions of the objective function and constraints, optimizing the conditions predicts a concentration of xylose of $C_{Xyl} = 22.68\ wt\%$ at a temperature of $T = 191.2$°C, $t = 18\ min$ and $C_{acid} = 1.12\ wt\%$. The resulting objective is only marginally smaller than the one predicted without considering uncertainty while maintaining the same operational conditions. Hence, it can be concluded that even with

varying feedstock composition, the operational conditions remain optimal, and on average, the output yield is reduced by about 3.3 %.

When hedging against the uncertainty with the mean value plus one standard deviation of the predictions of objective function and constraints, optimizing the conditions predicts a concentration of xylose of $C_{Xyl} = 20.88\, wt\%$ at a temperature of $T = 182.6$°C, $t = 18\, min$ and $C_{acid} = 0.84\, wt\%$. The yield for the output concentration is now reduced by around 10.8 % compared to the deterministic optimum. The chosen operational conditions vary significantly: the selected temperature is reduced, and the dilute acid concentration. This is because the formation of particularly acetic acid and partially furfural are influenced by the dilute acid concentration and the temperature in the case of acetic acid, as seen in the first-order sensitivity indices presented in Figure 5. In order to not infringe the constraints with the hedge against uncertainty, milder conditions are selected.

In conclusion, the optimization under uncertainty showed both that the assumption taken about the feedstock composition and the intrinsic uncertainty of the model due to the parameter estimation do not significantly influence the prediction of the objective to produce a maximum amount of xylose. A study with corn stover as feedstock confirms that feedstock variability – apart from trivial effects – does not significantly influence the monomer yields [54]. Moreover, the optimization under uncertainty with a variability of the feedstock composition only moderately influences the xylose concentration, as the chosen operational conditions are milder.

2.4 Conclusions

In the scope of this paper, a model for biomass pretreatment was developed based on a case study and validated for optimizing the operational conditions of the pretreatment to assist the conceptual process design of a biorefinery. Wheat straw was chosen as feedstock, and pretreatment experiments with both hydrothermal and dilute acid pretreatment were performed. The analysis of the experiments showed the clear favorability of the dilute acid pretreatment. A response surface model, a Gaussian process regression model, and a mechanistic model based on mass and energy balances and first-order reaction kinetics were fitted to the data. All models show considerably good validation metrics. However, the optimal conditions found with the three models differ, which is relatable to each model's properties. A comparative analysis shows that the predictions of the mechanistic model are most reliable for the underlying case study, which is thus recommended for use in conceptual process design. Furthermore, optimizing the operational conditions under different uncertainty scenarios shows that the model assumptions do not affect the predicted performance and consequently confirm the robustness of the model for use in process design applications under uncertainty.

Furthermore, more fundamental knowledge about the reaction mechanisms and the factors influencing them will significantly improve the predictive quality of knowledge-driven pretreatment models and broaden their validation range. Moreover, this allows for evaluating different pretreatment methods and feedstocks with the same knowledge-driven model, facilitating the conceptual process design even further and helping to assess viable biorefinery concepts for more sustainable production of chemicals.

2.5 Acknowledgements

The authors would like to acknowledge funding by the Novo Nordisk Foundation under grant no. NNF17SA0031362 and NNF20SA0066233. Furthermore, they would like to acknowledge Torbjørn Ølshøj Jensen for providing the wheat straw.

2.6 References

1. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. Available at: https://sdgs.un.org/2030agenda. (Accessed: 3rd February 2022)
2. Cherubini, F. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Convers. Manag.* **51**, 1412–1421 (2010).
3. Straathof, A. J. J. *et al.* Grand Research Challenges for Sustainable Industrial Biotechnology. *Trends Biotechnol.* **37**, 1042–1050 (2019).
4. Chaturvedi, T., Torres, A. I., Stephanopoulos, G., Thomsen, M. H. & Schmidt, J. E. Developing process designs for biorefineries-definitions, categories, and unit operations. *Energies* **13**, 1493 (2020).
5. van der Wielen, L. A. M., Mussatto, S. I. & van Breugel, J. Bioprocess intensification: Cases that (don't) work. *N. Biotechnol.* **61**, 108–115 (2021).
6. Noorman, H. J. & Heijnen, J. J. Biochemical engineering's grand adventure. *Chem. Eng. Sci.* **170**, 677–693 (2017).
7. Hassan, S. S., Williams, G. A. & Jaiswal, A. K. Lignocellulosic Biorefineries in Europe: Current State and Prospects. *Trends in Biotechnology* **37**, 231–234 (2019).
8. Hassan, S. S., Williams, G. A. & Jaiswal, A. K. Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities. *Renewable and Sustainable Energy Reviews* **101**, 590–599 (2019).
9. Koutinas, M., Kiparissides, A., Pistikopoulos, E. N. & Mantalaris, A. Bioprocess systems engineering: Transferring traditional process engineering principles to industrial biotechnology. *Comput. Struct. Biotechnol. J.* **3**, e201210022 (2012).
10. Moncada B, J., Aristizábal M, V. & Cardona A, C. A. Design strategies for sustainable biorefineries. *Biochem. Eng. J.* **116**, 122–134 (2016).

11. Ulonska, K., König, A., Klatt, M., Mitsos, A. & Viell, J. Optimization of Multiproduct Biorefinery Processes under Consideration of Biomass Supply Chain Management and Market Developments. *Ind. Eng. Chem. Res.* **57**, 6980–6991 (2018).

12. Darkwah, K., Knutson, B. L. & Seay, J. R. A Perspective on Challenges and Prospects for Applying Process Systems Engineering Tools to Fermentation-Based Biorefineries. *ACS Sustainable Chemistry and Engineering* **6**, 2829–2844 (2018).

13. Galbe, M. & Wallberg, O. Pretreatment for biorefineries: a review of common methods for efficient utilisation of lignocellulosic materials. *Biotechnol. Biofuels 2019 121* **12**, 1–26 (2019).

14. Mussatto, S. I. & Dragone, G. M. *Biomass Pretreatment, Biorefineries, and Potential Products for a Bioeconomy Development*. *Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery* (Elsevier Inc., 2016). doi:10.1016/B978-0-12-802323-5.00001-3

15. Saini, R. *et al.* Lignocellulosic Biomass-Based Biorefinery: an Insight into Commercialization and Economic Standout. *Curr. Sustain. Energy Reports* **7**, 122–136 (2020).

16. Saini, J. K. *et al.* Integrated Lignocellulosic Biorefinery for Sustainable Bio-Based Economy. in *Sustainable Approaches for biofuels Production Technologies* 25–46 (Springer, Cham, 2019). doi:10.1007/978-3-319-94797-6_2

17. Choi, S., Song, C. W., Shin, J. H. & Lee, S. Y. Biorefineries for the production of top building block chemicals and their derivatives. *Metabolic Engineering* **28**, 223–239 (2015).

18. Albuquerque, T. L. De, Da Silva, I. J., De MacEdo, G. R. & Rocha, M. V. P. Biotechnological production of xylitol from lignocellulosic wastes: A review. *Process Biochem.* **49**, 1779–1789 (2014).

19. Venkateswar Rao, L., Goli, J. K., Gentela, J. & Koti, S. Bioconversion of lignocellulosic biomass to xylitol: An overview. *Bioresource Technology* **213**, 299–310 (2016).

20. Dasgupta, D., Bandhu, S., Adhikari, D. K. & Ghosh, D. Challenges and prospects of xylitol production with whole cell bio-catalysis: A review. *Microbiol. Res.* **197**, 9–21 (2017).

21. Felipe Hernández-Pérez, A. *et al.* Xylitol bioproduction: state-of-the-art, industrial paradigm shift, and opportunities for integrated biorefineries. *Crit. Rev. Biotechnol.* **39**, 924–943 (2019).

22. Sluiter, A. *et al. Determination of Structural Carbohydrates and Lignin in Biomass. Laboratory Analytical Procedure (LAP)*. *Biomass Analysis Technology Team Laboratory Analytical Procedure* **2011**, (2004).

23. Sluiter, A., Ruiz, R., Scarlata, C., Sluiter, J. & Templeton, D. Determination of Extractives in Biomass: Laboratory Analytical Procedure (LAP); Issue Date 7/17/2005. (2008).

24. Sluiter, A. *et al. Determination of Sugars , Byproducts , and Degradation Products in Liquid Fraction Process Samples Laboratory Analytical Procedure (LAP) Issue Date : 12 / 08 / 2006 Determination of Sugars , Byproducts , and Degradation Products in Liquid Fraction Proce*. (2008).

25. Box, G. E. P. & Wilson, K. B. On the Experimental Attainment of Optimum Conditions. *J. R. Stat. Soc. Ser. B* **13**, 1–38 (1951).

26. Bas, D. & Boyaci, I. H. Modeling and optimization i: Usability of response surface methodology. *J. Food Eng.* **78**, 836–845 (2007).

27. Lenth, R. V. Response-surface methods in R, using RSM. *J. Stat. Softw.* **32**, 1–17 (2009).

28. Council, N. R. *Assessing the reliability of complex models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification*. *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification* (National Academies Press, 2012). doi:10.17226/13395

29. Vollmer, N. Pretreatment Model. *GitHub Repository* (2021).

30. McBride, K. & Sundmacher, K. Overview of Surrogate Modeling in Chemical Process Engineering. *Chemie-Ingenieur-Technik* **91**, 228–239 (2019).

31. Rasmussen, C. E. Gaussian Processes in machine learning. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*

Bioinformatics) (eds. Bousquet, O., von Luxburg, U. & Rätsch, G.) 63–71 (Springer Verlag, 2004). doi:10.1007/978-3-540-28650-9_4

32. Shen, J. & Wyman, C. E. A novel mechanism and kinetic model to explain enhanced xylose yields from dilute sulfuric acid compared to hydrothermal pretreatment of corn stover. *Bioresour. Technol.* **102**, 9111–9120 (2011).

33. Jacobsen, S. E. & Wyman, C. E. Cellulose and Hemicellulose Hydrolysis Models for Application to Current and Novel Pretreatment Processes. in *Twenty-First Symposium on Biotechnology for Fuels and Chemicals* 81–96 (Humana Press, 2000). doi:10.1007/978-1-4612-1392-5_6

34. Prunescu, R. M., Blanke, M., Jakobsen, J. G. & Sin, G. Dynamic modeling and validation of a biomass hydrothermal pretreatment process-a demonstration scale study. *AIChE J.* **61**, 4235–4250 (2015).

35. Mosier, N. *et al.* Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour. Technol.* **96**, 673–686 (2005).

36. Gürkan, S. & Krist, G. *Data Handling and parameter estimation*. *Experimental methods in Wastewater Treatment* **9781780404**, (IWA Publishing, 2016).

37. Sin, G., Gernaey, K. V. & Lantz, A. E. Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis. *Biotechnol. Prog.* **25**, 1043–1053 (2009).

38. Brun, R., Kühni, M., Siegrist, H., Gujer, W. & Reichert, P. Practical identifiability of ASM2d parameters - Systematic selection and tuning of parameter subsets. *Water Res.* **36**, 4113–4127 (2002).

39. Al, R., Behera, C. R., Zubov, A., Gernaey, K. V. & Sin, G. Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model. *Comput. Chem. Eng.* **127**, 233–246 (2019).

40. Saltelli, A. *et al. Global Sensitivity Analysis. The Primer*. *Global Sensitivity Analysis. The Primer* (John Wiley & Sons, Ltd, 2008). doi:10.1002/9780470725184

41. Saltelli, A. *et al.* Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* **181**, 259–270 (2010).

42. Al, R., Behera, C. R., Gernaey, K. V. & Sin, G. Stochastic simulation-based superstructure optimization framework for process synthesis and design under uncertainty. *Comput. Chem. Eng.* **143**, 107118 (2020).

43. Vassilev, S. V., Baxter, D., Andersen, L. K., Vassileva, C. G. & Morgan, T. J. An overview of the organic and inorganic phase composition of biomass. *Fuel* **94**, 1–33 (2012).

44. Wyman, C. *Handbook on Bioethanol*. *Handbook on Bioethanol* (Routledge, 2018). doi:10.1201/9780203752456

45. Pérez, J. A., González, A., Oliva, J. M., Ballesteros, I. & Manzanares, P. Effect of process variables on liquid hot water pretreatment of wheat straw for bioconversion to fuel-ethanol in a batch reactor. *J. Chem. Technol. Biotechnol.* **82**, 929–938 (2007).

46. Pérez, J. A. *et al.* Optimizing Liquid Hot Water pretreatment conditions to enhance sugar recovery from wheat straw for fuel-ethanol production. *Fuel* **87**, 3640–3647 (2008).

47. Kärcher, M. A., Iqbal, Y., Lewandowski, I. & Senn, T. Comparing the performance of Miscanthus x giganteus and wheat straw biomass in sulfuric acid based pretreatment. *Bioresour. Technol.* **180**, 360–364 (2015).

48. Kootstra, A. M. J., Beeftink, H. H., Scott, E. L. & Sanders, J. P. M. Comparison of dilute mineral and organic acid pretreatment for enzymatic hydrolysis of wheat straw. *Biochem. Eng. J.* **46**, 126–131 (2009).

49. Guerra-Rodríguez, E., Portilla-Rivera, O. M., Jarquín-Enríquez, L., Ramírez, J. A. & Vázquez, M. Acid hydrolysis of wheat straw: A kinetic study. *Biomass and Bioenergy* **36**, 346–355 (2012).

50. Morinelly, J. E., Jensen, J. R., Browne, M., Co, T. B. & Shonnard, D. R. Kinetic characterization of xylose monomer and oligomer concentrations during dilute acid pretreatment of lignocellulosic biomass from forests and switchgrass. *Ind. Eng. Chem. Res.* **48**, 9877–9884 (2009).

51. Liu, X., Lu, M., Ai, N., Yu, F. & Ji, J. Kinetic model analysis of dilute sulfuric acid-catalyzed hemicellulose hydrolysis in sweet sorghum bagasse for xylose production. *Ind. Crops Prod.* **38**, 81–86 (2012).

52. Jin, Q., Zhang, H., Yan, L., Qu, L. & Huang, H. Kinetic characterization for hemicellulose hydrolysis of corn stover in a dilute acid cycle spray flow-through reactor at moderate conditions. *Biomass and Bioenergy* **35**, 4158–4164 (2011).

53. Bas, D. & Boyaci, I. H. Modeling and optimization II: Comparison of estimation capabilities of response surface methodology with artificial neural networks in a biochemical reaction. *J. Food Eng.* **78**, 846–854 (2007).

54. Weiss, N. D., Farmer, J. D. & Schell, D. J. Impact of corn stover composition on hemicellulose conversion during dilute acid pretreatment and enzymatic cellulose digestibility of the pretreated solids. *Bioresour. Technol.* **101**, 674–678 (2010).

8

2.7 Supplementary Material

2.7.1 Design of Experiments

Both pretreatment methods are highly influenced by the pretreatment temperature and the residence time. Furthermore, for the dilute acid pretreatment, the amount of used acid is influential. This results in an experimental design with two factors for the autothermal and three factors for the dilute acid pretreatment.

In the light of the desired optimization of operational conditions and the development of a mechanistic model for the pretreatment, a central composite design (CCD) is chosen. The properties and setup of a CCD are described by Dean et al. as follows: The CCD consists of a factorial design with two levels (-1,1) for reach investigated factor, a set of center points, with the median value (0) of each factorial and a set of axial or star points, which are identical to the center points except for one factor but take values both below and above the median of the two factorial levels $(-\alpha, \alpha)$. The center point is replicated in order to calculate the variance of the experimental data. The CCD can either be circumscribed, inscribed, or face-centered. With the circumscribed design, the factorial points are on the two levels, and the axial points lie on a circle, circumscribing the factorial block with $\alpha > 1$. For the inscribed design, the star points lie on the two levels (-1,1), and the factorial points lie on $(-\alpha, \alpha)$ with $\alpha < 1$. For the face-centered design, both the factorial and the axial points lie at (-1,1) [1]. In this study, the CCD is calculated with the RSM library in R [2].

2.7.2 Reaction Equations of the Mechanistic Model

All indicated reactions are stoichiometrically formulated in Table S1, with water as reactant being excluded due to it being the main component of the reaction phase and hence assuming its concentration constant.

Table S1: Stoichiometric reaction equations of the mechanistic model

1.	$Xyn \rightarrow 1.136\ Xyo$
2.	$Cel \rightarrow 1.111\ Glu$
3.	$Arn \rightarrow 1.136\ Ara$
4.	$Act \rightarrow 1.365\ Aac$
5.	$Glu \rightarrow HMF$
6.	$HMF \rightarrow Fac$
7.	$Xyl \rightarrow Fur$
8.	$Xyl \rightarrow Deg$

2.7.3 Parameter Estimation & Identifiability Analysis

2.7.3.1 Parameter Estimation

According to Sin et al. (2016), the model parameters θ are treated as true, their estimators $\hat{\theta}$are treated as random variables. Parameters, in general, have to be estimated from data, which is inherently error-prone. Therefore, a stochastic process is assumed:

$$y = f(\theta) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \tag{1}$$

The model output y is a function of the model parameters θ plus assumed measurement errors with Gaussian white noise ε. Due to the mentioned error-prone measurements, the estimation has a degree of uncertainty. This uncertainty of the parameters propagates to the model output when using it for simulation purposes. In order to obtain values for θ, they are estimated from a likelihood estimation L following a multivariate normal distribution:

$$L(y,\theta) = \frac{1}{\sigma\sqrt{2\pi}} \cdot exp\,(-\frac{(y-f(\theta)^2}{2\sigma^2}). \quad (2)$$

The most likely estimate of the parameters θ are the parameters $\hat{\theta}$, which maximize the likelihood function:

$$\hat{\theta} = \arg\max L(y,\theta). \quad (3)$$

A particular case of the maximum likelihood estimation assumes independent measurement errors with Gaussian white noise. Then the likelihood function becomes the following loss function:

$$S(y,\theta) = \sum \frac{(y-f(\theta))^2}{\sigma^2}. \quad (4)$$

The minimum of this function again equals the estimators:

$$\hat{\theta} = \arg\min S(y,\theta). \quad (5)$$

To assess the quality of the estimated parameters, the covariance matrix of the estimators is calculated and analyzed:

$$COV(\hat{\theta}) = s^2 \cdot (\mathcal{J}' \cdot \mathcal{J})^{-1}, \quad (6)$$

Where $\mathcal{J}$ is the Jacobian of the estimated parameters:

$$\mathcal{J} = \frac{\partial f(\theta)}{\partial \theta}, \quad (7)$$

And σ^2 the unbiased estimation of the variance, obtained from the residuals of the minimization of the loss function:

$$\sigma^2 = \frac{S_{min}(y, \theta)}{n - p}, \tag{8}$$

With n as the number of measurements and p as the number of estimated parameters. Based on the covariance matrix, the confidence interval of the prediction is calculated. For this, the covariance matrix of the prediction is determined by a linear error propagation approach:

$$COV(y) = \mathcal{J} \cdot COV(\theta) \cdot \mathcal{J}', \tag{9}$$

which then yields the $(1 - \alpha)$ confidence interval of the predictions:

$$y_{1-\alpha} = y \pm \sqrt{\overline{diag(COV(y))}} \cdot t(N - m, \alpha/2), \tag{10}$$

Assuming a student-t distribution [3]. The estimation problem is solved with the *lsqnonlin* solver of the Optimization toolbox in MATLAB.

2.7.3.2 Identifiability Analysis

According to Brun et al. (2002), the identifiability analysis uses the first-order derivative of an output y with respect to the inputs $\boldsymbol{x}$. The absolute sensitivity sa is defined as the effect on perturbing x around its nominal value x^0:

$$sa = \frac{\partial y}{\partial x}. \tag{11}$$

Consequently, the relative sensitivity sr is defined as the effect on perturbing x around a nominal value x^0 with a fixed fraction of the nominal value:

$$sr = \frac{\partial y}{\partial x} \cdot \frac{x^0}{y^0}. \tag{12}$$

The values for the absolute and the relative sensitivity are calculated numerically with a small perturbation Δx of the model input:

$$\frac{\partial y}{\partial x} = \frac{f(x^0 + \Delta x) - f(x^0 - \Delta x)}{2\Delta x}. \tag{13}$$

The following methodology consists of two steps. The first step is a parameter significance ranking. The parameter significance δ^{msqr} is calculated as follows:

$$\delta^{msqr} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (sr_i)}. \tag{14}$$

As the second step, the collinearity indices γ_k of all parameter subsets θ_k for $k \in K$ as the subset in the set K of full factorial combinations of all parameters are calculated. The collinearity index is defined as follows:

$$\gamma_k = \frac{1}{\sqrt{\min \lambda_k}}, \tag{15}$$

with λ_k being the following eigenvalue:

$$\lambda_k = \text{eigen}(snorm_k^{'} \cdot snorm_k), \tag{16}$$

where the normalized sensitivity is defined as:

$$snorm_k = \frac{sr_k}{\|sr\|}. \tag{17}$$

The interpretation of these calculations is twofold: Ideally, all parameters, which are identified to have high cumulated significance δ^{msqr} are found in a parameter subset k^*, which then possibly shows a low collinearity index. In that way, all the significant parameters can be

uniquely identified in a parameter analysis, being favorable for a robust model [4]. The entire procedure is implemented in MATLAB.

7.4 Uncertainty and Sensitivity Analysis

7.4.1 Monte Carlo-Based Uncertainty Analysis

In detail, Monte Carlo method described by Sin et al. (2009) is structured in the following way:

1. Input Uncertainty Definition

For all the parameters m, which are considered uncertain, a range or distribution for each is defined. The distribution is usually normal or uniform, but other distributions are also possible. With estimated parameters, favorably the covariance matrix and standard deviations are provided.

2. Sampling

For the sampling, a sufficiently high number N is set, as well as an appropriate sampling technique. In this case, Latin Hypercube Sampling is chosen due to its reliability for a high-dimensional input space. The output is a sampled input space $X_{N \times m}$.

3. Monte Carlo Simulation

The sampled input space $X_{N \times m}$ is now used to create the respective output space $Y_{N \times n}$ for all output variables n by evaluating the model for all samples N.

4. Analysis

From the output, the mean value, standard deviations, and percentiles are calculated. In the case of priorly estimated parameters, both have to be considered contextually to interpret the uncertainties conclusively [5]. The entire procedure is implemented in MATLAB.

7.4.2 Variance-Based Sensitivity Analysis

Saltelli et al. (2010) describe the method as follows: For independent inputs, the variance of the model output $var(y)$ can be partitioned in the following way:

$$var(y) = \sum_{i}^{k} V_i + \sum_{i} \sum_{k} V_{ij} + \ldots + V_{123\ldots k}. \quad (18)$$

The variance results to:

$$V_i = \int f(\theta_i)^2 \, d\theta_i, \quad (19)$$

and the corresponding function evaluation to:

$$f(\theta_i) = E(y|\theta_i) - f_0. \quad (20)$$

For the sensitivity analysis, both the first-order sensitivity index $S_i = V_i$ of an input parameter θ_i, as well as its total sensitivity index S_{Ti} can be calculated. The first order sensitivity explains this parameter's single effect, indicating the expected variance reduction if this parameter could be fixed. Hence, the total sensitivity is the expected variance if all the parameters except the respective one could be fixed [6].

7.5 Model Validation & Metrics

Each model in the presented study is validated differently. The particular procedures of each model are listed in the following section:

- Analysis of Variance

The RSM model as a statistical model is classically validated by analyzing the significance of each term through an analysis of variance (ANOVA) [8]. As the underlying statistical hypothesis of RSM is that there is a relationship between the input variables and the predicted output variable, the ANOVA utilizes the mean and variance of each estimated model term in order to determine the quality of the fit [9]. For

a detailed explanation of the statistical backgrounds on ANOVA, the F-test, and the p-value, the reader is referred to the book by Dean et al. (2012) [9].

- Cross-Validation

In machine learning, a standard model validation practice is cross-validation and, in particular k-fold cross-validation [10]. Here, the set of data points is divided into k equal parts, and subsequently, $k-1$ parts are used for training the model – in this case, the GPR – and the k-th part of the data is held back for testing the model. This procedure is repeated k times so that each of the equal parts is used once for testing and $k-1$ times for training the model. For each iteration, the desired validation metrics – in this case, the coefficient of determination and the root mean square error – are calculated. A typical value for cross-validation is $k = 5$ [10].

- Train & Test Single-Split Validation

In contrast to cross-validation, where one data set is split into several parts and repeatedly used, for the classic train & test split validation, the data set is split into one dedicated training data set and one dedicated testing data set; the split ratio depends on the underlying case [10], [11].

The used metrics for quantifying both the goodness of fit and the predictive quality of the model, both the coefficient of determination R^2 and the root mean square error $RMSE$ are suitable metrics. The former is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i \in I} y_i - \hat{y}_i}{\sum_{i \in I} y_i - \bar{y}}, \tag{21}$$

With y_i being the measured data points in the set I, $\hat{y}_i$ being the equivalent data points predicted by the model and $\bar{y}$being the mean value of the measured data in set I. The latter is calculated with the following equation:

$$RMSE = \sqrt{\frac{\sum_{i \in I} (y_i - \hat{y})^2}{|I|}}, \tag{22}$$

With the cardinality of the set of data points $|I|$ in order to calculate the mean value. Due to the particularities of each introduced model, the precise procedure for their validation differs in certain aspects. A detailed description of each model can be found in the supplementary material.

7.6 Optimization

For solving the constrained nonlinear problem with each presented model in order to optimize the operational conditions for the pretreatment, different functionalities and toolboxes are used. They are presented in the following section:

- RSM

The RSM model is two times differentiable and can be optimized analytically by calculating the gradient ∇ of the polynomial and its Hessian matrix $\mathcal{H}$ to global optimality [12]. The operational conditions x^* that maximize the response variable need to fulfill the following conditions:

$$\nabla f_{RSM}(x^*) = 0, \quad \mathcal{H}(x^*) < 0. \tag{23}$$

This optimization is constrained to the upper and lower bounds of the operational variables, which do not fulfill the criteria in equation (23); these have to be evaluated separately. Equally, for the optimization study, the *rsm* library in R is chosen

- GPR

In contrast to the RSM model, a GPR model is not per se differentiable; this depends on the employed kernel and basis function. Hence, a numerical solution of the NLP with the GPR model instead of an analytical solution is calculated in this study. In the scope of this study, a genetic algorithm as representative of metaheuristics is chosen. Genetic Algorithms are population-based solvers that mimic biological

evolution and impose constraints on the objective function of the underlying optimization problem. They do not necessarily guarantee global optimality; however, this shortcoming is alleviated due to the population-based approach [13]. For the optimization, the *ga* function as the genetic algorithm of the Global Optimization toolbox of MATLAB is used with the fitted GPR.

- Mechanistic Model

For the mechanistic model, the underlying functional relations are differentiable. However, obtaining an analytical solution to the optimization problem is far from straightforward due to the model being a system of partially coupled ordinary differential equations. Hence, for this NLP, a numerical solver based on sequential quadratic programming is employed, being a gradient-based solver that works efficiently for constrained NLPs [12]. The fmincon solver is used for the optimization, employing a sequential quadratic programming algorithm optimizer in MATLAB. Furthermore, the *multi-start* function of the Global Optimization toolbox in MATLAB is used to assure global optimality.

- Optimization under Uncertainty

Al et al. (2020) describe the solver as follows: The solver itself utilizes Monte Carlo simulations in connection with a stochastic kriging metamodel. The metamodel is closely related to the presented GPR but extended with a term that incorporates extrinsic uncertainty. The solver utilizes an infill criterion which calculates an expected improvement for the next iteration upon which this new point is evaluated and added as an input to the metamodel. The solver is benchmarked thoroughly and excels in combination with the introduced infill criterion compared to others [14].

7.7 Parameter Estimation

The initially estimated parameters for the activation energies and the reaction order exponents, which are fixed for the estimation of the activation energies, are listed in Table S2:

Table S2: Values of the pretreatment model's fixed parameters based on an initial estimation.

Parameter	Value	Parameter	Value
$A_{Xyn,Xyl}$	$7.2492 \cdot 10^{11}$	$n_{Xyn,Xyl}$	1.3029
$A_{Cel,Glu}$	$8.3404 \cdot 10^{1}$	$n_{Cel,Glu}$	0.45938
$A_{Arn,Ara}$	$2.8249 \cdot 10^{5}$	$n_{Arn,Ara}$	0.77198
$A_{Act,Aac}$	$1.8263 \cdot 10^{3}$	$n_{Act,Aac}$	0.9038
$A_{Glu,HMF}$	$6.6523 \cdot 10^{5}$	$n_{Glu,HMF}$	0.44371
$A_{HMF,Fac}$	$1.7066 \cdot 10^{3}$	$n_{HMF,Fac}$	1.2807
$A_{Xyl,Fur}$	$2.1498 \cdot 10^{1}$	$n_{Xyl,Fur}$	0.042991
$A_{Xyl,Deg}$	$2.914 \cdot 10^{2}$	$n_{Xyl,Deg}$	0.46321

The scripts for the parameter estimation are provided through a GitHub repository [15].

7.8 Convergence of the Uncertainty Analysis

The convergence of the uncertainty analysis is visualized in S 1 .

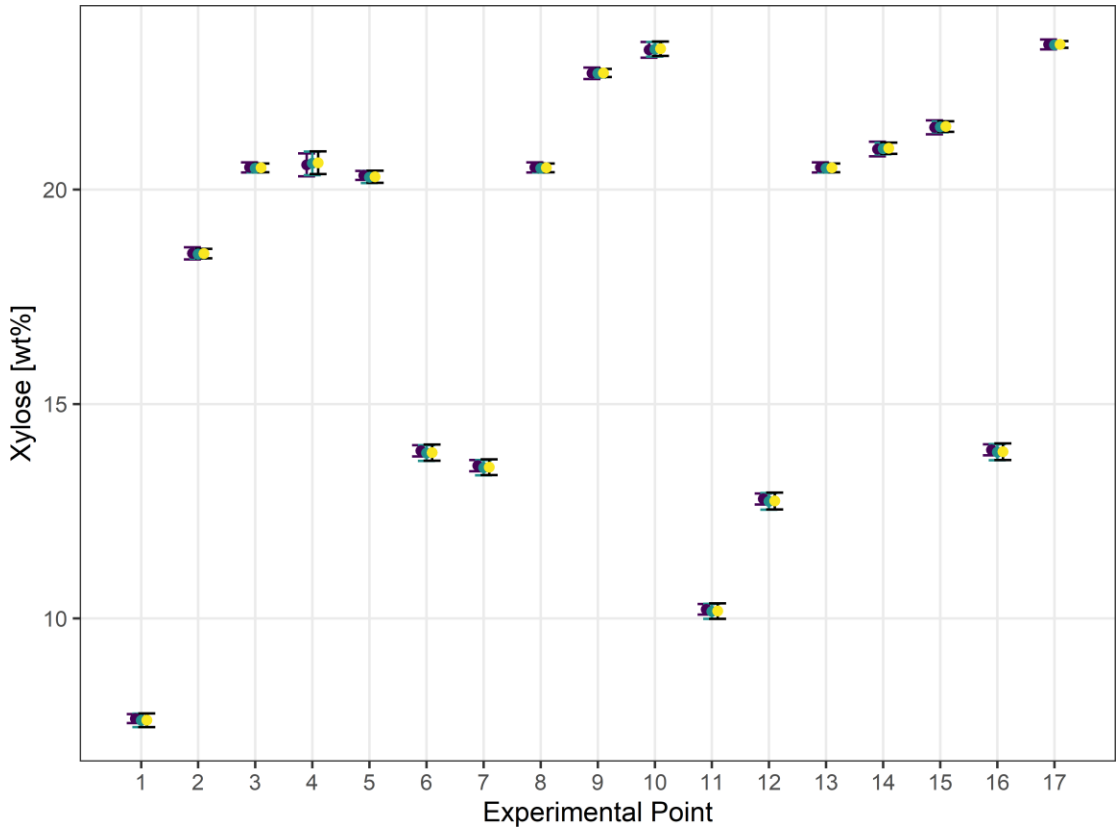


Figure S1: Convergence of the uncertainty analysis in the prediction of xylose for N=10 (●), N=100 (●) and N=1000 (●) Monte Carlo samples

For all 17 design points, the Monte Carlo simulations are repeated for $N = 10$, $N = 100$ and $N = 1000$ samples. It is visible, that the convergence is reached at fairly low sampling levels.

References

[1] A. Dean, D. Voss, and D. Draguljić, "Response Surface Methodology," in *Design and Analysis of Experiments*, Angela Dean, Daniel Voss, and Danel Draguljić, Eds. Springer, Cham, 2017, pp. 565–614.

[2] R. V. Lenth, "Response-surface methods in R, using RSM," *Journal of Statistical Software*, vol. 32, no. 7, pp. 1–17, Oct. 2009, doi: 10.18637/jss.v032.i07.

[3] S. Gürkan and G. Krist, *Data Handling and parameter estimation*, vol. 9781780404. London: IWA Publishing, 2016.

[4] R. Brun, M. Kühni, H. Siegrist, W. Gujer, and P. Reichert, "Practical identifiability of ASM2d parameters - Systematic selection and tuning of parameter subsets," *Water Research*, vol. 36, no. 16, pp. 4113–4127, Sep. 2002, doi: 10.1016/S0043-1354(02)00104-5.

[5] G. Sin, K. V. Gernaey, and A. E. Lantz, "Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis," in *Biotechnology Progress*, Jul. 2009, vol. 25, no. 4, pp. 1043–1053, doi: 10.1002/btpr.166.

[6] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," *Computer Physics Communications*, vol. 181, no. 2, pp. 259–270, Feb. 2010, doi: 10.1016/j.cpc.2009.09.018.

[7] R. Al, C. R. Behera, A. Zubov, K. V. Gernaey, and G. Sin, "Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model," *Computers and Chemical Engineering*, vol. 127, pp. 233–246, Aug. 2019, doi: 10.1016/j.compchemeng.2019.05.015.

[8] G. Leonzio and E. Zondervan, "Innovative application of statistical analysis for the optimization of CO2 absorption from flue gas with ionic liquid," in *Computer Aided Chemical Engineering*, vol. 46, Elsevier B.V., 2019, pp. 151–156.

[9] A. Dean, D. Voss, and D. Draguljić, *Design and Analysis of Experiments*, vol. 3. 2012.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY: Springer, 2009.

[11] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, 1st ed. Springer, 2013.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*. 2006.

[13] P. Siarry, *Metaheuristics*. Cham, Switzerland: Springer International Publishing, 2016.

[14] R. Al, C. R. Behera, K. V. Gernaey, and G. Sin, "Stochastic simulation-based superstructure optimization framework for process synthesis and design

under uncertainty," *Computers and Chemical Engineering*, vol. 143, p. 107118, Dec. 2020, doi: 10.1016/j.compchemeng.2020.107118.

[15] N. Vollmer, "Pretreatment Model," *GitHub Repository*, 2021. https://github.com/NikolausVollmer/Pretreatment-Model.

Chapter 3

EFFECTS OF INHIBITORY COMPOUNDS PRESENT IN LIGNOCELLULOSIC BIOMASS HYDROLYSATES ON THE GROWTH OF *BACILLUS SUBTILIS*

Jasper L. S. P. Driessen [1,†],
Lucas van der Maas[1,†], Solange I. Mussatto [2,*]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, 220, 2800 Kongens Lyngby, Denmark

[2] Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 223, 2800 Kongens Lyngby, Denmark

*** Corresponding author**
Solange I. Mussatto smussatto@dtu.dk

† These authors contributed equally to this work.

