Characterization of new parts for synthetic biology applications in bacteria and yeast supported by automated methods.

Charakterisierung neuer Parts für Anwendungen in der synthetischen Biologie in Bakterien und Hefen durch Unterstützung automatisierter Methoden.
Zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)
Genehmigte Dissertation von Thomas Zoll aus Beckendorf-Neindorf
Tag der Einreichung: 04. April 2022, Tag der Prüfung: 07. Oktober 2022

1. Gutachten: Prof. Dr. Johannes Kabisch
2. Gutachten: Prof. Dr. Beatrix Süß
Darmstadt, Technische Universität Darmstadt

TECHNISCHE UNIVERSITÄT DARMSTADT

Biology Department
Computer-aided Synthetic Biology

Characterization of new parts for synthetic biology applications in bacteria and yeast supported by automated methods.
Charakterisierung neuer Parts für Anwendungen in der synthetischen Biologie in Bakterien und Hefen durch Unterstützung automatisierter Methoden.

Accepted doctoral thesis by Thomas Zoll

Date of submission: 04. April 2022
Date of thesis defense: 07. Oktober 2022

Darmstadt, Technische Universität Darmstadt

“In a dark place we find ourselves, and a little more knowledge lights our way.”
– *Yoda*

Erklärungen laut Promotionsordnung

§ 8 Abs. 1 lit. c PromO

Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§ 8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§ 9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§ 9 Abs. 2 PromO

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 04. April 2022

T. Zoll

Ehrenwörtliche Erklärung:

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den

.....

Zusammenfassung

Der kürzlich vorgestellte IPCC Bericht mahnt uns erneut zu einer Verstärkung der Ambitionen, eine nachhaltige Wirtschaft aufzubauen. Auch die UN betonen, dass der größte Teil der Wirtschaft noch nach einem linearen Modell funktioniert und dringend reformiert werden muss. Einen wertvollen Beitrag können biotechnologische Anwendungen leisten, da sie energieintensive Prozesse der chemischen Industrie ersetzen können. Zusätzlich haben sie den Charme, dass die eingesetzten Organismen nicht auf Erdölerzeugnisse angewiesen sind, sondern kostengünstige Energieträger aus industriellen Nebenerzeugnissen als Energiequelle nutzen können. Besonders die Ölhefe *Yarrowia lipolytica* entwickelt sich zu einem wertvollen Arbeitstier für die Industrie aber auch für die Wissenschaft.
Leider steht für die Hefe bisher nur ein begrenztes Repertoire für das Metabolic Engineering oder für den Aufbau von Schaltkreisen in der synthetischen Biologie zur Verfügung. Durch die natürliche Anreicherung von Lipiden stellt die Hefe einen interessanten Produktionsorganismus für die Industrie dar. Diese Arbeit zielt unter anderem darauf ab, das Portfolio an SynBio Parts für diese Hefe zu erweitern. Dazu werden in chapter 3 Möglichkeiten vorgestellt, den Organismus weiter zu erschließen. So konnte eine Multiorganismen-Plasmid (section 3.3.1) entwickelt werden, das das schnelle Testen von Konstrukten in drei Organismen ermöglicht. Dieses Plasmid konnte dazu eingesetzte werden, ein Tetracycline Aptamer in *Y. lipolytica* zu testen (section 3.3.2). Das Aptamer wurde zuvor für *Saccharomyces cerevisiae* entwickelt und konnte dort erfolgreich für die Abschaltung der Translation verwendet werden. *Y. lipolytica* konnte erfolgreich mit einem Monomer des Aptamers modifiziert werden. Bedauerlicherweise war der Stamm, der das Aptamer trug, deutlich in ihrer Fitness beeinträchtigt. Dadurch war eine eindeutige Aussage, über die Funktionalität nicht möglich.
Eine Erweiterung des Repertoires für *Y. lipolytica* und andere Organismen wurde durch die Etablierung von artifiziellen Landing-Pads verfolgt (section 3.3.3). Die Computer-generierten DNA-Sequenzen wurden so generiert, dass sie in den Organismen *Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae* und *Yarrowia lipolytica* möglichst wenig Off-Target Effekte beim Einsatz von Cas9 hervorrufen. Dazu wurden in einer bis zu 900 bp langen Sequenz bis zu 25 artifizielle Protospacer aufgereiht. Diese konnten erfolgreich in das Genom der Organismen integriert werden. Als Proof-of-Concept dienten Versuche mit CRISPRi und CRISPRa, sodass durch die Adressierung unterschiedlicher Protospacer eine abgestufte Expressionsstärke erreicht wird, die sich nach dem Abstand zwischen Protospacer und Reportergen richtet. Eine fein abgestufte Aktivierung der Expression konnte in *S. cerevisiae* gezeigt werden. Durch den Einsatz von CRISPRa konnte die Basalaktivität des Minimalpromotors um das 3.4-fache gesteigert werden. In *Y. lipolytica* gab es keinen klaren Gradienten, wie in *S. cerevisiae*. Der Effekt von CRISPRi in *B. subtilis* war abhängig vom eingesetzten Promotor.
Zur Unterstützung der molekularbiologischen und mikrobiologischen Arbeiten wurden Methoden zur Automation entwickelt (chapter 4). Durch einen klar erkennbaren „*Edge-Effekt*“ bei der Nutzung des Inkubators der CompuGene Robotics Plattform, wurde eine Methode entwickelt, die die Proben in den Mikrotiterplatten zufällig verteilt (section 4.4). Diese Methode kann zwar nicht den Randeffekt unterbinden, kann aber die Vergleichbarkeit zwischen den Proben verbessern. Außerdem konnte eine Möglichkeit entwickelt werden, um den Pipettierroboter aus der Automationsanlage zum Picken von Kolonien zu verwenden (section 4.5).

Abstract

The recently presented IPCC report once again urges us to intensify efforts to build a sustainable economy. The UN also emphasizes that most of the economy still operates in a linear model and urgently needs to be reformed. Biotechnological applications can offer a valuable contribution, as they can replace energy-intensive processes in the chemical industry. In addition, they have the charm that the growth of the used organisms is not dependent on crude oil sources. Further, they can use inexpensive energy sources from industrial by-products as energy sources. The oil yeast *Yarrowia lipolytica* in particular is emerging as a valuable workhorse for industry and science.
Due to the natural accumulation of lipids, the yeast represents an interesting production host for industrial applications. To date, unfortunately, only a limited repertoire of metabolic engineering approaches or SynBio circuits is available for yeast. This thesis aims, among other things, to expand the portfolio of SynBio parts for this yeast. To this end, chapter 3 presents ways to further exploit the organism. Thus, a broad-host-range plasmid (section 3.3.1) was developed, which allows rapid testing of constructs in three organisms. This plasmid could be used to test a tetracycline aptamer in *Y. lipolytica* (section 3.3.2). The aptamer was previously developed for *Saccharomyces cerevisiae* and could be successfully used for translation inactivation. *Y. lipolytica* was successfully modified with a monomer of the aptamer. Unfortunately, growth assays revealed that the strain carrying the aptamer was severely impaired in fitness. Thus, a clear conclusion about the functionality was not possible.
An expansion of the repertoire of *Y. lipolytica* and other organisms was pursued by establishing artificial landing pads (section 3.3.3). The computer-generated DNA sequences were designed to cause as little off-target effects as possible when using Cas9 in the organisms *Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae*, and *Yarrowia lipolytica*. To this end, up to 25 artificial protospacers were lined up in a sequence up to 900 bp long. These were successfully integrated in the genome of the organisms. Experiments with CRISPRi and CRISPRa served as proof-of-concept. By addressing different protospacers, different expression levels should be achieved depending on the distance between protospacer and reporter gene. Fine-graded activation of expression was demonstrated in *S. cerevisiae*. The use of CRISPRa increased the basal activity of the minimal promoter by 3.4-fold. In *Y. lipolytica*, there was no clear gradient as in *S. cerevisiae*. The effect of CRISPRi in *B. subtilis* was dependent on the promoter used.
Automation methods were developed to support the molecular biology and microbiology work (chapter 4). Due to a clearly visible edge effect when using the incubator of the CompuGene Robotics platform, a method was developed to randomize the samples in the microtiter plates (section 4.4). Here it was shown that this method cannot eliminate the edge effect, but it can improve comparability between samples. In addition, a method was developed to use the liquid handler of the automation platform to pick colonies from agar plates (section 4.5).

Acknowledgments

Diese Arbeit ist nicht die Leistung einer einzelnen Person, sondern eine Arbeit von vielen. Ohne die Mithilfe von Kollegen, Freunden und meiner Familie, die mich während der gesamten Bearbeitungszeit unterstützt haben, wäre diese Arbeit nicht möglich gewesen. Deshalb möchte ich an dieser Stelle allen Beteiligten nur eines sagen: Danke!

Insbesondere möchte ich mich bei Prof. Johannes Kabisch für die Bereitstellung des Themas, die vielen spannenden Diskussionen, die Unterstützung und die Betreuung der Thesis bedanken. Ohne seine herausfordernden und guten Ideen, sowie seine Motivationsreden hätte es diese Arbeit wohl nicht gegeben.

Ein großer Dank gilt Prof. Beatrix Süß, die als Zweitgutachterin diese Arbeit beurteilt. Ein weiteres Danke gilt speziell Ihr und ihrer gesamten Arbeitsgruppe für die viele Unterstützung nach dem Umzug von der Universität Greifswald an die TU Darmstadt, insbesondere durch Marc Vogel und Michael Vockenhuber und des Weiteren auch für die Möglichkeit, die Infrastruktur der Arbeitsgruppe mitzuverwenden.

Vielen Dank an Prof. Dominik Niopek und Prof. Viktor Stein als Prüfende bei der Disputation der Thesis.

Außerdem möchte ich mich für die Unterstützung von allen Beteiligten aus dem CompuGene Projekt des HMWK bedanken. Insbesondere die vielen wertvollen, interdisziplinären Unterhaltungen bei Projekttreffen und Symposien haben mir einen Blick über den Tellerrand der Biologie hinaus ermöglicht. Danke auch an Dr. Brigitte Held und Dr. Anja Hofmann, für die vielen Einblicke und Gespräche auf der gemeinsamen Pendelstrecke nach Frankfurt. Danke an Anja Hofmann für die Versorgung mit Stämmen und Konstrukten für weiterführende Arbeiten.

Herzlichen Dank an alle ehemaligen Mitglieder der Arbeitsgruppe Computer-aided Synthetic Biology. Es hat echt viel Spaß gemacht, mit euch zusammenzuarbeiten. Ein riesiges Danke an die beiden dienstältesten Greifswalder Kollegen Silke Hackenschmidt und Florian Nadler, aber auch an unsere Postdocs Dr. Stefan Bruder und Dr. Felix Bracharz, für die vielen wertvollen Diskussionen und aufbauenden Gespräche, sonst hätte die Doktoranden-Zeit wohl schon früher geendet.

Vielen herzlichen Dank an Dr. Marianna Karava für die Hinweise und Korrekturen der Thesis und die diversen aufbauenden Gespräche.

Ein riesiges Dankeschön an alle, die mich in meine heutige berufliche Ausrichtung geschubst haben. Vielen Dank an Dr. Mark Dörr von der Universität Greifswald für die ersten Gehversuche in der Laborautomation. Danke an Dr. Niels Schlichting und Prof. Johannes Kabisch für die vielen Diskussionen über die Etablierung unserer Anlage. Nicht vergessen werden soll die andauernde Unterstützung durch das Team der Analytik Jena GmbH, allen voran Albert Koch, aber auch alle anderen Teammitglieder. Herzlichen Dank an Aron Eiermann für die vielen wertvollen Gespräche.

Außerdem möchte ich mich bei unseren externen Partnern bedanken, die mit Rat und Tat, aber auch mit Stämmen und Konstrukten diese Arbeit ermöglicht haben. Danke an Dr. Alexander Elsholz, Dr. Knut Finstermeier und Vanessa Munoz an das MPUSP in Berlin, sowie Leonie Baumann von der AG Boles an der Goethe-Universität Frankfurt.

Ein herzliches Dankeschön gilt meiner Familie, ohne deren Unterstützung ich heute nicht da wäre, wo ich

bin und es auch diese Arbeit nicht gegeben hätte. Danke, dass ihr für mich da wart, auch wenn ich mal wieder gestresst und schwierig war und so nicht den Weg in die Heimat gefunden habe.

Benim Askim, danke, dass du mich die ganze Zeit über ausgehalten hast, auch wenn ich zuweilen genervt und gestresst war. Danke, dass du zum Schluss wirklich nicht müde geworden bist, mich zu pushen, um die Arbeit fertig zu stellen.

Mit Sicherheit habe ich jetzt viele vergessen, die mich bei der Erstellung der Thesis unterstützt haben, sei es durch wertvolle Tips oder emotionalen Beistand und daher noch einmal einfach nur: **Danke!**

Contents

1. Introduction **1**
- 1.1. Metabolic engineering . . . 1
- 1.2. Synthetic Biology . . . 3
- 1.3. Microbial cell factories . . . 5
  - 1.3.1. *Escherichia coli* . . . 6
  - 1.3.2. *Saccharomyces cerevisiae* . . . 7
  - 1.3.3. *Bacillus subtilis* . . . 8
  - 1.3.4. *Yarrowia lipolytica* . . . 9
- 1.4. General research objective . . . 11

2. Materials and Methods **13**
- 2.1. Technical equipment, chemicals and consumables . . . 13
  - 2.1.1. Technical Equipment . . . 13
  - 2.1.2. Chemicals and Kits . . . 15
  - 2.1.3. Consumables . . . 16
  - 2.1.4. Software . . . 19
- 2.2. Strains and Plasmids . . . 20
- 2.3. Buffers and Media . . . 23
- 2.4. General methods . . . 27
  - 2.4.1. Molecularbiological methods . . . 27
  - 2.4.2. Microbiological methods . . . 30

3. FLEXpress - Methods for rapid prototyping in biotechnological applications **35**
- 3.1. Introduction . . . 35
  - 3.1.1. Riboswitches . . . 36
  - 3.1.2. Artificial riboswitches and a Tetracycline riboswitch . . . 37
  - 3.1.3. Plasmid based expression and it's limitations . . . 38
  - 3.1.4. Landing pads for stable Integration . . . 39
  - 3.1.5. CRISPR-Cas – A short history . . . 40
  - 3.1.6. CRISPR-Cas – A Tool for several purposes . . . 41
  - 3.1.7. Cas9 and Off-Target cleavage . . . 43
- 3.2. Scope of FLEXpress . . . 44
  - 3.2.1. FLEXpress-vector . . . 44
  - 3.2.2. Extending the application of the Tetracycline aptamer to a new host . . . 44
  - 3.2.3. CRISPRpads for cross-species applications . . . 45
- 3.3. Results . . . 47
  - 3.3.1. FLEXpress-Vector: Testing parts in up to three different organisms . . . 47
  - 3.3.2. Tetracycline aptamers for *Yarrowia lipolytica* . . . 48
  - 3.3.3. CRISPRpads in *E. coli*, *B. subtilis*, *S. cerevisiae* and *Y. lipolytica* . . . 52

3.4. Discussion . . . 66
3.4.1. FLEXpress . . . 66
3.4.2. Tetracycline aptamer in *Y. lipolytica* . . . 67
3.4.3. CRISPRpads in two prokaryotes and two eukaryotes . . . 68
3.5. Conclusion and Outlook . . . 73

4. Laboratory automation in microbiological applications 75
4.1. Introduction . . . 75
4.1.1. Lab automation as solution for replicability . . . 76
4.1.2. Advantages and disadvantages of laboratory automation . . . 77
4.2. CompuGene Robotics platform and a first workflow for characterization of microorganisms 79
4.3. Evolution of workflow development . . . 82
4.3.1. 1st Generation: Rigid loop-structures . . . 83
4.3.2. 2nd Generation: Separation into different tasks . . . 85
4.3.3. 3rd Generation: Encapsulated jobs for improved recycling . . . 89
4.3.4. Summary and concluding remarks . . . 95
4.4. Randomization to improve comparability . . . 95
4.4.1. Introduction and problem description . . . 95
4.4.2. Objective . . . 96
4.4.3. Results . . . 96
4.4.4. Summary and Outlook . . . 100
4.5. Colony Picking . . . 100
4.5.1. Introduction . . . 100
4.5.2. Current status and objective . . . 101
4.5.3. Results . . . 101
4.5.4. Discussion and concluding remarks . . . 105

Bibliography 109

A. Supplemental Information 133
A.1. Cloning . . . 133
A.1.1. Oligos . . . 133
A.1.2. PCR Fragments . . . 136
A.1.3. Constructed strains with relevant genotype . . . 145
A.1.4. Aptamers . . . 146
A.1.5. CRISPRpads . . . 146
A.2. TCapt: Additional methods and results . . . 150
A.2.1. FLEXpress vector cultivation . . . 150
A.2.2. Additional figures: FLEXpress vector cultivation . . . 151
A.2.3. Additional methods: cultivation conditions for automated aptamer cultivation experiment . . . 153
A.2.4. Additional methods: cultivation conditions for manual aptamer experiment . . . 153
A.2.5. Additional figures: Aptamer cultivation . . . 155
A.3. CRISPRpads: Additional Results and Methods . . . 159
A.3.1. CRISPRpad design and calculation . . . 159
A.3.2. Strain construction to evaluate CRISPRpads in *S. cerevisiae* BY4742 . . . 164
A.3.3. Cultivation of CRISPRpad strains derived from *S. cerevisiae* BY4742 . . . 164
A.3.4. Strain construction to evaluate CRISPRpads in *Y. lipolytica* Po1f . . . 165

A.3.5. Cultivation of CRISPRpad strains derived from *Y. lipolytica* Po1f 165
A.3.6. Strain construction to evaluate CRISPRpads in *E. coli* BW25113 166
A.3.7. Cultivation of CRISPRpad strains derived from *E. coli* BW25113 167
A.3.8. Cultivation of CRISPRpad strains derived from *B. subtilis* PY79 168
A.3.9. Additional figures: CRISPRpad cultivation 169
A.4. Automation: Additional figures 181

List of Figures

1.1. Workflow cycle of metabolic engineering . . . 2
1.2. Synthetic Biology DBTL Cycle . . . 4
1.3. Emerging organism *Y. lipolytica*. . . . 11

3.1. Scheme for mode of action of a riboswitch . . . 37
3.2. Scheme of CRISPR associated protein 9 (Cas9) and nuclease dead Cas9 (dCas9) . . . 41
3.3. Scheme of CRISPRa and CRISPRi . . . 42
3.4. Scheme of the CRISPRpad integration module . . . 45
3.5. FLEXpress vector for three different organisms . . . 47
3.6. Cultivation of *E. coli*, *S. cerevisiae* and *Y. lipolytica* with FLEXpress vector p26061 and p26062 48
3.7. Schematic representation of the mode of action of the Tetracycline aptamer . . . 49
3.8. Corrected fluorescent data from aptamer containing *S. cerevisiae* and *Y. lipolytica* . . . 50
3.9. Flow cytometry measurement of *Y. lipolytica* containing the Tetracycline aptamer . . . 51
3.10. Flow cytometry measurement of *S. cerevisiae* containing the Tetracycline aptamer . . . 52
3.11. Microscopy pictures of *Y. lipolytica* cells containing the Tetracycline aptamer . . . 53
3.12. Scheme of the protospacer distribution in the CRISPRpads Rank1 and Rank4 . . . 54
3.13. Protospacer context and cultivation results for CRISPRa in *S. cerevisiae* . . . 56
3.14. Protospacer context for CRISPRa in *E. coli* . . . 58
3.15. Cultivation results for CRISPRa in *E. coli* using different inducer concentration . . . 60
3.16. Protospacer context and cultivation results for CRISPRa in *Y. lipolytica* . . . 62
3.17. Protospacer context for CRISPRi in *B. subtilis* . . . 64
3.18. Cultivation results for CRISPRi in *B. subtilis* containing reporter with P_{43} . . . 65
3.19. Cultivation results for CRISPRi in *B. subtilis* containing reporter with P_{ylb} . . . 65
3.20. Flow cytometry results for CRISPRa in *E. coli* containing selected sgRNAs . . . 71
3.21. Selected *E. coli* strains with constitutive mRFP expression . . . 72

4.1. Scheme of the CompuGene Robotics Platform . . . 80
4.2. Workflow scheme for the characterization workflow . . . 81
4.3. Screenshot of Analytik Jena Composer Scripting Studio . . . 82
4.4. Workflow of 1st Generation . . . 83
4.5. Workflow of 2nd Generation . . . 90
4.6. Workflow of 3rd Generation . . . 94
4.7. Illustration of the evaporation problem . . . 97
4.8. Scheme for the randomization approach . . . 98
4.9. Example for the usage of the 3DpickO Server for colony picking . . . 103
4.10. Transformation of coordinate systems for colony picking to correct x,y values for colony picking . . . 104
4.11. Composer screenshot of the parallelized colony picking workflow . . . 105
4.12. Scheme for colony picking using two different workflows . . . 106

A.1. Fold change of mTFP fluorescence in *E. coli*, *S. cerevisiae* and *Y. lipolytica* with FLEXpress vector . . . 151
A.2. Growth of *E. coli*, *S. cerevisiae* and *Y. lipolytica*with FLEXpress vector . . . 151
A.3. Codon usage analysis of the mTFP gene for codon usage distribution in *E. coli* . . . 152
A.4. OD corrected fluorescent data for aptamer cultivation with *S. cerevisiae* and *Y. lipolytica* . . 155
A.5. OD data for aptamer cultivation with *S. cerevisiae* and *Y. lipolytica* . . . 155
A.6. OD data of a manual cultivation of *S. cerevisiae* and *Y. lipolytica* with aptamer . . . 156
A.7. Flow cytometry data of *Y. lipolytica* after 24 h of the aptamer cultivation . . . 157
A.8. G4 motif analysis of the aptamer (monomer) DNA sequence . . . 157
A.9. G4 motif analysis of the aptamer (dimer) DNA sequence . . . 158
A.10. G4 motif analysis of the aptamer (trimer) DNA sequence . . . 158
A.11. Sequence alignment of the CRISPRpads Rank1 and Rank4 . . . 161
A.12. Growth of *E. coli* with CRISPRpads and CRISPRa in different medias . . . 169
A.13. Growth of *E. coli* with CRISPRpads and CRISPRa in MSM using different glucose concentrations 170
A.14. Comparison of growth for *E. coli* with CRISPRpads and CRISPRa in MSM using different inducer concentrations . . . 171
A.15. Normalized FI for *E. coli* with CRISPRpads and CRISPRa in MSM using different inducer concentrations . . . 172
A.16. Subtracted normalized FI for *E. coli* strains with CRISPRpads in MSM using different inducer concentrations . . . 173
A.17. Growth of *S. cerevisiae* with CRISPRpads and CRISPRa . . . 174
A.18. Normalized FI of *S. cerevisiae* with CRISPRpads and CRISPRa . . . 174
A.19. Fold change of activation over the time of *S. cerevisiae* with CRISPRpads and CRISPRa . . . 175
A.20. Growth of *Y. lipolytica* with CRISPRpads and CRISPRa . . . 176
A.21. Normalized FI of *Y. lipolytica* with CRISPRpads and CRISPRa . . . 176
A.22. Fold Change of activation over the time of *Y. lipolytica* with CRISPRpads and CRISPRa . . . 177
A.23. Growth of *B. subtilis* with CRISPRpads and CRISPRi in MSM with glucose . . . 178
A.24. Growth of *B. subtilis* with CRISPRpads and CRISPRi in MSM with fructose . . . 179
A.25. Normalized FI of *B. subtilis* with CRISPRpads and CRISPRa in MSM with fructose . . . 180
A.26. Colony Picking using 3DpickO OpenCV. Camera: Canon, Pattern: 96 Well . . . 181
A.27. Colony Picking using 3DpickO OpenCV. Camera: Canon, Pattern: random . . . 182
A.28. Colony Picking using 3DpickO OpenCV. Camera: Sony, Pattern: 96 Well . . . 183
A.29. Colony Picking using 3DpickO OpenCV. Camera: Sony, Pattern: random . . . 184
A.30. Colony Picking using Platescan. Multiple target plates - 1/2 . . . 185
A.31. Colony Picking using Platescan. Multiple target plates - 2/2 . . . 186
A.32. Colony Picking using Platescan. One target plate . . . 187

List of Tables

2.1. Technical equipment . . . 13
2.2. Chemicals . . . 15
2.3. Consumables . . . 16
2.4. Restriction endonuclease and polymerases . . . 18
2.5. Software . . . 19
2.6. Original vectors and strains . . . 20
2.7. Relevant genotypes of used strains . . . 22
2.8. Buffer recipes . . . 23
2.9. Culturing media and supplements . . . 23
2.10. Amino acid stock solutions . . . 25
2.11. Antibiotic stock solutions . . . 25
2.12. Standard PCR program for Polymerase X . . . 27
2.13. Standard PCR program for Q5® Polymerase . . . 28
2.14. Standard PCR program for *Taq* . . . 28
2.15. Typical PCR reaction . . . 28

3.1. Change of fluorescence with/without Tetracycline aptamer . . . 38

4.1. Automation levels for biological examples . . . 77
4.2. Advantages and disadvantages of lab automation . . . 78
4.3. Picking results after detection with 3DPickO and Platescan . . . 106
4.4. Picking times for FeliX SELECT . . . 107

A.1. Oligonucleotides for sequencing, PCR and cloning . . . 133
A.2. PCR Fragments for Cloning . . . 136
A.3. PCR Fragments used for given plasmids . . . 141
A.4. Constructed strains . . . 145
A.5. Available protospacer in Rank1 and Rank4 . . . 159
A.6. Off-Target sequences identified by cas-offinder . . . 162

Abbreviations

5-FOA 5-Fluoroorotic Acid
ARS autonomously replicating sequences
bp base pair
Cas9 CRISPR associated protein 9
CDS coding DNA sequence
CDW cell dry weight
CEN centromere
CPad CRISPRpad
CRISPR clustered regularly interspaced short palindromic repeat
CRISPRa CRISPR activation
CRISPRi CRISPR interference
dCas9 nuclease dead Cas9
DSB double-strand break
DBTL design-test-build-learn
DNA deoxyribonucleic acid
FDA Food and Drug Administration
FI fluorescence intensity
FSC forward scatter
GOI gene of interest
GRAS generally recognized as safe
HR homologous recombination
IPC inter-process communication
MCF microbial cell factories
MTP microtiter plate
NHEJ nonhomologous end joining
OE-PCR overlap extension PCR
OMICS the collective characterization and quantification of pools of biological molecules
ori origin of replication
OD OD_{600}
PAM protospacer adjacent motif
PCR polymerase chain reaction
RBS ribosome-binding site
RNA ribonucleic acid
RT room temperature
SBS Society for Biomolecular Sciences
scRNA scaffold RNA
SD Shine–Dalgarno sequence
SELEX systematic evolution of ligands by exponential enrichment
sgRNA single-guide RNA
SLiCE seamless ligation cloning extract
SpCas9 Cas9 from *Streptococcus pyogenes*
SSC side scatter
SynBio synthetic biology
TALEN transcription activator-like effector nucleases
TAR transformation associated recombination
TSS transcriptional start site
UTR untranslated region
TCapt Tetracycline aptamer
UAS upstream activation sequence
ZFN zinc finger nucleases

1. Introduction

Today's society is characterized by a continuously increasing consumption of resources. Thus, it is not surprising, that Intasian and colleagues use a total of three pages in a review on metabolic engineering about global resource needs and opportunities that are being ignored [160]. But the United Nations also makes clear that "take, make, waste" or better a "linear" economy no longer has a future [160][350]. The UN report again highlights that resource extraction has tripled since the year 1970[350]. But many possibilities are being missed by not using resources in a circular fashion. For example, only 8.6 % of the world economy operates on circular principle [350].
Summarized by Intasian *et al.* (2021)[160]: "As the world is facing a serious threat from the climate change crisis, a more sustainable solution for manufacturing, i.e., circular economy in which waste from the same or different industries can be used as feedstocks or resources for production offers an attractive industrial/business model." How many opportunities is left untapped is clearly illustrated by the amount of organic material in municipal waste, which is currently about 53 % [160] and in some cases even more [11]. In total, this amounts to about 2.6 t per day worldwide [160].
In addition, a major problem of industrial production is still that a large part of the products is generated by excavating new resources from ground resources such as oil, gas or minerals. But valuable building blocks on the way to a circular economy can be a changed product design by using different materials or reuse and even repair [350]. In addition, biotechnology offers important key technologies on the way to a sustainable circular economy. For example, industrial by-products, such as waste from agriculture or lipids from the food industry, can serve as energy sources for microbiological processes, allowing new high-value products to be generated [9][160]. In addition, biotechnology can replace large-scale industrial processes by using bio-catalysts, making them more sustainable [98]. For example, processes to produce pharmaceuticals, flavors, cosmetics and many other products are now being replaced, since these were mostly produced by oil derived products from the chemical industry [270].
However, bio-catalysts are used both for using additional feedstocks or carbon sources and for the production of high-value chemicals. These can be purified enzymes, cell-free systems or whole cells. But before they can be used for biotechnological processes, they must be made capable of performing their task. This requires methods of metabolic engineering and synthetic biology, which will be described briefly in the next two sections.

1.1. Metabolic engineering

The substitutions of chemical processes can be achieved by different methods. Some processes can already be accomplished by using purified enzymes [32][160][302]. For example, transesterification of vegetable oils to biodiesel can be done by using lipases rather than alkaline lysis with short-chain alcohols [233][266]. However, the use of enzymes can also have disadvantages, such as expensive purification of the bio-catalysts [160], degradation of the enzymes during the process [367], or limited reusability if the enzymes are not immobilized [221].
As an alternative for enzyme reactions, whole-cell systems are available that enable continuous processes. In addition, the expensive purification of the bio-catalysts is no longer necessary here [160]. Organisms

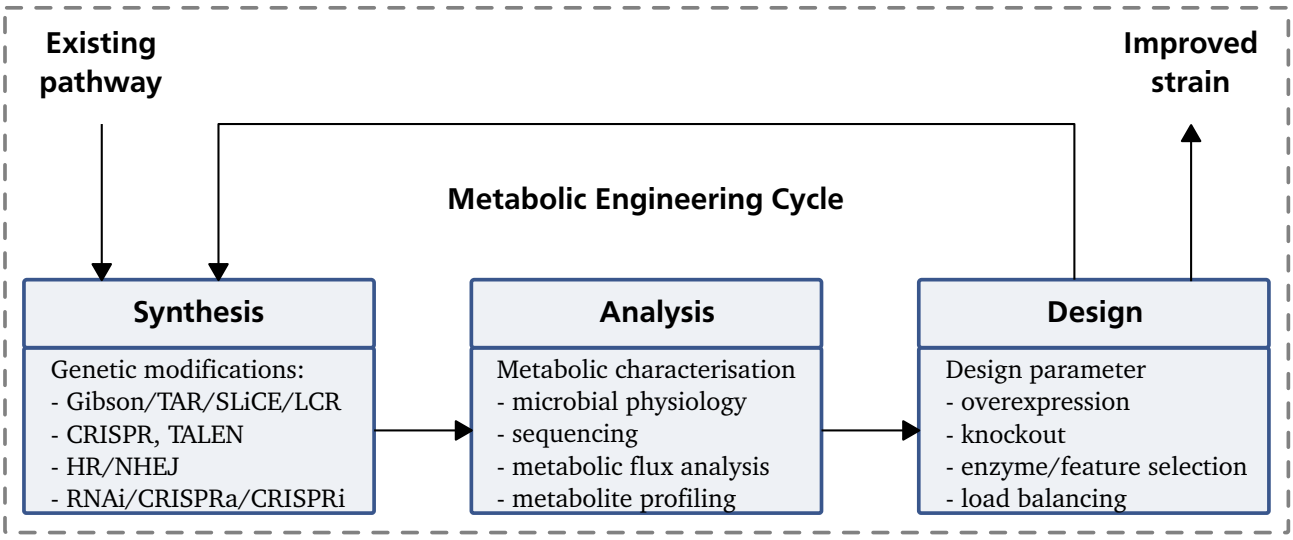


Figure 1.1.: Workflow of metabolic engineering. Modified from [358]

can be cultivated that already produce the desired product naturally [187] or organisms that have been modified using genetic engineering methods [358]. An overview of the advantages and disadvantages has been provided by Intasian *et al.* [160]. In this context, classical strain evolution tended to be carried out through untargeted interventions, e.g., the use of mutagenic radiation or chemicals [338]. With the rapid development of molecular biology methods, e.g., CRISPR-Cas9, TALEN, etc. [115][358]. Another advantage of targeted modification is the knowledge of the relationship between the resulting phenotype and the genetic modification carried out. This connection is important as it enables the basic principle of metabolic engineering. The basic principle was well summarized by Thykaer *et al.* in one sentence: "The essence of metabolic engineering is to apply analytical techniques for detailed phenotypic characterization of cells grown under industrial-like conditions to design directed genetic modifications that may be obtained through recombinant DNA technology, resulting in cells with improved properties."[358] Additionally, the workflow of metabolic engineering, illustrated in Fig. 1.1, is already very similar to the design-test-build-learn (DBTL) cycle of the synthetic biology (SynBio).
As the citation makes clear, metabolic engineering originates from industrial and applied research. Here, the optimization of existing production strains is usually carried out so that productivity and yield can be increased [270]. In addition, strain optimization is carried out under fixed criteria, i.e., a modified strain is reevaluated using the same analytical procedures as the precursor strain [358]. Additional experiments provide insight into the relationship between the new genotype and the resulting phenotype.
A key component of metabolic engineering is the metabolic analysis of the resulting mutants, which can be supported and modeled with mathematical models. For example, genome-scale metabolic (GEMs) models use sequencing data to obtain insights into metabolism [270]. This involves linking information from annotated gene sequences and experimental data to obtain predictions and descriptions for stoichiometry and mass balances in metabolic networks [133]. Another possibility is metabolic flux analyses (MFA), to determine fluxes in the steady state. These are supported by, for example, the carbon isotope C13 to measure flux in the system [270]. The analyses are used together with additional data from metabolic databases (MetaCyc, KEGG), enzyme databases (BRENDA), or organism-specific databases (HMDB, ECMDB). Other tools for metabolic flux analysis have been listed by Otero-Muras *et al.* [270], Fernández-Castané *et al.* [101], and Wang *et al.* [379]. Often, however, models may also introduce some uncertainty due to a lack

of knowledge, for example, insufficient annotation of the genome. In addition, these models can also be used to optimize fermentation processes [270], for example, when product accumulation in the medium has a negative impact on production.
In metabolic engineering, the metabolism itself is modified by methods of molecular biology. Different possibilities are available, such as deletion, overexpression, heterologous expression (enzymes/pathways) [358] or fine-tuning of the expression. For heterologous pathway expression, either entire pathways can be transplanted or completely reassembled [270]. In either case, detailed knowledge of the organism, as well as its interconnections between primary and secondary metabolism, is necessary in the design process [358]. Many other influences must be considered, such as the supply of energy (e.g., ATP), the supply of redox partners (e.g., NADH) or other important co-factors [270] and precursors. But also, the pathway itself has to be planned accordingly. Longer metabolic pathways have diverse intermediates, and these can accumulate in the cell. Some have an impact on flux through the pathway, others may be toxic to the host metabolism [105][115][192][270]. Thus, optimization of metabolic pathways also requires attention to the interplay of individual components, which requires detailed knowledge of enzyme properties, such as turn-over rate, total turnover number, product inhibition, and stability [143][192]. Furthermore, the competition with the host metabolism has to be considered and care has to be taken that intermediates do not drain into other metabolic pathways or that a too strong load [126][390] is created by the heterologous pathway.
Thus, it is necessary to decide for the best enzyme selection having the best flux through the pathway. Here, the models mentioned above assist and enable flux analyses through the pathway, identify interfering components of the host metabolism (potential target for deletion or overexpression) [358] and highlight potential balance problems. This can be used to influence expression, e.g., by using other promoters [320][270], changing gene dosage and alternate the ribosome-binding site (RBS) in bacteria [270]. The last mentioned points are strongly supported by the capabilities of synthetic biology [160].
SynBio and the introduction of automated OMICS approaches are creating an immense number of possible combinations. Sometimes, there is the problem that data from OMICS approaches are still insufficient and sometimes lack valuable information [115][358], but they are sufficient to model with sufficient accuracy and to make suggestions for optimization [270]. The amount of new combination possibilities exceeds the capacity to assemble and test the created designs [270]. The introduction of automation and the compatibility of some cloning methods to automation [316] can significantly accelerate this process. SynBio approaches also continue to offer new opportunities. Furthermore, from the basic idea, both metabolic engineering and SynBio are closely related [160].

1.2. Synthetic Biology

Perhaps, both fields cannot be clearly separated, meaning that they overlap in numerous areas [60]. Both disciplines use similar methods to achieve their goals, and these are often indistinguishable from each other. Cheng and Lu therefore propose a distinction: Metabolic engineering is the maximization of yield in biosynthetic pathways whereas SynBio is the integration of new pathways and optimization of flux and the ability to generate specific outputs [60]. Optimizing yields and productivity by optimizing flux was already the focus of metabolic engineering. Perhaps the selective control of expression and the generation of specific outputs, through the selective influence by external signals, is a characteristic that is unique to SynBio. Another point of distinction may also be the starting point of SynBio, which Stephanopoulos links to either the start of cost-effective deoxyribonucleic acid (DNA) synthesis or the development of the first counter [91] and toggle switches [116][339]. An interesting review about the history of SynBio was published by Cameron *et al.*[53] However, the current view on SynBio seems to be more technical as summarized by Carbonell *et al.*: ”Synthetic biology goes beyond the historical practice of a biological

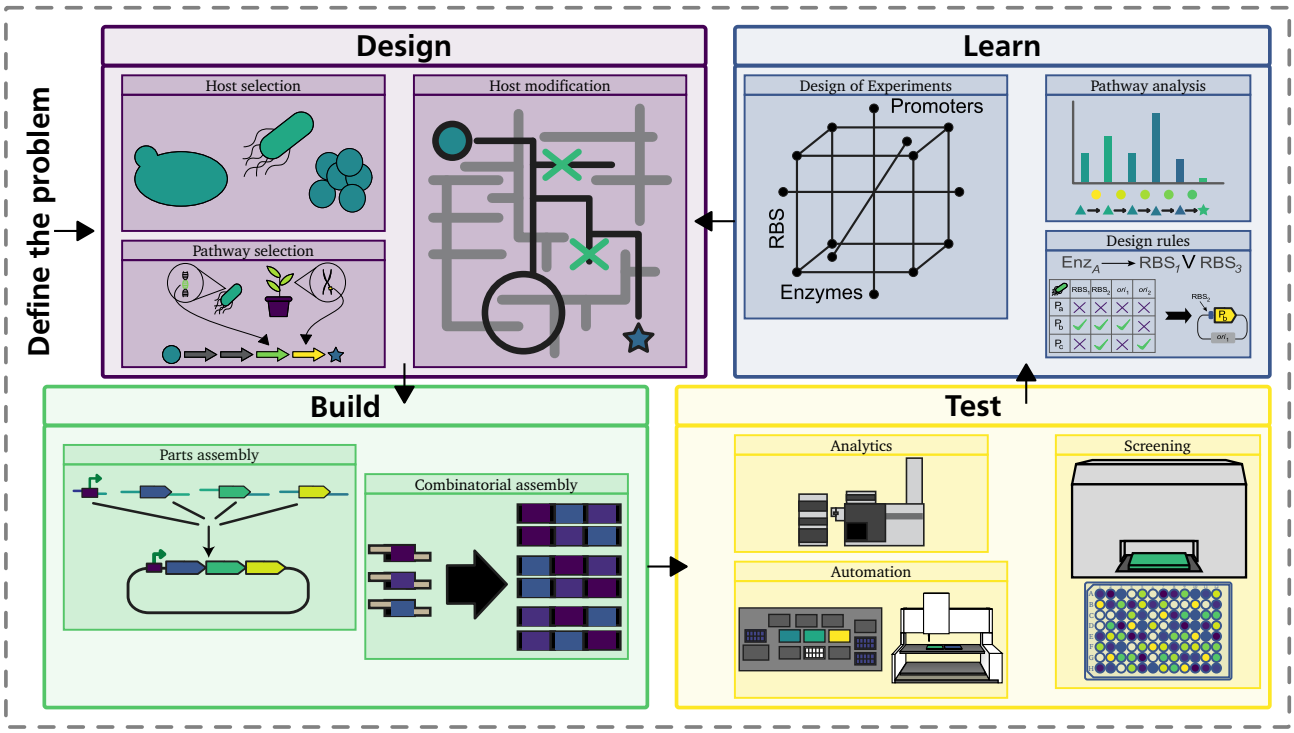


Figure 1.2.: Workflow of SynBio starting from problem definition. Modified from [109]

research based on describing and cataloguing [...], and aims to design biological systems to a given specification [...].”[55] ”Synthetic biology aims to capitalize on nature’s chemical and biological diversity by enabling the introduction of orthogonal, scalable, and robust functionalities into any living system and expands our ability to harness native biological systems for a wide range of applications.”[170]
To this end, SynBio brings together diverse groups of scientists from the fields of biological sciences, including biophysics, biochemistry, medicine, biomedical engineering, and molecular cell biology [60]. These are divided into two groups. One is developing biological components and systems to achieve specific output in biological systems whereas the other aims to provide self-replicating systems with useful functions [106]. This is done by adapting approaches from engineering and applying them to the design of biological systems [131]. This involves working in a similar principle to that already used in metabolic engineering (Fig. 1.1). However, since the focus here is on generating a new design rather than optimizing an established system, here the design is the starting point of a circular workflow that iteratively moves through Design → Build → Test and Learn (Fig. 1.2) [109].

The beginnings of circuit development probably originate from developments in microbiology [53], through the development of individual so-called parts, such as the toggle switch from the bacteriophage λ mentioned above [116][147]. The development of other switches originated from the natural control mechanisms of the model organism *E. coli*[53][131]. Today, the portfolio is much more diverse and allows intervention and control at various levels, thus enabling multiple layers of control over gene expression and signal transduction. For example, transcription factor-based systems are available that interact directly with their DNA counterpart [261][262][330]. In addition, proteins from natural signal transduction pathways are also available that can serve as sensing modules [337]. In addition, intervention and regulation options are available that are realized by aptamers, ribozymes, and TOEHOLD switches [127][129][381]. The aforementioned switches mediate targeted intervention using chemical and physical stimuli [60][142].

The general progress in modeling [147] and the wide use of OMICS technologies [109], makes the algorithm-driven design of new switches and circuits possible today. The basic prerequisite for this is a broad knowledge of the organism used [51] but also a good characterization of the parts and/or bio bricks [60][339]. The characterization of SynBio parts has been [60] and still is [55] a bottleneck. Despite the extensive use of OMICS technology, prediction and modeling of output is still error-prone, especially during industrial scale-up [55]. In addition, OMICS technologies are often too expensive for widespread use, so systems with simple readout (e.g., fluorescent proteins) are often employed for part the characterization [60].
But, the problem with generating parts libraries was the lack of a standard for characterizing the parts [60]. An attempt at standardization was pursued by Canton [54] and Arkin [7]. However, by using well-characterized parts, the development of complex circuits becomes possible in an automated way, as impressively shown by the development of Cello by Nielsen *et al.*[262] Using machine learning methods, design of experiments, and of OMICS technologies, the design can be further refined [109]. However, the use of machine learning cannot yet fully realize its potential because most data still contain too much noise. But these limitations can be compensated with the automation capabilities available today [55].
A further limitation was the widespread use of *E. coli* and *S. cerevisiae* in SynBio, but in recent years the portfolio has been extended significantly, e.g., to *Bacillus subtilis*, *Pseudomonas putida* but also to mammals and plants [51][131]. Some of these organisms have already been used as microbial cell factories for biotechnological applications.

1.3. Microbial cell factories

The use of microorganisms for the production of everyday products has a long history. Findings in Egypt from about 1500-1350 BC suggest that yeasts were consciously used to make bread [309]. Other explorations show the use of yeasts and *Lactobacillus sp.* to produce sourdough and fermented products in China from the 8th century B.C.[325]. In modern times, the portfolio in the use of cell factories has expanded drastically. In addition to their use for fermentation [377] and use probiotics [45], microorganisms are used as microbial cell factories (MCF) – both prokaryotes and eukaryotes [264] – in diverse industrial processes to produce fuels [166][374], pharmaceutics [123][264], cosmetics [136], amino acids [150], and many other products. Consequently, the term cell factory is used in biotechnology to refer to organisms that can produce a wide variety of molecules and materials [264].
Some organisms – such as *Aspergillus niger* [187][359] – already produce the desired product under certain fermentation conditions. Other organisms must be equipped with a heterologous pathway to produce the desired product using metabolic engineering or SynBio methods. As addressed in section 1.2, SynBio circuits and production pathways are often used in *E. coli* because the model organism is easy to modify, and the literature provides a lot of information about its metabolism. However, it is often questionable whether *E. coli* is a suitable host for the respective processes [197]. For some processes – e.g., for the production of substances for food industry – it does not seem to be a suitable host, due to the production of endotoxins and the formation of inclusion bodies [222][282][314]. Therefore, for use as MCF, some important characteristics can be listed for the selection as a production host. Important properties are: life-style, robust cell encapsulation, extensive knowledge of the metabolism, interaction possibilities between heterologous modules and host metabolism [106][197]. In addition, genetic accessibility plays an important role, i.e., a widely sequenced genome and molecular biological methods to modify the genome. But, SynBio methods have made organisms accessible that are considered to be rather difficult to manipulate [264][357]. Furthermore, non-biological aspects also play a role in biotechnological applications. For microbial production of food ingredients and therapeutics, classification as generally recognized as safe (GRAS) by the Food and Drug Administration (FDA) is an important aspect [222][323].

However, some MCFs, which will be used in the following chapters, are briefly introduced in the following sections. Two model organisms – *E. coli* and *S. cerevisiae* – that have prominent role in SynBio are briefly introduced. In addition, two organisms – *B. subtilis* and *Y. lipolytica* – are presented that have biotechnological relevance and are opened up for SynBio.

1.3.1. *Escherichia coli*

The workhorse in many areas of science and industry is *E. coli*. It was described as early as 1885 by Theodor Escherich and is part of the natural flora of animals and humans [97][225]. The bacterium is Gram-negative, rod-shaped, and facultative anaerobic. In addition to various pathogenic strains, the strains K12 and MG1655 and their derivatives are mainly recruited in the sciences [400]. The organism has become a model organism because it has been used for extensive research since the 1950s, e.g., it helped to elucidate the genetic code [68], DNA replication [218], and transcription [342], among others. Another milestone on the way to a model organism was the completely sequenced genome, which is available since 1997[27]. Nevertheless, *E. coli* was already being used biotechnologically in the 1970s. For example, scientists at Genentech succeeded in creating the first biotechnological process for the production of Somatostatin using *E. coli*[307]. The next major process was the production of Insulin [123][307] and many more processes were to follow [59].
The main advantages of *E. coli* for industrial processes are its short doubling time, and thus high growth rates, high cell densities during fermentation, it does not clump, it grows on many substrates and chemically defined media, and it has comparatively low production costs [3][59][69]. However, there are also disadvantages when using *E. coli*. Its use to produce food-grade products is limited by the formation of endotoxins [314]. Further, for the production of human proteins, the post-translational modification of recombinant proteins is lacking [103], as well as an efficient secretion system to facilitate downstream processing [243]. In addition, the formation of inclusion bodies leads to other problems [314][334].
However, genetic amenability is another key feature. The bacterium can be modified very easily by using plasmids. For this purpose, there are origin of replication (ori) with different copy numbers available, starting from about 1 copy up to 700 copies [69][334]. If a stable copy number is required, modifications of the genome are also possible, by RecET mediated recombination [304][403], the use of group II introns [189] or Lambda Red mediated homologous recombination [76]. The discovery and use of Cas9 has led to the development of additional methods of modification in combination with Lambda Red [291][304]. But modification of the *E. coli* genome is still difficult, especially when integrating larger constructs [206]. Nevertheless, the method of Datsenko and Wanner was used to create a knockout library (Keio collection) of K12 derivatives, allowing valuable insights on essential genes [10][395]. Various parts (promoters, RBS [376], insulators [275][262], riboswitches [78], etc.) are available for conditional gene expression and genetic circuit construction. In addition, the use of CRISPR activation (CRISPRa) [83] and CRISPR interference (CRISPRi) [25] is possible.
Furthermore, a variety of OMICS studies are available for the organism to create metabolic and regulatory models [3]. For example, Ishii *et al.* used a series of K12 mutants selected from the Keio collection to study changes in metabolism resulting from knockouts using DNA arrays, qRT-PCR, TOF-MS, and LC/MS [161]. This allowed the creation of metabolic models and flux analyses, providing access points for metabolic engineering and SynBio designs [138][269]. Further studies on transcriptomics [184], proteomics [335] and metabolomics [399] are available to investigate the response of the bacterium to a variety of stimuli [3].
The bacterium is used in SynBio and metabolic engineering for diverse purposes. One branch is the production of alternative fuels, such as hydrogen [239][361] or bioethanol [333]. In addition, the organism is used to produce amino acids, sugar alcohols, diols, and polymers [59], as well as a variety of

therapeutics [14]. Another important organism for biotechnology but also for SynBio is *S. cerevisiae*. The yeast is briefly introduced in the following section.

1.3.2. *Saccharomyces cerevisiae*

The yeast also known as Baker's yeast has been known for thousands of years. For example, a beer-like beverage was produced in China in 10,000 BC. Later, Lousi Pasteur found the connection between ethanol production and budding bodies. His method for purifying yeast strains led to the selection of certain yeasts for beer production at the Carlsberg Brewery in Copenhagen [210][263][369]. *S. cerevisiae* has become the most widely used yeast and the status as model organism due to its special characteristics. The high degree of conservation of certain cell properties and functions is remarkably similar to human cells, such as "autophagy, protein translocation and secretion, endoplasmic reticulum-associated protein degradation, heat shock, and protein folding and chaperone functions."[263] For example, findings about protein signaling peptides influencing cellular localization, and the protection of chromosome ends, through the attachment of DNA by telomerases, originated from studies with *S. cerevisiae* [152].
Since 1996, the genome of the reference strain S288C is available completely sequenced, which is considered as reference genome since then, which has been updated meanwhile [92][124]. Later, more than 1000 different isolates were sequenced and compared with each other, and minor difference was found between wild type isolates and industrial isolates [263][281]. As for *E. coli* with the Keio collection, deletion experiments were done with *S. cerevisiae*. These mutants were studied under different cultivation conditions, to gain valuable insights into the genotype-fitness relation under certain conditions such as salt or pH stress, allowing conclusions about gene function [34][119][388]. Such studies are made possible by the relative ease of modification of yeast by homologous recombination (HR), since polymerase chain reaction (PCR) fragments with short homology sites (35-51 bp) are sufficient for gene deletion [20]. The capabilities for efficient HR can be easily exploited for the cloning of plasmids, by the transformation associated recombination (TAR) method [212]. The efficiency is even sufficient for marker-less transformation [173], or for multiplexing of transformations by using CRISPR-Cas9 [241]. However, plasmids with different control sequences, as well as auxotrophic and dominant markers, are available for transformation, as summarized in an older review by Romanos *et al.*[303]
Other OMICS data have been collected, e.g., translation efficiency of mRNA to proteins was studied [209]. Further important studies have been conducted on the proteome. Since eukaryotic cells divide into different compartments, information on protein localization is particularly essential for optimizing flux in metabolic engineering and SynBio approaches. Thus, the fluorescent protein GFP has been used to perform localization and interaction experiments, by fusing GFP with the target protein [159]. Similar studies have been carried out with immuno tags [118]. The OMICS data were used to create various models of the yeast metabolism. For example, the GEM model of yeast metabolism was already published in version 8 (Yeast8) [148][232].
The data collected through laboratory experiments and models were already successfully used in metabolic engineering and SynBio [34][263]. The data collected on the sequence-to-protein relationship will support the creation of new designs and the optimization of existing ones [170][252]. Diverse well-characterized parts are available for yeast. The portfolio includes constitutive promoters, which were often derived from housekeeping genes, and inducible promoters (various GAL promoters, CUP1, Tet-ON/OFF, etc.) [157]. Additionally, heterologous promoters (e.g., from *Ashbya gossypii* [157]) or synthetic promoters [280] are available [295]. Some of the promoters were characterized under different fermentation conditions. Because this is a eukaryote, terminators, and the 3'UTR have role in circuit design as they have considerable influence on mRNA stability [81]. To this end, Yamanishi *et al.* characterized various terminators to be able to use them for design processes [396]. Another level of regulation can be provided by transcription

factors, which have sensing and regulatory functions [170]. For example, Naseri and colleagues were able to generate artificial transcription factors based on plant proteins, thereby enabling orthologous design [256]. Other orthogonal possibilities for regulation include the use of CRISPRa [401] and CRISPRi [122], as well as the use of sensors and regulators on RNA level, by using riboswitches [129][204].
S. cerevisiae not only brings a long history, but also important properties such as: robustness to fermentation conditions, stability to phages and inhibitors, and relatively high salt and pH tolerance [170][263]. One drawback is the substrate spectrum, as wild type strains are unable to utilize pentoses, a major component of lignocellulose [263]. However, utilization of xylose could be achieved by metabolic engineering [37]. Due to its history, *S. cerevisiae* is often used for the production of bioethanol [263] but also to produce n-butanol and isobutanol [211]. The host is also used for the production of flavors, amino acids, steroids, opioids, enzymes, and therapeutics [263]. Important therapeutics include the antimalarial drug Artemisinin, heterologously produced by metabolic engineering [271], and the anti-inflammatory Hydrocortisone [351]. A vivid example of metabolic engineering has been achieved by scientists from Amyris, which have re-engineered the central carbon metabolism for the production of isoprenoids [249].
Thus, baker's yeast represents an important production host for biotechnology, but it is also a model organism for basic research and SynBio. Some in the past less used organisms in SynBio, are *B. subtilis* and *Y. lipolytica*, which will be briefly introduced in the next two sections.

1.3.3. *Bacillus subtilis*

Like the previously described organisms, *B. subtilis* also has a long biotechnological history [406]. In the Asian region, *B. subtilis* has been used to produce fermented products, having different names depending on the country [100][406]. *B. subtilis* also appears to have a probiotic effect that strengthens the natural intestinal flora in humans [89]. The Gram-positive soil bacterium is rod-shaped, but can also form long chains separated by septa under certain growth conditions [95][345]. Like many other *Bacillus spp.*, *B. subtilis* can form endospores [89]. Because spores are easily transported by environmental forces, natural habitats are difficult to determine, so both terrestrial and marine isolates have been found [89]. The secretion of a wide variety of enzymes, helps the bacterium to survive in a wide variety of habitats [345]. Due to certain environmental factors, *B. subtilis* can form biofilms which are held together by proteins and polysaccharides. Three different types of biofilms are known [8]. Due to its natural occurrence on plants, the bacterium is also commonly found in the gastrointestinal tract of animals [8][89]. Due to the formation of biofilms and endospores, and secretion, as well as the ease of genetic modification, *B. subtilis* is used as model organism for physiology and metabolism studies [8][89][345].
The genetic information needed for the experiments is provided by the complete genome sequence of reference strain 168 [208], which was manually curated and updated, later [33]. The ease of genetic accessibility allowed extensive genome reduction studies that identified 271 essential genes under the growth conditions tested [201]. Similar to *E. coli* and *S. cerevisiae*, knockout libraries were created for *B. subtilis* to assess their gene functions [203]. However, gene analysis also revealed a variety of genes that show a strain-specific expression and are likely to be habitat-dependent [89]. The ease of genetic accessibility of *B. subtilis* is based on the formation of a natural competence which is strain-specific and influenced by external factors (e.g., nutrient deficiency) and growth phase [372][406]. The bacterium can bind and incorporate linear extracellular DNA for this purpose [345][406]. For genetic modification, homology sites of 400 to 500 bp in length are required for efficient homologous recombination, with integration efficiency decreasing as the length of the DNA being transferred decreases [345]. In addition, other OMICS data have been collected that are important for basic research as well as for metabolic engineering and SynBio. For example, a broad transcriptome study was conducted by Nicolas *et al.* with a total of 100 different conditions tested [260]. Other OMICS studies focused on the transition from

glucose to malate [44] or the influence of salt stress [139]. With the data, models were generated that cover the metabolism of *B. subtilis*. Among others, the iYO844 model uses protein content and turnover rates to predict ways to optimize the production of the industry-relevant substance poly-γ-glutamic acid in *B. subtilis* [246].
The large amount of information available and other important properties made *B. subtilis* a widely used organism in industry. With a doubling time of about 20 min, it grows comparatively fast [95]. In addition, *B. subtilis* has two secretion systems that can specifically and non-specifically release proteins into the surrounding medium [71][406]. Secretion via the sec-dependent transportation system or the twin-arginine translocation system is achieved by fusion with the respective signal peptides [71]. An important feature for protein production, for example of proteases or amylase, are the control mechanisms to recognize misfolded and incompleted proteins. However, these control systems can lead to bottlenecks in heterologous protein production [406]. Nevertheless, the bacterium is widely used to produce chemicals and antimicrobial materials in industry and medicine [345]. For example, chemicals such as vitamins, inositol, acetoin, riboflavin, subtilisin, and enzymes such as nattokinase and alpha-amylases are produced using *B. subtilis* [71][246][345]. In addition, the production of antibiotics, like sublancin, is achieved by metabolic engineering. Since the regulation of each pathway component is difficult to control, synthetic promoters have been specifically designed for this purpose [175].
In general, various tools are available for metabolic engineering and SynBio approaches. Constitutive (e.g., P_{43}, P_{luxS} [175]), inducible (e.g., P_{xyl}, P_{spac} [171], P_{glv} [175]), auto inducible [71][216], and synthetic promoters (e.g., P_{v1} [141]) are available to construct genetic circuits. Interestingly, the latter promoter allows constant increase in expression over time [141]. A listing and characterization of a wide variety of promoters was carried out by Yang *et al.* [397] and Song *et al.*[332] A library of RBS was provided by Guiziou *et al.* [134] Another tool for transcriptional control is the availability of CRISPRa [234] and CRISPRi [283]. For post-transcriptional control of expression, riboswitches were developed by Süß *et al.* [347] But like for the other organisms, standardization is lacking. Thus, Popp *et al.* suggested the introduction of BioBricks 2.0 for *B. subtilis* [288].
The potential use cases of *B. subtilis* in metabolic engineering and SynBio are versatile, especially due to the large amount of data and parts available. In the next section, *Y. lipolytica*, an organism which is still evolving into a broadly applicable organism in metabolic engineering and SynBio, is introduced.

1.3.4. *Yarrowia lipolytica*

The industrial workhorse and non-conventional yeast *Y. lipolytica* had a longer history and some reclassification of taxonomy [227][375]. The first description of the yeast *Mycotorula lipolytica* goes back to F.C. Harrison and was assigned to *Candida* meanwhile and finally to the new genus *Yarrowia* [375]. The species name '*lipolytica*' is derived from its ability to hydrolyze lipids [258]. *Y. lipolytica* is strictly aerobic and is typically found in habitats with high lipid and protein content. In addition, isolates were found on foods such as cheese, yogurt, and sausages. Here they are associated with the formation of typical aroma and fragrance compounds, which are produced by the secretion of lipases and esterases [405]. However, isolates have been collected from habitats with high salt concentration like the Dead Sea [50], as well as wastewater and oil-contaminated media [22]. The isolates from different habitats highlight the capabilities of the yeast to grow on a wide variety of hydrophilic and hydrophobic substrates such as sugars, lignocellulose, fatty acids, fats, oils, glycerol, n-paraffins, and other favorable renewable carbon sources [235][237]. As *S. cerevisiae*, wild type strains are unable to utilize pentoses [294]. But genome analyses by Ryu *et al.* revealed a cryptic xylose utilization pathway that could be activated by directed evolution [308].
The wild type strains differ in their morphology depending on growth conditions, media composition and

genetic background. Colonies can be heavily convoluted and matt or smooth and glistening [22][258]. *Y. lipolytica* is a dimorphic yeast, i.e., it can adopt different cell forms under different conditions: cell, pseudohyphae and hyphae [237][375]. Transition is often triggered by environmental conditions such as carbon or nitrogen sources, temperature, pH, or oxygen supply [193]. Some wild type strains accumulate lipids under nitrogen-limited conditions – typically 20 to 40 % of CDW [213]. Here, lipid accumulation is thought to be a type of stress response [253]. Due to oil accumulation, dimorphic transition, and secretion, *Y. lipolytica* is proposed as a model organism [237].

With the increasing industrial relevance came efforts to collect OMICS data from *Y. lipolytica*. Since 2004, the genome of CLIB99 is available [85]. The sequence of other isolates was added later [80][79]. The strains W9 and CLIB122 were sequenced and manually annotated [214]. *Y. lipolytica* has only six chromosomes and no 2-micron-like plasmids were identified. The CEN/ARS sequences found allow for extrachromosomal replication of DNA, but only limited suitability for heterologous expression due to the high loss frequency of the plasmids [227][235]. Here, integration into the genome turns out to be challenging because the yeast prefers nonhomologous end joining (NHEJ) for double-strand repair, so integration into the genome is rather random [174][205]. It was shown that deletion of *Ku70* and/or *Ku80* leads to a ratio change towards HR. This can be achieved by targeted knockout or by temporary knockdown via CRISPRi [205][318]. A significant increase in integration efficiency was achieved by transplanting *RAD52* from *S. cerevisiae*, where it is a core component of the HR machinery. Using homology sites of 1000 bp, this increased the targeting of a locus by 6.5-fold compared to wild type and 1.6-fold compared to *Ku70* deletion [174].

To further exploit the organism and make it accessible for metabolic engineering and SynBio, additional OMICS data were collected. For example, a transcriptome study was able to show that the transcript pool can be divided into four phases during fermentation: growth phase, transition phase, early and late lipid accumulation phase. Whereby, the last two phases are characterized by substantial flux changes [214][253]. Other multiomics studies focused on growth under nitrogen-limited conditions. For example, Pomraning *et al.* identified over 1200 new phosphorylation sites for regulation as well as an enrichment of TCA cycle intermediates and a downregulation of β-oxidation by metabolome, proteome, and phosphoproteome analysis [287]. Kerkhoven *et al.* performed metabolite and lipid analysis as well as RNAseq studies under nitrogen and carbon limited conditions and were unable to show direct transcriptional regulation of lipid metabolism but could show a shift away from amino acid synthesis towards lipid synthesis. The data collected were used to create a GEM [193]. Other models addressed the possibilities of metabolic engineering and used GEM to show ways to increase lipid production [191] or optimize the flux towards the synthesis of malonyl-CoA [393]. The lipid production capabilities of *Y. lipolytica* are also an important characteristic for its industrial utilization. The unique acetyl-CoA/malonyl-CoA supply is an important feature for producing high-quality secondary metabolites from low-value carbons [235]. In addition, *Y. lipolytica* has the ability for post-translational modification of proteins and natural secretion mechanisms, which it uses natively for the secretion of lipases and proteases [86][237]. Due to the already named ability to enrich oils and fats, *Y. lipolytica* is used for the production of oleochemicals and drop-in biofuels [235]. Thus, the palette of lipid products ranges from free fatty acids over fatty alcohols to polyunsaturated fatty acids [235]. For example, the yeast was enabled to produce omega-3 eicosapentaenoic acid by metabolic engineering [391]. DuPont uses *Y. lipolytica* for industrial production of docosahexaenoic acid [74]. Other product classes include polyols, terpenes, polyketides, and surfactants [235][405]. The widespread utilization for industrial applications is also reflected in the increasing number of publications. Fig. 1.3 shows the PubMed statistics for the respective organisms. In absolute numbers, *Y. lipolytica* is still underrepresented, but has increased in recent years, as the normalized graph shows.

The possibilities for metabolic engineering and for the construction of SynBio designs have been considerably expanded recently. Both, dominant (phleomycin, hygromycin, and nourseothricin) as well as

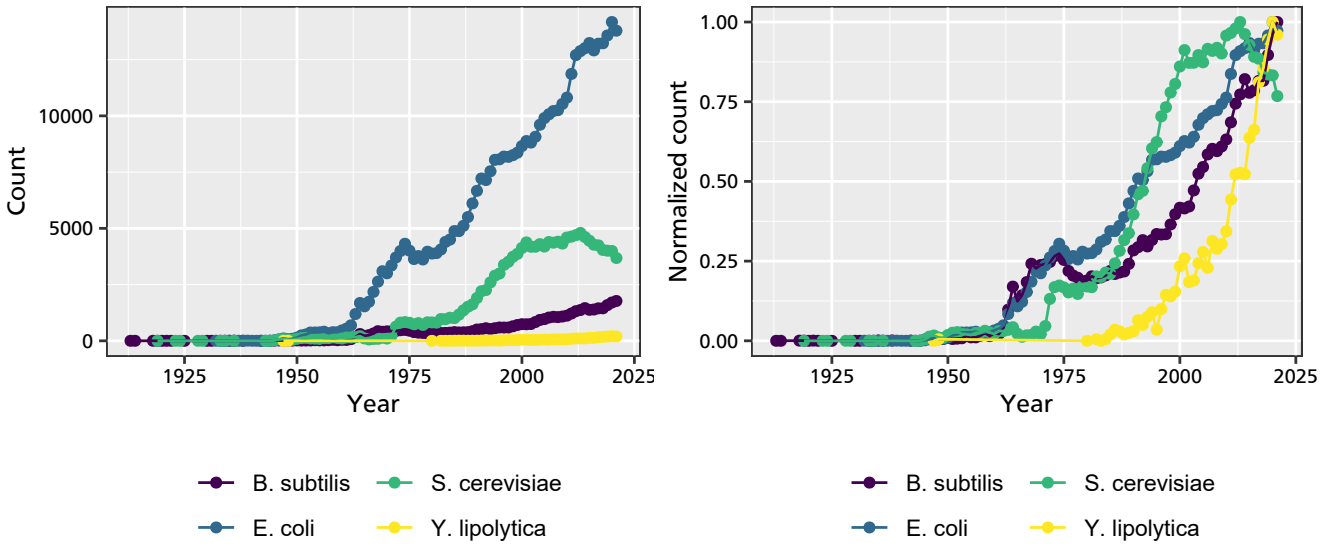


Figure 1.3.: Emerging organism *Y. lipolytica* in comparison to established organisms in SynBio. PubMed.gov stats downloaded on 24 February 2022 which contain counts until 2021. Normalized count was calculated as $Normalizedcount = (count - min(count))/(max(count) - min(count))$.

auxotrophic markers (*LEU2* and *URA3*) are available for genetic modification [235]. Recently, the *AMD1* marker was added, which enables Δ*AMD1* strains to grow again in defined medium by using acetamide as the sole nitrogen sources. A counter selection option with fluoroacetamide is also available for this marker [140]. Constitutive (P_{TEF1}, P_{FBA1}, P_{GPD}), inducible (P_{POX2}, P_{LIP2}, P_{EYK1}) and various hybrid promoters (P_{hp4d}) are available for construction [86][235][360][392]. Additional hybrid promoters resulted from the combination of $UAS1_{XPR2}$ elements with the P_{LEU2} core promoter [235]. A 7-fold enhancement of the P_{TEF1} activity was achieved by combining it with $UAS1_{TEF1}$ elements [28]. The palette was expanded by Curran *et al.* with synthetic promoters which can be used in both *S. cerevisiae* and *Y. lipolytica* [72]. The number of characterized terminators is rather limited with T_{XPR2} and T_{LIP2} [235]. However, the T_{CYC1} of *S. cerevisiae* can be used as well [319]. The palette for targeted gene activation or repression is expanding with the use of CRISPRi [318] and CRISPRa [319]. Nevertheless, the possibilities for targeted expression and load-balancing of metabolic pathways and circuits remains rather limited. But, Guo *et al.* succeeded in creating a synthetic chromosome that enabled *Y. lipolytica* to utilize both xylose and cellobiose [135].
Although the possibilities for modifications of *Y. lipolytica* are rather limited compared to *S. cerevisiae*, the range of tools has recently expanded noticeably. Since the yeast seems to become a model organism – as mentioned above – and as there is a continuing interest to further develop the yeast, this work aims to contribute to the expansion of the part portfolio.

1.4. General research objective

The upcoming transformations towards a circular economy and the associated reduction of resource utilization, as well as the reduction of greenhouse gas emissions, must be accelerated significantly. The widespread use of crude oil products, both as fuel and as a raw material in the petrochemical industry, must therefore be further reduced. The use of biotechnological processes can assist to replace energy-intensive

processes in chemical industry [98].
Bio-catalysts are frequently used for this purpose, whereby these can consist of individual enzymes, cell lysates or whole-cell systems. Bio-catalysts can be used to replace existing chemical processes or to introduce entirely new products, like new compounds from natural product research. Natural product research uses metagenome studies and multi omics techniques to discover entirely new substances and metabolic pathways to convert them into new products. Metabolic pathways are often expressed heterologously for the production of existing but also for the production of new substances. In some cases, individual metabolic pathway components are not balanced to each other, leading to the accumulation of intermediates or having a negative impact on the metabolism [105][192]. The methods of metabolic engineering and SynBio offer versatile possibilities to optimize the yield and quality of the products. This can be achieved by load balancing, genome reduction, and a variety of other techniques.
For the organisms *B. subtilis*, *E. coli* and *S. cerevisiae*, many well-characterized parts are available which can be used for a wide variety of designs. The industrially relevant oleaginous yeast *Y. lipolytica* is still poorly exploited for SynBio applications. Recently published methods to increase the efficiency of genome modifications and the development of Cas9/dCas9 methods clearly increase the amenability of the organism [174][205][318].
The pool of usable parts for *Y. lipolytica* is still limited. This thesis aims to contribute to add additional parts to the pool. For this purpose, a multi-organism system shall be developed to test the function of designs in different organisms (*E. coli*, *S. cerevisiae*, *Y. lipolytica*) simultaneously. This system will be used to test parts from *S. cerevisiae* in *Y. lipolytica* and to characterize their functionality in the new host. In addition, a CRISPR-Cas9 based system will be developed to allow for straightforward load balancing in *B. subtilis*, *E. coli*, *S. cerevisiae* and *Y. lipolytica* while providing a standardized DNA context for heterologous expression. To support the work on the four organisms, automated methods should be developed to support both construct and strain creation, but also characterization.
The objectives, which are intentionally kept short here, are intended to set the context of the work and will be described again in more detail in the respective chapters FLEXpress and Automation.

