THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

# CRISPR-based technologies for high-throughput metabolic engineering of yeast

CHRISTOS SKREKAS


Department of Biology and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2022

CRISPR-based technologies for high-throughput metabolic engineering of yeast

CHRISTOS SKREKAS
ISBN 978-91-7905-702-2



Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5168
ISSN 0346-718X

Department of Biology and Biological Engineering
Division of Systems and Synthetic Biology
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover: Schematic representation of the gRNA library screenings performed in the thesis

Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

**CRISPR-based technologies for high throughput metabolic engineering of yeast**

Christos Skrekas
Department of Biology and Biological Engineering
Chalmers University of Technology

# Abstract

The yeast *Saccharomyces cerevisiae* is a commonly used microorganism for metabolic engineering applications since it has a very well-studied metabolism and it can be easily genetically modified. This led to its use as a cell factory for the production of a wide variety of industrially relevant chemicals such as fuels, cosmetics and food additives. However, there is always space for improvement of the productivity metrics of the products of interest. Some of the challenges that have to be addressed are the efficient rewiring of the metabolism for improvement of the metabolic fluxes and the performance improvement of the enzymes of interest. The development of metabolite biosensors that can connect the levels of a metabolite of interest to a readable output, has allowed the use of high-throughput screening methods in yeast such as Fluorescence Activated Cell Sorting (FACS) to identify cells with higher metabolite levels. Moreover, the CRISPR/Cas9 technology that has emerged the last decade has not only sped up the introduction of genomic modifications for strain engineering, but it has been further developed for other purposes such as gene expression fine tuning or base editing. Those applications of CRISPR/Cas9 can be coupled to high-throughput screening methods and can give new insights into metabolic engineering challenges.

This study aimed to develop various CRISPR/Cas9-based tools in yeast along with their implementation in high-throughput setups for solving metabolic engineering challenges. At the same time, a modular cloning system for CRISPR/Cas9-based tools was developed for making the molecular cloning of those tools more fast, flexible and simple to use. Hyperactive variants of cytidine and adenine deaminases were explored for the construction of broad range CRISPR base editors in yeast for *in vivo* mutagenesis. Also, gRNA libraries were used in two different setups: with transcriptional activator dCas9-VPR for transcription optimization and with broad range base editors for directed evolution of a gene of choice.

In summary, this work explores some of the possibilities that CRISPR tools can offer when combined with gRNA libraries and at the same time it aims to contribute to the systematization of the experimental workflow for CRISPR applications in yeast.

**Keywords:** *Saccharomyces cerevisiae,* metabolic engineering, library, screening, directed evolution, CRISPR, selection, enrichment, cloning

# Table of contents

# List of publications

The thesis is based on the work contained in the following papers and manuscripts:

**I.** Ferreira, R., **Skrekas, C.,** Hedin, A., Sánchez, B. J., Siewers, V., Nielsen, J., & David, F. (2019). Model-assisted fine-tuning of central carbon metabolism in yeast through dCas9-based regulation. ACS Synthetic Biology, 8(11), 2457-2463.

**II.** Otto, M.*, **Skrekas, C.*,** Gossing, M., Gustaffsson, J., Siewers, V., David, F. (2021) An expansion of the yeast modular cloning toolkit for CRISPR-based applications, chromosomal integration and library construction. ACS Synthetic Biology, 10(12), 3461-3474.

**III.** **Skrekas, C.,** Limeta, A., Siewers, V., David, F. (2022) Targeted *in vivo* mutagenesis in yeast using CRISPR/Cas9 and hyperactive cytidine and adenine deaminases [manuscript].

**IV.** **Skrekas, C.*,** Liu, D*., Siewers, V., David, F. (2022) *In vivo* evolution of malonate transport in *Saccharomyces cerevisiae* [manuscript].

Additional publications not included in the thesis:

**V.** Ferreira, R., **Skrekas, C.,** Nielsen, J., & David, F. (2018). Multiplexed CRISPR/Cas9 genome editing and gene regulation using Csy4 in Saccharomyces cerevisiae. ACS Synthetic Biology, 7(1), 10-15.

**VI.** Dabirian, Y., **Skrekas, C.,** David, F., & Siewers, V. (2020). Does co-expression of *Yarrowia lipolytica* genes encoding Yas1p, Yas2p and Yas3p make a potential alkane-responsive biosensor in *Saccharomyces cerevisiae*?. PLOS One, 15(12), e0239882.

**VII.** **Skrekas, C.,** Ferreira, R., & David, F. (2022). Fluorescence-Activated Cell Sorting as a Tool for Recombinant Strain Screening. In: Yeast Metabolic Engineering (pp. 39-57). Humana, New York, NY [Book chapter].

**VIII.** Zrimec, J., Fu, X., Muhammad, A. S., **Skrekas, C.,** Jauniskis, V., Speicher, N. K., Börlin, C. S., Verendel, V., Chehreghani, M. H., Dubhashi, D., Siewers, V., David, F., Nielsen, J., Zelezniak, A. (2022). Controlling gene expression with deep generative design of regulatory DNA. Nature Communications, 13(1), 1-17.

*** contributed equally**

# Contribution summary

**I.** Performed parts of the experimental design, wet lab work and data analysis

**II.** For the CRISPR-based applications part: Designed the experiments, performed the research and wet lab work, analyzed and interpreted the data, wrote the manuscript

**III.** Designed the experiments, performed all wet lab work, interpreted the results, wrote the manuscript

**IV.** For the CRISPR-based evolution: Designed and performed the experiments, analyzed and interpreted the data, wrote the manuscript

**V.** Performed parts of the experimental design, wet lab work and data analysis

**VI.** Performed parts of the experiments

**VII.** Wrote the manuscript

**VIII.** Performed parts of the experiments

# Preface

This dissertation serves as partial fulfilment of the requirements to obtain the degree of Doctor of Philosophy at the Department of Biology and Biological Engineering at Chalmers University of Technology. The PhD studies were carried out from March 2018 until October 2022 under the supervision of Florian David and the co-supervision of Verena Siewers. The thesis was examined by Jens Nielsen. Funding partners were Formas, Åforsk and Knut and Alice Wallenberg Foundation.

Christos Skrekas
October 2022

# Abbreviations

| | |
|---|---|
| **ABE** | Adenine Base Editor |
| **AID** | Activation-Induced cytidine deaminase |
| **BE** | Base Editor |
| **BH** | Bridge Helix of Cas9 protein |
| **Cas9** | Crispr-Associated protein 9 |
| **CBE** | Cytidine Base Editor |
| **cfu** | colony forming units |
| **CRAIDE** | CRISPR- and RNA- Assisted *In vivo* Directed Evolution |
| **CRISPR** | Clustered Regularly Interspaced Short Palindromic Repeats |
| **CRISPRi** | CRISPR interference |
| **CRISPRa** | CRISPR activation |
| **CRISPRai** | CRISPR activation/interference |
| **crRNA** | CRISPR RNA |
| **DBTL** | Design-Build-Test-Learn |
| **DNA** | Deoxyribonucleic acid |
| **DR** | Direct Repeat |
| **dCas9** | dead Cas9 |
| **FACS** | Fluorescence Activated Cell Sorting |
| **GEM** | Genome-scale metabolic model |
| **GC** | Gas Chromatography |
| **GFP** | Green Fluorescent Protein |
| **HPLC** | High Performance Liquid Chromatography |
| **HR** | Homologous Recombination |
| **ICE** | *In vivo* Continuous Evolution |
| **LC** | Liquid Chromatography |
| **MoClo** | Modular cloning |
| **NGS** | Next Generation Sequencing |
| **NHEJ** | Non-homologous End Joining |
| **NUC** | Nuclease lobe of Cas9 protein |
| **OrthoRep** | Orthogonal replication |
| **ORF** | Open Reading Frame |

| | |
|---|---|
| **PCR** | Polymerase Chain Reaction |
| **REC** | Recognition lobe of Cas9 protein |
| **RNA** | Ribonucleic acid |
| **TRACE** | T7-driven continuous evolution |
| **tracrRNA** | trans-activating RNA |

# Acknowledgments

First of all, I would like to thank my supervisor Florian, not only for the giving me the opportunity to do this PhD, but for the constant support since I first came to SysBio as a master student. I felt many times that you believed in me more than me myself and this thought gave me the strength to continue many times. I would like to also thank my co-supervisor (and main supervisor in the beginning) Verena, for the support but also her comments and the feedback in all the stages of the PhD. I want to also thank Jens for being my examiner.

During those years I have collaborated with many people who helped me and taught me a lot. Raphael, thank you for introducing me to the world of CRISPR and for teaching me a lot of things that were precious for the rest of the PhD. Max, thank you for joining forces for the MoClo paper, this project saved the PhD in many ways. Michi, your comments and discussions during all those years have been more than helpful. Angelo, the contribution of you and your bioinformatics skills maybe needs a special chapter! Thanks for everything, you have been there whenever I needed your help even if your PhD has nothing to do with my projects. Johan, thank you for making my idea on gRNA multiplexing a nice tool. Dany, thanks a lot for the nice collaboration we had on the malonate transporter project. I want to thank also Yasaman for the collaboration in the alkane biosensor project, which might not be a part of this thesis but it was an interesting side project. Louis, thanks for sharing your experience especially on THI4 promoter, even if it didn't end up being a part of the thesis. Finally, I want to thank Aleksej and Xiaozhi for being a part of a very interesting paper on regulatory DNA design.

Our research engineers make a very good and demanding job all those years to keep the lab in a well-functioning state. Angelica, Emelie, Marie, Abder and Yili, thanks for your hard work! Thanks to SysBio which has been a nice working environment. Some special thanks to SysBio people that we shared knowledge, help, support or just some nice conversations: Andrea, Dimitra, Lucy, Marta, Oliver and Veronica.

I want to also thank my family: my mother, my father, my aunt Vasso and my sister Theodora for being my roots wherever I go and supporting me unconditionally. You are the ones that made it possible to finish a PhD now and I never forget it. And finally, thanks to my two closest friends. Matina, thank you for what we have shared those 10 years in biology, politics, personal issues but also for the moments of fun. Giorgos, just thank you for everything. You have been a great support during this PhD by all means possible.

“The only thing I know, is that I know nothing.”
- Socrates, Greek philosopher (470 BC – 399 BC)

# Chapter 1 – Introduction

This chapter is aiming to introduce the basic background knowledge regarding the topic of the thesis. The starting point is the concept of metabolic engineering, the central role that the yeast *Saccharomyces cerevisiae* has played in this, and how the advances in molecular biology have enabled us to engineer microbial cells for the production of a wide variety of industrially relevant chemicals. Next, the most common challenges in metabolic engineering are described, along with the contribution of high-throughput methods in overcoming them. The chapter continues with a description of CRISPR/Cas9 technology and the different tools that are derived from it throughout the last decade. Then, there is a brief discussion on synthetic biology and how standardization of the various genetic parts would facilitate the construction of advanced molecular tools. Finally, the scope of this thesis is presented, which is briefly to explore how different CRISPR tools can be combined with high-throughput approaches in order to facilitate yeast metabolic engineering.

## The yeast *Saccharomyces cerevisiae* as a cell factory

Evidences that indicate alcoholic beverage production via microorganism fermentation can be traced back to around 5400 B.C., and the use of the baker's yeast *Saccharomyces cerevisiae* for wine fermentations is reported via ribosome DNA analyses from at least 3150 B.C. (1). There are also indications that yeast was used since around 10000 B.C. for breadmaking (2). In 1876, Louis Pasteur identified yeast as the microorganism catalysing alcohol fermentation (3). Production of alcoholic beverages and bread are the first examples of the use of microorganisms by humans. During World War I, *S. cerevisiae* was used also for the production of glycerol via microbial fermentation (4).

In the beginning of the 20$^{th}$ century, there was a series of basic findings that sculptured the field of genetics and the concept that deoxyribonucleic acid (DNA) carries the genetic information for all the basic functions of every living organism (5). When it comes to *S. cerevisiae*, an important step to understand its metabolism was the identification of genes that encode enzymes that catalyse sugar catabolizing reactions (6, 7). The elucidation of the structure of proteins (8) and the DNA (9) in the early 50s and also the definition of the central dogma of molecular biology in 1970 (10) gave a significant understanding on how the DNA is organised and how proteins are synthesized in the cells. During the early 70s, the emerge of the recombinant DNA technology made it possible to artificially introduce DNA parts into a host organism, first in *Escherichia coli* (11). This technology enabled in 1977 the production of a human hormone from *E. coli*, somatostatin (12). In 1982, *S. cerevisiae* was successfully engineered for the production of human interferon-alpha (13), and in 1987 for the production of human insulin (14).

The advances in molecular biology and heterologous gene expression led to the development of a new field, which was first described in 1991 with the term **metabolic engineering.** Its first definition was "the improvement of cellular activities by manipulations of enzymatic, transport and regulatory functions with the use of recombinant DNA technology" (15). The advances in metabolic engineering led to the development of cell factories, which are engineered microorganisms that produce compounds of interest, which were initially mostly industrial enzymes and pharmaceutical proteins (16). Soon, the advances in metabolic engineering led to the heterologous expression and rewiring of whole metabolic pathways in order to produce a wider variety of products. The importance of metabolic fluxes and their control on metabolic engineering was later pointed out (17) since it is not always enough to just introduce heterologous pathways in order to achieve sufficient production of the target molecule. The quantification of the metabolic fluxes and the identification of the control mechanisms of them became a fundamental part of metabolic engineering (16, 17).

*S. cerevisiae* became one of the most common host organisms used for engineering of cell factories due to a number of advantages that it possesses. It was the first eucaryotic organism that had its genome fully sequenced and annotated (18, 19). Moreover, there is a wide variety of datasets available for this organism including transcriptome, proteome, metabolome and flux analyses (20–23), something that gives a better insight on the functions of the cell (24, 25). Homologous recombination has a high efficiency in *S. cerevisiae* (26), something that makes this organism very susceptible to genetic modifications (27). Large-scale production of the antimalarial drug artemisinin reported in 2006 (28) was a breakthrough and it was one of the first examples of application of the growing knowledge in yeast metabolic engineering. Since then, many industrially relevant compounds have been produced in yeast and some examples are fatty acid derived biofuels (29–32), terpenoids (33, 34), the value-added chemical 3-hydroxyproprionic acid (3-HP) (35) and many more. By August 2022, the most recent achievement in yeast metabolic engineering is the production of the two precursors of the anticancer drug vinblastine (36).

## Workflow for strain engineering and optimization

The whole process of strain engineering and strain optimization for cell factory construction has been systematized in the **Design**-**Build**-**Test**-**Learn** (DBTL) cycle (37–39) (Figure 1). The **design** step usually begins with the selection of the compound of interest and the organism that will serve as a cell factory for its production. A common practice is to search in metabolic pathway and enzyme databases to retrieve biosynthetic pathways from the organisms that originally produce the compound of interest (38). Mathematical models of cell metabolism called genome-scale metabolic models (GEMs) can also contribute to the strain design. GEMs can also exploit transcriptomics, genomics and metabolomics data to predict key reactions that will

favour the production of the desired compound (37, 40–42). When the design step is finished there are usually defined: the host organism, the desired pathway(s) to be inserted and the possible modifications in the endogenous metabolism of the host that will have a positive effect in the production of the compound of interest (39).

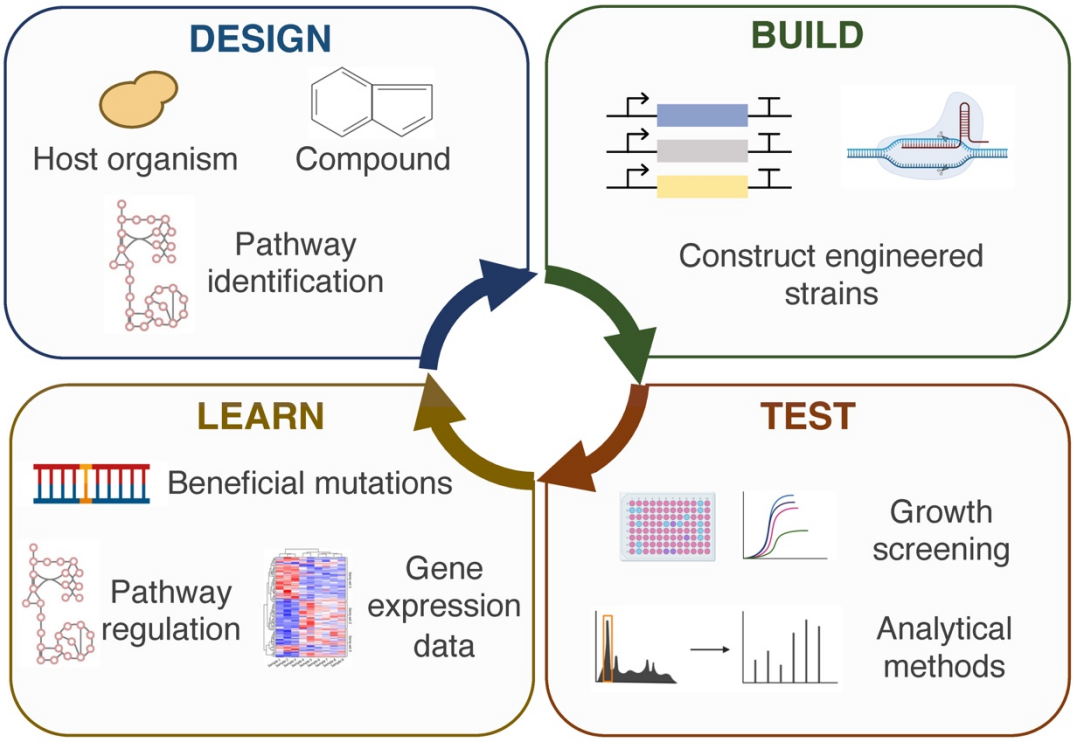


**Figure 1. The Design-Build-Test-Learn cycle for cell factory construction.**

The **building** step involves genetic engineering of the host organism usually by introducing the genes that encode the desired enzymes. It can also involve gene expression fine-tuning, mutagenesis of selected genes that can improve the efficiency of the encoded enzymes etc. (39). The **testing** phase refers to the assessment of the performance of the engineered strains. The most common testing approach is to define yield, rate and titer of the target compound by analytical methods such as liquid chromatography (LC, HPLC) or gas chromatography (GC). Growth rates and cell physiology parameters of the engineered strains can also be determined through the use of bioreactors (38, 39). Finally, the **learn** step wraps up the knowledge that occurs at the end of each DBTL cycle. For example, transcriptome analysis of the best performing strains of an engineering cycle can give novel information about the impact an engineering strategy has on the cellular metabolism and in the regulatory mechanisms of gene expression. Additionally, if several engineering approaches were tried, after the testing we can have a better insight on which approach was more successful and use those data for further improvement. Generally, the outcomes and the data from a learning phase can be incorporated into the previous knowledge and start a new DBTL cycle.

## Biosensors and methods for high-throughput metabolic engineering

Metabolite biosensors play a critical role in connecting the levels of a desired compound with an easily readable output, for example fluorescence. Transcription factor based biosensors are a widely used category of biosensors in metabolic engineering. If connected to an output such as Green Fluorescent Protein (GFP), they can respond to increasing concentrations of a metabolite with increasing green fluorescence (Figure 2). Then a cell population can be easily sorted by Fluorescence Activated Cell Sorting (FACS) and the most promising candidates can be further tested via traditional analytical methods. Biosensor-based library screening approaches have already been successfully applied, and those approaches can shorten the so-called test bottleneck (43, 44). Examples of biosensors developed in *S. cerevisiae* are for sensing malonyl-CoA (45–47), acyl-CoA (48, 49) and naringenin or *cis,cis*-muconic acid (50). As an example of combination of metabolite biosensors and library screening, the acyl-CoA biosensor has been successfully used to screen a gene overexpression library and to identify genes whose overexpression led to increased fatty acid and fatty alcohol levels (51).

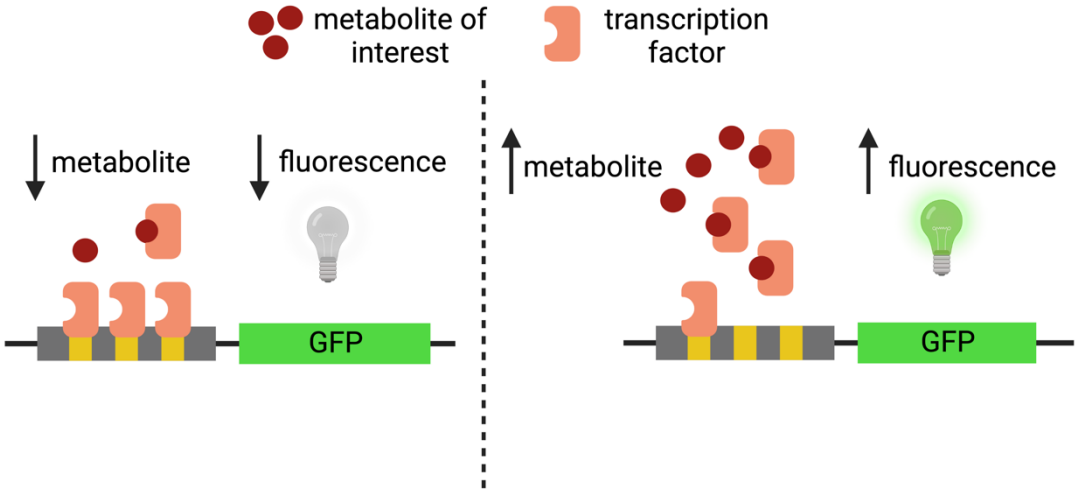


**Figure 2. Description of a transcription factor-based biosensor.** This is an example of a biosensor that positively responds to increasing concentrations of a metabolite. The output is Green Fluorescent Protein (GFP). Its promoter is controlled by a transcription factor which interacts with the metabolite. At increasing concentrations of the metabolite, the transcription factor does not bind to the promoter and enables the expression of the fluorescent protein. (Figure created with BioRender.com)

Since cost of DNA synthesis has been lowering over the last decade (52–54), something that allows the construction of various DNA libraries. Those libraries can consist of gene-encoding cassettes, mutagenized enzymes, promoters etc. (55–57). If phenotype of interest (e.g. the production levels of a desired metabolite) can be linked to the growth of the organism and, a library can be screened through growth-coupled assays (55, 58). Additionally, high-throughput cell screening approaches such as Fluorescence Activated Cell Sorting (FACS) (57, 59) or droplet-based microfluidic sorting (60) allow the screening and sorting of large single cell populations in a short amount of time.

Metabolite biosensors with fluorescent outputs can be used for library screening with FACS, allowing for screening of large libraries in a short time scale.

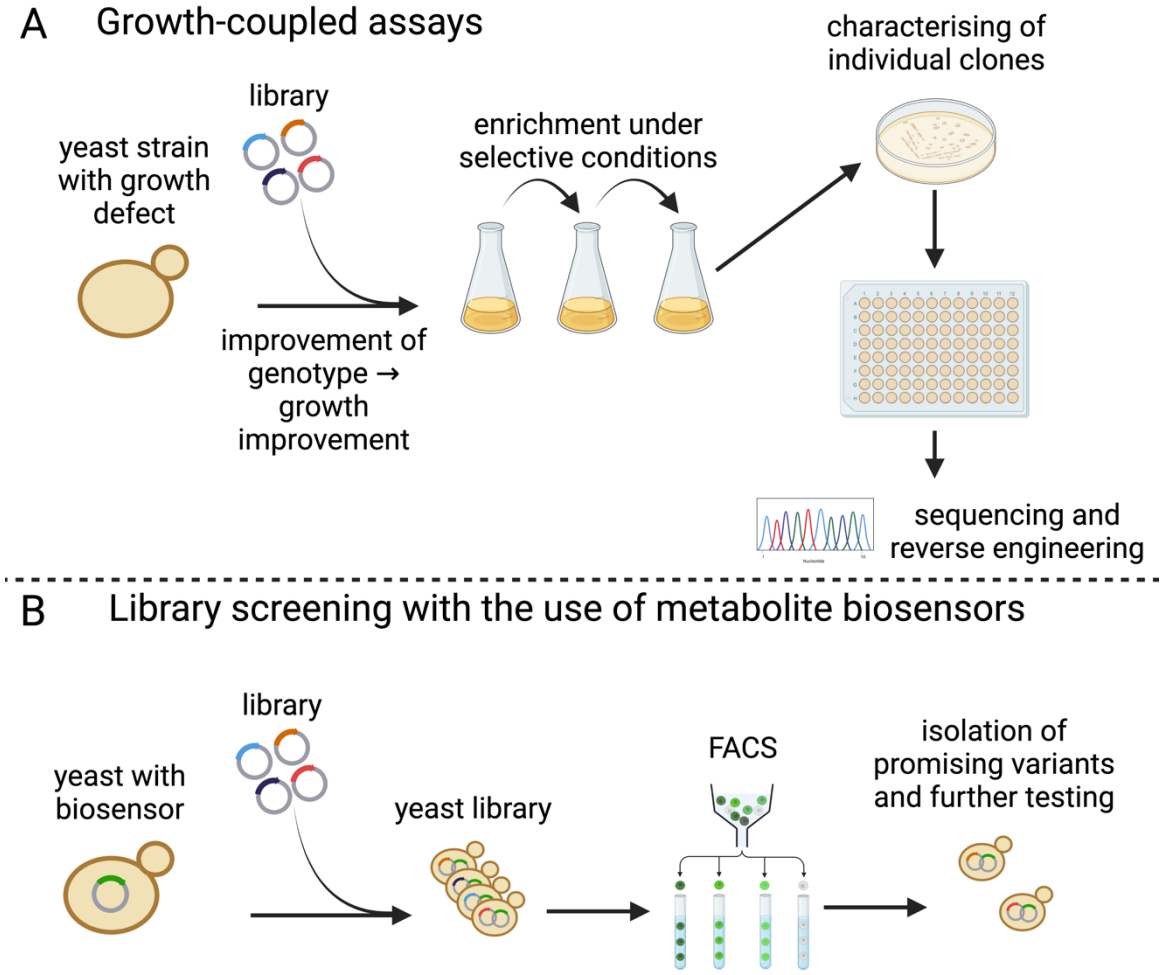


**Figure 3. Library screening approaches in yeast.** Libraries can be screened with growth-coupled assays **(A)** if the trait to be improved can be linked to the growth of the strain. In this case, upon the transformation of the library enrichments under selective conditions follow. Next, individual clones can be screened, sequenced and the most promising genotypes can be reverse engineered to the original strain for final verification. In case that the goal is to improve the production of a metabolite which has a biosensor available, then a library can be screened with FACS **(B)**. After several rounds of FACS, promising variants can be isolated and further tested. (Figure created with BioRender.com)

## CRISPR: Principles and applications

During the last decade, a new technology has emerged named Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR). The discovery of CRISPR began when in 1987 repetitive sequences with unknown function or homology with other known sequences were found at the 3'-end flanking region of a gene in *E. coli* (61). Similar palindromic DNA repeats were later discovered also in archaea (62). An increasing number of such repeats was discovered in the next years in other species of bacteria and archaea (63) until it was found that they were functionally related (64) and they were given the name CRISPR in 2002 (65).

