THE UNIVERSITY
of EDINBURGH

Engineering Microbes For Consolidated Bioprocessing: New Approaches In The Light Of Synthetic Biology

Marcos Valenzuela Ortega

This thesis is presented for the degree of Doctor of Philosophy

The University of Edinburgh

2021

Declaration

I declare that the work presented in this thesis has been composed and completed by myself, unless otherwise stated. This work has not previously been submitted for any other degree or personal qualification.

Marcos Valenzuela Ortega

Acknowledgements

First, I must thank my supervisor, Prof. Christopher French, for supporting me, spurring my scientific curiosity and always having an open door for advice and helpful criticism. I would also like to thank the suggestions and encouraging feedback of my PhD committee (Dr Andrew Free, Prof. Steve E. Fry, and Prof. Alistair Elfick). And of course, to the BBSRC and Eastbio for funding this project and making it possible.

Many thanks to all the French lab members, who have supported me and made this journey more enjoyable. To Chao-Kuo and Alejandro for setting the groundwork that allowed this project. To Darek, Paulina, Jaime providing so much advice. To Felipe, Andreas, and Prabu for creating a great environment and thoughtful (and sometimes silly) conversations. And to Maxine and Flo for making the lab merrier and brighter.

To Joan and Grant for their support and shared drams, especially during the toughest months. To Martí, Alberto, Kino and Ismael for reminding me of life outside academia, and keeping me sane and healthy. To my friends from Vilassar de Mar and Barcelona for their emotional support. To Darwin for being a source of wisdom. And to Maria for her encouragement and for everything we learnt together in the past.

Finally, I must thank my family above everyone. To Raquel and Dani, always my role models. And to my parents, Luisa and Antonio, because if everyone received as much support as they have given me, the world would be a better place.

Abstract

Mitigating climate change calls for a reduction in emissions of greenhouse gases, mainly CO_2. Twenty-seven per cent of all CO_2 emissions come from sources hard to eliminate, including aviation, shipping, and long-distance land transportation. To scale-up production, a transition from first-generation biofuels (edible plant biomass feedstock) to second-generation biofuels (non-edible lignocellulosic plant biomass) is required. However, lignocellulosic bioprocesses struggle to reach economic viability due to the high cost of added enzymes required to digest cellulose. A microbe that was able to generate the cellulase to digest cellulose and able to generate desired products (in what is called a consolidated bioprocess) is not known to exist in nature. Decades of research in synthetic biology have not been successful in creating it due to multiple challenges, such as generating candidate microbial strains and testing them. In this work, novel methodological approaches were developed to overcome these challenges, thus increasing the likelihood of future research finding the ideal microbe.

Firstly, the generation of candidate strains was improved by a re-design of the paradigm of modular cloning, based on a vector design called JUMP (Joint Universal Modular Plasmids). Complex multi-gene plasmids can be built from standard DNA parts in a reliable and automation-friendly way using modular cloning systems, based on Golden Gate cloning (which uses type IIS restriction enzymes). However, current standards lack the flexibility to change the microbial host and to perform assemblies to optimise genes of a group. JUMP vector backbones are based on the Standard European Vector Architecture (SEVA), and also have additional modular cloning sites to modify vectors for specific purposes. The experimental results presented here showed that these features allowed use of the vector system in different organisms and reduced the number of assemblies required to optimise and test multi-gene constructs.

Secondly, among different explored approaches to test strains, using fluorescent growth reporters was found to have the properties required to screen strains in a faster and more relevant way. Cellulose is insoluble in water, and consequently, previous analytical methodologies to assess the cellulolytic capacity of microorganisms used soluble analogues of cellulase substrates or depended on separation steps which are difficult to do in a fast and high-throughput way. The data

presented here showed that expression of fluorescence genes, providing a direct measure of growth, could be measured without separation of cellulose. An *Escherichia coli* strain expressing cellulases CenA and Chu2268, previously shown to bestow cellulolytic ability on *E. coli*, was confirmed to be able to grow using cellulose as sole carbon source, which demonstrated the use of fluorescent growth reporters to detect cellulolytic activity.

Finally, a strain benchmark was built and characterised to allow screening of expression of cellulolytic genes. A collection of modular parts encoding cellulolytic proteins and secretion signal peptides was built to allow future screening, and the effect of genetic and environmental factors in the measurement of fluorescence as a reporter of growth was investigated. It was found that volumetric conditions and medium additives critically affected the capacity of *E. coli* to grow at the expense of cellulose. A methodology was developed to allow growth measurement in 96-well plates.

In conclusion, this study lays the foundations to establish a faster research cycle to generate and screen microbes that can utilise cellulosic biomass to produce valuable bioproducts and biofuels.

Lay Summary

It is necessary to leave fossil fuels behind because the carbon dioxide released from their use is the main cause of global warming. The best replacement for some uses of fossils fuels (such heavy land vehicle transportation) is biofuels because they are made from plant material, obtained from plants that capture carbon dioxide from the air.

The best example of biofuel is bioethanol. It is produced by yeast from glucose sugar using biotechnology —the living organism transforms the starting material into the final product. Glucose is easily obtained from food crops, but it would be better to generate biofuels from non-edible plant material because it is very abundant, and it would not affect food availability. Glucose is also present in non-edible plant material but locked away as lignocellulose. Before it can be used by the organisms it needs to be broken down to glucose using enzymes called cellulases. But these are expensive to add and make the process too costly. This problem could be solved if the microbes themselves could make the cellulases as well as turning glucose into biofuels. Unfortunately, this ideal microorganism does not exist yet, and researchers have not succeeded in modifying existing organisms to this end.

This project aimed at solving the problems that make it so difficult to modify microbes to break down lignocellulose into glucose. First, a new technique to build DNA molecules with the instructions for microbes to make cellulases was developed. This aimed to make it easier and faster to generate a variety of new microbial prototypes. Second, a new way of testing these prototypes that are producing cellulases and breaking down lignocellulose correctly was developed. This should allow testing many more candidates at the same time. Lastly, a prototype microbe was generated that demonstrates these two new technologies work together to generate and test more potential microbes to make biofuels from lignocellulose.

The project's results will allow future researchers to generate the ideal biofuel-producing microorganism sooner, reducing fossil fuels dependency.

Resumen para público no experto

Es necesario dejar atrás los combustibles fósiles porque el dióxido de carbono liberado por su uso es la principal causa del calentamiento global. La mejor substitución para algunos usos de combustibles fósiles (como el transporte por tierra de vehículos pesados) son los biocombustibles, ya que se generan con materia vegetal, obtenida de plantas que capturan dióxido de carbono del aire.

El mejor ejemplo de biocombustible es el bioetanol. Se produce a partir del azúcar glucosa usando la biotecnología (un organismo vivo transforma la materia prima en el producto final). La glucosa es fácil de obtener de cultivos comestibles, pero es preferible generar biocombustibles de materia vegetal no comestible, ya que esta es muy abundante y no afectaría a la disponibilidad de alimentos. La glucosa también está presente en la parte no comestible de plantas, pero está bloqueada en forma de lignocelulosa. Para que los microorganismos puedan usar la lignocelulosa, hay que descomponerla en glucosa usando unas enzimas llamadas celulasas, pero estas son muy caras de añadir y hacen el proceso demasiado costoso. Este problema se solucionaría si los mismos microorganismos pudieran producir las celulasas y también pudieran transformar la glucosa en combustible. Desafortunadamente, este microorganismo ideal aún no existe y los investigadores no han conseguido modificar organismos existentes para este fin.

El objetivo de este proyecto ha sido resolver los problemas que dificultan la modificación de microbios para descomponer lignocelulosa en glucosa. Para ello, primeramente, se ha desarrollado una nueva técnica para construir moléculas de ADN con instrucciones para que los microorganismos produzcan celulasas, generando una diversidad de nuevos microbios prototipos más fácil y rápidamente. Segundo, se desarrolló una nueva manera de probar si estos prototipos generan celulasas y descomponen la lignocelulosa correctamente, permitiendo ensayar más candidatos simultáneamente. Finalmente, se ha generado un microorganismo prototipo que demuestra como estas dos nuevas tecnologías se combinan para crear y probar más potenciales microbios que permitan producir biocombustibles a partir de lignocelulosa.

Los resultados de este proyecto permitirán a futuros investigadores generar el microorganismo productor de biocombustible ideal antes, reduciendo la dependencia de los combustibles fósiles.

Resum per a públic no expert

És necessari deixar enrere els combustibles fòssils perquè el diòxid de carboni que s'allibera amb el seu ús és la principal causa de l'escalfament global. El millor substitut per alguns usos de combustibles fòssils, com ara el transport de vehicles pesants per terra, són els biocombustibles, perquè es generen amb matèria vegetal, obtinguda de les plantes que capturen el diòxid de carboni de l'aire.

El millor exemple de biocombustible és l'etanol. Es produeix a partir del sucre glucosa fent servir biotecnologia (un organisme viu transforma la matèria primera en el producte final). La glucosa és fàcil d'obtenir a partir de cultius comestibles, però seria preferible generar biocombustibles amb matèria vegetal no comestible, ja que és molt abundant i no afecta la disponibilitat d'aliments. La glucosa també es troba a la part no comestible de plantes, però està bloquejada en forma de lignocel·lulosa. Per a què els microorganismes puguin fer servir la lignocel·lulosa primer cal descompondre-la en glucosa fent servir uns enzims anomenats cel·lulases, però afegir-les és molt car i fan el procés massa costós. Aquest problema es resoldria si els mateixos microorganismes poguessin produir cel·lulases i també poguessin transformar la glucosa en combustible. Malauradament aquest microorganisme encara no existeix i els investigadors no han aconseguit modificar organismes existents per a aquesta finalitat.

L'objectiu d'aquest projecte ha estat resoldre els problemes que dificulten la modificació de microbis per a què descomponguin lignocel·lulosa a glucosa. Per això, primerament s'ha desenvolupat una nova tècnica per construir molècules d'ADN amb les instruccions per tal que els microorganismes produeixin cel·lulases, generant una diversitat de nous microbis prototip amb més facilitat i rapidesa. En segon lloc, s'ha desenvolupat una nova forma de provar si aquests prototips generen cel·lulases i trenquen la lignocel·lulosa correctament, el qual permet assajar més candidats a la vegada. Finalment, s'ha generat un microorganisme prototip que demostra com aquestes dues noves tecnologies es combinen per a crear i provar més microbis potencials que permetin produir biocombustible a partir de lignocel·lulosa.

Els resultats d'aquest projecte permetran a futurs investigadors generar el microorganisme productor de biocombustible ideal més aviat, reduint la dependència dels combustibles fòssils.

Contents

List of Figures

List of Tables

Abbreviations

A/C	Autoclaved
AbR	Antibiotic resistance gene
Amp	Ampicillin
AU	Arbitrary units
AUC	Area under the curve
BG	β-glucosidase
Bp	Base pairs
BSA	Bovine serum albumin
BX	β-xylosidase
Camp	Chloramphenicol
Carb	Carbenicillin
CASFR	Cellulase activity screening using fluorescent reporters
CBD	Cellulose binding domain
CBH	Cellobiohydrolase
CBM	Carbohydrate-binding module
CBP	Consolidated Bioprocess
CCS	Carbon capture and storage
CDS	Coding sequence
CFU	Colony forming unit
CMC	Carboxymethyl-Cellulose
DBTL	Design-Build-Test-Learn

DMSO	Dimethyl sulphoxide
DNA	Desoxyribonucleic acid
EB	Elution buffer
EC	Enzyme Comission number, given by the International Union of Biochemistry and Molecular Biology
EG	Endoglucanase
EX	Endoxylanase
F/S	Filter-sterilised
GHG	Greenhouse gas
IBPM	Ideal biofuel-producing microorganism
iGEM	International Genetically Engineered Machine Competition
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JUMP	Joint Universal Modular Plasmids
Kan	Kanamycin
LB	Lysogeny Broth medium
LPMO	Lytic Polysaccharide Monooxygenases
M9t	Minimal medium M9 salts complemented with thiamine
M9y	Minimal medium M9 salts complemented with yeast extract
MABEL	Mutagenesis with blunt-end ligation
MoClo	Modular cloning
Mw	Molecular weight
NEB	New England Biolabs

OD	Optical density
OriT	Origin of Conjugation
OriV	Origin of Replication
PASC	Phosphoric acid-swollen cellulose
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PNK	Polynucleotide Kinase
RBS	Ribosome binding site
RNA	Ribonucleic acid
rpm	Revolutions per minute
SEVA	Standard European Vector Architecture
SN	Supernatant
Sp	Spectinomycin
SP	Signal peptide
SSF	Simultaneous saccharification and fermentation
TAE	Tris-Acetate-EDTA buffer
TE	Tris-EDTA buffer
Tet	Tetracycline
TSS	Transformation and storage solution
v/v	Volume per volume
w/v	Weight per volume
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactoside

Chapter 1 Introduction

1.1 Climate change and fossil fuels

1.1.1 Clear causes, dark consequences

The global temperature of Earth is rising at unprecedented rate due to the increase of greenhouse gases in the atmosphere (Figure 1.1). Direct effects of global warming vary between regions and include increase in frequency and severity of extreme weather events, decreased precipitation, droughts, sea-level rise, acidification and decrease of oxygen levels in the oceans (Victor, Zhou et al. 2014). These will cause impacts in human and natural systems, including biodiversity loss, risks to food and water availability, human health risks, and reduced economic growth.

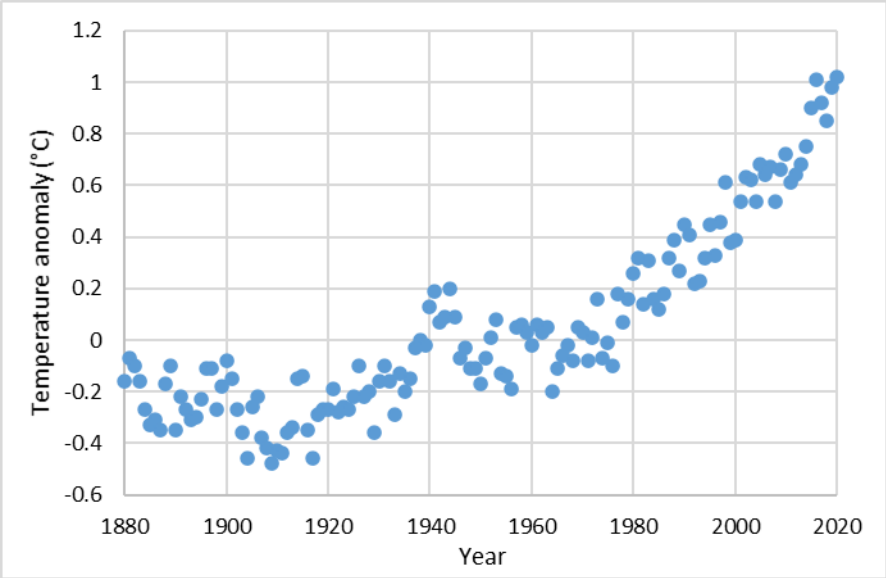


Figure 1.1. Change in global surface temperature relative to the average temperatures between 1951 and 1980. Data obtained from NASA's Goddard Institute for Space Studies (GISS), available on: https://climate.nasa.gov/vital-signs/global-temperature/.

The cause of global warming is the accumulation of greenhouse gases in the atmosphere due to anthropogenic activities (Victor, Zhou et al. 2014). The primary greenhouse gas (GHG) is carbon dioxide, released from fossil fuel combustion, followed by industrial activities and land use (Figure 1.2). In addition to CO_2, the gases with the next strongest greenhouse potential (methane, nitrous oxide, and fluorinated gases), represent 24 % of emissions in units of equivalent CO_2 (which is calculated considering lifetime in the atmosphere and warming effect). The

extensive use of fossil carbon (as petroleum, coal, and natural gas), which are the primary source of GHG, is associated with the historical economic development of the last centuries; fossil fuels are abundant, widespread, with low cost to use and without penalties to dispose of and provide energy and materials. Consequently, they are deeply rooted in society.

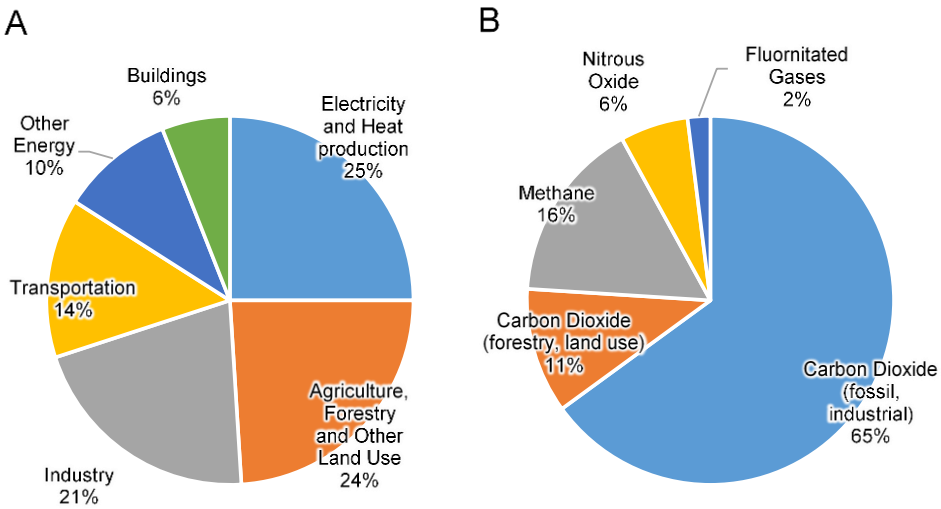


Figure 1.2. Emissions of greenhouse gases (GHG) in 2010. A: Percentage of Direct GHG emissions in CO_2-equivalent units, allocated per sector. B: Percentage of emissions in CO_2-equivalent units per groups of gases. Data from Victor, Zhou et al. (2014).

The impacts of global warming become more significant with the increase in temperature: there will be more substantial effects on natural and human systems, and adaptation to the new climate will be more difficult (Masson-Delmotte, Zhai et al. 2018). Not taking action on global warming is not acceptable and most UN members signed the Paris Agreement in 2016 placing the long-term objective of maintaining the global average temperature increase below 2 °C with respect to pre-industrial levels and pursue efforts to limit it to 1.5 °C. Although the Paris agreement was ambitious, the signing nations did not agree on putting in place any binding mechanism to apply specific measures.

Climate change mitigation measures (defined as “interventions to reduce emissions or enhance the sinks of greenhouse gases” (Victor, Zhou et al. 2014)) should be implemented as soon as possible to avoid more severe or irreversible consequences of global warming. Consequently, all economic sectors should become carbon-neutral (here understood as GHG-neutrality), as soon as possible and no later than the middle of the 21st century (Davis, Lewis et al. 2018). Achieving

this represents approximately an annual reduction of 3% of emissions between 2020 and 2050 (Haszeldine, Flude et al. 2018). However, various technical and non-technical challenges must be met to achieve carbon-neutrality.

1.1.2 Climate change mitigation

Non-technical solutions to mitigate climate change exist: the whole population could cease all non-essential activities that emit GHG (e.g. not travel long distances). However, substantial socioeconomic limitations make this solution not likely to occur: this approach would require the whole population to relinquish its lifestyle. Similarly, socioeconomic challenges apply to any technical solution. The costs of applying mitigation measures (or developing the technology needed) are much more immediate and unshared than the impacts of climate change. Hence we can frame climate change as a “tragedy of the commons” (Hardin 1968): the environment is a common good, individuals benefit from over-using it (by emitting GHG), the consequences of over-use are shared, and there is no way of increasing the capacity of the common good. Therefore, social, political, economic, environmental, and technological considerations must be taken to decide which mitigation measures are to be implemented and how they will be developed and applied.

For achieving carbon neutrality in a timely manner, mitigation measures must cover all sectors, but some measures could have an impact on the emissions from different sources. Geoengineering solutions (such as increasing Earth’s reflectivity) are not the preferred approach because of the complex and unknown impacts they could have (Kolstad, Urama et al. 2014). Another measure that could be applied to different sectors is carbon capture and storage (CCS), either coupled to processes producing CO_2 or coupled to direct air capture of CO_2. CCS can have a great beneficial effect in the long term if the technology matures and financial incentives are put in place, but it will not be sufficient to achieve carbon-neutrality by mid-century (Haszeldine, Flude et al. 2018). Therefore, it is still necessary to reduce emissions across all economic sectors.

In addition to the reduction of demand for GHG-intensive processes, it is evident that the refinement of processes to increase efficiency would reduce the environmental effect of multiple sectors. For example, land use-related activities can reduce their carbon-impact with existing technologies by improving agricultural efficiency and making the human diet more sustainable (by, for example, consuming

less red meat) (Willett, Rockström et al. 2019). Land use-related activities could reduce their carbon-impact with measures such as afforestation and reforestation, which can have a strong net-negative carbon balance (Bastin, Finegold et al. 2019).

Technical solutions exist for other sectors. The impact of transportation can be reduced by increasing use of public transport, walking, cycling and use of electric vehicles; heating could be electrical; and generation of electricity can become completely carbon-neutral using wind, solar, hydroelectric, and nuclear energy (Davis, Lewis et al. 2018). The challenges for this include increasing the infrastructure for increasing electricity requirement and generating it precisely matching the demand, which variates strongly along the day and year. Of the renewable energy sources mentioned, hydro and nuclear can provide a controllable supply, but they depend on costly infrastructure and availability of resources. Improvement of energy storage technology could allow better control in the supply of electricity using different sources (Davis, Lewis et al. 2018).

However, emission sources hard to eliminate or replace represent 27 % of global CO_2 emissions (Davis, Lewis et al. 2018). Industrial production of steel and cement do not have carbon-neutral alternatives. Increasing efficiency in the production and the use of these materials can reduce the impact, and CCS could be implemented along the production to make both processes carbon neutral. Aviation, shipping, and long-distance land transportation require high-density energy storage to avoid an increase in transport costs (Figure 1.3). Synthetic liquid hydrocarbons can be produced from hydrogen and atmospheric carbon; however, the energy required to capture CO_2 and obtain dihydrogen by electrolysis constitutes a significant barrier (Davis, Lewis et al. 2018).

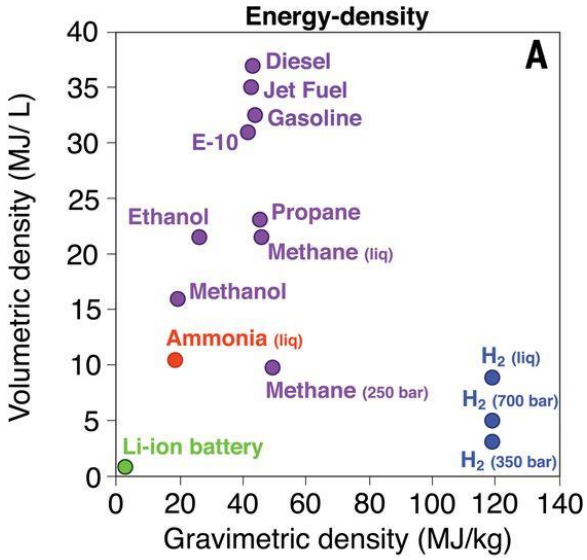


Figure 1.3. Comparisons of energy sources for transportation, including fossil and biofuel hydrocarbons (purple), hydrogen (blue), ammonia (red), electric lithium-ion batteries (green). From Davis, Lewis et al. (2018). Reprinted with permission from the American Association for the Advancement of Science.

An appealing alternative to fossil carbon is plant biomass, representing a potential source of energy, fuels and chemical feedstocks with a potential for neutral or negative GHG emissions. The uses of biomass as a replacement for fossil carbon and the related challenges are explained in the next section.

1.2 Plant biomass as an alternative to fossil carbon

Plant biomass is generated by CO_2 fixation during photosynthesis. Substituting biomass for production of energy and chemicals typically obtained from fossil carbon (often referred as bioenergy) is possible because plant biomass represents a reduced and energy-rich form of carbon (although less energy-rich than fossil fuels) that can be turned into energy (electricity and heat), fuels and other chemicals. It is also renewable, highly abundant, and widely accessible. The potential sources of plant biomass include edible crops (such as sugarcane and corn, currently used to produce some biofuels), non-edible crops with high productivity (or “energy crops”), forest biomass, and wastes or residues from agriculture, industry and consumers (Zabed, Sahu et al. 2017). It is estimated that 8 – $20*10^9$ tons/year of plant biomass are accessible for bioenergy (Zabed, Sahu et al. 2017).

However, it is essential to point out that GHG-neutrality is not guaranteed by replacing fossil-based fuels and chemical by the equivalents produced from biomass. Multiple factors must be considered in life-cycle analysis to ensure that

GHG emissions are effectively reduced (Dunn 2019). First, if the biomass feedstock is cultivated, production requires land, water and possibly fertiliser. Therefore, it is vital to evaluate the impact on GHG emissions due to land-use change, fertiliser production and use-related emissions (N_2O), and cultivation and harvesting. Additionally, the land use also might compete with food production and can have effects on natural systems (more specifically, on biodiversity, or soil nutrient loss). Second, biomass feedstock must be transported, processed, and the product transported again. All these steps require energy that, ideally, should be generated in a non-GHG-emitting way (Davis, Lewis et al. 2018). Third, direct emissions associated with use vary enormously. Fuel will release CO_2 in short timespans, while biochar will hold the carbon for thousands of years. If the aforementioned parts of the process have low or no emissions, and the CO_2 generated is associated with CCS, the whole process can reach negative emissions balances (Haszeldine, Flude et al. 2018).

1.2.1 Bioenergy pathways

Biomass can be directly combusted for the generation of energy (electricity and heat), or it can be transformed into fuels and chemicals through different pathways. Thermochemical processing of biomass with high heat generates some liquid and gaseous fuels, and other materials, but requires high energy inputs (Nanda, Mohammad et al. 2014). First-generation biofuels are generated from edible plant biomass, such as biodiesel produced by transesterification of oils or fats (Fukuda, Kondo et al. 2001) or biotechnological fermentation of sugars to ethanol, the principal current biofuel (Davis, Lewis et al. 2018).

Biotechnology represents a pathway to transform biomass into not only ethanol but also other fuels and chemicals. In biotechnological processes, chemical transformations are catalysed by living organisms and their enzymes. Consequently, they can take place at mild temperatures without the need for chemicals expensive to produce, dispose of or recycle. The foremost example is ethanol production from first-generation feedstocks (sugar or starch-rich crops), obtained by the efficient anaerobic fermentation of *Saccharomyces cerevisiae* (baker's yeast). However, edible feedstocks raise concerns regarding land use and food competition, adding ethical and economic constraints that make biofuels compete poorly against their fossil counterparts. Therefore, it is necessary to make efficient processes that use all the biomass carbon available, including the non-edible forms "locked" as

lignocellulose. The use of land-sourced lignocellulose for production of fuels is often referred to as second-generation biofuels, while third-generation biofuels refer to those processes based on algal biomass. Third-generation biofuels include algal biodiesels, but algae also generate cellulosic carbon that could potentially be used biotechnologically. The production of fourth-generation biotechnological fuels directly from CO_2 has also been suggested, but this requires a high energy input to produce a reduced material such as H_2.

Unfortunately, lignocellulosic biomass presents a central challenge to replacing fossil carbon biotechnologically: the difficulty of transforming it into reactive intermediates (Lynd, Liang et al. 2017). The nature of this challenge and how it has been addressed is described in the following sections.

1.2.2 Lignocellulose complexity

Lignocellulose is formed of different polymers with a complex structure that makes its monomeric sugars hard to access for biotechnology processes. Lignocellulose forms the structural material of plant cell walls and has evolved to be mechanically resistant and hard to digest to protect and support plants. The main three structural polymers are cellulose, hemicellulose and lignin (Figure 1.4), with proportions varying between plant species and part. Of energy crops and other relevant biomass sources, hemicellulose represents 16 - 33 % of the dry weight, lignin 9-32 %, and cellulose 37 – 52 % (Ragauskas, Beckham et al. 2014). Structurally, cellulose forms long fibres embedded in an amorphous matrix of lignin and hemicellulose (Lynd, Weimer et al. 2002).

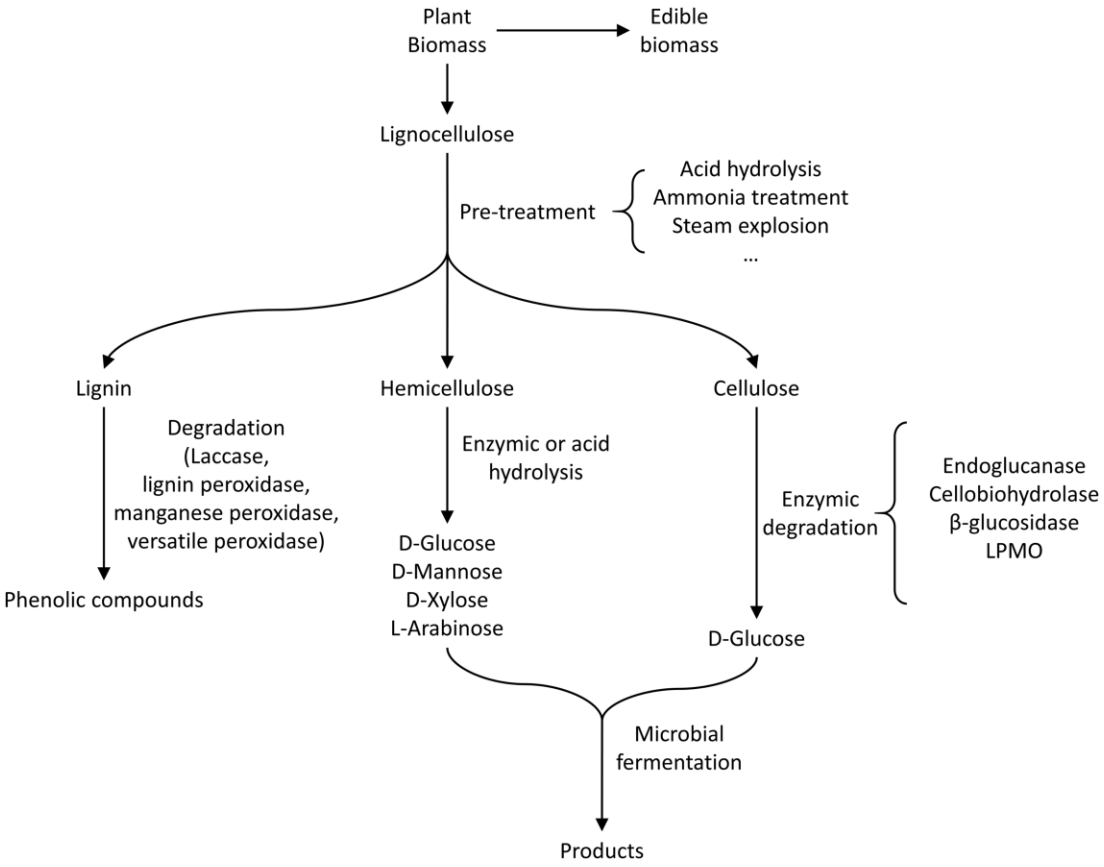


Figure 1.4. Major components of non-food plant biomass and biological routes to their utilisation. Reprinted from Valenzuela-Ortega and French (2019).

Hemicellulose is a branched heteropolysaccharide mainly formed of pentoses and hexoses, including D-xylose, D-glucose, L-arabinose, D-mannose and D-galactose, (among others). Hemicellulose is covalently bound to lignin (Terrett and Dupree 2019) and cellulose (Herburger, Franková et al. 2020), but is much more soluble than lignin and cellulose and can be readily digested by hydrolytic enzymes, such as xylanase (endo-xylanase or EX) or β-xylosidase (BX). However, its monosaccharides are less commonly metabolised by industrially relevant organisms than glucose. (Donev, Gandla et al. 2018). Lignin is a complex three-dimensional branched polymer formed by different phenylpropanoids and forms a matrix between fibres. Lignin is extremely difficult to degrade, but it is a potential source of phenolic compounds as chemical feedstocks (Ragauskas, Beckham et al. 2014).

Cellulose is the most abundant and valuable compound in lignocellulose. It is a linear homopolymer of D-glucose linked by β-(1,4)-glycosidic bonds. The repeating subunit in cellulose is the glucose dimer cellobiose because each glucose subunit is oriented 180° with respect to the adjacent subunits (Figure 1.5). Hydrogen bonds

within and between cellulose chains give crystallinity to the core of cellulose fibres, with some degree of amorphous (less organised) structure in the periphery. Cellulose digestion to its glucose monomer is complex due to this crystallinity; the digestion mechanisms known in nature are explained in the next section.

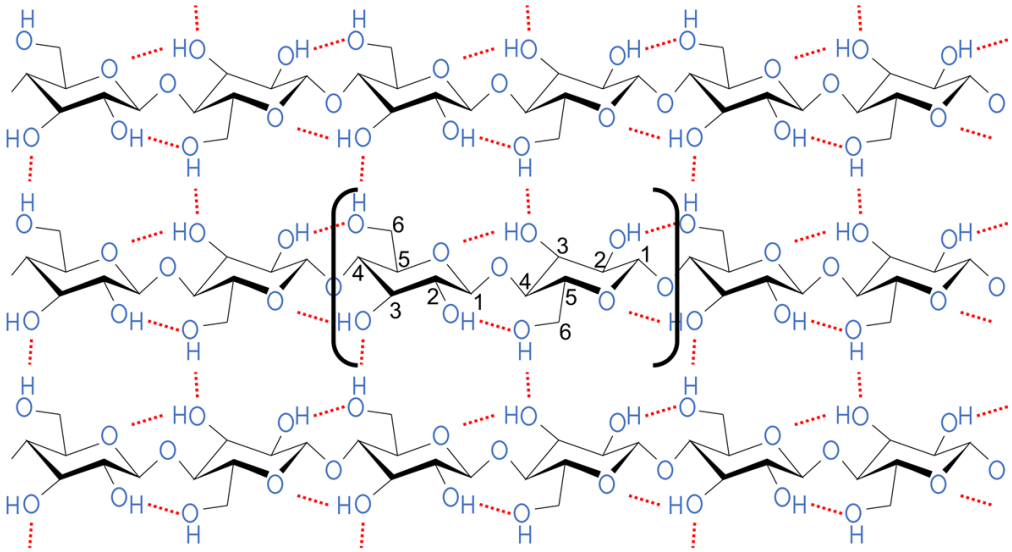


Figure 1.5. Structure of cellulose chains in a microfibril, showing intra-chain and inter-chain hydrogen bonds (red dotted lines). Reprinted from Valenzuela-Ortega and French (2019).

1.2.2.1 Cellulose digestion

Much of the knowledge of how cellulose is digested in nature comes from studying the enzymatic mechanism used by the fungus *Trichoderma reesei*. This filamentous mesophilic fungus was discovered during World War II and has been used for producing cellulases since then for the food and textile industries and production of laundry detergents (Bischof, Ramoni et al. 2016).

All hydrolytic cellulases display similar catalytic mechanisms using two carboxyl amino acids in the catalytic centre, but they differ in the cellulose substrate attacked, and that has been used to place them in three classic categories (Figure 1.6). First, endoglucanases (or EG, with Enzyme Commission number (EC) 3.2.1.4) cleave cellulose chains in random locations in the surface of fibres, thus generating a reducing end (on the C1 carbon of a glucose subunit) and a non-reducing end (C4 carbon). Second, cellobiohydrolases (or CBH, also named exoglucanases) attack the ends of cellulose chains in a processive manner, releasing cellobiose. Exoglucanases are end-specific, with enzymes type I (EC 3.2.1.176) active on

reducing ends and type II (EC 3.2.1.91) on non-reducing ends. And third, cellobiose is cleaved to glucose by β-glucosidases (BG, EC 3.2.1.21).

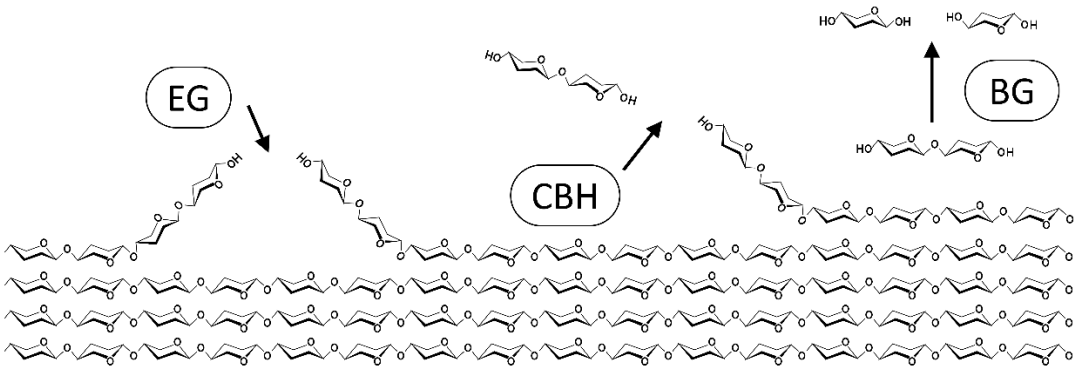


Figure 1.6. Enzymatic deconstruction of cellulose and activity of endoglucanases (EG), cellobiohydrolases (CBH) and β-glucosidases (BG). Reprinted from Valenzuela-Ortega and French (2019).

T. reesei has been useful for research and industrial production of cellulase because it secretes all the enzymes required to hydrolyse cellulose completely to glucose; however, the mechanisms to digest cellulose are more diverse in different organisms. Cellodextrin glucohydrolases (also named cellodextrinases, EC 3.2.1.74) can release glucose directly from cellooligosaccharides (Liu, Bevan et al. 2010). The bacterium *Cytophaga hutchinsonii* glides along cellulose fibres, degrading it with enzymes attached to the cell-wall or secreted to the periplasm. C*. hutchinsonii* does not possess any identifiable exoglucanase; instead, it uses EG and cellodextrin glucohydrolases such as Chu2268 (Zhu and McBride 2017). Other organisms are capable of importing and degrading cellooligosaccharides or cellobiose intracellularly, either by hydrolysis or by phosphorolysis (Parisutham, Chandran et al. 2017).

Cellulosomes are a multi-enzyme complex commonly found in cellulolytic species from the anerobic genus *Clostridium*. In the cellulosome of the model organism *Clostridium thermocellum* (Barth, Hendrix et al. 2018), cellulases have a cohesin domain that binds to one of the multiple dockerin domains of a scaffoldin protein, which has a cohesin domain that attaches the complex to the bacterial cell wall (Figure 1.7). Co-localisation of enzymes and proximity to the cell wall makes cellulosomes very efficient at biomass degradation and absorption of resulting soluble sugars.

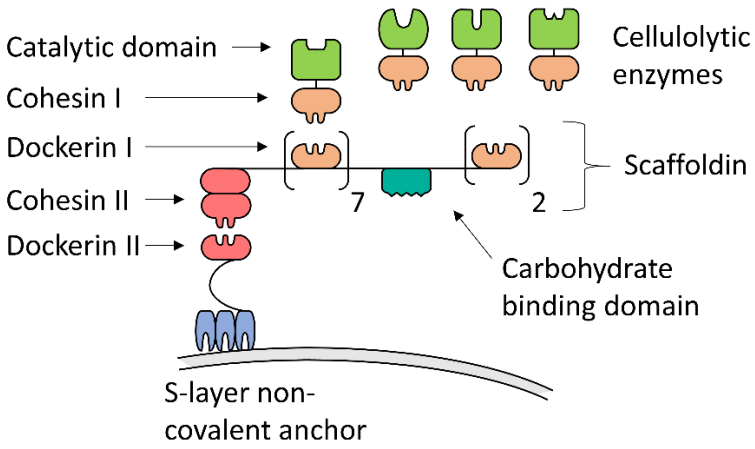


Figure 1.7. Natural cellulosome structure in *Clostridium thermocellum* and recombinantly expressed in *Bacillus subtilis* (Chang, Anandharaj et al. 2018). Reprinted from Valenzuela-Ortega and French (2019).

Non-hydrolytic proteins are now known to have very relevant roles in cellulose digestion in nature. Lytic polysaccharide monooxygenases (LPMO) are a type of enzyme widely distributed in fungi and bacteria that uses a different mechanism to cleave cellulose and other polysaccharides (Bissaro, Várnai et al. 2018). LPMO enzymes are copper-containing oxidoreductases that cleave cellulase by oxidation of either the C1 or C4 carbon of a β-1,4-glycosidic bond, which then spontaneously breaks (Figure 1.8). In addition to O_2, the LPMO catalytic cycle requires a reductant, which could include lignin degradation products (Brenelli, Squina et al. 2018) or electrons received from reductant oxidases such as cellobiose dehydrogenase (Loose, Forsberg et al. 2016). Alternatively, a recent study has also shown that these enzymes might replace O_2 and the additional reductant with H_2O_2, just requiring reductant to reduce Cu^{2+} to Cu^{1+} (Müller, Chylenski et al. 2018, Kuusk, Kont et al. 2019). LPMO acts in synergy with other cellulases increasing the overall hydrolysis (Müller, Chylenski et al. 2018); however, the need for oxygen might limit their activity to aerobic conditions.

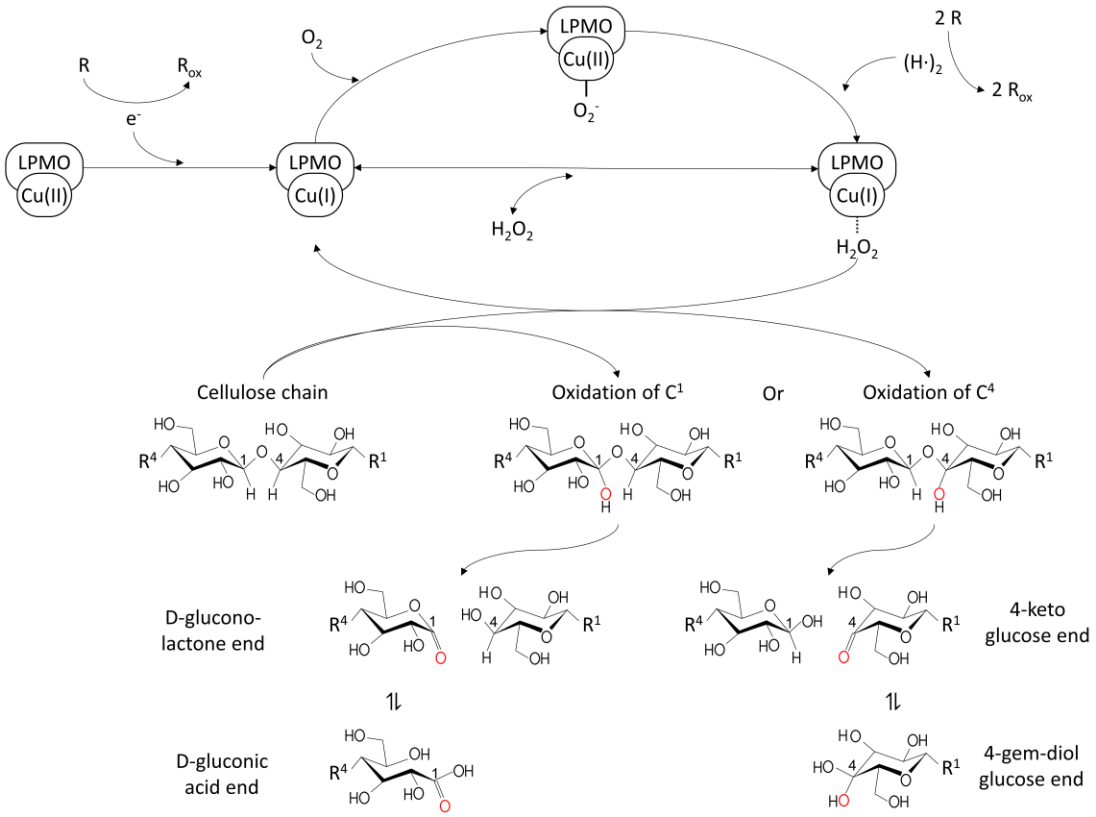


Figure 1.8. Proposed reaction cycle and products of Lytic Polysaccharide Monooxygenases. From Valenzuela-Ortega and French (2019), based on information in Bissaro, Várnai et al. (2018), Wang, Walton et al. (2019).

Non-catalytic proteins also play a role in cellulose digestion. Plant expansins and their homologues in fungi (designated swollenins) and bacteria are related to the cellulase family GH45 in CAZy (The carbohydrate-active enzymes database, Lombard, Golaconda Ramulu et al. (2014)), which includes various endo-hydrolases; however, expansins lack enzymatic activity (Cosgrove 2017). In plants, expansins reduce cellulose crystallinity to loosen the plant wall allowing cell growth; in cellulolytic organisms, a reduction in crystallinity facilitates attack by enzymes, shown with commercial cellulases (Cosgrove 2017) and other enzymes (Herburger, Franková et al. 2020).

The complex mechanisms that natural cellulolytic organisms use to digest cellulose illustrate why it is so challenging for biotechnological processes, but this understanding is the foundation to overcome the challenge.

1.2.3 Cellulosic bioprocesses

Commercial-scale processes generating products from fermenting cellulosic material have been trying to adapt the ethanol production process from glucose to use cellulose. Fermentation of glucose to ethanol is very efficiently done via organisms such as *Saccharomyces cerevisiae* or *Zymomonas mobilis*, which turn glucose to pyruvate generating energy in the form of ATP, and then turn pyruvate to ethanol and CO_2 using pyruvate decarboxylase and alcohol dehydrogenase to recycle NADH (Lynd, Liang et al. 2017). From a chemical point of view, part of the glucose carbon is oxidised while the other part is reduced obtaining energy.

First-generation ethanol is obtained from plant sucrose (or glucose from starch). Second-generation biofuels (at the commercial scale) aim to obtain glucose from cellulose and then follow the same process from glucose. For obtaining this glucose, biomass is pretreated thermochemically to increase solubility and surface of the polysaccharides, and to disrupt and separate lignin. After pretreatment, the cellulose fraction is digested with enzymes produced separately, mainly in fungi such as *T. reseei.* The resulting glucose is fermented, and the non-fermentable fraction is often used for heat and electricity generation through combustion. However, this process paradigm (called “Sequential Saccharification and Fermentation”) has found limited commercial success due to the high cost of the pretreatment and added enzymes, and due to the low returns caused by incomplete utilisation of all the biomass fractions, and efforts have been made to overcome these challenges by innovating within the existing paradigm and especially by moving to new process strategies (Lynd, Liang et al. 2017).

1.2.3.1 Innovations

New pretreatment approaches are required due to the currently used ones not reaching satisfactory results in a cost-effective and environmentally-friendly manner (Hassan, Williams et al. 2018). They require harsh chemicals or large amounts of energy for heating to high temperatures and chemical recovery. They are limited by the formation of growth inhibitors (for example, acid hydrolysis can lead to furfural and 5-hydroxymethyl furfural from pentoses and hexoses, respectively). Novel processes incorporate new solvents and technologies developed to an industrial scale in the food processing sector (Hassan, Williams et al. 2018). For example, the British company Bio-Sep Ltd is scaling up a separation process that uses an organic solvent mixture and ultrasonication to separate the fractions containing lignin,

hemicellulose and cellulose (Gates, McGarel-Groves et al. 2018). A recent metal-catalysed method showed separation of lignin by deconstructing it to phenolic oligomers, phenol, and propylene, which facilitates the valorisation of the lignin (Liao, Koelewijn et al. 2020).

In addition to improving pretreatment of the feedstock and valorisation of lignin, new process variants should also reduce the cost of the whole transformation procedure. One of the new approaches is “Simultaneous Saccharification and Fermentation”, which combines enzymatic hydrolysis with microbial fermentation. This strategy increases cellulase activity by continuously removing their products, which would otherwise inhibit the enzymes. A progression of this approach is “Simultaneous Saccharification and Co-Fermentation”, where non-glucose sugars from hemicellulose are also fermented. This approach included efforts giving *S. cerevisiae* the capacity to ferment these sugars (Ko, Um et al. 2016), or engineering other microorganisms, such as *Escherichia coli*, to more efficiently produce ethanol (Lewicka, Lyczakowski et al. 2014) or other products (Bokinsky, Peralta-Yahya et al. 2011).

1.2.3.2 Consolidated Bioprocessing

Despite the improvement in pretreatment of biomass and production of fuels from sugars, the cost of added cellulases remains a considerable barrier to making biorefineries viable (Lynd, Liang et al. 2017). This problem would be solved by using fermenting microorganisms that also produce the biomass-digesting enzymes, a strategy often called “Consolidated Bioprocessing” (CBP) (Lynd, Weimer et al. 2002, French 2009, Lynd, Liang et al. 2017).

To make CBP viable, the ideal biofuel-producing microorganism (IBPM) should have four characteristics (French 2009): it should secrete enzymes that digest cellulose efficiently, reducing pretreatment requirements; it must efficiently convert sugars to the desired fuels or chemicals; it must be able to withstand high concentrations of the product; it should do so and grow fast in industrial-scale conditions, to make processes efficient and economically viable.

However, known microorganisms in nature do not fulfil these criteria and generating the IBPM has remained the “Holy Grail” of industrial biotechnology and synthetic biology for more than 20 years.

1.3 Synthetic Biology

Synthetic biology can be defined as a multidisciplinary approach to science or engineering based on the generation of living systems (or systems based on biological components) with new-to-nature properties. When synthetic biology is applied to industrial biotechnology, it takes a top-down approach where an existing microorganism (or host) is manipulated, altering its biochemical behaviour to generate the desired product from the available substrate. This is achieved via multiple means: modification of genetic elements native to the host, introducing metabolic pathways or functions present in other organisms or entirely new to nature (Peretó 2020), and sometimes changes to the biochemical conditions beyond those encoded genetically (Sadler 2020).

Synthetic biology approaches the engineering of the biological host through the "design-build-test-learn" (DBTL) research cycle. The "Design" step comprises selecting a relevant microorganism or "chassis" and planning which set of genetic modifications are required to achieve the desired function and how to incorporate them. Previous knowledge (such as genomics and biochemical data) and modelling (systems biology) allow prediction of which (genetic) modifications are likely to result in the desired behaviour. However, because of the complexity of living systems, it is not possible to predict the effect of specific genetic modifications (Del Vecchio 2015, Gyorgy, Jiménez et al. 2015, Chao, Mishra et al. 2017). The complexity of interactions will increase with the complexity of the genetic construct introduced into the host, and consequently, multiple iterations of the research cycle might be necessary to achieve the desired effect in an optimal way (Naseri and Koffas 2020).

"Building" methods to generate new DNA molecules are used to introduce the genetic changes to the host. Despite the tremendous reduction in the cost of synthesising DNA molecules with custom sequences, researchers still depend on various methodologies to generate complex multi-gene constructs (Casini, Storch et al. 2015). The multi-gene DNA constructs are built and then introduced to the host. These constructs or devices are often referred to as pathways or circuits because often they contain metabolic pathways or genetic circuits (gene regulatory networks). In synthetic biology, approaching living systems from an engineering perspective, simple DNA elements (or parts) are characterised and rebuilt to generate new, more complex circuits (Cameron, Bashor et al. 2014). The simple parts include, for example, sequences coding for promoters, ribosome binding sites

(RBS), coding sequences, and transcription terminators. By incorporating modularity and standardisation in synthetic biology, the DNA assembly methods improve sharability of the parts, reliability of the assembly, and characterisation of the parts (Beal, Goñi-Moreno et al. 2020). Building (and designing) the candidates benefit from other advances in synthetic biology, such as new tools for computer-assisted design (Nielsen, Der et al. 2016, Du Lac, Duigou et al. 2020), and automation (Chao, Mishra et al. 2017); however, achieving the desired goal remains challenging due to the complexity of the manipulations required.

Generating the microbe with the desired functions can be framed as a problem with multiple variables or dimensions to be optimised qualitatively and quantitatively (Jeschek, Gerngross et al. 2017). Taking as an example the construction of a new metabolic pathway with multiple genes coding for enzymes, qualitative (or identity) variables include the enzymatic activities and the homologous genes chosen to be tested, and quantitative variables include the factors controlling the stages in the expression of those genes: gene dosage, transcription, and translation. The different variables must be tuned to achieve maximum product formation while reducing side-products and metabolic burden, which will reduce the host growth and the product yield (Jeschek, Gerngross et al. 2017). A sequential optimisation approach (of one gene at a time) might only reach a local optimum, and it requires an in-depth knowledge of the interactions in the system and ways to measure them. The sequential approach also requires more iterations of the cycle, making it costly and time-consuming. Recently, different strategies have been developed to optimise multiple variables of a system at the same time, such as combinatorial optimisation (Figure 1.9). This strategy is based on building and testing multiple combinations of the DNA elements controlling each variable.

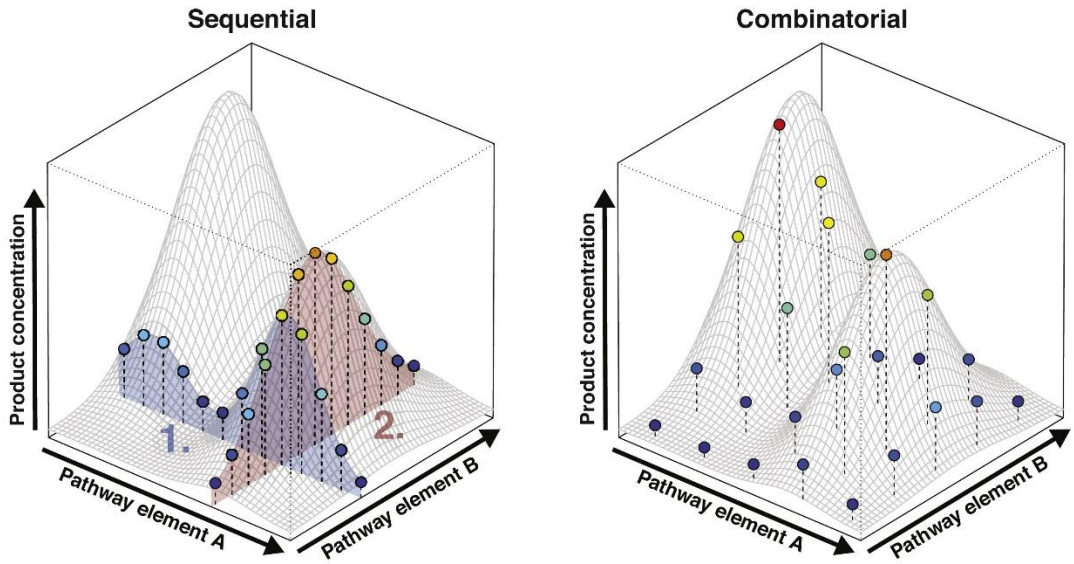


Figure 1.9. Sequential and combinatorial (non-sequential) optimisation as maximisation of product synthesised by a metabolic pathway. Sequential optimisation (left) involves finding the local optimum varying one factor at a time, and using the first local optimum as a starting point to optimise a second variable. Non-sequential or combinatorial optimisation (right) involves a simultaneous variation of multiple factors. Only combinatorial optimisation might reach a global maximum (red dot). Reprinted from Current Opinion in Biotechnology, Vol 47, Jeschek, Gerngross et al. (2017), Combinatorial pathway optimization for streamlined metabolic engineering, Pages 142-151, Copyright (2017), with permission from Elsevier.

Qualitative variation is reached through variation of coding DNA sequences and is more complicated to perform (Jeschek, Gerngross et al. 2017). DNA mutagenesis generates large libraries that require high-throughput screening methods, and testing homologues for coding sequence is complicated because preparing the DNA parts is costly and very laborious. On the other side, quantitative variation of gene expression is easily reached using different promoters controlling transcription (Moore, Lai et al. 2016) and ribosome binding sites (RBS) controlling translation (Salis, Mirsky et al. 2009). DNA assembly from modular DNA elements (Naseri and Koffas 2020) can easily generate combinatorial libraries that contain all possible combinations of elements. The size of combinatorial libraries (the different individuals that can exist) increases exponentially with the number of pathway components being co-optimised, leading to a “combinatorial explosion” (Jeschek, Gerngross et al. 2017). Libraries can only be as diverse as the “test” method allows screening of the candidates generated, often making this the bottleneck of the research cycle.

To “test” if the candidates behave in the desired way (and to learn if the modification caused the expected result), a measurable or selectable phenotype that can be linked to the genetic modification (or genotype) used. A relevant value of the system is measured in a quantifiable way, generally through optical means (spectrometry or fluorometry) or methods associated with chromatography. The method is chosen depending on the physicochemical properties of the system and the changes made. Therefore, the “test” step is specific for the pathway being engineered. The whole system can be assessed through a global measurement, which can be the same output being optimised (or a surrogate). Combinatorial optimisation approaches require global measurements, but a more profound understanding of the system might only be achieved by characterising specific genes or metabolic intermediates.

In many cases, the screening of candidates remains the bottleneck in the research cycle. To increase the throughput capacity of the screening step, some researchers have used biosensors where the system transforms variables which are hard to measure into others easier to detect, such as a fluorescent protein gene induced by a small intermediary molecule in a metabolic pathway being engineered (Fang, Wang et al. 2016). Alternatively, to maximise the benefit of combinatorial optimisation when still restricted by limited screening capacity, several strategies have been used to reduce the combinatorial library size without reducing the chance of reaching the global optimum (Jeschek, Gerngross et al. 2017): pathway modularisation breaks the whole pathway into modules treated as black boxes, reducing the dimensions of the problem; combinatorial libraries can be designed with fewer parts that remain functionally diverse; or using a model of the system which is trained with the first iterations of the cycle to predict the best designs that should be tested in the following cycles. However, these strategies are based on knowledge of the system and the internal interactions, which might not be available.

While synthetic biology advances have led to some successes that reached commercial-scale (Voigt 2020), the IBPM remains elusive. Substantial progress has been made engineering different hosts to make fuels and other useful molecules through different metabolic pathways, such as fatty acid synthesis (Wang, Ledesma-Amaro et al. 2020), isoprenoid pathway (Walls and Rios-Solis 2020) or polyketide synthesis (Barajas, Blake-Hedges et al. 2017). However, engineering and optimising of the capacity of microorganisms to break down the substrate is especially

challenging. The next section reviews some of the progress made in engineering non-cellulolytic organisms to digest biomass, emphasising cellulose.

1.3.1 Synthetic biology for consolidated bioprocessing

Different microorganisms are better suited for synthesis of different types of molecules. Similarly, different hosts might be better suited for different production locations due to the variability of feedstock available locally and the pretreatment being used (Chaves, Presley et al. 2019). Consequently, it would be useful to be able to equip different hosts with cellulolytic capacity. Earlier reports demonstrated the secretion of single cellulases from relevant hosts, and most recent publications describe the engineering of microorganisms to secrete enzymes with diverse activity, matching the new knowledge of the diversity in lignocellulose degradation mechanisms in nature. Even though most studies have demonstrated the enzymatic activity of the secreted proteins (either towards cellulose or soluble analogue substrates), only a few have achieved simultaneous saccharification and use of sugars (for either growth or product formation). Additionally, the product must be generated to a minimum concentration to make the process viable economically, which is between 4 and 5 % (v/v) for ethanol (Zacchi and Axelsson 1989). In all those cases, further optimisation is still required due to poor efficiencies.

Much research has gone into making the ethanol production champion, *S. cerevisiae*, not only cellulolytic but also able to use pentoses from hemicellulose digestion (Endalur Gopinarayanan and Nair 2019). The limited capacity of *S. cerevisiae* to secrete proteins only allowed sufficient enzyme secretion to reduce the amounts of added cellulases to produce some ethanol from biomass (20.4 g / L (Mo, Chen et al. 2015) and 14 g / L (Lee, Sung et al. 2017)). To improve secretion, (Davison, den Haan et al. 2016) isolated a wild strain with improved secretion (among other desirable properties for CBP), and Tang, Song et al. (2017) genetically modified the expression of factors involved in secretion in the endoplasmic reticulum and Golgi apparatus. A different strategy consisted of displaying the hydrolytic enzymes as a cellulosome attached to the cell wall, which led to 9.97 g / L ethanol being formed from cellulose and additional sugar (20 g / L galactose) (Fan, Zhang et al. 2016), or to 1.52 g / L ethanol from cellulose and rich medium (20 g / L of tryptone and 5 g / L yeast extract) (Tang, Wang et al. 2018). A recent study presented an industrial *S. cerevisiae* strain engineered to secrete seven hydrolytic enzymes and able to simultaneously ferment glucose, cellobiose

and xylose to ethanol; however, the authors did not report fermentation or growth from more complex substrates (Claes, Deparis et al. 2020).

Other yeasts are potential CBP hosts. *Pichia pastoris* (recently reassigned to the genus *Komagataella*) is commonly used for recombinant protein production and secretion. Expressing a mini-cellulosome with an EG and a xylanase, *P. pastoris* was able to produce more ethanol from the energy crop Miscanthus than the wildtype strain (1.08 g / L), yet in the presence of additional glucose (Shin, Hyeon et al. 2015). In another report, the expression of an EG, a CBH and a BG allowed growth from CMC and cellobiose, but not from cellulose (although sugar release was detected) (Kickenweiz, Glieder et al. 2018). Another relevant yeast is the oil-producer *Yarrowia lipolytica.* This yeast was found to express native BG genes and a consortium of strains secreting different cellulases (EG and CBH) displayed a small production of fatty acids (188 mg / L fatty acid methyl-esters) and growth from microcrystalline cellulose (Avicel) as carbon source(Wei, Wang et al. 2014).; growth was measured by weighing the solid fraction of the fermentation (cells plus cellulose) and subtracting the cellulose which was estimated by quantification of sugars after hydrolysis. It was found that 23 % of the Avicel had been consumed after 120 h. A single *Y. lipolytica* strain secreting 4 cellulases (2 EG and 1 CBH) and over-expressing a BG also displayed growth (measured the same way) from industrial pretreated cellulose pulp and Avicel (with 30.2 % Avicel consumption in the same time) (Guo, Duquesne et al. 2017). The additional expression of a xylanase and an LPMO in the same strain increased the growth from cellulose (Guo, Duquesne et al. 2017), reaching 36.5 % of consumption of the Avicel substrate (increased to a 42 % with a swollenin-pretreatment), yet a high initial cell load (1 g of dry cell weight per litre) was still necessary.

Among bacteria, *Escherichia coli* is the best-known host and has the most engineering tools available. *E. coli* is able to metabolise glucose and pentoses from hemicellulose, but not cellobiose. However, it naturally transports cello-dextrins (up to cellohexaose), and it only requires the expression of one BG or cellodextrin glucohydrolase (such as Chu2268 from *Cytophaga hutchinsonii*) to be able to grow using them as only carbon source (Liu 2012). By also expressing a secreted EG, it can digest non-pretreated pure cellulose filter paper enough to grow from it as main carbon source (Salinas 2017). In that experiment, the minimal medium M9 was supplemented with 0.2 g / L of yeast extract, which provided a very small carbon

source, but the strain expressing Chu2268 and the endoglucanase CenA was able to grow significantly more than the control strain without cellulases (Salinas 2017). Similarly, expression of 1 EG and 1 BG allowed generation of some useful products from cellulose as sole carbon source (28 mg / L n-butanol, 71 mg / L biodiesel, 35 mg / L D-limonene) (Bokinsky, Peralta-Yahya et al. 2011, Frederix, Mingardon et al. 2016). However, in those cases, both product titres and growth were insufficient for commercial scaling. One challenge that *E. coli* presents is limited ability for secretion of proteins, and some research has focused on identifying signal peptides for improved secretion (Gao, Luan et al. 2015). A strain engineered to produce high ethanol titers was engineered with a secreted multifunctional cellulase was able to produce 8 g / L from pretreated plant biomass with an initial cellular concentration of 5 Optical Density (or OD) units(Loaces, Schein et al. 2017). Recently, an *E. coli* strain was engineered to secrete and display a mini-cellulosome scaffoldin, yet the binding cellulases were generated separately (Vita, Borne et al. 2020). *Citrobacter freundii* is a close relative of *E. coli* that could benefit from tools developed for *E. coli* and can naturally use cellobiose; it has shown growth (albeit slowly and requiring 1 g / L of yeast extract) from Avicel and filter paper when expressing an endoglucanase and an exoglucanase from *Cellulomonas fimi* (*cenA* and *cex* genes, respectively) (Lakhundi, Duedu et al. 2017).

Other bacteria have been investigated as potential CBP hosts. *Pseudomonas putida* has various characteristics interesting for industrial production, such as resistance to various stresses (including pretreatment inhibitors and potential products), ability to metabolise different aromatics and organic acids, and availability of some genetic tools. Despite not naturally being able to use other sugars than glucose, Dvořák and de Lorenzo (2018) have engineered this host to use xylose and cellobiose. Additional work with this host showed auto-transporter-based display of active cellulase (Tozakidis, Brossette et al. 2016), but this was done as a way for enzyme production rather than for CBP application, and the authors did not show any use of sugars simultaneous with cellulose digestion.

Gram-positive bacteria are good candidates for protein secretion because of lacking an outer membrane and have been explored as IBPM. For example, a strain of the most studied gram-positive bacterium, *Bacillus subtilis,* was engineered to display a cellulosome with six different enzymes, which showed microbial hydrolysis of Napier grass, a potential energy crop (Chang, Anandharaj et al. 2018), yet using a high

initial cell concentration (with an OD of 10 units) and rich medium (lysogeny broth, LB) to sustain growth. The less-characterised thermophilic *Geobacillus* genus offers attractive candidates. *G. thermodenitrificans* engineered to secrete cellulase (EG or CBH) showed hydrolysis of the soluble cellulose analogue carboxymethyl-cellulose (CMC) (Daas, Nijsse et al. 2018).

Lactic acid bacteria and *Corynebacterium glutamicum* are widely used for industrial production of lactic acid and amino acids, respectively, and therefore are attractive CBP candidates. *Lactobacillus plantarum* could secrete two recombinant endoglucanases and two xylanases as a cellulosome (although as consortium of strains secreting a single enzyme each) that could be used to release soluble sugars from wheat straw, but not enough to grow (Stern, Moraïs et al. 2018); *Corynebacterium glutamicum* expressing recombinant EG and a recombinant BG could grow from cellobiose or CMC, but not from cellulose (Anusree, Wendisch et al. 2016).

The research towards finding the IBPM is progressing, but this is much slower than the progress made in other areas of synthetic biology. This could be caused by two challenges that emanate from the complexity of lignocellulose. Firstly, cellulose must be digested extracellularly by multiple secreted enzymes and, secondly, the crystalline nature of cellulose makes it insoluble which makes its suspensions turbid and non-homogeneous. Because of these two complications, most advances in synthetic biology can not be applied to engineer an IBPM.

The IBPM-candidate “building” is especially challenging because engineering the secretion of multiple enzymes presents more variables than the intracellular expression of genes. In addition to substrate hydrolysis, additional traits of the host that could require further genetic manipulation include transport of sugars, tolerance to inhibitors, and product synthesis. Different DNA assembly technologies for improved strain building, with a focus on those enabling work automation and part standardisation, are reviewed in Chapter 3.

Screening of IBPM candidates is particularly difficult by cellulose crystallinity. Turbidity blocks spectroscopic measurements and non-homogeneity impedes most analytical techniques. These issues are often circumvented using cellulose analogues that allow easier enzyme activity detection (driving experiments away from process-like conditions) or separation steps to isolate soluble analytes from the

insoluble substrate (strongly reducing the experimental throughput capacity). The different “screening” approaches and their limitations are reviewed in depth in Chapter 4.