2. Materials and Methods

2.1. Technical equipment, chemicals and consumables

2.1.1. Technical Equipment

Table 2.1.: Technical equipment

Equipment	Manufacturer	Model
Autoclave	Systec GmbH (Linden, Germany)	VX-75
Camera	Canon Inc. (Tokyo, Japan)	EOS 350D
	Sony (Tokyo, Japan)	DSC-QX10
Camera objectives	Canon Inc. (Tokyo, Japan)	EFS 18-55mm
	SIGMA Corporation (Kanagawa, Japan)	DC 18-125 mm
Centrifuges	B. Braun Biotech Int. GmbH (Melsungen, Germany)	6K15
	Biozym Scientific GmbH (Hessisch Oldendorf, Germany)	Sprout-Minizentrifuge
	Sigma Laborzentrifugen GmbH (Osterode,Germany)	1K1S
	VWR (Darmstadt, Germany)	Micro Star 17 Micro Star 17R Mega Star 3.0R
Centrifuge Rotors	Thermo Fisher Scientific Inc. (Waltham, MA, USA)	#3057 #6445
Electroporation chamber	Bio-Rad Laboratories Inc. (Hercules, CA, USA)	MicroPulser
FACS	Sony Biotechnology Inc. (San Jose, USA)	SH800S
Gel electrophoresis	febikon Labortechnik GmbH (Wermelskirchen, Germany)	PB-0
Heat Incubator	Eppendorf AG (Hamburg, Germany)	Thermomixer comfort
	Thermo Fisher Scientific Inc. (Waltham, MA, USA)	Drybath
Incubator	Infors HT (Bottmingen, Schweiz)	Minitron Multitron II
	New Brunswick™ (Upland, California, USA)	Innova™ 44
	VWR (Darmstadt, Germany)	Incu-Line Tower
Laboratory scale	Sartorius AG (Goettingen, Germany)	Quintix®

Continued on next page

Table 2.1 – continued from previous page

Equipment	Manufacturer	Model
	Shimadzu (Gießen, Germany)	AUW120D
Magnetic stirrer (heatable)	IKA-Werke GmbH & Co. KG (Staufen, Germany)	IKAMAG RCT basic
Microplate reader	BMG Labtech (Ortenberg, Germany)	CLARIOstar® Plus
	BMG Labtech (Ortenberg, Germany)	PHERAstar® FSX
Microscope	Carl Zeiss Jena GmbH (Jena, Germany)	AXIO VertA.1
pH-meter	HANNA Instruments Germany GmbH (Kehl am Rhein, Germany)	HI 2211 pH/ORP Meter
PCR cycler	Bio-Rad Laboratories Inc. (Hercules, CA, USA)	Tetrad2
Pipettes	Eppendorf AG (Hamburg, Germany)	Research® plus
Power supply for electrophoresis	Amersham Biosciences Europe GmbH (Freiburg, Germany)	EPS 301
Robot platforms	Eppendorf AG (Hamburg, Germany)	epMotion 5075
	Analytik Jena GmbH (Jena, Germany)	CompuGene Robotics Platform (Fig. 4.1)
Safety workbench	Thermo Fisher Scientific Inc. (Waltham, MA, USA)	HeraSafe KS15
UV/VIS-photometer	Amersham Biosciences Europe GmbH (Freiburg, Germany)	Ultrospec 10
	Dr. Bruno Lange GmbH & Co. KG (Duesseldorf, Germany)	CADAS 50S
	Mettler-Toledo GmbH (Gießen, Germany)	UV5 Nano
UV-table	CAMAG (Berlin, Germany)	Reprostar III
Vortex	Heidolph Instruments GmbH & Co. KG (Schwabach, Germany)	REAX top
Water purification	Thermo Fisher Scientific Inc. (Waltham, MA, USA)	Barnstead™ GenPure™ Pro UV/UF
CompuGene Robotics Platform by Analytik Jena GmbH		
Cytometer	Beckman-Coulter (Brea, USA))	CytoFlex S
Incubator	Thermo Fisher Scientific Inc. (Waltham, MA, USA)	Cytomat™ 2
Liquid handler	Analytik Jena GmbH (Jena, Germany)	CyBio FeliX with SELECT Head
		CyBio FeliX with R96
	Dispendix GmbH (Stuttgart, Germany)	I-DOT one
Microplate reader	BMG Labtech (Ortenberg, Germany)	PHERAstar® FSX
		+ Filter: FI 485/520
		+ Filter: FI 575/620
		+ Filter: FP 485/520/520

Continued on next page

Table 2.1 – continued from previous page

Equipment	Manufacturer	Model
Robotic arm (gripper)	PreciseAutomation (Livermore, USA)	PreciseFlex PF400

2.1.2. Chemicals and Kits

Table 2.2.: Chemicals

Substance	Chemical formula	Manufacturer
1,4 - Dithiothreitol (DTT)	$C_4H_{10}O_2S_2$	Carl Roth GmbH + Co. KG
5-Fluoroorotic acid (FOA)	$C_5H_3FN_2O_4$	Carl Roth GmbH + Co. KG
Acetic acid 100 %	CH_3COOH	Carl Roth GmbH + Co. KG
Agar-Agar, Kobe I	-	Carl Roth GmbH + Co. KG
Agarose Standard for DNA/RNA electrophoresis	-	Carl Roth GmbH + Co. KG
Ampicillin sodium salt	$C_{16}H_{18}N_3NaO_4S$	Carl Roth GmbH + Co. KG
Ammoniumchloride	NH_4Cl	Carl Roth GmbH + Co. KG
Biotin	$C_{10}H_{16}N_2O_3S$	Sigma-Aldrich Co. LLC.
Boric acid	H_3BO_3	Carl Roth GmbH + Co. KG
Caseinhydrolysate, standard	-	Carl Roth GmbH + Co. KG
Calcium chloride dihydrat	$CaCl_2 \cdot 2\,H_2O$	Carl Roth GmbH + Co. KG
Calcium chloride hexahydrate	$CaCl_2 \cdot 6\,H_2O$	Carl Roth GmbH + Co. KG
CelLytic™ B Cell Lysis Reagent	-	Sigma-Aldrich Co. LLC.
Chloramphenicol	$C_{11}H_{12}Cl_2N_2O_5$	Carl Roth GmbH + Co. KG
Cobalt(II) chloride hexahydrate	$CoCl_2 \cdot 6\,H_2O$	Carl Roth GmbH + Co. KG
Copper(II) chloride dihydrate	$CuCl_2 \cdot 2\,H_2O$	Carl Roth GmbH + Co. KG
D(-)-Fructose	$C_6H_{12}O_6$	Carl Roth GmbH + Co. KG
D(+)-Glucose	$C_6H_{12}O_6$	Sigma-Aldrich Co. LLC.
D(+)-Xylose	$C_5H_{10}O_5$	Carl Roth GmbH + Co. KG
Di-sodiumhydrogenphosphate dihydrate	$Na_2HPO_4 \cdot 2\,H_2O$	Carl Roth GmbH + Co. KG
DMSO	C_2H_6OS	Carl Roth GmbH + Co. KG
Erythromycin	$C_{37}H_{67}NO_{13}$	Carl Roth GmbH + Co. KG
Ethylenediaminetetraacetic acid disodium salt dihydrate (Na_2EDTA)	$C_{10}H_{14}N_2Na_2O_8 \cdot 2\,H_2O$	Sigma-Aldrich Co. LLC.
G418 disulfate (Geneticin)	$C_{20}H_{40}N_4O_{10} \cdot 2\,H_2SO_4$	Sigma-Aldrich Co. LLC.
Glycerol	$C_3H_8O_3$	Carl Roth GmbH + Co. KG
Hygromycin B	$C_{20}H_{37}N_3O_{13}$	Carl Roth GmbH + Co. KG
Iron(III) chloride hexahydrate	$FeCl_3 \cdot 6\,H_2O$	Carl Roth GmbH + Co. KG
Kanamycin sulphate	$C_{18}H_{36}N_4O_{11} \cdot H_2SO_4$	Carl Roth GmbH + Co. KG
L(+)-Arabinose	$C_5H_{10}O_5$	Carl Roth GmbH + Co. KG
L-Histidine hydrochloride monohydrate	$C_6H_9N_3O_2 \cdot HCl \cdot H_2O$	Carl Roth GmbH + Co. KG

Continued on next page

Table 2.2 – continued from previous page

Substance	Chemical formula	Manufacturer
L-Leucine	$C_6H_3NO_2$	Sigma-Aldrich Co. LLC.
L-Tryptophane	$C_{11}H_{12}N_2O_2$	Carl Roth GmbH + Co. KG
Lincomycin hydrochlorid	$C_{18}H_{34}N_2O_6S \cdot HCl$	Sigma-Aldrich Co. LLC.
LB broth (10 g/L NaCl)	-	Carl Roth GmbH + Co. KG
Lithium acetate dihydrate	$C_2H_3LiO_2 \cdot 2\,H_2O$	Carl Roth GmbH + Co. KG
Magnesium sulphate heptahydrat	$MgSO_4 \cdot 7\,H_2O$	Carl Roth GmbH + Co. KG
Magnesium sulphate hydrate	$MgSO_4 \cdot H_2O$	Carl Roth GmbH + Co. KG
Manganese(II) chloride tetrahydrate	$MnCl_2 \cdot 4\,H_2O$	Carl Roth GmbH + Co. KG
Polyethylenglykol 3350	$H(OCH_2CH_2)nOH$	Sigma-Aldrich Co. LLC.
Potassium dihydrogenphosphat	KH_2PO_4	Carl Roth GmbH + Co. KG
Serva DNA Stain Clear G	-	SERVA Electrophoresis GmbH
Sodium chloride	NaCl	Sigma-Aldrich Co. LLC.
Sodium hydroxide	NaOH	Sigma-Aldrich Co. LLC.
Spectinomycin dihydrochlorid pentahydrat	$C_{14}H_{24}N_2O_7 \cdot 2\,HCl \cdot 5\,H_2O$	Sigma-Aldrich Co. LLC.
Streptomycin sulphate	$C_{42}H_{84}N_{14}O_{36}S_3$	Carl Roth GmbH + Co. KG
Tetracycline hydrochloride	$c_{22}h_{25}ClN_2O_6$	Carl Roth GmbH + Co. KG
Thiamin	$C_{12}H_{17}ClN_4OS$	abcr GmbH Deutschland
TRIS hydrochloride	$C_4H_{11}NO_3 \cdot HCl$	Carl Roth GmbH + Co. KG
Trizma®-Base (TRIS Base)	$C_4H_{11}NO_3$	Sigma-Aldrich Co. LLC.
Tryptone/Pepteon from Caseine	-	Carl Roth GmbH + Co. KG
Uracil	$C_4H_4N_2O_2$	AppliChem GmbH
Yeastextract	-	Carl Roth GmbH + Co. KG
Yeast Synthetic Drop-out Medium Supplements	-	Sigma-Aldrich Co. LLC.
YNB w/o amino acids w ammonium sulfate	-	Sigma-Aldrich Co. LLC.
YPD-Media	-	Carl Roth GmbH + Co. KG
Zinc chloride	$ZnCl_2$	Carl Roth GmbH + Co. KG

2.1.3. Consumables

Table 2.3.: Consumables

Consumable	Manufacturer	Type (Order-No.)
Plastic consumables		
Cuvettes	SARSTEDT AG & Co. KG (Nümbrecht, Germany)	Semi-micro cuvette, PS (67.742)
	VWR (Darmstadt, Germany)	Electroporation cuvettes (732-1136)
Centrifuge tubes	SARSTEDT AG & Co. KG (Nümbrecht, Germany)	Reaction tube, 1.5 ml (72.690.001)

Continued on next page

Table 2.3 – continued from previous page

Consumable	Manufacturer	Type (Order-No.)
		Reaction tube, 2 ml (72.691)
		SafeSeal reaction tube, 2 ml, brown (72.695.001)
		Screw cap tube, 15 ml (62.554.502)
		Screw cap tube, 50 ml (62.547.254)
		Screw cap tube, stand, 50 ml (62.559.001)
Cryo tubes	SARSTEDT AG & Co. KG (Nümbrecht, Germany)	CryoPure tubes, 2 ml (72.379.004)
Filter	SARSTEDT AG & Co. KG (Nümbrecht, Germany)	Syringe filter, Filtropur S, 0.2 µm (83.1826.001)
		Syringe filter, Filtropur S, 0.45 µm (83.1826)
Microtiter plate (SBS format)	Greiner Bio-One International GmbH (Frickenhausen, Germany)	CELLCULTURE MICROPLATE, 96 WELL F-BOTTOM, LID WITH CONDENSATION RINGS (655180)
		FOURWELL PLATE, PS, LID (96077307)
		LID, HIGH PROFILE (9 MM) (656101)
		MICROPLATE, 96 WELL, F-BOTTOM (655101)
		MICROPLATE, 96 WELL, V-BOTTOM (651101)
		MULTIWELL PLATE, 24 WELL (662102)
		ONEWELL PLATE, LID (670190)
Microtiter plate	SARSTEDT AG & Co. KG (Nümbrecht, Germany)	PCR plate without skirt, 96 well (72.985)
		Lid for PCR plate (65.989.002)
Petri dish	SARSTEDT AG & Co. KG (Nümbrecht, Germany)	Petri dish, 92 x 16 mm (82.1473)
Pipette tips	Eppendorf SE (Hamburg, Germany)	Pipette tip, 5000 µl (0030000650)
		Pipette tip, 10000 µl (0030000765)
	SARSTEDT AG & Co. KG (Nümbrecht, Germany)	Pipette tip, 10 µl (70.1130)
		Pipette tip, 20 µl (70.3020)
		Pipette tip, 200 µl, yellow (70.760.012)
		Pipette tip, 1000 µl, blue (70.3050.020)
Pipette tips (Liquid handling robots)	Analytik Jena GmbH (Jena, Germany)	RoboTipTray (250 µL) for R96 Head (OL3810-25-664)
		TipBox (250 µL) for SELECT Head (OL3811-25-637-S)
	Eppendorf SE (Hamburg, Germany)	epT.I.P.S.®, 50 µl (0030014405)
		epT.I.P.S.®, 300 µl (0030014448)
		epT.I.P.S.®, 1000 µl (0030014480)

Continued on next page

Table 2.3 – continued from previous page

Consumable	Manufacturer	Type (Order-No.)
Kits		
DNA clean-up	Roche Molecular Systems Inc. (Mannheim, Germany)	High Pure PCR Template Preparation Kit (11796828001)
Gel clean-up	New England Biolabs GmbH (Frankfurt am Main, Germany)	Monarch DNA Gel Extraction Kit (T1020L)
PCR clean-up	Analytik Jena GmbH (Jena, Germany)	innuPREP PCRpure Kit (845-KS-5010250)
	MACHEREY-NAGEL GmbH & Co. KG (Düren, Germany)	NucleoSpin Gel and PCR Clean-up Kit (740609.250)
	New England Biolabs GmbH (Frankfurt am Main, Germany)	Monarch PCR & DNA Cleanup Kit (T1030L)
Plasmid clean-up	Analytik Jena GmbH (Jena, Germany)	innuPREP Plasmid Mini Kit 2.0 (845-KS-5041250)
	New England Biolabs GmbH (Frankfurt am Main, Germany)	Monarch Plasmid Miniprep Kit (T1010L)
other		
Sheath fluid	Sony Biotechnology Inc. (San Jose, USA)	10x ClearSort™ Sheath Fluid
ssDNA	Thermo Fisher Scientific Inc. (Waltham, MA, USA)	UltraPure™ Herring Sperm DNA Solution

Table 2.4.: Polymerases and restriction endonuclease. Unless otherwise stated, enzymes were used according to manufacturer protocols.

Enzyme	Manufacturer
Polymerases	
Polymerase X	Roboklon GmbH (Berlin, Germany)
Q5® High-Fidelity DNA Polymerase	New England BioLabs GmbH (Frankfurt am Main, Germany)
One*Taq*®	New England BioLabs GmbH (Frankfurt am Main, Germany)
Opti*Taq* DNA Polymerase	Roboklon GmbH (Berlin, Germany)
Taq DNA Polymerase	Roboklon GmbH (Berlin, Germany)
Buffers for polymerases	
5x One*Taq* Standard Reaction Buffer	New England BioLabs GmbH (Frankfurt am Main, Germany)
10x Buffer C	Roboklon GmbH (Berlin, Germany)

Continued on next page

Table 2.4 – continued from previous page

Enzyme	Manufacturer
10x Hybrid Buffer	Roboklon GmbH (Berlin, Germany)
Restriction endonuclease	
*ApaL*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*BamH*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*BsaH*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Dpn*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*EcoR*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Hind*III	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Mlu*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Nde*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Nhe*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Pme*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Pst*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Sca*I-HF	New England BioLabs GmbH (Frankfurt am Main, Germany)
*Xba*I	New England BioLabs GmbH (Frankfurt am Main, Germany)
Buffers and loading dyes	
1X T4 DNA Ligase Reaction Buffer	New England BioLabs GmbH (Frankfurt am Main, Germany)
6x Gel Loading Dye, Purple	New England BioLabs GmbH (Frankfurt am Main, Germany)
10x CutSmart®	New England BioLabs GmbH (Frankfurt am Main, Germany)
10x 1.1 Buffer®	New England BioLabs GmbH (Frankfurt am Main, Germany)
10x 2.1 Buffer®	New England BioLabs GmbH (Frankfurt am Main, Germany)
10x 3.1 Buffer®	New England BioLabs GmbH (Frankfurt am Main, Germany)

2.1.4. Software

Table 2.5.: Software

Manufacturer	Software	Purpose
Date evaluation and processing		
Biomatters Ltd.	Geneious® R10 (10.2.6) Geneious® Prime 2021 (2021.2) (trail)	genomic data processing sgRNA design
Inkscape Authors	Inkscape 1.1 (c68e22c387)	vector graphics
R Core team	R 4.0 (4.0.1)	data evaluation and plotting
Rstudio Inc.	Rstudio (free) 1.4.1717	IDE for R scripting

Continued on next page

Table 2.5 – continued from previous page

Manufacturer	Software	Purpose
TeXstudio Authors	TeXstudio 4.2.1	IDE for LaTeX
TeX live	texlive 2021 (full)	LaTeX distribution
CompuGene Robotics platform		
Analytik Jena	CyBio Composer (2.67.00.00)	IDE for robot
Beckman Coulter	CytExpert 2.2 (2.2.0.97)	Cytoflex S measurement software
BMG Labtech	MARS (3.32)	data evaluation and export
	PHERAstar software (5.41)	PHERAstar® measurement software
Microsoft	SQL server 2014 (12.0.2269)	data storage and processing
	SQL server Management Studio (12.0.2269.0)	IDE for SQL server
R Core team	R 3.5 (3.5.3)	data processing and control
Rstudio Inc.	Rstudio (free) 1.4.1717	IDE for R scripting

2.2. Strains and Plasmids

Table 2.6.: The following plasmids and strains were kindly provided by the given working groups, were constructed earlier or purchased from the given manufacturer. (p – plasmid; s – strain)

ID	Name	Origin
Plasmids		
p11023	pINA443	Gerold Barth; AG Barth (TU Dresden)
p11032	pFA6-kanMX4	Hans-Joachim Schüller; AG Schüller (Uni Greifswald)
p11055	spINA_P_{hp4d}-cre-T_{CYC1}	Silke Hackenschmidt; AG Kabisch (TU Darmstadt)
p13008	pCRISPRyl_Hyg_Ku70	Stefan Bruder; AG Kabisch (TU Darmstadt)
p14019	pJET-SSS	Florian Nadler; AG Kabisch (TU Darmstadt)
p26009	pTAR_LS92.1	Thomas Hofmeyer; AG Kabisch (Uni Greifswald)
p26025	pXK_PMT	Thomas Zoll (Masterthesis); AG Kabisch (Uni Greifswald)
p26033	pHAGT_INTB_V2	Thomas Zoll (Masterthesis); AG Kabisch (Uni Greifswald)
p26038	pXK_INTB_V2	Thomas Zoll (Masterthesis); AG Kabisch (Uni Greifswald)
p26039	pINA_PPP_LEU2	Thomas Zoll (Masterthesis); AG Kabisch (Uni Greifswald)
p26042	pPK421-Mono-3HA	Marc Vogel; AG Suess (TU Darmstadt)
p26043	pPK421-Dimer-3HA	Marc Vogel; AG Suess (TU Darmstadt)
p26043	pPK421-Trimer-3HA	Marc Vogel; AG Suess (TU Darmstadt)
p26112	pMK-RQ_Rank1	Invitrogen Inc.
p26113	pMK-RQ_Rank4	Invitrogen Inc.

Continued on next page

Table 2.6 – continued from previous page

ID	Name	Origin
p26126	pJZC518	Anja Hofmann; AG Kolmar (TU Darmstadt)
p26128	pJZC522	Anja Hofmann; AG Kolmar (TU Darmstadt)
p26129	pJZC588	Anja Hofmann; AG Kolmar (TU Darmstadt)
p26134	pFRP1642	Anja Hofmann; AG Kolmar (TU Darmstadt)
p26161	p426MET25-Envy	Leonie Baumann; AG Boles (Goethe-Universität, Frankfurt am Main)
p26193	pKD3_BR322_Rank1	Vanessa Munoz; AG Charpentier (MPUSP, Berlin)
p26194	pKD3_BR322_Rank4	Vanessa Munoz; AG Charpentier (MPUSP, Berlin)
p99010	pCD227_Cm	Addgene: #113318
p99011	pCD315_W108-scRNA.b1	Addgene: #113321
p99012	pKD4	Addgene: #45605
p99013	pCD227_Spec	Georg Schmidt; AG Kabisch (TU Darmstadt)
p99016	pCD315_Pro_4-scRNA.b1	Georg Schmidt; AG Kabisch (TU Darmstadt)
p99017	pKD4_sfGFP_noPromoter	Georg Schmidt; AG Kabisch (TU Darmstadt)
p99018	pKD4_BBa_J23117_sfGFP	Georg Schmidt; AG Kabisch (TU Darmstadt)
p99019	pKD46	Vanessa Munoz; AG Charpentier (MPUSP, Berlin)
Strains		
Comm.	*E. coli* GT115	InvivoGen Europe (www.invivogen.com)
Comm.	*E. coli* NEB® 10-beta	New England BioLabs GmbH (Frankfurt am Main, Germany)
Comm.	*E. coli* NEB® turbo	New England BioLabs GmbH (Frankfurt am Main, Germany)
s11041	*S. cerevisiae* BY4741	Hans-Joachim Schüller; AG Schüller (Uni Greifswald)
s11070	*Y. lipolytica* 63	Frieder Schauer; AG Schauer (Uni Greifswald)
s11084	*Y. lipolytica* H222	Silke Hackenschmidt; AG Kabisch (TU Darmstadt)
s11085	*Y. lipolytica* Po1f	Thomas Hofmeyer; AG Kolmar (TU Darmstadt)
s11135	*B. subtilis* PY79	Silke Hackenschmidt; AG Kabisch (TU Darmstadt)
s26419	*S. cerevisiae* CEN.PK2-1C	Eckhard Boles; AG Boles (Goethe-Universität, Frankfurt am Main)
s26435	*S. cerevisiae* BY4742	Eckhard Boles; AG Boles (Goethe-Universität, Frankfurt am Main)
s26436	*S. cerevisiae* S288C	Eckhard Boles; AG Boles (Goethe-Universität, Frankfurt am Main)
s55036	*Y. lipolytica* W29	Rodrigo Ledesma-Amaro; IC-CSynB Group (Imperial College, London)
s99001	*B. subtilis* PY79	Georg Schmidt; AG Kabisch (TU Darmstadt)
s99015	*E. coli* BW25113	Vanessa Munoz; AG Charpentier (MPUSP, Berlin)

Table 2.7.: Relevant genotypes of used strains

Strain	Designation	Genotype
s11041	*S. cerevisiae* BY4741	MATa *his3Δ0 leu2Δ0 met15Δ0 ura3Δ0*
s11084	*Y. lipolytica* H222	*ura3*-302::*SUC2*
s11085	*Y. lipolytica* Po1f	MATA *ura3*-302 *leu2-270 xpr2-322 axp2-ΔNU49 XPR2*::*SUC2*
s11135	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H*
s26419	*S. cerevisiae* CEN.PK2-1C	MATa *his3Δ1 leu2-3_112 ura3-52 trp1-289 MAL2-8c SUC2*
s26435	*S. cerevisiae* BY4742	MATa *his3Δ0 leu2Δ0 lys2Δ0 ura3Δ0*
s99001	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R)
s99003	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{ylb}-sfGFP-$Spec^R$)
s99004	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{43}-sfGFP-$Spec^R$)
s99005	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{ylb}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P21-Cm^R
s99006	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{ylb}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P2-Cm^R
s99007	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{ylb}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P4-Cm^R
s99008	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{ylb}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P11-Cm^R
s99009	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{ylb}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-empty-Cm^R
s99010	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{43}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P21-Cm^R
s99011	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{43}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P2-Cm^R
s99012	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{43}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P4-Cm^R
s99013	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{43}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-P11-Cm^R
s99014	*B. subtilis* PY79	Δ*sigF* Δ*skfA-H* Δ*ganA*::(P_{xyl}-dCas9-Er^R) Δ*pksX*::(Rank4-P_{43}-sfGFP-$Spec^R$) Δ*amyE*::P_{veg}-sgRNA-empty-Cm^R
s99015	*E. coli* BW25113	Δ(*araD-araB*)567 Δ*lacZ*4787(::*rrnB*-3) λ^- *rph*-1 Δ(*rhaD-rhaB*)568 *hsdR*514
s99017	*E. coli* GT115	*uidA*(Δ*MluI*)::*pir*-116

2.3. Buffers and Media

Table 2.8.: Buffer recipes

Buffer		Compound
L(+)-arabinose stock solution	1 M	L(+)-arabinose
Agarose gel for electrophoresis	4 g	Agarose
	400 mL	1x TAE
DTT (Dithiothreitol)	1 M	Dithiothreitol
FCC (frozen competent cell) solution	5 % (v/v)	Glycerol
	10 % (v/v)	DMSO
Lithium acetate	1 M	Lithium acetate
PEG 3350	50 % (w/v)	Polyethylenglycol 3350
TRIS-Acetat-EDTA (TAE) Buffer (50x)	2 M	TRIS-Acetat
	50 mM	Na_2EDTA
	1 M	Acetic acid
TE Buffer (10x); pH 8.0	100 mM	TRIS
	10 mM	Na_2EDTA

Table 2.9.: Culturing media and supplements. For preparation of agar plates, 1.5 g Agarose / 100 mL medium was added. Antibiotics were added after autoclaving and cooling to ~65 °C.

Culturing media		Compound
Culturing media		
5-Fluoroorotic acid (5-FOA) plates[1]	0.67 %	Yeast nitrogen base w/o aminoacids w ammonium sulfate
	0.2 %	SC–Ura drop-out mix
	2 %	D-glucose
	0.1 %	5-FOA
	50 µg/mL	Uracil
Acidic casein hydrolysate (ACH)	6.8 g/L	Yeast nitrogen base w/o aminoacids w ammonium sulfate
	14 g/L	Caseinhydrolysate
	10 g/L	D-glucose
Lysogeny broth (LB)	50 g/L	LB media (10 g/L NaCL)
M9 mineral medium (Helmholtz Munich, [117])	100 mL	M9 salt solution (10x)
	20 mL	20 % D-glucose
	1 mL	1 M $MgSO_4$
	0.3 mL	1 M $CaCl_2$

Continued on next page

[1] http://cshprotocols.cshlp.org/content/2016/6/pdb.rec086637.short (date of access: 03.04.2022)

Table 2.9 – continued from previous page

Culturing media		**Compound**
	1 mL	Biotin (1 mg/mL)
	1 mL	Thiamin (1 mg/mL)
	10 mL	Trace elements solution (100X)
	867 mL	sterile ddH_2O
MSM Media [257]	62.5 mL	4x MSM-base
	0.5 mL	500x Trace elements solution (TES)
	0.5 mL	1 M $MgSO_4$
	variable	Carbon source
	ad.	sterile ddH_2O
	250 mL	
Yeast extract pepton dextrose (YPD)	50 g/L	YPD
Yeast extract pepton dextrose (YPD) - 2x	100 g/L	YPD
Yeast nitrogen base (YNB)	6.7 g/L	Yeast nitrogen base w/o aminoacids w ammonium sulfate
	10 g/L	D-glucose
YT - 2x	16.0 g/L	Trypton
	10.0 g/L	Yeast extract
	5.0 g/L	Sodium chloride
M9 mineral medium - stock solutions; prepared in ddH_2O and sterilized by filtration		
M9 salt solution (10x), pH 7.2	75.2 g/L	$Na_2HPO_4 \cdot 2\,H_2O$
	30.0 g/L	KH_2PO_4
	5 g/L	NaCl
	5 g/L	NH_4Cl
20 % (w/v) D-glucose	200 g/L	D-glucose
1 M $MgSO_4$	24.65 g/100mL	$MgSO_4 \cdot 7\,H_2O$
1 M $CaCl_2$	14.70 g/100mL	$CaCl_2 \cdot 2\,H_2O$
Biotin	1 mg/mL	Biotin
Thiamin	1 mg/mL	Thiamin-HCl
0.1 M $CuCl_2$	1.70 g/100mL	$CuCl_2 \cdot 2\,H_2O$
0.2 M $CoCl_2$	4.76 g/100mL	$CoCl_2 \cdot 6\,H_2O$
0.1 M H_3BO_3	0.62 g/100mL	H_3BO_3
1 M $MnCl_2$	19.8 g/100mL	$MnCl_2 \cdot 4\,H_2O$
100X trace elements solution	5 g/L	Na_2EDTA
	0.83 g/L	$FeCl_3 \cdot 6\,H_2O$
	84 mg/L	$ZnCl_2$
	13 mg/L	$CuCl_2 \cdot 2\,H_2O$
	10 mg/L	$CoCl_2 \cdot 6\,H_2O$
	10 mg/L	H_3BO_3

Continued on next page

Table 2.9 – continued from previous page

Culturing media		Compound
	1.6 mg/L	$MnCl_2 \cdot 4\,H_2O$
MSM mineral medium - stock solutions; prepared in ddH_2O and sterilized by filtration		
MSM base (4x)	8 g/L	Na_2SO_4
	10.72 g/L	$(NH_4)_2SO_4$
	2 g/L	NH_4Cl
	4 g/L	$(NH_4)_2$-citrate
	58.4 g/L	K_2HPO_4
	16.28 g/L	$NaH_2PO_4 \cdot 2\,H_2O$
100X trace elements solution	0.5 g/L	$CaCl_2 \cdot 2\,H_2O$
	0.18 g/L	$ZnSO_4 \cdot 7\,H_2O$
	0.1 g/L	$MnSO_4 \cdot H_2O$
	20.1 g/L	Na_2EDTA
	16.7 g/L	$FeCl_3 \cdot 6\,H_2O$
	0.16 g/L	$CuSO_4 \cdot 5\,H_2O$
	0.18 g/L	$CoCl \cdot 6\,H_2O$

Table 2.10.: Amino acid stock solutions. Amino acids were solved in ddH_2O. If required, NaOH was added drop wise to solve substance completely. Amino acid stock solutions were sterilized by filtration (0.22 µm).

Amino acid stock		Compound
Histidin (100x)	5 mg/mL	Histidin
Leucin (100x)	10 mg/mL	Leucin
Tryptophan (100x)	10 mg/mL	Tryptophan

Table 2.11.: Antibiotic stock solutions. Antibiotics were solved in ddH_2O or Ethanol according to manufacturer. Aqueous solutions were sterilized by filtration (0.22 µm).

Antibiotic stock	Stock solution concentration	Final concentration
Bacteria		
Ampicillin	100 mg/mL	100 µg/mL
Chloramphenicol	34 mg/mL	34 µg/mL (Selection) 134 µg/mL (Plasmid amplification)
Erythromycin	1 mg/mL	50 µg/mL (*E. coli*) 1 µg/mL (*B. subtilis*)

Continued on next page

Table 2.11 – continued from previous page

Antibiotic stock	**Stock solution concentration**	**Final concentration**
Kanamycin	50 mg/mL	50 µg/mL
Lincomycin	25 mg/mL	25 µg/mL
Spectinomycin	100 mg/mL	100 µg/mL
Streptomycin	25 mg/mL	25 µg/mL
Tetracycline	10 mg/mL	10 µg/mL
Yeast		
Geneticin (G418)	200 mg/mL	200 µg/mL (*S. cerevisiae*, liquid) 400 µg/mL (*S. cerevisiae*, plate) 600 µg/mL (*Y. lipolytica*)
Hygromycin B	200 mg/mL	200 µg/mL (*S. cerevisiae*) 600 µg/mL (*Y. lipolytica*)

2.4. General methods

The following section contains the common methods for molecular biology and microbiology work that were performed as standard. Some methods include additional notes that provide guidance on the standard protocol. In the appendix, some methods, such as the preparation of individual constructs and strains, and robotic cultivations, are described in more detail. These methods and descriptions are then cross-referenced separately in the respective results sections.

2.4.1. Molecularbiological methods

Standard plasmid isolation from *E. coli* using clean-up kits

Kits from different manufacturers were used during the preparation time of the presented thesis. The following kits have been used: High Pure Plasmid Isolation Kit (Roche Diagnostics GmbH, Germany), Monarch® Plasmid Miniprep Kit (New England BioLabs GmbH, Germany) and innuPREP Plasmid Mini Kit 2.0 (Analytik Jena GmbH, Germany). All kits were used according to manufacturer protocol. Plasmid isolation was performed with cultures which have been grown over night (~16 h). The elution buffer of all kits was preheated to 70 °C as suggested by Roche to improve elution of DNA.

Polymerase chain reaction

DNA fragments for cloning purpose or sequencing have been amplified by proof-reading polymerases. Used polymerase were: NEB Q5®, NEB One*Taq*®(mixture of *Taq* and *Deep Vent* ®polymerase), Roboklon Opti*Taq* (mixture of *Taq* and *Pfu* polymerase) and Roboklon Polymerase X. Unless stated otherwise, a PCR was performed in a 50 µL reaction according to the manufacturer protocol. The components of a standard PCR reaction are listed in Tab. 2.15. Typical PCR cycling programs are given in Tab. 2.12, Tab. 2.13 and Tab. 2.14 . Temperature and elongation-time were adjusted, depending on the utilized oligonucleotide and expected fragment length. The T_m for the high fidelity polymerases of NEB and Roboklon were calculated using the T_m-calculators of NEB and Roboklon. The primer 3 version implemented in Geneious R10 was used to calculate the T_m for the other polymerases. To verify the PCR reaction, 3 µL were used for a agarose gel electrophoresis. If the fragments purpose was cloning, than a *Dpn*I digest was done by adding 1 µL of the restriction endonuclease to the PCR mixture. The digest was incubated for 90 min at 37 °C and 20 min at 80 °C for heat inactivation.

Table 2.12.: Cycling program of a standard PCR reaction with Polymerase X. X = T_m calculated

Step	**Temperature [°C]**	**Time [min]**	
Initial denaturation	93	2:00	
Denaturation	93	0:20	} 35x
Annealing	X	0:20	
Elongation	72	0:30 s/kb	
Final elongation	72	5:00	
Final hold	15	∞	

Table 2.13.: Cycling program of a standard PCR reaction with Q5® Polymerase. X = T_m calculated

Step	Temperature [°C]	Time [min]	
Initial denaturation	98	0:30	
Denaturation	98	0:10	35x
Annealing	X	0:10	
Elongation	72	0:30 s/kb	
Final elongation	72	2:00	
Final hold	15	∞	

Table 2.14.: Cycling program of a standard PCR reaction with NEB One*Taq*®, Roboklon Opti*Taq* or Roboklon *Taq*. X = T_m calculated

Step	Temperature [°C]	Time [min]	
Initial denaturation	95	10:00	
Denaturation	95	0:30	35x
Annealing	X	0:30	
Elongation	72	1:00 s/kb	
Final elongation	72	7:00	
Final hold	15	∞	

Table 2.15.: Typical PCR reaction.

Component	Volume	Final concentration
5x Buffer / 10x Buffer	10 µL / 5 µL	1x
dNTP mix (10 mM each)	1 µL	0.2 mM each
Upstream primer	Variable	0.2 mM - 0.5 mM
Downstream primer	Variable	0.2 mM - 0.5 mM
Template DNA	Variable	~0.1 µg
Polymerase	0.5 µL	1 U
Total volume	ad. 50 µL	-

Colony PCR

These method is a modified version of a PCR method by Güssow and Clarkson [137]. The method was used to determine the correct assembly of DNA fragments after performing a seamless ligation cloning

extract (SLiCE), TAR or transformation of an organism. The obtained single colonies served as PCR templates and were individually examined. For yeasts, the cells were re-suspended in 30 µL of a 20 mM NaOH solution by aid of toothpicks and cooked for 45 min at 98 °C. 2 µL of the clear supernatant were used as template. Bacterial cells were directly added to the PCR mix. The PCR was scaled to 15 µL. The program was changed towards a longer first denaturation step of 10 min in comparison to the standard program in Tab. 2.14.
Note: Colony PCR of yeasts may result in the formation of a white precipitate. This often leads to interference with the colony PCR. A short centrifugation leads to separation and the clear supernatant can be used for the reaction.

Clean-up of PCR products and restriction reaction mixes

The clean-up of DNA from PCR reactions and restriction reaction mixes was performed with different commercial clean-up kits: Macherey-Nagel NucleoSpin Gel and PCR-Clean-up-Kit, NEB Monarch® PCR & DNA Cleanup Kit and Analytik Jena innuPREP PCRpure Kit. The kits were used according to manufacturers protocol. Contrary to the manufacturer protocols, the elution buffer was pre-headed to 80 °C for at least 10 min. After adding the elution buffer, the columns were incubated for 1 min. This procedure was suggested by a note of Macherey-Nagel and was applied to all kits to improve elution of DNA.

Isolation of genomic DNA from bacteria and yeast

Genomic DNA was isolated with the Roche High Pure PCR Template Preparation Kit. The protocol was used according to the manufacturer protocol, except one step. The re-suspension buffer of the kit was replaced by the buffer P1 from QIAGEN® Plasmid Mini kit because it already contains RNase A. Standard elution volume was 100 µL with pre-heated elution buffer as suggested by a note of Macherey-Nagel.

Determination of quality and quantity of DNA samples

DNA concentration was measured by using the UV5nano photometer. Besides DNA concentration, the quality ratios 260/280 nm and 260/230 nm were determined to estimate the purity of the sample. For the measurement, 2 µL were loaded into the optical path of the device. To verify the NanoDrop results, a agarose gel electrophoresis was performed.

Agarose gel electrophoresis

Agarose gel electrophoresis was used to verify roughly quality and quantity of plasmids and PCR products. The agarose was solved in TAE buffer with a concentration of 1 % (w/v) by using a microwave (or appropriate concentrations for smaller or bigger DNA fragments). The agarose-solution was stored at 70 °C. Before pouring a gel, 0.5 µL Serva DNA Stain Clear G/10 mL was added to the liquid agarose. The Gel was incubated for about 15 min at RT for polymerisation. In the following step, the gel was put into a electrophoresis gel chamber, which was filled with TAE buffer. Finally, the agarose gel was loaded with the samples, which was already mixed with 10x Buffer C or Purple Loading Dye. Electrophoresis was performed for about 25 min at 100 V and 400 mA. Documentation was performed by using an UV-light table and a digital camera.

Sequencing of PCR products and plasmids

The purified DNA was sequenced at Eurofins MWG GmbH (Ebersberg, Germany) or Microsynth Seqlab GmbH (Göttingen, Germany). For this purpose, DNA solution was mixed with the appropriate oligo

nucleotide and shipped to the company. The preparation was carried out according to the manual of the respective company. Within two days, the sequencing results were available online and were evaluated using Geneious R10.

Digest DNA with restriction endonuclease

Restriction enzymes were used according to the protocol of the manufacturer. A typical digest was set to a final volume of 25 µL. Usually, 1 µg of DNA was used. The reaction mixture was incubated for 60 min at 37 °C and 20 min at 80 °C. The success was determined by agarose gel eletrophoresis and comparison with the undigested DNA.

2.4.2. Microbiological methods

Strain maintenance

Strains were stored on agar plates at 4 °C. For long term storage at −80 °C, an overnight culture was grown in appropriate media and supplements. On the next day, 800 µL were taken and transferred into a cryo tube which was filled with 200 µL of 100 % glycerol, resulting in 20 % (v/v) glycerol and stored at −80 °C.

Overnight culture

Overnight cultures were inoculated either from an agarose plate or cryo culture. Usually, a volume of 4 mL was used, which was inoculated with a sterile pipette tip or an inoculation loop. The media contained appropriate supplements such as antibiotics, nucleotides or amino acids to fulfill auxotrophy needs or to keep selection pressure. Incubation was performed over night (13-16 h) at either 30 °C and 200 rpm for yeasts or 37 °C and 180 rpm for bacteria.

Preparation of chemocompetent *S. cerevisiae*

The used protocol of Gietz and Schiestl [121] was slightly adapted by the Boles Group (Goethe-Universität, Frankfurt am Main). The desired strain was inoculated in 20 mL of 2x YPD and incubated over night (15-16 h) at 30 °C and 200 rpm. The main culture of 100 mL 2x YPD in 500 mL Erlenmeyer flasks were inoculated four times with 1 mL of the pre-culture. The cultures were grown at 30 °C and 120 rpm until they reached a OD_{600} (OD) of 0.5-0.7. Harvesting was done by centrifugation in 50 mL reaction tubes at 3000xg for 2 min. After removal of the supernatant, the cells were re-suspended in ddH_2O (room temperature (RT), 0.5x volume) and centrifuged again. The supernatant was removed and the pellet re-suspended in 0.01x volume FCC solution. Aliquots of 50 µL were distribute in 1.5 mL reaction tubes and slowly frozen using a NALGENE™ Cryo 1 C Freezing Container.

Chemical transformation of *S. cerevisiae*

The transformation mix was prepared while thawing the frozen cells on ice. Every time, the transformation mix was made freshly by mixing 260 µL PEG3350 (50 %), 36 µL Lithium acetate (1 M) and 10 µL ssDNA (10 mg/mL Hering sperm DNA). The ssDNA was boiled for 10 min at 95 °C and stored directly on ice until used. To remove residual FCC solution, the cells were centrifuged at 3000xg and RT for 30 s. The supernatant was removed and the pellet re-suspended in 306 µL transformation mix. Different amounts of DNA were added to the suspension depending on the DNA-type (Plasmid: 25 fmol, PCR-fragment: 250 fmol). The calculated amount of DNA was mixed with ddH_2O with a final volume of 54 µL, added to

the cell-suspension and vortexed again for 45 s. The cells were incubated at 42 °C for 30 min. For plating, the mix was centrifuged at 3000xg and RT for 30 s and the supernatant was removed. If an auxotrophic marker was used, the pellet was re-suspended in ddH_2O and plated subsequently on appropriate plates. In case of an antibiotic selection marker, the cells were regenerated in YPD for 2 h at 30 °C before plating. The plates were incubated at 30 °C until colonies were visible.

TAR cloning in *S. cerevisiae*

The assembly of large plasmids was performed by simultaneous transfer of DNA fragments into *S. cerevisiae* as described by Kuijpers *et al.*[207]. Fragments were amplified by PCR. Primer extensions were used to provide homologous overhangs of 35 bp to the neighboring fragment. A common problem is plasmid carry over from the PCR reactions. For this reason, the backbone was split into two parts within the marker gene. The DNA mixture was prepared by adding 100 fmol of both backbone fragments and the inserts were added in 5 to 10-fold excess. The assembly is performed via HR by the DNA repair machinery of *S. cerevisiae*. The transformation was performed as described above using *S. cerevisiae* BY4741 cells. The colonies obtained were evaluated by colony PCR. The plasmid of the putative positive single clones were isolated and used for a transformation of *E. coli* to amplify the plasmids. Final characterization was carried out by further specific PCRs and sequencing.

Preparation of electrocompetent *E. coli* cells

The preculture was prepared by adding 50 µL of an 'inoculation aliquot' of NEB®10-beta to 30 mL LB media with appropriate antibiotic. The culture was grown for 16 h at 180 rpm and 37 °C. Two 2 L Erlenmeyer flasks were filled with 1 L LB media and pre-warmed for 1 h in a shaking incubator at 37 °C. The main-culture was inoculated by adding 10 mL of the preculture to each of the pre-warmed 2 L Erlenmeyer flasks. Incubation took place at 180 rpm and 37 °C. The cells were harvested at an OD between 0.6 and 0.8. The following steps were performed on ice or a centrifuge at 4 °C. The cell suspension was transferred into centrifuge cups and subsequently cooled on ice for 30 min. Afterwards, the suspension was centrifuged at 3488xg for 15 min. The supernatant was removed and the pellet re-suspended in 1.0x volume 15 % (w/v) glycerol of the main-culture and centrifuged again at 3488xg. The washing steps were repeated with reduction of the 15 % (w/v) glycerol volume: 0.5x and 0.1x of the initial volume. Afterwards, the cells were gently re-suspended by adding 750 µL per 0.1 initial OD unit. Finally, the cell suspension was aliquoted by transfering 50 µL into sterile 1.5 mL reaction tubes and subsequent freezing in liquid nitrogen. Tubes were stored in a −80 °C freezer.

Transformation of *E. coli* by electroporation

The electrocompetent *E. coli* cells were incubated with 0.5 µL of plasmid DNA (∼100 µL) or up to 5 µL of SLiCE reaction for 5 min on ice . In the next step, the cell/DNA suspension was transferred into an ice-cold electroporation cuvette and placed in an eletroporation chamber. Electroporation was done with preset program Ec2 of the MicroPulse Device (2.5 kV, ∼5 ms). Immediately after pulsing, 1 mL of LB media was added and the suspension was transferred into a new 1.5 mL reaction tube. After 1 h of regeneration at 37 °C, the cells were streaked out on selective LB agar plates containing the appropriate antibiotic. The agar plates were finally incubated over night at 37 °C until colonies were visible.

Chemical transformation of *E. coli*

The protocol was slightly modified from NEB[2]. From cryo stock, an overnight culture was set in LB media and incubated over night at 37 °C and 200 rpm. On next morning, the main culture was set by inoculating 50 mL with 0.5 mL of the preculture in an 250 mL Erlenmeyer flask. The culture was grown until it reached an OD between 0.5 and 0.7. The cell suspension was centrifuged in 50 mL reaction tube at 3488xg and 4 °C for 5 min. The pellet was re-suspended in 5 mL ice-cold 30 mM $CaCl_2$. Subsequently, the supension was divided among three 1.5 mL reaction tubes and centrifuged again at 3500xg and 4 °C for 5 min. Each pellet was re-suspended in 0.5 mL ice-cold 30 mM $CaCl_2$ and distributed among 1.5 mL reaction tubes with 50 µL each. Afterwards, DNA was added and incubated for 30 min on ice. In the next step, a heat shock was carried out. This was done for 30 s in a 42 °C water bath. The tube was then immediately placed on ice for another 5 min. Contrary to the NEB protocol, 1 mL LB media was added and the cell mixture regenerated for 1 h. Finally, dilutions were plated on selective agar plates.

Preparation of Seamless Ligation Cloning Extract (SLiCE)

The SLiCE cloning method was developed by Zhang *et al.* [402] To get the SLiCE extract, *E. coli* cells were cultivated overnight in 20 mL 2x YT medium at 37 °C and 200 rpm. On the following day, the main culture of 90 mL 2x YT was inoculated using 3 mL of the overnight culture. The culture was grown until an OD of ∼3, followed by harvesting in two 50 mL centrifuge tubes and centrifuging at 3488xg and 4 °C for 10 min. The supernatant was discarded and the cell pellet re-suspended in 40 mL of ice-cold ddH_2O and centrifuged again. This washing step was repeated. Afterwards, the pellet was re-suspended in 20 mL of ice-cold ddH_2O, pooled and centrifuged again. The supernatant was fully discarded and the pellet weighted and subsequently frozen over night at −80 °C for better lysis. The pellet was thawed and 120 µL CelLytic™ B per 100 µg cell pellet was added. After 10 min of incubation at RT, the lysate was centrifuged (10 min, 17000xg, 4 °C) and the supernatant was transferred into a microfuge tubes, to get a defined volume of cell extract by measuring with pipette. Finally, one volume of 100 % glycerol was added, mixed and aliquoted in PCR tubes (∼10 µL). Finally, the extract was flash frozen in liquid nitrogen and stored in a −80 °C freezer.

Concatenation of DNA fragments by SLiCE

The method is based on HR, which allows the concatenation of DNA fragments [402]. The different DNA fragments have to contain overlapping homology sites of 35 bp, which can be added via primer extension. The reaction mixture contained the vector backbone and inserts in a molar ratio of 1:10 based on 100 fmol vector backbone, as well as 1.0 µL T4 DNA ligase buffer and 1.0 µL SLiCE extract. After 60 min incubation at 37 °C, one half of the reaction was used to transform electrocompetent *E. coli* NEB® 10-beta cells. The second half was stored at −20 °C.
Note: SLiCE extract and DNA fragments contain a lot of salts, which can short circuit electroporation. Either add 1x volume ddH_2O to cell-suspension or de-salt DNA fragments by dialysis.

Genomic integration into the *E. coli* genome

The used protocol was slightly adapted in comparison to the original protocol from Datsenko and Wanner [76]. The protocol was adapted by Shawn Douglas and can be found on OpenWetWare[3]. The transformation of helper plasmid pKD46 was carried out chemically as described above. For a genomic integration, four

[2] https://www.neb.com/protocols/2012/06/21/making-your-own-chemically-competent-cells (date of access: 06.11.2021)
[3] https://openwetware.org/wiki/Recombineering/Lambda_red-mediated_gene_replacement (date of access: 06.11.2021)

cultures were inoculated from single colonies in 2 mL LB-Amp media and incubated at 30 °C. Several controls were included here. Thus, the following cultures were set [Culture (DNA | Inducer)]: 1. culture (no DNA | ddH_2O), 2. culture (no DNA | 10 mM Ara), 3. culture (DNA | ddH_2O), **4. culture (DNA | 10 mM)**. When the bacteria reached an OD of 0.1, the respective inducer (ddH_2O); 10 mM (final) L(+)-arabinose) was added to activate the Lambda Red system. The incubation was continued until an OD of 0.4. The overnight culture tubes were incubated on ice for 10 min. Afterwards, 1 mL of each culture was transferred into 1.5 mL reaction tubes and centrifuged at 4000xg and 4 °C for 10 min. The supernatant was discarded and the pellet re-suspended in ice-cold ddH_2O and centrifuged again using the same conditions. Again, the supernatant was removed and the cells re-suspended in 50 µL ddH_2O. Subsequently, PCR amplified DNA or ddH_2O in equal amounts were added (5 pg up to 0.5 µg DNA). Finally, electroporation of the cells was carried out with 2 h of regeneration as described above. After plating, the bacteria were incubated at 37 °C. After confirmation of integration, the pKD46 helper plasmid can be removed by incubation at 37 °C and non-selective plates. Additional notes can be found in section A.3.6
Note: Formation of colonies can take noticeably longer than transformation with plasmid DNA.

Chemical transformation of *Y. lipolytica*

The strain to be transformed was spread on YPD plate. Usually, YPD plates were used but altered when selection was required. Plates were incubated over night at 30 °C. The highest transformation efficiency was achieved with yeasts not older than two days. 5 µL ssDNA (10 mg/mL Hering sperm DNA) was boiled at 96 °C for 5 min and put subsequently on ice. The ssDNA was mixed with ~0.5 µg DNA. 100 µL of transformation solution per transformation was prepared. This was mixed from 50 µL of 1 M Lithium Acetate, 25 µL of 10x TE and 25 µL of 1 M DTT. For the transformation, cells were scraped from plate and dissolved in 40 µL sterile ddH_2O. 10 µL of the cell suspension was mixed with transformation solution and the ssDNA/DNA mix. Afterwards, 400 µL PEG 3350 was added and mixed. Subsequently, a heat shock was performed by incubation at 40 °C for 30 min. After heatshock, the suspension was centrifuged at 2000xg for 5 min. The supernatant was removed and the pellet resuspended in 100 µL ddH_2O. Cells were plated on appropriate plates and incubated at 30 °C. For dominant marker, cells were resuspended in complex media like YPD and regenerated for 3 h at 30 °C and plated afterwards.

Preparation of colony picking experiments

Initially, *E. coli* NEB 10-beta were used for electrical transformation with the plasmid p10012. The plasmid contains an RFP and an GFP, which are expressed constitutively. The cells were grown in 25 mL LB + Amp at 180 rpm and 37 °C for 16 h. On the next day, the cells were harvested in a 15 mL reaction tube by centrifugation at 4 °C and 3488xg. The supernatant was discarded, and the pellet was resuspended in 15 % (w/v) glycerol. Again, the OD was measured. The Agilent OD Calculator was used to calculate the cell count. The result was used to calculate the dilution to get 20,000 CFU/mL. The cells were aliquoted and frozen at −80 °C.
For a picking experiment, the aliquot was thawed on ice and was 1:5 and 1:10 diluted in 15 % (w/v) glycerol. The dilutions were plated on LB agar (500 µL). The agar plates were GBO Onewell plates (670190) which were filled with 50 mL LB + Amp media. The plates were incubated either in the Cytomat 2 in the CompuGene Robotics Platform or in a standard incubator at 37 °C. Afterwards, the plates were used in the respective picking pipeline as described in section 4.5.3.

3. FLEXpress - Methods for rapid prototyping in biotechnological applications

Strain development is a wide field with many possibilities to speed up the whole process and to lower the costs. Developing new strains is required in biotechnology, metabolic engineering as well as in SynBio. A lot of different applications can are available, which require modifications of MCF.

3.1. Introduction

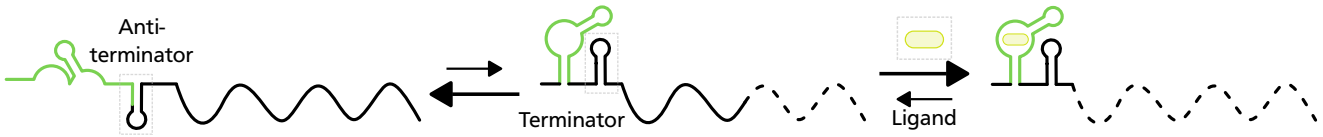
Screening for new natural products and drugs – like secondary metabolites as well as proteins – is done by genomic studies using bioinformatical approaches. This can be done for a single genome, but also for environmental samples from soil or marine environments in metagenome studies. Modern sequencing methods are sophisticated, so they deliver enormous amounts of data in a very short time for a lower price. In the past, natural products from environmental samples were obtained by cultivation of the found samples [190]. But just a small per percentage of the organisms can be cultivated. One possibility is simply a lack of knowledge about cultivation conditions. The other possibility is the cell state of the organisms, which perhaps needs certain factors to switch back into a growing state. For example, the fraction of cultivable organisms from soil samples is estimated to be between 2.5 and 19% [290]. Additionally, getting new products can be hampered by the pathway itself, which is not in an active state. It is possible that the required factors to activated the pathway are simply unknown or are not provided. Antimicrobial agents on the one hand are agents to prevent other microbes from proliferation. On the other hand, this agents are cooperative signals in microbial communities, which are required to activate the above mentioned expression [1][394].
To overcome culture-dependent techniques, multiple techniques have been established. E.g., a knowledge-based approach looks for desired functions on DNA level using homology-based algorithms. The knowledge about already known and characterized motifs and functions can be used to analyze the metagenome sequences [102]. This results can be used to clone genes with desired functions into expression hosts, which are subsequently screened for the secondary metabolites or expected functions. An other possibility is the extraction of DNA from environmental samples which can be transferred into libraries with subsequent functional screening of the derived strains [190][394]. The heterologous expression of the desired genes is usually the most challenging part. The transcription and translation of such genes includes proper folding [113]. Well-established model organisms are widely used, which are often extensively characterized and can easily be genetically modified, like *E. coli*. But, the correct folded and active version of a protein requires several factor which are possibly just provided by the origin organism, like co-factors, chaperons or protein-modifiying enzymes [113][190]. Modification of the expression host can already lead to improved product yields. But in some cases the modification of host organisms is laborious. The bakers yeast is easy to modify because of the highly efficient HR system [120][268]. On the other hand the widely used working-horse *E. coli* is difficult to modify [76]. Especially, when the prokaryote should be modified by a genomic knockout or insertion. Even modern approaches recruiting CRISPR-Cas9 still need the λ-Red

recombination system [61][178].
Over the years, *E. coli* has been modified in several ways to improve heterologous expression of proteins. For example, the Merck Group developed a strain which contains a plasmid to improve the expression of eukaryotic proteins and proteins with higher amount of disulfid bonds by providing tRNAs for rare codons[1]. However, the number of standard organisms is limited to a number of good characterized organisms. Even with modification is the product-yield low or below the detection limit. This could result in some products not being detected at all, especially when libraries are analyzed. However, using model organisms goes along with some obstacles [190]. Usually vectors have to be cloned for each host to test the expression. Thus, there have been attempts to develop plasmids which can be used in multiple host organisms, e.g., pMBD14 for *Streptomyces lividans* and *Pseudomonas putida*[244] or pJWC1 for *Agrobacterium tumefaciens*, *Burkholderia graminis* and some other [67]. However, there are other disciplines which require rapid approaches to test different constructs. The pharmaceutical industry uses genetic engineering and heterologous expression of proteins like human Insulin. Since 1973, plasmids can be assembled artificially by restriction endonucleases [64]. The development of this technique paved the way for the production of recombinant Insulin, which is expressed in *E. coli* and *S. cerevisiae*. But in *E. coli* it is necessary to perform additional downstream processing steps to get the functional protein, because the prokaryote produces insoluble inclusion bodies [15]. This problem is similar to the problem described above, that folding helper and other posttranslational modifications are missing or are different from the origin. A rapid prototyping approach as well as a rapid testing of different expression hosts could be beneficial for time and costs which have to be spent to develop new biotechnological products.
A branch of biotechnology has evolved in a way that it uses engineering-like techniques to specifically modify or designs organisms, respectively. Like state above, the SynBio (section 1.2) applies their DBTL cycle (Fig. 1.2) to develop new modules [109]. In comparison to 'classical' metabolic engineering involves SynBio prediction driven approaches to improve flux and product yields in the host organism [186]. This approach requires knowledge about the parts and organisms used. So far, the model organisms *E. coli* and *S. cerevisiae* are mainly used – but as already mentioned – these are not always the best organisms for the task [186]. Thus, other organisms have to be explored which are possibly better suited to express a certain compound. Several parts have already been developed, like promoters, insulators, artificial transcription factors, riboswitches and many more.

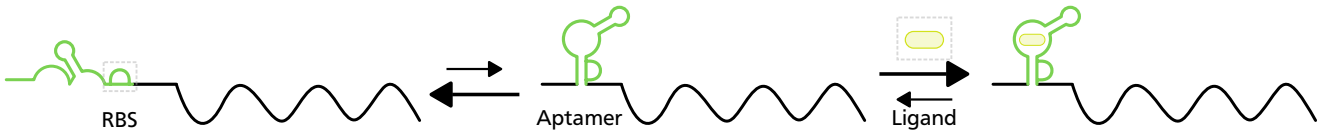
3.1.1. Riboswitches

Riboregulators are naturally occurring regulators of different cellular tasks. They can act in different manners and have been found in bacteria, archaea and as well as eukaryotes [277]. For example, this regulators are partly responsible for imprinting in the human genome, like silencing. The silencing of one X chromosome copy is done by a 16.5 kbp RNA which does not directly interact with the DNA of the chromosome which is silenced but acts like a coating complex [94][154]. Another example is the maintenance system for plasmids in bacteria, such as the R1 plasmid. The so called *hok/sok* gene encodes for a surpression RNA (*sok*) and a host killing protein (*hok*). Only if the daughter cell inherits the R1 plasmid, than enough of the translation inhibiting Sok-RNA is produced [278].
Another type of regulating RNA is found in *E. coli*. Small untranslated RNAs are sometimes located together with there respective target, leading to complementary sequences between regulator and target mRNA. An example for such a regulator is *OxyS*, which is induced during oxidative stress [4] and regulates the intracellular H_2O_2 titer [125]. This RNA blocks the access of the ribosome to initiate the translation of *fhlA* by masking the RBS[4][94]. Beside interfering RNAs, switches were found to regulate the transcription by

[1] https://www.merckmillipore.com/DE/de/product/Rosetta-gami-BDE3-Competent-Cells-Novagen,EMD_BIO-71136 (date of access: 12.11.2020)



(a) Regulation on transcriptional level



(b) Regulation on translational level

Figure 3.1.: Scheme for mode of action of a riboswitch. Aptamers can be placed within the 5'untranslated region (UTR) of a mRNA. The aptamer has a loose unstable conformation. If no ligand is present, then the equlibrium is shifted to a loose conformation. The loose conformation forms an anti-terminator (a) and does not interfere with the RNA-polymerase. If the ligand is present, the conformation of the bound state is stabilized by the ligand and a terminator is formed, preventing synthesis of full length RNA. The translation (b) is hindered by sequestration of the RBS when the ligand is bound. Modified from Findeiß *et al.* [104] and Tucker *et al.*[363]

sensing metabolites and other small molecules. Usually, these elements are located in the 5'UTR of the mRNA [277]. When no protein is required to sense and regulate the expression and it is directly modulated by the riboregulator, than this element is called 'riboswitch'[255]. Riboswitches on transcriptional level consist of a so called 'aptamer' and an 'expression platform'[39]. The aptamer domain is responsible for ligand binding. The conformational change during ligand-binding leads to a structural change of the expression platform (Fig. 3.1). The expression platform forms a intrinsic terminator, which hinders the RNA-polymerase from strand-elongation [39][57]. The absence of a ligand leads to formation of an antiterminator, which does not interfere with the RNA-polymerase [39]. The same system can sequester the Shine–Dalgarno sequence (SD)[40]. The binding of the ligand is achieved by structural orientation of the nucleobases. The small sequence space of four nucleobases leads to a high conservation of the aptamer sequence, since the threedimensional orientation has to be maintained [40][57].
An example for a bacterial riboswitch has been studied by Winkler *et al.*[386] and Nahvi *et al.* [255]. They could show a conformational change of the *thiM* mRNA in *E. coli*. The structural change occurs in presence of thiamine pyrophosphate, leading to a sequestration of the SD. In eukaryotes, riboswitches have an influence on different slicing pattern [40]. Donovan *et al.* [84] described a family of thiamine pyrophosphate binding riboswitches in *Candida spp*, influencing the splicing pattern of thiamine transporters depending on the thiamine concentration.
In summary, riboswitches are part of the balancing mechanisms of the cell metabolism. Besides the naturally occurring riboswitches, synthetic riboswitches have already been developed.

3.1.2. Artificial riboswitches and a Tetracycline riboswitch

Interestingly, the first artificial riboswitches have been developed earlier than the reports of the naturally occurring riboswitches [384]. Werstuck and Green presented riboswitches which bind kanamycin and tobramycin [383]. The majority of riboswitches has been found and were developed in bacteria. But, artificial riboswitches have been developed for eukaryotes as well [384]. The development of an aptamer is

Table 3.1.: Change of Fluorescence with/without Tetracycline aptamer (TCapt) [204].

Construct	Relative fluorescence (%) no tc	Relative fluorescence (%) 250 µM	Regulatory factor
pADH1-tc1-GFP	28.0	3.6	8
pADH1-tc2-GFP	23.8	1.1	21
pADH1-tc3-GFP	20.5	0.6	37

usually done by systematic evolution of ligands by exponential enrichment (SELEX) which was developed by Turk and Gold [77][364]. The enrichment of aptamers is achieved by an iterative process. The process starts with a randomized library of oligonucleotides which are incubated with the target molecule and separated by chromatography. After elution, the RNA molecules are enriched by PCR and enter the cycle again. After several cycles, the remaining oligonucleotides are further analyzed by cloning, sequencing and other assay [29]. The approach delivers new derived aptamers, which can be enriched by the desired target. Thus, they are promising tools for controlling gene-expression by adding or omitting the ligand.
Usually, fine-tuning of gene expression is achieved by transcription factors or degron-tags, which influence protein concentration upon adding or omitting small molecules. But sometimes, transcription factors have the problem that they are not tight and probably interfere with the host metabolism [204].
Suess *et al.* [346] used an aptamer which can coordinate a Tetracycline molecule in order to stabilize the structure. The aptamer was described earlier by Berens and colleagues to elucidate the interaction of Tetracycline with ribosomal RNA [23]. The aptamer cb32 was probably part of a pool of aptamers, which were selected in an SELEX experiment. Members of this aptamer group, which were isolated at least twice, were in vitro transcribed and further characterized [23]. The resulting aptamers were used to elucidate their regulatory ability. Thus, they have been inserted in the 5'UTR in front of the start-codon, resulting already in an up to 100-fold decrease of fluorescence. The addition of Tetracycline decreased further the fluorescence by up to 3-fold for aptamer cb32. Different efficiencies of regulations were observed when the positions in the 5'UTR were altered. The most efficient repression (6-fold) was observed close to the start codon [346]. The regulation by the cb32 aptamer was further analized by placing up to three aptamers into the 5'UTR. The integration of one aptamer behind a *ADH1* promoter reduced the activity by 28 %. A residual activity of 24 % and 21 % was measured for two and three aptamer copies, respectively (Tab. 3.1) [204].
The regulation by the aptamer is probably achieved by interference with the ribosomal machinery. If the aptamer is placed close to the start codon, than it hinders RNA scanning of the ribosome. Introduced in the cap-structure it blocks binding of the small ribosomal subunit [144].

3.1.3. Plasmid based expression and it's limitations

Unlike simple transformation with vectors, integration into the genome offers several advantages. For example, it can be useful to have the target DNA only as a single copy in the cell, since the number of copies per cell can vary substantially when replicative vectors are used. Depending on the ori used, the number of plasmid copies ranges from one to up to 700 copies in *E. coli* [111]. Other studies shows copy numbers between two and 40 copies per cell [167]. However, the number of copies per cell does not only depend on the used ori. Further, the copy number can depend on environmental conditions, such as media, temperature, deficiency, and the growth phase [111][320][380]. But even under constant growth conditions, cell-to-cell variability appears to be high. Even in a population of transformants, copy number

varies between individual cells and over time [206][240]. A common problem associated with plasmids is variation in gene dosage [240]. However, the number of copies can also be deliberately increased, e.g., by using higher antibiotic concentrations of chloramphenicol [21].
However, the number of plasmids per cell has an influence on the gene dose [111]. The amount of produced protein depends on the number of copies available for expression and thus also on the plasmid copy number [110][111]. Recombinant proteins are therefore often produced with high copy plasmids. Plasmids and their products can also affect the host metabolism. Depending on the copy number, the metabolism of the host changes, leading to induction of heatshock responses, which are usually a indicator for stress [26]. The described influence on host metabolism can also lead to a dysfunction of the replication mechanisms of the plasmid, which can result in uncontrolled replication of the extrachromosomal DNA [240]. However, since the high replication costs can also have negative impact on cell growth, low copy plasmids can be used for protein synthesis [240][305]. When selecting vectors, however, attention must be paid to compatibility groups when the genes to be expressed are located on different plasmids [240][305]. For the expression of heterologous genes, the selection of the ori is therefore only one component, in addition to promoter and RBS, which have an influence on expression efficiency.
Another effect of the permanent selection for the extrachromosomal DNA is that downstream processing (removal of antibiotics from products and waste) is more difficult, and together with the antibiotic itself, a cost factor during production [93][111][275]. The auxotrohpy markers, frequently used in yeasts, have the consequence that their use is only possible in combination with chemically defined media [105]. Both antibiotics and defined media can thus quickly become a cost factor for production which should not be underestimated. In addition, yeasts like *Y. lipolytica* develop resistances for antibiotics rather quickly, so that the selective effect decreases [19]. In addition, there are no harmonized design guidelines for the development and construction of plasmids [167]. This creates various uncharacterized context dependencies which can influence expression. Context dependency has also become a problem in synthetic biology with regard to the stability of designed switches and circuits. Nielsen and colleagues successfully addressed this by adding insulators to their design [262]. Besides insulators, another way to reduce context dependency would be to standardize vectors [328]. To this end, the third version of the Standard European Vector Architecture (SEVA), which aims to harmonize plasmid design, was published in 2020 [245].

3.1.4. Landing pads for stable Integration

One way to avoid some of the described disadvantages of plasmids, such as the influence of copy number, context dependency or the use of antibiotics, would be to integrate the constructs into the genome of the host. Integration allows stable transformation of the host without the need to maintain selection pressure through antibiotics or withholding amino acids [93]. In some organisms, the selection marker can be left out for the transformation [173] or the selection marker can be removed again by using a Cre recombinase or flippase [196][331].
But, in the end, integration into the genome is not context independent. Thus, integration into the genome of prokaryotes as well as eukaryotes shows context dependency. For example, studies in yeast showed that the proximity to open reading frames or the proximity to telomeres can have an influence on expression [105][296]. These effects are also known in prokaryotes, for example, position-dependent expression was observed in *Pseudomonas putida* [336]. In *E. coli*, other studies showed that the expression depends on the position in the genome, which is, among other things, dependent on the positioning relative to the replication origins [336]. Additionally, up to 300-fold variation in expression levels was measured by reporter gene readouts [43]. In addition, the proximity to DNA gyrase sites showed a positive effect on expression. Furthermore, it could be shown that neighboring genes influence the expression, independent of the orientation to the adjacent gene [43].

Insulating the expression cassette can reduce the noise and the context-dependent influence. Insulation has already been developed for circuit plasmids and was already used for integration cassettes [262][275]. Park and colleagues have integrated artificial landing pads into the *E. coli* genome having strong double terminators to efficiently block interference from outside in the sense and antisense direction [275].
Other landing pads have the advantage that they can be designed differently depending on the task. Kuhlmann and colleagues created artificial sequences to improve recombination into the *E. coli* genome. For this purpose, the pad sequence was randomly generated and tested against the *E. coli* genome to verify the uniqueness of the sequence. The pads were subsequently integrated into the genome. Finally, these serve as unique landing sites for integration with the Lambda Red system [206]. For integration in proteobacteria, a landing pad with mutated lox sites was designed to integrate heterologous genes into the genome using the Cre recombinase [378]. For the use of CRISPR-Cas9, another system was developed for yeast. Here, per landing pad, different numbers of copies were integrated into the genome. Each landing pad can be targeted with a unique sgRNA sequence. Thus, different landing pads can be targeted depending on the desired number of copies in the genome. If the heterologous gene is supposed to be integrated four times into the genome, the guide that belongs to the landing pad that has been integrated four times into the genome is selected. This method can be used to test quickly the load balancing which is regulated by the gene dose [35]. A minimalistic form of synthetic sequences for targeted integration are artificial Cas9 target sites integrated into the genome along with the PAM. The target sites, integrated into highly expressing loci, are used for standardized integration of heterologous genes into the yeast genome using Cas9 [13]. Cas9 has become a widely used tool in metabolic engineering as well as in SynBio.

3.1.5. CRISPR-Cas – A short history

The first indications of what is currently one of the most efficient tools for targeted genome editing already existed in the 1980s. The repetitive DNA sequences were recognized for the first time by Ishino *et al.* in *E. coli*[162]. Although the first sequencing methods were already available, they were not comparable to the methods used today [247][310]. Thus, it was difficult to classify and interpret these sequences in the early beginning of the genomics era. With increasing availability of partial and complete genome sequences, these repetitive motifs began to show up in other bacteria and archaea [163]. The term clustered regularly interspaced short palindromic repeats (CRISPR) was introduced by Jansen and colleagues in 2002 [168]. These sequences showed the same characteristics across species boundaries. The repeat regions are found between gene sequences at multiple loci and share the same short segments with little variation. The repeats vary in length depending on the organism [168]. They are separated by short pieces of DNA. The entire array is prefixed by a leader sequence [163].
Homologous genes, termed CRISPR-associated genes (Cas), were found within a short distance to the CRISPR loci in many species [168]. Interestingly, some species with multiple CRISPR loci having different repeat sequences, also showed different sets of Cas genes at these loci. Their function was still unknown, but helicase and DNA-binding domains were already identified [168]. As more and more sequences became available, also from bacteriophages and prophages, Mojica *et al.* and Pourcel *et al.* discovered similarities to spacers between repeats and bacteriophages and plasmids [251][289]. ”In this view, it is tempting to further speculate that CRISPRs may represent a memory of past ‘genetic aggressions’.” [289] The work of Barrangou and colleagues showed clearly the role as a defense mechanism by inserting or removing spacers from the genome of *Streptococcus thermophilus*. By doing this, the prokaryote could be made resistant or vulnerable to phages respectively [18]. The Cas genes used in this process have different functions. However, it was shown that Cas9 has a key role in the immune defense and that the ability to defend against foreign DNA is lost when Cas9 is knocked out [311]. Finally, an in vitro trial led to the programmable genetic engineering tool, which is commonly used today. The Cas9 from *Streptococcus*

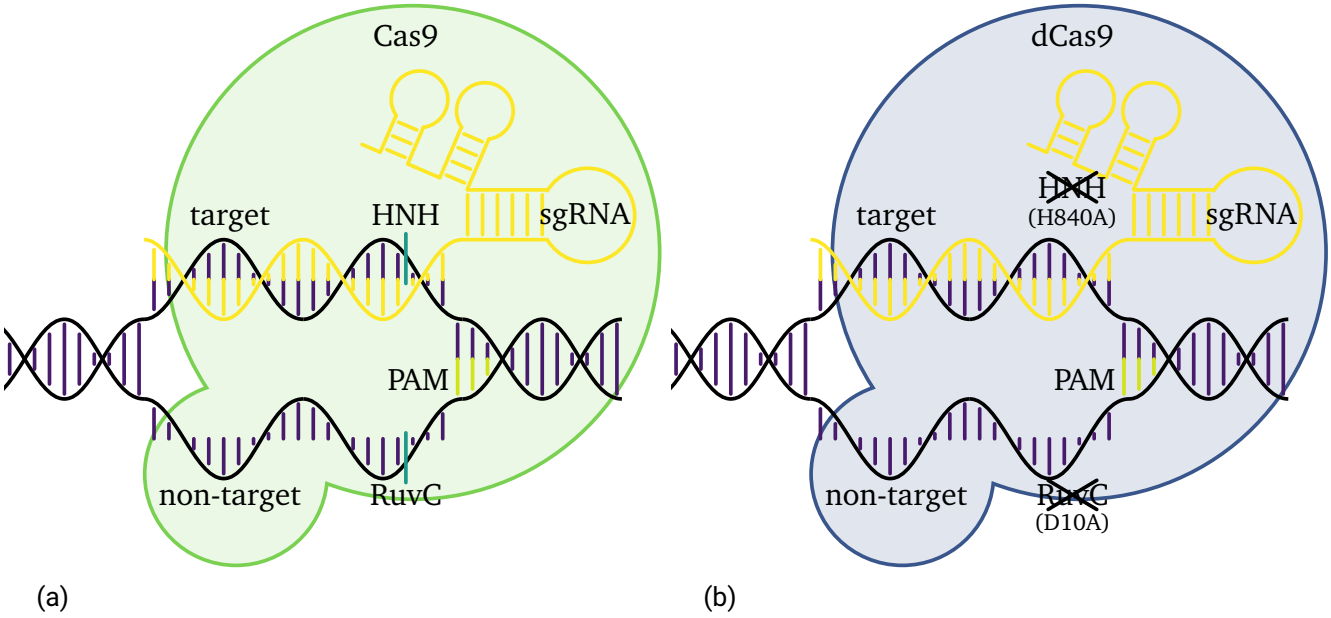


Figure 3.2.: Scheme of (a) Cas9 and (b) dCas9 DNA:RNA complex. Both enzymes are shown with sgRNA hybridized to the target strand. The cleavage sites of the nuclease domains are located three bases away from the PAM. In case of dCas9, the nuclease domains were deactivated by amino acid exchange with alanine.

pyogenes showed the ability to cut specifically DNA with its domains HNH and RuvC (Fig. 3.2a). The Cas9 enzyme is directed to its protospacer using a tracrRNA:crRNA duplex. A significant simplification of its use as a genetic engineering tool was the development of a ”chimera” of tracrRNA and crRNA. This single-guide RNA (sgRNA) molecule is used to direct the nuclease to the sequence specific cutting site [179].

3.1.6. CRISPR-Cas – A Tool for several purposes

The use of these chimera, or more precisely the modification of only 20 bp to cut the target, is one of the major advantages of Cas9. Setting a double-strand break can significantly shift the ratio of the used DSB repair mechanisms towards HR in some organisms [87][368]. Previous tools for inducing DSBs in a targeted manner, such as zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN), showed intrinsic non-specificity. The zinc fingers are often fused with nucleases so that the latter does not cut precisely at the DNA-zinc finger interaction site, but at some distance [114]. ZFNs and TALENs can be equipped with other modules due to their modularity, such as nucleases, transcription activators or repressors [114]. However, the design for ZFN seems to be quite inconvenient from today's point of view, as a specific amino acid sequence has to be generated and tested for each triplet interaction [56]. In addition, the use of TALEN is also associated with a high cloning effort due to a high proportion of repetitive sequences [114][149].
Cas9 has greatly simplified programmability and at first glance seems to have only the limitation of a PAM adjacent to the protospacer [284]. A limitation that has already been addressed by protein engineering [156]. Here, Pickard-Oliver and Gersbach provide a good overview of the variants with improved specificity and alternative PAM sequences that have been developed to date [284]. The main role of Cas9 is to cut DNA from invaders, so the enzyme is widely used for targeted genome editing. Insertion of blunt end double strand breaks triggers DNA repair, which is mainly NHEJ or HR. For replacement, knockout or

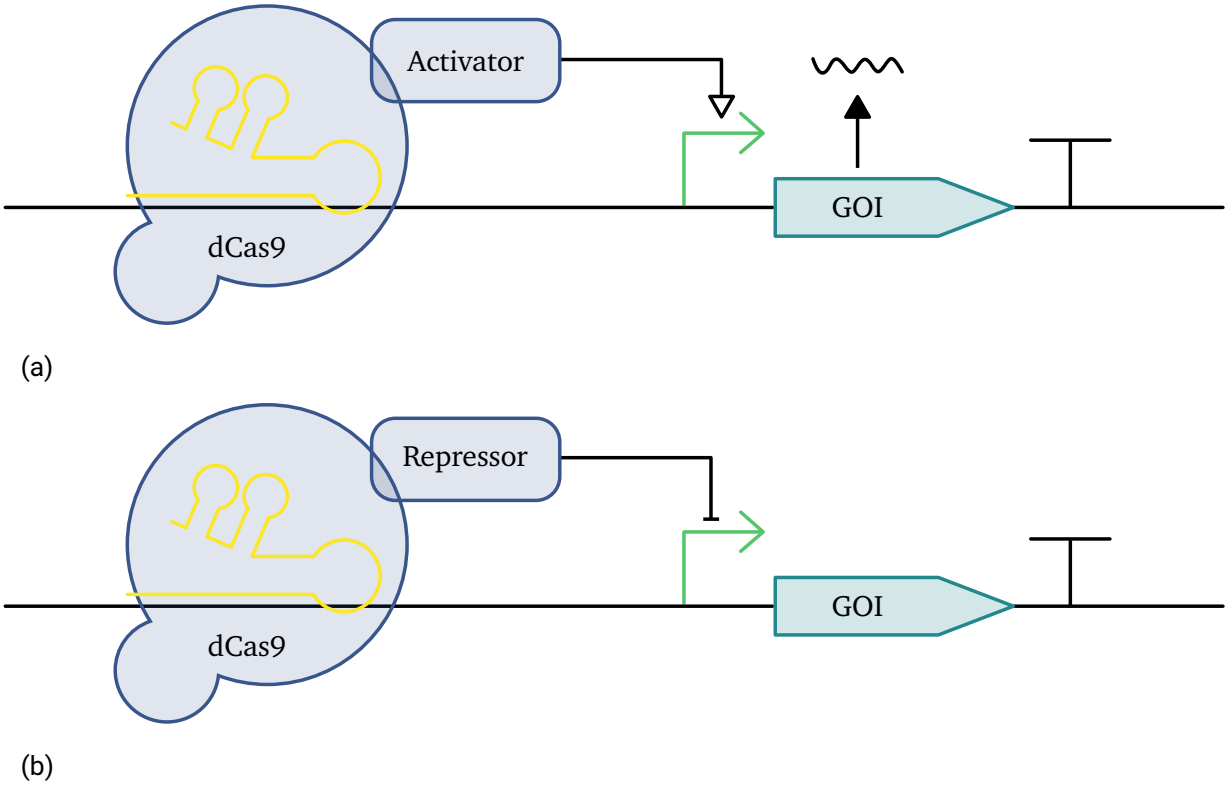


Figure 3.3.: Scheme of CRISPRa (a) and CRISPRi. For CRISPRa, the inactive nuclease is equipped with an activator domain which enables dCas9 to mediate the transcription activation. For CRISPRi, the enzyme can be coupled with a repressor to enhance the repression capabilities. Finally, dCas9 can be used to interfere with the transcription activation or with elongation by a road-blocking mechanism.

knockin, HR is often preferred as it allows targeted editing without scars. For this purpose, a corresponding repair template has to be provided [41]. On the other hand, NHEJ can cause indels or, when cut at two sites, deletions of several kilobases [169]. To reduce the frequency of NHEJ as a repair mechanism, Cas9 can be converted into a nickase by inserting either the D10A or H840A mutation into SpCas9. Accordingly, either the target or the non-target strand is cut [169][312]. Furthermore, Cas9 was used in combination with other enzymes. The enzyme was modified or fused with other proteins for various applications. For example, to insert single point mutations, dCas9 has been fused with a cytidine deaminase to insert mutations in a defined region close to the target side [202]. But not only DNA can be targeted for cutting. A system called Rcas9 can bind RNA molecules and offers the possibility to interfere the metabolism on transcriptional level [267][284].
Cas9 can be used in other ways to interfere with metabolism or cellular processes. Qi and colleagues introduced two alanine mutations into the HNH (H840A) and the RuvC1 (D10A) domains to obtain the nuclease inactive variant of Cas9 (dCas9) mentioned above (Fig. 3.2b). Using this simple road blocking method, an up to 300-fold repression was achieved with the CRISPRi system (Fig. 3.3b). This acted both in the genome and on a plasmids [292]. Repression can be further enhanced by using dCas9 repressor fusions. This was shown in cell culture using a Krüppel associated box (KRAB) domain and in a yeast using Mix1 fusions [122][382].
Not only efficient repression can be achieved by using CRISPRi. dCas9 can be used in combination with

transcriptional activators for targeted activation of expression (Fig. 3.3b. In yeast, activation was achieved by using viral activators. For this purpose, in eukaryotes, a herpes simplex activator is commonly used, which is used as monomer (VP16) or in multiple copies (VP48, VP64, VP192) [122][219]. To provide activating function, extended sgRNAs were developed. These elongated sgRNAs were extended by a module which is recognized by a binding-protein/activator fusion. MCP, PCP, and Com were fused with VP64 and used in *S. cerevisiae*. Thus, in terms of genetic switches and circuits, an additional switching layer was included [401]. The research group also succeeded in transferring the technology to the model organism *E. coli*. Because a suitable activator was not available, different activators were tested. With the MCP-SoxS fusion, an endogenous activator was found that showed good activation performance [83]. Activation via interaction with RNA polymerase complex in bacteria is still challenging, as the range of efficient activation seems to be relatively small in bacteria [25][83][228][229]. Therefore, proper design of guides and reduction off-targeting is of immense importance.