CRISPR arrays are a part of bacterial and archaeal immune systems, which face invasions from foreign DNAs such as phages and conjucative plasmids (63, 66). The other part of this system is a series of CRISPR-associated genes (*cas*) that encode the Cas proteins. The CRISPR arrays consist of short direct repeat (DR) sequences, interspaced with unique ones, which are called spacers. The stages of the CRISPR immunity are two. First comes the adaptation stage, when parts of the invading DNA are incorporated as spacers into the CRISPR locus. In this stage, commonly the genes *cas1* and *cas2* are involved. The second stage is called expression and interference, when the different spacers are transcribed and the resulting RNAs are matured in the so called CRISPR RNAs (crRNAs). The crRNAs along with Cas proteins target and cleave the invading DNA. The classification of CRISPR-Cas systems is an ongoing work and by the time this thesis is written, two classes have been described with several types each (67, 68). In this thesis, I will focus to a class 2 system which was first discovered in *Streptococcus pyogenes* and has been adapted for genetic engineering approaches (69, 70). In this system, DNA is targeted by the dual RNA-guided endonuclease Cas9. Cas9 forms a complex with a dual RNA which consists of two parts: (i) crRNA (spacer) which is complementary with the target DNA and guides Cas9 and (ii) trans-activating RNA (tracrRNA) which is essential for crRNA maturation and is required for the formation of the Cas9-RNA complex (71). The complex of Cas9 with the dual crRNA::tracrRNA is guided to the target DNA and it is activated only in the case that a 5'-NGG-3' Protospacer Adjacent Motif (PAM) is present directly after the target sequence (72–74). The PAM-based recognition mechanism prevents the self-destruction of the CRISPR array, given the fact that the PAM sequence is present only in the target DNA (75).

Jinek *et al.* (72) proved that the dual crRNA::tracrRNA complex with Cas9 can target a specific sequence *in vitro* and also that those two RNAs can be linked into a single RNA (sgRNA). In later literature, this single RNA is referred also simply as guide RNA (gRNA). This study was the first proof that a system consisting of a single RNA and Cas9 can be used for targeted cleavage of a DNA sequence. In January 2013, two reports (76, 77)

proved that this CRISPR/Cas9 approach can be used for *in vivo* targeted engineering of human cell lines. Cas9 can be guided by custom-made gRNAs to a desired locus of the chromosomal DNA and create double stranded breaks (DSBs) three base pairs upstream of the PAM sequence. Those breaks initiate the endogenous repair mechanisms of the cell, either homologous recombination or non-homologous end joining (NHEJ) and -thus- genome engineering can be performed in this locus. In this approach, the length of the spacer is standardized at 20 bp.

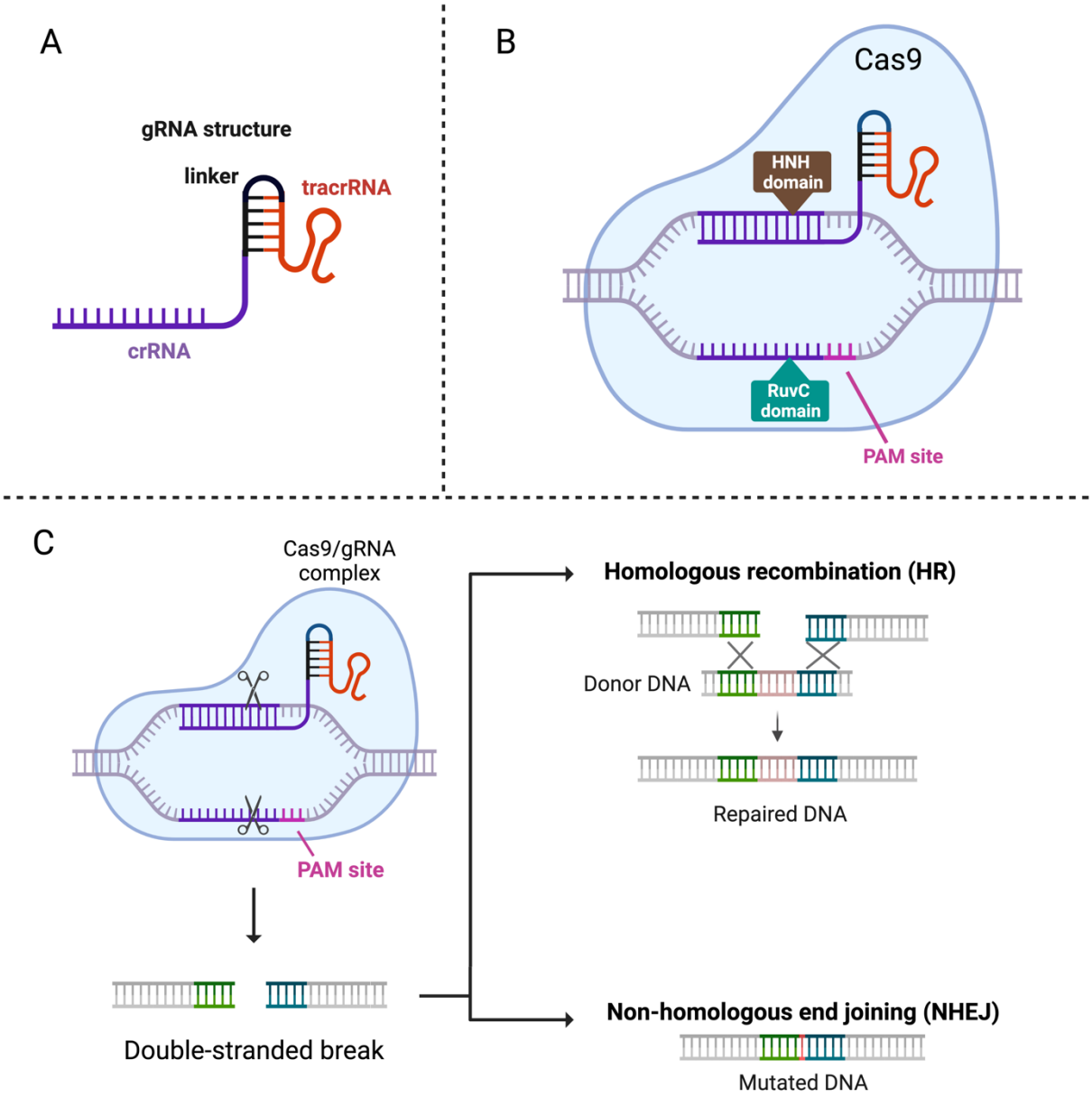


**Figure 4. The basic characteristics of CRISPR/Cas9. (A):** Structure of the single gRNA which consists of crRNA which guides the Cas9/gRNA complex, the structural tracrRNA which interacts with Cas9 and a linker RNA between them. **(B):** The Cas9/gRNA complex bound on the target DNA. A PAM site should follow after the 20 bp sequence recognized by the crRNA in order for Cas9 to activate. Two endonuclease domains (HNH and RuvC) cut both DNA strands 3 bp upstream of the PAM site. **(C):** The double stranded break created by Cas9 can be repaired either by homologous recombination if a donor DNA fragment is present or by non-homologous end joining. (Figure created with BioRender.com)

The crystal structure of the complex of Cas9-gRNA and the target DNA (78, 79) revealed that Cas9 consists of two distinct lobes: the target recognition lobe (REC), which is involved in target DNA recognition, and the nuclease lobe (NUC), which contains the endonucleases HNH and RuvC. The two lobes are connected with the bridge helix (BH).

In an attempt to increase the targeting potential of CRISPR, Cas9 was evolved to recognize as PAM sequences apart from the NGG motif the motifs NG, GAA and GAT (80). Moreover, alternatives to Cas9 have been employed such as Cpf1 that recognizes different PAM sequences and needs a shorter sgRNA sequence (81). Cas13 (known also as C2c2) is another alternative to Cas9 that targets RNA instead of DNA and has been adapted for precise transcriptome engineering (82–84).

## Genome engineering

CRISPR/Cas9 has been implemented in *S. cerevisiae* for cell factory construction. It has facilitated gene deletions, gene replacements, mutation introduction and chromosomal integration of genes for heterologous expression (85–87). Targeted double strand breaks can be repaired by homologous recombination with linear fragments (donor repair fragments) that contain the genetic alteration of choice. Donor repair fragments are introduced in yeast along with gRNA and Cas9. In combination with the very high efficiency homologous recombination in yeast, this approach allows fast and less time-consuming engineering. CRISPR/Cas9 has been also used for multiplexed pathway engineering on both genes and promoters (88).

This approach can also be combined with DNA libraries of donor repair fragments, enabling high-throughput metabolic engineering approaches (85). Those libraries can carry a variety of genetic alterations (deletions, substitutions, insertions). Such a high-throughput approach allowed the characterization of the DNA helicase Sgs1 (89). Additionally, the vitality for growth under specific conditions of 315 poorly characterized open reading frames (ORFs) was assessed (89).

CRISPR/Cas9 has also allowed the development of a marker-free kit for chromosomal integration of genes in *S. cerevisiae* (90). Previous chromosomal integration approaches (91, 92) exploited the high efficiency of homologous recombination (HR) in yeast but they used auxotrophic or antibiotic resistance markers. Auxotrophic markers have been reported to affect cell physiology (93) whereas the use of antibiotic selection markers bears the risk of spreading drug resistance. Marker-free integration allows also to skip the marker recycling step, which is needed especially when we aim for multiple gene edits, allowing faster building of complex cell factories.

One of the major drawbacks of CRISPR/Cas9 is the off-target activity that Cas9-gRNA complexes show (94–96) and this effect comes mainly from gRNA homology with other

loci in the genome apart from the targeted one. Two mutations - one in each nuclease domain - have been identified to convert Cas9 to a nickase, cutting only one DNA strand. $Cas9^{D10A}$ variant cleaves the gRNA-targeting strand and $Cas9^{H840A}$ cleaves the non-targeted strand. Those variants are referred as Cas9 nickases or nCas9, and genome engineering with them has led to more precise genome edits with lower off-target effects (76, 97–99). Among other strategies employed to minimize the off-target effects of CRISPR/Cas9 are the use truncated gRNAs less than 20 bp long (100) and development of Cas9 variants with higher fidelity (101, 102).

CRISPR/Cas9 activity is dependent also on the target DNA accessibility. DNA exists in two chromatin states, the highly condensed heterochromatin and the less condensed euchromatin. It has been shown that Cas9 targeting is more efficient in euchromatin areas rather than heterochromatin (103, 104). Also, CRISPR editing has been proved to be more efficient in the non-transcribed DNA strands rather than the transcribed ones (105). The GC content of the gRNA affects both off-target and on-target binding efficiency, with a 50-70% of GC content being the optimal for higher on-target efficiency and minimized off-target activity (106). Internal interactions within the gRNA sequence can lead to the formation of secondary structures that can inhibit Cas9 activity (104, 107). Various algorithms for gRNA design have been developed, calculating both the on-target and the off-target score for each gRNA based on the factors described previously and also on the target organism. Cleavage efficiency has been reported to vary significantly among different organisms and cell lines (108).

## Transcriptional regulation

Inactivation mutations of the two nuclease domains of Cas9 ($RuvC^{D10A}$ and $HNH^{H840A}$) has led to a variant that is called catalytically inactive Cas9 or dead-Cas9 (dCas9). This has converted Cas9 to a programmable RNA-dependent DNA-binding protein, which was initially used as a targeted gene expression downregulation tool (109). This technology was named CRISPR interference (CRISPRi), where dCas9 can be guided by a gRNA to a promoter of interest and block the transcription of the gene. This technology has been further expanded by fusing dCas9 to transcriptional repressors like Mxi (110). This technology has been adapted also for targeted transcriptional activation (CRISPR activation or CRISPRa) by fusing dCas9 with transcriptional activators like VP64 or the tripartite activator VP64-p65-RTA (VPR) which shows even higher transcription activation efficiency (110, 111). An alternative to protein fusion of Cas9 with transcription regulator effectors is the recruitment of them through RNA-binding proteins. In this approach, certain hairpin motifs are added to the sgRNA and the desired effectors are fused with RNA recognising domains. Such an approach is the recruitment of transcriptional activators via MS2 domains and MS2 recognising RNA hairpins added to the sgRNA (112). CRISPRi and CRISPRa have been successfully used for transcriptional reprogramming in *S. cerevisiae* (113, 114).

## Base editing and prime editing

Based on dCas9 and nCas9, another generation of CRISPR-based genomic engineering tools emerged, which are called base editors (BEs) (115). BEs are used for targeted base alteration in a desired genomic locus without introducing double strand breaks. They combine dCas9 (or nCas9) with adenine or cytidine deaminase. Those base editors attracted interest mainly because they were thought to be very promising tools for gene therapy. Therefore, the experimental focus was mainly on human cell lines (115). The first CRISPR-based base editor was a cytidine base editor (CBE), a protein fusion of dCas9 with the rat cytidine deaminase rAPOEBC1, which achieved C→T mutagenesis within the spacer site without DNA cleavage (116). Additionally, adenine base editors (ABEs) were developed, following a similar strategy like CBEs. Fusions of dCas9 with various adenine deaminases yield mostly A→T transversions and the most commonly used adenine deaminase is an adapted version of the *E. coli* TadA (117). Editing windows vary within ABEs and CBEs but typically they are between 5 to 17 base pairs from the PAM site (115–117).

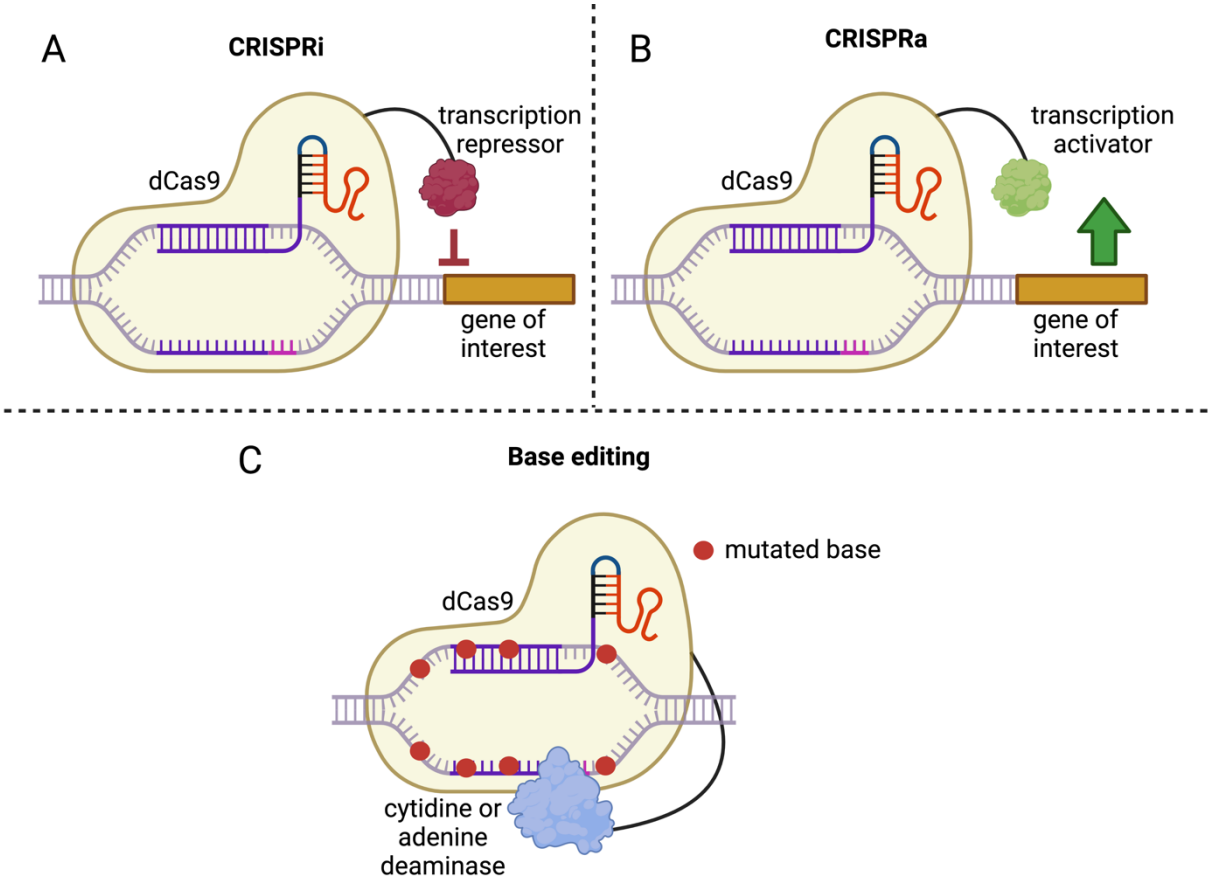


**Figure 5. Genetic tools derived from CRISPR.** In CRISPR interference (CRISPRi) **(A)** dCas9 fused with a transcription repressor can target a promoter of choice and downregulate the expression of the gene that this promoter controls. Alternatively, dCas9 can be fused with a transcription activator (CRISPR activation or CRISPRa) **(B)** and upregulate the expression of a target gene. dCas9 can be fused with a cytidine or adenine deaminase and act as an on-target base editor **(C)** (Figure created with BioRender.com)

Activation-induced cytidine deaminase (AID) is responsible for targeted hypermutation of the variable region of the immunoglobulin locus (118) and it is active on single stranded, actively transcribed DNA (119, 120). Nishida *et al.* (121) wanted to verify if a targeted base editing system based on AID can work in *S. cerevisiae.* They tried a sea lamprey ortholog of AID named PmCDA1 in fusion with dCas9 and nCas9 and they verified that it mutated G and C bases in an editing window from -13 bp to -20 bp from the PAM site. In parallel with this study, two more studies of Ma *et al.* (122) and Hess *et al.* (123) explored the application of human AID in targeted mutagenesis in mammalian cells. A hyperactive version of AID (AID*Δ) rendered a-100 bp mutagenesis window (+/- 50 bp from the PAM site) using a single gRNA, when AID was recruited to the genetic locus via RNA-mediated recruitment (123).

Another CRISPR-based tool has been first developed in human cells by Anzalone *et al.* (124) and it is called prime editing (PE). This tool combines nickase nCas9 fused with a reverse transcriptase. The desired edit is introduced via an extension of the gRNA. The edit is then introduced to the nicking site via reverse transcription. This method introduces deletions up to 44 bp, insertions up to 80 bp, and all kinds of base edits in the nicking site (115). PE is mostly focused in correcting insertions, deletions and point mutations in human cells associated with genetic diseases such as sickle cell disease and Tay-Sachs disease (124)

CRISPR base editing enabled small scale genomic alterations without the need for introduction of a donor DNA repair fragment. Moreover, the dCas9-AID*Δ hyperactive variant showed a base editing range in a window around 100 bp around a gRNA binding site, something that showed that CRISPR can potentially be used in directed evolution applications. Other CRISPR tools for directed evolution have also been developed and they will be presented in a following section of the chapter.

## Multiplexed gRNA expression

Simultaneous expression of multiple gRNAs will expand the possibilities of all CRISPR applications, because it will allow targeting of multiple genomic loci. Expression of multiple gRNAs has been done initially by expressing each gRNA individually (125). Such an approach in yeast has achieved simultaneous deletion of two genes with efficiency up to 100% (126). Another gRNA multiplexing strategy is expression of all the gRNAs in a single transcript flanked by RNA cleavage sequences. Such method is Csy4 multiplexing. Csy4 is an RNA endonuclease present in some native CRISPR systems and it recognizes a 28-bp stem-loop sequence in RNA transcripts and cleaves it after the 20$^{th}$ nucleotide (127). In yeast, Csy4-mediated gRNA multiplexing has been used for simultaneous deletion of four genes with an efficiency of 96% and successful simultaneous activation of three promoters using CRISPRi (128). Another gRNA multiplexing method is to flank each gRNA with the Hammerhead and hepatitis delta virus (HDV) ribozymes. This strategy has been successfully applied in yeast for simultaneous multiplexing of four gRNAs for CRISPRi (129). gRNA-tRNA arrays are

another multiplexing method, which exploits the endogenous mechanism of the cell for tRNA maturing. gRNAs are flanked with pre-tRNA genes, which are recognized and cleaved by endogenous ribonucleases. In yeast, a toolkit using gRNA-tRNA arrays called Lightning GTR-CRISPR has been developed and it achieved simultaneous disruption of four genes with 96% efficiency and six genes with 60% efficiency (130).

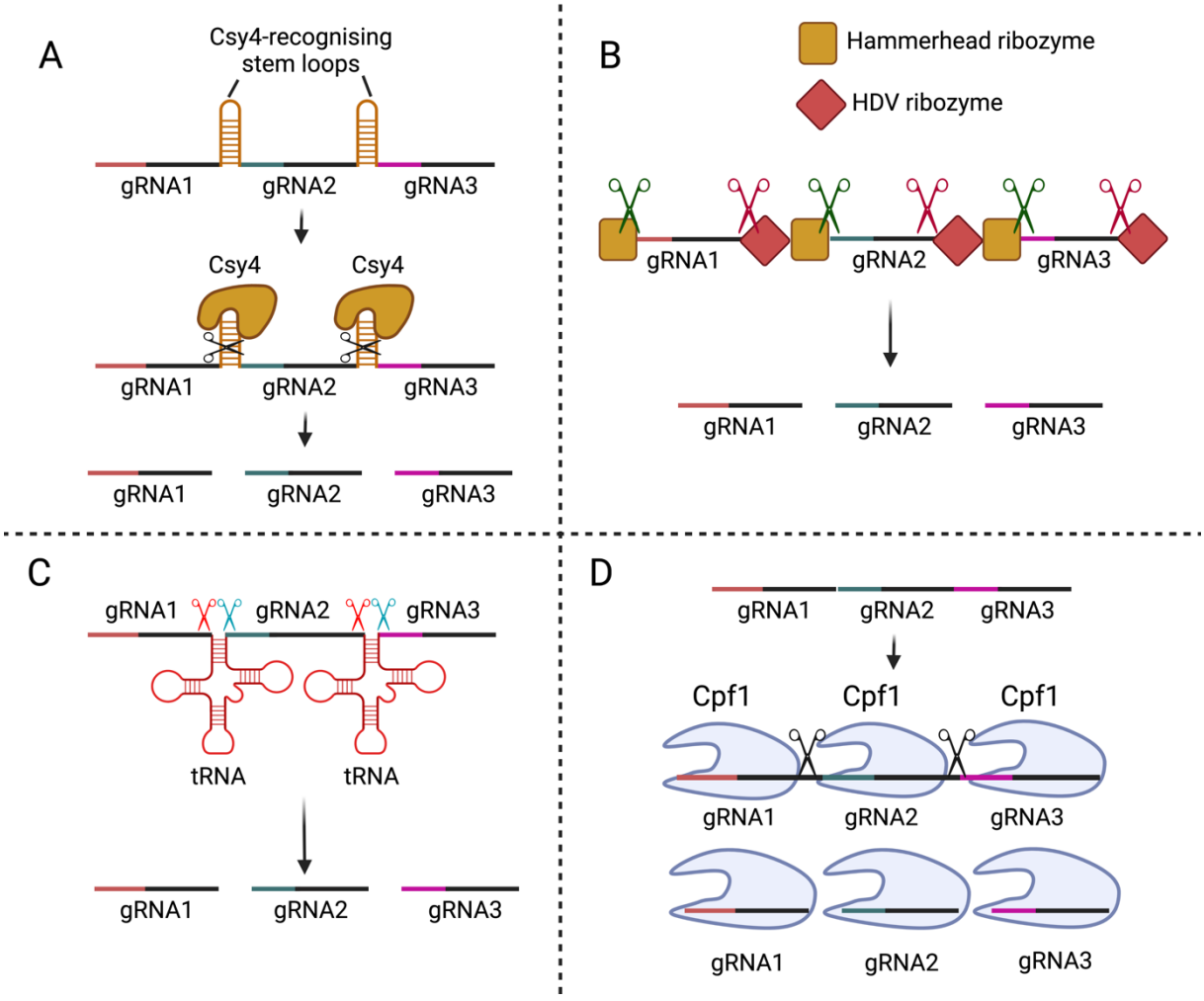


**Figure 6. Strategies for expression of multiple gRNAs from a single transcript. (A):** The gRNAs can be separated by a stem loop, which is recognized and cut by the ribonuclease Csy4. **(B):** Multiplexing via 3' and 5' self-cleaving ribozymes. The gRNAs are flanked by ribozymes that can be self-cleaved without the action of additional enzymes. **(C):** In tRNA-mediated multiplexing, the gRNAs are separated by a tRNA which is cleaved by endogenous tRNA-processing RNAses. **(D):** Cpf1 multiplexing. Cpf1 is an endonuclease similar to Cas9, which can additionally process transcripts with multiple gRNAs. (Figure created with BioRender.com)

There are also endonucleases that can both mature a long transcript with multiple gRNAs and in parallel have the same function as Cas9. An example is the endonuclease Cas12a (previously named Cpf1), which recognizes and cleaves specific hairpin structures in the RNA transcript resulting in the production of mature individual gRNAs (131). In yeast, it was possible with Cas12a multiplexing to edit up to four genomic loci (132). Cas12a multiplexing in mammalian cells has been successfully used for cleavage of five genes and transcriptional regulation of ten genes (133).

The construction of multiplexed gRNA arrays can be challenging. The repetitive elements that those arrays contain can lead to wrong assembling when we use methods that are based on overlap sequences such as Gibson assembly (134) or homologous recombination. Golden Gate cloning (135, 136) is an assembly method that uses type IIS restriction endonucleases that have distinct recognition and cleavage sites (137). This allows scarless ligation of multiple fragments in a one-pot reaction. McCarty et al. (138) successfully used Golden Gate assembly to construct Csy4 gRNA arrays that have up to 12 gRNAs multiplexed.

In yeast, gRNA multiplexing has been used for editing of multiple genes for metabolic engineering purposes (88, 126, 128, 130), simultaneous regulation of expression of multiple genes with CRISPRi (128) and also fine-tuning of the transcriptional repression effect of CRISPRi with multiple gRNAs targeting the same promoter (138). Recently, Csy4-based gRNA multiplexing arrays have allowed parallel transcription activation and repression of selected genes. This approach was named CRISPRai and it was successfully used for pathway engineering for succinic acid overproduction (139).

## Directed evolution

Directed evolution is commonly used for improving the properties of a desired protein. It consists of cycles of diversification of the gene encoding the protein of interest, followed by screening and selection of improved variants (140). Directed evolution has been used for applications such as discovery of improved antibodies (141, 142) and for the development of enzymes with improved properties (143–145). One of the most common genetic diversification strategies is to mutate the gene of interest ***in vitro*** by error-prone PCR (146, 147). Although it results in a much higher mutagenesis efficiency ($10^{-4}$ mutations per bp) compared to the natural DNA replication error rate ($10^{-10}$ mutations per base pair) (140, 148), it relies a lot on the transformation efficiency of the organism of interest and also on the maximal experimentally manageable library size.

What follows the library construction is usually selection if the protein to be evolved can be connected with some kind of selective pressure. Libraries can also be sorted with methods like FACS if there is a biosensor linked with the levels of the metabolite of interest. Sorting can also take place with droplet microfluidics methods. Clone-by-clone screening can be performed using microtiter plates, a method that usually has lower throughput and it is laborious and time consuming (140).