Altogether, these challenges have limited most IBPM-development research to sequential optimisation work, where expression enzymes of different activity are tested individually (and most frequently with substrate analogues), and then very few combinations of expressed enzymes are tested. Without a combinatorial optimisation, only local activity maxima are found, which is insufficient for reaching substrate hydrolysis with enough efficiency for an industrial process. While the successes of such studies have been reviewed prior in this section, their experimental approaches have been revised in more detail in Chapter 5.

1.4 Aims and scope of this work

The main aim of the work presented here is to facilitate the creation of an ideal bioprocessing microorganism by assessing the main challenges found in the research cycle (in candidate “building”, and in candidate “testing”). This global aim is to be achieved through the following specific objectives:

First, developing optimised strain-generation techniques that allow flexibility to test all the relevant variables in different hosts and also allow incorporation of heuristic approaches to reduce the complexity of strain construction. Building upon existing DNA assembly methodologies, a new system was designed to meet this target not only for the development of microorganisms for consolidated bioprocessing but for synthetic biology in general. The design, construction and demonstration of this system is explained in Chapter 3.

Second, developing a screening method capable of increased throughput. In Chapter 4, different approaches are explored to find one that permits screening or evaluating the IBPM candidates’ full capacity to digest and use cellulose.

Third and finally, combining the new approaches for developing a model strain to demonstrate a faster research cycle. Chapter 5 describes the building of a model strain to allow screening of cellulolysis genes. A design of experiment will be presented to allow the use of the benchmark strain to screen cellulase genes using a new collection of standard parts encoding cellulolytic proteins and secretion signals.

The methodologies have been developed with a broad usage in mind, but the bacterium *Escherichia coli* has been used to generate them and as an experimental host model, which has been engineered towards the goal of degrading biomass using pure cellulose as model substrate.

Chapter 2 Materials and Methods

Some of the methodologies here described have been published elsewhere as protocols (Valenzuela-Ortega and French 2020). Additionally, a selection of protocols was made available in the online depository protocols.io (dx.doi.org/10.17504/protocols.io.brszm6f6).

2.1 Materials

2.1.1 Chemicals and reagents

Table 2.1. List of chemicals and reagents used.

Supplier	Reagent	Cat.no.
Acros Organics	Agar	443570010
Aldrich	Sodium Carboxymethyl cellulose (CMC) (DS 0.7)	419276
	Sodium Carboxymethyl cellulose (CMC) (DS 0.9)	222321
Alfa Aesar	4-chloro-DL-phenylalanine (Cl-Phe)	A13323
	Cellobiose	A14553
	Iodonitro-tetrazolium violet	B22301
	Spectinomycin dihydrochloride pentahydrate	J61820
Fisher Scientific	$CaCl_2$	C/1400/60
	Dimethyl sulfoxide (DMSO)	BP231-4
	Kanamycin Sulphate	BP906-5
	NaCl	S/3160/63
	β-Nicotinamide Adenine Dinucleotide (NAD+)	BPL746-212
	Sodium hydroxide (pellets)	C/4880/53
	Yeast Extract	10225203
Fluka Analytical	Tryptone	95039
Gibco	Tris buffer ph 7.5, 1 M	15567-027
	Tris buffer ph 8, 1 M	15568-025
Melford Laboratories	5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal)	MB1001

Table 2.1. List of chemicals and reagents used.

Supplier	Reagent	Cat.no.
	Isopropyl β-D-1-thiogalactopyranoside (IPTG)	MB1008
Scientific Laboratory Supplies	D-Glucose, anhydrous	CHE 1804
	Glycerol	CHE2068
	Magnesium sulfate heptahydrate ($MgSO_4$ $7H_2O$)	CHE2456
	Tween 20 (Polysorbate 20)	CHE3852
Sigma	Agarose	A9539-500g
	Calcium chloride dihydrate ($CaCl_2$)	C-7902
	Cellulase from *Trichoderma reesei* ATCC 26921	C8546-5KU
	Chloramphenicol	C0378-25g
	Phenazine methosulfate	P9625
	Sodium Chloride (NaCl)	S7653
Sigma-Aldrich	3,5-Dinitrosalicylic acid (DNS)	D0550
	Avicel cellulose, PH-101 (~ 50 µm)	11365
	Bovine Serum Albumin (BSA)	A7906-100G
	Congo red	C6767-25G
	Phosphoric Acid 85 %	438081-500ML
	Sodium carbonate, anhydrous	222321
	Thiamine HCl	T1270
	Yeast Alcohol Dehydrogenase	A7011-30KU
Thermo Fisher	Carbenicillin Disodium salt	10177-012
VWR	PEG 4000	26606.293

2.1.2 Bacterial strains

All experiments and manipulation of DNA were done, unless otherwise specified, using *Escherichia coli* JM109. All strains used are listed in Table 2.2.

Table 2.2. List of strains used.

Strain	Genotype	Source
E. coli JM109	*F´ traD36 proA+B+ lacIq Δ(lacZ)M15/ Δ(lac-proAB) glnV44 e14- gyrA96 recA1 relA1 endA1 thi hsdR17*	New England Biolabs (Cat no. E4107)
E. coli BW25141	F-*, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), Δ(phoB-phoR)580, λ-, galU95, ΔuidA3::pir+, recA1, endA9(del-ins)::FRT, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Coli Genetics Stock Center
E. coli BW25142	*F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), Δ(phoB-phoR)580, λ-, galU95, ΔuidA4::pir-116, recA1, endA9(del-ins)::FRT, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Coli Genetics Stock Center
Pseudomonas putida KT2440	Standard *P. putida* strain without TOL plasmid. rmo- mod+	ATCC® 47054
E. coli HB101 RK600	*E. coli* HB101 helper strain to mobilise the *oriT* plasmids	Aitor de las Heras
Cupriavidus metallidurans CH34	Type strain	DSM 2839, LMG 1195
Escherichia coli MG1655	K-12 *F– λ– ilvG– rfb-50 rph-1*	Lab Stock
Citrobacter freundii SBS197	Wildtype strain.	Biology Teaching Laboratory of the School of Biological Sciences (SBS), University of Edinburgh.
Citrobacter freundii NCIMB 11490	Type strain	ATCC 8090
NEB® 5-alpha Competent *E. coli*	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs (NEB) (Cat no. C2987I)
E. coli BW25113	*F- DE(araD-araB)567 lacZ4787(del)::rrnB-3 LAM- rph-1 DE(rhaD-rhaB)568 hsdR514*	Coli Genetics Stock Center

2.1.3 Restriction enzymes / molecular biology reagents

Most restriction enzymes and molecular biology reagents were obtained from New England Biolabs (NEB). Restriction enzyme AarI was obtained from Thermo Fisher Scientific. Additional reagents and materials are listed in the Table 2.3 below.

Table 2.3. List of additional molecular biology reagents used.

Supplier	Reagent	Cat.no.
Formedium	TAE 50x	TAE1000
NBS biologicals	EZ-10 Spin Column & Collection Tube	SD5005.SIZE.100
New England Biolabs	2-log ladder	N3200L
New England Biolabs	Hot Start Taq	M0495S
New England Biolabs	Q5® High-Fidelity DNA Polymerase	M0491S
New England Biolabs	T4 DNA ligase	M0202L or M0202S
New England Biolabs	T4 Polynucleotide Kinase	M0201S
Promega	dNTPs (100 mM set)	U1330
Promega	GoTaq® Master Mix	M7123
QIAGEN	Elution Buffer (EB)	19086
QIAGEN	QIAprep Spin Miniprep Kit	27104
QIAGEN	QIAquick Gel Extraction Kit	28706
QIAGEN	QIAquick PCR Purification Kit	28104
QIAGEN	Tris-EDTA buffer (TE buffer)	NA
Thermo Fisher Scientific	Qubit kit dsDNA HS	Q32851
Thermo Scientific	BigDye Terminator v3.1 Cycle Sequencing Kit	4337456
Thermo Scientific	Nuclease Free Water	J71786-K2

2.1.4 Plasmids

BioBrick parts used for the construction of new parts were obtained from the iGEM annual distribution kits of the Registry of Standard Biological Parts (https://igem.org/Registry, Cambridge, MA, USA).

EcoFlex (Moore, Lai et al. 2016) and CIDAR (Iverson, Haddock et al. 2016) toolkits were obtained from Addgene.

Plasmids built by Dr Alejandro Salinas and Dr Chao-Kuo Liu (Salinas (2017) and unpublished work) were used to amplify parts coding for some enzymes. Those genes had been previously amplified or synthesised as codon optimised version of the gene from *Cytophaga hutchinsonii* ATCC 33406 (for gene *chu2268*), *Cellulomonas fimi* ATCC 484 (for *cenA*), *Micromonospora lupini* str. Lupac 08 (for gene ML12A and its codon-optimised version), *Actinoplanes missouriensis* 431 (for

gene AM6B and its codon-optimised version AM6Bo), *Teredinibacter turnerae* T7901 (for TT3D), *Bacillus subtilis* 168 (for *yoaJ*).

SEVA vectors were kindly provided by Victor de Lorenzo and his research group (22-24).

pCP20 was obtained from internal lab stock, but it can be obtained from the Coli Genetic Stock Center.

All JUMP plasmids were built as explained in Chapter 3. Alex Orr built plasmid pJUMP18-dCas9_O, under my supervision, following the same methodologies.

2.1.5 Synthesis of oligonucleotides and DNA gBlocks

All DNA oligonucleotides used were ordered from Sigma-Aldrich and Integrated DNA Technologies.

When oligonucleotides were ordered to build parts, those were annealed and, if necessary, phosphorylated as explained in Section 2.2.2.3.

Double-stranded “gene fragments” were obtained from IDT and Twist Bio. For fragments smaller than the minimum size (300 bp), multiple parts were combined in one order to reach the minimum size.

2.1.6 Growth media

Lysogeny Broth (LB) was prepared by Roger Land Building technicians: 10 g / L tryptone (Bacto), 5 g / L yeast extract (Oxoid), and 5 g / L sodium chloride (Sigma-Aldrich) were dissolved in distilled water, the pH was adjusted to 7.2 with sodium hydroxide, and the broth autoclaved. LB agar was prepared the same way containing 10 g / L agar (Formedium).

SOC broth (super optimal broth with catabolite repression) was prepared by Roger Land Building technicians: 20 g / L tryptone (Bacto), 5 g / L yeast extract (Oxoid), 0.58 g / L sodium chloride (Sigma-Aldrich), 0.186 g / L potassium chloride (Fisher Scientific), 2.47 g / L magnesium sulphate ($MgSO_4$ $7H_2O$) (Fisher Scientific), 2.03 g / L magnesium chloride ($MgCl_2$ 6 H_2O) (Fisher Scientific), and 3.6 g / L glucose were dissolved in distilled water, and then autoclaved.

LB with carboxymethyl-cellulose (LB+CMC) was prepared by combining and then autoclaving: 2 g / L CMC, 10 g / L tryptone, 5 g / L yeast extract, 5 g / L sodium chloride, and 15 g / L agar.

M9 salts 4X solution was prepared by Roger Land Building technicians: 28 g / L disodium hydrogen orthophosphate (Fisher Scientific), 12 g / L potassium dihydrogen orthophosphate (Fisher Scientific), 2 g / L sodium chloride (Sigma-Aldrich), and 4 g / L ammonium chloride (Fisher Scientific) were dissolved in distilled water and autoclaved.

Minimal medium M9 was prepared by combining pre-sterilised ingredients listed in Table 2.4. All ingredients were sterilised by autoclaving (a/c) or filter-sterilising (f/s). If this medium was supplemented with thiamine, it was referred to as M9t; if it was supplemented with yeast extract, it was referred to as M9y. If prepared as agar gel, it was prepared by adding the same reagents to melted agar, previously autoclaved.

Table 2.4. Preparation of minimal medium M9.

Reagent	Amount per litre	Final Concentration
M9 salts (4X)	250 mL	1X
1 M $MgSO_4$ (f/s or a/c)	2 mL	2 mM
0.1 M $CaCl_2$ (f/s or a/c)	1 mL	0.2 mM
10 g / L Thiamine HCl (f/s) (for M9t)	1 mL	0.01 g / L
200 g / L Yeast extract (a/c) (for M9y)	1 mL	0.2 g / L
Carbon Source (100 g / L) (f/s or a/c)	40 mL	4 g / L
dH_2O (a/c)	Up to final volume	

Glycerol, glucose and citrate stock solutions were prepared by Roger Land Building technicians, as 20 % (v/v) (glycerol) or 20 % (w/v) solutions (glucose and tri-sodium citrate) in distilled water, which were then autoclaved. Cellobiose (10 %) was prepared as 10 % (w/v) in distilled water and then autoclaved.

Media were supplemented with antibiotics when cultured strains carried a plasmid, and with other additives as specified. Solutions are listed in Table 2.5.

Table 2.5. Antibiotics and other supplements commonly used with growth media.

Antibiotic	Stock conc. (mg/mL or specified)	Final conc. (µg/mL or specified)	Stock solvent
Chloramphenicol (Camp)	35	35	Ethanol
Carbenicillin disodium salt (Carb), f/s	100	100	dH2O
Kanamycin sulphate (Kan), f/s	50	50	dH2O
Spectinomycin dihydrochloride pentahydrate (Spec), f/s	50	50	dH2O
Isopropyl β-D-1-thiogalactopyranoside (IPTG), f/s	90	90 (~0.4 mM)	dH2O
5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal)	40	40	DMSO
Bovine serum albumin (BSA), f/s	40	As specified	1x M9 salts buffer
Cellulases from *T. reesei*, f/s	40	As specified	1x M9 salts buffer
Tween 20 (Polysorbate 20), f/s	40 % (v/v)	1.6 % (v/v)	dH2O

2.1.7 Buffers and Solutions

Transformation and storage solution (TSS) (Chung, Niemela et al. 1989) was prepared by combining 18 mL LB (a/c), 1 mL MgCl2 (a/c), 1mL DMSO, 4 mL PEG 4000 40% (w/v) (a/c).

Elution buffer (EB) was obtained from Qiagen with composition: 10 mM Tris-Cl, pH 8.5.

Tris-EDTA buffer (TE) was obtained from Qiagen with composition: 10 mM Tris-Cl, 1 mM EDTA, pH 8.

Tris-Acetate-EDTA buffer (TAE) was prepared from a 50X concentrate stock solution (Formedium). The working solution (1X) composition was: 40 mM Tris Base, 20 mM Glacial Acetic Acid and 1mM Na_2EDTA

2.1.8 Laboratory equipment used

DNA was quantified with a Nanodrop 2000 (ThermoFisher).

Agarose gels were imaged with a BioRad Universal Hood II (BioRad).

Visualisation of plates with fluorescent colonies was done using an Invitrogen Safe Image Blue-light Transilluminator (Invitrogen, S37102) with an amber filter unit.

Microplate cultures were done using 96-well “Multiwell plates with Lid” (CytoOne, CC7672-7596). For cultures, they were covered with Breathe-Easy Gas Permeable Sealing Membrane for Microtiter plates (DiversifiedBiotech, catalogue number BEM-1).

Centrifugation was done with a microcentrifuge Micro Star 17 (VWR) for volumes of up to 2 mL, and with a centrifuge 3-16KL (Sigma) for bigger volumes.

Incubation of cultures was done using Infors HT (shaking incubator), SciQuib Incu-Shake MAXI (shaking incubator), or Blood Tube Rotator SB1 (rotating mill used in a 37 °C room, Stuart Scientific).

2.2 Methods

2.2.1 Microbiological techniques

2.2.1.1 Cultivation

All manipulation of sterile media and cultures was done under aseptic technique to avoid contamination.

Unless otherwise specified, all *E. coli* and *Citrobacter freundii* cultures were grown in LB medium, incubated shaking at 37 °C at 200 rpm (liquid culture) or without agitation (agar cultures). The same conditions but a reduced temperature of 30 °C were used for *Pseudomonas putida*. Unless otherwise specified, liquid cultures of up to 10 mL were incubated in 50-mL centrifuge tubes, and bigger cultures were incubated in shake flasks.

If strains grown were carrying a plasmid, cultures were supplemented with the antibiotic corresponding to the plasmid selection marker.

2.2.1.2 Preparation of chemically competent cells

Preparation of chemically competent cells was based on the method developed by Chung, Niemela et al. (1989), and is described here for the routine preparation of 100 aliquots of competent cells. When different numbers of aliquots were prepared, volumes were adjusted proportionately.

A starter culture was prepared by inoculating 10 mL of LB with a single colony from a fresh plate streaked with the strain of interest. The starter culture was grown overnight with agitation at 30 °C. The following morning, 1 mL of starter culture was used to inoculate 100 mL of LB pre-warmed at 37 °C, which was incubated with agitation at 37 °C until it reached an OD between 0.4 and 0.6. The culture was incubated on ice for 30 to 60 minutes, then centrifuged (3900 g, 4 °C, 10 min). The supernatant (SN) was carefully decanted, and the pellet resuspended (gently, by inversion) in 10 mL of TSS pre-chilled on ice. Cells resuspended in TSS were then transferred in 100 µL aliquots to pre-frozen 1.5 mL centrifuge tubes and stored at -80 °C.

2.2.1.3 Transformation of chemically competent *E. coli* cells

Aliquots of competent cells were thawed on ice. If aliquots were split, it was done into pre-chilled 1.5 mL tubes. The DNA used for the transformation was added, and cells were incubated on ice for 30 min, then heat-shocked for 90 seconds in a heat block at 42 °C. After the heat-shock, cells were incubated on ice for 2-5 minutes. An outgrowth step was done by adding 900 µL of SOC (or 9 times the aliquot volume) and then incubating cells with agitation at 37 °C for 1 h. If a thermosensitive plasmid was used for the transformation, the outgrowth was done at 30 °C for 3 h. After the outgrowth step, 100 µL of the transformation reaction was spread on an LB plate with the appropriate antibiotic.

Transformation efficiency was calculated using 1 µL of pUC19 plasmid (50 pg / µL) obtained from NEB. With *E. coli* JM109 used for routine cloning, transformation efficiency ranged between 10^7 and 10^8 CFU (colony forming units) per µg of pUC19.

The volume plated and the amount of DNA used for the transformation depended on the DNA type and transformation efficiency. For routine assemblies and cloning reactions, 1 µL of the cloning reaction was used, and 100 µL were plated. For more difficult assemblies, up to 5 µL of reaction were used, and the whole transformation could be plated. When transforming with plasmid DNA minipreps, DNA or volume plated were reduced to 1 ng or 10 µL, respectively.

2.2.1.4 Tri-parental conjugation

Conjugation of strains was performed based on the protocol described by de Lorenzo and Timmis (1994), and consisted of three “parental” strains: a recipient

strain, a donor strain (containing the plasmid to be transferred), and a helper strain *E. coli* HB101 RK600, that contains the mobilisable and conjugative plasmid RK600.

One hundred microlitres of overnight cultures of each of the three strains were combined with 5 mL of fresh LB. The mixture was incubated at 30 °C for 5 h, and then streaked to agar plates with medium selective for the recipient strain with antibiotic selective for the plasmid to be conjugated. *Pseudomonas putida* was selected using M9y medium with 0.4 % (w/v) tri-sodium citrate as a carbon source.

2.2.1.5 Preparation of cryogenic glycerol stocks

Relevant strains were stored long term in cryogenic stocks. Fresh cultures were combined in a 1:1 ratio with sterile 20 % (v/v) glycerol, mixed and incubated at room temperature for 10 – 20 minutes, then moved to -80 °C.

2.2.2 Molecular biology techniques

2.2.2.1 DNA purification via miniprep

Plasmid DNA was extracted and purified from cells carrying the DNA using the QIAprep Spin Miniprep Kit (Qiagen), following the manufacturer's protocol but adjusting volume of culture and elution buffer depending on the copy number of the plasmid being purified. EZ-10 Spin Columns were occasionally used instead of Qiagen columns.

For vectors with a low-copy number (such as SEVA/JUMP OriV #2, #3, #5, #6, #7, #9), 10 mL overnight culture was used, and DNA was eluted twice in the same 50 µL nuclease-free water.

For medium-copy number vectors (such as SEVA/JUMP OriV #9), 10 mL culture was used, and DNA was eluted in 50 µL nuclease-free water twice (total 100 µL).

For high-copy number vectors (such as SEVA/JUMP OriV #4, #8) 5 mL culture was used, DNA was eluted in 100 µL nuclease-free water.

When using high volumes of culture (10 or more mL), volumes of resuspension, lysis and neutralisation solutions (P1, P2, N3 solutions respectively) were increased (proportionately to the culture volume relative to the normal volume of 5 mL). For example, if the culture volume was 10 mL, the volumes used of P1, P2 and N3 were doubled; with a culture of 20 mL, P1, P2, N3 were quadrupled. With increased

volumes, lysates were carried out separately and then combined in the same silica column in multiple loadings.

DNA was quantified with a NanoDrop 2000 spectrophotometer. If the concentration was too low for analysis or the intended use, DNA was concentrated via RotaVap as explained in Section 2.2.2.13. If DNA concentration was still insufficient, the miniprep was repeated using a culture with bigger volume or improved agitation (i.e. incubated in an Erlenmeyer flask instead of a 50-mL centrifuge tube).

2.2.2.2 Other DNA purifications (PCR-cleanup and gel extraction)

When linear DNA, such as PCR products or restriction digestions, had to be purified for cloning purposes or Sanger sequencing, it was done using the QIAquick PCR Purification Kit or QIAquick Gel Extraction Kit (QIagen), following the manufacturer's protocol.

PCR products were analysed via agarose gel electrophoresis as described in section 2.2.2.6. If only one band corresponding to the desired molecular weight was observed, the product was purified from the reaction using the PCR Purification Kit. If multiple bands were observed, the desired size band was excised and the DNA purified with the Gel Extraction Kit.

The DNA of restriction digestions was routinely purified with the Gel Extraction Kit and only purified from the reaction mixture with the PCR Purification Kit if no additional DNA fragment would cause interference in later cloning steps.

DNA purified via the PCR Purification Kit was quantified using a NanoDrop spectrophotometer, and DNA purified via the Gel Extraction Kit was quantified using a Qubit 4 Fluorometer and Qubit dsDNA HS Assay Kit (Thermo Fisher), following the manufacturer’s protocol.

If the concentration was too low for the intended use, DNA was concentrated via RotaVap as explained in Section 2.2.2.13.

2.2.2.3 Preparation of DNA fragments by annealing and phosphorylation of oligonucleotides

To prepare short DNA double-stranded parts, complementary oligonucleotides were phosphorylated (if necessary) and annealed. 5 µL of each oligonucleotide (100 µM) were combined with: 40 µL of nuclease-free water if no phosphorylation was

necessary; 34 µL of nuclease-free water, 5 µL T4 DNA ligase buffer, and 1 µl T4 DNA polynucleotide kinase (PNK) if phosphorylation was necessary. This mixture was incubated at 37 °C for 30 minutes (if phosphorylation was needed), and then 5 minutes at 95 °C, after which the temperature was dropped slowly (0.2 °C / s) to room temperature. Parts were diluted in EB buffer to a stock concentration of 0.5 µM in TE buffer.

Some parts were designed as a chain of sub-parts with long intervening sticky ends (16 bp) to avoid long oligonucleotides and their increased cost and avoid unwanted secondary structures. These sub-parts were combined in the stock solution without further ligation steps.

2.2.2.4 Preparative PCR

PCR for preparative purposes (i.e. cloning) was done using the high fidelity Q5 polymerase (NEB), following the manufacturer's protocol. Special attention was paid to use an amount of template DNA below 1 ng, as recommended by the manufacturer. The annealing temperature was adjusted to the sequence of the oligonucleotides using NEB's online tool (https://tmcalculator.neb.com/).

2.2.2.5 Colony PCR

Colonies were individually resuspended in 40 µL of sterile water. Two µL of the suspension were combined with 8 µL of reaction mastermix prepared as explained below, with NEB's Hot Start Taq (Table 2.6) or Promega's GoTaq Green Mastermix (Table 2.7).

Table 2.6. Reaction set up for colony PCR with Hot Start Taq (NEB).

Reagent	Initial Conc.	Vol. (µL)	Final Conc.
dNTPs	10 mM	0.2	0.2 mM
Primer F	10 µM	0.2	0.2 µM
Primer R	10 µM	0.2	0.2 µM
Taq Sandard Buffer	10X	1	1X
H_2O (nuclease free)	-	6.35	-
Polymerase	5 U/µL	0.05	0.025 U/µl
Template	-	2	-

Table 2.7. Reaction set up for colony PCR with GoTaq Green Master Mix (Promega)

Reagent	Initial Conc.	Vol. (µL)	Final Conc.
Gotaq Green 2X MasterMix	2X	5	1X
Primer F	10 µM	0.2	0.2 µM
Primer R	10 µM	0.2	0.2 µM
H_2O (nuclease free)	-	2.6	-
Template	-	2	-

Reactions were incubated with the thermocycler programme as shown in Table 2.8. The annealing temperature was adjusted to the oligonucleotide primer sequences using NEB's online tool (https://tmcalculator.neb.com/).

Table 2.8. Thermocycling programme used for colony PCRs

95 °C	5 min	
95 °C	20 s	x30 cycles
Adjusted °C	30 s	
68 °C	1 min /kb	
68 °C	5 min	
10 °C	Hold	

Colony PCR reactions were analysed by agarose gel DNA electrophoresis as described in Section 2.2.2.6.

2.2.2.6 Agarose gel DNA electrophoresis

Agarose gel DNA electrophoresis was used for analytical and preparative purposes. Unless specified otherwise, gels were prepared with a concentration of 0.8% w/v agarose in 1x TAE buffer, pre-stained with SybrSafe (1X) (Thermo Fisher) and run at 100 V for 20 minutes. The electrophoresis was carried out in tanks (BioRad PowerPac 300, Biorad Mini-Sub Cell GT or Biorad Wide Mini-Sub Cell GT) containing 1X TAE buffer.

The molecular weight marker used, unless specified differently, was NEB's "2-Log DNA Ladder".

2.2.2.7 Restriction digestions (analytical and preparative)

Digestions of DNA with restriction endonucleases were prepared as recommended by the enzyme manufacturers (NEB or Thermo). For analytical digestions, at least 200 ng of DNA were digested. For preparative digestions, at least 1000 ng were digested.

2.2.2.8 DNA sequencing

Sanger DNA sequencing was done by performing the amplification reaction "in-house" and sending the reaction product to an external analysis provider (Edinburgh Genomics). The amplification reactions were done with BigDye™ Terminator v3.1 Cycle Sequencing Kit as shown in Table 2.9 and incubated as shown in Table 2.10. Reactions were stored at 4 °C without light exposure until analysed.

Table 2.9. Reaction set up for Sanger sequencing.

Reagent	Volume
BigDye® Terminator v3.1 Ready Reaction Mix	0.5 µL
Primer 10 µM	0.32 µL
5X Sequencing Buffer	2 µL
DNA template	At least 200 ng
Nuclease free water	Up to 10 µL

Table 2.10. Incubation conditions for Sanger sequencing.

95 °C	30 s	x50 cycles
50 °C	20 s	
60 °C	4 m	
10 °C	Hold	

2.2.2.9 Mutagenesis by blunt-end ligation (MABEL)

Mutagenesis by blunt-end ligation was used for single-site site-directed mutagenesis of plasmids. Primers were designed to amplify around the mutation site and outwards, then PCR amplifications were done with Q5 polymerase as described in Section 2.2.2.4. A phosphorylation and ligation 20 µL reaction was prepared with the purified DNA as shown in Table 2.11, then incubated for 30 minutes at 37 °C and 2 h at 16 °C. Transformation of *E. coli* was done using 1 µL of the mixture as

explained in Section 2.2.1.3 and screening of clones was done as explained in Section 2.2.2.12.

Table 2.11. Reaction set up for MABEL.

Reagent	Volume
T4 ligase buffer (10X)	2 µL
T4 ligase (400 units / µL)	1 µL
T4 PNK	1 µL
DpnI	0.25 µL
25-200 ng of PCR of product	-
Nuclease free water	Up to 20 µL

2.2.2.10 Gibson DNA assembly

Gibson DNA assembly (Gibson, Young et al. 2009) was done for multiple-site site-directed mutagenesis, to combine fragments coming from different plasmids, or both at the same time. For mutagenesis, fragments were amplified from mutation to mutation, with primers containing the new sequence and generating products with an overlap of at least 20 bp.

Assembly reaction was done using NEB's "NEBuilder HiFi DNA Assembly" Cloning Kit, following the manufacturer recommendations. One to five µL of assembly reaction were used to transform the competent cells provided with the cloning kit (NEB® 5-alpha Competent *E. coli*) as explained in Section 2.2.1.3 and screening of clones was done as explained in Section 2.2.2.12.

2.2.2.11 Golden Gate assembly

Golden Gate assembly conditions were based on those described by Iverson, Haddock et al. (2016). Generally, in a 20 µL reaction volume (with T4 Ligase Buffer 1X), 20 fmol of each part and destination vector in the assembly were cyclically digested —with 0.5 µL BsaI-HF-v2 (NEB) at 37 °C or 0.5 µL BsmBI-v2 (NEB) at 42 °C— and ligated with 0.25 µL T4 DNA ligase (NEB, at 400.000 units / mL) at 16 °C. Generally, 30 cycles of 3 and 3 minutes were done, but this was extended to 60 cycles of 5 and 5 minutes to increase the efficiency of more difficult assemblies. Temperature cycling was followed by a final digestion step of 15 minutes and an enzyme-inactivation step of 5 minutes at 65 °C (BsaI) or 80 °C (BsmBI).

Assemblies into the upstream or downstream secondary sites of JUMP vectors were performed in two steps, without purification. In the first step, equimolar amounts of the inserts (40 µM) were assembled without destination vector as in an assembly as described above. Meanwhile, the destination vector (~200 fmol) was digested with 1 µL of AarI or BbsI-HF and 0.5 µL of Shrimp Alkaline Phosphatase (NEB), with a final inactivation step of 20 minutes at 65 °C. In the second step, equimolar amounts (20 fmol) of insert assembly and destination vector were ligated with freshly added T4 ligase enzyme and T4 ligase buffer (NEB) for one hour at 16 °C.

Transformation of *E. coli* was done using 1 µL of the assembly mixture as explained in Section 2.2.1.3 and screening of clones was done as explained in Section 2.2.2.12.

2.2.2.12 Screening of cloning candidates

Upon transformation of any cloning reaction and growth in LB agar plates with antibiotic selecting for the transformed plasmid, colonies were screened to confirm that they carried the correct constructed plasmid.

Colony PCR was done as first step (Section 2.2.2.5), with primers flanking the insert in the plasmid (if the resulting amplicon would display a different size than that of the parental plasmids), or one primer annealing inside the insert and the other primer outside. SEVA primers PS1 (AGGGCGGCGGATTTGTCC) and PS2 (GCGGCAACCGAGCGTTC) bind in all SEVA and JUMP vectors and were used routinely as flanking primers (with an annealing temperature of 60 °C). PCR products were analysed by electrophoresis as described in Section 2.2.2.6.

For the clones that displayed a correct PCR product, the colony resuspension was used to inoculate an overnight culture for plasmid DNA miniprep (see Section 2.2.2.1) and spot LB agar plates with antibiotics selecting for the plasmids used during the cloning. The spotting was done to confirm that the colonies had not resulted from co-transformation of the parental plasmids used during the cloning. If clones displayed a correct resistance phenotype the following day, the minipreps were carried out and used for analytical digestion with restriction enzymes (Section 2.2.2.7) and analysed by electrophoresis (Section 2.2.2.6). Clones giving a correct restriction pattern were finally confirmed by Sanger sequencing (Section 2.2.2.8).

2.2.2.13 RotaVap concentration of DNA

If the concentration of purified DNA was not sufficient for the intended uses, the DNA was concentrated by evaporation of the solvent (H_2O) using an Eppendorf™ Concentrator Plus and its "Vacuum Aqueous" function at 45 °C.

2.2.3 Other techniques

2.2.3.1 Preparation of phosphoric acid-swollen cellulose (PASC)

Cellulose was treated with phosphoric acid to reduce crystallinity and increase the surface accessible to enzymes, with a methodology based Bokinsky, Peralta-Yahya et al. (2011). Two grams of Avicel cellulose were wetted with 6 mL of dH_2O, then mixed with 100 mL of phosphoric acid (85 % w/w) and stirred overnight at 4 °C. The following morning, 200 mL of cold dH_2O were added and mixed, causing the cellulose to precipitate. The cellulose was washed by steps of centrifugation (3900 g, 20 minutes) and resuspension of the pellet in water (first three – four washes), 50 mM Tris pH 8 (two washes), and water again (at least two additional washes after Tris). The pH of the suspension was confirmed to be ~7 (measured with strips pH-Fix 0-14, Fischer Scientific, 109642751)

Phosphoric acid swollen cellulose (PASC) was autoclaved in a 1 % (w/v) suspension in dH_2O.

2.2.3.2 Ethanol quantification assay

Ethanol concentration from microbial cultures was quantified with a chemoenzymatic assay, based on the one developed by Lewicka, Lyczakowski et al. (2014), but adapted for use with the 96-well microplate and plate reader.

A reaction mastermix was prepared as shown in Table 2.12 with 50 mM tris pH 7.5 as buffer. Ten µL samples of culture supernatant were combined with 200 µL of reaction mastermix and mixed. After an incubation (15 min at room temperature without light exposure) absorbance at 500 nm was measured for all wells. Using buffer as blank and an ethanol standard (multiple concentrations between 0 and 400 mM of ethanol in buffer), the ethanol concentration of the samples was calculated.

Table 2.12. Reaction master mix set up for ethanol quantification assay.

Reagent	Mw	Stock		Master mix		
		Conc. mM	Conc. mg / mL	µL stock per mL master mix	Conc. mg/mL	Conc. mM
Yeast Alcohol Dehydrogenase	-	-	5	5	0.03	-
NAD^+	685.41	20	13.71	1	0.01	0.02
Phenazine methosulfate	306.34	20	6.13	1	0.01	0.02
Iodonitro-tetrazolium violet	505	50	25.25	20	0.51	1

2.2.3.3 Congo Red assay

Endoglucanase activity was determined by the formation of haloes in Congo red-stained LB agar plates containing carboxymethyl-cellulose, prepared as described in Section 2.1.6. Two microlitres of a new culture of the strains of interest were spotted on the agar surface. After overnight incubation at 37 °C, the plates were stained with 5 mL of 1 mg /mL Congo red (in dH_2O) for 20 min followed by a wash step with 10 mL 1 M NaCl for 30 min. Halo formation was visualised directly and recorded by conventional photography.

2.2.3.4 Absorbance and Fluorescence techniques

Absorbance measurements used to quantify culture optical density (OD) were taken with a cuvette spectrophotometer (Jenway 7300, Jenway) when preparing competent cells and cell concentration was being normalised to inoculate fresh media. When OD was measured over time, it was done with a plate reader (FLUOstar Omega Microplate Reader, BMG Labtech), unless otherwise specified.

All fluorescence measurements were taken with a plate reader (FLUOstar Omega Microplate Reader, BMG Labtech). Green fluorescence (for sfGFP or eGFP) was measured with excitation and emission wavelengths of 485 and 520 nm. Red fluorescence (for mCherry) was measured with excitation and emission wavelengths of 584 and 610 nm, respectively. When fluorescence was normalised to OD, OD was measured with the same plate reader. The “gain” setting of the fluorescence measurement was calibrated for each experiment using a fresh culture of the strain expected to have the highest fluorescence for that experiment.

If fluorescence was measured of cultures growing outside the plate reader, the microplate was incubated in the plate reader for 15 minutes at 37 °C to ensure that the temperature differences did not affect fluorescence.

2.2.4 Computational techniques

2.2.4.1 DNA sequence editing software

For planning any DNA cloning, restriction digestion, PCR and Sanger sequencing SnapGene software (GSL Biotech LLC) was used. Sanger sequencing results were analysed with SnapGene.

Design of primers for part domestication was done manually using Snapgene and the tools developed by Sarrion-Perdigones, Vazquez-Vilar et al. (2013) (https://gbcloning.upv.es/tools/domestication/) (modifying overhangs if they were palindromic or repeated).

2.2.4.2 Analytical software – R and Excel

Microsoft PowerPoint was used to draw cartoons and schematic diagrams.

Microsoft Excel and R (RStudio, PBC, Boston, MA, USA) were used to explore, analyse, and plot data. R packages used included: tidyverse, viridis, RColorBrewer, ggpubr, and pracma, available from The Comprehensive R Archive Network (https://cran.r-project.org/); and growthcurves, developed by Hall, Acar et al. (2014) and available from https://sourceforge.net/projects/growthrates/.

Growth rates were calculated using the function “fit_easylinear” of the growthrates package. This function calculates the maximum growth rate with a minimum of h adjacent time points (depending on the experiment, at least 6), and the given growth rate is calculated with all bordering timepoints that have a growth rate of at least 95% of the maximum.

The area under the curve (AUC) of growth curves was calculated by trapezoidal integration with the function “trapz” of the pracma package.

2.2.4.3 Bioinformatics tools

Signal peptides of proteins were predicted using Phobius (Käll, Krogh et al. 2007) (http://phobius.sbc.su.se/), SignalP 4.0 (Petersen, Brunak et al. 2011, Almagro Armenteros, Tsirigos et al. 2019) (http://www.cbs.dtu.dk/services/SignalP/), and InterPro (Blum, Chang et al. 2020) (https://www.ebi.ac.uk/interpro/).

Additional tools used included BLAST (for retrieval of origin of sequences), the EcoCyc database (to access curated information of *E. coli* genes), GenBank (for accessing sequenced genomes to domesticate parts).

Chapter 3 Strain building: JUMP

3.1 Foreword

This chapter comprises the work completed to optimise technologies for genetically engineering microorganisms to carry out consolidated bioprocesses (CBP). It was found that increasing the flexibility of modular DNA assembly systems would be beneficial for “building” Ideal Biofuel Producing Microorganisms (IBPM), and these needs were found to be shared with other synthetic biology applications. Consequently, it was decided to generate a flexible DNA assembly system with an accompanying set of vectors and standard DNA parts of general use for synthetic biologists.

The work was published as a research article entitled “Joint Universal Modular Plasmids (JUMP): A flexible vector platform for synthetic biology” in the peer-reviewed journal Synthetic Biology (Valenzuela-Ortega and French 2021). All experimental work was conceived and performed by me. Sections 3.2 through 3.6 encompass the published manuscript, with the addition of Figure 3.3. Appendix I incorporates the Supplementary Information provided with the manuscript. The DNA sequences of vector and standard parts are available with the Supplementary Information of the article and in the repository Addgene, where plasmids were made available for distribution (https://www.addgene.org/browse/article/28203402/).

3.2 Abstract

Generation of new DNA constructs is an essential process in modern life science and biotechnology. Modular cloning systems based on Golden Gate cloning, using Type IIS restriction endonucleases, allow assembly of complex multi-part constructs from re-usable basic DNA parts in a rapid, reliable, and automation-friendly way. Many such toolkits are available, with varying degrees of compatibility, most of which are aimed at specific host organisms. Here we present a vector design which allows simple vector modification by using modular cloning to assemble and add new functions in secondary sites flanking the main insertion site (used for conventional modular cloning). Assembly in all sites is compatible with the PhytoBricks standard, and vectors are compatible with the Standard European Vector Architecture (SEVA) as well as BioBricks. We demonstrate that this facilitates the construction of vectors with tailored functions and simplifies the workflow for generating libraries of constructs with common elements. We have made available a

collection of vectors with ten different microbial replication origins, varying in copy number and host range, and allowing chromosomal integration, as well as a selection of commonly used basic parts. This design expands the range of hosts which can be easily modified by modular cloning and acts as a toolkit which can be used to facilitate the generation of new toolkits with specific functions required for targeting further hosts.

3.3 Introduction

Synthetic biology is facilitated by DNA assembly systems which allow rapid generation of new constructs (Cameron, Bashor et al. 2014). Despite DNA synthesis becoming increasingly more affordable, the function of new DNA is hard to predict due to strong context dependency caused by interaction between recombinant genes and with host elements (Klumpp and Hwa 2014, Del Vecchio 2015, Gyorgy, Jiménez et al. 2015). Synthetic biology researchers have to re-iterate the design-build-test cycle several times until the new system is optimized or tuned and, consequently, DNA assembly standards that are robust, automatable and accept re-usable DNA parts are still necessary (Chao, Mishra et al. 2017).

While Gibson assembly (Gibson, Young et al. 2009) and other overlap-based methodologies (as reviewed by Casini, Storch et al. (2015)) are very efficient in assembling multiple DNA parts, standards based on Golden Gate cloning are the best fit for automation and part re-usability (Casini, Storch et al. 2015). In Golden Gate cloning (Engler, Gruetzner et al. 2009), DNA parts are flanked by sites recognised by type IIS restriction enzymes, which cut outside the recognised site, leaving user-specified overhangs (or fusion sites). Parts are ligated in an ordered manner, and restriction sites are removed during assembly, with assemblies of up to 24 parts (Potapov, Ong et al. 2018). Different hierarchical standards based on Golden Gate have been published (Sarrion-Perdigones, Vazquez-Vilar et al. 2013, Engler, Youles et al. 2014, Iverson, Haddock et al. 2016, Moore, Lai et al. 2016, Schindler, Milbredt et al. 2016, Andreou and Nakayama 2018, Chiasson, Giménez-Oya et al. 2019, Pollak, Cerda et al. 2019, Vasudevan, Gale et al. 2019, Pollak, Matute et al. 2020), often generically referred to as modular cloning or “MoClo” after one of the first systems (Engler, Youles et al. 2014). In these standards, multiple level 0 parts (“basic” parts such as promoter elements, ribosome binding sites, coding sequences, N- or C-terminal tags, transcription termination sequences, etc.) are initially assembled in a specific order specified by the DNA overhangs generated

by cleavage by a Type IIS restriction endonuclease and are inserted into level 1 acceptor vectors, to generate level 1 assemblies such as a transcription unit (TU) with promoter, ribosome binding site, open reading frame, and transcription termination sequence. Typically, multiple level 1 assemblies can then be joined to form level 2 assemblies, level 2 assemblies to form level 3 assemblies, and so on (Figure 3.1A). Prior to such use, DNA parts need to be "domesticated", a process that removes internal "forbidden" restriction sites recognised by the enzymes used in the assembly process and adds flanking restriction sites which, on cleavage, will generate the appropriate overhangs for assembly, normally also introducing the part into a plasmid for amplification and distribution. After a part has been domesticated, Golden Gate-based standards are PCR-independent, allow re-use of parts in different assemblies and allow rapid generation of large and complex constructs. This is a big advantage over overlap-based assembly systems, as PCR reactions can fail and are prone to introduce mutations and, consequently, impose the need for more extensive quality control of the constructs generated.

The different Golden Gate-based standards have a similar hierarchical design, but differ in the restriction enzymes used, fusion sites at different junctions, number of vectors needed, use of linkers, and elements present in the fixed parts of the vectors, such as replication origins, selection markers, and other elements required for particular hosts (Figure 3.1).

To allow compatibility of basic parts, multiple toolkits have adopted the PhytoBrick (Patron, Orzaez et al. 2015) common syntax that dictates that BsaI must flank level 0 parts with specific fusion sites (Appendix I, Figure 7.2). The PhytoBrick Standard was originally agreed among the plant synthetic biology community but is also used in standards compatible with bacteria (Andreou and Nakayama 2018, Pollak, Cerda et al. 2019, Pollak, Matute et al. 2020).

For specific applications, some Golden Gate-based toolkits include special features in their vectors. For example, plant toolkits include left and right border sequences flanking the cloning site to allow *Agrobacterium*-mediated plant transformation (Sarrion-Perdigones, Vazquez-Vilar et al. 2013, Engler, Youles et al. 2014, Andreou and Nakayama 2018, Chiasson, Giménez-Oya et al. 2019, Pollak, Cerda et al. 2019), and some kits include unique flanking nucleotide sequences to allow the combination of multiple assemblies via Gibson assembly (Pollak, Cerda et al. 2019).

Alternatively, some toolkits extend the possible uses by allowing the addition of features to vectors as additional parts, such as selection markers for target hosts or homology arms for chromosomal integration (Chiasson, Giménez-Oya et al. 2019, Pollak, Matute et al. 2020). This strategy, however, increases the complexity of the assembly and might require multiple assembly steps. A similar situation occurs when multiple (or libraries of) constructs are to be generated which include common parts as well as variable parts. Examples include when inducible promoters are to be tested in the presence of a common transcription factor (or vice versa), or a library of guide RNAs with constant Cas9 or dCas9, or a library comprising a metabolic pathway with some genes variable and others constant.

An interesting feature partially addressing this is found in the EcoFlex toolkit (Moore, Lai et al. 2016). A subset of vectors includes a "secondary module" to simplify pathway optimisation. Multiple components (level 1 assemblies) can be introduced in the secondary module (which uses a different Type IIS restriction enzyme to that used for assembly in the main site), essentially modifying the vector, followed by a further assembly in the main site. Moore, Lai et al. (2016) showed that secondary sites increase assembly efficiency by decreasing the number of parts required in each assembly reaction, which is highly desirable when generating libraries of assemblies. While the potential of secondary sites was demonstrated, it was limited by its design in EcoFlex plasmids, being restricted to special level 2 vectors that could only receive 2 or 3 level 1 assemblies in the main site, and these had to be pre-assembled into a level 2 vector before sub-cloning them into the secondary site. Similarly, the original MoClo standard (Engler, Youles et al. 2014) also allowed serial addition of level 2 assemblies into a single vector; this required a large library of linker parts and was restricted to level 2 acceptor vectors.

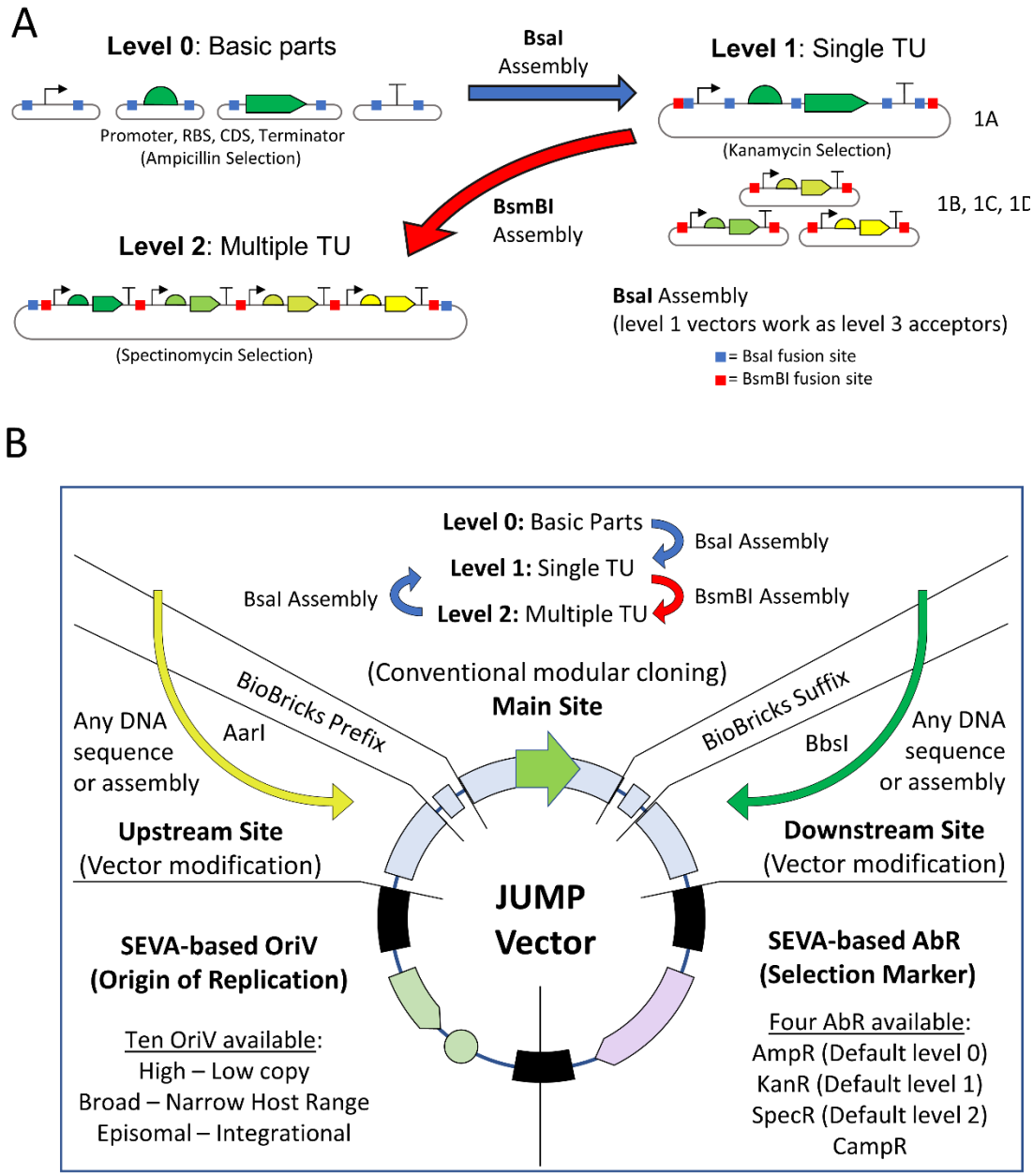


Figure 3.1. Modular cloning and general design of Joint Universal Modular Plasmids. A: Basic parts are contained in level 0 vectors, which are assembled to form a single Transcription Unit (TU) in a level 1 vector (either 1A, 1B, 1C or 1D). To simplify cloning and screening, the destination vector contains a different selection marker than insert-donor plasmids and a reporter gene replaced by the assembled TU. Using a second type IIS restriction enzyme and selection marker, multiple level 1 assemblies (1A, 1B, 1C and 1D) can be combined in a level 2 vector (2A, 2B, 2C, or 2D). In JUMP, level 1 plasmids can be used as level 3 assembly destination vectors. B: The design of JUMP vectors combines modular cloning in the main site (as shown in Figure 3.1A), compatibility with BioBricks and SEVA, and two orthogonal secondary modular cloning sites to introduce new features to any vector.

Here, we present "Joint Universal Modular Plasmids" (JUMP), a vector standard designed to improve the flexibility of current modular cloning systems (Figure 3.1B). JUMP vectors allow streamlined assembly and combine compatibility with PhytoBricks (Patron, Orzaez et al. 2015) and BioBricks (Shetty, Endy et al. 2008), with backbones based on the Standard European Vector Architecture (SEVA) (Silva-Rocha, Martínez-García et al. 2013, Martínez-García, Aparicio et al. 2015, Martínez-García, Goñi-Moreno et al. 2020). The design of SEVA allows the simple exchange of different elements of the vector (origins of replication (OriV), antibiotic selection markers (AbR) and cargo), and the SEVA repository comprises an ever-growing collection of standard vectors with different but compatible OriV and AbR. (Figure 3.2A). Moreover, all JUMP vectors include two orthogonal secondary sites, one upstream of the main insertion site and the other downstream, that can directly receive inserts from any modular level (Figure 3.2B). Thus, researchers can easily modify the vector chassis to reduce the steps needed for iterative assemblies and to give vectors new features (Figure 3.2C and D). While no toolkit can be suitable for all purposes, JUMP facilitates the modification of vectors to generate new toolkits which can be used for a wide variety of purposes.

3.4 Methodology

3.4.1 Strain and media

Escherichia coli JM109 (New England Biolabs Ltd (NEB)) was used for transformations with assemblies. Selection and proliferation of strains were done with LB medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) with the corresponding antibiotic depending on the plasmid (100 µg/L carbenicillin for level 0 vectors, 50 µg /L kanamycin for level 1 vectors, and spectinomycin at 50 µg/L for level 2 vectors). Vectors with chloramphenicol resistance were selected at 18 µg/mL. Homemade chemically competent cells were prepared based on the protocol of Chung, Niemela et al. (1989).

Conjugation of plasmids was performed as described by de Lorenzo and Timmis (1994): 0.1 mL of overnight LB cultures of the donor, recipient and helper strain (*E. coli* with plasmid RK600 coding *tra* genes, kindly provided by Dr. Aitor de las Heras), were mixed with 5 mL of fresh LB and incubated at 30 °C for 5 hours, then 0.1 mL were streaked in selective medium (M9 medium with citrate as carbon source for *Pseudomonas putida*).

3.4.2 Vector construction

SEVA vectors were kindly provided by Prof. Victor de Lorenzo and his research group (Silva-Rocha, Martínez-García et al. 2013, Martínez-García, Aparicio et al. 2015, Martínez-García, Goñi-Moreno et al. 2020). JUMP vectors were generated by modifying the replication origins (OriV) and antibiotic selection markers (AbR) of SEVA vectors, and creating a new "cargo" region as shown in Appendix I, Figure 7.1. "Forbidden" restriction sites (ie, those that might be used in assembly reactions; specifically, BsaI, BsmBI, BbsI, AarI, BtgZI) were removed from SEVA components by site directed mutagenesis as shown in Appendix I, Table 7.1 and 7.2. All vectors created are listed in Appendix I, Table 7.3 and are available through AddGene (https://www.addgene.org/browse/article/28203402/) and their sequences are available in the supplementary material.

3.4.3 Part domestication

Basic parts used here were domesticated by removing BsaI and BsmBI sites for compatibility with JUMP level 1 and 2 assemblies. We also routinely removed AarI, BbsI, BtgZ sites to allow assemblies using those enzymes for compatibility with other Golden Gate-based systems. We generated short parts (up to 120 bp) with annealed oligonucleotides and longer parts either by PCR or by DNA synthesis (Integrated DNA Technologies). To domesticate parts via PCR and remove internal forbidden sites we followed a protocol based on Sarrion-Perdigones, Vazquez-Vilar et al. (2013), successfully domesticating parts with up to 5 internal forbidden sites (Appendix I, Figure 7.3B). Parts are PCR-amplified between ends and internal forbidden sites as different sub-parts, which are assembled in the part-acceptor vector using BsmBI. The parts deposited in Addgene (including promoters, ribosome binding sites, open reading frames, transcription terminators, and others) are listed in Appendix I, Table 7.4, and their sequences are available in the supplementary material.

We domesticated parts in the universal part-acceptor vector pJUMP18-Uac, which generates basic parts with any fusion site with overlapping BsaI and BtgZI sites. DNA fragments flanked by overlapping BsaI and BsmBI sites can be used for both part domestication into pJUMP18-Uac and directly for level 1 assembly (Appendix I, Figure 7.3A). Promoters and terminators can be domesticated into the universal part acceptor or the specialised acceptors pJUMP19-Pac (promoter acceptor) and

pJUMP19-Tac (terminator acceptor), which allow part characterisation in *E. coli* as described in Results.

Basic (level 0) parts in the JUMP toolkit follow the PhytoBrick standard (18) as shown in Appendix I, Figure 7.2. Most parts in the toolkit were built from BioBrick parts (Shetty, Endy et al. 2008) or amplified from the genome of *E. coli* JM109. Terminators, as well as RBS PET, were adapted from the EcoFlex toolkit (Moore, Lai et al. 2016). The sequence of counter-selection coding sequence PheS A294G was obtained from Meyer, Segall-Shapiro et al. (2019), the Lambda Red coding sequences were obtained from pKD46 and cI^{ts} (temperature-sensitive cI repressor) was obtained from pCP20 (Cherepanov and Wackernagel 1995, Datsenko and Wanner 2000).

The correct sequence of all parts and modified vectors was confirmed by Sanger sequencing carried out by Edinburgh Genomics.

3.4.4 Assembly conditions

Assembly conditions in the main site were as described for CIDAR-MoClo (13) and are explained in detail in Supplementary Methodology. Briefly, 20 fmol of all parts and destination vector were cyclically digested —with BsaI (NEB) at 37 °C or BsmBI (NEB) at 42 °C— and ligated with T4 DNA ligase (NEB) at 16 °C (hereafter named BsaI assembly and BsmBI assembly, respectively). Double-stranded oligonucleotide linker “dummy parts” were used to replace parts and reduce the number of inserts in some assemblies.

Assemblies into the upstream or downstream secondary sites of JUMP vectors were performed in two steps, without purification. In the first step, equimolar amounts of the inserts are assembled without destination vector as in conventional modular cloning assemblies. Meanwhile, the destination vector is digested with AarI (Thermo Fisher Scientific Inc.) or BbsI-HF (NEB) and Shrimp Alkaline Phosphatase (NEB). In the second step, equimolar amounts of insert assembly and destination vector are ligated with freshly added T4 ligase enzyme and T4 ligase buffer (NEB) for one hour at 16 °C.

One microlitre of assembly reactions was used to transform *E. coli*, and transformed colonies appearing in the presence of the antibiotic corresponding to the destination plasmid were screened by colony PCR and restriction digestion of plasmid DNA

minipreps. Assembly into the main site (but not the upstream or downstream secondary sites) of JUMP vectors replaces an sfGFP cassette, with constitutive promoter J23100, which can be detected in colonies using a blue-light source such as ImagerTM Blue-Light Transilluminator (Invitrogen) with an amber filter unit. During the screening, we additionally checked for antibiotic resistance of transformed clones to ensure that only the destination vector had been transformed, due to the serendipitous finding that insert-carrying plasmids co-transform simultaneously with the destination plasmid even without the presence of their specific selecting antibiotic. We found that this also occurred using other toolkits, and co-transformed colonies, containing two different plasmids, were sometimes more than 10% of the total (data not shown).

Detailed protocols for JUMP plasmids have been published separately (Valenzuela-Ortega and French 2020). SnapGene software (GSL Biotech LLC) was used in the design processes of vector construction, part domestication and assembly planning. Open access design-automation tools, such as those made available by the Edinburgh Genome Foundry (https://cuba.genomefoundry.org/) can also be used to plan assemblies and quality-control checks.

3.4.5 Fluorescence measurements

Expression of sfGFP was measured to determine gene expression from different JUMP vectors. Absorbance and fluorescence were measured in a plate reader (FLUOstar Omega Microplate Reader, BMG Labtech). Overnight cultures were diluted 5-fold, and green fluorescence was measured in triplicate, with excitation and emission wavelengths of 485 and 520 nm, respectively, and was normalised by optical density measured at 600 nm. To characterise promoters and terminators, eGFP expression was measured as explained above, and relative fluorescence was normalised to that of the J23100 promoter (Kelly, Rubin et al. 2009). The characterisation of constructs with mCherry as reporter was done by fluorescence with excitation and emission wavelengths of 584 and 610 nm.

3.5 Results and discussion

3.5.1 Design and construction of JUMP vectors

The philosophy of JUMP was to generate a new Golden Gate-based assembly system which would build on the advantages offered by various different systems currently available while also allowing direct use in multiple species of bacteria, as

well as easy modification of vectors for new hosts and new applications, offering maximum flexibility and applicability. To this end, it was decided: that vectors would be based on the SEVA system, which offers wide flexibility and applicability to different bacterial species; that level 0 parts should be based on the PhytoBrick standard, already accepted by iGEM and used in multiple other Golden Gate-based kits; that BioBrick compatibility would be included, for backwards compatibility with constructs in the Registry of Standard Biological Parts; and that the main cloning site in all level 1 and level 2 vectors would be flanked by upstream and downstream secondary sites, which would also allow both level 1 and level 2 assembly directly into these sites. The design is summarised in Figure 3.1 and Figure 3.2.

Construction of JUMP vectors is explained in detail in Appendix I, Figure 7.1. Restriction sites relevant to cloning in JUMP and other PhytoBrick-compatible systems (BsaI, BsmBI, BbsI, AarI and BtgZI) were removed from SEVA components as shown in Appendix I, Tables 7.1 and 7.2. The JUMP “cargo” (Figure 3.2B), consisting of upstream cloning site flanked by AarI sites, BioBrick prefix, main modular cloning site consisting of marker gene sfGFP flanked by BsaI and BsmBI sites, BioBrick suffix, and downstream cloning site flanked by BbsI sites, was then introduced. Level 1 (with main modules 1A, 1B, 1C, 1D) and level 2 (with main modules 2A, 2B, 2C, 2D) vectors were built as a "core set" using SEVA's medium copy-number origin of replication #9 (pBRR322/ROP), and alternative vectors with nine different replication origins were built with the main module 1A and 2A as shown in Appendix I, Table 7.3. This architecture allows researchers to build intermediate assemblies with the core set and then perform the final assembly (either level 1 or level 2) with the OriV of choice. An example of a level 2 assembly appears in Appendix I, Figure 7.4. The default antibiotic markers are ampicillin/carbenicillin for level 0, kanamycin for level 1 and spectinomycin / streptomycin for level 2. Additional alternative vectors have been included in the toolkit with a different marker gene for screening (*lacZ’α* instead of sfGFP), and antibiotic marker (a broad-host-range chloramphenicol resistance gene adapted from the shuttle vector pSEVA3b61, which we named AbR #3 in JUMP vectors for the sake of simplicity) (Wright, Delmans et al. 2015). Additionally, we included two pairs of vectors that replace the vectors 1D (with 1D' and 1E) and 2D (with 2D' and 2E), which give the option of combining 5 inserts rather than the standard 4 in level 2 and level 3 assemblies. To replace the main site, the flanking BioBrick sites were used.

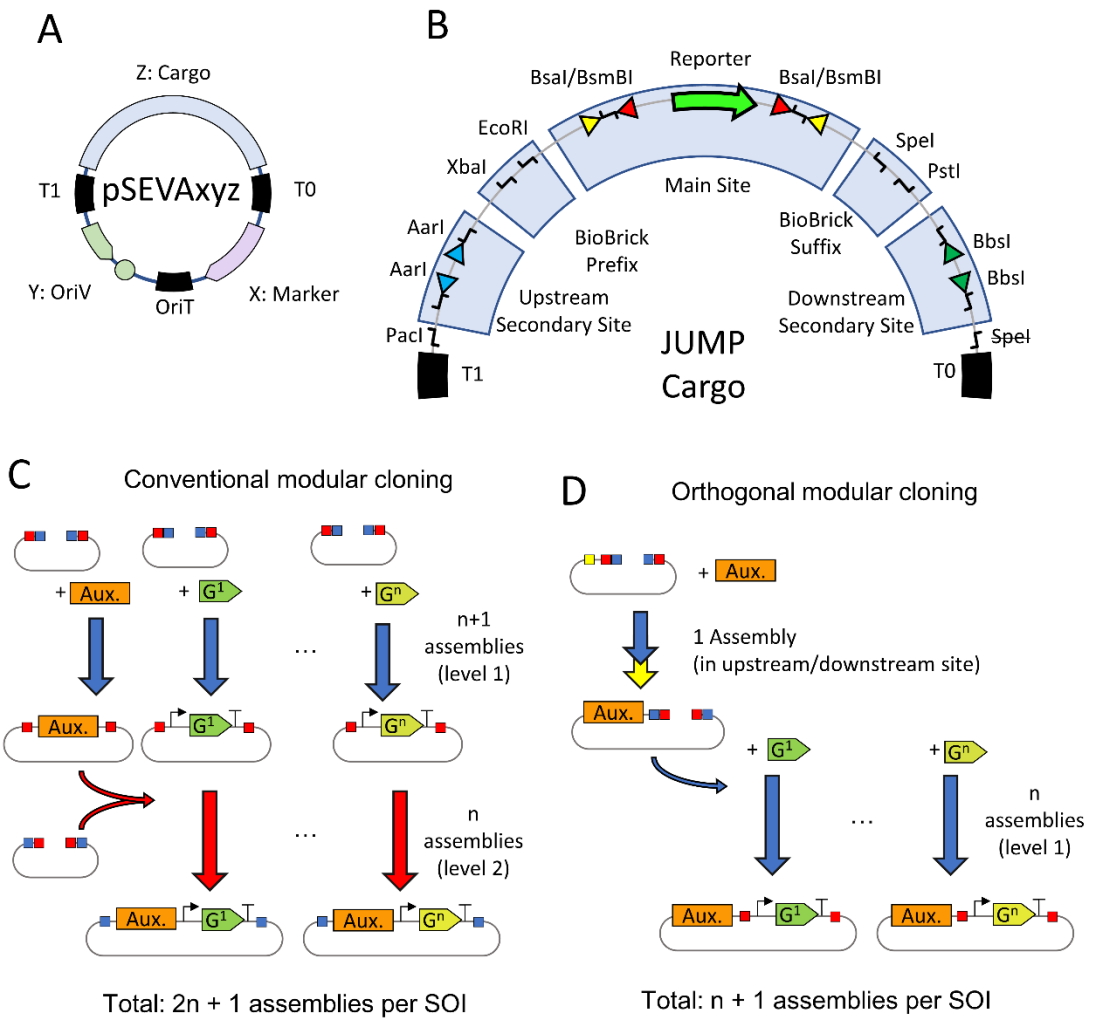


Figure 3.2. JUMP design and secondary sites. A: In SEVA (Standard European Vector Architecture) plasmids, three common short DNA sequences (black) flank three variable regions (coloured). Variable regions are the OriV (origin of replication), AbR (antibiotic selection marker) and "cargo" (any expression cassette). The invariable regions are two transcription terminators flanking the cargo (T1 and T0,) and origin of conjugation (oriT). Invariable regions also contain rare cutting sites, forbidden in the sequence of variable regions. B: JUMP is designed as special cargo of SEVA vectors to allow compatibility with future OriV's and AbR's of the collection. The cargo contains upstream modular site (with outward-facing AarI sites); BioBrick prefix (XbaI, EcoRI); main modular site (a screening reporter gene flanked by outwards-facing BsaI and inwards-facing BsmBI sites for level 1, and vice-versa for level 2); BioBrick suffix (SpeI, PstI); and downstream modular site (with outwards-facing BbsI sites). SEVA's canonical SpeI site was removed to allow BioBrick compatibility.C: Building constructs to test similar genes (G1 to Gn) as sequences of interest (SOI) that depend on common auxiliary factors (Aux.) with conventional modular cloning might require multiple assembly steps per SOI: the SOI is first assembled by itself and then combined with the auxiliary elements. D: Introduction of the auxiliary factors in vector chassis using orthogonal use of secondary sites reduces number of assembly steps to combine the SOI with the auxiliary factor. Squares indicate restriction sites for BsaI (blue), BsmBI (red), AarI (yellow) and BbsI (green).

The nomenclature of JUMP vectors (Figure 3.4A) is conservative with SEVA's (Silva-Rocha, Martínez-García et al. 2013, Martínez-García, Aparicio et al. 2015, Martínez-García, Goñi-Moreno et al. 2020) to facilitate the combination of new SEVA OriV and AbR with JUMP vectors. Of a vector "pJUMPxy-z", x and y indicate the selection marker and origin of replication as indicated by SEVA rules, with the difference of relevant restriction sites being removed in the JUMP version (see Figure 3.4B and C). The characteristics of the main module sites —level and position— are indicated by the index z (Figure 3.4D). For example, vector pJUMP29-1A(sfGFP) has AbR #2 (Kanamycin), OriV #9 (pBBR322/ROP) and main site module level 1 that would take position A in a level 2 assembly. Note that JUMP vectors are not SEVA vectors, and should be considered as SEVA “siblings”, as they do not strictly follow the SEVA rules, due to the SEVA native SpeI site being removed to make JUMP vectors compatible with BioBricks.