3.1.7. Cas9 and Off-Target cleavage

Offset targeting has been frequently observed and is a serious problem, especially with regard to its use as therapeutical agent. First, however, the question arises how the recognition of the target takes place. The *apo* enzyme is catalytically inactive and is not able to bind DNA or can only bind non-specifically, respectively [340]. The binding of crRNA:tracrRNA or sgRNA results in a major structural change of the multidomain protein [274]. In this conformation, Cas9 is able to search for targets. This is done by three-dimensional collision and testing for a valid PAM. If no PAM is detected dissociation from the DNA occurs immediately [176]. Binding of sgRNA results in an structural change, allowing two arginine residues to detect two guanine bases in the major groove [177]. It is important to note at this point that the detection of the PAM occurs on the non-target strand and serves to distinguish between self and non-self DNA [179][340]
However, in the next step, local melting of the target and non-target strands takes place. The nucleation and formation of the seed sequence binding occurs PAM-proximal at the two solvent-exposed bases. Apparently, at these positions, mismatches are not tolerated, so the Cas9:RNA complex loses the ability for further recognition and strand infiltration [177][340]. Further formation of the DNA:RNA duplex occurs in a seed sequence. This 12 bp long sequence is held in an A configuration by various hydrogen bonds with the protein framework [177]. In this region, mismatches with the target DNA are less tolerated in comparison to the 5' end of the RNA [165][177][179][340]. Further formation of the R-loop structure, the kinetics as wells as intermediate states are not yet fully understood [165]. Full formation resulted in a conformational change in the Cas9 enzyme resulting in activation of the nuclease domains [176]. This is an additional checkpoint to prevent non-specific cleavage [299].
However, off-target cutting may occur, in particular when the seed sequence is very similar or homologous [176]. A good definition is given by Ricci *et al.*: "At the molecular level, off-target effects are the unselective cleavages of DNA sequences that do not fully match the guide RNA, bearing base pair mismatches within the DNA:RNA hybrid" [299] Lin and colleagues grouped them in three categories: 1.) same length but with mismatches, 2.) at off-target site, some bases are missing and 3.) at off-target site, there are additional bases [226]. To prevent off-target cleavage, there are additional security mechanisms available. The so called checkpoint state is kept until about mm@17-20[2]. At mismatches mm@18-20 of the guide sequence, the conformation swings to the active state, with cutting rates comparable to on-target binding [299]. The Watson Crick mismatches appear to keep the PAM-distal end apart from each other, so new interactions with the nuclease domains prevent flipping to active state [299]. Nonetheless, this checkpoint provides additional protection from cleavage at off-target sites but not from binding and resting at off-target sites, e.g., the same is true for an inactive Cas9. Mismatches also seem to be tolerated in a sequence-

[2] 1 = PAM proximal, 20 = PAM distal; mm = mismatch; mm@17-20 = four mismatches at positions 17, 18, 19, and 20

specific manner [112]. Whereas PAM-proximal mismatches - especially the first two bases - are mostly not tolerated [112][177][179]. The same is true for seed sequence, where mismatches tend to be less tolerated [112][177][176][341][352].
Furthermore, time-resolved experiments show that when the heteroduplex is formed beyond 6-8 bases, the residence time on DNA increases significantly. Such off-targets are discarded very slowly [329]. Conclusively, off-target sites are a serious problem. In some cases, up to 150 off-target sites have been found in U2OS or HEK293 cells [362]. This is a serious problem especially for the clinical application of CRISPR-Cas.

3.2. Scope of FLEXpress

3.2.1. FLEXpress-vector

Modifying bacteria and yeasts but also mammalian and plant cells respectively, requires usually cloning of vectors for the respective host. In most cases, each organism has a specific system to modify the genome, to introduce knockouts, knockins as well as introduction of cassettes for heterologous gene expression. Transformation of organisms with replicative vectors is a relatively simple method, as it is extra chromosomal and forgoes recombination into the genome. Depending on the application, simultaneous presence of multiple copies of the introduced DNA can be beneficial for expression [315]. SynBio claims to be an engineering science and uses its principles to design and tune switches, circuits and pathways. To do this, the different components/parts must be well characterized to predict their behavior and combine them to functional units. Usually, these parts are developed and characterized in well known model organisms, like *E. coli* and *S. cerevisiae*. However, when SynBio principles should be applied to non-model organisms, these parts have to be characterized again, in order to predict their behavior under different metabolic conditions [5]. As described above, the usage of plasmids is a simple and efficient method to modify organisms. Vectors for multiple hosts have already been developed for this purpose. Taton and colleagues developed a broad-host range plasmid to test different parts like ribozyme-insulators, promoters and resistance cassettes in multiple distinct cyanobacteria strains [355].
The availability of well characterized parts for conditional gene expression in the industrial relevant yeast *Y. lipolytica* is rather limited. Thus, expanding parts for *S. cerevisiae* could be a valuable source for parts to be used for fine-tuned gene-expression, like the Tetracycline aptamer in section 3.2.2. To rapidly test parts in the non-conventional yeast, a broad-host range vector should be developed.

3.2.2. Extending the application of the Tetracycline aptamer to a new host

As stated in section 1.3.4, the yeast *Y. lipolytica* was used as host to produce more than 100 heterologous proteins. The yeast is an attractive expression host, because it is able to grow on fatty acids or alkanes and is able to secrete large amounts of proteins and metabolites [360]. Applications in SynBio as well as in metabolic engineering depend on fine-tuned expression of genes. The toolbox for the model-yeast *Y. lipolytica* is rather limited. Some constitutive promoters, such as P_{TEF} or P_{hp4d}, are available, but their application is limited, since their strength is influenced by the growth phase [258]. The probably best studied promoter P_{XPR2} requires cultivation condition which are usually unfavorable for industrial scales or SynBio respectively [360]. Other inducible promoters like P_{LIP2} and P_{POX2} are difficult to utilize, since their inducers are insoluble in water [360]. Recently, different variants of an Erythritol inducible promoter have been developed, which can be used under standard cultivation conditions [276][360].
Although there have been new developments in conditional gene expression, it is still challenging to achieve a fine-tuned regulation in *Y. lipolytica*. As mentioned in the previous section 3.1.2, the Tetracycline aptamer provides a valuable alternative for SynBio application. The aptamer developed by Berens *et al.*[23] was

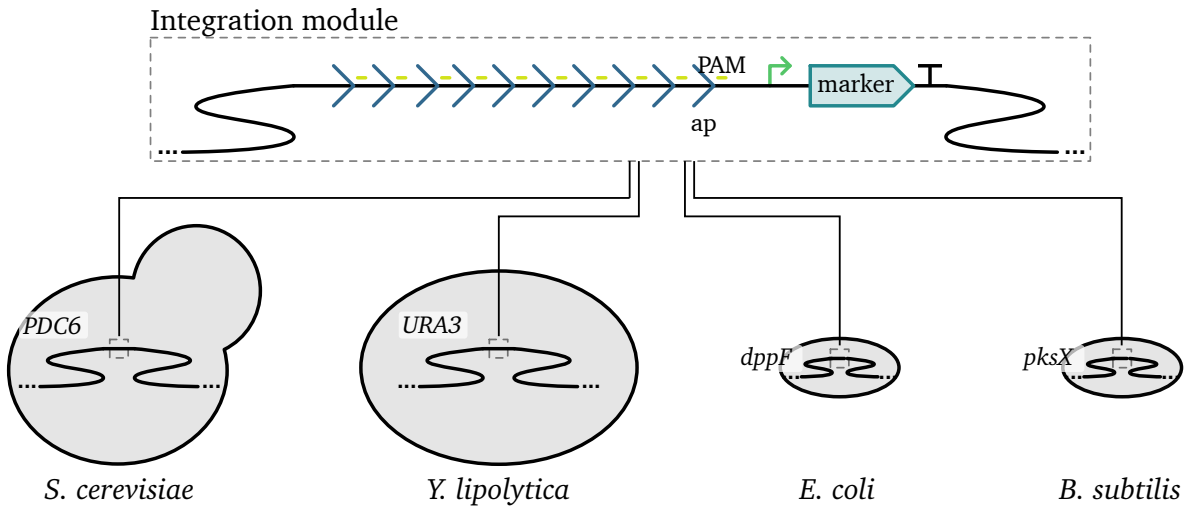


Figure 3.4.: The scheme shows the integration module for a CPad. The integration module consists of the pad itself and homology sites as well as marker cassette with promoter for the respective organism. The artificially designed pad comprises a series of computer-generated artificial protospacers (ap), whereas each spacer is followed by a PAM. The protospacers were designed to have as little as possible off-targets in the four target organisms *B. subtilis*, *E. coli*, *S. cerevisiae*, and *Y. lipolytica* when spCas9 is used. The respective integration locus is indicated for each organism.

already used as a translational switch in *S. cerevisiae* and thus represents a promising alternative to the promoter engineering approach of Trassaert *et al.* [360]. For this reason, the behavior of the aptamer should be investigated in industrial relevant yeast *Y. lipolytica*.

3.2.3. CRISPRpads for cross-species applications

The CRISPRpad (CPad)s address several challenges that have been described above. The CPads are artificial, new to nature DNA sequences developed by computer-aided design. These contain up to 25 artificial protospacer including a PAM and are separated by a spacer of 7 bp (Fig. 3.4). The design of the protospacers was carried out in such a way that these protospacers have no or as few as possible off-target effects to the host genome as well as to themselves [82]. The genomes of the model organisms *B. subtilis*, *E. coli* and *S. cerevisiae* as well as of the industrial relevant yeast *Y. lipolytica* were used as references for the calculations. Thus, the orthogonal sequences are optimal to be used in different organisms to provide a consistent DNA context for the potential heterologous DNA. The CPad will serve as artificial landing pad in the genome of the respective organism to be used for integration using the nuclease Cas9 or activation/repression by dCas9.

Subsequently, the nuclease can potentially be used to more easily integrate heterologous constructs into the host genome. Here, the harmonization of the sgRNA delivery and the harmonization of the repair templates support the idea of an orthogonal and easy to use landing pad system in multiple hosts. In case of a multi-gene pathway, all modules could be integrated into one CPad, which would otherwise be difficult, since twice as many sgRNAs are needed as genes to be integrated [35]. Furthermore, if different loci are addressed, the efficiency of the integration is only as high as the ”worst” sgRNA [35]. Therefore, the computational design of the artificial protospacers aims for minimal off-target activity, as well as maximum efficiency of Cas9 applications [82].

Since the application of Cas9 is not limited to nuclease function, the design also focused on conditional and fine-tuned gene expression. ”When expressing multi-gene pathways, balancing of the expression levels of the single genes becomes important in order to achieve an optimal flux through the pathway and to avoid the accumulation of undesired intermediates or by-products.”[105] To demonstrate addressing of individual protospacers in the CPad, CRISPRi and CRISPRa will be used to achieve different expression levels of a reporter gene. For this purpose, the CPads should be integrated into the genome of the four target organisms (Fig. 3.4). The already established CRISPRa systems for *E. coli*[83], *S. cerevisiae*[401] and *Y. lipolytica*[319] should serve as Proof-of-Concept, aiming for a graded activation of the reporter gene. A sufficiently efficient CRISPRa system was not available for *B. subtilis*. But, CRISPRi [283] could be as substitute for this prokaryote.

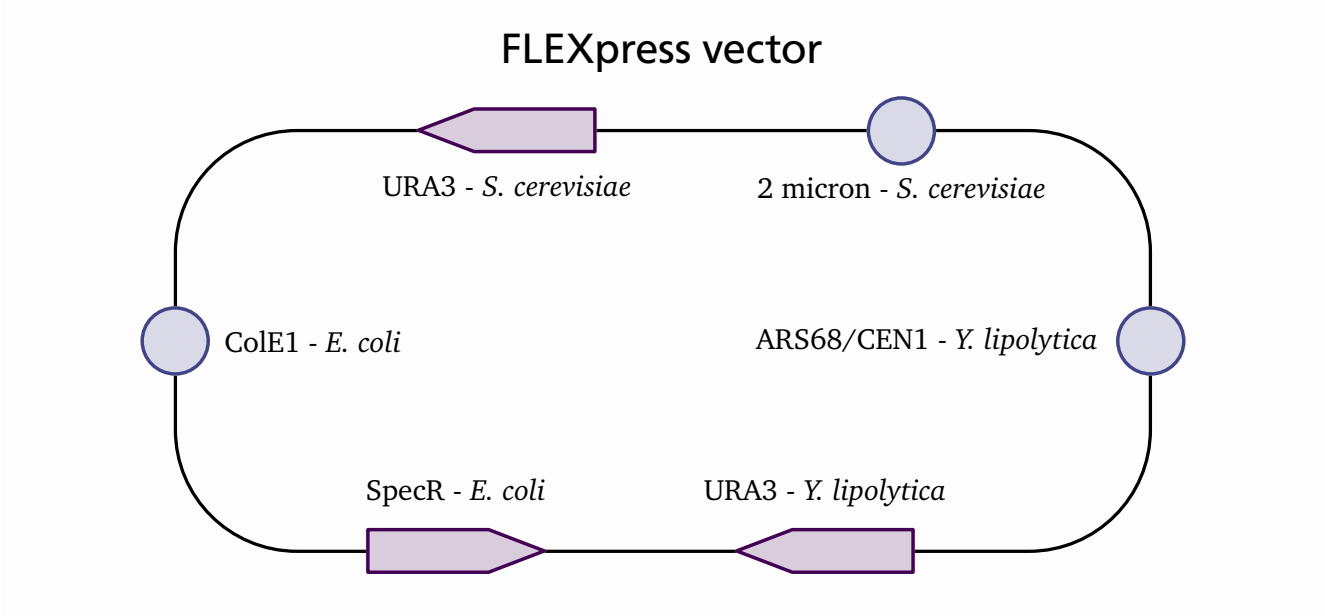


Figure 3.5.: For each organism, a marker was integrated as well as a ori. A URA3 marker was used for both yeasts, whereas a Spec^R cassette was used for *E. coli*. The replicative extrachromosomal DNA was maintained by a 2-micron ori in *S. cerevisiae*, a ARS68/CEN1 sequence in *Y. lipolytica* and a ColE1 ori in *E. coli*.

3.3. Results

3.3.1. FLEXpress-Vector: Testing parts in up to three different organisms

Previous work with *Y. lipolytica* suggests, that genetic modification could be challenging because the yeast prefers NHEJ over HR by integration into the genome [205][300]. But, the ratio can be shifted towards HR by deletion of *Ku70* or conditional suppression by CRISPRi [318][370]. However, there are also replicative plasmids available for the non-conventional yeast. The expression of heterologous parts in bacteria, such as *E. coli*, is usually carried out by transformation with replicative DNA rather than integration into the genome. This technique is also commonly used for *S. cerevisiae*, although integration into the genome is possible without considerable difficulty.
The FLEXpress vector was designed for a use case where a suitable expression host is not known before or to check, whether heterologous expression is possible at all. In order to reduce the cloning efforts for multiple organisms, the vector was designed to be maintained in three different organisms. The backbone contains marker and ori for *E. coli*, *S. cerevisiae* and *Y. lipolytica*. For maintenance in budding yeast, a URA3 marker and a 2-micron sequence was chosen. The elements for *Y. lipolytica* were an autonomously replicating sequences (ARS) in combination with a centromere (CEN) sequence and a URA3 marker [107][108][371]. A ColE1 ori in combination with the dominant marker cassette for Spectinomycin was used for *E. coli*. A scheme of the vector can be found in Fig. 3.5.
The construction of the empty FLEXpress vector p26061 is described in section A.1.4. Another plasmid was constructed to verify and test functionality in all three target organisms. The Plasmid p26062 was equipped with a constitutive promoter that drives the expression of the reporter protein mTFP. The used hybrid promoter P_{TEFag} from *Ashbya gossypii* originates from the *kanMX* marker which was described earlier by Wach *et al.*[373] The marker provides resistance to G418 (Geneticin) in yeasts and to Kanamycin

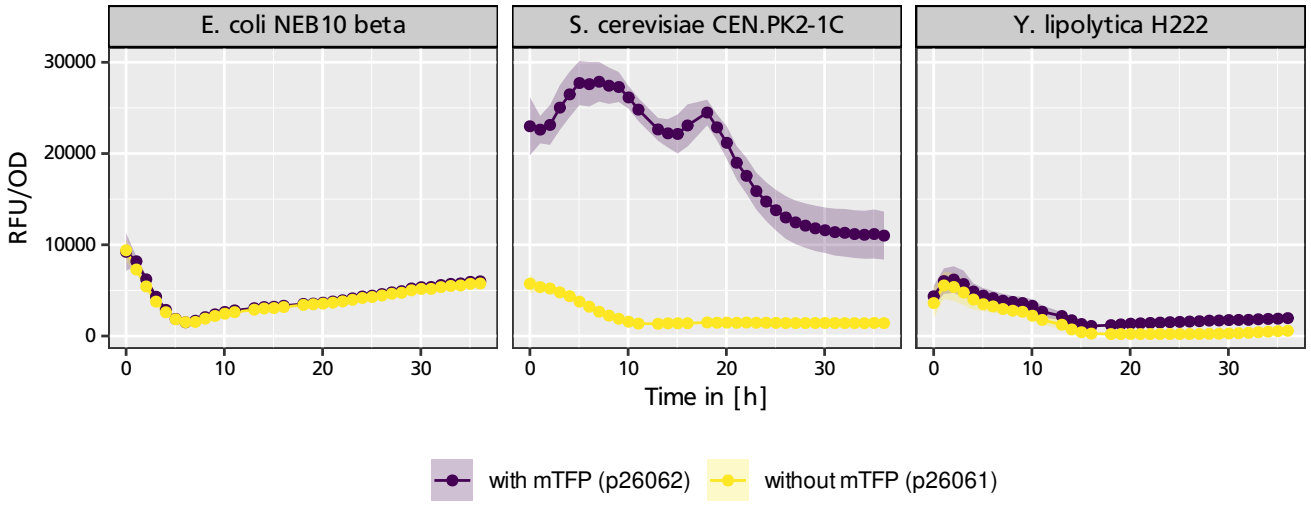


Figure 3.6.: Cultivation of *E. coli*, *S. cerevisiae* and *Y. lipolytica* with FLEXpress vector with (p26062) and without (p26061) fluorescent protein. Cultivation for 36 h in minimal medias (M9 media for *E. coli* and YNB media for *S. cerevisiae* and *Y. lipolytica*). Ribbon indicates standard deviation of n=40 samples for *E. coli* and *Y. lipolytica* and n=37 samples for *S. cerevisiae*, respectively.

in *E. coli*. The construction of the mTFP containing plasmid p26062 is described in section A.1.4. The plasmids p26061 (empty vector) and p26062 (P_{TEFag}-mTFP) were used to transform *E. coli* NEB® 10-beta, *S. cerevisiae* CEN.PK2-1C (s26419) and *Y. lipolytica* H222 (s11084). The clones obtained were used for cultivation in 96 well plates (detailed information can be found in section A.2.1). The fluorescence data were normalized to the optical density (Fig. 3.6). The R script for data evaluation removed automatically three wells with cultures of *S. cerevisiae* because no growth was detected (threshold was OD of 0.1 over the whole experimental time; additional information for data evaluation can be found in section A.2.1). Thus, standard deviation was calculated with n=40 for *E. coli* and *Y. lipolytica* and n=37 for *S. cerevisiae*, respectively and indicated as ribbon. The fold change (Fig. A.1) between mTFP and negative control was ~0,12 for *E. coli*. In *S. cerevisiae*, the fold change spanned a range between ~3 and ~17 depending on the growth phase. The highest fold change was reached at the beginning of the exponential phase (Fig. A.2) and decreased by entry into the stationary phase. *Y. lipolytica* also showed a behavior which was dependent on the growth phase. But in comparison to *S. cerevisiae*, the increase of the fold change was visible after entry into the exponential growth phase and had a peak of ~6,2 after 27 h. In *E. coli*, the OD-normalized fluorescence values (Fig. 3.6) showed only a small difference between the strain with and without mTFP. Nevertheless, $p - value < 0.05$ is true for most timepoints for *E. coli*.

3.3.2. Tetracycline aptamers for *Yarrowia lipolytica*

Strain construction for Tetracycline aptamer (TCapt) characterization in *Y. lipolytica*

The empty plasmid p26061 was the basis to construct three different vectors p26063, p26064 and p26065. Fig. 3.7 shows a scheme for the different vectors. The construction is described in detail in section A.1.4. The vector p26061 served as negative control and whereas p26062 was the positive control. The plasmids

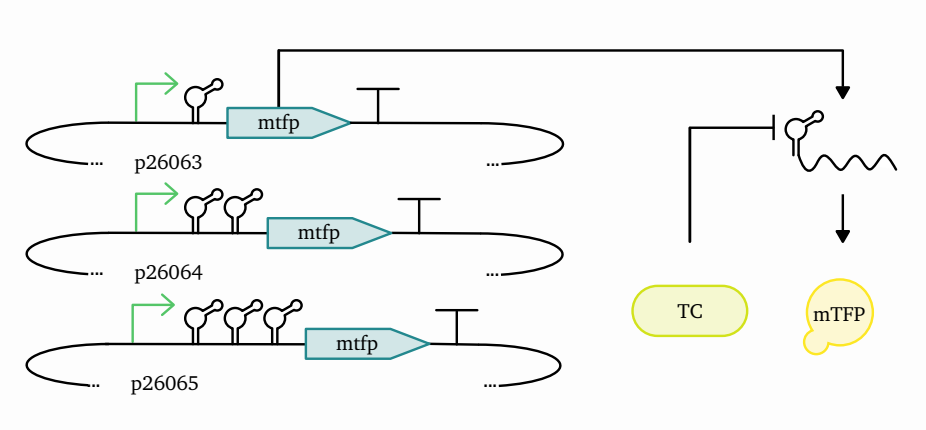


Figure 3.7.: Scheme of vectors which have been constructed to test the tetracycline aptamer for *S. cerevisiae* in *Y. lipolytica*. The expression is controlled by a constitutive promotor derived from the fungi *Ashbya gossypii*. The aptamer is placed between promoter and CDS of the target gene, which is *mTFP*. In presence of Tetracycline, the aptamer constitutes a rigid structure which blocks the translation of the target protein. Indicated are plasmid IDs for the corresponding constructs, containing either a monomer, a dimer or a trimer of the aptamer.

p26061 till p26065 have been used for the transformation of three different *Y. lipolytica* strains (H222, Po1f, and 63). The strain *Y. lipolytica* H222 is a industrial relevant strain, whereas *Y. lipolytica* PO1f is a typical laboratory strain and *Y. lipolytica* 63 a wild type strain. After the transformation procedure, the behavior was different between yeasts which have been transformed with p26061 and p26062 and plasmids which contained at least one TCapt (p26063-p26065). Usually, the first colonies are visible after two days of incubation at 30 °C on selective YNB media, which was the case for the control plasmids p26061 and p26062. But, cells which have been transformed with p26063, p26064 and p26065 showed another behavior. After one week, colonies were visible on plates for p26063 (TCapt-monomer). But, for the TCapt-dimer and -trimer, no colonies were obtained, even after extending the incubation time to four weeks. The transformation was repeated three times to exclude any mistakes during the transformation procedure. However, the transformation with a plasmid that contained more than one TCapt was not successful. The result was similar in all *Y. lipolytica* strains which were used. In comparison, *S. cerevisiae* formed colonies after three days of incubation at 30 °C for all transformations. The strains obtained were used for a cultivation experiment, which is described in the following section.

Cultivation of derived strains using lab automation

The cultivation experiment was carried out in the CompuGene Robotics platform using the randomization workflow (section A.2.3). The cultivation was done with *S. cerevisiae* CEN.PK2-1C (s26419) and *Y. lipolytica* H222 (s11084) containing either p26061 (negative control), p26062 (positive control), or p26063 (TCapt-monomer). As expected and already described by Kötter *et al.*, the normalized fluorescence dropped for the TCapt-monomer in *S. cerevisiae* in comparison to the positive control. The ratio observed was ~0.6 for the monomer compared to the positive control (Fig. 3.8). Whereas the fluorescence was unaffected by the induction for negative and positive control, the normalized fluorescence for the TCapt-monomer decreased significantly after induction with 250 µM Tetracycline (indicated by *, calculated by ANOVA and Tukey-test). The ratio between uninduced and induced condition was ~0.38 after 48 h for the TCapt-monomer. The

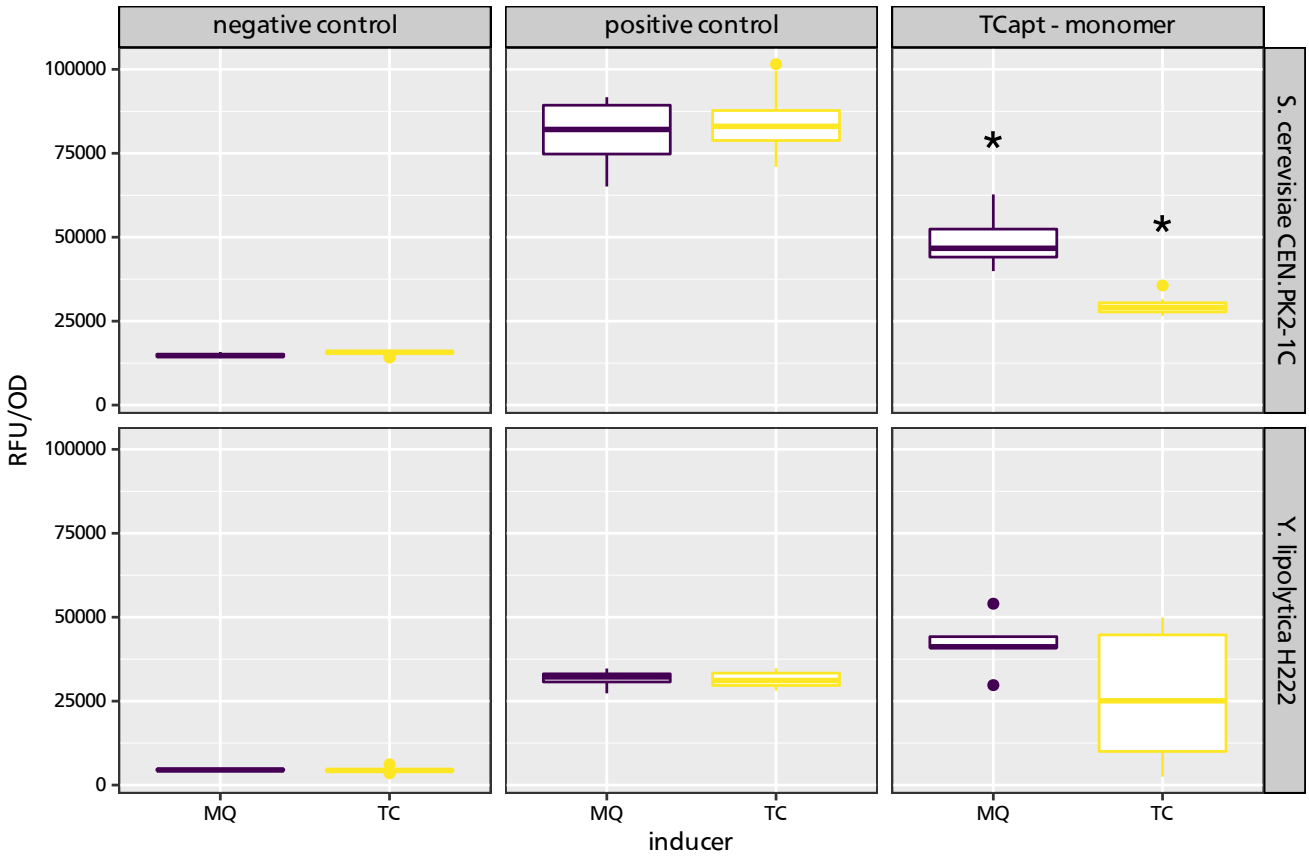


Figure 3.8.: OD-corrected values for aptamer cultivation in the CompuGene Robotics platform. Strains were cultivated in YNB minimal media. Fluorescence was measured every hour (GFP filterset for the PHERAstar®). Plot shows plate reader data for timepoint 48 h. Negative control was p26061 (empty vector) and positive control p26062 (P_{TEFag}-mTFP). Inducer: MQ – ddH_2O; TC – Tetracycline (250 µM). All strains are represented by at least 10 samples, except for *Y. lipolytica* H222 with TCapt-monomer (p26063) was represented by 7 samples. * indicates significance $p < 0.05$.

difference was stable after entry into the stationary phase after ~24 h (normalized FI over the time Fig. A.4; growth over the time Fig. A.5). Having a look at the time resolved fluorescence, the positive control had a higher normalized fluorescence, when induced with Tetracycline, showing two peaks after 10 h and 20 h (Fig. A.4). But finally, the induced and uninduced cultures aligned each other after entry into stationary phase.

For the new target *Y. lipolytica*, no significant effect was observed. The induced and uninduced controls were not effected by Tetracycline. After 48 h, the variance between the cultures with and without inducer and TCapt-monomer was to high to evaluate an effect (Fig. 3.8). In the time resolved plot (Fig. A.4), a bigger difference was visible at the beginning of the experiment. But contrary to *S. cerevisiae*, the normalized fluorescence for the TCapt-monomer was higher than for the positive control. For better insights, samples for flow cytometry were taken after 24 h and 48 h and were evaluated in the following section.

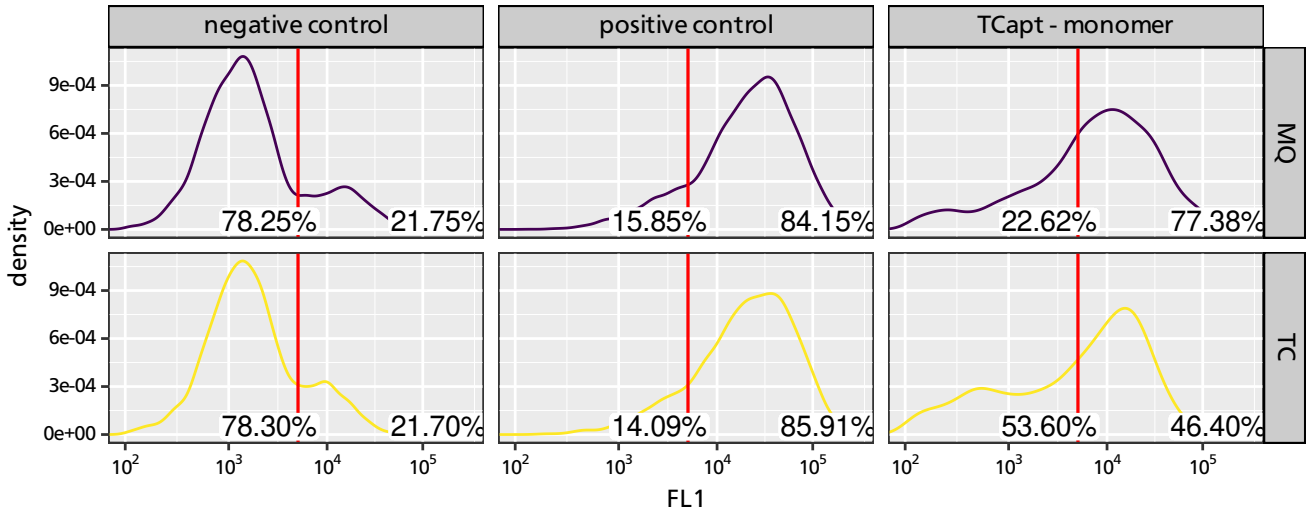


Figure 3.9.: Density plot of a flow cytometry measurement after 48 h for *Y. lipolytica*. *Y. lipolytica* strains were cultivated in YNB minimal media and 1:20 diluted in Sony Sheath Fluid. Negative control was p26061 (empty vector) and positive control p26062 (P_{TEFag}-mTFP). Inducer: MQ – ddH_2O; TC – Tetracycline (250 µM). All strains are represented by at least 10 samples, except for *Y. lipolytica* H222 with TCapt-monomer (p26063) was represented by 7 samples. The red line indicates the cutoff value which classified events into non-fluorescent (left) and fluorescent (right) events. Proportion is indicated as %.

Flow cytometry measurements of the tetracycline aptamer strains

To get the samples during cultivation, 10 µL of the main culture were transferred into a new 96 well microtiter plate (MTP) which was already filled with 190 µL Sony Sheath Fluid. The plates with the dilutions were measured in the Beckman-Coulter CytoFlex S and evaluated using R. The following gating strategy was applied to all samples, whereas organisms were evaluated separately. Initially, artifacts were removed, such as negative values, for forward scatter (FSC) and side scatter (SSC) as well as extreme outliers. To reduce subjectivity during gating, the singletGate() function of the flowStats package (V 4.2.0) was used to remove doublets by applying the function separately to the FSC and SSC channel. Since the strains showed a different growth behavior, all samples were reduced to the same number of events using the sampleFilter() function of the flowCore package (V 2.2.0).
Finally, the evaluation was supported by a linear gate in the FL1 channel, which was used to measure the fluorescence signal of the mTFP. This gate is indicated as a red line in Fig. 3.9 and Fig. 3.10. For positive and negative control, the proportion of induced and uninduced cells remained unaffected by the inducer. For induced *S. cerevisiae* CEN.PK2-1C as well as *Y. lipolytica* H222, a shift was observed towards the population which was designated as non-fluorescent cells (left side). For the TCapt-monomer, The shift was higher for *Y. lipolytica* as for *S. cerevisiae*. The data obtained support the plate reader data, but show a higher shift for *Y. lipolytica*. The shift was observed for the whole population in the plate reader data, but the difference was not significant (Fig. A.4), due to the high variance.

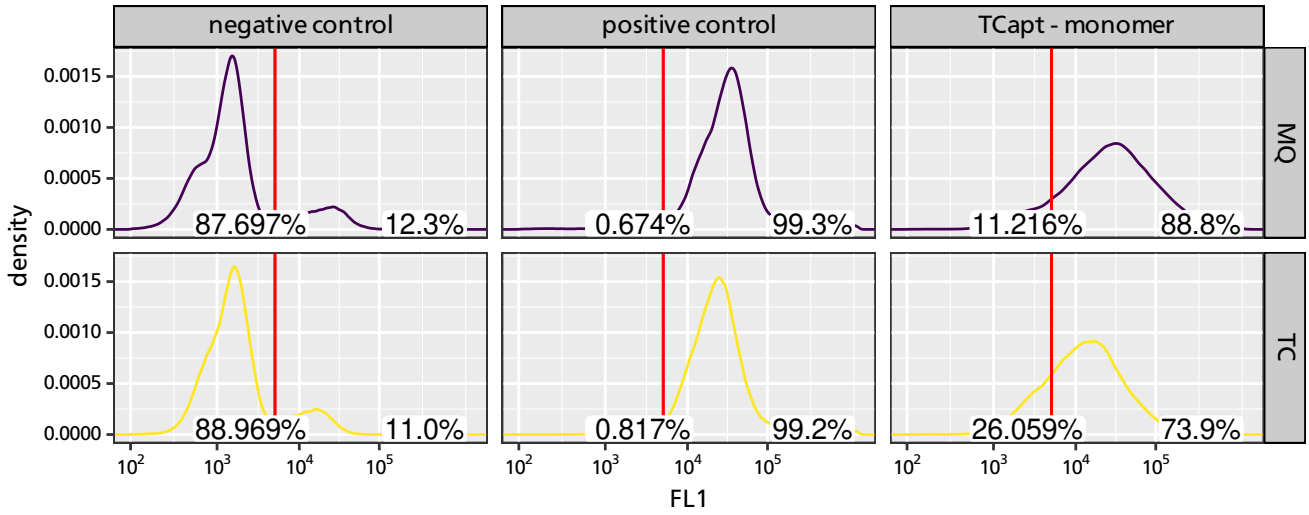


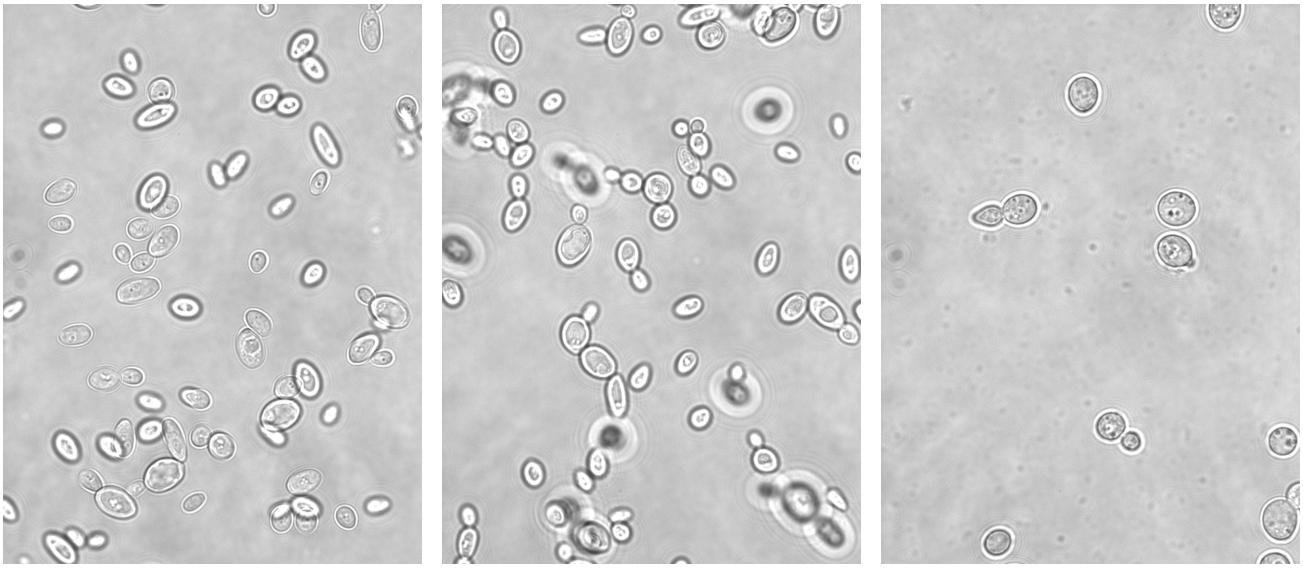
Figure 3.10.: Density plot of a flow cytometry measurement after 48 h for *S. cerevisiae*. *S. cerevisiae* strains were cultivated in YNB minimal media and 1:20 diluted in Sony Sheath Fluid. Negative control was p26061 (empty vector) and positive control p26062 (P_{TEFag}-mTFP). Inducer: MQ – ddH_2O; TC – Tetracycline (250 µM). All strains are represented by at least 10 samples. The red line indicates the cutoff value which classified events into non-fluorescent (left) and fluorescent (right) events. Proportion is indicated as %.

Growth of the *Y. lipolytica* strain with TCapt-monomer

As already mentioned above, the transformation of *Y. lipolytica* with at least two TCapts was not successful. However, transformation with only one aptamer causes the strain to be restricted in its growth. As shown in Fig. A.5, the strain having only the TCapt-monomer showed a slow increase in the absorbance values, measured with the PHERAstar. Additionally, this effect was observed in a pre-experiment during a manual cultivation performed in rich media (section A.2.4). The strain with the TCapt-monomer reached a OD of ~0.8 whereas the other *Y. lipolytica* strains reached a final OD of ~21 after 48 h. During the manual cultivation, samples were taken for flow cytometry measurements after 24 and 48 h. Fig. A.7 shows a plot with unprocessed FSC and SSC data for *Y. lipolytica* after 24 h. The strains without aptamer showed comparable shaping among each other in front and side scatter. But, the strain with TCapt-monomer had a higher rate for events with higher values for FSC and SSC. Additionally, the population is much more diverse in comparison to the other strains. The higher values for the FSC was confirmed in the microscope (Fig. 3.11) because the FSC is a rough measure for cell size. The strain containing the control construct without reporter and TCapt as well as the strain with the reporter control had their usual form, which is typically spherical and oval. Cells which have been transformed using p26063 had a higher portion of cells which had a different shape. They were bigger and completely round.

3.3.3. CRISPRpads in *E. coli*, *B. subtilis*, *S. cerevisiae* and *Y. lipolytica*

As stated above, the CPads are a new to nature artificial DNA sequence. They are rendered to reduce the off-target effects among all target organisms as well as to reduce cloning efforts for all organisms. Furthermore, the CPads provide different optimized artificial protospacer sequences, including a PAM as well as a 7 bp



negative control (p26061) positive control (p26062) TCapt-monomer (p26063)

Figure 3.11.: Microscope pictures (x63) of *Y. lipolytica* strains. Cells were cultivated in ACH media. After 24 h, 5 µL were taken and transferred to a microscopic slide.

spacer sequence. The main aim of the CPad is to reduce the amount of cloning work which has to be done for each organism. Thus, the expression of heterologous proteins can be easily achieved by integration into the same sequence context. Additionally, there are CRISPRa as well as CRISPRi technologies available, allowing fine-tuning of expression rates by targeting different protospacer upstream of the integrated protein sequence. In the following section is described, how the CPads have been designed as well as a Proof-of-Concept using CRISPRa in *E. coli*, *S. cerevisiae* and *Y. lipolytica* and CRISPRi in *B. subtilis*.

Computer-aided design off the CRISPRpads

The computer-aided design of the CPads was done by Knut Finstermeier (Charpentier Group, MPI for the Science of Pathogens, Berlin). The artificial sequences have been calculated against the genome of four industrial relevant organisms: *B. subtilis* PY79, *E. coli* BL21, *S. cerevisiae* S288C and *Y. lipolytica* W29. The protospacers were designed to have 20 bp without PAM sequence. The *in silico* design resulted in 50 functional protospacer with optimized folding behavior as well as reduced off-target activity to the genomic sequences. The protospacers were used to generate in total ten CPad sequences with a maximum length of 930 bp. The number of available protospacers per CPad varies between 25 and 29. The *in silico* design is described in detail in section A.3.1. The designs Rank1 and Rank4 (Fig. 3.12) were selected and synthesized at Thermo Fisher Scientific GENEART GmbH. The CPads were integrated into the different target organisms and evaluated using CRISPRa and CRISPRi.

CRISPRa to drive expression of a GFP in *S. cerevisiae*

Although the CPad was designed against the *S. cerevisiae* S288C genome, another strain was used to evaluate the Proof-of-Concept, since S288C is a strain without auxotrophic markers which can be used for

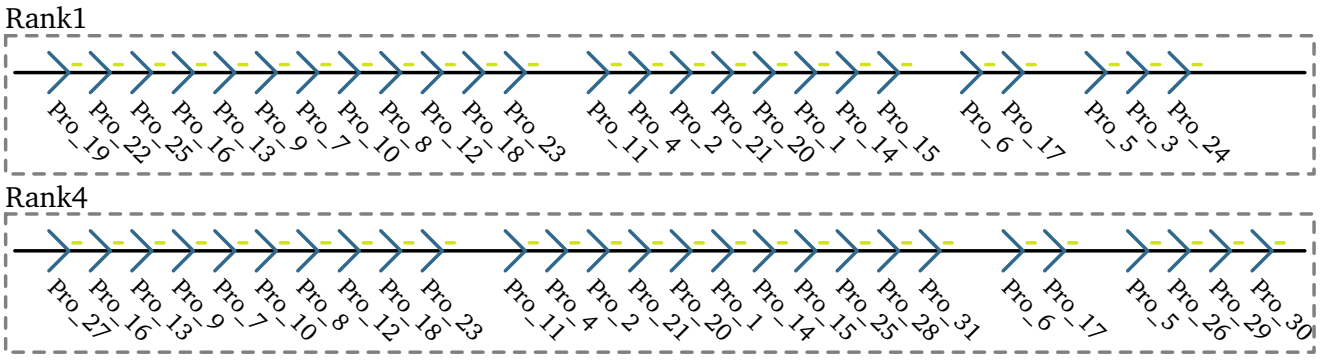


Figure 3.12.: Protospacer distribution in CPad designs ordered at Thermo Fisher Scientific GENEART GmbH. Each protospacer has a length of 20 bp. Every protospacer is equipped with a PAM, which has a NGG motive. The standard distance between PAM and the next protospacer is 7 bp. Each protospacer sequence got a unique ID with suffix Pro_ followed by a consecutive number.

cloning. Thus, different strains were evaluated, whether they could serve as host for the Proof-of-Concept. The CEN.PK2-1C is a widely used laboratory strain and is deficient for HIS, LEU, TRP and URA. But, a complete genome sequence is not available yet, as well as comparative genomic studies. Some studies showed significant difference in protein expression and lipid metabolism. Probably, the different expression pattern is caused by a different genotype. Daran-Lapujade and colleagues used oligonucleotide microarrays to compare both strains, which revealed some lacking genes but also a high degree of similarity. Additionally, a length difference for chromosome X was found [75]. Because differences in the target genome can lead to unexpected off-target effects of Cas9, other strains were evaluated for the Proof-of-Concept. As alternative, different designer deletion strains were evaluated, which have been developed by Brachmann and colleagues [36]. All strains can be traced back to the strain FY2, which is a *URA3* (*ura3*-52) deficient descendant of the S288C [387]. All strains were developed by PCR-mediated disruption, which was used for a clean knockout of marker genes [36]. Finally, *S. cerevisiae* BY4742 was chosen from the designer collection because of the available marker knockouts and the direct lineage to S288C.

The CPads were meant to be integrated into the genome of the target organism. However, it has been known for some time that the expression of genes is influenced by the locus. For example, genes have a lower expression which are located close to telomeres. Flagfeldt *et al.* [105] used a *lacZ* reporter assay to evaluate different loci for their expression rates in *S. cerevisiae*. Most of the loci were located in long terminal repeats, but three sites were coding regions: *SPB1* (AdoMet-dependent methyltransferase), *URA3* (OMP decarboxylase) and *PDC6* (pyruvate decarboxylase). According to the authors, the integration of *lacZ* did not affect the growth of the yeast. In comparison to the other non-LTR loci, the PDC6-locus had a relatively high activity in the *lacZ* assay [105]. Additionally, the PDC6 locus was studied earlier and did not find a difference in the phenotype, when PDC6 was knocked out. But PDC6 lacking strains were found to have a reduced pyruvate decarboxylase activity in media with ethanol as carbon source [151]. The results of Flagfeldt *et al.* and Hohmann *et al.* led to the decision for integrating the CPads into the PDC6 locus, as integration did not influence growth and also the phenotype remained unchanged except for the aforementioned change.

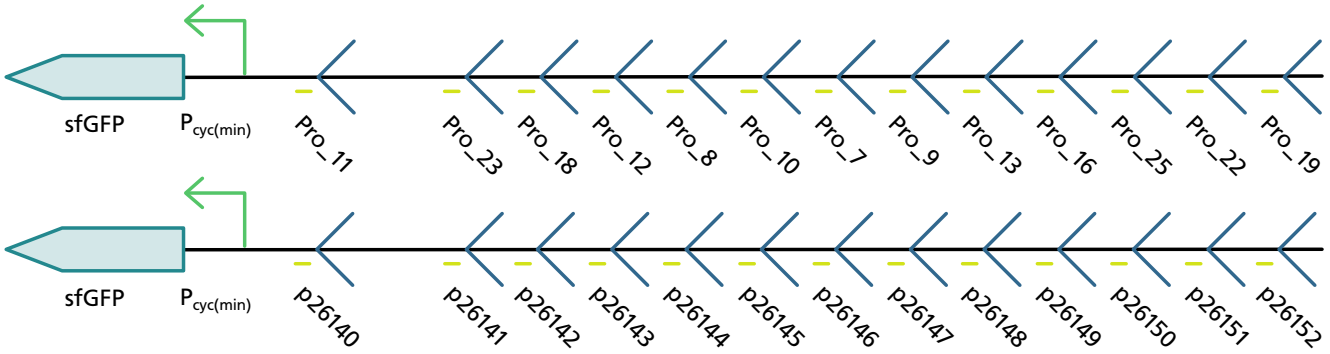
As Proof-of-Concept for specific targeting of different protospacers within the CPad, CRISPRa was used to modulate the expression of a reporter protein. The used method by Zalatan *et al.* is different from other methods which have been developed earlier. Here, dCas9 is not directly fused to a transcriptional activator. The authors used a modular design for the sgRNA which is used to direct the nuclease to the

target site. The sgRNA is extended by a so-called scaffold module. The resulting scaffold RNA (scRNA) is able to recruit an RNA-binding module. Additionally, the binding module is fused to a viral VP64 domain, which can be used to activate transcription in *S. cerevisiae*. They tested three different RNA hairpins with their corresponding proteins. Each recognition protein/activator fusion had a different activation efficiency [401]. Since the MS2 hairpin together with the MCP recognition protein showed the best performance, it was selected in the presented work. In the mentioned publication, Zalatan *et al.* used a modified version of the *S. cerevisiae* P_{CYC1} promoter. The minimal version of the promoter was extended by TetO sequences in different numbers to achieve a fine-tuning of the gene expression. The TetO boxes have been targeted with the CRISPRa machinery [401]. Here, a different strategy was used, where the TetO containing sequence was replaced by the CPad (Fig. 3.13a). In the CPad, the fine-tuning should be achieved by addressing different protospacer instead of TetO boxes. The mentioned CRISPRa-system still requires a minimal promoter to enable the binding of the transcription machinery. The version provided is still a rather long version of the P_{CYC1} promoter. A shorter version, the $P_{CYC1min}$, has been used by Machens *et al.*, which was used for CRISPRa in combination with synthetic transcription factors [236]. The promoter has a low basal activity and should drive the expression of a sfGFP reporter protein. Here, $P_{CYC1min}$-sfGFP was placed in the center of the CPad Rank1. The resulting strain *S. cerevisiae* s26458 was transformed with plasmids carrying different sgRNAs to address different protospacer upstream of the $P_{CYC1min}$ promoter. The available protospacer as well as the corresponding plasmids to address the protospacer are shown in scheme Fig. 3.13a. The strain construction is described in detail in section A.3.2. The derived strains were used for a cultivation in the CompuGene Robotics platform (section A.3.3).

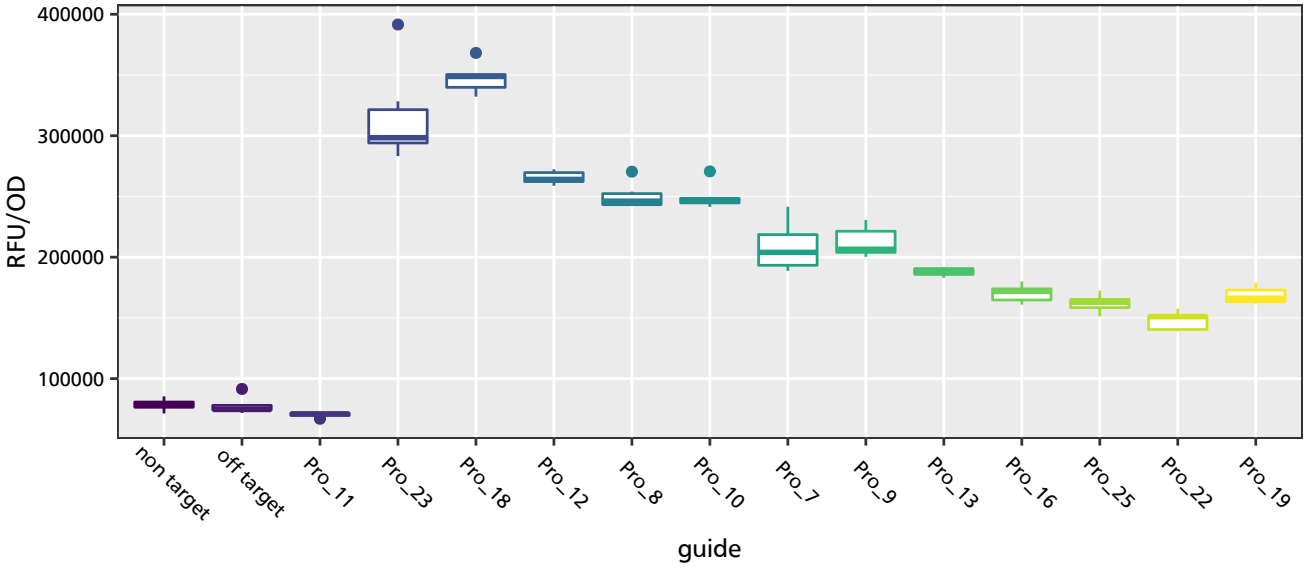
The data obtained were evaluated using an R pipeline to import and pre-process the plate reader data. The R script automatically removes wells without growth, as these are mostly a result of the randomization procedure. The threshold applied for growth was OD > 0.1. For this reason, all samples were represented by n = 6 samples except the strain with plasmid p26151 (n = 5 samples). The boxplot Fig. 3.13b shows the normalized values after 16 h (the normalized values over the time are shown in Fig. A.18). The experiment was done with two negative controls p26129 and p26139, where just background fluorescence was expected. The plasmid p26129 contained the original guide sequence to target a TetO box, which is not available in the target strain, and p26139 contained no guide sequence. Interestingly, the first protospacer Pro_11, which is located directly upstream of the $P_{CYC1min}$ promoter, showed a fluorescence which was comparable to the controls. The first protospacer showing fluorescence was Pro_23. The target with the highest fluorescence was Pro_18, with a distance of 87 bp to the promoter. As the distance to the promoter increases, the fluorescence also continues to decrease. In summary, all protospacer tested – with exception of Pro_11 – showed significant activation of fluorescence compared to controls. Over the time, the normalized fluorescence values decreased, especially by entry into the stationary phase (Fig. A.18). A similar effect was observed when the fold change of activation was calculated against the non-target control. The fold change reached a maximum of ~3.4 after 16 h for Pro_18, but decreased over the time (Fig. A.19).

CRISPRa to drive expression of a GFP in *E. coli*

The integration and Proof-of-Concept in the model organism *E. coli* is presented below. The industrially relevant expression strain *E. coli* BL21, originally developed from strain B and now frequently used in variant DE3 for expression with T7 polymerase, served as one of the hosts for calculating the artificial protospacers and CPads [344][389]. This strain is characterized by a high expression rate of recombinant protein, which makes it one of the most widely used industrial strains [321]. In this context, it can also compete against K derivatives in terms of protein yield, but seems to have disadvantages regarding system and plasmid stability, respectively [243].



(a) Top shows the distribution of protospacers upstream of $P_{CYC1min}$ promoter, as well as the plasmid IDs (bottom) containing the guide sequence to address the target protospacer.



(b) Boxplot shows the OD-normalized RFU values after 16 h of the cultivation. x-axis indicates the addressed protospacer.

Figure 3.13.: Protospacer distribution and normalized fluorescence values for CRISPRa in *S. cerevisiae*. The RFU was normalized to the OD values. Non-target control (p26129) has a guide sequence for a TetO box. Non-guide control (p26139) contained only scaffold RNA without guide sequence. Samples are represented by n=6, except Pro_22 with n=5.

The concept of landing pads is that they are integrated into the genome of the respective target organism. For *E. coli* BL21, it was reported earlier, that integration into the genome can be challenging [185]. Even the use of Cas9 supported engineering showed little improvement for BL21 [63]. For this reason, the K12 derivative BW25113 was used for the Proof-of-Concept [76]. This strain was the parent strain for the Keio gene knockout collection, and it is well characterized and in addition, it was not treated with mutagens in its strain history [10][395]. The change in strain background necessitated an analysis of potential off-targets when using the artificial protospacers. Since the genome of BW25113 has already been sequenced and is available as a ref-seq sequence, the off-targets were analyzed using the Cas-offinder (V 2.4.1) [12][128]. Although a whole-genome alignment showed substantial differences (data not shown), the Cas-offinder only showed off-targets from a threshold of four and more mismatches. Thereby, two artificial protospacers (Pro_5, Pro_22) showed the same off-targets in BL21 and BW25113 (Tab. A.6). In addition, other protospacers with potential off-target behavior were found in each of the two strains, Pro_7 in BL21 and Pro_16 in BW25113, respectively (Tab. A.6). As noted above, the modification of BL21 is rather difficult. Furthermore, the integration efficiency in *E. coli* drops considerably from a fragment size above 1500 bp [206]. Finally, for the reasons mentioned above, the well characterized laboratory strain K12 BW25113 was chosen for the Proof-of-Concept.

As the integration locus, the intergenic region between *yhjV* and *dppF* was selected, whereby the integration site is located in the terminator region of both genes and thus does not interfere with the promoter regulation of the two transporter genes. Moreover, this region has already been used for integration [200][322].

As a Proof-of-Concept, the fine-tuning of the gene expression of a reporter gene should be carried out with the help of CRISPRa. A modified version of the technique, which was already used in *S. cerevisiae*, was used again. First, Dong and colleagues tested several activator domains with different linker for use in *E. coli*. They found SoxS, an activator that achieved good activation performance when fused with MCP in combination with dCas9. Finally, the modified variant SoxS (R93A) was able to increase the expression even further. The only drawback of the technology is the narrow range of efficient activation. This was found to be between 70 and 80 bases on the template strand and between 80 and 90 bases on the non-template strand (relative to the transcriptional start site (TSS)) [83]. However, the basic design of CRISPRa technology is comparable to that for yeast. The gene of interest (GOI) is cloned downstream of a minimal promoter (BBa_J23117). Dong *et al.* placed a synthetic sequence (J1) in front of it, which presents a PAM every 10 bp, on both, sense and anti-sense strand. This sequence allows screening for activation on sense and anti-sense strand. Here, this artificial DNA part was replaced by the CPad allowing to address protospacers at different distances to the TSS (Fig. 3.14). A sfGFP was used as a reporter gene, which was integrated with the entire cassette into the genome of BW25113. Since the protospacers are not 10 but 30 bp separated compared to J1 sequence, the promoter-sfGFP cassette was cloned into one of the artificial protospacers (Pro_21) to optimize the distance between TSS and sgRNA target site. This was necessary because of the small range of efficient activation, which does not allow a larger margin. Details about the cloning procedure can be found in the Appendix (section A.3.6). Additionally, a second variant was cloned, where the expression cassette was placed directly downstream of one of the artificial protospacers (Pro_21), so that addressing of a directly adjacent protospacer could be investigated. However, the integration of this construct, where the cassette would have been in an optimal distance to the PAM, was not successful. As an alternative, a guide was designed that binds at a distance of 84 bp from the TSS on the template strand between Pro_2 and Pro_4. This guide was designed using Geneious Prime 2021 against the BW25113 genome and designated as guide_182 by the software (See Fig. 3.14).

The derived strains were cultivated and measured in the randomization worklfow (section A.3.8). The cultivation was first performed in different media to find a media with good growth and optimal measurement performance. For this purpose, two minimal media with different C-sources (MSM + 1 % glucose and MSM + 1 % fructose [257]) and the unfavored LB media were tested. LB was not favored because

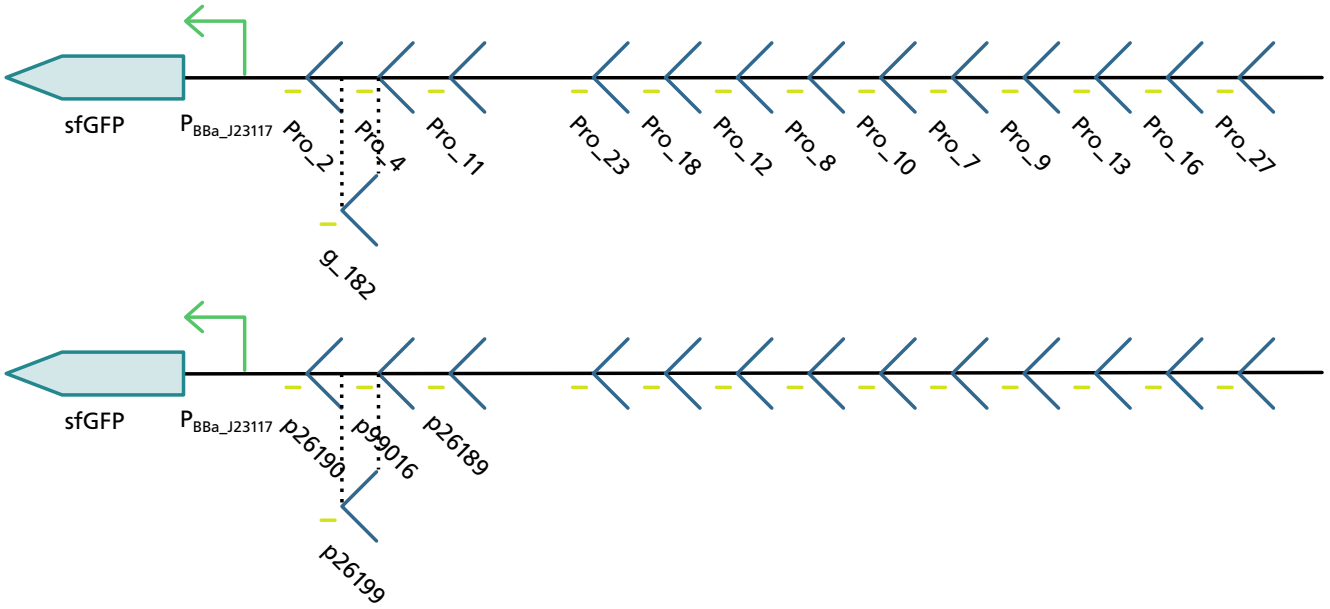


Figure 3.14.: Scheme of protospacer distribution upstream of minimal promoter BBa_J23117. Protospacer ID is given including corresponding plasmid number. guide_182 represents additionally designed sgRNA.

auto-fluorescence of LB was reported earlier [38].
Comparing the different media, there was no difference between strains s26902 (control without sfGFP) and s26903 (CPad Rank1 with sfGFP). In non-induced conditions, both strains reached an OD of ~2.1 in LB, ~3.5 in MSM with fructose and a maximum OD of ~2.75 in MSM with glucose (Fig. A.12). It was observed that the growth of the cultures induced with 1.5 % arabinose was restricted, with differences between the respective media. In LB media, the growth curve flattened early and remained at the same level and did not reach the OD of the controls without arabinose within the observation period. When using the minimal media with glucose, the situation was similar to LB, but here the increase in the induced cultures was slightly flatter than in the non-induced ones, but they finally reached a similar final OD. The slope of the growth curve for *E. coli* was much flatter in fructose media. Here, the induced cultures of three guide sequences (non-target, Pro_21, Pro_4) reached the same OD as the non-induced ones, but with a delay. Interestingly, for the uninduced cultures in fructose media, *E. coli* showed a flattening of the growth curve, followed by a further increase until finally entering the stationary phase. The induced cultures grew slower compared to the uninduced cultures, even compared to LB and MSM with glucose.
However, although the manufacturer specifies a range of 0 to 4 for the absorbance measurements, the precision drops by almost 1 % for an OD >2.0[3]. That means that the linear relationship is no longer ensured. Additionally, plate reader have the characteristic that they no longer provide reliable values at higher absorption values due to multiple scattering, so that the OD normalization could be biased [343]. For this reason, different glucose concentrations were tested allowing sufficient growth, but not allowing growth higher than 1.5 (OD).
Fig. A.13 shows growth experiments of the different strains with MSM minimal media with 0.05, 0.1 and 0.25 % (w/v) glucose as carbon source. The strains in 0.05 % (w/v) reached a maximum OD of ~0.3. A doubling of the glucose concentration also resulted in a doubling of the OD to ~0.6. A glucose concentration of 0.25 % (w/v) gave a maximum OD of ~1.4 in non-induced cultures. This OD was closest

[3] https://www.bmglabtech.com/de/pherastar-fsx/ (date of access: 05.12.2021)

to the target OD of 1.5, thus the MSM media with 0.25 % glucose was used for further experiments. In further experiments, it was observed that the growth behavior was also influenced by the addition of the inducer arabinose. This was particularly noticeable in the media with fructose as carbon source. Here, the growth curve had only a slight increase compared to glucose media and entered the stationary phase shortly before the end of the observation period of 48 h. In LB media, both conditions grew similarly but from an OD of ~1, the induced cultures entered the stationary phase, whereas the others continued to grow. The effect was least pronounced in MSM media with glucose. Here, the entry into the stationary phase was delayed by about five hours, but a similar final OD was achieved. At this point, it was noticeable that induction of the CRISPRa machinery had an effect on growth. Since growth was negatively affected compared to the uninduced control, different arabinose concentrations were tested. A description of the experimental procedure can be found in section A.3.8. The influence of the arabinose concentration on the growth of the different cultures is shown in Fig. A.14. The figure compares the control strain s26902 without sfGFP cassette and the strain with CPad and reporter cassette s26903. Both strains behaved very similar at different inducer concentrations. However, there were differences depending on the target of the sgRNA. For example, the growth of cultures targeting protospacer Pro_4 was little affected, whereas similar differences were seen with all other guide sequences, in the sense that growth was slightly negatively affected. Another exception was the additionally designed guide_182, whose influence on growth appeared to be noticeably higher when the CRISPRa machinery was induced.
However, the strains were also examined for their fluorescence values in the GFP channel. Fig. 3.15 shows a boxplot of the OD-normalized fluorescence values after 36 h. The timepoint was chosen because the growth process was quite different, especially in case of induced cultures, but here all cultures were already in the stationary phase. The figure shows the comparison of the normalized fluorescence values between the control without sfGFP (s26902) and the target strain with sfGFP (s26903). The plots are aligned with increasing distance to the TSS. The non-target control has no guide sequence and the off-target guide sequence targets another site outside the cassette.
It was noticeable that the values for non-target and off-target changed with higher arabinose concentrations ($\geq$ 0.25 % (w/v)), so that the sfGFP containing strain had higher values than the control strain. For Pro_2, there was a decrease in fluorescence values with increasing arabinose concentration. For the protospacers Pro_4 and Pro_11, only a small difference was observed. For the guide_182, a difference was expected, as it is located at the boundary of the region for which Dong *et al.* obtained a strong activation. Overall, the normalized fluorescence of the two induced guide_182 strains increased noticeably for arabinose concentrations $\geq$ 0.25 % (w/v) compared to the other cultures. However, it should be noted, that the difference between s26902 and s26903 with guide_182 was detectable but was smaller in comparison to the non-guide and off-target control. The drop in relative fluorescence in *E. coli* cultures containing the guide for Pro_2 was most noticeable. This already dropped clearly after induction with 0.05 % (w/v) arabinose. Overall, the relative fluorescence continued to increase after entering stationary phase, independent of strain and guide sequence. In most cases, s26902 and s26903 can be barely distinguished from each other. Only in the non-target control were the values slightly further apart as the other guide sequences (time resolved plot Fig. A.15). For a better comparison, the non-induced samples are also shown. These showed a very uniform course of fluorescence development, which, however, continued to increase steadily even after entry into the stationary phase. In GFP experiments with *E. coli*, it seems to be common practice to include a control without a sfGFP cassette (s26902) in order to subtract it as background from the otherwise identical strains [250][365]. This was also done for the presented data. But, this type of data evaluation did not provide any further insights (Fig. A.16).

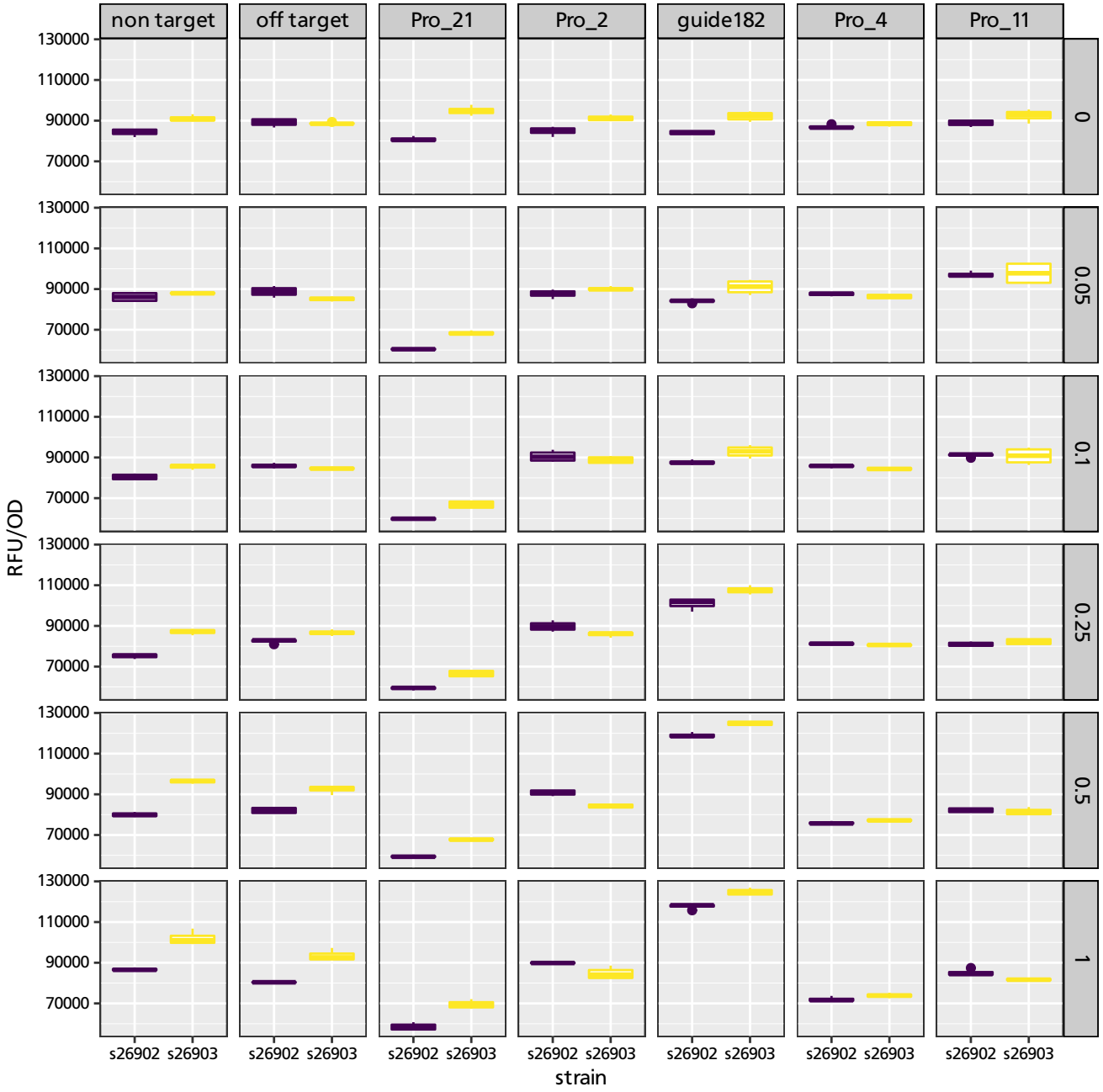
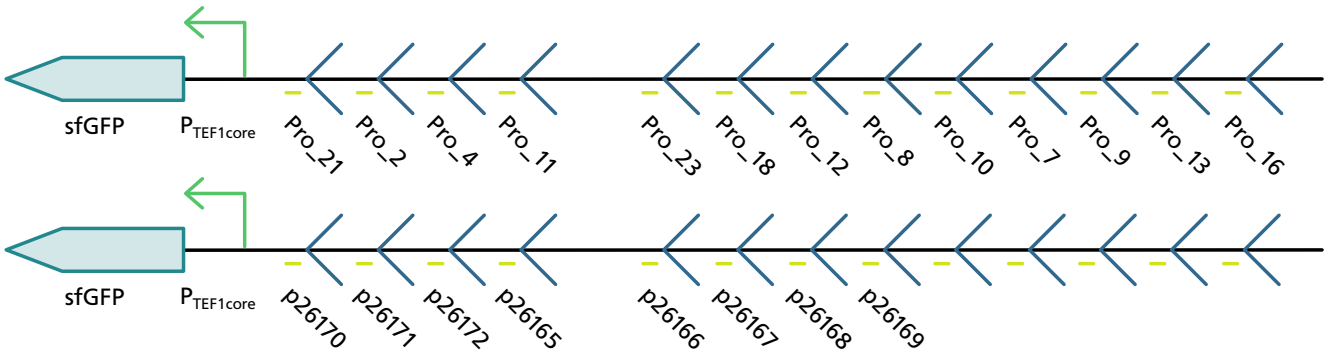


Figure 3.15.: The plot shows a boxplot of a cultivation with the strains s26902 (without sfGFP) and s26903 (with sfGFP) in the CompuGene Robotics platform. The chosen timepoint was after 36 h, when all strains reached stationary phase. The columns are aligned according to the distance to the TSS, and the rows are aligned with increasing arabinose concentration in % (w/v). Non-target control has no guide sequence, and off-target has a guide sequence outside the heterologous sequences. Every strain and concentration is represented by n = 4 samples.