*In vitro* mutagenesis does not allow continuous evolution experiments and if an additional evolution round is needed, another round of library construction has to be performed (149). Regarding this, ***in vivo*** mutagenesis could be an alternative way for genetic diversification since mutagenesis happens in parallel with cell division. The first *in vivo* mutagenesis systems in microbes relied on random diversification of the whole genome of the organism based on the expression of mutagenic enzymes (150, 151).

However, those methods target the whole genome of the organism. They can cause both lethal mutations or mutations that can bypass the selection or sorting method by creating false positive variants (152, 153). Targeted *in vivo* mutagenesis could be an alternative that could contribute to continuous evolution of genes of interest without the negative effects of off-targeting.

## Targeted *in vivo* mutagenesis

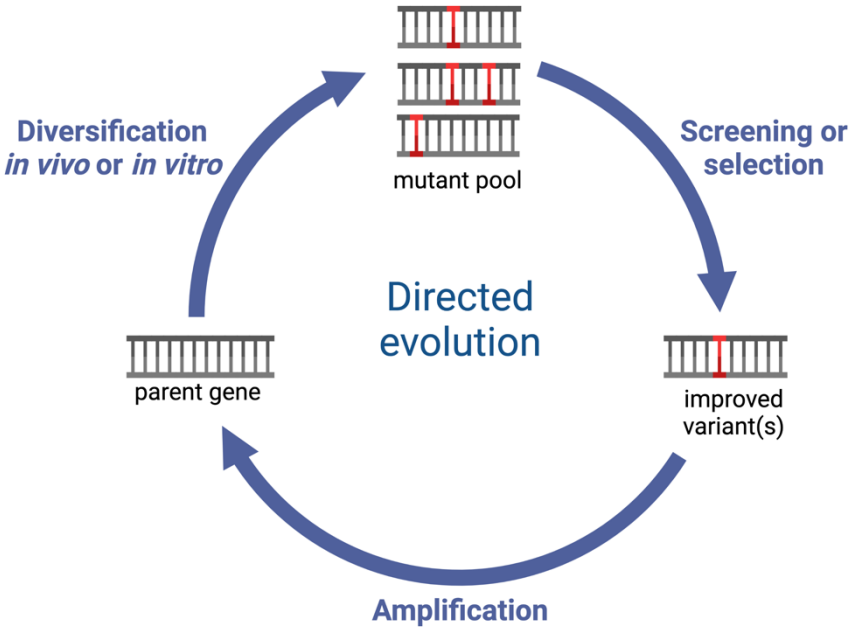


**Figure 7. The cycle of directed evolution.** First, the gene of interest is diversified with either *in vitro* or *in vivo* methods. This creates a mutant pool that can be screened or selected with e.g. growth-based assays. The variants that are enriched can be further tested and one or more improved variants occur. Those variants can be further improved by one or more additional directed evolution cycles.

Throughout the last decade, several systems for targeted *in vivo* mutagenesis have been developed. Those systems can be divided in two categories: those that are based on error-prone transcription or replication and those that are based on adaptations of the CRISPR technology. A summary of the most prominent targeted *in vivo* mutagenesis systems can be found in Table 1.

In the first category fall systems such as *In vivo* Continuous Evolution (ICE), Orthogonal Replication (OrthoRep) and T7 polymerase-driven continuous editing (TRACE). ICE is based on the native yeast retrotransposon Ty1. Ty1 replication is error-prone, so a gene of interest can be inserted into the retrotransposon and be evolved. The method shows similar mutagenesis rates with error-prone PCR, but strain and culturing optimization are needed to reach its best performance (154). OrthoRep connects the replication of an orthogonal plasmid with an error-prone DNA polymerase. A gene of interest can be evolved by introduction in this orthogonal plasmid (155, 156). TRACE combines the T7 transcription system with base editing. The gene of interest was cloned under a T7

**Table 1.** Tools for targeted *in vivo* mutagenesis and their basic characteristics

| Category | Tool name | Organism | Based on | Mutagenesis/ editing efficiency | Mutagenesis window | Ref. |
|---|---|---|---|---|---|---|
| Transcription or replication based | ICE | *S. cerevisiae* | Ty1 retrotransposon | 1.5 x $10^{-4}$ per base | up to 3.5 kb | (154) |
| | OrthoRep | *S. cerevisiae* | Orthogonal plasmid replication by error-prone DNAP | around $10^{-5}$ per base | a whole gene of choice cloned in an orthogonal plasmid | (155, 156) |
| | T7-DIVA | *E.coli* | base deaminases (BD) fused with T7RNAP combined with dCas9 | N/A (only phenotypical frequencies calculated) | a whole gene of choice | (157) |
| | TRACE | mammalian cells | T7-driven transcription by T7RNAP fusion with base editors | up to approximately $10^{-5}$ per base | around 2 kb | (153, 158) |
| CRISPR-based | CRISPR-X | mammalian cells | targeted AID*Δ recruitment by dCas9 and MS2 RNA hairpins | 5-10 x $10^{-4}$ per base | approximately 100 bp per gRNA | (123) |
| | TAM | mammalian cells | dCas9-AIDx fusion | 4 x $10^{-4}$ per base for G and C substitutions | 100 bp when 7 or 10 gRNAs targeting this area expressed simultaneously | (122) |
| | Target-AID | *S. cerevisiae*, mammalian cells | fusion of nCas9 or dCas9 with cytidine deminase PmCDA1 | around $10^{-2}$ for dCas9 and $10^{-1}$ for nCas9 per base | 3-5 bp per gRNA | (121) |
| | EvolvR | *E.coli* | nCas9 fused with error-prone DNA polymerase | around $10^{-5}$ per generation per nucleotide | approximately 60 bp per gRNA | (159) |
| | yEvolvR | *S. cerevisiae* | nCas9 fused with error-prone DNA polymerase | around $10^{-5}$ per generation per nucleotide | approximately 40 bp per gRNA | (160) |
| | CRAIDE | *S. cerevisiae* | Cas9 + extended cgRNA transcribed by epT7RNAP | 3.26 x $10^{-6}$ per base | up to 660 bp per cgRNA | (161) |
| *In vitro* mutagenesis | Taq DNA polymerase | N/A | Error-prone PCR | 2 x $10^{-4}$ - 2 x $10^{-5}$ per base | a whole gene of choice | (148) |

promoter and the T7 RNA polymerase was fused with a hyperactive cytidine deaminase. This method can mutate genes up to 2 kb downstream of the end of the T7 promoter (158).

TRACE is established in mammalian cells but a similar system has been first developed in *E.coli* (153). This system was later further developed by adding a dCas9-gRNA complex that stops the activity of the mutagenic T7 RNA polymerase in an attempt to protect downstream regions that do not need to be evolved. This system was named T7-targeted dCas9-limited *in vivo* mutagenesis (T7-DIVA) (157).

As it was described previously in this chapter, CRISPR was adapted for targeted base editing. As it was revealed in the studies performed mainly by Ma *et al.* (122) and Hess *et al.* (123), when hyperactive variants of the cytidine deaminase AID are used for targeted base editing, the occurring editing windows are enlarged with mainly C and G mutations. Those systems are developed in mammalian cells, but also Nishida *et al.* have developed a similar system (Target-AID) in *S. cerevisiae* which gave high mutagenesis efficiencies in a limited window of 3-5 bp per gRNA (121). EvolvR is another system developed in *E.coli* that combines CRISPR with error-prone replication. The mutator in this system is a chimeric protein consisting of a Cas9 nickase (nCas9) fused with an error-prone DNA polymerase. The chimeric editor is guided to the region of interest by a gRNA and mutagenesis occurs in a distance roughly 40 bp from the nicking site (159). EvolvR has been adapted successfully also for *S. cerevisiae* (yEvolvR) (160).

Recently, Jensen *et al.* have developed another CRISPR-based system for *in vivo* mutagenesis named CRISPR- and RNA-assisted *in vivo* directed evolution (CRAIDE) (161). This system uses, conventional Cas9 and extended gRNAs named chimeric donor gRNAs (cgRNAs). A cgRNA consists of the conventional targeting gRNA plus an extension of 600 bp, which is homologous with the target sequence. The cgRNA is transcribed by an error-prone T7 RNA polymerase and acts as donor for the DSB introduced by Cas9. This tool resulted in mutagenesis in a region of maximum 600 bp but in a mutagenesis rate lower than error-prone PCR or other targeted *in vivo* mutagenesis methods ($3.26 \times 10^{-6}$ mutations per base).

In summary, the systems based on transcription or replication by synthetic promoters or retrotransposons can show similar mutation rates with *in vitro* mutagenesis methods such as error-prone PCR. With those systems though, native genes cannot be targeted directly on the chromosome and evolution cannot be focused in certain areas of the gene, for example in the catalytic sites of an enzyme. CRISPR-based methods have more narrow editing windows but they can be more focused in certain genetic loci of interest.

## Synthetic biology and standardization of genetic parts

The aforementioned approaches regarding strain engineering, high throughput library screening and all the different CRISPR and *in vivo* mutagenesis tools, require extensive workload of molecular cloning. Apart from traditional molecular cloning, advanced technologies such as Gibson cloning (26, 134) have emerged and have allowed faster assembly of extensive molecular constructs. Gibson cloning allows for rapid assembly of multiple genetic fragments with overlapping regions. Type IIS restriction enzymes have distinct recognition and cleavage sites (137), something that allows for additional flexibility in cloning. As it was mentioned before, Golden Gate cloning takes advantage of the properties of the type IIS restriction enzymes and allows rapid one-pot molecular cloning.

The advances in molecular biology that were also described in the beginning of the chapter, have given rise to a distinct field named **synthetic biology**. There is no solid consensus on a commonly accepted precise definition of the term, but its most precise description is "the use of molecular biology tools and techniques to forward-engineer cellular behaviour" (162). It seems to overlap the definition of metabolic engineering, but in particular synthetic biology can be considered as a tool that can facilitate metabolic engineering (163). Examples of synthetic biology are the development of genetic circuits that can alter the transcriptional control of desired genes (162) or the development of cell-based programmable biosensors (164) such as transcription factor-based biosensors which were described earlier in this chapter. The various CRISPR-derived tools and targeted *in vivo* mutagenesis tools can be also considered as examples of synthetic biology. Standardization of the various genetic parts (e.g. promoters, coding sequences, terminators, regulators etc.) would be a valuable contribution to synthetic biology, but it is still in its early stages (165). Among of the first standardization approaches in synthetic biology were BioBricks and BglBricks (166, 167). Those were the first collections of genetic parts combined with a standard cloning routine.

Later on, Modular Cloning (MoClo) toolkits were introduced, first in *E. coli* (168) and later in *S. cerevisiae* (169). Those systems are organised basically in three levels. The different genetic parts are cloned and organised at the first level (level-0), which consists of the so-called part plasmids. The parts have type IIS recognising sequences, which enables them to be assembled by Golden Gate cloning. The parts can be assembled as functional expression cassettes in the second level (level-1) plasmids. Multiple cassettes can be assembled together in the third level (level-2) plasmids. As a result, MoClo systems can facilitate the construction of different types of synthetic biology constructs with minimizing the need of several PCR and cloning rounds. They also allow the different combinations between genetic parts, something that can shorten the construction time and allow for testing of a higher number of molecular constructs.

## Scope of the thesis

The continuous advances in yeast biology and metabolic engineering are opening the way to new possibilities and new applications regarding the use of yeast as a cell factory. The possibilities of yeast-based production of more industrially relevant compounds can be explored with an increasing variety of tools and strategies. At the same time, those advances can contribute to the improvement of the production of compounds that are already made by yeast cell factories. The Design-Build-Test-Learn cycle sums up the systematization of strain engineering. High-throughput screening can speed up this cycle since it enables the fast testing of large genetic libraries.

CRISPR/Cas9 was initially adapted almost a decade ago as a novel molecular tool that can facilitate conventional genomic engineering such as gene deletion or integration of heterologous genes in the chromosomal DNA. Since then, the applications of CRISPR have been expanded to many kinds of targeted genetic manipulation such as transcriptional regulation, base editing and directed evolution. CRISPR can be the basis to develop many kinds of DNA-binding proteins that can be targeted to a desired genomic locus by a gRNA, which is determined by the user. Moreover, gRNA multiplexing enables the simultaneous targeting of more than one genetic locus, something that allows the use of CRIPSR for more advanced experimental settings.

This thesis aims to combine the variety of the applications of CRISPR-derived tools with high-throughput methods that can be used to screen gRNA libraries. This can open the way to multiple types of screenings and applications. Simplification of molecular cloning for CRISPR applications will facilitate those methods as it will make them attractive and user-friendly. Moreover, as we described earlier, CRISPR can also be a powerful tool for targeted *in vivo* mutagenesis.

Those tools can be exploited for directed evolution assays in combination with gRNA libraries. This thesis aims to contribute to all the above-mentioned issues by:

- Developing toolkits that facilitate molecular cloning for all CRISPR-based applications (**paper II**).
- Exploring and deeply analyzing the effect of broad-range base editors in yeast (**paper III**).
- Performing gRNA library screenings for transcriptional regulation fine-tuning (**paper I**) and directed evolution of a gene of interest (**paper IV**).

More specifically, in **Chapter 2** the expansion of the already existing yeast MoClo kit (169) for CRISPR applications is presented. This work is also included in **paper II** along with other expansions of the yeast MoClo kit for chromosomal integrations and library construction. A two-plasmid system was developed, which can be used for any CRISPR application. One plasmid can be used to construct and express Cas9 or any other Cas9-derived chimeric protein. The other plasmid serves as a vector for single or multiplexed

gRNA cloning in one step using Golden Gate cloning. In the chapter is also described the development of a computer-based platform for one step multiplexing of up to 12 gRNAs in a single transcript.

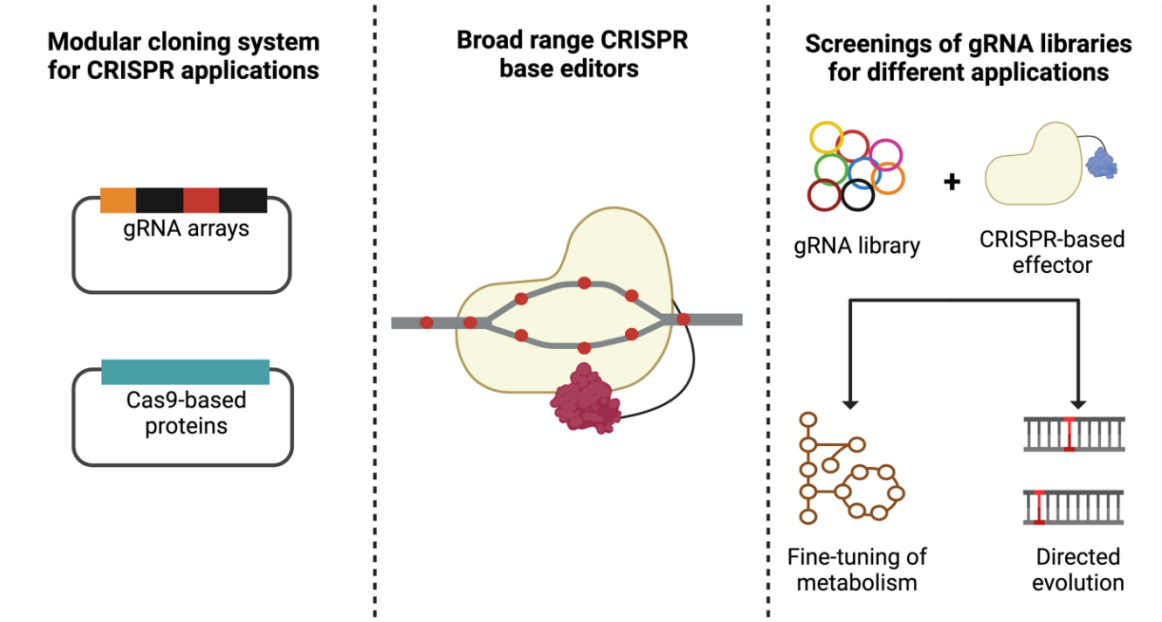


**Figure 8. Graphical summary of the thesis.** The main purpose was to contribute to CRISPR-based high-throughput screening by designing systematized cloning tools, exploring broad range base editors in yeast and applying gRNA library screenings for gene expression fine-tuning and directed evolution.

**Chapter 3** focuses on the development of CRISPR-based base editors for broad range *in vivo* mutagenesis. This work is also incorporated in **paper III**. The mutagenic activity of a protein fusion of dCas9 with the hyperactive cytidine deaminase AID*Δ was characterized in depth. Both phenotypical screenings and Next Generation Sequencing (NGS) analysis were used for this characterization. The base editor was used with one gRNA at a time or multiple gRNAs simultaneously to determine whether gRNA multiplexing has an effect on mutagenesis patterns. A similar -but less in-depth- strategy was followed for characterizing similar base editors using hyperactive variants of the adenine deaminase TadA.

**Chapter 4** includes applications where gRNA libraries were used along with different CRISPR-based tools depending on the biological problem that needs to be solved. **Paper I** presents a gene expression fine-tuning approach with the use of a gRNA library and the transcriptional activator dCas9-VPR in order to improve the metabolic fluxes towards the key metabolite malonyl-CoA. Later in this chapter a directed evolution approach for a malonate transporter is described, which is also incorporated in **paper IV**. For the evolution of the transporter, 60 gRNAs that target the gene encoding it are designed and a combinatorial gRNA library of 3600 variants of 2x multiplexed gRNAs is constructed. This library is used in combination with the broad-range editors constructed in paper III in growth-coupled evolution assays. Promising mutants were identified and

incorporated in the gene with reverse engineering and their beneficial effect was confirmed

# Chapter 2 – Modular cloning for CRISPR applications

The yeast modular cloning (MoClo) developed by Lee *et al.* (169) is a system that enables the organization of the different genetic parts in eight main categories, and their Golden Gate based assembly into single cassette and multicassette vectors (Figure 9). This system allows the quick combination of different genetic parts (promoters, terminators, coding sequences etc.) without the need of design of new primers and PCR cycles for every new cloning.

As it was described previously, CRISPR can be adapted for various uses beyond introducing double stranded breaks in the DNA. Single strand cutting nCas9 or non-cutting dCas9 can be fused to other protein domains with functions such as transcriptional regulators or base editors. This feature can create chimeric protein effectors that can be guided to a genetic locus of choice via gRNAs.
MoClo can shorten the engineering time of CRISPR effectors since the different domains can be efficiently combined by Golden Gate cloning and the need for PCRs and primer design for each construct can be minimized. In **paper II**, we adapt the yeast MoClo system for CRISPR applications. We propose a two-plasmid based system, with one plasmid designed to express the Cas9-based effector and the other to express one or multiple gRNAs. In this chapter are presented the main findings of the study. Technical details and additional results can be found in the paper and its supplementary material.

In the introduction of the thesis, several ways of expressing multiple gRNAs were discussed. More than one gRNA can be expressed via a single transcript which undergoes a maturation process and the single gRNAs can be released. In those transcripts, gRNAs are separated by a repetitive sequence which is recognised by ribonucleases. Moreover, the tracrRNA (referred to also as scaffold) is also a sequence common for all gRNAs. Those tandem repeats make the synthesis and cloning of those arrays challenging. In our lab, we have observed challenges in constructing such arrays with methods like Gibson cloning or classical restriction/digestion. The distinct recognition and cleavage sites of type IIS restriction enzymes can offer an alternative that will allow the construction of multiple gRNA arrays.

## Adapting MoClo for CRISPR

Our goal was to build a system of two plasmids that are compatible with the MoClo modular cloning approach. In this system the CRISPR effector of choice is cloned in one plasmid and the gRNA expression cassette on the other.

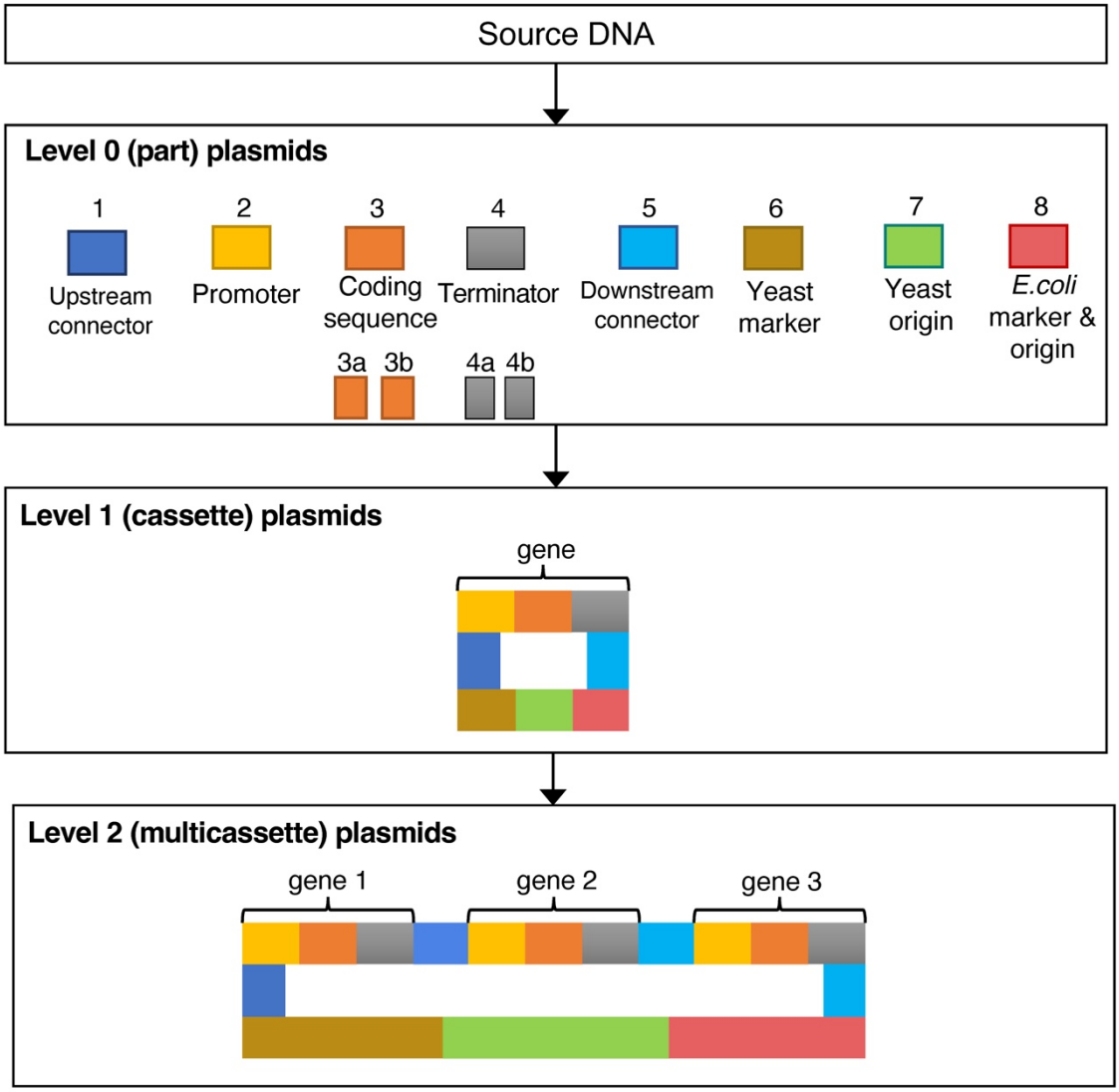


**Figure 9. The three organization levels of the yeast MoClo system.** Level 0 plasmids are the part plasmids, organized in 8 different types. Any source DNA coming from DNA synthesis or PCR can be introduced as a part. The parts 3 and 4 can be split in two sub-parts (a and b) and this feature can be used for building chimeric proteins. The 8 parts can be combined and form a plasmid, which is the level 1 plasmid containing one gene expression cassette. Up to four expression cassettes can be merged into level 3 plasmids.. Adapted from Lee *et al.* (169).

A gRNA cloning vector compatible with the type IIS restriction enzyme BsmBI was constructed (Figure 10A). The cleavage sites are located in the polIII promoter *SNR52*p and a tracrRNA sequence. Between the promoter and the tracrRNA is a marker gene, which can be the *yeGFP* gene (pMCL9 vector) or the gene *ccdB*, which is toxic for *E. coli* (pMCL11 vector). *E. coli* colonies that express yeGFP have a light green colour that is visible even in natural light. When *E. coli* cells are transformed with a plasmid that is expressing the *ccdB* killer gene (170), bacterial cell growth is blocked. This vector is therefore ideal for gRNA library construction. as the occurrence of false-positive clones is minimized.

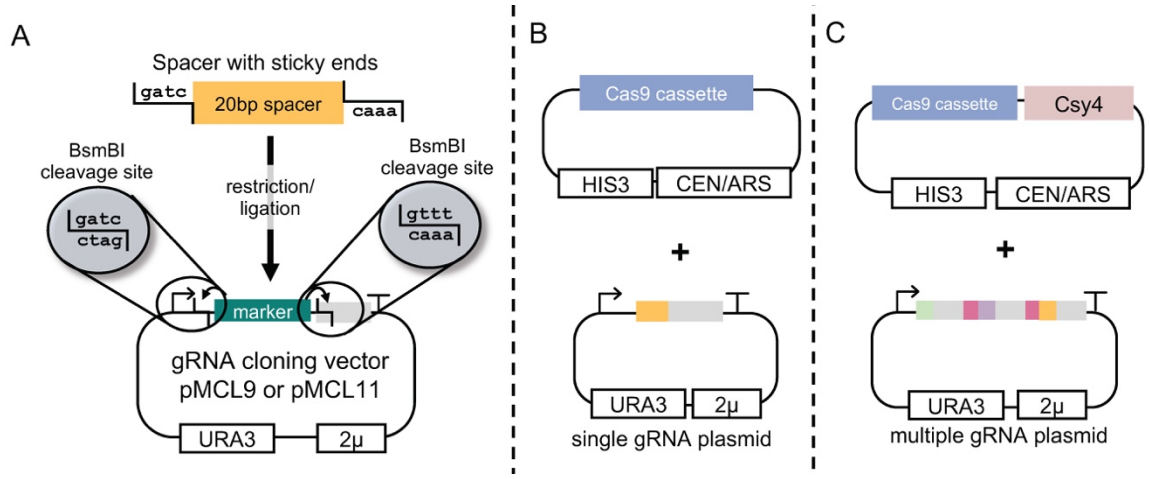


**Figure 11. Overview of the MoClo-compatible two-plasmid system for CRISPR applications. (A):** The 2µ gRNA cloning vector has two BsmBI cleavage sites flanking a marker gene, which can be either *yeGFP* (pMCL9) or the *E. coli* toxic *ccdB* (pMCL11). One single 20-bp spacer can be cloned in this vector or even multiple gRNAs. **(B):** A single gRNA plasmid can be combined with a centromeric single cassette plasmid that contains a Cas9 cassette of choice. **(C):** A multiplexed gRNA plasmid can be combined with a centromeric single cassette plasmid that contains a Cas9 cassette of choice and the *csy4* gene.