This chapter comprises a published manuscript.
DOI.org/10.3390/en14248419

Abstract

This study evaluated the individual and combined effects of inhibitory compounds formed during pretreatment of lignocellulosic biomass on the growth of *Bacillus subtilis*. Ten inhibitory compounds commonly present in lignocellulosic hydrolysates were evaluated, which included sugar degradation products (furfural and 5-hydroxymethylfurfural), acetic acid, and seven phenolic compounds derived from lignin (benzoic acid, vanillin, vanillic acid, ferulic acid, p-coumaric acid, 4-hydroxybenzoic acid, and syringaldehyde). For the individual inhibitors, syringaldehyde showed the most toxic effect, completely inhibiting the strain growth at 0.1 g/L. In the sequence, assays using mixtures of the inhibitory compounds at a concentration of 25% of their IC50 value were performed to evaluate the combined effect of the inhibitors on the strain growth. These experiments were planned according to a Plackett-Burman experimental design. Statistical analysis of the results revealed that in mixture, benzoic acid and furfural were the most potent inhibitors affecting the growth of *B. subtilis*. These results contribute to a better understanding of the individual and combined effects of inhibitory compounds present in biomass hydrolysates, on the microbial performance of *B. subtilis*. Such knowledge is important to advance the development of sustainable biomanufacturing processes using this strain cultivated in complex media produced from lignocellulosic biomass, supporting the development of efficient bio-based processes using *B. subtilis*.

3.1 Introduction

Due to its low cost, large availability (181.5 billion tons/year)[1], and sugar-rich composition, lignocellulosic biomass has attracted great interest to be used as a feedstock for the production of a wide range of bio-based products. One of the main advantages of using these materials in bioprocesses is that they do not compete with the food supply or existing arable land, as lignocellulosic materials include residues and side streams from agriculture, forestry, energy crops, biorefineries, and pulp mills[2,3]. In the last decade, lignocellulosic biomass has gained increased attention as a feedstock for many industrial processes, among which includes the production of enzymes. Although the global market for enzymes in industrial applications is expected to grow from USD 6.4 billion in 2021 to USD 8.7 billion by 2026[4], high production costs and high levels of competition have put pressure on the enzyme industry to seek new sustainable alternatives for their processes. Using lignocellulose as a feedstock can improve the sustainability of the production chain and reduce substrate costs, which, next to capital investment (50%), make up a third of the total costs[5].

Lignocellulose mainly consists of cellulose, hemicellulose, and lignin fractions. In spite of the potential of lignocellulosic biomass as a source of sugars for bioprocesses, some key problems related to their utilization still must be overcome to accelerate the transition towards a society less dependent on fossil fuels. One of these main problems is due to the fact that sugars present in lignocellulosic biomass are not freely available for biochemical conversion by microorganisms. Therefore, a pretreatment step is commonly required to break the physical and chemical bonds between the main constituents of the biomass, releasing sugars that can be used for fermentation[6]. However, pretreatment is not a selective reaction, and besides solubilizing hemicellulose sugars, it also promotes the formation and release of several other compounds from the lignocellulose structure, which have negative effects on fermentation, affecting the microbial metabolism and reducing the efficiency of the strain to convert sugars into products. The severity and mechanism of inhibition depend on the chemistry of the specific

compound, the environment during microbial fermentation, and the tolerance of the microorganism to each toxic compound[7]. The by-products generated from cellulose and hemicellulose fractions range from weak acids, for example from the acetyl groups present in hemicellulose, to furans formed during the degradation of sugars. Part of the lignin also breaks down during pretreatment, generating other inhibitory (phenolic) compounds[8]. The presence of inhibitory compounds together with sugars is one of the major challenges hindering the efficient utilization of lignocellulosic biomass hydrolysates in bioprocesses. Elucidating the individual and combined effects of these compounds on microbial performance to grow is essential to finding solutions to this problem, and obtaining an efficient and cost-competitive product formation from complex media produced from plant-based materials[9]. In this study, *Bacillus subtilis* was the microbial strain considered to evaluate the effect of inhibitory compounds, as it is a major workhorse in the production of industrial enzymes. This status is due to several reasons: it has high growth rates, is generally regarded as safe (GRAS) by the FDA, and is able to secrete high levels of protein[10]. Moreover, *B. subtilis* has a number of validated and putative transporters, which enable it to take up several types of monomeric sugars, making it a prime candidate for enzyme production based on biomass. Furthermore, the physiology and biochemistry of *B. subtilis* have been extensively studied due to its relevance to industry. The aim of this study was to elucidate the individual and combined effects of the inhibitory compounds present in lignocellulosic biomass hydrolysates on the growth performance of *B. subtilis*. The ten most common inhibitory compounds present in lignocellulosic hydrolysates were tested at different concentration levels, and the strain growth was monitored. The half maximal inhibitory concentration (IC50) was determined for each inhibitor. Then, assays using a mixture of the inhibitors were performed according to a Plackett–Burman experimental design. In the end, the results allowed us to conclude which individual inhibitory compounds and mixtures are the most toxic to *B. subtilis*.

3.2 Material and Methods

3.2.1 Microorganism and Inoculum Preparation

Bacillus subtilis BS168 was the microorganism used in the experiments. The strain was grown in M9 minimal media (Sigma-Aldrich, Burlington, MA, USA) supplemented with glucose (10 g/L) and tryptophan (0.05 g/L), at 250 rpm and 37 ◦C. Cryopreserved cells were grown overnight on LB agar plates at 37 ◦C, after which a single colony was used to inoculate a 12 mL culture tube containing 5 mL of LB, and the culture was kept for 8 h. Then, 10 µL of this culture was grown for 15 h in a 250 mL shake flask containing 20 mL of M9 medium (supplemented with 1 mL of 10 g/L yeast extract solution) to serve as inoculum for the growth experiments.

3.2.2 Culture Media and Conditions

The following inhibitory compounds were tested in the experiments: furfural, 5-hydroxymethylfurfural (5-HMF), acetic acid, vanillin, vanillic acid, benzoic acid, ferulic acid, p-coumaric acid, 4-hydroxybenzoic acid, and syringaldehyde (Sigma-Aldrich). Stock solutions of all inhibitors were prepared in 98% (v/v) ethanol at concentrations close to their respective solubility. Subsequently, dilutions were made in M9 minimal medium to obtain the inhibitors at appropriate concentrations for use in the experiments. To determine the effect of inhibitors on microbial growth rate, cells were grown in microtiter plates incubated in a Growth Profiler 960 (EnzyScreen, Heemstede, The Netherlands) at 250 rpm, 37 ◦C. Microtiter plate wells (280 µL) were inoculated with 20 µL of diluted preculture, so that the starting optical density (OD600) was approximately 0.05. The OD600 was measured every 20 min. All tests were performed in triplicate.

3.2.3 Determination of Growth Rate and Lag Phase

Growth rates were determined following the method described by Hemmerich et al. [11], which is based on an iterative procedure: first, the growth curve is transformed by a natural logarithm, followed by weighted linear regression to obtain the logarithmic growth phase. Then, three stopping criteria are applied. If the stopping criteria are not met, a new iteration is started. The three stopping criteria are defined as follows: firstly, the set R^2 value has to be met. The second criterion dictates that the increase in biomass in the last data point needs to be higher than the increase in the previous point to exclude data from the transition to the stationary phase. The final stopping criterion dictates that there must be an overall positive increase in biomass to rule out any measurement artefacts. If all these stopping criteria are met, the μ is calculated according to Equation (1).

Equation 1

$$\mu = \frac{1}{c_X} \times \frac{dc_X}{dt} \approx \frac{\Delta\, ln(c_x)}{\Delta t}$$

The lag phase of bacterial growth can be mathematically defined as the time up to the maximum of the second derivate of the growth curve[12]. This is the time point at which the growth rate increase is maximal. Determination of the lag phase was performed manually in Excel (Microsoft, Richmond, VA, USA) by taking the moving average of 7 points in the growth. The half-maximal inhibitory concentration (IC50) was defined as the concentration of the inhibitor at which the growth rate μ was half that of the control[13].

3.2.4 Plackett–Burman Experimental Design

The Plackett–Burman experimental design was chosen to evaluate the effect of mixtures of inhibitors, since it allows the screening of numerous parameters in a relatively small number of experiments[14]. In these experiments, 12 different combinations of 10 different inhibitors were tested in triplicate. The concentration at which the growth rate was 50% of the control without inhibition (IC_{50}) was calculated to each inhibitor by linear regression of the linear part of the inhibition curves. As combining multiple inhibitors at their corresponding IC_{50} values resulted in no growth, lower concentrations were also tested. One-eighth of the IC_{50} values yielded an optimal response for the determination of the main effects of inhibitors on growth rate. The design and statistical analysis of the Plackett–Burman experimental design were performed using the software Minitab (Minitab Inc., State College, PA, USA). Additional data about the Plackett–Burman design are shown in the Supplementary Materials.

3.3 Results and Discussion

3.3.1 Individual Effects of Inhibitory Compounds

Experimental growth curves and lag phases of *B. subtilis* grown in varying concentrations of the 10 different inhibitory compounds are shown in Figures 1 and 2. To obtain an appropriate range of growth for all inhibitory compounds, multiple independent experiments were performed with varying concentrations. The results obtained for each category of inhibitor compound (furan derivatives, weak acids, and phenolic compounds) are presented and discussed below.

3.3.1.1 Furan Derivatives

The degradation of pentoses and hexoses during pretreatment results in the formation of furfural and 5-HMF, respectively[8]. Both furan derivatives are toxic to the cell as they inhibit the glycolytic and fermentative enzymes essential to central metabolic pathways and the cross-linking of proteins, and cause DNA damage[15,16]. In addition, furan derivatives deteriorate membrane integrity because of their high hydrophobicity, causing membrane leakage/disruption and ultimately leading to lower ATP production and a drop in the growth rate[17].

Furfural inhibited the growth of *B. subtilis* at concentrations as low as 0.05 g/L (Figure 1A). The growth rate was about half the value of the control at a concentration of 0.2 g/L and was further reduced to roughly 35% at concentrations of 0.5 g/L and higher (Figure 2A). Higher concentrations resulted in a longer lag phase, while growth rates remained constant. *B. subtilis* appears to be more sensitive to furfural compared to certain yeast species, as it is shown that furfural concentrations below 0.5 g/L had a positive effect on the cell growth of *Pichia stipitis*, for example,[18]. Zheng et al.[19] discovered a *B. subtilis* strain (DS3) capable of growing on and utilizing furfural as sole carbon source. There is a possibility that other *B. subtilis* strains, including the one used in this study, could also utilize furfural as their sole carbon source, but this has not yet been proven and was also not investigated in this study. 5-HMF is commonly present in lower concentrations than furfural in hemicellulose hydrolysates; due to the low amount of hexoses usually present in hemicelluloses, the conditions usually applied for pretreatment (which do not degrade large amount of hexoses), and the

high reactivity of 5-HMF[8]. In the present study, 5-HMF showed an overall similar response to that of furfural: at higher concentrations, the microbial growth rate was reduced to a low, constant level, while the lag phase was prolonged (Figures 1B and 2A). Zhang et al.[20] investigated the effect of furfural and 5-HMF on the growth rate of a *Bacillus coagulans* species and found stronger growth inhibition of furfural compared to 5-HMF at concentrations below 3 g/L. However, at higher concentrations, the inhibition by 5- HMF was more severe than furfural. Although inhibition occurred at lower concentrations in the current work, the trend of inhibition found by Zhang and co-authors matches with the results obtained in this study: *B. subtilis* is more resistant to higher concentrations of furfural than 5-HMF (Figure 2A), but inhibition by furfural occurs even at low concentrations. In contrast, Pereira et al.[21] found that the growth of *B. subtilis* NCCB 70064 was inhibited less by 5-HMF compared to furfural. Moreover, the authors observed growth even at 2 g/L of 5-HMF, while in our study, no growth was observed at 5-HMF concentrations over 1 g/L (Figure 2A).

3.3.1.2 Weak Acids

During pretreatment of lignocellulosic biomass, weak acids such as acetic acid, formic acid, levulinic acid, and benzoic acid can be formed or released from the material structure. Undissociated weak acids are generally liposoluble and able to diffuse across the plasma membrane into the cytosol. Due to the neutral intracellular pH, weak acids dissociate, lowering the pH of the cell [8]. Multiple explanations have been proposed to explain the inhibitory effect of weak acids entering the cell. Active transport and ATPase can remove the dissociated acids and protons, respectively, but both are at the expense of ATP. As the proton-pumping capacity of the cell falls short at higher acid concentrations, the depletion of the ATP content, lower proton motive force, and acidification of the intracellular environment will lead to low cell viability [22]. In addition, it has been suggested that enzymes are not only inhibited by internal acidification, but also by the accumulation of the anionic form of the acid [23]. The inhibitory effects of weak acids are highly dependent on the ratio of dissociated to undissociated forms, which is dictated by the pH of the environment and pKa of the compound.

The two weak acids investigated in this study included acetic acid and benzoic acid. Acetic acid is released from the hemicellulose structure, while benzoic acid is a lignin degradation product [9]. For an acetic acid concentration of 0.75 g/L, no significant inhibition was observed in terms of lag phase or maximal growth rate (Figures 1C and 2B). However, when observing the growth curve (Figure 1C), an inhibitory effect is clear at this concentration. It seems that *B. subtilis* has a biphasic growth pattern when acetic acid is present in the medium. *B. subtilis* cultures are able to consume acetate and produce acetoin when extracellular acetate levels rise to toxic levels. Acetoin is a non-toxic pH-neutral overflow metabolite that can be used as a carbon source in later growth stages [24], which could explain the biphasic growth pattern observed in the present study. At a concentration of 2 g/L, growth only occurs after 48 h at 29% of the rate of the control.

The inhibitory effect of acetic acid is highly dependent on the pH of the medium. The productivity of *P. stipitis*, for example, dropped by 50% when the strain was grown in 0.8 g/L or 13.8 g/L of acetic acid at pH 5.1 or 6.5, respectively [25]. On the other hand, an increase in productivity at concentrations up to 1 g/L was observed for *Candida guilliermondii*, while other authors observed the same effect up to 10 g/L for *Saccharomyces cerevisiae* [26,27]. This allows us to conclude that the inhibiting effect of acetic acid is dependent on the species of microorganism and experimental conditions used for cultivation.

Unlike the acetic acid, the inhibitory effect of benzoic acid was clearly visible from 0.5 g/L onwards, while the lag phase was fairly constant (Figure 1D). The constant lag phase might indicate that the cell does not have a specific coping mechanism for benzoic acid. The cell does need to expend energy to reduce the intracellular concentration of the acid, as well as maintain a suitable intracellular pH; this energy cannot be used for growth, thus reducing the growth rate [28]. According to the literature, the minimal inhibitory concentration of benzoic acid for several yeast species varies between 0.17 and 1.25 g/L [29]. Benzoic acid was also found to be a potent inhibitor of the growth of *Rhodosporidium toruloides*, increasing the lag phase of this yeast by 60% compared to the control when present in concentrations higher than 1 mM [9].

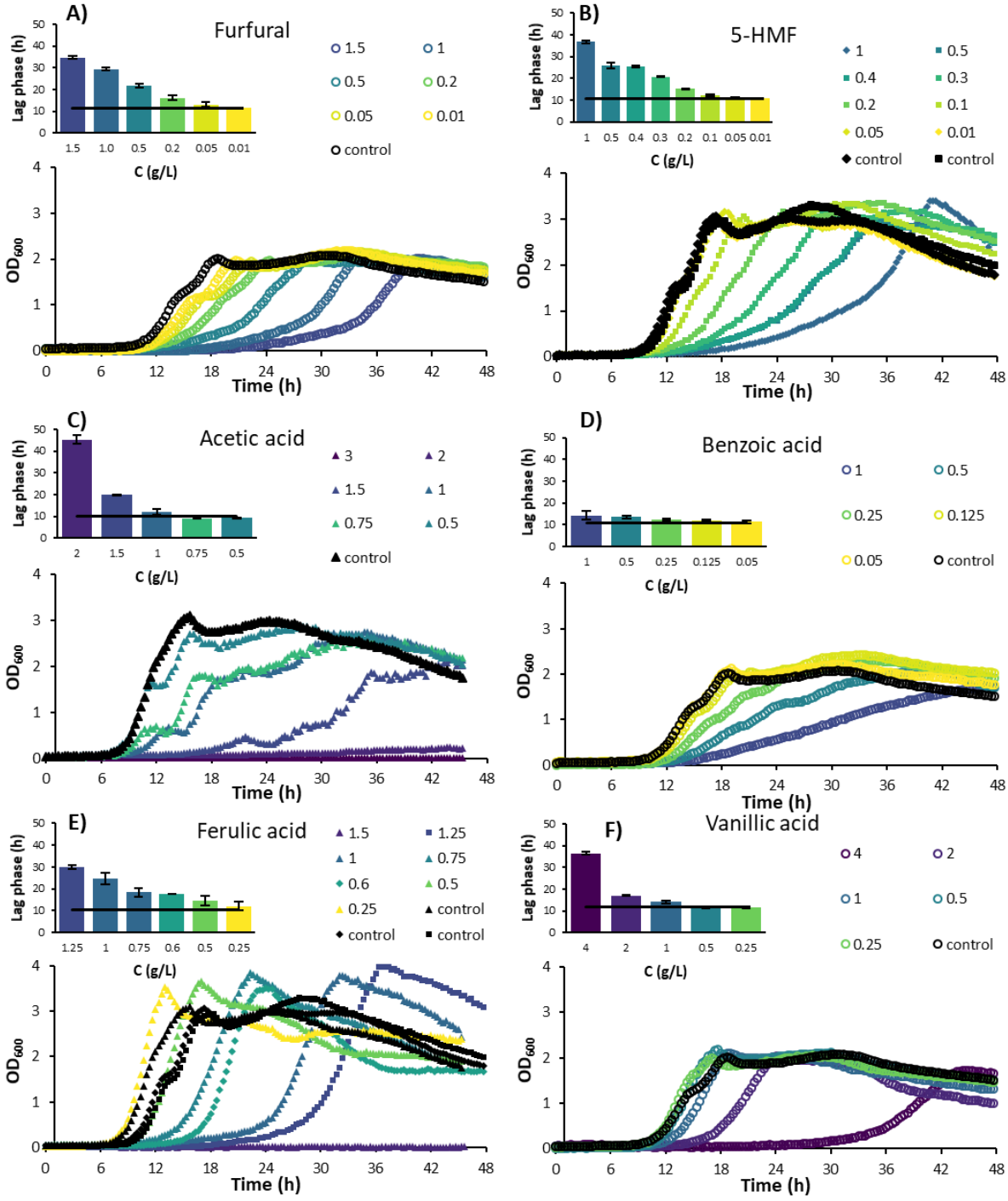


Figure 1. (continues on the next pages)

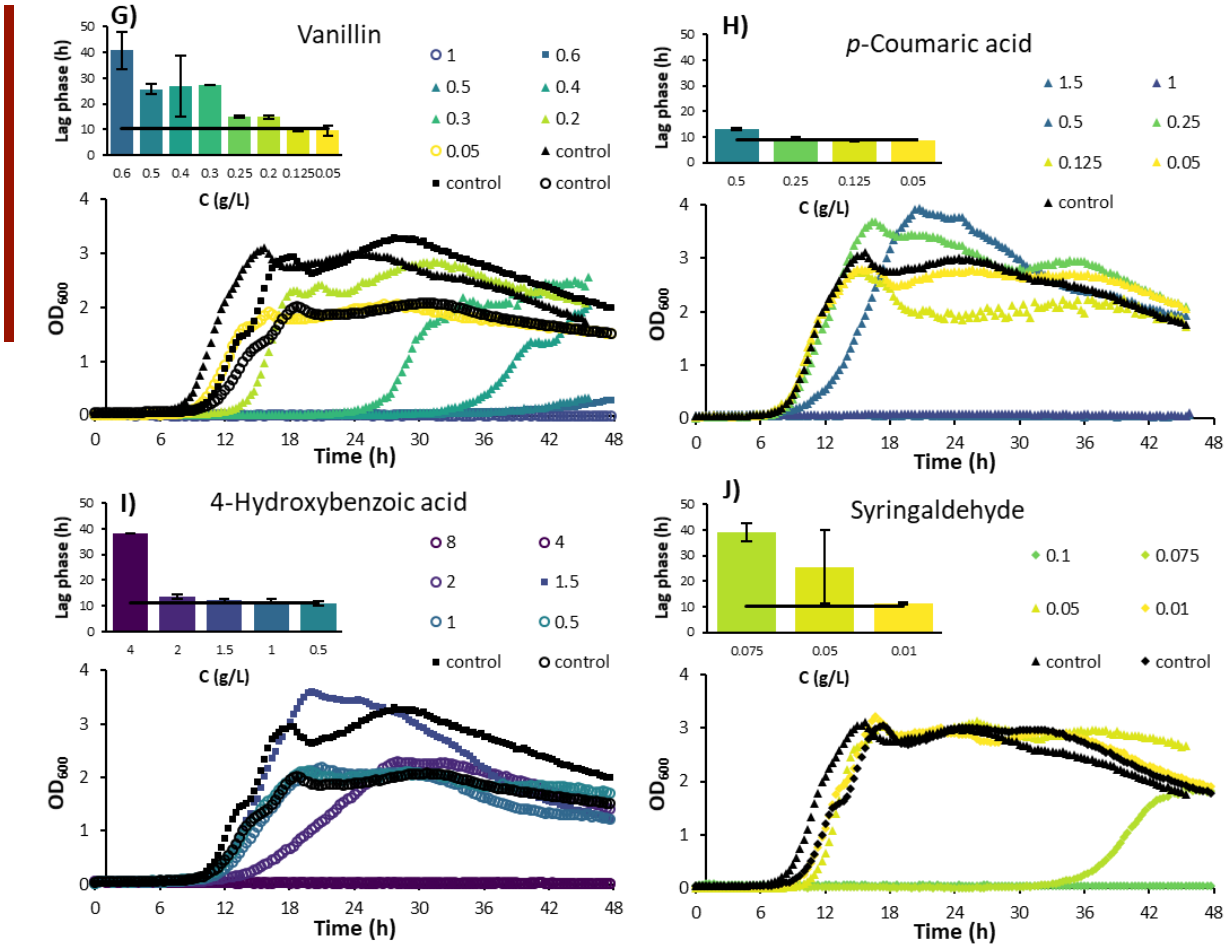


Figure 1. Experimental growth curves of *Bacillus subtilis* in M9 minimal media containing different concentrations of inhibitory compounds. (A) Furfural; (B) 5-Hydroxymethylfurfural (5-HMF); (C) Acetic acid; (D) Benzoic acid; (E) Ferulic acid; (F) Vanillic acid; (G) Vanillin; (H) p-Coumaric acid; (I) 4-Hydroxybenzoic acid; (J) Syringaldehyde. Average standard deviation of the data points was 17%. The lag phases shown in the bar graphs are average values of multiple experiments. The black line corresponds to the average lag phase of the controls without inhibitors present.

3.3.1.3 Phenolic Compounds

Phenolic compounds include acids (e.g., ferulic acid, p-coumaric acid, vanillic acid, 4-hydroxybenzoic acid), alcohols (e.g., guaiacol, catechol, vanillyl alcohol), and aldehydes (e.g., vanillin, syringaldehyde), some of which are considered the most potent inhibitors of microbial growth [30]. Molecular weight, polarity, and side groups dictate the specific inhibitory effect of each phenolic compound. Phenolic compounds generally cause a loss of integrity of cell membranes, leading to a loss of barrier capacity of the membrane. Consequently, a change in the intracellular environment occurs, reducing ATP levels, impairing proton motive force, and reducing protein function and nutrient transport[8]. In addition, phenolic compounds can cause enzyme denaturation, damage the cytoskeleton, as well as cause DNA damage by enhancing the formation of reactive oxygen species, and induce programmed cell death[31].

Ferulic acid, vanillic acid, and vanillin are derivatives of the guaiacyl building block of lignin[32]. For the range of 0.25–1.5 g/L of ferulic acid, the growth of *B. subtilis* was increasingly inhibited with a concomitant prolonged lag phase (**Figures 1E** and **2C**). The gradual increase in the lag phase might indicate that the organism has a way to adapt and survive under conditions with increasing levels of ferulic acid. For vanillic acid, a concentration of 2 g/L caused a reduction in growth to 55% compared to the control, with a 50% increase in the lag phase (**Figure 1F** and **2C**). At 4 g/L of vanillic acid, the growth rate was still at 35% of the control, while the lag phase was increased up to 36 h. This lag phase was longer than that with the chemically similar ferulic acid at the highest concentration, indicating that for vanillic acid, a coping mechanism might also be present. It was found that ferulic acid is a stronger inhibitor to the growth of *Clostridium beijerincki* than vanillic acid[33].

The results match with the effect of vanillic acid reported on the growth of other bacteria. For *Escherichia coli* and *Bacillus cereus*, for example, vanillic acid started to show a negative effect on the strain growth from 0.84 g/L and 0.42 g/L, respectively[34]. Although vanillin is chemically closely related to vanillic acid, it appears to be considerably more toxic. *B. subtilis* could not grow at vanillin concentrations higher than 0.6 g/L

(Figures 1G and 2C), while growth was still observed at 4 g/L of vanillic acid (Figures 1F and 2C).

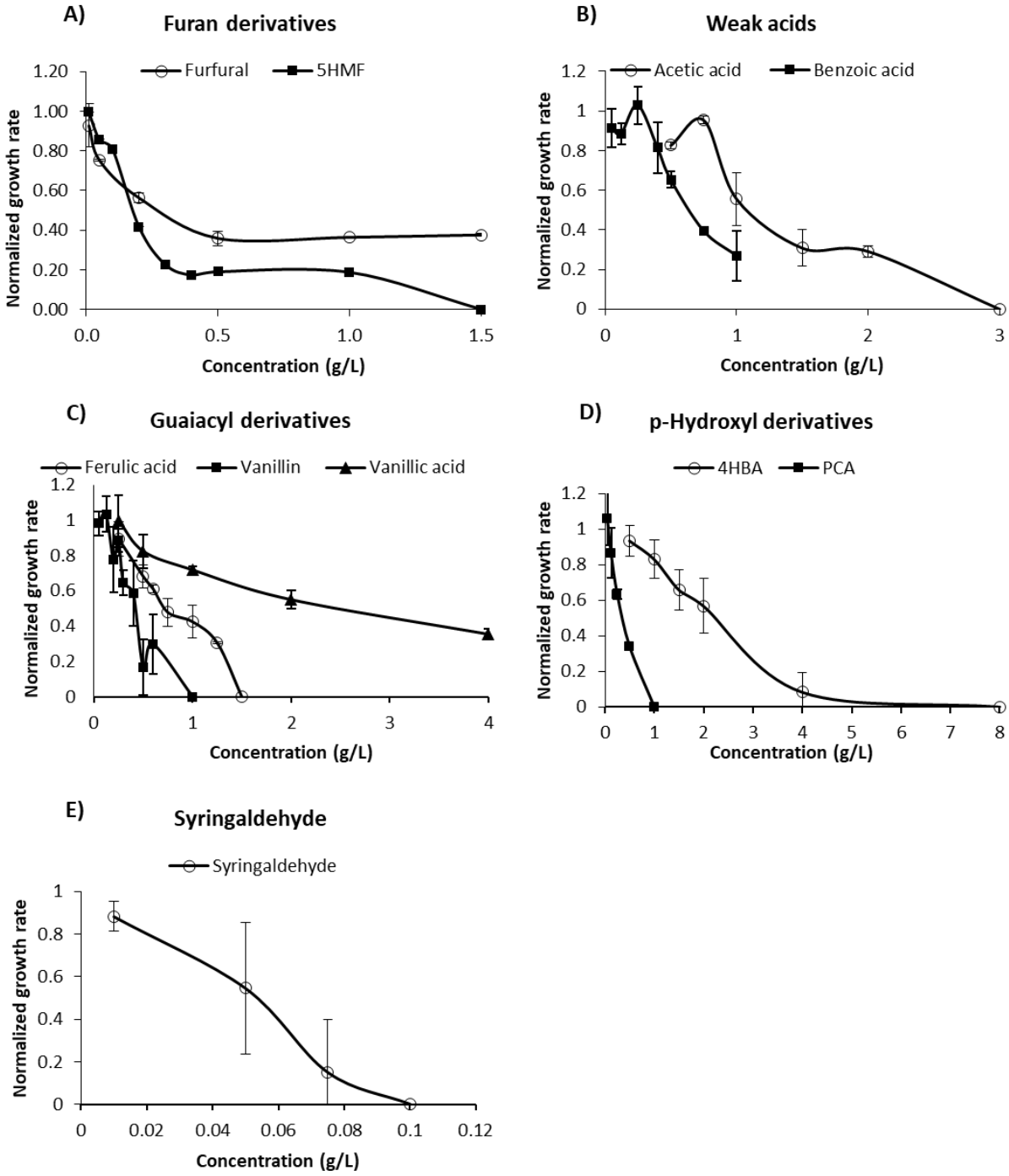


Figure 2. Averaged growth rates of Bacillus subtilis cultivated in M9 minimal media containing different concentrations of inhibitory compounds. Data from 4 independent experiments were normalized to the corresponding control and average values are shown.

p-Coumaric acid and 4-hydroxybenzoic acid are derivatives from the *p*-hydroxyl building block of lignin[32]. While *p*-coumaric acid significantly inhibited the growth rate of *B. subtilis* in concentrations higher than 0.25 g/L, the lag phase did not increase as much as with other inhibitory acids such as ferulic, vanillic, or acetic acids (Figures 1H and 2D). However, no growth was visible at concentrations of 1 g/L or higher. Herald and Davidson[35] studied the effect of *p*-coumaric acid on the growth of *E. coli* and *B. cereus* for different pH levels. Near-complete (99.5%) inhibition of *B. cereus* was found at a concentration of 0.5 g/L at all pH levels tested (6, 6.5, and 7). In contrast, a reduction of only 9% in the growth rate of *E. coli* was observed at 0.5 g/L and pH 7, whereas growth was completely inhibited at pH 5 at the same concentration. Similar to the toxicity of weak acids discussed previously, the inhibiting effect of *p*-coumaric acid for *E. coli* appeared to be heavily dependent on pH. When compared to other inhibitors such as ferulic, acetic, and vanillic acids, *B. subtilis* was inhibited at lower concentrations of p-coumaric acid without a strong increase in the lag phase. This might indicate that *B. subtilis* does not have a mechanism to adapt to the toxic effects of p-coumaric acid, while it does for ferulic, acetic, and vanillic acids.

4-Hydroxybenzoic acid is a phenolic derivative of benzoic acid. The inhibitory effect of 4-hydroxybenzoic acid on the growth rate of *B. subtilis* gradually increased with the increase in the concentration of 4-hydroxybenzoic acid to 2 g/L, while the lag phase was constant (Figures 1I and 2D). However, at 4 g/L, the growth rate was significantly reduced, and the lag phase was increased to approximately 3.5 times the control lag phase. Unlike the present study, Cho et al.[36] found that the growth of *B. subtilis* was reduced to 50% of the control at a concentration of 0.956 g/L of 4-hydroxybenzoic acid. However, in their study, a paper disk assay was used, which was mentioned as showing inconsistent results compared to liquid cultures due to the solubility and polarity of 4-hydroxybenzoic acid.

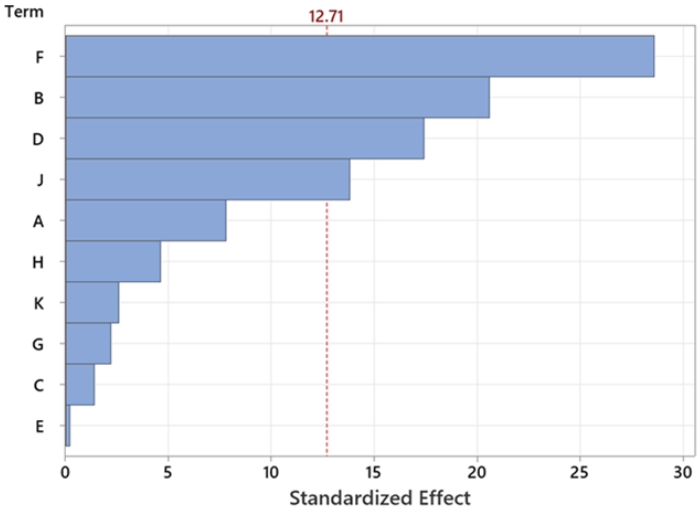
Syringyl propanoids are abundant lignin building blocks in angiosperm plants, which includes all grain-type plants[31]. During pretreatment, syringyl propanoid building blocks of lignin can degrade to syringaldehyde, among other compounds. Syringaldehyde is known to play an important role in membrane disruption. In the present study, syringaldehyde was the most toxic compound to *B. subtilis* as no growth was observed at 0.1 g/L or higher concentrations (**Figures 1J** and **2E**). Syringaldehyde has also been reported to fully inhibit butanol production and significantly reduce the growth of *Clostridium* species at a concentration of 1 g/L[37].

3.3.2 Combined Effect of Inhibitory Compounds

For the optimization of the pretreatment conditions, or effectively engineering tolerant microbial strains, knowledge of the contribution of individual and combined inhibitors to the overall toxic level of lignocellulosic hydrolysates and the underlying toxicity mechanisms are important information still lacking in the literature. Some studies have focused on the interactions of inhibitors by factorial designs or binary combinations[17,27,38]. However, testing interactions of a multitude of inhibitors present in lignocellulosic biomass hydrolysates via factorial designs is not a feasible task when considering the vast number of experimental runs to be performed (2^k, where k = number of inhibitors). Therefore, the main effect of each compound in a mixture of 10 inhibitors was tested in the present study using a Plackett–Burman experimental design. Although the Plackett–Burman design does not provide one-on-one interactions, this analysis does provide an indication of the main inhibitory effect when a large number of process parameters (presence of inhibitors) are used. Residual plots were used to verify whether the model is adequate and meets three general assumptions of the analysis: residuals are randomly distributed, residuals are independent of one another, and residuals are normally distributed (graphs not shown).

The relative magnitude of the main effects of each compound when present in a mixture are presented in the Pareto Chart plotted in **Figure 3A**. As can be seen, when present in a mixture, benzoic acid, furfural, vanillin, and syringaldehyde were the main inhibitory compounds affecting the growth rate of *B. subtilis* (results significant at $p < 0.05$).

A)



B)

Inhibitor	IC-50 (g L^{-1})
5-HMF	0.10
Furfural	1.15
Acetic acid	1.26
Vanillin	0.37
Vanillic acid	2.77
Benzoic acid	0.68
4HBA	1.83
TFA	0.63
Syringaldehyde	0.06
PCA	0.38

Figure 3. (A) Pareto chart of standardized effects to estimate the effect of inhibitors on the reduction in the growth rate of B. subtilis. Bars exceeding the dashed red reference line have a significant main effect for $p < 0.05$. A: 5-HMF; B: Furfural; C: Acetic acid; D: Vanillin; E: Vanillic acid; F: Benzoic acid; G: 4-Hydroxybenzoic acid (4HBA); H: Ferulic acid (TFA); J: Syringaldehyde; K: p-Coumaric acid (PCA). (B): IC50 values determined to each inhibitory compound.

It is worth noting that while all inhibitors were diluted equally in terms of inhibitory effect (12.5% of the IC50 value, shown in **Figure 3B**, results clearly show that the impact of each inhibitor is completely different when present in a mixture. A similar result was observed when evaluating the effect of inhibitory compounds on the growth of *Rhodosporidium toruloides*[9]. According to the authors, benzoic acid was the most potent individual inhibitor affecting the growth of the yeast, while in a mixture, furfural presented the highest toxicity. Zaldivar et al.[17] also found that binary combinations of inhibitors with furfural usually resulted in a stronger inhibition than that caused by the individual inhibitory compounds. Since biomass hydrolysates often contain multiple inhibitors, understanding the combined effect of these compounds on microbial performance is of paramount importance to design an efficient strategy to maximize the yield of a bioprocess from lignocellulosic hydrolysates. In this sense, the experimental design performed in this study gives useful indications of the main effects of the tested compounds on the growth rate of *B. subtilis* when multiple inhibitors are present.