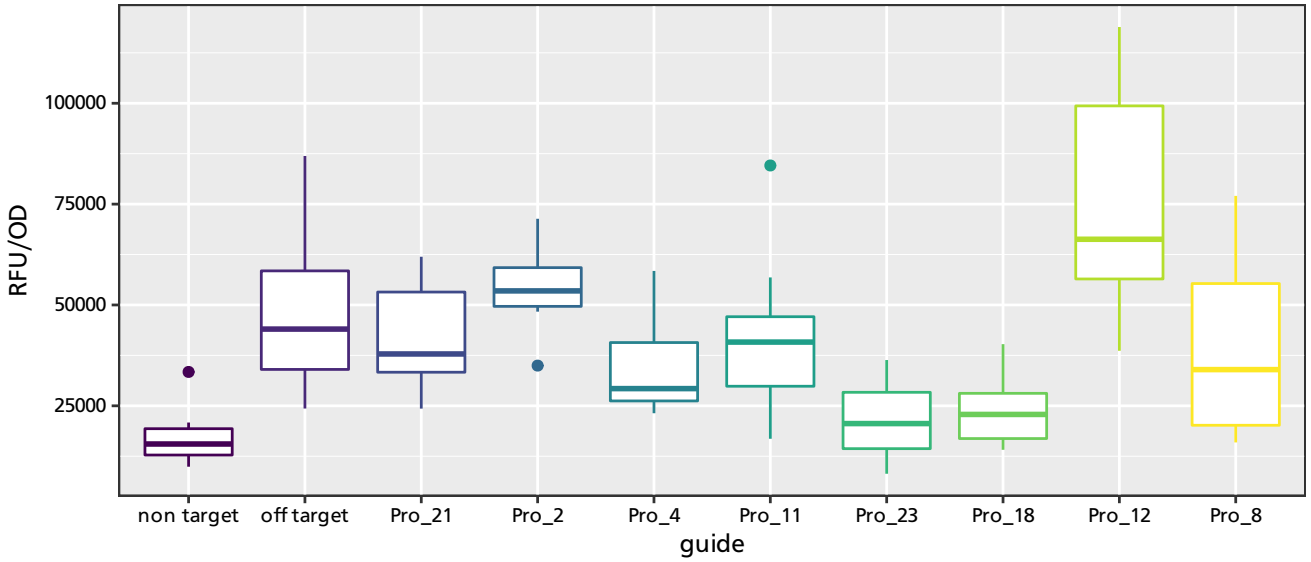
CRISPRa to drive expression of a GFP in *Y. lipolytica*

As described in section A.3.1, the CPads have been designed against the genome of multiple organisms. For *Y. lipolytica*, the strain W29 (CLIB89) was chosen. Like for *S. cerevisiae*, the target strain has no selectable auxotrophic marker to select for the components of the Proof-of-Concept system. The French wild type strain W29 has been modified by Madzak *et al.* [238] and Nicaud *et al.* [259] to improve the genetic accessibility as well as the ability for heterologous protein expression. First, the *URA3* gene was inactivated by insertion of the *SUC2* gene from *S. cerevisiae*. In the next step, *LEU2* was knocked out using a pop in/pop out event. Finally, AXP (acidic protease) was deleted, resulting in strain Po1f. Until today, a complete reference genome for Po1f is not available[4]. Although, a comparative study is not possible, the knockouts made by Madzak and Nicaud resulted in only minor changes in comparison to W29, leading to the decision to use the strain Po1f for the Proof-of-Concept, as it is a direct descendant of W29 and has selectable markers.
Until today, a comparative study to evaluated different loci for protein expression like in *S. cerevisiae* is not available yet. The decision was made to integrate the CPad with the reporter protein into the *URA3* locus because it was used earlier for chromosomal integration. But, the specific integration into one locus is challenging, since *Y. lipolytica* prefers NHEJ over HR for DNA strand repair. Thus, a knockout strain was created first. In the past, a deletion of *ku70* showed a shift towards HR [370]. Details about knockout strain creation can be found in section A.3.4.
Again, a CRISPRa system was used to show targeting of different protospacers within the CPad. In this case, the system of Schwartz *et al.* was used [319]. Again, a viral activation domain was used to recruit the DNA transcription machinery. The system contains a $P_{TEF1core}$ promoter, which is good characterized and has a rather low basal activity in absence of a UAS. Additionally, a GAL1 UAS from *S. cerevisiae* was placed upstream of the $P_{TEF1core}$ promoter by Schwartz *et al.*, which has no homologous counterpart in *Y. lipolytica*. The GAL1 UAS was targeted with the CRISPRa system. Here, contrary to Zalatan *et al.*[401], the activation domain was fused directly to dCas9 rather than to an additional RNA binding protein. The authors tested different activation domains. Again, the transcriptional activators VP16 and VP64 from the herpes simplex virus were tested. Additionally, VPR, a fusion of VP64, p65 and Rta, was used which was described earlier [58]. The ability for activation was shown in human HEK 293T cells, *Drosophila melanogaster* S2R+ cells and *Mus musculus* Neuro-2A cells, but also in *S. cerevisiae*. In the latter, up to 78-fold higher relative RNA expression was achieved compared to the non-target control. For the original VP64 domain, a 14-fold higher RNA expression was measured [58]. However, the VPR was part of the pool tested by Schwartz and colleagues and showed a significantly higher mean fluorescence as VP16 and VP64, respectively. Thus, the dCas9-VPR fusion was used to show CRISPRa with the CPad. The entire CRISPRa system is plasmid-based and does not require further integration of components into the genome, such as dCas9-VPR and sgRNA, as it is the case in Zalatan *et al.*[401] Only, the artificial target, consisting of the reporter $P_{TEF1core}$-sfGFP which was pre-integrated into the CPad (Fig. 3.16a), needs to be integrated into the genome. Here, the CPad replaces the function of the GAL1 UAS in Schwartz *et al.*[319], which is used to direct the transcription machinery to the binding sites within the $P_{TEF1core}$ promoter.
The strain development is described in detail in section A.3.4. The strain *Y. lipolytica* Po1f s26241 was used for transformation with multiple plasmids containing different guide sequence to address their respective protospacer within the CPad. The derived yeast strains were used for a cultivation in the CompuGene Robotics platform (section A.3.5). The data import as well as the evaluation was done by an automated R pipeline. Furthermore, the script assigns growth and non-growth depending on a threshold for the optical density, which was set to OD > 0.1. From this experiment, no cultures were discarded due to insufficient growth. The boxplot in Fig. 3.16b shows the OD-normalized fluorescence values measured by the plate

[4]NCBI Genome Browser: 05. May 2021



(a) Top shows the distribution of protospacer upstream of $P_{TEF1core}$ promoter as well as the plasmid IDs (bottom) containing the guide sequence to address the different protospacer.



(b) Boxplot shows the OD-normalized RFU values after 42 h of the cultivation. x-axis indicates the different guide plasmids.

Figure 3.16.: Protospacer distribution and normalized fluorescence values for CRISPRa in *Y. lipolytica*. The RFU was normalized to the OD values. Non target control (p26120) contained only scaffold RNA without guide sequence. Off target control (p26173) has a guide sequence for the *Ku70* gene. Samples are represented by n=8.

reader after 42 h. Since no sample was dropped, all graphs are represented by n = 8 samples. Additionally, the normalized data were plotted over the time (Fig. A.21). The x-axis of the boxplot shows the protospacer which was addressed by transformation of *Y. lipolytica* with the respective sgRNA plasmid. The experiment was done with two negative controls. While plasmid p26120 did not contain a guide sequence, p26173 had a guide sequence for the gene of Ku70, which most likely has no target within the heterologous sequences of the CPad as well as in the integrated reporter cassette. Compared to the non-target negative control without guide sequence, Pro_23 and Pro_18 showed no difference. But, the other guides showed a higher relative fluorescence compared to the non-target control. The guide sequence with the highest RFU/OD values was p26168 addressing Pro_12 which is located 210 bp upstream of the $P_{TEF1core}$ promoter. Contrary to *S. cerevisiae*, there was no clear gradient of activation. The activation pattern looked more like a wave, not like a clear gradient.
For comparison, the fold change was calculated and plotted over the time (Fig. A.22). The fold change was calculated against the non-target negative control. Comparing all guides, the maximum fold change was ~3.4 and was reached after 42 h for Pro_12. After 42 h, the fold change reached a plateau and started to decrease as the cultures entered the stationary phase (development of the OD is shown in Fig. A.20).

CRISPRi to drive expression of a GFP in *B. subtilis*

As a bacterial host with high industrial relevance, *B. subtilis* was part of the pool of organisms. Furthermore, *B. subtilis* is a model organism for sporulation and is widely used for genetic as well as metabolic engineering.
The CPads were rendered against the genome of the strain PY79. One of the advantages of this strain is the lack of auxotrophies, which thus makes the strain interesting for large-scale cultivation. The evaluation of the CPads functionality was the topic of the bachelor thesis of Georg Schmidt [317]. For the other organisms, CRISPRa was used to demonstrate the possibility to address different protospacer with the aim to tune the gene expression. Until the beginning of Georg Schmidt's bachelor thesis, a CRISPRa system for *B. subtilis* was not available to have a system which is comparable to the other organisms above. But as an alternative system, CRISPRi was successfully demonstrated for *B. subtilis* [283]. The system was used for the functional analysis of essential genes in *B. subtilis* and has been described earlier [292]. CRISPRi was shown to be highly efficient, since an up to 1000-fold repression was shown by the authors. In comparison to knockout experiments, the CRISPRi system allows a fast and efficient switching of gene expression.
The system consists of a dCas9 whose expression is controlled by a xylose inducible promoter (P_{xyl}) and a constitutively expressed guide RNA (P_{veg}). The design of Peters *et al.* [283] required a genomic integration of the sgRNA expression module and of dCas9. The target for the dCas9 should again be the CPad which contained again a reporter protein. The reporter sfGFP was controlled by different constitutive promoters. Two integration cassettes were prepared, containing the CPad Rank4 with pre-integrated reporter and either P_{43} or P_{ylb}. The reporter cassette was cloned behind Pro_21 in the center of the CPad as shown in Fig. 3.17. In total, eight strains were developed. Per promoter, four strains were constructed, allowing to address different protospacer upstream of each constitutive promoter. All CRISPRi components have been cloned into the genome of *B. subtilis* PY79 by Georg Schmidt (genotypes can be found in Tab. 2.7).
The constructed strains were used for a cultivation in the CompuGene Robotics platform using the randomization workflow with induction (section A.3.8). Since the dCas9 is controlled by a xylose inducible promoter, two microtiter plates were used, whereby one was induced with water as control and the other one was induced with 1 % (w/v) D-xylose. Initially, 0.3 % (w/v) D-glucose was used as carbon source for the defined MSM media [257]. The use of D-glucose as sole carbon source was based on the basic MSM media formulation of Neubauer *et al.* [257] For the first cultivation, it was observed that the use of D-glucose as carbon source led to a second exponential growth phase (diauxic growth) (Fig. A.23). This

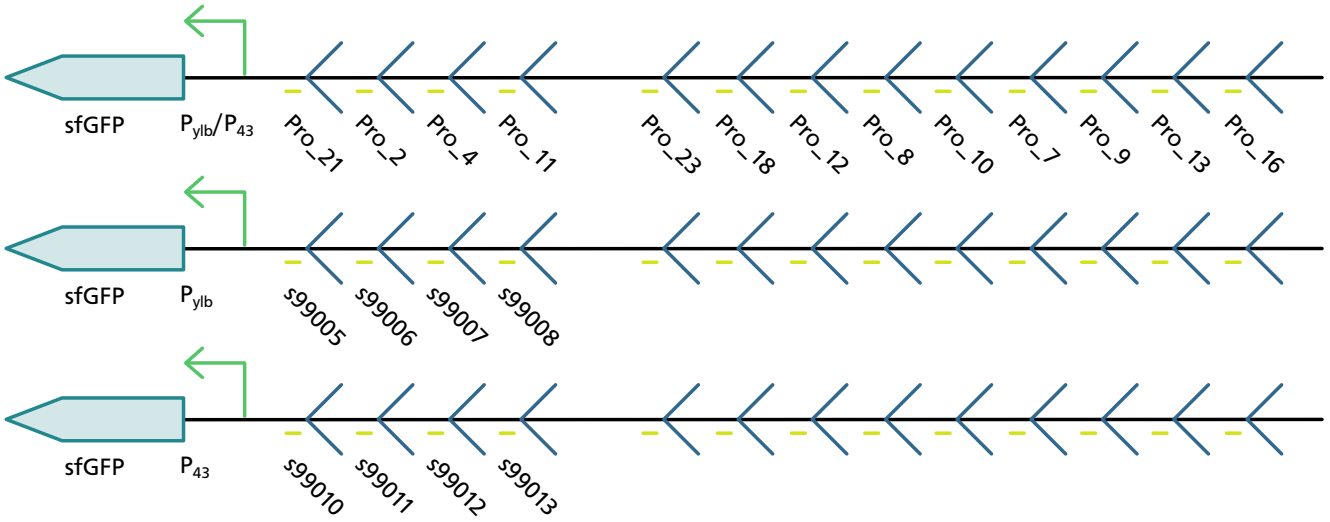


Figure 3.17.: Top shows the distribution of protospacer upstream of P_{ylb} and P_{43} respectively as well as the strain IDs (bottom) containing the guide sequence to address the different protospacer. All strains were constructed by Georg Schmidt (bachelor thesis) [317].

behavior was clearly shown when comparing the uninduced control and the induced cultures. In addition, both xylose inducible promoters which are commonly used in *B. subtilis* (P_{xylA} from *B. subtilis* and the P_{xyl} from *Bacillus megaterium*) are prone to catabolite repression [230]. Consequently, the strains constructed by Georg Schmidt were further studied using a different carbon source, which should not affect the xylose inducible promoter driving the expression of dCas9. In a pre-experiment (not shown here), the utilization of D-fructose seemed to be promising and was used for the further characterization of the constructed strains.

As shown in Fig. A.24, the different strains grew synchronously after inoculation until the end of the exponential phase. The used constitutive promoters showed differences in OD-normalized fluorescence values, independent of CRISPRi. For example, the P_{43} promoter showed an increase in fluorescence already at the beginning of the exponential phase (Fig. A.25). The P_{ylb} promoter, which was also used, showed very low activity until transition into the stationary phase, which then increased rapidly.

Addressing individual protospacers in the CPad was as different as the growth phase behavior of the promoters. At the beginning, the values still fluctuated clearly for both promoters, as shown in the time resolved plot in Fig. A.25. In case of P_{43}, until transition into the stationary phase, only protospacer Pro_21 showed a visible reduction in fluorescence compared to the non-target control s99009. After entering the stationary phase, also another protospacer (Pro_11) showed lower fluorescence as the control s99009 (Fig. 3.18). No difference was observed for Pro_4 when compared to the control. Pro_2 seemed to have a positive effect on the expression of the reporter gene. The pattern was comparable between induced and uninduced condition.

Looking at the strains for P_{ylb}, all strains initially showed the previously described effect that fluorescence increased with transition to stationary phase. The transition to stationary phase occurred after approximately 24 h (Fig. A.24). For this promoter, the two protospacers Pro_4 and Pro_11 seemed to have an effect on fluorescence when addressed with an inactive Cas9 (Fig. 3.19). In addition, Pro_11 is the protospacer with the largest tested distance to the start codon, and it seemed to have the biggest negative influence on expression. Induction with xylose affected all addressed protospacer, but the highest impact was observed for Pro_11. The ratio between the other protospacers remained largely unaffected.

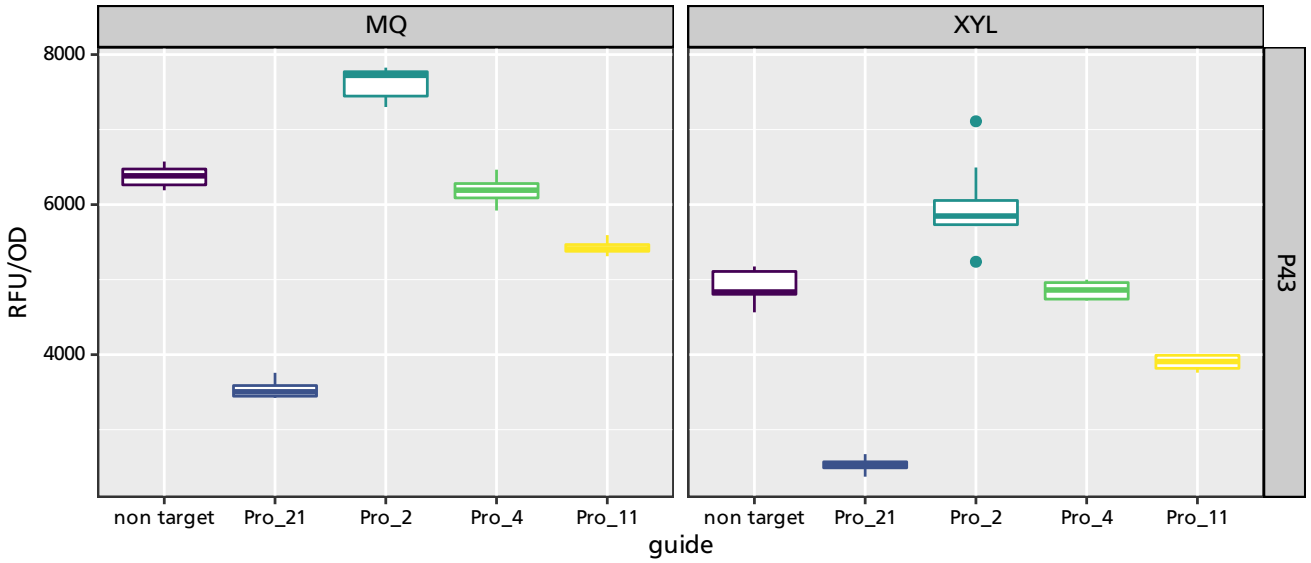


Figure 3.18.: Normalized fluorescence for CRISPRi in *B. subtilis* PY79 for P_{43}. The developed strains [317] have been cultivated in the CompuGene Robotics platform using the randomization workflow. The culture in MSM media were induced directly after randomization procedure with 1 % (w/v) D-xylose. Inducer: MQ - ddH_2O, XYL - 1 % D-xylose. The ribbon indicates standard deviation of n=8 samples. Timepoint: 36 h

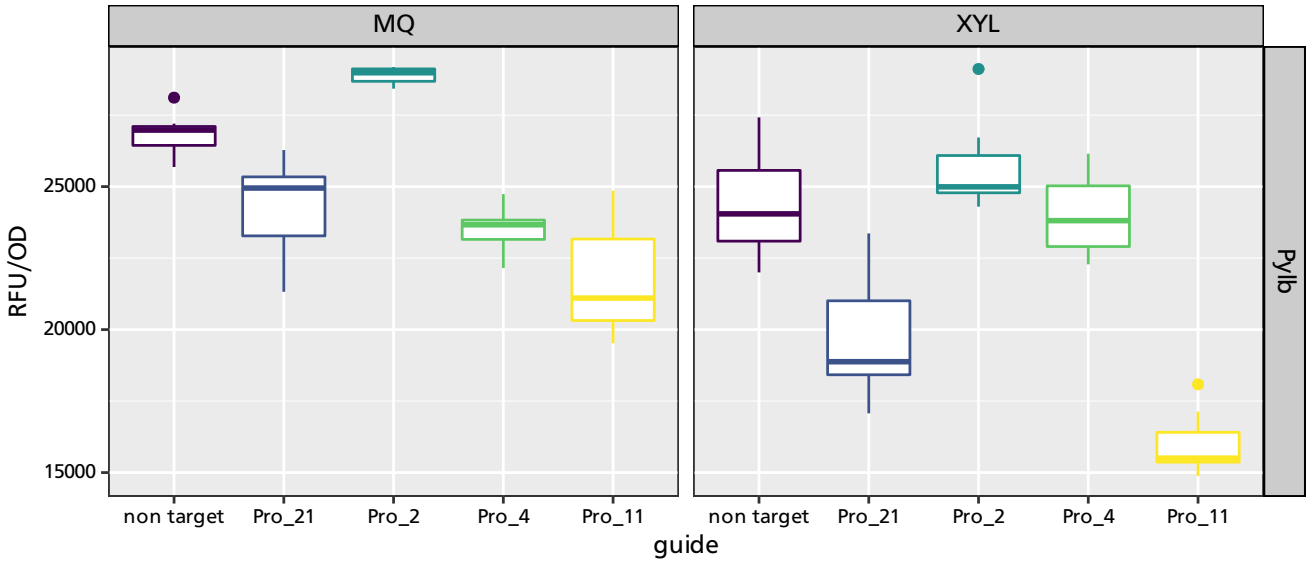


Figure 3.19.: Normalized fluorescence for CRISPRi in *B. subtilis* PY79 for P_{ylb}. The developed strains [317] have been cultivated in the CompuGene Robotics platform using the randomization workflow [317]. The culture in MSM media were induced directly after randomization procedure with 1 % (w/v) D-xylose. Inducer: MQ - ddH_2O, XYL - 1 % D-xylose. The ribbon indicates standard deviation of n=8 samples. Timepoint: 36 h

3.4. Discussion

3.4.1. FLEXpress

The advantages of using a single vector for ”multi-organism” expression were already outlined in section 3.2.1. The created FLEXpress backbone was successfully cloned and used for transformation of two yeasts (*S. cerevisiae*, *Y. lipolytica*) and one bacterium (*E. coli*). The performed cultivation with the mTFP containing vector, as well as with the empty control vector showed that all tested organisms were capable of propagation. Both, a vector, and parts, which can be used in multiple organisms, are particularly useful for metabolic engineering and SynBio. To this end, the FLEXpress vector was equipped with a reporter gene and a constitutive promoter. The promoter was derived from the KanMX marker developed by Wach *et al.*[373] Here, the promoter drives the expression of the resistance gene of the dominant marker. The marker confers resistance to G418 in *S. cerevisiae* and *Y. lipolytica* as well as to Kanamycin in *E. coli*. Thus, the promoter from *Ashbya gossypii* should be active in the aforementioned organisms and can be used for heterologous expression. To provide a new part for multi-platform usage, the promoter was investigated using the FLEXpress vector in *E. coli*, *S. cerevisiae*, and *E. coli*. To this end, the P_{TEFag} was cloned into the FLEXpress vector together with the reporter mTFP (p26062) and cultured together with the empty vector (p26061).
Unfortunately, the data suggest, that the used promoter can be employed in both yeasts, whereas the expression in *E. coli* was rather low. However, the utilization always depends on the scenario, e.g., a low protein expression rate is required. The promoter behaves very differently in the tested organisms. For example, in *E. coli*, it appears to be almost inactive, as shown in Fig. 3.6. However, there is a significant difference between the expression vector and the empty vector. However, the activity is apparently sufficient to confer resistance to Kanamycin, when used in the KanMX marker. Thus, the characteristics of the promoter in *E. coli* are fundamentally different from those in the two yeasts. Regardless of the low activity of the promoter, it could be used in a pathway, where fine-tuning of the expression levels of several proteins is required. For example, the P_{TEFag} could be used for a protein, which is known to cause the formation of inclusion bodies through high expression rates. The balancing of heterologous pathways can be highly beneficial for the obtained yield, since it prevents the cell from unnecessary metabolic burden [181] and accumulation of potentially toxic intermediates [62][285]. In addition, it should be noted that the lack of codon optimization of the mTFP CDS may have resulted in lower fluorescence. The used mTFP was originally optimized for the use in *Y. lipolytica*. Analysis with the *E. coli* Codon Usage Analyzer 2.1 developed by Morris Maduro[5] revealed several problematic codons that are rarely used in *E. coli*. The codons which have been identified as problematic are marked red in Fig. A.3. Codon optimization has a non-negligible impact on the expression of heterologous genes, as shown earlier [48].
However, differences in fluorescence values were also apparent in both tested yeasts. Moreover, the P_{TEFag} had high fluorescence values, suggesting high expression of the mTFP reporter. Thus, it could be used for protein expression in *S. cerevisiae*, as already done by Edwards and Wandless [90]. In addition, in *Y. lipolytica*, the activity of the promoter was already known from experiments of Thomas Hofmeyer (AG Kabisch), who compared various endogenous promoters of *Y. lipolytica*.
However, the differences mentioned in section 3.3.1 are seemingly dependent on the growth phase of the yeasts during cultivation. This can be seen in the normalized fluorescence data (Fig. 3.6), but also in the fold change values (Fig. A.1). *S. cerevisiae* showed two peaks in the fold change values, one at the beginning and a smaller one at the end of the exponential phase, whereas *Y. lipolytica* showed a strong increase after entering the exponential phase. This continued until the growth curve began to flatten and the fold change dropped sharply. In the past, it was assumed that the endogenous P_{TEF1} promoter

[5] https://faculty.ucr.edu/~mmaduro/codonusage/usage.htm (date of access: 26.06.2021)

of *S. cerevisiae* allows stable, constitutive expression [348]. But, the P_{TEF} promoters are also subject to fluctuations as shown by Peng *et al.* [279]. It seemed to be the case for the heterologous P_{TEFag} from *A. gossypii*. The presented data illustrate, that individual values for a certain timepoint are not meaningful. Furthermore, the characterization of promoters in terms of growth phases and environmental conditions (like media, pH, carbon source etc.) with temporal resolution appears to be useful, which was done for several yeast promoters [296].
However, growth phase dependence of promoters is not necessarily disadvantageous. For example, they can be used for biotechnological applications. For example, the synthesis of products can take place in two phases. In the first phase, the main focus is on growth, and in the second phase, on production, so that growth and production are not in competition with each other [47][385]. Such approaches also exist using synthetic biology, whose tools are able to dynamically regulate the metabolism [47].

3.4.2. Tetracycline aptamer in *Y. lipolytica*

Cross-platform vectors and promoters are valuable parts towards orthogonal designs in the DBTL cycle of SynBio. In addition to use promoters for conditional gene expression at transcriptional level, riboswitches can add an extra layer of control. Furthermore, different types of riboswitches can be used at the level of translation, allowing regulation on an additional level.
An example of such a riboregulator is the Tetracycline aptamer of Berens *et al.*[23] which was described above. The design by Süß *et al.*[346] successfully used the TCapt in the 5'UTR of yeast mRNAs, whose structure is stabilized at RNA level by binding of the ligand Tetracycline. As a result, translation is no longer carried out, so that gene expression can be switched off conditionally. A further development was carried out by Kötter *et al.*[204] aiming to strengthen the switching effect by "series switching" of the TCapt.
The aptamers were to be tested with regarding their orthogonality both in the model organism *S. cerevisiae* and in the non-conventional yeast *Y. lipolytica*. For this purpose, test constructs were cloned in FLEXpress background to be used for the transformation of the two yeasts. There were no problems during the transformation procedure of *S. cerevisiae* with the five constructs (empty vector, positive control, as well as monomer, dimer, and trimer of the TCapt). Clones were obtained for all constructs.
Unfortunately, the transformation of *Y. lipolytica* with TCapt-dimer and TCapt-trimer was not successful. Furthermore, even for the construct with only one TCapt, there was a severe impairment of the growth of *Y. lipolytica*. Since the only difference between the constructs was the integration of the Tetracycline aptamer, it is likely that the aptamer or its DNA sequence is a potential impairment for the yeast. In addition, this effect was observed on plates as well as in cultivation experiments. The cultures, that carried the aptamer, grew noticeably slower, compared to the controls. In addition, their morphology also changed. They look clearly rounder and larger than the wild type, positive control and negative control. This observation was supported by the change in front and side scatter in the cytometer during cultivation in shake flasks (Fig. A.7).
For some time, it is known that aptamers are involved in the regulation of expression [40][354]. They naturally occur in eukaryotes as well as in prokaryotes. However, due to their nature, they can form strong secondary structures. Therefore, it is not surprising when the DNA, that encodes an aptamer, can also constitute strong secondary structures. Since the cells seem to have problems with proliferation in this case, it is reasonable to conclude that the aptamer influences the replication of the plasmid. For some time, it has been known that DNA secondary structures have an influence on replication and translation [353]. During replication and transcription, the duplex is unwound and separated so that the DNA is single-stranded and can form secondary structures such as stem-loops, hairpins and G4 structures [353]. Secondary structures have various physiological functions, but also have an influence on genome stability

[31][73]. But they can also ensure that replication slows down in regions of high transcription [73][220]. There are a few ways to overcome fork barriers. First, helicases are used to resolve secondary structures. Secondly, the fork is stabilized, and the blockage is processed by DNA repair mechanisms. If the blockage resolution is unsuccessful, it can also lead to replication termination and restart. Thus, replication stalling can cause errors during replication [31][73]. Especially G4 structures seem to be highly conserved and involved in the regulation of transcription and replication but also in telomere maintenance [73][313]. An analysis of the aptamer sequence shows the presence of a G4 structure[6]. The results for one, two, and three aptamers can be found in the appendix (Fig. A.8, Fig. A.9 and Fig. A.10). Nevertheless, the cells must be able to break up G4 structures. This is done by various helicases that have been described for both prokaryotes and eukaryotes [31][73][327]. In *S. cerevisiae*, PifI seems to be specialized in breakdown of G4 structures on the lagging strand during DNA synthesis [73][273]. Proteins of the PifI family have also been found in other eukaryotes and prokaryotes [30]. PifI has been associated with replication problems in deletion experiments. In particular, there was a slowing of replication. Other helicases seem to be able to compensate for this task on the leading strand, but not on the lagging strand [73]. For *S. cerevisiae*, other helicases were also identified that can resolve G4 structures (Sgs1, Hrq1) [273][272]. On the DNA level, no homologous genes could be found using blastn on default settings in Geneious (Local BLAST DB for *Yarrowia lipolytica* W29[7]). Also searching with BLASTp (Standard Settings; Search Set: Organism: taxid 284591), there were no noteworthy matches for *HRG1*, *PIF1* and *SGS1* (protein sequences from yeastgenome.org). However, this does not mean that there are no helicases of the Pif1 family in *Y. lipolytica*, because at least there are some hits that could be associated with a helicase family. However, this could be a reason for impaired growth of *Y. lipolytica* with the TCapt and is maybe subject for further studies.
Nevertheless, the results of Kötter *et al.*[204] were confirmed for a single insertion of the aptamer. Since the focus was on porting the aptamer switches to *Y. lipolytica*, the dimer and the trimer were not investigated further in *S. cerevisiae* because there were no corresponding clones for in *Y. lipolytica*. Interestingly, *Y. lipolytica* with one aptamer had a higher relative fluorescence (compared to the positive control). This is possibly due to the inhibition of growth while the constitutive P_{TEFag} still produces mTFP, maybe causing the fluorescence protein to accumulate, resulting in a higher fluorescence per cell.

3.4.3. CRISPRpads in two prokaryotes and two eukaryotes

The use of artificial landing pads offers many advantages, such as identical DNA context in the genome. These sequences are known and therefore do not require a fully sequenced genomic sequence. The integration of landing pads into genomic regions, which have only contig resolution, would be conceivable – albeit with caution. Additionally, they offer a certain standardization of DNA parts. Depending on the design and integration locus, they would be orthogonal to the host and in addition, would not affect it. Landing pads have already been developed for diverse applications [35][206][275]. The CPads were specifically designed to be used with the nuclease Cas9 from *S. pyogenes*. They were designed to be artificial DNA which does not influence the host. In this context, the pads were rendered against the genome of four industrially relevant organisms – *B. subtilis*, *E. coli*, *S. cerevisiae* and *Y. lipolytica*. In particular, the yeast *Y. lipolytica* and the bacterium *E. coli* are classically difficult to modify at genomic level. Once integrated, the pads should simplify modification using an active Cas9 and increase transformation efficiency. However, cloning efficiency was not the only focus for the CPads, but also a streamlined cloning procedure for all four organisms. Furthermore, the CPads should allow for the use of the same guide RNAs and homology sites in all four organisms, and thus reduce the cloning effort per organism. Today, conditional gene expression using dCas9 is also available, allowing to use the CPads for standardized CRISPRi and CRISPRa applications.

[6] https://bioinformatics.ramapo.edu/QGRS/index.php
[7] *Y. lipolytica* W29 (BioProject: PRJNA295780); BLAST-version: ncbi-blast-2.11.0+-x64-linux

For example, fine-tuned gene expression in multigene pathways in *Y. lipolytica* is still challenging because the number of inducible promoters is rather limited [360]. In general, many promoters are inducible but not titratable, making targeted pathway control difficult, which can lead, for example, to the accumulation of potentially toxic intermediates [105]. In addition, effects such as catabolite repression effects, like the P_{xylA} promoter in *B. subtilis*, have to be considered for promoter selection [230]. Thus, CRISPRa or CRISPRi could allow for host-independent gene regulation.
In addition, imbalances might be related to gene-dose effects. Thus, integration directly into the genome – independent of the landing pads – is a conceivable possibility to eliminate gene-dose effects. When using plasmids, growth phases and environmental changes can have a severe influence on copy numbers and thus on gene dose [320][324]. Many of the aforementioned problems have been addressed with the CPads. To this end, targeted gene expression using CRISPRa and CRISPRi has been studied in this thesis.
These two methods are convenient for studying the targeting of different protospacers in the CPad. Different sgRNAs should be used, ideally leading to a stepwise activation of reporter expression. In other words, with increasing distance to the promoter, a decrease in reporter gene expression was expected for CRISPRa. In contrast, CRISPRi-induced repression should be less effective with increasing distance to the promoter. First, the utilization in *S. cerevisiae* and *Y. lipolytica* was investigated. As illustrated in Fig. 3.13b, the sfGFP reporter gene expression decreases with increasing distance in *S. cerevisiae*. Furthermore, the flexibility of the CPad became noticeable, as the position of Pro_11 was apparently too close to the $P_{CYC1min}$ promoter, resulting in an CRISPRi effect. This circumstance was particularly visible in comparison to the non-target and off-target controls. Here, however, it can also be observed that it is not a clear gradient, but more of a staircase-like decrease. The maximum activation was reached at 90 bp upstream of the minimal promoter or 229 bp upstream of the start codon. This is also consistent with similar studies using a VP64- and a VPR-dCas9 fusion that showed activation at 277 bp and 351 bp upstream of the TSS [52]. Probably, the start codon was meant by the authors, as the TSS is difficult to determine. Additionally, *S. cerevisiae* can use multiple TSSs within a promoter [293]. The maximum in this area is well conceivable, since Harbison *et al.* found a high density of transcription factor binding sites in a region around ~200 bp upstream of the start codon in yeast promoters [145]. On the other hand, the graded activation in the oleaginous yeast *Y. lipolytica* was not as clear as it was in *S. cerevisiae*.
The non-target control was taken as baseline, so that the protospacers Pro_23 and Pro_18 showed almost no activation(Fig. 3.16b). Clear activation was seen when addressing Pro_12, which is located 210 bp upstream of the $P_{TEF1core}$ promoter (or 322 bp from the start codon). The original design of Schwartz *et al.* had a maximum expression approximately 160 bp upstream of the start codon [319]. This would correspond to a distance which would be between Pro_2 and Pro_4 in the CPad in *Y. lipolytica*. Thus, the values for Pro_2 suggest activation of expression. However, the limiting factor in Schwartz *et al*. is a short range of only 235 bp, which is covered in the $P_{TEF1core}$-GAL1-UAS construction. However, a range of 350 bp (Pro_8) was examined with the CPad, showing fluorescence above baseline. But, further discussion of the activation efficiency, depending on the distance to the promoter of the two yeasts, is not meaningful at this point, since both yeasts clearly differ in terms of regulatory features. Moreover, the normalized values showed clear fluctuations, caused by OD normalization. Unfortunately, this is due to the formation of hyphae and pseudo-hyphae of the *Y. lipolytica* Po1f strain (confirmed by microscopy), which was described earlier [22]. The formation of hyphae substantially affects the OD measurement and makes it less accurate due to the formation of agglomerates. The results for *B. subtilis* are probably more difficult to interpret, as different results were obtained depending on the promoter.
A reduction of fluorescence directly upstream of the promoter was expected, since CRISPRi should interfere with the assembly of the RNA polymerase complex. Furthermore, the interference effect should decrease with increasing distance [6]. Two different constitutive promoters were studied in *B. subtilis*, which also differ in length (P_{43}: 78 bp; P_{ylb}: 103 bp), having probably an effect on CRISPRi efficiency. For P_{43}, Pro_21

showed the strongest interference (Fig. 3.18) whereas for P_{ylb}, the strongest interference was achieved for Pro_11 (Fig. 3.19). Interestingly, upstream of P_{43}, Pro_2 was able to enhance fluorescence. Perhaps a facilitated access of the RNA polymerase complex would be conceivable here, since the DNA double strand is already opened and destabilized by the infiltrating dCas9 complex. For CRISPRi in bacteria, a good guideline for sgRNA design is usually a target around the -35 and -10 box, as well as in the first third of the coding DNA sequence (CDS) [292]. Seemingly, binding of dCas9 can have an impact on gene expression further away from the TSS. Perhaps dCas9 perturbs the binding of regulatory proteins near the promoter, causing an impact on expression. An example would be the lac repressor in *E. coli*, which enhances repression in a DNA loop structure [66]. However, in *B. subtilis*, there were also differences after induction with ddH_2O as control. As expected, overall normalized fluorescence reduced depending on the used protospacer. Furthermore, the pattern of normalized fluorescence were comparable between induced and uninduced cultures. This is most likely due to the leaky P_{xyl} promoter, controlling the CRISPRi component dCas9. As result, the CRISPRa system would be active even without induction.
In *E. coli*, however, there were no clear answers when using the CPads with CRISPRa. In previous observations, the non-target control was always used as reference. But, there was an increased fluorescence for the induced condition as well as for the control condition. Unfortunately, this difference was also observed for the non-GFP strain s26902 (Fig. A.15). In addition, there is just a narrow range of activation, which was already observed in Dong *et al.* [83]. The optimal distance was difficult to realize in the current version of the CPad due to large distances between the artificial protospacers. Moreover, the additional guide_182 did not produce detectable activation. Especially in comparison to the controls, the difference between GFP and non-GFP strain was much smaller than for the controls. This effect suggests other reasons for the differences in the fluorescence values. In various experiments, it has been shown that some *E. coli* strains tend to develop both intracellular and extracellular autofluorescence [250]. Mihalcescu *et al.* were able to show that flavins accumulate in the culture medium over time, emitting at a wavelength at which is commonly used for GFP measurements. The degradation rate is rather low, causing an accumulation of such substances in the media. As a result, the detection range for GFP has a lower limit for some *E. coli* strains [250]. Additionally, the stress response to selection markers, such as Ampicillin, can enhance autofluorescence [298][349]. Often, autofluorescence is not a problem because medium and high copy plasmids are used, which provide a sufficiently high gene dose, and therefore background is negligible [24]. Since the CPad is a single integration, it might be the case that even with optimal activation, the gene dose is not sufficient to overcome the autofluorescence background.
Some technical methods have been developed to overcome autofluorescence [183][223]. The manufacturer of the PHERAstar FSX[8] suggests fluorescence polarization measurements to correct for autofluorescence. But these did not succeed in several experiments and may only be suitable for adherent cells. As an alternative, selected samples were taken after cultivation and measured in the Sony SH800SA to remove the autofluorescence which is caused by secreted metabolites and components of the media. Unfortunately, there was a negligible difference between GFP (s26903) and non GFP (s26902) strain (Fig. 3.20). Again, the difference was biggest for the non-target control condition. The GFP signal might be masked by autofluorescence, or that the CRISPRa system does not function optimally under the tested conditions, especially because guide_182 is at the edge of efficient activation range [83]. Unfortunately, no suitable PAM was available to further optimize the distance.
Although the gene dose of only one copy probably has a negative effect on the measurements in *E. coli*, it offers an overall benefit that should not be neglected. It is often difficult to estimate expression levels correctly, as they vary depending on many factors, such as growth rate, gene dose – which can vary significantly for plasmids in different growth phases – or environmental conditions [111][265][320][380]. However, to improve the readout by fluorescence measurements, an exchange of the reporter gene could

[8] https://www.bmglabtech.com/de/fluorescence-polarization/ (date of access: 10.01.2022)

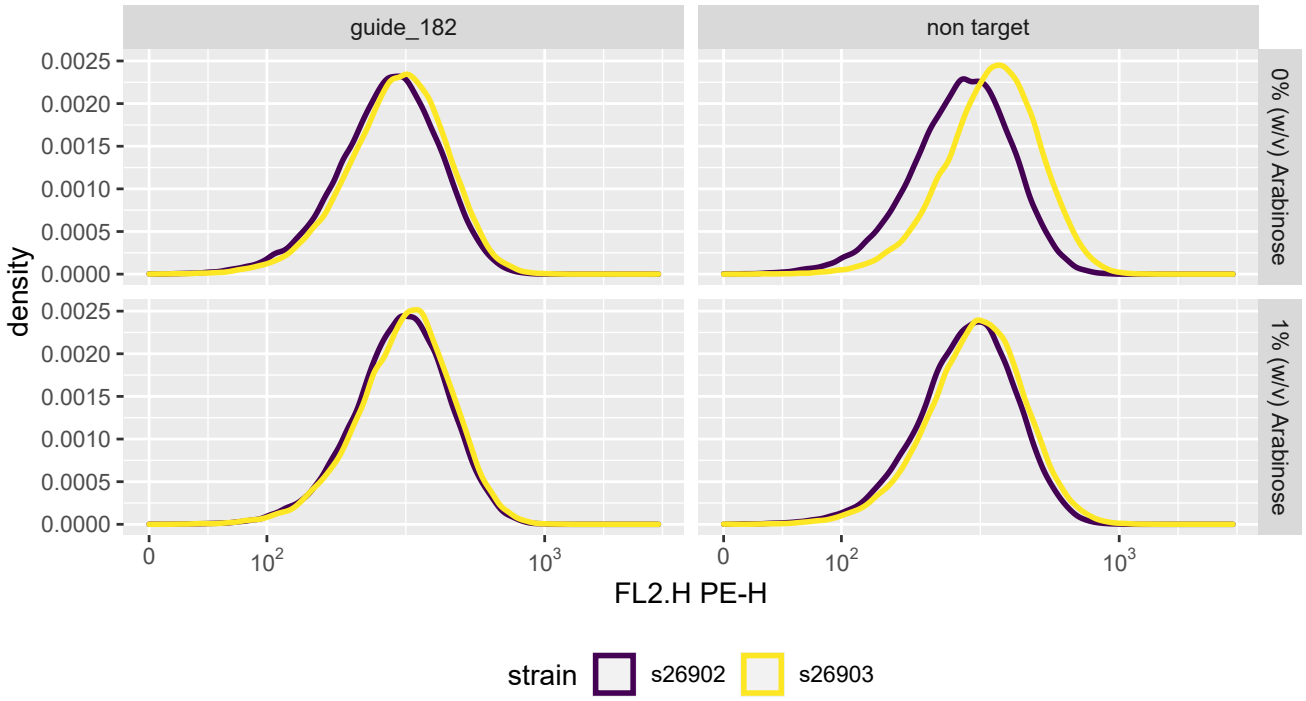


Figure 3.20.: Density plots of *E. coli* strains transformed with plasmids p26199 (guide_182) and p26200 (non-target) measured in Sony SH800SA. Strains selected were taken from well plate of the cultivation after 48 h. s26902 is control strain without GFP and s26903 strain with GFP. Samples were 1:20 diluted in Sony Sheath fluid and measured in the FACS.

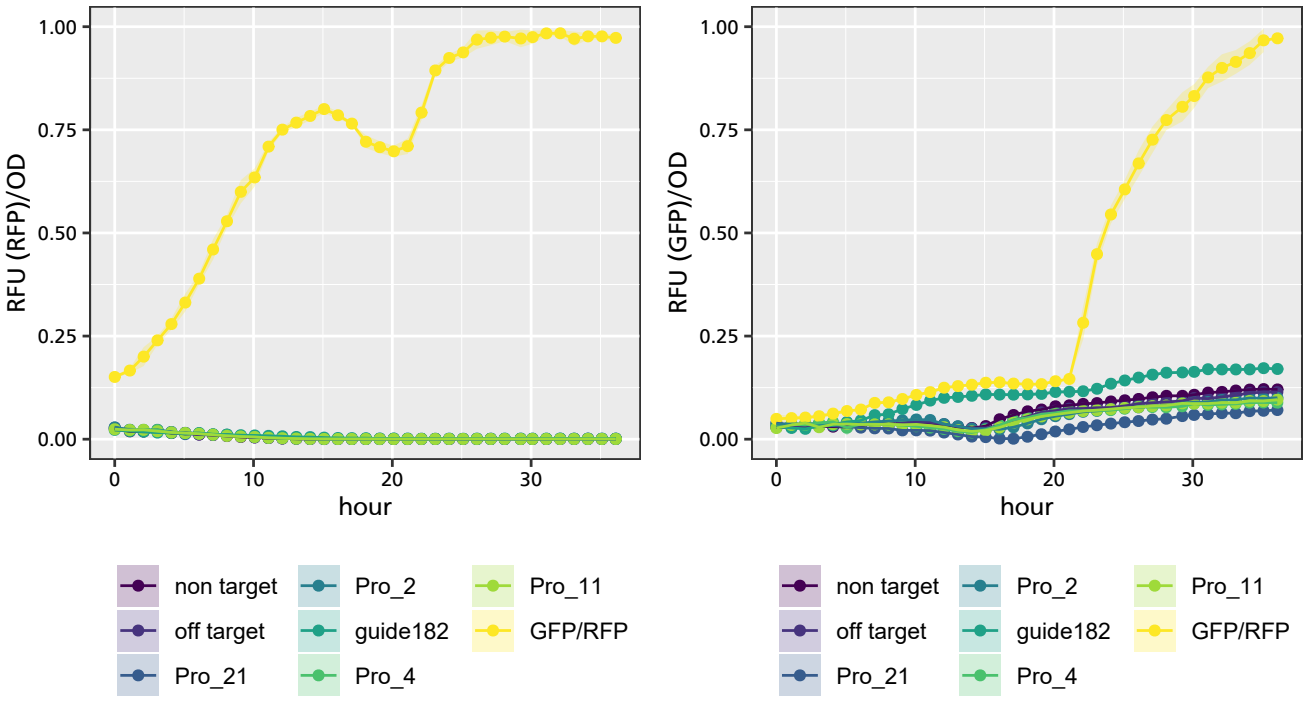


Figure 3.21.: Time course of fluorescence intensity normalized to optical density for *E. coli* strain expressing RFP as well as GFP. The strain (GFP/RFP) was cultivated together with other strains containing sgRNA delivery plasmids. OD corrected data were scaled to values between 0 and 1 to improved readability between RFP and GFP plot. The GFP/RFP strain contained the plasmid p10012, which expresses GFP as well as RFP under control of a strong promoter to evaluate autofluorescence development.

be considered. Regarding sensitivity, the usage of a luciferase would be conceivable. However, studies by Mihalcescu *et al.* suggested red fluorescent proteins as alternatives to GFP [250]. In addition, this was confirmed by cultivation experiments which were carried out to prepare an exchange of the reporter cassette. Unfortunately, the reporter exchange was not part of this thesis (Fig. 3.21).

As mentioned in section A.3.6, two variants were cloned for *E. coli* (1. in Pro_21 and 2. after Pro_21). Although the integration of the reporter cassette into the CRISPRpad was only shifted by four bases, it was not possible to integrate the CPad which had the integrated reporter after Pro_21. Since the cassette consisting of CPad, reporter and marker is about 3300 bp long, it is much larger than constructs where efficient recombination with Lambda-Red was measured. This drops to almost zero for fragments above 2500 bp [206]. Moreover, only a few clones could be generated for the integration of the variant, which was used for the experiments (integrated in Pro_21). The size dependence of transformation efficiency that Kuhlman *et al.* had already noted, was also observed here. For the CPad with selection marker (2469 bp), 16 clones could be generated in only one transformation, of which seven were positive in colony PCR and three were positive in sequencing. Contrary, integration of the Rank1-sfGFP cassette could only be achieved after eight transformations, with five clones generated of which three were colony PCR and two sequencing positive. This was achieved by using five times more DNA (500 ng) as suggested by the protocol. To date, transformation with Lambda Red represents the standard for chromosomal integration in *E. coli*. Alternatives, such as the use of Cas9, are still being combined with Lambda Red [291]. In conclusion, the model organism *E. coli* remains difficult to be modified.

In addition to the autofluorescence, it was observed that the cultures were subjected to increasing growth stress with increasing concentration of the inducer arabinose (Fig. A.14). Since the P_{BAD} promoter is excellent to titrate, lower arabinose concentrations were tested as the 1.5 % (w/v) in the Dong *et al.*[83][297] As the study by Cui *et al.* proposed, lower dCas9 concentrations are beneficial and improve the on-target positioning of dCas9. Additionally, higher concentrations of dCas9 lead to an increased risk for off-target binding. Further, the off-target binding may lead to interference with the regulation of essential genes. In other words, off-targeting may result in CRISPRa and CRISPRi of essential genes [70]. This mechanism could be a reason for a stronger autofluorescence, which differed depending on the used guide, regardless of GFP or non-GFP strain background. Thus, depending on the guide, off-target binding could still occur, and thus influence the metabolism. After establishing an easily measurable reporter system (mRFP, luciferase, etc.), it would be useful to determine the optimal inducer concentration to improve the on-target binding of dCas9.
Additionally, it has to be mentioned, that the Proof-of-Concept could only be shown with an inactive Cas9. Certainly, there is still much room for optimization for the next generation of CPads. This is already the second generation of pads, and it was further refined, aiming to reduce intramolecular homology. The number of repetitive sequences was again significantly reduced in the presented design, but sill led to problems during cloning, so that whole parts were dropped, when the selection markers was not set appropriately. A reduction of spacing or nesting of the artificial protospacers is maybe useful for further designs. This would mainly result in an improvement for CRISPRa in bacteria, as the range of activation is very small. Experiments with CRISPRa and CRISPRi have been chosen as the first access point to study the artificial sequences. In the end, however, these do not provide any information whether the goal of simplified cloning has been achieved. Here, studies still need to be carried out to assess transformation efficiencies while using Cas9-assisted cloning. This would be particularly interesting for organisms that are difficult to modify, such as *E. coli* and *Y. lipolytica*.

3.5. Conclusion and Outlook

SynBio continues to expand into biotechnological applications. But in the past, the focus of SynBio centered heavily on the model organisms *E. coli* and *S. cerevisiae* [131]. The wide availability of molecular biology tools, such as Cas9, is making a broad range of organisms accessible for designing switches, circuits, and entire synthetic pathways [51]. Biotechnological exploitation of the oleaginous yeast *Y. lipolytica* is increasingly the focus of industry and SynBio (Fig. 1.3). With the help of Cas9 and dCas9, modifications in the yeast genome can now be performed easily. However, the availability of well-characterized parts is still limited. Despite the availability of OMICS data for the organisms *B. subtilis*, *E. coli*, *S. cerevisiae*, and *Y. lipolytica*, predicting the response of the organism to genetic modifications remains challenging. Heterologous expression of proteins and pathways may result in changes in phenotype, metabolism, and fitness. Whether a host is suitable for production or a design, can often only be determined by trial and error. In addition, imbalanced metabolic pathways can lead to perturbations in metabolism and accumulation of intermediates [105].
To address these problems, several approaches were explored in this thesis. Thus, 1.) a plasmid was developed that allows for rapid testing of designs in *E. coli*, *S. cerevisiae*, and *Y. lipolytica* (section 3.3.1). 2.) the plasmid was used to test the transferability of a TCapt (section 3.3.2). Finally, 3.) artificial landing pads were tested to streamline the use of (d)Cas9 applications in the four organisms (section 3.3.3).
The FLEXpress plasmid (Fig. 3.5) was successfully tested in *E. coli*, *S. cerevisiae*, and *Y. lipolytica*. The use of the plasmid was demonstrated by selection and expression of the reporter gene mTFP. But, the used promoter from *A. gossypii* appeared to be limiting. The promoter showed a clearly detectable expression in *S. cerevisiae* and *Y. lipolytica*, which was strongly growth phase dependent in *S. cerevisiae*. Since the

promoter was derived originally from the KanMX marker, which can be used in *E. coli*, it was presumed that the promoter is active in *E. coli*. Although a significant difference was calculated, it was hardly visible in the plots (Fig. 3.6). Apparently, the weak expression of the KanMX marker is sufficient to confer resistance to Kanamycin.
Finally, the FLEXpress vector was used to verify the TCapt of Kötter *et al.* [204] in the yeast *Y. lipolytica*. The results of the authors were confirmed for *S. cerevisiae*, although not all designs were tested. The original design, containing up to three aptamers, was not established in *Y. lipolytica*. For the dimer and the trimer of the TCapt, the transformation of *Y. lipolytica* failed. Thus, only the monomer was tested in *S. cerevisiae* and *Y. lipolytica*. Further, it was challenging to cultivate the strain with the monomer. The aptamer had a clear effect on the fitness of the yeast, so a complete characterization was not possible (Fig. 3.8).
In addition to the transfer of the TCapt, the use of the new-to-nature CPad was shown. These were designed to have minimal off-target effect when using (d)Cas9 in the four organisms mentioned above. Most notably, in the yeast *S. cerevisiae*, a finely graded activation of the reporter gene was achieved, showing a decrease with increasing distance to the $P_{CYC1min}$ promoter. Similar targeted gene expression was demonstrated in *Y. lipolytica*, although the results were affected by the formation of hyphae. As an alternative, a change of the strain background, which is less prone to form hyphae, could be considered for future experiments. In contrast to the other organisms, a CRISPRi system was used in *B. subtilis*. For the prokaryote, two promoters were tested showing different patterns of repression, which was most likely due to the distance to the start codon. In *E. coli*, no targeted activation could be detected by using the CPads, neither by plate reader measurements nor by cytometry measurements. Since the distance of the additional guide guide_182 was designed to be in an optimal distance, it was assumed that autofluorescence masked the fluorescence on the GFP channel. Consequently, replacing the reporter, e.g., by mRFP or luciferase, could provide additional insights in the future. Further, for a revision of the design of the CPad itself, the distance between the artificial spacers should be shortened. This could potentially improve the usability in bacteria, since the range of efficient activation is rather small [83].

4. Laboratory automation in microbiological applications

The modern world of work is undergoing a fundamental process of change. In many parts, automation is replacing workers while creating new fields of employment. This process has not stopped at the life sciences either, even if the motivation is different. In the manufacturing industry, automation and robots are used to increase productivity. In addition, their usage has advantages, such as higher precision and the reduction of quality fluctuations.

4.1. Introduction

The latter point is probably also decisive for automation in the life sciences. Here, however, the comparability and reproducibility of results have a prominent role. Differences and a lack of reproducibility can have different sources. Starting with human causes induced by the experimenter, such as fatigue, injuries, time of the day, emotional state or other physical and mental handicaps [153]. Thus, although the protocols are similar, they are not performed exactly the same and the samples are treated differently than the day before. Perhaps the experiment or the analysis was carried out by another technician which performed some steps differently because of his/her different work experience. Another source of variation may be caused by chemicals used [248]. Particularly with cell cultures, batch-to-batch variability in components of the media can lead to significant changes in results [248][301]. But other causes should be mentioned briefly, such as defective equipment, insufficient calibration, as well as environmental factors which cannot be changed, such as temperature fluctuations or humidity [356].
Reproducibility and replicability are serious problems in the life sciences. The issue is intensified in SynBio and in metabolic engineering by the fact that not only qualitative but nowadays, quantitative statements are made [153][199][217]. Some are already talking about a ”reproducibility crisis” [17]. Leek and Peng distinguished between reproducibility (re-analysis of results) and replicability (chance of independently generating the same results) [217]. The difficulty of replicating scientific work has a much more trivial reason, which is illustrated by a paper from cancer research. Errington and colleagues set themselves the goal of replicating the 193 main experiments from 53 high-impact studies. They succeeded in replicating 50 experiments from 23 studies. The main reasons given were: incomplete protocols, incomplete list of used materials, no information on statistical evaluations. Furthermore, they were often unable to access the original data and the code for their analysis [96]. “And when asked “what percentage of published results in your field are reproducible?”, biologists estimated that only 59 % of results are reproducible.”[172] One way to improve replicability and reproducibility would be to pre-register studies so that both negative and positive results must be published [96].

4.1.1. Lab automation as solution for replicability

Another possibility to improve replicability is formalization through computer-aided biology (CAB). This allows for description of biological processes in a standardized way: ”digital representation of biological processes” [199]. Another way of abstracting processes is the systems biology markup language (SBML) [158]. A standard language for parts, including their properties, has been developed for SynBio [199]. To formalize the design principles in SynBio, the Synthetic Biology Open Language (SBOL) was developed by an academia/industry consortium and was published in version 3. It was developed as a standard for the specification and exchange of design information. It provides a ”[...] modular and hierarchical representation of the structure and function of a genetic design, as well as its relationship to and use within experiment plans, data, models, etc.”[16]
Using these design and formalization principles is certainly a first way to improve traceability. However, this does not reduce the variation of protocol execution, which can vary per experimenter. For example, the protocol is described, but not in all the details that are intuitively performed by each experimenter based on its knowledge and experience. However, hardware automation offers a solution. A literature analysis by Growth and Cox revealed that of 1454 analyzed studies, about 89 % of the used protocols have an automatable alternative [130]. Automation eliminates – at least for the automated part – the variance introduced by the experimenter. In addition, the use of automation leads to protocols becoming more standardized and to a computer-like description of the protocols through the programming of the systems [130]. Thus, automation enables not only the standardized description of constructs (SBOL) or processes (SBML), but also the standardized description of the executed protocols [153][199]. In SynBio in particular, many repetitive techniques are carried out, e.g., repeated cloning or repeated analysis of the circuits and switches that were created. It is precisely these repetitive processes that are ideal for automation [172]. There are different levels of automation. It should be critically noted that for some protocols, the use of electrical dispensers already represents partial automation. The Tab. 4.1 from Holland *et al.*, gives a short overview about the levels of automation [153].
However, the highest degree of automation is achieved by systems that can process entire work packages fully automated. Such systems offer the advantage that they process the specified protocol in exactly the same way as in the previous program run. Adam can be used as an example here. This automation solution cultivates yeasts fully automatically, whereby only consumables need to be refilled or the waste emptied by the user [130][198]. In addition, these systems offer logging, usually down to millisecond accuracy. Thus, levels of experimental logging can be achieved, which would be unattainable for manual experiments. Particularly in the regulated sector, these systems are connected to laboratory information management systems (LIMS), which log an exact trace of all steps, which were performed with each sample. Thus, the data is linked inseparably to the used protocol [172]. But, it is not only the reduction of variance between repetitions of the experiments that can be reduced in this way. Through automation of experiments, the throughput can be easily increased. Especially for the application of artificial intelligence, large amounts of data with low variance are of particular importance in order to train the algorithms [55]. However, increasing throughput by automation or through the use of microfluidic chips, brings new challenges. In this context, the analysis of the data is increasingly becoming a bottleneck, rather than the generation of the data [2][55].
For this reason, the field of automated data evaluation is becoming increasingly important. Since the focus here is on hardware automation, automated data evaluation should only be mentioned here, and references made to examples in the literature [2][88][215][326]. Further advantages and disadvantages of laboratory automation are mentioned below. Occasionally, it is the case that advantages and disadvantages are often inseparable.

Table 4.1.: Automation levels for biological examples. Modified from [153]

Level	Description	Biology research lab
1	Totally manual – Totally manual work, no tools are used, only the user's own muscle power. E.g., the user's own muscle power	Glass washing, tip stacking
2	Static hand tool – Manual work with support of static tool. E.g., screwdriver	Dissection scalpel
3	Flexible hand tool – Manual work with support of flexible tool. E.g., adjustable spanner	Pipette
4	Automated hand tool – Manual work with support of automated tool. E.g., hydraulic bolt driver	Stripette and electrical dispensers
5	Static machine/workstation – Automatic work by machine that is designed for a specific task. E.g., lathe	Centrifuge, PCR cycler, spectrophotometer, gel doc, plate reader
6	Flexible machine/workstation – Automatic work by machine that can be reconfigured for different tasks. E.g., CNC-machine	Motorized stage microscope, semi automated liquid handler
7	Totally automatic – Totally automatic work, the machine solve all deviations or problems that occur by itself. E.g., autonomous systems	Liquid handling robot, automated cell culture system, fully integrated robotic platform

4.1.2. Advantages and disadvantages of laboratory automation

Some examples given in Tab. 4.2 are not a clear advantage or disadvantage. Here it depends strongly on the respective use case. Certainly, the gain in system stability, reproducibility, and comparability outweighs the disadvantages. However, a user study must be carried out in advance to clarify whether the desired automation makes sense at all, or whether there is perhaps a simpler option to realize the project [404]. However, automation in the regulated pharmaceutical environment with highly repetitive tasks is probably easier to implement as automation in the academic environment, with its constantly changing tasks and questions [254]. This can be explained with software and hardware limitations of the implemented devices. Depending on the new research question, adjustments to the hardware might be necessary to meet the new requirements. Particularly in the project-funded research environment of universities, this can lead to not inconsiderable costs. But, if the problem is not solved, hardware limitations can lead to a reduction of creativity to improve available protocols, e.g., because a device does not offer the needed capabilities or is completely missing [153].

It is not only hardware that is limited, also labware and consumables have to be considered. For example, the introduction of SBS microplates has led to a kind of standardization of labware.[1] But, it is also a fact that the plates differ from manufacturer to manufacturer. Starting from the outer dimensions over the material to surface coating [224][242]. Due to the large market of consumables for automation applications, it might be useful to improve the reporting in a way which includes manufacturer and model of the labware [172]. The same must be applied for reporting of the automation itself. Often, protocols for automation systems are not transferable, which again limits replicability, e.g., when lab 1 has company A's liquid handler and lab 2 has company B's liquid handler. In addition, the used consumables are typically

[1] https://www.slas.org/education/ansi-slas-microplate-standards/ (date of access: 23.01.2022)

Table 4.2.: Advantages and disadvantages of lab automation in life science context

	Comment	Advantages (+) and Disadvantages (-)
Precise protocols and method execution	automation system follows exactly the programmed protocol	+ protocol reporting - mostly limited to device - hardware and software limitations reduce flexibility and creativity
Prevention of contamination	achieved by reduction of human interaction with samples [153]	+ increasing the validity of the data - hepa filter systems and enclosures required for sensitive samples
Standardization of labware	usually SBS standard plates used	+ improved replicability - not completely identical (re-teaching of robot) - sometimes special labware for one device
Non-standardized consumables	e.g., pipette tips of liquid handlers are not interchangeable	- limited to one manufacturer (business model of the manufacturers) - higher costs - replicability difficult on device from other manufacturers - different consumable compounds could affect results
Fast identification of errors	”fail fast, fail often” principle [153][194]	+ short repetition time of experiments + increase in result quality and reliability - higher costs for error identification
Reduction of human work	less manual work required to perform experiments	+ employees have time for other experiments or creative work (academia) [153] + reduction of variability caused by humans [46][286] - employees replaced by machines (industry)
Safety at work	robots and machines can work with substances which can have serious consequences for health	+ improved safety - possibly more difficult to handle due to enclosure
Usage of reagents	highly precise handling of liquids in automated systems	+ better replicability of experiments + reduction of reagent costs - sometimes high dead volumes depending on the system

available from the respective manufacturer only, as these are a part of their business model [153]. Possibly at this point, a standardized minimum requirement for specifying parameters of an automation experiment would be useful, as it is already available for RT-qPCR experiments[49].
Regarding liquid handling systems, there is another point which is both advantage and disadvantage at the same time. Liquid handling systems offer the possibility to pipette many samples and reagents with very high precision. They can perform this task not only with simple patterns, but also with very complex algorithms and simultaneous sample tracking. Additionally, the gain in precision means that reagents can be used more sparingly. But, some systems have large dead volumes, which a technician, for example, would not have by applying the pipette differently.[2] To overcome this specific problem, nanolitre dispensing systems could be employed to reduce the dead volume.
A general concern in life sciences is the high amount of plastic waste [366]. This issue is drastically increased through the use of automation equipment. Some manufacturers attempting to address this by making tips reusable. For example, this can be done by tip washing systems. An additional option is the use of special tips with reduced surface adhesion, that is achieved by surface coating [65][164]. But depending on the use case of the system, strong cleaning agents might be necessary, leading to a reduction of plastic consumption, but regarding sustainability, another problem is generated. In addition, depending on the application, contaminations may still be carried over or cleaning solution could remain in the tips [164]. Nevertheless, there is still a lot of potential for optimization regarding sustainability of automation solutions.
As the previous paragraphs show, automation has not only advantages. In the end, there are many aspects that must be considered for automation concepts. Nevertheless, laboratory automation in the biological and life sciences context offers an advantage that should not be neglected. Finally, as already mentioned at the beginning of the chapter, the use of robotic systems enables the replacement of repetitive work, so that scientists are significantly less time-bound to the experiments which are already automated. Now, they can carry out other experiments in parallel or develop new experiments for other questions [153].

4.2. CompuGene Robotics platform and a first workflow for characterization of microorganisms

Among other things, the aim of this work was to establish and evaluate various automation concepts in the context of SynBio and metabolic engineering. For this purpose, the working group had access to a fully integrated robotics platform from the Analytik Jena GmbH (Fig. 4.1), which was acquired as part of the CompuGene project funding. This platform was planned and developed for various automation scenarios. The aim was to use the robot for fully automated cloning and characterization of designed circuits, switches, and strains. For this purpose, different devices were available in the platform to enable these tasks. Integrated were two CyBio liquid handlers with an 8 channel (SELECT Head) and an 96 channel head (R96), as well as the nanoliter dispenser I-DOT from Dispendix. Furthermore, two analytical instruments were available: a plate reader (BMG PHERAstar® FSX) and a flowcytometer (Beckman Coulter Cytoflex S). For incubation, there was an automated plate incubator (Thermo Fisher Cytomat 2), a PCR cycler (Biometra TRobot), and a cooling position. All devices were connected by a rail-based robot from Precise Automation, which can transport labware in SBS format by means of a gripper.
Initially, two workflows were programmed by Analytik Jena to cover both cloning and characterization. The automation concept for cloning will not be discussed further, as it was part of the PhD thesis of Niels Schlichting [316]. The ordered workflow for characterization of strains is illustrated schematically in

[2] https://www.hamiltoncompany.com/automated-liquid-handling/everything-you-need-to-know-about-liquid-handling/manual-pipetting-vs-semi-automation-vs-automation (date of access: 23.01.2022)

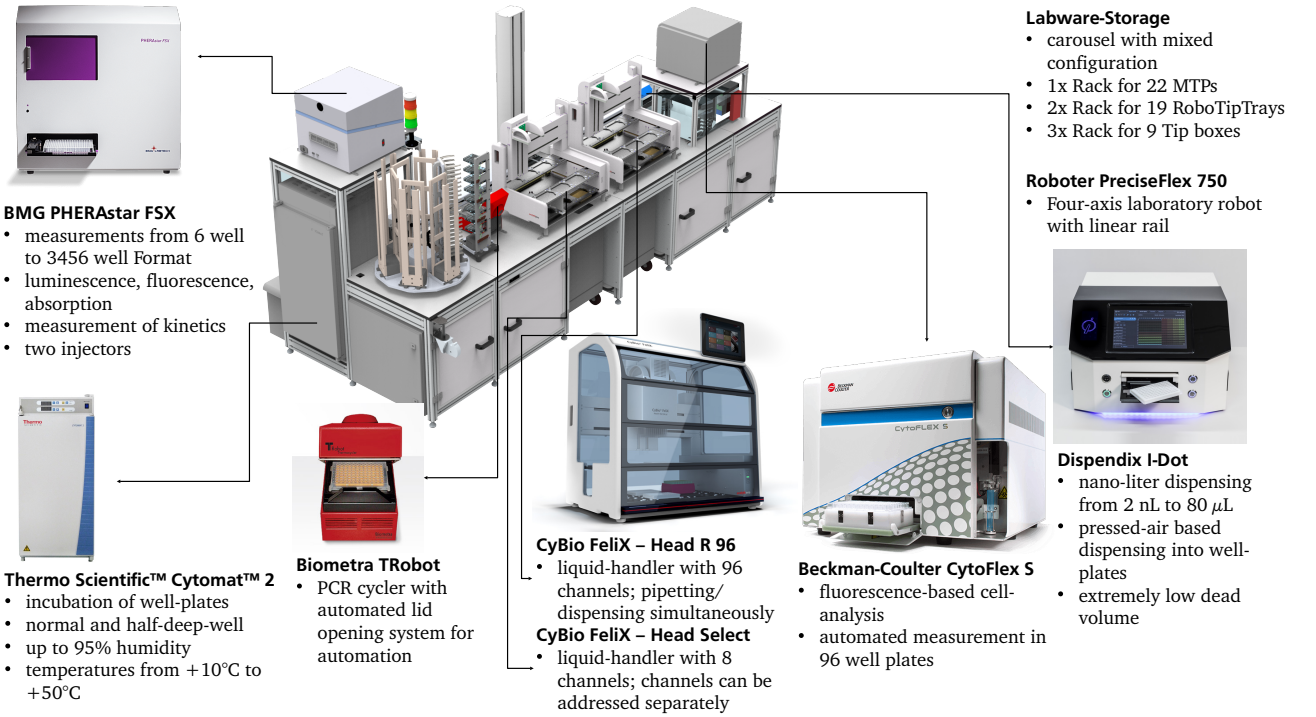


Figure 4.1.: Scheme of the CompuGene Robotics Platform

Fig. 4.2 and is briefly described below.
The basic idea of the workflow was inspired from a publication by Nielsen *et al.*, where the protocol was executed manually to analyze artificial genetic circuits [262]. The workflow starts with the incubation of the preculture plate in 96 well format, which is grown overnight in the incubator. After 16 h, the plate is transported to the PHERAstar®, and the absorbance is measured. The measured values are imported into the database by a .vbs script. A function of the SQL server uses the measured values to calculate the volume for the hit picking from the preculture plate to create a master plate with a desired OD. This so-called master plate is then transported from FeliX SELECT to the FeliX R96 to inoculate four equal main-culture plates with a preset OD. The plates were grown in the incubator and measured regularly until an OD between 0.4 and 0.6 is reached. The plate that reaches this range is induced on the Felix R96 with one of the four inducers and enters the measurement cycle. In the measurement cycle, OD and fluorescence are measured every hour in the plate reader and a sample of 10 µL is diluted in PBS-Kan at timepoint 0 h, 5 h and 10 h (in cycle 0, 5 and 10 after induction) and measured in the cytometer until a pre-defined stop criterion is reached. The copied plate is subsequently discarded. The process is terminated after all four plates have been induced and run through ten measurement cycles.
The ordered and previously described workflow thus offers the possibility to test four different inducers and to add them conditionally at a previously defined OD. In addition, different measurement methods are applied to the samples in fixed intervals. Furthermore, the illustrated workflow is sufficiently complex. Unfortunately, the underlying concept is too rigid to be easily adapted for a different question. For example, if more than four inducers should be used, changes would have to be made in many parts of different sub scripts to get the desired result. However, the concept underlying here was adopted and in the following subsections first simplified and further developed and replaced by other programming concepts, which are explained in detail by means of examples.

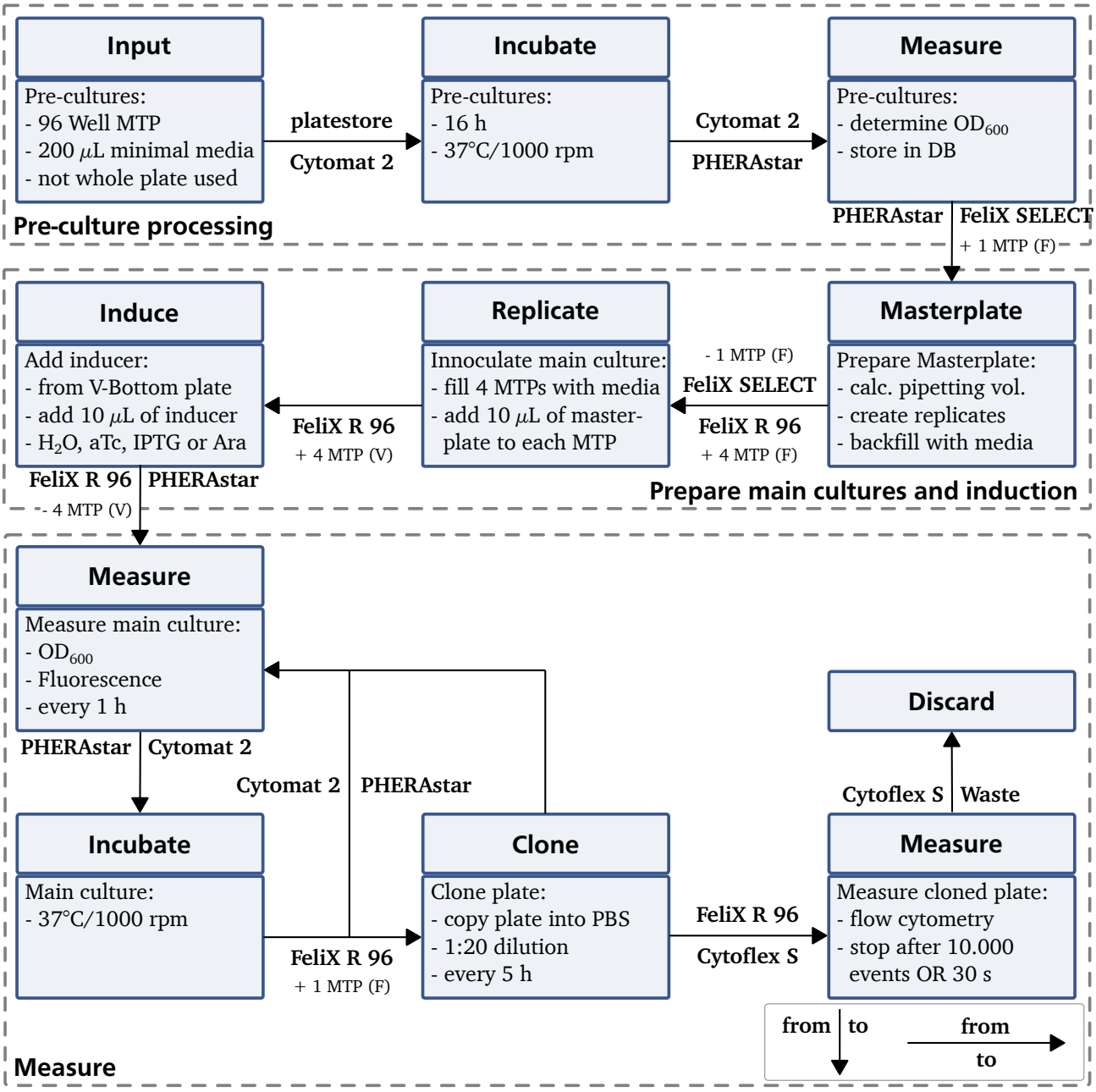


Figure 4.2.: Workflow scheme of the standard characterization experiment.