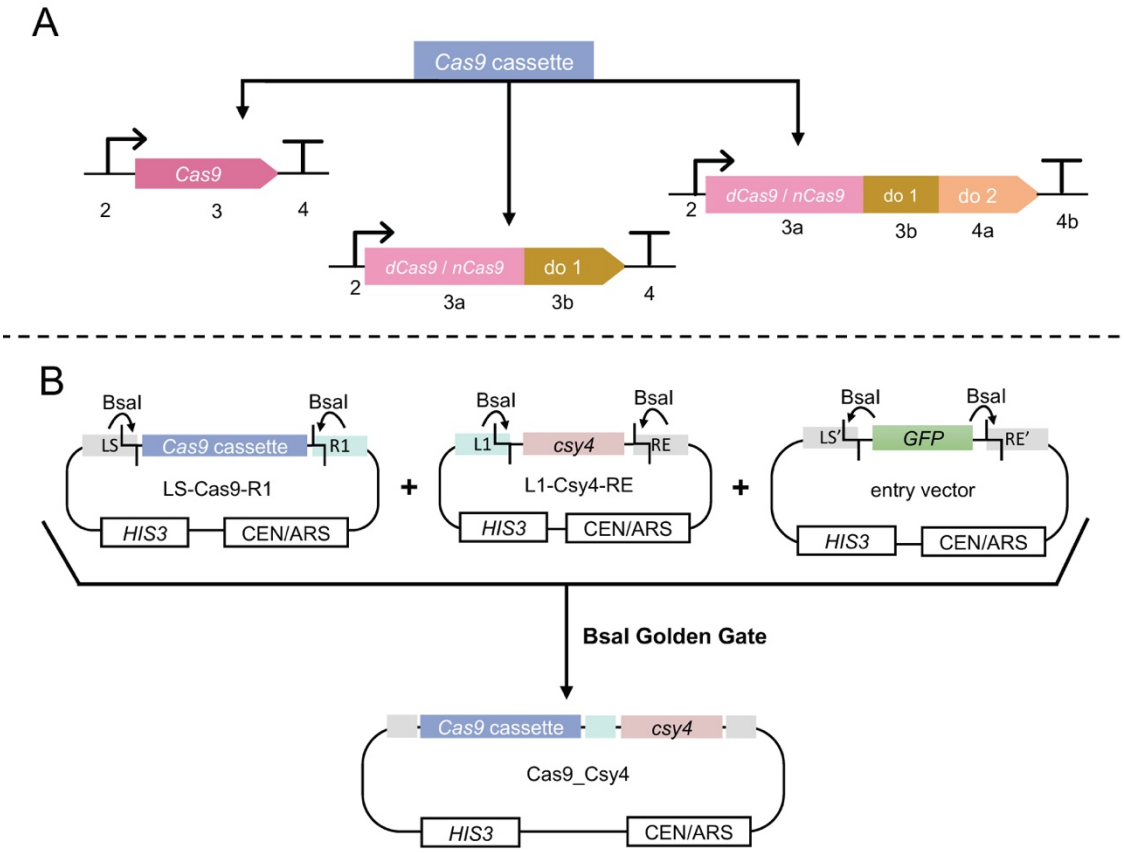


**Figure 10. Cas9 cassette and option for Csy4 expression for multiplexed gRNA expression. (A):** Cas9 cassettes can be built following the MoClo principles. Cas9 can be cloned standalone, or fusion expression cassettes based on dCas9 or nCas9 can be built with either one domain (do 1) or two domains (do 1 and do 2). **(B):** The Cas9 cassette and the *csy4* gene can be fused together in one multicassette plasmid in case a Csy4-multiplexed gRNA array is to be used.

In case a single gRNA is introduced in the gRNA cloning vector, the plasmid can be used along with a Cas9 expression plasmid (Figure 10B). The Cas9 cassette can encode either just a standalone *cas9* gene or a fusion expression cassette with dCas9 or nCas9 fused with one or two domains of choice (Figure 11A). Those plasmids can be built easily following the MoClo principles.

In case a Csy4-multiplexed gRNA array is to be expressed, a multicassette plasmid combining the Cas9 cassette of choice and the *csy4* gene can be used (Figure 10C). The multicassette plasmid can be constructed following MoClo workflow for multicassette level 2 plasmids (Figure 11B).

## Developing a platform for gRNA multiplexing

The tandem repeats of the multiple gRNA arrays make their construction challenging. One of the most efficient methods developed for the construction of those arrays involve two rounds of PCR and cloning (138). Based on the Golden Gate approach we sought to develop a single-round cloning method, which is also computer-aided and compatible with the MoClo-compatible two-plasmid system. The aim was to have a complete CRISPR package which enables easy construction of CRISPR effectors, multiplexing of gRNAs and also flexibility between the different multiplexing methods.

The result is a complete computer-aided toolkit that enables the creation of multiplexed gRNA arrays of up to 12 gRNAs (Figure 12). It is an online tool that is based on BsmBI Golden Gate cloning. The toolkit consists of three components:

1. **gRNA cloning vector.** It should include two BsmBI cloning sites, one site after a promoter of choice and one site just at the beginning of a tracrRNA (scaffold) sequence. The cloning site may contain a counter-selection cassette. In the default set of the tool, the cloning vector is pMCL9, which has the *SNR52*p and a *yeGFP* gene as a counter-selection cassette in the cloning site.
2. **PCR template,** which is a scaffold RNA sequence followed by a gRNA multiplexing sequence. In the default set of the toolkit, the gRNA multiplexing sequence is the 28 bp Csy4-recognisable stem loop and the PCR template is on a plasmid (pMCL8_28bp).
3. **Primer designing online tool MultigRNA** (https://ytkprimerdesign.shinyapps.io/multigrna/) . This is a RShiny app (171) that creates primer pairs for multiplexing up to 12 gRNAs in a single transcript. Its inputs consist of:
    a. The BsmBI cleavage sites of the destination vector, both in 5'->3' direction
    b. The 20-bp spacers (crRNAs) that we wish to multiplex. 2 to 12 spacers can be multiplexed by this application.
    c. The binding parts of the primers (forward and reverse). The binding parts are adapted to the template plasmid pMCL8_28bp, but they can be

changed if the template changes. This feature gives flexibility in the multiplexing method, given that another multiplexing sequence than the Csy4-recognising loop can be used.

The multigRNA tool is designed in a way so that the digested PCR products that occur do not have common BsmBI overhang sequences. The spacers are being cut in such a way that sticky ends with 3 or more nucleotides in common will not be present in the cloning mix. This is to ensure the correct cloning of the gRNA array.

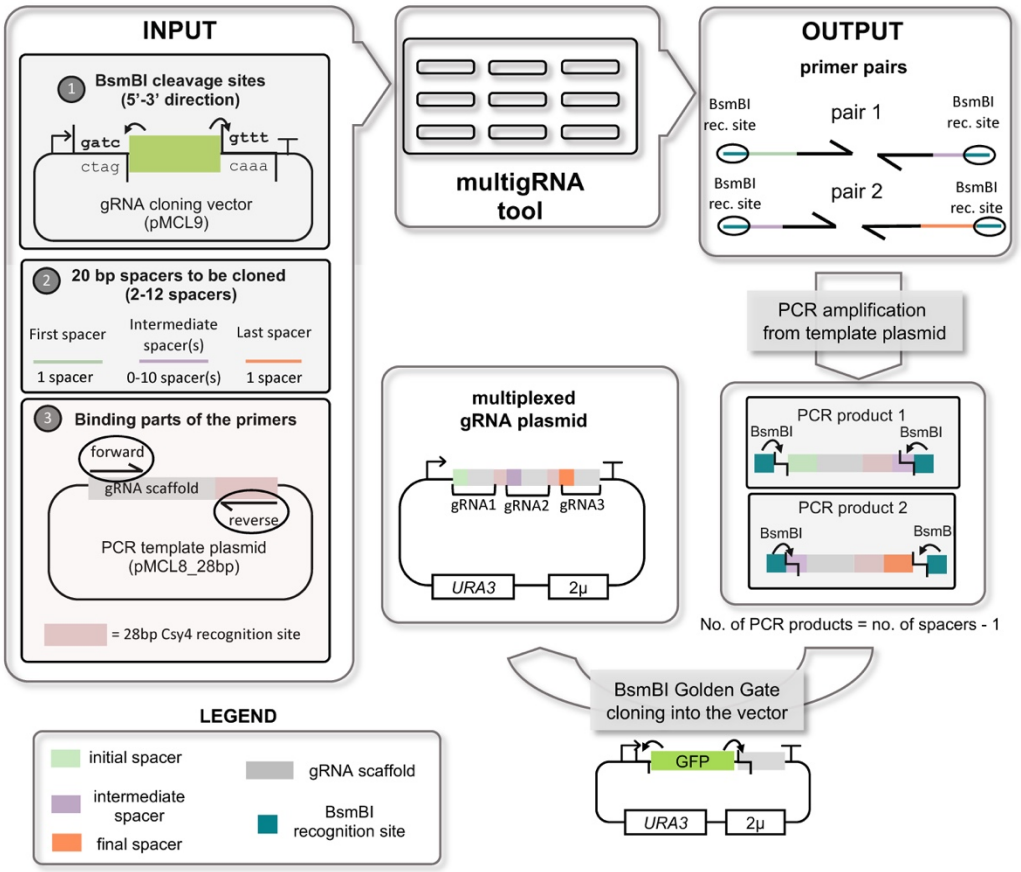


**Figure 12. Overview of the multigRNA tool for automated creation of multiple gRNA arrays.** The tool is designed to build pairs of primers for gRNA multiplexing using as a template a plasmid containing a gRNA scaffold (tracrRNA) followed by a 28-bp Csy4 recognition sequence. The input of the tool consists of (1) the BsmBI cleavage sites of the destination vector, (2) the 20-bp spacer sequences to be multiplexed and (3) the binding parts of the primers depending on the template. The outputs of the tool are pairs of primers (#primer pairs = #spacers – 1). Each primer is used with the template of choice to create one PCR product. The PCR products are purified and used for BsmBI Golden Gate cloning to create the multiplexed gRNA plasmid.

The output of the tool is primer pairs which can be used for PCR amplification using the template. The resulting PCR products are the building blocks of the gRNA array and their number is one less than the gRNAs to be multiplexed. For example, if three gRNAs are to be multiplexed, two PCR products will be needed.

Then, the PCR products are purified and used for BsmBI based Golden Gate cloning using the cloning vector of choice. The assembled plasmids are introduced into competent *E. coli* cells by chemical transformation and the colonies can be screened for successful building of the multiple gRNA array.

To examine the efficiency of the toolkit, I aimed to construct arrays with three (3x), five (5x) and nine (9x) gRNAs. I chose to multiplex gRNAs used for the gene deletions performed for the construction of the fatty acid overproducer strain MLM1.0. (31) (Figure 13). For the 3x and the 5x arrays, four out of four colonies screened had the correct construct, which was verified by sequencing (100% efficiency). For the 9x array, six out of eight colonies screened had the correct construct and they were also verified by sequencing (75% efficiency). The functionality of the toolkit was thus experimentally verified.

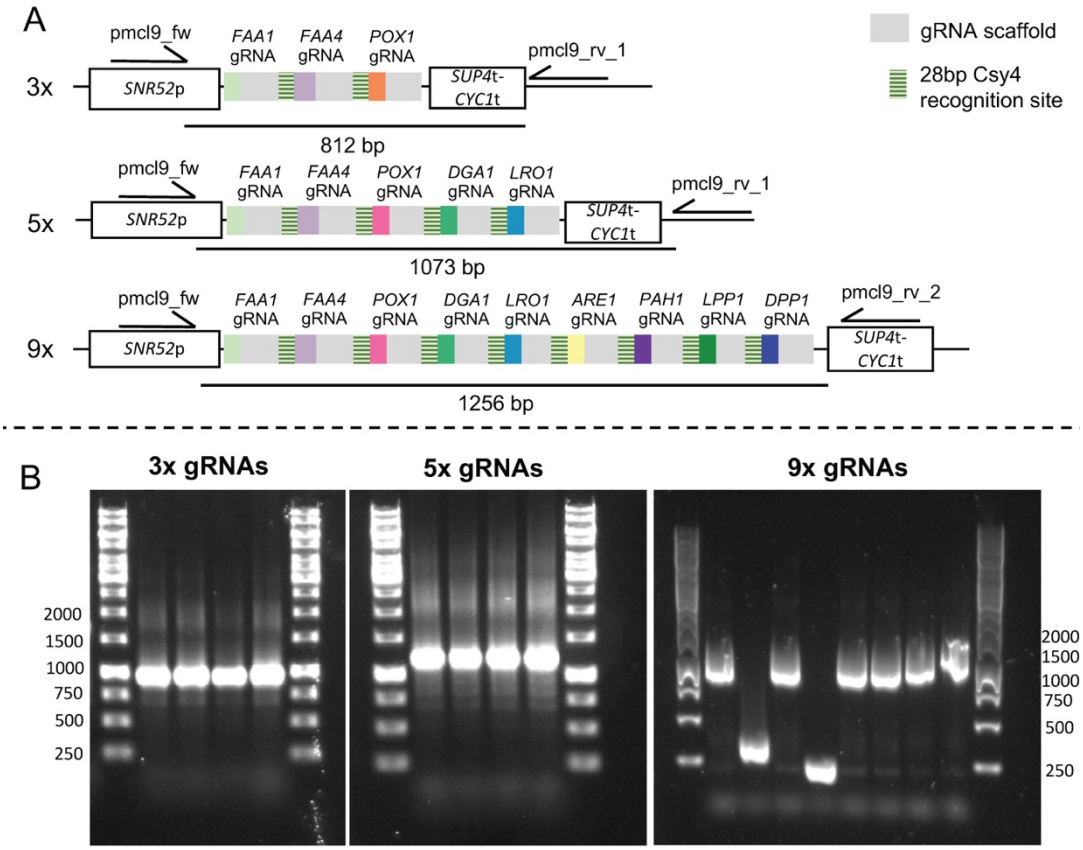


**Figure 13. Efficiency of gRNA multiplexing using the multigRNA tool. (A):** Arrays with three (3x), five (5x) and nine (9x) gRNAs were constructed using the tool. After *E. coli* transformation random clones were screened with colony PCR using the primers indicated in each construct. The expected band size for each construct is also indicated. **(B):** Results of the colony PCRs. For the 3x and 5x constructs four colonies were screened and for the 9x construct eight colonies were screened.

Since the multiplexed gRNAs were successfully used as single gRNAs for gene deletion, I tested the capacity of the constructed arrays to delete multiple genes in one round of transformation. The vector set-up for Csy4-multiplexed gRNAs was followed as indicated in Figure 10C.

A centromeric Cas9 and Csy4 expression plasmid was introduced in CEN.PK113-11C yeast cells. In a second transformation round, the three plasmids with the multiple gRNA arrays (3x, 5x and 9x) were introduced together with the respective DNA repair fragments for gene deletion. For example if the 3x gRNA plasmid was introduced which contained the gRNAs targeting the genes *FAA1, FAA4* and *POX1*, the 120 bp linear repair fragments for those genes was co-introduced as described previously (128, 172). From each multiple gene deletion experiment, eight clones were screened for gene deletions with colony PCRs.

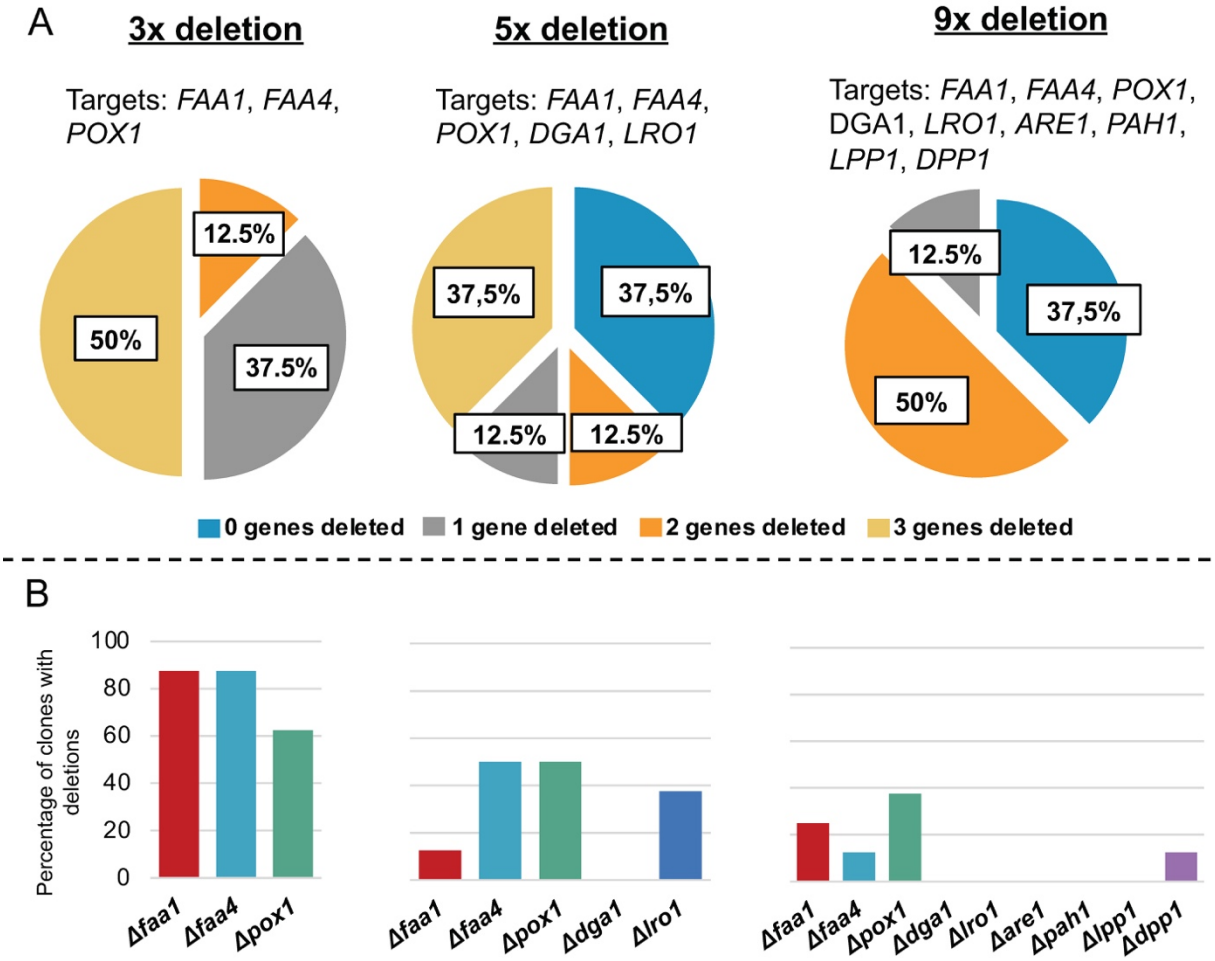


**Figure 14. Overview of the multiple gene deletion capacity of the three multiple gRNA arrays constructed when combined with Cas9, Csy4 and linear DNA repair fragments. (A):** Distribution of clones according to their total number of each deletions in each of the gRNA array used. **(B):** Percentage of clones with detected deletions of each gene for the three gRNA arrays.

The detailed results can be seen in Figure 14. No more than three genes could be deleted simultaneously in any experimental design. The most efficient multiple deletion was with the 3x gRNA array, where 50% of the clones were found with all the three genes deleted. The worst efficiency was observed in the multiple deletion using the 9x gRNA array, where 50% of the clones had only two genes deleted, which was the highest number of gene deletions in a clone. It has to be noted that among the clones transformed with the 5x and 9x gRNA arrays, there was a 37.5% of colonies that had no gene deletions, whereas in 3x array no colonies without a deletion detected.

Moreover, in Figure 14B we can see the deletion efficiencies per gene among the clones transformed with the three arrays, in the order the targeting gRNAs are located in each array. In the multiple deletion experiment with the 3x array all genes are deleted with frequencies over 60%. In the 5x and 9x arrays we can see some extent of deletion for the first three gRNAs of each array and for the last one. However, the deletion efficiencies are lowering when the number of gRNAs multiplexed in a single transcript increases.

Those findings can indicate a bias of gRNA efficiency regarding the position of each gRNA in the array, at least when it comes to gene deletion. In spite of this observation, similar Csy4-multiplexed gRNA arrays have been used successfully for enhanced gene regulation by targeting a promoter with multiple gRNAs (138) or combinatorial activation and repression of genes of choice (CRISPRai) (139). A possible explanation of this difference can be the fact that multiple gene deletion requires multiple DSBs that shall be repaired by homologous recombinations in all the targeted genomic loci. DSBs can be repaired by two mechanisms. Either by homologous recombination with the linear repair fragments we introduce, or by intrinsic DNA repair mechanisms of the cell. So it can be more challenging to achieve multiple successful gene deletions than locating a CRISPR-based effector to multiple genomic positions as it happens in CRISPRi and CRISPRa.

## Conclusions

In this chapter, a novel plasmid architecture was proposed, which makes CRISPR compatible with the yeast MoClo system. It consists of one plasmid for the expression of any CRISPR-based effector and one plasmid for the expression of a single or multiple gRNAs. The MoClo-compatible cloning architecture can enable a faster and more reliable building of CRISPR-based effectors with applications in gene deletion, gene regulation and base editing. Additionally, a computer-aided approach is proposed for the construction of multiplexed arrays with up to 12 gRNAs. This cloning kit aims to minimize the design and cloning time needed using CRISPR applications and facilitates gRNA multiplexing. This study has been also published in **paper II** along with other expansions of the yeast MoClo toolkit regarding library building and chromosomal integrations.

# Chapter 3 – Broad-range CRISPR base editors in yeast

As described in **Chapter 1**, CRISPR gave birth to a new way for on-target base editing. This was achieved by fusing dCas9 with cytidine or adenine deaminases. It enabled precise DNA editing without introducing any breaks and it is considered a powerful tool for gene therapy. The dCas9-mediated recruitment of hyperactive variants of the cytidine deaminase AID has rendered broadened editing regions (up to +/- 50 bp from the PAM site) and evolution of genes such as GFP to the enhanced EGFP (122, 123). Those findings pave the way for the use of broad-range base editors for directed evolution.

In this chapter are presented results of a deep characterization of a CRISPR base editor made by fusing dCas9 and the hyperactive cytidine deaminase AID*Δ. The characterization of the mutagenesis is done by two ways. First by phenotypical screenings and second by analysis of Next Generation Sequencing (NGS) data. We explore the activity of the base editor both with single gRNAs but also with multiple gRNA arrays, in an attempt to explore if gRNA multiplexing has an effect on mutagenesis efficiency.

In addition, we sought to construct broad-range base editors with adenine deaminase instead of cytidine deaminase. We performed a small-scale characterization of two hyperactive adenine deaminase base editors using phenotypical screenings and single clone Sanger sequencing. The findings that are presented in this chapter and also in **paper III.** Here, I present the experimental designs followed for the characterization of the base editors and the main results. Technical details and additional results can be found in the paper and its supplementary material.

## Constructing and exploring a broad-range yeast cytidine deaminase base editor

We constructed the chimeric base editor dCas9-AID*Δ, which is made by fusing the C-terminus of dCas9 with a 100-aa flexible linker followed by AID*Δ. The flexible linker was used by a previous study of a short-range yeast cytidine deaminase base editor, which has been shown to increase base editing efficiency (121).

The base editor was cloned into a centromeric plasmid with *HIS3* marker, which was constructed using the modular cloning approach described in Chapter 2. The base editor was put under the control of the galactose inducible promoter *GAL1*p. The reason for this was to make the system inducible and – thus - to be able to separate mutagenesis phase from selection phase (Figure 15A).

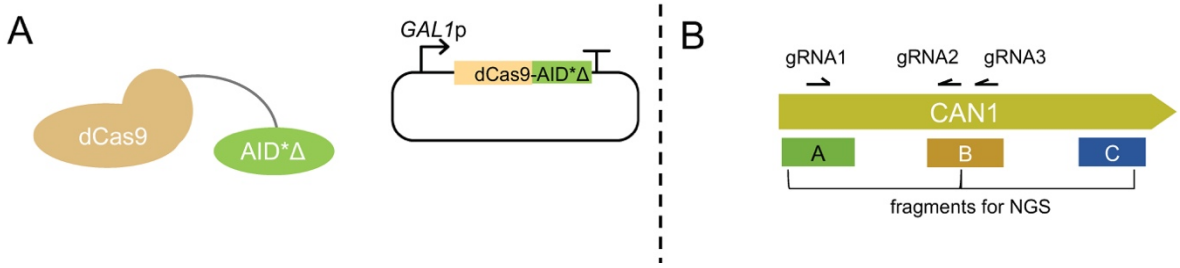


**Figure 15. Overview of the experimental design for the characterization of the base editor dCas9-AID*Δ. (A):** The base editor consists of dCas9 without stop codon with a C-terminus fusionfusion of a 100 aa flexible linker and AID*Δ. The expression cassette is cloned in a centromeric plasmid under the control of *GAL1*p. **(B):** Schematic representation of the *CAN1* gene and the gRNAs designed for the base editor testing, along with their direction. The 200-bp fragments (A, B, C) used for NGS are also indicated.

At a first stage, we aimed to use gene target that would be easily screened and indicate the mutagenesis efficiency of the system. We chose the endogenous yeast gene *CAN1* (total length 1773 bp) as a target, which encodes a plasma membrane arginine permease. Any loss of function mutation of this gene confers resistance to canavanine, an easily screenable phenotypic trait (121, 173).

Three gRNAs were designed to target the *CAN1* gene (also **paper III**: table S1). **gRNA2** (binding position: 787-767 bp) was previously used to characterize the base editor dCas9-PmCDA1 in yeast (121). **gRNA3** (binding position: 826-806 bp) was chosen to be very close to gRNA2 (19 bp gap between them) and in the same orientation in order to check whether multiplexing of gRNAs in close proximity has an effect on mutagenesis. **gRNA1** (binding position: 88-108 bp) is close to the start of the open reading frame (ORF) (PAM site 107 bp from the beginning of the ORF) but it was chosen not to be very close with the promoter in order to avoid potential canavanine-resistant clones that would occur in case of promoter mutation. . The C-terminal part of the gene was chosen to not be targeted by any gRNA in order to be able to screen for the off-target activity of the base editor with NGS (Figure 15B).

Three fragments of a size around 200 bp were defined for NGS. Fragment A covers the binding region of gRNA1, fragment B the binding region of gRNA2 and gRNA3 and fragment C covers the positions 1580-1681 close to C-terminus where no gRNA binds. Fragment C is designed for estimating the off-target effect of the base editor (Figure 15B).
The three guide RNAs were expressed as single constructs but also multiplexed employing Csy4. We constructed two multiple gRNA arrays, the **2x** which contains gRNA1 and gRNA2 and the **3x** which contains all the three gRNAs designed in the study. The plasmid setting used is similar to the one described in Chapter 2 (Figure 10B, 10C). The gRNAs were cloned in a 2μ vector which was combined with a centromeric plasmid harboring the *dCas9-AID*Δ* and *csy4* genes (Figure 16A, 16B).