3.4 Conclusions

Using a systematic approach, this study provides a solid base on the individual and combined effects of 10 inhibitors on the growth of *Bacillus subtilis*. When considering individual effects, syringaldehyde was the most toxic compound affecting microbial performance, whereas benzoic acid and furfural had the biggest main effects from a mixture of inhibitors. By combining the information provided in this study with the compositional analysis of a lignocellulosic hydrolysate, better predictions related to the potential toxicity of the hydrolysate and poor microbial performance can be made. This will allow scientists to prioritize strategies to overcome toxicity, not only for detoxification purposes but also for adaptive laboratory evolution experiments, to increase the tolerance of the strain to specific toxic compounds present in the hydrolysate to be used, accelerating the development of efficient fermentation processes using complex media produced from lignocellulosic biomass.

3.5 Acknowledgements

The authors would like to thank Francesco Reggianini for his support with some of the experimental work and data analysis.

This research was funded by the Novo Nordisk Foundation (NNF), Denmark, grant number NNF20SA0066233. PhD fellowships were covered by the NNF grant NNF17SA0031362 and supported by The Technical University of Denmark (DTU).

3.6 References

1 Dahmen, N., Lewandowski, I., Zibek, S., Weidtmann, A. Integrated lignocellulosic value chains in a growing bioeconomy: status quo and perspectives. GCB Bioenergy 2018, 11, 107–117. https://doi.org/10.1111/gcbb.12586.

2 Mussatto, S.I., Yamakawa, C.K., van der Maas, L., Dragone, G. New trends in bioprocesses for lignocellulosic biomass and CO2 utilization. Renew. Sustain. Energy Rev. 2021, 152, 111620. https://doi.org/10.1016/j.rser.2021.111620

3 Dragone, G., Kerssemakers, A.A.J., Driessen, J.L.S.P., Yamakawa, C.K., Brumano, L.P., Mussatto, S.I. Innovation and strategic orientations for the development of advanced biorefineries. Bioresour. Technol. 2020, 302, 122847. https://doi.org/10.1016/j.biortech.2020.122847.

4 BCC Research. Global Markets for Enzymes in Industrial Applications. Available online: https://www.bccresearch.com/market-research/biotechnology/global-markets-for-enzymes-in-industrial-applications.html. (Accessed on 23.10.2021)

5 Ravindran, R., Hassan, S.S., Williams, G.A., Jaiswal, A.K. A review on bioconversion of agro-industrial wastes to industrially important enzymes. Bioengineering 2018, 5, 93. https://doi.org/10.3390/bioengineering5040093

6 Mussatto, S.I., Dragone, G. Biomass pretreatment, biorefineries, and potential products for a bioeconomy development. In: Biomass Fractionation Technologies for a Lignocellulosic Feedstock Based Biorefinery, Mussatto, S.I. (Ed.); Elsevier: Amsterdam, 2016. pp. 1–22. https://doi.org/10.1016/B978-0-12-802323-5.00001-3.

7 Sivagurunathan, P., Kumar, G., Mudhoo, A., Rene, E.R., Saratale, G.D., Kobayashi, T., Xu, K., Kim, S.-H., Kim, D.-H. Fermentative hydrogen production using lignocellulose biomass: An overview of pre-treatment methods, inhibitor effects and detoxification experiences. Renew. Sustain. Energy Rev. 2017, 77, 28-42. https://doi.org/10.1016/j.rser.2017.03.091

8 Mussatto, S.I., Roberto, I.C. Alternatives for detoxification of diluted-acid lignocellulosic hydrolyzates for use in fermentative processes: A review. Bioresour. Technol. 2004, 93, 1–10. https://doi.org/10.1016/j.biortech.2003.10.005

9 Liu, Z., Fels, M., Dragone, G., Mussatto, S.I. Effects of inhibitory compounds derived from lignocellulosic biomass on the growth of the wild-type and evolved oleaginous yeast Rhodosporidium toruloides. Ind. Crops Prod. 2021, 170, 113799. https://doi.org/10.1016/j.indcrop.2021.113799

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

11 Hemmerich, J., Wiechert, W., Oldiges, M. Automated growth rate determination in high-throughput microbioreactor systems. BMC Res. Notes 2017, 10, 617. https://doi.org/10.1186/s13104-017-2945-6.

12 Buchanan, R.L., Cygnarowicz, M.L. A mathematical approach toward defining and calculating the duration of the lag phase. Food Microbiol. 1990, 7, 237–240. https://doi.org/10.1016/0740-0020(90)90029-H

13 Klyachko, K.A., Schuldiner, S., Neyfakh, A.A. Mutations affecting substrate specificity of the Bacillus subtilis multidrug transporter Bmr. J. Bacteriol. 1997, 179, 2189-2193. https://doi.org/10.1128/jb.179.7.2189-2193.1997

14 Beres, D.L., Hawkins, D.M. Plackett-Burman technique for sensitivity analysis of many-parametered models. Ecol. Model. 2001, 141, 171-183. https://doi.org/10.1016/S0304-3800(01)00271-X

15 Hadi, S.M., Shahabuddin, Rehman, A. Specificity of the interaction of furfural with DNA. Mutat. Res. Lett. 1989, 225, 101–106. https://doi.org/10.1016/0165-7992(89)90125-5

16 Modig, T., Lidén, G., Taherzadeh, M.J. Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. Biochem. J. 2002, 363, 769–776. https://doi.org/10.1042/0264-6021:3630769

17 Zaldivar, J., Martinez, A., Ingram, L.O. Effect of selected aldehydes on the growth and fermentation of ethanologenic Escherichia coli. Biotechnol. Bioeng. 1999, 65, 24–33. https://doi.org/10.1002/(sici)1097-0290(19991005)65:1<24::aid-bit4>3.0.co;2-2

18 Roberto, I.C., Lacis, L.S., Barbosa, M.F.S., Mancilha, I.M. Utilization of sugar cane bagasse hemicellulosic hydrolysate by Pichia stipitis for the production of ethanol. Process Biochem. 1991, 26, 15-21. https://doi.org/10.1016/0032-9592(91)80003-8

19 Zheng, D., Bao, J., Lu, J., Lv, Q. Biodegradation of furfural by Bacillus subtilis strain DS3. J. Environ. Biol. 2015, 36, 727–732. PMID: 26387346.

20 Zhang, Y., Chen, X., Luo, J., Qi, B., Wan, Y. An efficient process for lactic acid production from wheat straw by a newly isolated Bacillus coagulans strain IPE22. Bioresour. Technol. 2014, 158, 396–399. https://doi.org/10.1016/j.biortech.2014.02.128

21 Pereira, J.P.C., Verheijen, P.J.T., Straathof, A.J.J. Growth inhibition of S. cerevisiae, B. subtilis, and E. coli by lignocellulosic and fermentation products. Appl. Microbiol. Biotechnol. 2016, 100, 9069–9080. https://doi.org/10.1007/s00253-016-7642-1

22 Imai, T., Ohno, T. The relationship between viability and intracellular pH in the yeast Saccharomyces cerevisiae. Appl. Environ. Microbiol. 1995, 61, 3604–3608. https://doi.org/10.1128/aem.61.10.3604-3608.1995

23 Pampulha, M.E., Loureiro-Dias, M.C. Activity of glycolytic enzymes of Saccharomyces cerevisiae in the presence of acetic acid. Appl. Microbiol. Biotechnol. 1990, 34, 375–380. https://doi.org/10.1007/BF00170063

24 Speck, E.L., Freese E. Control of metabolite secretion in Bacillus subtilis. J. Gen. Microbiol. 1973, 78, 261-275. http://doi.org/10.1099/00221287-78-2-261

25 van Zyl, C., Prior, B.A., du Preez, J.C. Acetic acid inhibition of d-xylose fermentation by Pichia stipitis. Enzyme Microb. Technol. 1991, 13, 82–86. https://doi.org/10.1016/0141-0229(91)90193-E

26 Felipe, M.G.A, Vieira, D.C., Vitolo, M., Silva, S.S., Roberto, I.C., Manchilha, I.M. Effect of acetic acid on xylose fermentation to xylitol by Candida guilliermondii. J. Basic Microbiol. 1995, 35, 171–177. https://doi.org/10.1002/jobm.3620350309

27 Palmqvist, E., Grage, H., Meinander, N.Q., Hahn-Hägerdal, B. Main and interaction effects of acetic acid, furfural, and p- hydroxybenzoic acid on growth and ethanol productivity of yeasts. Biotechnol. Bioeng. 1999, 63, 46–55. https://doi.org/10.1002/(sici)1097-0290(19990405)63:1<46::aid-bit5>3.0.co;2-j

28 Warth, A.D., Effect of Benzoic Acid on Growth Yield of Yeasts Differing in Their Resistance to Preservatives. Appl. Environ. Microbiol. 1988, 54, 2091–2095, https://doi.org/10.1128/aem.54.8.2091-2095.1988

29 Chipley, J.R. Sodium benzoate and benzoic acid. In: Food Science and Technology, v.145. Marcel Dekker: New York, 2005, pp. 11.

30 Ibraheem, O., Ndimba, B.K. Molecular adaptation mechanisms employed by ethanologenic bacteria in response to lignocellulose-derived inhibitory compounds. Int. J. Biol. Sci. 2013, 9, 598–612. https://doi.org/10.7150/ijbs.6091

31 van der Pol, E.C., Bakker, R.R., Baets, P., Eggink, G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. Appl. Microbiol. Biotechnol. 2014, 98, 9579-9593. https://doi.org/10.1007/s00253-014-6158-9

32 Gu, H., Zhu, Y., Peng, Y., Liang, X., Liu, X., Shao, L., Xu, Y., Xu, Z., Liu, R., Li, J. Physiological mechanism of improved tolerance of Saccharomyces cerevisiae to lignin-derived phenolic acids in lignocellulosic ethanol fermentation by short-term adaptation. Biotechnol. Biofuels 2019, 12, 268. https://doi.org/10.1186/s13068-019-1610-9

33 Cho, D.H., Lee, Y.J., Um, Y. et al. Detoxification of model phenolic compounds in lignocellulosic hydrolysates with peroxidase for butanol production from

Clostridium beijerinckii. Appl Microbiol Biotechnol. 2009, 83, 1035–1043. https://doi.org/10.1007/s00253-009-1925-8

34 Merkl, R., Hrádková, I., Filip, V., Šmidrkal, J. Antimicrobial and antioxidant properties of phenolic acids alkyl esters. Czech J. Food Sci. 2010, 28, 275–279. https://doi.org/10.17221/132/2010-cjfs

35 Herald, P.J., Davidson, P.M. Antibacterial activity of selected hydroxycinnamic acids. J. Food Sci. 1983, 48, 1378–1379. https://doi.org/10.1111/j.1365-2621.1983.tb09243.x

36 Cho, J.-Y., Moon, J.-H., Seong, K.-Y., Park, K.-H. Antimicrobial activity of 4-hydroxybenzoic acid and trans 4-hydroxycinnamic acid isolated and identified from rice hull. Biosci. Biotechnol. Biochem. 1998, 62, 2273–2276. https://doi.org/10.1271/bbb.62.2273

37 Baral, N.R., Shah, A. Microbial inhibitors: formation and effects on acetone-butanol-ethanol fermentation of lignocellulosic biomass. Appl. Microbiol. Biotechnol. 2014, 98, 9151-9172. https://doi.org/10.1007/s00253-014-6106-8

38 Franden, M.A., Pilath, H.M., Mohagheghi, A., Pienkos, P.T., Zhang, M. Inhibition of growth of Zymomonas mobilis by model compounds found in lignocellulosic hydrolysates. Biotechnol. Biofuels, 2013, 6, 99. https://doi.org/10.1186/1754-6834-6-99

3.7 Supplementary Material

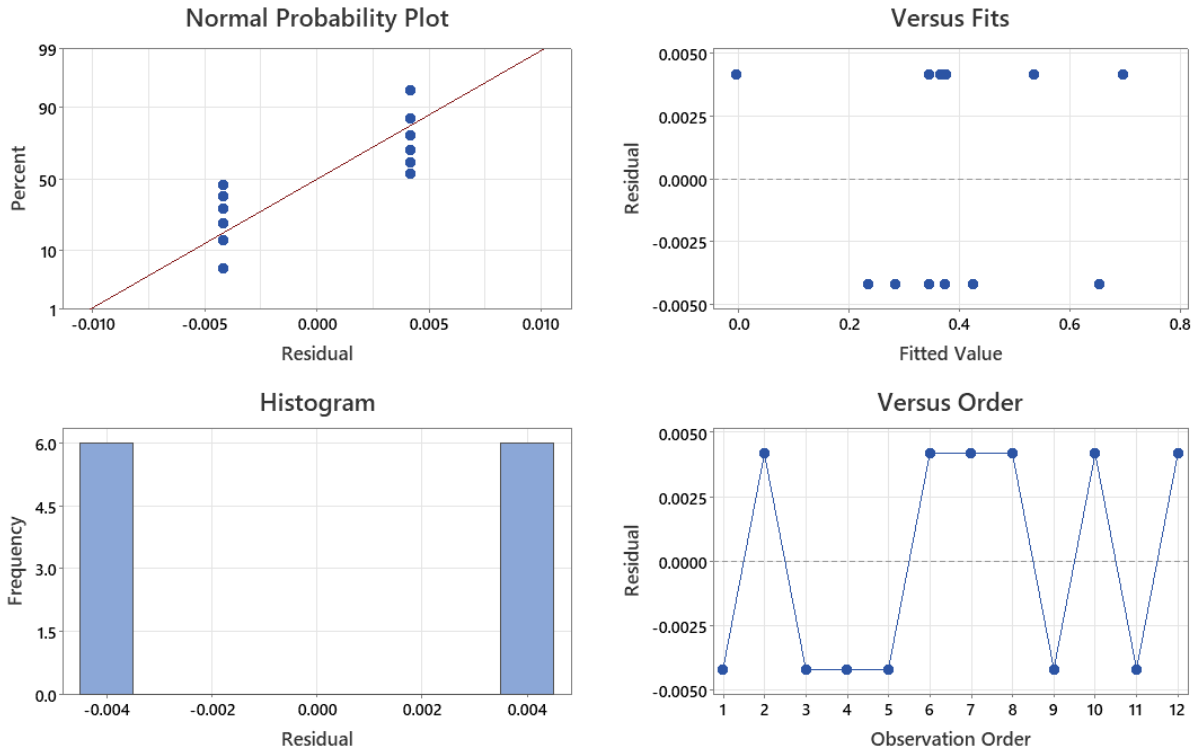


Figure S1. Residual plots from Plackett Burman design.

Table S1. Coded coefficients of Plackett Burman

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		0.38583	0.00417	92.6	0.007	
5-HMF	0.065	0.0325	0.00417	7.8	0.081	1
Furfural	−0.17167	−0.08583	0.00417	−20.60	0.031	1
Acetic acid	−0.01167	−0.00583	0.00417	−1.40	0.395	1
Vanillin	−0.14500	−0.07250	0.00417	−17.40	0.037	1
Vanillic acid	−0.00167	−0.00083	0.00417	−0.20	0.874	1
Benzoic acid	−0.23833	−0.11917	0.00417	−28.60	0.022	1
4HBA	−0.01833	−0.00917	0.00417	−2.20	0.272	1
TFA	0.03833	0.01917	0.00417	4.6	0.136	1
Syringaldehyde	−0.11500	−0.05750	0.00417	−13.80	0.046	1
PCA	−0.02167	−0.01083	0.00417	−2.60	0.234	1

Table S2. Analysis of variance from Plackett Burman.

Source	DF	AdjSS	AdjMS	F-Value	P-Value
Model	10	0.381483	0.038148	183.11	0.057
Linear	10	0.381483	0.038148	183.11	0.057
5HMF	1	0.012675	0.012675	60.84	0.081
Furfural	1	0.088408	0.088408	424.36	0.031
Acetic acid	1	0.000408	0.000408	1.96	0.395
Vanillin	1	0.063075	0.063075	302.76	0.037
Vanillic acid	1	0.000008	0.000008	0.04	0.874
Benzoic acid	1	0.170408	0.170408	817.96	0.022
4HBA	1	0.001008	0.001008	4.84	0.272
TFA	1	0.004408	0.004408	21.16	0.136
Syringaldehyde	1	0.039675	0.039675	190.44	0.046
PCA	1	0.001408	0.001408	6.76	0.234
Error	1	0.000208	0.000208		
Total	11	0.381692			

Chapter 4

ADAPTIVE LABORATORY EVOLUTION OF *BACILLUS SUBTILIS* TO OVERCOME TOXICITY OF LIGNOCELLULOSIC DDGS BIOMASS HYDROLYSATE

Jasper L. S. P. Driessen[1],
Josefin Johnsen[1], Elsayed T. Mohamed[1], Solange I. Mussatto[3],
Adam Feist[1,2], Sheila I. Jensen[1]*, Alex T. Nielsen[1]*

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, 220, 2800 Kongens Lyngby, Denmark

[2] Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA

[3] Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 223, 2800 Kongens Lyngby, Denmark

*** Corresponding authors**
Alex T. Nielsen atn@biosustain.dtu.dk
Sheila I. Jensen shin@biosustain.dtu.dk

This chapter is prepared for submission

Abstract

Low microbial tolerance to toxic compounds formed during biomass pretreatment is a significant challenge for the cost-effective production of bio-based products from lignocellulosic biomass. Engineering tolerance through rational design can be problematic due to a lack of prerequisite mechanistic knowledge. Therefore, we applied adaptive laboratory evolution to obtain tolerant *Bacillus subtilis* strains (BS168, KO7, KO7-S), capable of using DDGS-derived hydrolysate as a fermentation medium. Unlike non-evolved strains, the evolved isolates were able to grow in hydrolysate-based medium, and showed up to 48% of amylase production in a hydrolysate-based medium compared to an optimized and richer expression medium. Whole-genome resequencing data of independently evolved isolates revealed key mutations in both global regulators and specific genes related to tolerance and the employed cultivation conditions. As Bacillus subtilis is a frequently-used chassis organism, this study contributes to the implementation of lignocellulose as a feedstock for industrial fermentation.

4.1 Introduction

The global production volume of bioethanol hit a record of more than 109 billion liters in 2019 and is expected to rise to 134.5 billion liters in 2024 [1]. As a consequence, the production of a major by-product of the bioethanol process, called Dried Distillers Grains with Solubles (DDGS), will rise concomitantly [2–4]. Currently, selling of DDGS as animal feed is of vital importance to the economic viability of the bioethanol industry [3]. However, as the animal feed market is expected to be saturated, there is a growing need to find alternative ways to convert DDGS into high-value products [2,3]. In recent years, there has been an increased focus on the potential use of DDGS as a substrate for microbial fermentation to generate value to the bioethanol production process. The rich nutritional composition of DDGS in terms of carbon, nitrogen and other micronutrients proved to be an ideal starting point for the bio-manufacturing of a variety of products, including organic acids, biofuels and hydrolytic enzymes [5].

Bacillus subtilis (and other closely related species like *Bacillus licheniformis*) is recognized as one of the working horses in both industrial biotechnology and academia for the production of platform chemicals, biopolymers and enzymes [6]. This species is a widespread candidate as a production host due to its robustness in industrial fermentations, well-defined endogenous metabolism, distinct genetic background combined with established and emerging genetic manipulation tools and is generally recognized as safe (GRAS) [6,7].

While DDGS contains a considerable amount of carbohydrates, pretreatment of DDGS is required to fractionate and hydrolyze the lignocellulosic fibers to release the fermentable sugars. A multitude of chemical, physical and biological pretreatment methods and conditions are described in literature [5,8–10]. Often, biomass pretreatment involves the undesirable formation of lignocellulose-derived by-products, which have negative effects on fermentation and lead to a decrease in overall sugar yield. These inhibitory compounds include, amongst others, furan derivatives, organic acids and phenolic compounds [11]. Detoxification of

lignocellulosic hydrolysate is possible, but leads to increased costs and concomitant loss of sugars. The development of microbial strains with increased tolerance has the potential to minimize the degree of detoxification required. Unfortunately, tolerance is a complex trait, involving the coordinated action of hundreds of genes [12]. As biomass hydrolysates contain a plethora of different toxic compounds, rational design is problematic due to the lack of prerequisite knowledge [13,14]. Even if the exact tolerance mechanism can be rationally engineered, it can be challenging to fine-tune the expression of the genes involved to deliver the desired phenotype [15].

In contrast, tolerance adaptive laboratory evolution (TALE) enables strain optimization without a priori knowledge about the genetic changes necessary to increase tolerance towards hydrolysate-associated inhibitory compounds. By serial passaging of cells in conditions with increasing selective pressure, one can select cells, which acquire beneficial mutations for the chosen environment. The use of automated liquid handler systems allows dynamic control of the applied stress during the experiment, so as to maintain a strong selection pressure while not crashing the cultures. Employing ALE to overcome stress-induced growth inhibition by hydrolysate-associated inhibitory compounds is a popular method, as it involves tuning expression of multiple genes to alter complex physiological stress responses. Although ALE has been used to obtain tolerant strains for species like *Saccharomyces cerevisiae*, *Escherichia coli*, *Clostridium thermocellum* and *Corynebacterium glutamicum*, this method scarcely has been applied to yield tolerant *B. subtilis* strains[16–21].

In this study, we applied TALE using hydrolysate-based medium to obtain a tolerant *B. subtilis* strain. Following the initial growth screening, we inserted an amylase expression cassette in the selected evolved isolates to demonstrate that the cells could use DDGS hydrolysate as a fermentation medium to manufacture of bio-based products. By comparing whole-genome-sequencing data of independent parallel replicates starting from the same ancestral strain, it was possible to identify key mutations related to an improved fitness (i.e. increased growth performance) and improved tolerance in DDGS-hydrolysate.

4.2 Material and Methods

4.2.1 Plasmid and Strain Construction

4.2.1.1 Strain Overview

The TALE experiments were performed using *B. subtilis* background strains *BS168, PY79 KO7 and PY79 KO7s* (Table 1) [22]. In both PY79 derivatives, seven deletions have been made (*ΔnprE, ΔaprE, Δepr, Δmpr, ΔnprB, Δvpr, Δbpr*) to prevent produced proteins from being degraded by native extracellular proteases [23]. Additionally, the sporulation gene *sigF* has been deleted in *PY79 KO7s* ("Bacillus Genetic Stock Center").

For each of the three background strains, a xylose-consuming variant was constructed to increase the utilization of C5 sugars present in the DDGS hydrolysate. In these variants, three alterations were made. First, the heterologous xylose isomerase gene *xylA* and the xylulokinase gene *xylB* from *E. coli* and the homologous xylose transport protein *AraE*, were co-overexpressed by using a replicative plasmid (*pHT315:pSCG7-SG46-araE:p43-xylA-xylB* [25]). Second, the genomic native promoter and Shine Dalgarno-sequence of the *araE* gene were exchanged with a promoter and Shine Dalgarno-sequence associated with high expression [26]. Third, the gene encoding for the transcriptional repressor protein of *araE*, called *araR*, was deleted. Hence, six strains were made available, of which five were selected for the TALE experiment, being the background BS168 wild type (wt), BS168 xyl, KO7 wt, KO7-S wt and KO7-S xyl.

During strain construction, bacteria were routinely grown in LB media at 37 °C at 250 RPM or on LB agar plates at 37 °C. When required, antibiotics were supplemented to the media (100 µg mL^{-1} ampicillin (Amp) for *E. coli* strains, and 5 µg mL^{-1} chloramphenicol (Cam), 5 µg mL^{-1} kanamycin (Kan), 10 µg mL^{-1} erythromycin (Ery), and 100 µg mL^{-1} spectinomycin (Spc) for *B. subtilis* strains).

4.2.1.2 User Cloning and ProUSER Vectors

The pJOE8999 and PHT315 plasmids were made by using the USER cloning system [27]. All DNA fragments, pJOE8999 and PHT315 backbones were amplified by PCR using Phusion U Hot Start DNA Polymerase, which is able to amplify DNA fragments containing uracil. The ProUSER plasmids were digested by AsiSI and Nt.BbvCI restriction enzymes (New England BioLabs, United States) as described in [28]. The specific digestion of the ProUSER plasmid by AsiSI and Nt.BbvCI, creates 6 and 8 bp single-stranded DNA overhangs, which enabled USER cloning using the nicked backbones. DNA fragments and plasmid backbones were gel purified (NucleoSpin PCR clean-up gel extraction kit, Macherey-Nagel) and used for USER cloning as previously described [28].

4.2.1.3 Transformation

E. coli transformations were performed using cloned vectors as described in [29] to yield the required plasmid material. *B. subtilis* strains (BS168, KO7 and KO7-S) were either transformed by natural or induced competence. The use of natural competence was carried out as described previously [30], with the exception that histidine was not added to the SM1 and SM2 media and cells were recovered for at least 2 hours prior to exposing them to antibiotics. To induce competence, a mannitol-inducible *comKS*-cassette was inserted downstream of the *glms* locus of *B. subtilis* (unpublished material) and transformation was carried out as previously described [31].

4.2.1.4 Start Strain Construction

Deletions of the *ΔaraR* gene and substitutions of the promoter region ($P_{SGC14\text{-}SG46}$-*araE*) were made using a shuttle vector (pJOE8999) [32]. The vector carries a mannose-inducible cas9 gene system, capable of introducing a double-strand break at its target site. Repair by homologous recombination, using an engineered template containing both upstream and downstream flanking regions of the target site can be used to delete or substitute a desired DNA sequence. As such, both upstream and downstream flanking regions of the *araR* gene and the promoter region of the native *araE* gene were amplified using compatible USER primers and cloned into an amplified pJOE8999 backbone by using USER cloning as described above (pJD2, pJD5, **Table 2**). The replicative plasmid pJD18 was constructed to co-overexpress of

xylA, xylB and *araE* and transformed into the xylose-consuming variants (Table 2).

Post evolution, a mannitol-inducible *comKS*-cassette was inserted downstream of the *glms* locus in both the start strains and selected evolved isolates by transformation of *pSIJ1005* (unpublished material). Subsequently, an amylase expression cassette (*amyQ, B. amyloliquefaciens)* was integrated into the same *glmS* region by transformation of *pJD21*, replacing the comKs-cassette. As *pJD21* contains the strong constitutive promoter P_{3P} and integration in the *glmS* region has been found to be associated with high expression levels, strong expression of *amyQ* was expected (Table 2) [33]. The genomic deletions/substitutions and plasmids sequences were verified using a Mix2Seq Sanger sequencing kit (Eurofins Genomics, Luxembourg).

Table 1 Overview of constructed strains used in this study

Strain	Genotype	Source
E. coli DH5α	ϕ80dlacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 56 hsdR17 (rk− mk+) supE44 thi-1 gyrA relA1	Lab collection
B. subtilis 168 WT	*trpC2*	DSM23778
BS168	*B. subtilis* 168 WT, pHT315	This study
BS168 xyl	BS168::*ΔaraR*::$P_{SGC14\text{-}SG46}$-*araE*, pJD18	This study
BS168 *amyQ*	BS168, *glms*::P_{3P}-*amyQ*::cm	This study
BS168 xyl *amyQ*	BS168 xyl, *glms*::P_{3P}-*amyQ*::cm	This study
B. subtilis KO7 WT	*ΔnprE, ΔaprE, Δepr, Δmpr, ΔnprB, Δvpr, Δbpr*	[23]
KO7	*B. subtilis* KO7 WT, pJD13	This study
KO7 xyl	KO7::*ΔaraR*:: $P_{SGC14\text{-}SG46}$-*araE*, pJD18	This study
KO7 *amyQ*	KO7, *glms*::P_{3P}-*amyQ*::cm	This study
KO7 xyl *amyQ*	KO7 xyl, *glms*::P_{3P}-*amyQ*::cm	This study
B. subtilis KO7-S	*ΔnprE, ΔaprE, Δepr, Δmpr, ΔnprB, Δvpr, Δbpr, ΔsigF*	[23]
KO7-S	*B. subtilis* KO7-S, pJD13	This study
KO7-S xyl	KO7-S::*ΔaraR*:: $P_{SGC14\text{-}SG46}$-*araE*, pJD18	This study
KO7-S *amyQ*	KO7-S, *glms*::P_{3P}-*amyQ*::cm	This study
KO7-S xyl *amyQ*	KO7-S xyl, *glms*::P_{3P}-*amyQ*::cm	This study

Table 2 Overview of constructed plasmids used in this study

Plasmid	Description	Source
pJOE8999	Plasmid containing P_{man}-cas9, pUC ori *E. coli*, PE194^{ts} ori *B. subtilis*, and kanamycin resistance marker	[32]
pJD2	pJOE8999 derivative containing *araR* homology regions	This study
pJD5	pJOE8999- derivative containing P_{araE} homology regions for indel $P_{SGC14\text{-}SG46}$	This study [26]
pHT315	Replicative plasmid containing pHT315 ori, ColE1 ori, ampicillin resistance marker, erythromycin resistance marker	[34]
pJD18	pHT315 derivative containing $P_{SCG7\text{-}SG46}$-*araE* and P_{43}-*xylA*-*xylB*	This study
pProUSER13C1B	ProUSER plasmid containing P_{3P}, *glmS* homology regions, and chloramphenicol resistance marker	[28]
pJD21	pProUSER13C1B derivative for integrating P_{3P}-*amyQ* downstream of *glmS*	This study
pSIJ1005	Unpublished material	Unpublished

4.2.2 DDGS Pretreatment and Enzymatic Hydrolysis

DDGS was kindly provided by United Wisconsin Grain Producers (Friesland, WI, USA). Prior to steam explosion, the DDGS biomass was soaked in 0.5% w/w sulfuric acid for 20 minutes. The DDGS slurry (solid/liquid ratio of 20% w/w) was filtered through cotton fabric using a pressure press (Fischer HP25-M). Wet solids were pretreated in batches of approximately 2.2 kg in a 10 L steam explosion reactor (190 °C, 5 minutes incubation, 12.5 bar, Knislinge Mekaniska Verkstad AB designed by Process- & Industriteknik AB, Lund University, Sweden) as described by Palmqvist et al. [35]. Batches of pretreated slurry were pooled and subsequently separated by solid-liquid filtration. Wet slurry was stored at -20 °C prior to being subjected to enzymatic hydrolysis.

The pretreated DDGS biomass was diluted to 25% solid content (w/V) by adding 0.1 M sodium citrate buffer (final pH 5.1). The enzymatic hydrolysis was performed using an enzyme loading of 25 FPU/ g cellulose (Optimash® F200, DuPont Nutrition and Bioscience). Suspended solutions were incubated in glass bottles (VWR) for 48 h, placed on a horizontal Bottle Roller system housed in an incubator (Thermo ScientificTM, 80 rpm, 50 °C). After the enzymatic hydrolysis, the sugar-rich liquid was filter-sterilized and used as a carbon and nitrogen source in subsequent experiments. Glucose, xylose, arabinose, acetic acid, 5-HMF and furfural concentrations were quantified using a HPLC system (RI, Dionex Ultimate 3000, Germany) equipped with an Aminex HPX-87H column (300 × 7.8 mm, Bio-Rad, USA) eluted with 0.005 M H_2SO_4 at 50 °C for 50 min. with a flow rate of 0.6 mL/min. Total phenolic compounds were determined by a colorimetric method using a Folin-Ciocalteu reagent [36]. The composition of DDGS-based hydrolysate can be found in Table 3.

4.2.3 TALE and ALE Experiments

During the TALE experiment, a modified M9-based medium (called M9extra) was supplemented with increasing amounts of DDGS hydrolysate. The M9extra medium was made as described previously and contained the following: monosaccharides as carbon source (approximately the same as DDGS-hydrolysate: 30 g L^{-1} glucose, 10 g L^{-1} xylose, 5 g L^{-1} arabinose), M9-salts (12.8 g L^{-1} $Na_2HPO_4 \cdot 7H_2O$, 3 g L^{-1} KH_2PO_4 , 0.5 g L^{-1} NaCl, 1 g L^{-1} NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$), supplemented with 60 µM $FeCl_3$, a trace element solution (1.25 µM $MnCl_2 \cdot 4H_2O$, 0.21 µM $CoCl_2 \cdot 6H_2O$, 0.85 µM $ZnSO_4 \cdot 7H_2O$, 0.06 µM $CuCl_2 \cdot 2H_2O$, 0,08 µM H_3Bo_3, 0.105 µM $NiCl_2 \cdot 6H_2O$, 0.125 µM $NaMoO_4 \cdot 2H_2O$), and 50 mg L^{-1} L-tryptophan. In addition, media were supplemented with antibiotics (10 µg mL^{-1} erythromycin) [28]. The added hydrolysate was supplemented with the same concentration of components present in the M9 extra medium ($FeCl_3$, trace element solution, tryptophan, antibiotics) to keep the concentration constant throughout the TALE for more targeted selection pressure.

The TALE experiments were started with four independent biological replicates for each of the five starting strains (Table S4). In parallel, 20 bacterial cultures were passaged during the late exponential growth phase for an average of 330 generations using an automated liquid handler platform (15 mL 37 °C, 1200 rpm magnetic stirrer) as described previously [37]. OD_{600} measurements were performed periodically to determine growth rates and time of passaging of all ALEs (Tecan Sunrise plate reader). Each ALE flask, i.e. batch, was passed into a new fresh flask around an OD_{600} of 0.9 (equal to 3.86 for a benchtop spectrophotometer with a 1 cm light path) at the late exponential phase. Once a set growth rate was reached, DDGS hydrolysate supplementation to the M9extra media was increased by steps of 5% (w/V). Periodically, aliquots of samples were frozen in 25% glycerol solution and stored at −80°C for analysis. At the end of the evolution, strains were able to grow reproducibly at 95-100% hydrolysate supplemented media. In addition, a control ALE was performed with M9 extra medium to identify mutations related to media- or cultivation-specific adaptation [38,39].

4.2.4 Growth Screening Experiment

Growth characterization was performed for both the start strains and evolved strains. After the TALE experiment, start, mid- and endpoint populations were inoculated from glycerol stocks in 4 mL of LB medium (supplemented with 10 µg mL^{-1} erythromycin) and grown overnight (16h, 37 °C) in 24 deep-well plates. In parallel, endpoint populations were streaked on LB agar plates (supplemented with 10 µg mL^{-1} erythromycin) and grown overnight (16h, 37 °C). Subsequently, 5 µL of overnight culture (end- and midpoint populations) and 21 single colonies (endpoint clones) from each starting strain (20 in total) were inoculated in 800 µL of LB medium (supplemented with 10 µg mL^{-1} erythromycin) and grown for 9 hours into 96 microtiter deep-well plates. Next, 5 µL of the cultures were transferred in 800 µL of 50% (V/V) M9-extra/hydrolysate and grown overnight (16h, 37 °C) in 96 plates to increase the selection pressure during the pre-culture progressively. The growth screening experiments were started by inoculating 5 µL of overnight pre-culture in 295 µL of 100% hydrolysate-based medium. Growth performance was assessed by incubation in a growth profiler 960 for 48 hours (250 rpm, 37 °C, CR1496dg, EnzyScreen BV, Leiden, The Netherlands).

4.2.5 Amylase Production Screening Experiment

Selected endpoint clones were inoculated from frozen glycerol stocks on LB agar plates (supplemented with 10 µg mL^{-1} erythromycin) and grown overnight (16h, 37 °C). Cell pre-cultures were prepared as described in section 1.4. For screening experiments, cultures were started by inoculation of 5 µL of overnight pre-culture in 195 µL of 100% hydrolysate-based medium. Cells were incubated in 96 well microtiter plates, using a microplate spectrophotometer (BioTek ELx808). After 24 hours, cultures were centrifuged at 6000 g for 5 minutes at 4 °C, the pellet was discarded, and the supernatant was used for further analysis.