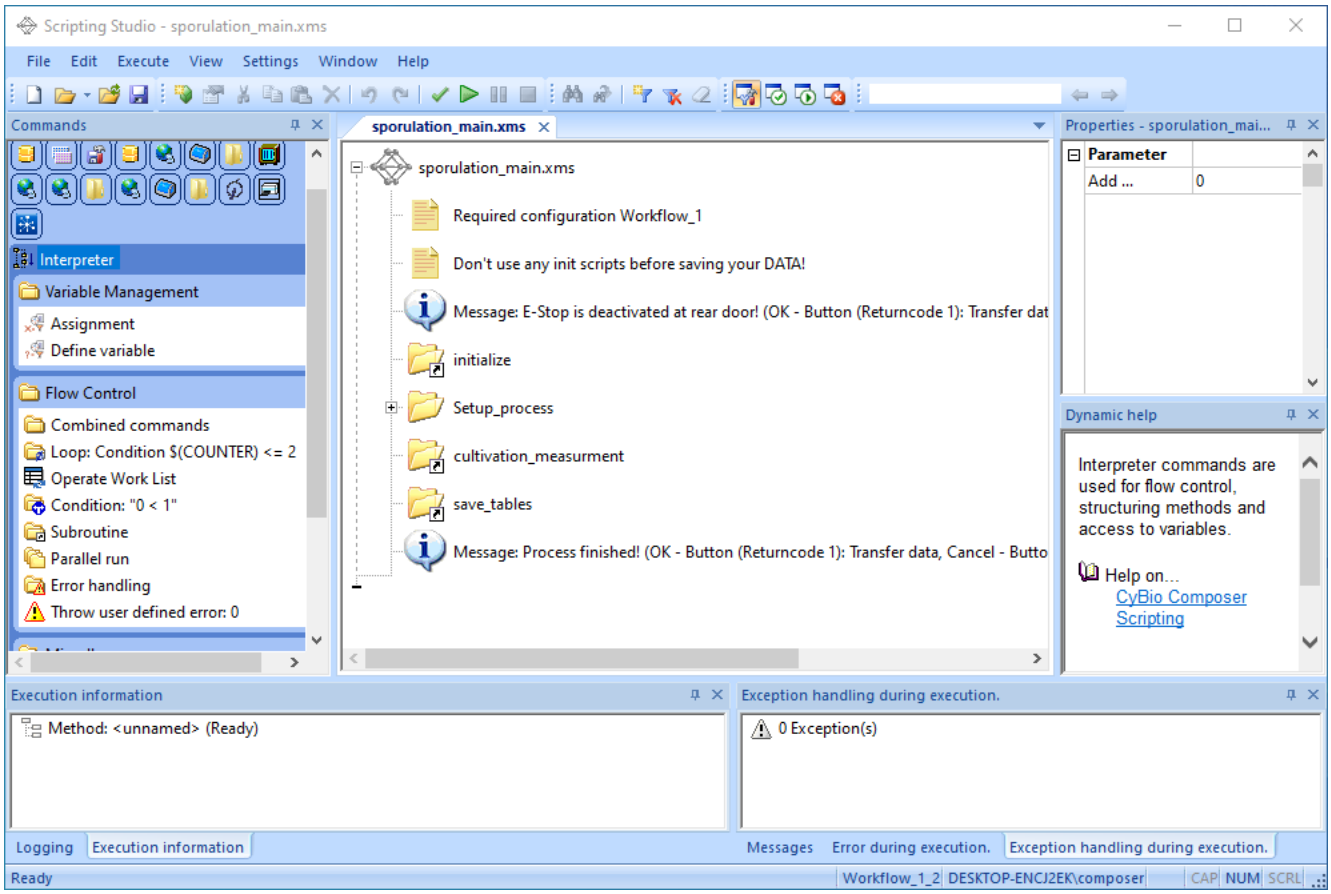
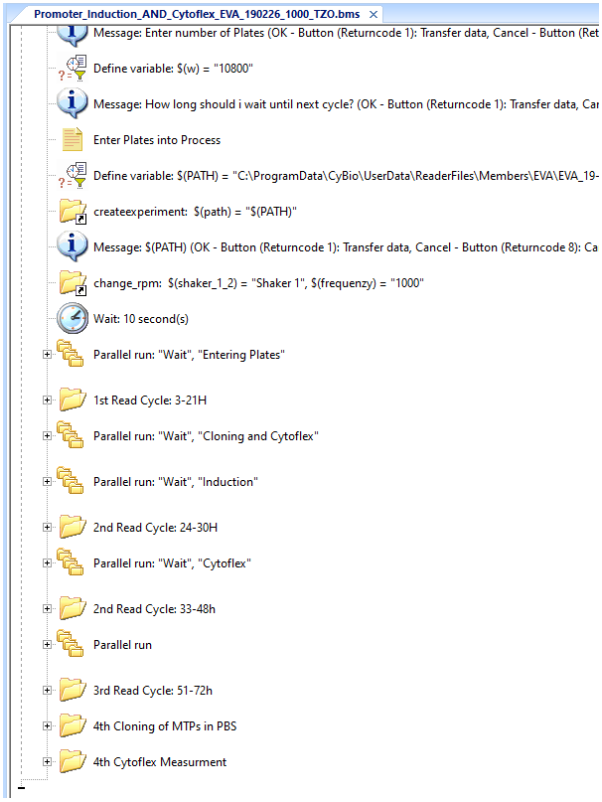


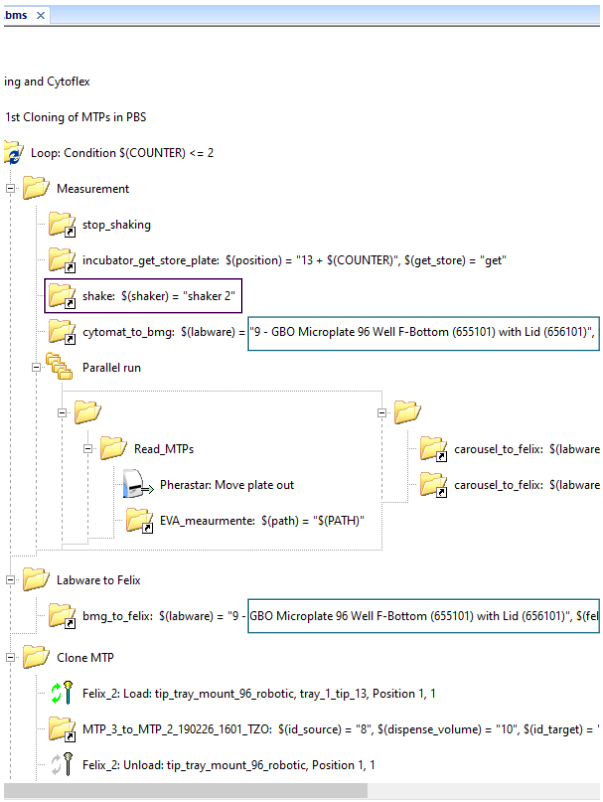
Figure 4.3.: Analytik Jena Composer Scripting Studio. The left panel shows basic commands, such as logic operations, loops, error handling etc. The device icons let the user switch to the respective device command list. The tabs at the bottom contain runtime information and error messages.

4.3. Evolution of workflow development

The system is programmed in the proprietary software Composer, which has been developed by the integrator Analytik Jena. The Composer Scripting Studio (Fig. 4.3) provides an interface for creating so-called scripts, which are interpreted and executed by Composer. In the scripts, logic operations, loops, conditional statements as well as so-called plug-in commands for the respective devices can be used to create workflows. The software offers the possibility of outsourcing of individual operations into so-called sub scripts, which can then be called in higher-level scripts. Furthermore, various other options are available, e.g., parallel execution, error handling options, and the creation and assignment of variables. The screenshot in Fig. 4.3 shows the internally available standard commands on the left side. The upper icon bar allows to switch to the so-called plugin commands, which are dependent on the platform's device equipment and enable the control or remote control of the respective device software. With the available on-board tools, complete workflows can be programmed in different ways. The respective concept offers advantages and disadvantages.



(a) Screenshot of the Promoter Induction Workflow. Sequential execution of loop structures containing mostly the same instructions.



(b) Expanded Cloning and Cytoflex with instructions. Marked are parameters which have to be changed in case another labware type should be used.

Figure 4.4.: Screenshots of the first generation of workflows for the CompuGene Robotics platform. **(a)** Shows the whole workflow with collapsed loop structures. **(b)** Illustrates the changes to be made, if another labware should be used. The parameter for the labware and the parameter for the towershaker have to be changed in every loop.

4.3.1. 1st Generation: Rigid loop-structures

Initially, a highly simplified workflow design was used that mainly consisted of a sequence of loop structures. The process was used to characterize a promoter from the yeast *Y. lipolytica* by culturing three plates with yeast strains of different genetic background. One plate was to be induced directly before the start, another one after 24 h and at the same time a plate with water as control. The OD and FI should be measured every 3 h in the plate reader, and a cytometer measurement should be performed at certain intervals. For this purpose, individual loop structures are executed in series, which actually have the same task (1st read cycle, 2nd read cycle, etc.) (Alg. 1 and Fig. 4.4a). Each step included incubation and measurement of multiple plates. But, this type of programming has some advantages and disadvantages, which will be discussed below.

First, the sequential execution of tasks is relatively easy to realize. However, the sequential execution of loops leads to the problem that the clarity of the script is severely impaired. The script does not use outsourcing of the same program parts to other sub scripts. This means that all identical program steps must be changed in all loops. An example would be, the change from a 96 well to a 24 well plate as

Algorithm 1 Pseudocode for Promotor Induction Composer Script - Part 1/2

```
1:  Composer Script PROMOTOR INDUCTION(pc, w)        ▷ pc = platecount , w = waiting time
2:      for i ← 1 to pc do                            ▷ Enter plates into process
3:          get plate from pos i
4:          measure plate in plate reader
5:          store plate in incubator pos i
6:      end for
7:      for n ← 1 to x do                             ▷ 1st Read cycle: 3-21 h
8:          for i ← 1 to pc do
9:              get plate from incubator pos i
10:             measure plate in plate reader
11:             store plate in incubator pos i
12:         end for
13:         wait w seconds                            ▷ runs parallel to for loop
14:     end for
15:     for i ← 1 to pc do          ▷ measure in plate reader, induce, copy for cytometer
16:         get plate from platestore pos i
17:         measure plate in plate reader
18:         induce main culture plate
19:         transfer 10 µL sample into new plate
20:         store main culture plate in incubator
21:         measure new plate in cytometer
22:     end for
23:     wait w seconds                                ▷ runs parallel to for loop
24:     for n ← 1 to y do                             ▷ 2nd Read cycle: 24-30 h
25:         for i ← 1 to pc do
26:             get plate from incubator pos i
27:             measure plate in plate reader
28:             store plate in incubator pos i
29:         end for
30:         wait w seconds                            ▷ runs parallel to for loop
31:     end for
32:     for i ← 1 to pc do                            ▷ Cloning and Cytometer Measurement
33:         get plate from platestore pos i
34:         measure plate in plate reader
35:         transfer 10 µL sample into new plate
36:         store main culture plate in incubator
37:         measure new plate in cytometer
38:     end for
39:     for n ← 1 to z do                             ▷ 3rd Read cycle: 33-48 h
40:         for i ← 1 to pc do
41:             get plate from incubator pos i
42:             measure plate in plate reader
43:             store plate in incubator pos i
44:         end for
45:         wait w seconds                            ▷ runs parallel to for loop
46:     end for
```

Algorithm 2 Pseudocode for Promotor Induction Composer Script - Part 1/2

```
47:     •
48:     •
49:     •
50:     for i ← 1 to pc do                ▷ Cloning and Cytometer Measurement; last round
51:         get plate from platestore pos i
52:         measure plate in plate reader
53:         transfer 10 µL sample into new plate
54:         store main culture plate in platestore
55:         measure new plate in cytometer
56:     end for
57: end procedure
```

parameter of the respective transport scripts, which was marked in Fig. 4.4b. Here, the parameter for the labware and the used tower in the incubator would have to be changed in each loop. Furthermore, when using loops, the number of runs between PHERAstar® measurements and additional measurements in the Cytoflex S must be changed manually, if other intervals are desired.
Another important factor is the system stability, e.g., due to a power failure or the E-Stop of the system which was installed for safety reasons. In addition, simple faults in the equipment can lead to interruptions and overall termination. If an interruption occurs, the experimenter must manually correct the error in the platform and deactivate the corresponding – already completed – program parts in the script. Loops that were already started must be set manually to the respective run, e.g., if two of the three plates were processed, the number of loop runs must be set to 1. However, since the position is calculated by using the counter variable of the loop (incubator_get_store_plate expects labware position as integer), this must be corrected manually. If the software was terminated, it might be necessary to use the log file first to obtain the termination point. Any parameters and variables changed during runtime will be lost with the termination and must be corrected manually. Manual changes are a common source of errors in program execution, and they reduce traceability. Although logging is still available, it also lacks clarity, especially for new users. In addition, the log must often be transferred to UNIX-like systems, since the Windows on-board tools are overwhelmed by the size of the log files.
Additionally, at this point, it is difficult to estimate the time sequence, since no parallelization of the cytometer measurement is included. The required time of cytometer measurements is difficult to predict because it depends on the stop condition which was set for each well of the plate. If the measurement takes longer as the 1 h time slot, then there might be delays in the scheduling of the entire program that cannot be compensated. Finally, due to the lack of flexibility of the loop structure, the backlog will accumulate, and measurement timepoints will no longer be kept.
But, the lack of flexibility significantly increases the reproducibility, as measurement intervals can always be precisely maintained, assuming the scheduling of the individual plates works out.

4.3.2. 2nd Generation: Separation into different tasks

The lack of flexibility led to a development based on the programming concept of Analytik Jena. A workflow with randomization will serve as example, whereby the randomization will be discussed in more detail in section 4.4. As the pseudocode in Alg. 3 shows, a subdivision of individual program sections was carried out in the main script. The script is divided into initialization, start, randomization, cultivation and tidy up. The summarizing of the individual tasks of the workflow makes a better readability of the program

Algorithm 3 Pseudocode for the main script of the randomization process

```
1: procedure RandomizationMain
2:     initialize robot                          ▷ reset DB, establish device connection
3:     import plates into process                ▷ requires number of plates and interval
4:     wait w seconds                            ▷ waiting time for preculture
5:     randomize plate into new plate
6:     cultivation                               ▷ cultivation and measurement of plates
7:     tidy up robot
8: end procedure
```

possible. The most considerable change from the previous concept is the subdivision of the workflow itself. The previous subdivision of the script was more sequential. This has the benefit that the desired protocol is reflected in the program flow. Here, the cultivation sub script abstracts the experimental flow and uses individual work packages to complete the tasks. As shown in the pseudocode in Alg. 4, the respective tasks are divided into individual functions, e.g., for measuring in the PHERAstar®, induction, or preparing a cytometer measurement.

Briefly summarized, the workflow enables a periodic measurement of the plates in the PHERAstar®. Furthermore, an induction with previously set OD or no induction can be carried out here. The use of conditional induction requires the use of the R script $\mathrm{PHERAread}$ that imports the .csv files of the PHERAstar® and stores the data including the mean values in the SQL database. However, here, the cytometer measurements can be carried out in parallel to the actual cultivation, since the cytometer measurement is a bottleneck which complicates the scheduling considerably (Fig. 4.5a).

The execution of the main program parts in the $\mathrm{Cultivation}$ script are shown as pseudocode in Alg. 4. Here, as in the first approach, loop structures were also used. However, here, a do-while instead of a for-loop was used. That means, a loop is not executed n times, but as long as a condition is fulfilled or no longer fulfilled. In this case, until both end conditions ($FINISHED\_CULTIVATION = 1$ and $FINISHED\_CYTOFLEX = 1$) have occurred. In this workflow, many runtime parameters are outsourced to the SQL database. For example, the plates involved in cultivation are entered in the $\mathrm{DIB\_Cultivation}$ table (Fig. 4.5b). Here, the platetracking ID (PT_ID), the cycle, the next measurement timepoint and some other parameters are entered for tracking purpose. The next measurement time specifies as integer value when the plate with the respective ID is to be measured again. Therefore, the script $\mathrm{next\_PTID}$ compares the $\mathrm{DIB\_Cultivation}$ table with the current system time. If this is greater than or equal to the next measurement time, the script returns the PTID with the smallest value for $\mathrm{NextMeasurement}$. As shown in line 14 (Alg. 4), the execution of the respective program part takes place. Here, the position of the plate in the incubator is determined automatically based on the PTID, and the plate is transported to the transfer position. Compared to the first generation, manual interventions in case of an abort are avoided here, since the position of the plate does not have to be calculated, as it was the case for the first generation. This means, that the position value of the plate does not have to be corrected manually, which reduces the risk of errors. The plate is measured in the PHERAstar® and the next measurement time is calculated ($system.time + interval$), and the cycle is incremented ($cycle = cycle + 1$) and updated in the SQL DB. At this point, it is checked whether the maximum number of measuring cycles for the respective plate was exceeded, and if necessary, the value $Finished$ is set to 1. The cycles are used at this point as a tool to determine the runtime of the plate and can be defined by the user. For example, if a plate should be incubated for 24 h and measured every 2 h, it will run through a total of 12 measurement cycles. Subsequently, a check is made whether the plate has to be induced or whether a measurement should be carried out in the Cytoflex. For this purpose, the variable IND or CYT is set to the value 1. The criterion

Algorithm 4 Pseudocode for cultivation script of randomization process - Part 1/2

```
1:  procedure RandomizationCultivation

2:      FINISHED_CULTIVATION ← 0
3:      FINISHED_CYTOFLEX ← 0
4:      PTID ← 0
5:      CYCLE ← 0
6:      IND ← 0
7:      CYT ← 0
8:      MOD                                                    ▷ to be set by user

9:      function PHERAstarReading
10:         PTID ← 0
11:         call next_PTID(FN, PTID)                           ▷ call sub next_PTID
12:                                  ▷ check DB; if plate found for measurement; return PTID
13:                                  ▷ FN is name of the requesting function
14:         if PTID > 0 then
15:             get plate from incubator with platetracking-ID PTID
16:             measure plate in plate reader and set next measurement time in DB
17:             if OD > 0.4 then
18:                 IND ← 1
19:                 transfer plate to FeliX R96
20:             else if plate is already induced and CYCLE mod MOD = 0 then
21:                 CYT ← 1
22:                 transfer plate to FeliX R96
23:             else
24:                 store plate in incubator
25:             end if
26:         else
27:             do nothing
28:         end if
29:     end function

30:     function Induce
31:         PTID ← 0
32:         call next_PTID(FN, PTID)                  ▷ check if IND = 1; return PTID
33:         if PTID > 0 then
34:             get plate with inducer and tip tray
35:             add inducer
36:             measure plate in plate reader and set next measurement time in DB
37:             store plate in incubator
38:             tidy up FeliX R96
39:         else
40:             do nothing
41:         end if
42:     end function
```

Algorithm 5 Pseudocode for cultivation script of randomization process - Part 2/2

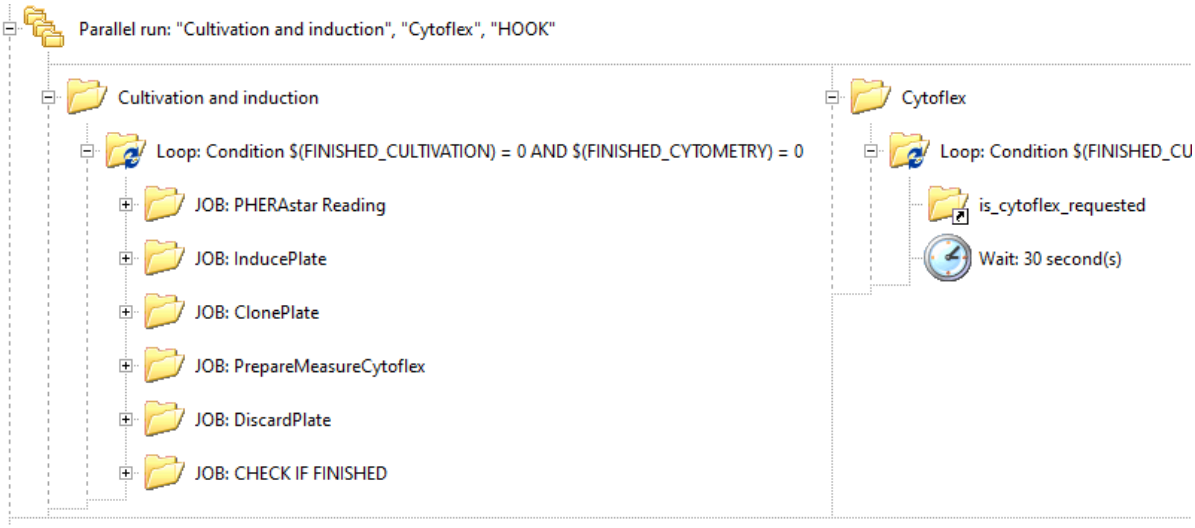
```
43:  function CLONE
44:      PTID ← 0
45:      call NEXT_PTID(FN, PTID)                    ▷ check if CYT = 1; return PTID
46:      if PTID > 0 then
47:          get pre-filled plate with PBS and tip tray
48:          transfer sample from main culture to PBS plate
49:          store main culture plate in incubator
50:          store PBS plate in platestore
51:          tidy up FeliX R96
52:      else
53:          do nothing
54:      end if
55:  end function

56:  function PREPARECYTOMETERMEASURMENT
57:      PTID ← 0
58:      call NEXT_PTID(FN, PTID)                    ▷ check if cytometer is empty AND
59:                                                  ▷ plate for measurement available; return PTID
60:      if PTID > 0 then
61:          transfer cloned plate from platestore to cytometer
62:      else
63:          do nothing
64:      end if
65:  end function

66:  function TIDYCYTOMETER
67:      PTID ← 0
68:      call NEXT_PTID(FN, PTID)                    ▷ check if cytometer has finished; return PTID
69:      if PTID > 0 then
70:          transfer cloned plate from platestore to cytometer
71:      else
72:          do nothing
73:      end if
74:  end function
75:  if cyometer is done then
76:      FINCUL ← 1
77:  end if
78:  if cultivation is done then
79:      FINCYT ← 1
80:  end if
81:  wait 10 seconds
82: end procedure
```

for induction is the OD. The criterion for the Cytoflex measurement is the resulting remainder in the division. Thus, the divisor can be determined by the user. As soon as the remainder 0 is generated during the division

of cycle and divisor, a cytometer measurement is performed. For both $IND = 1$ and $CYT = 1$, the main culture plate is transported to the FeliX R96, otherwise back to the incubator and the PHERAstar® reading section is complete. As the pseudocode further shows, either the Clone or Induce program section is then executed. In case of induction, an entry is made in the SQL table (Fig. 4.5b) that the plate was induced, and which inducer was used. If a plate was cloned, i.e., a sample was transferred to a new plate with PBS, then the plate is temporarily stored in the platestore. Subsequently, it is checked whether the cytometer is free or whether an already measured plate occupies the cytometer. If the cytometer is occupied, the plate is removed with $\mathrm{TidyCytometer}$. If the cytometer is free, the plate is transported to the cytometer with the $\mathrm{PrepareCytometerMeasurement}$ function. The measurement in the cytometer is done in a parallel branch, which is not included in the pseudocode, but is shown in the screenshot in Fig. 4.5a. Finally, it is checked, if all main culture plates were set to $Finished = 1$. If this is the case, $FINISHED\_CULTIVATION$ is also set to 1 and one of the end conditions is fulfilled. Another function checks, if there are still plates waiting for a cytometer measurement or if a plate is currently measured. If there are no more plates, $FINISHED\_CYTOFLEX$ is also set to 1 and the process is terminated.
This workflow enables a noticeable increase in fail-safety. If a failure occurs, the runtime information, such as the next measurement time or how long the process has been running, is stored in the database. Thus, the process can ”simply” be restarted, and the robot continues to work at that point it was interrupted. However, if the failure has occurred in one of the subtasks, it is necessary to restore to the state before the start of the last executed function. Additionally, since there is a 10-second wait at the end of the While-loop, this can also be used to interrupt the process deliberately, for example to refill tips, plates, labware and liquids. As a disadvantage, it has to be noted that the flexibility was limited by the fact that the timepoint for the cytometer measurement is calculated. This means that only regular intervals are possible. But by changing the calculation rule, this circumstance could be compensated. A benefit at this point is probably the flexibility of the workflow, since the cytometer measurement is carried out in parallel. This means that other tasks can be performed even though the measurement has not yet been completed.
However, optimizing the timing of the workflow also reveals disadvantages. To save time, the cytometer measurement/induction was nested to the PHERAstar measurement. Thus, the transport to the FeliX R96 for preparing the cytometer measurement or induction, is already carried out from the $\mathrm{PHERAstar\ measurement}$ function. This saves time in the process, but again reduces the clarity of the code. In addition, the script next_PTID becomes unnecessarily complicated and unclear. The latter points severely limit the reusability of the code, which is also further complicated by the GUI of Composer.

4.3.3. 3rd Generation: Encapsulated jobs for improved recycling

The focus of the further development was on the reusability, clarity, and stability of the workflows. The previous concept already used a subdivision into individual subtasks of the workflow. The nesting of individual tasks has a negative effect on the reusability of individual tasks. The interconnection consisted in the fact that transport operations for later tasks were already performed by the previous subtask. The same applies to runtime variables which were already set in previous subtasks. To this end, in each case, self-contained application parts were provided, whose entire processing is carried out by a single subscript. This principle is explained in the following paragraph by means of an example.
The example workflow was used to characterize the sporulation of *B. subtilis*. The aim was to optimize media for the sporulation for the projects of Marianna Karava and Benedict Spannenkrebs. The workflow here is based on the method for the quantification of spores using the dye SYBR® Green [188]. The workflow was planned to incubate plates in 24 well format. During this process, the plates should be measured every 3 h in the PHERAstar®. Every 6 h, in addition to the PHERAstar® measurement, a measurement in the Cytoflex was planned. Unfortunately, for the Cytoflex, the transfer from 24 well to 96



(a) Screenshot of the randomization workflow. All functions were collapsed. The left path shows the actual cultivation, whereas the right, parallel path shows the cytoflex measurement.

SQLQuery1.sql - DE...2EK\composer (55))*

```
SELECT TOP 1000 [ID]
      ,[PT_ID]
      ,[Cycle]
      ,[NextMeasurment]
      ,[EndMeasurment]
      ,[InductionRequested]
      ,[Induced]
      ,[Inducer]
      ,[Finished]
      ,[Preculture]
      ,[Notes]
  FROM [CSP1750].[dbo].[DIB_Cultivation]
```

100 %

Results | Messages

	ID	PT_ID	Cycle	NextMeasurment	EndMeasurment	InductionRequested	Induced	Inducer	Finished	Preculture	Notes
1	19	79	17	1643480416	NULL	0	0	NULL	1	0	NULL
2	20	80	17	1643480470	NULL	0	0	NULL	1	0	NULL
3	21	81	17	1643480523	NULL	0	0	NULL	1	0	NULL

(b) SQL Table DIB_Cultivation contains runtime information, like $PTID$ (platetracking ID), $cycle$, $NextMeasurement$ (as UNIX time), $InductionRequested$ (0 or 1), $Induced$ (0 or 1) and other.

Figure 4.5.: Workflow of 2nd Generation. Subdividing tasks into separate functions, like PHERAstar Reading or Induction

well plates was necessary due to technical reasons. This was done on the FeliX SELECT. First, PBS and SYBR® Green were added to the wells and afterwards the sample was added. The main culture plate was transferred back to the incubator. The new plate, filled with PBS, SYBR® and sample, was incubated for 20 min and subsequently measured in the cytometer.
To increase the stability of the entire workflow, the important variables and parameters were stored in different tables of the SQL database. This was realized by creating a setup section in a position, which is intuitive for the user (Alg. 6 and Fig. 4.6a). For example, the interval between the measurements or the value modulo for the interval of the Cytoflex measurements can be managed here. In contrast to the Analytik Jena script and the previous workflows, entries are already made in the platetracking table at this point. Thus, this script section helps the user to load the correct positions. In addition, a possibility was created to provide the measurement protocols for the PHERAstar® in the setup section. This central setup possibility increases clearly the usability, so that the most important changes only take place in the main script and not in many different sub-scripts.
Alg. 7 shows the main script, which is now no longer divided into different subtasks, but into self-contained

Algorithm 6 Pseudocode for Sporulation Optimization Main script

```
1:  Composer Script SporulationMain
2:      initialize robot                          ▷ reset DB, establish device connection
3:      StartCombinedCommands                     ▷ collection of required parameters/variables
4:          Check if SQL tables available and reset tables
5:          Enter plates into platetracking
6:          Add global variables to SQL table
7:          Add BMG reader protocols to SQL table
8:          Setup incubator and BMG reader
9:          Initialize sample tracking
10:     EndCombinedCommands
11:     Cultivation
12:     Tidy up robot
13: end Composer Script
```

work packages. Thus, the PHERAstar® measurement is in one package and the PHERAstar® measurement together with the preparation of the Cytoflex measurement is in another package. In addition, the scripts that take care of the task assignment have been revised. In order to conduct the nested tasks, this function was previously outsourced to a single script. Now every job has its own, clearly arranged script, which handles the assignments through a database query. In this way, the job scripts including the assignment script can be recycled for other workflows without much effort. As long as the required entries and tables can be found in the database.
As mentioned at the beginning of the section, attention was paid to the outsourcing of runtime information to the database. In case of a failure, usually only the plate tracking table has to be reset to the state before the start of the job. In addition, the plates have to be placed accordingly in the real world by the user and the process can be restarted. No further user intervention is required at this point. The use of the database enables an increase of fail-safety, and also drastically increases the sample tracking within the process.
Sample tracking is not a problem for PHERAstar® measurements, since the Composer plugin offers the possibility to pass both the platetracking ID and the barcode. This meta information is stored together with other meta information in the export file of the reader. Tracking for cytometer measurements, however, turned out to be more challenging. Analytik Jena's solution was to rename the folder, containing the measurement data so that the name was supplemented with the barcode of the plate. However, the

Algorithm 7 Pseudocode for Sporulation Optimization Cultivation script - Part 1/2

```
1:  procedure SporulationCultivation

2:      FIN_CUL ← 0                          ▷ Finished Cultivation FIN_CUL ← 1 if all plates processed
3:      FIN_CYT ← 0                          ▷ Finished Cytoflex FIN_CYT ← 1 if cytoflex is done
4:      FIN_ENT ← 0                ▷ Finished Entering FIN_ENT ← 1 all plates have entered the process
5:      PTID_GLOBAL ← 0

6:      while FIN_CUL = 0 OR FIN_CYT = 0 OR FIN_ENT = 0 do
7:          JOB Enter Plates                                  ▷ Enter plates with defined interval
8:              PTID_GLOBAL ← 0
9:              call CheckForPlateToEnter(PTID_GLOBAL)
10:             if PTID_GLOBAL > 0 then
11:                 call EnterPlates(PTID_GLOBAL)
12:             else
13:                 do nothing
14:             end if
15:         end JOB

16:         JOB BMG Measurement                              ▷ measures plate in BMG reader
17:             PTID_GLOBAL ← 0
18:             call CheckForPlateForBMG(PTID_GLOBAL)
19:             if PTID_GLOBAL > 0 then
20:                 call measureBMG(PTID_GLOBAL)
21:             else
22:                 do nothing
23:             end if
24:         end JOB

25:         JOB BMG Measurement, ClonePlate, Add SYBR
26:                 ▷ measures plate in BMG reader, transfer from 24 to 96 well plate and add SYBR
27:             PTID_GLOBAL ← 0
28:             call CheckForCytoflexCloning(PTID_GLOBAL)
29:             if PTID_GLOBAL > 0 then
30:                 call bmg_SYBR_cloning(PTID_GLOBAL)
31:             else
32:                 do nothing
33:             end if
34:         end JOB
```

Algorithm 8 Pseudocode for Sporulation Optimization Cultivation script - Part 2/2

```
35:         JOB Discard Plate                          ▷ removes measured plate from cytoflex
36:             PTID_GLOBAL ← 0
37:             call CheckForWasteOrRecycle(PTID_GLOBAL)
38:             if PTID_GLOBAL > 0 then
39:                 call PutInWastOrRecycle(PTID_GLOBAL)
40:             else
41:                 do nothing
42:             end if
43:         end JOB

44:         JOB Transfer To Cytoflex          ▷ Gets Plate from storage and puts into cytoflex
45:             PTID_GLOBAL ← 0
46:             call CheckForPlateForCytoflex(PTID_GLOBAL)
47:             if PTID_GLOBAL > 0 then
48:                 call TransferToCytoflex(PTID_GLOBAL)
49:             else
50:                 do nothing
51:             end if
52:         end JOB

53:         StartCombinedCommands
54:                 call CheckIfCultFinished(FIN_CUL)
55:                 call CheckIfCytoflexFinished(FIN_CYT)
56:                 call CheckIfEnteringFinished(FIN_ENT)
57:                 wait 10 seconds
58:         EndCombinedCommands
59:     end while
60: end procedure
```

cytometer software requires that the .xit file has the same name as the folder containing the respective measurement data in order to open the measurement data for viewing in the software. The name is automatically generated from the protocol name and a timestamp, and cannot be affected. Thus, renaming just the folder breaks the possibility to open the data in the CytExpert Software of the manufacturer. Another problem was the transfer from 24 well to 96 well plates. Here, only one half of the plate is needed for each measurement. To address the sustainability concern mentioned above, the possibility was created to use the 96 well plates twice, so that both plate halves can be used for measurements. For this purpose, the table PlateUsedCols in the DB was used, which tracks the used columns of each plate (Fig. 4.6c). However, this required that measurement protocols for the left or the right side are executed in the cytometer, which further complicates tracking. Using the tables PlateUsedCols and Cytoflex_SourceTarget (Fig. 4.6c), a trace can now be created to assign the folder containing the measurement data to the respective source plate. With the help of a timestamp, which is created shortly before the Cytoflex measurement, the folder name (also contains timestamps) can be estimated with a difference of 1-2 s. This meta information is summarized in a .csv file (Fig. 4.6b) and simplifies the assignment of the files to the respective source plate, e.g., by an R or Python script.

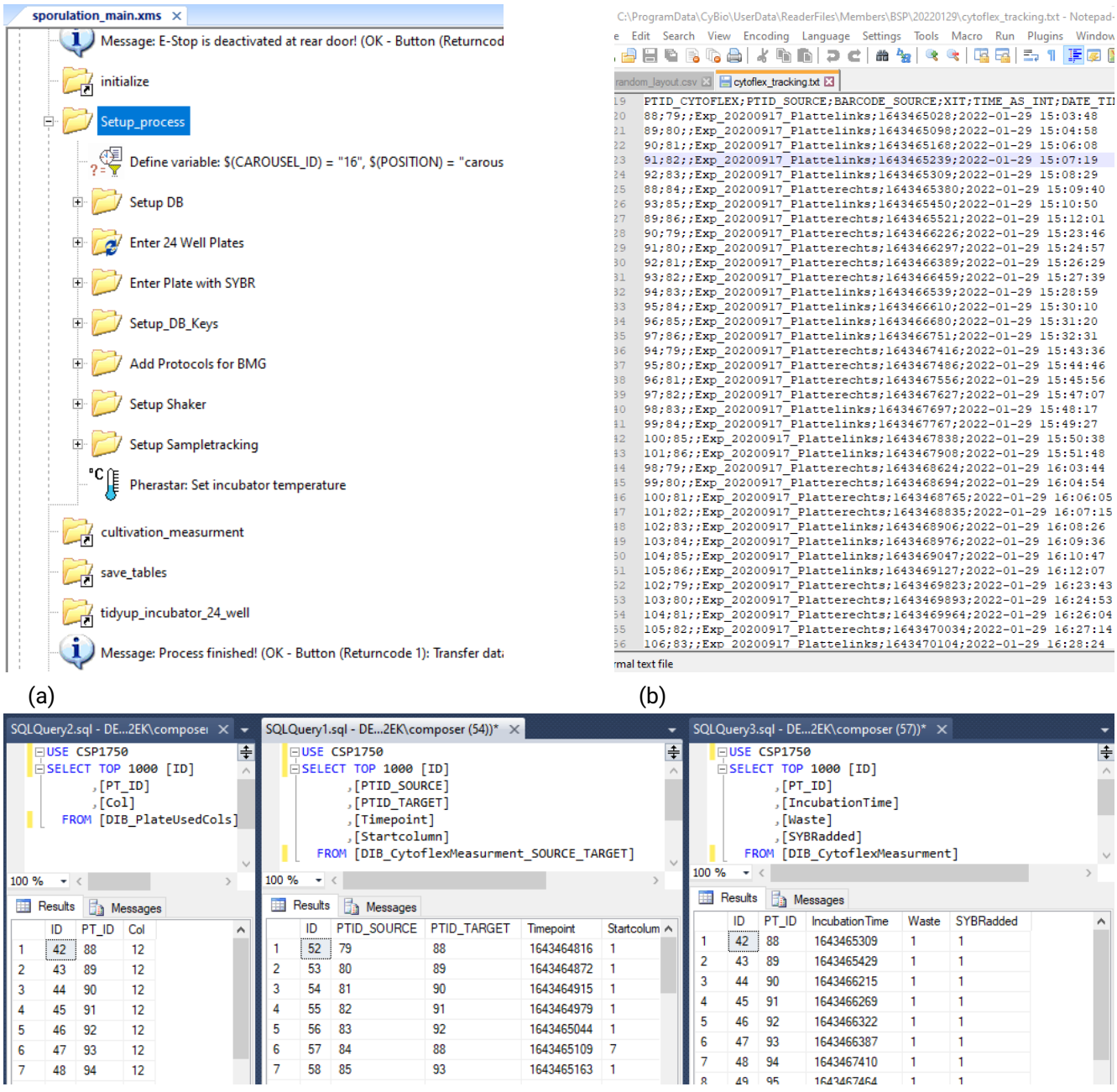


Figure 4.6.: **(a)** Main Script of the sporulation workflow. The setup section is marked, which can be used to change the main parameters of the workflow; **(b)** Output file for the Cytoflex Tracking. File summarizes several meta information: PTID of the plate in the cytoflex, PTID of the source plate, Name of the .xit file, time as UNIX time and time as data_time format; **(c) Left:** Helper table to recycle plates with PTID and until which column the plate was used; **Mid:** Helper table to generate .csv file in **(b)**; **Right:** Table to determine waiting time after SYBR® was added. In addition, contains marker column to transfer plate into waste.

4.3.4. Summary and concluding remarks

In summary, the last presented script concept reduces the time needed to set up new workflows. This is possible by reusing the jobs in other workflows. Furthermore, the parameterization of intervals, volumes, measurement protocols, etc., which are stored in the database, simplifies the reuse. Unfortunately, the use of encapsulated work packages increases the redundancy of code, so that in the last example the plate reader measurement appeared twice. However, overall, this can significantly speed up the creation of new workflows. As a positive side effect, downtime caused by programming and dry runs is significantly minimized. The presented three concepts have advantages and disadvantages that must be weighed against each other. Although the exact reproducibility of the workflows decreases when work packages are used, but it is possible to react much more flexible to any disruptions that may occur, since the system stabilizes itself again.

4.4. Randomization to improve comparability

4.4.1. Introduction and problem description

As described in section 4.1.2, the SBS standard plates introduced a certain degree of standardization. The plates can be acquired in a wide variety of designs. Basically, they are available with round or square wells. The bottoms can be V-, U- or F- shaped. There are also differences in the used compound, which naturally determines the properties. In addition, the plates can differ in the type of surface coating. The manufacturers optimize their plates depending on the application for which they are intended. For example, there are special plates which were designed for ELISA, PCR, cell culture or measurement methods like luminescence, to name just a few examples. Some plates are only used for a short time and discarded afterwards. For the cultivation of mammals, yeasts or bacteria, however, the plates are used for a longer period. For this purpose, they are incubated – depending on the equipment of the laboratory – in normal incubators with sticky pads, thermo shakers (e.g., ThermoMixer® from Eppendorf) or special incubators for microtiter plates. Many of the used incubators – like the plate holder for the ThermoMixer® – do not provide sufficient protection against evaporation. Even if a protection concept is available, there is still an increased risk of evaporation.
Foils, that are applied to the plates, work better in this context. However, the foils have to be removed each time for measurements, e.g., for absorption or fluorescence measurements. In addition, breathable foils are available on the market, but gas exchange is significantly reduced, especially when condensation of water occurs on the membrane. For automation, the use of foils requires peelers and sealers, which can be automated, so this may involve costs for additional equipment. The use of lids also reduces evaporation and protects samples. But, lids do not completely protect against evaporation. Condensation rings added to the lid by some manufacturers can further reduce but not eliminate the effect. Some models offer the possibility to fill the spaces between the wells with liquids. However, filling liquid leads to cross-contamination in the plate incubator with high shaking frequency, such as Cytomat 2 of the CompuGene Robotics platform.
Another method to reduce evaporation is used by Thermo Fisher in the incubators of the Cytomat series. According to the manufacturer, a water bath at the bottom of the incubator increases the humidity up to 94 %. However, even with high humidity, liquid still evaporates from the culture plates. In addition, the evaporation of liquid does not happen evenly. This effect is colloquially known as “edge effect” in the scientific community [155][306]. The effect appears to be a general problem, as a quick search on ResearchGate revealed. A recent study confirms the edge effect again by investigating it, using an MTS assay. Here, the metabolic activity of cell cultures was studied. The authors could clearly show that the metabolic activity of the cells decreased towards the outside [242].

The issue could be observed in own experiments and was illustrated in Fig. 4.7, where the evaporation after 72 h of incubation is shown. A reduction in volume can be clearly seen in the corner wells. In addition, a halving of the volume was observed in wells H8 and H9, which was even stronger when compared to the edges. Why these wells were severely affected cannot be clearly explained. Possibly, the active circulation of the air in the incubator has an influence on the uneven evaporation.

However, as some experiments show, this effect does not seem to have a strong influence on the growth itself. A similar observation was made by Mansoury and colleagues [242]. When evaporation varies greatly, the concentrations of media components change, decreasing possibly the comparability of the wells. Perhaps, as a result, the regular arrangement of biological replicates in columns or rows can have a negative effect on comparability [306]. Omitting the outer wells is often discussed as a solution for this problem. However, this drops considerably the number of usable wells to only 60 wells. Another solution is discussed in the following paragraph. Here, the comparability of the wells should be improved by randomizing the sample arrangement.

4.4.2. Objective

Randomization by the experimenter is very difficult to implement for well plates, especially when multiple components – like culture, media, and inducer – have to be added [155]. Online tools, like Pipette Show [99][3] or PlatR[4], can assist the pipetting step, but need to be programmed in a complex way. For this purpose, a full automation of the randomization should be implemented. For this purpose, a workflow component for Composer should be written, which handles the pipetting. Secondly, an R-function is to be created, which generates the randomized plate layout. The initial plate layout of the preculture plate should be provided by the user.

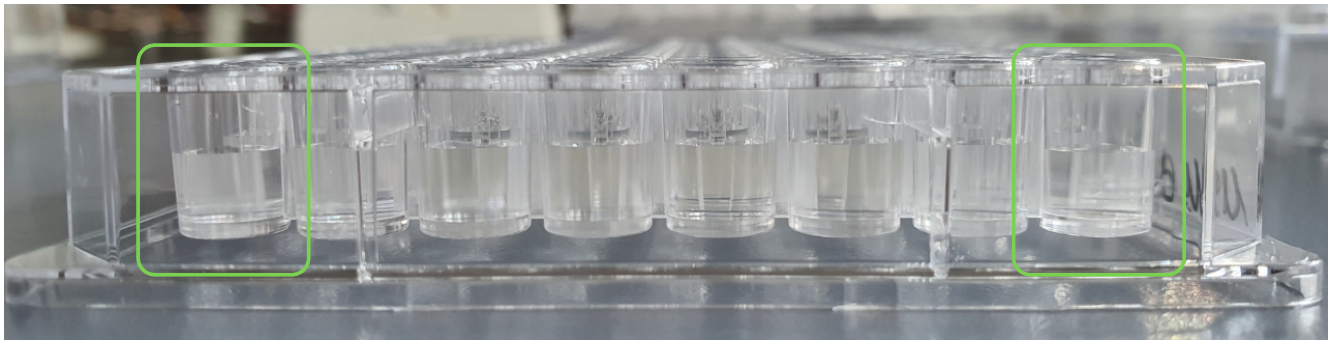
4.4.3. Results

For automated randomization of samples in high throughput, an R script for randomization as well as a Composer script for execution of the pipetting steps were developed. Both scripts communicate via the SQL server of the control computer. For illustration, the experimental procedure is shown in Fig. 4.8 and is described in the next paragraph.

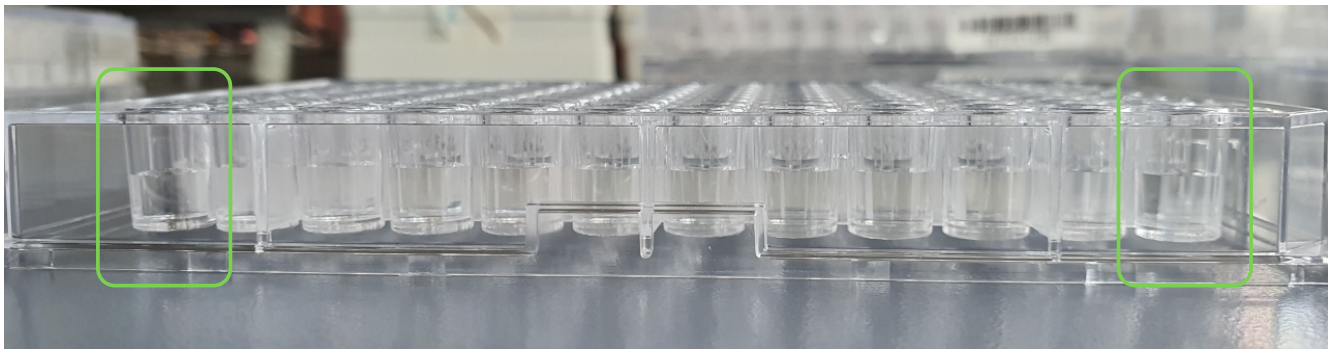
First, a preculture plate is created. The preculture plate contains the respective samples in the center of the plate, so that the outer wells of the plate can be filled with media or water to prevent evaporation. Next, the user has to enter the wet lab plate layout into a .csv file where the plate number, well, and a unique name for the sample is specified. It is sufficient to enter the wells which were filled. The plate layout is imported into the SQL database when the R script is started (see Alg. 9). Before starting, the While-loops of the R scripts PHERAread main and randomize.plate main have to be started. PHERAread waits for newly created .csv files by the PHERAstar®, whereas randomize.plate main waits for input from the Composer script. Both While-loops in the scripts are terminated when the $SHUTDOWN.REQUESTED$ variable in the DB is set to $FALSE$. When the workflow is started, the plate number from the plate layout .csv file is updated with the respective PTID. The user cannot enter the PTID into the plate layout beforehand, since the PTID is not known until the plate is introduced into the process. The precultures are incubated for a predefined interval and measured in the PHERAstar® afterwards. The R script PHERAread imports the measurement and metadata and stores them in the SQL database.

After the measurement in the PHERAstar®, the Composer script requests a randomization via the IPC

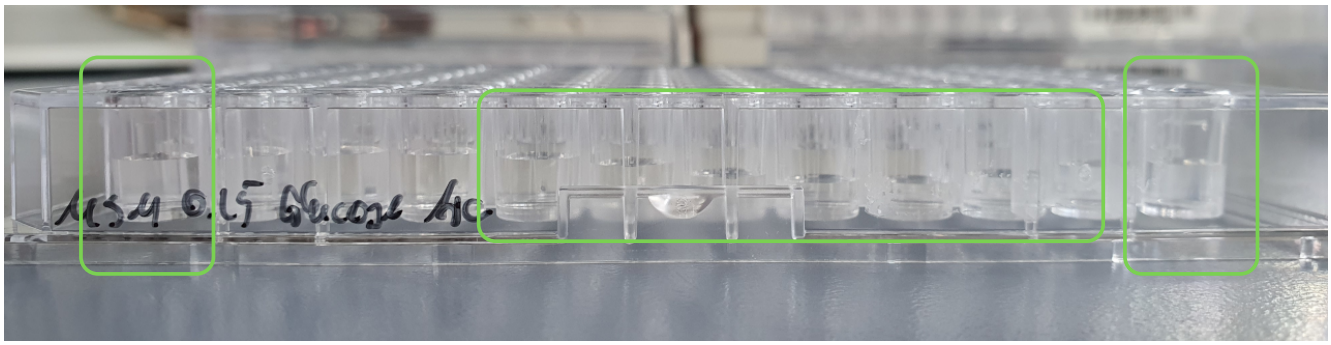
[3] https://pipette-show.de/ (date of access: 03.02.2022)
[4] https://biosistemika.com/platr/ (date of access: 03.02.2022)



(a) Col 1 with most problematic wells indicated in green rectangles.



(b) Row A with most problematic wells indicated in green rectangles.



(c) Row H with most problematic wells indicated in green rectangles.

Figure 4.7.: Illustration of the evaporation problem in the Cytomate 2 of the CompuGene robotics platform. The images were taken after 48 h of incubation. During the experiment, the OD and the fluorescence were measured every hour. Labware: Greiner Microplate, PS, F-Bottom (Order-No.: 655101) together with Greiner Lid, PS, high profil (Order-No: 656101)

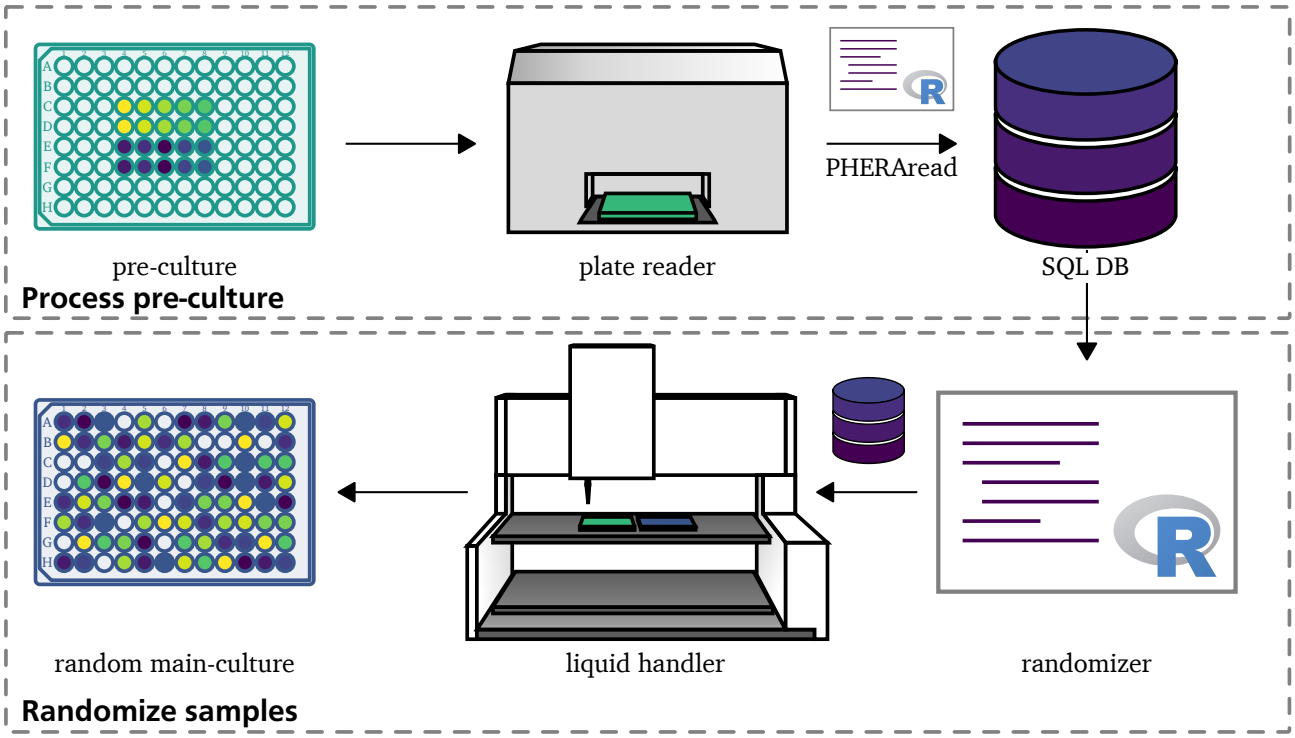


Figure 4.8.: Scheme for randomization of samples. The preculture plate is prepared by the experimenter and grown overnight in the incubator. The plate is measured in the PHERAstar® and the data is imported into the database using the R script PHERAread. Another R script uses the plate layout of the experimenter to randomly distribute samples across the plate. The randomizer script automatically creates bio replicates and blanks. The randomized layout is stored in the DB and used by Composer to control the FeliX and to pipette the main culture plate.

Algorithm 9 R pseudocode for randomization main script and the randomizer function

```
1: R script SporulationMain
2:     load required libraries
3:     establish DB connection

4:     vars ⇐ random_pos, wells, random_od...                ▷ list of required vars in DB
5:     for all vars do
6:         DB_return ⇐ create query for vars[i] and query DB
7:         if DB_return == 0 then
8:             create insert statement
9:             query DB
10:        else
11:            show info, that var is in DB
12:        end if
13:    end for

14:    Update vars to desired values                          ▷ e.g.: wells = 96, random_od = 0.05
15:    Reset tables Platelayout and PlatelayoutRandomized
16:    Import platelayout from .csv and store in Platelayout

17:    while shutdown.requested == FALSE do
18:        randomization.requested ← DB query random_start                ▷ Set by Composer
19:        if randomization.requested == TRUE then
20:            call randomizer
21:            store randomized layout in DB table PlatelayoutRandomized and as .csv
22:            while felix.done == FALSE do
23:                felix.done query from DB
24:                wait
25:            end while
26:            query PlatelayoutRandomized to get meta-information ▷ Composer enters additional infos
27:            store in .csv
28:        else
29:            show info, that randomization.requested == FALSE; do nothing
30:        end if
31:        shutdown.requested ⇐ query DB
32:        wait
33:    end while
34: end R script

35: function randomizer
36:    retrieve PTID, platelayout, plateposition, wellcount, OD values
37:    vol.preculture ⇐ calculate preculture volume from od values
38:    vol.preculture ⇐ calculate backfill volume
39:    ran.layout ⇐ create randomized layout
40:    merged ⇐ merge vol.preculture and ran.layout
41:    return merged
42: end function
```

variable $random\_start$. The randomizer function creates a new random plate layout. The following parameters are considered: target volume, target OD, 24 or 96 well plate. These parameters have to be set by the user at the beginning of the script. When the layout is created, bio replicates and blanks are automatically created. The calculation rule is developed in such a way that it creates blank wells in every case. For example: The preculture plate contains 12 different samples. This would result in 8 bio replicates on a 96 well plate. At this point, the script would automatically reduce the number of replicates to 7 and fill them up with blanks. Based on the target OD and the target volume, the amount of preculture is calculated. In addition, the backfill volume for filling with media is calculated. The created data frame is then transferred to the DB. A Composer script for the FeliX SELECT handles the pipetting. To speed up the process, a multi dispense option was created, since Composer does not offer a native multi dispense function. An additional feature tracks the volume of the preculture. If this is no longer sufficient to inoculate the target wells, a corresponding error code (Error 1) for the well is entered into the database. The finished, newly created plate (blue plate in Fig. 4.8) is then cultivated further and the preculture plate is transported to the platestore. An IPC switch signals the randomizer R script that the FeliX has completed pipetting. The R script then exports the $PlatelayoutRandomized$ table, where the randomized layout and error codes were entered, and saves it as a .csv file for later data analysis.

4.4.4. Summary and Outlook

The presented R script for randomization allows creation of a randomized layout starting from preculture plates. The purpose of randomization was to improve the comparability of samples, which would be limited by evaporation and a regular sample layout. Of course, the evaporation phenomenon only occurs in long cultivation of more than 48 h and can be reduced by reducing the measurement timepoints. Mansoury and colleagues showed that growth was not affected that much. But, the position on the plate had a significant effect on the metabolism of the studied samples [242]. Randomization by the experimenter is difficult to implement, especially when the well number increases. Nevertheless, to achieve a randomization, a block-type randomization was already developed [306].
Since the so-called edge effect cannot be prevented by randomization, the script could be further developed accordingly. Conceivable would be a switch that makes it possible to omit the edge wells. Or the outer wells could be selectively occupied by blanks. This would be particularly helpful for longer cultivation in well plates. Additionally, the Composer script should be a target for further development, since not all cones are being used for pipetting. By using the array function for the FeliX plugin, the pipetting speed could be increased. However, the complexity and readability of the script would decrease.

4.5. Colony Picking

4.5.1. Introduction

In molecular biology, situations frequently occur in which large libraries have to be created. For example, complete cDNA libraries are generated for transcriptome analyses [182]. Large quantities of plasmids are generated in biotechnology, e.g., for the creation of mutagenesis libraries in protein engineering [180]. In the latter, degenerate oligonucleotides are used to generate mutations in the target gene. Subsequently, the generated plasmids are introduced into the storage host and later into the expression host. Much larger libraries are created, when error-prone PCR is used, where entire genes are mutated [132]. In the methods mentioned so far, microorganisms are subsequently transformed by the created DNA constructs. In this process, the number of generated mutants/transformants can quickly increase to multiple 10000 [398].

The clones must then be further processed. Often, the first step is to create plates for long-term storage. At the same time, however, the first analyses are carried out. For this purpose, the obtained colonies have to be picked and transferred to new plates. Mostly, this is done manually using toothpicks, pipette tips or inoculation loops. Two problems arise with manual processing. First, sample tracking is error-prone due to manual recording in laboratory journals or LIMS systems. Second, the throughput of manual work is too low [182]. Thus, in 1992, Jones and colleagues already developed a camera-based picking system to support the human genome project [182].
For this reason, various commercial systems from different manufacturers are now established on the market. These are briefly presented below – without any claim to completeness. There are stand-alone systems but also add-on products for existing devices, like liquid handling robots. One such add-on is marketed by SciRobotics Ltd.[5] The Pickolo™ enhances existing liquid handlers manufactured by Tecan. The camera-based system brings a light table on which SBS plates, or Petri dishes can be photographed and processed. The software directly communicates with Freedom EVOware® and let the liquid handler pick with disposable tips [42]. Molecular Devices offers a complete solution with the QPix 400 series. These devices offer the possibility of both plating and subsequent detection and picking of the resulting colonies. The current versions can be integrated into existing large-scale systems for further processing of the picked clones.[6]

4.5.2. Current status and objective

A low-cost alternative was already developed by Hartley and colleagues [146]. They used a liquid handler for picking colonies. For this purpose, a camera was attached to the head of the robot, like SciRobotics did for the Pickolo™. The camera took six separate images of the agar plate. Image analysis and derivation of coordinates for the liquid handler was done using a plugin for the open-source software ImageJ [146].
The Marburg iGEM team of 2019 developed a way to convert an OT-2 from opentrons into a colony picker. This system is also camera-based. A trained network was used to identify colonies and convert them into coordinates for the liquid handler. The team trained the network for white *E. coli* colonies. However, they stated that it could be easily adapted for yeast or fluorescence, for example.[7]
At this point, a number of other open-source alternatives, both for detection and for picking, should be developed. The focus was on the reuse of existing devices that are already available in most molecular biology laboratories. Two concepts on the software side and two concepts on the hardware side should be explored. On the software side, an R pipeline was used, which should analyze exported measurement data from the BMG PHERAstar® FSX. Furthermore, a web-based server solution was planned, which would be able to analyze images from a geldoc. On the hardware side, a modified 3D printer and the FeliX SELECT from the CompuGene Robotics platform should be used for colony picking.

4.5.3. Results

Different software solutions and scripts were used depending on the measuring or recording equipment. For this purpose, tasks were divided among different groups accordingly. The Platescan script for processing the data from the BMG PHERAstar® FSX was mainly written by Dr. Felix Bracharz. The solution for image

[5] http://www.scirobotics.com/products/pickolo (date of access: 03.02.2022)
[6] https://de.moleculardevices.com/products/clone-screening/microbial-screening/qpix-400-series-microbial-colony-pickers (date of access: 03.02.2022)
[7] https://2019.igem.org/Team:Marburg/Colony_Picking (date of access: 03.02.2022)

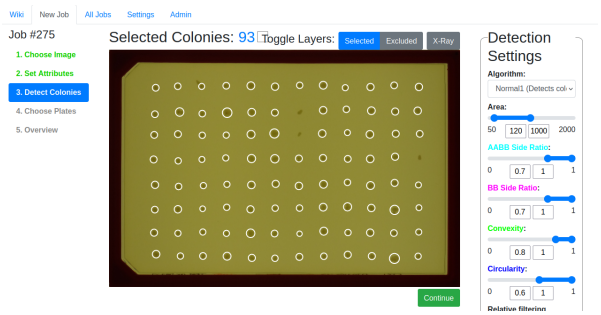
recognition (3DpickO), as well as the modification of the 3D printer were developed by the computer science students Harald Gültig, Lorenz Himmel, Patrick Mayer, Florian Szymanski and Oliver Tale-Yazdi. The import and the connection to the SQL DB of the CompuGene Robotics platform and the development of the FeliX SELECT pick script and the corresponding workflow was done by Thomas Zoll. Below, the software solutions Platescan and 3DpickO are briefly presented for completeness.

The Platescan R pipeline

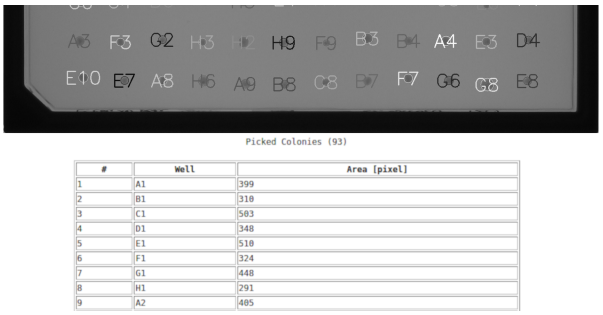
The Platescan script by FeliX Bracharz was used in combination with the PHERAstar® FSX. Through a modification – with kind support of BMG – a plate type was created that covered the largest possible area of the plate. This was achieved by creating an angular well layout, with minimal loss of area between each well. Using the matrix scan function of the PHERAstar®, the plate was scanned with a 15x15 matrix measurement. Onewell plates from Greiner were used, which were filled with 50 mL LB agar to get a constant height of the agar. The agar itself had a certain permeability for absorbance measurements. If a bacterial colony blocks the path of the light beam, the values in the absorbance measurement increased at this point. For the experiments carried out here, *E. coli* was used as test organism. After the measurement, the matrix scan data were exported as a .csv file. For some reason, the raw data could only be exported through the MARS data evaluation software, having a severe impact on full automation. The measurement software itself allowed online export of mean values for a respective well. The export of the mean values was not suitable for the detection of colonies. The exported raw data could then be imported and processed with the Platescan R package. The pipeline generated a data frame which contained additional information about the detected colonies: od.mean, od.sd, colony id, pixel (area), width, height, x, y and some others. This information was used for picking in the CompuGene Robotics Platform.

Gel documentation and 3DpickO

As an alternative for the plate reader, a geldoc was employed. This had a modular table, so that not only UV light but also transmitted light could be used. As in the previous example, Greiner Onewell plates were used, on which *E. coli* had been plated. In the first attempt, the images were taken with the Sony QX10 digital camera. The pictures were used by the computer science students to train the OpenCV image recognition library with annotated images of the agar plates. Additionally, a web application was developed as front end for the OpenCV pipeline. This web application supports the entire workflow, from reading the images until sending the job to the modified 3D printer. For this purpose, the user creates a new job by uploading a defined image of a plate to the web application. Defined image in this context means, that predefined criteria must be met, which are listed on a wiki page of the server. For better recognition of the edges of the plate, a 3D-printed frame must be placed on the plate before images are taken. Furthermore, it has to be ensured that the resolution and aspect ratio correspond to the criteria of the software. The imported image is then analyzed for colonies and subsequently appears in a mask in which the user can further influence the colony selection. Here, various parameters can be set, such as size, rounding, length-to-width ratio, etc. (Fig. 4.9a). In addition, the user can manually add or deselect colonies if colonies are not marked for picking with the help of the adjustable parameters. Afterwards, the order is created by selecting the target plate, which can be either a 96 well or a 384 well plate.
Finally, the user has the choice to download only the report or to send the pick job to the 3D printer, whereby the 3DpickO server can control the printer directly. The report contains again the respective information of the pick order including a picture, as well as an overview, into which well the colonies were transferred (Fig. 4.9b). To use the FeliX, an additional file is created with colony ID and x, y coordinate pairs, which can be imported into the SQL database of the CompuGene Robotics platform.



(a) The center shows the photographed plate with selected colonies (white circles). The right panel shows settings for the selection of colonies.

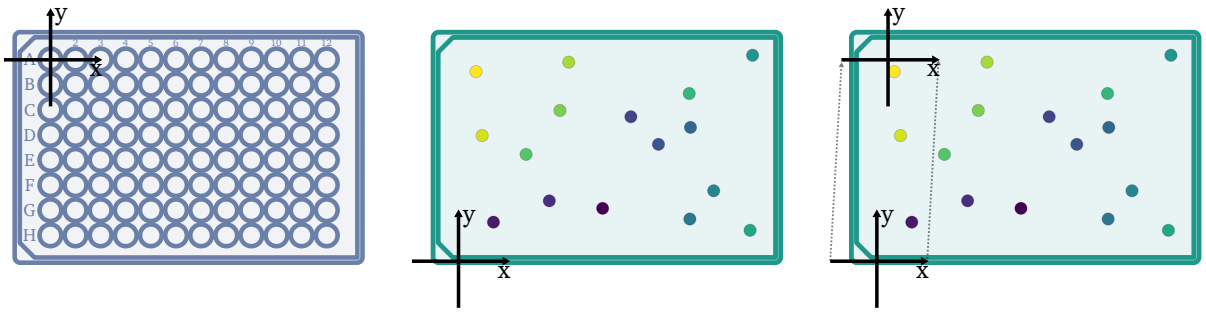


(b) The report shows once again the plate with overlay of the corresponding wells. The report also shows the colony area in px.

Figure 4.9.: Example for the usage of the 3DpickO Server for colony picking. **(a)** Step three of creating an order with 3DpickO. **(b)** Shows a report created by the 3DpickO Server.

Import functions for the CompuGene Robotics platform and the Colony Picking workflow

In the CompuGene Robotics platform, the FeliX SELECT should be used to pick colonies with its pipette tips and to transfer them to the desired target plates. For this purpose, a function of the FeliX Composer plugin is used, which makes it possible to apply an offset to the well position. In other words, the well A1 can be addressed and the target position can be adjusted in x, y and z direction by setting offsets in mm. A standard 96 well plate was used as reference layout and the well A1 was used as center of the coordinate system (Fig. 4.10a).
Both software solutions return text files with x and y values for the coordinates of the colonies. By using the 3D-printed mask, the OpenCV server is able to output the coordinates in mm, regardless of differences in zoom factor or camera resolution. The Platescan script returns values in a relative unit (hereafter referred to as units), so that a correction of the values had to be carried out first. This conversion was done experimentally. To this end, pipette tips were picked and 96 holes were made in the agar in 96 well format. The plate was measured, and the coordinates determined using the Platescan script. As a result, a mean distance of 150.17 units between the wells was determined. In the customs drawings from Greiner, the spacing of the well centers is given as 9 mm. This results in a unit/mm ratio of 16.68, which was used as a correction factor for the x and y coordinates. The shift of the 0,0 point to A1 was done by inversion of the y coordinate. Depending on the measurement method, the original 0,0 value was set differently (Fig. 4.10b), so two different R scripts were developed for the DB import. The reason was that the inversion alone was not sufficient to shift the coordinate system accordingly. For Platescan, the values still had to be shifted 3.6 mm in x and −3.74 mm in y direction. For 3DpickO, the correction was −14 mm in x direction and 11 mm in y direction (Fig. 4.10c). The respective scripts returned a data frame which was imported into the SQL DB (ColonyPicking table) of the CompuGene Robotics platform.
A Composer script, which can be embedded in cultivation workflows, was written for picking using the automation platform. Initially, the script was executed independently before a cultivation process was started. Here, a subscript checks the ColonyPicking table to determine whether further colonies should be picked. The ColonyPicking table contains the following information: row_names (ID of the colony), X, Y, Picked (colony picked (1) or not picked (0)), SourceBarcode, TargetBarcode and TargetWell. If the colony_pick_next_PTID script finds a colony that has not been picked, then the PTID of this source plate is returned, and the required material is transported to the FeliX SELECT. This includes Tip Boxes, 96 well plates with media and if requested a Onewell plate with agar to create a master plate. The picking script



(a) Coordinate system of FeliX SELECT with virtual 96 well plate

(b) Coordinate system of Platescan and 3DpickO illustrated with an Onewell plate.

(c) Shifting the coordinate system to the well A1 of the FeliX SELECT.

Figure 4.10.: Transformation of coordinate system.

retrieves a maximum of 96 unpicked colonies from the ColonyPicking table for the respective source plate, which is located on the FeliX SELECT. During the picking process, the barcode of the target plate and the current well of the target plate are entered into the ColonyPicking table. A distinction between cultivation plate and master plate is possible by entering both barcodes in the respective column: TargetBarcodeCult and TargetBarcodeMast.
During the first picking experiments, it became obvious that picking takes a relatively long time of about 18 min for a full 96 well plate, whereby transport processes are not considered (Tab. 4.4). For a cultivation with *E. coli*, where the OD and FI should be measured hourly, it is possible that from the second full plate onward, a delay in the cultivation occurs. To overcome this problem, a workflow was created that relied heavily on parallelization to address the scheduling problem. The scheduling problem was due to the Precise Flex which cannot act independently of the other tasks. To realize the free movement of the transport arm, the tasks were split into the subtasks Colony Picking and Cultivation while the transport operations were outsourced (Fig. 4.11). These paths request a transport task by making an entry in the DB. In the Assign Movement path, the request is processed, and the movement is assigned for the requesting path and executed in the Execute Next Movement path.
This optimized software could be used for various picking experiments, which are briefly presented in the next section.

Picking experiments with the FeliX SELECT

Both picking concepts were briefly summarized in Fig. 4.12. Starting from an agar plate, the plates are either scanned in the PHERAstar® or photographed in the geldoc. For the photography, it should be noted that a detection frame has to be placed on the plate beforehand to improve the detection of the transparent plate edges. On the other side, there is only one option available for the scanned plates. Here, the data are exported as raw data using the MARS software and processed via the Platescan R pipeline. Both OD and FI are available, whereby the FI depends on the installed filter set of the PHERAstar®. The coordinates are uploaded to the DB via the Platescan importer script. For the geldoc, the images are manually transferred to the 3DpickO server, and a new job is created. By means of different parameters, the user can influence the selection of the colonies (Fig. 4.9a). Afterwards, picking can be started directly via the 3D printer, or the report can be transferred to the robot DB.
Several experiments were performed to evaluate the picking success. To this end, *E. coli* was plated on LB agar plates. For the 3DpickO server, both regular patterns (96 well) and random patterns were evaluated.

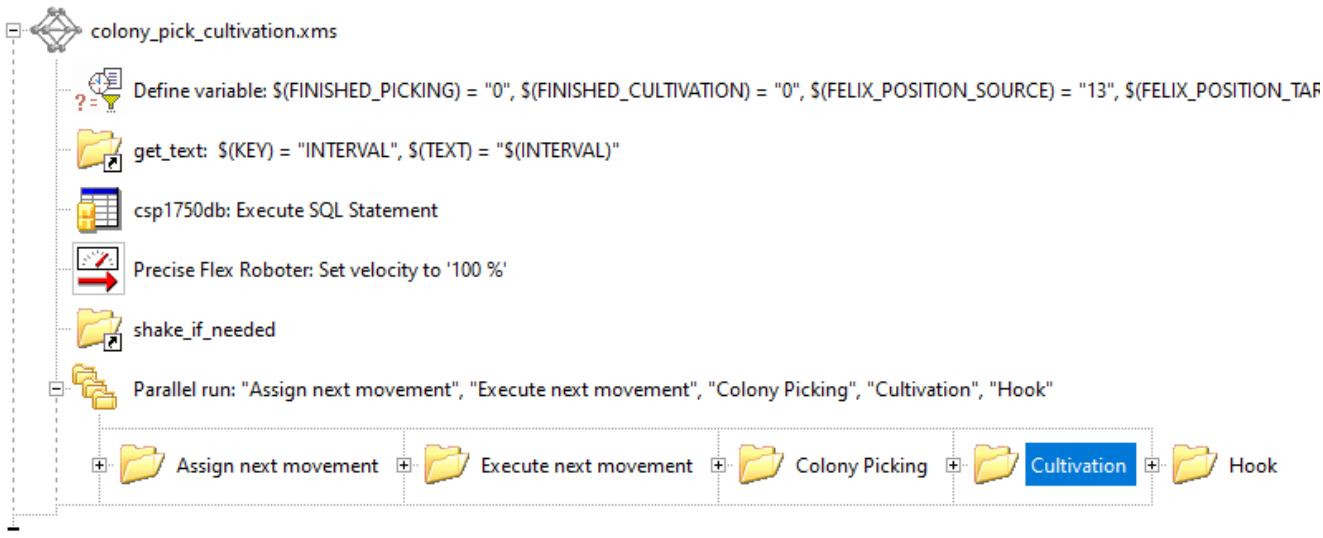


Figure 4.11.: Parallelized colony picking workflows. Tasks were divided in assigning movement, executing movement, colony picking and cultivation. The paths colony picking and cultivation requesting movements, which are assigned and executed by the respective paths.

In addition, two cameras of different price categories were used for the experiment. A Canon EOS 350D single-lens reflex camera and a Sony QX10 compact camera were compared to evaluate the influence of the camera. The results of the test were listed in Tab. 4.3. The comparability was improved by calculating the success and fail rate of each run. For this purpose, the order file, i.e., the colonies to be picked, and the cultures grown in LB were compared. An R script automatically generated a classification into grown and not grown cultures using the mean value of the OD per well for the total cultivation time. If this was above 0.05, the well was classified as grown. Detailed plots of the picked colonies and the corresponding growth plots can be found in Fig. A.26 until Fig. A.29. The picking experiment showed that there were no major differences in success rates between the two cameras. For example, both cameras delivered a success rate of around 90 % for colonies arranged in 96 well format. In contrast, the rate dropped to around 60 % for random patterns. When using 3DpickO, it was noticeable that the pick success decreased from left to right when the plate is viewed from above (Fig. A.26 and Fig. A.28). However, Platescan had higher picking success rates. Here the rate of grown cultures was about 98 % (individual evaluation in Fig. A.30, Fig. A.31 and Fig. A.32).

In addition, the time for picking was also examined for the FeliX SELECT (Tab. 4.4). Both the time with and without transport processes was extracted from the log files, since they are important to schedule the workflow accordingly. To this end, the corresponding times were taken in ms from the log files and converted. For better assessment, the duration per colony was calculated, which is about 10.7 s on average. This increased by about 4 s per colony when the transport processes for the culture plates, agar plate (source and master) and pipette tips are considered.