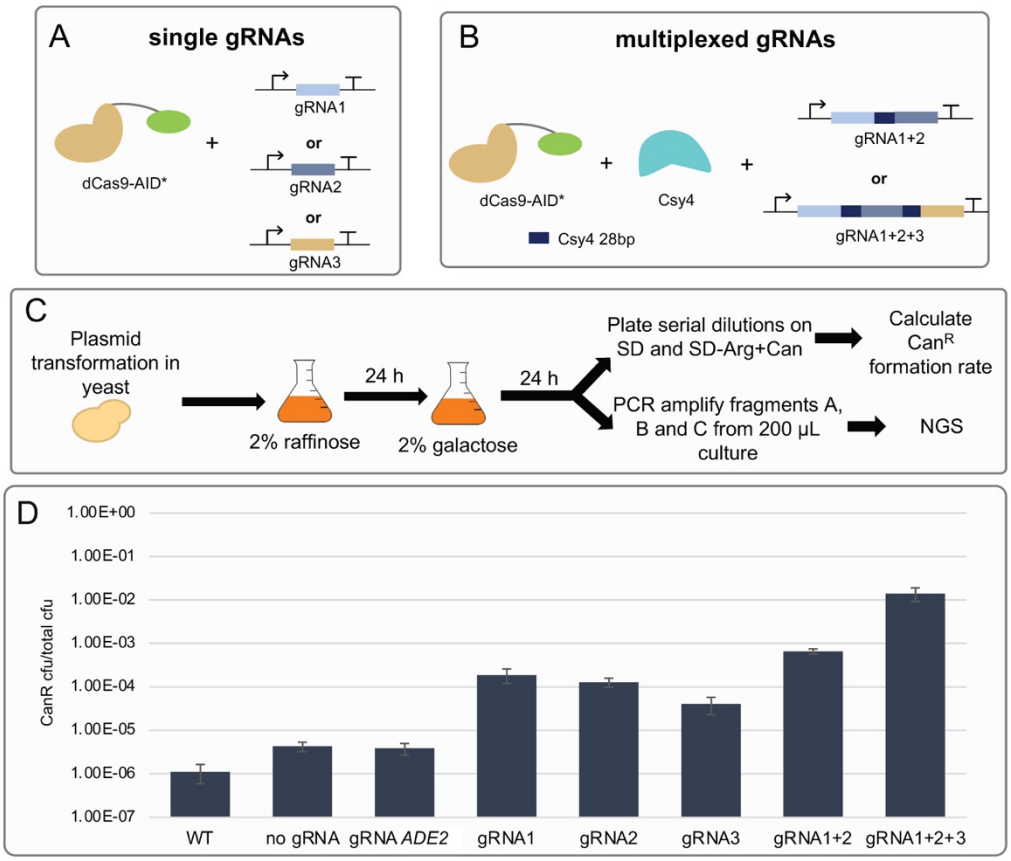


**Figure 16. The experimental assay followed for dCas9-AID*Δ characterization. (A):** Set-up when single gRNAs were used. *dCas9-AID*Δ* was expressed from a centromeric plasmid and one gRNA at a time was expressed from a high-copy 2µ plasmid. **(B):** Set-up when multiple gRNAs were used. *dCas9-AID*Δ* and *csy4* were expressed from a centromeric plasmid and multiple gRNA arrays were expressed from a high-copy 2µ plasmid. **(C):** Cultivation protocol followed after the introduction of the base editor and gRNA plasmids. The cells were precultured in raffinose, then cultured in galactose for 24 h to induce mutagenesis. The Can$^R$ formation rate was calculated by serial dilution plating on selective and non-selective media. Additionally, 200 µL samples were collected from each culture for NGS analysis. **(D):** Can$^R$ formation rate for all the conditions tested. For off-target activity estimation a gRNA targeting the gene *ADE2* was used. Experiments were performed in biological triplicates and the error bar indicates the standard deviation.

After the introduction of the plasmids, a protocol inducing the mutagenesis phase was followed. For every gRNA cassette (single or multiple) the experiments described below were performed in biological triplicates. Delft media was used for all cultures with variable carbon sources. First the cells were precultured in 2% raffinose for 24 hours. The reason behind this is that glucose leads to repression of *GAL1*p even in the presence of galactose. The repression of *GAL1*p by glucose and its activation by galactose are two separate mechanisms (174). This promoter shows also carbon source memory, which makes the activation of *GAL1*p slower when the cells pass directly from galactose to glucose (175). Raffinose is a neutral carbon source and the pretreatment aims to activate more efficiently the expression of dCas9-AID*Δ when galactose is introduced.

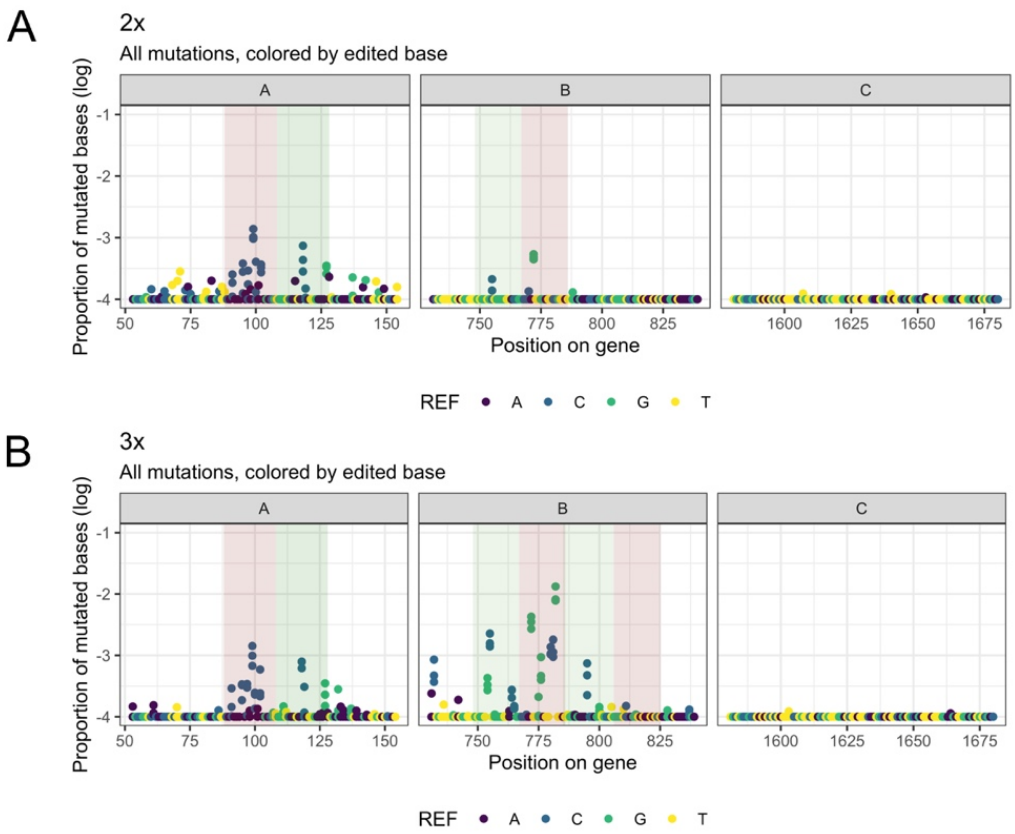


**Figure 17. NGS results of the three fragments of the *CAN1* gene after dCas9-AID*Δ mutagenesis when multiple gRNA arrays were used.** Mutation spectra of the gRNAs 1 and 2 multiplexed **(A)** and all three gRNAs multiplexed **(B)** are shown. The x axis of each graph denotes the gene position and on the y axis the proportion of each mutation over the wild type control in logarithmic scale. The -20 bp region from the PAM site of each gRNA is shown in red and the +20 bp region from the PAM site is shown in green. Each reference base that was mutated is shown with different colour.

Following raffinose preculture, cells were inoculated in 2% galactose media for mutagenesis induction. After 24 h of growth, serial dilutions were plated on the non-selective SD plates and the selective for *CAN1* loss of function SD-Arg+Can (**paper III:** materials and methods) (Figure 16C). The $Can^R$ colony formation rate was calculated as an estimate of the mutagenesis efficiency. It represents the ratio of canavanine resistant colony forming units (cfu) divided by the total cfu number grown on non-selective plates. As an off-target control, we introduced the gRNA that targets the gene *ADE2* (**paper III:** table S1)

The phenotypic screening results are shown in Figure 16D. The background $Can^R$ formation rate in the wild type (WT) strain was around $10^{-6}$ $Can^R$/total cfu which was slightly increased when no gRNA or a gRNA that targets the *ADE2* gene were expressed. In case one of the three on-target gRNAs was expressed alone, the $Can^R$ formation rates were similar for all tested gRNAs, around $10^{-4}$ $Can^R$/total cfu. When the two first gRNAs were multiplexed the $Can^R$ formation rate was increased to $10^{-3}$ $Can^R$/total cfu and when all the three gRNAs were multiplexed it was further increased to $10^{-2}$ $Can^R$/total cfu.

Those data indicate that gRNA multiplexing significantly elevates the mutagenesis efficiency of dCas9-AID*Δ.

In parallel with the phenotypical screening, 200 µL of galactose-grown cultures were taken and subjected to DNA extraction. Then the three NGS fragments A, B and C were PCR amplified and prepared for NGS MiSeq. DNA sequencing reads were aligned to the *CAN1* locus and the proportion of mutated bases was calculated for each position on the *CAN1* sequence (**paper III**: materials and methods).

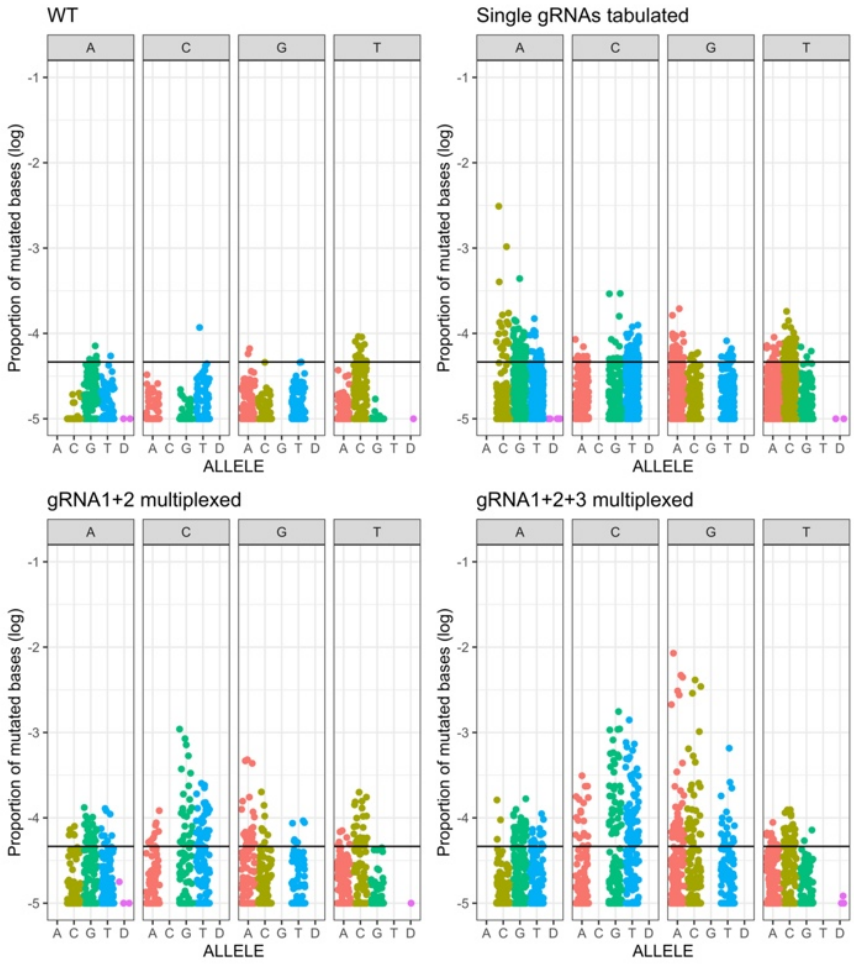


**Figure 18. Base substitution frequencies of the different targeted mutagenesis experiments.** The black line indicates the region in which 95% of the events are located in the WT samples. Plotted is the proportion of mutated bases for each position in the *CAN1* target sequence grouped by base substitution type, with each facet representing the reference sequence base and each color representing the resulting mutated base. D denotes base deletions. Order clockwise: WT samples, tabulated data from all the single gRNA samples, all three gRNAs multiplexed, gRNAs 1 and 2 multiplexed.

When a single gRNA was expressed, the mutagenesis efficiencies were at low levels. In case two distant gRNAs (gRNA1 and gRNA2) were expressed simultaneously, the mutagenesis effect was increased in both loci, in a region +/- 20 bp from the PAM site of each gRNA (Figure 17A). When all three gRNAs were multiplexed, the mutagenesis efficiency was further elevated especially with regard to gRNAs 2 and 3 that are in close proximity. Those gRNAs bind in the same direction and their PAM sites have a distance of 39 bp from each other. The mutation frequency in the B fragment where gRNA2 and gRNA3 are binding was highly increased especially in the region between the two PAM sites (Figure 17B). This finding indicates a synergistic effect when dCas9-AID*Δ binds in two proximal genomic loci.

Additionally, the different mutations were grouped by type of nucleotide exchange for the different gRNA expression schemes (Figure 18). In the case of single gRNAs being expressed, we observed a slight increase in all kinds of base substitutions. When two gRNAs were multiplexed, we observe a further increase mostly with regard to C to G and T and G to A and C substitutions. When all three gRNAs were multiplexed, the C and G mutagenesis rates were elevated even further. The highest increase in mutagenesis rate was observed in the case of G to all other three bases. This could be explained by the fact that the two gRNAs with close proximity to each other bind both to the coding strand since they are reverse complement to the gene sequence. G mutations on the coding strand, can be as C mutations on the non-coding strand. So the prevalence of G mutations can indicate C mutagenesis in the non-coding strand. This, along with the alignment results presented above, could be an additional indication that binding of two dCas9-AID*Δ in close proximity could have a synergistic effect and boost mutagenesis efficiency.

Attempting to have a closer insight on the mutagenesis patterns, we examined the unique mutations that are present in our NGS data on a read-by-read basis. We calculated the number of unique reads in each of our samples compared with the wild type NGS data normalized to sequencing depth. We compared the number of unique reads that occurred in each dCas9-AID*Δ/gRNA combination with the unique reads in the strain that expressed solely the base editor plasmid (Figure 19). We can see that gRNA1 had a site-specific effect when expressed single but gRNAs 2 and 3 do not significantly increase the number of unique reads compared with when only the base editor is expressed. Combinatorial expression of gRNA1 (targeting NGS fragment A) and gRNA2 (targeting NGS fragment B), increases even further the number of unique reads only in fragment A. It remains unclear whether this is a result of poorer binding efficiency of gRNA2 compared with gRNA1. Combining the three gRNAs increases the number of unique reads in both fragments A and B. The increase is more pronounced in fragment B than in fragment A, something that could indicate that maybe the proximity of the base editor binding sites can increase the base editing efficiency. However, the boost of unique reads that is observed in fragment A indicates that the proximity might

not be the sole reason for the enhanced efficiency that is observed in the case of the triple gRNA expression.

Focusing on the 50-bp hotspot region on site B (positions 750-800 bp on the *CAN1* gene), we next sought to quantify the mutation rate for each editor across this region. In order to quantify editing efficiencies across all editors, the proportions of mutated bases were normalized by sequencing depth and were summed up in this hotspot region, with the exclusion of positions with a depth normalized proportion of $10^{-4}$ or less (in order to correct for background noise). These values were then divided by the length of the hotspot region (0.05 kb) to estimate mutation rate per kb. Mutation rates, both normalized by length and non-normalized, are available in Table 2. We see that only in the 3x sample (and two proximal gRNAs targeting the area of fragment B) the mutation rate is almost seven times higher, from around 0.2 mutations/kb to almost 1.5 mutations/kb.

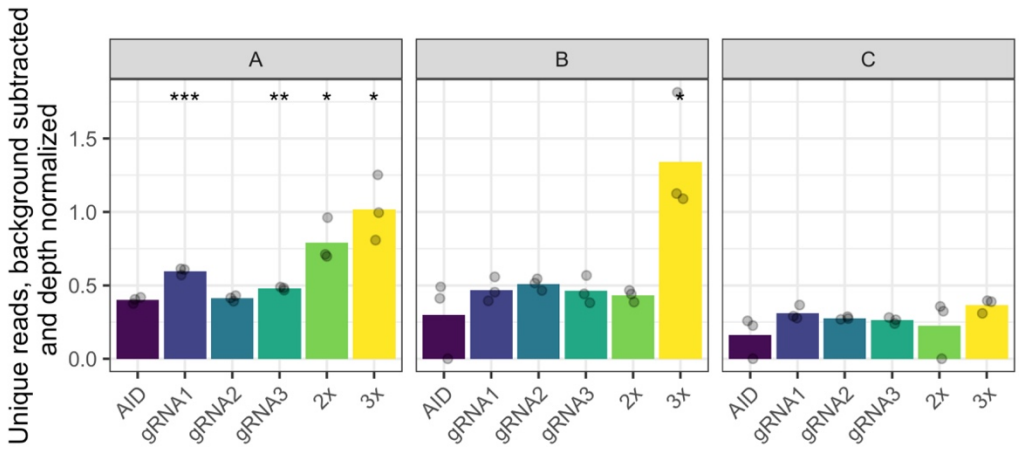


**Figure 19. Number of unique NGS reads having a mutation in the dCas9-AID*Δ experimental dataset.** As unique reads we define the reads that are different from the reads that occurred in the WT strain without base editor and gRNA(s). The number of unique reads was normalized to the sequencing depth of each sequencing run. On the graph is the sample with base editor only annotated as AID, the three single gRNAs, the 2x (gRNA1+gRNA2) array and 3x array. * p-value < 0.05, ** p-value < 0.01 *** p-value < 0.001

**Table 2. Mutation rate on NGS fragment B.** The number of fragment B mutated bases in each sample with proportion higher than $10^{-4}$ were summed up and corrected by sequencing depth (second column). Then they were also corrected by the length of the target region to calculate the mutation rate per kb.

| EDITOR | Mutation rate on site B (no. mutated bases / sequencing depth) | Mutation rate on site B per kb (no. mutated bases / sequencing depth / length of target region [kb]) |
|---|---|---|
| WT | 0.012 | 0.247 |
| AID w/o gRNA | 0.009 | 0.175 |
| gRNA1 | 0.012 | 0.239 |
| gRNA2 | 0.014 | 0.275 |
| gRNA3 | 0.013 | 0.260 |
| 2x | 0.013 | 0.267 |
| 3x | 0.073 | 1.470 |

## Exploring high efficiency yeast adenine deaminase base editors

Base editors based on AID*Δ can create mutations in a relatively extended window, but their drawback is that they mutate almost exclusively C and G bases. In an attempt to find solutions to end up in less biased base editors, we sought to investigate whether similar broad range adenine base editors can be developed in yeast.

We chose to construct two base editors based on the high efficiency adenine deaminase TadA8e, which has been used in adenine base editors (ABEs) in human cells (176). It is a variant of the *E.coli* originated tRNA adenine deaminase(177)was later evolved to the DNA editing enzyme TadA*(117). TadA8e was evolved from TadA* along with dCas9. TadA8e variant has improved Cas9 compatibility and higher editing efficiency, in a window from -4 to -8 bp from the PAM site when used with the *Streptococcus pyogenes* dCas9 variant we also used in all experiments (176). The mutant V106W has been identified to lower the off-target activity of TadA* towards RNA and DNA (178), an effect that was observed also in TadA8e (176). We constructed two base editors for testing, TadA8e-dCas9 and TadA8eV106W-dCas9. Their design is similar with dCas9-AID*Δ but the base editors are N-terminus fused, in accordance with previous studies. The dCas9 and the base editors are connected with the same 100 aa flexible linker used in dCas9-AID*Δ.

The experimental setup for phenotypical screening was the same as followed for the characterization of dCas9-AID*Δ. *CAN1* was again the marker gene and the same three gRNAs were tested. The gRNAs were expressed individually or in multiplexed cassettes, in the same setup that was used in the previous experiment (Figure 20A). The same gRNA targeting *ADE2* was again used as an off-target activity estimate. After plasmid transformation we performed raffinose de-repression and galactose induction of the base editor, followed by $Can^R$ screenings. The results of this assay are shown in Figure 20B.

The first observation to be noted, is that TadA8eV106W-dCas9 along with the off-target gRNA targeting *ADE2* showed the same $Can^R$ formation rate as the WT strain with no base editor. TadA8e-dCas9 with the same direction showed an about 10-fold increase in $Can^R$ formation.

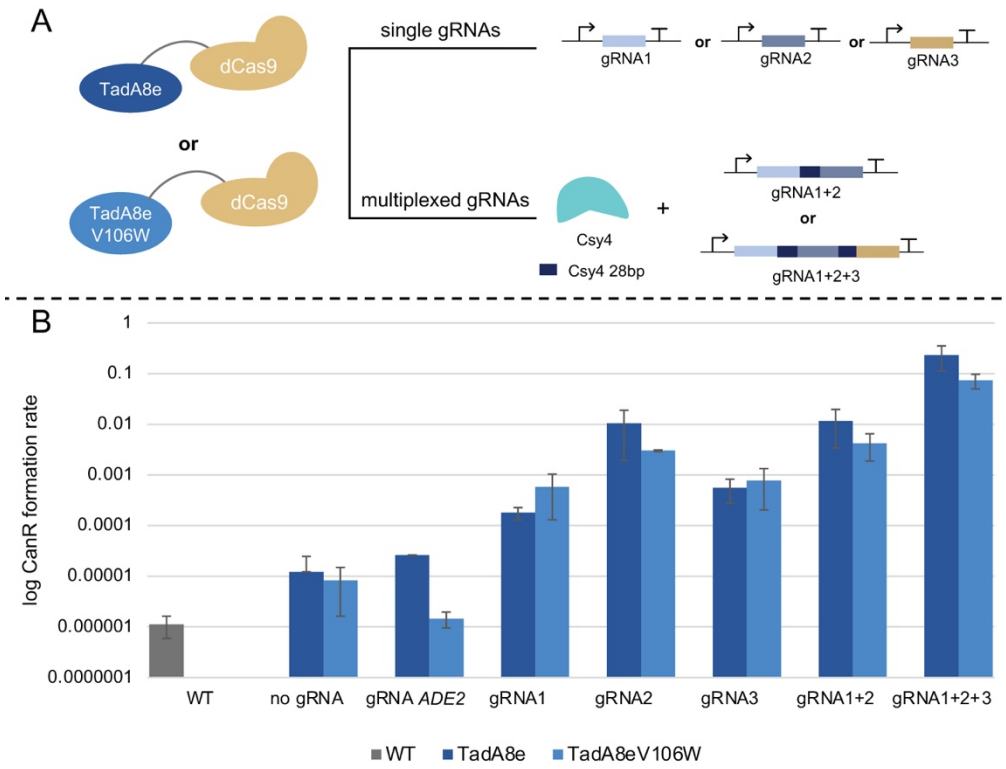


**Figure 20. Experimental assay for the screening of adenine base editors. (A)** The hyperactive adenine deaminase variants TadA8e and TadA8e V106W were fused with a 100 aa flexible linker and dCas9. TadA8e V106W is reported to show decreased off-target activity. The same experimental strategy and the same gRNAs as in the AID*Δ assay were used. **(B):** Can$^R$ formation rates for both TadA8e-dCas9 and TadA8eV106W-dCas9. A gRNA targeting *ADE2* was used for off-target activity estimation.

The three individual gRNAs showed similar Can$^R$ formation rates for both base editors, around $10^{-3}$ Can$^R$ cfu/total cfu. Only gRNA2 showed a bit higher Can$^R$ formation rate that reached $10^{-2}$ Can$^R$ cfu/total cfu, and similar Can$^R$ formation rate was observed when gRNA1 and gRNA2 were multiplexed. Can$^R$ formation rate reached $10^{-1}$ Can$^R$ cfu/total cfu when all the three gRNAs were multiplexed,.

In order to check the mutation window and the types of the mutation, we performed a small-scale screening based on Sanger sequencing of individual clones. We selected a 500-bp sequence in the center of the *CAN1* gene (534-1011 bp) for Sanger sequencing. In the middle of this fragment, gRNA2 and gRNA3 are binding (Fig. 21A). We chose to sequence this region to investigate the base editing window and efficiency when gRNA2 only or gRNAs 1, 2 and 3 are expressed together. We wanted to investigate whether we see the same boost in editing efficiency as in the case of dCas9- AID*Δ when the gRNAs are in close proximity.

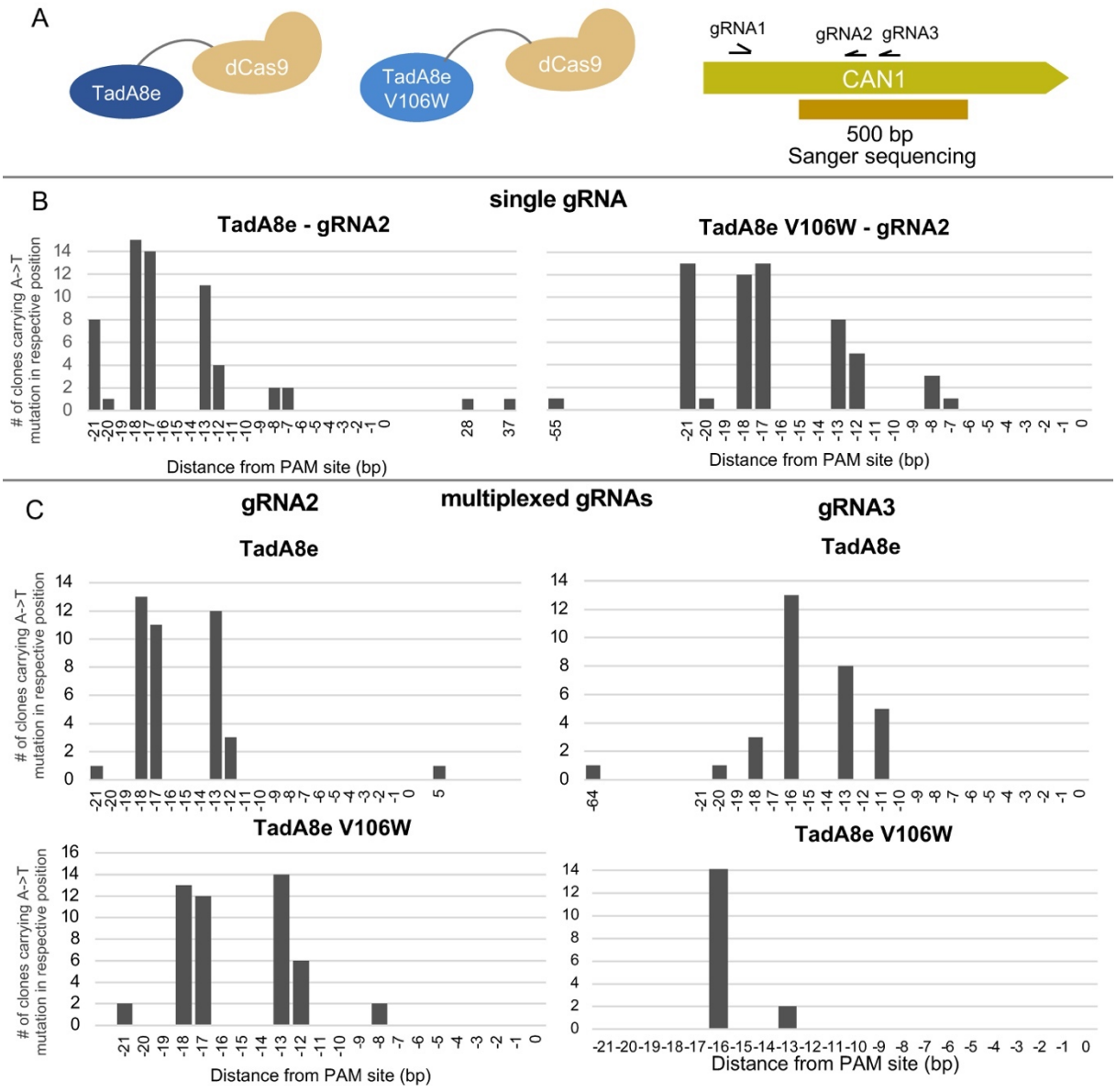


**Figure 21. Mutagenesis profile of the base editors TadA8e-dCas9 and TadA8eV106W-dCas9 in yeast. (A):** A 500 bp part around the binding sites of gRNA2 and gRNA3 was amplified from individual Can$^R$ clones and each PCR product was Sanger sequenced. For each condition 15 clones were screened. **(B):** A→T mutation frequencies for both base editor variants when a single gRNA (gRNA2) was expressed. **(C):** A→T mutation frequencies for both base editor variants in gRNA2 and gRNA3 loci when gRNAs 1, 2 and 3 were multiplexed (3x gRNA array).