To allow compatibility with the widely used PhytoBrick level 0 parts, all acceptor fusion sites are constant (GGAG and CGCT) in all main and secondary sites. In the main sites, hierarchical conventional modular cloning works similarly as in current standards (depicted in Figure 3.1A). PhytoBrick level 0 parts are assembled in any JUMP level 1 vector (of type 1A, 1B, 1C or 1D) using BsaI (indicated by the removal of the marker gene); four level 1 products (assembled in vectors of type 1A, 1B, 1C, and 1D) can be combined in any level 2 vector using BsmBI. As with Mobius (Andreou and Nakayama 2018) and Loop assembly (Pollak, Cerda et al. 2019, Pollak, Matute et al. 2020), JUMP level 1 vectors can be used as level 3 assembly destinations for four level 2 products (assembled in vectors of type 2A, 2B, 2C, 2D) using BsaI. In addition to the compatibility with other toolkits, the use of constant acceptor sites offers the possibility to introduce sequences from any assembly level into either secondary site in any JUMP vector.

A selection of PhytoBrick-compatible level 0 parts was built to allow diverse synthetic biology applications (as shown in), including constitutive and inducible promoters, RBS, terminators, and sequences coding for various transcription factors and reporters. Appendix I, Table 7.4 includes additional information about these parts.

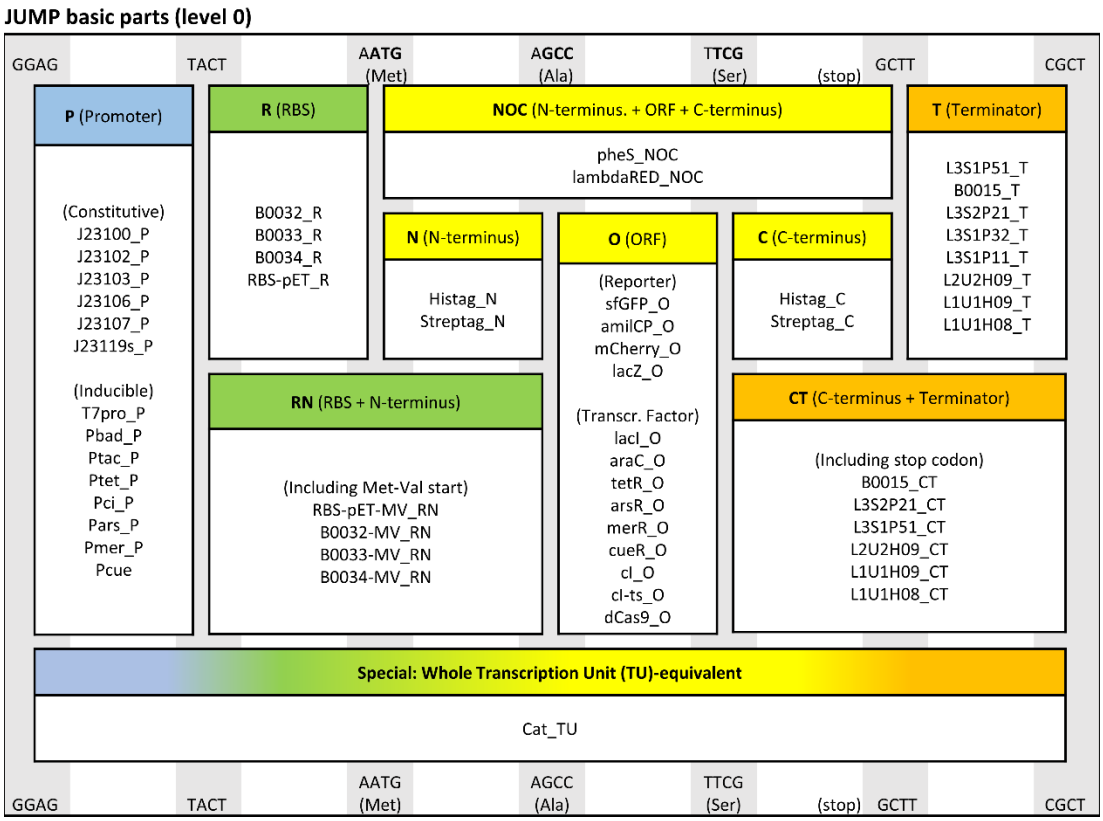


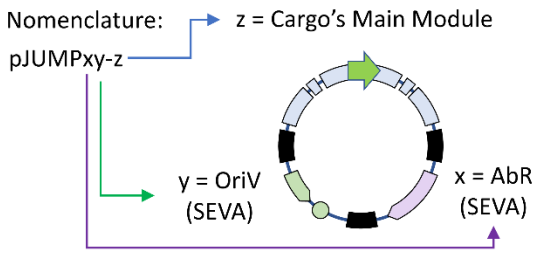
Figure 3.3. PhytoBrick standard parts included in JUMP toolkit in Addgene.

3.5.2 Use of different JUMP vectors

Most Golden Gate-based toolkits only offer vectors with one origin of replication. Two recent toolkits offer multiple microbial OriV; the uLoop kit (Pollak, Matute et al. 2020) offers four OriV with different copy number, and the MK toolkit (Chiasson, Giménez-Oya et al. 2019) offers four bacterial broad-host range OriV as well as two for yeast. A recent publication (Damalas, Batianis et al. 2020), offers four OriV in SEVA-derived plasmids named “SEVA 3.1” that allow Golden Gate cloning, but requires PCR-amplification of parts and vectors for each assembly, which is avoided by other Golden Gate-based assembly systems. The JUMP toolkit includes ten OriV, including broad-host range and variable copy-number OriV as well as conditional vectors to allow chromosomal integration. Four replicative OriV allow its use with other hosts beyond *E. coli*, the *Pseudomonas* shuttle OriV #4 (pRO1600 /ColE1) and three broad host range vectors: origins #2x (RK2), #3 (pBBR1) and 5# (RSF1010). RK2 and pBBR1 are known to be replicative in a wide number of gram-negative bacteria, and the hyper-promiscuous RSF1010 has been shown to replicate in some gram-positive bacteria and yeast species (Jain and Srivastava 2013). We confirmed that JUMP vectors can be trans-conjugated from *E. coli* to

Pseudomonas putida KT2440 (data not shown), using a tri-parental conjugation (de Lorenzo and Timmis 1994, Martínez-García, Aparicio et al. 2014). We successfully conjugated *P. putida* with level 1 vectors with OriV #2, #3 and #5. Unexpectedly, the *Pseudomonas* shuttle OriV #4 did not yield any conjugants in our tests, although it contains no changes in the sequence from the original SEVA OriV that has previously been introduced into *P. putida* (Wirth, Kozaeva et al. 2019)**.** Direct transformation of *P. putida* by electroporation was not tested.

A



B

AbR	Antibiotic Selection	Notes
1	Ampicillin /Carbenicillin	Default level 0
2	Kanamycin	Default level 1
3	Chloramphenicol	Broad host range marker, from pSEVA3b61
4	Spectinomycin /Strepomycin	Default level 2

C

OriV	Origin of replication	Notes
1	R6K	Narrow-host range, pir-dependent replication
2x	RK2	Low copy number, broad host range. X = native SEVA, contains BsmBI, AarI sites
3	pBBR1	Medium-low copy number, Broad host range
4	pRO1600 /ColE1	Shuttle E.coli/Pseudomonas *sp*
5	RSF1010	Broad host range
6	p15A	Medium-low copy number
7	pSC101	Low copy number
7ts	pSC101-ts	Thermosensitive replication, low copy number
8	pUC	High copy number
9	pBBR322-ROP	Medium-high copy number (Default level 1 and 2)

D

Main Module index		Donor fusion sites of product
Level 1	Level 2	
1A	2A	GGAG ... AATG
1B	2B	AATG ... AGCC
1C	2C	AGCC ... TTCG
1D	2D	TTCG ... CGCT

E

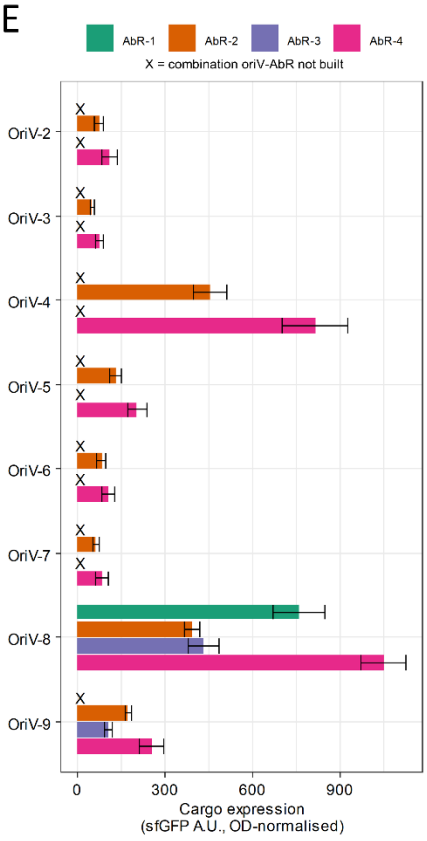


Figure 3.4. JUMP nomenclature and backbones. A: JUMP follows SEVA nomenclature for origins of replication (OriV) and selection marker (AbR), while cargo nomenclature is replaced by JUMP's main site module index. B: Antibiotic selection markers used in JUMP vectors. C: Origins of replication in JUMP vectors. D: Default main site modules and their donor fusion sites. E: OD-normalised

fluorescence of sfGFP cloning reporter in available combinations of OriV-AbR in the distributed JUMP toolkit. Error bars indicate standard deviation, n=3.

The availability of multiple vector backbones allows optimisation of the expression of the genetic construct. This is important as the characteristics of the vector strongly affect the expression of the gene it carries through context dependency (the effect of interactions of biological host complex environment and the function of the recombinant gene of interest) (Iverson, Haddock et al. 2016, Kim, Cavaleiro et al. 2016, Moore, Lai et al. 2016, Vasudevan, Gale et al. 2019). Kim, Cavaleiro et al. (2016) showed the relevance of "tuning" the vector chassis by testing different constructs with multiple combinations of OriV and AbR's. While the highest copy-number vectors were found to impose a deleterious burden on host cells, other OriV's showed different expression levels depending on the AbR they were paired with. The JUMP toolkit allows researchers to choose among ten different origins of replication and four antibiotic selection markers (Fig 3B and 3C). To show the effect of the vector chassis, we measured the expression of the constitutive sfGFP reporter, which is driven by promoter J23100 (Kelly, Rubin et al. 2009), in all empty vectors in the toolkit (Figure 3.4E). As expected, we found strong differences between different OriV and AbR. Expression was higher with the high-copy OriV #4 and #8, followed by the medium-copy #9, and the remaining OriV showed lower expression. The antibiotic resistance marker used also showed an effect on expression. AbR #4 (spectinomycin resistance), showed higher reporter expression than AbR #2 (kanamycin resistance) with all OriV. The fact that the difference increased with the copy number of the OriV suggests that this was caused by the expression of the resistance gene rather than the antibiotic itself.

The actual copy number of the vector was not quantified as it is known to vary depending on host strain, among other factors (Jahn, Vorpahl et al. 2016). These results were overall similar to those showed by the copy-number study of Jahn, Vorpahl et al. (2016), and the small differences (OriV #2 and #5 displaying slightly higher expression here) could be due to a different strain and different cargo being used. Nevertheless, these results show the value of testing different vectors if the expression of the assembled construct requires tuning for optimization.

The JUMP toolkit also offers conditional OriV's to allow chromosomal integration. The SEVA OriV #1 (R6K) depends on the *pir* gene, not present in most *E. coli* strains. We inserted the vector pJUMP21-1A in *E. coli* JM109 (*pir*-) by introducing a

sequence homologous to the 2kb upstream of *ldhA* in the upstream secondary site, obtaining kanamycin-resistant colonies in the *pir- E.coli* strain JM109 upon transformation. No transformants were obtained when the homology region was not included. Additionally, we built a thermosensitive oriV (#7ts) by replacing the Rep101 protein gene of SEVA's oriV #7 (pSC101-based) with its thermosensitive counterpart from the plasmid pCP20 (Cherepanov and Wackernagel 1995, Datsenko and Wanner 2000). After confirming that replication was temperature dependent (Appendix I, Figure 7.6), we introduced a counter-selection marker, *pheS* mutant A294G (Meyer, Segall-Shapiro et al. 2019), in the downstream secondary site to allow curable chromosomal integration, which we demonstrated with a scarless transcriptional fusion of *gltA* in *E. coli* MG1655 to RBS and sfGFP parts from the toolkit (Appendix I, Figure 7.7). The presence of this cassette, correctly inserted at the *gltA* locus, was confirmed by PCR.

3.5.3 Use of upstream and downstream sites

While the main site enables modular cloning as with other Golden Gate-based systems, the advantage of JUMP resides in the presence of additional Golden Gate cloning sites in all vectors. The upstream and downstream secondary cloning sites use AarI and BbsI to introduce any sequence into the vectors, thus not disrupting the function of the main site of any vector (level 0, 1 or 2). By being able to introduce any sequence in either secondary site of any vector, researchers can simplify the cloning steps needed to test the gene, genetic device or sequence of interest (SOI) (Figure 3.2C and D). Many recombinant SOI require other common sequences to be present (auxiliary elements such as a transcription factor for inducible expression, a biosensor or reporter gene, Cas9/dCas9 for CRISPR, homology sequences for chromosomal integration, etc). JUMP allows such auxiliary elements to be introduced in a secondary site, thus reducing the number of assemblies when the auxiliary elements are common but multiple SOI will be assembled or require optimisation.

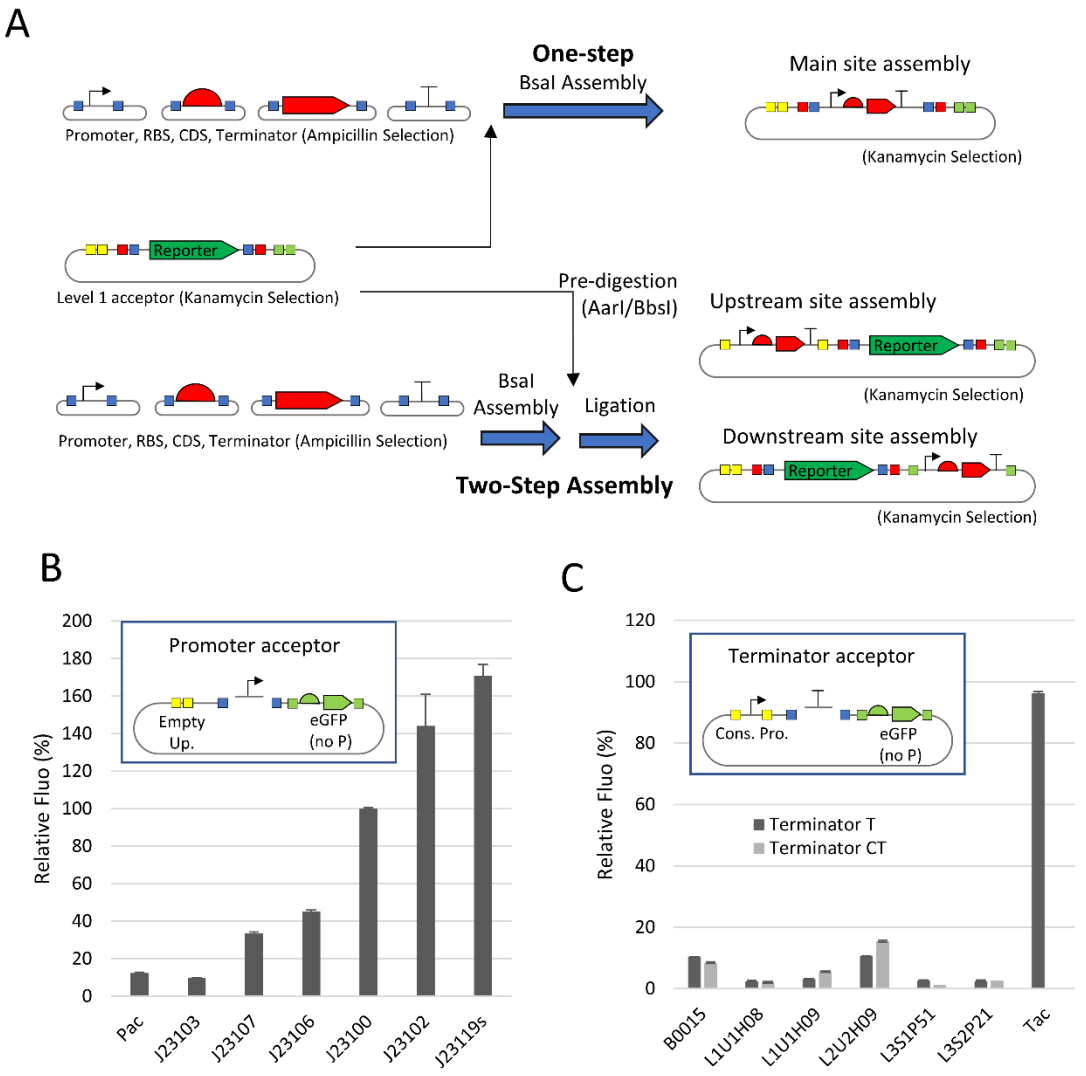


Figure 3.5. Use of secondary sites with two-step assembly. A: Two-step assembly works by assembling inserts first and then ligating the assembly reaction with the destination vector. Squares indicate restriction sites for BsaI (blue), BsmBI (red), AarI (yellow) and BbsI (green). B: Domestication and characterisation of promoters using level-0 promoter acceptor. C: Domestication and characterisation of terminators using level-0 terminator acceptor. T and CT terminators differ in the 5' end of the part (Appendix I, Figure 7.2), with CT terminators including a stop codon. In B and C, fluorescence was normalised to OD (600 nm) and is shown as % of that of the J23100 promoter. Error bars indicate standard deviation, n=9 (biological and technical triplicate).

The potential of secondary sites resides in how easily they can be used. By having the same receiving fusion sites, they are capable of receiving inserts assembled in the main site via a classic modular cloning approach (in addition to conventional restriction-ligation cloning). A conventional Golden Gate assembly (one-pot and one-endonuclease) does not allow assembly of inserts into secondary sites because

the inserts are assembled with BsaI/BsmBI and this would cut the destination vector in the main site. JUMP level 0 parts were originally designed with overlapping BsaI and BtgZI sites, to allow double-enzyme assemblies as shown by Sarrion-Perdigones, Vazquez-Vilar et al. (2013); however, one-pot one-step assembly with BtgZI digesting level 0 inserts and AarI or BbsI digesting the destination vector showed very poor efficiency, possibly due to BtgZI remaining attached to the DNA and interfering with the assembly (Radeck, Meyer et al. 2017). Therefore, we use an efficient two-step assembly approach. (Figure 3.5A). In the first step, the basic parts were assembled in the same conditions as for classic level 1 assemblies, and the destination vector was digested separately with AarI or BbsI for the upstream or downstream site, respectively, and dephosphorylated. In the second step, the digested backbone is ligated with the insert assembly mix. This assembly method was tested by assembling an mCherry TU from level 0 parts in the secondary sites of a level 1 destination vector. The transformation of the assembly resulted in more than 90% correct (mCherry+) colonies (Appendix I, Figure 7.5). We expect that this approach to modify vectors outside the conventional modular cloning site can easily be automated, as there is no need to perform any purification steps or PCR reactions to prepare inserts, since inserts can come from the modular cloning assembly pipeline without further modification. Additionally, AarI and BbsI sites do not have to be removed from inserts, as the restriction sites used for assembly on the main site (BsaI/BsmBI) are used on the sequence introduced in secondary sites. The two-step assembly can be extended to any insert and destination vector provided that the external fusion sites of the insert match the receiving fusion site of the destination vector (conserved for all JUMP vectors donor and acceptor sites), and the destination vector is selected with a different antibiotic than the insert donor vectors. If the antibiotic selection of the donor and recipient vector match, clones with recipient vector should be distinguishable by expressing the cloning marker still present in the main site.

We used secondary sites to characterise promoters and terminators (listed in Appendix I, Table 7.4 and available in Addgene) directly in the domestication (level 0) vector. We generated a level-0 promoter acceptor vector (pJUMP19-Pac) that reports the activity of the promoter by introducing eGFP without promoter in the downstream secondary site. During domestication of promoters, the eGFP indicated presence of the promoter and allowed characterization of the promoter strength (Figure 3.5B). We applied the same principle to domesticate and characterize

terminators by introducing a constitutive promoter in the upstream secondary site, in such a way that the expression of the downstream eGFP would be disrupted by the presence of terminators in the main site (Figure 3.5C). Thus, part characterization can be done directly in the level 0 donor vector, rather than requiring level 1 or level 2 assemblies (Iverson, Haddock et al. 2016, Moore, Lai et al. 2016).

A second example shows how the screening of a combinatorial assembly can be simplified using the secondary site (Figure 3.6). We generated a library of TU expressing different levels of the lambda phage repressor cI^{ts}, which we domesticated from pCP20 (Cherepanov and Wackernagel 1995, Datsenko and Wanner 2000). This was done to optimise this TU before using it to regulate other genes in a later level 2 assembly (not shown). To characterise the clones of the library, a reporter with the promoter controlled by cI must be present. We reduced the number of assembly steps needed to one per cI^{ts} by building the library in a level 1 vector where we had previously introduced the reporter gene in the upstream secondary site (Figure 3.6A). The clones from the library should vary in translation and stability of transcription factor, therefore we expected variability in the expression of the mCherry reporter at 30 °C and 37 °C, which we confirmed (Figure 3.6B).

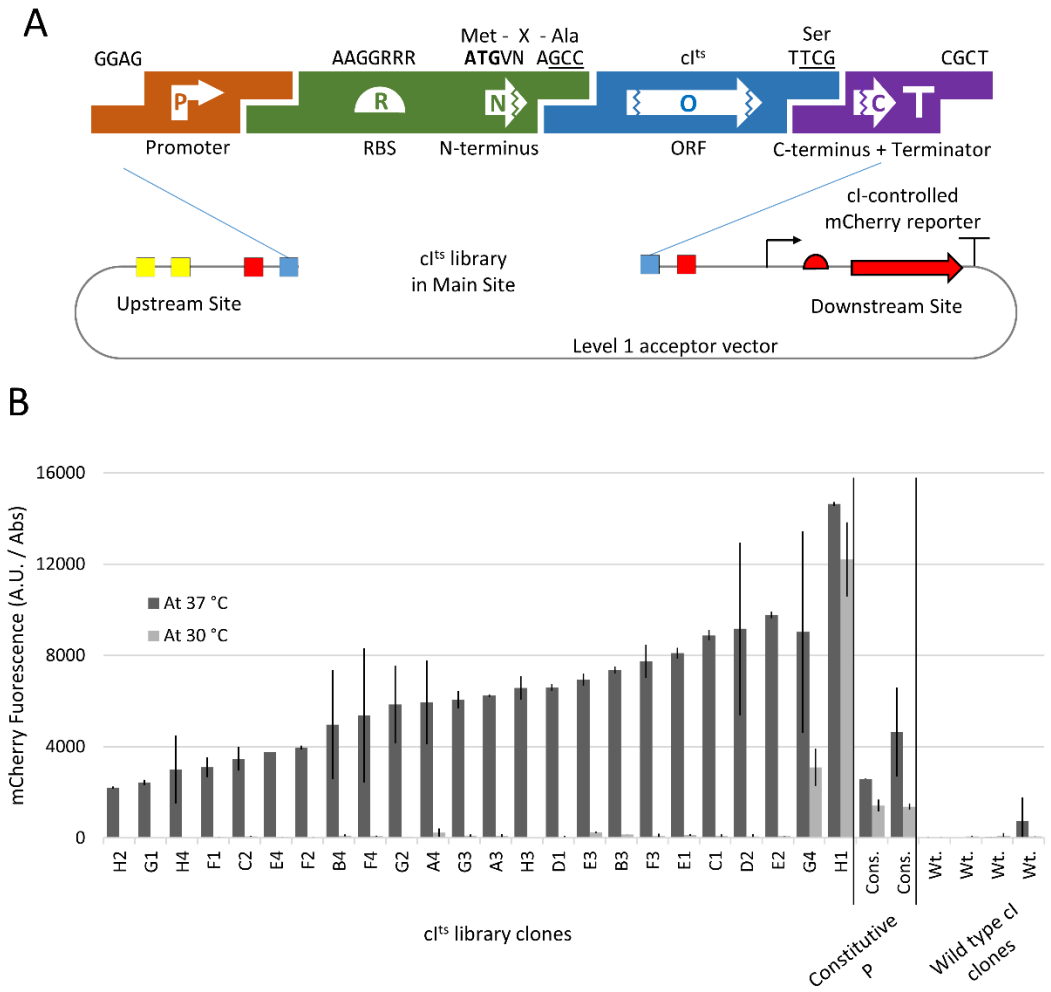


Figure 3.6. Use of secondary sites to screen variants of a transcription factor gene. A: The parts used for the construction of the cIts library were: a constitutive promoter, an RBS+N-terminus part, cIts CDS without N or C terminus, and C-terminus+Terminator. The RBS+N-terminus part was PCR-built from a degenerate oligonucleotide variable for the Shine-Dalgarno consensus and for the codon following the start ATG, thus giving the library diversity in translation initiation and half-life of protein (Looman, Bodlaender et al. 1987). B: Colonies that were white at 30 °C were tested for expression of the mCherry reporter at 37 °C. As control, we analysed the fluorescence of two colonies from the library that were unrepressed at 30 °C (clones G4 and H1), two colonies with a constitutive mCherry gene, and four colonies from an equivalent library assembled with the wild-type (thermostable) cI CDS part.

It is important to note that virtually any sequence can be introduced in secondary sites. Some recent publications (Iverson, Haddock et al. 2016, Schindler, Milbredt et al. 2016, Andreou and Nakayama 2018) have shown that new housekeeping elements normally present in the vector chassis (e.g. selection markers) can be introduced as assembly parts, thus increasing toolkit flexibility at the expense of assembly complexity. By using secondary sites, any new feature can be introduced

into JUMP vectors without increasing the complexity of later assemblies using the main cloning site. Therefore, tailored plasmids can be built for any application: auxiliary elements to test the assembly in the main site, recombinase recognition sites, counter-selection markers, homology sequences for chromosomal integration, alternative origins of replication, etc.

3.5.4 Toolkit distribution

A collection of 96 vectors is available in Addgene (https://www.addgene.org/browse/article/28203402/) (Appendix I, Tables 7.3 and 7.4). This collection includes a selection of PhytoBrick-compatible level 0 parts for synthetic biology applications in bacteria (as shown in), a basic part Universal Acceptor, the level 1 and level 2 vector "core set", and alternative vectors. We have removed relevant type IIS restriction sites from all these basic parts and vectors, with the exception of the OriV #2x, (unmodified SEVA broad host range oriV #2/R2K, which is BsaI free but contains other forbidden sites). Addgene kindly sequenced all the deposited plasmids confirming the absence of unwanted mutations.

3.6 Conclusions

JUMP has been designed as a platform to allow Type IIS-RE-based assemblies to be used in multiple bacterial hosts as well as allowing easy modification of the vectors to suit the requirements of new hosts. The capabilities of Golden Gate-based assembly are combined with the vector flexibility of SEVA and multiple cloning sites. The toolkit generated allows any assembly level be done with ten different origins of replication. The vectors range from low to high copy number, include integrative vectors and can be used with a very wide list of non-model microorganisms. Since they are based on SEVA, future OriV's of the collection and AbR's added to the collection can be easily incorporated into JUMP vectors, and researchers can easily create new origins of replication and markers for the toolkit or use JUMP vectors with any PhytoBrick-compatible part. We have demonstrated that the upstream and downstream secondary sites expand the paradigm of modular cloning to allow orthogonal modification of the vector chassis.

3.7 Addendum

The example level 2 assembly used to demonstrate multi-gene plasmid construct with JUMP vectors was done using genes coding for the cellulases CenA, Chu2268,

alcohol dehydrogenase Adh and pyruvate decarboxylase Pdc. The results appear in Appendix I, Figure 7.4 as they were originally published with the manuscript but they are explained more in detail in Chapter 4. The domestication and sequence of those parts are shown in Appendix II. Some additional standard parts of general interest are shown in Appendix II, which include the silica-binding tags based on the work of Taniguchi, Nomura et al. (2007) (part Si-tag_N) and Abdelhamid, Ikeda et al. (2016) (part SB7_C), and linker parts with the 100-bp sequences of the unique flanking nucleotide sequences used by the Loop assembly system (Pollak, Cerda et al. 2019) to allow compatibility with Gibson assembly (parts UNS-2, UNS-3, UNS-4, UNS-5, UNS-7).

The two-step assembly protocol, developed to use secondary sites in a fast and automation-friendly way, was found to be quite efficient, but the efficiency of assemblies in the main site was not determined. The efficiency of any cloning assembly reaction (including conventional Golden Gate cloning in JUMP's main module) is expected to depend on several factors. The first set of factors includes the sequence of the fusion sites, restriction enzymes and protocol used, which are not different between JUMP and other assembly systems which have shown good yields (Sarrion-Perdigones, Vazquez-Vilar et al. 2013, Iverson, Haddock et al. 2016, Moore, Lai et al. 2016, Andreou and Nakayama 2018). The second set of factors expected to affect cloning efficiency include the size of insert, biological functions of elements coded in insert and vector backbone, and even the quality of cloning enzyme and competent cells used for transformation. The latter factors vary between laboratory supplies and cloning application, and finding their specific effect on cloning was found to be beyond the scope of this work.

JUMP vectors and modular assembly should be useful to generate IBPM candidates, allowing optimisation of gene expression in the quantitative regulatory elements (promoter, RBS, terminator, copy number) and optimisation of qualitative variables such as gene and secretion signal. To facilitate N- and C-terminal fusion of enzymes to secretion parts, coding sequences were mainly designed in the open format ORF, which lacks start and stop codon (Appendix I, Figure 7.2 shows formats of parts generated, matching the PhytoBrick standard). For compatibility with a closed coding sequence format (part type NOC, standing for N-ter + ORF + C-ter), RBS parts were designed in two ways, with and without N-ter (part types R, RN); and terminator parts were designed with and without C-ter end (part types T and

CT). These formats of parts were designed to allow flexibility using coding parts of the open and closed formats.

This dissertation adopts *E. coli* as an experimental model, but other potential organisms are also attractive chassis to develop IBPM candidates. The flexibility of JUMP vector backbones should facilitate the transfer of tools and standard parts to additional hosts beyond *E. coli*. In addition to broad-host-range vectors, secondary sites allow introduction of additional features required to use new microbial chassis. Standard parts comprising an origin of replication and selection marker have been used to introduce JUMP vectors in *Saccharomyces cerevisiae* (Florentina Winkelmann, unpublished). Moreover, secondary sites should be beneficial to introduce additional cellulases or screening-related genes in a vector to generate combinatorial assemblies of a single gene of interest, which will be shown in Chapter 5 with a demonstration of a two-step assembly introducing multiple TU in the secondary site of a level 1 vector.

Chapter 4 Strain Screening: new approaches

4.1 Abstract

Different alternatives were explored to find a methodology that could be applied in a high throughput way to screen the cellulolytic capabilities of engineered microbes. Detection of a synthesis product (ethanol) was found to not be high-throughput nor sensitive enough when engineering non-cellulolytic candidates. Artificial evolution was explored to select the expression strength of an intracellular cellulase Chu2268, and was found to be too slow in removing *chu2268*-negative clones, but the selected clones could be useful for future more complex assemblies. It was discovered that fluorescence could be measured directly in the presence of cellulose without separation steps, and it could be used to indirectly detect the growth of microbes digesting cellulose. Fluorescence from mCherry was found to correlate linearly with cell concentration, but the proportionality depended on cellulose form and concentration. In conditions of (exponential) balanced growth, fluorescence could be used to describe growth well in ideal conditions with extended exponential growth in microplate cultures.

4.2 Introduction

Screening of synthetic biology candidates for a desired function is based on linking a genotype that has the desired function with a measurable or selectable phenotype. Isolation of clones allows them to be analysed and characterized individually. This isolation can be a physical separation on an agar plate, separation of individual clones in individual liquid cultures (ranging from shake flasks to microtiter plates), or in microdroplets in a culture emulsion. As mentioned in Section 1.3, the screening step is often the bottleneck of the Design-Build-Test-Learn cycle, and increasing the screening throughput would be desirable. The throughput of a technique is the number of different clones that can be screened or selected, which depends on the trait of interest (Jeschek, Gerngross et al. 2017). The throughput is higher when candidate cells can be characterised in smaller volumes (e.g. a microtiter plate allows many more simultaneous cultures than shake-flasks)(Mühlmann, Kunze et al. 2017). A special case occurs when in directed evolution experiments the trait of interest is linked to growth and survival because this allows selection of candidates in the same culture without physical separation (Zhou and Alper 2019).

Screening is sometimes the limiting step in the research cycle (Jeschek, Gerngross et al. 2017). This is especially true for developing Ideal Biofuel Producing Microorganism (IBPM) candidates because the selection of candidates for their cellulolytic capability is limited by the microcrystalline nature of cellulose (French 2009). Since it is not soluble in culture media, cellulose is often introduced in the form of suspensions which cause optical turbidity and heterogeneity. Because cellulose must be extracellularly digested, lytic enzymes must be secreted, which impedes directed evolution experiments where clones are not separated physically. Moreover, microdroplet-based screening has not been applied in the presence of cellulose due to its non-homogeneous nature in cultures, and it is necessary to isolate clones in individual cultures. Turbidity blocks most optical measurements and analytical techniques, which is often circumvented using soluble cellulose analogues (Ostafe, Prodanovic et al. 2013) or applying separation steps (Mühlmann, Kunze et al. 2017), such as centrifugation to partition the insoluble cellulose from the sample supernatant (SN) containing soluble molecules. And unfortunately, every required step reduces the throughput capacity.

Assessment of IBPM candidates is done by quantifying elements of the CBP process (Figure 4.1), but experimental conditions can be far from the real process. Similar to the screening of microbes engineered with multiple genetic modifications, the measured trait can provide information on the expression and activity of individual gene modifications or it can be a trait that depends on several of the gene modifications, thus giving a global picture of the system’s alteration.

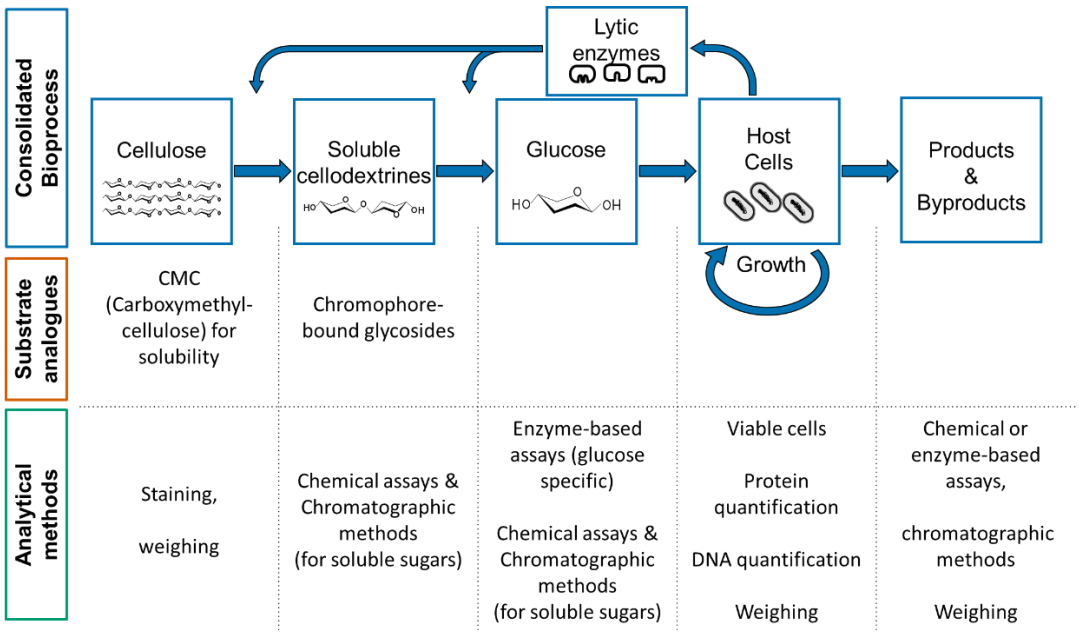


Figure 4.1. Existing cellulase characterisation strategies applied at different steps of a consolidated bioprocess.

Functional characterisation of individual cellulases often uses substrate analogues with higher solubility than cellulose or linked to chromophores to directly quantify enzymatic cleavage (Table 4.1). For example, carboxymethyl-cellulose has some hydroxyl groups of glucose monomers substituted by -CH_2-COOH groups, which impede chain-to-chain interactions thus avoiding crystallinity. Chromogenic substrates are cleaved by specific enzymes releasing a measurable chromophore. For example, *para*-nitrophenyl-β-D-glucopyranoside (p-NPG) is hydrolysed by BG enzymes releasing *p*-nitrophenol, which exhibits absorbance at 400 nm.

Table 4.1. Substrate analogues recognised by cellulases.

Substrate	Enzymatic activity detected	Reference
CMC	EG	(Fan, Zhang et al. 2016)
CMC	CBH	(Tang, Wang et al. 2018)
Azo-CMC	EG (requires substrate separation)	(Bokinsky, Peralta-Yahya et al. 2011) (Lakhundi, Duedu et al. 2017)
Azo-xylan	Endoxylanase (requires substrate separation)	(Bokinsky, Peralta-Yahya et al. 2011)
4-Methylumbelliferyl-β-d-glucopyranoside	CBH (on agar plates or in liquid)	(Lakhundi, Duedu et al. 2017)
4-Methylumbelliferyl-β-d-glucopyranoside	Total cellulase activity	(Chang, Anandharaj et al. 2018)
4-Methylumberiferyl β-D-lactopyranoside	CBH	(Davison, den Haan et al. 2016)
4-nitrophenyl-β-D-cellobioside	CBH	(Tozakidis, Brossette et al. 2016)
4-nitrophenyl-β-D-glucopyranoside	BG	(Anusree, Wendisch et al. 2016, Tozakidis, Brossette et al. 2016)
Azurine-cross-linked hydroxyethylcellulose (AZCL-HE-Cellulose)	EG	(Sasaki, Mitsui et al. 2019)
4,6-*O*-(3-Ketobutylidene)-4-nitrophenyl-β-D-cellopentaoside	EG and BG	(Wightman, Kroukamp et al. 2020)
4,6-O-(3-Ketobutylidene)-4-nitrophenyl-β-D-cellopentaoside	EG (supplied with additional BG)	(Claes, Deparis et al. 2020)

Using model substrates has allowed some HTS to screen cellulases (Table 4.2). Nonetheless, activity towards analogue substrates might not correlate with activity towards real substrates and may also overestimate activity compared towards more realistic conditions where the substrate will be less soluble and accessible.

Table 4.2. High-throughput screening methods applied to cellulases and cellulolytic organisms.

Enzyme or candidates screened	Screening scale	Analytical or selection principle	Reference
Intracellular cellulase (different classes)	Survival and selection	Yeast three-hybrid system adapted to detect a cello-oligosaccharides with ligands bound at each end	(Peralta-Yahya, Carter et al. 2008)
Secretion of cellulases	Microdroplets and fluorescence-assisted sorting	Detection of CMC hydrolysis to glucose using hexose oxidase, which generates H_2O_2 which acts on a reagent generating fluorophore.	(Ostafe, Prodanovic et al. 2013)
Bioprospecting of soil bacteria with CBH activity	Microdroplets and fluorescence-assisted sorting	Fluorogenic substrate (β-D-cellobioside-6,8-difluoro-7-hydroxycoumarin-4-methanesulfonate)	(Najah, Calbrix et al. 2014)
Biomining enzyme screening from metagenomics-based libraries of recombinant clones.	Robotics-assisted Microplate assays	Recombinant expression assay of cell lysates with Azo-CMC, CMC (sugars detected with PAHBAH), and MUC	(Mühlmann, Kunze et al. 2017)
Selection of EG library generated by random mutagenesis	Agar and microplate assays	CMC and Congo red staining (agar), Azo-CMC (microplate cultures)	(Cecchini, Pepe et al. 2018)

The activity of one or multiple cellulases towards realistic substrates can be characterised by quantifying the soluble sugars released from the substrate (Table 4.3). Colorimetric techniques generate an optical measure proportional to the concentration of the analyte, such as 3,5-dinitrosalicylic acid (DNS) which reacts with all soluble reducing sugars generating a signal at 450 nm. However, these methods depend on the analyte not being consumed, hence they cannot be applied in CBP-like conditions where the microbial host should be using the sugars. Other approaches assess the global cellulolytic activity by indirectly quantifying the sugars consumed by the microbial candidate, such as measuring product generated (e.g., ethanol) with methods that often have limited throughput due to the need for separation steps (Table 4.3). Another way of quantifying use of sugars released from cellulose is by measuring microbial growth (Table 4.3). The archetypal way of measuring microbe concentration through optical density (OD) caused by scattering

(Stevenson, McVey et al. 2016) is completely impeded by cellulose turbidity. Researchers have turned to counting colony forming units (Lakhundi, Duedu et al. 2017) or quantifying specific biomolecules. Total protein concentration has been quantified with colorimetric assay (Salinas 2017) and DNA can be precisely quantified in the presence of cellulose using fluorescent dyes (Duedu and French 2017). This last method was especially interesting as it showed that fluorescence could be measured despite turbidity.

Table 4.3. Analytical methods used to detect elements in the CBP process.

Analyte	Assay	Reference
Avicel	Congo red staining in agar plates	(Wei, Wang et al. 2014)
CMC	Congo red staining in agar plates	(Gao, Luan et al. 2015)
CMC	Viscosity measurement of liquid	(Fan, Zhang et al. 2016)
Cellobiose	Iodine staining in agar plates	(Anusree, Wendisch et al. 2016)
Reducing sugars	Dinitrosalicylic acid (DNS) assay of supernatants	(Gao, Luan et al. 2015, Anusree, Wendisch et al. 2016, Tozakidis, Brossette et al. 2016)
Reducing sugars	4-Hydroxybenzhydrazide (PAHBAH) assay of supernatants	(Mühlmann, Kunze et al. 2017)
Hydrolysed soluble sugars (xylose, xylobiose, xylotriose, glucose, cellobiose, and cellotriose)	High-performance anion exchange chromatography of supernatants	(Stern, Moraïs et al. 2018)
Glucose	High-performance Liquid Chromatography of supernatants	(Lee, Sung et al. 2017)
Cellobiose	High-performance Liquid Chromatography of supernatants	(Hetzler, Bröker et al. 2013)
Glucose	Enzymatic assay commercial kit (glucose oxidase) of supernatants	(Liang, Si et al. 2014)
Cell mass and solids (Avicel)	Weighing centrifuged pellet without supernatant (the cellulose fraction was calculated with posterior acid hydrolysis and HPLC analysis)	(Guo, Duquesne et al. 2017, Guo, Duquesne et al. 2017)
Microbial cells	Haemocytometer counting	(Liang, Si et al. 2014)
Viable cells	Counting colony forming units after plating in agar	(Bokinsky, Peralta-Yahya et al. 2011)

Table 4.3. Analytical methods used to detect elements in the CBP process.

Analyte	Assay	Reference
Cell density	OD (with CMC as substrate)	(Kickenweiz, Glieder et al. 2018)
Total protein	Bradford Assay	Salinas thesis
DNA	Fluorescent staining with SYBR I Green nucleic acid gel stain and Propidium Iodide	(Duedu and French 2017)
Fatty acids	GC/MS of supernatants	(Bokinsky, Peralta-Yahya et al. 2011)
Pinene, Butanol	GC of supernatants	(Bokinsky, Peralta-Yahya et al. 2011)
Ethanol	HPLC of supernatants	(Tang, Wang et al. 2018)
Ethanol	GC of supernatants	(Shin, Hyeon et al. 2015)
Ethanol	Enzymatic assay (ADH) of supernatants	(Lewicka, Lyczakowski et al. 2014)
CO_2	Weighting whole culture (CO_2 release reduces weight)	(Claes, Deparis et al. 2020)
Cellulase detection after polyacrylamide gel electrophoresis	Zymogram (gel is submerged in 4-Methylumbelliferyl-β-d-cellobioside solution to detect cellulase activity, or laid over Agar-CMC and stained with Congo red)	(Chang, Anandharaj et al. 2018)

An ideal screening method should be applicable in experimental conditions that are relevant and informative of the behaviour of microbial candidates in the final production conditions. In the case of engineering IBPM candidates, those conditions should include *in situ* secretion of enzymes, cellulolysis and use of the resulting sugars. Such methodology should allow combined assessment of additional relevant characteristics of the candidate in addition to cellulase activity (such as secretion capacity, growth, and genetic stability), namely a “global” measurement. Additionally, the screening approach should be of high-throughput capacity to allow screening for non-sequential optimisation approaches (such as combinatorial assembly). Unfortunately, none of the existing methods fulfil all these requirements for an ideal screening system.

4.2.1 Aims

The aim of the work in this chapter is the development of a high-throughput screening methodology for cellulolytic activity from engineered microorganisms in CBP-relevant conditions. Different approaches will be explored: increasing the throughput of existing methods that are CBP-relevant, applying directed evolution and, finally, trying to develop real-time measurements of microbial growth in the presence of cellulose.

A paradigm of cellulose degradation consists of a complex combination of different cellulases working synergistically (Introduction, Section 1.2.2.1). However, a simpler mechanism would be preferable in the first instance to facilitate the optimisation of the screening methodology. The work in this chapter will be based on the simplified system used by Dr Salinas (Salinas 2017) which demonstrated the cellulolytic capacity of *E. coli* when expressing CenA and Chu2268 (Section 1.3.1).

4.3 Results

4.3.1 Screening via product synthesis

The final value to optimise in a CBP is the amount of product formed from the starting cellulose. This requires maximising cellulase activity and directing most carbon available from depolymerised cellulose to product synthesis, with just enough being used for cellular growth and maintenance.

The medium LB does not contain carbohydrates that can be fermented to ethanol by *E. coli* (Sezonov, Joseleau-Petit et al. 2007), and consequently ethanol produced must be generated from additionally available sugars. In unpublished experiments by Teresa Garcia Ybarra and Prof. Chris French, an *E. coli* strain was engineered to express CenA, an endoglucanase from *Cellulomonas fimi* which is apparently secreted from *E. coli* by an unknown mechanism, and Chu2268, a cellodextrinase from *Cytophaga hutchinsonii* which has been shown to allow growth of *E. coli* on cellooligosaccharides even when expressed intracellularly, in addition to Pdc and Adh, two enzymes known to increase the yield of ethanol from glucose in bacteria (Lewicka, Lyczakowski et al. 2014). The resulting strain was apparently able to generate a much higher concentration of ethanol when growing in LB medium with cellulose filter paper than in medium without cellulose. However, these results were inconsistent and difficult to reproduce.

The work of this section attempted to reproduce that result to determine whether ethanol production can be used as a high-throughput screening method for cellulase activity. Additionally, the plasmid generated was used and published as an example of a level 2 assembly using JUMP (Valenzuela-Ortega and French 2021).

4.3.1.1 Strain construction

Based on the results from Alejandro Salinas and Teresa Garcia Ybarra, the four gene combination (encoding CenA, His-tagged Chu2268, Pdc, Adh) was constructed using JUMP vectors (Figure 4.2). The level 1 single-gene constructs were built with different promoters. The promoters used with CenA and H6-Chu2668 were the inducible P_{lac} (R0010) and the constitutive J23116, J23106 and J23100; while for Pdc and Adh only J23106 and J23100 were used. No colonies with correct assembly could be isolated for the cellulase genes with the strongest constitutive promoter (J23100), which could be due to a high burden from the constitutive expression of these genes.

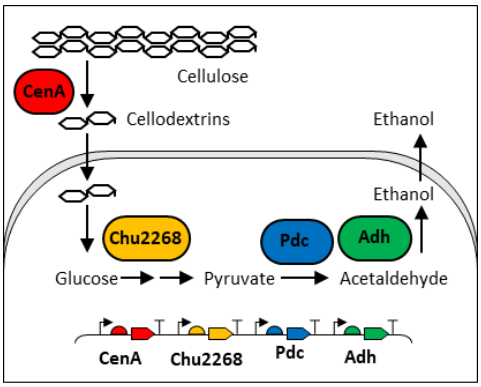


Figure 4.2. Design of a cellulolytic ethanologenic construct (pMV170). CenA endoglucanase is secreted and cleaves cellulose releasing soluble cellodextrins. Cellodextrins are intracellularly digested to glucose by His-tagged Chu2268 cellulase. Glucose is used by *E. coli* metabolic machinery, and glycolysis generates pyruvate. Pdc (pyruvate decarboxylase) and Adh (acetaldehyde dehydrogenase) transform pyruvate into ethanol. Reprinted from Valenzuela and French (2021) (accepted for publication).

Consequently, based on the possibility that stronger expression would cause a stronger burden, the level 2 assembly (named pMV170) was only built with the inducible P_{lac} for *cenA* and medium strength J23106 promoters for the rest of the genes (Table 4.4). Some of the resulting clones could be confirmed with the correct assembled sequences and were confirmed to have endoglucanase activity as shown in Figure 4.3.

Table 4.4. Vectors and inserts used to assemble the cellulolytic ethanologenic construct (pMV170).

#	Plasmid	Description
pMV033	pJUMP49-2A (sfGFP)	Level 2 JUMP destination vector.
pMV146	pJUMP29-1A: J23106-H6-Chu2268	Level 1 assembly (1A): J23106 - pET_RBS_R-His6_N - Chu2268_O - BBa_B0012_CT
pMV152	pJUMP29-1B: Plac-cenAo	Level 1 assembly (1B): R0010(plac) - pET_RBS_R - cenAo_NOC - L3S2P21_T
pMV136	pJUMP29-1C: J23106-Pdc	Level 1 assembly (1C): J23106 - pET_RBS_MV_RN - zmPdc_O - L3S1P51_CT
pMV138	pJUMP29-1D: J23106:Adh	Level 1 assembly (1D): J23106 - pET_RBS_MV_RN - zmAdh_O – L2U2H09_CT

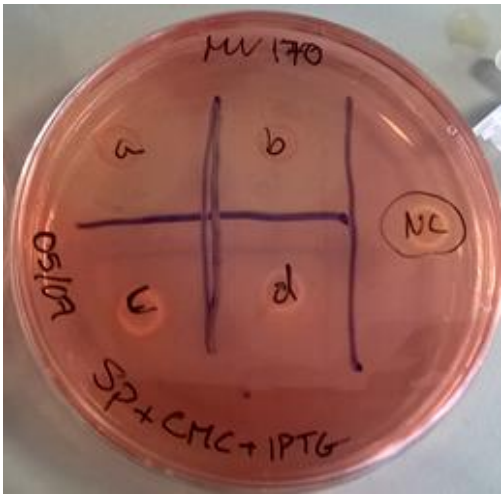


Figure 4.3. Endoglucanase activity in candidate clones of cellulolytic ethanologenic construct (pMV170). LB agar plate containing CMC was stained with Congo red, showing halos around spotted clones that have endoglucanase activity. Letters a,b,c,d indicate four candidate clones of assembly pMV170. NC is negative control with empty destination vector (pJUMP49-2A(sfGFP)). Clones “a” and “b” were confirmed by Sanger sequencing, and clone “a” was selected for further experiments.

E. coli JM109 with pMV170 (strain MV170) showed much slower growth than with empty level 2 destination vector. With triplicate overnight cultures, strain MV170 reached OD values of 2.04, 2.04 and 2.12 while the strain carrying the equivalent empty vector reached OD values of 4.15, 3.9 and 3.97. This could be due to the sum of the burden caused by each of the three constitutive genes in pMV170.

4.3.1.2 Characterisation of ethanol production

Strain MV170 was tested for production of ethanol from cellulose in LB medium. This strain and a control carrying the empty pJUMP49-2A(sfGFP) were grown overnight, then resuspended (initial OD of 0.16) in fresh LB medium with 0.1 mM IPTG and 0.75 % (w/v) of different carbon substrates to determine activity of the enzymes (which is 42 mM for glucose). Medium was supplemented with 25 mM MOPS buffer (pH 7.3) to avoid acidification by glucose fermentation. The cultures, of 10 mL volume, were incubated in 13-mL flasks in a rotating-mill rather than on a shaker to reduce aeration. Samples were taken at different times, and the supernatant was used to quantify ethanol with the method used by Lewicka, Lyczakowski et al. (2014), which was modified for use in a microtiter plate reader to increase throughput.

The experimental strain produced more ethanol than the control, and this was higher in the presence of glucose or cellobiose, but not phosphoric acid-swollen cellulose (PASC). This proved activity of all genes except *cenA*, probably the bottleneck in this pathway. The activity of CenA, while present (as shown in Figure 4.3), was apparently not enough to digest real cellulose at a rate that would allow ethanol to accumulate. Hence, without a more active cellulose combination, the quantification of cellulolytic activity from product formed is not sensitive enough. Additionally, while the ethanol assay had been scaled down to a plate reader format, it still required sampling, preparation (cellulose separation) and reaction incubation before a measurement could be taken, and consequently it was found to not be a viable high throughput screening assay.

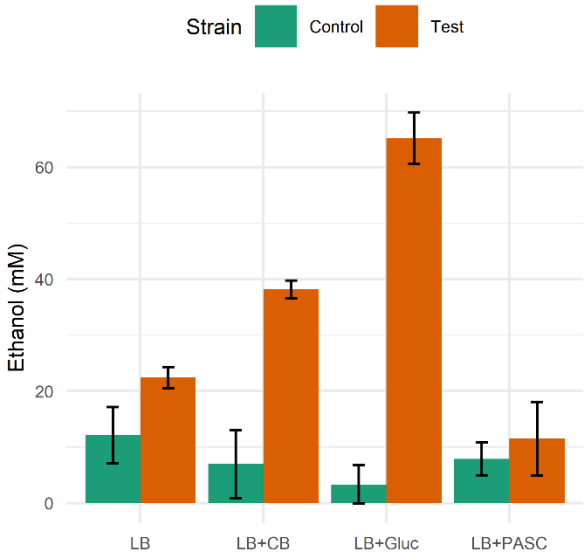


Figure 4.4. Ethanol formed by cellulosic ethanol-producing strain (MV170) on LB medium supplemented with different carbon sources after 48 h of incubation. Control strain is *E. coli* JM109 carrying the equivalent empty vector pJUMP49-2A(sfGFP). Bars indicate mean of biological triplicates; error bar indicates SD. Reprinted from Valenzuela and French (2021) (accepted for publication).

4.3.2 Screening via artificial evolution

Application of directed evolution to select cellulolytic microbes is impeded by the fact that cellulose is degraded extracellularly. This breaks the connection between genotype and phenotype needed for selection: the cellulases secreted by a single cell benefit the whole population. Interestingly, microbes that digest cellulose in competitive environments tend to display cellulases on their surface, thus physical proximity gives them an advantage in absorbing the sugars released from the extracellular digestion (Zhu and McBride 2017).

Inspired by this natural phenomenon, directed evolution was explored to select for cellulolytic candidates. The experimental model was simplified by using a soluble substrate (cellobiose) known to be digested intracellularly by a recombinant cellulase (Chu2268). A library of clones with different expression levels of this gene were grown on minimal medium with cellobiose as principal carbon source, which should enrich for the clones expressing just enough enzyme (so carbon is not the bottleneck) but not enriching for clones expressing so much Chu2268 that it becomes a burden (Figure 4.5A). Medium M9y (supplemented with yeast extract) was used as it had previously been shown to allow a recombinant *E. coli* strain to grow from cellulose (Salinas 2017). It would be expected that, from an initial

population varying in the strength of expression of Chu2268, clones that had the optimal expression should divide faster and be enriched after multiple selection passages (Figure 4.5B).

Additionally, the resulting selected Chu2268 level 1 assemblies should be useful for later level 2 assemblies.

4.3.2.1 Chu2268 library design

Previously generated JUMP parts were used to make two combinatorial libraries of cassettes expressing Chu2268 with diverse promoters, RBSs and protein N-terminus (Figure 4.5C and D). To avoid homology with the constitutive promoters of the J23100 family, the assemblies were done with Ptet_P, Pars_P, and Pcue_P. Despite being considered inducible promoters, they were expected to behave as constitutive in the experimental conditions because of the lack of their effectors (tetracycline and heavy metals). Previously characterisation of promoters indicated that, in *E. coli* JM109 (Section 3.5.3), P_{tet}, P_{ars} and P_{cue} had a relative strength of 106, 21, and 7 % compared to J23100, respectively in the uninduced state in the same strain and using the same OriV (although in LB medium).

The ribosome binding site (RBS) of the libraries was a partially degenerate Shine-Dalgarno sequence AAGRRR, with a 15-fold strength variation as calculated with the RBS calculator (Salis, Mirsky et al. 2009). The degenerate RBS had been used previously to build the library of thermosensitive cI-coding genes (Chapter 3). The N-terminus varied between the two built libraries. Following the design by Dr Salinas (Salinas 2017), the library "H6" had a hexa-histidine tag in the N-terminus of Chu2268 (part type O, without signal peptide); and library X was built with a degenerate codon after the methionine start codon, affecting protein half-life and translation initiation with between 3 and 15-fold different strength (Looman, Bodlaender et al. 1987, Salis, Mirsky et al. 2009).

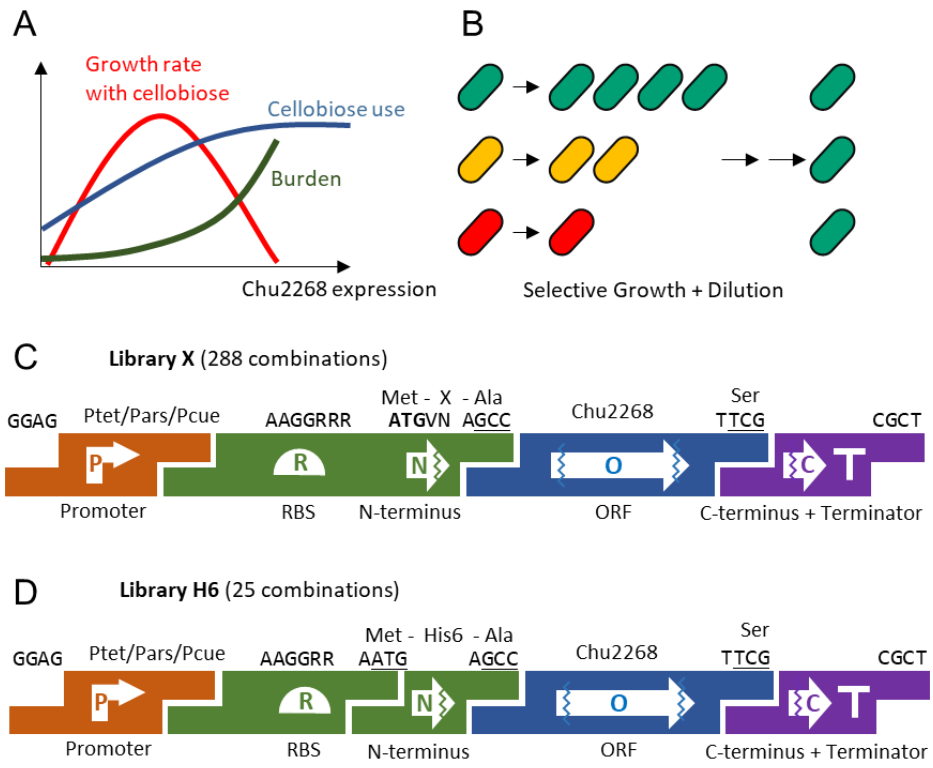
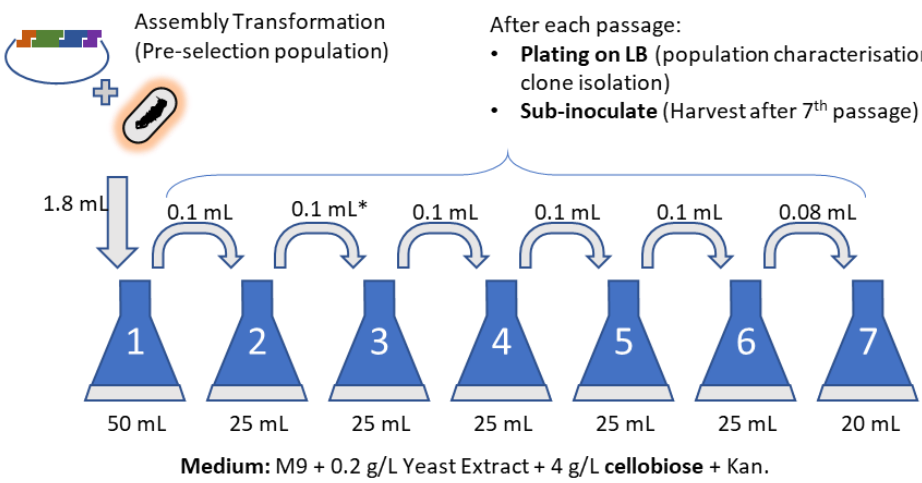


Figure 4.5. Artificial evolution to tune Chu2268. A: expected effect of expression strength of Chu2268 on growth rate, with a maximum caused by a balance between the positive effect of the cellobiose usage capacity and the negative effect of the expression burden. B: enrichment of clones with improved growth by applying selective growth and dilution passages. Multiple passages will enrich fastest growing clones. C: assembly of Library X for degenerate RBS and second codon (M-X-Chu2268). D: assembly of Library H6 for degenerated RBS and histidine tag (M-H6-Chu2268).

4.3.2.2 Library Selection

E. coli JM109 was transformed (in duplicate) with the combinatorial assemblies. Ninety percent of the duplicate transformation mixes were used to inoculate the first passage of the selection process (Figure 4.6A) (each library separately). The remaining 10 % of the transformation was plated to evaluate the assembly (considering white colonies successful assemblies and GFP+ colonies parental destination vector). The libraries were expected to contain all the possible combinations: library X was estimated (based on number of colonies from the 10 % that was plated) to have 732 different clones in the minimal medium inoculation (for 288 possible combinations), and library H6 had 45 clones (for 24 combinations).

A



Medium: M9 + 0.2 g/L Yeast Extract + 4 g/L **cellobiose** + Kan.

B

$$Growth\ per\ culture = \frac{Biomass_{final}}{Biomass_{initial}} = \frac{Vol_{final} \times Conc._{final}}{Vol_{initial} \times Conc._{initial}} \rightarrow \frac{Vol_{final}}{Vol_{initial}}$$

$$\frac{50}{1.8} \times \left(\frac{25}{0.1}\right)^{6} = 27.7\hat{7} \times (250)^{6} = 6.78 \times 10^{15}$$

$\log_2 6.78 \times 10^{15}$ = 52.5 generations for Library X (50.5 generations for library H6*)

C

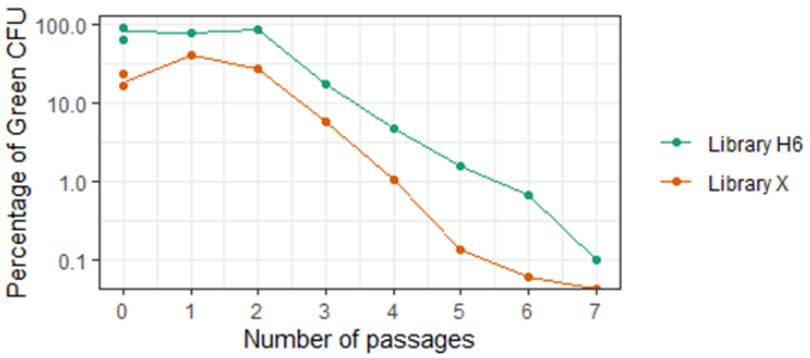


Figure 4.6. Selection of the assembly libraries with selective growth medium. A: assembly transformation was used to inoculate the selective medium (M9y+cellobiose+Kan), seven sequential passages were done. B: The number of divisions per passage was calculated by assuming that the cell concentration of the inoculum was the same as the one at the end of the passage, and therefore the cell divisions per passage were the log2 of the ratio of total volume to inoculum volume. C: fraction of total population of cells expressing GFP after each passage. After each passage, a culture sample was plated on LB+Kan to determine what fraction of the population of CFU was GFP+ (carrying parental level 1 vector without Chu2268). Dots indicate percentage of GFP+ CFU for individual plates, and line indicates mean (Two plates for transformation or passage 0, 1 plate for the rest). *After 2nd passage of Library H6, 0.4 mL instead of 0.1 mL were used as sub-inoculum due to a low OD.

The selection process (Figure 4.6A) consisted of seven sequential culture passages with selection medium (M9y + 4 g / L cellobiose +Kan), incubated at 37 °C for 24 hours. Medium M9y (with 0.2 g / L yeast supplementation) was chosen because it was the one previously shown to allow *E. coli* to grow from cellulose by Dr Salinas. At the end of each passage, cultures were used to inoculate the next passage. It was calculated that after seven passages, both libraries had gone through more than 50 generations (Figure 4.6B).

At the end of each passage, cultures were spread on LB plates to characterise how the sub-population of GFP+ cells (carrying the parental destination vector) were removed by the growth selection (Figure 4.6C). It was found that the fraction of GFP+ cells were decreasing slowly and only after the second passage. Surprisingly, the GFP+ sub-population grew during the first two passages. This could be due to a combination of different possible reasons. The first passage contained a fraction of rich medium (1800 µL transformation mix containing SOC broth were used to inoculate the 50 mL of the first culture) allowing some growth regardless of the ability to assimilate cellobiose. Additionally, the clones with the strongest Chu2668 expression could be suffering a burden that led to cell lysis, with the subsequent disappearance of this sub-population and simultaneous release of cellulase to the extracellular medium, which in turn allowed *chu2268*- cells to grow. The sub-population of GFP+ colonies decreased after the second passage with a logarithmic trend, demonstrating the enrichment of cells expressing Chu2268. However, this counter-selection of Chu2268- cells was “slow”, suggesting that either *chu2268*+ clones were releasing enzyme, glucose or other nutrients that allowed *chu2268*- clones to grow.

Even though the selection was found to be of limited success due to the slow disappearance of *chu2268*- clones, the selected clones could still be useful for later level 2 assemblies that combined this with additional genes.

4.3.2.3 Genotypic characterisation of selected clones

Four clones from each library were sequenced, suggesting selection to a similar expression strength (Table 4.5). No clone had the strongest promoter (P_{tet}). Three of the four clones of Library X had the weakest promoter (P_{cue}); and one clone with the medium-strength P_{ars} (clone “7X1”) had a much weaker RBS than the other 3 clones. Interestingly, the clone with P_{cue} and the strongest RBS (clone “7X3”) had a

short half-life N-terminus amino acid (lysine) which might have led to a similar cellular amount of Chu2268 as in the other clones (Tobias, Shrader et al. 1991). All successfully sequenced clones from Library H had the weakest promoter (P_{cue}) and had a stronger RBS than those found in clones from Library X, which could be due to the translation of this protein with a histidine tag requiring a stronger RBS to achieve the same level of protein per cell.

Table 4.5. Sequence analysis of four clones from each library after selection. The sequence corresponding to the RBS and N-terminus was identified and analysed for translation initiation strength (Salis, Mirsky et al. 2009). Clones 7X1-7X4 came from Library X, and clones 7H1-7H4 come from Library H.