Additionally, amylase production experiments were performed in Axygen 24-well deep well plate (Corning Life Science, Corning, New York, USA) for both hydrolysate-based medium and a rich synthetic media, called Cal18-2. Cal18-2 media [40] contained the following: (40 g L^{-1} yeast extract (LP0021B, Thermofisher Scientific), 1.3 g L^{-1} $MgSO_4 \cdot 7H_2O$, 50 g L^{-1} maltodextrin (DE 13-17 (Sigma-Aldrich, Saint Louis, MO, USA)), 20 g L^{-1} $NaH_2PO_4 \cdot 2H_2O$, 6.7 mL L^{-1} Na_2MoO_4 stock solution (2.0 g L^{-1}), 6.7 mL L^{-1} trace metal solution (consisting of 4.48 g L^{-1} $MnSO_4 \cdot H_2O$, 3.33 g L^{-1} $FeCl_3 \cdot 6H_2O$, 0.625 g L^{-1} $CuSO_4 \cdot 5H_2O$, 7.12 g L^{-1} $ZnSO_4 \cdot 7H_2O$), and 100 µL L^{-1} Pluronic L-61 (Sigma-Aldrich, Saint Louis, MO, USA), adjusted to pH = 6). Strains were streaked on LB agar plates (supplemented with 10 µg mL^{-1} erythromycin) and grown overnight (16h, 37 °C). Cell pre-cultures were prepared as described above, with the exception that Cal 18-2 pre-cultures were grown in Cal 18-2 medium instead of 50% hydrolysate-based medium. Amylase production experiments started at an initial OD_{600} of 0.1, and after 72 h, cultures were centrifuged at 6000 g for 5 minutes at 4 °C, the pellet was discarded, and the supernatant was used for further analysis.

The amylase assay was adapted from Xiao et al. [41]. Culture supernatants were diluted in 100 mM phosphate buffer (pH = 5.9) solution. Subsequently, 10 µL of the diluted supernatant was mixed with 40 µL starch solution (2g L^{-1} in 100 mM phosphate buffer (pH = 5.9). The samples were incubated at 65 °C for 10 min and stopped by adding 50 µL of 1 M HCl. The reactions were mixed with 50 µL iodine solution consisting of 5mM I_2 and 50 mM KI. The absorbance values were measured at 580 nm.

4.2.6 Whole-genome Re-sequencing and Analysis

For each TALE and ALE experiment, start clones, endpoint populations and selected evolved isolates were selected and prepared for whole-genome re-sequencing. Genomic DNA was extracted from overnight cultures (grown in LB medium) using the MasterPure Gram Positive DNA Purification Kit (Lucigen). Quality was assessed by evaluating Abs_{260nm}/Abs_{280nm} using a Nanodrop (Thermo Fisher scientific, USA). DNA concentration was measured using a Qubit ds-DNA broad range assay (Thermo Fisher scientific, USA), and paired-end sequencing libraries were generated using the Illumina 300 cycle (150 bp x 2) kit (San Diego, CA, USA). Sequencing was performed on an Illumina NextSeq 500/550 system (Illumina, USA). The average coverage for each sample was over 60. Genome sequencing reads were analyzed using the in-house mutations analysis pipeline called "ALE mut pipeline" to generate lists of mutations for each evolved strain [42]. The reference strain for this analysis was BS168 with the GenBank accession number NC_000964 and PY79 with the GenBank accession number CP006881.1. The variant calling data is available through the public ALEdb platform at http://aledb.org.

4.3 Results and Discussion

In this study, TALE experiments were performed to generate *B. subtilis* strains, more tolerant to the toxic compounds present in DDGS-hydrolysate, and as such, contribute to the development of manufacturing bio-based products using second-generation carbon (**Figure 1**). The TALE experiment was performed using *B. subtilis* background strains BS168, KO7 and KO7-S. Seven deletions have been made in the KO7 and KO7-S background strains (*ΔnprE, ΔaprE, Δepr, Δmpr, ΔnprB, Δvpr, Δbpr*), to prevent produced proteins from being degraded by native extracellular proteases 23. Additionally, the sporulation gene *sigF* has been deleted in PY79 KO7s. For each of the three background strains, a xylose-consuming variant was constructed to increase the utilization of C5 sugars present in the DDGS hydrolysate. Of the six available strains, five were selected for the TALE experiment, being BS168 wt, BS168 xyl, KO7 wt, KO7-S wt and KO7-S xyl.

Prior to the TALE experiment, steam explosion pretreatment and enzymatic hydrolysis of the DDGS biomass were performed to produce the DDGS-hydrolysate (**4.3.1**). Subsequently, the inhibitory effect of compounds present in the pretreated hydrolysate medium was assessed (**4.3.2**). The TALE experiments were executed with four independent biological replicates for each of the five start strains (**4.3.3**). The characterization of the evolved strains was done by a growth screening (**4.3.4**) and a protein production screening (**4.3.5**). By comparing whole-genome-sequencing data of independent replicates, we aimed to identify mutations related to the improved phenotypes (**4.3.6**).

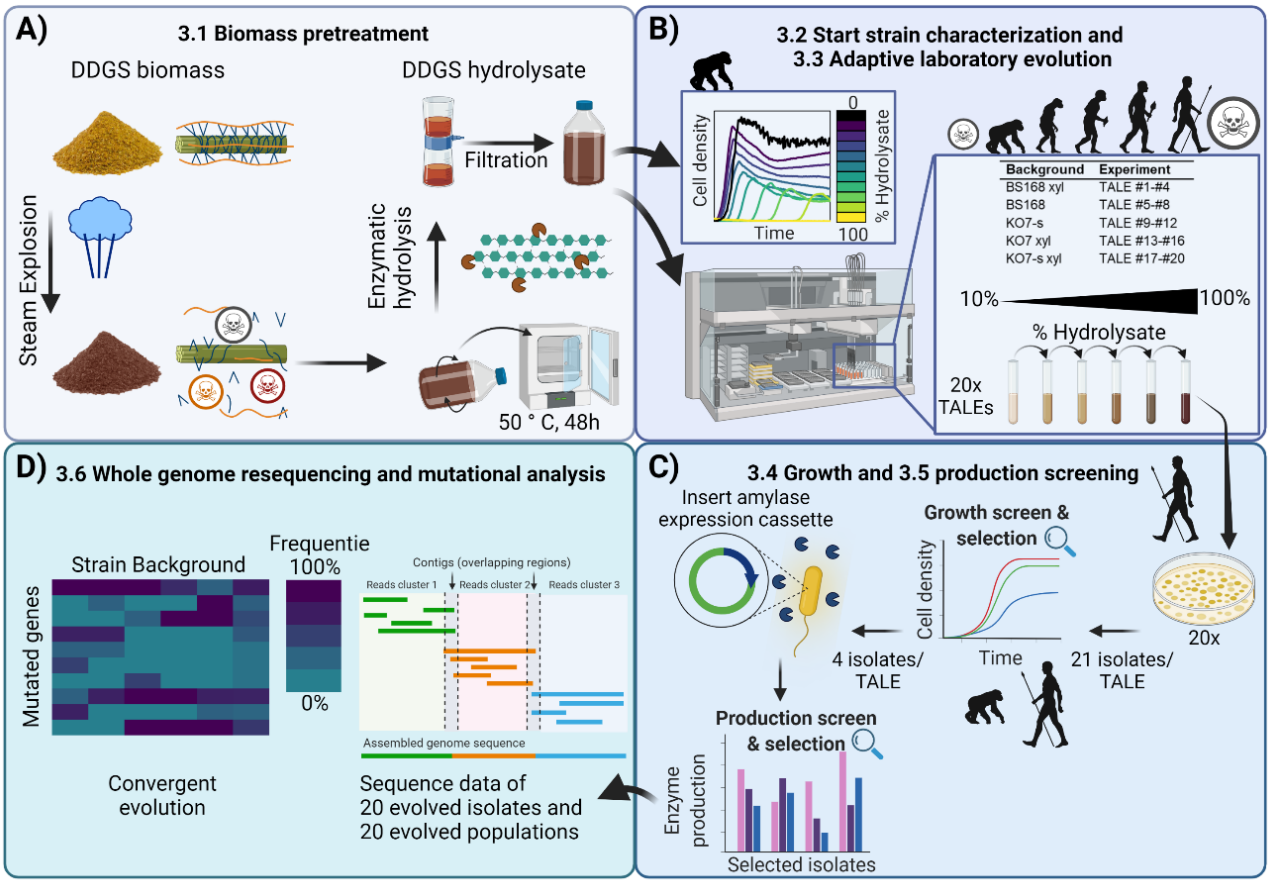


Figure 1 The overall workflow of the study. A) DDGS biomass was subjected to a steam explosion pretreatment, after which polymeric fibers were converted to monomers by enzymatic hydrolysis. B) DDGS-hydrolysate was used to perform an initial characterization of the tolerance of the starting strain and to perform the TALE experiments. C) Evolved populations were grown on LB plates and evolved isolates were picked for the initial growth screening. Selected isolates were inserted with an amylase expression cassette and a production screening was performed. D) The best producing evolved isolates and evolved end populations were sent for whole-genome resequencing and mutational analysis was performed.

4.3.1 DDGS-based Hydrolysate

Steam explosion pretreatment and enzymatic hydrolysis yielded a DDGS-based hydrolysate containing sugars, protein and inhibitory compounds, including furan derivatives, organic acid and phenolic compounds (Table 3). The presence of both sugars and protein is one of the reasons for the increased interest in using DDGS as a starting point for microbial fermentation [5]. Steam explosion is one of the most widely studied pretreatment strategies in both lab-scale and different pilot plants and is considered a cost-effective method near commercialization [43–45]. Nonetheless, the formation of by-products derived from lignocellulosic biomass during the process is confirmed in our experiments and poses a well-known hurdle, which needs to be overcome.

Table 3 Composition of hydrolysate-based medium

Compound	Concentration (g L^{-1})
Glucose	29.1
Xylose	9.9
Arabinose	5.1
Protein	10.5
5-HMF	0.3
Furfural	0.9
Acetic acid	2.0
Phenolic compounds	3.3

4.3.2 Characterization of Start Strains

All start strains showed gradually impaired growth performance with increasing hydrolysate supplementation (Figure 2). It is noteworthy that cells start to grow earlier at 10% hydrolysate supplementation (V/V) compared to the control, which could indicate some nutrients present in the hydrolysate medium that are beneficial to the cells. Overall, it is clear that all strains do not cope with high amounts of hydrolysate supplementation, confirming the need for strains with improved tolerance. These results are in line with the previous chapter, that showed that compounds commonly present in lignocellulosic hydrolysate have an inhibitory effect on *B. subtilis* [14].

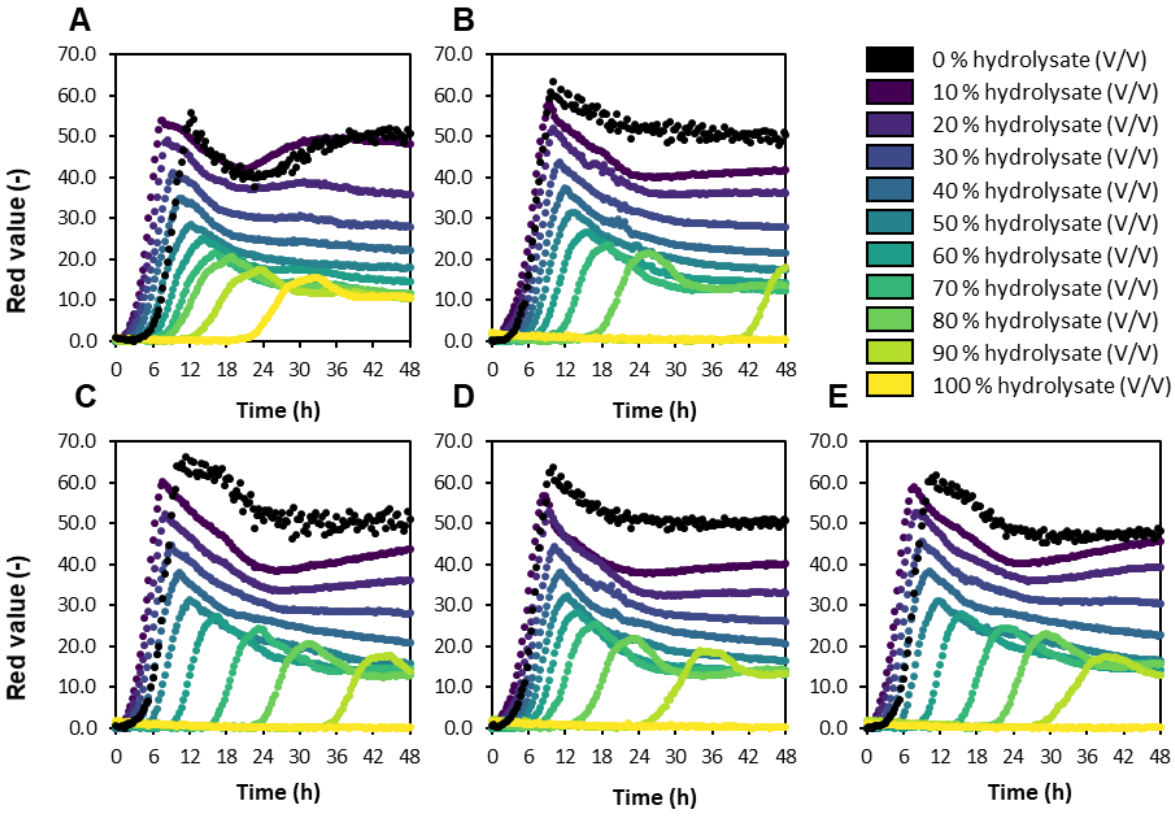


Figure 2 Growth performance of A) BS168, B) BS168 xyl, C) KO7s, D) KO7 xyl, and E) KO7s xyl in M9extra minimal media supplemented with increasing amounts of DDGS-hydrolysate. Lines show the mean growth curves of 3 biological replicates, the average standard deviation of all data points is equal to 1.01.

4.3.3 Fitness Trajectory During TALE and ALE Experiments

The TALE experiments successfully yielded strains with increased tolerance to lignocellulosic hydrolysate (Figure 3). During the experiment, the hydrolysate concentration was increased stepwise for 28 days, corresponding to an average of 330 generations of evolution. Cultures were passed into a new fresh flask around an OD_{600} of 0.9 nm (equal to 3.86 for benchtop reader) at the late exponential phase. During the TALE experiments, the population growth rate fluctuated in response to increasing hydrolysate supplementation. These fluctuations have been previously described in TALE experiments as restorative shifts [46]. General stress responses will be induced upon exposure to higher amounts of hydrolysate, reallocating resources from cell growth to metabolic detoxification of inhibitors by means of differential gene expression. Over time, adaptive mutations arise that are specific for present toxic compounds and allow global gene expression to return to its pre-perturbed state. As the specific tolerance responses should be more energy-efficient than global stress responses, more resources and energy are available for cell growth [12]. Evolved populations were able to grow reproducibly at 95-100% hydrolysate supplemented media by the end of the experiment. Cryogenic stock samples were made of both midpoint evolution populations (60-65% hydrolysate) and the endpoint populations (95-100% hydrolysate).

In addition to the TALE experiments, an evolution experiment employing a constant condition was performed (ALE) with BS168 xyl and KO7-S xyl backgrounds (Figure 3FG). Cells were grown in M9extra minimal media supplemented with the same glucose, xylose and arabinose concentrations as the TALE experiment to identify adaptive mutations related to media components and/or cultivation conditions. While KO7s xyl started the experiment at a lower growth rate than the BS168 xyl background, both had similar final growth rates at the end of the experiment.

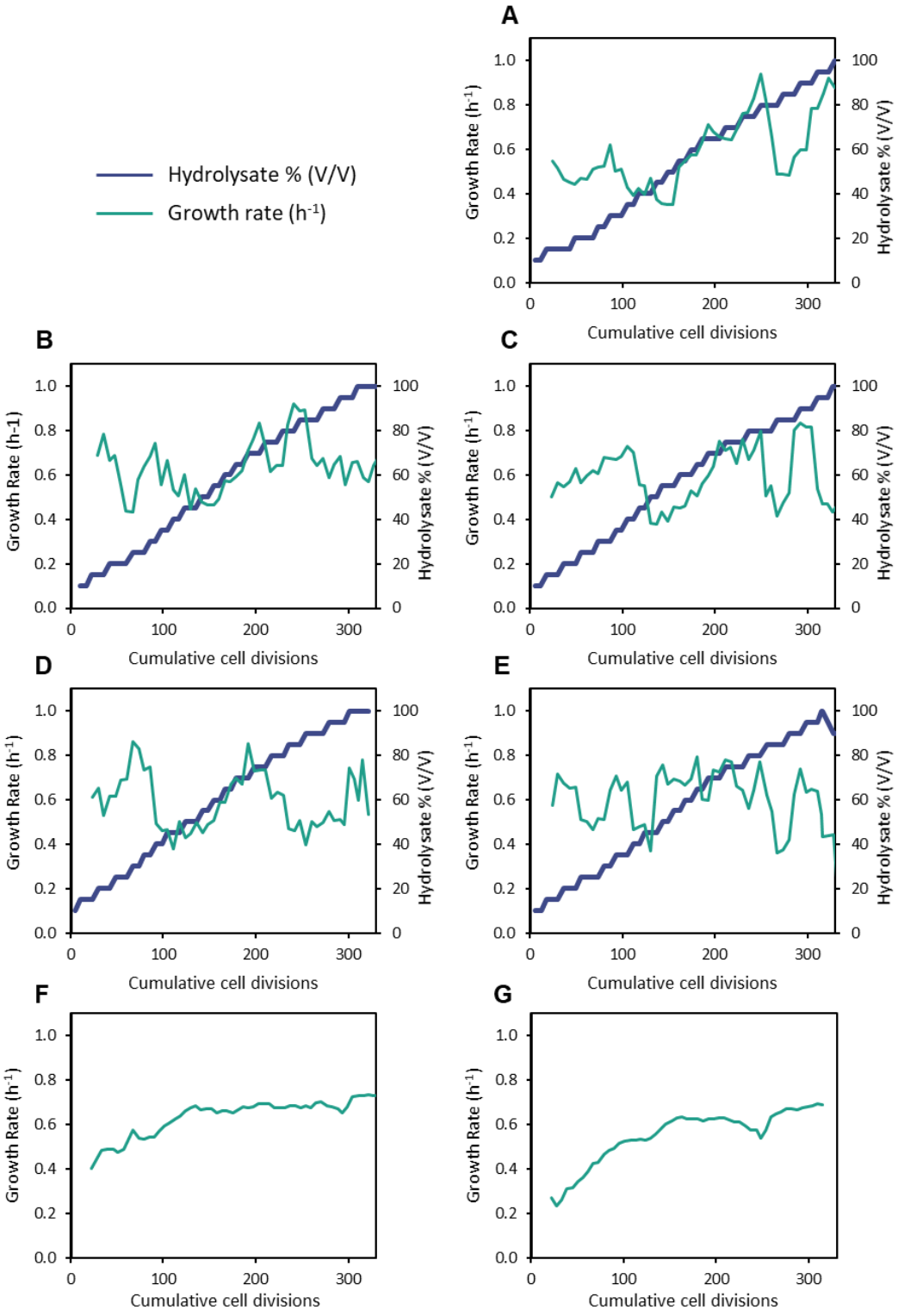


Figure 3 Population growth rate and hydrolysate supplementation over the course of representative TALE experiments of A) TALE #5 BS168, B) TALE #1 BS168 xyl, C) TALE #9 KO7s, D) TALE #13 KO7, E) TALE #17 KO7s xyl, F) ALE #21 BS168 xyl and G) ALE #26 KO7s xyl. Plots for the remaining replicates are shown in Figure S8 and Figure S9.

4.3.4 Growth Screening

During the initial screen in a 100% hydrolysate-based medium, the growth of 21 isolates of each evolved population was qualitatively assessed to select tolerant strains for further production experiments (Figure 4). For all 20 TALE experiments, growth in 100% could be reproduced for multiple end-isolates, while no growth was observed in any of the start strains. Endpoint populations of TALE experiments #9, #11-18, and #20 did not show growth, which most likely indicates the tolerant evolved isolates were not represented or lost during the preculture of the evolved population (Figure S9). While midpoint populations of the BS168 strain showed inferior growth compared to the endpoint populations/isolates, the midpoint population of the KO7/KO7-S strains did not show any growth at all. Up to 4 isolates of each of the 20 end populations were selected for further screening based on qualitative assessment .

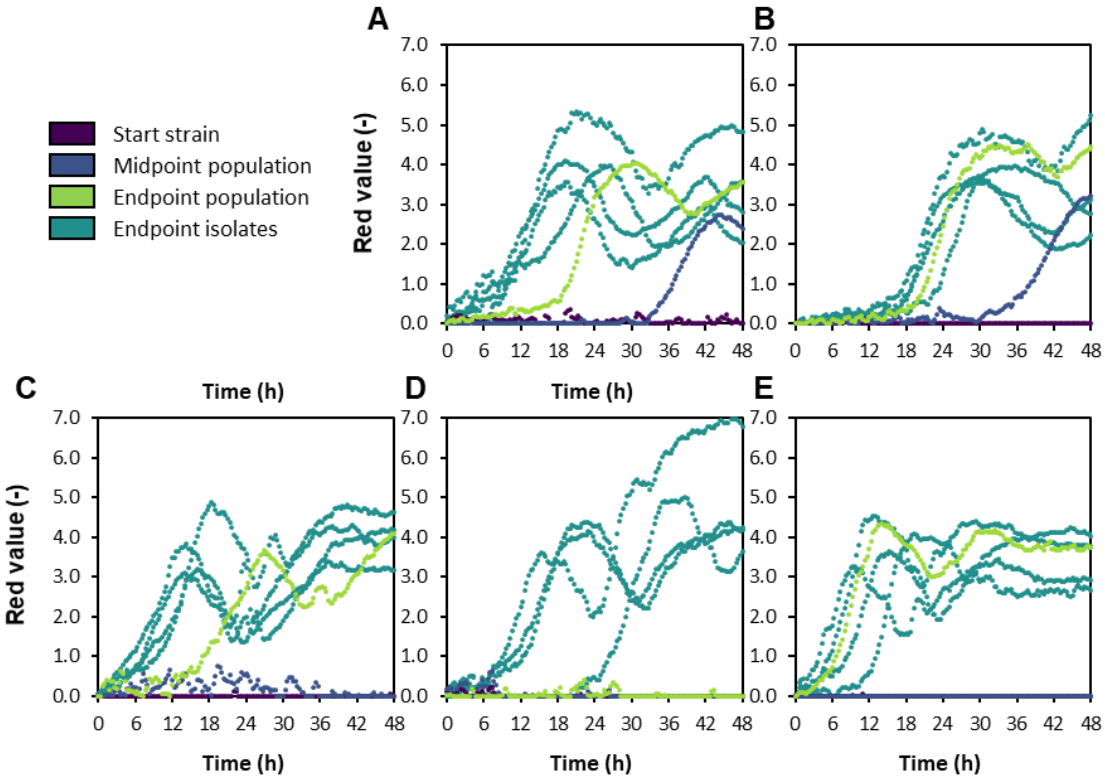


Figure 4 Growth performance of start strains, midpoint populations, endpoint populations and evolved isolates of representative ALE experiments. Graphs show 1 of the 4 replicates for A) BS168, B) BS168 xyl, C) KO7s, D) KO7, and E) KO7s xyl backgrounds grown in in 100% hydrolysate-based medium.

4.3.5 Production Screening

As a proof of the principle that evolved strains are not only able to grow, but also capable of producing protein using second-generation carbon sources, an amylase expression cassette was inserted in the evolved isolates. Of the selected clones, 60 were successfully transformed, representing all TALE experiments except for one replicate (TALE #4, BS168 xyl). Subsequently, protein production was evaluated based on the amylase activity of the culture supernatant in both a hydrolysate-based medium and a rich synthetic media (Cal18-2).

Most of the selected evolved strains showed amylase production in a 100% hydrolysate-based medium (**Figure 5** A2). For the evolved BS168 strains, 4 out of 23 cultures grew but did not show amylase activity in the culture supernatant. For the KO7/KO7-S background, only 1 out of the 37 evolved isolates did not show any amylase activity (TALE#12 I11). As most of the extracellular proteases are knocked out for the KO7/KO7-S background, the amylase-expression cassette might have been mutated, hampering production in some of the isolates. When assessing the averaged amylase activities for the different backgrounds (**Figure 5** A1), two significant differences were observed. First of all, the averaged amylase activity of evolved strains of the KO7-S background were higher ($p<0.05$) than those of the BS168 background. Again, extracellular proteases present in the BS168 background could degrade part of the produced amylase and likely play a key role. Second, the evolved strain variants which were constructed with the aim to increase the consumption of C5 sugars (BS168 xyl, KO7-S xyl) show a higher average amylase activity compared to the wild type strains. As arabinose and xylose represent a third of the monosaccharides present in the DDGS-hydrolysate, increasing C5 consumption could lead to more available resources for protein production.

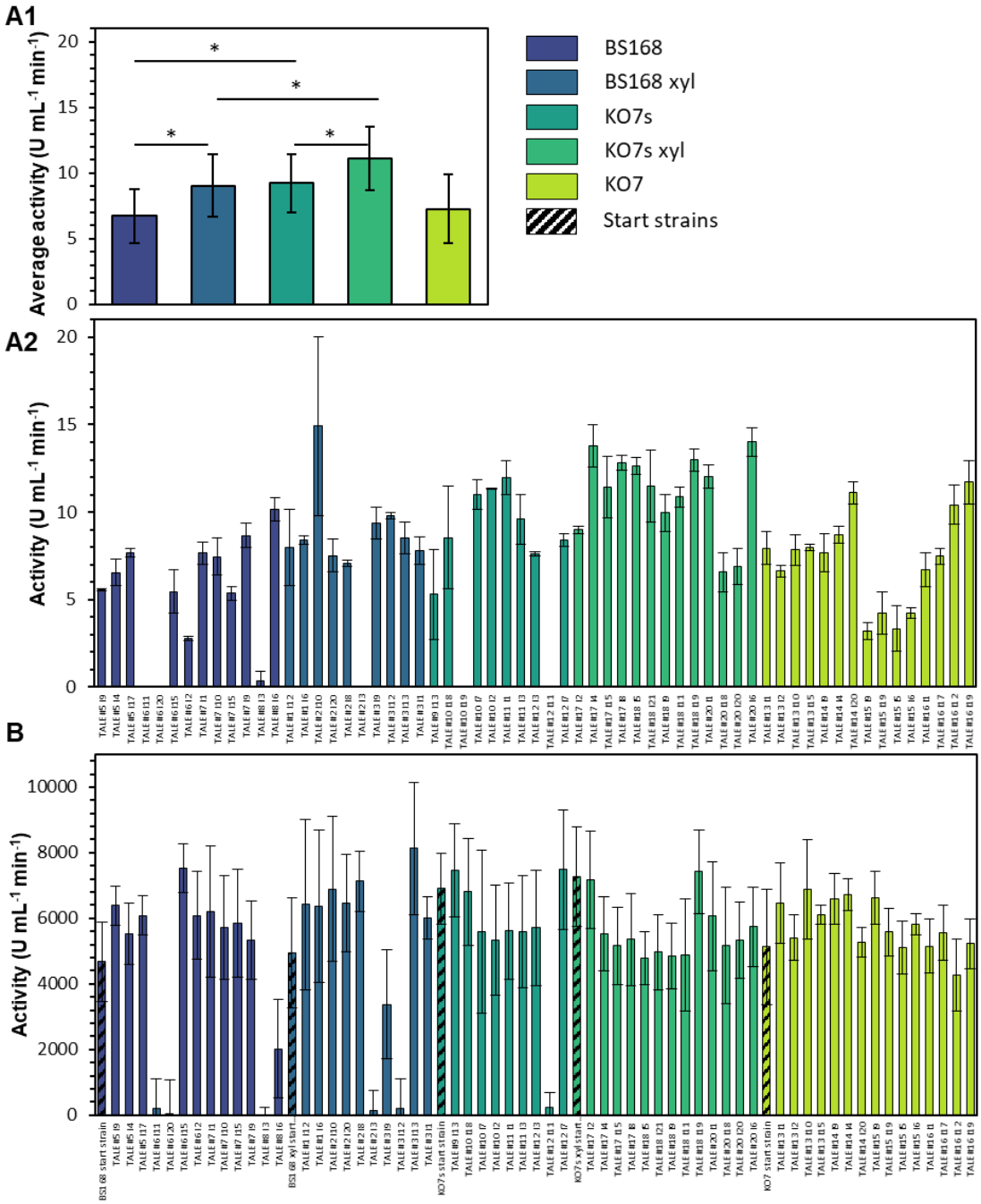


Figure 5 Amylase production of evolved isolates measured by amylase activity of culture supernatant in A2) 100% hydrolysate-based medium in 96 well plate-format after 24 h and B) Cal18-2 media in a 24-well plate format after 72 h. A1) show average amylase activities for the different strain backgrounds of A2. Evolved isolates which did not show any production were excluded from the average. * $p<0,05$

In addition, start strains and evolved isolates were grown in non-toxic, more optimal conditions for protein expression to verify that start strains were actually capable of producing amylase (**Figure 5B**). Apart from the fact that the overall amylase activity is higher as a result of the higher oxygen transfer, more prolonged incubation and a non-toxic, richer medium, start strains are in the same range compared to their evolved descendants.

Lastly, evolved isolates representing each strain background were used for a production experiment in 24-well deep well plates containing 100% hydrolysate medium. Surprisingly, cells showed up to 48% amylase productivity in DDGS hydrolysate compared to Cal 18-2 medium (**Figure S11**). Considering that Cal 18-2 is optimized for protein expression and contains 268% of protein and 116% of sugars compared to the DDGS hydrolysate, it is very promising to find that amylase production is in the same range.

4.3.6 Whole-genome Re-sequencing

To determine the genetic make-up of the improved phenotypes of each TALE experiment, start strains and 20 evolved isolates TALE isolates showing the highest amylase productivity and eight evolved ALE control isolates were sent for whole-genome resequencing (**Table S4**). Start strain-specific mutations were excluded from the analysis. During the quality assessment, isolate TALE#14 I4 was excluded from further analysis due to sequencing issues. The remaining 19 evolved isolates had an average number of mutations of eight with a standard deviation of two. Functional annotations of all mutated genes or genetic regions are summarized based on GO terms in a Sankey diagram (**Figure 6**[47]). Almost two-third of the mutated genes are involved in stress response mechanisms, metabolism, transcriptional regulation, sporulation, and cell motility, highlighting their importance for the improved phenotype.

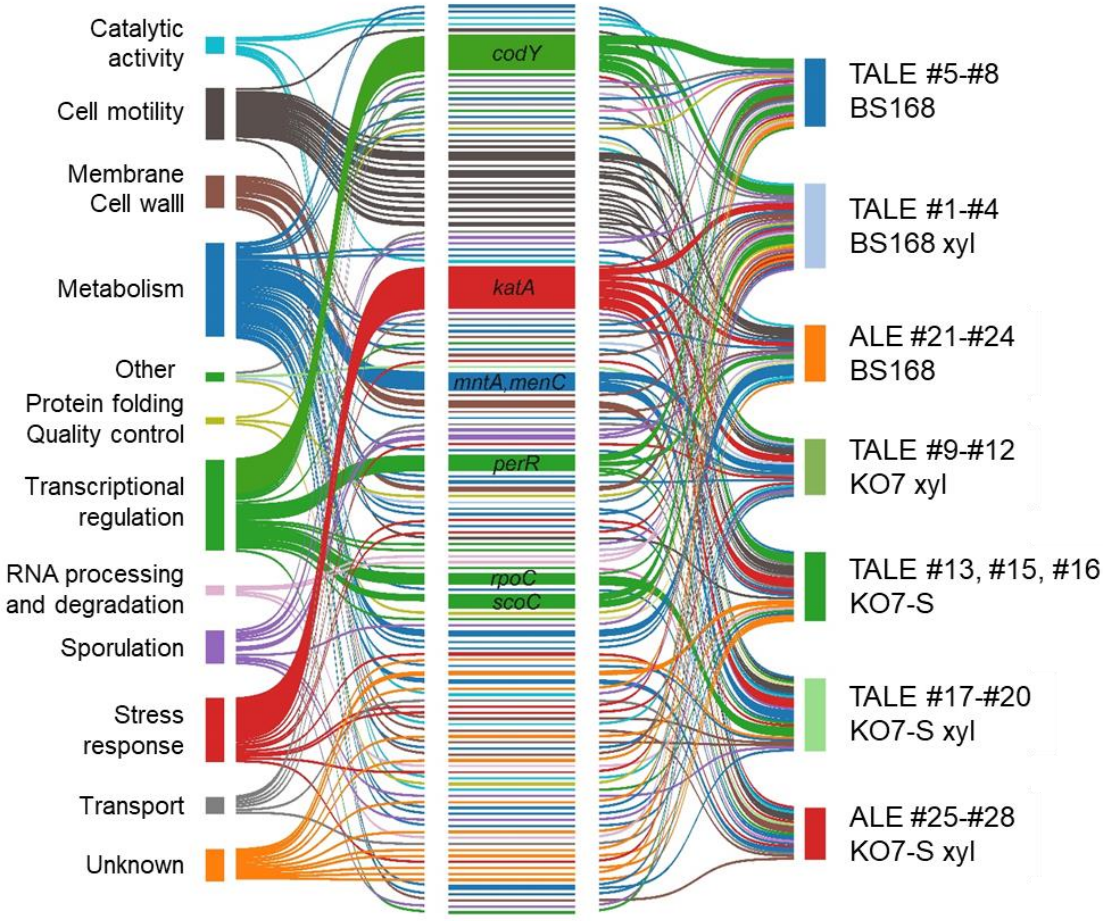


Figure 6 The Sankey diagram showing the functional allocation for each of the mutations of the different strain backgrounds.

Converged mutations were identified by comparing all clonal isolates of all 19 TALE experiments and 8 ALE M9extra control experiments. Genes or genetic regions that showed a mutation in >2 independent TALE or ALE experiments, in either replicates from the same or different backgrounds were considered converged mutations and are summarized in **Figure 7**. A more detailed overview of all observed mutations per isolate is given in **Table S6**. Mutations of genes or genetic regions across the independent replicates strongly suggest that they are adaptive [48]. Evolved isolates acquired TALE-specific converged mutations (*codY, mntA*, *scoC*, *rpoC*, *ponA, veg* and *oppD*), as well as converged mutations shared by both TALE and M9 ALE experiments (*katA*, *perR* and genes associated to the flagellar structure). While it is

likely the TALE-specific mutations are associated with increased tolerance, the shared mutations point towards media- and cultivation-specific adaptations.

	BS168 WT	BS168 XYL	KO7s WT	KO7 XYL	KO7-s XYL	Cumulative TALE	BS168 XYL M9 control	KO7-s XYL M9 control
codY	4/4	4/4	4/4	2/3	1/4	15/19	0/4	0/4
rpoC	0/4	1/4	0/4	0/3	4/4	5/19	0/4	0/4
mntA	0/4	0/4	1/4	3/3	4/4	8/19	0/4	0/4
scoC	3/4	3/4	0/4	0/3	0/4	6/19	0/4	0/4
oppD	1/4	1/4	0/4	0/3	0/4	2/19	0/4	0/4
ponA	2/4	0/4	0/4	0/3	0/4	2/19	0/4	0/4
Veg	0/4	0/4	2/4	0/4	0/4	2/19	0/4	0/4
katA, ssuB (intergenic)	1/4	3/4	4/4	3/3	3/4	14/19	2/4	1/4
perR	3/4	0/4	0/4	0/3	1/4	4/19	2/4	1/4
Flag*	0/4	0/4	4/4	3/3	4/4	11/19	3/4	3/4

Figure 7 Heat map of converged mutated genes across independent TALE and ALE experiments for different backgrounds. Numbers indicate the number of evolved isolates in which mutations in the given genes or genetic regions were observed. In the Cumulative TALE column, the total number of evolved isolates in which mutations were observed are added up for all TALE experiments. *Genes associated to the flagellar structure are pooled, these include fliG (TALE#9), fliR (TALE#9, #10, ALE #28), fliH (TALE #12, ALE #21), fliF (TALE #15, ALE #23), fliI (TALE #13), fliK (TALE #16, ALE #27), fliY, (TALE #17), fliZ (TALE #18, ALE #24), fliP (TALE #19), flhA (ALE #26)

The most frequently mutated gene, specific for only the TALE experiments, was the transcription factor *codY*. The pleiotropic regulator CodY is involved in the expression of several hundred genes, including those related to degradative enzymes, transport systems, intracellular catabolic systems, chemotaxis, motility, genetic competence and sporulation [49,50]. Remarkably, all eight BS168 strains were found to have amino acid substitutions for the residue arginine-214 (R214X). As R214 is part of the highly conserved DNA-binding domain (202-222), called the HTH-motif, the observed substitutions will likely influence the transcriptional regulation by CodY (**Figure S12**). Previous studies showed that substitutions of arginine-214 led to a reduced ability of CodY to bind to the target genes. In contrast, it is also shown that the context of the HTH-motif in DNA binding domains may influence the recognition and specificity of HTH-mediated protein-DNA interactions [51]. Considering the conserved nature of the *codY* gene, it is possible that the SNPs observed outside of the HTH-domain in the KO7, KO7-S and KO7-S xyl backgrounds (T125I, A186T, E193G, 2xY241C, L245P, L245P), could still influence the regulatory function of the protein.