4.5.4. Discussion and concluding remarks

Some of the presented solutions already exist on the market and have been adopted here. For example, concepts for the use of liquid handling robots were presented earlier. For example, the Pickolo™ add-on is a commercial alternative which can pick Petri dishes and well plates. However, further investments have to

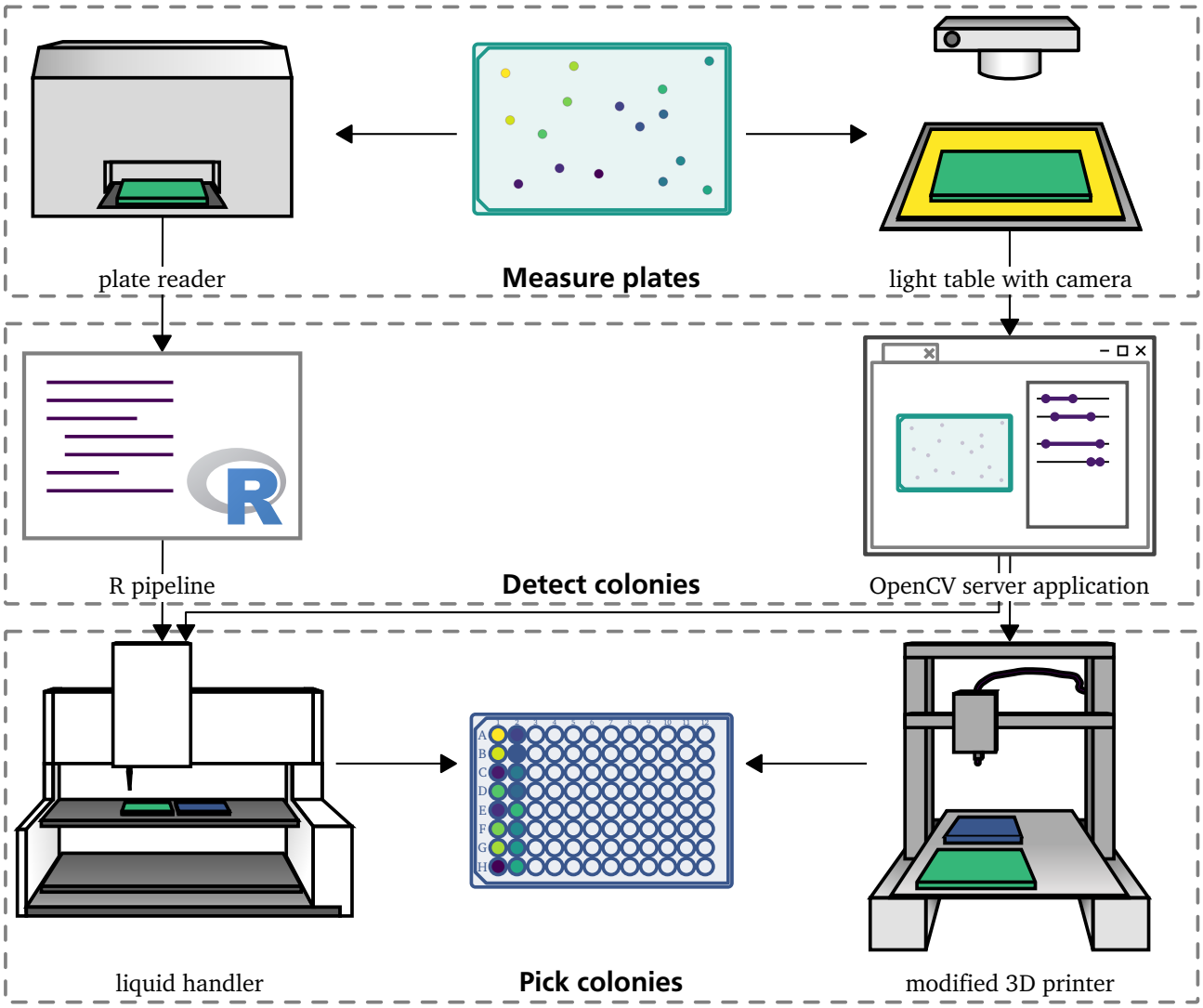


Figure 4.12.: Workflow scheme for colony picking

Table 4.3.: Picking results after detection with 3DPickO and Platescan, followed by picking with CyBio FeliX SELECT

Device	Detection	Pattern	Success in [%]	Fail in [%]
Sony QX10	3DpickO	random	61.0	39.0
Sony QX10	3DpickO	regular (96 well)	91.4	8.6
Canon EOS 350D	3DpickO	random	60.0	40.0
Canon EOS 350D	3DpickO	regular (96 well)	88.0	12.0
BMG PHERAstar FSX	Platescan	random (one target)	98.4	1.6
BMG PHERAstar FSX	Platescan	random (two targets)	98.1	1.9

Table 4.4.: Picking times for FeliX SELECT. Start and End times were taken from log files which have ms.

	without transport			with transport		
Colony Count	**61**	**91**	**94**	**61**	**91**	**94**
Duration/plate [min]	10.08	16.78	17.35	15.09	21.16	24.49
Duration/colony [s]	9.91	11.07	11.07	14.84	13.95	15.63
Mean duration/colony [s]	10.68			14.81		

be made for the existing Tecan Freedom liquid handler, including technician deployment.
Hartley and colleagues' solution offers another way to extend an existing platform [146]. The camera-based system creates images of six parts of the plate, which are then assembled using the open-source software ImageJ. But it creates six output files that are used for picking. The picking is carried out by means of the tips of the liquid handling system. A similar concept to 3DpickO was presented by the iGEM team from Marburg in 2019, where an Opentrons 2 was equipped with a light table. Here, a plate was photographed and analyzed using artificial intelligence to derive the coordinates. Finally, the colonies were also picked with the tips of the pipetting head. This solution offers the advantage that everything takes place in one device, as in one of the commercial pickers, like the QPix from Molecular Devices. In case of the 3DpickO workflow, the images have to be captured externally in the geldoc, so manual steps are performed at this point. However, if picking is done with the 3D printer, no further manual steps are necessary. Like the Opentrons solution, however, only one plate can be processed before the next user intervention is necessary. When using the FeliX, considerably less user intervention is required. For Platescan, full automation is hindered unfortunately because the export of the measurement data must be done by employing the BMG MARS software. The actual measurement software of the reader only allows the export of mean values per well. Additionally, BMG uses an unspecified DB to store the measurement data, not allowing for a workaround. However, the biggest advantage of Platescan is the sample tracking, starting from the measurement in the plate reader until the cultivation.
The previously mentioned concepts require many manual steps by the user, which could be potential sources of error. However, there are also possible sources of errors for the FeliX SELECT, possibly explaining the low success rates for the random experiment. It was noticed that the source plate has a gap of about 1 mm in x and y direction on the deck, allowing the plate to move. Additionally, the arm does not always place the plate on exactly the same spot. Depending on the colony size, this may or may not result in the colony being hit. In full automation – where the plates are replaced by the arm – the use of a clamping mechanism on the deck is therefore not practical, unfortunately. The presented options – including 3DpickO – offer a fast solution to determine the coordinates on the plates. Although Platescan offers a good hit rates, it has the disadvantage that the PHERAstar® needs about 20 min for a Onewell plate.
The presented solutions for colony picking in the CompuGene Robotics platform proved to be promising alternatives to commercial systems. Here, the quite low effort required to reuse already existing devices became apparent. Thus, the plate reader and the liquid handler from the CompuGene platform could be used without modifications to the hardware. For the use of the 3DpickO server, only a suitable camera and a geldoc with transmitted light table were necessary to pick colonies.

Bibliography

[1] M. I. Abrudan et al. “Socially mediated induction and suppression of antibiosis during bacterial coexistence”. In: *Proceedings of the National Academy of Sciences* (2015). DOI: 10.1073/pnas.1504076112.

[2] M. Adam, H. Fleischer, and K. Thurow. “Generic and Automated Data Evaluation in Analytical Measurement”. In: *SLAS TECHNOLOGY: Translating Life Sciences Innovation* (2017). DOI: 10.1177/2211068216672613.

[3] P. A. Adamczyk and J. L. Reed. “*Escherichia coli* as a model organism for systems metabolic engineering”. In: *Current Opinion in Systems Biology* (2017). DOI: 10.1016/j.coisb.2017.11.001.

[4] S. Altuvia. “The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding”. In: *The EMBO Journal* (1998). DOI: 10.1093/emboj/17.20.6069.

[5] V. Amarelle et al. “Expanding the Toolbox of Broad Host-Range Transcriptional Terminators for Proteobacteria through Metagenomics”. In: *ACS Synthetic Biology* (2019). DOI: 10.1021/acssynbio.8b00507.

[6] D. A. Anderson and C. A. Voigt. “Competitive dCas9 binding as a mechanism for transcriptional control”. In: *Molecular Systems Biology* (2021). DOI: 10.15252/msb.202110512.

[7] A. Arkin. “Setting the standard in synthetic biology”. In: *Nature Biotechnology* (2008). DOI: 10.1038/nbt0708-771.

[8] S. Arnaouteli et al. “*Bacillus subtilis* biofilm formation and social interactions”. In: *Nature Reviews Microbiology* (2021). DOI: 10.1038/s41579-021-00540-9.

[9] O. Awogbemi et al. “Advances in biotechnological applications of waste cooking oil”. In: *Case Studies in Chemical and Environmental Engineering* (2021). DOI: 10.1016/j.cscee.2021.100158.

[10] T. Baba et al. “Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection”. In: *Molecular Systems Biology* (2006). DOI: 10.1038/msb4100050.

[11] R. Babu, P. M. Prieto Veramendi, and E. R. Rene. “Strategies for resource recovery from the organic fraction of municipal solid waste”. In: *Case Studies in Chemical and Environmental Engineering* (2021). DOI: 10.1016/j.cscee.2021.100098.

[12] S. Bae, J. Park, and J.-S. Kim. “Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases”. In: *Bioinformatics* (2014). DOI: 10.1093/bioinformatics/btu048.

[13] S. Baek et al. “The yeast platform engineered for synthetic gRNA-landing pads enables multiple gene integrations by a single gRNA/Cas9 system”. In: *Metabolic Engineering* (2021). DOI: 10.1016/j.ymben.2021.01.011.

[14] M. N. Baeshen et al. “Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives”. In: *Journal of Microbiology and Biotechnology* (2015). DOI: 10.4014/jmb.1412.12079.

[15] N. A. Baeshen et al. “Cell factories for insulin production”. In: *Microbial Cell Factories* (2014). DOI: 10.1186/s12934-014-0141-0.

[16] H. Baig et al. “Synthetic Biology Open Language (SBOL) Version 3.0”. In: *Journal of Integrative Bioinformatics* (2020). DOI: 10.1515/jib-2020-0017.

[17] M. Baker. “1,500 scientists lift the lid on reproducibility”. In: *Nature News* (2016). DOI: 10.1038/533452a.

[18] R. Barrangou et al. “CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes”. In: *Science* (2007). DOI: 10.1126/science.1138140.

[19] G. Barth and C. Gaillardin. “*Yarrowia lipolytica*”. In: *Nonconventional Yeasts in Biotechnology: A Handbook*. 1996. DOI: 10.1007/978-3-642-79856-6_10.

[20] A. Baudin et al. “A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*”. In: *Nucleic Acids Research* (1993). DOI: 10.1093/nar/21.14.3329.

[21] S. Begbie et al. “The Effects of Sub-Inhibitory Levels of Chloramphenicol on pBR322 Plasmid Copy Number in *Escherichia coli* DH5α Cells”. In: *Journal of Experimental Microbiology and Immunology* (2005).

[22] A. Beopoulos et al. “The Hydrocarbon-Degrading Oleaginous Yeast *Yarrowia lipolytica*”. In: *Handbook of Hydrocarbon and Lipid Microbiology*. 2010. DOI: 10.1007/978-3-540-77587-4_152.

[23] C. Berens, A. Thain, and R. Schroeder. “A tetracycline-binding RNA aptamer”. In: *Bioorganic & Medicinal Chemistry* (2001). DOI: 10.1016/S0968-0896(01)00063-3.

[24] S. Berthoumieux et al. “Shared control of gene expression in bacteria by transcription factors and global physiology of the cell”. In: *Molecular Systems Biology* (2013). DOI: 10.1038/msb.2012.70.

[25] D. Bikard et al. “Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system”. In: *Nucleic Acids Research* (2013). DOI: 10.1093/nar/gkt520.

[26] S. Birnbaum and J. E. Bailey. “Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*”. In: *Biotechnology and Bioengineering* (1991). DOI: 10.1002/bit.260370808.

[27] F. R. Blattner et al. “The Complete Genome Sequence of *Escherichia coli* K-12”. In: *Science* (1997). DOI: 10.1126/science.277.5331.1453.

[28] J. Blazeck et al. “Generalizing a hybrid synthetic promoter approach in *Yarrowia lipolytica*”. In: *Applied Microbiology and Biotechnology* (2013). DOI: 10.1007/s00253-012-4421-5.

[29] M. Blind and M. Blank. “Aptamer Selection Technology and Recent Advances”. In: *Molecular Therapy - Nucleic Acids* (2015). DOI: 10.1038/mtna.2014.74.

[30] M. L. Bochman, C. P. Judge, and V. A. Zakian. “The Pif1 family in prokaryotes: what are our helicases doing in your bacteria?” In: *Molecular Biology of the Cell* (2011). DOI: 10.1091/mbc.e11-01-0045.

[31] M. L. Bochman, K. Paeschke, and V. A. Zakian. “DNA secondary structures: stability and function of G-quadruplex structures”. In: *Nature Reviews Genetics* (2012). DOI: 10.1038/nrg3296.

[32] U. T. Bornscheuer. “Enzymes in Lipid Modification”. In: *Annual Review of Food Science and Technology* (2018). DOI: 10.1146/annurev-food-030117-012336.

[33] R. Borriss et al. “*Bacillus subtilis*, the model Gram-positive bacterium: 20 years of annotation refinement”. In: *Microbial Biotechnology* (2018). DOI: 10.1111/1751-7915.13043.

[34] D. Botstein and G. R. Fink. “Yeast: An Experimental Organism for 21st Century Biology”. In: *Genetics* (2011). DOI: 10.1534/genetics.111.130765.

[35] L. Bourgeois, M. E. Pyne, and V. J. J. Martin. “A Highly Characterized Synthetic Landing Pad System for Precise Multicopy Gene Integration in Yeast”. In: *ACS Synthetic Biology* (2018). DOI: 10.1021/acssynbio.8b00339.

[36] C. B. Brachmann et al. “Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications”. In: *Yeast* (1998). DOI: 10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2.

[37] D. Brat, E. Boles, and B. Wiedemann. “Functional Expression of a Bacterial Xylose Isomerase in *Saccharomyces cerevisiae*”. In: *Applied and Environmental Microbiology* (2009). DOI: 10.1128/AEM.02522-08.

[38] B. P. Bratton, B. Barton, and R. M. Morgenstein. “Three-dimensional Imaging of Bacterial Cells for Accurate Cellular Representations and Precise Protein Localization”. In: *JoVE (Journal of Visualized Experiments)* (2019). DOI: 10.3791/60350.

[39] R. R. Breaker. “Prospects for Riboswitch Discovery and Analysis”. In: *Molecular Cell* (2011). DOI: 10.1016/j.molcel.2011.08.024.

[40] R. R. Breaker. “Riboswitches and the RNA World”. In: *Cold Spring Harbor Perspectives in Biology* (2012). DOI: 10.1101/cshperspect.a003566.

[41] E. K. Brinkman et al. “Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks”. In: *Molecular Cell* (2018). DOI: 10.1016/j.molcel.2018.04.016.

[42] T. Brown. *Gene Cloning and Colony Picking*. 2011.

[43] J. A. Bryant et al. “Chromosome position effects on gene expression in *Escherichia coli* K-12”. In: *Nucleic Acids Research* (2014). DOI: 10.1093/nar/gku828.

[44] J. M. Buescher et al. “Global Network Reorganization During Dynamic Adaptations of *Bacillus subtilis* Metabolism”. In: *Science* (2012). DOI: 10.1126/science.1206871.

[45] M. Bull et al. “The life history of *Lactobacillus acidophilus* as a probiotic: a tale of revisionary taxonomy, misidentification and commercial success”. In: *FEMS Microbiology Letters* (2013). DOI: 10.1111/1574-6968.12293.

[46] I. Burckhardt. “Laboratory Automation in Clinical Microbiology”. In: *Bioengineering* (2018). DOI: 10.3390/bioengineering5040102.

[47] J. M. Burg et al. “Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations”. In: *Current Opinion in Chemical Engineering* (2016). DOI: 10.1016/j.coche.2016.09.008.

[48] N. A. Burgess-Brown et al. “Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study”. In: *Protein Expression and Purification* (2008). DOI: 10.1016/j.pep.2008.01.008.

[49] S. A. Bustin et al. “The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments”. In: *Clinical Chemistry* (2009). DOI: 10.1373/clinchem.2008.112797.

[50] L. Butinar et al. “Yeast diversity in hypersaline habitats”. In: *FEMS Microbiology Letters* (2005). DOI: 10.1016/j.femsle.2005.01.043.

[51] P. Calero and P. I. Nikel. “Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms”. In: *Microbial Biotechnology* (2019). DOI: 10.1111/1751-7915.13292.

[52] E. Cámara, I. Lenitz, and Y. Nygård. “A CRISPR activation and interference toolkit for industrial *Saccharomyces cerevisiae* strain KE6-12”. In: *Scientific Reports* (2020). DOI: 10.1038/s41598-020-71648-w.

[53] D. E. Cameron, C. J. Bashor, and J. J. Collins. “A brief history of synthetic biology”. In: *Nature Reviews Microbiology* (2014). DOI: 10.1038/nrmicro3239.

[54] B. Canton, A. Labno, and D. Endy. “Refinement and standardization of synthetic biological parts and devices”. In: *Nature Biotechnology* (2008). DOI: 10.1038/nbt1413.

[55] P. Carbonell, T. Radivojevic, and H. García Martín. “Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation”. In: *ACS Synthetic Biology* (2019). DOI: 10.1021/acssynbio.8b00540.

[56] D. Carroll et al. “Design, construction and in vitro testing of zinc finger nucleases”. In: *Nature Protocols* (2006). DOI: 10.1038/nprot.2006.231.

[57] P. Ceres et al. “Modularity of Select Riboswitch Expression Platforms Enables Facile Engineering of Novel Genetic Regulatory Devices”. In: *ACS Synthetic Biology* (2013). DOI: 10.1021/sb4000096.

[58] A. Chavez et al. “Highly efficient Cas9-mediated transcriptional programming”. In: *Nature Methods* (2015). DOI: 10.1038/nmeth.3312.

[59] X. Chen et al. “Metabolic engineering of *Escherichia coli*: A sustainable industrial platform for bio-based chemical production”. In: *Biotechnology Advances* (2013). DOI: 10.1016/j.biotechadv.2013.02.009.

[60] A. A. Cheng and T. K. Lu. “Synthetic Biology: An Emerging Engineering Discipline”. In: *Annual Review of Biomedical Engineering* (2012). DOI: 10.1146/annurev-bioeng-071811-150118.

[61] A. Choudhury et al. “CRISPR/Cas9 recombineering-mediated deep mutational scanning of essential genes in *Escherichia coli*”. In: *Molecular Systems Biology* (2020). DOI: 10.15252/msb.20199265.

[62] V. Chubukov et al. “Synthetic and systems biology for microbial production of commodity chemicals”. In: *npj Systems Biology and Applications* (2016). DOI: 10.1038/npjsba.2016.9.

[63] M.-E. Chung et al. “Enhanced integration of large DNA into *E. coli* chromosome by CRISPR/Cas9”. In: *Biotechnology and Bioengineering* (2017). DOI: 10.1002/bit.26056.

[64] S. N. Cohen et al. “Construction of Biologically Functional Bacterial Plasmids In Vitro”. In: *Proceedings of the National Academy of Sciences* (1973). DOI: 10.1073/pnas.70.11.3240.

[65] J. Comley. *Automated Liquid Handler Vendor vs Third Party Plastic Company – Drug Discovery World (DDW)*. 2016.

[66] A. Cournac and J. Plumbridge. “DNA Looping in Prokaryotes: Experimental and Theoretical Approaches”. In: *Journal of Bacteriology* (2013). DOI: 10.1128/JB.02038-12.

[67] J. W. Craig et al. “Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria”. In: *Applied and Environmental Microbiology* (2010). DOI: 10.1128/AEM.02169-09.

[68] F. H. C. Crick et al. “General Nature of the Genetic Code for Proteins”. In: *Nature* (1961). DOI: 10.1038/1921227a0.

[69] J. E. Cronan. “*Escherichia coli* as an Experimental Organism”. In: *eLS*. 1st ed. 2014. DOI: 10.1002/9780470015902.a0002026.pub2.

[70] L. Cui et al. “A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9”. In: *Nature Communications* (2018). DOI: 10.1038/s41467-018-04209-5.

[71] W. Cui et al. “Exploitation of *Bacillus subtilis* as a robust workhorse for production of heterologous proteins and beyond”. In: *World Journal of Microbiology and Biotechnology* (2018). DOI: 10.1007/s11274-018-2531-7.

[72] K. A. Curran et al. “Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast”. In: *ACS Synthetic Biology* (2015). DOI: 10.1021/sb5003357.

[73] D. Dahan et al. “Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures”. In: *Nucleic Acids Research* (2018). DOI: 10.1093/nar/gky1065.

[74] H. G. Damude et al. “Docosahexaenoic acid producing strains of *Yarrowia lipolytica*”. 2014.

[75] P. Daran-Lapujade et al. “Comparative genotyping of the *Saccharomyces cerevisiae* laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays”. In: *FEMS Yeast Research* (2003). DOI: 10.1016/S1567-1356(03)00156-9.

[76] K. A. Datsenko and B. L. Wanner. “One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products”. In: *Proceedings of the National Academy of Sciences* (2000). DOI: 10.1073/pnas.120163297.

[77] M. Debiais et al. “Splitting aptamers and nucleic acid enzymes for the development of advanced biosensors”. In: *Nucleic Acids Research* (2020). DOI: 10.1093/nar/gkaa132.

[78] S. K. Desai and J. P. Gallivan. “Genetic Screens and Selections for Small Molecules Based on a Synthetic Riboswitch That Activates Protein Translation”. In: *Journal of the American Chemical Society* (2004). DOI: 10.1021/ja048634j.

[79] H. Devillers and C. Neuvéglise. “Genome Sequence of the Oleaginous Yeast *Yarrowia lipolytica* H222”. In: *Microbiology Resource Announcements* (2019). DOI: 10.1128/MRA.01547-18.

[80] H. Devillers et al. “Draft Genome Sequence of *Yarrowia lipolytica* Strain A-101 Isolated from Polluted Soil in Poland”. In: *Genome Announcements* (2016). DOI: 10.1128/genomeA.01094-16.

[81] N. Dhillon et al. “Permutational analysis of *Saccharomyces cerevisiae* regulatory elements”. In: *Synthetic Biology* (2020). DOI: 10.1093/synbio/ysaa007.

[82] J. G. Doench et al. “Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9”. In: *Nature Biotechnology* (2016). DOI: 10.1038/nbt.3437.

[83] C. Dong et al. “Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria”. In: *Nature Communications* (2018). DOI: 10.1038/s41467-018-04901-6.

[84] P. D. Donovan et al. “TPP riboswitch-dependent regulation of an ancient thiamin transporter in *Candida*”. In: *PLOS Genetics* (2018). DOI: 10.1371/journal.pgen.1007429.

[85] B. Dujon et al. “Genome evolution in yeasts”. In: *Nature* (2004). DOI: 10.1038/nature02579.

[86] R. Dulermo et al. “Using a vector pool containing variable-strength promoters to optimize protein production in *Yarrowia lipolytica*”. In: *Microbial Cell Factories* (2017). DOI: 10.1186/s12934-017-0647-3.

[87] S. Durai et al. “Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells”. In: *Nucleic Acids Research* (2005). DOI: 10.1093/nar/gki912.

[88] A. Dutta. “Adding automated data analysis and biological evaluation to affyanalysisQC”. In: *Nature Precedings* (2011). DOI: 10.1038/npre.2011.5969.1.

[89] A. M. Earl, R. Losick, and R. Kolter. “Ecology and genomics of *Bacillus subtilis*”. In: *Trends in Microbiology* (2008). DOI: 10.1016/j.tim.2008.03.004.

[90] S. R. Edwards and T. J. Wandless. “Dicistronic regulation of fluorescent proteins in the budding yeast *Saccharomyces cerevisiae*”. In: *Yeast* (2010). DOI: 10.1002/yea.1744.

[91] M. B. Elowitz and S. Leibler. “A synthetic oscillatory network of transcriptional regulators”. In: *Nature* (2000). DOI: 10.1038/35002125.

[92] S. R. Engel et al. “The Reference Genome Sequence of *Saccharomyces cerevisiae*: Then and Now”. In: *G3: Genes|Genomes|Genetics* (2013). DOI: 10.1534/g3.113.008995.

[93] J. A. Englaender et al. “Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in *E. coli*”. In: *ACS Synthetic Biology* (2017). DOI: 10.1021/acssynbio.6b00350.

[94] V. A. Erdmann et al. “Regulatory RNAs”. In: *Cellular and Molecular Life Sciences CMLS* (2001). DOI: 10.1007/PL00000913.

[95] J. Errington and L. T. v. d. 2. Aart. “Microbe Profile: *Bacillus subtilis*: model organism for cellular development, and industrial workhorse”. In: *Microbiology* (2020). DOI: 10.1099/mic.0.000922.

[96] T. M. Errington et al. “Challenges for assessing replicability in preclinical cancer biology”. In: *eLife* (2021). DOI: 10.7554/eLife.67995.

[97] T. Escherich. “The Intestinal Bacteria of the Neonate and Breast-Fed Infant”. In: *Reviews of Infectious Diseases* (1988). DOI: 10.1093/clinids/10.6.1220.

[98] M. Etschmann, P. Gebhart, and D. Sell. “Economic Benefits of the Application of Biotechnology-Examples”. In: *Engineering and Manufacturing for Biotechnology*. 2002. DOI: 10.1007/0-306-46889-1_23.

[99] J. Falk, M. Mendler, and J. Kabisch. “Pipette Show: An Open Source Web Application to Support Pipetting into Microplates”. In: *ACS Synthetic Biology* (2022). DOI: 10.1021/acssynbio.1c00494.

[100] E. R. (Farnworth. *Handbook of Fermented Functional Foods*. 2003. DOI: 10.1201/9780203009727.

[101] A. Fernández-Castané et al. “Computer-aided design for metabolic engineering”. In: *Journal of Biotechnology* (2014). DOI: 10.1016/j.jbiotec.2014.03.029.

[102] M. Ferrer, F. Martínez-Abarca, and P. N. Golyshin. “Mining genomes and ‘metagenomes’ for novel catalysts”. In: *Current Opinion in Biotechnology* (2005). DOI: 10.1016/j.copbio.2005.09.001.

[103] N. Ferrer-Miralles et al. “Microbial factories for recombinant pharmaceuticals”. In: *Microbial Cell Factories* (2009). DOI: 10.1186/1475-2859-8-17.

[104] S. Findeiß et al. “Design of Artificial Riboswitches as Biosensors”. In: *Sensors* (2017). DOI: 10.3390/s17091990.

[105] D. B. Flagfeldt et al. “Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*”. In: *Yeast* (2009). DOI: 10.1002/yea.1705.

[106] P. Foley and M. Shuler. “Considerations for the design and construction of a synthetic platform cell for biotechnological applications”. In: *Biotechnology and Bioengineering* (2010). DOI: 10.1002/bit.22575.

[107] P. Fournier et al. “Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*”. In: *Proceedings of the National Academy of Sciences* (1993). DOI: 10.1073/pnas.90.11.4912.

[108] P. Fournier et al. “Scarcity of ars sequences isolated in a morphogenesis mutant of the yeast *Yarrowia lipolytica*”. In: *Yeast* (1991). DOI: 10.1002/yea.320070104.

[109] P. S. Freemont. “Synthetic biology industry: data-driven design is creating new opportunities in biotechnology”. In: *Emerging Topics in Life Sciences* (2019). DOI: 10.1042/ETLS20190040.

[110] C. French and J. M. Ward. “Production and Modification of *E. coli* Transketolase for Large-Scale Biocatalysisa”. In: *Annals of the New York Academy of Sciences* (1996). DOI: 10.1111/j.1749-6632.1996.tb33171.x.

[111] K. Friehs. “Plasmid Copy Number and Plasmid Stability”. In: *New Trends and Developments in Biochemical Engineering*. 2004. DOI: 10.1007/b12440.

[112] Y. Fu et al. “High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells”. In: *Nature Biotechnology* (2013). DOI: 10.1038/nbt.2623.

[113] E. M. Gabor, W. B. L. Alkema, and D. B. Janssen. “Quantifying the accessibility of the metagenome by random expression cloning techniques”. In: *Environmental Microbiology* (2004). DOI: 10.1111/j.1462-2920.2004.00640.x.

[114] T. Gaj, C. A. Gersbach, and C. F. Barbas. “ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering”. In: *Trends in Biotechnology* (2013). DOI: 10.1016/j.tibtech.2013.04.004.

[115] R. García-Granados, J. A. Lerma-Escalera, and J. R. Morones-Ramírez. “Metabolic Engineering and Synthetic Biology: Synergies, Future, and Challenges”. In: *Frontiers in Bioengineering and Biotechnology* (2019). DOI: 10.3389/fbioe.2019.00036.

[116] T. S. Gardner, C. R. Cantor, and J. J. Collins. “Construction of a genetic toggle switch in *Escherichia coli*”. In: *Nature* (2000). DOI: 10.1038/35002131.

[117] A. Geerlof. *M9 mineral medium*. https://www.helmholtz-muenchen.de/pepf/protocols/expression/index.html. 2010.

[118] S. Ghaemmaghami et al. “Global analysis of protein expression in yeast”. In: *Nature* (2003). DOI: 10.1038/nature02046.

[119] G. Giaever et al. “Functional profiling of the *Saccharomyces cerevisiae* genome”. In: *Nature* (2002). DOI: 10.1038/nature00935.

[120] D. G. Gibson et al. “One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome”. In: *Proceedings of the National Academy of Sciences* (2008). DOI: 10.1073/pnas.0811011106.

[121] R. D. Gietz and R. H. Schiestl. “High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method”. In: *Nature Protocols* (2007). DOI: 10.1038/nprot.2007.13.

[122] L. A. Gilbert et al. “CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes”. In: *Cell* (2013). DOI: 10.1016/j.cell.2013.06.044.

[123] D. V. Goeddel et al. “Expression in *Escherichia coli* of chemically synthesized genes for human insulin”. In: *Proceedings of the National Academy of Sciences* (1979). DOI: 10.1073/pnas.76.1.106.

[124] A. Goffeau et al. “Life with 6000 Genes”. In: *Science* (1996). DOI: 10.1126/science.274.5287.546.

[125] B. González-Flecha and B. Demple. “Role for the *oxyS* Gene in Regulation of Intracellular Hydrogen Peroxide in *Escherichia coli*”. In: *Journal of Bacteriology* (1999). DOI: 10.1128/JB.181.12.3833-3836.1999.

[126] T. E. Gorochowski et al. “A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes”. In: *ACS Synthetic Biology* (2016). DOI: 10.1021/acssynbio.6b00040.

[127] A. A. Green et al. “Toehold Switches: De-Novo-Designed Regulators of Gene Expression”. In: *Cell* (2014). DOI: 10.1016/j.cell.2014.10.002.

[128] F. Grenier et al. “Complete Genome Sequence of *Escherichia coli* BW25113”. In: *Genome Announcements* (2014). DOI: 10.1128/genomeA.01038-14.

[129] A.-C. Groher et al. “Tuning the Performance of Synthetic Riboswitches using Machine Learning”. In: *ACS Synthetic Biology* (2019). DOI: 10.1021/acssynbio.8b00207.

[130] P. Groth and J. Cox. “Indicators for the use of robotic labs in basic biomedical research: a literature analysis”. In: *PeerJ* (2017). DOI: 10.7717/peerj.3997.

[131] N. P. Group. “Milestones in synthetic (micro)biology”. In: *Nature Reviews Microbiology* (2014). DOI: 10.1038/nrmicro3261.

[132] A. Gruet, S. Longhi, and C. Bignon. “One-step generation of error-prone PCR libraries using Gateway technology”. In: *Microbial Cell Factories* (2012). DOI: 10.1186/1475-2859-11-14.

[133] C. Gu et al. “Current status and applications of genome-scale metabolic models”. In: *Genome Biology* (2019). DOI: 10.1186/s13059-019-1730-3.

[134] S. Guiziou et al. “A part toolbox to tune genetic expression in *Bacillus subtilis*”. In: *Nucleic Acids Research* (2016). DOI: 10.1093/nar/gkw624.

[135] Z.-p. Guo et al. “An artificial chromosome ylAC enables efficient assembly of multiple genes in *Yarrowia lipolytica* for biomanufacturing”. In: *Communications Biology* (2020). DOI: 10.1038/s42003-020-0936-y.

[136] P. L. Gupta et al. “Eminence of Microbial Products in Cosmetic Industry”. In: *Natural Products and Bioprospecting* (2019). DOI: 10.1007/s13659-019-0215-0.

[137] D. Güssow and T. Clackson. “Direct clone characterization from plaques and colonies by the polymerase chain reaction”. In: *Nucleic Acids Research* (1989). DOI: 10.1093/nar/17.10.4000.

[138] O. Hädicke and S. Klamt. “EColiCore2: a reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model”. In: *Scientific Reports* (2017). DOI: 10.1038/srep39647.

[139] H. Hahne et al. “A Comprehensive Proteomics and Transcriptomics Analysis of *Bacillus subtilis* Salt Stress Adaptation”. In: *Journal of Bacteriology* (2010). DOI: 10.1128/JB.01106-09.

[140] M. Hamilton et al. “Identification of a *Yarrowia lipolytica* acetamidase and its use as a yeast genetic marker”. In: *Microbial Cell Factories* (2020). DOI: 10.1186/s12934-020-1292-9.

[141] L. Han et al. “Fabrication and characterization of a robust and strong bacterial promoter from a semi-rationally engineered promoter library in *Bacillus subtilis*”. In: *Process Biochemistry* (2017). DOI: 10.1016/j.procbio.2017.06.024.

[142] M. M. Hanczyc. “Engineering Life: A Review of Synthetic Biology”. In: *Artificial Life* (2020). DOI: 10.1162/artl_a_00318.

[143] A. D. Hanson et al. “The number of catalytic cycles in an enzyme’s lifetime and why it matters to metabolic engineering”. In: *Proceedings of the National Academy of Sciences* (2021). DOI: 10.1073/pnas.2023348118.

[144] S. Hanson et al. “Molecular analysis of a synthetic tetracycline-binding riboswitch”. In: *RNA* (2005). DOI: 10.1261/rna.7251305.

[145] C. T. Harbison et al. “Transcriptional regulatory code of a eukaryotic genome”. In: *Nature* (2004). DOI: 10.1038/nature02800.

[146] T. Hartley et al. “Cost-Effective Addition of High-Throughput Colony Picking Capability to a Standard Liquid-Handling Platform”. In: *JALA: Journal of the Association for Laboratory Automation* (2009). DOI: 10.1016/j.jala.2008.03.004.

[147] J. Hasty, D. McMillen, and J. J. Collins. “Engineered gene circuits”. In: *Nature* (2002). DOI: 10.1038/nature01257.

[148] B. D. Heavner et al. “Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance”. In: *Database* (2013). DOI: 10.1093/database/bat059.

[149] F. Heigwer et al. “E-TALEN: a web tool to design TALENs for genome engineering”. In: *Nucleic Acids Research* (2013). DOI: 10.1093/nar/gkt789.

[150] T. Hirasawa and H. Shimizu. “Recent advances in amino acid production by microbial cells”. In: *Current Opinion in Biotechnology* (2016). DOI: 10.1016/j.copbio.2016.04.017.

[151] S. Hohmann. “Characterization of PDC6, a third structural gene for pyruvate decarboxylase in *Saccharomyces cerevisiae*.” In: *Journal of Bacteriology* (1991). DOI: 10.1128/jb.173.24.7963-7969.1991.

[152] S. Hohmann. “Nobel Yeast Research”. In: *FEMS Yeast Research* (2016). DOI: 10.1093/femsyr/fow094.

[153] I. Holland and J. A. Davies. “Automation in the Life Science Research Laboratory”. In: *Frontiers in Bioengineering and Biotechnology* (2020). DOI: 10.3389/fbioe.2020.571777.

[154] Y.-K. Hong, S. D. Ontiveros, and W. M. Strauss. “A revision of the human XIST gene organization and structural comparison with mouse Xist”. In: *Mammalian Genome* (2000). DOI: 10.1007/s003350010040.

[155] J. W. L. Hooton and V. Paetkau. “Random assignment of treatments in a 96-well (8 × 12) microtiter plate A practical method”. In: *Journal of Immunological Methods* (1986). DOI: 10.1016/0022-1759(86)90218-8.

[156] J. H. Hu et al. “Evolved Cas9 variants with broad PAM compatibility and high DNA specificity”. In: *Nature* (2018). DOI: 10.1038/nature26155.

[157] G. Hubmann, J. M. Thevelein, and E. Nevoigt. “Natural and Modified Promoters for Tailored Metabolic Engineering of the Yeast *Saccharomyces cerevisiae*”. In: *Yeast Metabolic Engineering*. 2014. DOI: 10.1007/978-1-4939-0563-8_2.

[158] M. Hucka et al. “The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models”. In: *Bioinformatics* (2003). DOI: 10.1093/bioinformatics/btg015.

[159] W.-K. Huh et al. “Global analysis of protein localization in budding yeast”. In: *Nature* (2003). DOI: 10.1038/nature02026.

[160] P. Intasian et al. “Enzymes, In Vivo Biocatalysis, and Metabolic Engineering for Enabling a Circular Economy and Sustainability”. In: *Chemical Reviews* (2021). DOI: 10.1021/acs.chemrev.1c00121.

[161] N. Ishii et al. “Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations”. In: *Science* (2007). DOI: 10.1126/science.1132067.

[162] Y. Ishino et al. “Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product”. In: *Journal of Bacteriology* (1987). DOI: 10.1128/jb.169.12.5429-5433.1987.

[163] Y. Ishino, M. Krupovic, and P. Forterre. “History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology”. In: *Journal of Bacteriology* (2018). DOI: 10.1128/JB.00580-17.

[164] M. Iten et al. “Reduction of Carry over in Liquid-Handling Systems with a Decontamination Step Integrated in the Washing Procedure”. In: *JALA: Journal of the Association for Laboratory Automation* (2010). DOI: 10.1016/j.jala.2010.05.006.

[165] I. E. Ivanov et al. “Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling”. In: *Proceedings of the National Academy of Sciences* (2020). DOI: 10.1073/pnas.1913445117.

[166] A. P. Jacobus et al. “*Saccharomyces cerevisiae* strains used industrially for bioethanol production”. In: *Essays in Biochemistry* (2021). DOI: 10.1042/EBC20200160.

[167] M. Jahn et al. “Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR”. In: *Microbial Cell Factories* (2016). DOI: 10.1186/s12934-016-0610-8.

[168] R. Jansen et al. “Identification of genes that are associated with DNA repeats in prokaryotes”. In: *Molecular Microbiology* (2002). DOI: 10.1046/j.1365-2958.2002.02839.x.

[169] M. Jasin and J. E. Haber. “The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair”. In: *DNA Repair* (2016). DOI: 10.1016/j.dnarep.2016.05.001.

[170] M. K. Jensen and J. D. Keasling. “Recent applications of synthetic biology tools for yeast metabolic engineering”. In: *FEMS Yeast Research* (2015). DOI: 10.1111/1567-1364.12185.

[171] D.-E. Jeong et al. “Genome engineering using a synthetic gene circuit in *Bacillus subtilis*”. In: *Nucleic Acids Research* (2015). DOI: 10.1093/nar/gku1380.

[172] M. M. Jessop-Fabre and N. Sonnenschein. “Improving Reproducibility in Synthetic Biology”. In: *Frontiers in Bioengineering and Biotechnology* (2019). DOI: 10.3389/fbioe.2019.00018.

[173] M. M. Jessop-Fabre et al. “EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9”. In: *Biotechnology Journal* (2016). DOI: 10.1002/biot.201600147.

[174] Q. Ji et al. “Improving the homologous recombination efficiency of *Yarrowia lipolytica* by grafting heterologous component from *Saccharomyces cerevisiae*”. In: *Metabolic Engineering Communications* (2020). DOI: 10.1016/j.mec.2020.e00152.

[175] S. Ji et al. “Improved production of sublancin via introduction of three characteristic promoters into operon clusters responsible for this novel distinct glycopeptide biosynthesis”. In: *Microbial Cell Factories* (2015). DOI: 10.1186/s12934-015-0201-0.

[176] F. Jiang and J. A. Doudna. “CRISPR–Cas9 Structures and Mechanisms”. In: *Annual Review of Biophysics* (2017). DOI: 10.1146/annurev-biophys-062215-010822.

[177] F. Jiang et al. “A Cas9–guide RNA complex preorganized for target DNA recognition”. In: *Science* (2015). DOI: 10.1126/science.aab1452.

[178] Y. Jiang et al. “Multigene Editing in the *Escherichia coli* Genome via the CRISPR-Cas9 System”. In: *Applied and Environmental Microbiology* (2015). DOI: 10.1128/AEM.04023-14.

[179] M. Jinek et al. “A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity”. In: *Science* (2012). DOI: 10.1126/science.1225829.

[180] H. Jochens, D. Aerts, and U. T. Bornscheuer. “Thermostabilization of an esterase by alignment-guided focussed directed evolution”. In: *Protein Engineering, Design and Selection* (2010). DOI: 10.1093/protein/gzq071.

[181] K. L. Jones, S.-W. Kim, and J. D. Keasling. “Low-Copy Plasmids can Perform as Well as or Better Than High-Copy Plasmids for Metabolic Engineering of Bacteria”. In: *Metabolic Engineering* (2000). DOI: 10.1006/mben.2000.0161.

[182] P. Jones et al. “Integration of image analysis and robotics into a fully automated colony picking and plate handling system”. In: *Nucleic Acids Research* (1992). DOI: 10.1093/nar/20.17.4599.

[183] H. de Jong et al. “Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria”. In: *BMC Systems Biology* (2010). DOI: 10.1186/1752-0509-4-55.

[184] N. Joudeh et al. “Transcriptomic Response Analysis of *Escherichia coli* to Palladium Stress”. In: *Frontiers in Microbiology* (2021). DOI: 10.3389/fmicb.2021.741836.

[185] M. Juhas and J. W. Ajioka. “Lambda Red recombinase-mediated integration of the high molecular weight DNA into the *Escherichia coli* chromosome”. In: *Microbial Cell Factories* (2016). DOI: 10.1186/s12934-016-0571-y.

[186] D. Jullesson et al. “Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals”. In: *Biotechnology Advances* (2015). DOI: 10.1016/j.biotechadv.2015.02.011.

[187] L. Karaffa and C. P. Kubicek. “*Aspergillus niger* citric acid accumulation: do we understand this well working black box?” In: *Applied Microbiology and Biotechnology* (2003). DOI: 10.1007/s00253-002-1201-7.

[188] M. Karava, F. Bracharz, and J. Kabisch. “Quantification and isolation of *Bacillus subtilis* spores using cell sorting and automated gating”. In: *PLOS ONE* (2019). DOI: 10.1371/journal.pone.0219892.

[189] M. Karberg et al. “Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria”. In: *Nature Biotechnology* (2001). DOI: 10.1038/nbt1201-1162.

[190] M. Katz, B. M. Hover, and S. F. Brady. “Culture-independent discovery of natural products from soil metagenomes”. In: *Journal of Industrial Microbiology & Biotechnology* (2016). DOI: 10.1007/s10295-015-1706-6.

[191] M. Kavšček et al. “Optimization of lipid production with a genome-scale model of *Yarrowia lipolytica*”. In: *BMC Systems Biology* (2015). DOI: 10.1186/s12918-015-0217-4.

[192] J. D. Keasling. “Manufacturing Molecules Through Metabolic Engineering”. In: *Science* (2010). DOI: 10.1126/science.1193990.

[193] E. J. Kerkhoven et al. “Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*”. In: *npj Systems Biology and Applications* (2016). DOI: 10.1038/npjsba.2016.5.

[194] R. Khanna, I. Guler, and A. Nerkar. “Fail Often, Fail Big, and Fail Fast? Learning from Small Failures and R&D Performance in the Pharmaceutical Industry”. In: *Academy of Management Journal* (2016). DOI: 10.5465/amj.2013.1109.

[195] O. Kikin, L. D’Antonio, and P. S. Bagga. “QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences”. In: *Nucleic Acids Research* (2006). DOI: 10.1093/nar/gkl253.

[196] J. Kim et al. “A versatile and highly efficient method for scarless genome editing in *Escherichia coli* and *Salmonella enterica*”. In: *BMC Biotechnology* (2014). DOI: 10.1186/1472-6750-14-84.

[197] J. Kim et al. “Properties of alternative microbial hosts used in synthetic biology: towards the design of a modular chassis”. In: *Essays in Biochemistry* (2016). DOI: 10.1042/EBC20160015.

[198] R. D. King et al. “The Automation of Science”. In: *Science* (2009). DOI: 10.1126/science.1165620.

[199] R. Kitney et al. “Enabling the Advanced Bioeconomy through Public Policy Supporting Biofoundries and Engineering Biology”. In: *Trends in Biotechnology* (2019). DOI: 10.1016/j.tibtech.2019.03.017.

[200] G. Klauck et al. “Spatial organization of different sigma factor activities and c-di-GMP signalling within the three-dimensional landscape of a bacterial biofilm”. In: *Open Biology* (2018). DOI: 10.1098/rsob.180066.

[201] K. Kobayashi et al. “Essential *Bacillus subtilis* genes”. In: *Proceedings of the National Academy of Sciences* (2003). DOI: 10.1073/pnas.0730515100.

[202] A. C. Komor et al. “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”. In: *Nature* (2016). DOI: 10.1038/nature17946.

[203] B.-M. Koo et al. “Construction and Analysis of Two Genome-Scale Deletion Libraries for *Bacillus subtilis*”. In: *Cell Systems* (2017). DOI: 10.1016/j.cels.2016.12.013.

[204] P. Kötter et al. “A fast and efficient translational control system for conditional expression of yeast genes”. In: *Nucleic Acids Research* (2009). DOI: 10.1093/nar/gkp578.

[205] A. Kretzschmar et al. “Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining”. In: *Current Genetics* (2013). DOI: 10.1007/s00294-013-0389-7.

[206] T. E. Kuhlman and E. C. Cox. “Site-specific chromosomal integration of large synthetic constructs”. In: *Nucleic Acids Research* (2010). DOI: 10.1093/nar/gkp1193.

[207] N. G. Kuijpers et al. “A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences”. In: *Microbial Cell Factories* (2013). DOI: 10.1186/1475-2859-12-47.

[208] F. Kunst et al. “The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*”. In: *Nature* (1997). DOI: 10.1038/36786.

[209] P.-J. Lahtvee et al. “Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast”. In: *Cell Systems* (2017). DOI: 10.1016/j.cels.2017.03.003.

[210] C. Lahue et al. “History and Domestication of *Saccharomyces cerevisiae* in Bread Baking”. In: *Frontiers in Genetics* (2020). DOI: 10.3389/fgene.2020.584718.

[211] E. I. Lan and J. C. Liao. “Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources”. In: *Bioresource Technology* (2013). DOI: 10.1016/j.biortech.2012.09.104.

[212] V. Larionov et al. “Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination”. In: *Proceedings of the National Academy of Sciences* (1996). DOI: 10.1073/pnas.93.1.491.

[213] Z. Lazar, N. Liu, and G. Stephanopoulos. “Holistic Approaches in Lipid Production by *Yarrowia lipolytica*”. In: *Trends in Biotechnology* (2018). DOI: 10.1016/j.tibtech.2018.06.007.

[214] R. Ledesma-Amaro and J.-M. Nicaud. “*Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids”. In: *Progress in Lipid Research* (2016). DOI: 10.1016/j.plipres.2015.12.001.

[215] H. Lee et al. “High-Throughput Analysis of Clinical Flow Cytometry Data by Automated Gating”. In: *Bioinformatics and Biology Insights* (2019). DOI: 10.1177/1177932219838851.

[216] S.-J. Lee et al. “Development of a stationary phase-specific autoinducible expression system in *Bacillus subtilis*”. In: *Journal of Biotechnology* (2010). DOI: 10.1016/j.jbiotec.2010.06.021.

[217] J. T. Leek and R. D. Peng. “Opinion: Reproducible research can still be wrong: Adopting a prevention approach”. In: *Proceedings of the National Academy of Sciences* (2015). DOI: 10.1073/pnas.1421412111.

[218] I. R. Lehman et al. “Enzymatic Synthesis of Deoxyribonucleic Acid: I. PREPARATION OF SUBSTRATES AND PARTIAL PURIFICATION OF AN ENZYME FROM *ESCHERICHIA COLI*”. In: *Journal of Biological Chemistry* (1958). DOI: 10.1016/S0021-9258(19)68048-8.

[219] A. Lek et al. “Nuclease-Deficient Clustered Regularly Interspaced Short Palindromic Repeat-Based Approaches for In Vitro and In Vivo Gene Activation”. In: *Human Gene Therapy* (2021). DOI: 10.1089/hum.2020.241.

[220] A. M. León-Ortiz, J. Svendsen, and S. J. Boulton. “Metabolism of DNA secondary structures at the eukaryotic replication fork”. In: *DNA Repair* (2014). DOI: 10.1016/j.dnarep.2014.03.016.

[221] L.-J. Li et al. “A study on the enzymatic properties and reuse of cellulase immobilized with carbon nanotubes and sodium alginate”. In: *AMB Express* (2019). DOI: 10.1186/s13568-019-0835-0.

[222] L. Li and N. S. Han. “Application of Lactic Acid Bacteria for Food Biotechnology”. In: *Emerging Areas in Bioengineering*. 2018. DOI: 10.1002/9783527803293.ch22.

[223] C. A. Lichten et al. “Unmixing of fluorescence spectra to resolve quantitative time-series measurements of gene expression in plate readers”. In: *BMC Biotechnology* (2014). DOI: 10.1186/1472-6750-14-11.

[224] S. Lilyanna et al. “Variability in Microplate Surface Properties and Its Impact on ELISA”. In: *The Journal of Applied Laboratory Medicine: An AACC Publication* (2018). DOI: 10.1373/jalm.2017.023952.

[225] J. Y. Lim, J. W. Yoon, and C. J. Hovde. “A Brief Overview of *Escherichia coli* O157:H7 and Its Plasmid O157”. In: *Journal of Microbiology and Biotechnology* (2010). DOI: 10.4014/JMB.0908.08007.

[226] Y. Lin et al. “CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences”. In: *Nucleic Acids Research* (2014). DOI: 10.1093/nar/gku402.

[227] L. Liu et al. “Increasing expression level and copy number of a *Yarrowia lipolytica* plasmid through regulated centromere function”. In: *FEMS Yeast Research* (2014). DOI: 10.1111/1567-1364.12201.

[228] Y. Liu, X. Wan, and B. Wang. “Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria”. In: *Nature Communications* (2019). DOI: 10.1038/s41467-019-11479-0.

[229] Y. Liu and B. Wang. “A Novel Eukaryote-Like CRISPR Activation Tool in Bacteria: Features and Capabilities”. In: *BioEssays* (2020). DOI: 10.1002/bies.201900252.

[230] G. L. Lorca et al. “Catabolite Repression and Activation in *Bacillus subtilis*: Dependency on CcpA, HPr, and HprK”. In: *Journal of Bacteriology* (2005). DOI: 10.1128/JB.187.22.7826-7839.2005.

[231] R. Lorenz et al. “ViennaRNA Package 2.0”. In: *Algorithms for Molecular Biology* (2011). DOI: 10.1186/1748-7188-6-26.

[232] H. Lu et al. “A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism”. In: *Nature Communications* (2019). DOI: 10.1038/s41467-019-11581-3.

[233] J. Lu et al. “Pretreatment of immobilized Candida sp. 99-125 lipase to improve its methanol tolerance for biodiesel production”. In: *Journal of Molecular Catalysis B: Enzymatic* (2010). DOI: 10.1016/j.molcatb.2009.08.002.

[234] Z. Lu et al. “CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in *Bacillus subtilis*”. In: *Nucleic Acids Research* (2019). DOI: 10.1093/nar/gkz072.

[235] J. Ma et al. “Synthetic biology, systems biology, and metabolic engineering of *Yarrowia lipolytica* toward a sustainable biorefinery platform”. In: *Journal of Industrial Microbiology and Biotechnology* (2020). DOI: 10.1007/s10295-020-02290-8.

[236] F. Machens et al. “Synthetic Promoters and Transcription Factors for Heterologous Protein Expression in *Saccharomyces cerevisiae*”. In: *Frontiers in Bioengineering and Biotechnology* (2017). DOI: 10.3389/fbioe.2017.00063.

[237] C. Madzak, C. Gaillardin, and J.-M. Beckerich. “Heterologous protein expression and secretion in the non-conventional yeast Yarrowia lipolytica: a review”. In: *Journal of Biotechnology* (2004). DOI: 10.1016/j.jbiotec.2003.10.027.

[238] C. Madzak, B. Tréton, and S. Blanchin-Roland. “Strong Hybrid Promoters and Integrative Expression/ Secretion Vectors for Quasi-Constitutive Expression of Heterologous Proteins in the Yeast *Yarrowia lipolytica*”. In: *Journal of Molecular Microbiology and Biotechnology* (2000).

[239] T. Maeda, V. Sanchez-Torres, and T. K. Wood. “Metabolic engineering to enhance bacterial hydrogen production”. In: *Microbial Biotechnology* (2008). DOI: 10.1111/j.1751-7915.2007.00003.x.

[240] J. Mairhofer et al. “Comparative Transcription Profiling and In-Depth Characterization of Plasmid-Based and Plasmid-Free *Escherichia coli* Expression Systems under Production Conditions”. In: *Applied and Environmental Microbiology* (2013). DOI: 10.1128/AEM.00365-13.

[241] R. Mans et al. “CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*”. In: *FEMS Yeast Research* (2015). DOI: 10.1093/femsyr/fov004.

[242] M. Mansoury et al. “The edge effect: A global problem. The trouble with culturing cells in 96-well plates”. In: *Biochemistry and Biophysics Reports* (2021). DOI: 10.1016/j.bbrep.2021.100987.

[243] K. Marisch et al. “Evaluation of three industrial *Escherichia coli* strains in fed-batch cultivations during high-level SOD protein production”. In: *Microbial Cell Factories* (2013). DOI: 10.1186/1475-2859-12-58.

[244] A. Martinez et al. “Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts”. In: *Applied and Environmental Microbiology* (2004). DOI: 10.1128/AEM.70.4.2452-2463.2004.

[245] E. Martínez-García et al. “SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts”. In: *Nucleic Acids Research* (2020). DOI: 10.1093/nar/gkz1024.

[246] I. Massaiu et al. “Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains”. In: *Microbial Cell Factories* (2019). DOI: 10.1186/s12934-018-1052-2.

[247] A. M. Maxam and W. Gilbert. “A new method for sequencing DNA”. In: *Proceedings of the National Academy of Sciences* (1977). DOI: 10.1073/pnas.74.2.560.

[248] N. McGillicuddy et al. “Examining the sources of variability in cell culture media used for biopharmaceutical production”. In: *Biotechnology Letters* (2018). DOI: 10.1007/s10529-017-2437-8.

[249] A. L. Meadows et al. “Rewriting yeast central carbon metabolism for industrial isoprenoid production”. In: *Nature* (2016). DOI: 10.1038/nature19769.

[250] I. Mihalcescu et al. “Green autofluorescence, a double edged monitoring tool for bacterial growth and activity in micro-plates”. In: *Physical Biology* (2015). DOI: 10.1088/1478-3975/12/6/066016.

[251] F. J. Mojica et al. “Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements”. In: *Journal of Molecular Evolution* (2005). DOI: 10.1007/s00239-004-0046-3.

[252] A. Momen-Roknabadi et al. “An inducible CRISPR interference library for genetic interrogation of *Saccharomyces cerevisiae* biology”. In: *Communications Biology* (2020). DOI: 10.1038/s42003-020-01452-9.

[253] N. Morin et al. “Transcriptomic Analyses during the Transition from Biomass Production to Lipid Accumulation in the Oleaginous Yeast *Yarrowia lipolytica*”. In: *PLOS ONE* (2011). DOI: 10.1371/journal.pone.0027966.

[254] P. Moutsatsou et al. “Automation in cell and gene therapy manufacturing: from past to future”. In: *Biotechnology Letters* (2019). DOI: 10.1007/s10529-019-02732-z.

[255] A. Nahvi et al. “Genetic Control by a Metabolite Binding mRNA”. In: *Chemistry & Biology* (2002). DOI: 10.1016/S1074-5521(02)00224-7.

[256] G. Naseri et al. “Plant-Derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae*”. In: *ACS Synthetic Biology* (2017). DOI: 10.1021/acssynbio.7b00094.

[257] P. Neubauer, L. Häggström, and S.-O. Enfors. “Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations”. In: *Biotechnology and Bioengineering* (1995). DOI: 10.1002/bit.260470204.

[258] J.-M. Nicaud. “*Yarrowia lipolytica*”. In: *Yeast* (2012). DOI: 10.1002/yea.2921.

[259] J.-M. Nicaud et al. “Protein expression and secretion in the yeast *Yarrowia lipolytica*”. In: *FEMS Yeast Research* (2002). DOI: 10.1016/S1567-1356(02)00082-X.

[260] P. Nicolas et al. “Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*”. In: *Science* (2012). DOI: 10.1126/science.1206848.

[261] A. A. K. Nielsen, T. H. Segall-Shapiro, and C. A. Voigt. “Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression”. In: *Current Opinion in Chemical Biology* (2013). DOI: 10.1016/j.cbpa.2013.10.003.

[262] A. A. K. Nielsen et al. “Genetic circuit design automation”. In: *Science* (2016). DOI: 10.1126/science.aac7341.

[263] J. Nielsen. “Yeast Systems Biology: Model Organism and Cell Factory”. In: *Biotechnology Journal* (2019). DOI: 10.1002/biot.201800421.

[264] P. I. Nikel and D. Mattanovich. “Microbial cell factories: a biotechnology journey across species”. In: *Essays in Biochemistry* (2021). DOI: 10.1042/EBC20210037.

[265] N. Nordholt et al. “Effects of growth rate and promoter activity on single-cell protein expression”. In: *Scientific Reports* (2017). DOI: 10.1038/s41598-017-05871-3.

[266] B. Norjannah et al. “Enzymatic transesterification for biodiesel production: a comprehensive review”. In: *RSC Advances* (2016). DOI: 10.1039/C6RA08062F.

[267] M. R. O’Connell et al. “Programmable RNA recognition and cleavage by CRISPR/Cas9”. In: *Nature* (2014). DOI: 10.1038/nature13769.

[268] T. L. Orr-Weaver, J. W. Szostak, and R. J. Rothstein. “Yeast transformation: a model system for the study of recombination”. In: *Proceedings of the National Academy of Sciences* (1981). DOI: 10.1073/pnas.78.10.6354.

[269] J. D. Orth et al. “A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism - 2011”. In: *Molecular Systems Biology* (2011). DOI: 10.1038/msb.2011.65.

[270] I. Otero-Muras and P. Carbonell. “Automated engineering of synthetic metabolic pathways for efficient biomanufacturing”. In: *Metabolic Engineering* (2021). DOI: 10.1016/j.ymben.2020.11.012.

[271] C. J. Paddon and J. D. Keasling. “Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development”. In: *Nature Reviews Microbiology* (2014). DOI: 10.1038/nrmicro3240.

[272] K. Paeschke and P. Burkovics. “Mgs1 function at G-quadruplex structures during DNA replication”. In: *Current Genetics* (2021). DOI: 10.1007/s00294-020-01128-1.

[273] K. Paeschke, J. A. Capra, and V. A. Zakian. “DNA Replication through G-Quadruplex Motifs Is Promoted by the *Saccharomyces cerevisiae* Pif1 DNA Helicase”. In: *Cell* (2011). DOI: 10.1016/j.cell.2011.04.015.

[274] G. Palermo et al. “CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations”. In: *Proceedings of the National Academy of Sciences* (2017). DOI: 10.1073/pnas.1707645114.

[275] Y. Park et al. “Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads”. In: *Molecular Systems Biology* (2020). DOI: 10.15252/msb.20209584.

[276] Y.-K. Park et al. “Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*”. In: *FEMS Yeast Research* (2019). DOI: 10.1093/femsyr/foy105.

[277] N. Pavlova, D. Kaloudas, and R. Penchovsky. “Riboswitch distribution, structure, and function in bacteria”. In: *Gene* (2019). DOI: 10.1016/j.gene.2019.05.036.

[278] K. Pedersen and K. Gerdes. “Multiple hok genes on the chromosome of *Escherichia coli*”. In: *Molecular Microbiology* (1999). DOI: 10.1046/j.1365-2958.1999.01431.x.

[279] B. Peng et al. “Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities”. In: *Microbial Cell Factories* (2015). DOI: 10.1186/s12934-015-0278-5.

[280] B. Peng et al. “An Expanded Heterologous GAL Promoter Collection for Diauxie-Inducible Expression in *Saccharomyces cerevisiae*”. In: *ACS Synthetic Biology* (2018). DOI: 10.1021/acssynbio.7b00355.

[281] J. Peter et al. “Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates”. In: *Nature* (2018). DOI: 10.1038/s41586-018-0030-5.

[282] C. Peterbauer, T. Maischberger, and D. Haltrich. “Food-grade gene expression in lactic acid bacteria”. In: *Biotechnology Journal* (2011). DOI: 10.1002/biot.201100034.

[283] J. M. Peters et al. “A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria”. In: *Cell* (2016). DOI: 10.1016/j.cell.2016.05.003.

[284] A. Pickar-Oliver and C. A. Gersbach. “The next generation of CRISPR–Cas technologies and applications”. In: *Nature Reviews Molecular Cell Biology* (2019). DOI: 10.1038/s41580-019-0131-5.

[285] D. J. Pitera et al. “Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*”. In: *Metabolic Engineering* (2007). DOI: 10.1016/j.ymben.2006.11.002.

[286] M. Plebani. “The detection and prevention of errors in laboratory medicine”. In: *Annals of Clinical Biochemistry* (2010). DOI: 10.1258/acb.2009.009222.

[287] K. R. Pomraning et al. “Multi-omics analysis reveals regulators of the response to nitrogen limitation in *Yarrowia lipolytica*”. In: *BMC Genomics* (2016). DOI: 10.1186/s12864-016-2471-2.

[288] P. F. Popp et al. “The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*”. In: *Scientific Reports* (2017). DOI: 10.1038/s41598-017-15107-z.

[289] C. Pourcel, G. Salvignol, and G. Vergnaud. “CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies”. In: *Microbiology* (2005). DOI: 10.1099/mic.0.27437-0.

[290] I. D. Puspita et al. “Are Uncultivated Bacteria Really Uncultivable?” In: *Microbes and Environments* (2012). DOI: 10.1264/jsme2.ME12092.

[291] M. E. Pyne et al. “Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in *Escherichia coli*”. In: *Applied and Environmental Microbiology* (2015). DOI: 10.1128/AEM.01248-15.

[292] L. S. Qi et al. “Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”. In: *Cell* (2013). DOI: 10.1016/j.cell.2013.02.022.

[293] C. Qiu et al. “Universal promoter scanning by Pol II during transcription initiation in *Saccharomyces cerevisiae*”. In: *Genome Biology* (2020). DOI: 10.1186/s13059-020-02040-0.

[294] J. Quarterman et al. “A survey of yeast from the *Yarrowia* clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate”. In: *Applied Microbiology and Biotechnology* (2017). DOI: 10.1007/s00253-016-8062-y.

[295] A. Rantasalo et al. “Synthetic Toolkit for Complex Genetic Circuit Engineering in *Saccharomyces cerevisiae*”. In: *ACS Synthetic Biology* (2018). DOI: 10.1021/acssynbio.8b00076.

[296] A. Reider Apel et al. “A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*”. In: *Nucleic Acids Research* (2017). DOI: 10.1093/nar/gkw1023.

[297] O. Rengby and E. S. J. Arnér. “Titration and Conditional Knockdown of the prfB Gene in *Escherichia coli*: Effects on Growth and Overproduction of the Recombinant Mammalian Selenoprotein Thioredoxin Reductase”. In: *Applied and Environmental Microbiology* (2007). DOI: 10.1128/AEM.02019-06.

[298] S. Renggli et al. “Role of Autofluorescence in Flow Cytometric Analysis of *Escherichia coli* Treated with Bactericidal Antibiotics”. In: *Journal of Bacteriology* (2013). DOI: 10.1128/JB.00393-13.

[299] C. G. Ricci et al. “Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics”. In: *ACS Central Science* (2019). DOI: 10.1021/acscentsci.9b00020.

[300] G.-F. Richard et al. “Comparative Genomics of Hemiascomycete Yeasts: Genes Involved in DNA Replication, Repair, and Recombination”. In: *Molecular Biology and Evolution* (2005). DOI: 10.1093/molbev/msi083.

[301] J. Richardson et al. “Metabolomics analysis of soy hydrolysates for the identification of productivity markers of mammalian cells for manufacturing therapeutic proteins”. In: *Biotechnology Progress* (2015). DOI: 10.1002/btpr.2050.

[302] P. K. Robinson. “Enzymes: principles and biotechnological applications”. In: *Essays in Biochemistry* (2015). DOI: 10.1042/bse0590001.

[303] M. A. Romanos, C. A. Scorer, and J. J. Clare. “Foreign gene expression in yeast: a review”. In: *Yeast* (1992). DOI: 10.1002/yea.320080602.

[304] C. Ronda et al. “CRMAGE: CRISPR Optimized MAGE Recombineering”. In: *Scientific Reports* (2016). DOI: 10.1038/srep19452.

[305] G. L. Rosano and E. A. Ceccarelli. “Recombinant protein expression in *Escherichia coli*: advances and challenges”. In: *Frontiers in Microbiology* (2014). DOI: 10.3389/fmicb.2014.00172.

[306] C. Roselle, T. Verch, and M. Shank-Retzlaff. “Mitigation of microtiter plate positioning effects using a block randomization scheme”. In: *Analytical and Bioanalytical Chemistry* (2016). DOI: 10.1007/s00216-016-9469-0.

[307] E. Russo. “Special Report: The birth of biotechnology”. In: *Nature* (2003). DOI: 10.1038/nj6921-456a.

[308] S. Ryu, J. Hipp, and C. T. Trinh. “Activating and Elucidating Complex Sugar Metabolism in *Yarrowia lipolytica*”. In: *Applied and Environmental Microbiology* (2015). DOI: 10.1128/AEM.03582-15.

[309] D. Samuel. “Investigation of Ancient Egyptian Baking and Brewing Methods by Correlative Microscopy”. In: *Science* (1996). DOI: 10.1126/science.273.5274.488.

[310] F. Sanger et al. “Nucleotide sequence of bacteriophage φX174 DNA”. In: *Nature* (1977). DOI: 10.1038/265687a0.

[311] R. Sapranauskas et al. “The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*”. In: *Nucleic Acids Research* (2011). DOI: 10.1093/nar/gkr606.

[312] A. Satomura et al. “Precise genome-wide base editing by the CRISPR Nickase system in yeast”. In: *Scientific Reports* (2017). DOI: 10.1038/s41598-017-02013-7.

[313] M. Sauer and K. Paeschke. “G-quadruplex unwinding helicases and their function *in vivo*”. In: *Biochemical Society Transactions* (2017). DOI: 10.1042/BST20170097.

[314] M. Sauer and N. S. Han. “Lactic acid bacteria: little helpers for many human tasks”. In: *Essays in Biochemistry* (2021). DOI: 10.1042/EBC20200133.

[315] R. O. Schlechter et al. “Chromatic Bacteria – A Broad Host-Range Plasmid and Chromosomal Insertion Toolbox for Fluorescent Protein Expression in Bacteria”. In: *Frontiers in Microbiology* (2018). DOI: 10.3389/fmicb.2018.03052.

[316] N. C. Schlichting. “Optimization and automation of the ligase cycling reaction”. 2020.

[317] G. Schmidt. “Beschleunigung von CRISPR- und Metabolic Engineering Anwendungen in *Bacillus subtilis* durch artifizielle Landing Pads”. 2020.

[318] C. Schwartz et al. “CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*”. In: *Biotechnology and Bioengineering* (2017). DOI: 10.1002/bit.26404.

[319] C. Schwartz et al. “Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables *Yarrowia lipolytica* Growth on Cellobiose”. In: *Biotechnology Journal* (2018). DOI: 10.1002/biot.201700584.

[320] T. H. Segall-Shapiro, E. D. Sontag, and C. A. Voigt. “Engineered promoters enable constant gene expression at any copy number in bacteria”. In: *Nature Biotechnology* (2018). DOI: 10.1038/nbt.4111.

[321] T. Selas Castiñeiras et al. “*E. coli* strain engineering for the production of advanced biopharmaceutical products”. In: *FEMS Microbiology Letters* (2018). DOI: 10.1093/femsle/fny162.

[322] D. O. Serra, G. Klauck, and R. Hengge. “Vertical stratification of matrix production is essential for physical integrity and architecture of macrocolony biofilms of *Escherichia coli*”. In: *Environmental Microbiology* (2015). DOI: 10.1111/1462-2920.12991.

[323] V. Sewalt et al. “The Generally Recognized as Safe (GRAS) Process for Industrial Microbial Enzymes”. In: *Industrial Biotechnology* (2016). DOI: 10.1089/ind.2016.0011.

[324] B. Shao et al. “Single-cell measurement of plasmid copy number and promoter activity”. In: *Nature Communications* (2021). DOI: 10.1038/s41467-021-21734-y.

[325] A. Shevchenko et al. “Proteomics identifies the composition and manufacturing recipe of the 2500-year old sourdough bread from Subeixi cemetery in China”. In: *Journal of Proteomics* (2014). DOI: 10.1016/j.jprot.2013.11.016.

[326] W. Z. Shou and J. Zhang. “Recent development in software and automation tools for high-throughput discovery bioanalysis”. In: *Bioanalysis* (2012). DOI: 10.4155/bio.12.51.

[327] K. Shukla et al. “*Escherichia coli* and *Neisseria gonorrhoeae* UvrD helicase unwinds G4 DNA structures”. In: *Biochemical Journal* (2017). DOI: 10.1042/BCJ20170587.

[328] R. Silva-Rocha et al. “The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes”. In: *Nucleic Acids Research* (2013). DOI: 10.1093/nar/gks1119.

[329] D. Singh et al. “Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9”. In: *Nature Communications* (2016). DOI: 10.1038/ncomms12778.

[330] A. L. Slusarczyk, A. Lin, and R. Weiss. “Foundations for the design and implementation of synthetic genetic circuits”. In: *Nature Reviews Genetics* (2012). DOI: 10.1038/nrg3227.

[331] C. W. Song and S. Y. Lee. “Rapid one-step inactivation of single or multiple genes in *Escherichia coli*”. In: *Biotechnology Journal* (2013). DOI: 10.1002/biot.201300153.

[332] Y. Song et al. “Promoter Screening from *Bacillus subtilis* in Various Conditions Hunting for Synthetic Biology and Industrial Applications”. In: *PLOS ONE* (2016). DOI: 10.1371/journal.pone.0158447.

[333] C.-S. Soo et al. “Co-production of hydrogen and ethanol by *Escherichia coli* SS1 and its recombinant”. In: *Electronic Journal of Biotechnology* (2017). DOI: 10.1016/j.ejbt.2017.09.002.

[334] H. P. Sørensen and K. K. Mortensen. “Advanced genetic strategies for recombinant protein expression in *Escherichia coli*”. In: *Journal of Biotechnology* (2005). DOI: 10.1016/j.jbiotec.2004.08.004.

[335] B. Soufi et al. “Characterization of the *E. coli* proteome and its modifications during growth and ethanol stress”. In: *Frontiers in Microbiology* (2015). DOI: 10.3389/fmicb.2015.00103.

[336] C. Sousa, V. de Lorenzo, and A. Cebolla. “Modulation of gene expression through chromosomal positioning in *Escherichia coli*”. In: *Microbiology* (1997). DOI: 10.1099/00221287-143-6-2071.

[337] V. Stein and K. Alexandrov. “Synthetic protein switches: design principles and applications”. In: *Trends in Biotechnology* (2015). DOI: 10.1016/j.tibtech.2014.11.010.

[338] G. Stephanopoulos. “Metabolic Fluxes and Metabolic Engineering”. In: *Metabolic Engineering* (1999). DOI: 10.1006/mben.1998.0101.

[339] G. Stephanopoulos. “Synthetic Biology and Metabolic Engineering”. In: *ACS Synthetic Biology* (2012). DOI: 10.1021/sb300094q.

[340] S. H. Sternberg et al. “DNA interrogation by the CRISPR RNA-guided endonuclease Cas9”. In: *Nature* (2014). DOI: 10.1038/nature13011.

[341] S. H. Sternberg et al. “Conformational control of DNA target cleavage by CRISPR–Cas9”. In: *Nature* (2015). DOI: 10.1038/nature15544.

[342] A. Stevens. “Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* B”. In: *Biochemical and Biophysical Research Communications* (1960). DOI: 10.1016/0006-291X(60)90110-8.

[343] K. Stevenson et al. “General calibration of microbial growth in microplate readers”. In: *Scientific Reports* (2016). DOI: 10.1038/srep38828.

[344] F. W. Studier and B. A. Moffatt. “Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes”. In: *Journal of Molecular Biology* (1986). DOI: 10.1016/0022-2836(86)90385-2.

[345] Y. Su et al. “*Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine”. In: *Microbial Cell Factories* (2020). DOI: 10.1186/s12934-020-01436-8.

[346] B. Suess et al. “Conditional gene expression by controlling translation with tetracycline-binding aptamers”. In: *Nucleic Acids Research* (2003). DOI: 10.1093/nar/gkg285.

[347] B. Suess et al. “A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo”. In: *Nucleic Acids Research* (2004). DOI: 10.1093/nar/gkh321.

[348] J. Sun et al. “Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*”. In: *Biotechnology and Bioengineering* (2012). DOI: 10.1002/bit.24481.

[349] J. Surre et al. “Strong increase in the autofluorescence of cells signals struggle for survival”. In: *Scientific Reports* (2018). DOI: 10.1038/s41598-018-30623-2.

[350] R. Swannell. *The Role of Business in Moving from Linear to Circular Economies*. 2021.

[351] F. M. Szczebara et al. “Total biosynthesis of hydrocortisone from a simple carbon source in yeast”. In: *Nature Biotechnology* (2003). DOI: 10.1038/nbt775.

[352] M. D. Szczelkun et al. “Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes”. In: *Proceedings of the National Academy of Sciences* (2014). DOI: 10.1073/pnas.1402597111.

[353] K. Szlachta et al. “Alternative DNA secondary structure formation affects RNA polymerase II promoter-proximal pausing in human”. In: *Genome Biology* (2018). DOI: 10.1186/s13059-018-1463-8.

[354] S. Tapsin et al. “Genome-wide identification of natural RNA aptamers in prokaryotes and eukaryotes”. In: *Nature Communications* (2018). DOI: 10.1038/s41467-018-03675-1.

[355] A. Taton et al. “Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria”. In: *Nucleic Acids Research* (2014). DOI: 10.1093/nar/gku673.

[356] J. A. Teixeira da Silva. “Room temperature in scientific protocols and experiments should be defined: a reproducibility issue”. In: *BioTechniques* (2021). DOI: 10.2144/btn-2020-0131.

[357] F. Thoma and B. Blombach. “Metabolic engineering of *Vibrio natriegens*”. In: *Essays in Biochemistry* (2021). DOI: 10.1042/EBC20200135.

[358] J. Thykaer and J. Nielsen. “Metabolic engineering of β-lactam production”. In: *Metabolic Engineering* (2003). DOI: 10.1016/S1096-7176(03)00003-X.

[359] Z. Tong et al. “Systems metabolic engineering for citric acid production by *Aspergillus niger* in the post-genomic era”. In: *Microbial Cell Factories* (2019). DOI: 10.1186/s12934-019-1064-6.

[360] M. Trassaert et al. “New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*”. In: *Microbial Cell Factories* (2017). DOI: 10.1186/s12934-017-0755-0.

[361] K. Trchounian and A. Trchounian. “Hydrogen production from glycerol by *Escherichia coli* and other bacteria: An overview and perspectives”. In: *Applied Energy* (2015). DOI: 10.1016/j.apenergy.2015.07.009.

[362] S. Q. Tsai et al. “GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases”. In: *Nature Biotechnology* (2015). DOI: 10.1038/nbt.3117.

[363] B. J. Tucker and R. R. Breaker. “Riboswitches as versatile gene control elements”. In: *Current Opinion in Structural Biology* (2005). DOI: 10.1016/j.sbi.2005.05.003.

[364] C. Tuerk and L. Gold. “Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase”. In: *Science* (1990). DOI: 10.1126/science.2200121.

[365] J. H. Urban and J. Vogel. “Translational control and target recognition by *Escherichia coli* small RNAs in vivo”. In: *Nucleic Acids Research* (2007). DOI: 10.1093/nar/gkl1040.

[366] M. A. Urbina, A. J. R. Watts, and E. E. Reardon. “Labs should cut plastic waste too”. In: *Nature* (2015). DOI: 10.1038/528479c.

[367] B. Valderrama, M. Ayala, and R. Vazquez-Duhalt. “Suicide Inactivation of Peroxidases and the Challenge of Engineering More Robust Enzymes”. In: *Chemistry & Biology* (2002). DOI: 10.1016/S1074-5521(02)00149-7.

[368] K. M. Vasquez et al. “Manipulating the mammalian genome by homologous recombination”. In: *Proceedings of the National Academy of Sciences* (2001). DOI: 10.1073/pnas.111009698.

[369] A. Vaughan-Martini and A. Martini. “Facts, myths and legends on the prime industrial microorganism”. In: *Journal of Industrial Microbiology* (1995). DOI: 10.1007/BF01573967.

[370] J. Verbeke, A. Beopoulos, and J.-M. Nicaud. “Efficient homologous recombination with short length flanking fragments in Ku70 deficient *Yarrowia lipolytica* strains”. In: *Biotechnology Letters* (2013). DOI: 10.1007/s10529-012-1107-0.

[371] L. Vernis et al. “An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica*.” In: *Molecular and Cellular Biology* (1997). DOI: 10.1128/MCB.17.4.1995.

[372] L. Vojcic et al. “An efficient transformation method for *Bacillus subtilis* DB104”. In: *Applied Microbiology and Biotechnology* (2012). DOI: 10.1007/s00253-012-3987-2.

[373] A. Wach et al. “New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*”. In: *Yeast* (1994). DOI: 10.1002/yea.320101310.