We screened 15 Can$^R$ clones from the strains expressing the adenine deaminase base editors and gRNA2, and 15 Can$^R$ clones from the strains with the base editors and the triplicated gRNAs. Mutations were almost exclusively detected in the gRNA binding site, and the only kind of mutations observed were T→C substitutions in the coding strand or A→G substitutions in the non-coding strand. In Figure 21B we can see the editing window of gRNA2 when TadA8e-dCas9 or TadA8eV106W-dCas9 are expressed. Editing starts at the position -7 from the PAM site and the highest mutation rate observed is at the position -18. In the case of TadA8eV106W-dCas9, high mutation rates are observed also at the position -21. Moreover, in the case of TadA8e-dCas9 two single mutations in

higher distance from the PAM site are observed at the positions +28 and +37 from the PAM site.

When the two proximal gRNAs 2 and 3 are expressed simultaneously, the positions of gRNA2 that are the closest to the PAM site are no longer mutated, but the mutagenesis rate of the positions -12/-13 and -17/-18 remains almost the same. Strains expressing gRNA3 show a similar mutagenesis profile, especially when TadA8e-dCas9 is expressed. In the case of TadA8eV106W-dCas9, only two positions in the gRNA3 binding site are mutated, and mostly position -16. The V106W mutation lowers significantly the off-target activity of TadA8e-dCas9 but it also seems to lower the on-target mutagenesis efficiency, especially in the context of multiplexed gRNA expression. The gRNA3 has slightly lower on-target score than gRNA2 (59.3 vs. 62.4 respectively, see also **paper III**: table S1) but it is unclear if this difference can fully explain the low mutagenesis rate observed in gRNA3 position when multiplexed gRNAs were expressed along with TadA8eV106W-dCas9.

## Conclusions

In this chapter we aimed to construct and characterize broad range CRISPR base editors for targeted *in vivo* mutagenesis. The goal was to investigate their activity both phenotypically and at the DNA level, their mutagenesis window and their mutagenesis efficiency.

The cytidine deaminase base editor dCas9-AID*Δ with single gRNAs showed poor mutagenesis efficiency in NGS analysis, although it increased about 100 times the occurrence of $Can^R$ clones in the canavanine resistance assays. Multiplexed expression of two distant gRNAs (distance between PAM sites 659 bp) increased the mutagenesis efficiency in both loci, forming a mutagenesis window of +/- 20 bp from each PAM site. The expression of two proximal gRNAs with 19 bp distance between them, resulted in an increase both of the editing window and the mutagenesis efficiency in the site. The expression of those two proximal gRNAs in this area renders a mutagenesis window of about 80 bp in total. This finding is an indication of a possible homodimerization effect of AID*Δ which boosts up its efficiency. It has been already shown that AID forms dimeric or even multimeric complexes (179).

We also performed a smaller scale characterization of two high efficiency adenine deaminase base editors, previously used in human cells. We showed that they work efficiently in yeast cells but their effect appeared to be restricted to the gRNA binding sites, even when the two proximal gRNAs that boosted up dCas9-AID*Δ efficiency were expressed simultaneously. However, since only $Can^R$ clones were sequenced, NGS analysis could give a better insight on the profile of the adenine base editors we developed. The base editors dCas9-AID*Δ and TadA8eV106W-dCas9 were later used for directed evolution assays and this is presented in **Chapter 4** and **Paper IV**.

# Chapter 4 – Combining CRISPR tools and gRNA libraries

In this chapter we explore different applications of gRNA library screening combined with different CRUSPR tools. Two projects are presented with a common aim to improve the supply of the key metabolic precursor malonyl-CoA in yeast. The first project is a gRNA library screening assay covered also in **paper I**. It combines a single gRNA library and the transcription activator dCas9-VPR. The aim of this project was to screen for altered transcription levels of genes that could increase the yeast metabolic fluxes towards malonyl-CoA. The gRNA library in this project was combined with a malonyl-CoA biosensor and screened by FACS.
The other project is included in **paper IV** which it involves the improvement of malonyl-CoA supply by external malonate. We sought to improve malonate import in the cell by evolving a dicarboxylic aid transporter. Broad range base editors that were constructed and characterized in **chapter 3/paper III** were used along with a duplexed gRNA library for the evolution of the gene that encodes the transporter. In this case the library was not screened but a growth-coupled assay was performed.

## gRNA library screening combined with dCas9-VPR for the screening of gene transcription setups

In this project, we sought to combine the targeted gene regulation that CRISPRi offers, with a high-throughput screening setup. Metabolite biosensors connect the intracellular levels of a metabolite as an input with defined outputs like fluorescence and they can be combined with screening methods like FACS. We selected the malonyl-CoA **FapR biosensor** (45–47) which connects malonyl-CoA levels with GFP expression (see also Figure 2). The biosensor is designed in a way that increasing cytosolic malonyl-CoA levels lead to increased GFP expression. Malonyl-CoA is a key precursor for valuable chemicals like fatty acids, 3-hydroxyproprionic acid (3-HP) etc. (180).

We wanted to investigate genes that changes in their transcription levels could increase the metabolic fluxes towards malonyl-CoA. dCas9-VPR can upregulate transcription in different levels or even downregulate it depending on its binding site in a promoter (114). By using it we could screen for the effect of both activation or repression of a certain gene. An overview of the screening strategy can be seen in Figure 22 and it is also described in detail below.

For the identification of the genes that would be the targets of our screening, we performed a prediction using a genome-scale metabolic model (GEM). GEMs are based on mathematical representation of the cell metabolism. The reactions, the metabolites and the possible interactions of them are modelled based on all knowledge that is available (40). Those models make it possible to perform predictions for example the production rate of a certain metabolite in given conditions. This is done by flux balance

analysis (FBA), where a biological objective of choice can be applied (e.g. maximizing the flux towards a metabolite of interest) and the optimal solution can be found in steady state growth (181).

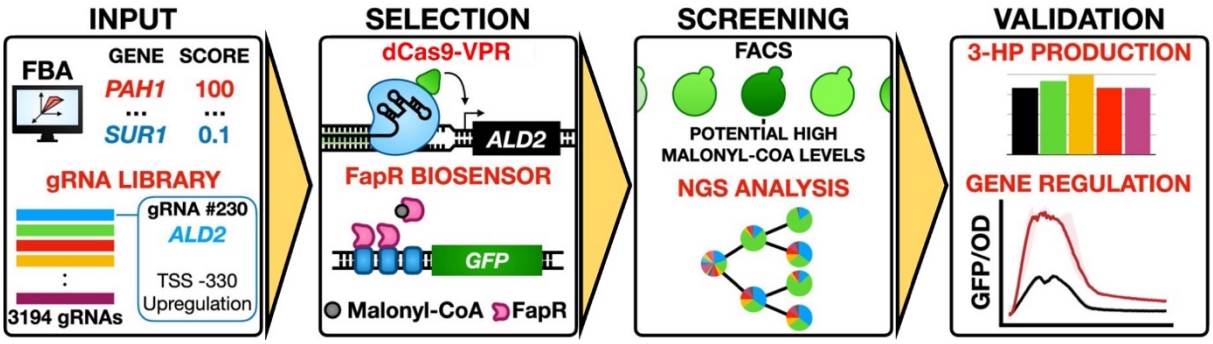


**Figure 22. Workflow for the gRNA library/dCas9-VPR screening for optimizing fluxes towards malonyl-CoA.** Candidate genes whose changes in expression levels could improve flux towards malonyl-CoA were identified and several gRNAs for each promoter were designed. A library of 3194 gRNAs was constructed in yeast cells that express dCas9-VPR and the malonyl-CoA biosensor. Screening of the library was performed with FACS, enriched gRNAs were verified and their effect on 3-HP production and gene regulation was identified.

An FBA was carried out, where 11 suboptimal growth rates (from 0.2*$\mu_{max}$ to 0.8* $\mu_{max}$) were applied in the yeast GEM and malonyl-CoA and acetyl-CoA fluxes were maximized (**paper I**: Figure 1a). For each solution, a k score was calculated for each reaction, which is the flux of each reaction divided by the flux of the same reaction when $\mu=\mu_{max}$. A score higher than 1 represented upregulation of the reaction, whereas a score lower than 1 represented downregulation. This analysis resulted in 70 upregulation targets, 80 downregulation targets and 18 genes that could be upregulation or downregulation targets depending on the carbon source.

Since the effect of a certain gRNA in gene regulation is not clearly predictable, we designed up to 21 gRNAs per target gene in order to cover multiple potential transcription levels of each gene. The final library consisted of 3194 gRNAs and it was cloned into a plasmid carrying dCas9-VPR (**paper I**: Figure S1). The yeast strain used expressed the FapR malonyl-CoA biosensor and a mutant version of ACC1 (ACC1**) which increases the flux from acetyl-CoA to malonyl-CoA (182). The library sorted with three rounds of FACS in total, using two different gates (Figure 23). 49 gRNAs were selected for further screening and characterization. Screening took place in yeast cells that expressed the enzyme Mcr which converts malonyl-CoA to 3-HP (35). We chose to screen for the production level of a malonyl-CoA product, in order to identify even temporary increases in the malonyl-CoA pool. Only eight gRNAs resulted in 3-HP yield increases higher than 15% compared with the strain carrying no gRNA (Figure 24).

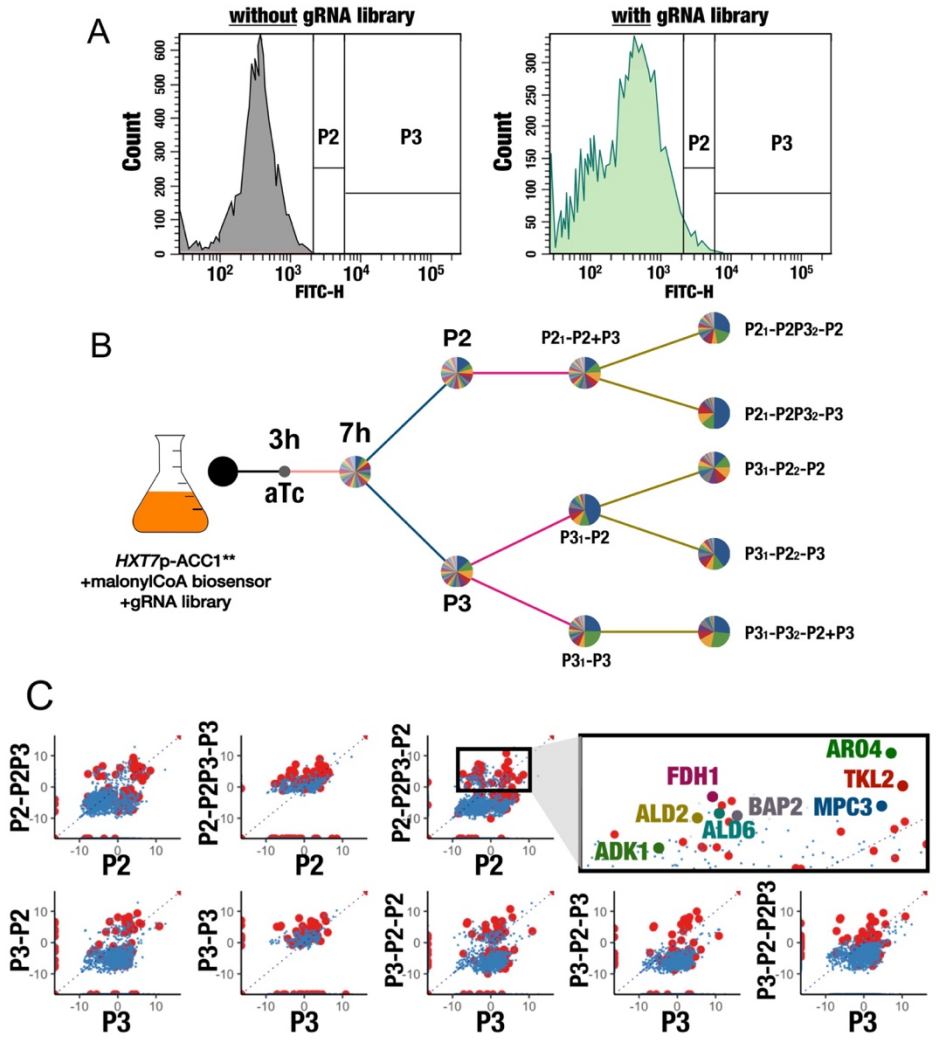


**Figure 23. Sorting strategy of the gRNA library. (A):** FACS gating strategy. After 7 hours of culture the cells were sorted in two gates, P2 and P3. The same gates were used for other two rounds of sorting. **(B):** Schematic representation of the sorting rounds. Each pie chart represents the gRNA distribution after sorting. **(C):** Plots of log2-fold change in the abundance of each gRNA over the initial library, red dots are gRNAs that were tested for 3-HP production.

The effect of each gRNAs on gene expression was tested on fusions of the yeast promoters that a particular gRNA is targeting with GFP. The gRNA/dCas9-VPR effect on transcription was reflected in the GFP levels without and with the gRNA (**paper I**: Figure S3). *ADK1* #15 (TSS -761) showed the highest 3-HP increase and it slightly downregulates the expression of the *ADK1*p-*GFP* fusion. *ADK1* encodes the adenylate kinase and influences ATP homeostasis. *ALD2* #1 (TSS -241) increases also significantly 3-HP and also *ALD2* expression under glucose. *ARO4* #1 (TSS -641) does not have an effect on the expression levels of *ARO4*p-*GFP* but it increases the expression levels of the neighboring gene *SPO23*. This was reflected on the effect of this gRNA on a *SPO23*p-*GFP* construct. *SPO23* encodes a protein of unknown function so far, making this an even more interesting and unexpected finding. Other potential gene targets identified by this screening are *PDX1*, *TKL2* and *AHP1*.

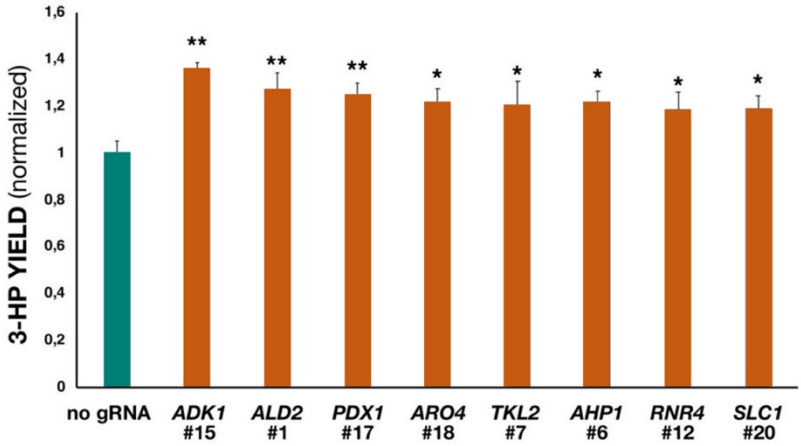


**Figure 24. Normalized 3-HP yields by the most promising gRNAs.** In the figure are shown the gRNAs that yielded more than 15% increased 3-HP levels than the strain with no gRNA expressed. The 3-HP are normalized by OD and 3-HP levels of the negative control (no gRNA). Biological triplicates were measured and the sampling took place after 72h of growth. * p-value < 0.05, ** p-value < 0.01 *** p-value < 0.001

## gRNA library and base editors for directed evolution of a malonate transporter

An alternative way to increase malonyl-CoA supply in yeast is by importing malonate extracellularly and converting it into the cell to malonyl-CoA. This has been achieved by heterologously expressing the dicarboxylic acid transporter and the bacterial malonate synthetase Mae1 (183). This pathway could increase malonyl-CoA supply in microbial consortia. Yeast could be cultivated with a malonate overproducing microorganism, import malonate, convert it to malonyl-CoA and produce the malonyl-CoA-derived products of choice. Microbial consortia can be more efficient since metabolic burden of long pathways is divided and they are also more robust to environmental changes (184, 185).

In **paper IV** we sought to improve the pathway of malonyl-CoA supply by extracellular malonate. First stage was to screen and identify the most efficient pair of dicarboxylic acid transporter and malonyl-CoA synthetase. To have a connection of the malonyl-CoA supply efficiency and the phenotype, we made our CEN.PK113-11C strains malonate-dependent by deleting *ACC1* (Figure 25). This made external malonate the sole source of cytosolic malonyl-CoA and the growth malonate dependent. After screening of 16 pairs of dicarboxylic acid transporter and malonyl-CoA synthetase (**paper IV**: Tables 1 and 2) we ended up proceeding with the dicarboxylic acid transporter SpMae1 from *Schizosaccharomyces pombe* and the malonyl-CoA synthetase RtMatB from *Rhizobium leguminosarium* bv. *trifolii*. This pair combined with *ACC1* deletion had the highest growth rate and the shortest lag phase (**paper IV**: Figures 1 and S1).

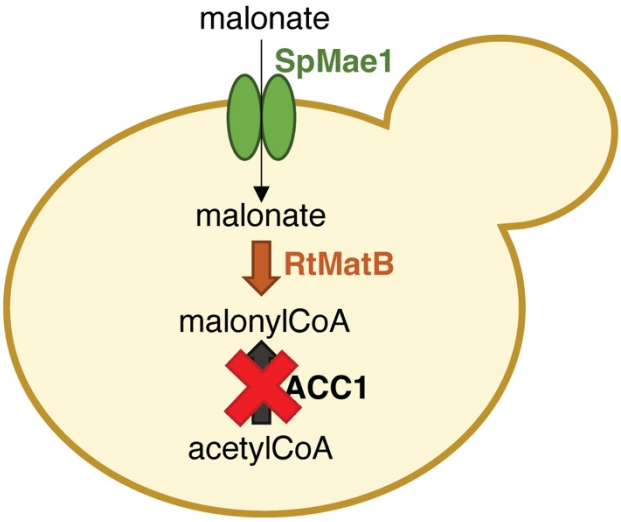


**Figure 25. Design of the malonate-dependent strain.** After the introduction of the dicarboxylic acid transporter (here SpMae1) and the malonyl-CoA synthetase (here RtMatB), the ACC1 gene was deleted. This is making external malonate the sole source of cytosolic malonyl-CoA and the growth of the strain dependent on the uptake of extracellular malonate.

Malonate is a dicarboxylic acid with pKa values of 2.8 and 5.7. Since SpMae1 is more efficient in the transport of monoanionic dicarboxylates (186), malonate transport is expected to more efficient at lower extracellular pH values where the monoprotonated form is prevalent. This was verified by our experimental data also in our *SpMae1-RtMatB-ΔACC1* strain (yFlav27) (**paper IV:** Figure 2), where growth in pH 4.5 is faster than in pH 6. In pH 6 the lag phase is also longer and the maximum OD is lower. In pH 6 there is no diauxic shift observed, and the same happens in pH 4.5 in low malonate concentrations (0-10 mM).

Since the malonate uptake pathway would be useful to be used in microbial consortia with malonate-producing bacteria, it would be valuable to improve the transport in higher pH values. It would be also beneficial to improve malonate uptake in low concentrations (10 mM and below). We chose to proceed in directed evolution of *SpMae1* in order to obtain variants that will perform better in higher pH and/or lower malonate concentrations. We chose to use the base editors dCas9-AID*Δ and TadA8eV106W-dCas9 for *in vivo* mutagenesis of *SpMae1*. The idea was to design a library of gRNAs that cover as much of the gene as possible, introduce it in yFlav27, induce mutagenesis on galactose and then enrich beneficial mutants on restrictive conditions (high pH or low malonate concentration).

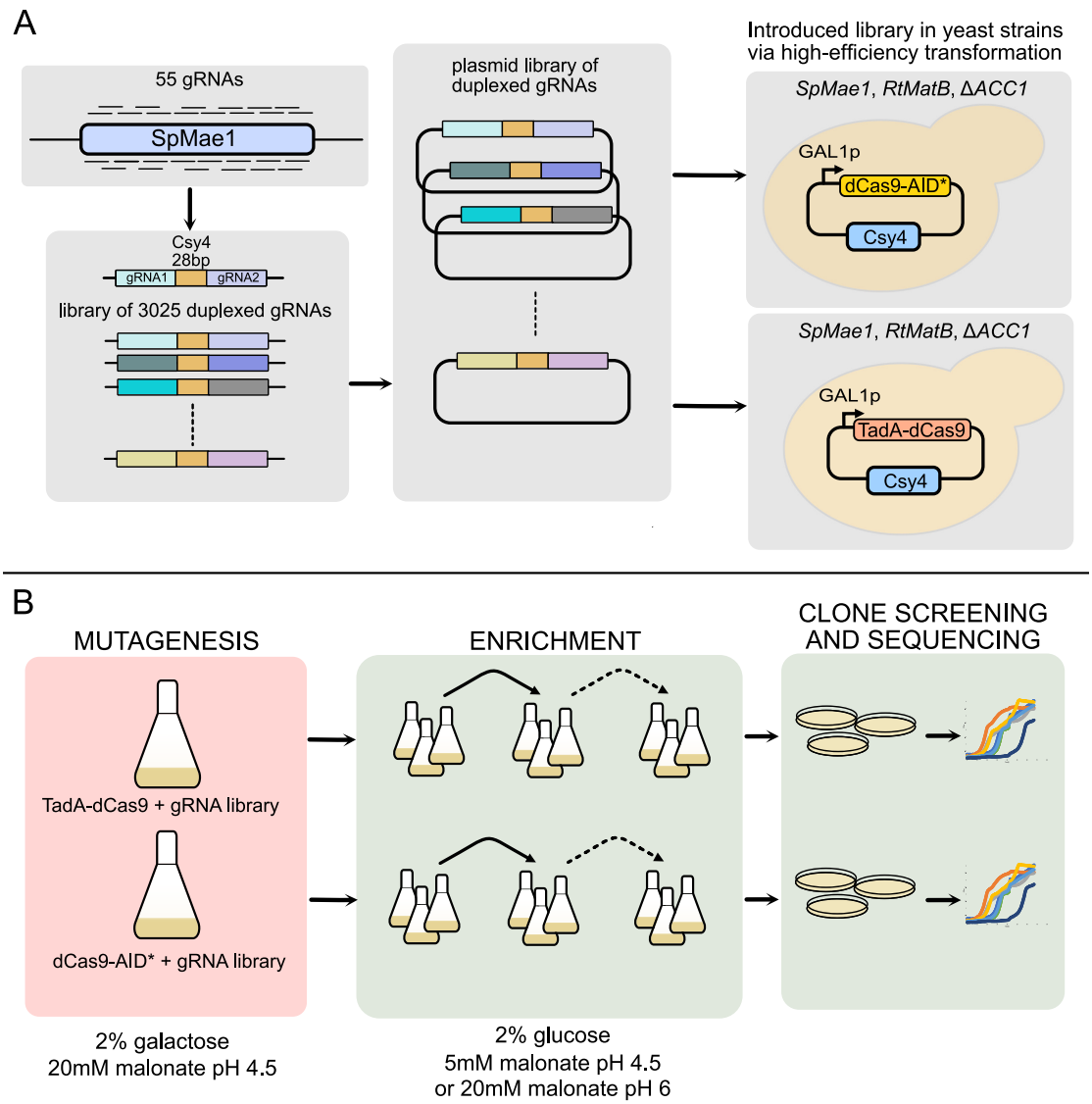


**Figure 26. The SpMae1 directed evolution assay. (A)** 55 gRNAs which cover *SpMae1* were Csy4-dublexed in a library of 3025 variants in total which was cloned into a 2µ vector. The plasmid library was then inserted into yFlav27 (*SpMae1-RtMatB-ΔACC1*) carrying a centromeric plasmid with either dCas9-AID* or TadA8eV106W-dCas9 under the control of *GAL1*p **(B)** The two libraries were cultivated in 2% galactose media with 20 mM malonate at pH 4.5 for mutagenesis. Next, cultures were transferred into 2% glucose media with low malonate concentration or high pH to enrich beneficial mutations. Individual clones were isolated on agar plates and screened. For spatial reasons TadA8eV106W is denoted as TadA and AID*Δ is denoted as AID*.

We designed 55 gRNAs that cover 67.5% of *SpMae1* (889 bp out of 1317 bp). Because of the poor performance of dCas9-AID*Δ with single gRNAs, we duplexed those gRNAs in any possible combination and the final library contained 3025 variants. The library was cloned in *E.coli* using the pMCL11 plasmid with the *ccdB* counterselection marker as dropout cassette (check also Chapter 2). Then the plasmid library was isolated by midiprep and was introduced in yFlav27 cells carrying a base editor/Csy4 coding centromeric plasmid(Figure 26A). Two yeast libraries were constructed, which differ in the base editor plasmid they carry (dCas9-AID*Δ or TadA8eV106W-dCas9).

Then, the mutagenesis/enrichment cycles followed. The media used in all the liquid cultures was Delft media (188) with variable carbon sources in 2% concentration. The optimal malonate/pH conditions were used if not stated otherwise (20 mM malonate/pH 6). Uracil and/or histidine were supplemented when needed in concentration 100 mg/L. The efficiency of our transformation was more than enough to cover at least 3*3025 variants (library size around $4*10^4$ cfu per transformation). After transformation, cells carrying gRNA library plasmids were enriched in raffinose media without added histidine and uracil. Then the mutagenesis was induced by inoculating library cells in 2% galactose in an initial OD of 0.1. When OD reached 3-4 after approximately 48h, the libraries were transferred in glucose media and restrictive conditions in order to enrich for *SpMae1* variants with beneficial mutations.

Each galactose grown cell population (set of gRNA library + base editor) was enriched in two conditions: **(i)** low malonate concentration (5 mM malonate/pH 4.5) and **(ii)** l high pH (20 mM malonate/pH 6). Four enrichment rounds were performed in the condition **(i)** and two enrichment rounds were performed for the condition **(ii)** (**paper IV**: Figure S2). Below are presented the enrichment assays in the two conditions.