As many independent converged mutations suggest to have a causal relation with the improved tolerance of the evolved isolates, it is interesting to reflect on the replicates which did not share the mutations in the *codY* gene (TALE#16, TALE#17, TALE#18, TALE#19) and attempt to find mutations which have a similar functional outcome related to tolerance. The end isolate of TALE #16 only showed eight SNPs, of which only five were not shared with other evolved isolates. Intriguingly, this included the HTH-domain global regulator gene c*ymR*. Transcriptome analysis of Δ*cymR* mutants showed that many genes related to stress response were altered [52]. Moreover, the evolved isolates of TALE #17, TALE#18, TALE#19 and TALE#20 all showed the exact similar SNP in the *rpoC* gene (125,619, C→T) (**Figure 7**, **Table S6**). The expression of *rpoC* is regulated by CodY and known to regulate stress response genes, including those of extracytoplasmic function (ECF), σ factors (σM, σW, and σX) and the general stress σ factor (σB) [53].

A total of 8 out of 19 evolved isolates shared mutations in the promoter region (-8, -15) of the manganese transporter lipoprotein, *mntA*. It has been shown that manganese plays a vital role in both PerR-mediated regulation of genes, which are involved in protection against reactive oxygen species and the activity of the general stress σ factor (σB) [54,55]. Moreover, overexpression of *mntA* in *Streptococcus oligofermentans* led to 12-fold higher survival rates upon exposure to peroxide stress, while the *mntA* knock-out showed 5.7 lower survival rates [56]. Notably, mutations in the *mntA* promoter region seemed to only occur in evolved isolates of the KO7, KO7-S and KO7-S xyl background.

Converged mutations observed in the transcriptional repressor *scoC*, were exclusively present in evolved isolates of the BS168 and BS168 xyl background. Mutations varied from SNPs to truncations which deleted about 50% of the gene. The ScoC regulator plays a crucial role in regulating extracellular protease expression and sporulation. As there are no mutations observed in the KO7/KO7-S strains, the regulation regarding to extracellular proteases is likely of importance. For evolved isolates of the BS168/BS168 xyl background, targets of ScoC (a*prE* and *nprE*) account for ~95% of the total extracellular protease activity [57]. Increased production of extracellular proteases may be advantageous, as it would enable strains to make use of the proteins present in biomass hydrolysate. In accordance, it has been shown that both CodY and ScoC act as direct repressors, forming a rare feed-forward regulatory loop. As such, both inactivations of *codY* and *scoC* are required for the de-repression of a*prE* and *nprE* in nutrient-rich environments [58]. The same feed-forward loop is utilized to regulate of the oligopeptide ABC transporter *oppD*. Notably, not only is *oppD* directly negatively regulated by both CodY and ScoC, mutations in *oppD* gene are mutually exclusive from *scoC* mutations (TALE#2, TALE#6) in the evolved isolates of the BS168/BS168 xyl background. Although *scoC* and *oppD* are in the same regulatory pathway, the exact mechanism of how mutations in either *scoC* or *oppD* are beneficial for growth in biomass hydrolysate remains enigmatic. Another TALE-specific converged mutation was a deletion of Δ39 bp at the 3' end of the Penicillin-binding protein 1A/1B, *ponA*. As the protein is involved in cell wall biogenesis, it could have a potential link to the tolerance of the strain, although there is no obvious link as the observed deletion does not alter any of

the enzymatic active sites of the protein. Lastly, protein Veg was mutated in 2 evolved isolates of the KO7-S background. Protein Veg is responsible for biofilm regulation via repression of *sinR* [59].

Next, we identified converged mutations in both TALE and control ALE experiments, which may be generally beneficial under the employed cultivation conditions. Except for TALE#2, all other 18 TALE experiments yielded evolved isolates which were either mutated in the promoter region of the vegetative catalase, *katA* or its transcriptional repressor *perR*. Mutations in the *katA* promoter were either around the -10/-35 regions, or in the regulatory domain called the Per box. In contrast, mutations in the *perR* gene led to substitutions across the whole length of the gene. Moreover, 6 out of the 8 evolved isolates derived from control ALE experiments also showed mutations in either the promoter region of *katA* or *perR*. This suggests that the observed mutations are linked to the cultivation conditions of the experiment and the M9extra-medium rather than tolerance towards biomass hydrolysates. As the mutations are so dominantly present across the different evolution experiments, a higher expression of the vegetative catalase *katA* may be advantageous by protecting the cell from stress-derived oxidative damage. Apparently, the increased ability to ward off reactive oxygen species is not only beneficial when cultivating cells in biomass hydrolysate, but also in the M9extra medium. Intriguingly, the evolved isolate from TALE#2 showed a deletion in the promoter region of the metalloregulation DNA-binding stress protein, called *mrgA*. Being part of the same regulon as *katA*, the role of MrgA is associated with protection against oxidative damage and is under transcriptional repression of PerR. As with some of the *katA-mutants*, the Per box of the promoter region of *mrgA* was altered [60]. Similarly, evolved strains of the M9extra control experiment which showed mutations in neither *katA* or *perR (*ALE#25 and ALE#26), were mutated in other genes related to protection against oxidative stress (*rsbX* and *trxB,* respectively) [61,62].

Eleven different genes encoding parts of the flagella structure were mutated in both TALE and ALE experiments, and hence are likely to play a role in improving growth in the applied cultivation conditions. Mutations included 8 deletions, 7 introductions of stop codons and 3 substitutions. It is worth mentioning that none of the evolved isolates of the BS168/BS168 xyl TALE experiments showed mutations in genes of the flagella structure, while this was the case for the ALE control. The absence/presence of flagella-associated mutations had severe consequences on the phenotype and ability to form a biofilm, as is clearly visible in Figure S13. It has been extensively described that both the biosynthesis and the operation of flagella represents a considerable cost for the host and that loss of flagella functions can be generally beneficial in well-mixed cultivations [63]. As such, improved growth has been observed in both evolutionary studies and rationally designed cell factories [64,65].

In short, mutations associated with increased tolerance included global regulators (*codY, scoC*, *rpoC, ponA, veg*) and more specific genes (transporters *mntA* and *oppD)*. Loss of flagella function and protection against oxidative damage appeared to be generally beneficial under the employed cultivation conditions. As commonly observed in stress tolerance ALE studies, point mutations in global regulators likely tune expression levels of many different stress-related genes to restore growth [15,66]. Although alterations of the global regulator genes render it challenging to provide a molecular mechanistic basis for the improved phenotype, mutations present in many independent experiments (like *codY*) are highly likely to be causal and can be used for future rational strain design. Current work can be expanded to evolutionary studies focusing on individual inhibitory compounds present in biomass hydrolysate to unravel general and specific chemical tolerance mechanisms systematically. Additional rational design and evolutionary engineering can generate strains even closer to their biological limit. Nevertheless, as industrial implementation requires high concentration of sugars and concomitantly higher amounts of inhibitors, the realization of a cost effective bioprocess using lignocellulose demands a multidisciplinary approach looking beyond the cell, fine-tuning both pretreatment, detoxification and fermentation conditions to keep the inhibitory compounds within the limits of biology.

4.4 Conclusions

This study demonstrated that TALE could efficiently be applied to obtain *B. subtilis* strains with increased tolerance to biomass hydrolysate-associated inhibitory compounds. Production experiments showed that evolved isolates could reach up to 48% of amylase productivity in DDGS hydrolysate when compared to optimized expression medium, containing more protein (268%) and sugars (116%). Whole-genome resequencing data of independently evolved isolates revealed key mutations related to tolerance and the employed cultivation conditions. As *B. subtilis* is a well-known industrial chassis organism, the obtained results are relevant for using lignocellulose as a feedstock for the manufacturing of bio-based products.

4.5 Acknowledgements

This work was supported by the Novo Nordisk Foundation (NNF), Denmark (grant number NNF17SA0031362 and NNF10CC1016517) and The Technical University of Denmark (DTU). We further acknowledge funding from the Independent Research Foundation Denmark (grant no. 7017-00321B).

We would like to thank International Flavors & Fragrances Inc. for kindly providing the Optimash® F200.

We would like to acknowledge the United Wisconsin Grain Producers for kindly providing the DDDS biomass.

The authors would like to thank Francesco Reggianini for his support with some of the experimental work.

The authors would like to acknowledge Christian Roslander and Mats Galbe for their advice and support during the steam explosion pretreatment (Lund University).

Figure 1 Created with BioRender.com.

4.6 References to Chapter 4

1. Fao & Oecd. *OECD-FAO Agricultural Outlook 2015-2024*. (2015).
2. Chatzifragkou, A. *et al.* Biorefinery strategies for upgrading Distillers' Dried Grains with Solubles (DDGS). *Process Biochemistry* (2015). doi:10.1016/j.procbio.2015.09.005
3. Chatzifragkou, A. & Charalampopoulos, D. Distiller's dried grains with solubles (DDGS) and intermediate products as starting materials in biorefinery strategies. in *Sustainable Recovery and Reutilization of Cereal Processing By-Products* (2018). doi:10.1016/B978-0-08-102162-0.00003-4
4. RFA. *Annual Industry Outlook 2020*. (2020).
5. Iram, A., Cekmecelioglu, D. & Demirci, A. Distillers' dried grains with solubles (DDGS) and its potential as fermentation feedstock. *Applied Microbiology and Biotechnology* **104**, 6115–6128 (2020).
6. Liu, L. *et al.* Developing Bacillus spp. as a cell factory for production of microbial enzymes and industrially important biochemicals in the context of systems and synthetic biology. *Appl. Microbiol. Biotechnol.* **97**, 6113–6127 (2013).
7. Gu, Y. *et al.* Advances and prospects of Bacillus subtilis cellular factories: From rational design to industrial applications. *Metabolic Engineering* **50**, 109–121 (2018).
8. Nghiem, N. P., Montanti, J. & Kim, T. H. Pretreatment of Dried Distiller Grains with Solubles by Soaking in Aqueous Ammonia and Subsequent Enzymatic/Dilute Acid Hydrolysis to Produce Fermentable Sugars. *Appl. Biochem. Biotechnol.* **179**, 237–250 (2016).
9. Liu, H. *et al.* Preparation of hydrolytic liquid from dried distiller's grains with solubles and fumaric acid fermentation by Rhizopus arrhizus RH 7-13. *J. Environ. Manage.* **201**, 172–176 (2017).
10. Dien, B. S. *et al.* Enzyme characterization for hydrolysis of AFEX and liquid hot-water pretreated distillers' grains and their conversion to ethanol. *Bioresour. Technol.* **99**, 5216–5225 (2008).
11. van der Pol, E. C., Bakker, R. R., Baets, P. & Eggink, G. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. *Applied Microbiology and Biotechnology* **98**, 9579–9593 (2014).
12. Lastiri-Pancardo, G. M. & Utrilla, J. Evolutionary engineering of microorganisms to overcome toxicity during lignocellulose hydrolysates utilization. in *Engineering of Microorganisms for the Production of Chemicals and Biofuels from Renewable Resources* 181–200 (Springer International Publishing, 2017). doi:10.1007/978-3-319-51729-2_7

13. Palmqvist, E., Grage, H., Meinander, N. Q. & Hahn-Hägerdal, B. Main and interaction effects of acetic acid, furfural, and p- hydroxybenzoic acid on growth and ethanol productivity of yeasts. *Biotechnol. Bioeng.* **63**, 46–55 (1999).

14. van der Maas, L., Driessen, J. L. S. P. & Mussatto, S. I. Effects of Inhibitory Compounds Present in Lignocellulosic Biomass Hydrolysates on the Growth of Bacillus subtilis. *Energies* **14**, 8419 (2021).

15. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56**, 1–16 (2019).

16. Wang, X., Khushk, I., Xiao, Y., Gao, Q. & Bao, J. Tolerance improvement of Corynebacterium glutamicum on lignocellulose derived inhibitors by adaptive evolution. *Appl. Microbiol. Biotechnol.* **102**, 377–388 (2018).

17. Wallace-Salinas, V. & Gorwa-Grauslund, M. F. Adaptive evolution of an industrial strain of Saccharomyces cerevisiae for combined tolerance to inhibitors and temperature. *Biotechnol. Biofuels* **6**, 151 (2013).

18. Qin, D. *et al.* An auto-inducible Escherichia coli strain obtained by adaptive laboratory evolution for fatty acid synthesis from ionic liquid-treated bamboo hydrolysate. *Bioresour. Technol.* **221**, 375–384 (2016).

19. Linville, J. L. *et al.* Industrial Robustness: Understanding the Mechanism of Tolerance for the Populus Hydrolysate-Tolerant Mutant Strain of Clostridium thermocellum. *PLoS One* **8**, e78829 (2013).

20. Koppram, R., Albers, E. & Olsson, L. Evolutionary engineering strategies to enhance tolerance of xylose utilizing recombinant yeast to inhibitors derived from spruce biomass. *Biotechnol. Biofuels* **5**, 32 (2012).

21. Almario, M. P., Reyes, L. H. & Kao, K. C. Evolutionary engineering of *Saccharomyces cerevisiae* for enhanced tolerance to hydrolysates of lignocellulosic biomass. *Biotechnol. Bioeng.* **110**, 2616–2623 (2013).

22. Zeigler, D. R. *et al.* The origins of 168, W23, and other Bacillus subtilis legacy strains. *J. Bacteriol.* **190**, 6983–6995 (2008).

23. Zeigler, D. R. CONSTRUCTION OF EXOPROTEASE-FREE, MARKER-FREE BACILLUS SUBTILIS HOSTS | Daniel R Zeigler | 2 updates | 1 publications | Research Project. (2016). Available at: https://www.researchgate.net/project/Construction-of-exoprotease-free-marker-free-Bacillus-subtilis-hosts. (Accessed: 16th September 2020)

24. Bacillus Genetic Stock Center. Available at: https://bgsc.org/index.php.

25. Arantes, O. & Lereclus, D. Construction of cloning vectors for Bacillus thuringiensis. *Gene* **108**, 115–119 (1991).

26. Guiziou, S. *et al.* A part toolbox to tune genetic expression in Bacillus subtilis.

Nucleic Acids Res. **44**, 7495–7508 (2016).

27. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34**, e122 (2006).

28. Falkenberg, K. B. *et al.* The ProUSER2.0 Toolbox: Genetic Parts and Highly Customizable Plasmids for Synthetic Biology in Bacillus subtilis. *ACS Synth. Biol.* acssynbio.1c00130 (2021). doi:10.1021/ACSSYNBIO.1C00130

29. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96**, 23–28 (1990).

30. Vojcic, L., Despotovic, D., Martinez, R., Maurer, K. H. & Schwaneberg, U. An efficient transformation method for bacillus subtilis DB104. *Appl. Microbiol. Biotechnol.* **94**, 487–493 (2012).

31. Rahmer, R., Heravi, K. M. & Altenbuchner, J. Construction of a super-competent Bacillus subtilis 168 using the PmtlA-comKS inducible cassette. *Front. Microbiol.* **6**, (2015).

32. Altenbuchner, J. Editing of the Bacillus subtilis genome by the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **82**, 5421–5427 (2016).

33. Falkenberg, K. B. General rights Synthetic Biology approaches for improved protein production in the bacterium Bacillus subtilis. *Downloaded from orbit.dtu.dk* (2021).

34. Cavin, J. F., Dartois, V. & Diviès, C. Gene cloning, transcriptional analysis, purification, and characterization of phenolic acid decarboxylase from Bacillus subtilis. *Appl. Environ. Microbiol.* **64**, 1466–1471 (1998).

35. Palmqvist, E. *et al.* Design and operation of a bench-scale process development unit for the production of ethanol from lignocellulosics. *Bioresour. Technol.* **58**, 171–179 (1996).

36. Singleton, V. L. & Rossi, J. A. Colorimetry of Total Phenolics with Phosphomolybdic-Phosphotungstic Acid Reagents. *Am. J. Enol. Vitic.* **16**, (1965).

37. Mohamed, E. T. *et al.* Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. *Microb. Cell Fact.* **16**, 204 (2017).

38. Mohamed, E. T. *et al.* Generation of an E. coli platform strain for improved sucrose utilization using adaptive laboratory evolution. *Microb. Cell Fact.* **18**, 116 (2019).

39. LaCroix, R. A. *et al.* Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. *Appl. Environ. Microbiol.* **81**, 17–30 (2015).

40. Rasmussen, M. D., Bjoernvad, M. E., & Diers, I. Patent No. WO 00/75344. World Intellectual Property Organization. (2000).

41. Xiao, Z., Storms, R. & Tsang, A. A quantitative starch-iodine method for measuring alpha-amylase and glucoamylase activities. *Anal. Biochem.* **351**, 146–148 (2006).

42. Phaneuf, P. V., Gosting, D., Palsson, B. O. & Feist, A. M. Aledb 1.0: A database of mutations from adaptive laboratory evolution experimentation. *Nucleic Acids Res.* **47**, D1164–D1171 (2019).

43. Galbe, M. & Wallberg, O. Pretreatment for biorefineries: a review of common methods for efficient utilisation of lignocellulosic materials. *Biotechnol. Biofuels 2019 121* **12**, 1–26 (2019).

44. Galbe, M. & Zacchi, G. Pretreatment of Lignocellulosic Materials for Efficient Bioethanol Production. *Adv. Biochem. Eng. Biotechnol.* **108**, 41–65 (2007).

45. Zheng, Y., Shi, J., Tu, M. & Cheng, Y. S. Principles and Development of Lignocellulosic Biomass Pretreatment for Biofuels. *Adv. Bioenergy* **2**, 1–68 (2017).

46. Sandberg, T. E. *et al.* Evolution of Escherichia coli to 42 °C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. *Mol. Biol. Evol.* **31**, 2647–2662 (2014).

47. BSubCyc | Encyclopedia of Bacillus Subtilis Subtilis Genes and Metabolism. Available at: https://bsubcyc.org/. (Accessed: 9th January 2022)

48. Bailey, M. J. & Markkanen, P. H. Use of mutagenic agents in improvement of α-amylase production by bacillus subtilis. *J. Appl. Chem. Biotechnol.* **25**, 73–79 (2007).

49. Cao, H., Villatoro-Hernandez, J., Weme, R. D. O., Frenzel, E. & Kuipers, O. P. Boosting heterologous protein production yield by adjusting global nitrogen and carbon metabolic regulatory networks in Bacillus subtilis. *Metab. Eng.* **49**, 143–152 (2018).

50. Sonenshein, A. L. Pentose-phosphate pathway Control of key metabolic intersections in Bacillus subtilis. (2007). doi:10.1038/nrmicro1772

51. Joseph, P., Ratnayake-Lecamwasam, M. & Sonenshein, A. L. A region of Bacillus subtilis CodY protein required for interaction with DNA. *J. Bacteriol.* **187**, 4127–4139 (2005).

52. Even, S. *et al.* Global control of cysteine metabolism by CymR in Bacillus subtilis. *J. Bacteriol.* **188**, 2184–2197 (2006).

53. Lee, Y. H., Nam, K. H. & Helmann, J. D. A mutation of the RNA polymerase β′ subunit (rpoC) confers cephalosporin resistance in Bacillus subtilis. *Antimicrob. Agents Chemother.* **57**, 56–65 (2013).

54. Fuangthong, M., Herbig, A. F., Bsat, N. & Helmann, J. D. Regulation of the Bacillus subtilis fur and perR genes by PerR: Not all members of the PerR regulon are peroxide inducible. *J. Bacteriol.* **184**, 3276–3286 (2002).

55. Chen, L., Keramati, L. & Helmann, J. D. Coordinate regulation of Bacillus subtilis peroxide stress genes by hydrogen peroxide and metal ions. *Proc. Natl. Acad. Sci.* **92**, 8190–8194 (1995).

56. Wang, X., Tong, H. & Dong, X. PerR-Regulated Manganese Ion Uptake Contributes to Oxidative Stress Defense in an Oral Streptococcus. (2014). doi:10.1128/AEM.00064-14

57. Kawamura, F. & Doi, R. H. Construction of a Bacillus subtilis double mutant deficient in extracellular alkaline and neutral proteases. *J. Bacteriol.* **160**, 442–444 (1984).

58. Barbieri, G., Albertini, A. M., Ferrari, E., Sonenshein, A. L. & Belitsky, B. R. Interplay of CodY and ScoC in the regulation of major extracellular protease genes of Bacillus subtilis. *J. Bacteriol.* **198**, 907–920 (2016).

59. Lei, Y., Oshima, T., Ogasawara, N. & Ishikawa, S. Functional analysis of the protein veg, which stimulates biofilm formation in bacillus subtilis. *J. Bacteriol.* **195**, 1697–1705 (2013).

60. Fuangthong, M. & Helmann, J. D. Recognition of DNA by Three Ferric Uptake Regulator (Fur) Homologs in Bacillus subtilis. *J. Bacteriol.* **185**, 6348–6357 (2003).

61. Smits, W. K., Dubois, J. Y. F., Bron, S., Van Dijl, J. M. & Kuipers, O. P. Tricksy business: Transcriptome analysis reveals the involvement of thioredoxin A in redox homeostasis, oxidative stress, sulfur metabolism, and cellular differentiation in Bacillus subtilis. *J. Bacteriol.* **187**, 3921–3930 (2005).

62. Voelker, U., Luo, T., Smirnova, N. & Haldenwang, W. Stress activation of Bacillus subtilis σ(B) can occur in the absence of the σ(B) negative regulator RsbX. *J. Bacteriol.* **179**, 1980–1984 (1997).

63. Lara, A. & Gossett, G. Minimal Cells: Design, Construction, Biotechnological Applications.

64. Mohamed, E. T. *et al.* Adaptive laboratory evolution of Pseudomonas putida KT2440 improves p-coumaric and ferulic acid catabolism and tolerance. *Metab. Eng. Commun.* **11**, e00143 (2020).

65. Liu, Y. *et al.* Developing rapid growing Bacillus subtilis for improved biochemical and recombinant protein production. *Metab. Eng. Commun.* **11**, (2020).

66. Lennen, R. *et al.* Adaptive laboratory evolution reveals general and specific chemical tolerance mechanisms and enhances biochemical production. *bioRxiv* 634105 (2019). doi:10.1101/634105

4.7 Supplementary material

Table S4 Strains selected for Whole Genome Resequencing

Strain	ID	Background
Start strain	0	BS168 xyl
Start strain	0	BS168
Start strain	0	KO7-S
Start strain	0	KO7 xyl
Start strain	0	KO7-S xyl
Evolved isolate	TALE#1 I6	BS168 xyl
Evolved isolate	TALE#2 I10	BS168 xyl
Evolved isolate	TALE#3 I12	BS168 xyl
Evolved isolate	TALE#4 I7	BS168 xyl
Evolved isolate	TALE#5 I17	BS168
Evolved isolate	TALE#6 I15	BS168
Evolved isolate	TALE#7 I1	BS168
Evolved isolate	TALE#8 I6	BS168
Evolved isolate	TALE#9 I13	KO7-S
Evolved isolate	TALE#10 I2	KO7-S
Evolved isolate	TALE#11 I1	KO7-S
Evolved isolate	TALE#12 I7	KO7-S
Evolved isolate	TALE#13 I15	KO7 xyl
Evolved isolate	TALE#14 I20	KO7 xyl
Evolved isolate	TALE#15 I6	KO7 xyl
Evolved isolate	TALE#16 I19	KO7 xyl
Evolved isolate	TALE#17 I4	KO7-S xyl
Evolved isolate	TALE#18 I5	KO7-S xyl
Evolved isolate	TALE#18 I19	KO7-S xyl
Evolved isolate	TALE#20 I6	KO7-S xyl
Evolved isolate	ALE#21 I1	BS168 xyl
Evolved isolate	ALE#22 I1	BS168 xyl
Evolved isolate	ALE#23 I1	BS168 xyl
Evolved isolate	ALE#24 I1	BS168 xyl
Evolved isolate	ALE#25 I1	KO7-S xyl
Evolved isolate	ALE#26 I1	KO7-S xyl
Evolved isolate	ALE#27 I1	KO7-S xyl
Evolved isolate	ALE#28 I1	KO7-S xyl

Table S5 Oligos used in this study

No.	Name	Sequence (5' to 3')
768	pJOE8999_BB-U-fw	AAGGCCUTTCTAGATTAAGAAATAATCTTCATC
765	pJOE8999_BB-U-rv	AGGTACAUTTTACTCAATTCTCTAATCACGG
JD01	gRNAaraR-U-fw	ATGTACCUACGCAAGGCGTGAAACGGATGAAGTTTTAGAGCTAGAAATAGCAAGTTAAAA
767	gRNAaraR-U-rv	ACCCTAUAGTGAGTCGTATTAAAAAGGCCC
JD02	araR-up-U-fw	ATAGGGUCGACGGCCAATGGCACCATTCCATAAGCAAGC
JD03	araR-up-U-rv	ATATTTGUACGTACTAATTAAATGTAATTTTCGTTAAATTTTAATATAAGTACGTACAAT
JD04	araR-down-U-fw	ACAAATAUAGAAAAAGCAATGTATGGGTCTCCCCG
JD05	araR-down-U-rv	AGGCCTUATCCAATTGACAGCCAGAGCGCTCC
JD44	ParaE-down-U-rv	AAAGGAAGUATTTGAAAATGAAGAATACTCCAACTCAATTAGAACCAAATGTTC
JD48	ParaE-up-U-rv	ACTTCCTTUCATGTCActgataatttaacaCACTTTcaaaagagtgtcaacgtgtattgacgcagtAAAAAGCAATGTATGGGTCTCCCCG
JD51	ParaE-up-U-fw	ATAGGGUCGACGGCCACCAATTGACAGCCAGAGCGCTC
JD52	ParaE-down-U-fw	AGGCCTUATTTTCTCCATTAATACGTGTCAGGATCTTTAAAGC
JD53	gRNAaraE_U_fv	ATGTACCUACGCCGCTTTACTCAACATTCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAA
JD16	pHT315_BB-U-fw	AGCGCAGCUGAAATAGCTGCGCTTTTTTGTGTCATAATTAAGCCAGCCCCGACACC
JD17	pHT315_BB-U-rv	AGCTGAAAUAGCTGCGCTTTTTTGTGTCATAACAGCTCACTCAAAGGCGGTAATACG

JD18	xylB-U-fw	ATTTCAGCUGCGCTTTTTTCGAAACAAACGCATTTGACCAAACAAGC
JD19	xylB-U-rv	ATTATCTGTUCGACAAATAATCGTGGCGATGCGCAACTG
JD20	xylA-U-fw	AACAGATAAUGGTTTACCAGATTTTCCAGTTGTTCC
JD21	xylA-U-rv	ATGCAAGCCUATTTTGACCAGCTCG
JD22	P43-U-fw	AGGCTTGCAUGTGTACATTCCTCTCTTACCTATAATGGTACCGC
JD23	P43-U-rv	ATTTTACAUTTTTAGAAATGGGCGTGAAAAAAAGCGC
JD24	PHT2-U-fw	ATGTAAAAUTACCGCTCGCCGCAGCCG
JD46	PHT2-U-rv	ACTTCCTTUCATGTCAaacttcattttaaatgtgtcggaaatacttgtcaagcttgccatcttaacGCCTGGGGTGCCTAATGAGTG
JD44	araE-U-fw	AAAGGAAGUATTTGAAAATGAAGAATACTCCAACTCAATTAGAACCAAATGTTC
JD29	araE-U-rv	AGCTGCGCUTTTTTTCATTTTATCCAAAGCTTTTCAATTTCCTCGAGCG
JD77	amyQ-U-fw	ggcgaUAGGAGGACAAACATGATTCAAAAACGA
JD78	amyQ-U-rv	ggtgcgaUTTATTTCTGAACATAAATGGAGACGGACCCT

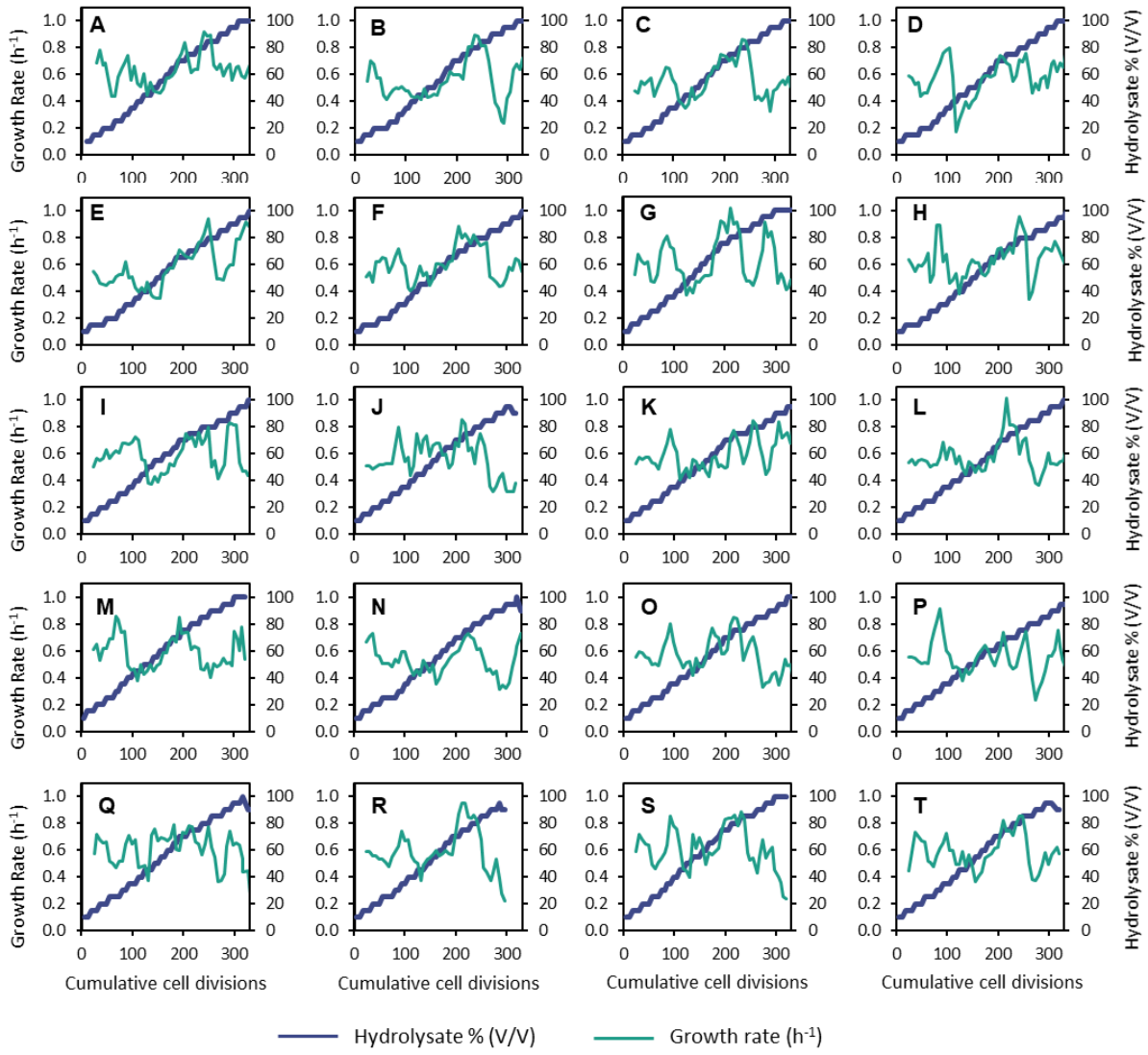


Figure S8 Population growth rate and hydrolysate supplementation over the course of all TALE experiments: A) TALE #1 BS168 xyl, B) TALE #2 BS168 xyl, C) TALE #3 BS168 xyl, D) TALE #4 BS168 xyl, E) TALE #5 BS168, F) TALE #6 BS168, G) TALE #7 BS168, H) TALE #8 BS168, I) TALE #9 KO7-s, J) TALE #10 KO7-s, K) TALE #11 KO7-s, L) TALE #12 KO7-s, M) TALE #13 KO7 xyl, N) TALE #14 KO7 xyl, O) TALE #15 KO7 xyl, P) TALE #16 KO7 xyl, Q) TALE #17 KO7-s xyl, R) TALE #18 KO7-s xyl, S) TALE #19 KO7-s xyl and T) TALE #20 KO7-s xyl

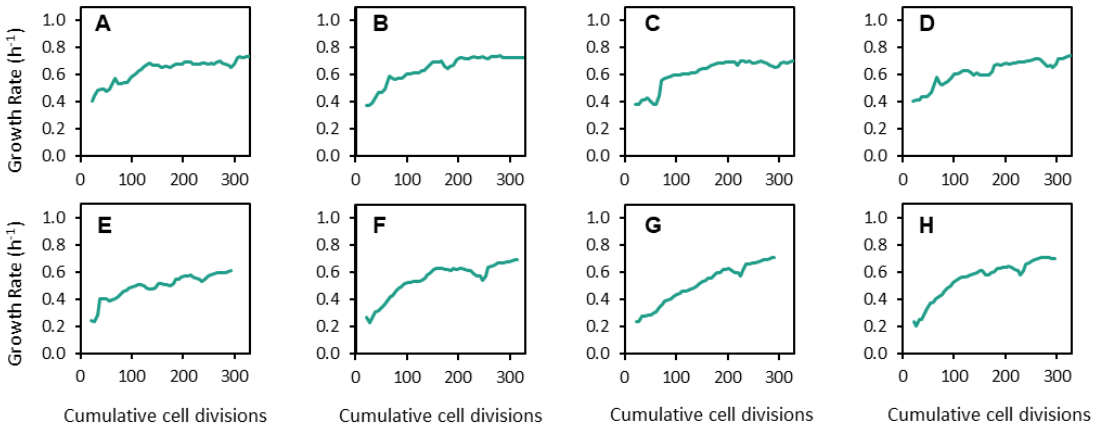


Figure S9 Population growth rate over the course of all control ALE experiments: A) ALE #21 BS168 xyl, B) ALE #22 BS168 xyl, C) ALE #23 BS168 xyl, D) ALE #24 BS168 xyl, E) ALE #25 KO7-S xyl, F) ALE #26 KO7-S xyl, G) ALE #27 KO7-S xyl, H) KO7-S xyl #28 BS168

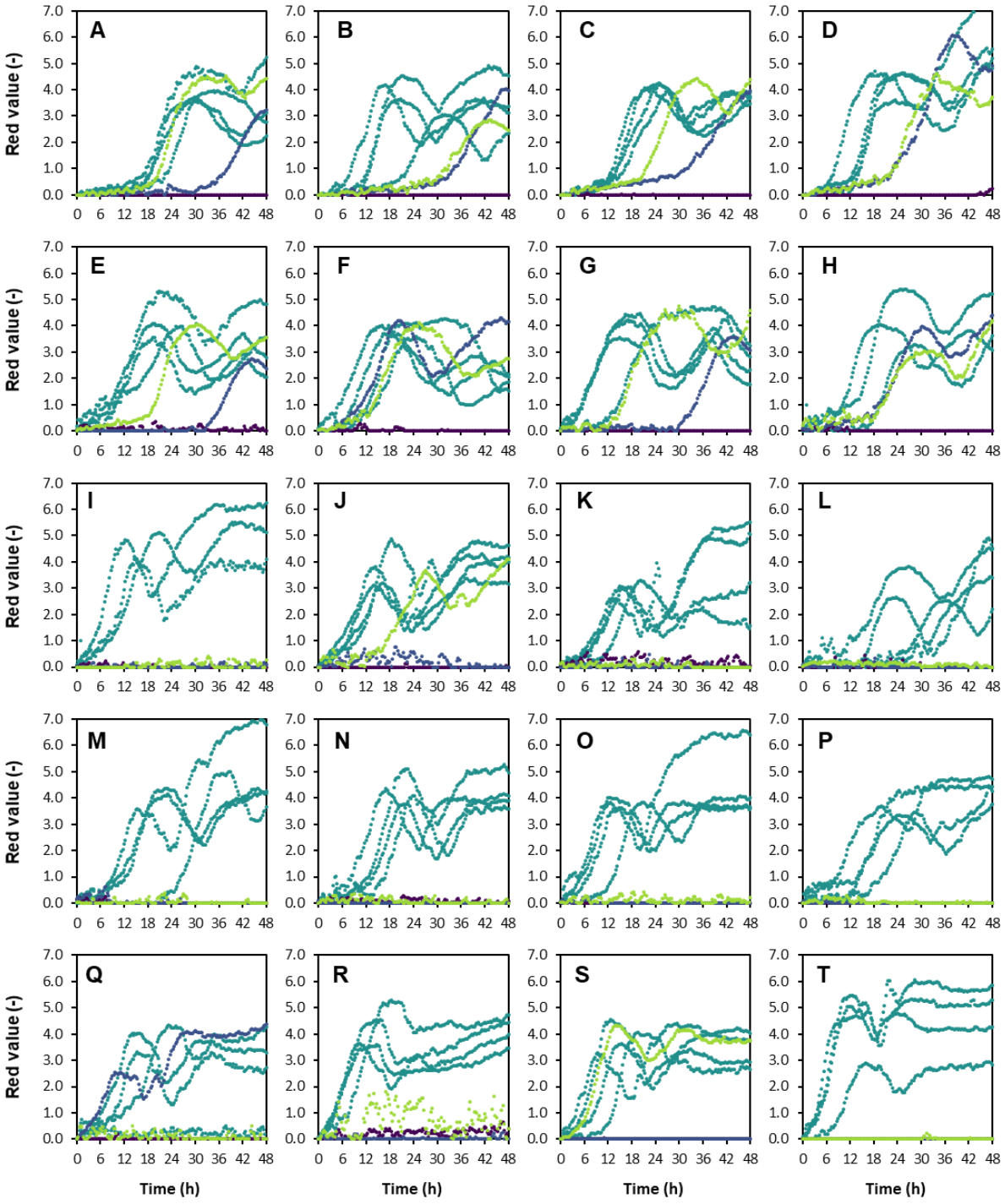


Figure S10 Growth performance of start strains, midpoint populations, endpoint populations and evolved isolates of all TALE experiments: A) TALE #1 BS168 xyl, B) TALE #2 BS168 xyl, C) TALE #3 BS168 xyl, D) TALE #4 BS168 xyl, E) TALE #5 BS168, F) TALE #6 BS168, G) TALE #7 BS168, H) TALE #8 BS168, I) TALE #9 KO7-s, J) TALE #10 KO7-s, K) TALE #11 KO7-s, L) TALE #12 KO7-s, M) TALE #13 KO7 xyl, N) TALE #14 KO7 xyl, O) TALE #15 KO7 xyl, P) TALE #16 KO7 xyl, Q) TALE #17 KO7-s xyl, R) TALE #18 KO7-s xyl, S) TALE #19 KO7-s xyl and T) TALE #20 KO7-s xyl. Start strains (●), midpoint population (●),endpoint population (●) and endpoint isolates (●).