Clone	Promoter	RBS	Start	Amino acid after start codon (and half life) (Tobias, Shrader et al. 1991)	Predicted RBS Strength
7X1	P_{ars}	AGGAAG	atgCAagcc	Glutamine (long HL)	3044
7X2	P_{cue}	AGGAAA	atgAAagcc	Lysine (short HL)	8725
7X3	P_{cue}	AGGAAA	atgGTagcc	Valine (long HL)	4171
7X4	P_{cue}	AGGAAA	atgAAagcc	Lysine (short HL)	8725
7H1	P_{cue}	AGGAGA	H6	H6	38703
7H2	Failed sequencing reaction				
7H3	P_{cue}	AGGGAA	H6	H6	19601
7H4	P_{cue}	AGGGAA	H6	H6	19601

4.3.2.4 Phenotypic characterisation of selected clones

It was expected that the clones selected were those with a higher growth rate. To confirm this, the growth of the three selected clones of Library X was compared to some clones pre-selection (Figure 4.7). Seven random white colonies were picked from a fresh assembly transformation to represent the pre-selection population (clones 01-07), and one GFP+ colony was used as negative control (clone “08”). The plasmids of post-selection clones (7X1, 7X2 and 7X3) were re-introduced into *E. coli* JM109 to rule out any genomic mutations. Washed overnight LB+Kan cultures were used to inoculate (with a dilution factor of 20) the selection medium, which was used to load triplicate wells of a 96-well microplate. Some clones were grown with the equivalent glucose medium as positive control. The cultures were incubated, and OD measurement was taken at 30 minutes intervals (Figure 4.7).

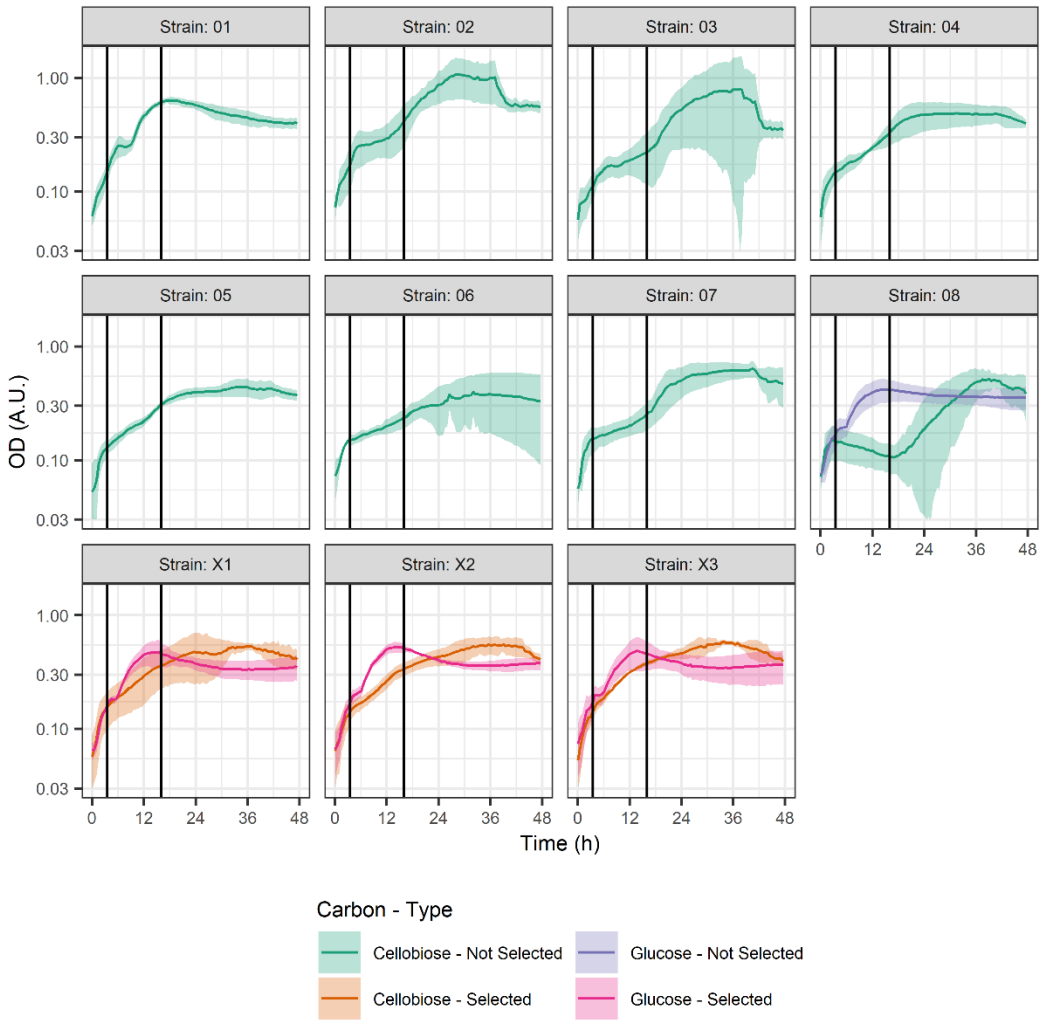


Figure 4.7. Growth curves of cellobiose-growth selection experiment clones separated by strain. Lines indicate mean of triplicate wells, coloured “ribbon” indicates SD. Vertical black lines indicate time exclusion used to calculate maximum growth rate (from 4.5 to 14.5 h). Growth rate was calculated using a minimum of 10 timepoints.

All clones displayed fast growth for the first 4 – 5 hours, then transitioning to different growth rates varying widely between clones (Figure 4.7). This initial fast growth (also displayed by the Chu2268- control clone “08” with cellobiose) could be due to the presence of 0.2 g / L of yeast extract. The inoculum culture medium was LB (which contains 5 g / L of yeast extract), which was expected to adapt cells to use the nutrients in this supplement. However, yeast extract was very dilute in M9y and any carbon sources should become exhausted very quickly, which explains the transition to different growth rates by clones with different Chu2268 expression levels.

Two growth metrics were calculated to clarify the differences between clones (Figure 4.8), due to the high variability between replicates. Due to the OD of the negative control also increasing after 18-22 hours (probably due to measurement artefacts, such as condensation on the film covering the lid) and due to the initial high growth, the growth rate calculation was limited to the period between 4.5 and 14.5 h.

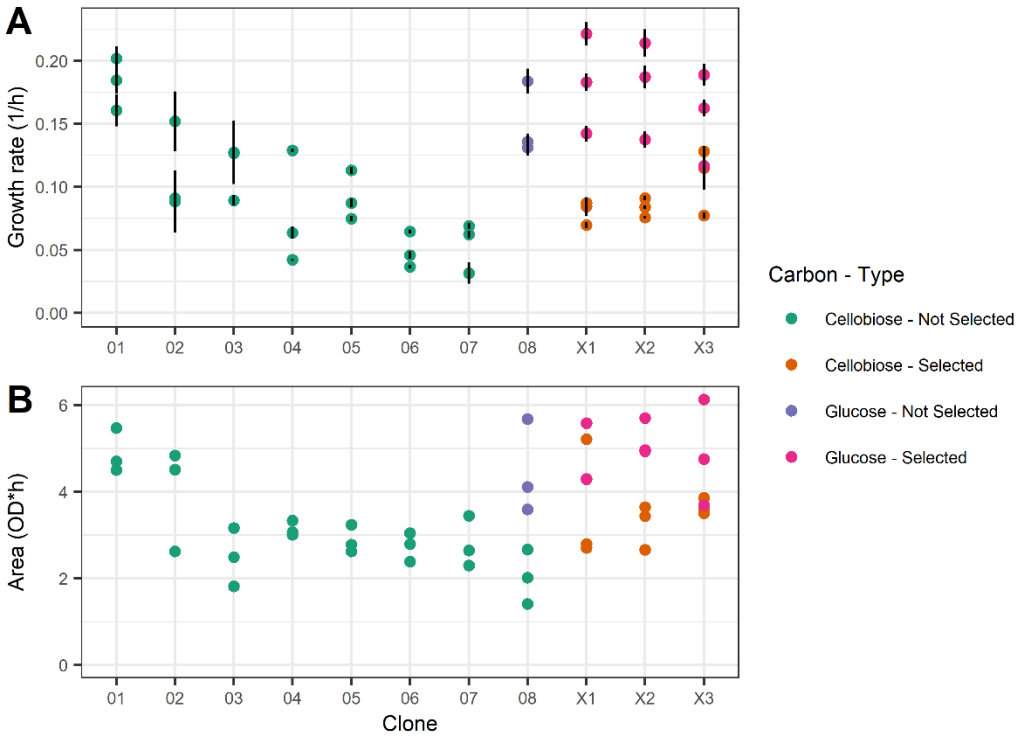


Figure 4.8. Growth characterisation of clones from Chu2268 library X selection. A: growth rates of clones. Individual points: growth rate estimation for individual sample, error bar: standard error of regression. Maximum growth rate was calculated between 4.5 and 14.5 h. In this time frame, the negative control strain with cellobiose did not show an OD increase and growth rate was not calculated. B: Area under the growth curves (AUC). Individual points indicate area under the growth curve for individual sample. Area under curve for timescale 0 to 14.5 h selected due to 14.5 being the final time used for possible growth rate calculation. Unselected clones: 01 to 07, selected clones: X1, X2, X3 (standing for 7X1, 7X2, 7X3); *chu2268*- clone: 08.

Growth rates varied highly between replicates (Figure 4.8A), reflecting the high variability observed in the curves. Despite the high variability, the growth was generally much lower for clones growing in cellobiose than any clone growing on glucose. Those growing on glucose displayed a growth rate estimated between 0.1077 and 0.1997 h^{-1}.

All clones grew more slowly on cellobiose than on glucose, except for the unselected clone “01”, which displayed a growth curve on cellobiose similar to other clones on glucose. The three selected clones still displayed a slower growth on cellobiose than in glucose, however they seemed to have an advantage over most of the unselected clones. The fact that the unselected clone #01 showed higher growth rate than the selected ones is surprising. It is not known how the expression of unselected clone “01” compares to the selected ones (unselected clones were not sequenced) and whether clones like “01” were present in the final pool. It is possible that clone “01” had a higher growth rate but was removed because of another reason, such as higher gene instability due to higher expression level.

Because of the growth rate being so variable within replicates, the area under the curve (AUC) was also considered as a metric to characterise growth (Figure 4.8B). The AUC was calculated for the first 14.5 hours because this was the upper time limit used for growth rate calculations. The AUC of growth curves still displayed high variability and showed same trend as the growth rates.

Taken together, cellulolytic phenotype selection seemed limited. Even with intracellular expression of the enzyme, multiple passages were necessary to remove the *chu2268*- clones. While it might be possible to apply selection to enzymes displayed in the cell envelope, results of this section indicate that selection pressure would be very slow (if it had any effect). Nonetheless, the selected *chu2268* assemblies should serve for later assemblies as they showed a growth slightly higher than most of the unselected clones.

4.3.3 Screening via fluorescent growth reporter

In CBP-like conditions, all sugars released from cellulose should be converted to cell biomass and product, with the latter being the variable to maximise for the final process. However, if product synthesis is initially “left aside” to first focus on optimising the cellulolytic activity, all the digested substrate will be turned into cell biomass, making microbial growth a good indicator of cellulolytic activity. If growth in CBP-like conditions is limited by the carbon source (glucose released from cellulose digestion) and glucose release is proportional to the cellulase present, growth should be proportional to cellulase activity and growth rate proportional to the cellulase per cell.

Based on the quantification of DNA in the presence of cellulose using fluorescent dyes (Duedu and French 2017), it was hypothesised that any source of fluorescence should be detectable despite cellulose turbidity. The work in this section explored the use of a fluorescent protein growth reporter to screen cellulolytic candidates.

4.3.3.1 Measuring fluorescence in cultures with cellulose

To determine whether mCherry expression could be measured in cultures using cellulose as carbon source, it was decided to replicate the experiment from Dr Salinas's thesis where *E. coli* expressing Chu2268 and CenA could grow on cellulose (Salinas 2017) in M9y-based medium. However, growth would not be determined measuring protein concentration by the Bradford assay, but through the expression of a constitutive mCherry reporter.

A constitutive *mCherry* gene (J23100_P – PET-RBS-MV_RN - mCherry_O - L3S1P51_CT) was assembled in plasmid pJUMP27-1A, resulting in plasmid pMV128. Plasmid pMV128 was co-transformed with pSB1C3-Plac-*lacZ'*-H6:Chu2268-*cenA* (the plasmid used by Dr Salinas) or with the empty pSB1C3 (negative control) in *E. coli* JM109. Starter overnight LB cultures were centrifuged and resuspended (with an initial OD of 0.1) in M9y + 1 % PASC medium with chloramphenicol, kanamycin, and 0.4 mM IPTG. The filter paper from the original experiment was replaced with PASC to allow easier sampling and measurement. As positive control, the strain with pMV128 and the empty pSB1C3 was supplemented with 2 mg / mL of commercially available cellulase mixture from *T. reesei*, while the other cultures had the same concentration of BSA. Ten-mL cultures in 50-mL centrifuge tubes, in triplicate for each condition, were incubated with shaking at 37 °C, and red fluorescence was measured every two days.

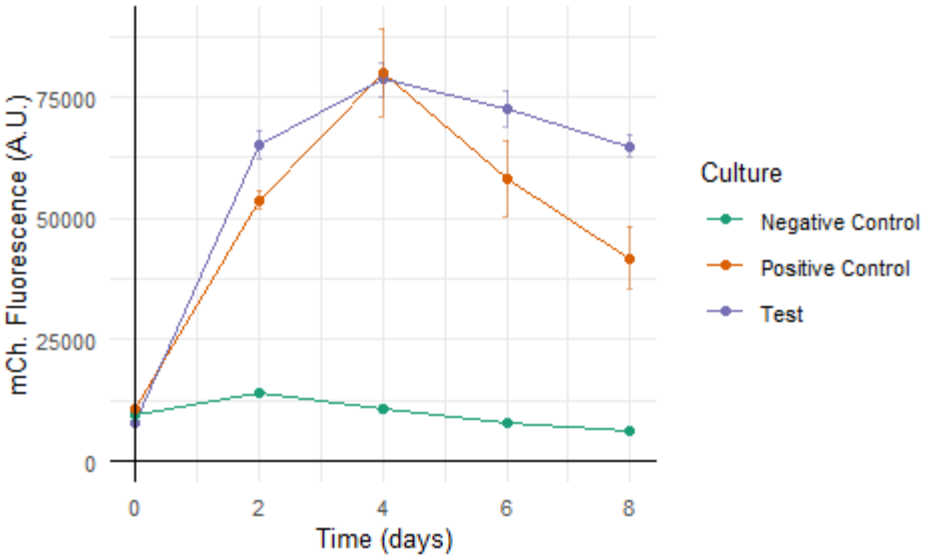


Figure 4.9. Growth of *E. coli* on cellulose carrying pSB1C3-Plac-*lacZ'*-H6:Chu2268-*cenA* and pMV128 (J23100-mCherry). Positive and negative control cultures strain contained plasmid pSB1C3 empty instead. Cultures were supplemented with 2 mg/mL of BSA (test and negative control) or cellulase from *T. reesei* (positive control). Dots indicate mean of triplicate cultures and error bars indicate SD.

Fluorescence measurements confirmed growth of *E. coli* with CenA and Chu2268 from cellulose (Figure 4.9). The signal from mCherry increased in both the test strain and the positive control during the first four days, while the negative control showed no growth over the whole experiment. These results demonstrated the cellulolytic activity from Dr Salinas's construct and the use of protein fluorescence as a means of direct growth measurement in the presence of cellulose.

4.3.3.2 Effect of cellulose on fluorescence

To accurately describe growth through protein fluorescence, it would be necessary to know if fluorescence measurements are affected by cellulose presence and in which way this occurs. An overnight *E. coli* JM109 culture expressing mCherry but no cellulases (with plasmid pJUMP29-1A: J23100_P - PET-RBS-MV_RN - mCherry_O - L3S1P51_CT, named pMV130) was resuspended in 1x M9 salts buffer, combined with different concentrations of Avicel (Figure 4.10) or PASC (Figure 4.11), and the fluorescence was measured.

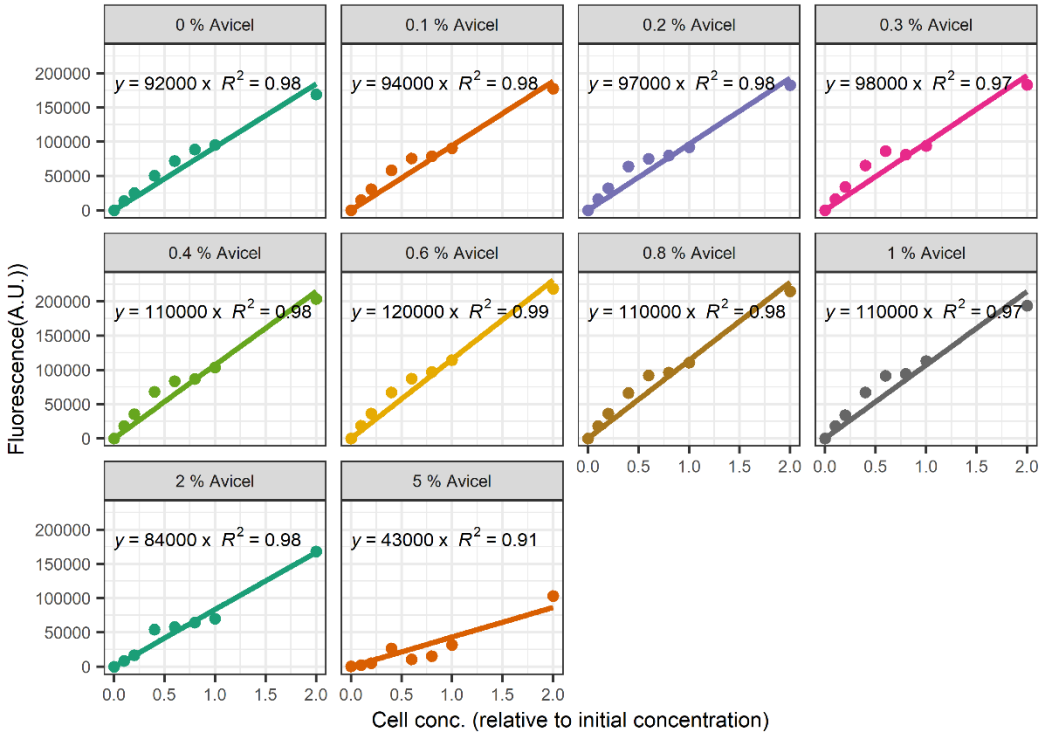


Figure 4.10. Effect of Avicel cellulose on fluorescence. Fluorescence – cell relationship for different cellulose (Avicel) concentrations. Dots indicate single measurements, line and equation indicate linear regression and parameters.

Fluorescence was always found to show a linear correlation with cell density at different concentrations of Avicel (Figure 4.10), with linear regressions showing always R^2 higher than 0.97. However, the proportionality of cell concentration to fluorescence varied with cellulose concentration. This proportionality (shown as the slope of the linear regression in Figure 4.10) decreased from 0 to 0.2% Avicel and increased from 0.2 to 0.4% and remained constant until 1%. With the highest concentrations (2 and 5%), the proportionality decreased.

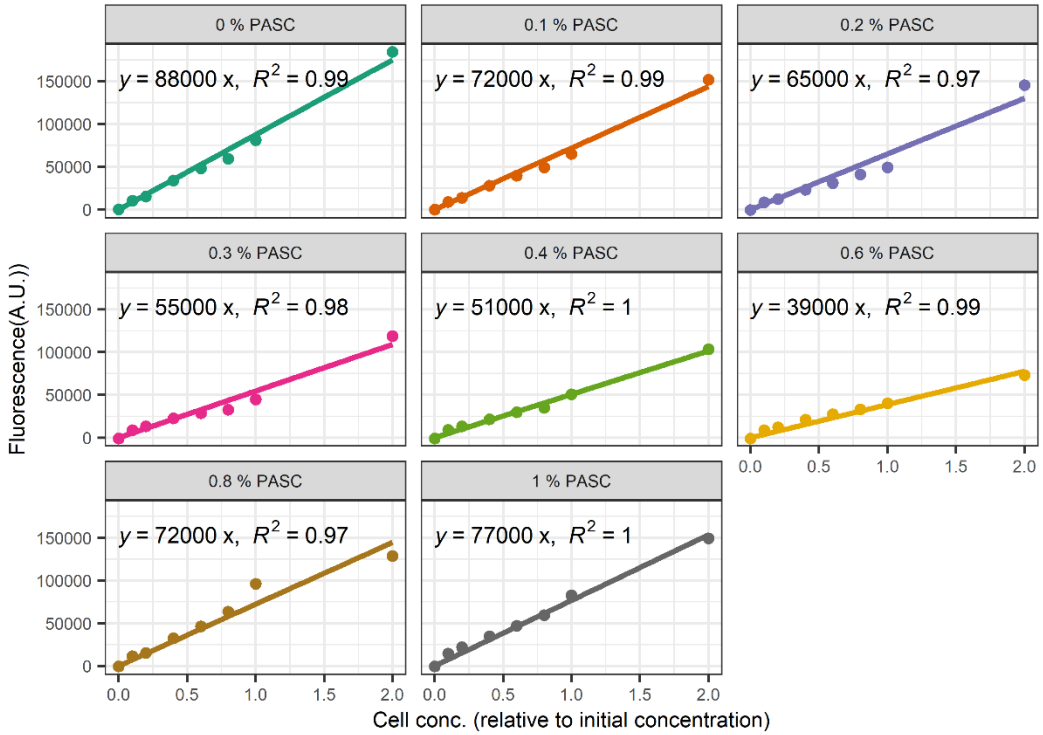


Figure 4.11. Effect of PASC cellulose on fluorescence. Fluorescence – cell relationship for different cellulose (Avicel) concentrations. Dots indicate single measurements, line and equation indicate linear regression and parameters.

PASC had a more complex effect on cellulose (Figure 4.11), with cell – cellulose proportionality being much more sensitive to cellulose concentration. The proportionality decreased from 0 to 0.6% PASC and then increased from 0.6 to 1% (Figure 4.11). Despite this, at all concentrations of cellulose fluorescence was proportional to cells ($R^2 > 0.97$). The difference in the trend between cellulose forms could be due to the difference in particle size and form, with Avicel being made of microcrystalline particles of approximately 50 µm and PASC made of much more irregularly structured amorphous cellulose.

The effect on fluorescence of a decreasing cellulose concentration in fermentation conditions deserves further examination, as it could cause an underestimation of growth. However, it should be expected that the effect should be small for the experimental conditions used in this work. Considering PASC fermentation, if a IBPM candidate achieved depolymerisation of 20% of cellulose (from an initial 1 % or 10 g / l to 8 g / L), this would cause fluorescence proportionality to drop approximately 6.5 % (considering slopes of 77000 to 72000 shown in Figure 4.11). At the same time, the 20 % depolymerisation would provide 2.2 g / L of glucose,

which is similar to concentrations (2 or 4 g / L) used to grow *E. coli* in M9 medium (Bokinsky, Peralta-Yahya et al. 2011). Based on this, it was decided to not “take into consideration” the effect of decreasing proportionality until highly cellulolytic strains are obtained.

4.3.3.3 Continuous measurement of fluorescence in microplate cultures containing cellulose.

If fluorescence could be used to measure microbial growth in the presence of cellulose to characterise IBPM candidates, it would be ideal to do in cultures incubated in 96-well microplates. Compared to shake flasks, using microplate cultures would allow an increase in the throughput of the technique by permitting a higher number of cultures and reducing the labour required for sampling.

An experiment was set up with conditions such that they would make measured fluorescence proportional to the cell concentration. If microbial growth could be characterised in such setting, it might be possible to do it with conditions closer to those expected in CBP. In complex media, different nutrients are exhausted at different times thus leading to nutritional conditions constantly changing, changing the growth rate and intra-cellular concentrations over time. Consequently, a minimal medium was used to provide a long exponential phase. Yeast extract was not added to the M9 medium in this experiment, but it was supplemented with 0.01 g / L of thiamine (medium M9t) due to *E. coli* JM109 and DH5α being auxotrophic for this vitamin.

E. coli DH5α carrying plasmid pMV130 (pJUMP29-1A: J23100_P - PET-RBS-MV_RN - mCherry_O - L3S1P51_CT) with constitutive mCherry expression was grown in starter overnight culture of M9t + 4 g / L glycerol to adapt the growth to the minimal medium and reduce the lag phase. The starter culture was used to inoculate M9t medium with different concentrations of glucose or glycerol to cause differences in growth, with and without Avicel (1 %). The cultures were loaded in a 96-well microplate in replicate wells (n=5), and incubated in the plate reader with measurements every 3 hours (Figure 4.12).

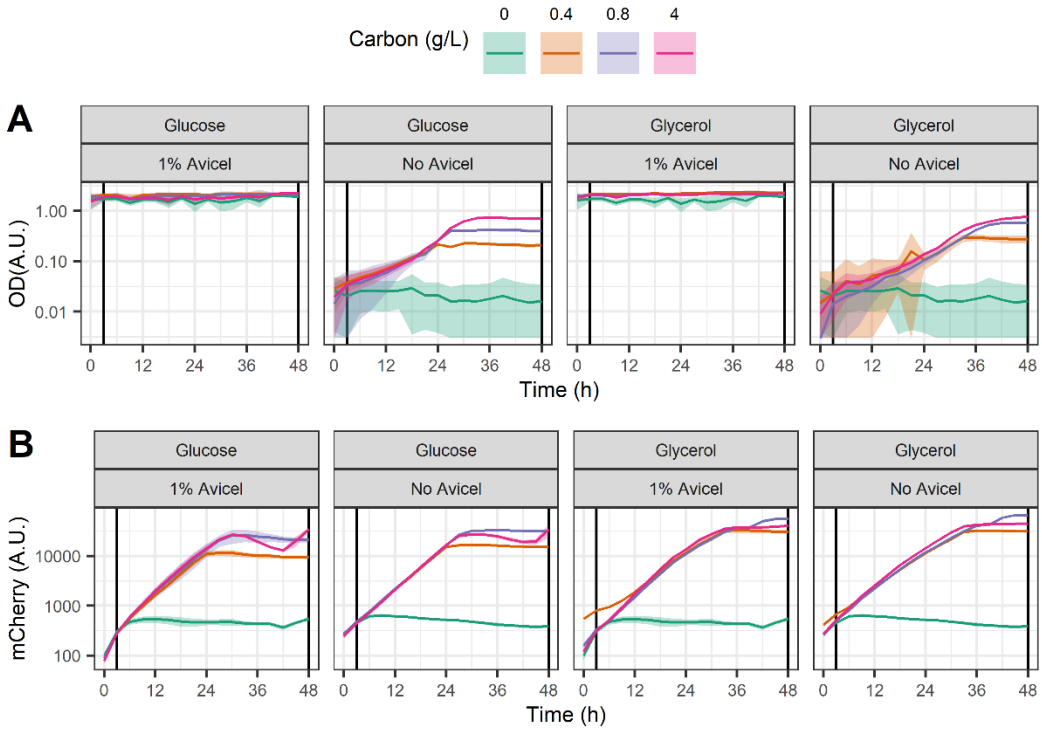


Figure 4.12. Growth curves of a strain with constitutive mCherry growing in minimal M9t medium with different concentrations of glucose or glycerol as carbon source, with and without 1% Avicel. Measurements taken at 3 h intervals. A: OD measurements, B: mCherry fluorescence measurements. Line indicates mean of replicate cultures (n=5), coloured ribbon indicates SD. Black vertical line indicates time exclusion to calculate growth rate (from 3 to 48 h timepoints).

Growth curves (Figure 4.12) show mCherry and OD development over 48 hours, with long exponential stage as originally intended. Cellulose turbidity caused OD to be very high from time 0 but, without cellulose, OD increased exponentially until 24-30 h for glucose cultures and 33-39 h for glycerol cultures, after which OD remained mostly flat. The mCherry curve showed a similar pattern, but for the highest glucose concentration fluorescence decreased temporarily after the exponential phase to then increase again. This could be due to cells displaying a diauxic growth behaviour with the high initial glucose concentration, which is described to cause changes in genetic expression in *E. coli* going from catabolising glucose to acetate (Enjalbert, Cocaign-Bousquet et al. 2015).

However, mCherry measurement had a more defined and clearer exponential phase and showed less well to well variation than OD. These differences could be to OD not being linear at higher cellular concentrations but also to the plate reader being

more precise measuring fluorescence than OD (which could be distorted by condensation in the microplate lid, for example).

Different growth metrics were used to characterise growth from OD and fluorescence measurements for each culture well separately (Figure 4.13). Maximum growth rate calculation was done limiting the time series to time points between 3 and 48 h, because between 0 and 3 h (first and second readings), fluorescence increased at a higher rate than the rest of the curve and it also increased for the 0-carbon control, which could be due to a carry-over of medium from the inoculum or measurement artefacts such as the plate increasing in temperature from room temperature to 37 °C. Growth rate was not calculated for the 0-carbon control because there was not enough increase in signal for these cultures. OD-based metrics were calculated for cultures without Avicel and fluorescence-based metrics for all cultures.

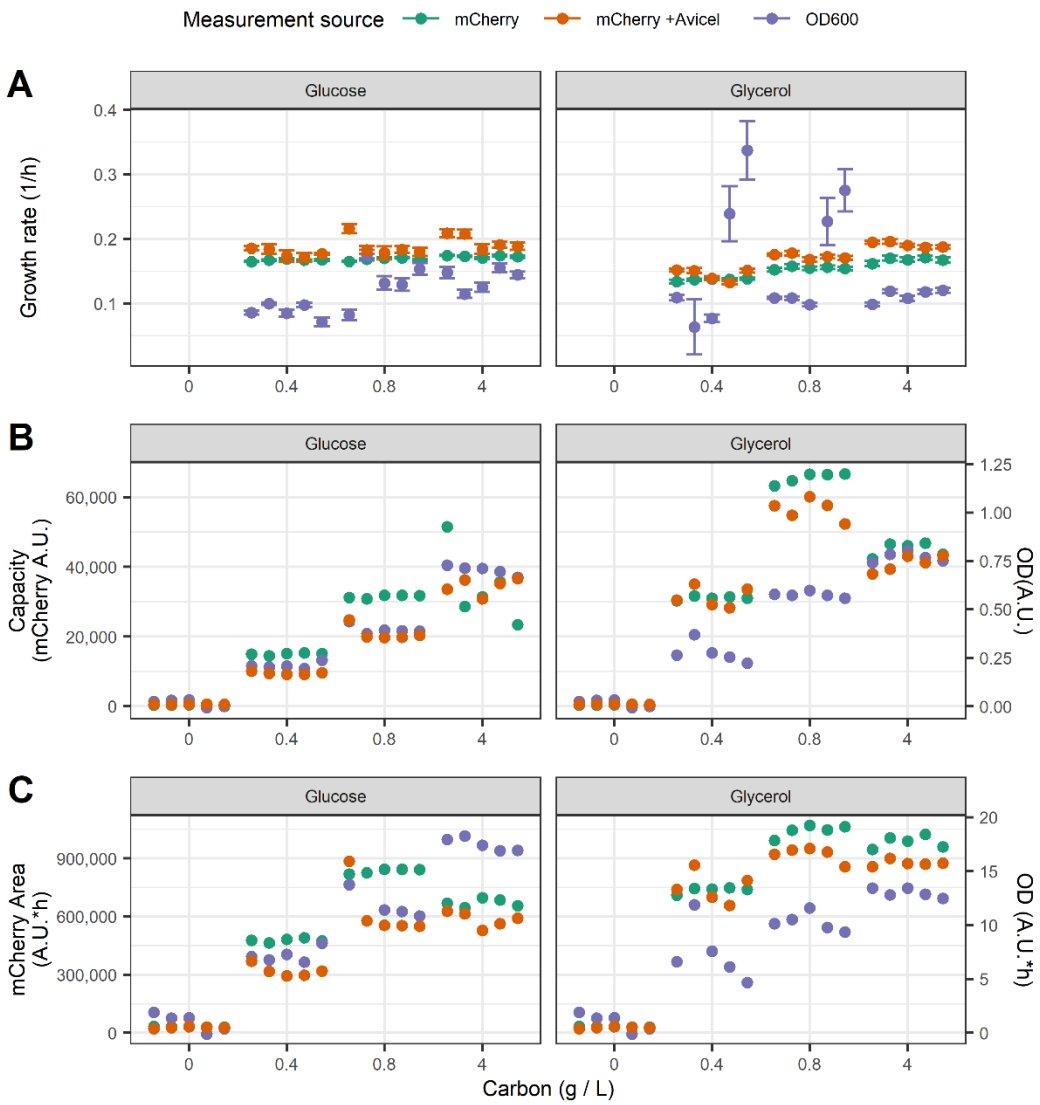


Figure 4.13. Growth metrics of a strain with constitutive mCherry growing in minimal M9t medium with different concentrations of glucose or glycerol as carbon source, with and without 1% Avicel. Single points indicate measurement for each well (n=5), colour indicate measurement source for the metric calculation (mCherry-based measurements with Avicel being done in a separate well). A: Maximum growth rate calculated after the 3 h timepoint, error bars indicate standard error of regression used to calculate growth rate. B: Capacity value calculated as signal achieved at 48 h. C: Area under the growth curve (AUC) from timepoint 0 to 48 h.

The variability of OD was reflected in the growth rate calculation, with fluorescence-based growth rate being much more similar between replicates even in the presence of Avicel. Interestingly, growth rate did not seem to be affected by the initial glucose concentration and only slightly by glycerol. Other metrics, AUC (area under the curve) and capacity (measured as maximum measurement reached by each well) diminished the variability of OD measurements, but should not be directly compared

to the values obtained based on fluorescence measurements due to the different units.

AUC and capacity showed more differences between carbon source than those found with growth rates, with an increase of both metrics with higher carbon concentrations. The exception was the metrics for cultures with 4 g / L carbon, which did not show an increase comparable to the increase in carbon available from 0.8 g / L. Possible explanation include carbon being limiting nutrient at low concentrations but not at 4 g / L and the diauxic growth reducing the amount of mCherry per cell.

All together, these results showed that growth could be characterised well with fluorescence measurements in these “ideal” conditions, showing similar trends when observing OD and mCherry fluorescence. Fluorescence readings were found to be more precise and less variable than OD readings, even in the presence of cellulose, which indicated this was a promising way of screening for cellulose candidates. However, it would be necessary to demonstrate that this observation remained true in CBP-like conditions, where the medium would not be as “defined” and strains might show slower growth.

4.4 Conclusions

Screening of cellulolytic candidates based on product quantification was found not to be a good screening method mainly due to its low throughput. Even though the ethanol assay had been adapted to a microplate format, the need for sampling, cellulose separation and reaction incubation made it too laborious to be applied on a higher scale. Additionally, the tested strain did not produce sufficient ethanol, in contrast to previous unpublished results; although it should be noted that different plasmids were used: pSB1C3-Plac-*lacZ'*-H6:Chu2268-*cenA* is a higher copy-number plasmid and provides inducible expression. Nonetheless, product maximisation will still be required for CBP optimisation and a method for quantifying the product formed by candidate strains should be useful in future when cellulolysis is much improved.

Directed evolution for the selection of candidates with increased cellulolytic capability is limited by the fact that cellulose is extracellularly digested, which allows “slackers” (clones without cellulase expression) to benefit from the performance of

the clones with the desired phenotype. The results showed that even when selecting candidates for intracellular digestion of cellodextrins, “slackers” were hard to remove from the population. This highlights the limited applicability of directed evolution to improve cellulolytic capabilities of organisms. Despite this, the clones selected in cellobiose medium for 50 generations showed a narrow range of predicted Chu2268 expression strength, which suggests that they had been selected for higher growth rate and possibly higher genetic stability, and the selected clones could be used in future work. Optimisation of other genotypes might benefit more from directed evolution, such as cellodextrin transport.

Additionally, characterisation of the clones was difficult due to OD variability in plate-based cultures. As observed in later experiments, OD plate readings were not as precise as fluorescence (unfortunately, the Chu2668 library clones lacked a fluorescent reporter). Furthermore, the complex supplementation might have distorted growth differences, especially in the first hours of culture before cellobiose becomes the limiting carbon source. Calculation of growth metrics allowed a clearer distinction between the library clones, lessening the noise caused by non-exponential growth and OD variability.

Future research could also benefit from microplates with membrane-controlled diffusion between wells, which have been used for separated co-culture of microbial species (Thøgersen, Melchiorsen et al. 2018, Jo, Bernstein et al. 2021). Using such plates for the compartmentalisation of cellulose and the IBPM candidate would enable these to grow using the insoluble substrate and direct growth measurement through OD. However, possible disadvantages of such an approach could include microbial growth limited by diffusion kinetics or throughput lower than CASFR due to the need for specialised, sometimes house-made plates.

Cellulase activity screening using fluorescent reporters (CASFR) was found to meet the requirements for a good high-throughput screening method. Fluorescence could be measured without cellulose-separation steps, and it allowed measurement of microbial growth in 96-well microplate cultures containing cellulose.

The fluorophore/fluorescence relationship remained “linear” in the presence of cellulose, but the signal-to-analyte ratio changed (decreasing at low concentrations and increasing at higher cellulose concentrations). This complex behaviour change in the apparent fluorescence was found to be small in the experimental conditions,

although this might require more attention when using strains that are more strongly cellulolytic.

This phenomenon could be caused by different scattering regimes taking place with different concentrations of cellulose. This change of regime is the reason that OD is only proportional to cell concentration at low cell densities (in a single scattering regime) but not a higher concentration (in a multiple scattering regime) (Stevenson, McVey et al. 2016). The effect of scattering on fluorescence has been studied more in detail in biomedical imaging. It has been reported that at low concentrations of scatterer apparent fluorescence intensity decreases exponentially (with increasing concentration of scatterer), but at larger concentrations light behaves diffusively and the measured fluorescence increases (with increasing concentrations of scatterer), then reaching a plateau (Müller, Georgakoudi et al. 2001). Additionally, scattering affects the fluorescence emission spectrum, causing (depending on multiple factors) possible shifts to the red (Müller, Georgakoudi et al. 2001). A finer characterisation of fluorescent cultures could take multiple fluorescence values from a single reporter to predict microbial growth and cellulose degradation. Another property of CASFR is that it represents a “global” measurement influenced by the activity of all the cellulases secreted, sugar incorporation and general microbial fitness, and consequently it could be applied to select diverse genetic modifications in a combinatorial optimisation approach.

An additional consideration for future application of CASFR is the fact that glucose might be turned into acetic acid (and lactic acid) causing pH to drop and some fluorophores are pH sensitive. The reason for using mCherry as reporter was that GFP fluorescence falls at pH 6 (Campbell and Choy 2001).

Other researchers have tried to quantify cell viability in turbid conditions (but not caused by cellulose) using fluorogenic dyes such as resazurin (Vukomanovic and Torrents 2019, Rajnovic and Mas 2020). Resazurin is converted to the green-fluorescent resorubin by cellular NADPH/NADH-dependent reductases. Using resazurin or similar dyes should allow the application of CASFR without genetic modification of the candidates, however it is important to notice that enzymatic conversion to the active resorubin is very quick and, in slow growing cultures, this might impede noticing differences between candidates growing at different speeds. In other words, in slow-growing cultures fluorescence appearance would not be

caused by cell growth but by the initial cell concentration. Sampling and dying could be used to overcome this, but at the expense of reducing potential technique throughput.

There are few instances where fluorescent proteins have been used to track microbial growth in the presence of other non-translucid substrates. The growth of *Yarrowia lipolytica* (a potential IBPM host as reviewed in Section 1.3.1) was measured using the red fluorescent RedStar2 in an emulsion of fatty acids or medium containing milk (Nicaud, Crutz-Le Coq et al. 2014, Leplat, Nicaud et al. 2015). The growth could be measured in such media in a plate reader, showing the effect on the growth of overexpressing an extracellular protease. While the authors pointed out an increased basal background signal and the need to validate conditions of growth and measurement, the effect of the scatterer on the fluorophore was not shown. Nonetheless, it was demonstrated that fluorescent growth reporters are a useful tool to compare strains.

Despite imperfections in the method, fluorescence could be measured well in microplate cultures, with a reduced variability compared to OD and the frequent measurements of cultures growing in the plate reader allowed calculation of different growth metrics, which could be useful to find differences between cellulolytic strains. The growth conditions used in the cellobiose cultures shown in Figure 4.7 were quite different to those used in the “ideal” exponential-growth experiment in Figure 4.12, with a much higher variability in the former. The growth expected with conditions closer to a CBP (with *in situ* secretion of cellulases and use of the resulting sugars) should be similar to the cellobiose cultures, because it would be preferable to maintain yeast extract to facilitate an initial cellulase production as compared to a more defined and minimum medium, and also to have similar conditions to those previously shown to allow *E. coli* to use cellulose for growth (Salinas 2017). Yeast extract might be able to provide a small amount of carbon source, but it is expected to be depleted very early and cells would depend on cellulose digestion to obtain more carbon. Despite the expected variability in the presence of yeast extract, fluorescence readings in microplate cultures should provide more precise measurements than OD and should facilitate the visualisation of growth (and growth metrics) to explore differences between IBPM candidates.

Chapter 5 A New Experimental Cycle

5.1 Abstract

Previous attempts to engineer microbial hosts to become cellulolytic were limited by analytical techniques and could just screen one or a few combinations of multiple cellulolytic genes. The use of fluorescence as growth proxy in the presence of cellulose turbidity could debottleneck analysis of microbial candidates. The work presented here tries to demonstrate the use of CASFR (cellulase activity-screening using fluorescent reporters) and set an example of screening consolidated bioprocess candidates assembled using JUMP vectors. Demonstrating CASFR requires building a benchmark strain with a minimum capacity to grow from cellulose in a detectable way in high throughput-friendly conditions and generating a collection of parts to make combinatorial libraries of cellulolytic genes. These key results were achieved, leading to an *E. coli* benchmark strain engineered to digest enough cellulose to show growth measurable in microtiter plates. These should be the pillars to build an improved research cycle to generate Ideal Biofuel-Producing Microorganisms (IBPM).

5.2 Introduction

The research cycle of synthetic biology is often limited by the screening step, as explained in the Introduction. This limitation has been especially evident for the development of CBP hosts, as reviewed in Chapter 4, because of the technical difficulties imposed by cellulose crystallinity. The CASFR (cellulase activity screening using fluorescent reporters) strategy could potentially solve this bottleneck in the development of IBPM candidates, but it remains necessary to demonstrate its use to screen cellulolytic strains in relevant conditions. Moreover, CASFR should demonstrate that it enables a systematic improvement in the research cycle compared to past attempts.

Previous efforts at engineering the IBPM have followed different approaches, but all suffer from similar limitations (Table 5.1). Generally, a set of genes encoding different enzyme types are introduced in the host and tested (as a sequential optimisation approach). Then, the best genes of each type are put together in one or a few combinations and tested. Individual enzymes, which are displayed in the cell envelope or secreted to the culture supernatant (SN), are often tested with substrate analogues to allow a higher throughput, which has the limitations described in

Chapter 4. The exception is BG enzymes, which can easily be screened since their substrate, cellobiose, is soluble, allowing growth measurements from the resulting glucose. Even using substrate analogues, few reports have tested more than a dozen candidate individual genes. An influential early report by Bokinsky, Peralta-Yahya et al. (2011) set an example of systematic screening of multiple variables. The authors tested 16 BG or xylobiosidase genes for the candidate's growth, and 10 EG fused to one secretion partner (SP) were tested for activity towards an analogue. The identified best sequences were tested with 4 – 5 different promoters, but only one combination EG+BG was built for cellulose degradation.

Table 5.1. Literature summary of systematic construction and screening of IBPM candidates. Activity testing of individual enzymes means characterisation using substrate analogues as those listed in Table 4.1.

Host (reference)	Single enzymes screened	Enzyme combinations screened
E. coli (Bokinsky, Peralta-Yahya et al. 2011)	10 EG combined with 1 secretion partner, tested for activity in SN. The best one was tested with 5 promoters. 4 BG tested for allowing growth from cellobiose. The best 2 were tested with 4 promoters. 12 xylobiosidases tested for growth from xylan. The best one was tested with 4 promoters.	1 EG+BG combination tested for growth and product formation from cellulose. 1 xylobiosidase+ 1 (previously characterised) EX combination tested for growth and product formation from hemicellulose.
B. subtilis (Chang, Anandharaj et al. 2018)	(No gene tested individually)	Two constructs built (with 8 genes of *Clostridium thermocellum* cellulosome) with different gene orders. Tested for protein expression and secretion, individual activity, and for sugar release from Napier grass.
Geobacillus thermoglucosidasius (Daas, Nijsse et al. 2018)	2 possible EG genes from the host characterised for activity (recombinant in *E. coli*) 1 of the two EG above, plus 3 EG, 3 CBH tested for CMC hydrolysis (from cell lysate).	(No combinations tested)
Lactobacillus plantarum (Stern, Moraïs et al. 2018)	2 EX and 2 EG combined with 2 secretion partners, tested for activity and for binding to cohesin.	4 consortia tested for hydrolysis of wheat straw (using concentrated SN)

Table 5.1. Literature summary of systematic construction and screening of IBPM candidates. Activity testing of individual enzymes means characterisation using substrate analogues as those listed in Table 4.1.

Host (reference)	Single enzymes screened	Enzyme combinations screened
Corynebacterium glutamicum (Anusree, Wendisch et al. 2016)	(No gene tested individually) 2 secreted EG cloned in one plasmid, 1 BG (secreted free or displayed in the cell wall) cloned in another compatible plasmid.	4 (2x2) combinations of plasmids built, expressing 1 EG and 1 BG respectively. Tested for individual enzymatic activity in the SN, tested for hydrolysis of cellulose (rice straw), and fermentation of cellobiose and CMC (measuring OD and lysine produced).
S. cerevisiae (Lee, Sung et al. 2017)	7 CBH type I, 3 CBH type II, 1 EG, and 1 BG were fused with 24 secretion partners; they were tested for secretion (SDS-PAGE) and activity.	A 4-strain consortium with the best strain expressing each of the four enzyme types was cultured and characterised for cellulosic substrates' hydrolysis. Different-ratios consortia were tested for fermentation of Avicel (in the presence of added cellulase).
S. cerevisiae (Sasaki, Mitsui et al. 2019)	Three libraries of 1 EG gene (varying integration, promoter or both) were tested by: screening 400 clones by halo formation in agar plates; 45-64 clones tested for activity in culture SN (96-well plate); best 8 clones tested for activity in culture SN (shake flask, triplicate). The same analogue, AZCL-hydroxyethyl-cellulose, used in three screening steps.	(No combinations tested)
S. cerevisiae (Fan, Zhang et al. 2016)	3 BG tested for growth from cellodextrins (presence of a transporter gene) 8 scaffoldin variants tested for GFP display. 8 EG and 8 CBH displayed in cellulosome individually tested for hydrolysis of CMC.	3 EG combined with 5 CBH and tested for ethanol fermentation from CMC, Avicel and PASC (in the presence of galactose).
S. cerevisiae (Tang, Wang et al. 2018)	1 BG and 3 CBH tested with and without dockerin domain (activity). 1 BG, 1 EG, 1 CBH with two different displays tested for activity. 1 BG, 1 EG, 1 CBH combined with 3 secretion partners, and tested for activity and immunofluorescence. To improve secretion, 3 secretion-related genes were co-expressed with BG or CBH genes, and activity was tested.	3 strain consortia (before improving secretion and after, and with improved ratios) tested for ethanol fermentation from PASC (in rich medium).

Table 5.1. Literature summary of systematic construction and screening of IBPM candidates. Activity testing of individual enzymes means characterisation using substrate analogues as those listed in Table 4.1.

Host (reference)	Single enzymes screened	Enzyme combinations screened
S. cerevisiae (Davison, den Haan et al. 2016)	30 strains tested for the secretion of 1 BG gene (for activity). 7 best strains additionally tested for secreted activity of 1 CBH or 1 EG (separately). Also tested the effect of various stresses on growth rate (such as growth inhibitors).	(No combinations tested)
S. cerevisiae (Claes, Deparis et al. 2020)	In a xylose-using strain, 1 BX and 1 BG were tested for SN activity and co-fermentation of glucose, xylose, cellobiose and xylan. 3 EX, 4 bifunctional acetylxylan esterases/EX, 6 CBH type I, 6 CBH type II, and 6 EG were tested for SN activity.	The best gene of each type was integrated sequentially and confirmed for activity. Strains were tested for: co-fermentation of cellobiose, glucose, xylose and xylan; and for hydrolytic activity (of culture SN) of PASC, FP, CMC and xylan.
S. cerevisiae (Wightman, Kroukamp et al. 2020)	(No gene tested individually)	Library of clones with different gene copy numbers of 1 EG and 1 BG were generated and tested for individual activity and synergistic activity (using substrate analogues).
P. pastoris (Kickenweiz, Glieder et al. 2018)	2 BG tested for growth from cellobiose.	3 strains with different combinations of BG, CBH and EG genes were tested for growth on CMC, and SN tested for hydrolysis of Avicel.
Y. lipolytica (Wei, Wang et al. 2014)	1 EG, 1 CBH type II and 4 CBH type I tested for secretion (SDS-PAGE and western blot), then for growth from CMC. 1 of the CBH I genes was also purified and tested for hydrolysis of Avicel (in the presence of added BG and EG).	1 consortium (1 EG + CBH II) tested for clearing of Avicel in rich-medium agar plates, and SN tested for PASC hydrolysis. 1 consortium (1 EG + 1 CBH I + 1 CBH II) compared with the previous consortium for Avicel fermentation (measuring growth and fatty acid produced).
Y. lipolytica (Guo, Duquesne et al. 2017)	2 EG, 2 CBH type I, and 1 CBH type II tested for secretion (SDS-PAGE and western blot) and purified enzyme activity (using a his-tag). 1 CBH I and 1 CBH II genes with a different promoter tested for secretion (SDS-PAGE)	6 strains (with different combinations of the 2 EG and 4 CBH genes, and always 2 BG) were tested for hydrolytic activity in the SN towards Avicel, PASC, and CMC, and for growth from Avicel.

Table 5.1. Literature summary of systematic construction and screening of IBPM candidates. Activity testing of individual enzymes means characterisation using substrate analogues as those listed in Table 4.1.

Host (reference)	Single enzymes screened	Enzyme combinations screened
Y. lipolytica (Guo, Duquesne et al. 2017)	1 LPMO purified and tested for activity alone and with added cellulases. 1 swollenin purified and tested with added cellulases (it was found that pre-incubation with swollenin was more effective than expression during saccharification)	Work continued with the strain from the previous report (see above). Authors changed the promoter of 1 EG; then they introduced a swollenin, an EX, an LPMO, or the LPMO and the EX. The five new strains were compared to the parental one in SN hydrolytic activity towards CMC, PASC, Avicel and industrial cellulose pulp, and compared in growth from Avicel.

In the reviewed studies, just a few combinations of cellulase genes in one host are tested. Consequently, emerging interactions due to context-dependency are ignored, and this sequential optimisation is only finding local activity maxima when a much-improved capacity is required. The best example of non-sequential optimisation is the recent publication of Wightman, Kroukamp et al. (2020), which shows other limitations. One hundred and sixty clones of a strain library with different copy numbers of an EG gene and BG gene were screened using a substrate analogue (4,6-O-(3-ketobutylidene)-4-nitrophenyl-β-D-cellopentaoside) that must be cleaved by both enzymes to generate the signal. However, it is unlikely that the BG/EG ratio that best cleaves a soluble 5-glucose substrate is the optimal ratio required for digesting more realistic cellulosic substrates, and this method can not be applied to combinations including other enzymes.

Few combinations of various secretion partners with enzyme-coding sequences have been examined. The report with most enzyme-secretion partner combinations explored was presented by Lee, Sung et al. (2017), where 11 enzyme sequences for different enzymes were screened with 24 secretion partners. The authors avoided constructing a strain expressing multiple genes by combining single enzyme-secreting strains as consortia, which they tried to optimise quantitatively using different ratios of the strains in the inoculum, but also noting that this ratio varied during fermentation.

An ideal research cycle strategy should build and test multiple combinations of cellulolytic genes in CBP-like conditions: an insoluble cellulose substrate is digested

by enzymes secreted *in situ,* and the resulting sugars are used by the microorganism. Multiple variables of each gene are relevant and must be included in the optimisation process: coding sequence of the mature enzyme, secretion partner, regulatory elements. Unfortunately, testing all these variables for the different genes becomes an intractable problem due to a combinatorial explosion. Therefore, an initial screen of individual genes is necessary to select the promising genes among all the possible combinations. Unlike in previous studies, this initial selection should be made in CBP-relevant conditions too: individual genes should be tested with natural substrates which require the presence of additional cellulases to be digested.

The use of JUMP along CASFR fulfils these requirements. JUMP secondary sites enable introduction of additional genes in the secondary site of a vector before a sequence of interest (i.e. the gene to be optimised) is assembled in the vector; and CASFR should potentially allow assessment of the use of sugars by the IBPM candidates regardless of the type of genes being tested. A benchmark set of genes known to provide a cellulolytic activity sufficient for growth —such as *cenA* and *chu2268* for *E. coli* (Section 1.3.1)— could be introduced in secondary sites along with other required genes, as the fluorescent growth reporter. This construct would allow testing of any gene of interest in the main site, for example, LPMO, expansin, CBH, or combinatorial libraries of these with different secretion partners. Additionally, the assembled gene of interest (and not the additional genes in the secondary site) could be combined with other relevant genes. In other words, JUMP design could allow a size reduction of combinatorial libraries in a level 1 assembly, and selected genes from different classes could be directly assembled in a combinatorial way in a level 2 construct and further screened. This ideal research cycle would require a benchmark cellulase host, a collection of parts, and CASFR to be demonstrated as a selection system.

Unpublished work of Dr Chao-Kuo Liu and Prof. Chris French offers a collection of cellulolytic proteins that were functional in *E. coli*. This extensive work started with the activity characterisation of 108 candidate genes identified from seven bacterial species (*Actinoplanes missouriensis* 431, *Cytophaga hutchinsonii* ATCC 33406, *Cellulomonas fimi* ATCC484, *Micromonospora lupine* lupac 08, *Sorangium cellulosum* so ce56 and *Teredinibacter turnerae* T7901) and two fungal species (*Neurospora crassa* OR74A and *Podospora anserine*). All the genes were

expressed in *E. coli,* and the activity of individual enzymes in cell lysates was tested against several substrate analogues, identifying the ones with the most activity. Then, 11 EG and 8 cellodextrinases were combined and tested for sugar release from cellulose. The most promising combinations were co-expressed in single strains, demonstrating partial hydrolysis of cellulose (filter paper) present in cultures. Additionally, LPMO enzymes were tested in the presence of added cellulases showing increased sugar release for some of the candidate enzymes. While growth or sugar utilisation was not measured, this foundational work demonstrated the activity from multiple enzymes.

In addition to the enzyme-coding sequence itself, multiple secretion partners should be examined for each enzyme. *E. coli* has been used extensively for recombinant protein production, but this bacterium is not generally considered a good protein secretor because it has an outer membrane. However, there are plenty of examples demonstrating good production of secreted recombinant proteins (Kleiner-Grote, Risse et al. 2018) where the peptide of interest is secreted due to the host cellular secretion machinery recognising the secretion signal peptide (SP). In some cases, the SP leads to the periplasmic space; in other cases, secretion to the extracellular medium occurs through known or unknown mechanisms. The signal peptide can be native to the protein of interest, or it can be a taken from a different gene and fused to the protein cargo to be secreted. In some cases, the protein of interest is fused to a whole protein known to be secreted to the extracellular medium by *E. coli*.

Different known mechanisms translocate proteins through the inner or outer membrane of bacteria (Green and Mecsas 2016, Burdette, Leach et al. 2018, Kleiner-Grote, Risse et al. 2018). The most successfully used for secretion of recombinant proteins are the Tat and Sec routes, found in all domains of life that, in *E. coli,* transport proteins into the periplasmic space. Both these pathways use N-terminal secretion signal peptides (SP) to recognise which proteins to secrete. The Tat pathway translocates proteins after folding in the cytoplasm, whereas the Sec route translocates proteins before folding with the SecYEG channel. The Sec route is often sub-divided depending on whether the protein is translocated post-translationally (SecA/SecB route, such as with *E. coli* proteins DsbA, TorT, TolB) or co-translationally (Signal Recognition Particle or SRP route, such as with MalE or maltose-binding protein) (Lee, Lee et al. 2016). The name of the Tat route comes from the twin-arginine motif in the signal peptide in proteins transported through this

route (twin-arginine translocation). Sec signal peptides are generally short (~20 residues) and contain a similar structure: an N-terminal region with positively charged residues, a central hydrophobic region recognised by either SRP or SecA, and a mainly polar C-terminal region that contains the signal cleavage motif.

Many secretion signal peptides are known to work with recombinant proteins in *E. coli*. These SP include some from proteins naturally expressed by *E. coli*, such as OsmY (Bokinsky, Peralta-Yahya et al. 2011), SfmC and PhoA (Huber, Boyd et al. 2005), OmpT (Kurokawa, Yanagi et al. 2001), DsbA, TorT and TolB (Lee, Lee et al. 2016), YebF (Natarajan, Haitjema et al. 2017), LamB (Pratap and Dikshit 1998), OmpA (Schlegel, Rujas et al. 2013), Spy (Tsirigotaki, De Geyter et al. 2017), OmpC (Wang, Ma et al. 2016), MalE (Zalucki and Jennings 2007), and the bacteriophage pIII protein (Sroga and Dordick 2002). Fusion to OmpF led to an impressive 5.6 g/L accumulation of the recombinant protein in the culture supernatant (Jeong and Lee 2002). Due to this secretion system's universality, heterologous SP have been found to lead to good periplasmic or extracellular secretion in *E. coli*, including the widely used PelB from *Erwinia carotovora* (Wang, Ma et al. 2016). Other heterologous SP include the signal of Vrg-6 from *Bordetella pertussis* (Knapp, Amann et al. 1996), the KP signal peptide from *Kocuria* sp 3-3 (Cui, Meng et al. 2017), and multiple SP from *Bacillus lehensis* (Jonet, Mahadi et al. 2012, Ling, Rahmat et al. 2017). A cellulase catalytic domain (“Cel-CD”) isolated from *Bacillus* sp. was found to be secreted without its native signal peptide, and it was later found that the first 20 amino acids of the domain acted a signal peptide (Gao, Luan et al. 2015, Gao, Wang et al. 2015, Gao, Luan et al. 2016). In other cases, SP modifications have led to improvements too (Jonet, Mahadi et al. 2012, Kim, Min et al. 2015).

Interestingly, the codon sequence of signal peptides strongly affects the translocation through Sec or Tat. It was observed that *E. col* secreted proteins contain more rare codons in the signal peptide than the genomic average (Power, Jones et al. 2004), and it was proved that replacing the SP codons for the most “optimal” ones (those most used by an organism) reduced secretion efficiency (Zalucki and Jennings 2007). It is believed that the reduction in translation speed caused by non-optimal codons facilitates the recognition of the signal. Many other factors (such as folding kinetics, and protein and RNA stability) affect the secretion and make it impossible to predict whether a SP will lead to high secretion of the

target protein in a specific host (Freudl 2018). In other words, an SP that is good for one cargo protein might be mediocre for another one and vice-versa.

The Tat and Sec routes only lead the protein cargo to the periplasmic space. However, in many of the mentioned reports, the protein was confirmed to be secreted to the extracellular medium, which could be caused by known and unknown mechanisms. Leaks through the outer membrane can occur and might be facilitated by surfactant additives (Liu, Tian et al. 2014) or specific mutations (Burdette, Leach et al. 2018). Another mechanism that could be involved is the formation of outer membrane vesicles, known to have a role in *E. coli* during pathogenesis (Salvachúa, Werner et al. 2020) and in *P. putida* for degradation of biomass-derived compounds (Salvachúa, Werner et al. 2020). Some proteins have a C-terminal autotransporter (AT, or Type 5 Secretion System) domain which tunnels the protein cargo through the outer membrane. The AT domain stays membrane-attached with the cargo protein being displayed in the cell envelope, but, in some cases, the cargo is released by a cleavage in the peptide that links the cargo to the AT domain. Examples of AT domains applied to extracellular secretion of recombinant proteins include Ag43 (Wargacki, Leonard et al. 2012), Pet and Pic from *E. coli* (Sevastsyanovich, Leyton et al. 2012).

5.2.1 Aims

This chapter's objective is to demonstrate the use of CASFR to screen IBPM candidates using *E. coli* as host and JUMP as platform to allow multi-gene combinatorial plasmid assemblies.

Meeting this objective requires collecting standardised parts coding for SP and enzymes active in *E. coli* so a benchmark cellulolytic construct can be built using JUMP vectors. This construct will be used to simplify the construction of multi-cellulolytic gene plasmids and their screening in CBP-like conditions (with cellulose as sole carbon source). The resulting strain will be used to find the best conditions to screen candidates using CASFR, aiming at cellulose-consuming cultures in 96-well microplates (or “microplate cultures”), as this will allow higher throughput in future experiments. These cultures should demonstrate whether fluorescence indicates growth in CBP-like conditions, allowing screening of candidates and giving a signal proportional to cellulose hydrolysis.

The work of this chapter will be based on the previous experiment that showed growth from cellulose as sole carbon source in Chapter 4 (Figure 4.9 in Section 4.3.3.1). The medium in that experiment was M9 minimal medium supplemented with 0.2 g / L yeast extract (M9y). The dilute supplementation should allow *E. coli* to secrete an initial amount of cellulase to jump-start growth before using cellulose as sole carbon source, and it was decided to use the same medium. In that experiment, the tested culture also contained BSA, which was included as a replacement to the cellulase added in the positive control culture. However, BSA likely had a beneficial effect on the cellulase activity, such as stabilising the enzyme (Kadhum, Rajendran et al. 2018). Consequently, BSA was also included in the culture media (unless otherwise specified).

5.3 Results and Discussion

5.3.1 A toolbox for building IBPM candidates

5.3.1.1 Cellulolytic-enzyme coding sequences

The first cellulases domesticated in JUMP format were the two genes used by Dr Alejandro Salinas demonstrating *E. coli* growth from cellulose as sole carbon source, including EG CenA with codon-optimised sequence and cellodextrinase Chu2268 (sequences in Appendix II and part details in Table 5.2).

Then, eight additional enzymes with activity confirmed by Dr Liu were selected and domesticated as part type O, to allow fusion to different SP parts (sequences shown in Appendix III). The gene nomenclature used here follows the style used by Dr Liu, with further genetic information listed in Table 5.2. EG cellulases ML12A and AM6B were domesticated via PCR using codon-optimised gene sequences, generating parts ML12Ao_O and AM6Bo_O. The potential cellodextrinases SC1B and TT3D (with activity against both MUC and MUG, and hence expected to release glucose from cellooligosaccharides) and the expansin *yoaJ* were PCR-domesticated from wildtype sequences generating SC1B_O, TT3D_O and YoaJ_O. Furthermore, some enzymes were domesticated using synthesised DNA due to their higher number of internal forbidden sites. Due to their high GC content, they were codon-optimised to allow correct synthesis, as per manufacturer requirements (Twist Bio). Synthesised enzyme parts included the CBH-xylanase Cex, the cellodextrinase SC1C, and the LPMOs AMAE, AMAG and CFAA, generating Cexo_O, SC1Co_O, AMAEo_O, AMAGo_O and CFAAo_O, respectively.

Table 5.2. List of cellulolytic protein-coding parts built. Changes are indicated with the 1-letter nomenclature of amino acids.

Part	Activity type	Organism	Gene Id	Changes
CenAo_NOC	EG	*Cellulomonas fimi* ATCC484	CP002666.1, locus Celf_3184	1 internal site removed.
CenAo_O	EG	*Cellulomonas fimi* ATCC484	CP002666.1, locus Celf_3184	Added 3'GS, 1 internal sites removed.
Chu2268_O	Cellodextrinase	*Cytophaga hutchinsonii* ATCC 33406	CP000383.1, locus CHU_2268	Added 5' A, 3'GS, 2 internal sites removed.
ML12Ao_O	EG	*Micromonospora lupine* lupac 08	NZ_HF570108.1, locus MILUP08_RS07045	Added 3' S
AM6Bo_O	EG	*Actinoplanes missouriensis* 431	NC_017093.1, locus AMIS_RS19855	Added 3' S
SC1B_O	Cellodextrinase	*Sorangium cellulosum* So ce 56	NC_010162.1, locus sce3455	Added 3' S
TT3D_O	Cellodextrinase	*Teredinibacter turnerae* T7901	NC_012997.1, locus TERTU_RS18160	Added 5' A, 3'GS, 3 internal sites removed.
YoaJ_O	Expansin	*B.subtilis* 168	CP053102.1, locus HIR77_10095	Added 3' GS
Cex_O	CBH, xylanase	*Cellulomonas fimi* ATCC484	CP002666.1, locus Celf_1271	Added 3' S
SC1C_O	Cellodextrinase	*Sorangium cellulosum* So ce 56	NC_010162.1, locus sce3455	Added 3' S
AMAE_O	LPMO	*Actinoplanes missouriensis* 431	NC_017093.1, locus AMIS_RS25835	Added 3' S
AMAG_O	LPMO	*Actinoplanes missouriensis* 431	NC_017093.1, locus AMIS_RS39000	Added 3' GS
CFAA_O	LPMO	*Cellulomonas fimi* ATCC 484	CP002666.1, locus Celf_0270	Added 3' GS

After domestication, part YoaJ_O was found to have an incorrectly designed 5' fusion site and would require re-domestication before it could be used.

Dr Salinas's cellulolytic construct used the wildtype codon sequence of *cenA*. Consequently, a domestication of this part was also attempted using Dr Salinas' plasmid pSB1C3-Plac-*lacZ'*-H6:Chu2268-*cenA*, as PCR template to remove two internal forbidden sites. However, the domestication was unsuccessful. Only five colonies with the potential part appeared from the domestication assembly

transformation, and they all contained either sequence deletions or point mutations changing the amino acid sequence. The appearance of this mutations could be due to the high GC content of the sequence (72%) which could cause the PCR to generate incorrect fragments or the sequence to be less stable *in vivo*, either in the template plasmid or in the JUMP part plasmid. The construction of CenAw_O was abandoned due to the difficulty found and part redundancy (with codon-optimised variant and other EG-coding parts).

5.3.1.2 Secretion parts

Based on the reported recombinant protein secretion in *E. coli*, a collection of SP parts was built. Part sequences appear in Appendix IV.