Regarding the condition **(i),** after the enrichment rounds, the liquid cultures were plated on SD-Ura-His plates of pH 4.5 with 5 mM malonate and 30 clones from each library population (60 clones in total) were screened in a 96-well plate growth profiler. The 10 fastest growing clones underwent PCR amplification of their *SpMae1* ORF followed by Sanger sequencing. No mutations were found in the dCas9-AID* library clones. Among the TadA8eV106W-dCas9 library clones, only one was mutated. This clone had the mutations L76S and V274A.

In the condition **(ii),** enriched liquid cultures were plated on SD-Ura-His plates of pH 6 with 20 mM malonate. Fewer clones appeared in this experiment, and they were screened in shake flask. 6 clones from the dCas9-AID* library and 13 clones from the TadA8eV106-dCas9 library were screened (**paper IV:** Figure S3). All 6 clones from the dCas9-AID* library and the 11 faster growing clones from the TadA8eV106-dCas9 library had both *SpMae1* and the gRNA coding locus sequenced. No mutations were found among the dCas9-AID* library derived clones. Among the TadA8eV106W-dCas9 library derived clones, 6 out of 11 sequenced clones carried a M43V mutation and in total 9 clones carried a missense mutation (**paper IV:** Table S4).

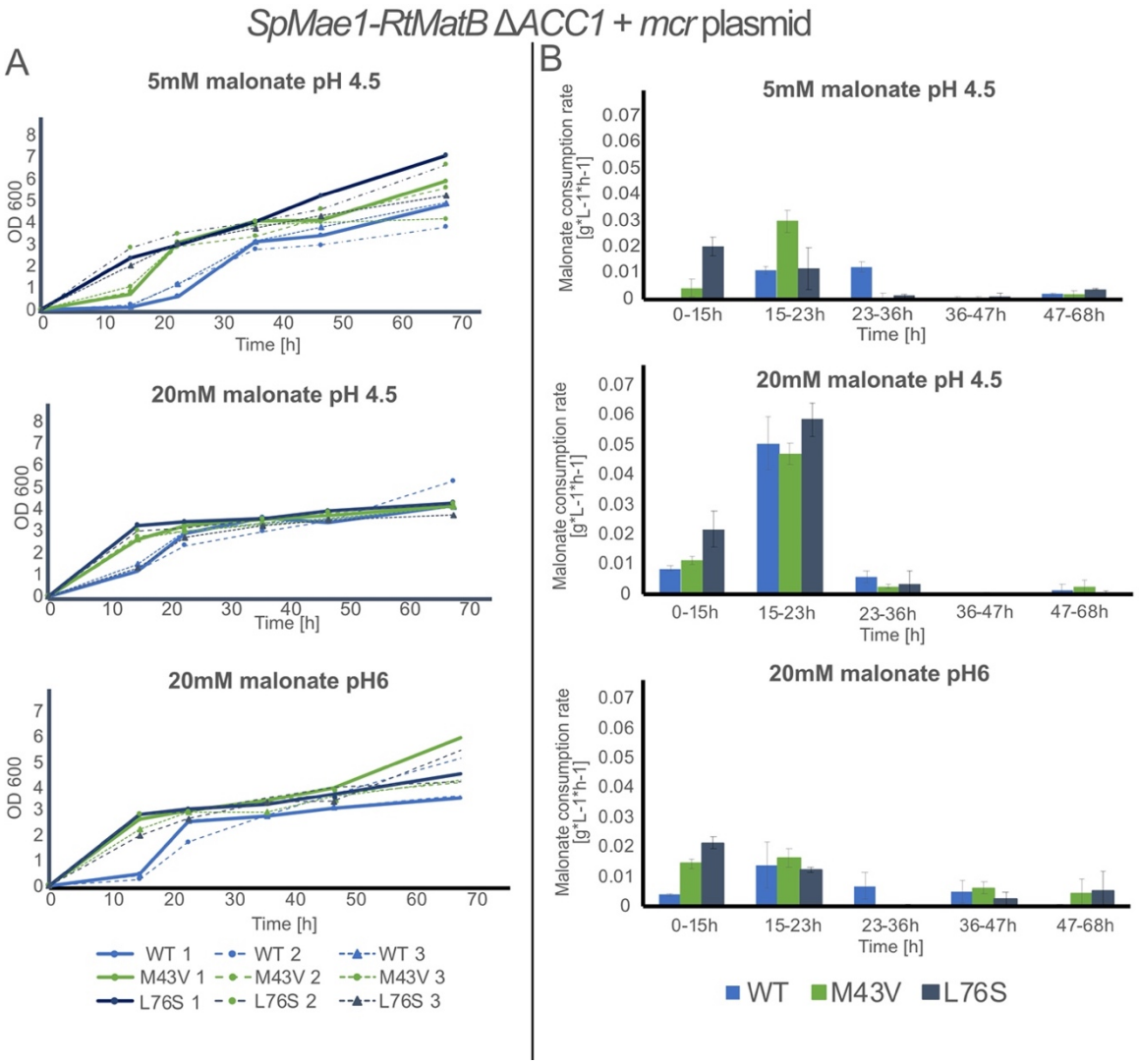


**Figure 27. Shake flask growth and malonate consumption of strains expressing different *SpMae1* variants.** Cells express either WT SpMae1, the M43V or the L76S variant. All strains also express *RtMatB* encoding malonyl-CoA synthetase and *mcr* encoding malonyl-CoA reductase which converts malonyl-CoA to 3-HP. *ACC1* is deleted in all strains. **(A):** $OD_{600}$ over time in shake flask growth. **(B):** Malonate consumption rate in different cultivation time intervals. Biological triplicates were measured and the error bars represent the standard deviation.

We chose to introduce the mutations M43V and L76S to the *SpMae1* gene of yFlav27 and check if they have a beneficial effect. In all the strains we introduced a plasmid that carries the gene *mcr* which encodes the enzyme malonyl-CoA reductase which converts malonyl-CoA to 3-HP (35). We introduced a malonyl-CoA consuming reaction in order to facilitate malonate uptake in the cells and be able to see more clearly the potential effects of the point mutations.

The shake flask growth experiments (Figure 27A) showed that the $SpMae1^{L76S}$-carrying strain shows improved growth compared with the strain carrying the WT gene in both low malonate concentration and high pH. The $SpMae1^{M43V}$-carrying strain shows improved growth mostly in high pH and in a lower extent in low malonate concentrations.

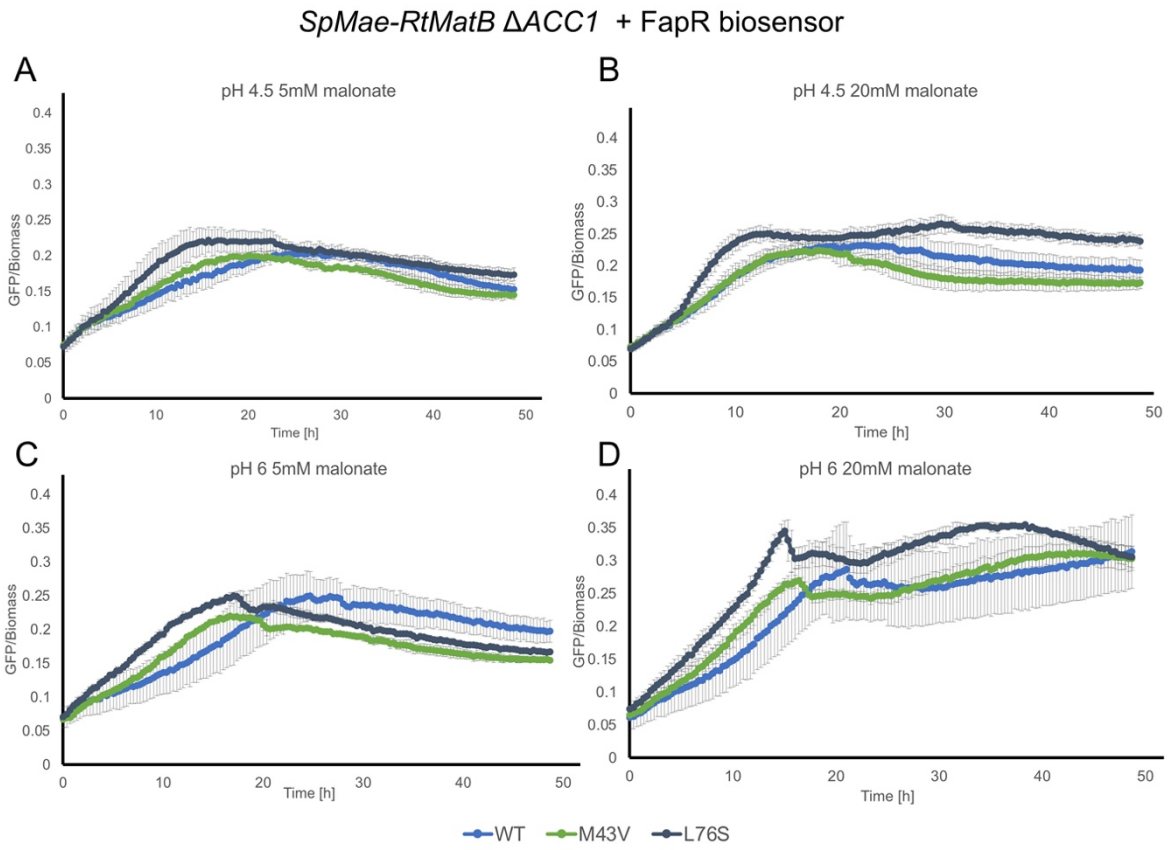


**Figure 28. GFP to biomass ratio over time for the three SpMae1-RtMatB-Δ*ACC1* + FapR biosensor strains with different SpMae1 variants.** The experiment was performed in a Biolector. The y-axis represents arbitrary units. Four conditions were tested: (A) pH 4.5 and 5 mM malonate, (B) pH 4.5 and 20 mM malonate, (C) pH 6 and 5 mM malonate and (D) pH 6 and 20 mM malonate. Measurements were performed in biological triplicates and the error bars represent the standard deviation.

In the shake flask experiments, we also measured the extracellular malonate concentration in each sampling point. Based on those data, we calculated the malonate consumption for the time intervals between the samplings (Figure 27B). The most interesting finding is that both mutant strains have a higher consumption rate of malonate than the WT in the first 15 h of culture. This trend is the same for all conditions tested and the $SpMae1^{L76S}$ mutant outperforms $SpMae1^{M43V}$. The $SpMae1^{L76S}$ mutant performed better than both $SpMae1^{M43V}$ and $SpMae1^{WT}$ also in 3-HP production in both high and low malonate concentrations (**paper IV:** Figure S4).

We performed another set of experiments to investigate the dynamics of malonyl-CoA supply in the mutant strains. Instead of introducing the the *mcr* plasmid in the strains, we introduced a FapR malonyl-CoA biosensor with expanded dynamic range as it was improved by Dabirian *et al*. (189). Then we monitored biomass and green fluorescence in real time. The GFP/Biomass ratio in this experimental set can work as an indicator of cytosolic malonyl-CoA levels.
The experiment was performed in four conditions: **(a)** pH 4.5 and 5 mM malonate, **(b)** pH 4.5 and 20 mM malonate, **(c)** pH 6 and 5 mM malonate and **(d)** pH 6 and 20 mM

malonate. The strain carrying the *SpMae1*$^{L76S}$ mutant showed higher GFP/Biomass ratio than the other two strains, in all the conditions tested. This trend is more obvious within the first 20 hours of growth. This is an indication of increased cytosolic levels of malonyl-CoA. Since the only difference between this strain is the *SpMae1* variant, we can conclude that the improved GFP/Biomass of the *SpMae1*$^{L76S}$ mutant is a result of improved malonate uptake. In combination with the improved malonate uptake - especially within the first 15h of growth- we can conclude that especially the L76S mutation improves the malonate uptake performance of SpMae1 transporter.

## Conclusions

In this chapter we use gRNA libraries in a variety of screening projects. In the first half of the chapter (**paper I**) we perform a gRNA library screening in combination with the dCas9-VPR transcriptional activator and FBA. This experiment reveals novel regulatory setups that increase the metabolic fluxes towards a metabolite of interest, in our case malonyl-CoA. The different effects that dCas9-VPR can have on transcription depending on the promoter binding site of the gRNA gives the opportunity to screen even for slight changes in the transcription level of a gene. Moreover, this assay revealed even unknown genes whose changes in their expression levels has an effect on the metabolic fluxes, such as *SPO23*.

In the second half of the chapter (**paper IV**) we used a gRNA library for directed evolution. In this way we explored the potential of two broad range base editors that we constructed previously. We sought to apply this approach in a project of improving the cytosolic malonyl-CoA supply in yeast by external malonate transport, and we selected to evolve the malonate transporter *SpMae1* of this pathway. TadA8eV106W-dCas9 base editor evolution resulted in two interesting mutants of *SpMae1* that seem to improve its malonate transport performance. However, dCas9-AID*Δ evolution did not result in any mutant, something that might have to do with its weak performance with single gRNAs that was observed earlier. Even though we constructed a duplexed gRNA library, this might not be enough to ensure high mutagenesis efficiencies for this base editor that could make it beneficial for directed evolution approaches. When it comes to the two *SpMae1* mutations that occurred from this assay, they can be further analyzed by protein modeling or rational engineering. The performance of a double mutant *SpMae1*$^{M43V,L76S}$ would be also interesting to be examined.

# Chapter 5 – Conclusions and future prospects

## The molecular cloning challenges of CRISPR

CRISPR has now a much broader field of applications than when it was first discovered. Initially it was a tool for introducing DSBs in genomic loci of choice, but now the non-cutting dCas9 or single stranding nCas9 can be fused with a wide variety of domains, enabling targeted transcription regulation of genes, base editing and more widely the targeted recruitment of any effector of choice in a genomic locus of choice. However, there was a lack of systematization regarding the cloning setup for CRISPR applications in yeast. Additionally, it has been observed that in many yeast CRISPR applications the Cas9-based protein and the gRNA were cloned the same plasmid (114, 121). This could not allow independent engineering of the Cas9-based gene and the gRNA. As a result, significant delays can occur when a wide variety of CRISPR effectors needs to be screened and/or many gRNAs need to be tested.

In **paper II** we attempted an extension of the already existing yeast MoClo system (169)for CRISPR applications. Now Cas9 and gRNA are expressed from two separate vectors and their cloning architecture is compatible with a well characterized system which enables the easy exchange of domains with Golden Gate cloning reactions. This makes it easier for a researcher to build up CRISPR effectors in a more fast and reliable way, avoiding costly and time-consuming sequencing verifications and primer orderings due to consecutive PCRs. Moreover, the multigRNA tool we built, allows for the quick and simple generation of multiple gRNA arrays. It has the capacity to multiplex up to 12 gRNAs and we have successfully verified *in vitro* building of arrays with up to 9 gRNAs with more than 75% success rate.

However, we did not succeed to develop a platform for creating multiplexed gRNA libraries. We aimed to establish a platform that could overcome the need to order pre-multiplexed gRNA libraries and could allow the multiplexing to happen during the cloning. We tried some proof of concept approaches to randomly multiplex gRNAs in a 5x array, for example by adding unique 4 bp overhangs similar to a previous approach for gRNA array construction (138) but it was not successful. Most of the clones had one or two gRNAs cloned and not five. The tandem repeats of Csy4 recognizing loop and gRNA scaffold might be the source of the problem and it was present even when we used commercially available *E.coli* competent cells. A good improvement of this toolkit, and a challenge for a future researcher on the field, could be to find a way to overcome those difficulties and establish a platform for randomized gRNA multiplexing. This could expand even more the possibilities of gRNA library screenings, as it will broaden the possibilities for combinatorial screenings.

Another limitation in our study was the number of possible simultaneous gene deletions using CRISPR. In our study we managed to perform successfully up to three gene deletions with one transformation. In a previous study of our lab it was possible to delete up to four genes simultaneously using Csy4 gRNA arrays (128) but we didn't try to delete four genes in the present study. Since multiplexed gRNA arrays have been proved to work efficiently in combinatorial CRISPRi and CRISPRa arrays (138, 139), a possible reason for this is the increased number of simultaneous homologous gene repairs that have to happen in order to have successful gene deletions. Moreover, the higher the number of deletions, the more repair fragments have to be introduced in the cell. However, we observed a trend that in arrays of more than 5 gRNAs, the first three and the last gRNA are the most functional, something that might be useful for the future research on the field.

## CRISPR broad range base editors

In **paper III** our goal was to develop CRISPR base editors with broad editing windows that can be used for directed evolution approaches. Our intention was to move further from the high-precision base editors that were developed in yeast (121, 190) to broad range base editors similar to the CRISPR-X tool (123)

Our dCas9-AID*Δ base editor seemed to lead to low mutagenesis rates with single gRNAs according to our NGS data, but some interesting combinatorial effects were observed, especially when two gRNAs with distance 19 bp with each other were expressed simultaneously. In the multiple gRNA samples we observed almost exclusively G and C mutations to all the possible nucleotides, with a preference of C→G, C→T, G→A and G→C mutations. The boost-up effect observed with multiplexing, indicates a combinatorial effect between two copies of dCas9-AID*Δ that are bound very closely to each other on a genomic locus. Previous studies have pointed out the capacity of AID (with its C-terminus deleted or not) to form dimers or even multimers (179, 191). On the other hand, *in vitro* atomic force microscopy (AFM) experiments have shown that AID along with single stranded DNA is predominantly present as a monomer and it shows deamination activity (192). It has been also reported that AID activity and transcription are strongly connected, since AID is acting only on transcriptionally active DNA regions which have open transcription bubbles with single stranded DNA exposed (193, 194).

A similar effect of multiplexing on AID activity has been also shown in the study of Ma *et al.* (122) where a similar base editor (dCas9-AIDx) was characterized. AIDx does not contain the point mutations that make AID*Δ hyperactive but the effect of gRNA multiplexing is similar. The CRISPR-X tool (123) combines dCas9 and AID*Δ but instead of fusing the cytidine deaminase domain with dCas9, the authors used a RNA hairpin mediated recruitment. They fused AID*Δ an MS2 domain and they added to the gRNA two MS2-recognising RNA hairpins. With this setting they achieved mutagenesis in a 100

bp region around a targeted genomic locus (+/- 50 bp from the PAM site) using only one gRNA. This setting maybe allows dimerization of AID*Δ which is synthesized freely and independently from dCas9. Dimerization of the free AID*Δ copies can also take place during the recruitment via the hairpins of the Cas9/gRNA complex which is bound on the targeted locus. Regarding our base editor dCas9-AID*Δ, multiplexing of two proximal gRNAs in a locus seems to increase the total number of mutations per read. Moreover, multiplexing appeared to increase the base editing efficiency even when two more distant genomic loci are targeted. The effect in this case is not as pronounced but it is existent when comparing the results with datasets from single gRNA experiments. It would be worth to test a base editor with a design similar to CRISPR-X in yeast and test if it shows better mutagenesis efficiency even with single gRNAs.

We also performed a brief characterization of two base editors which combined dCas9 and high activity adenine deaminases (TadA8e and TadA8eV106W). Since we did not perform NGS analysis and we analyzed only $Can^R$ clones, they were not characterized in the same depth, but they resulted in mutations only whithin the gRNA binding site and they did not show the same combinatorial effect as dCas9-AID*Δ. Those base editors are also more biased as we only observed A→T mutations.

Other constructs that could be tried out are dCas9 fusions with AID*Δ dimers, maybe connected with flexible linkers of different length. In order to overcome the bias that those deaminase-based base editors have, combinatorial base editors with both TadA8e and AID*Δ could be tested, especially if AID*Δ dimerization proves to solve the efficiency problem. Similar combinatorial base editors have been developed for human cells (195), however they show narrow editing windows (roughly an area from -10 to -20 relative to the PAM site). We attempted to construct the base editors TadA8eV106W-dCas9-AID*Δ and dCas9-AID*Δ-AIΔ*Δ but the cloning attempts in *E.coli* were unsuccessful. Cloning problems occurred also during the construction of dCas9-AID*Δ, mostly in the form of point mutations occurring within the dCas9 domain. Those cloning problems happened only in constructs combining dCas9 and AID*Δ and this could indicate some possible toxic effects in *E.coli*, which potentially occur from leaky expression of the dCas9-AID*Δ fusion. This might be a useful observation for future researchers that wish to work with those constructs.

## The possibilities of gRNA libraries

In the other two papers of the thesis, we explored different experimental setups that involve gRNA library screenings for the improvement of malonyl-CoA supply in yeast. In **paper I** we used a single gRNA library combined with the transcriptional activator dCas9-VPR to uncover transcriptional setups that improve the fluxes towards the key metabolite malonyl-CoA. In **paper IV** we used a duplexed gRNA library along with dCas9 base editors in order to evolve a gene that encodes a malonate transporter.

### Transcriptional regulation screenings

The first project (**paper I**) combines along with gRNA libraries and dCas9-VPR, FBA predictions using a genome-scale metabolic model (GEM). This is an example of how synergies between metabolic modeling and high throughput screening can help in revealing insights of yeast cell metabolism that can be valuable for strain engineering. The availability of metabolite biosensors and high-throughput screening methods like FACS opens new opportunities for screening approaches. They can also give surprising results, like the hit of the unknown function gene *SPO23* we got in our screening.

Those screenings need very well curated GEMs and well-developed biosensors for the compound of interest. Moreover, CRISPRa approaches still lack predictability. The relation between the gRNA binding site and the effect dCas9-VPR remains unclear in many cases. For some promoters that encode genes participating in central carbon metabolism rational gRNA design assays were performed, based on transcription factor binding data (196). Those experiments revealed a trend of dCas9-VPR upregulating a gene when bound next to a transcription factor (TF) binding motif, whereas it was downregulating or having no effect when bound on the TF motif. More knowledge on the TF binding motifs on yeast promoters would facilitate the gRNA design and allow for a more focused and predictable gRNA design which can facilitate CRISPRa screenings.

## Directed evolution

In the second project (**paper IV**) we used a gRNA library and the base editors dCas9-AID*Δ and TadA8eV106W-dCas9 in order to evolve thr dicarboxylic acid transporter SpMae1 for improved malonate transport. We managed to obtain two interesting variants with the adenine base editor. The beneficial effect of those point mutations could be verified by reverse engineering experiments. However, we didn't manage to get any mutants with dCas9-AID*Δ, something that might be related with its poor performance with single gRNAs. Probably the gRNA duplexing design of the library did not result in increased mutagenesis efficiency of dCas9-AID*Δ and improvements on the design dCas9-AID*Δ should be considered.

Looking at the recent literature, we can see that even base editors with narrow range window were used in directed evolution experiments. For example, Liu *et al.* (197) managed to improve tolerance of *S.cerevisiae* in various stress conditions by rationally targeting the transcription factor encoding gene *SPT15*. They predicted computationally the sites of the transcription factor that could play a role in the functionality of the transcription factor and they targeted them with individual gRNAs and the base editor nCas9-PmCDA1 that shows a narrow editing window (121). Another interesting approach, more similar to what we followed, has been carried out in rice. Kuang *et al.* (198) evolved the OsALS1 protein to be tolerant the herbicide bispyribac-sodium by using gRNA libraries and the AID*Δ-based and TadA*-based base editors. According to those examples, those gRNA-mediated directed evolution assays can be improved either by the help of computational predictions or by applying a "die or live" selection method, like a resistance to a totally killing herbicide. Our selection on either low malonate concentration or increased pH might have put pressure on the enrichment cycles but it was leaky in some extent, since those conditions enabled some growth of the strain carrying the WT *SpMae1* gene.

Transcription or replication based evolution methods like ICE (154), OrthoRep (155, 156) and TRACE (153, 158) could be an alternative when potential mutagenesis hotspots cannot be determined beforehand. They can also be an alternative to error-prone PCR since they allow continuous evolution rounds and there is no need for multiple error-prone PCR and transformation cycles. They can also be a valuable method for identifying mutational hotspots in a given selection condition which can be further explored by approaches combining CRISPR base editors and gRNA libraries approaches. Another yeast CRISPR-based evolution approach that can give an alternative to cytidine or adenine base editors is yEvolvR (160) which can mutate a region up to 40 bp upstream of a gRNA and with a less biased way since it mutates all the four DNA bases in more balanced distribution.

# References


1. Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsinelli M. 2003. Evidence for *S. cerevisiae* Fermentation in Ancient Wine. Journal of Molecular Evolution 57:S226–S232.

2. Money NP. 2018. The Rise of Yeast: How the sugar fungus shaped civilisation. Oxford University Press.

3. Pasteur L. 1876. Études sur la bière, ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable: avec une théorie nouvelle de la fermentation. Gauthier-Villars.

4. Wang Z, Zhuge J, Fang H, Prior BA. 2001. Glycerol production by microbial fermentation: A review. Biotechnology Advances 19:201–223.

5. Barnett JA. 2007. A history of research on yeasts 10: foundations of yeast genetics. Yeast 24:799–845.

6. Winge O, Roberts C. 1950. The polymeric genes for maltose fermentation in yeast and their mutability. Comptes rendus des Travaux du Laboratoire Carlsberg, Ser Physiol 25:35–83.

7. Winge Ö. 1952. The genetic situation concerning fermentation in yeasts. 2. Heredity 6:263–269.

8. Pauling L, Corey RB, Branson HR. 1951. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. Proceedings of the National Academy of Sciences 37:205–211.

9. Watson JD, Crick FHC. 1953. The Structure of Dna. Cold Spring Harbor Symposia on Quantitative Biology 18:123–131.

10. Crick F. 1970. Central Dogma of Molecular Biology. 5258. Nature 227:561–563.

11. Cohen SN, Chang ACY, Boyer HW, Helling RB. 1973. Construction of Biologically Functional Bacterial Plasmids In Vitro. Proceedings of the National Academy of Sciences 70:3240–3244.

12. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, Boyer HW. 1977. Expression in *Escherichia col*i of a chemically synthesized gene for the hormone somatostatin. Science 198:1056–1063.

13. Tuite MF, Dobson MJ, Roberts NA, King RM, Burke DC, Kingsman SM, Kingsman AJ. 1982. Regulated high efficiency expression of human interferon-alpha in *Saccharomyces cerevisiae*. The EMBO Journal 1:603–608.

14. Thim L, Hansen MT, Sørensen AR. 1987. Secretion of human insulin by a transformed yeast cell. FEBS Letters 212:307–312.

15. Bailey JE. 1991. Toward a Science of Metabolic Engineering. Science 252:1668–1675.

16. Nielsen J. 2001. Metabolic engineering. Applied Microbiology and Biotechnology 55:263–283.

17. Stephanopoulos G. 1999. Metabolic Fluxes and Metabolic Engineering. Metabolic Engineering 1:1–11.

18. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 Genes. Science 274:546–567.

19. Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, Dunn B, Dwight S, Riles L, Mortimer RK, Botstein D. 1997. Genetic and physical maps of *Saccharomyces cerevisiae*. Nature 387:67–73.

20. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proceedings of the National Academy of Sciences 94:13057–13062.

21. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. 2001. Global Analysis of Protein Activities Using Proteome Chips. Science 293:2101–2105.

22. Villas-Boas SG, Moxley JF, Åkesson M, Stephanopoulos G, Nielsen J. 2005. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. Biochemical Journal 388:669–677.