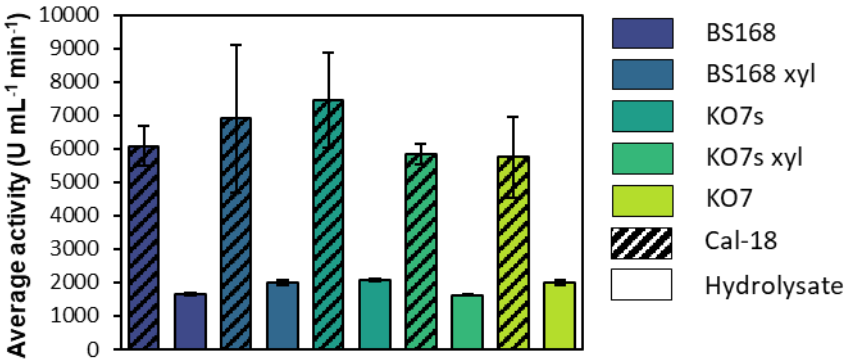


Figure S11 Amylase production of evolved isolates measured by amylase activity of culture supernatants in 100% hydrolysate-based medium and Cal18-2 media in a 24-well plate format after 72 h.

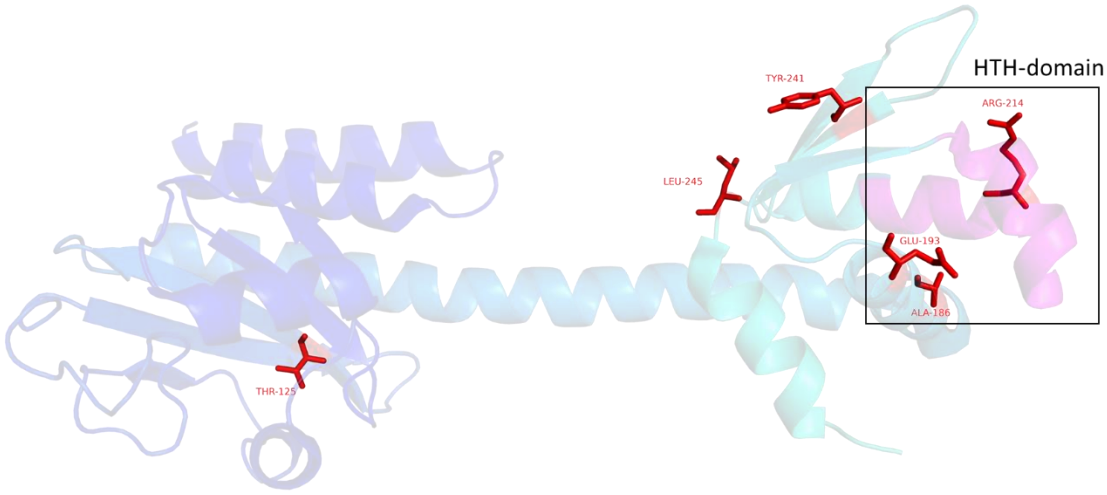


Figure S12 Structure of the CodY protein (exemplified by the PDB ID:5loe. Mutated amino acids and their side-chains are indicated in red and the HTH DNA-binding domain is colored pink. Structure was visualized using PyMOL *(Schrödinger LLC)*.

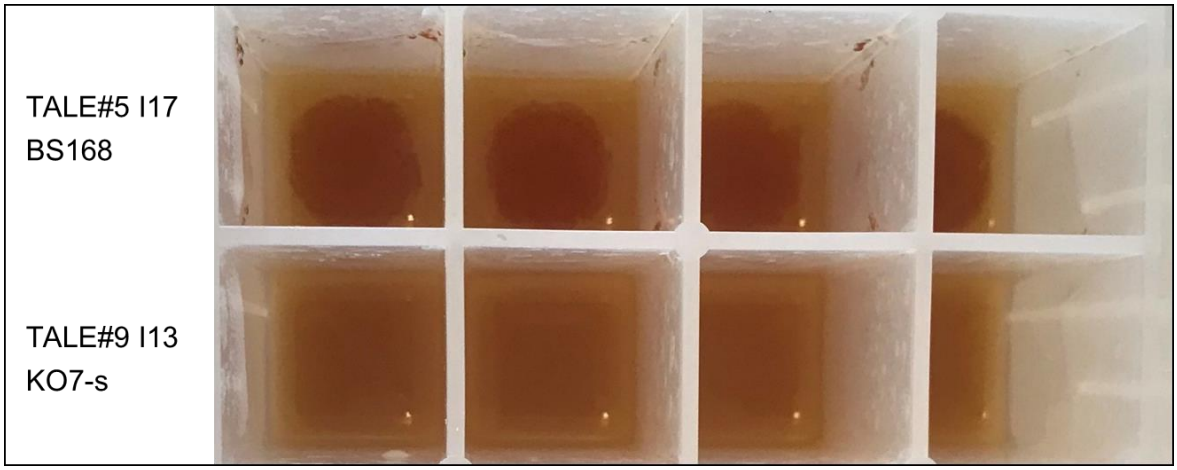


Figure S13 Pictures of end isolates TALE#5 I17 and TALE#9 I17 after 24h growth in cal-18 medium. Presence/absence of biofilm is clearly visible.

Table S6 Full list of mutations in the endpoint isolates of TALEs and ALEs shown in table 3. For each TALE experiment, "A" indicates the ALE experiment number, "F" indicates the flask number, "I" indicates the isolate number, and "R" indicates the technical replicate.

BS168 xyl TALE #1-4									
Position	**Mutation Type**	**Sequence Change**	**Gene**	**Product**	**Details**	**A1 F58 I6 R1**	**A2 F60 I10 R2**	**A3 F57 I12 R3**	**A4 F59 I7 R4**
1,672	SNP	T>C	*dnaA*	chromosomal replication initiator informational ATPase	H421H (CAT>CAC)		1		
73,935	SNP	T>C	*prkT*	serine/threonine-protein kinase	Y43H (TAT>CAT)			1	
78,236	SNP	A>T	*ftsH*	ATP-dependent cytoplasmic membrane protease	N418I (AAT>ATT)				1
127,648	SNP	A>G	*rpoC*	RNA polymerase (beta' subunit)	E696G (GAA>GGA)			1	
197,804	SNP	C>G	*cdaR*	regulator of diadenylate cyclase activity	P260A (CCC>GCC)	1			
506,071	SNP	C>A	*ydbM*	putative acyl-CoA dehydrogenase	P307Q (CCG>CAG)			1	
842,292	SNP	A>G	*ltaSB*	enzyme responsible for polyglycerolphosphate LTA synthesis	Y569H (TAC>CAC)			1	
961,037	INS	(A)5>6 R1 R3	*katA, ssuB*	vegetative catalase 1/aliphatic sulfonate ABC transporter (ATP-binding protein)	intergenic (-51/-357)	1		1	
961,062	DEL	Δ1 bp R4	*katA, ssuB*	vegetative catalase 1/aliphatic sulfonate ABC transporter (ATP-binding protein)	intergenic (-76/-332)				1
961,067	SNP	T>A R4	*katA, ssuB*	vegetative catalase 1/aliphatic sulfonate ABC transporter (ATP-binding protein)	intergenic (-81/-327)				1

961,069	SNP	G>T R4	*katA, ssuB*	vegetative catalase 1/aliphatic sulfonate ABC transporter (ATP-binding protein)	intergenic (-83/-325)				1
1,073,374	SNP	C>T (R1)	*scoC*	transcriptional regulator of extracellular protease production, sporulation and bacilysin production (MarR family)	W115* (TGG>TAG)	1			
1,073,401	SNP	A>T (R4)	*scoC*	transcriptional regulator of extracellular protease production, sporulation and bacilysin production (MarR family)	L106* (TTG>TAG)				1
1,073,672	SNP	A>T (R3)	*scoC*	transcriptional regulator of extracellular protease production, sporulation and bacilysin production (MarR family)	F16I (TTC>ATC)			1	
1,224,521	SNP	A>G	*oppD*	oligopeptide ABC transporter (ATP-binding protein)	E356G (GAA>GGA)		1		
1,248,775	SNP	A>T	*cotO*	spore outer coat protein	K37N (AAA>AAT)				1
1,323,881	SNP	C>T	*xkdD*	phage PBSX	P27L (CCG→CTG)		1		
1,481,481	SNP	T>C	*ykzU, ykuH*	conserved hypothetical protein/conserved protein of unknown function	intergenic (+30/-66)			1	
1,524,173	SNP	T>G	*rnjA*	ribonuclease J1	T205P (ACC>CCC)			1	
1,569,011	SNP	G>A	*ylbG*	conserved hypothetical protein	V30I (GTA>ATA)		1		
1,690,758	SNP	C>G	*codY*	transcriptional regulator, GTP and BCAA-dependent	R214G (CGT>GGT)	1			1
1,690,759	SNP	G>A	*codY*	transcriptional regulator, GTP and BCAA-dependent	R214H (CGT>CAT)		1	1	
1,764,843	SNP	A>G	*recA*	multifunctional SOS repair factor	S67G (AGC>GGC)		1		
2,063,773	SNP	T>A	*yobHc*	fragment of putative DNA phage repair protein	pseudogene (391/654 nt)				1

2,134,705	SNP	T>C	*ldcB*	D-alanyl-D-alanine carboxypeptidase lipoprotein (D-ala releasing from tetrapeptide hydrolysis)	Y228C (TAT>TGT)	1			
2,291,897	SNP	C>A	*scuA*	assembly factor BSco of the Cu(A) site of cytochrome c oxidase	E130* (GAG>TAG)			1	
2,435,685	SNP	T>C	*ppiB*	peptidyl-prolyl isomerase	E36G (GAA>GGA)		1		
2,485,359	SNP	T>C	*yqjF*	conserved protein of unknown function	F150S (TTC>TCC)	1			
2,778,062	SNP	C>T	*yrhH*	putative methyltransferase	D120N (GAT>AAT)	1			
2,861,836	DEL	(A)9>8	*mreB, ysxA*	cell-shape determining protein/conserved nucleotide-related metabolism protein	intergenic (-88/+4)			1	
3,383,490	DEL	(ATTATAATT)2>1	*yusZ, mrgA*	putative short-chain acyl dehydrogenase/metalloregulation DNA-binding stress protein	intergenic (+15/-67)		1		
3,799,612	SNP	G>A	*ywkB*	putative metabolite transporter	A242V (GCG>GTG)		1		
3,803,671	SNP	C>A	*rho*	transcriptional terminator Rho	G338V (GGC>GTC)		1		
4,000,709	SNP	T>C	*yxjF*	putative hydroxyacid dehydrogenase	N202D (AAT>GAT)		1		
4,124,455	SNP	T>C	*liaM*	permease for export of regulatory peptide LiaD*	Y163C (TAT>TGT)			1	

BS168 WT TALE #5-8									
Position	Mutation Type	Sequence Change	Gene	Product	Details	A5 F60 I17 R1	A6 F61 I15 R2	A7 F57 I1 R3	A8 F60 I6 R4
423,163	SNP	A>G	*gerKB*	spore germination receptor subunit	Y61C (TAT>TGT)			1	
718,125	SNP	T>C	*yerD*	putative osmotic shock glutamate synthase subunit (flavoprotein subunit, ferredoxin-dependent)	Y78C (TAT>TGT)				1
944,526	SNP	A>G	*perR*	transcriptional regulator (Fur family)	K14E (AAG>GAG)		1	1	1
944,551	SNP	C>T	*perR*	transcriptional regulator (Fur family)	P22L (CCT>CTT)		1		
944,586	SNP	T>C	*perR*	transcriptional regulator (Fur family)	S34P (TCT>CCT)				1
961,037	INS	(A)5>6	*katA, ssuB*	vegetative catalase 1/aliphatic sulfonate ABC transporter (ATP-binding protein)	intergenic (-51/-357)	1			
1,073,459	DEL	Δ1,532 bp	*scoC,yhaH,yhzF,trpP*	[scoC],yhaH,yhzF,[trpP]				1	
1,073,489	SNP	C>T	*scoC*	transcriptional regulator of extracellular protease production, sporulation and bacilysin production (MarR family)	A77T (GCA>ACA)	1			
1,073,527	SNP	G>C	*scoC*	transcriptional regulator of extracellular protease production, sporulation and bacilysin production (MarR family)	S64C (TCT>TGT)	1		1	1
1,224,521	SNP	A>G	*oppD*	oligopeptide ABC transporter (ATP-binding protein)	E356G (GAA>GGA)		1		
1,605,538	SNP	T>A	*sigE, sigG*	RNA polymerase sporulation-specific sigma-29 factor (sigma-E)/RNA polymerase sporulation-specific sigma factor (sigma-G)	intergenic (+48/-92)		1		
1,690,758	SNP	C>A	*codY*	transcriptional regulator, GTP and BCAA-dependent	R214S (CGT>AGT)	1	1	1	1
1,749,907	SNP	G>T	*rnjB*	dual activity 5' exo-and endoribonuclease J2	D164Y (GAC>TAC)			1	

1,763,199	SNP	A>G	*pgsA*	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase	N193D (AAC>GAC)		1		
1,994,711	SNP	C>T	*ppsA*	non-ribosomal plipastatin synthetase A involved in synthesis of plipastatin	E1083K (GAA>AAA)		1		
2,308,413	SNP	A>G	*degR, ypzA*	activator of degradative enzymes (aprE, nprE, sacB) production or activity/putative spore coat protein	intergenic (-74/-82)	1			
2,344,013	DEL	Δ39 bp	*ponA*	peptidoglycan glycosyltransferase (penicillin-binding proteins 1A and 1B)	coding (2570-2608/2745 nt)		1	1	
2,435,324	DEL	(A)7>6	*ypzD, ppiB*	putative germination protein/peptidyl-prolyl isomerase	intergenic (+100/+36)		1		
2,626,240	SNP	T>C	*dnaK*	molecular chaperone, ATP-dependent	M570V (ATG>GTG)	1			
2,775,128	SNP	T>C	*cypB*	cytochrome P450 CYP102A3	R643G (AGG>GGG)	1			
3,059,489	DEL	(C)8>7	*ytnP, trmB*	putative quorum-quenching lactonase/tRNA (guanine-N(7)-)-methyltransferase	intergenic (-88/+58)				1
3,324,382	SNP	T>C	*yunD*	putative nuclease/nucleotidase/phosphoesterase	H226R (CAT>CGT)			1	
3,474,002	SNP	A>G	*yvbI*	conserved protein of unknown function	I211V (ATC>GTC)				1
3,975,557	DEL	(A)5>4	*cydC*	ABC membrane transporter (ATP-binding protein) required for cytochrome bb' function (reductant efflux pump)	coding (1235/1704 nt)		1		
3,994,666	SNP	G>A	*yxjL*	two-component response regulator [YxjM]	A100T (GCG>ACG)			1	

KO7S WT TALE #9-12									
Position	**Mutation Type**	**Sequence Change**	**Gene**	**Product**	**Details**	**A9 F56 I13 R1**	**A10 F54 I2 R2**	**A11 F58 I1 R3**	**A12 F59 I7 R4**
52,896	SNP	T>C	*veg*	Protein veg	V45A (GTT>GCT)		1	1	
62,658	DEL	Δ2 bp	*mfd*	Transcription-repair-coupling factor	coding (2229-2230/3534 nt)				1
484,682	SNP	A>G	*dctP*	C4-dicarboxylate transport protein	Q325R (CAG>CGG)			1	
924,071	INS	(A)5>6 R9 R10 R11 R12	*katA, SsuB*	Vegetative catalase/Aliphatic sulfonates import ATP-binding protein SsuB	intergenic (-51/-228)	1	1	1	1
1,158,754	SNP	A>G	*argC*	N-acetyl-gamma-glutamyl-phosphate reductase	T229T (ACA>ACG)	1			
1,282,039	INS	(G)6>7	*yjqA, yjqB*	Uncharacterized protein yjqA/UPF0714 protein yjqB	intergenic (-97/-9)				1
1,380,183	SNP	T>C	*ktrD*	Ktr system potassium uptake protein D	M360T (ATG>ACG)				1
1,473,627	SNP	A>G	*fruA*	PTS system fructose-specific EIIABC component	Q445R (CAA>CGA)	1			
1,574,708	SNP	A>G	*ylmH*	Putative RNA-binding protein ylmH	Q6R (CAG>CGG)				1
1,653,530	SNP	C>T	*codY*	GTP-sensing transcriptional pleiotropic repressor CodY	T125I (ACA>ATA)				1
1,653,878	SNP	A>G	*codY*	GTP-sensing transcriptional pleiotropic repressor CodY	Y241C (TAT>TGT)		1	1	
1,653,890	SNP	T>C	*codY*	GTP-sensing transcriptional pleiotropic repressor CodY	L245P (CTA>CCA)	1			
1,657,920	SNP	T>A	*fliG*	Flagellar motor switch protein FliG	V255D (GTC>GAC)	1			
1,658,610	SNP	C>T	*fliH*	putative flagellar assembly protein fliH	Q149* (CAA>TAA)		1	1	

1,669,467	SNP	C>T	*fliR*	Flagellar biosynthetic protein fliR	R204* (CGA>TGA)				1
2,235,729	SNP	C>T	*resD*	Transcriptional regulatory protein resD	E26K (GAA>AAA)	1			
2,383,370	SNP	T>C	*yqgX*	Uncharacterized protein yqgX	I14V (ATC>GTC)			1	
2,700,146	SNP	C>T	*lonA*	Lon protease 1	V167I (GTT>ATT)			1	
2,963,941	SNP	A>G	*mntA, menC*	Manganese-binding lipoprotein mntA/o-succinylbenzoate synthase	intergenic (-128/+152)	1			
2,972,835	SNP	T>C	*ytdA*	Putative UTP--glucose-1-phosphate uridylyltransferase	N192D (AAC>GAC)				1
3,182,935	SNP	T>C	*yusE*	Thioredoxin-like protein yusE	A103A (GCA>GCG)	1			
3,191,622	SNP	T>C	*yuzM*	Uncharacterized protein yuzM	M9T (ATG>ACG)	1			
3,407,798	SNP	A>G	*yvpB*	Uncharacterized protein yvpB	S112G (AGC>GGC)	1			

KO7 XYL TALE #13, #15, #16								
Position	Mutation Type	Sequence Change	Gene	Product	Details	A13 F56 I15 R1	A15 F58 I6 R3	A16 F57 I19 R4
681,397	SNP	A>G	*yerD*	putative membrane protein yerD/Hypothetical Protein	intergenic (-6/-103)		1	
683,800	SNP	A>G	*pcrA*	ATP-dependent DNA helicase pcrA	Q466R (CAG>CGG)			1
924,071	INS	(A)5>6	*katA*	Vegetative catalase/Aliphatic sulfonates import ATP-binding protein SsuB	intergenic (-51/-228)	1	1	1
1,653,712	SNP	G>A	*codY*	GTP-sensing transcriptional pleiotropic repressor CodY	A186T (GCA>ACA)		1	
1,653,890	SNP	T>C	*codY*	GTP-sensing transcriptional pleiotropic repressor CodY	L245P (CTA>CCA)	1		
1,655,979	DEL	Δ1 bp	*fliF*	Flagellar M-ring protein fliF	coding (431/1596 nt)		1	
1,659,332	DEL	Δ141 bp	*fliI*	Flagellum-specific ATP synthase fliI	coding (418-558/1317 nt)	1		
1,662,121	DEL	(AGCTGAAAGCC)2>1	*fliK*	putative flagellar hook-length control protein fliK	coding (806-816/1464 nt)			1
1,803,627	SNP	G>T	*pksN*	frameshift (+1) in homopolymer run at nucleotide position 6839 in gene	pseudogene (6260/16546 nt)		1	
2,096,896	SNP	T>C	*yodT*	putative aminotransferase YodT	I350M (ATA>ATG)			1
2,529,258	SNP	T>C	*yrkH*	Uncharacterized protein yrkH	Y280C (TAT>TGT)	1		
2,629,680	SNP	A>G	*cymR*	HTH-type transcriptional regulator CymR	L104P (CTC>CCC)			1
2,688,788	SNP	T>C	*ysxE*	Uncharacterized protein ysxE	E19G (GAG>GGG)	1		
2,963,941	SNP	A>G	*mntA*	Manganese-binding lipoprotein mntA/o-succinylbenzoate synthase	intergenic (-128/+152)			1

2,963,948	SNP	G>A	*mntA*	Manganese-binding lipoprotein mntA/o-succinylbenzoate synthase	intergenic (-135/+145)	1	1	
3,010,076	SNP	A>G	*yubD*	putative MFS-type transporter yubD	L386P (CTG>CCG)	1		
3,248,155	SNP	C>T	*nhaK*	Sodium, potassium, lithium and rubidium/H(+) antiporter	A15T (GCG>ACG)			1
3,414,697	SNP	A>G	*nagBA*	Glucosamine-6-phosphate deaminase 1	I101V (ATT>GTT)		1	
3,617,990	DEL	Δ580 bp	*ywkB*	[U712_18650],[U712_18655]	1			ywkB
3,772,635	SNP	A>G	*dltB*	Protein dltB	Y334C (TAT>TGT)		1	
3,831,557	SNP	G>A	*licT*	Transcription antiterminator LicT/Uncharacterized protein yxiP	intergenic (-6/+90)	1		

KO7S XYL TALE #17-20									
Position	Mutation Type	Sequence Change	Gene	Product	Details	A17 F55 I4 R1	A18 F53 I5 R2	A19 F53 I19 R3	A20 F53 I6 R4
125,619	SNP	C>T	*rpoC*	DNA-directed RNA polymerase subunit beta'	R21C (CGT>TGT)	1	1	1	1
907,728	SNP	G>T	*perR*	Peroxide operon regulator	G70C (GGT>TGT)	1			
908,027	SNP	T>C	*ygzB*	UPF0295 protein ygzB	D107G (GAT>GGT)		1		
924,071	INS	(A)5>6	*katA*	Vegetative catalase/Aliphatic sulfonates import ATP-binding protein SsuB	intergenic (-51/-228)		1		1
924,079	SNP	T>A	*katA*	Vegetative catalase/Aliphatic sulfonates import ATP-binding protein SsuB	intergenic (-59/-220)			1	
1,364,839	DEL	(A)8>7	*ykoS*	putative membrane protein ykoS	coding (30/1695 nt)				1
1,395,470	SNP	A>G	*kinD*	Sporulation kinase D	L186P (CTA>CCA)		1		
1,653,734	SNP	A>G	*codY*	GTP-sensing transcriptional pleiotropic repressor CodY	E193G (GAG>GGG)				1
1,663,565	SNP	G>A	*flgE*	Flagellar basal-body rod protein flgG	G116R (GGG>AGG)				1
1,666,671	SNP	T>A	*fliY*	Flagellar motor switch phosphatase FliY	L321* (TTA>TAA)	1			
1,667,358	SNP	G>A	*fliZ*	Flagellar biosynthetic protein fliZ	W37* (TGG>TAG)		1		
1,668,072	DEL	Δ12 bp	*fliP*	Flagellar biosynthetic protein fliP	coding (172-183/666 nt)			1	
1,969,141	SNP	T>C	*proH*	Pyrroline-5-carboxylate reductase 1	I259V (ATC>GTC)				1
2,199,002	SNP	C>T	*cheR*	Chemotaxis protein methyltransferase/Nucleoside diphosphate kinase	intergenic (-20/+220)			1	

2,307,456	SNP	T>C	*yqjA*	Uncharacterized protein yqjA	Y305C (TAT>TGT)		1		
2,391,082	SNP	C>T	*glpG*	Rhomboid protease gluP	R72K (AGA>AAA)			1	
2,572,416	SNP	T>C	*adhB*	putative zinc-type alcohol dehydrogenase-like protein AdhB	C371R (TGT>CGT)			1	
2,963,948	SNP	G>A	*mntA*	Manganese-binding lipoprotein mntA/o-succinylbenzoate synthase	intergenic (-135/+145)	1	1	1	1
2,996,490	SNP	T>C	*rnpA*	16S ribosomal RNA	noncoding (5/1554 nt)			1	
3,586,470	SNP	C>T	*ureB*	Urease subunit beta	R60H (CGT>CAT)		1		
3,617,854	SNP	A>G	*ywkB*	putative transporter YwkB	M113T (ATG>ACG)			1	
3,813,837	SNP	T>C	*pepT*	Peptidase T	L304P (CTT>CCT)				1
3,892,629	SNP	C>T	*iolI*	Inososе isomerase	E12K (GAA>AAA)	1			

BS168 XYL ALE #21-24									
Position	**Mutation Type**	**Sequence Change**	**Gene**	**Product**	**Details**	**A21 F66 I1 R1**	**A22 F68 I1 R1**	**A23 F66 I1 R1**	**A24 F66 I1 R1**
121760	SNP	T>C	*rlmG, rpoB*	23S rRNA m2G1835 methyltransferase/RNA polymerase (beta subunit)	intergenic (+87/-159)				1
529425	SNP	G>A	*trnS-Leu2, ydcL*	tRNA-Leu/ICEBs1 mobile element: integrase	intergenic (+3/+80)				1
944623	SNP	C>T	*perR*	transcriptional regulator (Fur family)	A46V (GCT>GTT)			1	
944637	SNP	T>A	*perR*	transcriptional regulator (Fur family)	F51I (TTT>ATT)		1		
961043	INS	(46 bp)1>2	*katA, ssuB*	vegetative catalase 1/aliphatic sulfonate ABC transporter (ATP-binding protein)	intergenic (-57/-351)				1
961054	SNP	A>G	*katA, ssuB*	vegetative catalase 1/aliphatic sulfonate ABC transporter (ATP-binding protein)	intergenic (-68/-340)	1			
1100592	SNP	C>A	*yhfK*	putative NAD-binding epimerase / hydratase	A125E (GCA>GAA)	1			
1221661	SNP	C>A	*oppB*	oligopeptide ABC transporter (permease)	T23K (ACA>AAA)				1
1445938	SNP	G>A	*ykvQ*	putative sporulation-specific glycosylase (HGT island)	pseudogene (301/699 nt)				1
1580038	DEL	(T)9>8	*bshC, mraZ*	malate glucosamine cysteine ligase/inhibitor of RsmH and transcriptional regulator	intergenic (+43/-83)				1
1589333	SNP	G>A	*murD*	UDP-N-acetylmuramoylalanyl-D-glutamate ligase	A145T (GCC>ACC)	1			
1684043	SNP	T>C	*topA*	DNA topoisomerase I	I128T (ATC>ACC)	1			
1684164	SNP	A>C	*topA*	DNA topoisomerase I	Q168H (CAA>CAC)		1		
1685128	SNP	C>G	*topA*	DNA topoisomerase I	R490G (CGC>GGC)			1	

1693526	INS	(A)7>8	*fliF*	flagellar basal-body M-ring protein	coding (1031/1611 nt)			1	
1695562	INS	(A)7>8	*fliH*	flagellar export apparatus subunit of cytoplasmic ATPase	coding (309/627 nt)	1			
1704223	SNP	C>T	*fliZ*	flagellar regulatory protein	Q5* (CAA>TAA)				1
1716709	DEL	Δ24 bp	*sigD*	RNA polymerase sigma-28 factor (sigma-D)	coding (217-240/765 nt)		1		
2373376	SNP	C>A	*trpB*	tryptophan synthase (beta subunit)	R44L (CGT>CTT)				1
2891153	SNP	G>T	*leuB*	3-isopropylmalate dehydrogenase	A322E (GCG>GAG)			1	
3075913	SNP	G>A	*ytfP*	putative NAD(FAD)-utilizing dehydrogenase	E183K (GAA>AAA)			1	
3315630	SNP	G>A	*hom*	homoserine dehydrogenase	T167I (ACC>ATC)	1			

BS168 XYL ALE #21-24									
Position	Mutation Type	Sequence Change	Gene	Product	Details	A25 F54 I1 R1	A26 F57 I1 R1	A27 F53 I1 R1	A28 F56 I1 R1
28,183	SNP	C>T	*dnaX*	DNA polymerase III subunit gamma/tau	A457V (GCC>GTC)			1	
149,756	SNP	C>T	*rpoA*	DNA-directed RNA polymerase subunit alpha	L277F (CTT>TTT)				1
507,506	SNP	G>A	*rsbX*	Phosphoserine phosphatase rsbX	C105Y (TGC>TAC)	1			
570,528	SNP	T>C	*ydgF*	putative transporter YdgF	N194S (AAC>AGC)		1		
907,885	SNP	G>T	*perR*	Peroxide operon regulator	G122V (GGC>GTC)				1
924,077	SNP	T>C	*katA*	Vegetative catalase/Aliphatic sulfonates import ATP-binding protein SsuB	intergenic (-57/-222)			1	
1,071,581	SNP	C>T	*yhfQ*	Putative ABC transporter substrate-binding lipoprotein yhfQ	P272L (CCG>CTG)	1			
1,659,714	DEL	Δ771 bp	*fliI, fliJ*	[U712_08530],[U712_08535]		1			
1,662,478	SNP	T>G	*fliK*	putative flagellar hook-length control protein fliK	I388S (ATC>AGC)			1	
1,669,460	INS	(ATTAGGTATC)1>2	*fliR*	Flagellar biosynthetic protein fliR	coding (603/780 nt)				1
1,670,899	SNP	C>T	*flhA*	Flagellar biosynthesis protein flhA	Q50* (CAG>TAG)		1		
1,928,885	SNP	C>T	*ppsD*	Plipastatin synthase subunit D	L1973L (CTG>CTA)				1
2,803,182	DEL	Δ4 bp	*pyk*	Pyruvate kinase	coding (1215-1218/1758 nt)				1
3,070,780	SNP	C>T	*comA*	Transcriptional regulatory protein ComA	A175T (GCA>ACA)		1		

3,389,130	SNP	A>G	*mgfK*	UPF0052 protein yvcK	L89P (CTT>CCT)	1			
3,391,818	SNP	G>A	*trxB*	Thioredoxin reductase	S65F (TCT>TTT)		1		
3,512,331	SNP	G>A	*ywtF*	Putative transcriptional regulator ywtF	G80E (GGA>GAA)				1
3,596,507	SNP	A>C	*murAA*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase 1	M203R (ATG>AGG)				1
3,596,615	SNP	G>A	*murAA*	UDP-N-acetylglucosamine 1-carboxyvinyltransferase 1	A167V (GCT>GTT)		1		
3,969,880	SNP	T>C	*walK*	Sensor histidine kinase WalK	D554G (GAT>GGT)	1			
3,970,969	SNP	G>A	*walK*	Sensor histidine kinase WalK	A191V (GCA>GTA)		1		

Chapter 5

TECHNO-ECONOMIC ASSESSMENT OF ON-SITE ENZYME PRODUCTION USING LIGNOCELLULOSIC DDGS

Jasper L. S. P. Driessen[1], Sheila I. Jensen[1], Solange I. Mussatto[2], Alex T. Nielsen[1], John M. Woodley[3*]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, 220, 2800 Kongens Lyngby, Denmark

[2] Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 223, 2800 Kongens Lyngby, Denmark

[3] Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads 228A, 2800 Lyngby, Denmark

*** Corresponding authors**
John M. Woodley jw@kt.dtu.dk

This chapter is prepared for submission

Abstract

One of the hurdles to realising industrial-scale second-generation biorefineries is the high operational cost associated with the use of enzymes. Replacing the current glucose-based substrates for the production of enzymes with cheaper alternatives can offer a way to lower the enzyme production cost. To this end, the major by-product of the corn-based bioethanol industry, called Distillers' dried grain with solubles (DDGS), is an interesting candidate. DDGS has a rich nutritional composition for fermentation, is increasingly available as production of liquid biofuels increases, and allows the possibility of integrating enzyme production on-site of the biorefinery. In this study, we processed DDGS by steam explosion pretreatment and enzymatic hydrolysis to obtain relevant conversion yields and a techno-economic assessment was performed. Key parameters to improve the economic potential of DDGS-based on-site enzyme production were identified using a sensitivity analysis and applied in an optimistic scenario. Although DDGS-based enzyme production was not economically viable for the base case assuming experimentally derived yields, possible modifications were identified that could significantly improve the economics of the process. These include omitting the costly drying of DDGS and improved extraction of the considerable amount of protein present in the feedstock. In parallel, continuous optimization of both the cell factory and fermentation process designs are key to ensure adequate enzyme yields.