[374] B. D. Wahlen et al. “Biodiesel from Microalgae, Yeast, and Bacteria: Engine Performance and Exhaust Emissions”. In: *Energy & Fuels* (2013). DOI: 10.1021/ef3012382.

[375] J. P. v. d. Walt and J. A. v. Arx. “The yeast genus *Yarrowia*”. In: *Antonie van Leeuwenhoek* (1980). DOI: 10.1007/BF00394008.

[376] B. Wang et al. “Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology”. In: *Nature Communications* (2011). DOI: 10.1038/ncomms1516.

[377] C. Wang et al. “Purification and Characterization of Nattokinase from *Bacillus subtilis* Natto B-12”. In: *Journal of Agricultural and Food Chemistry* (2009). DOI: 10.1021/jf901861v.

[378] G. Wang et al. “CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria”. In: *Nature Microbiology* (2019). DOI: 10.1038/s41564-019-0573-8.

[379] L. Wang et al. “A review of computational tools for design and reconstruction of metabolic pathways”. In: *Synthetic and Systems Biotechnology* (2017). DOI: 10.1016/j.synbio.2017.11.002.

[380] G. Węgrzyn. “Replication of Plasmids during Bacterial Response to Amino Acid Starvation”. In: *Plasmid* (1999). DOI: 10.1006/plas.1998.1377.

[381] J. E. Weigand et al. “Screening for engineered neomycin riboswitches that control translation initiation”. In: *RNA* (2008). DOI: 10.1261/rna.772408.

[382] L. Wensing et al. “A CRISPR Interference Platform for Efficient Genetic Repression in *Candida albicans*”. In: *mSphere* (2019). DOI: 10.1128/mSphere.00002-19.

[383] G. Werstuck and M. R. Green. “Controlling Gene Expression in Living Cells Through Small Molecule-RNA Interactions”. In: *Science* (1998). DOI: 10.1126/science.282.5387.296.

[384] M. Wieland and J. S. Hartig. “Artificial Riboswitches: Synthetic mRNA-Based Regulators of Gene Expression”. In: *ChemBioChem* (2008). DOI: 10.1002/cbic.200800154.

[385] T. C. Williams et al. “Quorum-sensing linked RNA interference for dynamic metabolic pathway control in *Saccharomyces cerevisiae*”. In: *Metabolic Engineering* (2015). DOI: 10.1016/j.ymben.2015.03.008.

[386] W. Winkler, A. Nahvi, and R. R. Breaker. “Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression”. In: *Nature* (2002). DOI: 10.1038/nature01145.

[387] F. Winston, C. Dollard, and S. L. Ricupero-Hovasse. “Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C”. In: *Yeast* (1995). DOI: 10.1002/yea.320110107.

[388] E. A. Winzeler et al. “Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis”. In: *Science* (1999). DOI: 10.1126/science.285.5429.901.

[389] W. B. Wood. “Host specificity of DNA produced by *Escherichia coli*: Bacterial mutations affecting the restriction and modification of DNA”. In: *Journal of Molecular Biology* (1966). DOI: 10.1016/S0022-2836(66)80267-X.

[390] G. Wu et al. “Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications”. In: *Trends in Biotechnology* (2016). DOI: 10.1016/j.tibtech.2016.02.010.

[391] D. Xie, E. N. Jackson, and Q. Zhu. “Sustainable source of omega-3 eicosapentaenoic acid from metabolically engineered *Yarrowia lipolytica*: from fundamental research to commercial production”. In: *Applied Microbiology and Biotechnology* (2015). DOI: 10.1007/s00253-014-6318-y.

[392] X. Xiong and S. Chen. “Expanding Toolbox for Genes Expression of *Yarrowia lipolytica* to Include Novel Inducible, Repressible, and Hybrid Promoters”. In: *ACS Synthetic Biology* (2020). DOI: 10.1021/acssynbio.0c00243.

[393] P. Xu et al. “Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA”. In: *Metabolic Engineering* (2011). DOI: 10.1016/j.ymben.2011.06.008.

[394] W. Xu et al. “Emerging molecular biology tools and strategies for engineering natural product biosynthesis”. In: *Metabolic Engineering Communications* (2020). DOI: 10.1016/j.mec.2019.e00108.

[395] N. Yamamoto et al. “Update on the Keio collection of *Escherichia coli* single-gene deletion mutants”. In: *Molecular Systems Biology* (2009). DOI: 10.1038/msb.2009.92.

[396] M. Yamanishi et al. “A Genome-Wide Activity Assessment of Terminator Regions in *Saccharomyces cerevisiae* Provides a "Terminatome" Toolbox”. In: *ACS Synthetic Biology* (2013). DOI: 10.1021/sb300116y.

[397] S. Yang et al. “Characterization and application of endogenous phase-dependent promoters in *Bacillus subtilis*”. In: *Applied Microbiology and Biotechnology* (2017). DOI: 10.1007/s00253-017-8142-7.

[398] B. Ye et al. “Random Mutagenesis by Insertion of Error-Prone PCR Products to the Chromosome of *Bacillus subtilis*”. In: *Frontiers in Microbiology* (Nov. 2020). DOI: 10.3389/fmicb.2020.570280.

[399] Y. Ye et al. “Global Metabolomic Responses of *Escherichia coli* to Heat Stress”. In: *Journal of Proteome Research* (2012). DOI: 10.1021/pr3000128.

[400] D. Yu, G. Banting, and N. F. Neumann. “A review of the taxonomy, genetics, and biology of the genus *Escherichia* and the type species *Escherichia coli*”. In: *Canadian Journal of Microbiology* (2021). DOI: 10.1139/cjm-2020-0508.

[401] J. G. Zalatan et al. “Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds”. In: *Cell* (2015). DOI: 10.1016/j.cell.2014.11.052.

[402] Y. Zhang, U. Werling, and W. Edelmann. “SLiCE: a novel bacterial cell extract-based DNA cloning method”. In: *Nucleic Acids Research* (2012). DOI: 10.1093/nar/gkr1288.

[403] Y. Zhang et al. “A new logic for DNA engineering using recombination in *Escherichia coli*”. In: *Nature Genetics* (1998). DOI: 10.1038/2417.

[404] D. Zielinski et al. “iPipet: sample handling using a tablet”. In: *Nature Methods* (2014). DOI: 10.1038/nmeth.3028.

[405] S. S. Zinjarde. “Food-related applications of *Yarrowia lipolytica*”. In: *Food Chemistry* (2014). DOI: 10.1016/j.foodchem.2013.11.117.

[406] J. C. Zweers et al. “Towards the development of *Bacillus subtilis* as a cell factory for membrane proteins and protein complexes”. In: *Microbial Cell Factories* (2008). DOI: 10.1186/1475-2859-7-10.

A. Supplemental Information

A.1. Cloning

A.1.1. Oligos

Table A.1.: Oligonucleotides used for cloning, PCR and sequencing, with name and sequence. The first two digits are a personal identifier (11 - Silke Hackenschidt, 14 - Florian Nadler, 26 - Thomas Zoll, 99 - Georg Schmidt) and the last three digits are a consecutive number. Binding sites are **highlighted**.

Name	Sequence (5' → 3')
o11016	**GGTTTCTTAGACGTCAGGTGG**
o11017	**CTCTGACACATGCAGCTCC**
o11064	**CTGACGGGCTTGTCTGC**
o14119	**TTACTTCTTCTGCCGCCTG**
o14120	**ACTCTGCGGGTGTATACAG**
o26093	CATCCGAACATAAACAA**ATGGTGTCTAAGGGCGAGGA**
o26094	CTCGCCCTTAGACACCAT**TTGTTTATGTTCGGATGTGATGTGAG**
o26096	**GCCTCCTGTCTGACTCGTC**
o26102	ATATACAGATCCCGCGG**CTCTGACACATGCAGCTCCC**
o26103	ATGTGTCAGAGCCGCGG**GATCTGTATATATATATATATATGCAAGCC**
o26107	**ATGGTGTCTAAGGGCGAGGAG**
o26108	TTTTTATTGTCAGTACT**TTACTTGTACAGCTCGTCCATGC**
o26109	ACGAGCTGTACAAGTAA**AGTACTGACAATAAAAAGATTC**
o26112	TCTTGTGCCGCGGCGGCCGC**TCTGGTGGTATTGTGACTGGG**
o26113	CACCAGAGCGGCCGCCGCGG**CACAAGAGGACGCTTTATTCTTCC**
o26114	AAGGATCTCAAGGCGCGCC**GCCTGCAGGTCGACTTAAAC**
o26115	GACCTGCAGGCGGCGCGCC**TTGAGATCCTTTTTTTCTGCGCG**
o26116	**TTGTTTATGTTCGGATGTGATGTGAG**
o26117	GATACAGTTCTCACATCACATCCGAACATAAACAA**GTACGGGCCTAAAACATACCAGATC**
o26118	ACGCCCATGGTGGTCTCCTCGCCCTTAGACACCAT**TTTGGCCTAGGTGGTCGTATTC**
o26119	GATACAGTTCTCACATCACATCCGAACATAAACAA**GTACGCCGGTAAAACATACCAG**
o26120	ACGCCCATGGTGGTCTCCTCGCCCTTAGACACCAT**ATGTTCTCGAGGCCTAGGTG**
o26121	AGGCTTAATTAAACTAGT**GGATCCGAATTTTCAGAACC**
o26122	GGATCCACTAGTTTAATTAA**GCCTCCTGTCTGACTCGTC**
o26123	**ATGCGTAAAGGCGAAGAGCT**
o26142	TACGGGCGACAGTCACAACTAGT**GGATCCGAATTTTCAGAACC**
o26143	TCTGAAAATTCGGATCCACTAGT**TGTGACTGTCGCCCGTAC**
o26144	CGAGTCAGACAGGAGGCTTAATTAA**GATATTACTTTCTGCGCACTTAACTTC**

Continued on next page

Table A.1 – continued from previous page

Name	Sequence (5' → 3')
o26145	GCGCAGAAAGTAATATCTTAATTAA**GCCTCCTGTCTGACTCGTC**
o26211	**TGCCACCTGACGTCTAAGAAAC**
o26218	ACCATGGGCAAATATTATAC**GCAAGGCGACAAGGTGC**
o26220	TCAGCACCTTGTCGCCTTGC**GTATAATATTTGCCCATGGTGAAAACGG**
o26226	AACAATTAATAGACTGGATG**GAGGCGGATAAAGTTGCAGGA**
o26227	CCTGCAACTTTATCCGCCTC**CATCCAGTCTATTAATTGTTGCCGG**
o26263	CTAGAGGATCCCCGGGTACC**GTAAGATCCTTGAGAGTTTTCG**
o26264	CCGGGAGCTGCATGTGTCAG**CGCTGTTGAGATCCAGTTC**
o26273	**GTTACTTCTTCTGCCGCCTG**
o26274	**GTACTCTGCGGGTGTATACAG**
o26316	AATAACTTCGTATAATGTATGCTATACGAACGGTA**TAGCGCTCCGGGACCCTAGG**
o26320	AATAACTTCGTATAATGTATGCTATACGAACGGTA**TATAGACCCTACCCGGACTACG**
o26323	GAGAGTGTAGAGAAGAGCGCGGCCGC**GCGGCTCTGACACATGCA**
o26324	GCATGTGTCAGAGCCGCGCGGCCGC**GCTCTTCTCTACACTCTCATATTCC**
o26325	**TTTGGTGGTGAAGAGGAGAC**
o26326	ACTAAATTTATTTCAGTCTCCTCTTCACCACCAAA**CGTTCGCTAACGTCGGGG**
o26329	ATAACTTCGTATAGCATACATTATACGAACGGTA**AGGTTAGACTATGGATATGTAATT-TAACTG**
o26330	TTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCGCGGCCGC**CTGTATGCT-GATCGTCTCA**
o26331	ACTAAATTTATTTCAGTCTCCTCTTCACCACCAAA**CTCGCACTTAGGTCGTCTAGG**
o26332	CATGTGTCAGAGCCGCGCGGCCGC**CTTTCAAGGGTGGGGGCG**
o26333	**GTTGAAGTCGCCTGGTAGCC**
o26334	TTAACACCATTTTTGGGCTACCAGGCGACTTCAAC**CGTTCGCTAACGTCGGGG**
o26335	TACCGTTCGTATAGCATACATTATACGAAGTTATT**TGTGACTGTCGCCCGTAC**
o26336	TACCGTTCGTATAATGTATGCTATACGAAGTTAT**GATATTACTTTCTGCGCACTTAAC**
o26337	ATAACTTCGTATAGCATACATTATACGAACGGTA**GGATCACCTCGCCCTG**
o26338	CCCGAAAAGTGCCACCTGACGTCTAAGAAACCGCGGCCGC**TGAACAACAGTCTCTCCCC**
o26339	CCCCCACCCTTGAAAGGCGGCCGC**GCGGCTCTGACACATGC**
o26340	TTAACACCATTTTTGGGCTACCAGGCGACTTCAAC**CTCGCACTTAGGTCGTCTAGG**
o26368	TACCGTTCGTATAGCATACATTATACGAAGTTATT**CCAAACAGCACAGACGAATG**
o26369	TACCGTTCGTATAATGTATGCTATACGAAGTTAT**CTACAATAGCTTTATTGGCCCTATTG**
o26379	GATACAGTTCTCACATCACATCCGAACATAAACAA**ATGTCCAATTTACTGACCGTAC**
o26380	GCGGCAATGACGAGTCAGACAGGAGGCTTAATTAA**GACGCAAGAGAAGCCG**
o26398	AAAACAAGAGTTTTATATACATACAGAGCACATGC**CCCTGCGTCGTCCCCTAG**
o26399	**GCATGTGCTCTGTATGTATATAAAACTC**
o26400	**TATTAATTTAGTGTGTGTATTTGTGTTTGTG**
o26401	GACACACAAACACAAATACACACACTAAATTAATA**ATGCGTAAAGGCGAAGAGCT**
o26402	GCGTGAATGTAAGCGTGACATAACTAATTACATGA**TCATTTGTACAGTTCATCCATACCATG**
o26403	**TCATGTAATTAGTTATGTCACGC**
o26404	**GCAAATTAAAGCCTTCGAGC**
o26405	AGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGC**GACATGGAGGCCCAGAATACC**
o26406	**CAGTATAGCGACCAGCATTCAC**
o26407	GCGTCAATCGTATGTGAATGCTGGTCGCTATACTG**TACGGGGCTTACCTCCG**

Continued on next page

Table A.1 – continued from previous page

Name	Sequence (5' → 3')
o26414	**GTATAAAAGACCACCGTCCCCG**
o26417	GCGGCCGC**GCGGCTCTGACACATGC**
o26418	GCGGCCGC**GGTTTCTTAGACGTCAGGTGG**
o26419	TCCGGGAGCTGCATGTGTCAGAGCCGCGCGGCCGC**CGTTCGCTAACGTCGGGG**
o26420	AAAGTGCCACCTGACGTCTAAGAAACCGCGGCCGC**TAGCGCTCCGGGACCCTAGG**
o26426	TCCGGGAGCTGCATGTGTCAGAGCCGCGCGGCCGC**GCATGTGCTCTGTATG-TATATAAAACTC**
o26429	TCCGGGAGCTGCATGTGTCAGAGCCGCGCGGCCGC**GTATAAAAGACCACCGTCCCCG**
o26430	ACGACACCAGTGAACAGCTCTTCGCCTTTACGCAT**TTTGAATGATTCTTATACTCAGAAG-GAAATG**
o26435	ATTTCTAGCTCTAAAAC**GATCATTTATCTTTCACTGCGG**
o26436	GTGAAAGATAAATGATC**GTTTTAGAGCTAGAAATAGCAAG**
o26437	**GTTTTAGAGCTAGAAATAGCAAG**
o26438	TTCTAGCTCTAAAACTGCGTCGTCCCCTAGGTAAG**GATCATTTATCTTTCACTGCGG**
o26439	TTCTAGCTCTAAAACTTAGGTGCCCACGGATCGCA**GATCATTTATCTTTCACTGCGG**
o26440	TTCTAGCTCTAAAACAGCCCCTATAGCGTCCCTAG**GATCATTTATCTTTCACTGCGG**
o26441	TTCTAGCTCTAAAACGTTAGCGAACGCGTCGTACC**GATCATTTATCTTTCACTGCGG**
o26442	TTCTAGCTCTAAAACGTAAGGGACTATAGGCCTAG**GATCATTTATCTTTCACTGCGG**
o26443	TTCTAGCTCTAAAACGTACGGGACCGAGCGTTAGC**GATCATTTATCTTTCACTGCGG**
o26444	TTCTAGCTCTAAAACGGACCCGTACCGTGCGTAAG**GATCATTTATCTTTCACTGCGG**
o26445	TTCTAGCTCTAAAACGATAGGGCTAGTAGCATTAG**GATCATTTATCTTTCACTGCGG**
o26446	TTCTAGCTCTAAAACTAACGGTTCGCACGACGCAT**GATCATTTATCTTTCACTGCGG**
o26447	TTCTAGCTCTAAAACCGTCGGGTAGGGACTACGCA**GATCATTTATCTTTCACTGCGG**
o26448	TTCTAGCTCTAAAACCTAGACGAGCGTACGCTTAC**GATCATTTATCTTTCACTGCGG**
o26449	TTCTAGCTCTAAAACTAGGGTGCTAAGTCGTCGCA**GATCATTTATCTTTCACTGCGG**
o26450	TTCTAGCTCTAAAACAGCCCCGACGTTAGCGAACG**GATCATTTATCTTTCACTGCGG**
o26451	AGGAAAGGTAATTCGGGGACGGTGGTCTTTTATAC**CCCTGCGTCGTCCCCTAG**
o26452	AGGAAAGGTAATTCGGGGACGGTGGTCTTTTATAC**CCGGTTAGCGTACGGGTC**
o26461	TGCGATCCGTGGGCACCTAA**GTTTTAGAGCTAGAAATAGCAAG**
o26462	CTAGGGACGCTATAGGGGCT**GTTTTAGAGCTAGAAATAGCAAG**
o26463	GGTACGACGCGTTCGCTAAC**GTTTTAGAGCTAGAAATAGCAAG**
o26464	CTAGGCCTATAGTCCCTTAC**GTTTTAGAGCTAGAAATAGCAAG**
o26465	GCTAACGCTCGGTCCCGTAC**GTTTTAGAGCTAGAAATAGCAAG**
o26466	CTTACGCACGGTACGGGTCC**GTTTTAGAGCTAGAAATAGCAAG**
o26467	CTAATGCTACTAGCCCTATC**GTTTTAGAGCTAGAAATAGCAAG**
o26468	ATGCGTCGTGCGAACCGTTA**GTTTTAGAGCTAGAAATAGCAAG**
o26469	TGCGTAGTCCCTACCCGACG**GTTTTAGAGCTAGAAATAGCAAG**
o26470	GTAAGCGTACGCTCGTCTAG**GTTTTAGAGCTAGAAATAGCAAG**
o26471	TGCGACGACTTAGCACCCTA**GTTTTAGAGCTAGAAATAGCAAG**
o26472	CGTTCGCTAACGTCGGGGCT**GTTTTAGAGCTAGAAATAGCAAG**
o26474	TCGGGTCCGTCTAACGTAAC**GTTTTAGAGCTAGAAATAGCAAG**
o26475	TTAGGGACCCGTACGCTAAC**GTTTTAGAGCTAGAAATAGCAAG**
o26476	**GCCAACACCCGCTGAC**
o26477	CCAATTCGCCCTATAGTGAG**GTTTTGCCAGCGGAATTCCAC**

Continued on next page

Table A.1 – continued from previous page

Name	Sequence (5' → 3')
o26478	GTAAAACGACGGCCAGTGAG**GACATCACCCGAAAAGAAGCTAAG**
o26479	TGGAATTCCGCTGGCAAAAC**CTCACTATAGGGCGAATTGGTTC**
o26480	GCTTCTTTTCGGGTGATGTC**CTCACTGGCCGTCGTTTTAC**
o26481	**GACATCACCCGAAAAGAAGCTAAG**
o26482	GCCTTTGTTACGTCTATATTCATTGAAACTGATTATTCGA**CTTTAATTTGCGGCCGGTAC**
o26483	AGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGC**CTTACCTCCGTAGTATCGCAC**
o26484	AGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGC**ATCCCTACTAGGATCCGGCA**
o26496	CTTACCTAGGGGACGACGCA**GTTTTAGAGCTAGAAATAGCAAG**
o26509	AACTGGTAGCTTGTTTTGCTGTTGATCAGGATGAAAATCCGCAGCGTTAA**CGATGGTAGTGTGGGGTCT**
o26510	CTGTTGAAGCTGATTGAGTAAACCGGAGCGCATGGCCCCGGTTTTGTGAG**TTAGGCACCCCAGGCTTTAC**
o26511	TTCTAGCTCTAAAACGTTAGCGTACGGGTCCCTAA**ACTAGTACAGTAGAGAGTTGCG**
o26512	TTCTAGCTCTAAAACGTTACGTTAGACGGACCCGA**ACTAGTACAGTAGAGAGTTGCG**
o26513	TTCTAGCTCTAAAACTGCGTCGTCCCCTAGGTAAG**ACTAGTACAGTAGAGAGTTGCG**
o26514	**ATCCCTACTAGGATCCGGCAC**
o26515	**TTAGCGTACGGGTCCCTAAC**
o26516	ACGGTCCCAGCGTCGTGCCGGATCCTAGTAGGGAT**TTATTTATAAAGTTCGTCCATACCG**
o26517	ACGGTCCCAGCGTCGTGCCGGATCCTAGTAGGGATCCGG**TTATTTATAAAGTTCGTCCATACCG**
o26522	TTCTAGCTCTAAAACGAGTGTATACCGTGCGATAC**ACTAGTACAGTAGAGAGTTGCG**
o26523	GTATCGCACGGTATACACTC**GTTTTAGAGCTAGAAATAGCAAG**
o26524	ATTTCTAGCTCTAAAAC**ACTAGTACAGTAGAGAGTTGCG**
o26525	CTCTCTACTGTACTAGT**GTTTTAGAGCTAGAAATAGCAAG**
o99034	GGTCCGTCTAACGTAACAGGGTACACGTTAGGGACCCGTACGCTAACCGG**TTGACAGCTAGCTCAGTCC**
o99036	CTCGGGTCCGTCTAACGTAACAGGGTACACGTTAGGGACCCGTACGCTAA**TTGACAGCTAGCTCAGTCC**

A.1.2. PCR Fragments

Table A.2.: PCR Fragments which were used for cloning with expected length in bp and used oligonucleotides (Tab. A.1)

Name	Template	Length [bp]	Oligonucleotides
Cloning of FLEXpress vector p26061			
f_pTAR_1v2	p26009	1842	o26115 & o14120
f_pTAR_2v2	p26009	2189	o14119 & o26102
f_URA3_Yl_empty	gDNA *Y. lipolytica* H222	1534	o26122 & o26112
f_ARS68_empty	p26039	2308	o26103 & o26121

Continued on next page

Table A.2 – continued from previous page

Name	Template	Length [bp]	Oligonucleotides
f_SpecR_eco	p14019	1009	o26113 & o26114
Cloning of pTCapt plasmids p26062, p26063, p26064 and p26065			
f_FLEXpress_1v2	p26061	4431	o26145 & o14120
f_FLEXpress_2v2	p26061	4526	o14119 & o26142
f_TCapt-mono	p26042	147	o26117 & o26118
f_TCapt-dimer	p26043	256	o26119 & o26120
f_TCapt-trimer	p26044	332	o26117 & o26120
f_mTFP_control	p26025	745	o26093 & o26108
f_mTFP_universal	p26025	728	o26107 & o26108
f_T_{tefag}	p11032	206	o26109 & o26144
f_P_{tefag}	p11032	323	o26143 & o26094
f_P_{tefag}_universal	p11032	305	o26143 & o26116
Cloning of CPad plasmids p26116 and p26117			
f_Rank1_Sc	p26112	910	o26334 & o26316
f_Rank4_Sc	p26113	953	o26340 & o26320
f_lox71_kanMX_lox61	p26033	1329	o26335 & o26336
f_PDC6_front	gDNA *S. cerevisiae* S288C	567	o26332 & o26333
f_PDC6_back	gDNA *S. cerevisiae* S288C	540	o26337 & o26338
f_FLEXpress_1v2_universal	p26061	3020	o11016 & o26274
f_FLEXpress_2v2_Sc	p26061	2218	o26273 & o26339
Cloning of CPad plasmids p26118 and p26119			
f_FLEXpress_1v2_universal	p26061	3020	o11016 & o26274
f_FLEXpress_2v2_Yl	p26061	2200	o26273 & o26323
f_lox71_HygB_lox61	p55001	1907	o26368 & o26369
f_URA3_front	gDNA *Y. lipolytica*W29	1525	o26324 & o26325
f_URA3_back	gDNA *Y. lipolytica*W29	1569	o26329 & o26330
f_Rank1_Yl	p26112	910	o26326 & o26316
f_Rank4_Yl	p26113	970	o26331 & o26320
Cloning of marker rescue plasmid p26125			
f_FLEXpress_1v2	p26062	4414	o26096 & o14120
f_FLEXpress_2v2	p26062	4971	o14119 & o26116
f_cre-T_{CYC1}	p11055	1352	o26379 & o26380
Cloning of CPad plasmids p26132, p26133, p26137 and p26153			
f_TAR1v2_Sc	p26038	3028	o26418 & o26274
f_TAR2v2_Sc	p26038	2202	o26273 & o26417
f_$P_{CYC1min}$	gDNA *S. cerevisiae* S288C	174	o26426 & o26400
f_T_{CYC1}	gDNA *S. cerevisiae* S288C	248	o26403 & o26404
f_sfGFP	p26052	787	o26401 & o26402
f_sfGFP_T_{CYC1}	OE-PCR (f_sfGFP, f_T_{CYC1})	1000	o26401 & o26404
f_Rank1_front	p26112	483	o26419 & o26398
f_Rank1_back_V2	p26112	497	o26407 & o26420

Continued on next page

Table A.2 – continued from previous page

Name	Template	Length [bp]	Oligonucleotides
f_$P_{CYC1min}$_sfGFP_T_{CYC1}	p26137	1104	o26399 & o26404
f_HygB	p26134	1646	o26405 & o26406
Cloning of sfGFP cassette plasmid p26138			
f_p26137_1v2	p26137	3321	o26123 & o26227
f_p26137_2v2	p26137	4331	o26226 & o26417
f_$P_{TEF1core}$	gDNA *Y. lipolytica* W29	187	o26429 & o26430
Cloning of sgRNA delivery plasmids p26139, p26140, p26141, p26142, p26143, p26144, p26145, p26146, p26147,p26148, p26149, p26150, p26151 and p26152			
f_p26139_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26436 & o26263
f_pJZC588_1v3_uni	p26129	2107	o26437 & o26263
f_p26141_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26461 & o26263
f_p26142_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26462 & o26263
f_p26143_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26463 & o26263
f_p26144_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26464 & o26263
f_p26145_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26465 & o26263
f_p26146_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26466 & o26263
f_p26147_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26467 & o26263
f_p26148_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26468 & o26263
f_p26149_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26469 & o26263
f_p26150_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26470 & o26263
f_p26151_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26471 & o26263
f_p26152_1v3	p26129 (*Nde*I/*Xba*I Digest)	2124	o26472 & o26263
f_pJZC588_2v3	p26129	992	o26264 & o11064
f_p26139_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2268	o11017 & o26435
f_p26140_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26438
f_p26141_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26439

Continued on next page

Table A.2 – continued from previous page

Name	Template	Length [bp]	Oligonucleotides
f_p26142_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26440
f_p26143_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26441
f_p26144_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26442
f_p26145_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26443
f_p26146_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26444
f_p26147_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26445
f_p26148_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26446
f_p26149_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26447
f_p26150_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26448
f_p26151_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26449
f_p26152_3v3	p26129 (*BsaH*I/*Pst*I Digest)	2286	o11017 & o26450
Cloning of integration plasmid p26162			
f_p426MET25	p26161	9381	o26480 & o26479
f_LYS2	gDNA *S. cerevisiae* S288C	4556	o26477 & o26478
Cloning of pre-integrated sfGFP cassette plasmids p26163, p26164			
f_p26118_1v2	p26118	6307	o26226 & o26451
f_p26118_2v2	p26118	4654	o26483 & o26227
f_p26119_1v2	p26119	6307	o26226 & o26452
f_p26119_2v2	p26119	4654	o26484 & o26227
f_$P_{TEF1core}$_sfGFP_T_{CYC1}	p26138	1077	o26414 & o26404
Cloning of CRISPRa plasmids for *Y. lipolytica*			
f_dCas9-VPR_universal	p26120 (*Avr*II/*Nde*I Digest)	6395	o26176 & o26212
f_p26165_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26496 & o33147
f_p26166_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26461 & o33147
f_p26167_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26462 & o33147
f_p26168_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26463 & o33147

Continued on next page

Table A.2 – continued from previous page

Name	Template	Length [bp]	Oligonucleotides
f_p26169_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26464 & o33147
f_p26170_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26475 & o33147
f_p26171_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26474 & o33147
f_p26172_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o26473 & o33147
f_p26173_1v3	p26120 (*Avr*II/*Nhe*I Digest)	4798	o13111 & o33147
f_p26165_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26497
f_p26166_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26498
f_p26167_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26499
f_p26168_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26500
f_p26169_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26501
f_p26170_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26502
f_p26171_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26503
f_p26172_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o26504
f_p26173_3v3	p26120 (*Avr*II/*Nde*I Digest)	2505	o14127 & o13110
Cloning of CPad-sfGFP integration plasmids for *E. coli*			
f_pKD3_BR322_Rank1_1v2	p26193 (*Nde*I/*Nhe*I Digest)	2746	o26514 & o26218
f_pKD3_BR322_Rank1_2v2	p26193 (*Hind*III/*Sca*I Digest)	1358	o26220 & o26515
f_J23117_sfGFP_ap	p99018	868	o99034 & o26516
f_J23117_sfGFP_ip	p99018	868	o99036 & o26517
OE-PCR: f_pKD3_Rank1_2v2-_J23117_sfGFP_ap	f_pKD3_BR322_Rank1_2v2 & f_J23117_sfGFP_ap	2176	o26220 & o26516
OE-PCR:f_pKD3_Rank1_2v2-_J23117_sfGFP_ip	f_pKD3_BR322_Rank1_2v2 & f_J23117_sfGFP_ip	2180	o26220 & o26517
Cloning of sgRNA delivery plasmids plasmids for *E. coli*			
f_p26189_1v2	p99016 (*Eco*RI/*Sca*I)	1696	o26496 & o26226

Continued on next page

Table A.2 – continued from previous page

Name	Template	Length [bp]	Oligonucleotides
f_p26190_1v2	p99016 (*EcoR*I/*Sca*I)	1696	o26474 & o26226
f_p26191_1v2	p99016 (*EcoR*I/*Sca*I)	1696	o26475 & o26226
f_p26199_1v2	p99016 (*EcoR*I/*Mlu*I)	2196	o26523 & o26263
f_p26200_1v2	p99016 (*EcoR*I/*Mlu*I)	2166	o26525 & o26263
f_p26189_2v2	p99016 (*Hind*III/*Xba*I)	1218	o26227 & o26513
f_p26190_2v2	p99016 (*Hind*III/*Xba*I)	1218	o26227 & o26512
f_p26191_2v2	p99016 (*Hind*III/*Xba*I)	1218	o26227 & o26511
f_p26199_2v2	p99016 (*BamH*I/*Sca*I)	745	o26264 & o26522
f_p26200_2v2	p99016 (*BamH*I/*Sca*I)	727	o26264 & o26524

Table A.3.: PCR Fragments used for given plasmids. Fragment name refers to Name in Tab. A.2

Plasmid	Name	Fragment Name
p26061	pFLEXpress_empty	f_pTAR_1v2
		f_pTAR_2v2
		f_URA3_Yl_empty
		f_ARS68_empty
		f_SpecR_eco
p26062	pFLEXpress_control	f_FLEXpress_1v2
		f_FLEXpress_2v2
		f_mTFP_control
		f_T_{tefag}
		f_P_{tefag}
p26063	pTCapt_tc-mono_mTFP	f_FLEXpress_1v2
		f_FLEXpress_2v2
		f_TCapt-mono
		f_mTFP_universal
		f_T_{tefag}
		f_P_{tefag}_universal
p26064	pTCapt_tc-dimer_mTFP	f_FLEXpress_1v2
		f_FLEXpress_2v2
		f_TCapt-dimer
		f_mTFP_universal
		f_T_{tefag}
		f_P_{tefag}_universal
p26065	pTCapt_empty_mTFP	f_FLEXpress_1v2
		f_FLEXpress_2v2
		f_TCapt-trimer
		f_mTFP_universal
		f_T_{tefag}

Continued on next page

Table A.3 – continued from previous page

Plasmid	Name	Fragment Name
p26116	pFLEXpress_Rank1_KanMX_PDC6	f_P_{tefag}_universal f_FLEXpress_1v2_universal f_FLEXpress_2v2_Sc f_PDC6_front f_PDC6_back f_lox71_kanMX_lox61 f_Rank1_Sc
p26117	pFLEXpress_Rank4_KanMX_PDC6	f_FLEXpress_1v2_universal f_FLEXpress_2v2_Sc f_PDC6_front f_PDC6_back f_lox71_kanMX_lox61 f_Rank4_Sc
p26118	pFLEXpress_Rank1_HygB_URA3	f_FLEXpress_1v2_universal f_FLEXpress_2v2_Sc f_URA3_front f_URA3_back f_lox71_HygB_lox61 f_Rank1_Yl
p26119	pFLEXpress_Rank4_HygB_URA3	f_FLEXpress_1v2_universal f_FLEXpress_2v2_Sc f_URA3_front f_URA3_back f_lox71_HygB_lox61 f_Rank4_Yl
p26125	pFLEXpress_P_{TEFag}-Cre-T_{CYC1}	f_FLEXpress_1v2 f_FLEXpress_2v2 f_cre-T_{CYC1}
p26137	pXK_$P_{CYC1min}$_sfGFP_T_{CYC1}	f_TAR1v2_Sc f_TAR2v2_Sc f_$P_{CYC1min}$ f_sfGFP_T_{CYC1} f_HygB
p26138	pXK_$P_{CYC1min}$_sfGFP_T_{CYC1}	f_p26137_1v2 f_p26137_2v2 f_$P_{TEF1core}$
p26139	p26139_pJZC588_empty	f_p26139_1v3 f_pJZC588_2v3 f_p26139_3v3
p26140	p26140_pJZC588_Pro_11	f_pJZC588_1v3_uni f_pJZC588_2v3 f_p26140_3v3
p26141	p26141_pJZC588_Pro_23	f_p26141_1v3 f_pJZC588_2v3

Continued on next page

Table A.3 – continued from previous page

Plasmid	Name	Fragment Name
		f_p26141_3v3
p26142	p26142_pJZC588_Pro_18	f_p26142_1v3
		f_pJZC588_2v3
		f_p26142_3v3
p26143	p26143_pJZC588_Pro_12	f_p26143_1v3
		f_pJZC588_2v3
		f_p26143_3v3
p26144	p26144_pJZC588_Pro_8	f_p26144_1v3
		f_pJZC588_2v3
		f_p26144_3v3
p26145	p26145_pJZC588_Pro_10	f_p26145_1v3
		f_pJZC588_2v3
		f_p26145_3v3
p26146	p26146_pJZC588_Pro_7	f_p26146_1v3
		f_pJZC588_2v3
		f_p26146_3v3
p26147	p26147_pJZC588_Pro_9	f_p26147_1v3
		f_pJZC588_2v3
		f_p26147_3v3
p26148	p26148_pJZC588_Pro_13	f_p26148_1v3
		f_pJZC588_2v3
		f_p26148_3v3
p26149	p26149_pJZC588_Pro_16	f_p26149_1v3
		f_pJZC588_2v3
		f_p26149_3v3
p26150	p26150_pJZC588_Pro_25	f_p26150_1v3
		f_pJZC588_2v3
		f_p26150_3v3
p26151	p26151_pJZC588_Pro_22	f_p26151_1v3
		f_pJZC588_2v3
		f_p26151_3v3
p26152	p26152_pJZC588_Pro_19	f_p26152_1v3
		f_pJZC588_2v3
		f_p26152_3v3
p26153	pFLEXpress_Rank1_$P_{CYC1min}$_sfGFP_T_{CYC1}	f_TAR1v2_Sc
		f_TAR2v2_Sc
		f_$P_{CYC1min}$
		f_sfGFP_T_{CYC1}
		f_HygB
		f_Rank1_front
		f_Rank1_back_V2
p26162	p426-P_{PDA1}-mCherry-T_{GND1}-LYS2	f_p426MET25
		f_LYS2
p26163	pFLEXpress_Rank1_$P_{TEF1core}$_sfGFP_T_{CYC1}	f_p26118_1v2

Continued on next page

Table A.3 – continued from previous page

Plasmid	Name	Fragment Name
p26164	pFLEXpress_Rank4_$P_{TEF1core}$_sfGFP_T_{CYC1}	f_p26118_2v2 f_$P_{TEF1core}$_sfGFP_T_{CYC1} f_p26119_1v2 f_p26119_2v2 f_$P_{TEF1core}$_sfGFP_T_{CYC1}
p26165	p26165_CRISPRyl_VPR_sgRNA_Pro_11	f_dCas9-VPR_universal f_p26165_1v3 f_p26165_3v3
p26166	p26166_CRISPRyl_VPR_sgRNA_Pro_23	f_dCas9-VPR_universal f_p26166_1v3 f_p26166_3v3
p26167	p26167_CRISPRyl_VPR_sgRNA_Pro_18	f_dCas9-VPR_universal f_p26167_1v3 f_p26167_3v3
p26168	p26168_CRISPRyl_VPR_sgRNA_Pro_12	f_dCas9-VPR_universal f_p26168_1v3 f_p26168_3v3
p26169	p26169_CRISPRyl_VPR_sgRNA_Pro_8	f_dCas9-VPR_universal f_p26169_1v3 f_p26169_3v3
p26170	p26170_CRISPRyl_VPR_sgRNA_Pro_21	f_dCas9-VPR_universal f_p26170_1v3 f_p26170_3v3
p26171	p26171_CRISPRyl_VPR_sgRNA_Pro_2	f_dCas9-VPR_universal f_p26171_1v3 f_p26171_3v3
p26172	p26172_CRISPRyl_VPR_sgRNA_Pro_4	f_dCas9-VPR_universal f_p26172_1v3 f_p26172_3v3
p26173	p26173_CRISPRyl_VPR_sgRNA_Ku70	f_dCas9-VPR_universal f_p26173_1v3 f_p26173_3v3
p26189	p26189_pCD315_Pro_11-scRNA.b1	f_p26189_1v2 f_p26189_2v2
p26190	p26190_pCD315_Pro_2-scRNA.b1	f_p26190_1v2 f_p26190_2v2
p26191	p26191_pCD315_Pro_21-scRNA.b1	f_p26191_1v2 f_p26191_2v2
p26195	p26195_KD3_Rank1_sfGFP_ap	f_pKD3_BR322_Rank1_1v2 f_pKD3_Rank1_2v2_J23117_sfGFP_ap
p26196	p26196_KD3_Rank1_sfGFP_ip	f_pKD3_BR322_Rank1_1v2 f_pKD3_Rank1_2v2_J23117_sfGFP_ip
p26199	p26199_pCD315_guide_182-scRNA.b1	f_p26199_1v2 f_p26199_2v2

Continued on next page

Table A.3 – continued from previous page

Plasmid	Name	Fragment Name
p26200	p26200_pCD315_empty-scRNA.b1	f_p26200_1v2 f_p26200_2v2

A.1.3. Constructed strains with relevant genotype

Table A.4.: Constructed strains with ID, designation and relevant genotype

Strain	Designation	Relevant genotype
s26235	*Y. lipolytica* Po1f	*ku70*
s26241	*Y. lipolytica* Po1f	*ku70* ura3::Rank1-$P_{TEF1core}$-sfGFP-HygB
s26445	*S. cerevisiae* BY4742	leu2::P_{TDH3}-dCas9-LEU2cg
s26446	*S. cerevisiae* BY4742	leu2::P_{TDH3}-dCas9-LEU2cg his3::P_{ADH1}-MCP-VP64-HIS3cg
s26447	*S. cerevisiae* BY4742	leu2::P_{TDH3}-dCas9-LEU2cg his3::P_{ADH1}-MCP-VP64-HIS3cg pdc6::Rank1-KanMX
s26449	*S. cerevisiae* BY4742	leu2::P_{TDH3}-dCas9-LEU2cg his3::P_{ADH1}-MCP-VP64-HIS3cg pdc6::Rank1-lox72
s26454	*S. cerevisiae* BY4742	leu2::P_{TDH3}-dCas9-LEU2cg his3::P_{ADH1}-MCP-VP64-HIS3cg pdc6::Rank1-lox72 $P_{CYC1min}$-sfGFP-HygB
s26458	*S. cerevisiae* BY4742	leu2::P_{TDH3}-dCas9-LEU2cg his3::P_{ADH1}-MCP-VP64-HIS3cg pdc6::Rank1-lox72 $P_{CYC1min}$-sfGFP-HygB lys2::P_{PDA1}-mCherry-LYS2
s26901.3	*E. coli* BW25113	Rank1-*sfGFP* Cm^R
s26902	*E. coli* BW25113 (s99015)	p99013($Spec^R$)
s26903	*E. coli* BW25113 (s26901.3)	Rank1-P_{BBa_J23117}-*sfGFP*(Cm^R) p99013($Spec^R$)
s26904	*E. coli* BW25113 (s26902)	p99013($Spec^R$) p99011(Amp^R)
s26905	*E. coli* BW25113 (s26902)	p99013($Spec^R$) p99016(Amp^R)
s26906	*E. coli* BW25113 (s26902)	p99013($Spec^R$) p26189(Amp^R)
s26907	*E. coli* BW25113 (s26902)	p99013($Spec^R$) p26190(Amp^R)
s26908	*E. coli* BW25113 (s26902)	p99013($Spec^R$) p26191(Amp^R)
s26909	*E. coli* BW25113 (s26902)	p99013($Spec^R$) p26199(Amp^R)
s26910	*E. coli* BW25113 (s26902)	p99013($Spec^R$) p26200(Amp^R)
s26911	*E. coli* BW25113 (s26903)	Rank1-*sfGFP*(Cm^R) p99013($Spec^R$) p99011(Amp^R)
s26912	*E. coli* BW25113 (s26903)	Rank1-*sfGFP*(Cm^R) p99013($Spec^R$) p99016(Amp^R)

Continued on next page

Table A.4 – continued from previous page

Strain	Designation	Relevant genotype
s26913	*E. coli* BW25113 (s26903)	Rank1-*sfGFP*(Cm^R) p99013($Spec^R$) p26189(Amp^R)
s26913	*E. coli* BW25113 (s26903)	Rank1-*sfGFP*(Cm^R) p99013($Spec^R$) p26190(Amp^R)
s26915	*E. coli* BW25113 (s26903)	Rank1-*sfGFP*(Cm^R) p99013($Spec^R$) p26191(Amp^R)
s26916	*E. coli* BW25113 (s26903)	Rank1-*sfGFP*(Cm^R) p99013($Spec^R$) p26199(Amp^R)
s26917	*E. coli* BW25113 (s26903)	Rank1-*sfGFP*(Cm^R) p99013($Spec^R$) p26200(Amp^R)

A.1.4. Aptamers

Empty plasmid: FLEXpress-vector p26061

The FLEXpress vector is mainly based on the plasmid pTAR_LS92.1 which was used for TAR [207] and constructed by Thomas Hofmeyer (Uni Greifswald). It provides a URA3 auxotrophy marker and a 2-micron sequence for *S. cerevisiae*, as well as a ColE1 ori. The ARS68 was integrated to maintain the plasmid in *Y. lipolytica*. The pTAR_LS9 plasmid provides Amp^R marker for *E. coli*. This marker was replaced by a $Spec^R$ to improve the selection efficiency in *E. coli*.
The DNA fragments were amplified by NEB One*Taq*® Polymerase using the protocol given by the manufacturer. To reduce the background, the URA3 marker for *S. cerevisiae* was split. The fragments were amplified using the templates and primer given in Tab. A.2. The fragments were designed to have a 35 bp overlap to each other. Thus, they were used for TAR cloning in *S. cerevisiae* CEN.PK2-1C. The obtained clones were verified by colony PCR and sequencing.

TCapt vectors: p26062, p26063, p26064 and p26065

The FLEXpress-vector p26061 was the basis for the vectors to test the different TCapt variants. Therefore, TAR cloning was used again [207]. The used PCR fragments can be found in Tab. A.2. The DNA fragments were amplified by NEB One*Taq*® polymerase using the protocol given by the manufacturer. T_m was calculated by Geneious R10. For the vector backbone fragments f_pTCapt_1v2 and f_pTCapt_2v2, a *Dpn*I digest was carried out directly in the PCR mixture. The clean-up was performed with the MACHEREY-NAGEL NucleoSpin Gel and PCR Clean-up Kit according to the manufacturer protocol with the suggested pre-warming of the elution buffer. The DNA was used to assemble the plasmids p26062, p26063, p26064 and p26065 via TAR. The obtained clones were verified by colony PCR and sequencing.

A.1.5. CRISPRpads

CPad integration into *S. cerevisiae* with p26116 and p26117

The integration plasmid for the CPad was assembled by TAR cloning [207]. The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase according to the protocol of the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by NEB Monarch PCR & DNA Cleanup Kit. The used backbone:insert-ratio was 1:5. The obtained clones were verified by colony PCR and sequencing.

CPad integration into *Y. lipolytica* with p26118 and p26119

The integration plasmid for the CPad was assembled by TAR cloning [207]. The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase according to the protocol of the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by NEB Monarch PCR & DNA Cleanup Kit. The ratio between backbone and homology site was 1:1.5 and between backbone and the other fragments 1:5. The obtained clones were verified by colony PCR and sequencing.

Marker rescue for *S. cerevisiae* and *Y. lipolytica* (p26125)

The plasmid p11055 was designed with a dominant marker gene. Usually, the plasmid is removed by serial plating on non-selecting plates. Additionally, the Cre recombinase was under control of a *Y. lipolytica* promoter. Thus, the Cre recombinase was cloned into the FLEXpress backbone under control of the P_{TEFag} promoter, which can be used in *S. cerevisiae* as well as in *Y. lipolytica*. Furthermore, the FLEXpress backbone contains a URA3 marker for *S. cerevisiae* and *Y. lipolytica* allowing a counter-selection with 5-Fluoroorotic Acid (5-FOA). The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase using the protocol given by the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by NEB Monarch PCR & DNA Cleanup Kit. The used backbone:insert-ratio was 1:10. Initially, SLiCE was used to clone the plasmid. Two experiments did not work because no clones were obtained by *E. coli* transformation. Thus, the Cre recombinase plasmid was assembled by TAR cloning while keeping the backbone:insert-ratio [207]. The obtained clones were verified by colony PCR and sequencing.

Preparation of sfGFP-cassette for *S. cerevisiae* (p26137)

The intermediate plasmid was used to assemble the sfGFP-cassette with $P_{CYC1min}$ [236] and T_{CYC1}. The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase using the protocol given by the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by NEB Monarch PCR & DNA Cleanup Kit. The used backbone:insert-ratio was 1:5 except for $P_{CYC1min}$ which was added in 10-fold excess. The obtained clones were verified by colony PCR and sequencing. Sequencing revealed three mutations in the T_{CYC1} terminator, but this occurred in all clones, so this was likely due to the PCR template (chromosomal DNA from *S. cerevisiae* S288C).

Preparation of sfGFP-cassette for *Y. lipolytica* (p26138)

Again, this plasmid was used as intermediate plasmid. This plasmid is similar to p26137, but the promoter was exchanged by the $P_{TEF1core}$ [319]. The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase using the protocol given by the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by Macherey-Nagel NucleoSpin Gel and PCR-Clean-up-Kit. A ratio of 1:10 for backbone and insert was used in the SLiCE reaction. The obtained clones were verified by colony PCR and sequencing.

Cloning of sgRNA delivery plasmids p26139, p26140, p26141, p26142, p26143, p26144, p26145, p26146, p26147, p26148, p26149, p26150, p26151 and p26152

The plasmid pJZC588 (p26129) is designed to code the sgRNA for the application in yeast. Compared to the original plasmid, only 20 bases are to be exchanged or removed for use with the CPads. First, the empty version (p26139) was cloned using SLiCE. Subsequently, p26140 was cloned using the universal fragment f_pJZC588_1v3_uni. The fragment was designed without extension to the neighboring fragment to be compatible to all plasmid versions. But, the cloning efficiency was rather low. Thus, the fragments were designed to have an extension to the neighboring fragment, containing the guide sequence. In addition, a considerable carryover of the original plasmid took place. For this reason, for the other plasmids, the plasmid backbone was digested with different restriction enzymes before being used as PCR template. Tab. A.2 shows the used restriction enzymes. The DNA fragments were amplified by NEB Q5® polymerase using the protocol given by the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were purified using the Analytik Jena innuPREP PCRpure Kit. The used backbone:insert-ratio for the SLiCE was 1:1.5. Plasmids were confirmed via plasmid purification, restriction digestion, and sequencing.

Integration vector for sfGFP in *S. cerevisiae* (p26153)

The sfGFP module from p26137 was used and integrated between homology sites for integration into the CPad in *S. cerevisiae*. The plasmid was slightly modified. The sfGFP-cassette was placed directly after Pro_11. The spacer behind Pro_11 together with another protospacer led to a sequence where the amplification primer bound twice. Hence, seven bases were removed behind the integration site for the sfGFP-cassette, leading to f_Rank1_back_V2. The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase using the protocol given by the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by NEB Monarch PCR & DNA Cleanup Kit. The used backbone:insert-ratio was 1:5 except for $P_{CYC1min}$ which was used in 10-fold excess. The obtained clones were verified by colony PCR and sequencing.

Integration vector for normalization cassette in *S. cerevisiae* (p26162)

The cassette, containing mCherry, was intended to normalize fluorescence of sfGFP by cloning the gene behind a promoter of the housekeeping gene *PDA1*. The plasmid p426MET25 was equipped with URA3 marker. Since, URA3 is used for the sgRNA delivery plasmids, a marker exchanges was required. The last remaining marker in *S. cerevisiae* BY4742 was the auxotrophy for lysine. Thus, the marker was amplified from the prototrophic strain *S. cerevisiae* S288C, whereas the other marker genes originated from *Candida glabrata*. The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase using the protocol given by the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by Macherey-Nagel NucleoSpin Gel and PCR-Clean-up-Kit. The used backbone:insert-ratio was 1:5 in the SLiCE reaction. The obtained clones were verified by colony PCR and sequencing.

Integration vector for CPad with pre-integrated sfGFP for *Y. lipolytica* (p26163, p26164)

As already mentioned, *Y. lipolytica* prefers NHEJ over HR for repair of DSBs. Thus, integration of the sfGFP in the CPad might be challenging. Therefore, the $P_{TEF1core}$-sfGFP-T_{CYC1} cassette from p26138 was

pre-integrated into the CPad with homology sites to the URA3 locus of *Y. lipolytica*. The fragments were amplified using the templates and primer given in Tab. A.2, and the NEB Q5® polymerase using the protocol given by the manufacturer. The annealing temperature was calculated by NEB T_m-calculator. All fragments were treated by *Dpn*I which was directly added to the PCR mixture. The fragments were extracted by Macherey-Nagel NucleoSpin Gel and PCR-Clean-up-Kit. The used backbone:insert-ratio was 1:10 in the SLiCE reaction. The obtained clones were verified by colony PCR and sequencing.

Integration vector for CPad with pre-integrated sfGFP for *E. coli* (p26195, p26196)

As it is rather challenging to modify the *E. coli* genome by integration, the CPads should be integrated with pre-integrated sfGFP like for *B. subtilis* and *Y. lipolytica*. First, the sfGFP cassette of Dong *et al.* [83] was stepwise reconstructed by Georg Schmidt (research assistant) using oligonucleotide extensions. First, he assembled p99017 via SLiCE and appended the Bujard RBS to the sfGFP within pKD4 background. In a second step, the P_{BBa_J23117} promoter was appended by oligonucleotide extensions using SLiCE (p99018). The p99018 was used as a template for assembling the plasmids with CPad and pre-integrated sfGFP cassette. The SLiCE fragments were amplified using the Roboklon Polymerase X according to manufacturer instructions. A first attempt to assemble the plasmids was unsuccessful. Thus, one backbone fragment was concatenated with the P_{BBa_J23117}-sfGFP cassette via overlap extension PCR (OE-PCR) using Polymerase X. All fragments have been digested with *Dpn*I which was directly added to the PCR-mix. The purification was done via Analytik Jena innuPREP PCRpure Kit. Since only two backbone fragments had to be connected, a 1:1 ratio was chosen for the reaction. The obtained clones were verified by colony PCR and sequencing.
Note: For cloning of pKD4 derivatives (R6K gamma *ori*), a strain with pir^+ genotype must be used. For this purpose, *E. coli* GT115 was purchased from InvivoGen in the variant with modified *pir* gene for enhanced plasmid amplification.

Cloning of sgRNA delivery plasmids for *E. coli* (p26189, p26190, p26191, p26199, p26200)

The plasmid pCD315 (p99011) was designed to carry the guide sequence as well as the MS2 RNA extension for MCP binding. Only 20 bases should be exchanged during cloning, resulting in a strong carryover of template plasmids. For this reason, the plasmid backbone was divided into two parts. The other half, which should not be amplified, was cut with restriction enzymes (see Tab. A.2 for the enzymes used). Furthermore, only a small amount of template (approx. 50 fmol) was used for PCR with NEB Q5® polymerase. Oligos and fragment names can be found in Tab. A.2. The polymerase was used according to the manufacturer's instructions. However, for the 2v2 fragments, a touch-down PCR from 74 down to 64 °C had to be performed to obtain a sufficient amount of PCR product. Subsequently, a *Dpn*I digestion was performed, whereby the enzyme was added directly to the PCR mixture. Purification was performed using the Analytik Jena innuPREP PCRpure Kit according to the supplier's instructions. Since only two backbone fragments were involved, the SLiCE reaction was performed in a 1:1 ratio. Plasmids were confirmed via plasmid purification, restriction digestion, and sequencing.

A.2. TCapt: Additional methods and results

A.2.1. FLEXpress vector cultivation

The cultivation was carried out in the CompuGene Robotics platform. The obtained clones were used to inoculate the preculture plates. One preculture plate was used for each organism. Six colonies per organism and plasmid were used to inoculate the precultures. The plate was filled with 200 µL of media (M9 media (section 2.3) for *E. coli* and YNB supplemented with HIS, LEU and TRP for *S. cerevisiae* and *Y. lipolytica*) and inoculated by means of toothpicks. The precultures were introduced in the automation platform and cultivated with the randomization workflow (section 4.4.3). Plate reader and incubator were set to 30 °C. The tower shaker of the incubator was set to 1000 rpm.
The following script versions were used for the cultivation of the shown data:

- Composer-Script: commit 46daeb92ee0468e33913907fe2a78223ade776a9
- Randomizer (R): commit dc9f2dc02618cf724975459249e196909ebc75be
- PHERAread (R): commit a48a2f43e970d182646df426a73563cc4123da39

The PHERAstar® measured OD and FI using a filter-set for GFP. The gain for the FI measurement was set to 50. The script PHERAread was used to import the .csv files which were generated by the BMG PHERAstar. The script flexpress_cultivation.R was used to evaluate the data and to generate the plots. Original Data and the R script are available for review in the git repository of the thesis.

A.2.2. Additional figures: FLEXpress vector cultivation

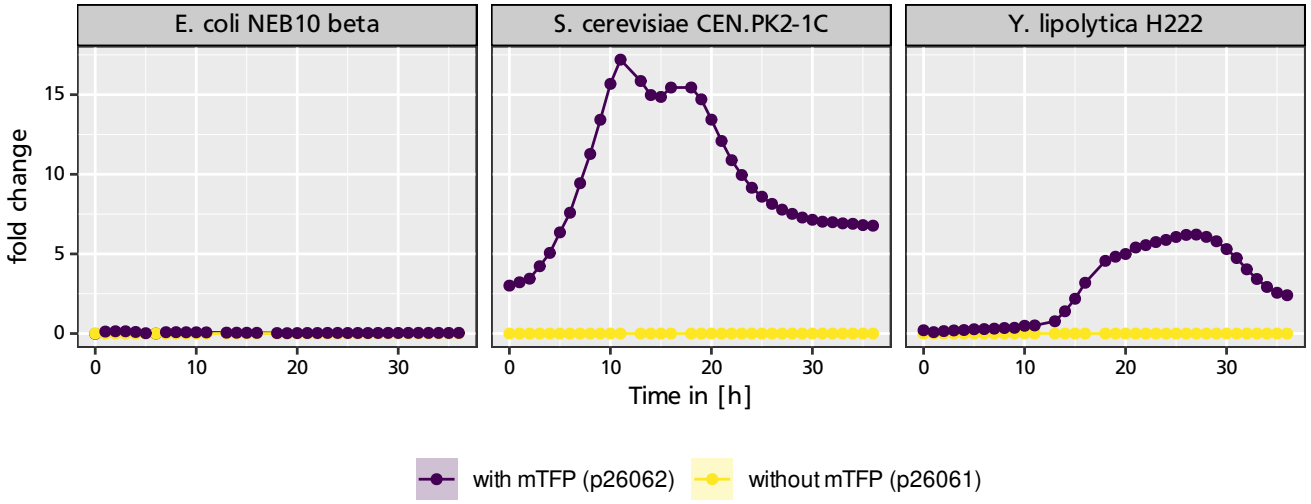


Figure A.1.: Fold change of mTFP (p26062) in *E. coli*, *S. cerevisiae* and *Y. lipolytica* with FLEXpress vector. Fold change was calculated against negative control (p26061) without mTFP by division of means. $foldchange = (\overline{FInorm_i} - \overline{FInorm_c})/\overline{FInorm_c}$ (i - target strain, c - control strain)

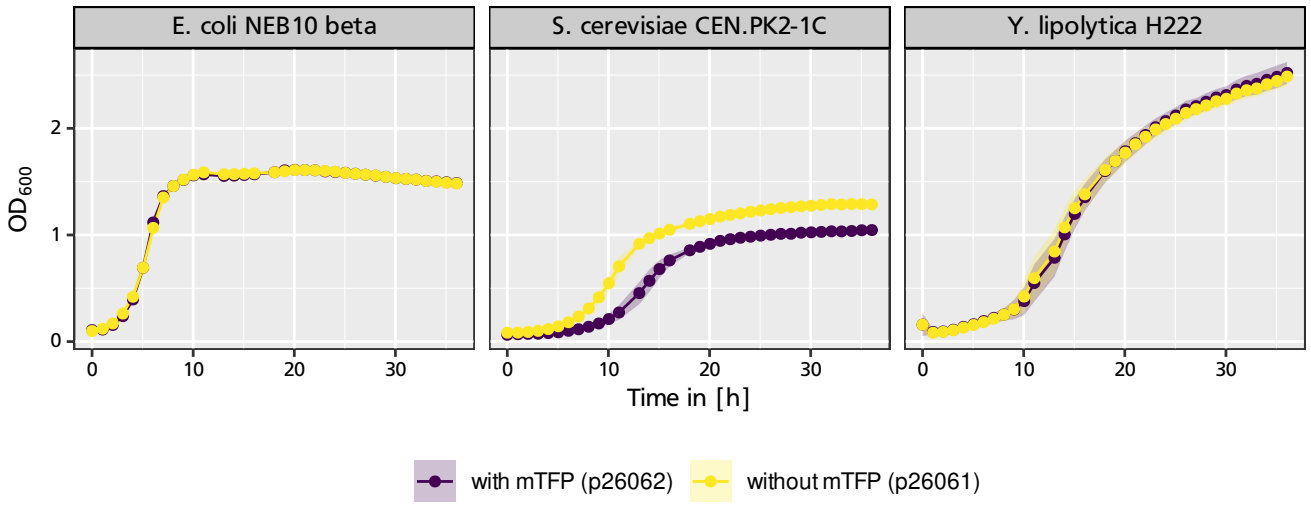


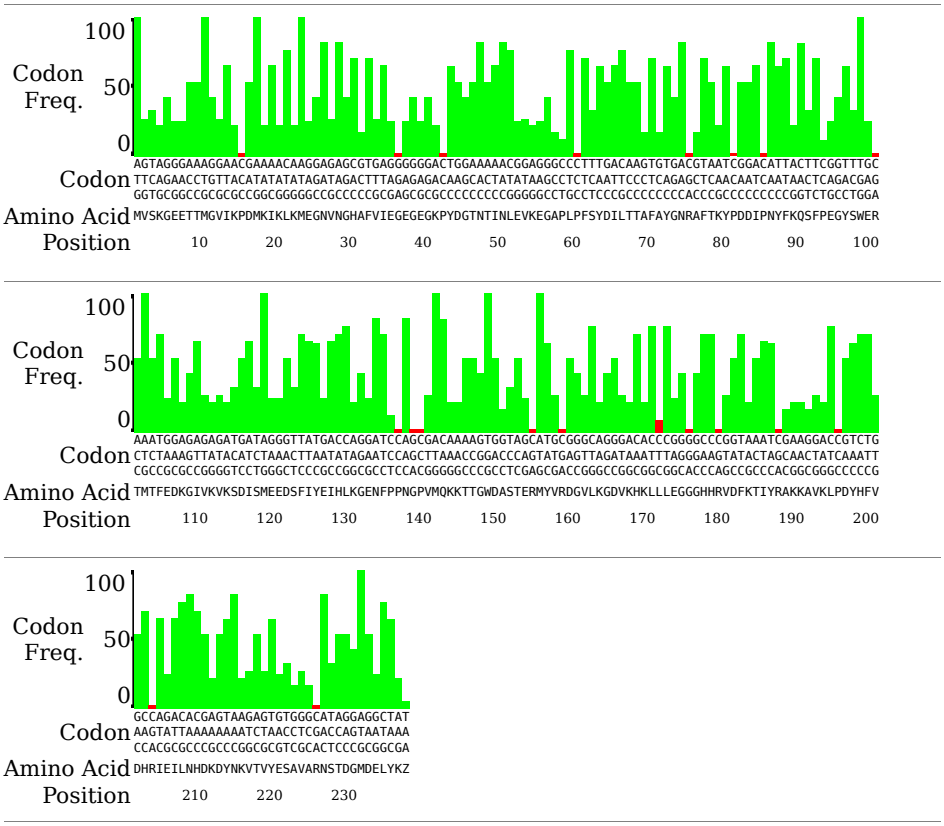
Figure A.2.: Growth of *E. coli*, *S. cerevisiae* and *Y. lipolytica* with p26061 (without mTFP) and p26062 (with mTFP). The ribbon indicates the standard deviation of n = 40 samples for *E. coli* and *Y. lipolytica* and n = 37 samples for *S. cerevisiae*, respectively.

***E. coli* Codon Usage Analysis 2.0** by Morris Maduro

Date: Sat 26 Jun 121 - 10:53 am **Sequence**: (mTFP)

Sequence entered (711 bases):

```
5'-ATGGTGTCTA AGGGCGAGGA GACCACCATG GGCGTGATCA AGCCCGACAT GAAGATCAAG CTGAAGATGG AGGGCAACGT
   GAACGGCCAC GCCTTCGTGA TCGAGGGAGA GGGCGAGGGC AAGCCCTACG ACGGCACCAA CACCATCAAC CTGGAGGTGA
   AGGAGGGCGC CCCTCTGCCC TTCTCTTACG ACATCCTGAC CACCGCCTTC GCCTACGGCA ACCGAGCCTT CACCAAGTAC
   CCCGACGACA TCCCCAACTA CTTCAAGCAG TCTTTCCCTG AGGGCTACTC TTGGGAGCGA ACCATGACCT TCGAGGACAA
   GGGCATCGTG AAGGTGAAGT CTGACATCTC TATGGAGGAG GACTCTTTCA TCTACGAGAT CCACCTGAAG GGCGAGAACT
   TCCCTCCCAA CGGACCCGTG ATGCAGAAGA AGACCACCGG CTGGGACGCC TCTACCGAGC GAATGTACGT GCGAGACGGC
   GTGCTGAAGG GCGACGTGAA GCACAAGCTG CTCCTGGAGG GCGGAGGCCA CCACCGAGTG GACTTCAAGA CCATCTACCG
   AGCCAAGAAG GCCGTGAAGC TGCCCGACTA CCACTTCGTG GACCACCGAA TCGAGATCCT GAACCACGAC AAGGACTACA
   ACAAGGTGAC CGTGTACGAG TCTGCCGTGG CCCGAAACTC TACCGACGGC ATGGACGAGC TGTACAAGTA A-3'
```



Colors: ■ = less than 10% of codons for same amino acid; ■ = at least 10%

Fraction of sense codons below threshold (=10.00): **20/236 (8%)**

-- End of report --

Done

Figure A.3.: Codon usage analysis of mTFP gene. E. coli Codon Usage Analyzer 2.1 was used by pasting nucleotide sequence in the text box on the website ucr.edu. Problematic codons marked red.

A.2.3. Additional methods: cultivation conditions for automated aptamer cultivation experiment

The cultivation of Kötter *et al.* was adapted for an automation workflow [204]. ACH medium was used for transformation because it contains no or only small amounts of uracil, and could therefore be used to select for uracil auxotrophy of the *S. cerevisiae* and *Y. lipolytica* strains. However, the yeasts have been transferred to YNB agar plates which were supplemented with HIS, LEU and TRP to adapt the metabolism to minimal media. For preparation of the experiment, the different yeast strains were used: *S. cerevisiae* CEN.PK2-1C (s26419) with p26061, p26062 and p26063 and *Y. lipolytica* H222 (s11084) with p26061, p26062 and p26063. To prepare the preculture plates, two standard F-bottom 96 well plates (GBO 655180) were filled with YNB + HIS, LEU, TRP). Each strain was inoculated four times on each preculture plate. The two plates were necessary because one plate was to be induced with water as a control and the other plate with the inducer tetracycline directly after randomization.
Plate reader and incubator were set to 30 °C. These two plates were used in a randomization workflow in the automation platform (section 4.4.3). The tower shaker of the incubator was set to 1000 rpm. The new created main-culture plates (GBO 655101) were subsequently induced by the 96 well head of the liquid handler (250 µM Tetracycline (final) and ddH_2O as control). The FI and OD was measured every 1 h in the BMG PHERAstar FSX®. A flow cytometry measurement was done every 24 h, which was done by cloning of the main-culture plate. Therefore, 10 µL were transferred in a new F-bottom 96 well plates (GBO 655101) which was filled with 190 µL Sony Sheath Fluid and measured in the Beckman-Coulter Cytoflex S.
The following script versions were used for the cultivation of the shown data:

- Composer-Script: commit 6293ba78b2ecef32ad4ffddfe2a4ceb91cceefa3
- randomizer (R): commit dc9f2dc02618cf724975459249e196909ebc75be
- PHERAread (R): commit a48a2f43e970d182646df426a73563cc4123da39

Finally, the data were evaluated and plotted using different R-based scripts.

- Plate reader data: aptamer_cultivation_drsudo.R
- Cytoflex data: evaluate_flowStats_aptamers.R

A.2.4. Additional methods: cultivation conditions for manual aptamer experiment

A precultures of *Y. lipolytica* H222 (s11084) and *S. cerevisiae* CEN.PK2-1C (s26419) with p26061, p26062 and p26063 were set in 25 mL ACH-media with TRP to complement auxotrophy of the used *S. cerevisiae* strain. Cultures were incubated for 48 h at 28 °C and 180 rpm. To have the same inoculation volume, the cultures were centrifuged at 3000xg for 5 min in a 50 mL reaction tube. The pellets were re-suspended in the same volume ACH + TRP. Subsequently, the OD was determined as 1:100 dilution. The cultures were diluted to an OD of 4 and measured again. The main cultures were set in triplicates with a starting OD of 0.01 in 30 mL ACH + TRP and incubated for 48 h at 28 °C and 180 rpm. The cultures were directly induced using ddH_2O and Tetracycline (250 µM final). The OD was measured every 3 h. A sample for flow cytometry was taken after 24 and 48 h. 1 mL was taken and stored subsequently on ice. The samples were centrifuged and 4000xg for 10 min at 4 °C. The supernatant was decanted, and the pellet re-suspended in Sony Sheath Fluid. As suggested by Sony, the samples were further diluted (1:4) and measured in Sony SH800SA. Finally, the data were evaluated and plotted using different R-based scripts.

- Cultivation data: aptamer_cultivation.R

- Sony flow cytometer data: yarrowia_samples.R

A.2.5. Additional figures: Aptamer cultivation

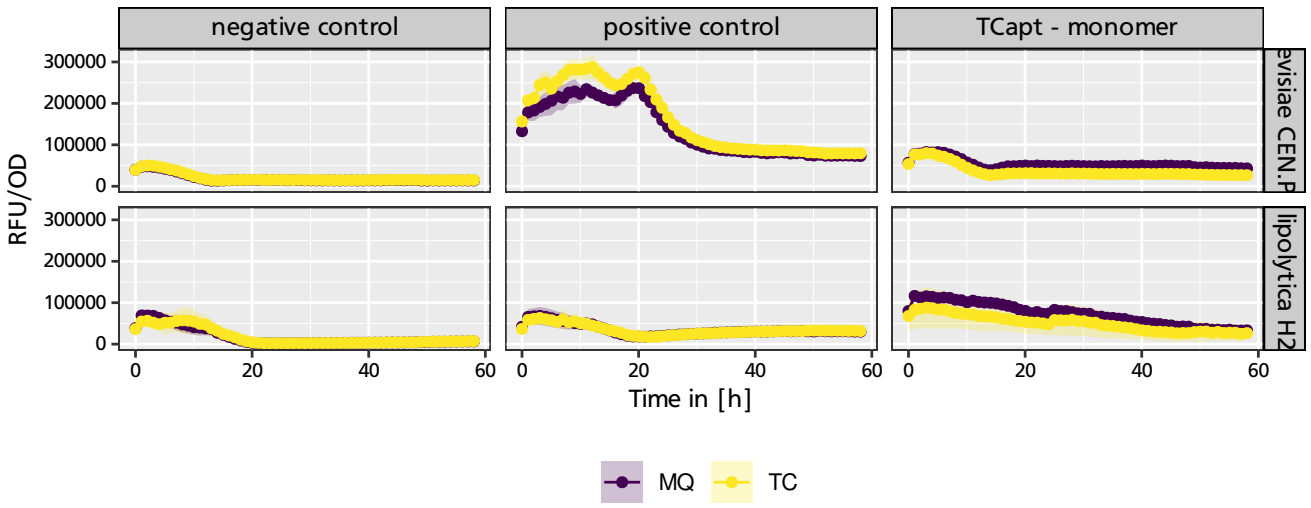


Figure A.4.: OD corrected FI data for the aptamer cultivation. The ribbon indicates the standard deviation of at least n = 10 samples for *Y. lipolytica* and *S. cerevisiae*, except *Y. lipolytica* H222 with p26063 (n = 7). Inducer: MQ – ddH_2O; TC – Tetracycline (250 µM).

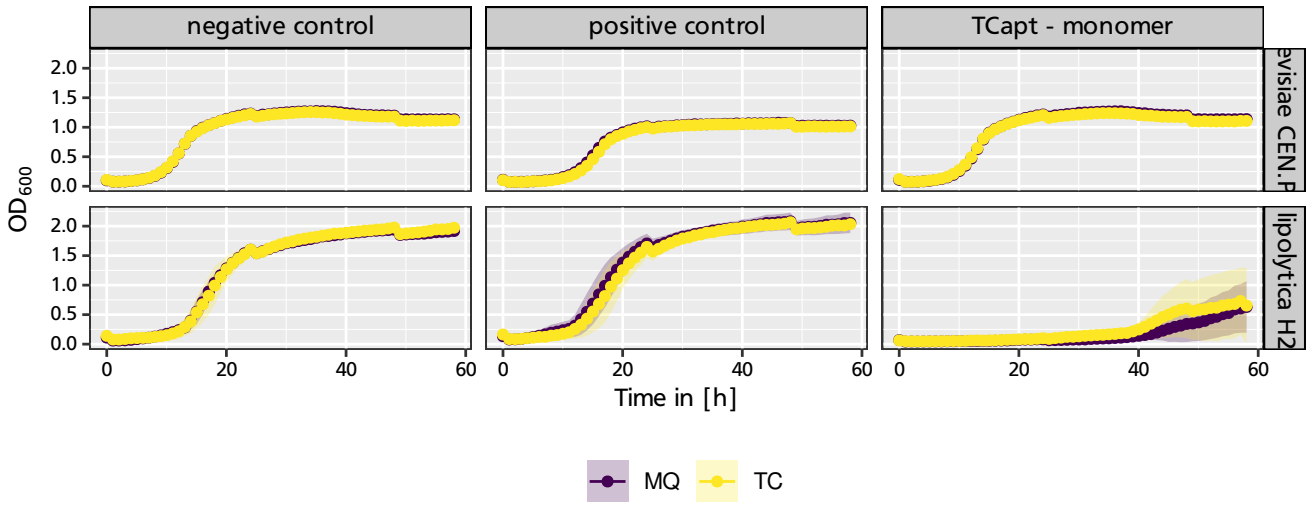


Figure A.5.: OD data for the aptamer cultivation. The ribbon indicates the standard deviation of at least n = 10 samples for *Y. lipolytica* and *S. cerevisiae*, except *Y. lipolytica* H222 with p26063 (n = 7). Inducer: MQ – ddH_2O; TC – Tetracycline (250 µM).

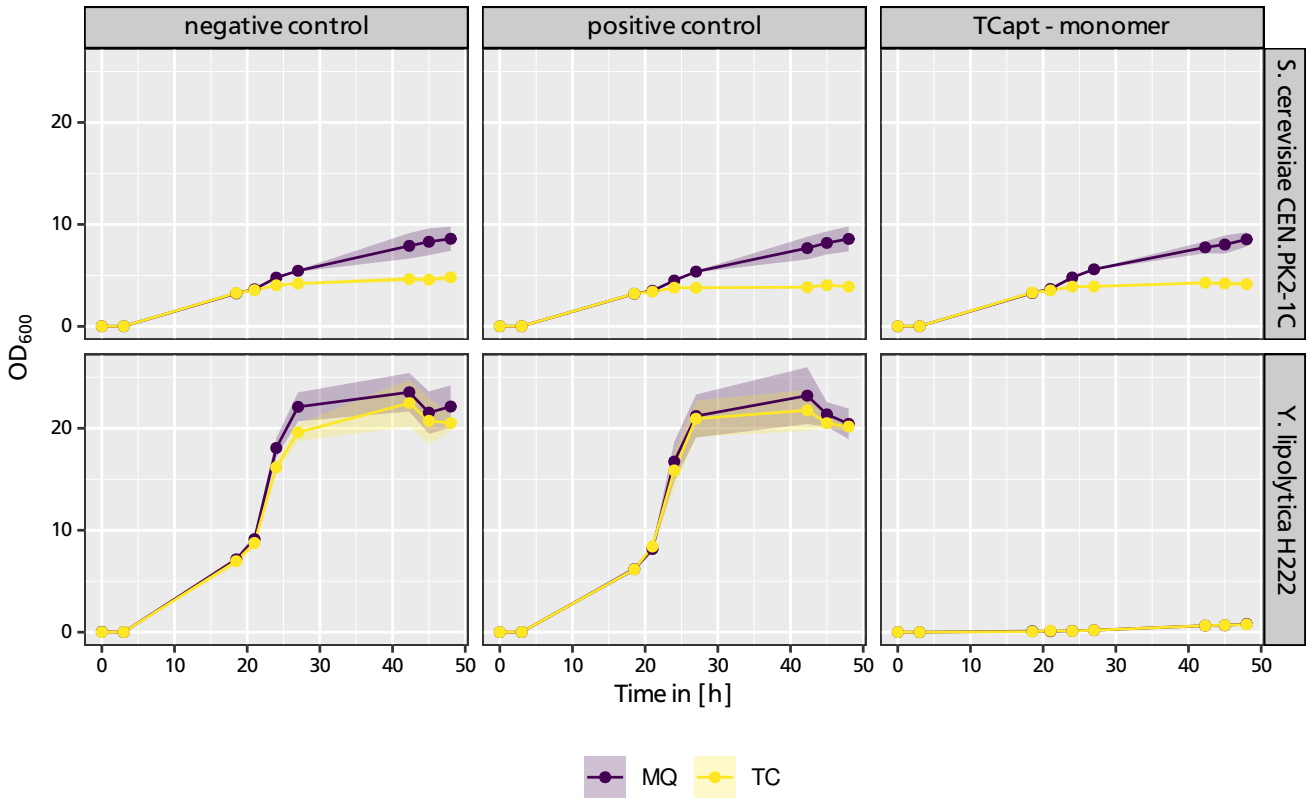


Figure A.6.: Cultivation of *S. cerevisiae* CEN.PK2-1C and *Y. lipolytica* H222 with TCapt-monomer in ACH + TRP media over 48 h. Negative control was p26061 (empty vector) and positive control p26062 (P_{tefag}-mTFP). Inducer: MQ – ddH_2O; TC – Tetracycline (250 µM). The ribbon indicates the standard deviation calculated by three independent samples.