23. Sauer U. 2006. Metabolic networks in motion: 13C-based flux analysis. Molecular Systems Biology 2:62.

24. Mustacchi R, Hohmann S, Nielsen J. 2006. Yeast systems biology to unravel the network of life. Yeast 23:227–238.

25. Nielsen J, Jewett MC. 2008. Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. FEMS Yeast Research 8:122–131.

26. Merryman C, Gibson DG. 2012. Methods and applications for assembling large DNA constructs. Metabolic Engineering 14:196–204.

27. Nielsen J, Larsson C, van Maris A, Pronk J. 2013. Metabolic engineering of yeast for production of fuels and chemicals. Current Opinion in Biotechnology 24:398–404.

28. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R, Keasling JD. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. 7086. Nature 440:940–943.

29. Buijs NA, Siewers V, Nielsen J. 2013. Advanced biofuel production by the yeast Saccharomyces cerevisiae. Current Opinion in Chemical Biology 17:480–488.

30. Zhou YJ, Buijs NA, Siewers V, Nielsen J. 2014. Fatty Acid-Derived Biofuels and Chemicals Production in *Saccharomyces cerevisiae*. Frontiers in Bioengineering and Biotechnology 2:32.

31. Ferreira R, Teixeira PG, Siewers V, Nielsen J. 2018. Redirection of lipid flux toward phospholipids in yeast increases fatty acid turnover and secretion. PNAS 115:1262–1267.

32. Steen EJ, Chan R, Prasad N, Myers S, Petzold CJ, Redding A, Ouellet M, Keasling JD. 2008. Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. Microbial Cell Factories 7:36.

33. Kampranis SC, Makris AM. 2012. Developing a yeast cell factory for the production of terpenoids. Computational and Structural Biotechnology Journal 3:e201210006.

34. Zhang Y, Nielsen J, Liu Z. 2017. Engineering yeast metabolism for production of terpenoids for use as perfume ingredients, pharmaceuticals and biofuels. FEMS Yeast Research 17:fox080.

35. Ji R-Y, Ding Y, Shi T-Q, Lin L, Huang H, Gao Z, Ji X-J. 2018. Metabolic Engineering of Yeast for the Production of 3-Hydroxypropionic Acid. Frontiers in Microbiology 9:2185.

36. Zhang J, Hansen LG, Gudich O, Viehrig K, Lassen LMM, Schrübbers L, Adhikari KB, Rubaszka P, Carrasquer-Alvarez E, Chen L, D'Ambrosio V, Lehka B, Haidar AK, Nallapareddy S, Giannakou K, Laloux M, Arsovska D, Jørgensen MAK, Chan LJG, Kristensen M, Christensen HB, Sudarsan S, Stander EA, Baidoo E, Petzold CJ, Wulff T, O'Connor SE, Courdavault V, Jensen MK, Keasling JD. 2022. A microbial supply chain for production of the anti-cancer drug vinblastine. Nature 609:1–7.

37. Carbonell P, Currin A, Jervis AJ, Rattray NJW, Swainston N, Yan C, Takano E, Breitling R. 2016. Bioinformatics for the synthetic biology of natural products: integrating across the Design–Build–Test cycle. Natural Product Reports 33:925–932.

38. Campbell K, Xia J, Nielsen J. 2017. The Impact of Systems Biology on Bioprocessing. Trends in Biotechnology 35:1156–1168.

39. Petzold C, Chan LJ, Nhan M, Adams P. 2015. Analytics for metabolic engineering. Frontiers in Bioengineering and Biotechnology 3:135.

40. Oberhardt MA, Palsson BØ, Papin JA. 2009. Applications of genome-scale metabolic reconstructions. Molecular Systems Biology 5:320.

41. Sánchez BJ, Nielsen J. 2015. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. Integrative Biology 7:846–858.

42. Zhang C, Hua Q. 2016. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. Frontiers in Physiology 6:413.

43. Rogers JK, Taylor ND, Church GM. 2016. Biosensor-based engineering of biosynthetic pathways. Current Opinion in Biotechnology 42:84–91.

44. Mahr R, Frunzke J. 2016. Transcription factor-based biosensors in biotechnology: current state and future prospects. Applied Microbiology and Biotechnology 100:79–90.

45. David F, Nielsen J, Siewers V. 2016. Flux Control at the Malonyl-CoA Node through Hierarchical Dynamic Pathway Regulation in *Saccharomyces cerevisiae*. ACS Synthetic Biology 5:224–233.

46. Li S, Si T, Wang M, Zhao H. 2015. Development of a Synthetic Malonyl-CoA Sensor in *Saccharomyces cerevisiae* for Intracellular Metabolite Monitoring and Genetic Screening. ACS Synthetic Biology 4:1308–1315.

47. Xu P, Li L, Zhang F, Stephanopoulos G, Koffas M. 2014. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. Proceedings of the National Academy of Sciences 111:11299–11304.

48. Teo WS, Hee KS, Chang MW. 2013. Bacterial FadR and synthetic promoters function as modular fatty acid sensor- regulators in *Saccharomyces cerevisiae*. Engineering in Life Sciences 13:456–463.

49. Zhang F, Carothers JM, Keasling JD. 2012. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. 4. Nature Biotechnology 30:354–359.

50. Skjoedt ML, Snoek T, Kildegaard KR, Arsovska D, Eichenberger M, Goedecke TJ, Rajkumar AS, Zhang J, Kristensen M, Lehka BJ, Siedler S, Borodina I, Jensen MK, Keasling JD. 2016. Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. 11. Nature Chemical Biology 12:951–958.

51. Dabirian Y, Gonçalves Teixeira P, Nielsen J, Siewers V, David F. 2019. FadR-Based Biosensor-Assisted Screening for Genes Enhancing Fatty Acyl-CoA Pools in *Saccharomyces cerevisiae*. ACS Synthetic Biology 8:1788–1800.

52. Carlson R. 2009. The changing economics of DNA synthesis. 12. Nature Biotechnology 27:1091–1094.

53. Kosuri S, Church GM. 2014. Large-scale de novo DNA synthesis: technologies and applications. 5. Nature Methods 11:499–507.

54. Hughes RA, Ellington AD. 2017. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. Cold Spring Harbor Perspectives in Biology 9:a023812.

55. Dietrich JA, McKee AE, Keasling JD. 2010. High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. Annual Review of Biochemistry 79:563–590.

56. Lin J-L, Wagner JM, Alper HS. 2017. Enabling tools for high-throughput detection of metabolites: Metabolic engineering and directed evolution applications. Biotechnology Advances 35:950–970.

57. Skrekas C, Ferreira R, David F. 2022. Fluorescence-Activated Cell Sorting as a Tool for Recombinant Strain Screening, p. 39–57. In Mapelli, V, Bettiga, M (eds.), Yeast Metabolic Engineering: Methods and Protocols. Springer US, New York, NY.

58. Dragosits M, Mattanovich D. 2013. Adaptive laboratory evolution – principles and applications for biotechnology. Microbial Cell Factories 12:64.

59. Mattanovich D, Borth N. 2006. Applications of cell sorting in biotechnology. Microbial Cell Factories 5:12.

60. Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret J-C, Marquez M, Klibanov AM, Griffiths AD, Weitz DA. 2010. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. Proceedings of the National Academy of Sciences 107:4004–4009.

61. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. Journal of Bacteriology 169:5429–5433.

62. Mojica FJM, Juez G, Rodriguez-Valera F. 1993. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Molecular Microbiology 9:613–621.

63. Ishino Y, Krupovic M, Forterre P. 2018. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. Journal of Bacteriology 200:e00580-17.

64. Mojica FJM, Diez-Villasenor C, Soria E, Juez G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Molecular Microbiology 36:244–246.

65. Jansen Ruud, Embden JanDA van, Gaastra Wim, Schouls LeoM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology 43:1565–1575.

66. Mojica FJM, Rodriguez-Valera F. 2016. The discovery of CRISPR in archaea and bacteria. The FEBS Journal 283:3162–3169.

67. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV. 2015. An updated evolutionary classification of CRISPR–Cas systems. 11. Nature Reviews Microbiology 13:722–736.

68. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas Č, White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV. 2020. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. 2. Nature Reviews Microbiology 18:67–83.

69. Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096.

70. Hsu PD, Lander ES, Zhang F. 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157:1262–1278.

71. Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V. 2013. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus.* RNA Biology 10:841–851.

72. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337:816–821.

73. Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences 109:E2579–E2586.

74. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155:733–740.

75. Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJJ. 2017. CRISPR-Cas: Adapting to change. Science 356:eaal5056.

76. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339:819–823.

77. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-Guided Human Genome Engineering via Cas9. Science 339:823–826.

78. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell 156:935–949.

79. Jiang F, Doudna J. 2017. CRISPR-Cas9 Structures and Mechanisms. Annual review of biophysics 46:483-503.

80. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, Liu DR. 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. 7699. Nature 556:57–63.

81. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F. 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell 163:759–771.

82. Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J, Zhang F. 2017. RNA editing with CRISPR-Cas13. Science 358:1019–1027.

83. Smargon AA, Cox DBT, Pyzocha NK, Zheng K, Slaymaker IM, Gootenberg JS, Abudayyeh OA, Essletzbichler P, Shmakov S, Makarova KS, Koonin EV, Zhang F. 2017. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Molecular Cell 65:618-630.e7.

84. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, Severinov K, Regev A, Lander ES, Koonin EV, Zhang F. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353:aaf5573.

85. DiCarlo JE, Conley AJ, Penttilä M, Jäntti J, Wang HH, Church GM. 2013. Yeast Oligo-Mediated Genome Engineering (YOGE). ACS Synthetic Biology 2:741–749.

86. Jakočiūnas T, Jensen MK, Keasling JD. 2016. CRISPR/Cas9 advances engineering of microbial cell factories. Metabolic Engineering 34:44–59.

87. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Research 41:4336–4343.

88. Jakočiūnas T, Bonde I, Herrgård M, Harrison SJ, Kristensen M, Pedersen LE, Jensen MK, Keasling JD. 2015. Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. Metabolic Engineering 28:213–222.

89. Guo X, Chavez A, Tung A, Chan Y, Kaas C, Yin Y, Cecchi R, Garnier SL, Kelsic ED, Schubert M, DiCarlo JE, Collins JJ, Church GM. 2018. High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast. 6. Nature Biotechnology 36:540–546.

90. Jessop-Fabre MM, Jakočiūnas T, Stovicek V, Dai Z, Jensen MK, Keasling JD, Borodina I. 2016. EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. Biotechnology Journal 11:1110–1117.

91. Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, Forster J, Nielsen J, Borodina I. 2014. EasyClone: method for iterative chromosomal integration of multiple genes *Saccharomyces cerevisiae*. FEMS Yeast Research 14:238–248.

92. Stovicek V, Borja GM, Forster J, Borodina I. 2015. EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains. Journal of Industrial Microbiology and Biotechnology 42:1519–1531.

93. Pronk JT. 2002. Auxotrophic Yeast Strains in Fundamental and Applied Research. Applied and Environmental Microbiology 68:2095–2100.

94. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. 9. Nature Biotechnology 31:822–826.

95. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. 9. Nature Biotechnology 31:839–843.

96. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, Aryee MJ, Joung JK. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. 2. Nature Biotechnology 33:187–197.

97. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim J-S. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Research 24:132–141.

98. Satomura A, Nishioka R, Mori H, Sato K, Kuroda K, Ueda M. 2017. Precise genome-wide base editing by the CRISPR Nickase system in yeast. 1. Scientific Reports 7:2095.

99. Standage-Beier K, Zhang Q, Wang X. 2015. Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. ACS Synthetic Biology 4:1217–1225.

100. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. 3. Nature Biotechnology 32:279–284.

101. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. 7587. Nature 529:490–495.

102. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. 2016. Rationally engineered Cas9 nucleases with improved specificity. Science 351:84–88.

103. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S, Jaenisch R, Zhang F, Sharp PA. 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. 7. Nature Biotechnology 32:670–676.

104. Verkuijl SA, Rots MG. 2019. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. Current Opinion in Biotechnology 55:68–73.

105. Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science 343:80–84.

106. Javaid N, Choi S. 2021. CRISPR/Cas System and Factors Affecting Its Precision and Efficiency. Frontiers in Cell and Developmental Biology 9:761709.

107. Thyme SB, Akhmetova L, Montague TG, Valen E, Schier AF. 2016. Internal guide RNA interactions interfere with Cas9-mediated cleavage. 1. Nature Communications 7:11750.

108. Liu G, Zhang Y, Zhang T. 2020. Computational approaches for effective CRISPR guide RNA design and evaluation. Computational and Structural Biotechnology Journal 18:35–44.

109. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell 152:1173–1183.

110. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell 154:442–451.

111. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, P R Iyer E, Lin S, Kiani S, Guzman CD, Wiegand DJ, Ter-Ovanesyan D, Braff JL, Davidsohn N, Housden BE, Perrimon N, Weiss R, Aach J, Collins JJ, Church GM. 2015. Highly efficient Cas9-mediated transcriptional programming. 4. Nature Methods 12:326–328.

112. Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, Lim WA. 2015. Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. Cell 160:339–350.

113. Deaner M, Alper HS. 2017. Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in *Saccharomyces cerevisiae*. Metabolic Engineering 40:14–22.

114. Jensen ED, Ferreira R, Jakočiūnas T, Arsovska D, Zhang J, Ding L, Smith JD, David F, Nielsen J, Jensen MK, Keasling JD. 2017. Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. Microbial Cell Factories 16:46.

115. Kantor A, McClements ME, MacLaren RE. 2020. CRISPR-Cas9 DNA Base-Editing and Prime-Editing. 17. International Journal of Molecular Sciences 21:6240.

116. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. 7603. Nature 533:420–424.

117. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551:464–471.

118. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. 2000. Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. Cell 102:553–563.

119. Brar SS, Sacho EJ, Tessmer I, Croteau DL, Erie DA, Diaz M. 2008. Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. DNA Repair 7:77–87.

120. Bransteitter R, Pham P, Scharff MD, Goodman MF. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proceedings of the National Academy of Sciences 100:4102–4107.

121. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z, Kondo A. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353(605):aaf8729

122. Ma Y, Zhang J, Yin W, Zhang Z, Song Y, Chang X. 2016. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nature Methods 13:1029–1035.

123. Hess GT, Frésard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC. 2016. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. 12. Nature Methods 13:1036–1042.

124. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. 7785. Nature 576:149–157.

125. Vad-Nielsen J, Lin L, Bolund L, Nielsen AL, Luo Y. 2016. Golden Gate Assembly of CRISPR gRNA expression array for simultaneously targeting multiple genes. Cell Mol Life Sci 73:4315–4325.

126. Generoso WC, Gottardi M, Oreb M, Boles E. 2016. Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*. Journal of Microbiological Methods 127:203–205.

127. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. 2010. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. Science 329:1355–1358.

128. Ferreira R, Skrekas C, Nielsen J, David F. 2018. Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation Using Csy4 in *Saccharomyces cerevisiae*. ACS Synthetic Biology 7:10–15.

129. Deaner M, Mejia J, Alper HS. 2017. Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode dCas9 Activator in Saccharomyces cerevisiae. ACS Synthetic Biology 6:1931–1943.

130. Zhang Y, Wang J, Wang Z, Zhang Y, Shi S, Nielsen J, Liu Z. 2019. A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in *Saccharomyces cerevisiae*. Nature Communications 10:1053.

131. Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E. 2016. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature 532:517–521.

132. Świat MA, Dashko S, den Ridder M, Wijsman M, van der Oost J, Daran J-M, Daran-Lapujade P. 2017. FnCpf1: a novel and efficient genome editing tool for *Saccharomyces cerevisiae*. Nucleic Acids Research 45:12585–12598.

133. Campa CC, Weisbach NR, Santinha AJ, Incarnato D, Platt RJ. 2019. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. Nature Methods 16:887–893.

134. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 6:343–345.

135. Engler C, Kandzia R, Marillonnet S. 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLOS ONE 3:e3647.

136. Engler C, Gruetzner R, Kandzia R, Marillonnet S. 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIS Restriction Enzymes. PLOS ONE 4:e5553.

137. Szybalski W, Kim SC, Hasan N, Podhajska AJ. 1991. Class-IIS restriction enzymes — a review. Gene 100:13–26.

138. McCarty NS, Shaw WM, Ellis T, Ledesma-Amaro R. 2019. Rapid Assembly of gRNA Arrays via Modular Cloning in Yeast. ACS Synthetic Biology 8:906–910.

139. Shaw WM, Studená L, Roy K, Hapeta P, McCarty NS, Graham AE, Ellis T, Ledesma-Amaro R. 2022. Inducible expression of large gRNA arrays for multiplexed CRISPRai applications. 1. Nature Communications 13:4984.

140. Wang Y, Xue P, Cao M, Yu T, Lane ST, Zhao H. 2021. Directed Evolution: Methodologies and Applications. Chemical Reviews 121:12384–12444.

141. Traxlmayr MW, Faissner M, Stadlmayr G, Hasenhindl C, Antes B, Rüker F, Obinger C. 2012. Directed evolution of stabilized IgG1-Fc scaffolds by application of strong heat shock to libraries displayed on yeast. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1824:542–549.

142. Doerner A, Rhiel L, Zielonka S, Kolmar H. 2014. Therapeutic antibody engineering by high efficiency cell screening. FEBS Letters 588:278–287.

143. Zhao H, Chockalingam K, Chen Z. 2002. Directed evolution of enzymes and pathways for industrial biocatalysis. Current Opinion in Biotechnology 13:104–110.

144. Arnold FH, Volkov AA. 1999. Directed evolution of biocatalysts. Current Opinion in Chemical Biology 3:54–59.

145. Arnold FH. 2018. Directed Evolution: Bringing New Chemistry to Life. Angewandte Chemie International Edition 57:4143–4148.

146. Cirino PC, Mayer KM, Umeno D. 2003. Generating Mutant Libraries Using Error-Prone PCR, p. 3–10. In Directed Evolution Library Creation. Humana Press, New Jersey.

147. Copp JN, Hanson-Manful P, Ackerley DF, Patrick WM. 2014. Error-Prone PCR and Effective Generation of Gene Variant Libraries for Directed Evolution, p. 3–22. In Gillam, EMJ, Copp, JN, Ackerley, D (eds.), Directed Evolution Library Creation. Springer New York, New York, NY.

148. Ricardo PC, Françoso E, Arias MC. 2020. Fidelity of DNA polymerases in the detection of intraindividual variation of mitochondrial DNA. Mitochondrial DNA Part B 5:108–112.

149. Cravens A, Jamil OK, Kong D, Sockolosky JT, Smolke CD. 2021. Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering. Nature Communications 12:1579.

150. Greener A, Callahan M, Jerpseth B. 1997. An efficient random mutagenesis technique using an *E.coli* mutator strain. Molecular Biotechnology 7:189–195.

151. Badran AH, Liu DR. 2015. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. Nature Communications 6:8425.

152. Tizei PAG, Csibra E, Torres L, Pinheiro VB. 2016. Selection platforms for directed evolution in synthetic biology. Biochemical Society Transactions 44:1165–1175.

153. Moore CL, Papa LJ, Shoulders MD. 2018. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo. Journal of the American Chemical Society 140:11560–11564.

154. Crook N, Abatemarco J, Sun J, Wagner JM, Schmitz A, Alper HS. 2016. In vivo continuous evolution of genes and pathways in yeast. 1. Nature Communications 7:13051.

155. Ravikumar A, Arrieta A, Liu CC. 2014. An orthogonal DNA replication system in yeast. Nature Chemical Biology 10:175–177.

156. Ravikumar A, Arzumanyan GA, Obadi MKA, Javanpour AA, Liu CC. 2018. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. Cell 175:1946-1957.e13.

157. Álvarez B, Mencía M, de Lorenzo V, Fernández LÁ. 2020. In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. Nature Communications 11:6436.

158. Chen H, Liu S, Padula S, Lesman D, Griswold K, Lin A, Zhao T, Marshall JL, Chen F. 2020. Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. Nature Biotechnology 38:165–168.

159. Halperin SO, Tou CJ, Wong EB, Modavi C, Schaffer DV, Dueber JE. 2018. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature 560:248–252.

160. Tou CJ, Schaffer DV, Dueber JE. 2020. Targeted Diversification in the *S. cerevisiae* Genome with CRISPR-Guided DNA Polymerase I. ACS Synthetic Biology 9:1911–1916.

161. Jensen ED, Laloux M, Lehka BJ, Pedersen LE, Jakočiūnas T, Jensen MK, Keasling JD. 2021. A synthetic RNA-mediated evolution system in yeast. Nucleic Acids Research 49:e88.

162. Cameron DE, Bashor CJ, Collins JJ. 2014. A brief history of synthetic biology. 5. Nature Reviews Microbiology 12:381–390.

163. Stephanopoulos G. 2012. Synthetic Biology and Metabolic Engineering. ACS Synthetic Biology 1:514–525.

164. Hicks M, Bachmann TT, Wang B. 2020. Synthetic Biology Enables Programmable Cell-Based Biosensors. ChemPhysChem 21:132–144.

165. Decoene T, De Paepe B, Maertens J, Coussement P, Peters G, De Maeseneire SL, De Mey M. 2018. Standardization in synthetic biology: an engineering discipline coming of age. Critical Reviews in Biotechnology 38:647–656.

166. Shetty RP, Endy D, Knight TF. 2008. Engineering BioBrick vectors from BioBrick parts. Journal of Biological Engineering 2:5.

167. Anderson Jc, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, Keasling JD. 2010. BglBricks: A flexible standard for biological part assembly. Journal of Biological Engineering 4:1.

168. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLOS ONE 6:e16765.

169. Lee ME, DeLoache WC, Cervantes B, Dueber JE. 2015. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synthetic Biology 4:975–986.

170. Bernard P, Gabarit P, Bahassi EM, Couturier M. 1994. Positive-selection vectors using the F plasmid ccdB killer gene. Gene 148:71–74.

171. Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B. 2021. shiny: Web Application Framework for R (1.6.0).

172. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M, Daran-Lapujade P, Pronk JT, van Maris AJA, Daran J-MG. 2015. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. FEMS Yeast Research 15(2).

173. Whelan WL, Gocke E, Manney TR. 1979. The CAN1 locus of *Saccharomyces Cerevisiae*: fine-structure analysis and forward mutation rates. Genetics 91:35–51.

174. Texari L, Dieppois G, Vinciguerra P, Contreras MP, Groner A, Letourneau A, Stutz F. 2013. The Nuclear Pore Regulates GAL1 Gene Transcription by Controlling the Localization of the SUMO Protease Ulp1. Molecular Cell 51:807–818.

175. R. Stockwell S, R. Landry C, A. Rifkin S. 2015. The yeast galactose network as a quantitative model for cellular memory. Molecular BioSystems 11:28–37.

176. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, Doudna JA, Liu DR. 2020. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nature Biotechnology 38:883–891.

177. 2002. tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli.* The EMBO Journal 21:3841–3851.

178. Rees HA, Wilson C, Doman JL, Liu DR. 2019. Analysis and minimization of cellular RNA editing by DNA adenine base editors. Science Advances 5:eaax5717.

179. Ta V-T, Nagaoka H, Catalan N, Durandy A, Fischer A, Imai K, Nonoyama S, Tashiro J, Ikegawa M, Ito S, Kinoshita K, Muramatsu M, Honjo T. 2003. AID mutant analyses indicate requirement for class-switch-specific cofactors. 9. Nature Immunology 4:843–848.

180. Krivoruchko A, Nielsen J. 2015. Production of natural products through metabolic engineering of *Saccharomyces cerevisiae*. Current Opinion in Biotechnology 35:7–15.

181. Orth JD, Thiele I, Palsson BØ. 2010. What is flux balance analysis? 3. Nature Biotechnology 28:245–248.

182. Shi S, Chen Y, Siewers V, Nielsen J. 2014. Improving Production of Malonyl Coenzyme A-Derived Metabolites by Abolishing Snf1-Dependent Regulation of Acc1. MBio 5:e01130-14.

183. Chen WN, Tan KY. 2013. "Malonate Uptake and Metabolism in *Saccharomyces cerevisiae*." Applied Biochemistry and Biotechnology 171:44–62.

184. Roell GW, Zha J, Carr RR, Koffas MA, Fong SS, Tang YJ. 2019. Engineering microbial consortia by division of labor. Microbial Cell Factories 18:35.

185. Brenner K, You L, Arnold FH. 2008. Engineering microbial consortia: a new frontier in synthetic biology. Trends in Biotechnology 26:483–489.

186. Camarasa C, Bidard F, Bony M, Barre P, Dequin S. 2001. Characterization of *Schizosaccharomyces pombe* Malate Permease by Expression in *Saccharomyces cerevisiae*. Applied and Environmental Microbiology 67:4144–4151.

187. Benatuil L, Perez JM, Belk J, Hsieh C-M. 2010. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Engineering, Design and Selection 23:155–159.

188. Verduyn C, Postma E, Scheffers WA, Van Dijken JP. 1992. Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8:501–517.

189. Dabirian Y, Li X, Chen Y, David F, Nielsen J, Siewers V. 2019. Expanding the Dynamic Range of a Transcription Factor-Based Biosensor in *Saccharomyces cerevisiae*. ACS Synthetic Biology 8:1968–1975.

190. Tan J, Zhang F, Karcher D, Bock R. 2019. Engineering of high-precision base editors for site-specific single nucleotide replacement. 1. Nature Communications 10:439.

191. Muramatsu M, Nagaoka H, Shinkura R, Begum NA, Honjo T. 2007. Discovery of Activation-Induced Cytidine Deaminase, the Engraver of Antibody Memory, p. 1–36. In Advances in Immunology. Academic Press.

192. Brar SS, Sacho EJ, Tessmer I, Croteau DL, Erie DA, Diaz M. 2008. Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. DNA Repair 7:77–87.

193. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 422:726–730.

194. Bransteitter R, Pham P, Calabrese P, Goodman MF. 2004. Biochemical Analysis of Hypermutational Targeting by Wild Type and Mutant Activation-induced Cytidine Deaminase. Journal of Biological Chemistry 279:51612–51621.

195. Sakata RC, Ishiguro S, Mori H, Tanaka M, Tatsuno K, Ueda H, Yamamoto S, Seki M, Masuyama N, Nishida K, Nishimasu H, Arakawa K, Kondo A, Nureki O, Tomita M,

Aburatani H, Yachie N. 2020. Base editors for simultaneous introduction of C-to-T and A-to-G mutations. 7. Nature Biotechnology 38:865–869.

196. Bergenholm D, Dabirian Y, Ferreira R, Siewers V, David F, Nielsen J. 2021. Rational gRNA design based on transcription factor binding data. Synthetic Biology 6:ysab014.

197. Liu Y, Lin Y, Guo Y, Wu F, Zhang Y, Qi X, Wang Z, Wang Q. 2021. Stress tolerance enhancement via SPT15 base editing in *Saccharomyces cerevisiae*. Biotechnology for Biofuels 14:155.

198. Kuang Y, Li S, Ren B, Yan F, Spetz C, Li X, Zhou X, Zhou H. 2020. Base-Editing-Mediated Artificial Evolution of OsALS1 In Planta to Develop Novel Herbicide-Tolerant Rice Germplasms. Molecular Plant 13:565–572.