Several secretion parts were domesticated via PCR or DNA synthesis and cloned in pJUMP18-Uac for sequence confirmation and replication of the part. Further details of the parts appear in Table 5.3, including genetic and literature references. Three C-ter autotransporter (AT) domains were domesticated along with the N-ter SP that had been used with each of them. Each N part – C part couple was cloned together in pJUMP18-Uac, generating vectors containing two parts that should fuse to the N- and C-ter ends of a part type O during a level 1 assembly. This resulted in pic_Ndom_Cdom, pet_Ndom_Cdom, and ag43_Ndom_Cdom. The sequence encoding Cel-CD was obtained via DNA synthesis and cloned in pJUMP18-Uac, but, unlike in the report using this sequence for secretion, this part included the native SP (part CelCD_Ndom). This sequence served as a template to clone additional parts via PCR: the 29 amino acid signal peptide of the native gene (Cel-CD-N29_N) and the sequences that had shown secretion in the original report: the cellulolytic domain without the native signal peptide (Cel-CD-Gdom_N) and the first 20 amino acids of the cellulolytic domain (Cel-CD-G20_N). Similarly, XynA_Ndom was synthesised as the whole endoxylanase gene sequence and was used as template to domesticate the 29-amino acid SP, which had been reported to be secreted (XynA-SP_N). Later on, it was found that the synthesised parts CelCD_Ndom and XynA_Ndom had an incorrectly designed 3' fusion site and would require re-domestication before they could be used. However, the parts domesticated using these as template were correct. Parts encoding PelB and KP-SP were originally built by oligo annealing as short parts following the EcoFlex standard (Moore, Lai et al. 2016), which were later used as templates to clone JUMP parts PelB_N and PelB-D5_N, and KP-SP_N, respectively. Parts YebF_N,

Spy_N and OsmY code for the whole proteins as N-ter fusions, and were generated by PCR using *E. col* MG1655 genome as template.

The SP of the cellulolytic genes domesticated as O parts (Section 5.3.1.1) were domesticated as N parts. Parts ML12Aw-SP_N and AM6Bw-SP_N were generated by PCR from the wildtype codon sequences of these two genes and were cloned in pJUMP18-Uac (Table 5.3). SP of CenA was PCR-domesticated from the codon-optimised sequence, generating CenAo-SP_N. PCR-domestication of the SP from the wildtype codon sequence of CenA was attempted without success, but domestication from annealed oligonucleotides worked, generating CenAw-SP_N.

Some of the shorter parts were obtained by DNA synthesis and were not cloned into any part-acceptor vector. Further details of the parts appear in Table 5.4, including genetic and literature references. The list of “linear” SP includes those from some of the genes domesticated as O parts: Cex-SP_N, TT3D-SP_N, SC1B-SP_N, SC1C-SP_N, YoaJ-SP_N, AMAE-SP_N, AMAG-SP_N and CFAA-SP_N. Additional linear parts included 11 SP from *E. coli,* 15 SP from *Bacillus lehensis*, Vrg-6_N from *Bordetella pertussis*, and pIII_N from page M13 (Table 5.4).

Table 5.3. Secretion parts cloned in pJUMP18-Uac. The changes to the amino acid sequence with respect to reported sequences are listed (1-letter code), but silent mutations to introduce standard fusion sites are omitted.

#	Part name	Organism source	GeneBank or Gene Id	Literature reference	Changes with respect to original sequence
SP01	pic_Ndom_Cdom	*E. coli* 042	AF056581.1	(Sevastsyanovich, Leyton et al. 2012)	Nter: initial 53 residues; Cter: S + final 286 residues
SP02	pet_Ndom_Cdom	*E. coli* 042	AF097644.1	(Sevastsyanovich, Leyton et al. 2012)	Nter: initial 52 residues, added A; Cter:final 286 residues
SP03	CelCD_Ndom	*Bacillus* sp.	M15743.1	(Gao, Luan et al. 2015, Gao, Wang et al. 2015, Gao, Luan et al. 2016)	Includes SP of native gene and catalytic domain (407 residues). BsaI site removed. *3' end overhang is faulty, 1 bp is missing right before, changing reading frame downstream.
SP04	XynA_Ndom	*Bacillus amyloliquefaciens* strain YB-1402	KX620015.1	(Jeong and Lee 2000)	29 residues signal peptide + 552 residues mature peptide. *3' end overhang is faulty, 1 bp is missing right before, changing reading frame downstream.
SP05	CenAo-SP_N	*Cellulomonas fimi* ATCC484	CP002666.1, locus Celf_3184	Liu unpublished work	31 residues from codon-optimised gene.
SP06	CenAw-SP_N31	*Cellulomonas fimi* ATCC484	CP002666.1, locus Celf_3184	Liu unpublished work	31 residues from wildtype codon gene.
SP07	PelB_N	*Erwinia carotovora*	Plasmid pET27b	(Wang, Ma et al. 2016)	22 residues as in pET27b. Added 3' A
SP08	PelB Asp5_N	*Erwinia carotovora*, modified	Plasmid pET27b, modified	(Kim, Min et al. 2015)	22 residues as in pet27b. Added 3' DDDDDGA
SP09	KP-SP_N	*Kocuria* sp. 3-3	KF277252.1	(Cui, Meng et al. 2017)	31 residues from gene.
SP10	XynA-SP_N	*Bacillus amyloliquefaciens* strain YB-1402	KX620015.1	(Jeong and Lee 2000)	29 residues from gene.

Table 5.3. Secretion parts cloned in pJUMP18-Uac. The changes to the amino acid sequence with respect to reported sequences are listed (1-letter code), but silent mutations to introduce standard fusion sites are omitted.

#	Part name	Organism source	GeneBank or Gene Id	Literature reference	Changes with respect to original sequence
SP11	Cel-CD-N29_N	*Bacillus* sp.	M15743.1	(Gao, Luan et al. 2015, Gao, Wang et al. 2015, Gao, Luan et al. 2016)	29 residue signal peptide of original gene
SP12	Cel-CD-G20_N	*Bacillus* sp.	M15743.1	(Gao, Luan et al. 2015, Gao, Wang et al. 2015, Gao, Luan et al. 2016)	Peptide used by authors: M+20 residues from catalytic domain + KRGA
SP13	Cel-CD-Gdom_N	*Bacillus* sp.	M15743.1	(Gao, Luan et al. 2015, Gao, Wang et al. 2015, Gao, Luan et al. 2016)	Domain as used by authors (M + 375 from catalytic domain), +GGA
SP14	ag43_Ndom_Cdom	*E. coli* MG1655	946540	(Wargacki, Leonard et al. 2012)	51 n-ter residues, 600 c-ter residues of the recombinant construct in reference.
SP51	ML12Aw-SP_N	*Micromonospora lupine* lupac 08	NZ_HF570108.1, locus MILUP08_RS07045	Liu unpublished work	No changes from predicted sequence.
SP52	AM6Bw-SP_N	*Actinoplanes missouriensis* 431	NC_017093.1, locus AMIS_RS19855	Liu unpublished work	No changes from predicted sequence.
SP53	OsmY_N	*E. coli* MG1655	948895	(Bokinsky, Peralta-Yahya et al. 2011)	Complete gene (201 residues) + GSGSA
SP54	YebF_N	*E. coli* MG1655	946363	(Natarajan, Haitjema et al. 2017)	Complete gene (118 residues) + SA
SP55	Spy_N	*E. coli* MG1655	946253	(Tsirigotaki, De Geyter et al. 2017)	Complete gene (161 residues) + SA

Table 5.4. Secretion parts not cloned in vector format. The changes to the amino acid sequence respect reported sequences are listed, but silent mutations to introduce standard fusion sites are omitted.

#	Part name	Organism source	GeneBank or Gene Id	Literature reference	Changes respect original sequence
SP15	G1_N	*B. lehensis*	AIC96492	(Jonet, Mahadi et al. 2012)	Changed start to ATG codon.
SP16	G1-M5_N	*B. lehensis*	AIC96492, modified	(Jonet, Mahadi et al. 2012)	No changes from reported sequence.
SP17	GAP_N	*B. lehensis*	AIC95922	(Ling, Rahmat et al. 2017)	Changed start to ATG codon.
SP18	CSP_N	*B. lehensis*	AIC95945	(Ling, Rahmat et al. 2017)	Internal BbsI sites removed. Finishes SKA; Added A.
SP19	Ble1853_N	*B. lehensis*	AIC94431	(Ling, Rahmat et al. 2017)	No changes from reported sequence.
SP20	Ble1848_N	*B. lehensis*	AIC94426	(Ling, Rahmat et al. 2017)	Contains SpeI site.
SP21	ECDS_N	*B. lehensis*	AIC92706	(Ling, Rahmat et al. 2017)	Changed start to ATG codon. Finishes VYA; Added A.
SP22	LytE_N	*B. lehensis*	AIC96238	(Ling, Rahmat et al. 2017)	Changed start to ATG codon.
SP23	Chit_N	*B. lehensis*	AIC93540	(Ling, Rahmat et al. 2017)	Finishes LWA, Added A.
SP24	BGL_N	*B. lehensis*	AIC93282	(Ling, Rahmat et al. 2017)	Finishes AGA, Changed G codon.
SP25	BGL(4)_N	*B. lehensis*	AIC96475	(Ling, Rahmat et al. 2017)	Finishes ANQ, Added A.
SP26	MEP_N	*B. lehensis*	AIC95833	(Ling, Rahmat et al. 2017)	Changed start to ATG codon.
SP27	TNPP_N	*B. lehensis*	AIC95918	(Ling, Rahmat et al. 2017)	No changes from reported sequence.
SP28	Ble3542_N	*B. lehensis*	AIC96089	(Ling, Rahmat et al. 2017)	Finishes ASS, added A.
SP29	Ble3571_N	*B. lehensis*	AIC96118	(Ling, Rahmat et al. 2017)	Finishes AYA, added A.
SP30	lamB_N	*E. coli* MG1655	948548	(Pratap and Dikshit 1998)	finishes HA, added A
SP31	MalE_N	*E. coli* MG1655	948538	(Zalucki and Jennings 2007)	Finishes LA, added A.

Table 5.4. Secretion parts not cloned in vector format. The changes to the amino acid sequence respect reported sequences are listed, but silent mutations to introduce standard fusion sites are omitted.

#	Part name	Organism source	GeneBank or Gene Id	Literature reference	Changes respect original sequence
SP32	OmpA_N	*E. coli* MG1655	945571	(Schlegel, Rujas et al. 2013)	No changes from predicted sequence.
SP33	OmpC_N	*E. coli* MG1655	946716	(Wang, Ma et al. 2016)	Finishes NA. Added A.
SP34	OmpF_N	*E. coli* MG1655	945554	(Jeong and Lee 2002)	Finishes NA. Added A.
SP35	OmpT_N	*E. coli* MG1655	945185	(Kurokawa, Yanagi et al. 2001)	Finishes FA, Added A
SP36	PhoA_N	*E. coli* MG1655	945041	(Huber, Boyd et al. 2005)	Changed start to ATG codon.
SP37	DsbA_N	*E. coli* MG1655	948353	(Lee, Lee et al. 2016)	No changes from predicted sequence.
SP38	SfmC_N	*E. coli* MG1655	945367	(Huber, Boyd et al. 2005)	finishes HA, Added A
SP39	TolB_N	*E. coli* MG1655	945429	(Lee, Lee et al. 2016)	finishes HA, Added A
SP40	TorT_N	*E. coli* MG1655	946289	(Lee, Lee et al. 2016)	No changes from predicted sequence.
SP41	Vrg-6_N	*Bortadetella pertussis*	M77374.1	(Knapp, Amann et al. 1996)	Finishes LA, Added A.
SP42	pIII_N	Enterobacteria phage M13	927334	(Sroga and Dordick 2002)	Finishes SA, changed S codon.
SP43	Cex-SP_N	*Cellulomonas fimi* ATCC484	CP002666.1, locus Celf_1271	Liu unpublished work	No changes from predicted sequence.
SP44	TT3D-SP_N	*Teredinibacter turnerae* T7901	NC_012997.1, locus TERTU_RS18160	Liu unpublished work	Contains PstI site. Finishes LVG. Added AA
SP45	SC1B-SP_N	*Sorangium cellulosum* So ce 56	NC_010162.1, locus sce7663	Liu unpublished work	Weak prediction.

Table 5.4. Secretion parts not cloned in vector format. The changes to the amino acid sequence respect reported sequences are listed, but silent mutations to introduce standard fusion sites are omitted.

#	Part name	Organism source	GeneBank or Gene Id	Literature reference	Changes respect original sequence
SP46	SC1C-SP_N	*Sorangium cellulosum* So ce 56	NC_010162.1, locus sce3455	Liu unpublished work	Weak prediction.
SP47	YoaJ-SP_N	*B.subtilis* 168	CP053102.1, locus HIR77_10095	Liu unpublished work	No changes from predicted sequence.
SP48	AMAE-SP_N	*Actinoplanes missouriensis* 431	NC_017093.1, locus AMIS_RS25835	Liu unpublished work	Internal change to remove BsaI site.
SP49	AMAG-SP_N	*Actinoplanes missouriensis* 431	NC_017093.1, locus AMIS_RS39000	Liu unpublished work	Finishes ALA, Added A.
SP50	CFAA-SP_N	*Cellulomonas fimi* ATCC 484	CP002666.1, locus Celf_0270	Liu unpublished work	No changes from predicted sequence.

The linear parts that were not cloned into a vector were designed as multiple parts in tandem (of two to four parts per molecule) to reduce the cost of synthesis, as described in Table 5.5. These parts could be used directly in level 1 assemblies (in a combinatorial way) or cloned into pJUMP18-Uac to isolate individual parts using the flanking BsaI and BsmBI sites (as shown in Appendix I, Figure 7.3).

Table 5.5. Tandems of linear parts.

Tandem	Parts
SP16-7-8	G1-M5_N, GAP_N, CSP_N
SP15-9-20-1	G1_N, Ble1853_N, Ble1848_N, ECDS_N
SP22-3-4	LytE_N, Chit_N, BGL_N
SP25-6-7	BGL(4)_N, MEP_N, TNPP_N
SP28-9-30	Ble3542_N, Ble3571_N, lamB_N
Sp31-2-3	MalE_N, OmpA_N, OmpC_N
SP34-5-6	OmpF_N, OmpT_N, PhoA_N
SP37-8-9	DsbA_N, SfmC_N, TolB_N
SP40-1-2	TorT_N, Vrg-6_N, pIII_N
SP43-4-5	Cex-SP_N, TT3D-SP_N, SC1B-SP_N
SP46-7-8	SC1C-SP_N, YoaJ-SP_N, AMAE-SP_N
SP49-50	AMAG-SP_N, CFAA-SP_N

5.3.2 Construction of an experimental benchmark strain

5.3.2.1 Mark I

The construction of a preliminary strain designated Mark I was attempted by transferring the cellulolytic genes of Dr Salinas's plasmid pSB1C3-Plac-*lacZ'*-H6:Chu2268-*cenA* to a JUMP plasmid vector with a constitutive mCherry gene (pJUMP29-1A:J23100-mCherry or pMV130). The aim of generating this strain was to generate a single-plasmid strain which should provide more reproducibility and stability than a two-plasmid system as used in Chapter 4.

The cellulolytic genes in pSB1C3-Plac-*lacZ'*-H6:Chu2268-*cenA* were controlled by a promoter BioBrick part that also contained a *lacZ* gene. They were digested using the flanking BioBrick sites and ligated into the BioBrick sites of the destination plasmid. Resulting clones were screened by LacZ+ phenotype (white/blue screening

using X-gal) and genotype (colony PCR/restriction analysis confirming size of the insert). The only clone with an insert of the expected size contained a point mutation introducing a stop codon at the beginning of the CenA coding sequence. Sequencing of clones with incorrect insert sizes revealed that at least two different deletions were present after cloning: one from the RBS of *chu2260* to the RBS of *cenA* (22 identical bases) or one from the middle of *chu2268* to the end of *cenA* (no repeated sequences found).

The construction of Mark I was abandoned due to the difficulties found, possibly caused by selection against the cellulase gene sequences.

5.3.2.2 Mark II and III

A series of level 2 JUMP plasmids were designed to contain the genes encoding CenA, Chu2268, mCherry as growth reporter and the transcription factors needed for inducible expression (Table 5.6). The level 2 assemblies were built on the medium-copy-number pJUMP49-2A vector, containing five level 1 assemblies: a *cenA* gene or another sequence (as shown in Table 5.7); a *chu2268* gene selected from the cellobiose-screened library in Chapter 4 (clone “7X3” in Table 4.5); a temperature-inducible gene coding mCherry and one encoding a temperature-sensitive cI transcription factor controlling the fluorescent reporter (from the assembly shown in Chapter 3, Figure 3.5); and a *lacI* gene to control the expression of the endoglucanase. The rationale to have an inducible mCherry-coding gene was to allow increased sensitivity: if the resulting strains displayed insufficient cellulolytic capacity to generate a fluorescent signal, it would be possible to inoculate cultures with a higher initial cell concentration that would lead to a higher secreted cellulase. A higher initial fluorescent signal would decrease the signal-to-noise ratio, but this would be avoided with an inoculum of non-fluorescent cells (grown at 30 °C).

The Mark II series of plasmids was built using different genes with an mCherry-coding gene controlled by the cI^{ts}-controlled promoter (Pci_P) and a very weak RBS (B0033_RN) to reduce the burden caused by this gene. Colonies with this gene had distinguishable colouration compared to the colonies without it. Nonetheless, the resulting level 2 constructs did not display sufficient fluorescence in liquid cultures. A new series of plasmids, Mark III, was built using a stronger RBS (B0034_RN), as shown in Table 5.6.

Table 5.6. Composition of Mark III plasmids.

Module	Plasmid Id	Level 1 assembly	Description
A	Endoglucanase or other, as shown in Table 5.7.		
B	pMV249	1B:Chu2268 Clone #7X1	Chu2268 for cellodextrin assimilation
C	pMV246	1C:cI^{ts} Clone #1A	Temperature-control of mCherry. Clone selected with high induction at 37 °C.
D	pMV290	1D':Pci_P - B0034_RN - mCherry_O - B0015_CT	Growth reporter (temperature-induced).
E	pMV242	1E - J23107_P - B0033_RN - lacI_O - L2U2H09_CT	LacI-coding gene with weak promoter and very weak RBS, for control of gene in 1A.

For the construction of Mark II and III constructs, a low-expression *lacI* gene was built using a weak constitutive promoter (J23107) and a weak RBS (B0033_RN) or very weak RBS (B0034_RN) to reduce the burden in the strains carrying the level 2 constructs. The colonies with the less weak RBS displayed extremely small size, possibly caused by LacI expression despite the reduced strength. Consequently, level 2 assemblies were performed with the very weak RBS-version of this gene (Table 5.6).

Table 5.7. Mark III strains. Genes in modules 1B, 1C, 1D and 1E are as shown in Table 5.6.

Mark III strain	Plasmid Id	Description	1A module (Level 1 plasmid id)
Mk3-O	pMV 292	Mark III strain with CenA with codon optimised SP.	Ptac_P - pet_RBS_R - CenAw_SP_N - CenAo_O - L1U1H09_CT (pMV244)
Mk3-W	pMV 293	Mark III strain with CenA with wildtype codon SP (but codon optimised enzyme sequence).	Ptac_P - pet_RBS_R - CenAo_SP_N - CenAo_O - L1U1H09_CT (pMV244)
Mk3-U	pMV 294	Mark III strain without CenA as negative control.	UNS-4 (dummy linker of 100 bp, pMV251)
Mk3-G	pMV 295	Mark II strain without CenA but with sfGFP to monitor induction.	Ptac_P - pet_RBS_R - His6_N - LacI_O - L1U1H09_CT (pMV291)

The level 2 constructs in Mark III strains were built with different genes as level 1A insert (Table 5.7). Two versions of an IPTG-inducible *cenA* were used, with a codon-

optimised SP and the wildtype codon SP (both had the same CenAo_O part with the codon-optimised sequence of the enzyme), generating Mk3-O and Mk3-W. A negative control plasmid without any EG was built (Mk3-U), using a short dummy linker (of the UNS series, Appendix II). Construct Mk3-G was built using a *sfGFP* gene with the same regulatory parts as *cenA* in Mk3-O and Mk3-W, but without the SP. Mk3-G was built to test how IPTG induction affected the expression of the growth reporter. Mark III constructs were built and confirmed by phenotype (cellobiose+ and mCherry+), colony PCR and restriction digestion analysis. Figure 5.1 shows construction of all Mk3 strains, including those with their construction explained in Section 5.3.4.

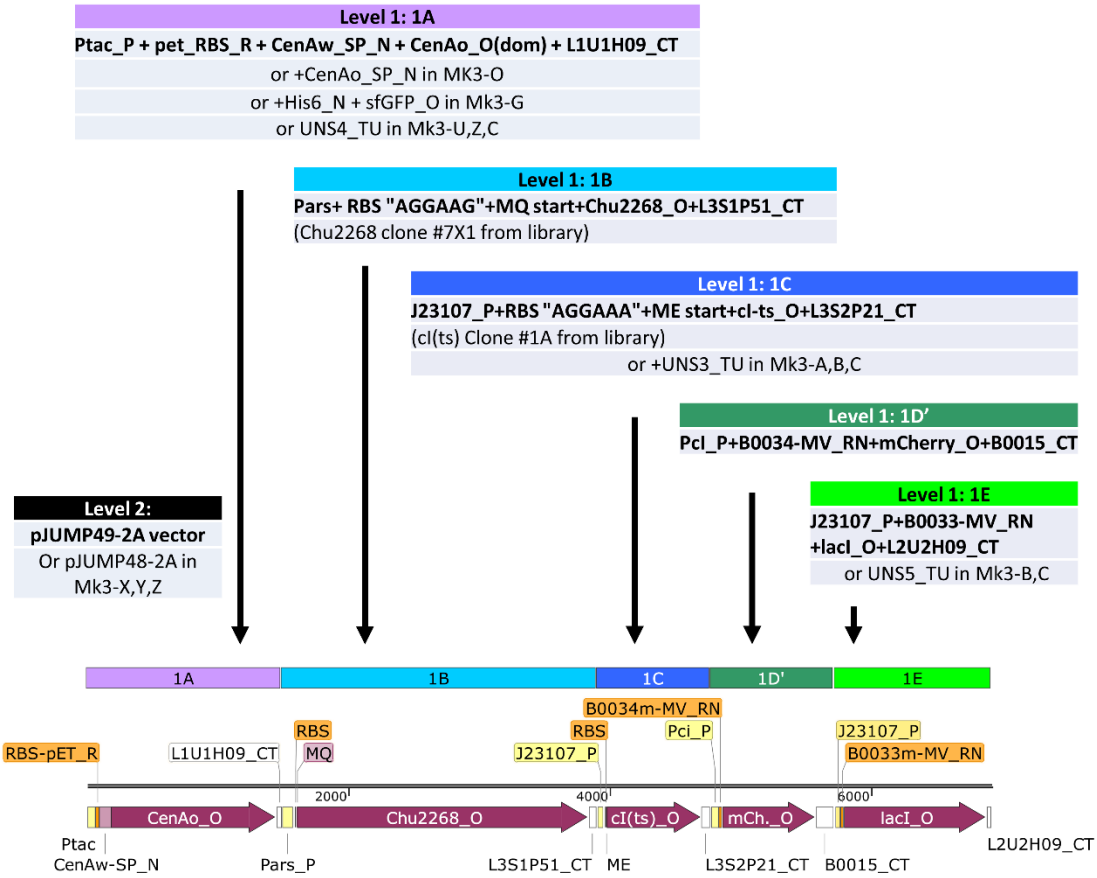


Figure 5.1. Diagram showing construction of Mk3 strains. Parts indicated are those for the “default” Mk3-W construct, with variations indicated for other Mk3 strains.

During the construction of Mark II and III strains, some selected combinations of level 1 assemblies were also combined in the secondary Downstream Site (BbsI) of a level 1 vector (pJUMP29-1A(*lacZ*)). These plasmids were built to allow screening of additional genes of interest in the presence of the “benchmark” genes. The two-

step assembly method (described in Section 2.2.2.11 and exemplified in Section 3.5.3) was used, with the difference that the final ligation was further digested with BbsI to reduce the background caused by the level 1 donor vectors, and colonies tested were picked among those expressing LacZ (the cloning marker in the Main Site of the destination vector). Mk2-W and later Mk3-W and Mk3-U were introduced into pJUMP29-1A(*lacZ*). Appendix V contains the DNA maps of the Mk3-W and Mk3-U and the confirmation of some clones via colony PCR and restriction digestion.

These results also demonstrated multi-gene assembly (and same-level same-antibiotic assembly) using the two-step modular cloning method with JUMP vectors.

5.3.3 Characterisation of the benchmark strain

The newly built Mk3 strains were used to confirm if CASFR provided a signal proportional to growth using cellulose as carbon source and whether it allowed distinguishing strains with different cellulolytic activity. LB cultures incubated at 30 °C overnight were used as inoculum, which were washed and resuspended normalising concentrations so the initial OD of cultures was 0.1. Unless otherwise specified, different cultures were prepared in 10 mL of M9y medium and incubated as 200-µL cultures in triplicate in a 96-well microplate, incubated at 37 ºC and monitored for fluorescence and OD over time.

5.3.3.1 Response to cellobiose

The effect of cellobiose concentration on growth was studied with the strain Mk3-U (*cenA*-) (Figure 5.2). OD measurements were much more variable between well replicates than mCherry readings. The noise in OD growth curves masked any differences between starting cellobiose concentrations, but the effect of cellobiose was clear on the mCherry-fluorescence curves. For all cultures (including the no-cellobiose control), fluorescence increased rapidly for the first hour of the experiment, followed by 5 hours of slower growth (and then plateau for the 0-cellobiose control). The initial quick growth was probably sustained by yeast extract and was so steep for fluorescence due to the change from 30 to 37 °C (which induces mCherry). This was the reason to exclude these first hours from the growth rate calculation. For the rest of the cultures, growth rate and overall total growth increased with initial cellobiose amounts, with differences apparent after the first hour.

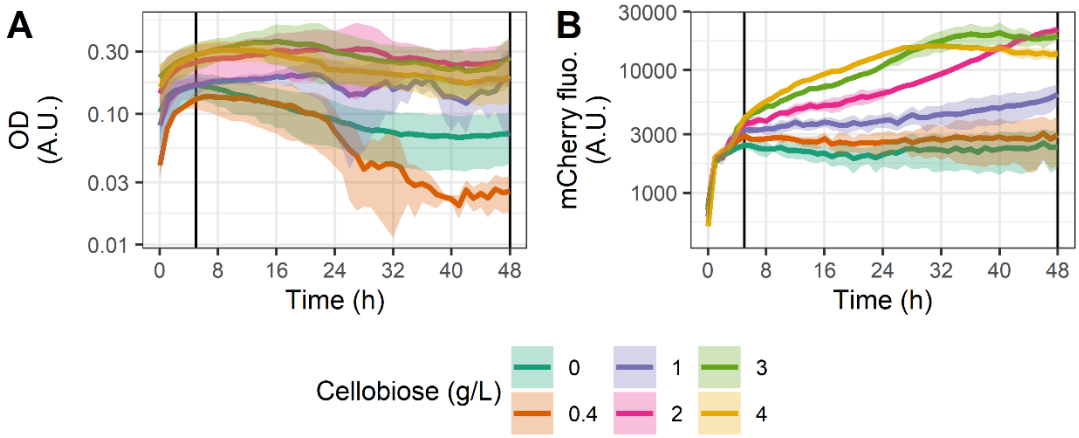


Figure 5.2. Growth curves of model strain Mk3-U in M9y medium with different concentrations of cellobiose as carbon source. Measurements were taken in 1 h intervals. A: OD, B: mCherry fluorescence. Coloured lines indicate mean of triplicate culture, coloured ribbon indicates SD. Black vertical line indicates time exclusion to calculate growth rate (from 5 h to 48 h).

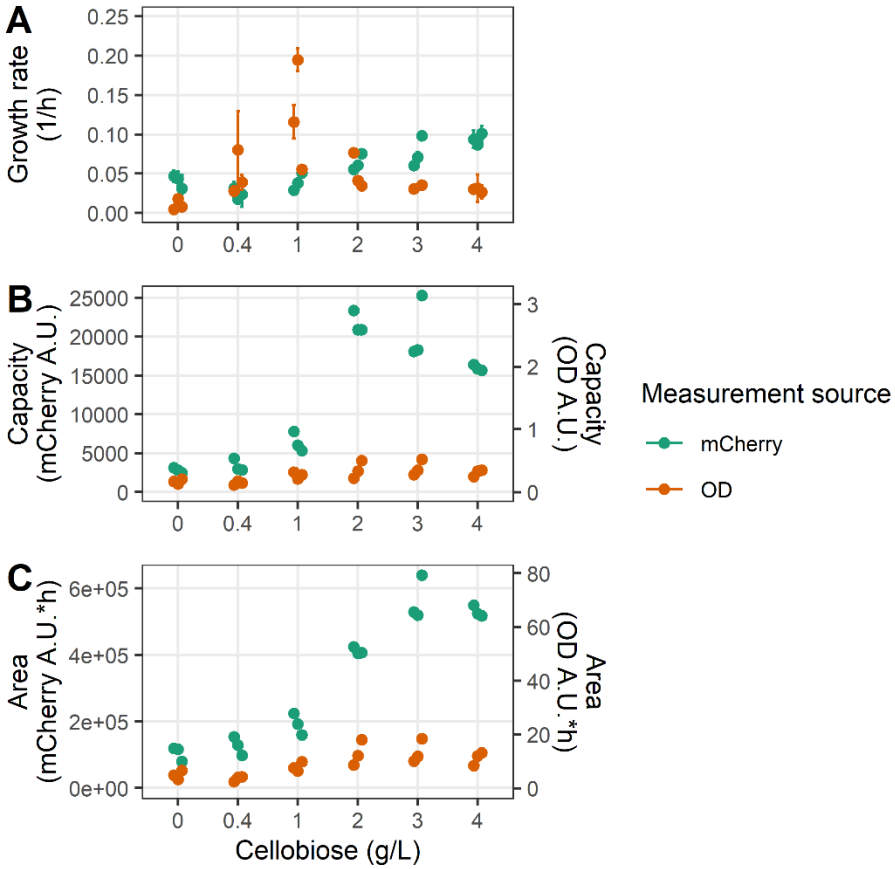


Figure 5.3. Growth metrics of model strain Mk3-U in M9y medium with different concentrations of cellobiose as carbon source. Single points indicate measurement for each measurement, colour indicate signal source for the measurement. A: Maximum growth rate calculated after the 5 h timepoint, error bars indicate standard error of regression. B: Capacity value was calculated as maximum signal achieved by 48 h. C: Area under the growth curve from timepoint 0 to 48 h.

Three growth metrics were calculated (Figure 5.3). Values calculated from mCherry reflected the same observations made from the growth curve: maximum growth rate (calculated after the 5th hour) and AUC increased with cellobiose concentration. However, capacity peaked at 2 g/L cellobiose and the fluorescence curve for this cellobiose concentration indicates that the microbial culture was growing slower than with higher cellobiose concentrations but reaching higher cell density after ~44 h.

These results, moreover, confirmed that the dilute yeast extract supplement (0.2 g / L) could not allow growth after the first hours of culture and cellulose would then become the sole carbon (or cellobiose in this experiment). Additionally, the growth curve was found to be less “exponential” than when *E. coli* was grown in M9 with

thiamine instead of yeast extract using glycerol or glucose (Figure 4.12), but growth metrics distinguished cellobiose availability well.

5.3.3.2 Response to cellulase

The effect of cellulase concentration on growth characterisation was studied using strain Mk3-U with M9y medium with 1% PASC as carbon source and different concentrations of added *T. reesei* cellulase (Figure 5.4).

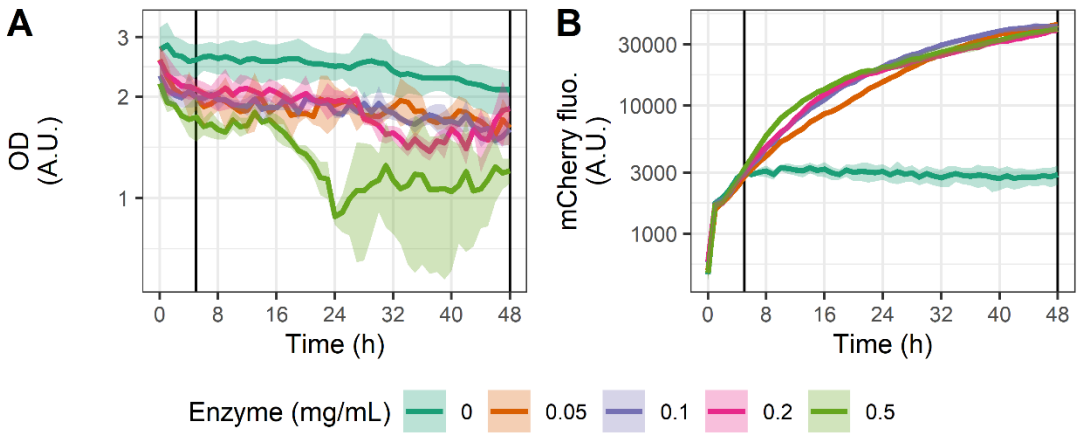


Figure 5.4. Growth curves of model strain Mk3-U in M9y medium with 1 % PASC as carbon source and different concentrations of added *T. reesei* cellulase. Measurements were taken at 1 h intervals. A: OD, B: mCherry fluorescence. Coloured lines indicate mean of triplicate cultures, coloured ribbon indicates SD. The black vertical line indicates time exclusion to calculate growth rate (from 5 h to 48 h).

OD measurements were very high from the beginning of the experiment due to cellulose turbidity; and even though an over-time OD reduction occurred in the presence of cellulase, there was no difference for intermediate cellulase loadings. The fluorescence curves showed more difference between cultures with and without added cellulase, with a small difference between cellulase concentrations, suggesting that glucose generated did not limit growth at these enzyme loadings. As in the cellobiose-response experiment, fluorescence increased during the first 5 h, without differences between cultures with and without added cellulases.

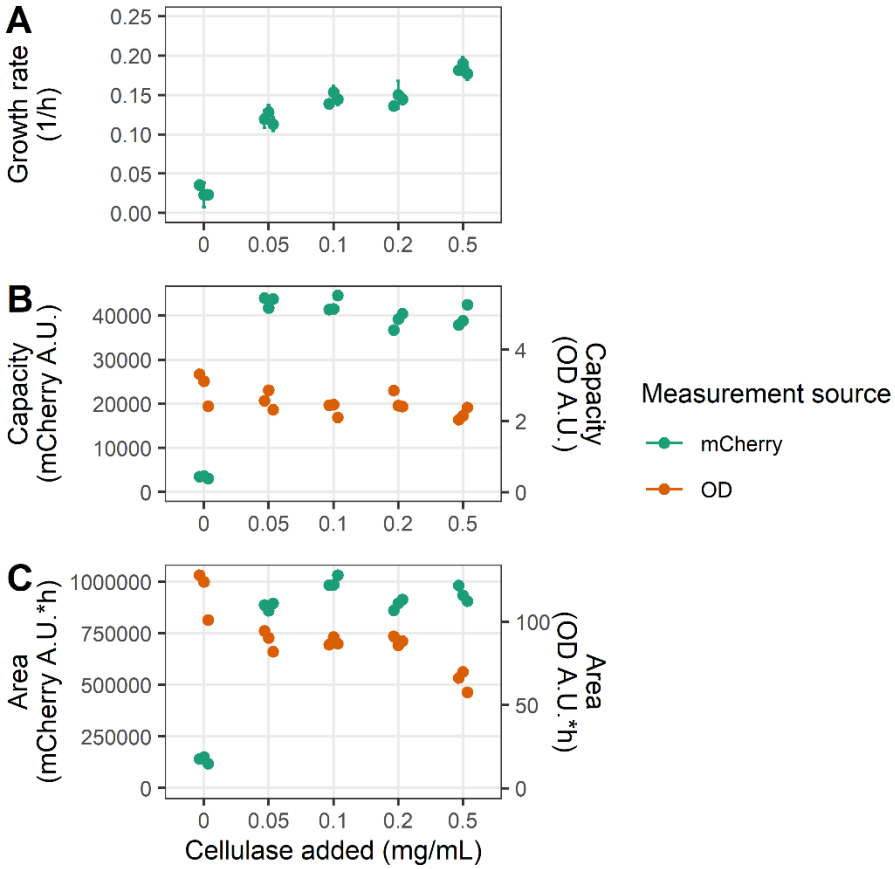


Figure 5.5. Growth metrics of model strain Mk3-U in M9y medium with 1 % PASC as carbon source and different concentrations of added *T. reesei* cellulase. Single points indicate calculations for each measurement, colour indicate signal source for the measurement. A: Maximum growth rate calculated after the 5 h timepoint, error bars indicate standard error of regression used to calculate growth rate. B: Capacity value was calculated as maximum signal achieved by 48 h. C: Area under the growth curve from timepoint 0 to 48 h.

Growth metrics based on mCherry measurements (Figure 5.5) reflected the same trend as the time-course curves, with clear distinction of the culture without cellulase. Maximum growth rate was the only metric that showed a slight but small difference between concentrations of cellulase. It is likely that glucose available was not limiting growth with any of these cellulase loadings.

5.3.3.3 Orthogonality to inducible expression

The effect of the induction of a second gene on the growth reporter was studied using the strain Mk3-G, carrying an IPTG-inducible P_{tac}-*sfGFP* on module 1A instead of an endoglucanase gene. This strain was built to determine whether the growth reporter (mCherry) would be affected by the expression of another gene.

As a preliminary test, both Mk3-G and Mk3-U strains were grown overnight in LB + Spec medium. Strain Mk3-G was grown in duplicate with and without IPTG. Triplicate measurements were taken of OD, mCherry fluorescence and sfGFP fluorescence in a plate reader (Figure 5.6). The induction did not cause a relevant decrease in expression of mCherry in strain Mk3-G and, as expected, sfGFP increased with induction (16.2-fold). Interestingly, expression of mCherry was slightly lower (~20 %) in Mk3-U.

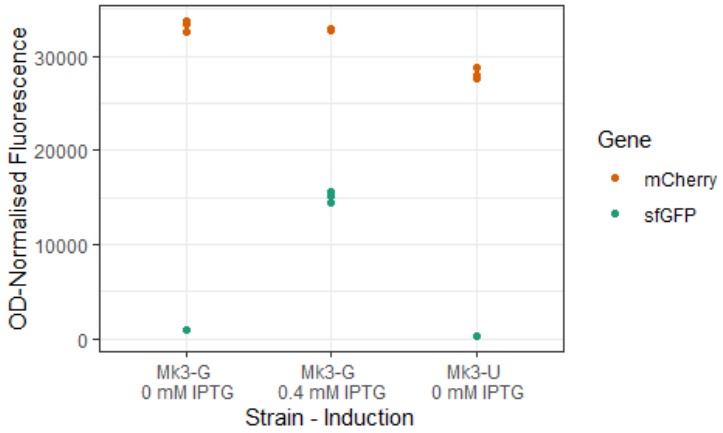


Figure 5.6. Effect of induction on mCherry growth reporter and inducible gene, using Mk3-G and M3k-U cultured in LB. Dots indicate technical triplicate measurements of one single LB culture.

It was further tested whether induction affected the growth reporter in 96-well microplate culture with M9y + 4 g/L cellobiose medium (Figure 5.7), using the same conditions as for experiments in Figure 5.2 and Figure 5.4.

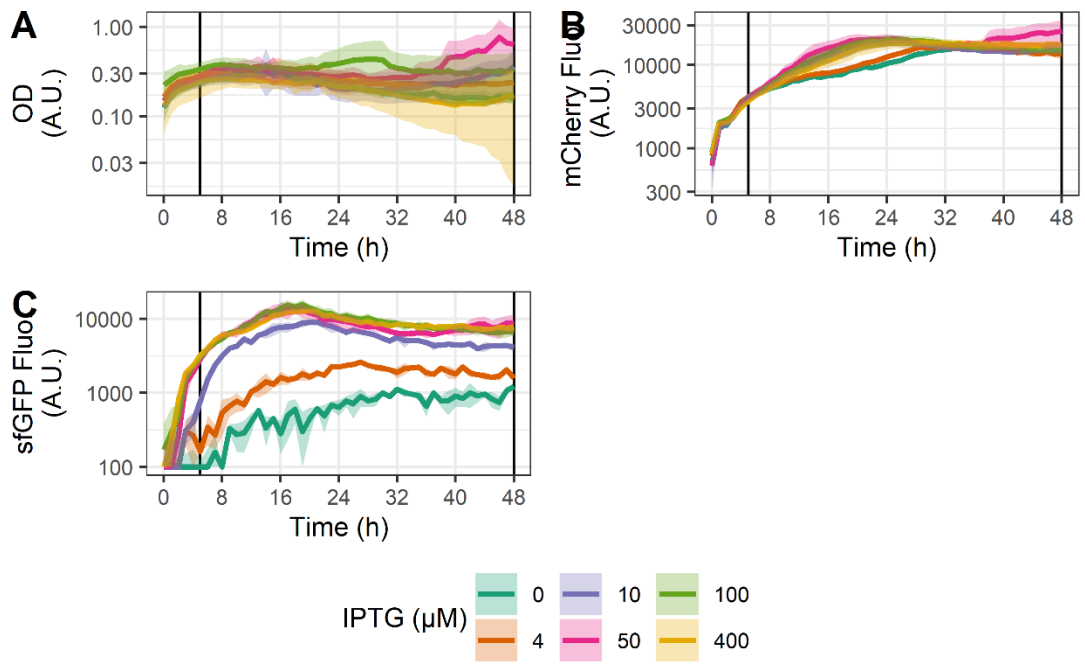


Figure 5.7. Growth curves of strain Mk3-G in M9y +4 g / L cellobiose medium with different levels of IPTG induction. Measurements taken at 1 h intervals. A: OD, B: mCherry fluorescence, C: sfGFP fluorescence. Coloured lines indicate mean of triplicate cultures, coloured ribbon indicates SD. Black vertical line indicates time exclusion to calculate growth rate (from 5 h to 48 h).

As in previous microplate-based cultures, OD was found to be much more variable between well replicates and much less “sensitive” than mCherry or sfGFP readings (Figure 5.7). OD growth curves noise did not allow distinction of differences between IPTG concentrations but both mCherry and sfGFP curves did. As expected, green fluorescence increased with IPTG. Surprisingly, the fluorescence of mCherry increased faster with higher IPTG concentrations but reaching similar maximum values, with the lowest IPTG cultures (with 0 and 4 µM) still rising towards the end of the experiment. It is important to note that sfGFP should not generate a fluorescent signal in the wavelength channels used to measure mCherry.

Growth metrics were calculated using mCherry and OD measurements (Figure 5.8). As in previous experiments, the first 5 h were not used for growth-rate calculations. The variability of OD-based metrics masked any possible difference between cultures, but values calculated from mCherry showed higher growth rates with higher IPTG, while the capacity and AUC values showed smaller differences between cultures.

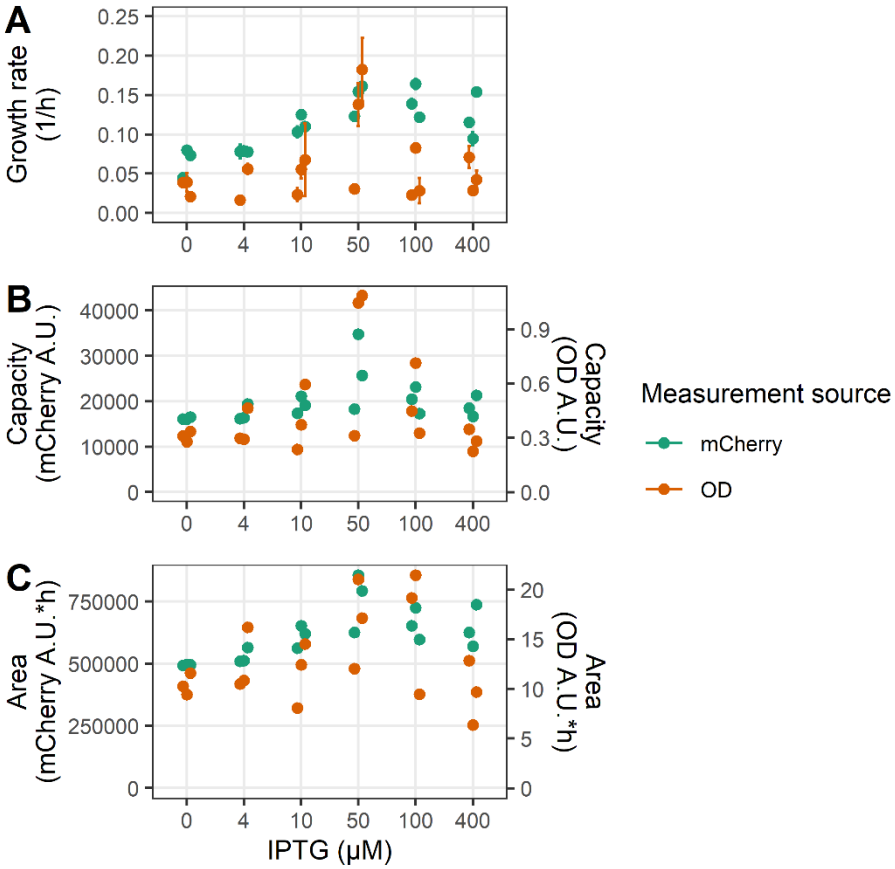


Figure 5.8. Growth metrics of strain Mk3-G in M9y +4 g/L cellobiose medium with different concentrations IPTG induction. Single points indicate measurement for each measurement, colour indicate signal source for the measurement. A: Maximum growth rate calculated after the 5 h timepoint, error bars indicate standard error of regression used to calculate growth rate. B: Capacity value was calculated as maximum signal achieved by 48 h. C: Area under the growth curve from timepoint 0 to 48 h.

The results observed could be due to IPTG increasing growth or mCherry expression. When this experiment was set, the initial 10 mL preparations of each culture were incubated with shaking at 37 °C in 50-mL centrifuge tubes after loading the microplate cultures. Every 8 hours, 200 µL samples were taken from the 10 mL cultures and measured in a plate reader (Figure 5.9). In these cultures, the OD curve showed clear differences between IPTG concentration. The effect of IPTG on OD was similar to that observed on mCherry: cultures had a faster growth but similar capacity with higher IPTG, which supported the explanation of IPTG not increasing mCherry expression but accelerating growth.

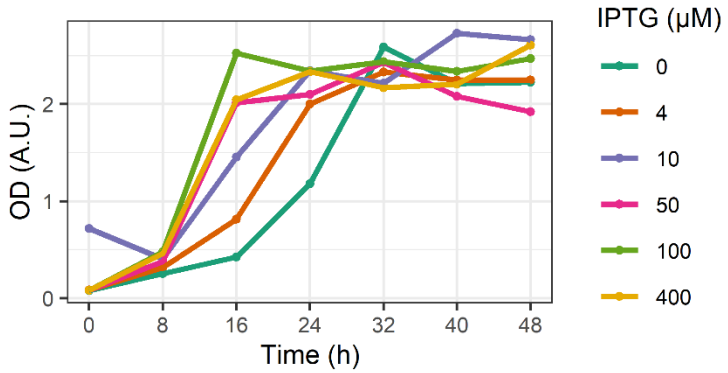


Figure 5.9. Growth curves of strain Mk3-G in M9y +4 g / L cellobiose medium with different concentrations IPTG induction. Line indicates single OD measurement taken in 8 h intervals.

5.3.3.4 Fluorescence from cellulose degradation by different Mark III strains

The development of fluorescence of Mark III strains growing using cellulose as the sole carbon source was studied. In microplate-based cultures prepared as in previous experiments, strains with CenA showed no difference to the negative control strain Mk3-U (Figure 5.10).

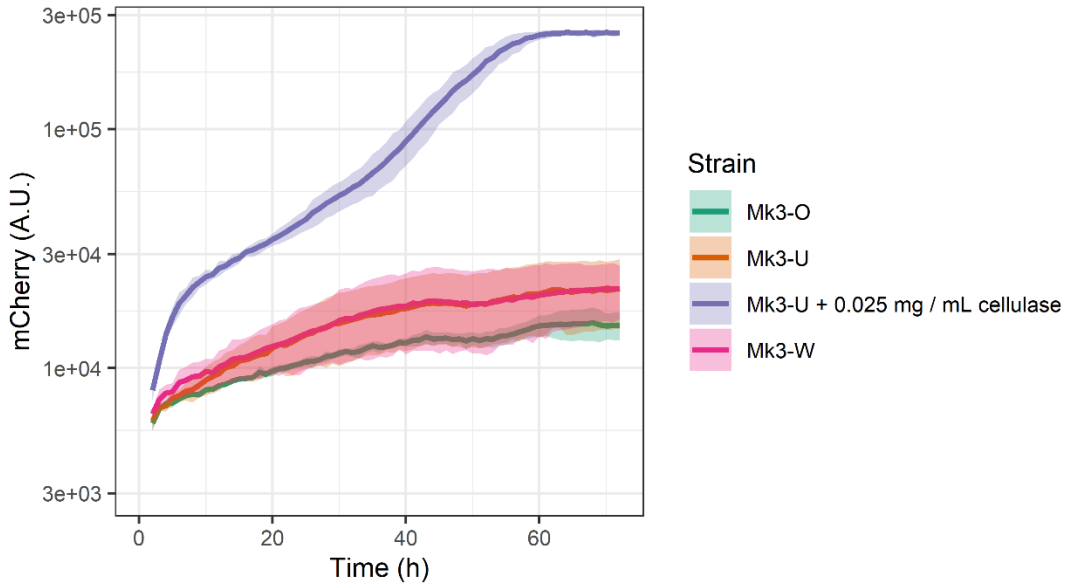


Figure 5.10. Growth curves of Mark III strains in M9y medium with 1 % PASC as carbon source in a 96-well microplate. Fluorescence measurements taken in 1 h intervals. Coloured lines indicate mean of triplicate culture, coloured ribbon indicates SD.

The lack of fluorescence-reported growth in this experiment could be due to insufficient cellulase secreted by the cells, as suggested by the positive control with added cellulase showing higher fluorescence. The previous experiment with

Dr Salinas's construct (Figure 4.9 in Section 4.3.3.1) was carried out with 10-mL cultures in 50-mL centrifuge tubes, which provided higher agitation and lower evaporation than the 200-µL microplate cultures.

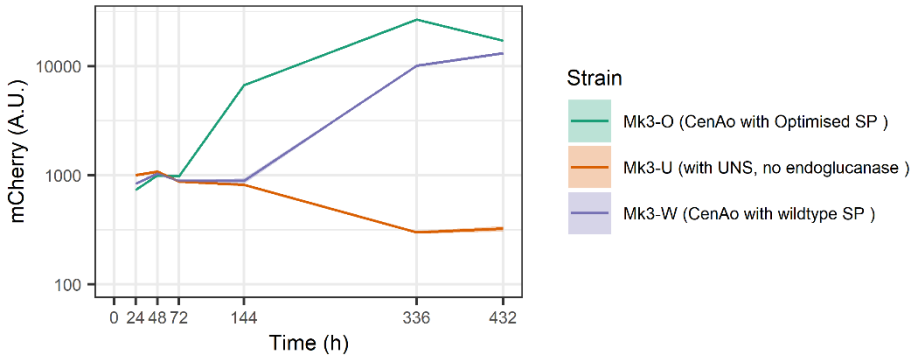


Figure 5.11. Growth of Mark III strains O, W and U in M9y medium with 1 % PASC as carbon source in 50-mL centrifuge tubes. Fluorescence measurements were taken at different intervals as indicated in X axis. Coloured lines indicate mean of technical triplicate measurements of single cultures, coloured ribbon indicates SD.

In order to determine whether the different culture conditions or the difference in strains was causing this lack of growth, Mark III strains were tested for growth from PASC as sole carbon source again but in 10-mL cultures in 50-mL centrifuge tubes (Figure 5.11). In those conditions, EG-secreting strains (Mk3-O and Mk3-W) displayed increased mCherry over time while the EG- negative control strain, Mk3-U, did not. Interestingly, Mk3-O (carrying *cenA* with codon-optimised SP) showed faster growth than Mk3-W (carrying *cenA* with wildtype-codon SP).

While this result would require confirmation (there was no culture replicate), this suggests that Mark III strains were cellulolytic, with growth being measurable at least in 10-mL cultures.

5.3.4 Alternative approaches

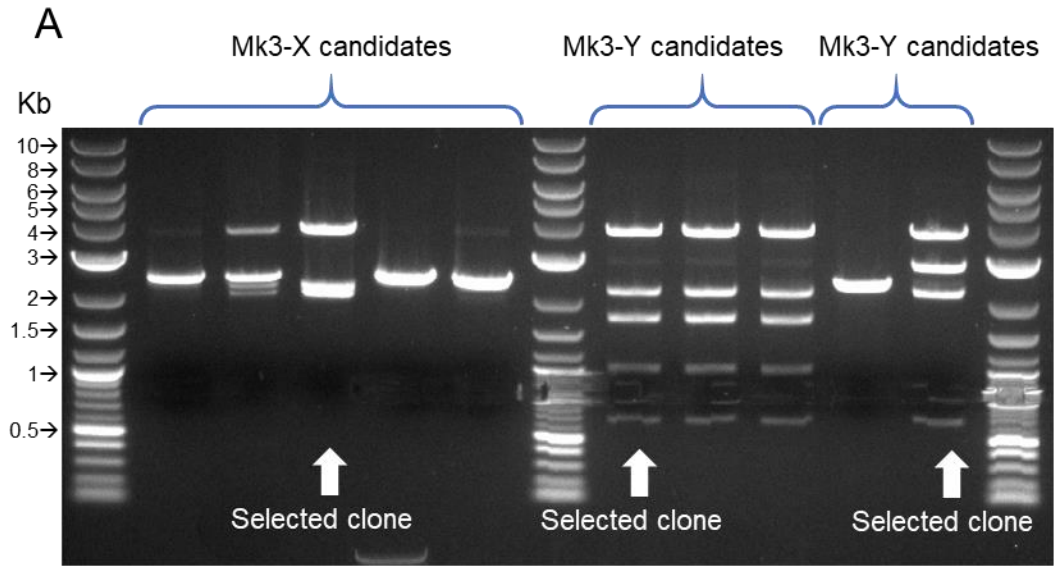
The results in Figure 5.11 indicated that Mark III strains displayed enough cellulolytic capacity to grow from cellulose as sole carbon source. However, a high-throughput strain screening system would preferably be based on microplate cultures rather than bigger-scale cultures and, consequently, different approaches were followed to get a benchmark strain showing growth from cellulose in microplate cultures.

5.3.4.1 Alternative benchmark plasmids with a higher copy number

A new set of level 2 constructs were built with the same design as Mk3-O, Mk3-W and Mk3-U (Table 5.6) but using a high copy number destination vector (pJUMP48-2A). Because an increased gene dosage would be expected to increase cellulase and growth reporter expression, these strains could potentially have improved growth and growth-reporting in microplate cultures. The resulting strains (Table 5.8), Mk3-X, Mk3-Y and Mk3-Z were built and confirmed by phenotype (cellobiose+ and mCherry+), colony PCR and restriction digestion analysis (Figure 5.12).

Table 5.8. High-copy Mark III strains. Genes in modules 1B, 1C, 1D and 1E are as shown in Table 5.6.

Mark III strain	Plasmid Id	Description	1A module (Level 1 plasmid id)
Mk3-X	pMV 301	High copy vector Mark III strain with without CenA as negative control.	UNS-4 (dummy linker of 100 bp, pMV251)
Mk3-Y	pMV 302	High copy vector Mark III strain with CenA with wildtype codon SP (but codon optimised enzyme sequence).	Ptac_P - pet_RBS_R - CenAw_SP_N - CenAo_O - L1U1H09_CT (pMV244)
Mk3-Z	pMV 303	High copy vector Mark III strain CenA with codon optimised SP.	Ptac_P - pet_RBS_R - CenAo_SP_N - CenAo_O - L1U1H09_CT (pMV244)



B

	Mk3-X (pMV301)	Mk3-Y (pMV302)	Mk3-Z (pMV303)
	BglI + NsiI	BglI + NsiI	BglI + NsiI
Expected band sizes:	1. 3788 bp	1. 3788 bp	1. 3788 bp
	2. 2209 bp	2. 2209 bp	2. 2774 bp
	3. 2014 bp	3. 1720 bp	3. 2209 bp
		4. 1054 bp	4. 680 bp
		5. 680 bp	

Figure 5.12. Restriction digestion of plasmid candidates for construction of Mk3-X, Mk3-Y and Mk3-Z. A: electrophoresis gel of digestion of miniprep plasmids with BglI and NsiI. B: Expected band sizes for correct constructs. Clones with correct band pattern selected for further tests are indicated by white arrow.

Before the strains were tested on cellulose medium, however, they were found to be genetically unstable. The strain used, *E. coli* JM109 is *recA1* and displays reduced homologous recombination, but higher copy plasmids increase the burden of genetic expression and also increase the likelihood of homologous recombination generating deletions. Streaking the strains to LB+Sp agar plates sequentially (Figure 5.13) showed appearance of mCherry- colonies descending from mCherry+ colonies, which indicated deletions of at least the fluorescent reporter gene. Consequently, the high-copy Mark III strains were abandoned.

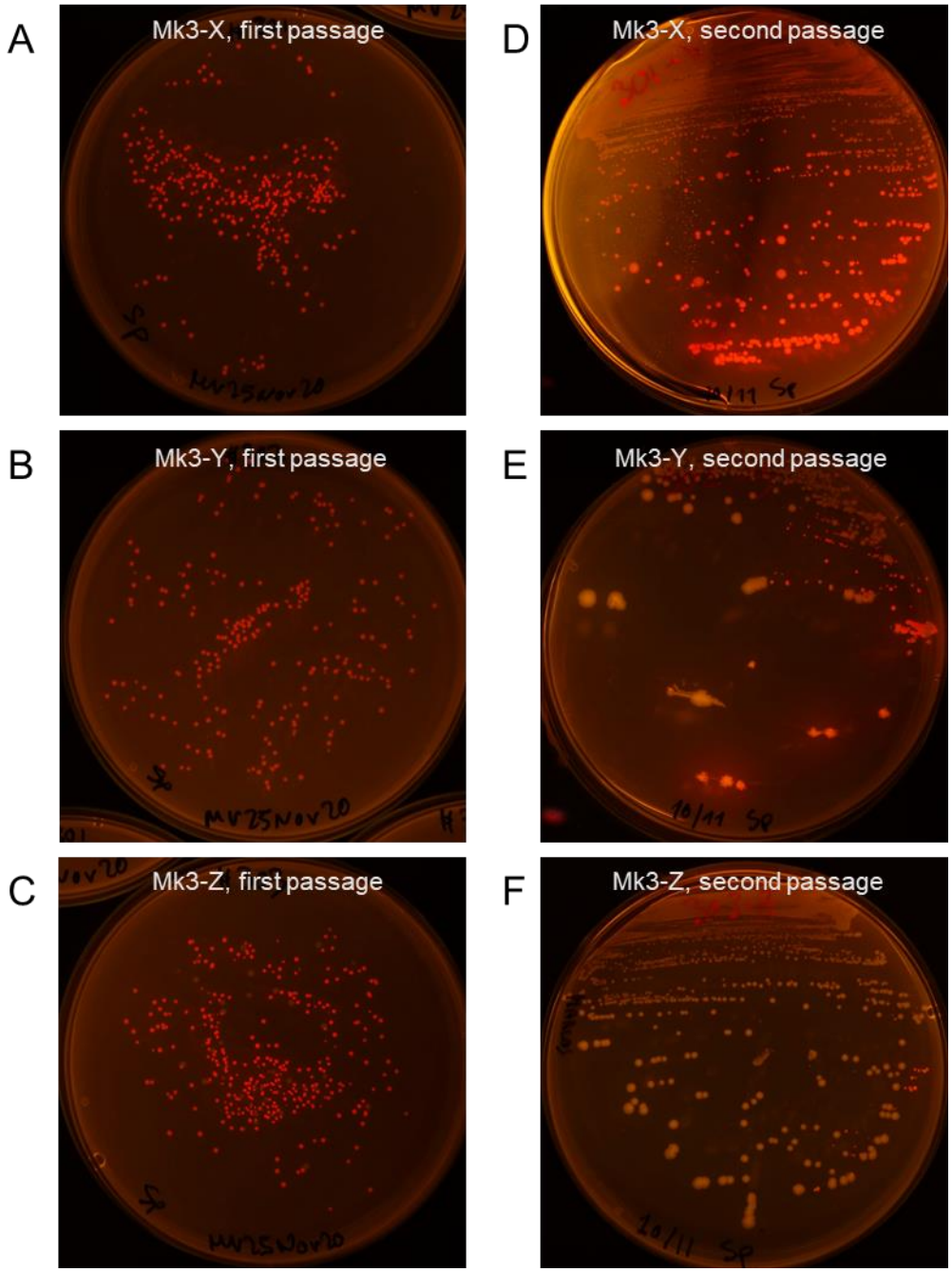


Figure 5.13. Sequential streaking of high copy Mark III strains in LB + Sp agar plates. A,B,C: fluorescence of plates with first streaking of strains Mk3-X, Mk3-Y and Mk3-Z, respectively. Bigger white (mCherry-) colonies appeared on the plate with Mk3-Z. A red colony (mCherry+) of each strain was streaked onto new plates. On the second passage, only after a 2-day incubation at 37 °C (D, E, F) some red colonies appeared on the plates streaked with Mk3-Y and Mk3-Z, which had displayed white colonies much sooner.

5.3.4.2 Alternative plasmid design

A new set of vectors was designed to increase the growth reporter and CenA expression while using the medium copy vector pJUMP49-2A. In case the expression of these two genes was too low due to their controlling transcription factors not allowing higher expression, it was decided to build new level 2 assemblies without the *cI*-ts transcription factor gene (to increase mCherry expression), or without *cI*-ts and *lacI* (to increase CenA). Even without the *lacI* gene in the plasmid, the strain used (*E. coli* JM109) should control expression of *lac*-promoters because it is supposed to be $lacI^q$. Based on the design of the level 2 plasmid Mk3-W (CenA with the wildtype signal peptide), the new plasmids were built by replacing *cI*-ts or *lacI* with different UNS linkers. A new *cenA*- negative control strain was built, also lacking *cI* and *lacI,*

Table 5.9. Mark III strains A, B and C built based on Mk3-W. The level 1 assemblies' composition is the same as for Mk3-W (see Table 5.6 and Table 5.7 in page 121).

Mark III strain	Plasmid Id	Module 1A	Module 1B	Module 1C	Module 1D'	Module 1E
Mk3-W	pMV 293	*cenA* (wt-SP)	*chu2268*	cI-ts	Pci-*mCherry*	*lacI*
Mk3-A	pMV 315	*cenA* (wt-SP)	*chu2268*	UNS-3	Pci-*mCherry*	*lacI*
Mk3-B	pMV 316	*cenA* (wt-SP)	*chu2268*	UNS-3	Pci-*mCherry*	UNS-5
Mk3-C	pMV 317	UNS-4	*chu2268*	UNS-3	Pci-*mCherry*	UNS-5

The development of fluorescence of Mark III strains A, B and C growing using cellulose as sole carbon source was tested (Figure 5.14). Microplate-based cultures were prepared as in previous experiments, with the variation that BSA was not added on this occasion. Strains with CenA showed no difference to the negative control strain Mk3-C. Control cultures with added *T. reesei* cellulase showed evident growth that became faster with higher cellulase concentrations.

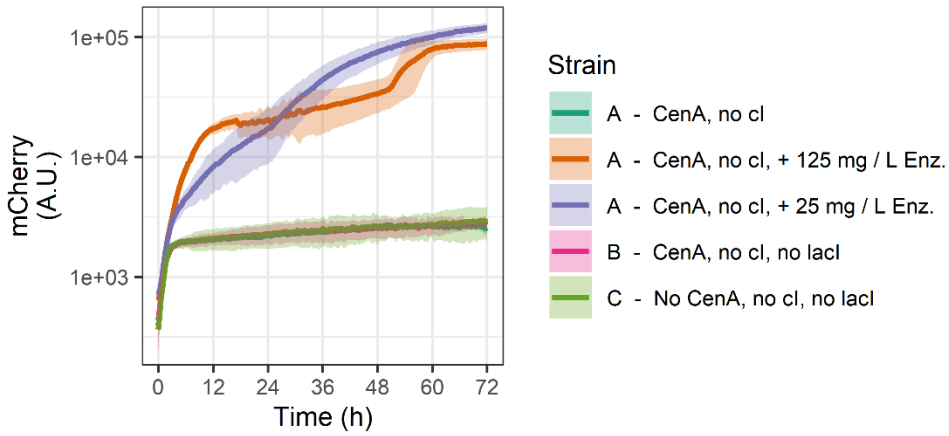


Figure 5.14. Growth curves of Mark III strains A, B and C in M9y medium with 1 % PASC as the sole carbon source in a 96-well microplate. No BSA was added to the cultures. Fluorescence measurements were taken in 0.5 h intervals. Coloured lines indicate mean of triplicate culture, coloured ribbon indicates SD.

Unlike previous experiments, the medium of these cultures (M9y+PASC) did not contain BSA, which could have been required to achieve sufficient cellulolysis in previous experiments. Determining this would require further testing.

5.3.4.3 Alternative experimental conditions

Cultures with intermediate volumes (5 and 10 mL) allowed *E. coli* to grow and express mCherry in cellulose cultures, but this mCherry-detected growth could not be transferred to microplate-based cultures. This volumetric discrepancy could be due to various causes. One possibility is that the lag phase of the cultures (the time that *E. coli* required to secrete enough cellulase to start growing in an observable way) was longer than the period that cultures could be maintained in microplates before evaporation halted growth. An alternative explanation was a reduced growth capacity caused by lower agitation and aeration, limiting energy obtained from each glucose molecule (which is lower in anaerobic conditions). Strong agitation (through vortexing) has also been shown to increase hydrolysis and disruption of cellulose filter paper by *C. freundii* (Lakhundi, Duedu et al. 2017).

It was decided to confirm the observed mCherry-detected growth of Mark III strains in 5 mL cultures; and at the same time to test the effect of additives in the medium, including BSA. All tested strains were grown in the standard CBP-like medium used in previous experiments (i.e. M9y + 1 % PASC + 0.4 mM IPTG + Sp), and one strain (Mk3-A) would be with medium modifications. Cultures were incubated in 50-mL

centrifuge tubes, and 200 µL were sampled to 96-well microplates every four days to measure fluorescence. Additionally, the samples taken from the tube-cultures would be incubated in the plate reader and measured over time to determine if the lag period was the limiting factor in microplate cultures.