Nomenclature

5-HMF	Hydroxymethylfurfural
C	Equipment cost of the unit operation
$CAPEX_{DDGS}$	Capital cost associated with DDGS hydrolysate production
$\mathrm{CAPEX_{enzyme}}$	Capital cost associated with enzyme production
$\mathrm{C_{CO}}$	Corn oil cost associated with enzyme production
$\mathrm{C_{CSL}}$	Corn steep liquor cost associated with enzyme production
C_{DDGS}	DDGS-based medium production cost
C_{DFC}	Direct fixed cost associated with DDGS hydrolysate production
CDS	Condensed Distillers's Solubles
$\mathrm{C_E}$	Annual electricity cost associated with enzyme production
$\mathrm{C_{enzyme}}$	Enzyme production cost
$CEPCI$	Chemical Engineering Plant Cost Index
$C_{facility}$	Facility-dependent cost associated with DDGS hydrolysate production
$\mathrm{C_{FO}}$	Fixed operational cost associated with enzyme production
$\mathrm{C_{Glucose}}$	Glucose cost associated with enzyme production
$\mathrm{C_{HN}}$	Residual nutrient cost associated with enzyme production
C_{labour}	Labour cost associated with DDGS hydrolysate production

$C_{materials}$	Material cost associated with DDGS hydrolysate production
C_{NH_3}	Ammonium cost associated with enzyme production
C_{SO_2}	Sulfur dioxide cost associated with enzyme production
$C_{start-up}$	Single investment required to prepare the new factory for operation and validation
$C_{utilities}$	Operational cost associated with DDGS hydrolysate production
C_{WC}	Funds tied up in operating the biorefinery
DDGS	Distillers' Dried Grain with Solubles
HPLC	High-performance liquid chromatography
NREL	National Renewable Energy Laboratory
$OPEX_{DDGS}$	Operational cost associated with DDGS hydrolysate production
$OPEX_{enzyme}$	Operational cost associated with enzyme production
SCP	Single cell protein
T	Temperature
TEA	Techno-economic assessment
TS	Thin Stillage
UWGP	United Wisconsin Grain Producers
v	Annual amount of biomass processed in the biorefinery
WDG	Wet distiller's grains
x	Scaling factor

5.1 Introduction

Biorefineries can significantly contribute to the more sustainable production of chemicals, materials, and fuels as they use renewable biomass as a feedstock instead of fossil fuels [1,2]. While first-generation biorefineries using easily accessible sugar- or starch-based carbon are economically viable, this practice has received criticism as it can compromise other sustainable development goals, such as zero hunger and global food security [3–6]. The realization of second-generation biorefineries circumvents this problem by using lignocellulosic biomass as a feedstock. Although lignocellulose is an abundant and cheap resource, polymeric carbohydrates are not freely accessible for microbial fermentation [7,8]. For this reason, pretreatment is required to break the molecular bonds between the main constituents of lignocellulose before (hemi)cellulases are applied to convert the biomass to fermentable sugars.

Despite the significant improvement in the performance of commercial cellulases during the last years, enzymes still represent a significant cost factor for cellulosic biorefineries [9]. To achieve the transition to a bio-economy and the associated replacement of bulk petrochemicals by bio-based alternatives, the availability of bulk amounts of low-cost enzymes is paramount [10]. A frequently proposed strategy to reduce enzyme production costs is to replace conventional substrates with cheaper alternatives, as the cost of carbon is often more than 50% of the total direct cost of production [11–14].

To this end, the major by-product of the conventional corn-based dry-grind bioethanol process, called Dried Distillers Grains with Solubles (DDGS), is an interesting candidate for several reasons. First, the global production volume of bioethanol is expected to rise to 134.5 billion liters in 2024, and the production of DDGS is estimated to increase concomitantly [15–18]. The selling of DDGS as a co-product is responsible for approximately 22% of the value output of a dry-grind bioethanol plant, making it crucial to the economic viability of the bioethanol industry [16,19]. However, as there is a risk that this market could be saturated, alternative ways to convert DDGS into high-value products

are currently explored [15,16]. Second, using DDGS as a substrate for microbial fermentation is regarded as promising, since the composition of DDGS is rich in carbon, organic nitrogen, and other micronutrients required for the bio-manufacturing of a variety of products, including organic acids, biofuels, and in this context hydrolytic enzymes [16,20–23].

As it is likely that future second-generation biorefineries will be either integrated or co-located near first-generation biorefineries as a result of feedstock availability (corn stover), this offers possibilities to produce the enzymes on-site [13]. In this setup, DDGS biomass could serve as a feedstock for enzyme production and the produced enzymes could then be used in both the first- and second-generation biorefineries (**Figure 1**). On-site enzyme production offers the possibility to omit the cost associated with purification, formulation, and transportation of the enzyme and use the whole fermentation broth instead [10,24,25].

The focus of this study is to explore potential cost reductions by replacing glucose as a feedstock for enzyme production by DDGS biomass, rather than comparing on-site and off-site enzyme production. A detailed techno-economic assessment, developed by the NREL, provides a reference for a glucose-based on-site enzyme production facility integrated in an second-generation corn stover biorefinery [26,27]. For simplicity, the comparison between glucose- and DDGS-based enzyme productions are made assuming a similar process design, i.e. on-site enzyme production integrated into a second-generation biorefinery.

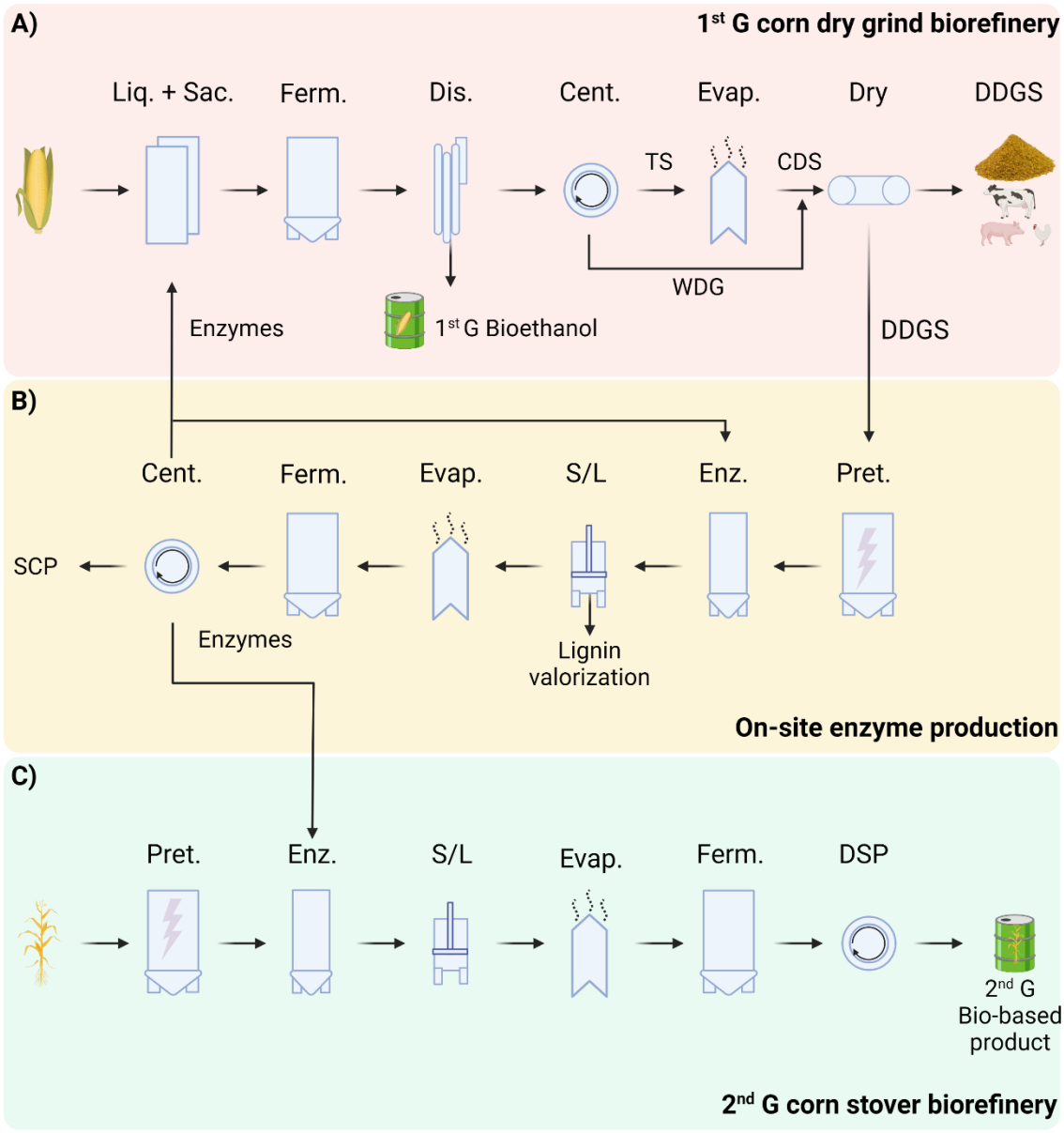


Figure 1 A) Schematic overview of the current dry-grind bioethanol production process and production of DDGS as the major by-product. B) An on-site enzyme production facility as a strategy to valorize DDGS. C) An adjacent 2nd generation corn stover biorefinery using low-cost enzymes. Abbreviations: Liq. Liquefaction, Sac. Saccharification, Ferm. Fermentation, Cent. Centrifugation, TS. Thin stillage, WDG. Wet distiller's grains, CDS. Condensed Distiller's Solubles, Evap. Evaporation, Dry. Drying, DDGS. Distiller's Dried Grains with Solubles, Pret. Pretreatment, Enz. Enzymatic hydrolysis, S/L. Solid Liquid separation, SCP. Single Cell Protein.

Previous techno-economic assessments have shown that reducing enzyme production costs could be reduced by replacing the relatively expensive glucose syrup by either less expensive sugar streams or lignocellusic biomass [13,28]. However, unlike the previous studies using lignocellulose, we determine biomass conversion yields in the process and account for the additional costs associated with biomass pretreatment and enzymatic hydrolysis. First, biomass compositional analysis, biomass pretreatment and enzymatic hydrolysis experiments were performed to provide the yields needed to estimate the required amount of DDGS biomass. Subsequently, a techno-economic analysis was performed to explore whether DDGS-based enzyme production can lead to cost reductions for on-site enzyme production compared to the use of glucose.

5.2 Material and Methods

5.2.1 Materials

Glucose, xylose, arabinose, acetic acid, formic acid, furfural, and 5-HMF were determined by high-performance liquid chromatography (HPLC), using a Dionex Ultimate 3000 UHPLC + Focused System (Dionex Softron GmbH, Germany) equipped with a Bio-Rad Aminex column HPX-87H (300mm× 7.8mm) ((60 C, mobile phase 5.0mM sulfuric acid, 0.6 mL/min). Sugars and acids were determined by a Shodex RI-101 refractive index detector, whereas 5-HMF and furfural were analyzed using ultraviolet measurements at 254nm. Chemicals were purchased from Sigma-Aldrich (Saint Louis, USA), unless otherwise stated.

5.2.1.1 DDGS Compositional Analysis

DDGS was kindly provided by the United Wisconsin Grain Producers (UWGP, Friesland, WI, USA). The compositional analysis of raw and pretreated DDGS biomass included the determination of hexane-, water-, and ethanol- extractable material, the main carbohydrate polymers (cellulose, hemicellulose, lignin, and starch), protein, and inorganic materials (ash). This was performed as follows. Sample preparation of dried biomass was executed using a laboratory mill (PX-MFC 90D, Kinematica™Polymix™, 1 mm sieve) as described by the NREL [29]. Subsequently, moisture content of the non-extracted DDGS biomass was measured by using a thermobalance MB 163-M (VWR International, Denmark). Crude protein content of non-extracted biomass was determined by the Dumas method as described by Kim et al. for DDGS (nitrogen factor 5.9) [30]. Residual starch present in the DDGS biomass was measured using a kit from Megazyme (K-TSTA-100A). The hexane-, water-, and ethanol extractives were determined using a Soxhlet extractive apparatus, as described by AOAC92039 (ether is substituted by hexane) and the NREL [31,32]. Protein concentration present in the water extractives were determined using a colorimetric assay (Biorad RC DC™).

Subsequently, structural carbohydrates, soluble/insoluble lignin, and ash fractions of the extracted biomass were determined as described previously by the NREL [33,34].

5.2.1.2 DDGS Steam Explosion Pretreatment

Prior to steam explosion pretreatment, the DDGS biomass was soaked in a dilute acid solution for 20 minutes (0.5% w/w sulfuric acid, 20% w/w dry mass). The DDGS slurry was filtered using a pressure press (Fischer HP25-M). Wet solids (44.8% dry mass) were processed in batches of approximately 2.2 kg using a 10 L steam explosion reactor as described by Palmqvist et al. [35] (190 °C, 5 minutes residence time, 12.5 bar, Knislinge Mekaniska Verkstad AB designed by Process- & Industriteknik AB, Lund University, Sweden). Batches of pretreated slurry were pooled and subsequently separated by solid-liquid filtration. Wet solids were stored at -20 °C for enzymatic hydrolysis.

5.2.1.3 Enzymatic Hydrolysis

The pretreated DDGS biomass was diluted by adding 0.1 M sodium citrate buffer, leading to 10, 17.5, and 25% solid content (w/V) (final pH 5.1). The enzymatic hydrolysis was performed using an enzyme load of 5, 25, and 45 FPU/ g cellulose (Optimash® F200, International Flavors & Fragrances Inc.). Suspended solutions were incubated in glass bottles (VWR) for 72h, placed on a horizontal Bottle Roller system housed in an incubator (Thermo ScientificTM, 80 rpm, 50 °C). After the enzymatic hydrolysis, the liquefied biomass was filter-sterilized and concentrations of glucose, xylose, arabinose, acetic acid, 5-HMF, and furfural were determined by HPLC.

5.2.2 Modelling

5.2.2.1 Overview

The aim of this techno-economic model is to compare an on-site enzyme production facility using DDGS biomass or glucose syrup as a fermentation feedstock. The glucose-based enzyme production model (Area 400) developed by the NREL is used as a reference [26,27]. Accordingly, the annual enzyme production rate was fixed on 2280 tons per year as assumed in the reference [26]. The DDGS-based model includes costs for all unit operations to convert DDGS biomass to the enzyme product. Wastewater treatment was not included in this model, as the focus of this study was to compare different substrates for on-site enzyme production facility as part of an overarching biorefinery process.

5.2.2.2 DDGS Feedstock

As DDGS is currently a by-product of the dry-grind bioethanol process, it was assumed readily available to be fed into the steam explosion reactor of an on-site enzyme facility without additional costs. Although DDGS contains a considerable amount of macromolecules, only glucan, xylan, arabinan, and crude protein are modelled to be converted in enzyme product. The annual feedstock needed to produce the required amount of enzyme was 38,933 dry ton for the base case scenario. Moisture content was assumed to be 10% as an average as described elsewhere [36].

5.2.2.3 Steam Explosion

DDGS biomass (0.45 w/w dry matter), water, and sulfuric acid are added to the steam explosion unit. Required chemical were assumed stored in storage takes and a heat exchanger was used to recover waste heat of the pretreated slurry (80% heat recovery) [37]. Biomass slurry and chemicals were transported through centrifugal pumps. The effects of different pretreatment methods is often roughly compared by using a severity factor [38]. This factor is based on the applied residence time and temperature according to Equation 1:

Equation 1

$$\log(R_0) = \log(t \times e^{\frac{T-100}{14.75}})$$

where t is the residence time in minutes, T is the temperature in ° C and 100 is the used reference temperature. The applied conditions in our steam explosion pretreatment led to a severity factor of 3.5. Conversion yields of glucose, xylose, and arabinose for the steam explosion pretreatment were based on experimental data (Figure 2). As biomass pretreatment leads to the unwanted formation of compounds that are inhibitory to the enzymatic hydrolysis and fermentation, the pretreated slurry is fed to a detoxification tank [39].

5.2.2.4 Detoxification Unit

Detoxification was assumed to be performed by an overliming step with 5% excess ammonia and a subsequent neutralization step with sulfuric acid as previously described [37].

5.2.2.5 Enzymatic Hydrolysis

Detoxified slurries fed to the enzymatic hydrolysis. Experimentally determined glucose, xylose, and arabinose yields (25% w/V, 72h) are assumed for the cost estimation. Enzyme loading of 20 mg per g cellulose (as previously described [26]) was assumed to be valid for the applied enzyme loading during the performed experiments (25 FPU / g cellulose). As the outgoing sugar stream is too dilute for industrial fed-batch applications, a mechanical vapour recompression (MVR) evaporator is assumed to concentrate the outgoing sugar stream.

5.2.2.6 Enzyme Fermentation

In this model, enzymes are assumed to be produced in 5 parallel aerobic submerged fermentation (120h, 300 m^3, 80% working volume) as described in detail by the NREL [14]. The enzyme yield during microbial fermentation is dependent on the substrate components, protein composition of the enzyme, and the generic cell mass composition of the used cell factory, and the maintenance requirements of the cell factory. The overall enzyme yield is defined as the sum of the yields on the various substrates present in the DDGS-hydrolysate. Substrates include glucose, xylose, arabinose, and protein. Enzyme yield on the monomeric

sugars was based on the NREL 2011 biorefinery model [14]. Using the assumptions of the NREL (an overall molar selectivity of carbon to 31% protein, 4% cell mass, and 65% CO_2), a mass yield of 0.24 g enzyme per g sugar is used for the base case. Furthermore, it is assumed proteins in the DDGS hydrolysate are readily available and converted into enzymes at an efficiency of 90%.

5.2.3 Economic Analysis

5.2.3.1 Enzyme Production Costs

Enzyme production costs include capital costs and operational costs (**Equation 2**), which are based on a previous study performed by the NREL [27].

Equation 2

$$C_{enzyme} = CAPEX_{enzyme} + OPEX_{enzyme}$$

$OPEX_{enzyme}$ consist of fixed operational cost (labor, maintenance, insurance, tax) and variable cost for glucose, corn steep liquor, corn oil, host nutrients, ammonia, sulfuric acid, and electricity (**Equation 3**).

Equation 3

$$OPEX_{enzyme} = C_{FO} + C_{Glucose} + C_{CSL} + C_{CO} + C_{HN} + C_{NH_3} + C_{SO_2} + C_E$$

In this study, the economics of replacing glucose, corn steep liquor, and corn oil with DDGS-based hydrolysate medium is evaluated. Hence, the question to be answered is whether the production cost of DDGS-based hydrolysate is less than the cumulative cost of glucose syrup, corn steep liquor, and corn oil (**Equation 4**).

Equation 4

$$C_{DDGS} < C_{Glucose} + C_{CSL} + C_{CO}$$

Capital and residual operational costs related to the enzyme fermentation are assumed equal to the values of the NREL study (Equation 5).

Equation 5

$$\Delta CAPEX_{enzyme} + \Delta C_{FO} + \Delta C_{HN} + \Delta C_{NH_3} + \Delta C_{SO_2} + \Delta C_E = 0$$

5.2.3.2 DDGS-hydrolysate Production Costs

The production of DDGS-based medium includes capital and operational costs ($CAPEX_{DDGS}$, $OPEX_{DDGS}$) associated to the additional unit of operations, being steam explosion and enzymatic hydrolysis (Equation 6).

Equation 6

$$C_{DDGS} = CAPEX_{DDGS} + OPEX_{DDGS}$$

The total capital investment is a sum of direct fixed capital (C_{DFC}), start-up cost ($C_{start\text{-}up}$), and working capital (C_{WC}) [40]. The C_{DFC} includes direct costs (equipment cost, installation, piping, instrumentation, electrical, insulation buildings, yard improvement, and auxiliary facilities cost) and the related indirect costs (construction, engineering, contractor's fee, and contingency). Direct cost is a fixed fraction of the equipment cost and indirect costs are dependent on the direct cost (Table S3). $C_{start\text{-}up}$ is the specific investment required to prepare a new plant for operation and validation.

C_{WC} is the funds (raw materials, consumables, labor, utilities) needed to operate the biorefinery. It is assumed these funds are retrieved upon closure of the factory and therefore working capital is not included in the estimation of $CAPEX_{DDGS}$ (Equation 7).

Equation 7

$$CAPEX_{DDGS} = C_{DFC} + C_{start-up}$$

Equipment cost was based on previous studies (**Equation 8**) [14,37,41].

Equation 8

$$\frac{C}{C_0} = \frac{v}{v_0}^{x}$$

With C as the equipment cost of the unit operation, v as the annual amount of biomass processed in the biorefinery (kg/year), x as the scaling factor, and all 0-indexed variables referring to the original capacity of the reference plant [24].

Start-up costs are defined as the one-time investment needed to prepare the new facility for operation. $CAPEX_{DDGS}$ per kg enzyme was calculated by assuming straight-line depreciation for a period of 15 years.

$OPEX_{DDGS}$ consist of feedstock costs, other material costs, labor dependent costs, facility dependent costs and utilities as previously described [37,40] (**Equation 9**,

Table S4).

Equation 9

$$OPEX_{DDGS} = C_{DDGS} + C_{materials} + C_{labour} + C_{facility} + C_{utilities}$$

C_{DDGS} was based on the determined yields of each unit of operation and fixed annual enzyme production (2280 tons per year) [26]. As DDGS is commonly sold as animal feed, the value is largely based on the protein content [42]. Therefore, it is assumed that the value of DDGS used in the current study depreciates linearly with protein content, i.e. one-third of the average market value of 2020 ($175 per dry ton, C_{DDGS}= $117) [36]. Other material costs include enzymes, ammonia, sulfuric acid, and water [26]. Cellulase loading was assumed to be 20 mg g^{-1} cellulose as previously described [14]. Other materials (sulfuric acid, ammonia, water), labor (including basic rate, benefits, supervision, operating supplies, and administration) and utilities (cooling-, heating agent, and electricity),

and requirements for steam explosion pretreatment are taken from Baral et al. [37].

All costs were adjusted for the analysis year 2020 (Equation 10).

Equation 10

$$C_{Year\ of\ analysis} = C_0 \frac{CEPCI_{Analysis\ year}}{CEPCI_0}$$

With C_0 being the original cost and the Chemical Engineering Plant Cost Index (CEPCI) indices for the corresponding years [43].

5.2.3.3 Sensitivity Analysis and Optimization

The base case was estimated based on economic and technical parameters from performed wet-lab experiments and literature data. As these inherently entail a degree of uncertainty, the sensitivity of selected parameters on the operational DDGS hydrolysate costs were evaluated by a change of $\pm 20\%$. Subsequently, the best-case scenario DDGS substrate cost was estimated using optimistic assumptions. Whenever Equation 4 was valid, (i.e. scenarios where the enzyme price decreased compared to glucose-based media), and operational cost calculations were iterated to account for the lower enzyme cost.

5.3 Results and Discussion

When estimating the amount of DDGS needed for an on-site enzyme production process, four conversion factors need to be determined: 1) the amount of carbohydrates (glucan, xylan, arabinan) and protein present in the DDGS, 2) the yield of these compounds after the pretreatment, 3) enzymatic hydrolysis, and 4) the yield from hydrolysate to enzyme during fermentation. First, the experimentally determined yields will be discussed. Subsequently, economic analysis for the use of DDGS for enzyme production for different scenarios is presented and compared to the glucose-based process.

5.3.1 DDGS Conversion Factors

5.3.1.1 Biomass Characterization

In general, the values for various components of the DDGS biomass determined in this study agree with the literature [44] (Table 1). The total extractable material, obtained by water-, ethanol-, and hexane-extraction, was a significant part of the total biomass, adding up to 45.3%. The majority of the water-extractable material originates from the thin stillage syrup sprayed on the dried distillers' grains [44]. This syrup includes mainly fermentation by-products, soluble oligo- and monosaccharides, proteins, and organic acids. Around 4.2% of the total 22.1% of the protein was present in the water extractives. Ethanol extractable material (4.6%) contains chlorophyll, waxes, and other minor components, while the hexane extractable material (11.2%) mostly contains lipids/corn oil. Total glucan content was 12.9%, of which 2.5% was found to be in the form of starch. Formic acid and 5-HMF are counted as glucan, as it is likely that these compounds are derived from glucose rather than other hexoses during the biomass characterization procedure. Glucan content is lower compared to what was reported previously [44]. Besides variation between DDGS sources, this deviation can be explained because extracted biomass was used in our approach for the carbohydrate determination, preventing water-

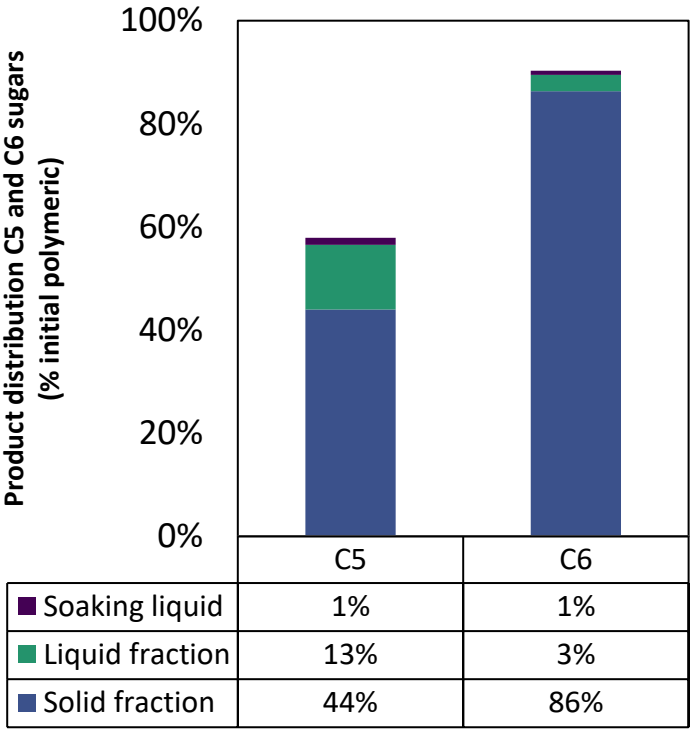
soluble sugars present in the water extractives to be counted double [44]. Xylan, arabinan, mannan, galactan, and furfural cumulatively accounted for 19.4% of the biomass. In the perspective of using DDGS as feedstock for enzyme production, the total carbohydrate content and protein content are important. The DDGS feedstock used in this study originates from a biorefinery process in which part of the protein is separated and sold as a high-value protein product [45]. As such, the DDGS biomass used contained less protein and more carbohydrates compared to conventional DDGS (21% vs 31% crude protein content) [46]. Soluble and insoluble lignin was relatively low in the DDGS biomass (7.4%), which is consistent with earlier reports [16]. The remaining dry biomass is composed of acetate (0.8%) and ash (0.1%).

Table 1 Compositional analysis of raw and pretreated DDGS biomass

	Raw DDGS wt.%	St. dev.	Pretreated DDGS wt.%	St. dev.
Dry matter	89.1	1.8	37.1	0.8
Water extractives	29.5	0.6	22.6	1.7
Hexane extractives	11.2	1.3	12.2	0.4
Ethanol extractives	4.6	1.0	7.3	0.7
Lignin	7.3	1.4	12.0	0.7
Glucan	12.9	0.6	16.0	0.6
Xylan	9.9	0.5	6.3	0.1
Arabinan	6.1	0.5	3.5	0.0
Mannan	0.4	0.1	0.3	0.2
Galactan	1.2	0.1	1.4	0.3
Furfural (anhydrous)	1.9	0.4	1.2	0.1
Acetate	0.8	0.0	0.7	0.0
Protein	20.9	0.8	30.4	0.8
- of which are water soluble	4.2	0.1	9.2	0.0
Ash	0.1	0.0	0.2	0.1
Total	102.6%		102.7%	

5.3.1.2 Steam Explosion Pretreatment

Steam explosion pretreatment with the addition of dilute sulfuric acid catalyst (0.5% w/w) is designed to solubilize the hemicellulosic fraction and improve the subsequent enzymatic hydrolysis step. The efficiency of the applied pretreatment was assessed by measuring the sugars liberated to the pretreatment liquid and the determination of the solid fraction. Data from the steam explosion have been grouped into C5 carbon molecules (xylose, arabinose, and furfural) and C6 carbon molecules (glucose, galactose, mannose, and 5-HMF). Most of the C6 sugars (92%) were accounted for as polymers in the solid fraction or as monosaccharides in the liquids, nearly closing the mass balance of the pretreatment (Figure 2). In contrast, this is not the case for the C5 group (58%). Steam explosion has been shown to lead to the formation of oligomers of xylan for both DDGS or other corn biomass such as corn stover [22,47,48]. In line with this, a significant part of the C5 sugars is likely present in the form of solubilized oligosaccharides in the liquid fraction. The conversion of monosaccharides and furan derivatives to humins, adsorption and/or condensation of furfural to other biomass components such as lignin and experimental errors can also contribute to the incomplete mass balance during pretreatment [49]. As expected, the pretreatment removed a considerable part of the hemicellulosic fraction from the biomass, while the C6 sugars were retained in the solid fraction, as is reflected in the composition of the pretreated material (Table 1).



	C5	C6
Soaking liquid	1%	1%
Liquid fraction	13%	3%
Solid fraction	44%	86%

Figure 2 Product distribution after steam explosion pretreatment

The severity factor (Section 5.2.2.3) is an often-used parameter to correlate pretreatment conditions to sugar yields. However, reliable correlations between severity factors and yields of both pretreatment and the subsequent enzymatic hydrolysis are hard to establish. Therefore, these correlations should be used with caution [37,50]. Nonetheless, the observed value of glucose yield (3%) in our study is in line with what was modelled by Baral and Shah [37]. On the other hand, yields for monomeric xylose were lower in our study (13% vs 37%), as most of the xylan is probably released as oligomeric sugars.

5.3.1.3 Enzymatic Hydrolysis

Enzymatic hydrolysis of the pretreated DDGS was performed for varying substrate (10-25% w/V) and enzyme loadings (5-45 FPU/g glucan) (**Figure 3**). After 72h, yields of glucose, xylose, and arabinose varied from 0.53-0.81 g glucose / g cellulose, 0.32-0.52 g xylose / g xylan and 0.19-0.29 g arabinose / g arabinan respectively. Glucose yields agree with previously reported values for enzymatic hydrolysis of DDGS pretreated by liquid hot water and AFEX methods (0.78 g glucose / g glucan, solid loading 15-17% w/V) [51]. Previous studies modelled

glucose and xylose yields based on the severity factor of the steam explosion pretreatment [37]. Glucose yields observed in our study closely resemble the modelled values (0.68 g glucose / g glucan), while xylose yields during enzymatic hydrolysis are lower than modelled (0.65 g xylose / g xylan) [37]. It has been previously stated that the efficient hydrolysis of the arabinoxylan fraction in the corn pericarp fiber of DDGS requires a multitude of enzymes, such as pectinase, feruloyl esterases, β- glucosidases, and xylanases [52]. Enzymatic hydrolysis of raw DDGS biomass (25% w/v, 25 FPU/g glucan) led to a 50% decrease in glucose yield, confirming the importance of the steam explosion pretreatment to make the cellulose more accessible to the hydrolytic enzymes.

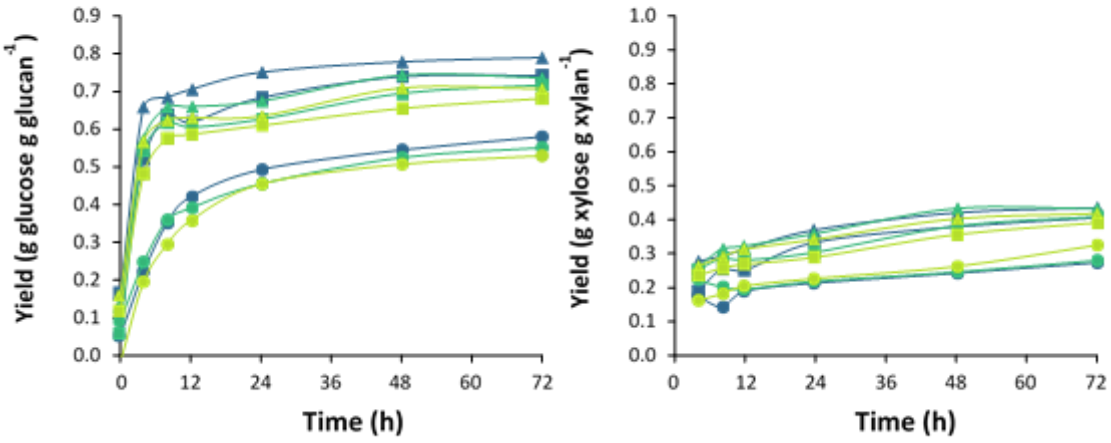


Figure 3 Enzymatic hydrolysis of steam explosion pretreated pulps. Solid loading (% w/V) is indicated by light green (10%), dark green (17. 5%) or blue (25%) color and enzyme load (FPU / g glucan) is indicated by ● (5), ■ (25) or ▲ (45).

5.3.1.4 Enzyme Fermentation

The amount of enzyme that can be produced from DDGS hydrolysate during fermentation is dependent on the flux distribution of substrate to cell growth, maintenance energy, and product. To model the use of complex media in such a fermentation, the overall enzyme yield on hydrolysate is defined as the sum of the yields on the various substrates. Enzyme yield on the monomeric sugars was based on the NREL 2011 biorefinery model [14]. This follows the assumption that cell factories are optimized to use C5-sugars as efficient as glucose. Without doubt, considerable efforts related to cell factory engineering are required for this assumption to be valid, as simultaneous consumption of different sugars by microbes is often regulated by catabolite repression [53]. Although models concerning glucose-based enzyme production assume the addition of protein to the fermentation in the form of corn steep liquor, enzyme yields are solely based on glucose or carbohydrates [14,28,54]. Stoichiometric coefficients for enzyme production reactions are depicted in most detail by Hong et al [54].

It is frequently observed that using complex organic carbon and nitrogen sources, such as peptones and yeast extract, lead to significantly higher enzyme production rates and titers compared to fermentations based solely on glucose and inorganic nitrogen source [55–57]. Previous studies for example showed significantly higher enzyme production could be reached when using protein-containing soybean hulls (12% crude protein) compared to sugarcane molasses [10,58]. Moreover, fed-batch fermentations with addition of 2% yeast extract (approximately 12.5-15 g L^{-1} protein, nitrogen factor 6.25) led to roughly three times higher enzyme titers compared to fermentation without an organic nitrogen source [13]. Unlike others, we consider the conversion of not only carbohydrates, but also protein to enzymes. It is assumed DDGS derived proteins are converted into enzymes at an efficiency of 90% to reach comparable increases in overall enzyme yield. The full conversion is considered too optimistic, as a part of the DDGS protein is likely directed to alternative requirements of the cell and the amino acids present in DDGS biomass likely differ from the exact amino acid composition of the enzyme product.