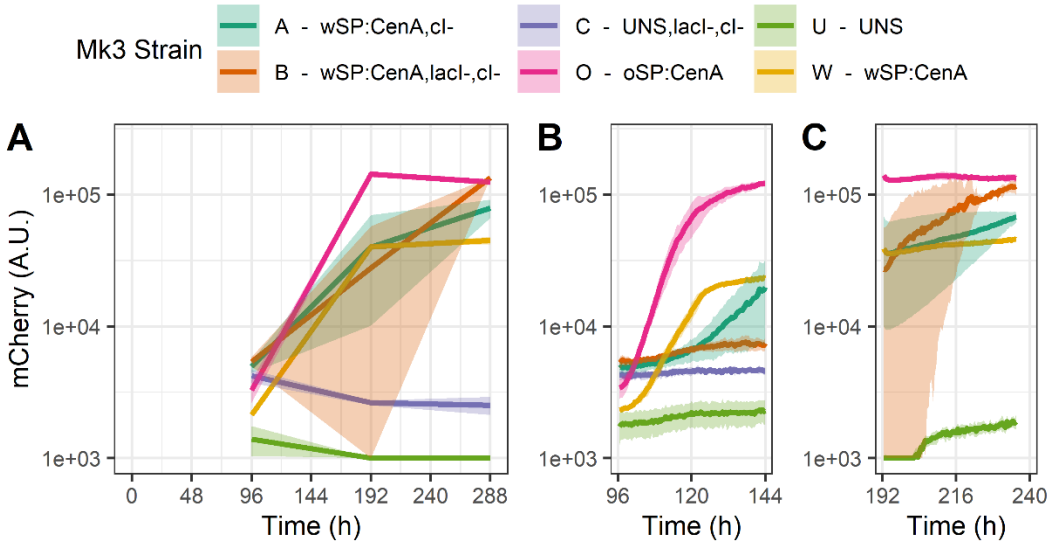


Figure 5.15. Growth of Mark III strains (A, B, C, O, W and U) in M9y medium with 1 mg / mL BSA and 1 % PASC as carbon source. A: cultures in 50-mL centrifuge tubes, with sampling and fluorescence measurement every 4 days. B: incubation and measurement (0.5 h interval) of microplate with samples from day 4 (= 96 h). C: incubation and measurement (0.5 h interval) of microplate with samples from day 8 (= 192 h). Lines show mean of triplicate[1] cultures, shadow indicates SD. Curves for individual cultures appear in Appendix V.

The fluorescence measurements showed evident differences between strains (Figure 5.15), with *cenA*+ strains (Mk3-O, Mk3-W, Mk3-A and Mk3-B) showing mCherry increasing over time while *cenA*- strains (Mk3-U, Mk3-C) did not. This observation was evident both in the 5-mL cultures and the 96-well microplate cultures derived from them (depending on the strain and sampling time), indicating that a pre-incubation allowed Mark III strains to grow from cellulose in microplate cultures. Interestingly, the initial quick growth previously observed in previous microplate cultures (such as Figure 5.2 or Figure Figure 5.14) was completely absent in the microplate cultures from day 4, which displayed exponential curves from the beginning for some strains. The initial growth previously observed was

[1] Some individual cultures were lost. One culture of Mk3-W was spilled before the first sampling, and this strain only appears as n=2 for the whole experiment. One culture of strain Mk3-B evaporated between day 4 and 8; this strain appears as n=3 for day 4 measurement and time-course, and n=2 for the other measurements.

probably caused by yeast extract supporting growth during the first hours and increase of temperature from 30 °C to 37 °C (from inoculum culture to main culture), and these two factors were probably left behind before the fourth-day microplate was set.

The increased growth (or at least increased mCherry fluorescence) of strain Mk3-O over Mk3-W was replicated, which was unexpected based on the literature evidence suggesting that non-codon-optimised SP improve secretion of recombinant proteins (see Section 5.2). Even though the RBS part used was identical, this difference could be due to different translation initiation rates because it is known that the codon after the ATG start affects translation initiation (Looman, Bodlaender et al. 1987). The RBS Calculator (Salis, Mirsky et al. 2009) indicated that the translation initiation rate was 21,128 arbitrary units for CenA with the codon-optimised SP (in Mk3-O) and 5,501 A.U. for CenA with the wildtype SP (in Mk3-W).

Mark III strains A and B also showed a clear difference with their CenA- control strain (Mk3-C); however, these strains showed much higher variability between culture replicates than that displayed by Mk3-W and Mk3-O. The absence of cI^{ts} or *lacI* in the vector did not seem to improve growth-reporting or cellulolysis. Strain Mk3-A, compared to its cI^{ts}+ counterpart Mk3-W, showed higher mCherry fluorescence at day 4 (as its control Mk3-C did) but the growth observed in the microplate was much slower at that timepoint, as if higher expression of mCherry reduced the secretion of CenA and consequently the glucose available. Mark III strain B (*cI*-, *lacI*-) showed similar mCherry expression at day 4 but even slower growth. Even though this strain eventually showed higher fluorescence in later days of the experiment, the variability was very high.

Mark III strain A was tested in different conditions, including “standard” medium (i.e. M9y + 1 % PASC + 0.4 mM IPTG + Sp), medium with 5 mg / L of *T. reesei* cellulase as positive control, medium without BSA, and medium with 1.6 % (v/v) Tween 20, a non-ionic surfactant (Figure 5.16). As expected, added cellulases led to a very high fluorescence from the first measurement, and not much relative growth later, probably caused by total consumption of the available glucose, limitation by another nutrient or acidification of the medium caused by glucose fermentation. Remarkably, Tween 20 caused a significant increase in fluorescence, with a much higher signal from day 4 and maintained growth until surpassing that of the added-cellulase

culture. The observed effect of surfactants could be explained by two described means: surfactants interacting with the outer membrane can increase protein secretion (Liu, Tian et al. 2014) and enhance enzymatic cellulose hydrolysis (stabilising the enzyme or making the substrate more accessible) (Kadhum, Rajendran et al. 2018).

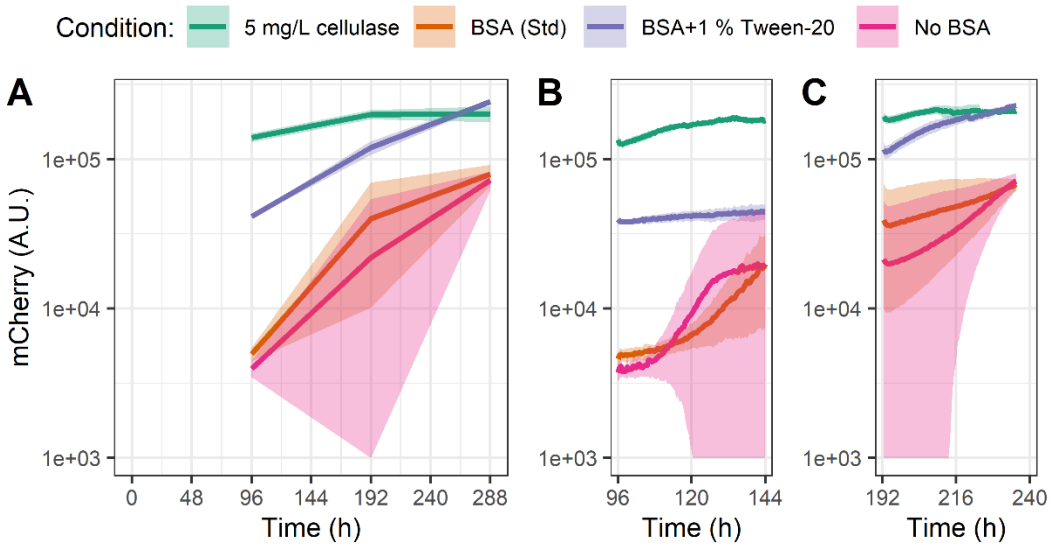


Figure 5.16. (p17nov2020_compund_ABr) Growth of Mark III strains A in M9y medium with 1 % PASC as carbon source and different additives. A: cultures in 50-mL centrifuge tubes, with sampling and fluorescence measurement every 4 days. B: incubation and measurement (0.5 h interval) of microplate with samples from day 4 (= 96 h). C: incubation and measurement (0.5 h interval) of microplate with samples from day 8 (= 192 h). Lines show mean of triplicate cultures; shadow indicates SD. Curves for individual cultures appear in Appendix V.

However, the high variability between replicates in the standard medium with and without BSA did not allow clear conclusions of the effect of BSA. Cultures with and without BSA had similar mCherry signals at days 4 and 12, but examination of individual cultures (shown in Appendix V) shows that two out of three cultures with BSA had higher growth at day 8, while only one of the three cultures without BSA had high growth then. These results are not conclusive, but they might have been caused by a delay in growth in the no-BSA cultures on top of the high variability between replicates.

5.4 Conclusions

The results presented in this chapter have set some foundations to perform an improved research cycle for generating and screening IBPM candidates. A collection of parts was generated to allow assemblies with diverse combinations of SP and cellulolytic proteins, and the introduction of multiple genes in a secondary site of a JUMP vector was demonstrating, opening a door to future combinatorial assemblies to generate SP/CDS diversity in plasmids ready to test. Furthermore, a successful *E. coli* benchmark strain was built and tested, despite challenges.

The effect of LacI and IPTG is particularly intriguing and would deserve elucidation for minimising metabolic burden during the construction of future strains. The genomic *lacI* gene provides just enough repressor to control the genomic lactose promoter (except for *lacI*^{q} strains as *E. coli* JM109 is reported to be). Because of this, overexpression of LacI is usually required to control the inducible expression of genes on multi-copy plasmids. Adverse effects on growth caused by LacI have not been reported, except for the isolated DNA-binding domain of the transcription factor being extremely toxic (Romanuka, van den Bulke et al. 2009). Nonetheless, the Mark III-encoded LacI has a valine after the start codon, leading to a much longer protein half-life compared to the lysine found in the wildtype LacI (Tobias, Shrader et al. 1991). Optimising the expression of LacI could follow similar work to that done for the cI-ts transcription factor in Chapter 3. Alternatively, the transcription factors might not be required if the strains used were capable of inducible expression of plasmids by themselves, as is the case with the *E. coli* Marionette strains (Meyer, Segall-Shapiro et al. 2019)

The difficulty of manipulating the wildtype-codon *cenA* impeded what should have been a simple cloning task (Mark I construct) and would have allowed improving culture methodologies before constructing Mark III strains. Nonetheless, Mark III strains were built and, after some trial and error, they demonstrated measurable microbial growth from cellulose as main carbon source in microplate-based cultures. Moreover, fluorescence measurements revealed clear differences between closely related strains (Mk3-W and Mk3-O) and differences between additives in the media, which might set an example of how fluorescence can be used to optimise culture conditions to reach a high-throughput screening system. An initial pre-incubation in 50-mL tubes allowed measurable growth in microplate culture, and it also removed the initial very quick growth which had been observed even in initial microplate

control cultures, probably caused by yeast extract or temperature change. Even though a pre-incubation in a bigger volume seems essential for the tested strains, this requirement should not impede high-throughput screening in microplate-based cultures; it is possible that the initial culture can be done in deep-well 96-well plates and samples transferred to measurable microplates in an automation-friendly way.

Once having found optimal culture conditions for a high-throughput screening system, it would be required to confirm that the fluorescence signal or a derived growth metric correlate with cellulose degradation of the strains. It was confirmed that fluorescence increased with cellobiose available and cellulase activity, but a confirmation using an alternative measurement of growth in the same conditions would more definitive. Such confirmatory growth measurement could be quantification of colony forming units (viable cells) or protein concentration, although the latter would be obstructed by the presence of BSA. Confirmation that the difference in fluorescence profiles of strains allows prediction of their cellulolytic capacity would be the definitive proof that CASFR is a viable and useful method to screen IBPM candidates.

Successful application of CASFR in a microplate culture format would provide IBPM development with a research throughput comparable to that found in other synthetic biology projects that require optimisation of multiple genes. As in projects where the desired product can be measured spectroscopically (Coussement, Bauwens et al. 2017), or even when product synthesis is linked to growth rate using a biosensor (Dietrich, Shis et al. 2013), it should be straightforward to do an initial screen of several hundreds of pathway library members to then confirm the activity of best performers via other methodologies. Additionally, automation could increase the research capacity even further, following the work of “biofoundries” such as that shown by Carbonell, Jervis et al. (2018), where a 500-fold increase in production of a target molecule was achieved after combinatorial optimisation of the expression strength of four genes (in addition to gene order, vector, host strain, medium and induction conditions). Reaching a research cycle with high-throughput capacity and laboratory and informatics automation should be the next step towards the development of an IBPM.

Chapter 6 Concluding remarks

The advances of synthetic biology of the last decades have been barely applied to engineering microbial hosts for a consolidated bioprocess. In this project, some of the challenges of developing an ideal biofuel-producing microorganism have been assessed. The new methodologies developed should allow future researchers to iterate the design-build-learn-cycle research cycle faster and in higher throughput.

Engineering microbial hosts to digest lignocellulose is a problem with a modular nature: multiple recombinant genes must be inserted in the host. Each of them has regulatory elements and coding sequence divided between secretion signals and the mature protein (which could further be divided into functional domains). Consequently, a modular approach to assemble DNA is highly convenient. JUMP combines the benefits of established modular cloning systems (part reusability, efficient assembly, automation-friendliness) with additional flexibility in the vector backbone (to adjust copy number or facilitate transfer to new hosts, to reduce assembly work with vector modifications). Because of these features, this project's outcomes will benefit future research more than if a non-modular and non-standardised approach had been taken.

The collection of standard parts created for general use, in addition to the vectors, should be useful for synthetic biology researchers, and the set of signal peptides and cellulolytic protein parts will allow a simple introduction to modular cloning to those engineering cellulolytic organisms. Future development of JUMP could include adapting vectors to new hosts, domesticating different classes of SP parts (such as Tat signals), and domesticating a more diverse set of cellulolytic parts (such as enzymes targeting hemicellulose or lignin).

Cellulase activity-screening via fluorescent reporter (CASFR) could overcome the main constraints in screening cellulolytic candidates. CASFR is the first analytical method that requires no cellulose separation and can be applied to microplate cultures with continuous measurements. Nonetheless, a more in-depth characterisation of the effect of the scatterer concentration on measured fluorescence would be desirable, especially before applying CASFR to microorganisms with higher cellulolytic capabilities. Reproducibility of the effect observed should be tested with multiple cellulose forms, including the ones used in this project and other industrially relevant forms. The knowledge of scatterer-

fluorescence interactions in the biomedical and biophysical sciences could be applied to characterise this interaction when applied to CASFR. For example, a possible redshift could potentially be used to quantify both scatterer and fluorophore. In addition to screening candidates in a CBP context, CASFR could also be applied to optimise other processes, such as simultaneous saccharification and fermentation.

JUMP and CASFR were successfully combined in experimental benchmark strains, with Mk3-O being the most successful. The new strains demonstrated that a future CBP research cycle could be based on microbial growth measurements in microplate cultures, although the work on this front is incomplete. Considering that pre-incubation of cultures in a 5-mL scale allowed growth measurements in microplate format, the next steps would be finding the minimum pre-incubation period and the minimum pre-incubation volume required (e.g. by sampling from large culture to microplate culture after 1, 2, 3, and 4 days, and trying the pre-incubation in deep-well microplates). Another experimental facet to optimise is the medium, including surfactants, BSA and nutritional supplements. Yeast extract was found to support initial growth but possibly distort the exponential curve of cultures. Although this influence might be removed in the culture pre-incubation, a rich but defined medium could be tested, like the one used by Bokinsky, Peralta-Yahya et al. (2011) with *E. coli* growing on cellulose. A defined medium would be preferable, as it could provide more reproducibility than yeast extract. Lastly, once the methodology had been optimised it would be necessary to prove that the differences in CASFR-growth profiles (and growth metrics) of different strains predict the cellulolytic capacity or actual growth of those strains (as measured via the concentration of proteins or colony-forming units).

After calibration and confirmation of CASFR, more advanced benchmark strains could be developed, either using *E. coli* or other potential hosts such as *S. cerevisiae* or *Y. lipolytica*. If continuing with *E. coli*, the Mk3-O gene set could be introduced in a secondary site of a JUMP vector to find additional beneficial genes with combinatorial assemblies in the vector's main site. Many more aspects of the IBPM candidate require optimisation, such as quantitative tuning of proven SP/CDS partners in combinatorial assemblies. Testing different vector systems (e.g. with lower copy or integrative vector) could lead to better use of cellular resources and a more orthogonal growth assessment (Klumpp and Hwa 2014, Gyorgy, Jiménez et

al. 2015). Additionally, product synthesis and tolerance should also be improved to achieve a realistic consolidated bioprocess ultimately.

Finally, this research aimed to facilitate a transition to a fossil-fuel-free society by bringing forward economically feasible biofuels and biochemicals based on lignocellulosic biomass. Nonetheless, the bioenergy sector should be developed with sustainability at its heart, considering the origin of the energy for processing and transportation and the lignocellulosic substrate source; and the role of the bioenergy sector should be second to energy technologies that are proven to be carbon-neutral and are currently established, such as solar and wind energy. Then, biofuels are just one of the many changes that must occur immediately. The Covid-19 pandemic has caused the most important crisis of our generation, but it should also serve as an opportunity to reshape our future, ahead of the even bigger crisis that climate change represents. These are severe challenges and many wish they need not have happened in our time. But that is not for them to decide. All we have to decide is what to do with the time that is given us.

Chapter 7 Appendices

7.1 Supporting information (Chapter 3)

Supplementary Methodology

General materials and methods

Molecular biology reagents were obtained from New England Biolabs except for AarI (ThermoFisher). Oligonucleotides were obtained from Sigma-Aldrich.

For colony PCR screening GoTaq® G2 Green Master Mix from Promega Corporation was used.

Plasmid DNA used for assemblies was purified with the QIAprep Spin Miniprep Kit (QIagen) (plasmid), using 10 mL of LB overnight culture for medium and low copy number vectors, and 5 mL for high-copy number vectors (oriV #4, #8). DNA from PCR or digestions was purified with the QIAquick PCR Purification Kit (QIagen) or QIAquick Gel Extraction Kit (QIagen).

Oligo-linker preparation

Oligo-linker dummies were prepared by combining 5 µL of each oligonucleotide (100 µM), 34 µL of nuclease free water, 5 µL of T4 DNA ligase buffer and 1 µL T4 DNA polynucleotide kinase (PNK). They were incubated for 30 minutes at 37 ˚C, 95 ˚C for 5 minutes and cool down slowly to room temperature (0.2 ˚C per second if using a thermocycler) to allow annealing.

Conventional Modular Cloning (Main Module)

Parts are assembled in equimolar concentrations, with reaction and incubation conditions as explained below. BsaI is used for level 1 (and level 3 assemblies) and BsmBI for level 0 and level 2 assemblies.

Set-up (20 µL Reactions)

Reagent	Initial Conc.	Vol. (µL)	Final Conc.
Backbone DNA	x	x	20 fmol/20µL
Insert DNA	x	x	20 fmol/20µL
T4 ligase buffer	10 X	2	1X
T4 DNA ligase (enzyme)	400 unit/ µL	0.25	100 u./20µL
BsmBI / BsaI-HF	10/20 u./µL	0.5 (or 1)	10-20 u./20µL
H_2O (nuclease free)	-	To give 20 µL	-

Thermocycling incubation:

BsmBI assembly		
42 °C	15 min	
42 °C	3 min	x30 cycles
16 °C	3 min	
55 °C	15 min	
80 °C	5 min	
10 °C	hold	

BsaI-HF assembly		
37 °C	15 min	
37 °C	3 min	x30 cycles
16 °C	3 min	
37 °C	15 min	
80 °C	5 min	
10 °C	hold	

For difficult assemblies (higher number of parts, parts of bigger size) or when a higher colony number is desirable (combinatorial assemblies to generate libraries), the incubation protocol can be extended. Below there is a long protocol for overnight assemblies.

Long thermocycling incubation:

Overnight BsmBI assembly		
42 °C	15 min	
42 °C	5 min	X60 cycles
16 °C	5 min	
55 °C	60 min	
80 °C	5 min	
10 °C	hold	

Overnight BsaI-HF assembly		
37 °C	15 min	
37 °C	5 min	X60 cycles
16 °C	5 min	
37 °C	60 min	
80 °C	5 min	
10 °C	hold	

Two-step assembly (Secondary modules)

All JUMP vectors enable introduction of flanking sequences on either side of the Main Module. Conventional cloning allows introduction of any sequence using AarI (in the Upstream Module) or BbsI (in the Downstream Module). Moreover, because the receiver fusion sites of both secondary sites are the same as the ones used by the Main Module across all JUMP vectors (GGAG and CGCT), the sequences introduced in secondary sites can originate from other JUMP vector Main Modules. To maximise the utility of this design principle, we have developed a two-step assembly method to efficiently assemble TU from basic parts in either secondary site.

First step:

- Assemble inserts, without destination vector, using the methodology described in section 3.7, but increasing concentration to 40 femtomoles of each part per 20 µL reaction.
- We recommend using the overnight protocol to increase efficiency.
- While the assembly reaction is running, pre-digest the destination vector:

Upstream Pre-Digestion	Downstream Pre-Digestion
200 femtomoles of destination vector	
1 µL AarI	1 µL BbsI-HF
0.4 µL AarI's oligonucleotide 50X (see manufacturer's recommendations)	-
2 µL AarI's Buffer 10X	2 µL Cutsmart Buffer 10X
0.5 µL Shrimp Alkaline Phosphatase	
Nuclease Free Water to a final volume of 20 µL	

(Incubate at 37 °C for as long as the half-assembly is running and 20 minutes at 65 °C).

A backbone pre-digestion can be used multiple times with different insert half-assemblies. Both first-step reactions can be stored frozen, but to store them for long periods, reactions should be purified with the PCR-purification kit.

Second step (mix and incubate).

- Ligate insert half-assembly and pre-digested destination vector:

- 20 femtomoles of assembly product (10 µL of insert half-assembly reaction)
- 20 femtomoles of pre-digested backbone (2 µL of reaction)
- 2 µL of T4 ligase buffer 10X
- 1 µL of T4 DNA ligase enzyme
- Nuclease Free Water to a final volume of 20 µL

- Incubate for 1 hour at 16 °C

To select correct clones, transform *E. coli* (see note 1) with 1 – 2.5 µL of assembly reaction, and plate 10% of the assembly on LB agar containing the antibiotic corresponding to the destination vector.

Supplementary tables

Table 7.1. Modification of SEVA selection markers.

AbR	Antibiotic Selection	Changes from original SEVA (bases from SEVA SwaI site)
1	Ampicillin /Carbenicillin	Unmodified.
2	Kanamycin	Unmodified
3	Chloramphenicol	PCR-amplified marker from pSEVA3b61 was introduced in SwaI/PshAI site of JUMP vectors. Modified from pSEVA3b61 to follow SEVA guidelines of format with primers: ccggttATTTAAATaacaaacgggattgacttttaaaaaagg + ttaattGACAAAAGTCccttcaaacttcccaaaggc).
4	Spectinomycin /Strepomycin	748C>T (BtgZI site)

Table 7.2. Modification of SEVA origins of replication.

OriV	Origin of replication	Changes from original SEVA (bases from SEVA FseI site)
1	R6K	Unmodified.
2x	RK2	Unmodified, contains 4 BtgZI, 2 AarI and 1 BsmBI sites. Deletion (56bp after position 2031) found in original SEVA OriV.
3	pBBR1	486C>T (BbsI site).
4	pRO1600 /ColE1	Unmodified.
5	RSF1010	417G>T (BtgZI site), 1189G>T (BtgZI site), 1453G>A (BbsI site), 1962C>G (BbsI site). Deletions (1 bp after position 533, 2bp after position 3219, 1bp after 3223) and insertion (1 bp after position 540) found in original SEVA OriV.

Table 7.2. Modification of SEVA origins of replication.

OriV	Origin of replication	Changes from original SEVA (bases from SEVA FseI site)
6	p15A	Unmodified. Deletion (1bp after 599) found in original SEVA OriV.
7	pSC101	Unmodified.
7ts	pSC101-ts	531C>A, 793C>A (from Rep101 present in pCP20)
8	pUC	Unmodified. Mutation (427T>A) found in original SEVA OriV.
9	pBBR322-ROP	1029C>A (BsmBI site).

Table 7.3. JUMP vectors in Addgene toolkit

Well	Plasmid Name	Vector level	Purpose
A1	pJUMP19-min2	Level 0.	"Minimal cargo" vector. Allows introducing any part using BsmBI; but lacks cloning reporter. Can be used to generate new promoter/terminator acceptor vectors.
A2	pJUMP19-Pac	Level 0.	Promoter acceptor, Minimal cargo vector; with Downstream eGFP without promoter. Allows introducing any part using BsmBI.
A3	pJUMP19-Tac	Level 0.	Terminator acceptor, Minimal cargo vector; with eGFP in downstream secondary site and constitutive promoter J23100 in upstream site. Allows introducing any part using BsmBI.
A4	pJUMP18-Uac	Level 0.	Universal Acceptor Plasmid. The fusion sites of the basic part generated are introduced during domestication. sfGFP reporter gene. High copy number oriV #8.
A5	pJUMP29-1A(sfGFP)	Level 1. Core set.	OriV 9 (pBBR322/ROP; medium copy number).
A6	pJUMP29-1B(sfGFP)	Level 1. Core set.	OriV 9 (pBBR322/ROP; medium copy number).
A7	pJUMP29-1C(sfGFP)	Level 1. Core set.	OriV 9 (pBBR322/ROP; medium copy number).
A8	pJUMP29-1D(sfGFP)	Level 1. Core set.	OriV 9 (pBBR322/ROP; medium copy number).
A9	pJUMP49-2A(sfGFP)	Level 2. Core set.	OriV 9 (pBBR322/ROP; medium copy number).
A10	pJUMP49-2B(sfGFP)	Level 2. Core set.	OriV 9 (pBBR322/ROP; medium copy number).
A11	pJUMP49-2C(sfGFP)	Level 2. Core set.	OriV 9 (pBBR322/ROP; medium copy number).
A12	pJUMP49-2D(sfGFP)	Level 2. Core set.	OriV 9 (pBBR322/ROP; medium copy number).

Table 7.3. JUMP vectors in Addgene toolkit

Well	Plasmid Name	Vector level	Purpose
B1	pJUMP21-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin R6K (pir conditional).
B2	pJUMP22x-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin RK2 (broad-host-range). *CONTAINS one BsmBI and multiple BtgZI and AarI forbidden sites.
B3	pJUMP23-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin pBBR1 (medium-copy; broad-host-range)
B4	pJUMP24-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin pRO1600/ColE1 (E.coli - Pseudomonas shuttle)
B5	pJUMP25-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin RSF1010 (Broad-host-range)
B6	pJUMP26-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin p15A (medium copy number)
B7	pJUMP27-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin pSC101 (low copy number)
B8	pJUMP27ts-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Thermosensitive origin pSC101 (low copy number)
B9	pJUMP28-1A(sfGFP)	Level 1 alternative vector.	Alt. vector with Origin pUC (high copy number)
B10	pJUMP27ts-1A(lacZ)	Level 1 alternative vector.	With lacZ as alternative cloning reporter. Thermosensitive origin pSC101 (low copy number)
B11	pJUMP29-1A(lacZ)	Level 1 alternative vector.	With lacZ as alternative cloning reporter. OriV 9 (pBBR322/ROP; medium copy number).
B12	pJUMP29[dCas9]-1A(lacZ)	Level 1 alternative vector.	With lacZ as alternative cloning reporter.OriV 9 (pBBR322/ROP; medium copy number). Constitutive dCas9 in downstream secondary site.
C1	pJUMP27ts[mCherry]-1A(lacZ)	Level 1 alternative vector.	With lacZ as alternative cloning reporter. Thermosensitive origin pSC101 (low copy number). Constitutive mCherry in downstream secondary site.
C2	pJUMP29-1D'(sfGFP)	Level 1 alternative vector.	OriV 9 (pBBR322/ROP; medium copy number).

Table 7.3. JUMP vectors in Addgene toolkit

Well	Plasmid Name	Vector level	Purpose
C3	pJUMP29-1E(sfGFP)	Level 1 alternative vector.	OriV 9 (pBBR322/ROP; medium copy number).
C4	pJUMP39-1A(sfGFP)	Level 1 alternative vector.	Level 1 vector with alternative selection marker (chloramphenicol). OriV 9 (pBBR322/ROP; medium copy number).
C5	pJUMP38-1A(sfGFP)	Level 1 alternative vector.	Level 1 vector with alternative selection marker (chloramphenicol). Origin pUC (high copy number). sfGFP cloning reporter.
C6	pJUMP41-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin R6K (pir conditional).
C7	pJUMP42x-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin RK2 (broad-host-range).*CONTAINS one BsmBI and multiple BtgZI and AarI forbidden sites.
C8	pJUMP43-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin pBBR1 (medium-copy; broad-host-range)
C9	pJUMP44-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin pRO1600/ColE1 (E.coli - Pseudomonas shuttle)
C10	pJUMP45-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin RSF1010 (Broad-host-range)
C11	pJUMP46-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin p15A (medium copy number)
C12	pJUMP47-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin pSC101 (low copy number)
D1	pJUMP47ts-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Thermosensitive origin pSC101 (low copy number)
D2	pJUMP48-2A(sfGFP)	Level 2 alternative vector.	Alt. vector with Origin pUC (high copy number)
D3	pJUMP49-2D'(sfGFP)	Level 2 alternative vector.	OriV 9 (pBBR322/ROP; medium copy number).
D4	pJUMP49-2E(sfGFP)	Level 2 alternative vector.	OriV 9 (pBBR+A15:D42322/ROP; medium copy number).

Table 7.4. JUMP basic parts in Addgene toolkit.

Well	Plasmid Name	Part type	Purpose
D5	pJUMP19-J23100_P	Basic Part P	Promoter (constitutive); J23100
D6	pJUMP19-J23102_P	Basic Part P	Promoter (constitutive); J23102.
D7	pJUMP19-J23103_P	Basic Part P	Promoter (constitutive); J23103.
D8	pJUMP19-J23106_P	Basic Part P	Promoter (constitutive); J23106.
D9	pJUMP19-J23107_P	Basic Part P	Promoter (constitutive); J23107.
D10	pJUMP19-J23119s_P	Basic Part P	Promoter (constitutive); J23119 (1bp short on 3').
D11	pJUMP19-T7pro_P	Basic Part P	Promoter (non-constitutive); T7 promoter; recognized by T7 polymerase
D12	pJUMP19-Pbad_P	Basic Part P	Promoter (non-constitutive); Pbad (I13453), arabinose induced; controlled by AraC.
E1	pJUMP19-Ptac_P	Basic Part P	Promoter (non-constitutive); Ptac; IPTG/lactose induced; controlled by LacI.
E2	pJUMP19-Ptet_P	Basic Part P	Promoter (non-constitutive); Ptet; induced by tetracycline/analogue; controlled by TetR
E3	pJUMP19-Pci_P	Basic Part P	Promoter (non-constitutive); Pci; controlled by cI.
E4	pJUMP19-Pars_P	Basic Part P	Promoter (non-constitutive); Pars; metal sensitive; controlled by arsR
E5	pJUMP19-Pmer_P	Basic Part P	Promoter (non-constitutive); Pmer; metal sensitive; controlled by merR
E6	pJUMP19-Pcue_P	Basic Part P	Promoter (non-constitutive); Pcue; metal sensitive; controlled by CueR
E7	pJUMP18-B0032_R	Basic Part R	RBS; B0032m Ribosome Binding Site.
E8	pJUMP18-B0033_R	Basic Part R	RBS; B0033m Ribosome Binding Site.
E9	pJUMP18-B0034_R	Basic Part R	RBS; B0034m Ribosome Binding Site.
E10	pJUMP18-RBS-pET_R	Basic Part R	RBS; pET vector Ribosome Binding Site.
E11	pJUMP18-RBS-pET-MV_RN	Basic Part RN	RBS+N-terminus; B0032m Ribosome Binding Site + Met-Val N-terminus start.
E12	pJUMP18-B0032-MV_RN	Basic Part RN	RBS+N-terminus; B0033m Ribosome Binding Site + Met-Val N-terminus start.
F1	pJUMP18-B0033-MV_RN	Basic Part RN	RBS+N-terminus; B0034m Ribosome Binding Site + Met-Val N-terminus start.
F2	pJUMP18-B0034-MV_RN	Basic Part RN	RBS+N-terminus; pET vector Ribosome Binding Site + Met-Val N-terminus start.
F3	pJUMP18-sfGFP_O	Basic Part O	ORF; sfGFP; fluorescent protein

Table 7.4. JUMP basic parts in Addgene toolkit.

Well	Plasmid Name	Part type	Purpose
F4	pJUMP18-amilCP_O	Basic Part O	ORF; amilCP; blue chromoprotein
F5	pJUMP18-araC_O	Basic Part O	ORF; araC; transcription factor
F6	pJUMP18-tetR_O	Basic Part O	ORF; TetR; transcription factor
F7	pJUMP18-arsR_O	Basic Part O	ORF; arsR; transcription factor
F8	pJUMP18-merR_O	Basic Part O	ORF; merR; transcription factor
F9	pJUMP18-lacI_O	Basic Part O	ORF; lacI; transcription factor
F10	pJUMP18-mCherry_O	Basic Part O	ORF; mCherry; fluorescent protein
F11	pJUMP18-cueR_O	Basic Part O	ORF; cueR; transcription factor
F12	pJUMP18-lacZ_O	Basic Part O	ORF; lacZ alpha-peptide; blue/white screening.
G1	pJUMP18-dCas9_O	Basic Part O	ORF; Catalytically dead mutant of the Cas9 from Streptococcus pyogenes.
G2	pJUMP18-cI_O	Basic Part O	ORF; cI; transcription factor from lambda phage.
G3	pJUMP18-cI-ts_O	Basic Part O	ORF; cI; transcription factor from lambda phage; thermosensitive.
G4	pJUMP18-pheS_NOC	Basic Part NOC	Complete ORF; PheS A294G; counter-selection marker; mutant phenylalanine tRNA synthetase.
G5	pJUMP18-lambdaRED_NOC	Basic Part NOC	Complete ORF; LambdaRED / λRED; Policistronic "NOC" part with coding regions for the three lambda re
G6	pJUMP18-Histag_N	Basic Part N	N-terminus; polyhistidine tag
G7	pJUMP18-Streptag_N	Basic Part N	N-terminus; Strep-tag II
G8	pJUMP18-Histag_C	Basic Part C	C-terminus; polyhistidine tag
G9	pJUMP18-Streptag_C	Basic Part C	C-terminus; Strep-tag II
G10	pJUMP19-L3S1P51_T	Basic Part T	Terminator; L3S1P51; synthetic terminator
G11	pJUMP19-B0015_T	Basic Part T	Terminator; B0015. Terminator.
G12	pJUMP19-L3S2P21_T	Basic Part T	Terminator; L3S2P21; synthetic terminator
H1	pJUMP19-L3S1P32_T	Basic Part T	Terminator; L3S1P32; synthetic terminator
H2	pJUMP19-L3S1P11_T	Basic Part T	Terminator; L3S1P11; synthetic terminator
H3	pJUMP19-L2U2H09_T	Basic Part T	Terminator; L2U2H09; synthetic terminator

Table 7.4. JUMP basic parts in Addgene toolkit.

Well	Plasmid Name	Part type	Purpose
H4	pJUMP19-L1U1H09_T	Basic Part T	Terminator; L1U1H09; synthetic terminator
H5	pJUMP19-L1U1H08_T	Basic Part T	Terminator; L1U1H08; synthetic terminator
H6	pJUMP19-B0015_CT	Basic Part CT	C-terminus + Terminator; B0015. C-termini STOP + Terminator.
H7	pJUMP19-L3S2P21_CT	Basic Part CT	C-terminus + Terminator; L3S2P21. C-termini STOP + terminator (synthetic)
H8	pJUMP19-L3S1P51_CT	Basic Part CT	C-terminus + Terminator; L3S1P51. C-termini STOP + terminator (synthetic)
H9	pJUMP19-L2U2H09_CT	Basic Part CT	C-terminus + Terminator; L2U2H09. C-termini STOP + terminator (synthetic)
H10	pJUMP19-L1U1H09_CT	Basic Part CT	C-terminus + Terminator; L1U1H09. C-termini STOP + terminator (synthetic)
H11	pJUMP19-L1U1H08_CT	Basic Part CT	C-terminus + Terminator; L1U1H08. C-termini STOP + terminator (synthetic)
H12	pJUMP18-cat_TU	Basic Part Whole TU	Codes for chloramphenicol resistance; including promoter and terminator. Broad host range. From *Bacillus/E.coli* shuttle pSEVA3b61.

JUMP supplementary figures

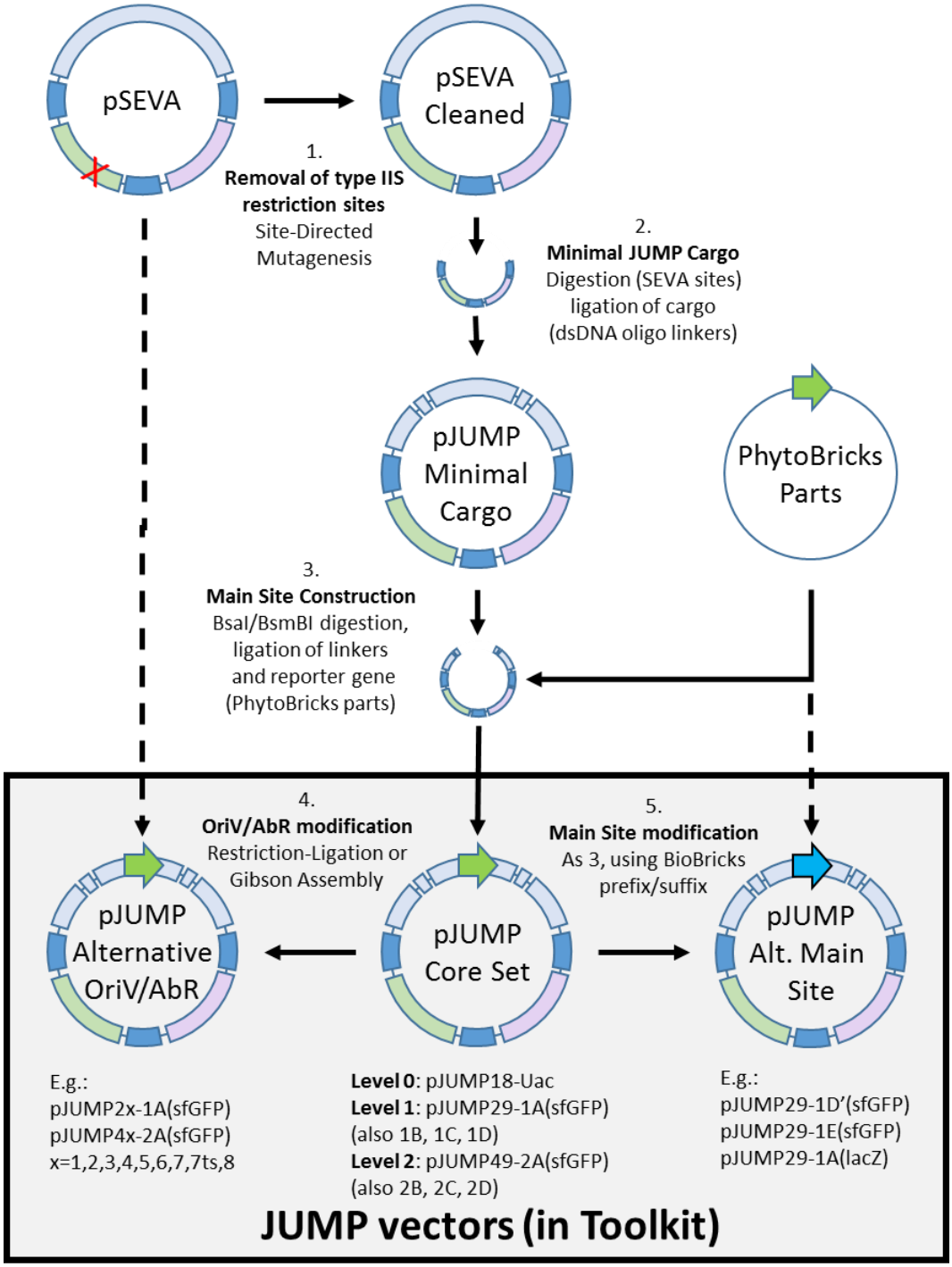


Figure 7.1. Construction of JUMP vectors. (1) For construction of the core set of vectors, forbidden sites were removed from SEVA vector using site-directed mutagenesis by blunt end ligation. (2) On the "cleaned" SEVA vectors, a minimal JUMP cargo was introduced between the SEVA PacI and SpeI sites. The minimal cargoes were built with short dsDNA linkers (oligonucleotides annealed and phosphorylated) that included Upstream and Downstream Module, Biobricks prefix and suffix, a minimal main module with only inwards BsaI (for building level 1 vectors) or BsmBI sites (for building level 0 and

level 2 vectors). (3) We digested the minimal main module BsaI/BsmBI site to then ligate linkers (containing outwards BsmBI/BsaI and the fusion site specific for each vector) and a cloning reporter gene (built from digested Phytobricks parts). 4) Vectors with alternativer OriV or AbR were built using SEVA's restriction sites or with Gibson assembly when SDM was used to remove forbidden sites. Thermosensitive OriV #7ts was built by combining the Rep101ts gene from pCP20 with SEVA's OriV #7 (pSC101 derivative). The chloramphenicol marker AbR (#3) was cloned from pSEVA3b61 changing orientation to follow SEVA guidelines. (5) We used BioBricks sites to generate alternative main module (fusion sites or alternative reporter), using linkers and PhytoBricks parts or PCR fragments.

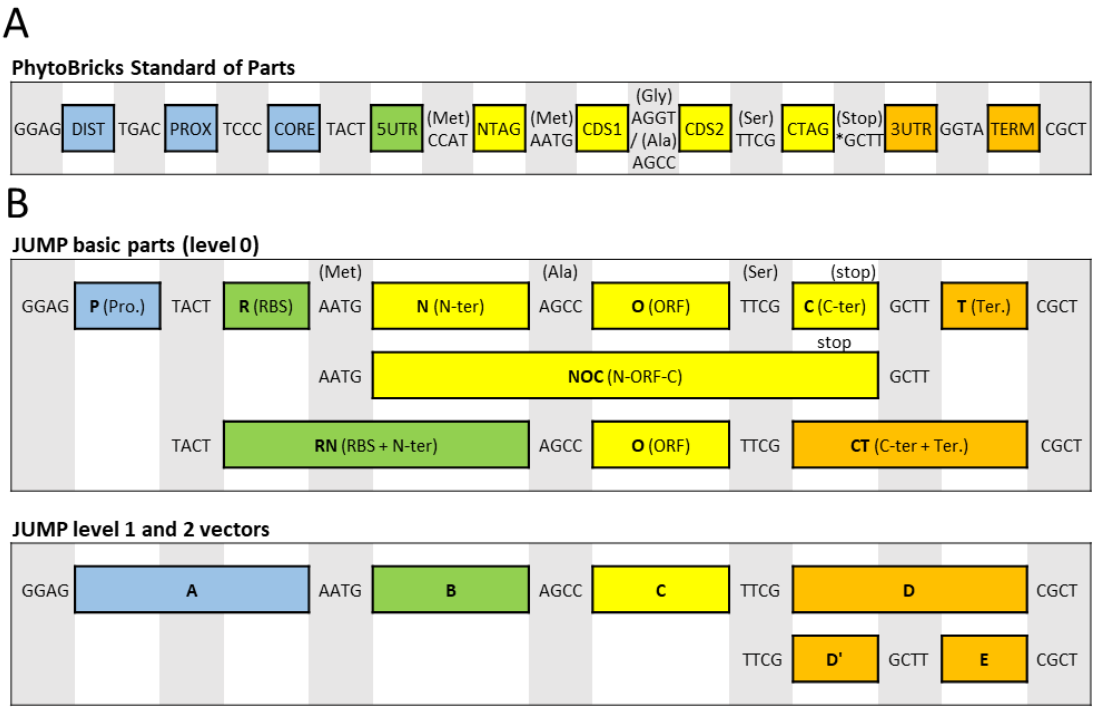


Figure 7.2. PhytoBrick standardisation of parts and JUMP fusion sites. A) PhytoBrick standard for fusion sites (18). Abbreviations of PhytoBricks stand for distal (DIST), proximal (PROX) and core/minimal (CORE) promoter regions, 5' untranslated region (5UTR), n-terminus tag (NTAG), coding sequences (CDS1, CDS2), c-terminus tag (CTAG), 3' untranslated region (3UTR) and terminator (TERM). B) Fusion sites used in JUMP parts and vectors (level 1 and level 2 "donor sites"). All "receiving" receiving fusion sites are GGAG – CGCT. Abbreviations of JUMP parts stand for promoter (P), ribosome binding site (R), n-terminus peptide (N), open-reading frame without start and stop codons (O), c-terminus peptide or end (C), terminator (T), or combination parts (RN for ribosome with n-terminus start, NOC for coding sequence with start and end, and CT for terminator with c-terminus end).

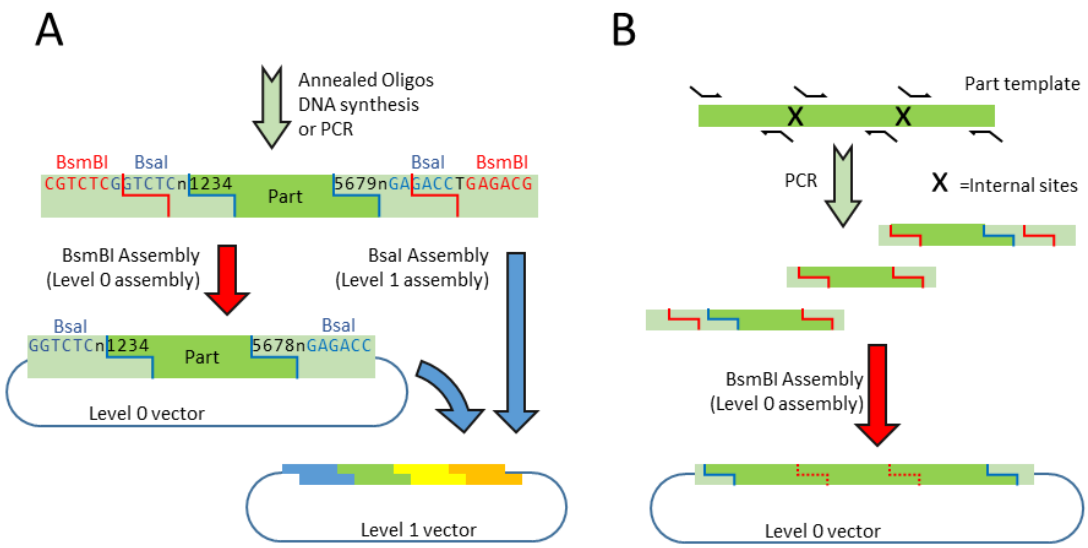


Figure 7.3. Domestication of parts. A) Domestication of parts using a Universal Acceptor (such as pJUMP18-Uac). Linear fragments with overlapping BsaI and BsmBI can be used for both a level 1 assembly and for domestication Golden Gate Reaction to introduce the basic part into a level 0 vector. The part-specific fusion sites are represented as “1234” and “5678”. B) PCR-based domestication allows removal of internal forbidden sites during cloning. Parts are amplified as multiple sub-part fragments with primers that introduce silent single-point mutations removing internal sites and flanking BsmBI site in each sub-part fragment. BsmBI is used to assemble the sub-parts in the Universal Acceptor.

A

Plasmid	Content
pNF49-2A (sfGFP)	Destination vector
pNF29-1A::J23106:His6:Chu2268	B-glucosidase from *Cytophaga hutchinsonii*, constitutive
pNF29-1B::R00106:cenAo	Endoglucanase from *Cellulomonas fimi*, lactose promoter
pNF29-1C::J23106:Pdc	Pyruvate decarboxylase from *Zymomonas mobilis*, constitutive
pNF29-1D::J23106:Adh	Alcohol dehydrogenase from *Zymomonas mobilis*, constitutive

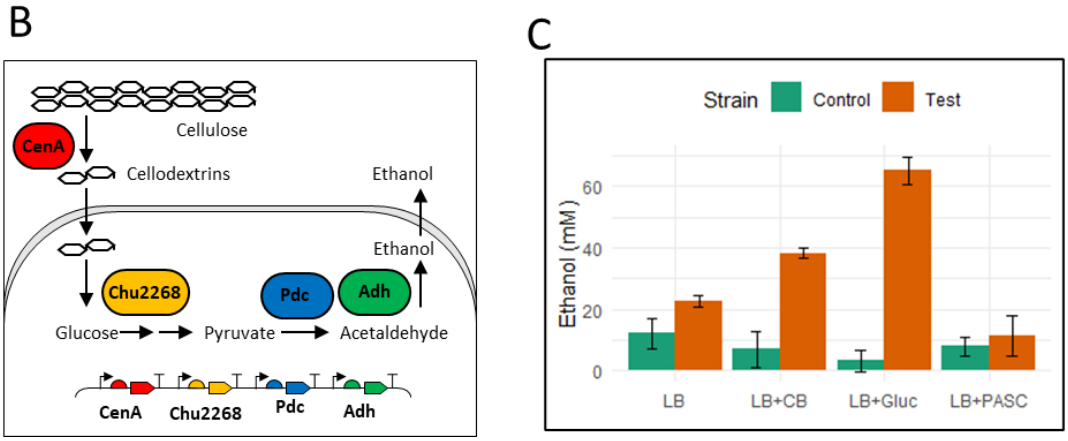


Figure 7.4. A level 2 assembly was performed to show the capability of JUMP vector for multi-gene vectors. A) Four level 1 assemblies were put together, each encoding an enzyme: two cellulases (CenA, Chu2268), a pyruvate decarboxylase (Pdc) and an aldehyde dehydrogenase (Adh). The basic parts coding for the enzymes were domesticated via PCR from plasmids containing these enzymes. The four genes were assembled correctly in pJUMP49-2A (confirmed by sequencing). B) The two cellulases allow *E. coli* to digest cellulose and to use it as carbon source (unpublished results). Adh and Pdc increase the ethanol yield in *E.coli* (Lewicka, Lyczakowski et al. 2014). C) Ethanol production of test strain (JM109+assembly) compared to control (JM109+empty pJUMp49-1A) in different media: LB, LB+CB (cellobiose), LB+Gluc (glucose), LB+PASC (phosphoric acid-swollen cellulose). Ethanol measured as explained by Lewicka et al (1): ethanol production was significantly increased from glucose and cellobiose. Even though we could not observe ethanol production from cellulose, the activity of CenA was confirmed by digestion of carboxymethyl-cellulose (data not shown).

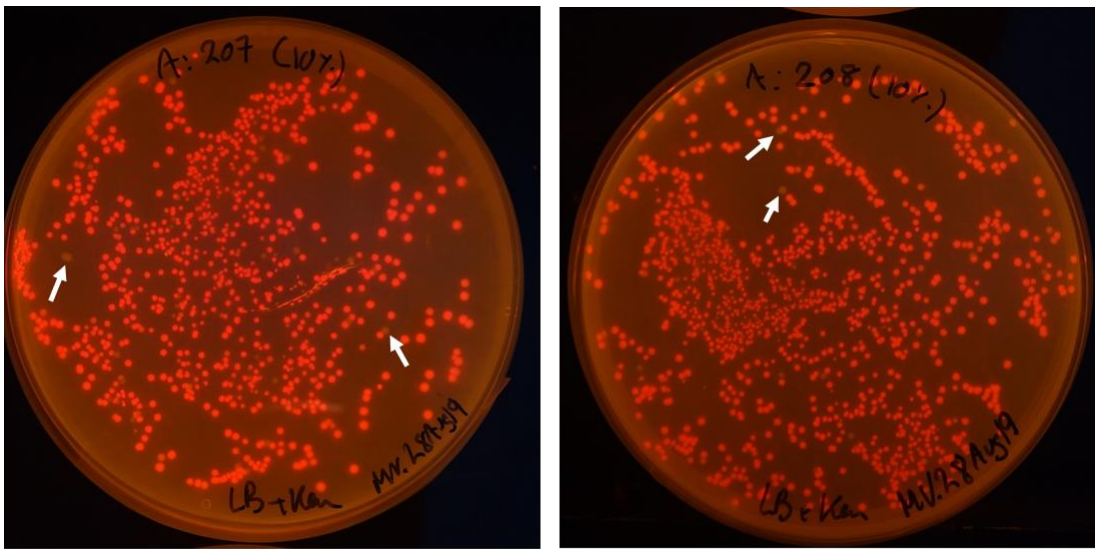
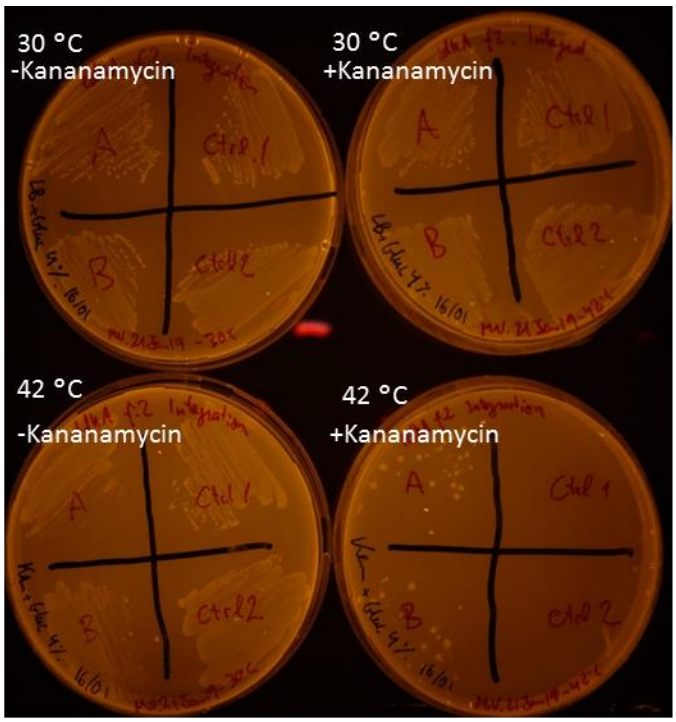


Figure 7.5. Two-step assembly of mCherry TU in secondary sites. Basic parts were combined to build a full transcription unit coding for a constitutive mCherry gene, which was assembled in the Upstream and Downstream Module of pJUMP29-1A(lacZ). Red fluorescence of colonies indicates correct assembly, while non-fluorescent colonies (as the ones pointed out by white arrows) are failed assemblies.



Ctrl 1 / 2:
pJUMP27ts-1A with no homology arms,

A and B:
pJUMp7ts-1A with homologous arms

	Homology Arm	-Kanamycin	+Kanamycin
30 °C	+HA	+	+
	No HA.	+	+
42 °C	+HA	+	+
	No HA.	+	-

Figure 7.6. Thermosensitivity of OriV #7ts and conditional integration. Conditional integration with OriV #7ts. Plasmid pJUMP27ts-1A(lacZ) (with thermosensitive replication) was tested in *Escherichia coli* MG1655 for plasmid stability and integration. While *E. coli* with the empty plasmid was able to grow carrying the plasmid at 30 °C or in the absence of kanamycin, only when homology sequence of 2 kb was inserted in the plasmid it was only to show colonies in the presence of kanamycin at 42 °C.

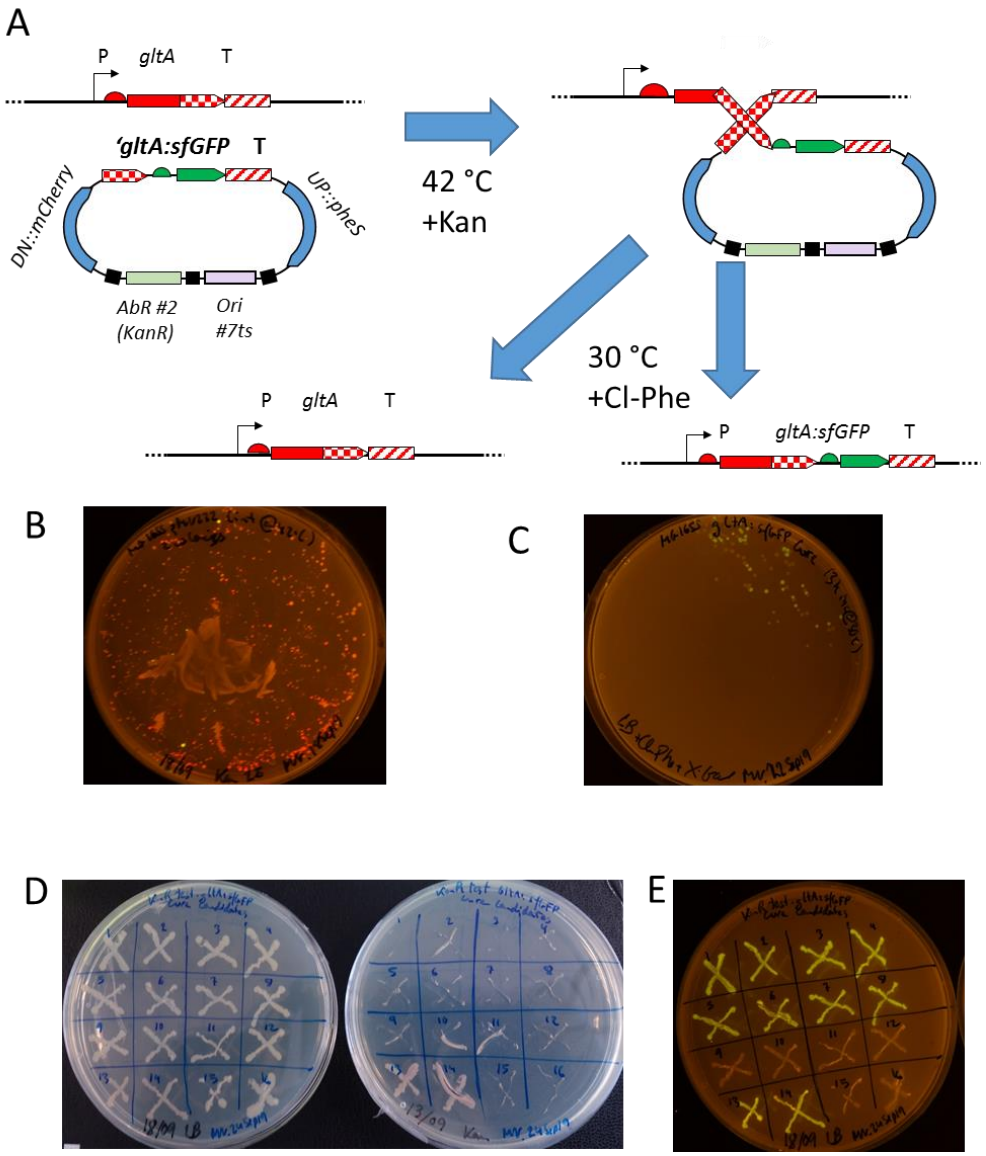


Figure 7.7. Chromosomal integration with conditional OriV #7ts. A) We introduced an mCherry reporter and *pheS* A294G counter-selection marker (1) in in upstream and downstream modules of pJUMP27-1A(lacZ), respectively. In the Main Module, we assembled a 400 bp left homology arm amplified from the end of the *E. coli* gltA gene, RBS and sfGFP parts, and 400 bp right homology arm amplified from the 400 bp following gltA. *E. coli* MG1655 was transformed and vector integration was forced by maintaining kanamycin selection at 42 °C. We incubated a GFP+ colony for 3 hours to promote a second recombination and selected curation by streaking cells in the presence of 8mM 4-chloro-L-phenylalanine. B) Plate with integrants. Homology of *pheS* with chromosomal wild-type gene led to high rate of off-site integrants (mCherry+ GFP-). C) Cure of a GFP+ integrant lead to circa 50% GFP+ and 50% GFP- colonies from recombination in right and left homology arms, respectively. D) Cured colonies were tested for kanamycin sensitivity. Left plate: LB only, right plate: LB+kanamycin. Samples: 1-8: GFP+ colonies, 9-12: GFP- colonies, 13-14: integrant pre-cure (KanR) control, 15-16: Wild-type (KanS) control. E) GFP expression of cured colonies (same samples). Correct integration was confirmed by PCR using flanking primers (data not shown).

7.2 Additional JUMP parts

The coding sequence of the cellodextrinase Chu2268 was domesticated via PCR as part type O to allow N- and C-terminal fusions, using as template Dr Salinas's plasmid used for cellulose-based growth (pSB1C3-Plac-*lacZ'*-H6:Chu2268-*cenA*, Salinas (2017)). The resulting part was named Chu2268_O. The *cenA* gene was expressed with its native signal peptide using the wild-type codon composition in Dr Salinas's experiment, but it was initially domesticated with the codon optimised sequence used by Dr Liu, generating a part of type NOC with the native SP and a part O without SP, thus generating parts CenAo_NOC and CenAo_O, respectively. Parts coding the enzymes Adh (alcohol dehydrogenase) and Pdc (pyruvate decarboxylase) were PCR-domesticated from psB1C3-Plac-*lacZ'-pdc-adhB* (Lewicka, Lyczakowski et al. 2014), generating zmAdh_O and zmPdc_O. Si-tag_N and Sb7-tag_C were domesticated via DNA synthesis.

Table 7.5. Sequences of additional JUMP parts coding for proteins. Sequences include both sides fusion sites. Amino acid additions to native sequences are listed.

Chu2268o_O. Added Ala at 5’ end, and Gly-Ser at 3’
AGCCTGTGAGAAAAAAACAGAAGCGGGGAGTAAACCTCAGGCCTTTGATAAAGAAGTACAGGA TCTGTTAAAAAATATGTCGCTGGAAGAAAAAGCCGGACAGATGACGCAGATCGATATCCGTAAT TTATTGAACAACGGATATGGAAATACCGATGAAAAATTAGATACCGCCAGATTAAAAGAAGCCAT TCAAACCTATCATGTCGGTTCTATATTAAATTGCATTCAGGCATACACACCTGAAAAATGGGTTG AATTGATTTCTCAGATCCAGAACGAAGCATTGCAATCACCTAATAAAATTCCGGTTTTATACGGA ACAGATGCCATGCATGGTGTTGGCTTTATTAAAGATGCCGTATTGTTTCCGCATAACATAGGTAT GGCTGCATCCCGCAACGATCAATTGGTTTCGCAGGCAGCTCAGGTAACGTCAACAGAAGCGCG TTCCGTTGGCCTTACCTGGAATTTTGCTCCGGTGCTTGACGTAGGCAGAGAACCTTACTGGTCG CGCTTTGAAGAAACCTTTGGTGAAGATGTATACATAACAACACAAATGGGTTCAGCAGCCGTAC AAATGATGGAAGGCAGTGACCTGACGTCTAAAACAAACATAGCTTCCTGCCTGAAACACTTTAT TGGATATTCAGCACCCAAAAATGGTATAGACCGCACACAATCACACATTCCTGAAATTGTGCTG CGCGAATATTATCTGCCCCCGTTCAGAGAAGCGATCAATAAAGGCGCTTCATCCATTATGATCA ACTCTGCTGAGATCAATGGCATTCCGTGTCACGGCAACAAATGGCTGTTAACAGATCTGCTCCG TACAGAGCTTGGATTTACCGGAATGGTTGTATCGGATTGGGAAGATGTGATCCGTTTGCACACC TGGCATAAAGTAGCTGCTACGCCGAAAGAAGCCGTAATGATGGCCGTTAACGCAGGCGTGGAT ATGAGCATGGTGCCGAATGATTACTCGTTTCCGAAATATCTGGTAGAATTGGTGAAAGAAGGAA AAGTTTCTATGGCGCGCATCGACGAAGCGGTGGGACGTATTCTGACCCTGAAAATCAAACTGG GCTTGATGAAAAATCCATTGCCATCCATTGCAGATGTTGGTGTAGTGGGTTCTGATGCACATCA ACAGATCGCACTAAATGCAGCGCGTGAATCAATTACATTATTGAAGAACGATAAAAATATTCTGC CTCTTGCAAAAGATAAAAAAATACTATTGGTTGGTCCGGCCGCTAACAGTTTATCTGCGCTGCA CAGTTCATGGTCATACACATGGCAGGGAAGCAATGAATCGCTGTATCCTGAAACAACTAAAACA ATCCGTGAGGCATTAGAAGCGAGCGGTAATAAAGCTAACATCAGAACGAATGCGACTACAGGA TTTGATGATGCAGCAAATTATGATGTTTCCTTTATTCAAAAAAATACAGCAGGGGTGGATGTAAT TATTGTGTGTGTAGGTGAAGCTGCCTATGCCGAACAGCCTGGAGTAATCAAGGATCTCAATCTG CCGGAAGCGCAGAAGCAATTGATTGTTGCTGCAAAGAAAACAGGCAAACCTGTCATTGTATGC CTGGTAGAGGGCAGGCCGCGTTTGTTTCCGGAAGAAGAAGCACTTGCAGATGCTGTAATAATG TGCTACCGTCCGGGCAGCAAAGGTGCAGACGCATTTGCAGAGATTCTATACGGTGACATCAAT CCCAGCGGTAAGCTTCCGTTCACCTATCCGAGATACGACGGAGATATTACTACATACGATTATA AATTTAAGGAAACCGAGCAGCAGTTAAAACCGGGCGTGAGTGAATTTGTTGCGTTCAATCCGCA ATGGCCATTCGGACATGGCTTGAGCTACACAACGTTTGCATATTCAAACCTTAACGTGAACAAA AGTAATTTTACGAAGAACGATAGTGTGCTGGTAACGGTAGACATAAGCAATACAGGAGCACGTA CAGGTAAGATAGCCGTAGAGTTATATTCCAGAGATCACTTCGCATCGATTACGCCTTCTGAAAG ACGCTTACGTAAATACACAAAGATTGAACTGAAAGCAGGAGAGAAGCAAACCGTTTCGTTTACG ATCAAAGCAGCGGATCTGCAATTTGTAAATAAAGATCTGAAAACAGTAACCGAAGCAGGCGCTT TTGATCTGATGATCGGAAATCTCCAGACAGAAATATATTTTAATGAGGGTTCG

Table 7.5. Sequences of additional JUMP parts coding for proteins. Sequences include both sides fusion sites. Amino acid additions to native sequences are listed.

CenAo_O. Added GS at 3’
AGCCCCGGGTTGTCGTGTTGATTATGCAGTTACCAATCAGTGGCCTGGTGGTTTTGGTGCAAAT GTTACCATTACCAATCTGGGTGATCCGGTTAGCAGCTGGAAACTGGATTGGACCTATACCGCA GGTCAGCGTATTCAGCAGCTGTGGAATGGCACCGCAAGCACCAATGGTGGCCAGGTTAGCGTT ACCAGCCTGCCGTGGAATGGTAGCATTCCGACCGGTGGCACCGCCAGCTTTGGTTTTAATGGT AGCTGGGCAGGTAGTAATCCGACACCGGCAAGCTTTAGCCTGAATGGTACAACCTGTACCGGC ACCGTTCCGACCACCAGTCCGACCCCGACCCCTACACCGACCACACCGACACCTACCCCAACC CCGACTCCGACACCAACGCCGACCGTTACACCGCAGCCGACCAGCGGTTTTTATGTTGATCCG ACAACCCAGGGTTATCGTGCATGGCAGGCAGCAAGCGGCACCGATAAAGCCCTGCTGGAAAA AATTGCACTGACACCGCAGGCATATTGGGTTGGTAATTGGGCAGATGCAAGCCATGCACAGGC CGAAGTTGCAGATTATACCGGTCGTGCAGTTGCAGCAGGTAAAACCCCGATGCTGGTTGTTTAT GCAATTCCGGGTCGTGATTGTGGTAGCCATAGTGGTGGTGGTGTTAGCGAAAGCGAATATGCA CGTTGGGTTGATACCGTTGCACAGGGTATTAAAGGTAATCCGATTGTTATTCTGGAACCGGATG CACTGGCACAGCTGGGTGATTGTAGCGGTCAGGGTGATCGTGTTGGTTTTCTGAAATATGCAG CAAAAAGCCTGACCCTGAAAGGTGCACGTGTTTATATCGATGCAGGTCATGCAAAATGGCTGA GCGTGGATACCGCAGTTAATCGTCTGAATCAGGTTGGTTTTGAATATGCGGTGGGTTTTGCACT GAATACCAGCAATTATCAGACCACAGCAGATAGCAAAGCATATGGTCAGCAGATTAGCCAGCGT CTGGGTGGCAAAAAATTCGTTATTGATACCAGCCGTAATGGCAATGGTAGTAATGGTGAATGGT GTAATCCGCGTGGTCGTGCACTGGGTGAACGTCCGGTTGCAGTTAATGATGGTAGTGGTCTGG ATGCCCTGCTGTGGGTTAAACTGCCTGGTGAAAGTGATGGTGCATGTAATGGTGGTCCGGCAG CAGGTCAGTGGTGGCAAGAAATCGCACTGGAAATGGCACGTAATGCCCGTTGG
CenAo_NOC.
AATGAGCACCCGTCGTACCGCAGCAGCACTGCTGGCAGCAGCAGCCGTTGCAGTTGGTGGTC TGACCGCACTGACCACCACAGCAGCACAGGCAGCACCGGGTTGTCGTGTTGATTATGCAGTTA CCAATCAGTGGCCTGGTGGTTTTGGTGCAAATGTTACCATTACCAATCTGGGTGATCCGGTTAG CAGCTGGAAACTGGATTGGACCTATACCGCAGGTCAGCGTATTCAGCAGCTGTGGAATGGCAC CGCAAGCACCAATGGTGGCCAGGTTAGCGTTACCAGCCTGCCGTGGAATGGTAGCATTCCGAC CGGTGGCACCGCCAGCTTTGGTTTTAATGGTAGCTGGGCAGGTAGTAATCCGACACCGGCAAG CTTTAGCCTGAATGGTACAACCTGTACCGGCACCGTTCCGACCACCAGTCCGACCCCGACCCC TACACCGACCACACCGACACCTACCCCAACCCCGACTCCGACACCAACGCCGACCGTTACACC GCAGCCGACCAGCGGTTTTTATGTTGATCCGACAACCCAGGGTTATCGTGCATGGCAGGCAGC AAGCGGCACCGATAAAGCCCTGCTGGAAAAAATTGCACTGACACCGCAGGCATATTGGGTTGG TAATTGGGCAGATGCAAGCCATGCACAGGCCGAAGTTGCAGATTATACCGGTCGTGCAGTTGC AGCAGGTAAAACCCCGATGCTGGTTGTTTATGCAATTCCGGGTCGTGATTGTGGTAGCCATAGT GGTGGTGGTGTTAGCGAAAGCGAATATGCACGTTGGGTTGATACCGTTGCACAGGGTATTAAA GGTAATCCGATTGTTATTCTGGAACCGGATGCACTGGCACAGCTGGGTGATTGTAGCGGTCAG GGTGATCGTGTTGGTTTTCTGAAATATGCAGCAAAAAGCCTGACCCTGAAAGGTGCACGTGTTT ATATCGATGCAGGTCATGCAAAATGGCTGAGCGTGGATACCGCAGTTAATCGTCTGAATCAGGT TGGTTTTGAATATGCGGTGGGTTTTGCACTGAATACCAGCAATTATCAGACCACAGCAGATAGC AAAGCATATGGTCAGCAGATTAGCCAGCGTCTGGGTGGCAAAAAATTCGTTATTGATACCAGCC GTAATGGCAATGGTAGTAATGGTGAATGGTGTAATCCGCGTGGTCGTGCACTGGGTGAACGTC CGGTTGCAGTTAATGATGGTAGTGGTCTGGATGCCCTGCTGTGGGTTAAACTGCCTGGTGAAA GTGATGGTGCATGTAATGGTGGTCCGGCAGCAGGTCAGTGGTGGCAAGAAATCGCACTGGAA ATGGCACGTAATGCCCGTTGGGGTTCGTAACTTCTGACTGAGTTGCAGCTT

Table 7.5. Sequences of additional JUMP parts coding for proteins. Sequences include both sides fusion sites. Amino acid additions to native sequences are listed.

zmAdh_O. Added 5’ Ala and 3’ Ser.
AGCCGCATCAAGCACCTTTTATATCCCGTTTGTTAATGAAATGGGCGAGGGTTCTCTGGAAAAA GCGATCAAAGACCTGAATGGTAGTGGGTTCAAAAATGCCCTGATTGTGTCAGACGCCTTCATGA ATAAAAGTGGCGTGGTGAAACAGGTTGCTGATCTGCTGAAAGCACAGGGTATCAATAGCGCCG TGTATGACGGTGTTATGCCGAATCCGACCGTTACAGCCGTACTGGAGGGACTGAAAATCCTGA AAGACAACAACAGCGACTTCGTGATTTCTCTGGGTGGTGGTTCTCCACATGATTGTGCCAAAGC CATTGCCCTGGTGGCTACAAATGGAGGTGAAGTGAAAGATTATGAGGGGATCGACAAAAGCAA AAAACCGGCACTGCCTCTGATGAGCATCAATACCACAGCGGGTACAGCATCTGAGATGACTCG TTTCTGTATTATCACCGACGAGGTTCGCCATGTGAAAATGGCCATCGTGGATCGTCATGTTACA CCGATGGTTAGCGTGAATGATCCTCTGCTGATGGTGGGTATGCCTAAAGGGCTGACTGCCGCT ACAGGAATGGACGCCCTGACTCATGCCTTCGAAGCCTATTCTAGCACCGCCGCTACTCCTATTA CCGACGCTTGTGCCCTGAAAGCTGCCTCTATGATTGCCAAAAATCTGAAAACCGCTTGTGACAA CGGCAAAGACATGCCTGCTCGTGAGGCCATGGCCTATGCCCAATTTCTGGCCGGTATGGCGTT TAATAACGCCTCTCTGGGCTATGTTCATGCTATGGCTCACCAACTGGGAGGCTATTATAACCTG CCACACGGCGTCTGTAATGCTGTGCTGCTGCCTCATGTTCTGGCCTATAATGCCAGCGTGGTT GCTGGTCGCCTGAAAGATGTTGGAGTGGCTATGGGTCTGGATATTGCCAATCTGGGGGACAAA GAGGGTGCTGAAGCGACAATTCAAGCCGTCCGTGACCTGGCTGCTTCTATCGGAATCCCGGCA AACCTGACAGAACTGGGCGCCAAAAAAGAAGATGTGCCTCTGCTGGCTGATCATGCTCTGAAA GACGCTTGTGCTCTGACTAATCCTCGTCAGGGAGATCAGAAAGAAGTCGAGGAGCTGTTTCTG AGTGCCTTTTCG
zmPdc_O. Added 5’ Ala and 3’ Ser
AGCCTCTTATACCGTGGGGACCTATCTGGCAGAACGCCTGGTTCAAATTGGGCTGAAACACCA CTTCGCTGTTGCCGGTGATTATAACCTGGTGCTGCTGGACAATCTGCTGCTGAATAAAAACATG GAGCAAGTCTATTGCTGTAACGAACTGAACTGTGGCTTTAGCGCTGAAGGCTATGCCCGTGCC AAAGGTGCCGCCGCTGCTGTTGTAACCTATAGTGTGGGTGCCCTGAGTGCTTTCGATGCTATT GGAGGGGCTTATGCTGAAAACCTGCCGGTAATCCTGATTTCTGGTGCCCCGAACAATAACGAT CATGCCGCCGGTCATGTGCTGCATCATGCTCTGGGTAAAACCGACTATCATTATCAGCTGGAAA TGGCCAAAAACATTACTGCCGCTGCCGAGGCGATTTATACTCCGGAAGAGGCTCCGGCCAAAA TTGATCATGTCATCAAAACCGCACTGCGTGAGAAAAAACCGGTCTATCTGGAAATTGCTTGTAA CATTGCCTCAATGCCGTGTGCCGCTCCTGGTCCGGCTAGTGCCCTGTTTAACGATGAAGCTTC GGATGAAGCAAGCCTGAATGCTGCCGTTGAAGAAACGCTGAAATTCATTGCCAACCGTGACAA AGTCGCTGTACTGGTGGGTTCTAAACTGCGCGCTGCTGGTGCTGAGGAAGCCGCTGTAAAATT CGCTGACGCCCTGGGTGGAGCTGTAGCAACCATGGCTGCCGCCAAATCCTTCTTTCCTGAGGA GAATCCTCATTATATCGGCACGAGCTGGGGCGAGGTTTCATATCCAGGGGTGGAGAAAACCAT GAAAGAGGCAGACGCCGTTATCGCTCTGGCTCCTGTGTTCAATGACTATTCCACCACCGGGTG GACTGATATTCCTGACCCGAAAAAACTGGTGCTGGCTGAACCTCGTTCTGTTGTGGTGAATGGC ATCCGTTTTCCGAGCGTTCACCTGAAAGACTATCTGACTCGTCTGGCCCAAAAAGTGTCCAAAA AAACTGGCGCTCTGGACTTCTTTAAAAGTCTGAATGCCGGGGAACTGAAAAAAGCTGCTCCGG CCGATCCTTCTGCTCCTCTGGTTAATGCCGAGATTGCTCGTCAAGTTGAAGCTCTGCTGACTCC GAATACCACCGTTATCGCCGAAACTGGTGATTCCTGGTTCAATGCCCAGCGTATGAAACTGCC GAATGGTGCTCGTGTTGAGTATGAGATGCAATGGGGCCATATTGGTTGGTCTGTGCCTGCTGC TTTTGGTTATGCCGTTGGTGCTCCTGAACGTCGTAACATCCTGATGGTAGGGGATGGATCATTT CAACTGACCGCCCAGGAAGTTGCTCAAATGGTTCGTCTGAAACTGCCTGTGATCATCTTTCTGA TTAACAACTATGGCTATACGATCGAGGTCATGATTCACGATGGTCCGTATAACAACATTAAAAAC TGGGACTATGCTGGTCTGATGGAGGTGTTCAATGGTAATGGCGGGTATGATTCTGGAGCCGGG AAAGGCCTGAAAGCGAAAACCGGGGGTGAGCTGGCTGAGGCTATTAAAGTTGCCCTGGCCAAT ACAGATGGGCCTACCCTGATCGAATGCTTTATTGGGCGTGAGGACTGTACAGAAGAACTGGTA AAATGGGGGAAACGTGTTGCCGCTGCTAACAGCCGTAAACCGGTGAATAAACTGCTTTCG

Table 7.5. Sequences of additional JUMP parts coding for proteins. Sequences include both sides fusion sites. Amino acid additions to native sequences are listed.

Si-tag_N (added Gly-Ser-Gly-Ala at 3')
AATGGCAGTTGTTAAATGTAAACCGACATCTCCGGGTCGTCGCCACGTAGTTAAAGTGGTTAAC CCTGAGCTGCACAAGGGCAAACCTTTTGCTCCGTTGCTGGAAAAAAACAGCAAATCCGGTGGT CGTAACAACAATGGCCGTATCACCACTCGTCATATCGGTGGTGGCCACAAGCAGGCTTACCGT ATTGTTGACTTCAAACGCAACAAAGACGGTATCCCGGCAGTTGTTGAACGTCTTGAGTACGATC CGAACCGTTCCGCGAACATTGCGCTGGTTCTGTACAAAGACGGTGAACGCCGTTACATCCTGG CCCCTAAAGGCCTGAAAGCTGGCGACCAGATTCAGTCTGGCGTTGATGCTGCAATCAAACCAG GTAACACCCTGCCGATGCGCAACATCCCGGTTGGTTCTACTGTTCATAACGTAGAAATGAAACC AGGTAAAGGCGGTCAGCTGGCACGTTCCGCTGGTACTTACGTTCAGATCGTTGCTCGTGATGG TGCTTATGTCACCCTGCGTCTGCGTTCTGGTGAAATGCGTAAAGTAGAAGCAGACTGCCGTGC AACTCTGGGCGAAGTTGGCAATGCTGAGCATATGCTGCGCGTTCTGGGTAAAGCAGGCGCTGC ACGCTGGCGTGGTGTTCGTCCGACCGTTCGCGGTACCGCTATGAACCCGGTAGACCACCCAC ATGGTGGTGGTGAAGGTCGTAACTTTGGTAAGCACCCGGTAACTCCGTGGGGCGTTCAGACCA AAGGTAAGAATACCCGCAGCAACAAGCGTACTGATAAATTCATCGTACGTCGCCGTAGCAAAG GTAGTGGAGCC
SB7-tag_C. This part has an faulty 5' site due to an additional excessive base pair between BsaI recognition site and cutting site.
ATTCGAGACAATCAAGTAGAGGAAGATAAGCTT

Table 7.6. Sequences of additional JUMP parts working as linkers. Sequences include both sides fusion sites. These parts are equivalent to a whole TU.

UNS-2
GGAGGCTGGGAGTTCGTAGACGGAAACAAACGCAGAATCCAAGCCGCT
UNS-3
GGAGGCACTGAAGGTCCTCAATCGCACTGGAAACATCAAGGTCGCGCT
UNS-4
GGAGCTGACCTCCTGCCAGCAATAGTAAGACAACACGCAAAGTCCGCT
UNS-5
GGAGGAGCCAACTCCCTTTACAACCTCACTCAAGTCCGTTAGAGCGCT
UNS-7
GGAGCAAGACGCTGGCTCTGACATTTCCGCTACTGAACTACTCGCGCT

7.3 Cellulolytic part collection

Additional enzymes-coding parts, PCR-domesticated and cloned into pJUMP18-Uac (Chapter 5),

Table 7.7. Sequences of additional JUMP parts coding for cellulolytic proteins that were PCR-domesticated and cloned into pJUMP18-Uac. Sequences include both sides fusion sites.

ML12Ao_O
AGCCGATACCCAGATTTGTGAACAGTATGGTAGCACCGTTGTTGGTGGTCGTTATGTTGTTCAG AATAATCGTTGGGGCACCACCGCACAGCAGTGTATTAATGCGACCAGCAATGGTTTTGAAATCA CCACCCTGAATGGTAGCAGCCCGACCAATGGTGCACCGACCGCATATCCGAGCATCTTTTTTG GTTGTCATTATACCAATTGCAGCCCTGGCACCAATCTGCCGATTCAGGTTAGCCAGATTAGCAG CGCAACCAGCTCAATTAGCTATCGTTATGTGAGCGGTGCAACCTATAATGCCAGCTATGATATT TGGCTGGACCCGAGCCCGAAACGTGATGGTGTTAATCAGATGGAAATCATGATCTGGCTGAAT CGTCAGGGTCCGATCCAGCCGATTGGTAGCGTTGTTGGCACAACCAATCTGGCAGGTCGTACC TGGGAAGTTTGGCGTGGTAGCAATGGTAGTAATAACGTGATTAGCTATGTTGCACAGAGCACCA CCAGTAGCCTGAATTTTAGCGTTCTGGATTTTATCAATGATACCCGTAATCGTGGTGCCATTAGC AATAGCTGGTATCTGACCAGCATTCAGGCAGGTTTTGAACCGTGGCAGGGTGGCGTTGGTCTG GCAGTTACCAGCTTTAGCGCAAATGTTAATGGTGGTGGCACCCCTCCGACCAATCCGAGCACA CCTACACCGCCTCCGAGTGGTGGTGCAGGTTGTCGTGTGAAATATACCGCAAATGCATGGAAT AATGGCTTTACCGCAGATGTGCAGATTACCAATACCGGTAGCAGCGCGATTAATGGTTGGACA CTGACCTATGGTCTGCCTGGTGGTCAGCAGGTTACCAGCGCATGGAATGCAACCGTGAGCCAG AGCGGTAGTACCGTTACCGCACGTAATGTTAGCCATAATGGTAGTGTTGCTCCGGGTGGCACC GCCAGCTTTGGTTATCAGGGCACCCTGAGCGGTAGCTATAGCAGTCCGACCTCATTTAGTCTG AATGGCACGACCTGTAGCCGTGCTTCG
AM6Bo_O
AGCCGGCACCATTAGCGGTAGCCTGTATCGTGAACCGGATACCAGCGTTAGCCGTTGGGTTGC AGCAAATCCGGCAGATAGCCGTATGCCGGTTATTCGTGATCGTATTGCAAGCCAGCCTGCAAG CCATTGGCTGAGCAATTTTAACCAGAGCACCATTCGTAGCGAAGTTAGCGGTTATGTTGGTGCA GCCAATGCAGCCGGTCAGATTCCGGTTCTGACCGTTTATGGTATTACCAATCGTGATTGTGGTG GTGCAAGCAGCGGTGGTGCACCGGATCTGACACAGTATCAGACCTGGATTAGCAATCTGGCAG GCGGTCTGGGTAATCGTACCGTTGTTATTATTCTGGAACCGGATTCAATTGCACTGCAAACCTG TCTGAGCGGTGCCGATCTGGCAGCACGTAATCAGGCACTGCATACCGCAGTTCGTACCCTGAA AAGCGGTAATCCGAGCAGCAAAATCTATCTGGATGCAGGTCATAGCACCTGGAATAGCGCAGC AGAACAGGCACGTCGTCTGATTGCAGCAGGTATTGCAGATGCAGATGGTTTTTATACCAACGTG AGCAACTTTAATCCGACCAGCGGTGAAGCAAGCTATGGTCGTAGCATTATTAGCAGCCTGAGC GCACAGGGTGTTAGCGGTAAACGTCAGGTTATTGATACCAGCCGTAATGGTGGTGCGAGCGGT GATTGGTGTGGTGATGATAACACCGATCGTCGTATTGGTCGTTATCCGACCGTTGCCACCGGT GATGACAATATTGATGGTTATCTGTGGGTTAAACCGCCTGGTGAAGCCGATGGTTGTCGTTATG CCGCAGGTAGCTTTCAGCCGGATCTGGCATATAGCCTGGCCAGCAGCGCAACCAATCCGCCTA GCCCGAGCCCGAGCAATCCGAGTAGCAGCCCGAGTAGCTCACCGAGCAGTAGCCCGTCAGGT AGCCCGAGTGCAAGTCCGAGCGCAAGTCAGCCGACCGGTGCATGTAGCGCAACCTTTCGTGT TACCGGTAGCTGGCAGGGTGGTTTTCAGGGCGAAGTTACCGTTACCGCAGGTACAGCCGCACT GAGCGGTTGGAGCGTTAGCTGGCCTCTGAGCGCAGGTCAGACCATTACCCAGGCATGGAATG GCACCCTGAGCGCCACCGGTAGCACCGTGACCGTTCGTAATGTTAGCTGGAATGGCGCACTG GCACCGCGTGCAACCGCACAGTTTGGTTTTCTGGCAAATGGTGCAGCAACCACCCCGACCCTG ACCTGTGCAGCAGCTTCG

Table 7.7. Sequences of additional JUMP parts coding for cellulolytic proteins that were PCR-domesticated and cloned into pJUMP18-Uac. Sequences include both sides fusion sites.

SC1B_O
AGCCCGCGAGGACGGCAAGGGGCCGTCCGTGTGGGACATGTTTTGTAGGAAGCCGGACGCC GTGTGGAACGGCCAGACGGGCGACGTCGCGTGCGACCACTACCACCGGTACCAGGAGGACG TCGCCCTCATGCGCGACATCGGCCTCCGCGCCTATCGACTCAGCACCTCGTGGCCGCGCGTC CTCCCGGAGGGCACGGGCGCCGTCAACGAGCGCGGGCTCGCTTTCTACGATCGCCTCGTGGA CGCGCTGCTCGAGGCGGGCGTCGAGCCCCACATCACGCTGTTCCACTGGGACTTTCCTCTCG CGCTCTACCACAAGGGCGGGTGGCTGAACCGCGACAGCGCCGACTGGTTCGCCGAGTACACG AGCGTCGTGGTCAGGAAGCTCGGCGATCGGGTGAAGGACTGGATGACGCTCAACGAGCCTCA GGTGTTCGCCGGCGCTGGCCACTACGAGGGGCGCCACGCCCCCGGCGACCGGCTGCGCTTC GCCGAGGTGCTCCGGGTCTGCCACCACGTCTTGCTCGCGCACGGCAAGTCGGTGCAGGCCAT CCGCGCGGCGAGCCCCGGGAGGAGCAACGTCTATTATGCGCCCGTCGCGCAGGTGAAATACC CCGCGACCTCGTCGCCGGAGGATCTGGAGGCGGCGCGCGCGGCCACGTGGTCGGTGAACGC GCGGACCACGTGGACCAACACGTGGTGGATGGATCCCGTGTTCTTCGGCAGGTACCCCGAGG ACGGGCTCCGCCTGTTCGCGAACGAGCTGCCCGAGGTCCGGAGCGGCGACATGGAGATCATC GCGCAGCCGGTCGATTTCTGCGGCGTGAACGTCTACTGGGGAGTCCCGGTGCGCGCCGGGG CGAACGGCGCGGCCGAGGTGGTGCCCCACCCGGTGGGACACCCGATCACCGCGTTCGAGTG GTCGGTCACCCCCGAGTGCCTCTACTGGGGGCCCAGGTTCTTCTATGAGCGGTACAAGCGGC CGATCGTGATCACCGAGAACGGCCTGTCGACCCGGGACTGGATCTCGCTCGACGGGAAGGTG CACGATCCCCAGCGCATCGATTTCATGGCGCGCCACCTGATCGAGCTGCGCCGGGCGATCGC CGAGGGCGCGGTCGTGAACGGCTACTTCCACTGGTCGATCCTGGACAACTTCGAGTGGGCTC ACGGCTACAAACACCGCTTCGGCCTCGTCTTCGTCGACTACCCGACGGGCCGGCGCGTCCTC AAGGATTCGGCGCACTGGTATCGTGAGCTCATCGCCTCGAACGGCGCGACCCTCGGGGGTTC G

Table 7.7. Sequences of additional JUMP parts coding for cellulolytic proteins that were PCR-domesticated and cloned into pJUMP18-Uac. Sequences include both sides fusion sites.

TT3D_O
AGCCAGTACCGACGACGCTAAATCCACGTCTGAGAGCCAGCGCCAACATGCAGATTCTCATTC AAAAAATATCCAATTGTGGCCCAAGGTAAAAAGCCCGATTAAGTCAGACCCGGCGCTCGAGGC GACGATCAATGATCGCTACCTGGCAAAAATGTCGATTGAAGAAAAAGTTGCGCAAATTATTCAA CCAGAAATTCGCAGTATTTCGCCCGCGCAATTTGCCGAATACCGTTTTGGTTCAATCCTTAATG GCGGTGGCGCATTCCCCCAAAACAATAAACACGCGAGCGTCGCCGATTGGGTGGCACTTGCG GATGCCTATTACGAGGCATCCAAAGCAGCGCCAAATGCCGAAACCGCTATTCCTGCTATCTGG GGAACCGATGCCGTCCACGGTCACAACAACGTAATCGGTGCAACGCTGTTCCCGCACAATATT GCTCTTGGCGCGACCCATAATGCTGAATTGATTAAGCAAATCGCAGCGGCGACTGCGCAGGAA GTTGCCGCAACCGGTATCGATTGGGTATTCGCACCAACCGTCGCGGTTGTGCGCGATGACCG CTGGGGCCGCACCTACGAAGGTTTCAGCGAAGATCCACAACTGGTTACTGCCTATGCCAAAGC CTACGTAAATGGTATGCAAGGCGAGGTAGACGCGGAGGATTTTCTGCGTGATGGTCATGTGAT TGGTACGGCCAAACACTTCCTTGGCGATGGCGGTACCGATAAAGGCGACGATCAGGGAAATAA TCTTGCCAGCGAAGATGCACTTATCCGTTTACACGCGCAGGGTTATGTCGCGGCAATTGAAGC CGGTGTGCAAACGATTATGGCTTCCTTTAATAGCTGGCACGGTTTGAAAATGCACGGCAATCAC TACCTGTTAACCGAAGTGTTGAAAAATCGTATGGGCTTCGATGGCTTTGTGGTTGGCGATTGGA ATGGACACGGCCAGGTTGACGGTTGTACCAATATCAGTTGCGCGGCGTCTATCAATGCGGGTG TGGATATGATTATGGTACCGGACGACTGGCAGGGCATGTACGAAAATACTGTCGCGCAAGTGA AGTCTGGTGAAATTTCCATGGCGCGTCTGGACGATGCCGTACGTAGAATTTTACGTGTAAAATT TCGTGCAGGCCTGTTTGACGATGTCGGCGCGCCGTCGACTCGTGCCTACGCAGGCAAAGCCG ATGTGCTGGCCAGTGACGCACATCGCGAAATCGCCCGCCAAGCTGTGCGCGAATCACTGGTG CTATTAAAAAATAAAGGCGCGTTATTGCCAATTTCGCCAACGGCCAATATTTTGGTTGCGGGCG ATGGTGCTGACAACATAGGTAAACAAAGTGGTGGCTGGACCATAACCTGGCAAGGTACCGGCA ATACCAATGCTGATTTCCCCAACGGCAGTTCTATTTACGCCGGGCTCGCTGCTCAGGTGGAGC AGGCGGGCGGTAAAATTACCCTGAGTGCGGATGGAAGTTTCGGCCAAAAGCCCGATGTGGCC ATCGTGGTGTTCGGCGAAAATCCCTATGCGGAGGGGCAAGGCGATCTTGGCTCGTTGGAATAC CAGGTTAATTCCCACAGCGATTTAGCACTGCTTAAAAAATTGAAAGCCGCTGGAATTCCGGTGG TATCGGTATTTCTGACCGGTCGCCCCTTGTGGATAAATCCAGAGCTGAATGCATCTGATGCATT CGTAGTTGCCTGGTTACCCGGGAGTGAAGGCGGTGCAGTGGCAGATGTGTTGTTGCGTAATTC CGCTGACAAAGTACAAACTGACTTTTCCGGTAAACTCAGTTATTCCTGGCCTGCCCACGAGTAT CAGTTGGCTAATCGCGGGGACAAACAAACGCCATTGTTCGCTTATGGTTATGGCTTAAATTACC GTACTACGGATAAAACCGTTTACCCGCTGCCCGAAGAACGGCAAACCAATTTAGGGGATGCCG CGCGTGGTGTTGATTTATTCCGGGGTCGCGCGTTGCCGCCTTGGCAAATGACGCTGCTATCCG GCACGCAGAGTAAGCCGATGGATTCCAGCGTGGTGACGTTAGACGGTCTTACCGTACGTACCT CCGATAGAGATGTTCAGGAGGATGCGCGGGAACTTAGCTGGCAGGGGGGCTCGACAGCAAGT CTCGCACTGACCGCTGAAAAAATGCAGGATCTTGCGGTGAATCTCAGTAACAACAGCTCACTAT TTTTCGATCTTAAATTGCAAACTGAAGGGCCGAAAAATTTGTTGCTGGGAATGACTTGTGGCGG CACCTGCGGCGGAAACCTGCTGTTGGATTCCGTGTTGGCGCCATTGAGTACTGGCGAGTGGC ACACGGTCGAAATCGATTTGCAGTGCTTGAATCTGGCGGGCGTGGAGTTTTATCACATTGACTC CCCTTTCCGATTAAGCTCGCAAGGCGACGCGAAATCTCTACAATTAAGTAATATCCGTATCAGT GCCAAATCTGGCACACCTTTAACGTGTCCGGGAGAAGGTTCG
YoaJ_O
AGCCAGCAGCATATGACGACCTGCATGAAGGTTATGCAACGTATACAGGGTCAGGCTATTCAG GAGGAGCTTTCCTGCTGGATCCCATTCCTTCCGATATGGAGATTACTGCAATAAATCCGGCGGA TCTCAATTACGGAGGAGTAAAAGCGGCACTTGCCGGCTCTTATTTGGAAGTTGAAGGGCCAAA AGGGAAAACAACCGTATATGTTACTGATCTTTATCCCGAAGGCGCTCGGGGAGCTCTTGATCTG TCACCTAATGCCTTCCGTAAAATCGGCAATATGAAAGACGGAAAAATCAATATTAAATGGCGTGT TGTCAAAGCCCCAATCACCGGCAATTTCACGTACCGGATCAAAGAAGGCAGCAGCAGGTGGTG GGCAGCAATCCAAGTCAGAAATCACAAGTATCCTGTTATGAAAATGGAATATGAAAAGGATGGT AAGTGGATCAACATGGAGAAAATGGACTATAACCATTTTGTGAGTACGAATTTAGGTACTGGCT CTCTCAAAGTCAGAATGACTGACATCCGCGGAAAAGTTGTGAAAGACACCATTCCAAAGCTGCC TGAAAGCGGAACGTCCAAAGCCTATACAGTACCGGGCCATGTTCAGTTTCCTGAAGGTTCG

Table 7.8. Sequences of additional JUMP parts coding for cellulolytic proteins that were synthesised and cloned into pJUMP18-Uac. Sequences include both sides fusion sites

Cexopt_O
AGCCGCAACCACTCTTAAAGAAGCCGCGGATGGCGCAGGGCGCGATTTTGGATTTGCCTTAGA TCCGAATCGTCTGAGTGAAGCCCAATATAAAGCAATTGCTGATTCCGAATTTAATTTGGTGGTG GCGGAAAATGCAATGAAATGGGATGCAACAGAACCTTCTCAAAATAGTTTCTCGTTTGGAGCCG GAGATCGTGTGGCCTCATATGCGGCGGATACTGGTAAAGAATTATATGGGCATACACTGGTGT GGCATTCACAATTGCCAGATTGGGCCAAGAATCTGAATGGAAGCGCATTTGAATCAGCAATGGT TAATCATGTTACAAAAGTCGCGGATCATTTTGAAGGGAAAGTTGCCAGCTGGGATGTGGTGAAT GAAGCCTTTGCGGATGGAGGAGGTCGTCGCCAAGATTCAGCATTTCAACAAAAGTTGGGTAAT GGATATATTGAAACGGCCTTTCGTGCCGCTCGCGCCGCCGATCCCACGGCGAAACTCTGTATT AATGATTATAATGTGGAAGGAATTAATGCCAAAAGTAATAGCCTGTATGATCTGGTTAAAGATTT TAAAGCCCGTGGAGTTCCACTGGATTGTGTGGGATTTCAAAGTCATCTGATTGTTGGTCAAGTC CCTGGAGATTTTCGTCAAAATTTGCAGCGTTTTGCAGATCTTGGGGTAGATGTCCGTATTACTG AACTGGATATTCGTATGCGTACACCTAGCGACGCCACAAAATTAGCCACGCAAGCCGCGGATT ATAAGAAAGTTGTTCAAGCGTGTATGCAAGTGACTCGTTGTCAAGGGGTAACTGTGTGGGGTAT TACGGATAAATATTCTTGGGTTCCCGATGTGTTTCCAGGTGAAGGCGCAGCACTCGTCTGGGAT GCCAGTTATGCAAAGAAACCTGCTTATGCTGCTGTCATGGAAGCGTTTGGTGCAAGTCCTACAC CTACCCCGACAACTCCTACGCCTACCCCGACCACCCCAACCCCAACGCCTACCAGTGGACCTG CTGGATGTCAAGTTTTATGGGGTGTTAATCAATGGAATACAGGTTTTACTGCCAATGTGACGGT TAAGAATACCTCATCGGCCCCTGTGGATGGTTGGACACTTACCTTTTCGTTTCCTAGTGGGCAA CAAGTGACTCAAGCTTGGTCCAGTACCGTGACACAAAGCGGTTCCGCGGTTACCGTACGTAAT GCGCCCTGGAATGGTTCAATTCCAGCGGGTGGGACGGCCCAATTTGGATTTAATGGATCACAT ACCGGTACTAATGCAGCACCTACCGCCTTTTCACTGAATGGTACCCCATGTACCGTTGGTTCG
SC1Copt_O
AGCCGCACCATTTCCAGCCGATTTCCTCTGGGGCACCGCAACGGCCGCATATCAAATTGAAGG GGCTGCTTCGGAAGATGGGCGTACCCCAAGCGTCTGGGATGTGTTTTCTAAGACGCCGGGTAA AGTCTTTGAGGGTCATACTGGTGACGTGGCATGTGATCATTATCATCGTTATAAAGAGGACGTG GGACTCCTGGCGGAACTGGGTGTGAAAAGTTATCGTTTTAGTGTATCCTGGACCCGTGTTTTGC CTGATGGAACTGGAAAAGTGAATCCAAAAGGACTGGACTTTTATGACCGTCTTGTAGATGAACT GCTTCGTGCAGGGATTGTACCGATGTGTACATTGTTTCATTGGGATTTTCCTCAAGCACTTCAA GACCGTGGTGGCTTTCTTCAACGTGATGTGGCGGATTGGTTTGCCGATTATACAACTGTTGTCG CACGCCGCTTAGGGGACCGTGTTCCATGGTGGGTAACCCAAAATGAACCTCAAGCCTTTATTG GAAATGCCCTGCTTAATGGGGTGTTTGCTCCTGGCCTGAAATTACCTTATCGCGAATATTTAAC GGCGGCGCATAATCAAATGCGCGCACATGGTAAGCAAGTTGATGCGCTCCGTGCGGCCGCGA CCGCAGCCAAAATTGGTTATGTATTGGCGACACAAGTCCAACGTCCTGCCACTGAAGATCCCG CTGATATTGAAGCTGCTCGTGAAGCCATCTTTTCCGTTTACGATAAGAATCCATGGAATAATGCA TGGTGGATTTGCCCTGTTTTGGAAGGTCGCTACCCAGAATCTGGTCTGCGTCTGTTTGGTGATG ATATGCCGGATTTTCCGGCTTCCGATTTTGATCAAATTAAACGCCCAATTGATTATCTGGGTCTT AATATGTATTCAGCTGGTACCTGGCGCCAAGGTAAAGATGGTCGTCCTGAACGTGTGGTTTCTC CACCTGGTTATCCCCGTGGTACTCTGGATTGGCTTCAGATTGTTCCGTCGACCTTGTATTGGGG TTCTCGTTACTTTTGGGAACGTTATAAACTTCCAATTGGCATTACTGAACATGGTTTGGCAACTC GTGATCAAGTGTTCTTGGATGGTAAAGTACATGATCCAAAACGTATTGATGTGATGCATCGTTAT CTTCTGGGTTTAGCTCGCGCAGTGCAAGAAGGTGTTCCAGTGGTAGGTTATTGGGCGTGGAGC CTTTTGGATAATTTTGAATGGGCAGAAGGATATAAAGACCGTTTTGGGTTAGTGTATGTGGACTA TGCCACACAACGTCGCATTCCTAAAGATTCTTTTAGTTGGTATCGCGAAGTTATTGCAACAGGC GGTCGTAGTCTGCTGGGTCCCACTGCAGTTGCATGGGATAATGTAACAGGTTCG

Table 7.8. Sequences of additional JUMP parts coding for cellulolytic proteins that were synthesised and cloned into pJUMP18-Uac. Sequences include both sides fusion sites

AMAEopt_O
AGCCCATGGTTATATTAGCAGCCCGCCATCCCGTCAAGCGAATTGTGCATCGGGTGCTGTGAG CGGGTGTGGTGATATTGTTTATGAGCCGCAATCGGTCGAAAATCCGAAAGGAAGCACCCAATG TAATGGTGGTGGTTCACGCTTTACAGTTCTTAATGATAATTCGAAAGCGTGGCCTGCCGCCAAT GTGGGTTCAAGTGTAACTTTTAATTGGGTCATTACTGCCCGCCATGCCACTAGCACGTGGGAAT ATTTTATTGGGAATAAACTTGTAGCATCGTTTAATGATAATGGGGCCCAACCAGGTGCGTCTAAA AGCCATACTGTTAACTTTTCTGGATATTCTGGTCGTCAAACTGTCCTGGCGCGTTGGAATGTGG CGGATACTATTAATGCCTTTTATTCTTGCGTAGATCTTAATTTTGGTGGTGGCACTTCAAATCCT ACCCCGACTCCAACGACACCCACTCCGACGCCGACTACTCCTGGCCCTACGTCGAGTCCGAC CCCAGGACCAACCGCAACGTCGAATCCGGGTGCTACGACCTGGGCAGCAGGTACTGCTTATAA AGTAGGGGATCGTGTTACCTATTCTGGTAAATCATATCAATGTCGCCAAGCCCATACGGCTATT ACAGGCTGGGAACCCCCATATGTGGCCGCCTTATGGACGGCGGTTTCG
AMAGopt_O
AGCCCATGGCTATGTGTCGTCACCGCCCAGTCGTCAAGCACTGTGTGCAGCGGGTACTGTCAG CGATTGTGGTGCAATCCAATTTGAACCTCAATCAGTGGAAGCCCCTAAAGGCTCCTCTCAGTGT AATGGCGGGAATGCGAATTTTAGTGTCTTGAATGATGAGTCGCGTAATTGGCCAGCCACATCTG TGGGTAATAGCGTTACATTTAATTGGGTTTTAACGGCTCGTCATCGTACGGCGACATGGGTATA CTCGATTGATGGCACAAAAGTTGCGACATTTGATGATAAGAATGCCATTCCGAATGCAACCGTC AGCCATAAAGTAGATTTGTCTGGCTTTTCTGGACGTAAAACTGTGCTGGCCGTTTGGAATATTG GTGACACTGCGGCAGCGTTTTATTCGTGTGTCGATCTTAATATTGGTGGTGGCGGCAGCTCCA CTACTCCGACGACGACCCCTACAACGGCACCCACCAAAGCTCCAACTACAGCGCCAACCACTT CCGCACCCACAACGGCACCGGGAACGACGGCCCCAACTAAAGCACCCACGAGTACGGCCACA GCCGCACCTACCTCTACAACCACAGCACCCGCCAATGGCACTGAATGGGCTCCCGGGGTTTCT TATAAAGTGGGAGATGTAGTGACGTATCAAGGAGTTAGTTATAAAACTCGTCAAGCACATACAA GCATTCGTTCGTGGGAACCGAGCATCTTTACTTTGGCGCTGTGGTTGCCTCTTGGTTCG
CFAAopt_O (+GS)
AGCCCATGGTAGCGTTACTGATCCGCCAACCCGTAATTATTCGTGTTGGGAGCGCTGGGGTTC AGATCATTTGAATCCGCAAATGGCAACCTTGGATCCTATGTGTTGGGGTGCATTTCAACATGAT CCTAATGCCATGTGGAATTGGAATGGTTTATATCGTGAAAATGTTGGCGGTCGCCATGAAGCAG TCATTCCGGATGGTCAATTATGTTCTGGTGGACGTACTTTTAGCCCTCGTTATGATTATCTTGAT ACGCCCGGCCCTTGGACAGCAAAAGCTGTTCCGGAGAAATTTACCCTTACCCTGACTGATGGT GCTAAACATGGTGCGGATTATCTCCGTATTTATGTGTCGAAACCCGGATTTGATCCAACTAAAG AAGCACTTGGGTGGGATGATATTACCCTTCTTAAAGAGACTGGACGTTATGGAACAACTGGACT GTATCAAACGGATGTAGATTTGACTGGTCGTTCAGGTCGCGCCGTATTGTTTACAATTTGGCAA GCGTCGCATCTGGATCAACCCTATTATTTGTGTAGTGATATTAATGTTGGTGGGACAACGACTA CTCCTACCCCTACCCCAACGGTAACCCCGACCGTAACCCCAACCCCAACCGTCACCCCCACCC CCACTGTTACCCCAACAGTGACCCCTACGCCTACTACCACTCCAACTCATGGAACAGGTGGAT GTACAGCCACGGTTAAAGTGGTTAATTCCTGGGGAAGCGGCTTTGTGGGAGAAGTCACGGTAA CAGCGGGAGCCAGTGCAATTACCGGTTGGCATACCTCGATTGGTGGCACTACGGTACAACAAG CCTGGTCATCGACCTTGAGCGCTGCAAATACATTAACATCAGTAGATTGGAACTCGAAATTGGC TGCCGGTGGAACGGCTACTGCTGGTTTTATTGGTTCGGGTAATGGTCAAGGGGTTACAGCGGT GTCTTGTGGTACCCATGGTTCG

7.4 Secretory part collection

Secretion partner-coding parts (Chapter 5).

Table 7.9. Sequences of additional JUMP parts coding for secretion peptides and secretion fusion partners that were cloned into pJUMP18-Uac. Sequences include both sides fusion sites

pic_Ndom_Cdom (SP01)
Ndom: AATGAATAAAGTTTATTCTCTTAAATATTGCCCCGTCACCGGGGGGCTTATTGCTGTCTCTGAAC TTGCCCGCAGGGTAATAAAAAAGACATGCCGAAGATTAACGCATATTCTTCTGGCTGGCATTCC AGCAATCTGTCTGTGTTACTCTCAGATAGCC; Cdom: TTCGTATAAAAACTTCATGACGGAAGTTAACAATCTGAACAAACGTATGGGTGACCTGCGTGAC ACAAACGGTGATGCCGGTGCCTGGGCGCGCATCATGAGTGGTGCCGGTTCTGCGGACGGTGG TTACAGTGATAATTACACCCATGTTCAGGTCGGCTTTGACAAAAAACATGAACTGGACGGTGTG GACCTGTTTACCGGTGTCACGATGACCTATACCGACAGCAGTGCAGACAGCCATGCATTCAGC GGAAAGACGAAATCGGTGGGGGGCGGTCTGTATGCTTCAGCATTGTTTGAGTCCGGTGCCTAT ATCGATTTGATTGGTAAATATATTCACCATGACAATGATTACACAGGTAACTTTGCTAGCCTGGG AACGAAACACTACAACACCCATTCCTGGTATGCCGGTGCTGAAACGGGTTACCGCTATCACCT GACAGAGGACACGTTCATTGAGCCGCAGGCTGAACTGGTTTACGGCGCCGTGTCCGGGAAAA CATTCCGCTGGAAAGACGGTGATATGGACCTGAGCATGAAGAACAGGGACTTCAGTCCGCTGG TTGGAAGAACAGGGGTTGAACTGGGCAAGACCTTCAGTGGTAAGGACTGGAGTGTGACGGCC CGTGCCGGAACCAGCTGGCAGTTTGACCTGCTGAATAATGGAGAAACCGTACTGCGTGATGCG TCCGGGGAGAAACGGATAAAAGGAGAGAAGGACAGCCGGATGCTGTTTAATGTTGGTATGAAT GCGCAGATAAAGGACAATATGCGCTTTGGTCTGGAGTTTGAGAAGTCAGCCTTTGGTAAATATA ACGTGGATAATGCGGTAAACGCGAATTTCCGGTATATGTTCTAAGCTT
pet_Ndom_Cdom (SP02)
Ndom: AATGAATAAAATATACTCCATTAAATATAGTGCTGCCACTGGCGGACTCATTGCTGTTTCTGAAT TAGCGAAAAAAGTCATATGTAAAACAAACCGAAAAATTTCTGCTGCATTATTATCTCTGGCAGTT ATTAGTTATACTAATATAATATATGCAGCC; Cdom: TTCGTATAAAGCCTTCCTTGCAGAGGTCAACAACCTCAACAAACGTATGGGTGATCTGCGTGAC ATTAACGGTGAGGCCGGTGCATGGGCCCGTATCATGAGTGGAACCGGGTCTGCCGGCGGTGG ATTCAGTGACAACTACACCCACGTTCAGGTCGGTGCGGATAACAAACATGAACTCGATGGCCTT GACCTCTTCACCGGGGTGACCATGACCTATACCGACAGCCATGCAGGCAGTGATGCCTTCAGT GGTGAAACGAAGTCTGTGGGTGCCGGCCTCTATGCCTCTGCCATGTTTGAGTCCGGAGCATAT ATCGACCTCATCGGTAAGTACGTTCACCATGACAACGAGTATACCGCAACTTTCGCCGGCCTTG GCACCAGAGACTACAGCTCCCACTCCTGGTATGCCGGTGCGGAAGTCGGTTACCGTTACCATG TAACTGACTCTGCATGGATTGAGCCGCAGGCGGAACTTGTTTACGGTGCTGTATCCGGGAAAC AGTTCTCCTGGAAGGACCAGGGAATGAACCTCACCATGAAGGATAAGGACTTTAATCCGCTGAT TGGGCGTACCGGTGTTGATGTGGGTAAATCCTTCTCCGGTAAGGACTGGAAAGTCACAGCCCG CGCCGGCCTTGGCTACCAGTTTGACCTGTTTGCCAACGGTGAAACTGTACTGCGTGATGCGTC CGGTGAAAAACGTATCAAAGGTGAAAAAGACGGCCGTATGCTCATGAATGTTGGTCTGAATGCT GAGATTCGTGACAACGTACGCTTTGGTCTTGAGTTTGAGAAATCGGCATTTGGTAAGTACAACG TGGATAACGCCATCAACGCCAACTTCCGTTACTCCTTCTAAGCTT

Table 7.9. Sequences of additional JUMP parts coding for secretion peptides and secretion fusion partners that were cloned into pJUMP18-Uac. Sequences include both sides fusion sites

CelCD_Ndom (SP03)
AATGCTCAGAAAGAAAACAAAGCAGTTGATTTCTTCCATTCTTATTTTAGTTTTACTTCTATCTTT ATTTCCGACAGCTCTTGCAGCAGAAGGAAACACTCGTGAAGATAATTTTAAACATTTATTAGGTA ATGACAATGTTAAACGCCCTTCTGAGGCTGGCGCATTACAATTACAAGAAGTCGATGGACAAAT GACATTAGTAGATCAACATGGAGAAAAAATTCAATTACGTGGAATGAGTACACACGGATTACAAT GGTTTCCTGAGATCTTGAATGATAACGCATACAAAGCTCTTGCTAACGATTGGGAATCAAATATG ATTCGTCTAGCTATGTATGTCGGTGAAAATGGCTATGCTTCAAATCCAGAGCTGATTAAAAGCA GAGTCATTAAAGGAATAGATCTTGCTATTGAAAATGACATGTATGTCATCGTTGATTGGCATGTA CATGCACCTGGTGATCCTAGAGATCCCGTTTACGCTGGAGCAGAAGATTTCTTTAGAGATATTG CAGCATTATATCCTAACAATCCACACATTATTTATGAGTTAGCGAATGAGCCAAGTAGTAACAAT AATGGTGGAGCTGGGATTCCAAATAATGAAGAAGGTTGGAATGCGGTAAAAGAATACGCTGAT CCAATTGTAGAAATGTTACGTGATAGCGGGAACGCAGATGACAATATTATCATTGTGGGTAGTC CAAACTGGAGTCAGCGTCCTGACTTAGCAGCTGATAATCCAATTGATGATCACCATACAATGTA TACTGTTCACTTCTACACTGGTTCACATGCTGCTTCAACTGAAAGCTATCCGCCTGAAACTCCTA ACTCTGAAAGAGGAAACGTAATGAGTAACACTCGTTATGCGTTAGAAAACGGAGTAGCAGTATT TGCAACAGAGTGGGGAACTAGCCAAGCAAATGGAGATGGTGGTCCTTACTTTGATGAAGCAGA TGTATGGATTGAGTTTTTAAATGAAAACAACATTAGCTGGGCTAACTGGTCTTTAACGAATAAAA ATGAAGTATCTGGTGCATTTACACCATTCGAGTTAGGTAAGTCTAACGCAACAAGTCTTGACCC AGGGCCAGACCAAGTATGGGTACCAGAAGAGTTAAGTCTTTCTGGAGAATATGTACGTGCTCG TATTAAAGGTGTGAACTATGAGCCAATCGACCGTACAAAATACACGAAGTACTTGGCGGAGCC
XynA_Ndom (SP04)
AATGTTTAAGTTTAAAAAGAAATTCTTAGTTGGATTAACGGCAGCTTTCATGAGTATCAGCATGT TTTCGGCAACCGCCTCTGCGGCTGGCACAGATTACTGGCAAAATTGGACAGATGGGGGCGGA ACAGTCAACGCAGTCAATGGATCTGGCGGGAATTACAGTGTTAATTGGTCTAATACCGGAAATT TCGTTGTTGGTAAAGGCTGGACTACAGGCTCGCCATTTAGAACAATAAACTATAATGCCGGAGT CTGGGCGCCGAATGGCAATGGATATTTGACTTTATATGGCTGGACGAGAGCACCTCTCATCGA ATATTATGTAGTGGATTCATGGGGTACTTACAGACCTACCGGAACGTATAAAGGTACTGTAAAG AGTGATGGAGGTACATATGACATATATACAACAACACGTTATAACGCACCTTCCATCGATGGCG ATAACACTACTTTTACGCAGTACTGGAGTGTTCGCCAGTCGAAGCGCCCGACCGGAAGCAACG CTGCAATCACTTTCAGCAATCATGTTAACGCATGGAAGAGCCATGGAATGAATCTGGGCAGTAA TTGGGCTTATCAAGTCTTAGCGACAGAAGGATATAAAAGCAGCGGAAGTTCTAATGTAACGTGT GGGGCGGAGCC
CenAo_N31 (SP05)
AATGAGCACCCGTCGTACCGCAGCAGCACTGCTGGCAGCAGCAGCCGTTGCAGTTGGTGGTC TGACCGCACTGACCACCACCGCAGCACAGGCAGCC
CenAw_N31 (SP06)
AATGTCCACCCGCAGAACCGCCGCAGCGCTGCTGGCGGCCGCGGCCGTCGCCGTCGGCGGT CTGACCGCCCTCACCACCACCGCCGCGCAGGCAGCC
PelB_N (SP07)
AATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCGAT GGCAGCC
PelB Asp5_N (SP08)
AATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCGAT GGCAGATGATGACGATGATGGAGCC
KP-SP_N (SP09)
ATGAGCAGACGAGCCCCCCTCCTCCGCGCCGGAGCGGCCACCGCAGTCGCCGTCCTGTACCT CACCGCGGTCCCCCAGGCCGCGAGCGCAGCC

Table 7.9. Sequences of additional JUMP parts coding for secretion peptides and secretion fusion partners that were cloned into pJUMP18-Uac. Sequences include both sides fusion sites

XynA-SP_N (SP10)
ATGTTTAAGTTTAAAAAGAAATTCTTAGTTGGATTAACGGCAGCTTTCATGAGTATCAGCATGTTT TCGGCAACCGCCTCCGCAGCC
Cel-CD-N29_N (SP11)
AATGCTCAGAAAGAAAACAAAGCAGTTGATTTCTTCCATTCTTATTTTAGTTTTACTTCTATCTTT ATTTCCGACAGCTCTTGCAGCC
Cel-CD-G20_N (SP12)
AATGGAAGGAAACACTCGTGAAGATAATTTTAAACATTTATTAGGTAATGACAATGTTAAACGCA AACGCGGAGCC
Cel-CD-Gdom_N (SP13)
AATGGAAGGAAACACTCGTGAAGATAATTTTAAACATTTATTAGGTAATGACAATGTTAAACGCC CTTCTGAGGCTGGCGCATTACAATTACAAGAAGTCGATGGACAAATGACATTAGTAGATCAACA TGGAGAAAAAATTCAATTACGTGGAATGAGTACACACGGATTACAATGGTTTCCTGAGATCTTG AATGATAACGCATACAAAGCTCTTGCTAACGATTGGGAATCAAATATGATTCGTCTAGCTATGTA TGTCGGTGAAAATGGCTATGCTTCAAATCCAGAGCTGATTAAAAGCAGAGTCATTAAAGGAATA GATCTTGCTATTGAAAATGACATGTATGTCATCGTTGATTGGCATGTACATGCACCTGGTGATCC TAGAGATCCCGTTTACGCTGGAGCAGAAGATTTCTTTAGAGATATTGCAGCATTATATCCTAACA ATCCACACATTATTTATGAGTTAGCGAATGAGCCAAGTAGTAACAATAATGGTGGAGCTGGGAT TCCAAATAATGAAGAAGGTTGGAATGCGGTAAAAGAATACGCTGATCCAATTGTAGAAATGTTA CGTGATAGCGGGAACGCAGATGACAATATTATCATTGTGGGTAGTCCAAACTGGAGTCAGCGT CCTGACTTAGCAGCTGATAATCCAATTGATGATCACCATACAATGTATACTGTTCACTTCTACAC TGGTTCACATGCTGCTTCAACTGAAAGCTATCCGCCTGAAACTCCTAACTCTGAAAGAGGAAAC GTAATGAGTAACACTCGTTATGCGTTAGAAAACGGAGTAGCAGTATTTGCAACAGAGTGGGGAA CTAGCCAAGCAAATGGAGATGGTGGTCCTTACTTTGATGAAGCAGATGTATGGATTGAGTTTTT AAATGAAAACAACATTAGCTGGGCTAACTGGTCTTTAACGAATAAAAATGAAGTATCTGGTGCAT TTACACCATTCGAGTTAGGTAAGTCTAACGCAACAAGTCTTGACCCAGGGCCAGACCAAGTATG GGTACCAGAAGAGTTAAGTCTTTCTGGAGAATATGTACGTGCTCGTATTAAAGGTGTGAACTAT GAGCCAATCGACCGTACAAAATACACGAAAGTACTTGGCGGAGCC

Table 7.9. Sequences of additional JUMP parts coding for secretion peptides and secretion fusion partners that were cloned into pJUMP18-Uac. Sequences include both sides fusion sites

ag43_Ndom_Cdom (SP14)
Ndom: AATGAAACGACATCTGAATACCTGCTACAGGCTGGTATGGAATCACATGACGGGCGCTTTCGT GGTTGCCTCCGAACTGGCCCGCGCACGGGGTAAACGTGGCGGTGTGGCGGTTGCACTGTCTC TTGCCGCAGTCACGTCACTCCCGGTGCTAGCC; Cdom: TTCGGGCGGTGTACTGCTGGCCGATTCCGGTGCCGCTGTCAGTGGTACCAATAACGGCGCCA TACTTACCCTTTCCGGtAAGACGGTGAACAACGATACCCTGACCATCCGTGAAGGtGATGCACTC CTaCAGGGAGGCTCTCTCACCGGTAACGGCAGCGTGGAAAAATCAGGAAGTGGCACACTCACT GTCAGCAACACCACACTCACCCAGAAAGCCGTCAACCTGAATGAAGGCACGCTGACGCTGAAC GACAGTACCGTCACCACGGATGTCATTGCTCAGCGCGGTACAGCCCTGAAGCTGACCGGCAG CACTGTGCTGAACGGTGCCATTGACCCCACGAATGTCACTCTCGCCTCCGGTGCCACCTGGAA TATCCCCGATAACGCCACGGTGCAGTCGGTGGTGGATGACCTCAGCCATGCCGGACAGATTCA TTTCACCTCCACCCGCACAGGGAAGTTCGTACCGGCAACCCTGAAAGTGAAAAACCTGAACGG ACAGAATGGCACCATCAGCCTGCGTGTACGCCCGGATATGGCACAGAACAATGCTGACAGACT GGTCATTGACGGCGGCAGGGCAACCGGAAAAACCATCCTGAACCTGGTGAACGCCGGCAACA GTGCGTCGGGGCTGGCGACCAGCGGTAAGGGTATTCAGGTGGTGGAAGCCATTAACGGTGCC ACCACGGAGGAAGGGGCCTTTGTCCAGGGGAACAGGCTGCAaGCCGGTGCCTTTAACTACTC CCTCAACCGGGACAGTGATGAGAGCTGGTATCTGCGCAGTGAAAATGCTTATCGTGCAGAAGT CCCCCTGTATGCCTCCATGCTGACACAGGCAATGGACTATGACCGGATTGTGGCAGGCTCCCG CAGCCATCAGACCGGTGTAAATGGTGAAAACAACAGCGTCCGTCTtAGCATTCAGGGCGGTCAT CTCGGTCACGATAACAATGGCGGTATTGCCCGTGGGGCCACGCCGGAAAGCAGCGGCAGCTA TGGATTCGTCCGTCTGGAGGGTGACCTGATGAGAACAGAGGTTGCCGGTATGTCTGTGACCGC GGGGGTATATGGTGCTGCTGGCCATTCTTCCGTTGATGTTAAGGATGATGACGGCTCCCGTGC CGGCACGGTCCGGGATGATGCCGGCAGCCTGGGCGGATACCTGAATCTGGTACACACGTCCT CCGGCCTGTGGGCTGACATTGTGGCACAGGGAACCCGCCACAGCATGAAAGCGTCATCGGAC AATAACGACTTCCGCGCCCGGGGCTGGGGCTGGCTGGGCTCACTGGAAACCGGTCTGCCCTT CAGTATCACTGACAACCTGATGCTGGAGCCACAACTGCAaTATACCTGGCAGGGACTTTCCCTG GATGACGGTAAGGACAACGCCGGTTATGTGAAGTTCGGGCATGGCAGTGCACAACATGTGCGT GCCGGTTTCCGTCTGGGCAGCCACAACGATATGACCTTTGGCGAAGGCACCTCATCCCGTGCC CCCCTGCGTGACAGTGCAAAACACAGTGTGAGTGAATTACCGGTGAACTGGTGGGTACAGCCT TCTGTTATCCGCACCTTCAGCTCCCGGGGAGATATGCGTGTGGGGACTTCCACgGCAGGCAGC GGGATGACGTTCTCTCCCTCACAGAATGGCACATCACTGGACCTGCAaGCCGGACTGGAAGCC CGTGTCCGGGAAAATATCACCCTGGGCGTTCAGGCCGGTTATGCCCACAGCGTCAGCGGCAG CAGCGCTGAAGGGTATAACGGTCAGGCCACACTGAATGTGACCTTCTGAGCTT
ML12Aw-SP_N (SP51)
AATGAAACGTCAACTTCGGGCCCTGGCCGCCGCCGGCCTGCTGGTCGCCAGCTCGCTCGTCG CCGTGGCCCTCGGCGGCAACGCCTCTGCC
AM6Bw-SP_N (SP52)
AATGTCCCGTATCCCCCTCGCCCTCGCTGCTGGCCTGCTGATAGCCACCGGAGCTGTCGCTTT CGCCGGGAACGCGTCAGCC

Table 7.9. Sequences of additional JUMP parts coding for secretion peptides and secretion fusion partners that were cloned into pJUMP18-Uac. Sequences include both sides fusion sites

OsmY_N (SP53)
AATGACTATGACAAGACTGAAGATTTCGAAAACTCTGCTGGCTGTAATGTTGACCTCTGCCGTC GCGACCGGCTCTGCCTACGCGGAAAACAACGCGCAGACTACCAATGAAAGCGCAGGGCAAAA AGTCGATAGCTCTATGAATAAAGTCGGTAATTTCATGGATGACAGCGCCATCACCGCGAAAGTG AAGGCGGCCCTGGTGGATCATGACAACATCAAGAGCACCGATATCTCTGTAAAAACCGATCAA AAAGTCGTGACCCTGAGCGGTTTCGTTGAAAGCCAGGCCCAGGCCGAAGAGGCAGTGAAAGT GGCGAAAGGCGTTGAAGGGGTGACCTCTGTCAGCGACAAACTGCACGTTCGCGACGCTAAAG AAGGCTCGGTGAAGGGCTACGCGGGTGACACCGCCACCACCAGTGAAATCAAAGCCAAACTG CTGGCGGACGATATCGTCCCTTCCCGTCATGTGAAAGTTGAAACCACCGACGGCGTGGTTCAG CTCTCCGGTACCGTCGATTCTCAGGCACAAAGTGACCGTGCTGAAAGTATCGCCAAAGCGGTA GATGGTGTGAAAAGCGTTAAAAATGATCTGAAAACTAAGGGCAGCGGATCAGCC
YebF_N (SP54)
AATGAAAAAAAGAGGGGCGTTTTTAGGGCTGTTGTTGGTTTCTGCCTGCGCATCAGTTTTCGCT GCCAATAATGAAACCAGCAAGTCGGTCACTTTCCCAAAGTGTGAAGATCTGGATGCTGCCGGA ATTGCCGCGAGCGTAAAACGTGATTATCAACAAAATCGCGTGGCGCGTTGGGCAGATGATCAA AAAATTGTCGGTCAGGCCGATCCCGTGGCTTGGGTCAGTTTGCAGGACATTCAGGGTAAAGAT GATAAATGGTCAGTACCGCTAACCGTGCGTGGTAAAAGTGCCGATATTCATTACCAGGTCAGC GTGGACTGCAAAGCGGGAATGGCGGAATATCAGCGGCGTTCAGCC
Spy_N (SP55)
AATGCGTAAATTAACTGCACTGTTTGTTGCCTCTACCCTGGCTCTTGGCGCGGCTAACCTGGCC CATGCCGCAGACACCACTACCGCAGCACCGGCTGACGCGAAGCCGATGATGCACCACAAAGG CAAGTTCGGTCCGCATCAGGACATGATGTTCAAAGACCTGAACCTGACCGACGCGCAGAAACA GCAGATCCGCGAAATCATGAAAGGCCAGCGTGACCAGATGAAACGTCCGCCGCTGGAAGAAC GCCGCGCAATGCATGACATCATTGCCAGCGATACCTTCGATAAAGTAAAAGCTGAAGCGCAGA TCGCAAAAATGGAAGAACAGCGCAAAGCTAACATGCTGGCGCACATGGAAACCCAGAACAAAA TTTACAACATCCTGACGCCGGAACAGAAAAAGCAATTTAATGCTAATTTTGAGAAGCGTCTGAC AGAACGTCCAGCGGCAAAAGGTAAAATGCCTGCAACTGCTGAATCAGCC

Table 7.10. Sequences of additional JUMP parts coding for secretion peptides obtained via DNA synthesis and not cloned into vector format. Sequences include both sides fusion sites

G1_N (SP15)
AATGAACGATTTAAATGATTTTTTGAAAACGATTTCATTAAGCTTTATCTTTTTCTTGCTTCTTTCT TTACCTACTGTTGCGGAAGCC
G1-M5_N (SP16)
AATGAACGATTTAAATGATTTTTTGAAAACGATTTCATTAAGCTTTGGCTTTTTCTTGCTTCTTTCT TTACCTACTGTTGCGGAAGCC
GAP_N (SP17)
AATGATTGGTAAAAAGAAAGGTCACTCGTTTCTTCAGTCAATTGGTTTTACTGTACTTGCATCTG CTTTGTTTGTTTTCGCAAACGCAGAATCGGCATCAGCC
CSP_N (SP18)
AATGAGTATCCAATGGAATAAAGTATGTCTCCTACTAGGGGCAACCTTTTTATTTCTGTCTGTAG TTTTCCCATCAGTTAGTAAGGCAGCC
Ble1853_N (SP19)
AATGAGGTATCAGAAAGAATTTTTTGCAACTGCAATGCTCTCGTTTATGGTGTTAAGTTTGCAAC AACCCGCTTTAGCC

Table 7.10. Sequences of additional JUMP parts coding for secretion peptides obtained via DNA synthesis and not cloned into vector format. Sequences include both sides fusion sites

Ble1848_N (SP20)
AATGAAAAAAATCAAAGTATTAACCATTACGGCTCTAGCATTTGCACTGATTACTAGTGGGAATACGGCATTAGCC
ECDS_N (SP21)
AATGAGAAAAGTGTTGAAAGTGGGGGTTACGCTTGCTGTATTGACCTCAGTAAGTTATCTGTTAAGTACGGATAAGCAAGTATATGCAGCC
LytE_N (SP22)
AATGAGCAAATCGAATCTAAACAAATTTTTATTCTCTTCAGCTGTTGTCGCTGGAGTTGTAGCAGTTGCTCCACAAGTGTCAGAAGCC
Chit_N (SP23)
AATGAGAGTAAAACGTACTTCTCTACTATTAATCGTTATAACGTTATTATTTTTCAATGGAAGTGATCAAGCACTATGGGCAGCC
BGL_N (SP24)
AATGAAGAAAGTAATAGGAGCACTATCGATAGCTGCTTGCACCACTTTGTTCGCCACTTCATATGCCGGAGCC
BGL(4)_N (SP25)
AATGAAAAAAAGATGGTTCTGGTTCGTTTTAGCACTTATTGTACTCTTTCCGTTTCAATCAAGTTTCGCCAATCAAGCC
MEP_N (SP26)
AATGAAAAAAACGACAAATCTTAACTGTACGAAATCACTATCTGTCGTCGCGCTATCATTTTGTATGGTGGCTAGTTCTTTTACTTCCGTGCAGTACGTAAAAGCC
TNPP_N (SP27)
AATGGAGAAATGGAAAAAGTGGGTAGGTGCTAGTCTCGTCGCATTGACTCTCCCGTTTAGTGTAAATGGAGAGGCGAAAGCC
Ble3542_N (SP28)
AATGAAAAAAGCTTGTTCTTTATTTGTTGCTGTTATAGTGTTTTTATTATTTATGCCATATTCTTTCGCATCATCAGCC
Ble3571_N (SP29)
AATGAAAAAGGTATTTAGCACATGTCTGATTCTGTTTTTTCTTTTCTTGTCTAAAGACGCTTATGCAGCC
lamB_N (SP30)
AATGATGATTACTCTGCGCAAACTTCCTCTGGCGGTTGCCGTCGCAGCGGGCGTAATGTCTGCTCAGGCAATGGCAGCC
MalE_N (SP31)
AATGAAAATAAAAACAGGTGCACGCATCCTCGCATTATCCGCATTAACGACGATGATGTTTTCCGCCTCGGCTCTCGCAGCC
OmpA_N (SP32)
AATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCAAGCC

Table 7.10. Sequences of additional JUMP parts coding for secretion peptides obtained via DNA synthesis and not cloned into vector format. Sequences include both sides fusion sites

OmpC_N (SP33)
AATGAAAGTTAAAGTACTGTCCCTCCTGGTCCCAGCTCTGCTGGTAGCAGGCGCAGCAAACGCAGCC
OmpF_N (SP34)
AATGATGAAGCGCAATATTCTGGCAGTGATCGTCCCTGCTCTGTTAGTAGCAGGTACTGCAAACGCAGCC
OmpT_N (SP35)
AATGCGGGCGAAACTTCTGGGAATAGTCCTGACAACCCCTATTGCGATCAGCTCTTTTGCAGCC
PhoA_N (SP36)
AATGAAACAAAGCACTATTGCACTGGCACTCTTACCGTTACTGTTTACCCCTGTGACAAAAGCC
DsbA_N (SP37)
AATGAAAAAGATTTGGCTGGCGCTGGCTGGTTTAGTTTTAGCGTTTAGCGCATCAGCC
SfmC_N (SP38)
AATGATGACTAAAATAAAGTTATTGATGCTCATTATATTTTATTTAATCATTTCGGCCAGCGCCCATGCAGCC
TolB_N (SP39)
AATGAAGCAGGCATTACGAGTAGCATTTGGTTTTCTCATACTGTGGGCATCAGTTCTGCATGCAGCC
TorT_N (SP40)
AATGCGCGTACTGCTATTTTTACTTCTTTCCCTTTTCATGTTGCCGGCATTTTCAGCC
Vrg-6_N (SP41)
AATGAAAAAGTGGTTCGTTGCTGCCGGCATCGGCGCTGCCGGACTCATGCTCTCCAGCGCCGCCCTGGCAGCC
pIII_N (SP42)
AATGAAAAAATTATTATTCGCAATTCCTTTAGTTGTTCCTTTCTATTCTCACTCAGCC
Cex-SP_N (SP43)
AATGATTCAGCACCTCCCGCGGACGGGCCCCCACGTCACAGGGTGCACCCGGCACTGGCTCGACGAGGAGGACATCATGCCTAGGACCACGCCCGCACCCGGCCACCCGGCCCGCGGCGCCCGCACCGCTCTGCGCACGACGCTCGCCGCCGCGGCGGCGACGCTCGTTGTCGGCGCCACGGTCGTGCTGCCCGCCCAAGCC
TT3D-SP_N (SP44)
AATGGAGTTTACTATGTCGATGTCGCATCTGTGGGCTGCAGTTCTTTGCCTGAGTGTATTGGTAGGTGCAGCC
SC1B-SP_N (SP45)
AATGACGCGACAGCAGAGCGAAGAGGGCTTTCCGGAAGGGTTCCTGTGGGGCGCCGCCGCGGCGTCGTACCAGATCGAGGGCGCAGCC
SC1C-SP_N (SP46)
AATGGCTCTCCATCCTTCCGAGTCGCTCCCCGCGAGCAACGCCTCTCCCGCCTCGCCCTCCTCCGTGCCCCCGCCTGTGACCGCCGCCTCGGCCGGCGCGGGCCGAGCAAACGCGACAGCC

Table 7.10. Sequences of additional JUMP parts coding for secretion peptides obtained via DNA synthesis and not cloned into vector format. Sequences include both sides fusion sites

YoaJ-SP_N (SP47)
AATGAAAAAGATCATGAGTGCATTTGTTGGTATGGTTTTGTTGACGATCTTCTGTTTTTCCCCGCAAGCTTCAGCC
AMAE-SP_N (SP48)
AATGAAGCGTAGATTCACCCTGCCCGCCCTGACCGTCGCCGCCGTCACCGGATCGATGTTCGTCGCGGTTTCTCCGGCCTCAGCC
AMAG-SP_N (SP49)
AATGCGCAGGAAGATCGCCTACCCGCTGGCTACCCTCGGGATGGTCGCTGCTTCGACCGCGGTGGTCGCGTCACCCGCCCTGGCAGCC
CFAA-SP_N (SP50)
AATGTCCGTCCGTGCAAGACCGCGCACCCTCCTGGCCGCGCTCGCAGCCGTCGCCGTGGCGATCGGCGCCGCCGTCGTCGCGTCCACCCCTGCCTCAGCC

7.5 Additional results for Chapter 5

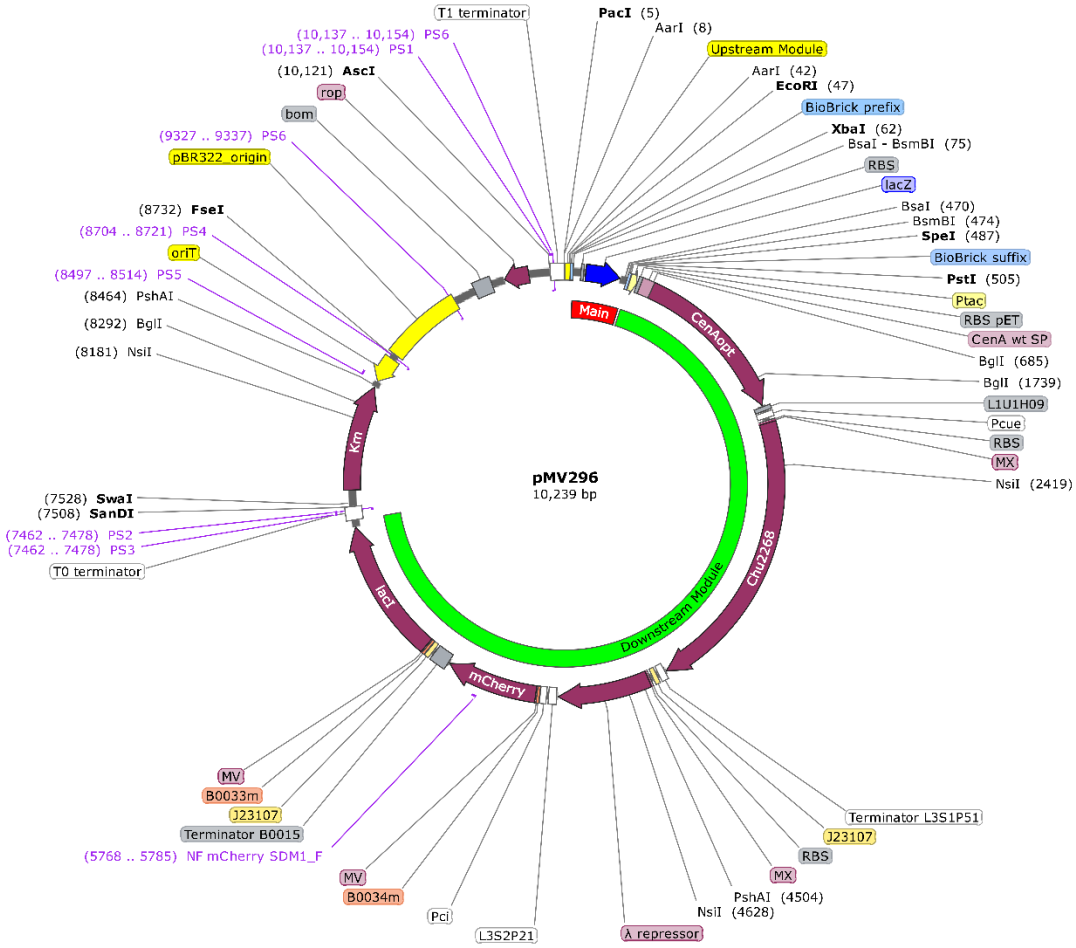


Figure 7.8. DNA map of plasmid pMV296 or pJUMP29-1A(lacZ)[Down:Mk3-W].