5.3.2 OPEX and CAPEX Estimation DDGS Hydrolysate

Using the compositional analysis and the yields for steam explosion pretreatment, enzymatic hydrolysis, and fermentation, the required annual amount of DDGS biomass was estimated for the fixed amount of enzyme production stated in the NREL report[27] (*v*, Equation 8) (Table 2). Subsequently, $CAPEX_{DDGS}$ and $OPEX_{DDGS}$ were determined as described in section 5.2.3 (Table S3 and Table S4). $CAPEX_{DDGS}$ amounted to \$23.1 M and translated into a base cost of \$ 0.68 kg^{-1} enzyme using DDGS-based medium. On the other hand, $OPEX_{DDGS}$ was a total of \$ 8.4 M, i.e. \$ 3.70 kg^{-1} enzyme. DDGS feedstock was found to be the major contributor to the operational costs, amounting to 54% of the $OPEX_{DDGS}$. The previous is in line with previous estimates for steam explosion pretreatment of corn stover [37].

With the provided information, it can be concluded that costs associated with producing DDGS-based hydrolysate C_{DDGS} were found to be greater than the cumulative cost of glucose, corn steep liquor, and corn oil (4.37 vs. 4.33 \$/kg enzyme) used in a traditional production setup. Therefore, it is clear one would not invest in an on-site enzyme facility using DDGS, under the current assumptions and experimentally determined yields, as it is not economically feasible. However, the costs for using DDGS in the base case scenario are very close to the cost of glucose-based enzyme production. Therefore, a sensitivity analysis was performed and more optimistic assumptions were evaluated.

Table 2 Conversion yields and cost estimation for the base case scenario

Glucan		
Biomass composition	0.13	g g^{-1}
Yield pretreatment	0.89	g g^{-1}
Yield enzymatic hydrolysis	0.68	g g^{-1}
Yield glucose / biomass (dw) [1)]	0.08	g g^{-1}
Xylan		
Biomass composition	0.10	g g^{-1}
Yield pretreatment	0.45	g g^{-1}
Yield enzymatic hydrolysis	0.42	g g^{-1}
Yield xylose / biomass (dw) [2)]	0.03	g g^{-1}
Arabinan		
Biomass composition	0.06	g g^{-1}
Yield pretreatment	0.41	g g^{-1}
Yield enzymatic hydrolysis	0.23	g g^{-1}
Yield arabinose / biomass (dw) [3)]	0.02	g g^{-1}
Yield sugar /biomass (dw) [1+2+3)]	0.13	g g^{-1}
Protein		
Biomass comp. protein/Biomass dw	0.21	g g^{-1}
Yield pretreatment, enzymatic hydrolysis	0.15	g g^{-1}
Yield protein / biomass (dw)	0.03	g g^{-1}
Enzyme		
Yield enzyme/sugar	0.24	g g^{-1}
Yield enzyme/protein	0.9	g g^{-1}
Yield enzyme/biomass (dw)	0.06	g g^{-1}

5.3.3 Sensitivity Analysis of C_{DDGS}

The analysis described in the previous section includes many assumptions regarding technical and economic parameters, which all entail a certain degree of uncertainty. Next to the uncertainties derived from experimental work, research efforts in developing more effective enzyme cocktails, cell factories, and optimized fermentation conditions can greatly affect the conversion of biomass to product in terms of yield and productivity [59–61]. Hence, evaluating the sensitivity of key parameters is crucial to identifying targets for future research efforts. Some parameters were selected based on their predicted impact on the process (enzyme yield during fermentation) or the fact that they are susceptible to change in the near future (DDGS market price, enzyme performance) (Figure 4).

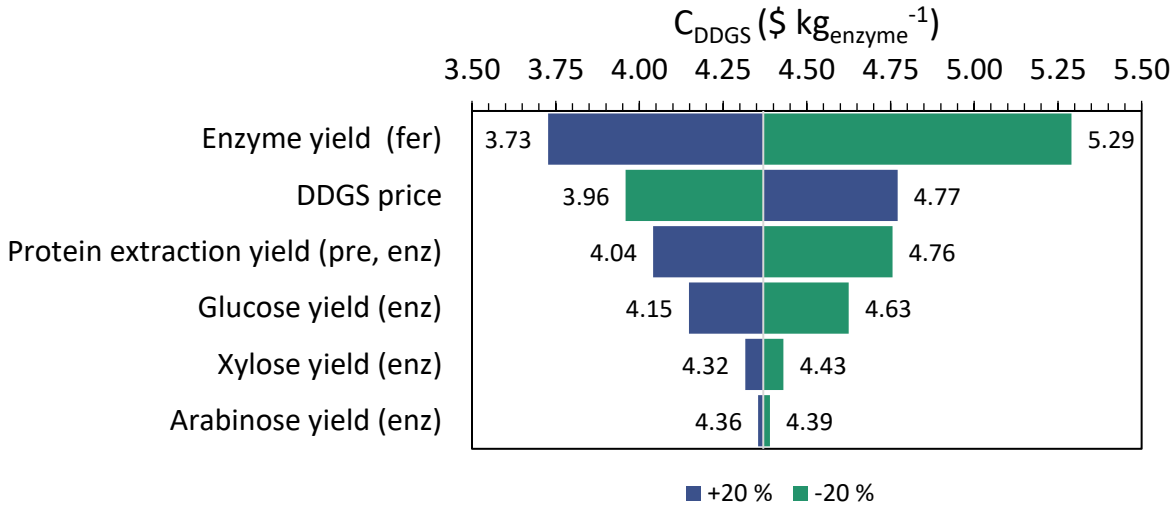


Figure 4 Sensitivity analysis for C_{DDGS}, using DDGS-based hydrolysate for on-site enzyme production. The output variable is defined as C_{DDGS} required to produce 1 kg of enzyme. Evaluated parameters refer to: enzyme yield on DDGS hydrolysate during fermentation (Enzyme yield (fer)), DDGS price, extracted fraction of total protein present in the original DDGS biomass (Protein yield (pre, enz), glucose, xylose and arabinose yields during enzymatic hydrolysis (Glucose yield, Xylose yield, Arabinose yield (enz).

The enzyme yield on hydrolysate was found to be the most sensitive parameter, as it affects the final cost in multiple ways. The enzyme yield directly relates to the required feedstock, influencing feedstock costs, operational costs, and capital costs. Unfortunately, it is also the most uncertain parameter to be estimated. Although metabolic models are increasingly applied to facilitate the estimation of conversion yields and rates to specific biochemicals or single amino acids, modelling anabolic reactions towards proteins is less explored, especially for complex media [62]. As such, iterations of scaled-down small-scale fermentation experiments using DDGS-based media and model adjustments are crucial to confirm the assumptions used in this study.

Although the fibrous DDGS used in this study is a less valuable by-product stream than conventional DDGS, its value is relatively high compared to other lignocellulosic biomass (such as corn stover) and represents a significant contribution to the operational costs [9,63]. As such, DDGS price was a sensitive input parameter, which agrees with previous reports for other biomass resources processed by steam explosion pretreatment [37,64].

The remaining parameters describe how efficient DDGS biomass is converted into the different substrates during pretreatment and/or enzymatic hydrolysis. As the protein to enzyme conversion during fermentation is assumed efficient, the amount of protein that is solubilized in the biomass hydrolysate during the pretreatment and enzymatic hydrolysis was found to be a sensitive parameter. The sensitivity of glucose yield is lower, as the subsequent conversion to enzyme is lower. Because only part of the xylose and arabinose are released during the enzymatic hydrolysis, sensitivity towards these yield parameters was lower than glucose.

5.3.4 Optimistic Scenario for Further Cost Reduction

Considering that the operational costs of DDGS utilization in the base case scenario are estimated from experimental and literature data, it is worthwhile to evaluate the applied assumptions to identify potential improvements. Selection of conditions that may change identified sensitive input parameters can lead to reduced operational costs.

Among the various costs of producing DDGS as a co-product during the dry-grind bioethanol process, drying of the WDGS is a very energy-intensive process and poses a considerable cost [20,42]. This step is necessary to prevent the DDGS from spoiling due to fungal infections and facilitate transport over long distances. However, in the current setup, DDGS is used as a feedstock for the on-site production of enzymes, offering a possibility to omit the costly drying step [20]. Accordingly, the mixed CDS and DWG are fed to the steam explosion pretreatment unit without further drying. Thus, the value of WDG is considered for the optimistic scenario (162 $/dry ton, 2020).

The pretreatment and enzymatic hydrolysis in this study focused on extracting fermentable sugars from the DDGS biomass for fermentation, with a concomitant extraction of only 14% of the original protein. Much effort has been invested in extracting proteins from DDGS using different methods, ranging from physical, chemical, and enzymatic treatments showing varying yields (30-90%) [23,65]. The thermal treatment during the drum drying stage has been discussed to decrease the solubility of the present protein. As discussed previously, omitting this step could have additional benefits for protein extraction purposes [16,20,65]. Protein extraction, mediated by enzymatic hydrolysis, was shown to be particularly promising in different studies, yielding up to 80-90% of the protein present in the DDGS biomass [66,67]. The addition of the proteases could offer a straightforward approach to increase the availability of proteins to serve as a rich peptone for fermentation. As such, a value of 80% was used for the optimistic scenario. Peptones are readily soluble in water, do not precipitate upon exposure to heat and are commonly used for industrial fermentation [68]. To prevent degradation of the (hemi)cellulases, proteases should only be added after the polymeric carbohydrates are hydrolysed. In addition,

immobilized proteases should be used or proteases should be inactivated (e.g. by short heat-treatment [69]) before fermentation to prevent the degradation of the enzyme products.

Lastly, the conversion yields for glucose, xylose, and arabinose used in the base case are based on experimental data of small-scale batch experiments. Process optimization in an industrial setting could improve the observed yields, and as such, many techno-economic studies consider higher yields for their process. In the latest assessment of the NREL, the used assumptions were 90% conversion of cellulose to glucose and xylan to xylose, as well as 85% arabinan to arabinose [63]. Although the NREL's assumptions are based on future targets, other experimental studies yielded 90% glucose conversion of DDGS biomass, by using a variety of enzymes and using low solid loading (5%) [51]. Similarly, a recent techno-economic assessment used sugar conversion yields of the same range [37,70–72]. Accordingly, the NREL assumptions were used for the optimistic scenario in this study.

Applying the optimistic values for the input parameters, the substrate cost is reduced to \$1.59 (**Figure 5**). The extraction of DDGS proteins for enzyme production was found to be the most critical parameter in reducing substrate cost. Although adding proteases leads to an increase in the raw material cost, the overall cost reduction is considerable. When accounting for the additional operational and capital costs related to enzyme fermentation (**Equation 6**), the resulting enzyme production cost (C_{enzyme}) in the optimistic scenario would amount to \$4.07 kg^{-1}, compared to the glucose-based enzyme fermentation cost of \$6.73 kg^{-1}. Although the likelihood that all parameters are as advantageous as presented may not be high, the significant cost reduction gap provides a considerable solution for a combination of parameters to lead to realistic cost reductions. For example, even when the conversion of protein to enzyme during fermentation is reduced from 90% to 50% for the optimistic scenario, the resulting total enzyme production costs (C_{enzyme}) would still reduce the cost to \$ 5.86 kg^{-1}.

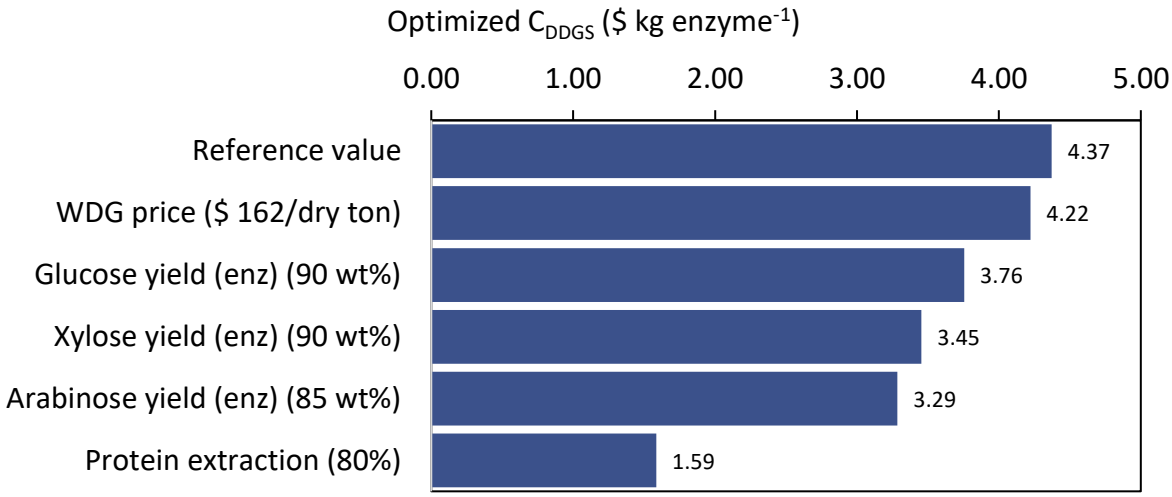


Figure 5 Reduction DDGS-substrate cost under optimistic scenario assumptions.

5.3.5 Process Evaluation

In this study, the annual enzyme production volume (2280 dry ton year^{-1}) was based on previous models concerning a glucose-based on-site enzyme facility integrated in a second-generation bioethanol biorefinery [26]. The feedstock requirements of DDGS-based enzyme production are approximately 39,000 dry ton for the base case scenario and 11,000 dry tons for the optimistic scenario. This corresponds to 36% and 11% of what is annually produced in a typical dry-grind bioethanol plant [42]. Produced enzymes in the current setup are directed to the dry grind bioethanol plant (amylases, glucoamylases), the enzymatic hydrolysis of the pretreated DDGS ((hemi)cellulases, proteases) and to local second-generation biorefineries ((hemi)cellulases). Only a minor part of the produced enzymes will be used for the hydrolysis of the DDGS components (base case: 89 dry ton cellulases; optimistic case 73 dry ton cellulases + proteases). The recent development of yeasts that are able to produce amylases and glucoamylases for the hydrolysis of starch will minimize the need for additional supplied enzymes [73,74]. Therefore, the amylases and glucoamylases used in the dry-grind bioethanol process will be minimal. As such, the majority of the produced enzymes can serve as a cheaper alternative for second-generation biorefineries.

5.4 Future Perspectives

Currently, enzymes used in biorefineries are purchased from a limited number of well-established multinational enzyme-producing companies. Compared to this conventional process, the model discussed in this study entails two main differences. The first is the location of production, being on-site at the biorefinery instead of an off-site centralized production facility. The second is the origin of the substrate, being lignocellulose in the proposed setup, while first-generation carbon is the status quo in conventional processes.

The main advantages associated with the off-site production are economies of scale, confidential fermentation processes and production hosts, optimized for high titer, rate, and yield (TRY indicators). On the other hand, on-site enzyme production benefits from cost savings associated with purification, concentration, formulation, and transportation of the enzyme product, as well as advantages related to shared utilities with the biorefinery in which it is integrated [14,59]. Besides costs, a life cycle assessment of enzyme production performed by Novozymes showed that 38% of the product's contribution to global warming is from formulation and recovery processes [75]. In addition, producing in close proximity to the place where the product is used can decrease supply risks.

Although DDGS-based enzyme production is compared to a glucose-based reference model in this study, the use of pure glucose syrup is not current industrial practice. Carbohydrates used in industrial enzyme production commonly originate from corn starch, sucrose, and glucose/maltose mixtures, while proteins derived from potato fruit juice, corn steep powder or soy bean meal are added [75]. Although the relative cost benefits associated with the use of DDGS in the optimistic scenario might be less prominent when it is compared to the substrates used in industrial practice, the use of DDGS as second-generation feedstock is still highly relevant from a sustainability perspective. Using DDGS-derived protein to replace soy-based feed is already common practice as the amino acid composition is considered to be relatively

comparable [23]. Likewise, this supports replacing soy-based media used in industrial enzyme processes by DDGS-based protein.

Although the opportunities to reduce enzyme production costs by producing on-site using DDGS are evaluated in this study, it would be a considerable undertaking, including five enzyme production reactors of 300 m^3. One may question whether the cost benefits outweigh the considerable investment in R&D and the operational and logistical risks on-site enzyme production entails. Biorefineries could be forgiven for "sticking to what works" and keep relying on externally purchased enzymes. Strong collaboration between biorefineries and enzyme-producing companies could lower the risks significantly. Enzyme-producing companies have the key advantage of decades of R&D efforts and first-hand industrial experience. Using highly engineered production hosts and optimized fermentation processes is crucial in obtaining the required performance needed to make on-site enzyme production feasible. In addition, enzyme-producing companies own highly relevant data that can be applied to further economic assessments, accurately estimating cost savings for each scenario. Accordingly, enzyme suppliers are able to adapt their strategy to lower enzyme costs and outrun the competition. The joint venture of POET-DSM is an example of this approach, in which an ethanol producer and enzyme manufacturer join forces [76]. However, the venture has experienced difficulties related to technological setbacks and uncertain policy, resulting in the announcement that they will pause their activities to focus on optimization of the process [77,78]. Nonetheless, the realisation of future on-site enzyme production strategies will likely be marked by similar collaborations.

5.5 Conclusions

Although this study shows that the replacement of glucose-based enzyme production by DDGS is not competitive under the experimentally determined yields, it is clear there are significant opportunities for further cost reductions by realistic adjustments of sensitive process parameters. The most straightforward parameters to optimize have a high impact and are relatively easy to incorporate. These include lowering the costs for DDGS by removing the drying step to produce the feedstock and optimising the extraction of protein present in DDGS feedstock. More impactful, and more challenging, is achieving a high substrate to product yield during fermentation. Iterative optimization cycles for both improvement of cell factories and fermentation process conditions will therefore be crucial. In addition, the ongoing development of enzyme cocktails with higher performance and innovation of pretreatment processes should lead to high recoveries of fermentable carbohydrates. In conclusion, on-site enzyme production using DDGS as a feedstock has a lot of potential to reduce enzyme production costs and can thereby contribute to the economic viability of future bio-based production processes.

5.6 Acknowledgements

This work was supported by the Novo Nordisk Foundation (NNF), Denmark (grant number NNF17SA0031362 and NNF10CC1016517) and The Technical University of Denmark (DTU).

We would like to thank International Flavors & Fragrances Inc. for kindly providing the Optimash® F200.

We would like to acknowledge the United Wisconsin Grain Producers for kindly providing the DDDS biomass.

The authors would like to thank Francesco Reggianini for his support with some of the experimental work.

The authors would like to acknowledge Christian Roslander and Mats Galbe for their advice and support during the steam explosion pretreatment (Lund University).

5.7 References to Chapter 5

1. Geoff Bell, A. *et al.* IEA BIOENERGY Task42 BIOREFINING Sustainable and synergetic processing of biomass into marketable food & feed ingredients, products (chemicals, materials) and energy (fuels, power, heat) The Netherlands (coordinator) Electronic copies. (2014).
2. IEA Bioenergy - Task 42. Biorefineries: adding value to the sustainable utilisation of biomass.
3. Naik, S. N., Goud, V. V, Rout, P. K. & Dalai, A. K. Production of first and second generation biofuels: A comprehensive review. doi:10.1016/j.rser.2009.10.003
4. Hazell, P. & Pachauri, R. K. BIOENERGY AND AGRICULTURE: PROMISES AND CHALLENGES. (2020).
5. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. Available at: https://sdgs.un.org/2030agenda. (Accessed: 3rd February 2022)
6. Straathof, A. J. J. *et al.* Grand Research Challenges for Sustainable Industrial Biotechnology. *Trends Biotechnol.* **37**, 1042–1050 (2019).
7. Popp, J., Lakner, Z., Harangi-Rákos, M. & Fári, M. The effect of bioenergy expansion: Food, energy, and environment. *Renew. Sustain. Energy Rev.* **32**, 559–578 (2014).
8. Deng, Y. Y., Koper, M., Haigh, M. & Dornburg, V. Country-level assessment of long-term global bioenergy potential. *Biomass and Bioenergy* **74**, 253–267 (2015).
9. Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A. & Blanch, H. W. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol. Bioeng.* **109**, 1083–1087 (2012).
10. Ellilä, S. *et al.* Development of a low-cost cellulase production process using Trichoderma reesei for Brazilian biorefineries. *Biotechnol. Biofuels* **10**, 30 (2017).
11. Chotani, G., Peres, C., Schuler, A. & Moslemy, P. Bioprocessing Technologies. *Bioprocess. Renew. Resour. to Commod. Bioprod.* **9781118175**, 133–166 (2014).
12. Sanford, K., Chotani, G., Danielson, N. & Zahn, J. A. Scaling up of renewable chemicals. *Curr. Opin. Biotechnol.* **38**, 112–122 (2016).
13. Ellilä, S. *et al.* Low-cost glucose-based cellulase production. *NWBC 2018 Proc. 8th Nord. Wood Biorefinery Conf.* 1–6 (2018).

14. Humbird, D. *et al.* Process design and economics for conversion of lignocellulosic biomass to ethanol. *NREL Tech. Rep. NREL/TP-5100-51400* **303**, 275–3000 (2011).

15. Chatzifragkou, A. *et al.* Biorefinery strategies for upgrading Distillers' Dried Grains with Solubles (DDGS). *Process Biochemistry* (2015). doi:10.1016/j.procbio.2015.09.005

16. Chatzifragkou, A. & Charalampopoulos, D. Distiller's dried grains with solubles (DDGS) and intermediate products as starting materials in biorefinery strategies. in *Sustainable Recovery and Reutilization of Cereal Processing By-Products* (2018). doi:10.1016/B978-0-08-102162-0.00003-4

17. RFA. *Annual Industry Outlook 2020*. (2020).

18. Fao & Oecd. *OECD-FAO Agricultural Outlook 2015-2024*. (2015).

19. Schwarck, R. *et al.* RFA ETHANOL INDUSTRY OUTLOOK 2021. (2021).

20. Iram, A., Cekmecelioglu, D. & Demirci, A. Distillers' dried grains with solubles (DDGS) and its potential as fermentation feedstock. *Applied Microbiology and Biotechnology* **104**, 6115–6128 (2020).

21. Perkis, D., Tyner, W. & Dale, R. Economic analysis of a modified dry grind ethanol process with recycle of pretreated and enzymatically hydrolyzed distillers' grains. *Bioresour. Technol.* **99**, 5243–5249 (2008).

22. Nghiem, N. P., Montanti, J. & Kim, T. H. Pretreatment of Dried Distiller Grains with Solubles by Soaking in Aqueous Ammonia and Subsequent Enzymatic/Dilute Acid Hydrolysis to Produce Fermentable Sugars. *Appl. Biochem. Biotechnol.* **179**, 237–250 (2016).

23. Zhao, J., Wang, D. & Li, Y. Proteins in dried distillers' grains with solubles: A review of animal feed value and potential non-food uses. *JAOCS, J. Am. Oil Chem. Soc.* **98**, 957–968 (2021).

24. Olofsson, J., Barta, Z., Börjesson, P. & Wallberg, O. Integrating enzyme fermentation in lignocellulosic ethanol production: Life-cycle assessment and techno-economic analysis. *Biotechnol. Biofuels* **10**, 1–14 (2017).

25. Barta, Z., Kovacs, K., Reczey, K. & Zacchi, G. Process design and economics of on-site cellulase production on various carbon sources in a softwood-based ethanol plant. *Enzyme Res.* **2010**, (2010).

26. Davis, R. *et al.* Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons. *Natl. Renew. Energy Lab.* 147 (2015).

27. R. Davis, L. Tao, C. Scarlata, and E. C. D. T. Biorefinery Analysis Process Models | NREL. Available at: https://www.nrel.gov/extranet/biorefinery/aspen-models/. (Accessed: 8th April 2022)

28. Johnson, E. Integrated enzyme production lowers the cost of cellulosic ethanol. *Biofuels, Bioprod. Biorefining* **10**, (2016).

29. Hames, B. *et al.* Preparation of Samples for Compositional Analysis: Laboratory Analytical Procedure (LAP); Issue Date 08/08/2008. (2008).

30. Kim, Y. *et al.* Composition of corn dry-grind ethanol by-products: DDGS, wet cake, and thin stillage. *Bioresour. Technol.* **99**, 5165–5176 (2008).

31. Sluiter, A., Ruiz, R., Scarlata, C., Sluiter, J. & Templeton, D. Determination of Extractives in Biomass: Laboratory Analytical Procedure (LAP); Issue Date 7/17/2005. (2008).

32. Association of Official Analytical Chemists. *AOAC Official Methods of Analysis*. (1984).

33. Sluiter, A. *et al.* Determination of Structural Carbohydrates and Lignin in Biomass: Laboratory Analytical Procedure (LAP) (Revised July 2011). (2008).

34. Sluiter, A. *et al.* Determination of Ash in Biomass: Laboratory Analytical Procedure (LAP); Issue Date: 7/17/2005. (2008).

35. Palmqvist, E. *et al.* Design and operation of a bench-scale process development unit for the production of ethanol from lignocellulosics. *Bioresour. Technol.* **58**, 171–179 (1996).

36. Hofstrand, D. Agricultural Marketing Resource Center | Agricultural Marketing Resource Center. Available at: https://www.agmrc.org/. (Accessed: 7th April 2022)

37. Baral, N. R. & Shah, A. Comparative techno-economic analysis of steam explosion, dilute sulfuric acid, ammonia fiber explosion and biological pretreatments of corn stover. *Bioresour. Technol.* **232**, 331–343 (2017).

38. Overend, R. & Chornet, E. Fractionation of lignocellulosics by steam-aqueous pretreatments. *Philos. Trans. R. Soc. London. Ser. A, Math. Phys. Sci.* **321**, 523–536 (1987).

39. Driessen, J. L. S. P., Maas, L. Van Der, Reggianini, F. & Mussatto, S. I. Effect of inhibitory compounds present in lignocellulosic biomass hydrolysates on growth of Bacillus subtilis. 1–23

40. Shah, A., Baral, N. R. & Manandhar, A. *Technoeconomic Analysis and Life Cycle Assessment of Bioenergy Systems*. *Advances in Bioenergy* **1**, (Elsevier, 2016).

41. Aden, A. *et al.* • • • • • • Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover. (2002).

42. Kwiatkowski, J. R., McAloon, A. J., Taylor, F. & Johnston, D. B. Modeling the process and costs of fuel ethanol production by the corn dry-grind process. *Ind. Crops Prod.* **23**, 288–296 (2006).

43. The Chemical Engineering Journal. The Chemical Engineering Plant Cost Index - Chemical Engineering. Available at: https://www.chemengonline.com/pci-home. (Accessed: 8th April 2022)

44. Kim, Y. *et al.* Composition of corn dry-grind ethanol by-products: DDGS, wet cake, and thin stillage. *Bioresour. Technol.* **99**, 5165–5176 (2008).

45. Jessen, H. Wisconsin ethanol plant will install Fluid Quip system. Available at: http://ethanolproducer.com/articles/12239/wisconsin-ethanol-plant-will-install-fluid-quip-system. (Accessed: 5th May 2020)

46. Belyea, R. L., Rausch, K. D. & Tumbleson, M. E. Composition of corn and distillers dried grains with solubles from dry grind ethanol processing. *Bioresour. Technol.* **94**, 293–298 (2004).

47. Öhgren, K., Galbe, M. & Zacchi, G. *Optimization of Steam Pretreatment of SO 2-Impregnated Corn Stover for Fuel Ethanol Production*. (2005).

48. Iram, A., Cekmecelioglu, D. & Demirci, A. Optimization of dilute sulfuric acid, aqueous ammonia, and steam explosion as the pretreatments steps for distillers' dried grains with solubles as a potential fermentation feedstock. *Bioresour. Technol.* **282**, 475–481 (2019).

49. Smit, A. & Huijgen, W. Effective fractionation of lignocellulose in herbaceous biomass and hardwood using a mild acetone organosolv process. *Green Chem.* **19**, 5505–5514 (2017).

50. Pedersen, M. & Meyer, A. S. Lignocellulose pretreatment severity – relating pH to biomatrix opening. *N. Biotechnol.* **27**, 739–750 (2010).

51. Kim, Y. *et al.* Enzyme hydrolysis and ethanol fermentation of liquid hot water and AFEX pretreated distillers' grains at high-solids loadings. *Bioresour. Technol.* **99**, 5206–5215 (2008).

52. Dien, B. S. *et al.* Enzyme characterization for hydrolysis of AFEX and liquid hot-water pretreated distillers' grains and their conversion to ethanol. *Bioresour. Technol.* **99**, 5216–5225 (2008).

53. Chen, R. & Dou, J. Biofuels and bio-based chemicals from lignocellulose: metabolic engineering strategies in strain development. *Biotechnol. Lett.* **38**, 213–221 (2016).

54. Hong, Y., Nizami, A. S., Pour Bafrani, M., Saville, B. A. & Maclean, H. L. Impact of cellulase production on environmental and financial metrics for lignocellulosic ethanol. *Biofuels, Bioprod. Biorefining* **7**, 303–313 (2013).

55. Nagar, S., Gupta, V. K., Kumar, D., Kumar, L. & Kuhad, R. C. Production and optimization of cellulase-free, alkali-stable xylanase by Bacillus pumilus SV-85S in submerged fermentation. *J. Ind. Microbiol. Biotechnol.* **37**, 71–83 (2010).

56. Chauhan, B. & Gupta, R. Application of statistical experimental design for

optimization of alkaline protease production from Bacillus sp. RGR-14. *Process Biochem.* **39**, 2115–2122 (2004).

57. Şahin, B., Öztürk, S., Çalık, P. & Özdamar, T. H. Feeding strategy design for recombinant human growth hormone production by Bacillus subtilis. *Bioprocess Biosyst. Eng.* **38**, 1855–1865 (2015).

58. Barbosa, F. F. *et al.* VARIATION IN CHEMICAL COMPOSITION OF SOYBEAN HULLS 1. (2008).

59. Dragone, G. *et al.* Innovation and strategic orientations for the development of advanced biorefineries. *Bioresource Technology* **302**, (2020).

60. Noorman, H. J. & Heijnen, J. J. Biochemical engineering's grand adventure. *Chem. Eng. Sci.* **170**, 677–693 (2017).

61. Woodley, J. M. Towards the sustainable production of bulk-chemicals using biotechnology. *N. Biotechnol.* **59**, 59–64 (2020).

62. Oddone, G. M., Mills, D. A. & Block, D. E. A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. *Metab. Eng.* **11**, 367–381 (2009).

63. Davis, R. *et al.* Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbon Fuels and Coproducts: 2018 Biochemical Design Case Update: Biochemical Deconstruction and Conversion of Biomass to Fuels and Products via Integrated Biorefinery Path. (2018).

64. Kumar, D. & Murthy, G. S. Impact of pretreatment and downstream processing technologies on economics and energy in cellulosic ethanol production. *Biotechnol. Biofuels* **4**, 1–19 (2011).

65. Chatzifragkou, A. *et al.* Extractability and characteristics of proteins deriving from wheat DDGS. *Food Chem.* **198**, 12–19 (2016).

66. Cookman, D. J. & Glatz, C. E. Extraction of protein from distiller's grain. *Bioresour. Technol.* **100**, 2012–2017 (2009).

67. Bandara, N., Chen, L. & Wu, J. Protein extraction from triticale distillers grains. *Cereal Chem.* **88**, 553–559 (2011).

68. Romero, E., Bautista, J., García-Martinez, A. M., Cremades, O. & Parrado, J. Bioconversion of corn distiller's dried grains with solubles (CDDGS) to extracellular proteases and peptones. *Process Biochem.* **42**, 1492–1497 (2007).

69. Deeth, H. C. Heat-induced inactivation of enzymes in milk and dairy products. A review. *Int. Dairy J.* **121**, 105104 (2021).

70. Liu, Z. H. *et al.* Effects of biomass particle size on steam explosion pretreatment performance for improving the enzyme digestibility of corn stover. *Ind. Crops Prod.* **44**, 176–184 (2013).

71. Bondesson, P. M., Galbe, M. & Zacchi, G. Ethanol and biogas production after steam pretreatment of corn stover with or without the addition of sulphuric acid. *Biotechnol. Biofuels* **6**, 1–11 (2013).

72. Zimbardi, F. *et al.* Acid impregnation and steam explosion of corn stover in batch processes. *Ind. Crops Prod.* **26**, 195–206 (2007).

73. Kumar, D. & Singh, V. Dry-grind processing using amylase corn and superior yeast to reduce the exogenous enzyme requirements in bioethanol production. *Biotechnol. Biofuels* **9**, 1–12 (2016).

74. IFF. IFF's Health & Biosciences Division Announces Agreement for Commercializing DSM Bio-based Products & Services Yeast Technology. Available at: https://www.dupontnutritionandbiosciences.com/news/iffs-health-biosciences-division-announces-agreement-for-commercializing-dsm-bio-based-products-services-yeast-technology.html. (Accessed: 26th April 2022)

75. Nielsen, P. H., Oxenbøll, K. M. & Wenzel, H. Cradle-to-gate environmental assessment of enzyme products produced industrially in denmark by novozymes A/S. *Int. J. Life Cycle Assess. 2007 126* **12**, 432–438 (2006).

76. Chandel, A. K., Garlapati, V. K., Singh, A. K., Antunes, F. A. F. & da Silva, S. S. The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization. *Bioresour. Technol.* **264**, 370–381 (2018).

77. Voegele, E. Poet-DSM halts Project Liberty production over RFS mismanagement | EthanolProducer.com. Available at: http://www.ethanolproducer.com/articles/16722/poet-dsm-halts-project-liberty-production-over-rfs-mismanagement. (Accessed: 14th April 2022)

78. Bioenergy International. POET-DSM to pause ethanol production and revert to R&D at Project LIBERTY | Bioenergy International. Available at: https://bioenergyinternational.com/poet-dsm-pause-ethanol-production-and-revert-to-rd-at-project-liberty/. (Accessed: 14th April 2022)

5.8 Supplementary material

Table S3 Cost breakdown of $CAPEX_{DDGS}$

DDGS base case		
Annual required biomass		38,932,869
1. Equipment Purchase Cost [37]	% purchase cost	$4,188,650
2. Installation	37	$1,560,069
3. Process Piping	35	$1,466,062
4. Instrumentation	40	$1,675,425
5. Insulation	5	$209,364
6. Electrical	10	$418,899
7. Buildings	45	$1,884,961
8. Yard Improvement	15	$628,263
9. Auxiliary Facilities	40	$1,675,425
10. TPDC (1-9)		$13,707,118
Total Plant Indirect Cost (TPIC)	% TPDC	$4,188,650
11. Engineering	20	$2,741,355
12. Construction	20	$2,741,355
13. TPIC (11+12)		$5,482,709
Total Plant Cost (TPC)		
14. TPC (10+13)		$19,189,827
Contractor's Fee & Contingency (CFC)	% TPC	
15. Contractor's Fee	5	$959,526
16. Contingency	10	$1,919,052
17. CFC (15+16)		$2,878,577
Direct Fixed Capital Cost		
18. DFC (14+17)		$22,068,405
	% DFC	
19. Working capital	10	$2,206,840
20. Start-up and validation cost	5	$1,103,420
$CAPEX_{DDGS}$ (18-20)		$23,171,825
CAPEX (kg^{-1} enzyme)		$0.68

Table S4 Cost breakdown of $OPEX_{DDGS}$

DDGS base case	Price ($ kg^{-1})	Amount (year^{-1})	Cost ($ year^{-1})
1. DDGS (Feedstock)	0.117	38932869	$4,549,025
2. Enzyme	6.29	89398	$601,940
3. Ammonia	0.55	423134	$249,192
4. Sulfuric acid	0.11	804393	$94,744
5. Water	0.0002	81094620	$17,367
6. Raw materials (1-5)			$5,512,268
7. Labor (h)	69	3470	$256,361
8. Electricity (kWh)	0.07	1257033	$94,219
9. Heating agent (kg)	0.012	22090850	$283,848
10. Cooling agent (kg)	0.0004	228809267	$98,000
11. Utilities (7-10)			$732,429
12. Maintenance	10% purchasing cost		$418,865
13. Miscellaneous costs (insurance, tax, overhead)	8% DFC		$1,765,472
14. Facility dependent cost (12+13)			$2,184,337
$OPEX_{DDGS}$ (6+11+14)			$8,429,034
OPEX (kg^{-1} enzyme)			$3.70