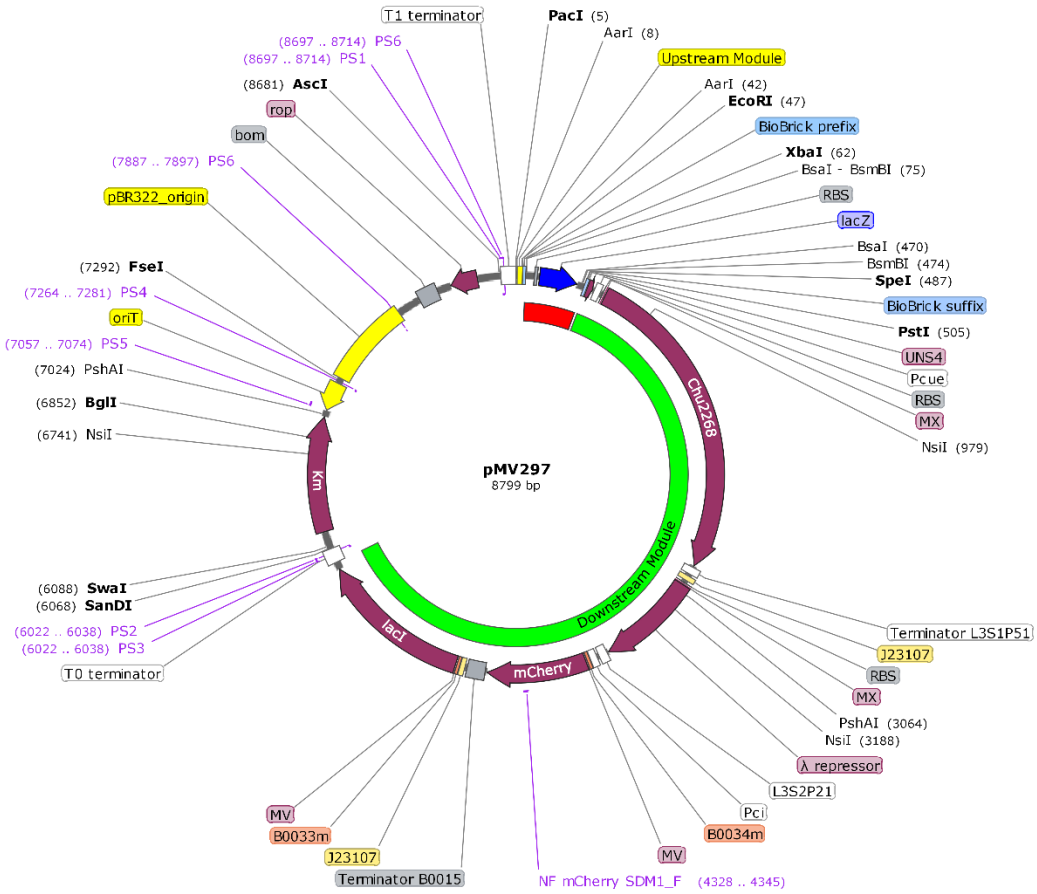


Figure 7.9. DNA map of plasmid pMV297 or pJUMP29-1A(lacZ)[Down:Mk3-U].

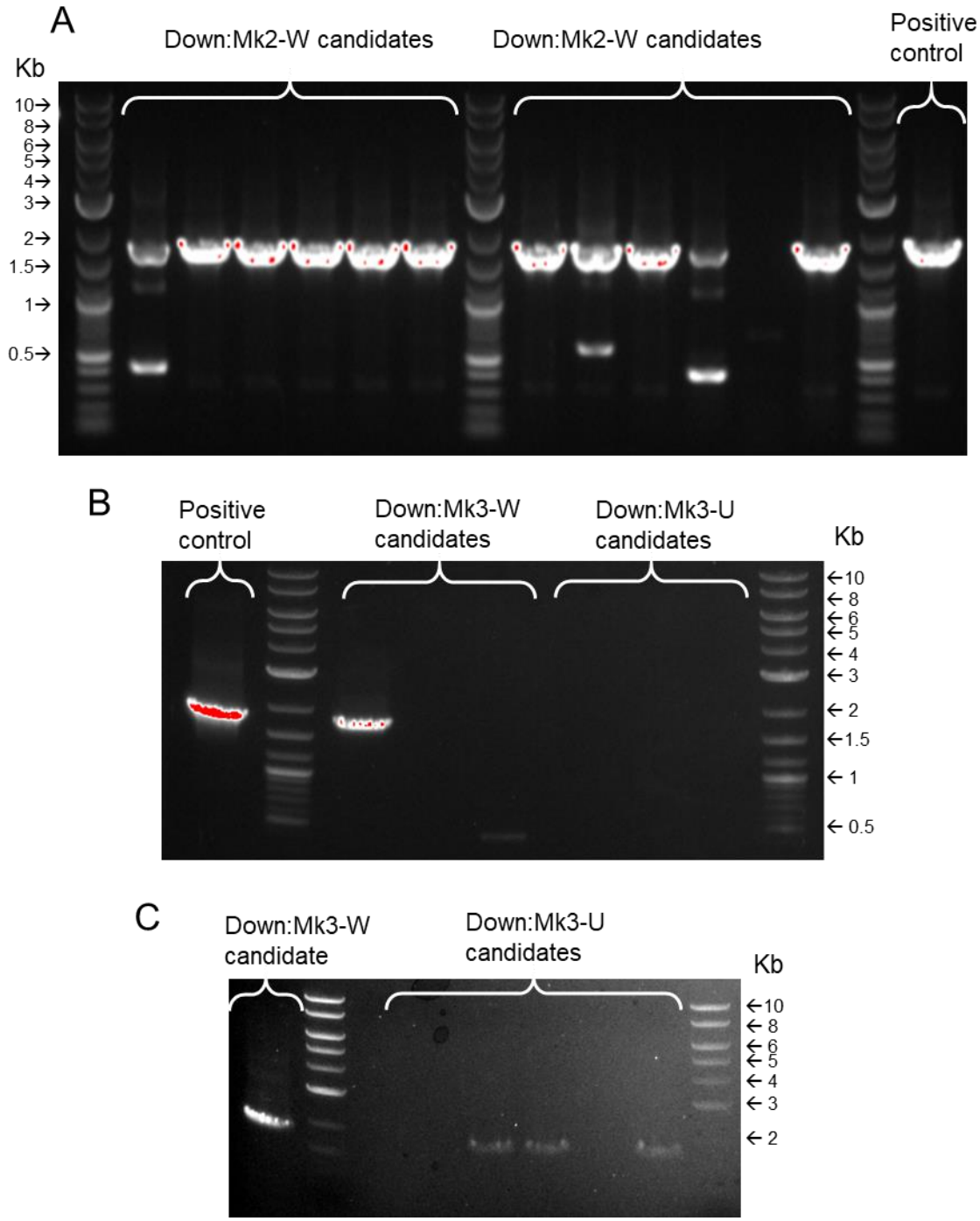
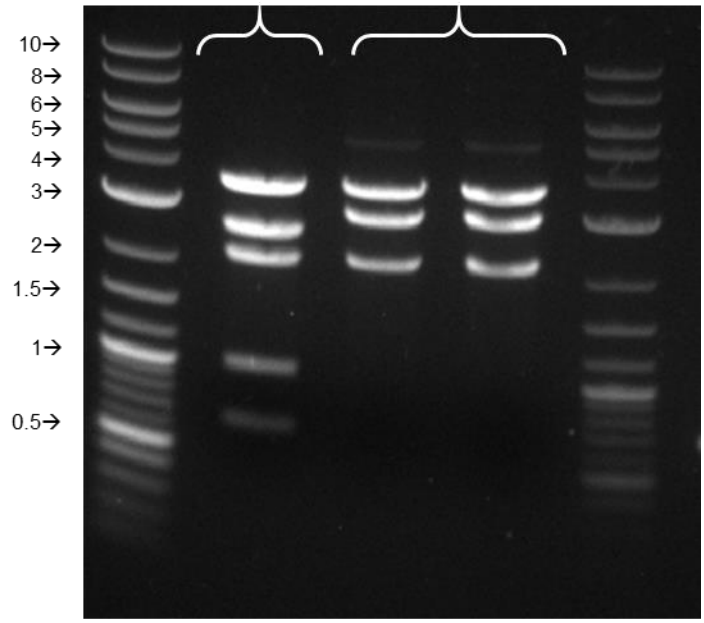


Figure 7.10. Colony PCR for Mark II and III constructs assembled in downstream site of level 1 vector. A: Mark II construct, pMV289 or pJUMP29-1A(lacZ)[Down:Mk2-W] (8 correct of 12). B and C: Mark III constructs, pMV296 or pJUMP29-1A(lacZ)[Down:Mk3-W] and pMV297 or pJUMP29-1A(lacZ)[Down:Mk3-U] (1 correct of 6 in B, 4 correct of 7 in C). Primers used: PS2 (GCGGCAACCGAGCGTTC) and “NF mCherry SDM1_F” (ACGTCTCAAACCATGGGCTGGGAGGC). Expected band size: 1767 bp.

A



B

Expected band sizes:	pMV296 BglI + NsiI	pMV297 NsiI + BglI
	1. 3553 bp	1. 3553 bp
	2. 2632 bp	2. 2926 bp
	3. 2209 bp	3. 2209 bp
	4. 1054 bp	4. 111 bp
	5. 680 bp	
	6. 111 bp	

Figure 7.11. Restriction digestion of three candidates of Mark III constructs assembled in downstream site of level 1 vector: pMV296 or pJUMP29-1A(lacZ)[Down:Mk3-W] and pMV297 or pJUMP29-1A(lacZ)[Down:Mk3-U]. A: Gel electrophoresis of plasmids digested with BglI and NsiI. B: Expected band sizes. The three clones were correct.

Individual-culture curves of experiment shown in Section 5.3.4.3.

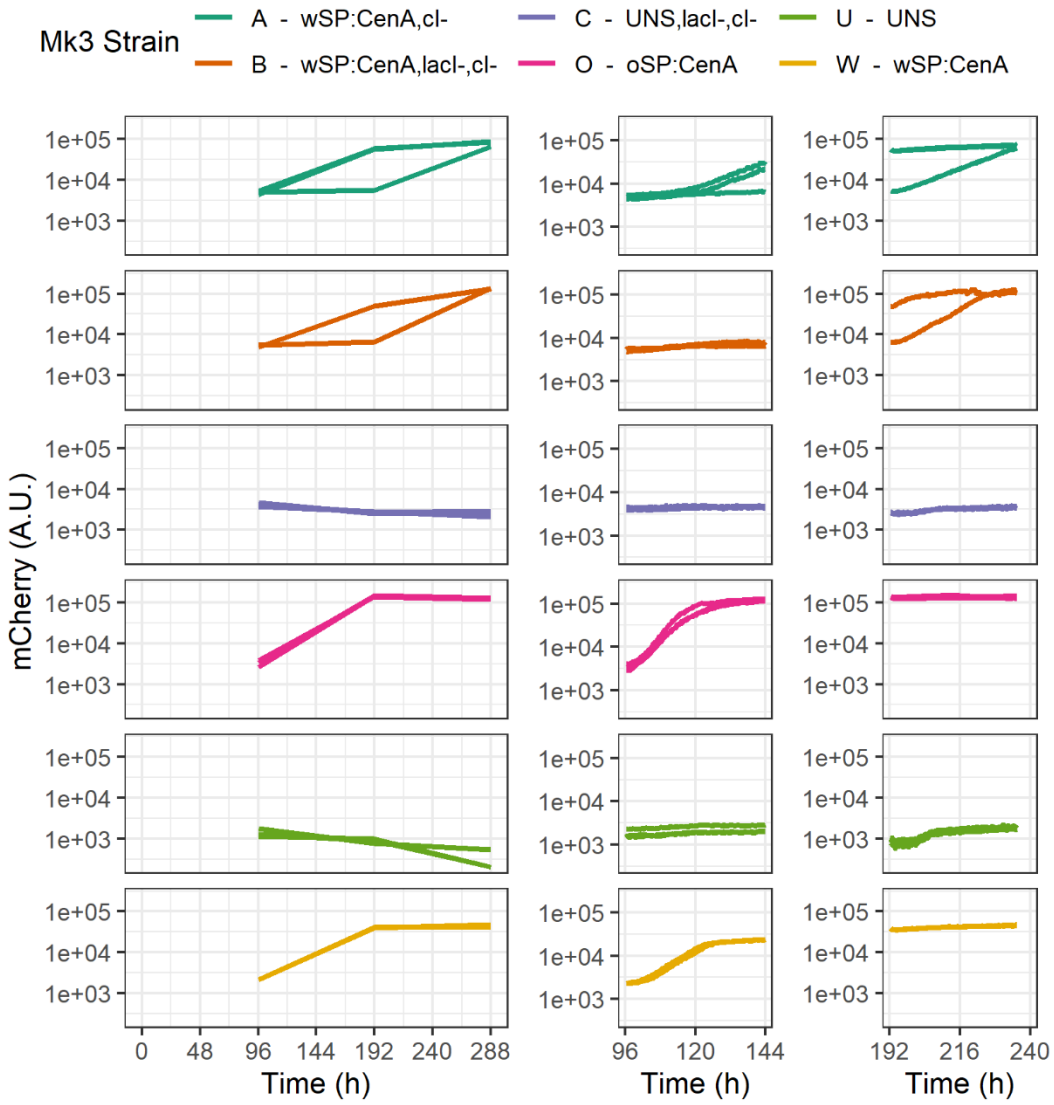


Figure 7.12. Growth of Mark III strains (A, B, C, O, W and U) in M9y medium with 1 mg / mL BSA and 1 % PASC as carbon source. Individual cultures of results shown in Figure 5.14.

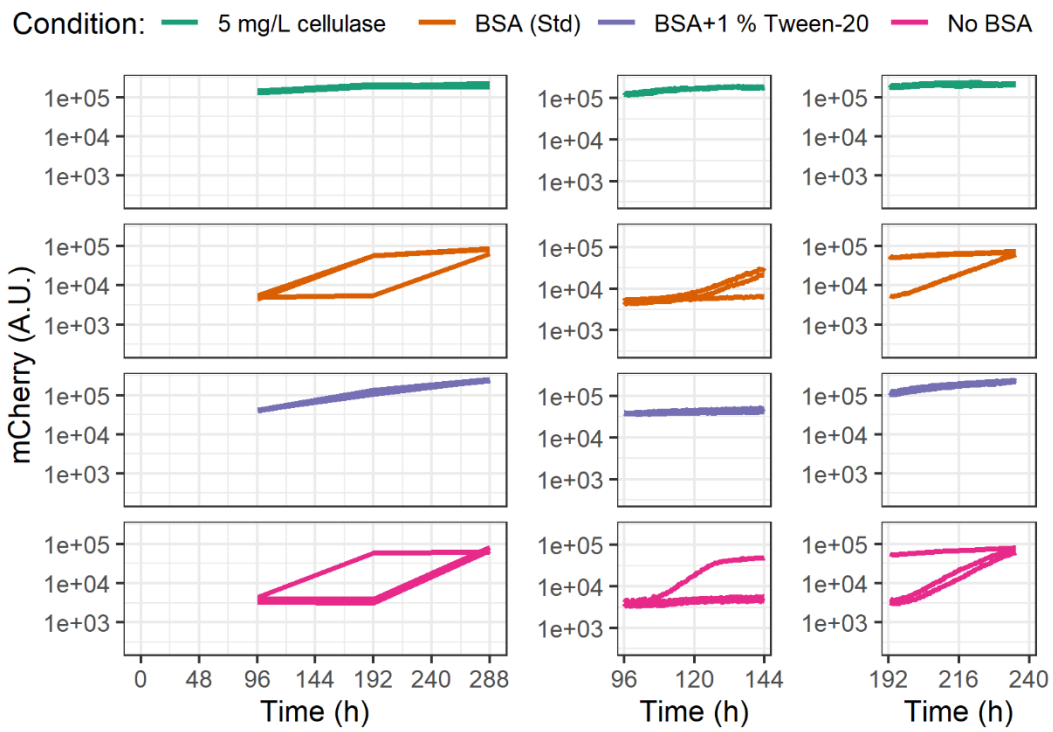


Figure 7.13. Growth of Mark III strains A in M9y medium with 1 % PASC as carbon source and different additives. Individual cultures of results shown in Figure 5.15.

References

Abdelhamid, M. A., T. Ikeda, K. Motomura, T. Tanaka, T. Ishida, R. Hirota and A. Kuroda (2016). "Application of volcanic ash particles for protein affinity purification with a minimized silica-binding tag." J Biosci Bioeng **122**(5): 633-638.

Almagro Armenteros, J. J., K. D. Tsirigos, C. K. Sønderby, T. N. Petersen, O. Winther, S. Brunak, G. von Heijne and H. Nielsen (2019). "SignalP 5.0 improves signal peptide predictions using deep neural networks." Nat Biotechnol **37**(4): 420-423.

Andreou, A. I. and N. Nakayama (2018). "Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly." PLoS One **13**(1): e0189892.

Anusree, M., V. F. Wendisch and K. M. Nampoothiri (2016). "Co-expression of endoglucanase and β-glucosidase in *Corynebacterium glutamicum* DM1729 towards direct lysine fermentation from cellulose." Bioresour Technol **213**: 239-244.

Barajas, J. F., J. M. Blake-Hedges, C. B. Bailey, S. Curran and J. D. Keasling (2017). "Engineered polyketides: Synergy between protein and host level engineering." Synth Syst Biotechnol **2**(3): 147-166.

Barth, A., J. Hendrix, D. Fried, Y. Barak, E. A. Bayer and D. C. Lamb (2018). "Dynamic interactions of type I cohesin modules fine-tune the structure of the cellulosome of *Clostridium thermocellum*." Proc Natl Acad Sci U S A **115**(48): E11274-e11283.

Bastin, J. F., Y. Finegold, C. Garcia, D. Mollicone, M. Rezende, D. Routh, C. M. Zohner and T. W. Crowther (2019). "The global tree restoration potential." Science **365**(6448): 76-79.

Beal, J., A. Goñi-Moreno, C. Myers, A. Hecht, M. D. C. de Vicente, M. Parco, M. Schmidt, K. Timmis, G. Baldwin, S. Friedrichs, P. Freemont, D. Kiga, E. Ordozgoiti, M. Rennig, L. Rios, K. Tanner, V. de Lorenzo and M. Porcar (2020). "The long journey towards standards for engineering biosystems: Are the Molecular Biology and the Biotech communities ready to standardise?" EMBO Rep **21**(5): e50521.

Bischof, R. H., J. Ramoni and B. Seiboth (2016). "Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*." Microb Cell Fact **15**(1): 106.

Bissaro, B., A. Várnai, K. Røhr Å and V. G. H. Eijsink (2018). "Oxidoreductases and Reactive Oxygen Species in Conversion of Lignocellulosic Biomass." Microbiol Mol Biol Rev **82**(4).

Blum, M., H. Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G. A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D. H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, C. A. Orengo, A. P. Pandurangan, C. Rivoire, C. J. A. Sigrist, I. Sillitoe, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, A. Bateman and R. D. Finn (2020). "The InterPro protein families and domains database: 20 years on." Nucleic Acids Res.

Bokinsky, G., P. P. Peralta-Yahya, A. George, B. M. Holmes, E. J. Steen, J. Dietrich, T. S. Lee, D. Tullman-Ercek, C. A. Voigt, B. A. Simmons and J. D. Keasling (2011). "Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*." Proc Natl Acad Sci U S A **108**(50): 19949-19954.

Brenelli, L., F. M. Squina, C. Felby and D. Cannella (2018). "Laccase-derived lignin compounds boost cellulose oxidative enzymes AA9." Biotechnol Biofuels **11**: 10.

Burdette, L. A., S. A. Leach, H. T. Wong and D. Tullman-Ercek (2018). "Developing Gram-negative bacteria for the secretion of heterologous proteins." Microb Cell Fact **17**(1): 196.

Cameron, D. E., C. J. Bashor and J. J. Collins (2014). "A brief history of synthetic biology." Nat Rev Microbiol **12**(5): 381-390.

Campbell, T. N. and F. Y. M. Choy (2001). "The effect of pH on green fluorescent protein: A brief review." Molecular Biology Today **2**(1): 1-4.

Carbonell, P., A. J. Jervis, C. J. Robinson, C. Yan, M. Dunstan, N. Swainston, M. Vinaixa, K. A. Hollywood, A. Currin, N. J. W. Rattray, S. Taylor, R. Spiess, R. Sung, A. R. Williams, D. Fellows, N. J. Stanford, P. Mulherin, R. Le Feuvre, P. Barran, R. Goodacre, N. J. Turner, C. Goble, G. G. Chen, D. B. Kell, J. Micklefield, R. Breitling, E. Takano, J. L. Faulon and N. S. Scrutton (2018). "An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals." Commun Biol **1**: 66.

Casini, A., M. Storch, G. S. Baldwin and T. Ellis (2015). "Bricks and blueprints: methods and standards for DNA assembly." Nat Rev Mol Cell Biol **16**(9): 568-576.

Cecchini, D. A., O. Pepe, A. Pennacchio, M. Fagnano and V. Faraco (2018). "Directed evolution of the bacterial endo-β-1,4-glucanase from *Streptomyces sp*. G12 towards improved catalysts for lignocellulose conversion." AMB Express **8**(1): 74.

Chang, J. J., M. Anandharaj, C. Y. Ho, K. Tsuge, T. Y. Tsai, H. M. Ke, Y. J. Lin, M. D. Ha Tran, W. H. Li and C. C. Huang (2018). "Biomimetic strategy for constructing *Clostridium thermocellum* cellulosomal operons in *Bacillus subtilis*." Biotechnol Biofuels **11**: 157.

Chao, R., S. Mishra, T. Si and H. Zhao (2017). "Engineering biological systems using automated biofoundries." Metab Eng **42**: 98-108.

Chaves, J. E., G. N. Presley and J. K. Michener (2019). "Modular engineering of biomass degradation pathways." Processes **7**(4).

Cherepanov, P. P. and W. Wackernagel (1995). "Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant." Gene **158**(1): 9-14.

Chiasson, D., V. Giménez-Oya, M. Bircheneder, S. Bachmaier, T. Studtrucker, J. Ryan, K. Sollweck, H. Leonhardt, M. Boshart, P. Dietrich and M. Parniske (2019). "A unified multi-kingdom Golden Gate cloning platform." Sci Rep **9**(1): 10131.

Chung, C. T., S. L. Niemela and R. H. Miller (1989). "One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution." Proc Natl Acad Sci U S A **86**(7): 2172-2175.

Claes, A., Q. Deparis, M. R. Foulquié-Moreno and J. M. Thevelein (2020). "Simultaneous secretion of seven lignocellulolytic enzymes by an industrial second-generation yeast strain enables efficient ethanol production from multiple polymeric substrates." Metab Eng **59**: 131-141.

Cosgrove, D. J. (2017). "Microbial Expansins." Annu Rev Microbiol **71**: 479-497.

Coussement, P., D. Bauwens, J. Maertens and M. De Mey (2017). "Direct Combinatorial Pathway Optimization." ACS Synth Biol **6**(2): 224-232.

Cui, Y., Y. Meng, J. Zhang, B. Cheng, H. Yin, C. Gao, P. Xu and C. Yang (2017). "Efficient secretory expression of recombinant proteins in *Escherichia coli* with a novel actinomycete signal peptide." Protein Expr Purif **129**: 69-74.

Daas, M. J. A., B. Nijsse, A. H. P. van de Weijer, B. Groenendaal, F. Janssen, J. van der Oost and R. van Kranenburg (2018). "Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes." BMC Biotechnol **18**(1): 42.

Damalas, S. G., C. Batianis, M. Martin-Pascual, V. de Lorenzo and V. A. P. Martins Dos Santos (2020). "SEVA 3.1: enabling interoperability of DNA assembly among the SEVA, BioBricks and Type IIS restriction enzyme standards." Microb Biotechnol **13**(6): 1793-1806.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-6645.

Davis, S. J., N. S. Lewis, M. Shaner, S. Aggarwal, D. Arent, I. L. Azevedo, S. M. Benson, T. Bradley, J. Brouwer, Y. M. Chiang, C. T. M. Clack, A. Cohen, S. Doig, J. Edmonds, P. Fennell, C. B. Field, B. Hannegan, B. M. Hodge, M. I. Hoffert, E. Ingersoll, P. Jaramillo, K. S. Lackner, K. J. Mach, M. Mastrandrea, J. Ogden, P. F. Peterson, D. L. Sanchez, D. Sperling, J. Stagner, J. E. Trancik, C. J. Yang and K. Caldeira (2018). "Net-zero emissions energy systems." Science **360**(6396).

Davison, S. A., R. den Haan and W. H. van Zyl (2016). "Heterologous expression of cellulase genes in natural *Saccharomyces cerevisiae* strains." Appl Microbiol Biotechnol **100**(18): 8241-8254.

de Lorenzo, V. and K. N. Timmis (1994). "[31] Analysis and construction of stable phenotypes in gram-negative bacteria with *Tn*5- and Tn10-derived minitransposons." **235**: 386-405.

Del Vecchio, D. (2015). "Modularity, context-dependence, and insulation in engineered biological circuits." Trends Biotechnol **33**(2): 111-119.

Dietrich, J. A., D. L. Shis, A. Alikhani and J. D. Keasling (2013). "Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis." ACS Synth Biol **2**(1): 47-58.

Donev, E., M. L. Gandla, L. J. Jönsson and E. J. Mellerowicz (2018). "Engineering Non-cellulosic Polysaccharides of Wood for the Biorefinery." Front Plant Sci **9**: 1537.

Du Lac, M., T. Duigou, J. Herisson, P. Carbonell, N. Swainston, V. Zulkower, L. Faure, M. Mahdy, P. Soudier and J.-L. Faulon (2020). "Galaxy-SynBioCAD: Synthetic Biology Design Automation tools in Galaxy workflows." BioRxiv.

Duedu, K. O. and C. E. French (2017). "Two-colour fluorescence fluorimetric analysis for direct quantification of bacteria and its application in monitoring bacterial growth in cellulose degradation systems." J Microbiol Methods **135**: 85-92.

Dunn, J. B. (2019). "Biofuel and bioproduct environmental sustainability analysis." Curr Opin Biotechnol **57**: 88-93.

Dvořák, P. and V. de Lorenzo (2018). "Refactoring the upper sugar metabolism of *Pseudomonas putida* for co-utilization of cellobiose, xylose, and glucose." Metab Eng **48**: 94-108.

Endalur Gopinarayanan, V. and N. U. Nair (2019). "Pentose Metabolism in *Saccharomyces cerevisiae*: The Need to Engineer Global Regulatory Systems." Biotechnol J **14**(1): e1800364.

Engler, C., R. Gruetzner, R. Kandzia and S. Marillonnet (2009). "Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes." PLoS One **4**(5): e5553.

Engler, C., M. Youles, R. Gruetzner, T. M. Ehnert, S. Werner, J. D. Jones, N. J. Patron and S. Marillonnet (2014). "A golden gate modular cloning toolbox for plants." ACS Synth Biol **3**(11): 839-843.

Enjalbert, B., M. Cocaign-Bousquet, J. C. Portais and F. Letisse (2015). "Acetate Exposure Determines the Diauxic Behavior of *Escherichia coli* during the Glucose-Acetate Transition." J Bacteriol **197**(19): 3173-3181.

Fan, L. H., Z. J. Zhang, S. Mei, Y. Y. Lu, M. Li, Z. Y. Wang, J. G. Yang, S. T. Yang and T. W. Tan (2016). "Engineering yeast with bifunctional minicellulosome and cellodextrin pathway for co-utilization of cellulose-mixed sugars." Biotechnol Biofuels **9**: 137.

Fang, M., T. Wang, C. Zhang, J. Bai, X. Zheng, X. Zhao, C. Lou and X. H. Xing (2016). "Intermediate-sensor assisted push-pull strategy and its application in heterologous deoxyviolacein production in *Escherichia coli*." Metab Eng **33**: 41-51.

Frederix, M., F. Mingardon, M. Hu, N. Sun, T. Pray, S. Singh, B. A. Simmons, J. D. Keasling and A. Mukhopadhyay (2016). "Development of an: *E. coli* strain for one-pot biofuel production from ionic liquid pretreated cellulose and switchgrass." Green Chemistry **18**(15): 4189-4197.

French, C. E. (2009). "Synthetic biology and biomass conversion: a match made in heaven?" J R Soc Interface **6 Suppl 4**(Suppl 4): S547-558.

Freudl, R. (2018). "Signal peptides for recombinant protein secretion in bacterial expression systems." Microb Cell Fact **17**(1): 52.

Fukuda, H., A. Kondo and H. Noda (2001). "Biodiesel fuel production by transesterification of oils." J Biosci Bioeng **92**(5): 405-416.

Gao, D., Y. Luan, Q. Liang and Q. Qi (2016). "Exploring the N-terminal role of a heterologous protein in secreting out of *Escherichia coli*." Biotechnol Bioeng **113**(12): 2561-2567.

Gao, D., Y. Luan, Q. Wang, Q. Liang and Q. Qi (2015). "Construction of cellulose-utilizing *Escherichia coli* based on a secretable cellulase." Microb Cell Fact **14**: 159.

Gao, D., S. Wang, H. Li, H. Yu and Q. Qi (2015). "Identification of a heterologous cellulase and its N-terminus that can guide recombinant proteins out of *Escherichia coli*." Microb Cell Fact **14**: 49.

Gates, M. G., A. R. McGarel-Groves, N. D. Anderson, K. J. C. Day, P. Leahy and G. N. Drage (2018). Process operations for biomass fractionation, Google Patents.

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd and H. O. Smith (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nat Methods **6**(5): 343-345.

Green, E. R. and J. Mecsas (2016). "Bacterial Secretion Systems: An Overview." Microbiol Spectr **4**(1).

Guo, Z. P., S. Duquesne, S. Bozonnet, G. Cioci, J. M. Nicaud, A. Marty and M. J. O'Donohue (2017). "Conferring cellulose-degrading ability to *Yarrowia lipolytica* to facilitate a consolidated bioprocessing approach." Biotechnol Biofuels **10**: 132.

Guo, Z. P., S. Duquesne, S. Bozonnet, J. M. Nicaud, A. Marty and M. J. O'Donohue (2017). "Expressing accessory proteins in cellulolytic *Yarrowia lipolytica* to improve the conversion yield of recalcitrant cellulose." Biotechnol Biofuels **10**: 298.

Gyorgy, A., J. I. Jiménez, J. Yazbek, H. H. Huang, H. Chung, R. Weiss and D. Del Vecchio (2015). "Isocost Lines Describe the Cellular Economy of Genetic Circuits." Biophys J **109**(3): 639-646.

Hall, B. G., H. Acar, A. Nandipati and M. Barlow (2014). "Growth rates made easy." Mol Biol Evol **31**(1): 232-238.

Hardin, G. (1968). "The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality." Science **162**(3859): 1243-1248.

Hassan, S. S., G. A. Williams and A. K. Jaiswal (2018). "Emerging technologies for the pretreatment of lignocellulosic biomass." Bioresour Technol **262**: 310-318.

Haszeldine, R. S., S. Flude, G. Johnson and V. Scott (2018). "Negative emissions technologies and carbon capture and storage to achieve the Paris Agreement commitments." Philos Trans A Math Phys Eng Sci **376**(2119).

Herburger, K., L. Franková, M. Pičmanová, J. W. Loh, M. Valenzuela-Ortega, F. Meulewaeter, A. D. Hudson, C. E. French and S. C. Fry (2020). "Hetero-trans-β-Glucanase Produces Cellulose-Xyloglucan Covalent Bonds in the Cell Walls of Structural Plant Tissues and Is Stimulated by Expansin." Mol Plant **13**(7): 1047-1062.

Hetzler, S., D. Bröker and A. Steinbüchel (2013). "Saccharification of cellulose by recombinant *Rhodococcus opacus* PD630 strains." Appl Environ Microbiol **79**(17): 5159-5166.

Huber, D., D. Boyd, Y. Xia, M. H. Olma, M. Gerstein and J. Beckwith (2005). "Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation." J Bacteriol **187**(9): 2983-2991.

Iverson, S. V., T. L. Haddock, J. Beal and D. M. Densmore (2016). "CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology." ACS Synth Biol **5**(1): 99-103.

Jahn, M., C. Vorpahl, T. Hübschmann, H. Harms and S. Müller (2016). "Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR." Microb Cell Fact **15**(1): 211.

Jain, A. and P. Srivastava (2013). "Broad host range plasmids." FEMS Microbiol Lett **348**(2): 87-96.

Jeong, K. J. and S. Y. Lee (2000). "Secretory production of human leptin in *Escherichia coli*." Biotechnol Bioeng **67**(4): 398-407.

Jeong, K. J. and S. Y. Lee (2002). "Excretion of human beta-endorphin into culture medium by using outer membrane protein F as a fusion partner in recombinant *Escherichia coli*." Appl Environ Microbiol **68**(10): 4979-4985.

Jeschek, M., D. Gerngross and S. Panke (2017). "Combinatorial pathway optimization for streamlined metabolic engineering." Curr Opin Biotechnol **47**: 142-151.

Jo, C., D. B. Bernstein, N. Vaisman, H. M. Frydman and D. Segrè (2021). "A co-culture microplate for real-time measurement of microbial interactions." bioRxiv.

Jonet, M. A., N. M. Mahadi, A. M. Murad, A. Rabu, F. D. Bakar, R. A. Rahim, K. O. Low and R. M. Illias (2012). "Optimization of a heterologous signal peptide by site-directed mutagenesis for improved secretion of recombinant proteins in *Escherichia coli*." J Mol Microbiol Biotechnol **22**(1): 48-58.

Kadhum, H. J., K. Rajendran and G. S. Murthy (2018). "Optimization of Surfactant Addition in Cellulosic Ethanol Process Using Integrated Techno-economic and Life Cycle Assessment for Bioprocess Design." ACS Sustainable Chemistry and Engineering **6**(11): 13687-13695.

Käll, L., A. Krogh and E. L. Sonnhammer (2007). "Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server." Nucleic Acids Res **35**(Web Server issue): W429-432.

Kelly, J. R., A. J. Rubin, J. H. Davis, C. M. Ajo-Franklin, J. Cumbers, M. J. Czar, K. de Mora, A. L. Glieberman, D. D. Monie and D. Endy (2009). "Measuring the

activity of BioBrick promoters using an in vivo reference standard." J Biol Eng **3**: 4.

Kickenweiz, T., A. Glieder and J. C. Wu (2018). "Construction of a cellulose-metabolizing *Komagataella phaffii* (*Pichia pastoris*) by co-expressing glucanases and β-glucosidase." Appl Microbiol Biotechnol **102**(3): 1297-1306.

Kim, S. H., A. M. Cavaleiro, M. Rennig and M. H. Norholm (2016). "SEVA Linkers: A Versatile and Automatable DNA Backbone Exchange Standard for Synthetic Biology." ACS Synth Biol **5**(10): 1177-1181.

Kim, S. K., W. K. Min, Y. C. Park and J. H. Seo (2015). "Application of repeated aspartate tags to improving extracellular production of *Escherichia coli* L-asparaginase isozyme II." Enzyme Microb Technol **79-80**: 49-54.

Kleiner-Grote, G. R. M., J. M. Risse and K. Friehs (2018). "Secretion of recombinant proteins from E. coli." Eng Life Sci **18**(8): 532-550.

Klumpp, S. and T. Hwa (2014). "Bacterial growth: global effects on gene expression, growth feedback and proteome partition." Curr Opin Biotechnol **28**: 96-102.

Knapp, S., E. Amann and K.-J. Abel (1996). Nucleic acid encoding a signal peptide, a recombinant molecule comprising the nucleic acid, methods of using the nucleic acid, and methods of using the signal peptide, Google Patents.

Ko, J. K., Y. Um, H. M. Woo, K. H. Kim and S. M. Lee (2016). "Ethanol production from lignocellulosic hydrolysates using engineered *Saccharomyces cerevisiae* harboring xylose isomerase-based pathway." Bioresour Technol **209**: 290-296.

Kolstad, C., K. Urama, J. Broome, A. Bruvoll, M. Cariño-Olvera, D. Fullerton, C. Gollier, W. M. Hanemann, R. Hassan and F. Jotzo (2014). "Social, economic and ethical concepts and methods."

Kurokawa, Y., H. Yanagi and T. Yura (2001). "Overproduction of bacterial protein disulfide isomerase (DsbC) and its modulator (DsbD) markedly enhances periplasmic production of human nerve growth factor in *Escherichia coli*." J Biol Chem **276**(17): 14393-14399.

Kuusk, S., R. Kont, P. Kuusk, A. Heering, M. Sørlie, B. Bissaro, V. G. H. Eijsink and P. Väljamäe (2019). "Kinetic insights into the role of the reductant in H(2)O(2)-driven degradation of chitin by a bacterial lytic polysaccharide monooxygenase." J Biol Chem **294**(5): 1516-1528.

Lakhundi, S. S., K. O. Duedu, N. Cain, R. Nagy, J. Krakowiak and C. E. French (2017). "*Citrobacter freundii* as a test platform for recombinant cellulose degradation systems." Lett Appl Microbiol **64**(1): 35-42.

Lee, C. R., B. H. Sung, K. M. Lim, M. J. Kim, M. J. Sohn, J. H. Bae and J. H. Sohn (2017). "Co-fermentation using Recombinant *Saccharomyces cerevisiae* Yeast Strains Hyper-secreting Different Cellulases for the Production of Cellulosic Bioethanol." Sci Rep **7**(1): 4428.

Lee, Y. J., R. Lee, S. H. Lee, S. S. Yim and K. J. Jeong (2016). "Enhanced secretion of recombinant proteins via signal recognition particle (SRP)-dependent secretion pathway by deletion of rrsE in *Escherichia coli*." Biotechnol Bioeng **113**(11): 2453-2461.

Leplat, C., J. M. Nicaud and T. Rossignol (2015). "High-throughput transformation method for *Yarrowia lipolytica* mutant library screening." FEMS Yeast Res **15**(6).

Lewicka, A. J., J. J. Lyczakowski, G. Blackhurst, C. Pashkuleva, K. Rothschild-Mancinelli, D. Tautvaišas, H. Thornton, H. Villanueva, W. Xiao, J. Slikas, L. Horsfall, A. Elfick and C. French (2014). "Fusion of pyruvate decarboxylase and alcohol dehydrogenase increases ethanol production in *Escherichia coli*." ACS Synth Biol **3**(12): 976-978.

Liang, Y., T. Si, E. L. Ang and H. Zhao (2014). "Engineered pentafunctional minicellulosome for simultaneous saccharification and ethanol fermentation in *Saccharomyces cerevisiae*." Appl Environ Microbiol **80**(21): 6677-6684.

Liao, Y., S. F. Koelewijn, G. Van den Bossche, J. Van Aelst, S. Van den Bosch, T. Renders, K. Navare, T. Nicolaï, K. Van Aelst, M. Maesen, H. Matsushima, J. M. Thevelein, K. Van Acker, B. Lagrain, D. Verboekend and B. F. Sels (2020). "A sustainable wood biorefinery for low-carbon footprint chemicals production." Science **367**(6484): 1385-1390.

Ling, H. L., Z. Rahmat, A. M. A. Murad, N. M. Mahadi and R. M. Illias (2017). "Data for proteome analysis of *Bacillus lehensis* G1 in starch-containing medium." Data Brief **14**: 35-40.

Liu, C. (2012). Cellulose degradation system of *Cytophaga hutchinsonii*. Doctor of Philosophy, University of Edinburgh.

Liu, W., D. R. Bevan and Y. H. Zhang (2010). "The family 1 glycoside hydrolase from *Clostridium cellulolyticum* H10 is a cellodextrin glucohydrolase." Appl Biochem Biotechnol **161**(1-8): 264-273.

Liu, Z., L. Tian, Y. Chen and H. Mou (2014). "Efficient extracellular production of κ-carrageenase in *Escherichia coli*: effects of wild-type signal sequence and process conditions on extracellular secretion." J Biotechnol **185**: 8-14.

Loaces, I., S. Schein and F. Noya (2017). "Ethanol production by *Escherichia coli* from *Arundo donax* biomass under SSF, SHF or CBP process configurations and in situ production of a multifunctional glucanase and xylanase." Bioresour Technol **224**: 307-313.

Lombard, V., H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat (2014). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic Acids Res **42**(Database issue): D490-495.

Looman, A. C., J. Bodlaender, L. J. Comstock, D. Eaton, P. Jhurani, H. A. de Boer and P. H. van Knippenberg (1987). "Influence of the codon following the AUG initiation codon on the expression of a modified *lacZ* gene in *Escherichia coli*." The EMBO journal **6**(8): 2489-2492.

Loose, J. S., Z. Forsberg, D. Kracher, S. Scheiblbrandner, R. Ludwig, V. G. Eijsink and G. Vaaje-Kolstad (2016). "Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase." Protein Sci **25**(12): 2175-2186.

Lynd, L. R., X. Liang, M. J. Biddy, A. Allee, H. Cai, T. Foust, M. E. Himmel, M. S. Laser, M. Wang and C. E. Wyman (2017). "Cellulosic ethanol: status and innovation." Curr Opin Biotechnol **45**: 202-211.

Lynd, L. R., P. J. Weimer, W. H. van Zyl and I. S. Pretorius (2002). "Microbial cellulose utilization: fundamentals and biotechnology." Microbiol Mol Biol Rev **66**(3): 506-577, table of contents.

Martínez-García, E., T. Aparicio, V. de Lorenzo and P. I. Nikel (2014). "New transposon tools tailored for metabolic engineering of gram-negative microbial cell factories." Front Bioeng Biotechnol **2**: 46.

Martínez-García, E., T. Aparicio, A. Goñi-Moreno, S. Fraile and V. de Lorenzo (2015). "SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities." Nucleic Acids Res **43**(Database issue): D1183-1189.

Martínez-García, E., A. Goñi-Moreno, B. Bartley, J. McLaughlin, L. Sánchez-Sampedro, H. Pascual Del Pozo, C. Prieto Hernández, A. S. Marletta, D. De Lucrezia, G. Sánchez-Fernández, S. Fraile and V. de Lorenzo (2020). "SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts." Nucleic Acids Res **48**(D1): D1164-d1170.

Masson-Delmotte, T., P. Zhai, H. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan and R. e. Pidcock (2018). IPCC, 2018: Summary for Policymakers. In: Global warming of 1.5 C. An IPCC Special Report on the impacts of global warming of 1.5 C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global. World Meteorological Organization, Geneva, Tech. Rep. V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield Geneva, Switzerland, 32 pp., Intergovernmental Panel on Climate Change.

Meyer, A. J., T. H. Segall-Shapiro, E. Glassey, J. Zhang and C. A. Voigt (2019). "*Escherichia coli* "Marionette" strains with 12 highly optimized small-molecule sensors." Nat Chem Biol **15**(2): 196-204.

Mo, C., N. Chen, T. Lv, J. Du and S. Tian (2015). "Direct ethanol production from steam-exploded corn stover using a synthetic diploid cellulase-displaying yeast consortium." BioResources **10**(3): 4460-4472.

Moore, S. J., H. E. Lai, R. J. Kelwick, S. M. Chee, D. J. Bell, K. M. Polizzi and P. S. Freemont (2016). "EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology." ACS Synth Biol **5**(10): 1059-1069.

Mühlmann, M., M. Kunze, J. Ribeiro, B. Geinitz, C. Lehmann, U. Schwaneberg, U. Commandeur and J. Büchs (2017). "Cellulolytic RoboLector - towards an automated high-throughput screening platform for recombinant cellulase expression." J Biol Eng **11**: 1.

Müller, G., P. Chylenski, B. Bissaro, V. G. H. Eijsink and S. J. Horn (2018). "The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail." Biotechnol Biofuels **11**: 209.

Müller, M. G., I. Georgakoudi, Q. Zhang, J. Wu and M. S. Feld (2001). "Intrinsic fluorescence spectroscopy in turbid media: disentangling effects of scattering and absorption." Appl Opt **40**(25): 4633-4646.

Najah, M., R. Calbrix, I. P. Mahendra-Wijaya, T. Beneyton, A. D. Griffiths and A. Drevelle (2014). "Droplet-based microfluidics platform for ultra-high-throughput bioprospecting of cellulolytic microorganisms." Chem Biol **21**(12): 1722-1732.

Nanda, S., J. Mohammad, S. N. Reddy, J. A. Kozinski and A. K. Dalai (2014). "Pathways of lignocellulosic biomass conversion to renewable fuels." Biomass Conversion and Biorefinery **4**(2): 157-191.

Naseri, G. and M. A. G. Koffas (2020). "Application of combinatorial optimization strategies in synthetic biology." Nat Commun **11**(1): 2446.

Natarajan, A., C. H. Haitjema, R. Lee, J. T. Boock and M. P. DeLisa (2017). "An Engineered Survival-Selection Assay for Extracellular Protein Expression Uncovers Hypersecretory Phenotypes in *Escherichia coli*." ACS Synth Biol **6**(5): 875-883.

Nicaud, J.-M., A.-M. Crutz-Le Coq, T. Rossignol and N. Morin (2014). Protocols for monitoring growth and lipid accumulation in oleaginous yeasts. Hydrocarbon and Lipid Microbiology Protocols, Springer**:** 153-169.

Nielsen, A. A., B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore and C. A. Voigt (2016). "Genetic circuit design automation." Science **352**(6281): aac7341.

Ostafe, R., R. Prodanovic, U. Commandeur and R. Fischer (2013). "Flow cytometry-based ultra-high-throughput screening assay for cellulase activity." Anal Biochem **435**(1): 93-98.

Parisutham, V., S. P. Chandran, A. Mukhopadhyay, S. K. Lee and J. D. Keasling (2017). "Intracellular cellobiose metabolism and its applications in lignocellulose-based biorefineries." Bioresour Technol **239**: 496-506.

Patron, N. J., D. Orzaez, S. Marillonnet, H. Warzecha, C. Matthewman, M. Youles, O. Raitskin, A. Leveau, G. Farré, C. Rogers, A. Smith, J. Hibberd, A. A. Webb, J. Locke, S. Schornack, J. Ajioka, D. C. Baulcombe, C. Zipfel, S. Kamoun, J. D. Jones, H. Kuhn, S. Robatzek, H. P. Van Esse, D. Sanders, G. Oldroyd, C. Martin, R. Field, S. O'Connor, S. Fox, B. Wulff, B. Miller, A. Breakspear, G. Radhakrishnan, P. M. Delaux, D. Loqué, A. Granell, A. Tissier, P. Shih, T. P. Brutnell, W. P. Quick, H. Rischer, P. D. Fraser, A. Aharoni, C. Raines, P. F. South, J. M. Ané, B. R. Hamberger, J. Langdale, J. Stougaard, H. Bouwmeester, M. Udvardi, J. A. Murray, V. Ntoukakis, P. Schäfer, K. Denby, K. J. Edwards, A. Osbourn and J. Haseloff (2015). "Standards for plant synthetic biology: a common syntax for exchange of DNA parts." New Phytol **208**(1): 13-19.

Peralta-Yahya, P., B. T. Carter, H. Lin, H. Tao and V. W. Cornish (2008). "High-throughput selection for cellulase catalysts using chemical complementation." J Am Chem Soc **130**(51): 17446-17452.

Peretó, J. (2020). "Transmetabolism: the non-conformist approach to biotechnology." Microb Biotechnol.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nat Methods **8**(10): 785-786.

Pollak, B., A. Cerda, M. Delmans, S. Álamos, T. Moyano, A. West, R. A. Gutiérrez, N. J. Patron, F. Federici and J. Haseloff (2019). "Loop assembly: a simple and open system for recursive fabrication of DNA circuits." New Phytol **222**(1): 628-640.

Pollak, B., T. Matute, I. Nuñez, A. Cerda, C. Lopez, V. Vargas, A. Kan, V. Bielinski, P. von Dassow, C. L. Dupont and F. Federici (2020). "Universal loop assembly: open, efficient and cross-kingdom DNA fabrication." Synth Biol (Oxf) **5**(1): ysaa001.

Potapov, V., J. L. Ong, R. B. Kucera, B. W. Langhorst, K. Bilotti, J. M. Pryor, E. J. Cantor, B. Canton, T. F. Knight, T. C. Evans and G. J. S. Lohman (2018). "Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly." ACS Synthetic Biology **7**(11): 2665-2674.

Power, P. M., R. A. Jones, I. R. Beacham, C. Bucholtz and M. P. Jennings (2004). "Whole genome analysis reveals a high incidence of non-optimal codons in secretory signal sequences of *Escherichia coli*." Biochem Biophys Res Commun **322**(3): 1038-1044.

Pratap, J. and K. L. Dikshit (1998). "Effect of signal peptide changes on the extracellular processing of streptokinase from *Escherichia coli*: requirement for secondary structure at the cleavage junction." Mol Gen Genet **258**(4): 326-333.

Radeck, J., D. Meyer, N. Lautenschläger and T. Mascher (2017). "Bacillus SEVA siblings: A Golden Gate-based toolbox to create personalized integrative vectors for *Bacillus subtilis*." Scientific Reports **7**(1): 14134.

Ragauskas, A. J., G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan and C. E. Wyman (2014). "Lignin valorization: improving lignin processing in the biorefinery." Science **344**(6185): 1246843.

Rajnovic, D. and J. Mas (2020). "Fluorometric detection of phages in liquid media: Application to turbid samples." Anal Chim Acta **1111**: 23-30.

Romanuka, J., H. van den Bulke, R. Kaptein, R. Boelens and G. E. Folkers (2009). "Novel strategies to overcome expression problems encountered with toxic proteins: application to the production of Lac repressor proteins for NMR studies." Protein Expr Purif **67**(2): 104-112.

Sadler, J. C. (2020). "The Bipartisan Future of Synthetic Chemistry and Synthetic Biology." Chembiochem **21**(24): 3489-3491.

Salinas, A. (2017). A synthetic biology approach for green macroalgal biomass depolymerization. Doctor of Philosophy, The University of Edinburgh.

Salis, H. M., E. A. Mirsky and C. A. Voigt (2009). "Automated design of synthetic ribosome binding sites to control protein expression." Nat Biotechnol **27**(10): 946-950.

Salvachúa, D., A. Z. Werner, I. Pardo, M. Michalska, B. A. Black, B. S. Donohoe, S. J. Haugen, R. Katahira, S. Notonier, K. J. Ramirez, A. Amore, S. O. Purvine, E. M. Zink, P. E. Abraham, R. J. Giannone, S. Poudel, P. D. Laible, R. L. Hettich and G. T. Beckham (2020). "Outer membrane vesicles catabolize lignin-derived aromatic compounds in *Pseudomonas putida* KT2440." Proc Natl Acad Sci U S A **117**(17): 9302-9310.

Sarrion-Perdigones, A., M. Vazquez-Vilar, J. Palací, B. Castelijns, J. Forment, P. Ziarsolo, J. Blanca, A. Granell and D. Orzaez (2013). "GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology." Plant Physiol **162**(3): 1618-1631.

Sasaki, Y., R. Mitsui, R. Yamada and H. Ogino (2019). "Secretory overexpression of the endoglucanase by *Saccharomyces cerevisiae* via CRISPR-δ-integration and multiple promoter shuffling." Enzyme Microb Technol **121**: 17-22.

Schindler, D., S. Milbredt, T. Sperlea and T. Waldminghaus (2016). "Design and Assembly of DNA Sequence Libraries for Chromosomal Insertion in Bacteria Based on a Set of Modified MoClo Vectors." ACS Synth Biol **5**(12): 1362-1368.

Schlegel, S., E. Rujas, A. J. Ytterberg, R. A. Zubarev, J. Luirink and J. W. de Gier (2013). "Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels." Microb Cell Fact **12**: 24.

Sevastsyanovich, Y. R., D. L. Leyton, T. J. Wells, C. A. Wardius, K. Tveen-Jensen, F. C. Morris, T. J. Knowles, A. F. Cunningham, J. A. Cole and I. R. Henderson (2012). "A generalised module for the selective extracellular accumulation of recombinant proteins." Microb Cell Fact **11**: 69.

Sezonov, G., D. Joseleau-Petit and R. D'Ari (2007). "*Escherichia coli* physiology in Luria-Bertani broth." J Bacteriol **189**(23): 8746-8749.

Shetty, R. P., D. Endy and T. F. Knight, Jr. (2008). "Engineering BioBrick vectors from BioBrick parts." J Biol Eng **2**: 5.

Shin, S. K., J. E. Hyeon, Y. I. Kim, D. H. Kang, S. W. Kim, C. Park and S. O. Han (2015). "Enhanced hydrolysis of lignocellulosic biomass: Bi-functional enzyme complexes expressed in *Pichia pastoris* improve bioethanol production from *Miscanthus sinensis*." Biotechnol J **10**(12): 1912-1919.

Silva-Rocha, R., E. Martínez-García, B. Calles, M. Chavarría, A. Arce-Rodríguez, A. de Las Heras, A. D. Páez-Espino, G. Durante-Rodríguez, J. Kim, P. I. Nikel, R. Platero and V. de Lorenzo (2013). "The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes." Nucleic Acids Res **41**(Database issue): D666-675.

Sroga, G. E. and J. S. Dordick (2002). "A strategy for in vivo screening of subtilisin E reaction specificity in *E. coli* periplasm." Biotechnol Bioeng **78**(7): 761-769.

Stern, J., S. Moraïs, Y. Ben-David, R. Salama, M. Shamshoum, R. Lamed, Y. Shoham, E. A. Bayer and I. Mizrahi (2018). "Assembly of Synthetic Functional Cellulosomal Structures onto the Cell Surface of *Lactobacillus plantarum*, a Potent Member of the Gut Microbiome." Appl Environ Microbiol **84**(8).

Stevenson, K., A. F. McVey, I. B. N. Clark, P. S. Swain and T. Pilizota (2016). "General calibration of microbial growth in microplate readers." Sci Rep **6**: 38828.

Tang, H., M. Song, Y. He, J. Wang, S. Wang, Y. Shen, J. Hou and X. Bao (2017). "Engineering vesicle trafficking improves the extracellular activity and surface display efficiency of cellulases in *Saccharomyces cerevisiae*." Biotechnol Biofuels **10**: 53.

Tang, H., J. Wang, S. Wang, Y. Shen, D. Petranovic, J. Hou and X. Bao (2018). "Efficient yeast surface-display of novel complex synthetic cellulosomes." Microb Cell Fact **17**(1): 122.

Taniguchi, K., K. Nomura, Y. Hata, T. Nishimura, Y. Asami and A. Kuroda (2007). "The Si-tag for immobilizing proteins on a silica surface." Biotechnol Bioeng **96**(6): 1023-1029.

Terrett, O. M. and P. Dupree (2019). "Covalent interactions between lignin and hemicelluloses in plant secondary cell walls." Curr Opin Biotechnol **56**: 97-104.

Thøgersen, M. S., J. Melchiorsen, C. Ingham and L. Gram (2018). "A Novel Microbial Culture Chamber Co-cultivation System to Study Algal-Bacteria Interactions Using *Emiliania huxleyi* and *Phaeobacter inhibens* as Model Organisms." Front Microbiol **9**: 1705.

Tobias, J. W., T. E. Shrader, G. Rocap and A. Varshavsky (1991). "The N-end rule in bacteria." Science **254**(5036): 1374-1377.

Tozakidis, I. E., T. Brossette, F. Lenz, R. M. Maas and J. Jose (2016). "Proof of concept for the simplified breakdown of cellulose by combining *Pseudomonas putida* strains with surface displayed thermophilic endocellulase, exocellulase and β-glucosidase." Microb Cell Fact **15**(1): 103.

Tsirigotaki, A., J. De Geyter, N. Šoštaric, A. Economou and S. Karamanou (2017). "Protein export through the bacterial Sec pathway." Nat Rev Microbiol **15**(1): 21-36.

Valenzuela-Ortega, M. and C. French (2021). "Joint universal modular plasmids (JUMP): a flexible vector platform for synthetic biology." Synth Biol (Oxf) **6**(1): ysab003.

Valenzuela-Ortega, M. and C. E. French (2019). "Engineering of industrially important microorganisms for assimilation of cellulosic biomass: towards consolidated bioprocessing." Biochem Soc Trans **47**(6): 1781-1794.

Valenzuela-Ortega, M. and C. E. French (2020). "Joint Universal Modular Plasmids: A Flexible Platform for Golden Gate Assembly in Any Microbial Host." Methods Mol Biol **2205**: 255-273.

Vasudevan, R., G. A. R. Gale, A. A. Schiavon, A. Puzorjov, J. Malin, M. D. Gillespie, K. Vavitsas, V. Zulkower, B. Wang, C. J. Howe, D. J. Lea-Smith and A. J. McCormick (2019). "CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax." Plant Physiology **180**(1): 39-55.

Victor, D. G., D. Zhou, E. H. M. Ahmed, P. K. Dadhich, J. Olivier, H. Rogner, K. Sheikho and M. Yamaguchi (2014). Mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T.

Zwickel and J.C. Minx Cambridge, United Kingdom and New York, NY, USA, Intergovernmental Panel on Climate Change. **1454:** 111-150.

Vita, N., R. Borne and H. P. Fierobe (2020). "Cell-surface exposure of a hybrid 3-cohesin scaffoldin allowing the functionalization of *Escherichia coli* envelope." Biotechnol Bioeng **117**(3): 626-636.

Voigt, C. A. (2020). "Synthetic biology 2020-2030: six commercially-available products that are changing our world." Nat Commun **11**(1): 6379.

Vukomanovic, M. and E. Torrents (2019). "High time resolution and high signal-to-noise monitoring of the bacterial growth kinetics in the presence of plasmonic nanoparticles." J Nanobiotechnology **17**(1): 21.

Walls, L. E. and L. Rios-Solis (2020). "Sustainable Production of Microbial Isoprenoid Derived Advanced Biojet Fuels Using Different Generation Feedstocks: A Review." Front Bioeng Biotechnol **8**: 599560.

Wang, B., P. H. Walton and C. Rovira (2019). "Molecular Mechanisms of Oxygen Activation and Hydrogen Peroxide Formation in Lytic Polysaccharide Monooxygenases." ACS Catal **9**(6): 4958-4969.

Wang, J., R. Ledesma-Amaro, Y. Wei, B. Ji and X. J. Ji (2020). "Metabolic engineering for increased lipid accumulation in *Yarrowia lipolytica* - A Review." Bioresour Technol **313**: 123707.

Wang, P., J. Ma, Y. Zhang, M. Zhang, M. Wu, Z. Dai and M. Jiang (2016). "Efficient Secretory Overexpression of Endoinulinase in *Escherichia coli* and the Production of Inulooligosaccharides." Appl Biochem Biotechnol **179**(5): 880-894.

Wargacki, A. J., E. Leonard, M. N. Win, D. D. Regitsky, C. N. Santos, P. B. Kim, S. R. Cooper, R. M. Raisner, A. Herman, A. B. Sivitz, A. Lakshmanaswamy, Y. Kashiyama, D. Baker and Y. Yoshikuni (2012). "An engineered microbial platform for direct biofuel production from brown macroalgae." Science **335**(6066): 308-313.

Wei, H., W. Wang, M. Alahuhta, T. Vander Wall, J. O. Baker, L. E. Taylor, 2nd, S. R. Decker, M. E. Himmel and M. Zhang (2014). "Engineering towards a complete heterologous cellulase secretome in Yarrowia lipolytica reveals its potential for consolidated bioprocessing." Biotechnol Biofuels **7**(1): 148.

Wightman, E. L. I., H. Kroukamp, I. S. Pretorius, I. T. Paulsen and H. K. M. Nevalainen (2020). "Rapid optimisation of cellulolytic enzymes ratios in *Saccharomyces cerevisiae* using in vitro SCRaMbLE." Biotechnol Biofuels **13**(1): 182.

Willett, W., J. Rockström, B. Loken, M. Springmann, T. Lang, S. Vermeulen, T. Garnett, D. Tilman, F. DeClerck, A. Wood, M. Jonell, M. Clark, L. J. Gordon, J. Fanzo, C. Hawkes, R. Zurayk, J. A. Rivera, W. De Vries, L. Majele Sibanda, A. Afshin, A. Chaudhary, M. Herrero, R. Agustina, F. Branca, A. Lartey, S. Fan, B. Crona, E. Fox, V. Bignet, M. Troell, T. Lindahl, S. Singh, S. E. Cornell, K. Srinath Reddy, S. Narain, S. Nishtar and C. J. L. Murray (2019). "Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems." Lancet **393**(10170): 447-492.

Wirth, N. T., E. Kozaeva and P. I. Nikel (2019). "Accelerated genome engineering of *Pseudomonas putida* by I-SceI-mediated recombination and CRISPR-Cas9 counterselection." Microb Biotechnol.

Wright, O., M. Delmans, G. B. Stan and T. Ellis (2015). "GeneGuard: A modular plasmid system designed for biosafety." ACS Synth Biol **4**(3): 307-316.

Zabed, H., J. N. Sahu, A. Suely, A. N. Boyce and G. Faruq (2017). "Bioethanol production from renewable sources: Current perspectives and technological progress." Renewable and Sustainable Energy Reviews **71**: 475-501.

Zacchi, G. and A. Axelsson (1989). "Economic evaluation of preconcentration in production of ethanol from dilute sugar solutions." Biotechnol Bioeng **34**(2): 223-233.

Zalucki, Y. M. and M. P. Jennings (2007). "Experimental confirmation of a key role for non-optimal codons in protein export." Biochem Biophys Res Commun **355**(1): 143-148.

Zhou, S. and H. S. Alper (2019). "Strategies for directed and adapted evolution as part of microbial strain engineering." Journal of Chemical Technology and Biotechnology **94**(2): 366-376.

Zhu, Y. and M. J. McBride (2017). "The unusual cellulose utilization system of the aerobic soil bacterium *Cytophaga hutchinsonii*." Appl Microbiol Biotechnol **101**(19): 7113-7127.