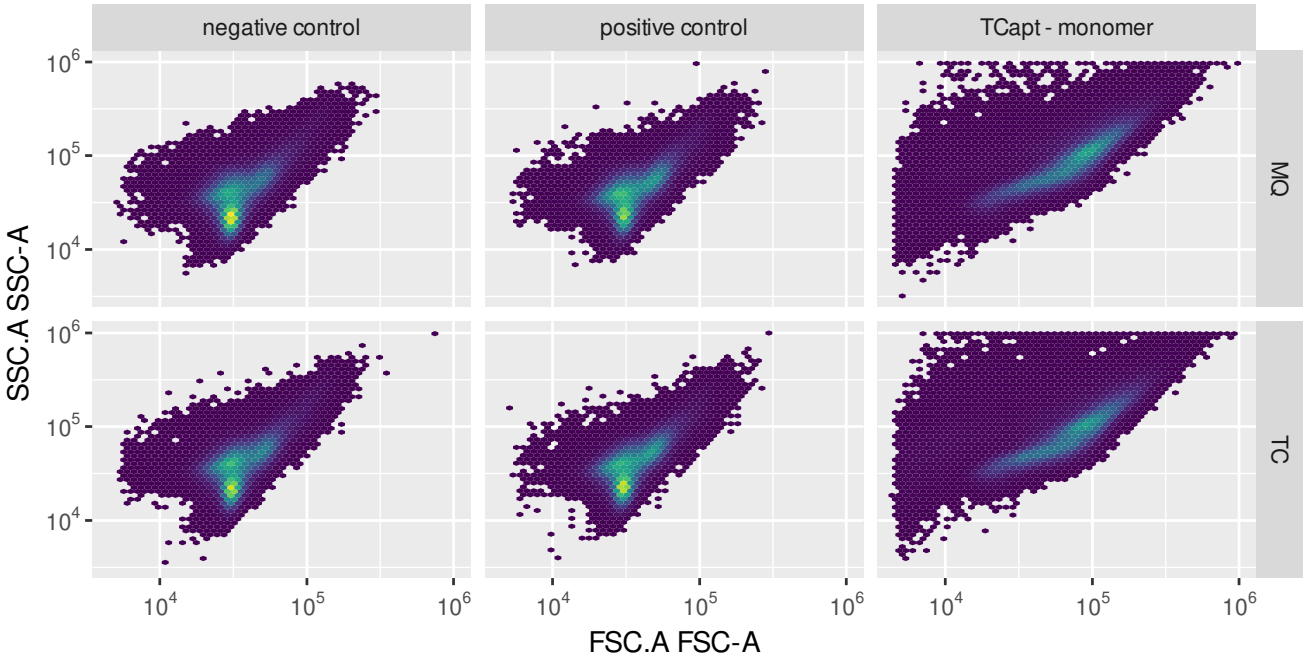


Figure A.7.: Additional flow cytometry data of *Y. lipolytica* H222 after 24 h after manual cultivation. Samples were measured in the Sony SH800SA in analyzing mode. Plot shows FSC versus SSC of unprocessed flow cytometry data. Negative control was p26061 (empty vector) and positive control p26062 (P_{tefag}-mTFP). Inducer: MQ – ddH_2O; TC – Tetracycline (250 µM).

Sequence View

Search Parameters: QGRS Max Length: 30 | Min G-Group Size: 2 | Loop size: from 0 to 36 | Loop search string:

Gene Information	
Gene ID:	**Number of Products:** 1
Gene Symbol:	**Number of poly A Signals:**
Gene Size: 69 nt.	**QGRS found:** 1
	QGRS found (including overlaps): 2

000001 GGCCTAGGTG GTCGTATTCT TCACCTCTCC AGATTAAAGC GC**GG**GT**GG**CG ATCT**GG**TATG TTTTA**GG**CC

Total time: 0.37813186645508 ms

Figure A.8.: Analysis report of the DNA sequence of the aptamer (monomer). The sequence (as well as reverse complement) was added to the web interface and the analysis was carried out at standard settings [195]. The found G4 motive was highlighted in yellow by the software.

Sequence View

Search Parameters: QGRS Max Length: 30 | Min G-Group Size: 2 | Loop size: from 0 to 36 | Loop search string:

Gene Information	
Gene ID:	**Number of Products:** 1
Gene Symbol:	**Number of poly A Signals:**
Gene Size: 171 nt.	**QGRS found:** 2
	QGRS found (including overlaps): 4

000001 GGCCTAGGTG GTCGTATTCT TCACCTCTCC AGATTAAAGC GCGGGTGGCG ATCTGGTATG TTTTAGGCCT
CGAGTTGTTG TTGTTGTTGT TGTTGTTGTT
000101 GACCGGTAGG TGGTCGTATT CTTCACCTCT CCAGATTAAA GCGCGGGTGG CGATCTGGTA TGTTTTACCG
G

Total time: 0.39815902709961 ms

Figure A.9.: Analysis report of the DNA sequence of the aptamer (dimer). The sequence (as well as reverse complement) was added to the web interface and the analysis was carried out at standard settings [195]. The found G4 motive was highlighted in yellow by the software.

Sequence View

Search Parameters: QGRS Max Length: 30 | Min G-Group Size: 2 | Loop size: from 0 to 36 | Loop search string:

Gene Information	
Gene ID:	**Number of Products:** 1
Gene Symbol:	**Number of poly A Signals:**
Gene Size: 247 nt.	**QGRS found:** 3
	QGRS found (including overlaps): 8

000001 GGCCTAGGTG GTCGTATTCT TCACCTCTCC AGATTAAAGC GCGGGTGGCG ATCTGGTATG TTTTAGGCCT
CGAGTTGTTG TTGTTGTTGT TGTTGTTGTT
000101 GACCGGTAGG TGGTCGTATT CTTCACCTCT CCAGATTAAA GCGCGGGTGG CGATCTGGTA TGTTTTACCG
GTTCTTAAGG CCTAGGTGGT CGTATTCTTC
000201 ACCTCTCCAG ATTAAAGCGC GGGTGGCGAT CTGGTATGTT TTAGGCC

Total time: 1.4200210571289 ms

Figure A.10.: Analysis report of the DNA sequence of the aptamer (trimer). The sequence (as well as reverse complement) was added to the web interface and the analysis was carried out at standard settings [195]. The found G4 motive was highlighted in yellow by the software.

A.3. CRISPRpads: Additional Results and Methods

A.3.1. CRISPRpad design and calculation

The following computational design was done by Knut Finstermeier (Charpentier Group, MPI for the Science of Pathogens, Berlin). The NCBI genome sequences for *B. subtilis* PY79, *E. coli* BL21, *S. cerevisiae* S288C and *Y. lipolytica* W29 were concatenated to one sequence. The reference sequence was circularized and split into overlapping segments. This was done for the reverse complement as well. Afterwards, common five-mers were identified for the reference sequence and the reverse complement. The next step was to prepare the mismatch matrix by pre-calculation of all possible combinations of five-mers while applying activity scores based on position and nucleotide mismatch activity [82]. To reflect overall mismatch activity, the segment-specific activities were multiplied with each other. In the next step, four-mers were created, and all possible designs were tested against the combined genome sequence. The obtained designs were filtered by inclusion of rare five-mers until a certain threshold for the reference sequence and the reverse complement was reached. All designs with GC >65 % were excluded from the pool as well as homopolymers with >4 bp. The Vienna 2.0 package [231] was used to analyze secondary structures interfering with protospacer activity. Designs, which are too similar, were excluded by calculating the Hamming distances. The remaining sequences were ranked based on a score which was calculated by considering folding, maximized on-target activity and minimized off-target activity. The ranked protospacer were used to design CRISPRpads by taking into account a minimized off-target activity, within-pad off-target activity and number of utilized protospacers while the structural variance was maximized. The protospacer were labeled as linker (L) and base (B). The criteria were 'GG' at position 7-8 or 'CC' at position 13-14 to be labeled as linker. Other protospacer were assigned as base. The labeled parts were paired based on identical end-sequences: L-B-L, B-L (left end), L-B (right end). Pairs with isolated left and right end were discarded. The iterative design started with left ends. The assembled chains were ranked according to their off-target activity against the reference sequence and selected for the lowest sum after each iteration. The cutoff was 100. The iteration procedure was terminated when either the requested length was reached or the requested number of protospacer. All designs <90 % of target length were discarded. The kept sequences were scored, sorted, and selected for low within-pad off-target activity as well as low reference genome off-target activity. The first 1000 sequences were kept. By counting the mismatches between base and linker elements, a distance matrix was calculated. Based on the highest rank, designs were selected and the distance to other designs was maximized. Finally, the CRISPRpad designs were aligned, and the similarity was determined using the Hamming distance and Ward clustering.
Next, the designs Rank1 and Rank4 were selected and synthesized at Thermo Fisher Scientific GENEART GmbH. Both designs share a larger part of homology in the center of the CPad (Fig. A.11) because the design resulted into a selection of similar protospacers. Although, Rank1 contains 25 protospacer and Rank4 contains 26 protospacer, only 31 unique protospacer were used from the set of 50 protospacer which can be found in Tab. A.5.

Table A.5.: Available protospacer in Rank1 and Rank4 without PAM sequence.

ID	Sequence (5' → 3')		ID	Sequence (5' → 3')
Pro_1	CTAAGCCTACTAGACCTATA		Pro_17	CTTACGTCGCACGGATCGGG
Pro_2	TCGGGTCCGTCTAACGTAAC		Pro_18	CTAGGGACGCTATAGGGGCT
Pro_3	CTTACGACGCACGCATACGC		Pro_19	CGTTCGCTAACGTCGGGGCT
Pro_4	CTTACCTCCGTAGTATCGCA		Pro_20	CTAGGATCCGGCACGACGCT

Continued on next page

Table A.5 – continued from previous page

ID	Sequence (5' → 3')		ID	Sequence (5' → 3')
Pro_5	CTTACCTCCGTTACGACGCA		Pro_21	TTAGGGACCCGTACGCTAAC
Pro_6	GTAACCGTACGTTCGCTAAC		Pro_22	TGCGACGACTTAGCACCCTA
Pro_7	CTTACGCACGGTACGGGTCC		Pro_23	TGCGATCCGTGGGCACCTAA
Pro_8	CTAGGCCTATAGTCCCTTAC		Pro_24	TACGGGTAGCCCTAGGGTCC
Pro_9	CTAATGCTACTAGCCCTATC		Pro_25	GTAAGCGTACGCTCGTCTAG
Pro_10	GCTAACGCTCGGTCCCGTAC		Pro_26	CCGGATCCGTACGCACCTAC
Pro_11	CTTACCTAGGGGACGACGCA		Pro_27	CTCGCACTTAGGTCGTCTAG
Pro_12	GGTACGACGCGTTCGCTAAC		Pro_28	TGCGTTACGTTAGACCCGTA
Pro_13	ATGCGTCGTGCGAACCGTTA		Pro_29	CGACGAGCGGCGTAGATAGC
Pro_14	CTCGGTAGGTAGTCCCTTAG		Pro_30	TAGGGTCGCGTAGTCCGGGT
Pro_15	TCCGTCGTGCACGTGTCGCA		Pro_31	CGCAGTGCTATAGCTCGCCG
Pro_16	TGCGTAGTCCCTACCCGACG			

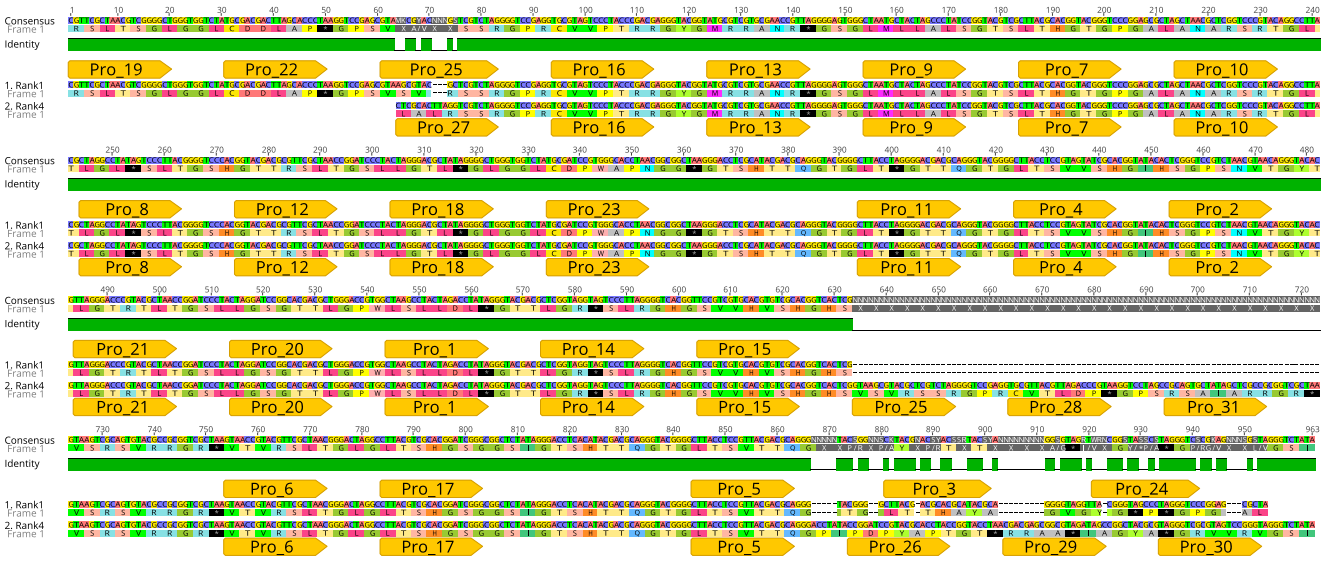


Figure A.11.: Sequence alignment of the ordered sequences Rank1 and Rank4 with Geneious R10 sequence alignment tool.

Table A.6.: Off-Target sequences identified by cas-offinder using version 2.4.1 as offline package for linux. The input- and output-file can be found in the git repository. The details column contains the organism name and strain as well as the GenBank IDs of the FASTA files which were used for the analysis. The sgRNA column contains the sgRNA without PAM sequence. Chromosomal sequence contains small letters which indicate mismatches to the sgRNA. The mismatch column contains the number of mismatches of the sgRNA::DNA duplex. The threshold was set from one to a maximum of four mismatches.

PS	Strain (GenBank ID)	sgRNA	genomic sequence	MM
Bacillus subtilis				
Pro_1	PY79 (NC_022898.1)	CTAAGCCTACTAGACCTATA	CTtAGCCTACcAaACCTcTAAGGATGA	4
Pro_9	PY79 (NC_022898.1)	CTAATGCTACTAGCCCTATC	tTAcaGCTACTAtCCCTATCCGGTGCC	4
Pro_23	PY79 (NC_022898.1)	TGCGATCCGTGGGCACCTAA	cGCGATCaGTGcGCAtCTAAAGGAGCA	4
Pro_25	PY79 (NC_022898.1)	GTAAGCGTACGCTCGTCTAG	GTAAGCGccCGCTCtTCaAGAGGTCTT	4
Pro_28	PY79 (NC_022898.1)	TGCGTTACGTTAGACCCGTA	TGCGTcAgGTTAcACCaGTAAGGAAAT	4
Escherichia coli				
Pro_5	BL21 (NZ_CP060121.1)	CTTACCTCCGTTACGACGCA	CTTAgCTCtGcTACGtCGCATGGATTT	4
Pro_7	BL21 (NZ_CP060121.1)	CTTACGCACGGTACGGGTCC	gTcACGCACGGTACcGGTaCCGGAGCA	4
Pro_22	BL21 (NZ_CP060121.1)	TGCGACGACTTAGCACCCTA	gGCGgCGACTgAGCACCCaAGGGTGAT	4
Pro_5	BW25113 (NZ_CP064677.1)	CTTACCTCCGTTACGACGCA	CTTAgCTCtGcTACGtCGCATGGATTT	4
Pro_16	BW25113 (NZ_CP064677.1)	TGCGTAGTCCCTACCCGACG	cGCGgAGTCgCTgCCCGACGCGGTGGT	4
Pro_22	BW25113 (NZ_CP064677.1)	TGCGACGACTTAGCACCCTA	gGCGgCGACTgAGCACCCaAGGGTGAT	4
Saccharomyces cerevisiae				
Pro_8	BY4742; chr. XVI (CP026290)	CTAGGCCTATAGTCCCTTAC	CTAtGatTgTAGTCCCTTACAGGATTC	4
Pro_25	BY4742; chr. VII (CP026294)	GTAAGCGTACGCTCGTCTAG	GTAAGCGTcCaCTCGTCatGCGGTCCA	4
Pro_2	BY4742; chr. IV (CP026298)	TCGGGTCCGTCTAACGTAAC	TttGGTCCcTtTAACGTAACTGGTAAA	4
Pro_1	BY4742; chr. V (CP026299)	CTAAGCCTACTAGACCTATA	CTAAcCCTtCTAGACgTATAGGGAATT	3
Pro_1	BY4742; chr. XV (CP026303)	CTAAGCCTACTAGACCTATA	CTAcctCTACTAtACCTATAGGGAACT	4
Pro_2	S288C; chr.IV (NC_001136)	TCGGGTCCGTCTAACGTAAC	TttGGTCCcTtTAACGTAACTGGTAAA	4
Pro_1	S288C; chr.V (NC_001137)	CTAAGCCTACTAGACCTATA	CTAAcCCTtCTAGACgTATAGGGAATT	3
Pro_25	S288C; chr.VII (NC_001139)	GTAAGCGTACGCTCGTCTAG	GTAAGCGTcCaCTCGTCatGCGGTCCA	4
Pro_1	S288C; chr.XV (NC_001147)	CTAAGCCTACTAGACCTATA	CTAcctCTACTAtACCTATAGGGAACT	4
Pro_8	S288C; chr.XVI (NC_001148)	CTAGGCCTATAGTCCCTTAC	CTAtGatTgTAGTCCCTTACAGGATTC	4
Yarrowia lipolytica				
Pro_31	W29; chr. A (CP017553.1)	CGCAGTGCTATAGCTCGCCG	aGCcGTGCTATAGCTCaCCaGGGTCAT	4
Pro_18	W29; chr. B (CP017554.1)	CTAGGGACGCTATAGGGGCT	CcAttGACGgTATAGGGGCTTGGAGTC	4
Pro_24	W29; chr. B (CP017554.1)	TACGGGTAGCCCTAGGGTCC	aAgGGGgAGCCtTAGGGTCCTGGGAGG	4

Continued on next page

Table A.6 – continued from previous page

PS	Strain (GenBank ID)	sgRNA	genomic sequence	MM
Pro_3	W29; chr. C (CP017555.1)	CTTACGACGCACGCATACGC	CTTgaGACGCACGtATgCGCCGGGAAC	4
Pro_5	W29; chr. C (CP017555.1)	CTTACCTCCGTTACGACGCA	CTTACCTCgGTTACcAgGCtTGGCCTC	4
Pro_15	W29; chr. D (CP017556.1)	TCCGTCGTGCACGTGTCGCA	TCCGTCGTGCAtaTaTaGCATGGGCGA	4
Pro_24	W29; chr. D (CP017556.1)	TACGGGTAGCCCTAGGGTCC	TAgGGGTAtCCCTgGGGcCCGGGGGCG	4
Pro_5	W29; chr. E (CP017557.1)	CTTACCTCCGTTACGACGCA	CTTACCTCgGTTACcAgGCtTGGAAGT	4
Pro_19	W29; chr. E (CP017557.1)	CGTTCGCTAACGTCGGGGCT	CGTTCGgTAACaTCGcGGCgAGGCTAT	4
Pro_22	W29; chr. E (CP017557.1)	TGCGACGACTTAGCACCCTA	ctCGACGACTTgGCtCCCTATGGCGGC	4
Pro_20	W29; chr. F (CP017558.1)	CTAGGATCCGGCACGACGCT	CTtGGATCtcGaACGACGCTGGGATCT	4
Pro_31	PO1f; chr. A (CM002778.1)	CGCAGTGCTATAGCTCGCCG	aGCcGTGCTATAGCTCaCCaGGGTCAT	4
Pro_18	PO1f; chr. B (CM002779.1)	CTAGGGACGCTATAGGGGCT	CcAttGACGgTATAGGGGCTTGGAGTC	4
Pro_24	PO1f; chr. B (CM002779.1)	TACGGGTAGCCCTAGGGTCC	aAgGGGgAGCCtTAGGGTCCTGGGAGG	4
Pro_3	PO1f; chr. C (CM002780.1)	CTTACGACGCACGCATACGC	CTTgaGACGCACGtATgCGCCGGGAAC	4
Pro_5	PO1f; chr. C (CM002780.1)	CTTACCTCCGTTACGACGCA	CTTACCTCgGTTACcAgGCtTGGCCTC	4
Pro_15	PO1f; chr. D (CM002781.1)	TCCGTCGTGCACGTGTCGCA	TCCGTCGTGCAtaTaTaGCATGGGCGA	4
Pro_24	PO1f; chr. D (CM002781.1)	TACGGGTAGCCCTAGGGTCC	TAgGGGTAtCCCTgGGGcCCGGGGGCG	4
Pro_5	PO1f; chr. E (CM002782.1)	CTTACCTCCGTTACGACGCA	CTTACCTCgGTTACcAgGCtTGGAAGT	4
Pro_19	PO1f; chr. E (CM002782.1)	CGTTCGCTAACGTCGGGGCT	CGTTCGgTAACaTCGcGGCgAGGCTAT	4
Pro_22	PO1f; chr. E (CM002782.1)	TGCGACGACTTAGCACCCTA	ctCGACGACTTgGCtCCCTATGGCGGC	4
Pro_20	PO1f; chr. F (CM002783.1)	CTAGGATCCGGCACGACGCT	CTtGGATCtcGaACGACGCTGGGATCT	4

A.3.2. Strain construction to evaluate CRISPRpads in *S. cerevisiae* BY4742

First, the *dCas9* gene was integrated in the strain *S. cerevisiae* BY4742. The yeast cells were prepared for chemical transformation. The plasmid pJZC518 (p26126) was prepared for integration by digestion with *Pme*I. The *S. cerevisiae* cells were transformed using the plasmid fragment and plated on YNB plates supplemented with HIS, LYS and URA. Integration into the *leu2* locus was verified by colony PCR and sequencing. The resulting strain was the s26445.
Afterwards, chemical competent cells were made from s26445. Again, the plasmid pJZC522 (p26128) was prepared for integration by digestion with *Pme*I to integrate the MCP-VP64 fusion into the *his3* locus. The s26445 cells were transformed using the plasmid fragment and plated on YNB plates supplemented with LYS and URA. Integration into the *his3* locus was verified by colony PCR and sequencing. The resulting strain was the s26446.
The next step was to integrate the CPad using a dominant marker flanked by *loxP* sites. The integration fragment was amplified by PCR using oligos o26332/o26338. The strain s26446 was transformed using 250 fmol of the cleaned PCR fragment by chemical transformation and plating on YPD-G418 (400 µg/mL). The clones were verified by PCR and sequencing. The obtained strain was s26447 and was subsequently used to remove the KanMX marker. This was done by transformation with the Cre recombinase plasmid p26125. For this purpose, chemically competent cells were prepared and used to be transformed with 20 fmol of plasmid DNA. The transformation mixture was plated on ACH + URA agar plates. Afterwards, the colonies were screened for clones without KanMX marker, using colony PCR. Clones that showed a negative signal in PCR were spread on 5-FOA plates to remove the Cre recombinase plasmid. The marker removal was validated again by isolation of genomic DNA, PCR with a high-fidelity polymerase, and sequencing. The ID s26449 was assigned to the strain.
The next step was the integration of the reporter gene cassette $P_{CYC1min}$-sfGFP-T_{CYC1}. The integration fragment was amplified by PCR using NEB Q5® polymerase with the oligos o11064/o26211 from p26153. Afterwards, the PCR mixture was purified with the Macherey-Nagel NucleoSpin Gel and PCR-Clean-up-Kit. The transformation was performed by chemical transformation with 250 fmol of the PCR fragment and plating on YPD-Hygromycin plates. The integration was verified by PCR and sequencing. The identifier s26454 was assigned to the strain.
Finally, a P_{PDA1}-mCherry cassette was integrated into the genome of s26454 to normalize the fluorescence values to reflect the current state of the metabolism. Thus, the promoter of the E1 alpha subunit of the pyruvate dehydrogenase complex was suggested by AG Boles (Goethe-Universität, Frankfurt am Main) and used for this purpose. The plasmid p426MET25-Envy from Leonie Baumann (AG Boles, Goethe-Universität, Frankfurt am Main) had only a marker which could not be used in the previously developed strain, so a marker exchange was carried out as described in section A.1.5. For integration, the cassette was amplified by PCR from plasmid p26162 using o26482 and o26481. Due to the 2-micron ori, the plasmid can be stably propagated in yeast. Thus, special treatment of the PCR mixture was required. *Dpn*I was directly added after the PCR reaction. After the cleanup, the product was used for a chemical transformation of *S. cerevisiae* s26454. The obtained clones were verified by PCR and sequencing. Finally, the ID s26458 was assigned, and the strain was used for the CRISPRa experiments.

A.3.3. Cultivation of CRISPRpad strains derived from *S. cerevisiae* BY4742

The cultivation was carried out in the CompuGene Robotics platform. The strain s26458 was used for a transformation with the respective sgRNA delivery plasmids. The obtained clones were used to inoculate the preculture plate. One colony per strain was used to inoculate the precultures. The plate (GBO 655180) was filled with 200 µL of media (YNB + 2 % D-glucose (section 2.3) and inoculated by means of toothpicks. The precultures were introduced in the automation platform and cultivated with the randomization workflow

(section 4.4.3). Plate reader and incubator were set to 30 °C. The tower shaker of the incubator was set to 1000 rpm.
The following script versions were used for the cultivation of the shown data:

- Composer-Script: commit 91f61e3b65fd6e59a9e66b85a71c280ebf13b357
- Randomizer (R): commit 8a03f5882105909b90a4d1279a8c99871fe7749e
- PHERAread (R): commit f77eadfdb026dfd4308f94ff14c13e577a1aeb26

The PHERAstar® measured OD and FI using a filter-set for GFP. The gain for the FI measurement was set to 500. The script PHERAread was used to import the .csv files which were generated by the BMG PHERAstar®. The script $\mathrm{crisprpads\_cultivation.R}$ was used to evaluate the data and to generate the plots. Original data and the R scripts are available for review in the git repository of the thesis.

A.3.4. Strain construction to evaluate CRISPRpads in *Y. lipolytica* Po1f

Initially, the strain *Y. lipolytica* Po1f (s11085) was modified to shift the ratio between HR and NHEJ as described in section 3.3.1. *Y. lipolytica* was transformed using the plasmid p13008. The plasmid contains an active Cas9 with a guide sequence to target *KU70*. For verification, 40 clones were picked and spread on non-selecting plates. From the master plate, eight clones were selected together with untransformed yeasts and used for colony PCR. The fragments obtained were purified and verified by sequencing. Clones containing a mutation were inoculated in 1 mL non-selecting YPD in a 24 well MTP and cultivated for 24 h to get rid of the Cas9 plasmid. Dilution streaking was performed, and single colonies were selected. Chromosomal DNA was isolated and the *KU70* locus was amplified by PCR and verified by sequencing. One clone had an additional adenine base, resulting in a frameshift, which was one of the favored events. The strain ID s26235 was assigned.
The strain was further modified by transformation. The DNA fragment was amplified by PCR using oligos o26476 and o26211 and p26164 as template. The amplified fragment contained homology sites for the URA3 locus, as well as the $P_{TEF1core}$-sfGFP cassette, which was pre-integrated into the CPad. The transformation was done by heat shock transformation with 250 fmol DNA. The obtained clones were verified by colony PCR and sequencing of the *URA3* locus. The ID s26241 was assigned to one of the positive clones which was used for the CRISPRa experiments.

A.3.5. Cultivation of CRISPRpad strains derived from *Y. lipolytica* Po1f

The cultivation was carried out in the CompuGene Robotics platform. The strain s26241 was used for a transformation with the respective sgRNA delivery plasmids. The obtained clones were used to inoculate the preculture plate. One colony per strain was used to inoculate the precultures. The plate (GBO 655180) was filled with 200 µL of media (YNB + URA + 2 % D-glucose) (section 2.3) and inoculated by means of toothpicks. The precultures were introduced in the automation platform and cultivated with the randomization workflow (section 4.4.3). Plate reader and incubator were set to 30 °C. The tower shaker of the incubator was set to 1000 rpm.
The following script versions were used for the cultivation of the shown data:

- Composer-Script: commit dc07bdd4a038fc24240bc7a85406a5ad8e189799
- Randomizer (R): commit 8a03f5882105909b90a4d1279a8c99871fe7749e

- PHERAread (R): commit f77eadfdb026dfd4308f94ff14c13e577a1aeb26

The PHERAstar® measured OD and FI using a filter-set for GFP. The gain for the FI measurement was set to 400. The script PHERAread was used to import the .csv files which were generated by the BMG PHERAstar®. The script crisprpads_yl_cultivation.R was used to evaluate the data and to generate the plots. Original data and the R scripts are available for review in the git repository of the thesis.

A.3.6. Strain construction to evaluate CRISPRpads in *E. coli* BW25113

Integration into the *E. coli* genome is often done by Lambda Red mediated recombination. This method was developed by Datsenko and Wanner to create knockout strains [76]. For this purpose, the target strain *E. coli* BW25113 was transformed with the pKD46 helper plasmid containing the Lambda Red recombination machinery. For this purpose, a protocol from NEB was used (section 2.4.2). After the transformation, the plates were incubated on Amp containing LB plates at 30 °C, since the plasmid contains a temperature sensitive origin of replication. The genomic integration is described in section 2.4.2. The PCR fragments were specially treated to prevent the transfer of plasmid DNA from the template plasmid. The PCR fragments were amplified from p26195 and p26196 using the oligos o26509 and o26510. Both oligos have 50 bp homology to the target locus in the genome. After the PCR, the fragments were purified using Analytik Jena innuPREP PCRpure Kit with subsequent *Dpn*I digest in CutSmart® buffer, since the in-PCR digest is not efficient enough to prevent plasmid carryover. Afterwards, the DNA was purified again using Analytik Jena innuPREP PCRpure Kit. In contrast to the protocol from OpenWetWare, the cell suspension was washed three times in 1 mL ddH_2O. In addition, the protocol suggests that plating should be done on LB plates with 10 µg/mL and 25 µg/mL Chloramphenicol. However, 10 µg/mL Chloramphenicol quickly showed a formation of background, so that only 25 µg/mL Chloramphenicol plates were used.
The obtained colonies were checked for integration by colony PCR and sequencing. After successful confirmation, clones were spread on LB without selection and incubated at 43 °C to remove the pKD46 helper plasmid. Loss of the plasmid was determined via plate assay by plating the clones on LB Amp and non-selective LB plates. Unfortunately, integration of the sfGFP cassette that was supposed to be integrated after Pro_21 was not successful. Only clones for integration into Pro_21 were found (fragment of p26196). The strain with the successful integration of the Rank1-sfGFP construct got strain number s26901.3.
Afterwards, s26901.3 was transformed with the plasmids for the CRISPRa system. The original strain s99015 was used as a negative control for the experiments. The transformation was carried out in steps, so that the dCas9 plasmid (p99013) was transferred first using chemical transformation. The integrity of the dCas9 CDS as well as other important components was reconfirmed in the transformed strains using plasmid isolation and sequencing. The derived strains got the IDs s26902 (s99015 + p99013) and s26903 (s26901.3 + p99013). Finally, strains s26902 and s26903 were transformed using the plasmids for sgRNA delivery (p99011, p99016, p26189, p26190, 26191, p26199 and p26200). Again, this was done by chemical transformation according to the protocol of NEB. Successful integration and integrity of the guide sequence was confirmed via colony PCR with a high-fidelity polymerase and sequencing. Subsequently, new IDs were assigned to the derived strains, which can be found in section A.1.3.

A.3.7. Cultivation of CRISPRpad strains derived from *E. coli* BW25113

Testing different media for *E. coli* cultivation

The cultivation was carried out in the CompuGene Robotics platform. The strains s26902 and 26903 were used for a transformation with the respective sgRNA delivery plasmids. The obtained clones were used to inoculate the preculture plates. Three colonies per strain and sgRNA were used to inoculate the precultures. Two plates (GBO 655180) per media were filled with 200 µL of media and inoculated by means of toothpicks. The precultures were introduced in the automation platform and cultivated with the randomization workflow (section 4.4.3). Plate reader and incubator were set to 30 °C. The tower shaker of the incubator was set to 1000 rpm. Directly after randomization, each new plate (GBO 655101) was induced either with ddH_2O or 1.5 % (w/v) arabinose (final).
Different medias were tested to cultivate the different *E. coli* strains (each media was supplemented with Amp and Spec):

1. LB, MSM + 1 % (w/v) D-glucose, MSM + 1 % (w/v) D-fructose
    - Composer-Script: Commit f7282283f7a9bd9a3a17c0aeaede897cf87cf740
2. MSM + 0.05 % (w/v) D-glucose, MSM + 0.10 % (w/v) D-glucose, MSM + 0.25 % (w/v) D-glucose, MSM + 1.00 % (w/v) D-glucose
    - Composer-Script: Commit 1b0742bab7997d0c3b140cb4bfd5217980ac9dd5

The following script versions were used for the cultivation of the shown data:

- Randomizer (R): commit dc9f2dc02618cf724975459249e196909ebc75be
- PHERAread (R): commit a48a2f43e970d182646df426a73563cc4123da39

The PHERAstar® measured OD and FI using a filter-set for GFP. The data were imported by PHERAread, pre-processed (Ec_AraInduction.R), and collected in one folder. The script crisprpads_ec_cultivation_pre-processed.R was used to evaluate the data and to generate the plots. Original data and the R scripts are available for review in the git repository of the thesis.

Testing different inducer concentrations for *E. coli* cultivation

The cultivation was carried out in the CompuGene Robotics platform. The strains s26902 and 26903 were used for a transformation with the respective sgRNA delivery plasmids. The obtained clones were used to inoculate the precultures in 4 mL overnight culture tubes with MSM + 0.25 % D-glucose. The cultures were grown overnight at 30 °C and 180 rpm. After 16 h, the OD was determined as 1:10 dilution. For each culture, a dilution in MSM + 0.25 % D-glucose was prepared with a target OD of 0.05. The cultures were distributed on plates (GBO 655101) using the EpMotion. Each strain was represented by 4 replicates. In addition to the culture, 10 µL inducer was added in different concentrations: 0.00, 0.05, 0.10, 0.25, 0.50 and 1.0 % (w/v) arabinose (final). The final volume was 200 µL. The main cultures were introduced in the automation platform and cultivated with the cultivation workflow for 96 well plates. Plate reader and incubator were set to 30 °C. The tower shaker of the incubator was set to 1000 rpm.
The following script versions were used for the cultivation of the shown data:

- Composer-Script: commit 9f2b96725ac8736266235c02538f909e18a39f86

The PHERAstar® measured OD and FI using a filter-set for GFP. The script mars_import.R was used to import the .csv files. The script crisprpads_ec_MARSexp.R was used to evaluate the data and to generate the plots. Original data and the R scripts are available for review in the git repository of the thesis.
After 48 h, samples were taken from the strain s26902 and s26903 with plasmids p26199 and p26200. The samples were 1:20 diluted in Sony Sheath Fluid and measured in the Sony FACS in Analyzer Mode. The R script crisprpads_ec_FACS.R was used to import and evaluate the data, and to generate the plots.

A.3.8. Cultivation of CRISPRpad strains derived from *B. subtilis* PY79

The cultivation was carried out in the CompuGene Robotics platform. The strains s99005 to s99099 (with P_{43}) and s99010 to s99014 (with P_{ylb}) were used to inoculate the preculture plates. Two plates (GBO 655180) per media were filled with 200 µL of media and inoculated by means of toothpicks. The precultures were introduced in the automation platform and cultivated with the randomization workflow (section 4.4.3). Plate reader and incubator were set to 30 °C. The tower shaker of the incubator was set to 1000 rpm. Directly after randomization, each new plate (GBO 655101) was induced either with ddH_2O or 1.0 % (w/v) xylose (final).
Different medias were tested to cultivate the different *B. subtilis* strains (each media was supplemented with Cm and Spec):

1. MSM + 0.3 % (w/v) D-glucose
    - Composer-Script: Commit 8cefef122de3176aa27e9ebe90daa5bfc50be293
2. MSM + 1.0 % (w/v) D-fructose
    - Composer-Script: Commit f6007757f5481f7698f445c53b0c5e83514646f8

The following script versions were used for the cultivation of the shown data:

- Randomizer (R): commit dc9f2dc02618cf724975459249e196909ebc75be
- PHERAread (R): commit a48a2f43e970d182646df426a73563cc4123da39

The PHERAstar® measured OD and FI using a filter-set for GFP. The data were imported by PHERAread. The script crisprpads_bs_cultivation.R was used to evaluate and plot the data from the experiment with D-glucose as carbon source and crisprpads_bs_cultivation_fructose.R was used to evaluate and plot the data from the experiment with D-fructose as carbon source. Original data and the R scripts are available for review in the git repository of the thesis.

A.3.9. Additional figures: CRISPRpad cultivation

CRISPRpads in *E. coli* BW25113

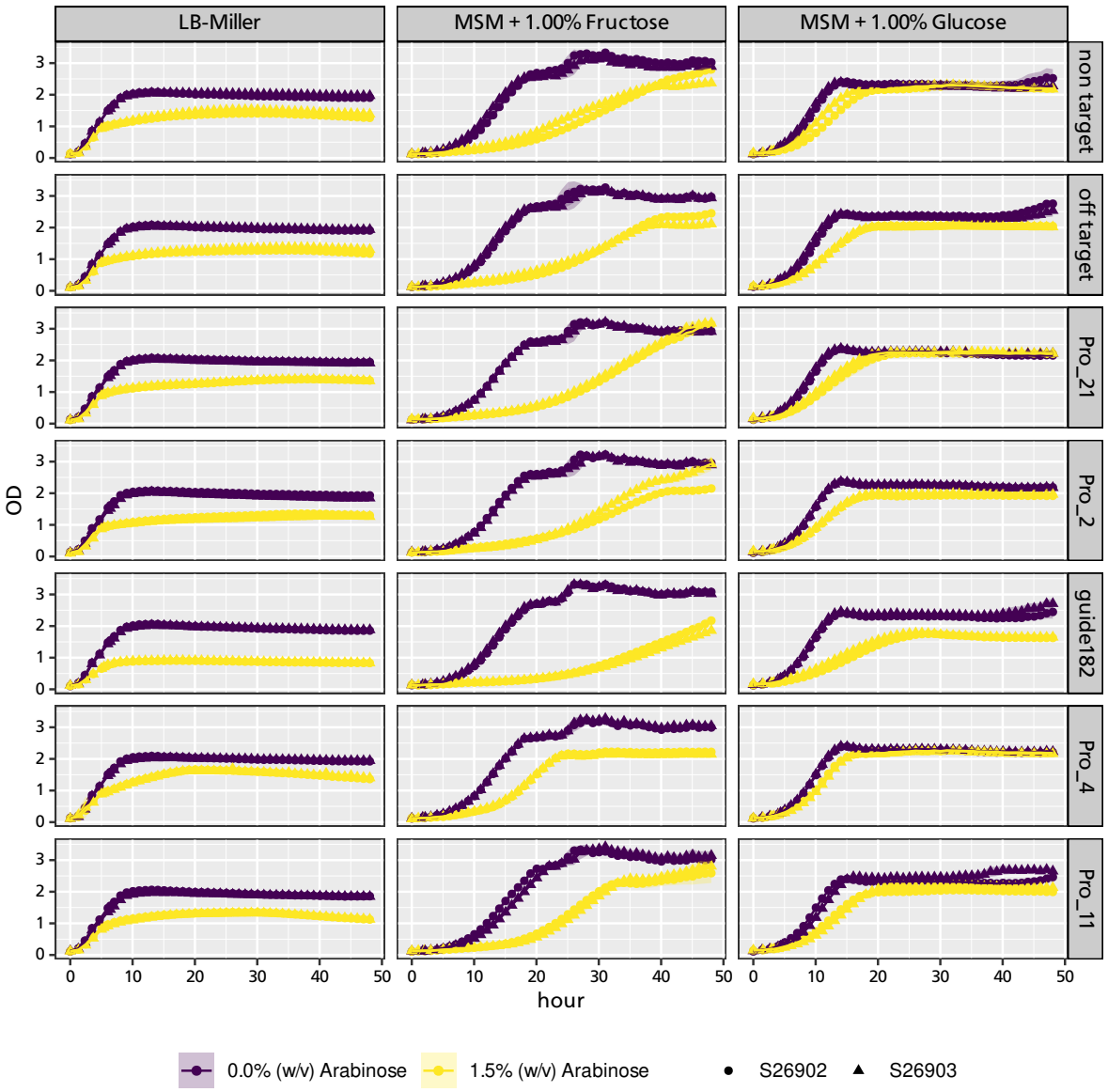


Figure A.12.: Growth of *E. coli* BW25113 strains in different media. s26902 (•) was the control without sfGFP integration. s26903(△) was the strain with integrated sfGFP reporter gene. Both contained the CRISPRa machinery, which consisted of the dCas9 plasmid (p99013) and the sgRNA delivery plasmids, targeting one of the protospacer in the CPad which are indicated on the right of the panel. The color represents the arabinose concentration used for induction. The ribbon indicates the standard deviation of n = 6 samples.

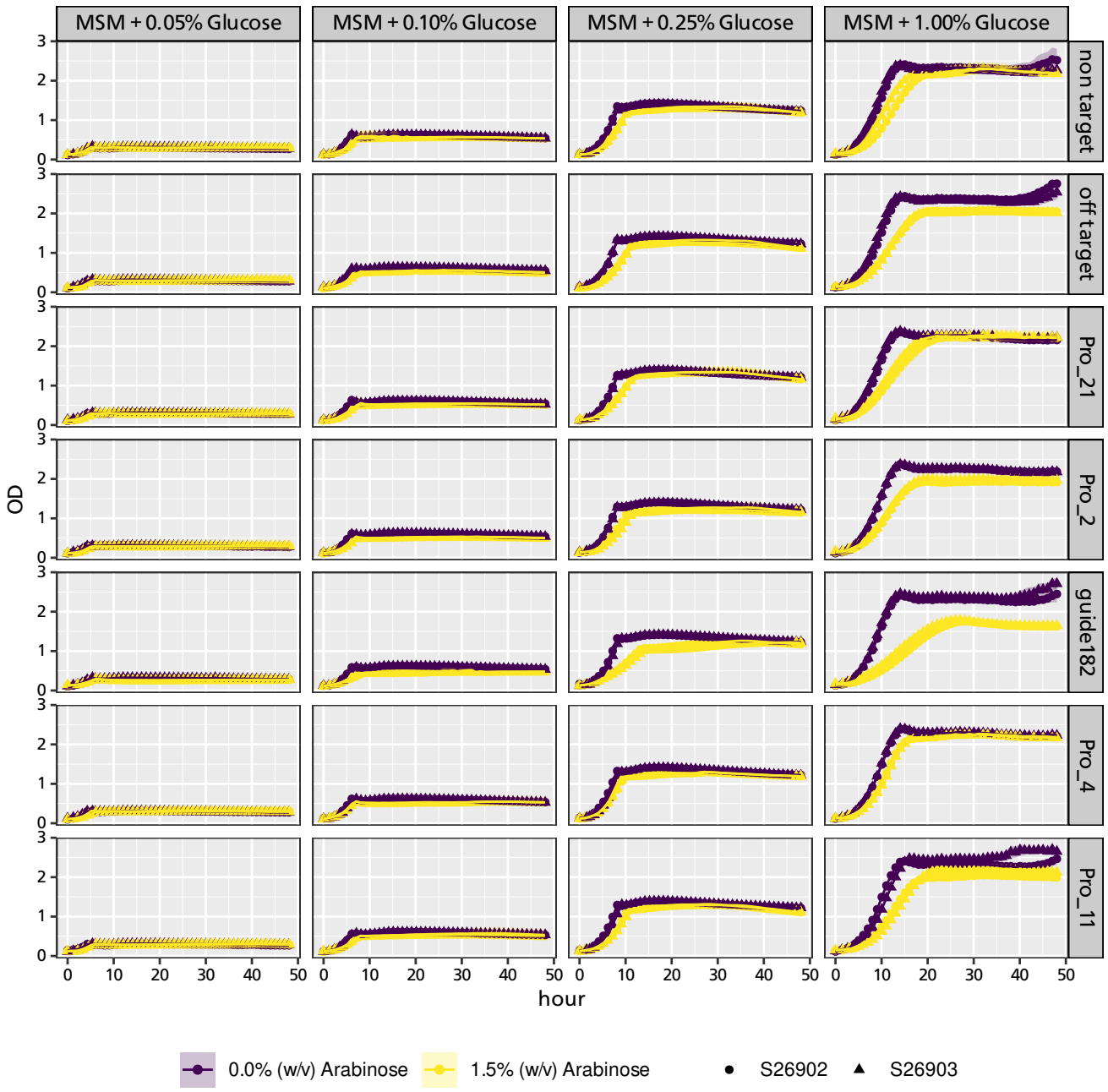


Figure A.13.: Growth of *E. coli* BW25113 strains in MSM media with different glucose concentrations. s26902 (•) was the negative control without sfGFP integration. s26903(△) was the strain with integrated sfGFP reporter gene. Both contained the CRISPRa machinery, which consisted of the dCas9 plasmid (p99013) and the sgRNA delivery plasmids, targeting one of the protospacer in the CPad which are indicated on the right of the panel. The color represents the arabinose concentration used for induction. The ribbon indicates the standard deviation of n = 6 samples.

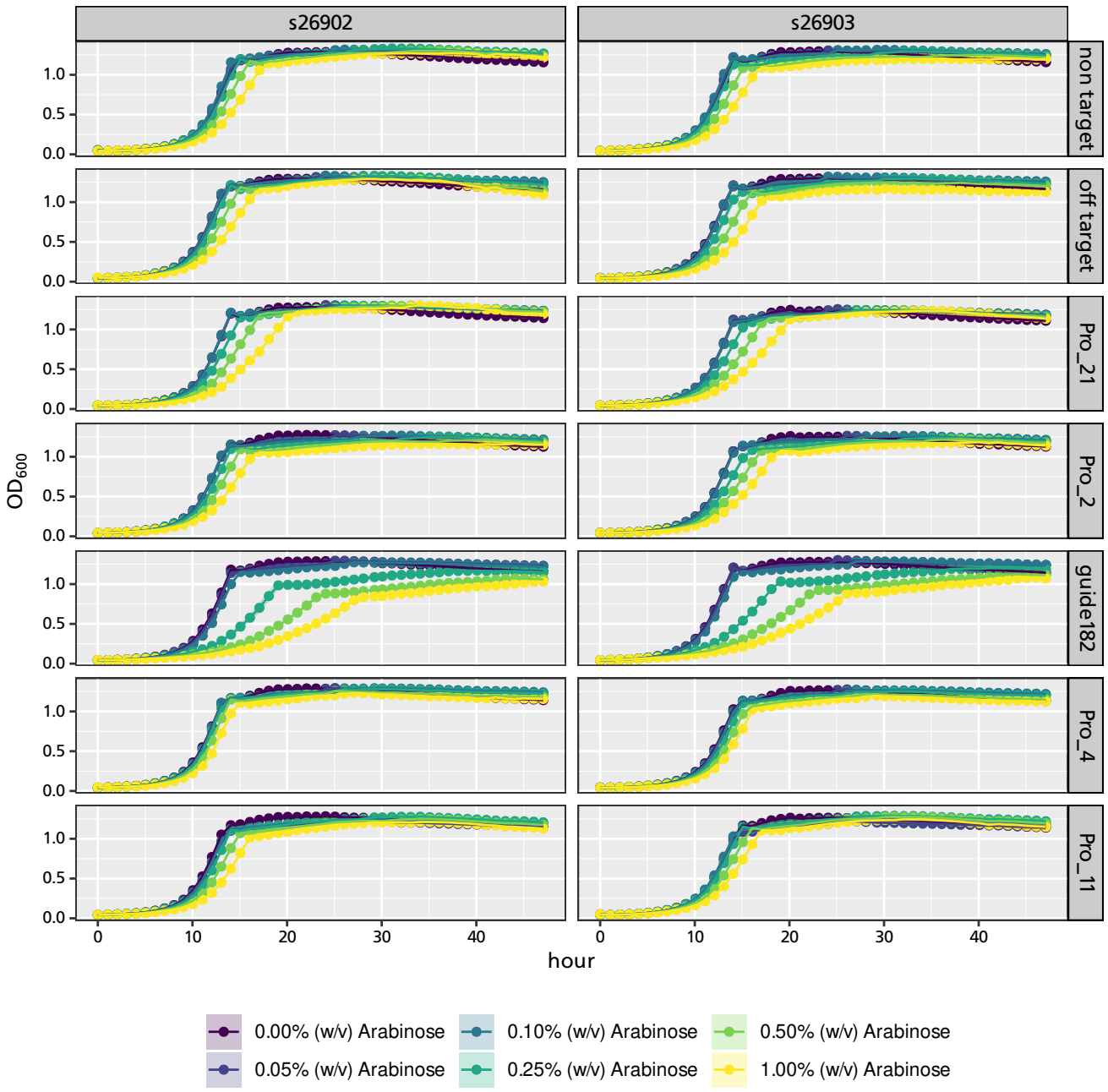


Figure A.14.: Growth of *E. coli* BW25113 strains in MSM media with 0.25 % (w/v) glucose and different inducer concentrations ranging from 0.00 % to 1.00 % (w/v). s26902 was the negative control without sfGFP integration. s26903 was the strain with integrated sfGFP reporter gene. Both contained the CRISPRa machinery, which consisted of the dCas9 plasmid (p99013) and the sgRNA delivery plasmids, targeting one of the protospacer in the CPad which are indicated on the right of the panel. The color represents the arabinose concentration used for induction. The ribbon indicates the standard deviation of n = 4 samples.

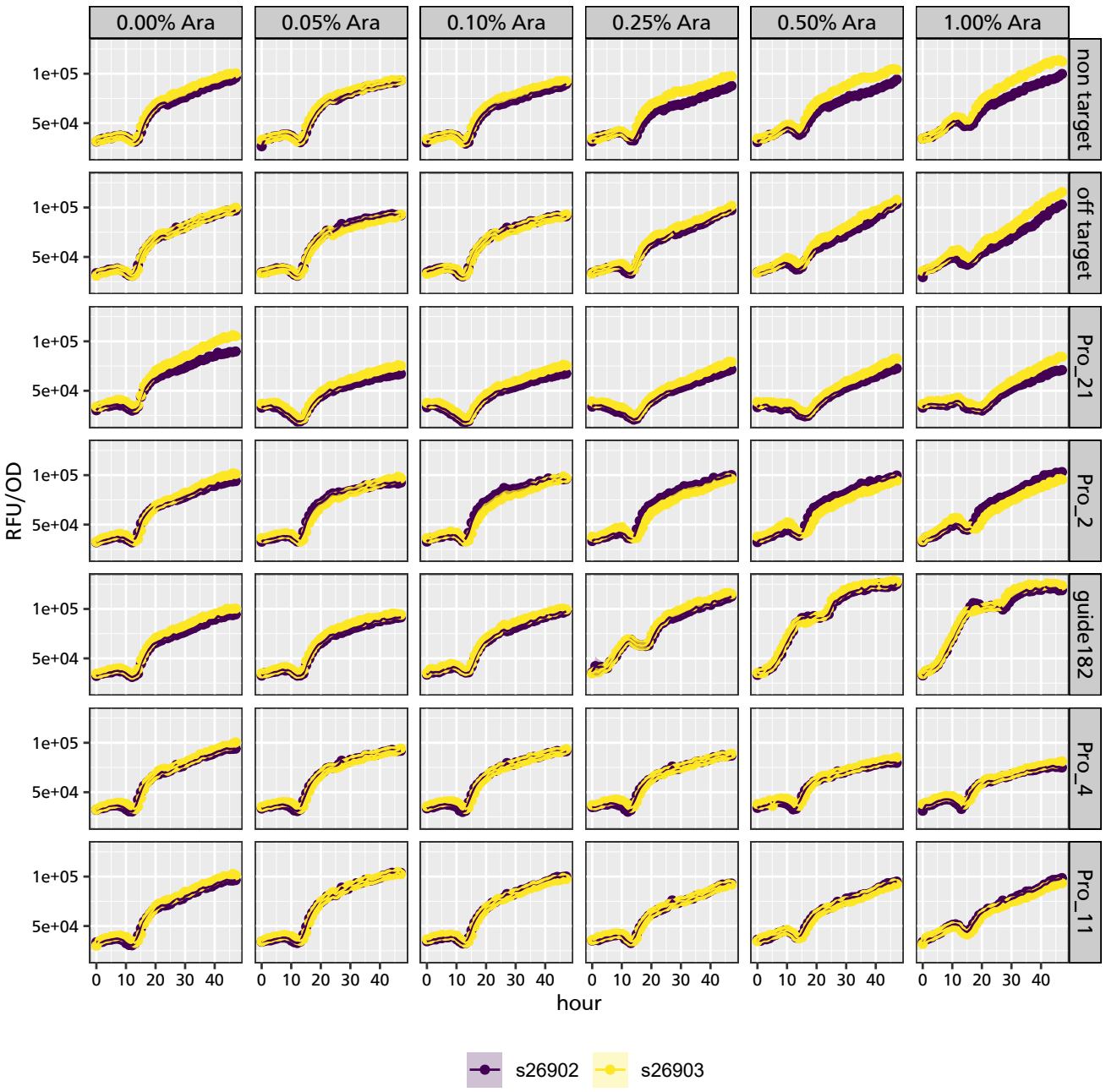


Figure A.15.: Normalized FI of *E. coli* BW25113 strains in MSM media with 0.25 % (w/v) glucose and different inducer concentrations ranging from 0.00 % to 1.00 % (w/v) – indicated as column header. s26902 (violet) was the negative control without sfGFP integration. s26903 (yellow) was the strain with integrated sfGFP reporter gene. Both contained the CRISPRa machinery, which consisted of the dCas9 plasmid (p99013) and the sgRNA delivery plasmids, targeting one of the protospacer in the CPad which are indicated on the right of the panel. The ribbon indicates the standard deviation of n = 4 samples.

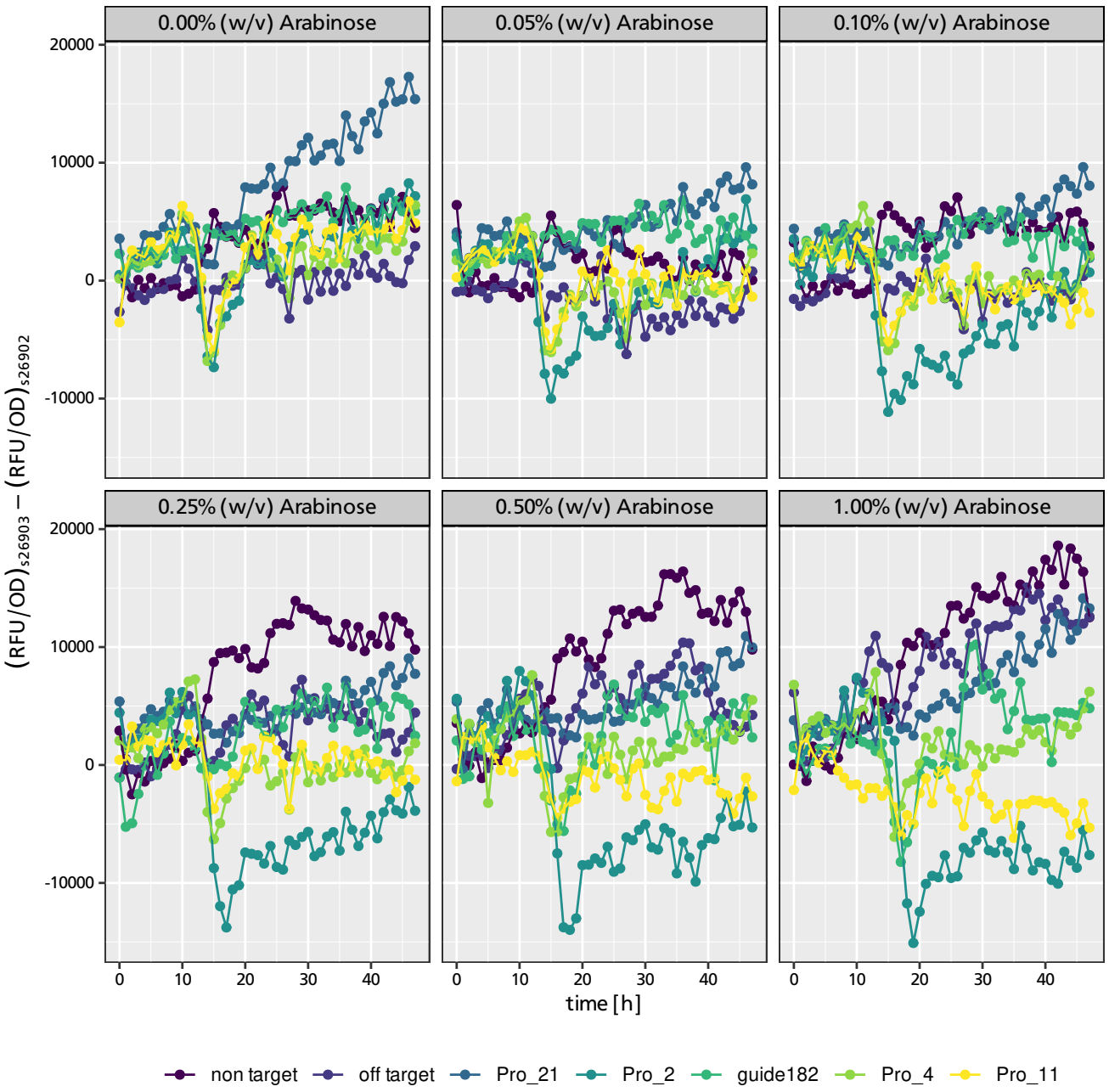


Figure A.16.: Subtracted s26902 from s26903 values in *E. coli* BW25113 strains in MSM media with 0.25 % (w/v) glucose and different inducer concentrations ranging from 0.00 % to 1.00 % (w/v). s26902 was negative control without sfGFP integration. s26903 was the strain with integrated sfGFP reporter gene. Both contain the CRISPRa machinery, which consisted of the dCas9 plasmid (p99013) and the sgRNA delivery plasmids, targeting one of the protospacer in the CPad which are indicated as different colors. The mean of the replicates (n = 4) was calculated and the mean of s26902 was subtracted from s26903.

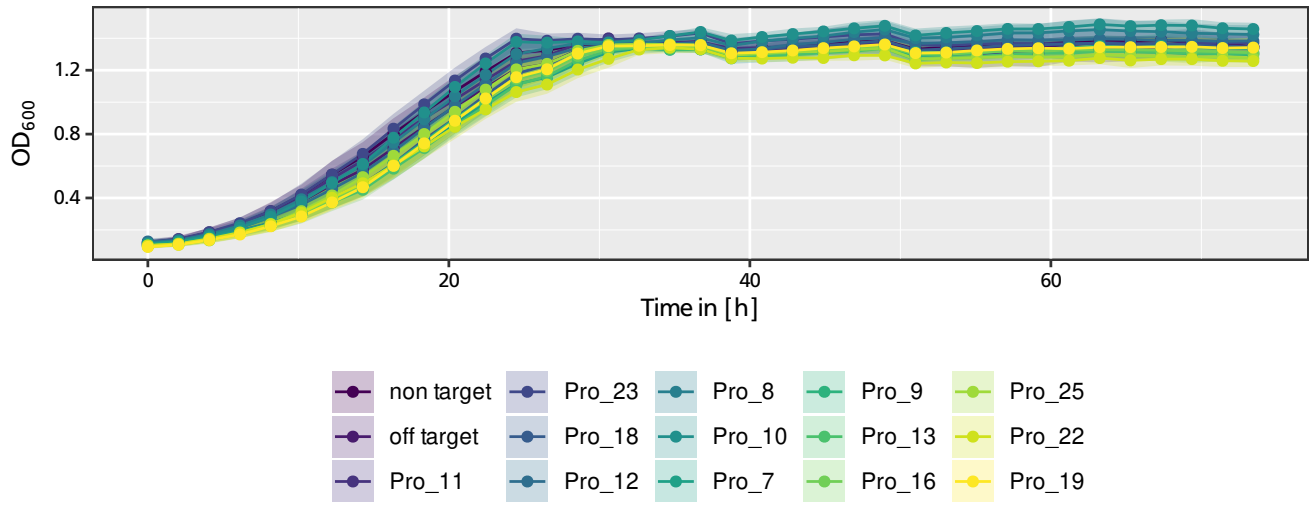


Figure A.17.: Growth of *S. cerevisiae* s26458 with different sgRNAs. The ribbon indicates the standard deviation of n = 6 samples. Control strain with off-target guide was *S. cerevisiae* s26458 with plasmid p26129 and control with non-targeting was *S. cerevisiae* s26458 with plasmid guide p26139.

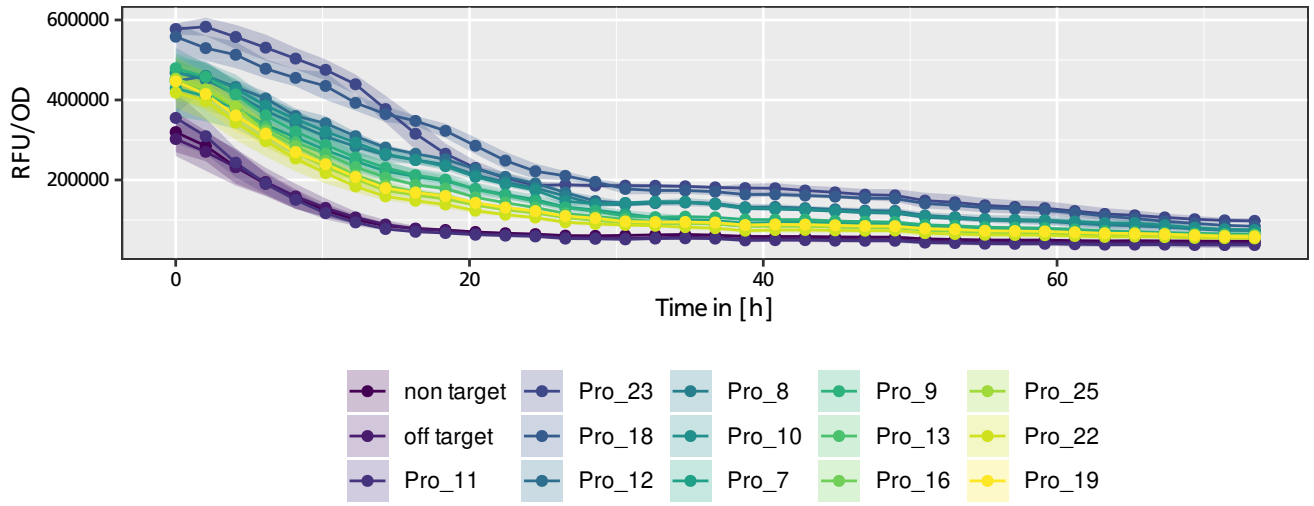


Figure A.18.: RFU normalized to OD of *S. cerevisiae* s26458 with different sgRNAs to activate expression of the target reporter gene. The ribbon indicates the standard deviation of n = 6 samples. Control strain with off-target guide was *S. cerevisiae* s26458 with plasmid p26129 and control with non-targeting was *S. cerevisiae* s26458 with plasmid guide p26139.

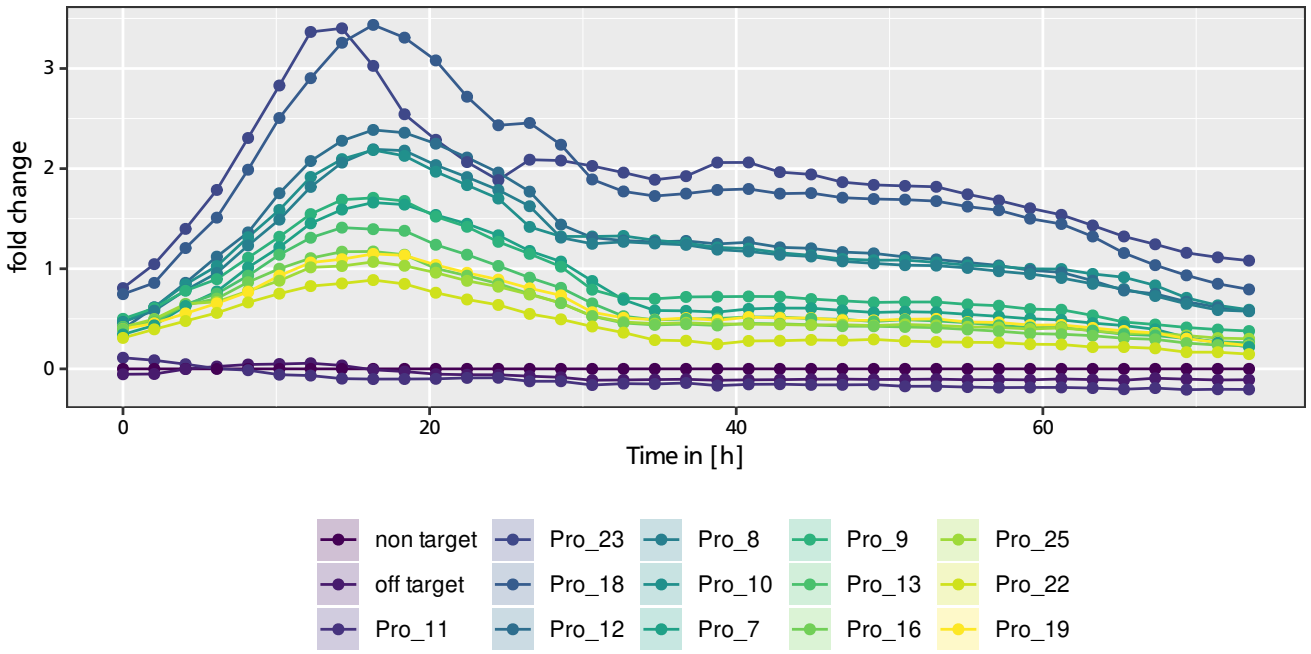


Figure A.19.: Fold change of activation over the time. Fold change was calculated as $foldchange = (\overline{FInorm_i} - \overline{FInorm_c})/\overline{FInorm_c}$ (i - strain with specific guide RNA, c - control with non-target guide RNA). Control strain with non-target guide was *S. cerevisiae* s26458 with plasmid p26139. Mean was calculated from n = 6 samples.

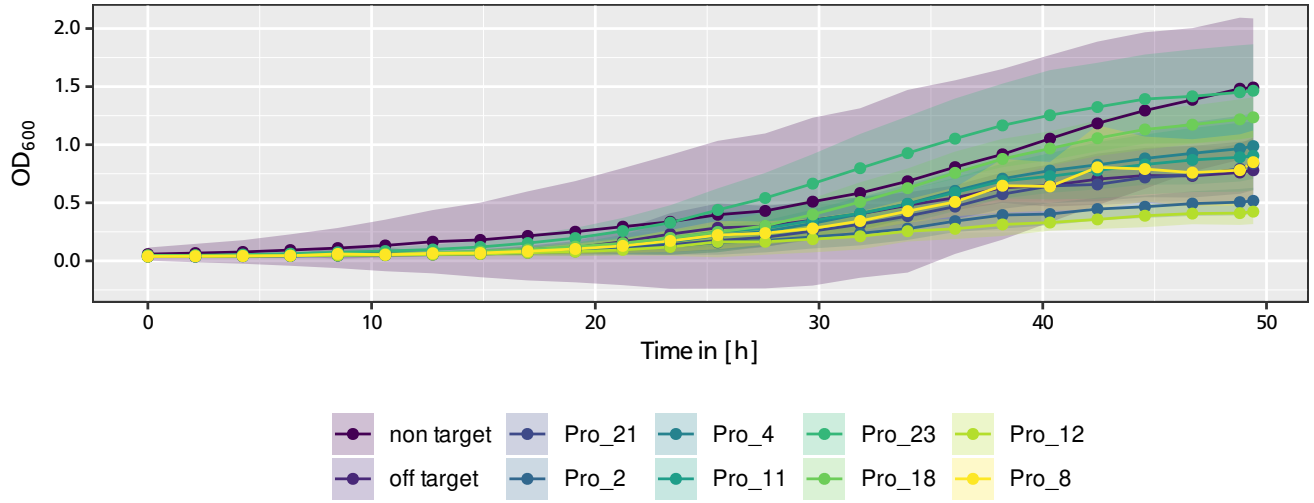


Figure A.20.: Growth of *Y. lipolytica* s26241 with different sgRNAs. The ribbon indicates the standard deviation of n = 8 samples. Control strain with non-target guide was *Y. lipolytica* s26241 with plasmid p26120 and control with off-targeting guide was *Y. lipolytica* s26241 with plasmid p26173.

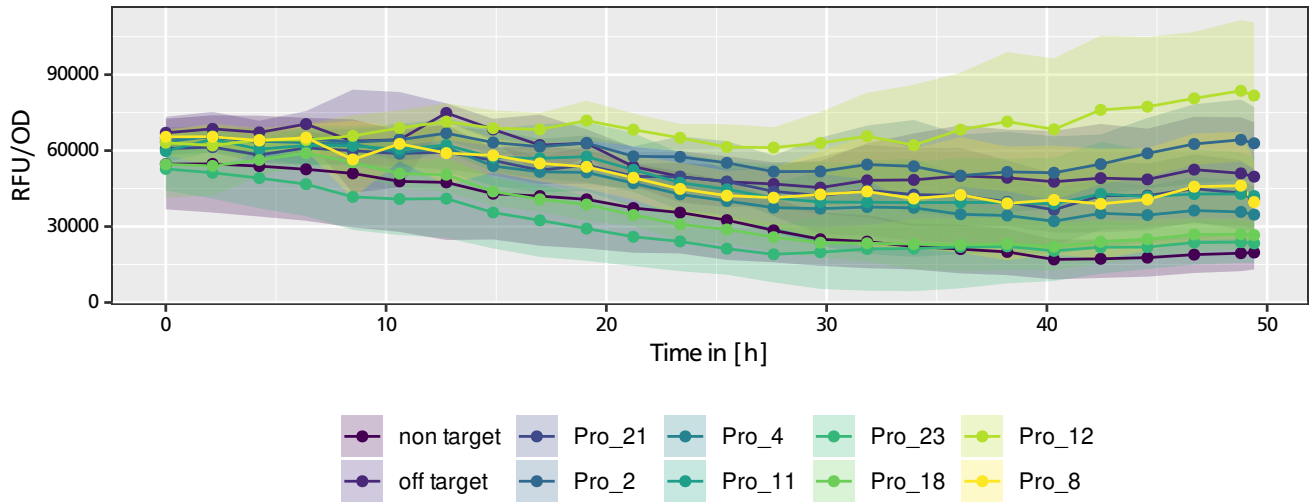


Figure A.21.: RFU normalized to OD of *Y. lipolytica* s26241 with different sgRNAs to activate expression of the target reporter gene. The ribbon indicates the standard deviation of n = 8 samples. Control strain with non-target guide was *Y. lipolytica* s26241 with plasmid p26120 and control with off-targeting guide was *Y. lipolytica* s26241 with plasmid p26173.

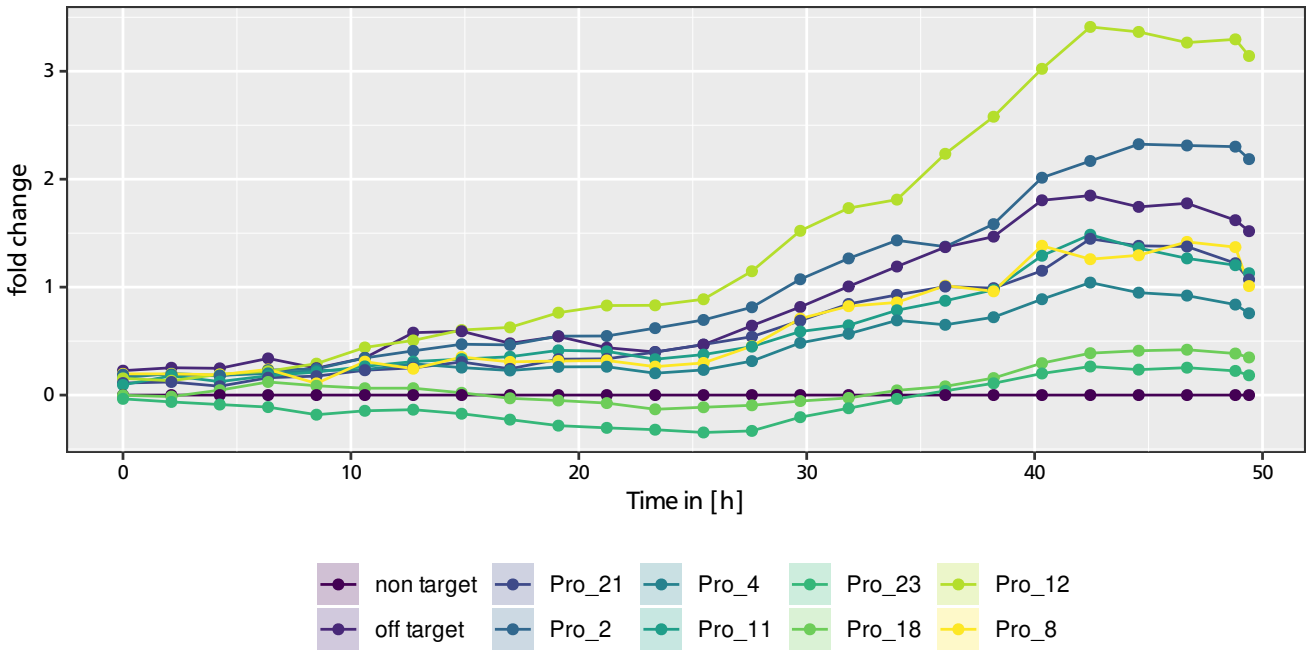


Figure A.22.: Fold change of activation over the time. Fold change was calculated as $foldchange = (\overline{FInorm_i} - \overline{FInorm_c})/\overline{FInorm_c}$ (i - strain with specific guide RNA, c - control with non-target guide RNA). Control strain with non-target guide was *Y. lipolytica* s26241 with plasmid p26120. Mean was calculated from n = 8 samples.

CRISPRpads in *B. subtilis* PY79

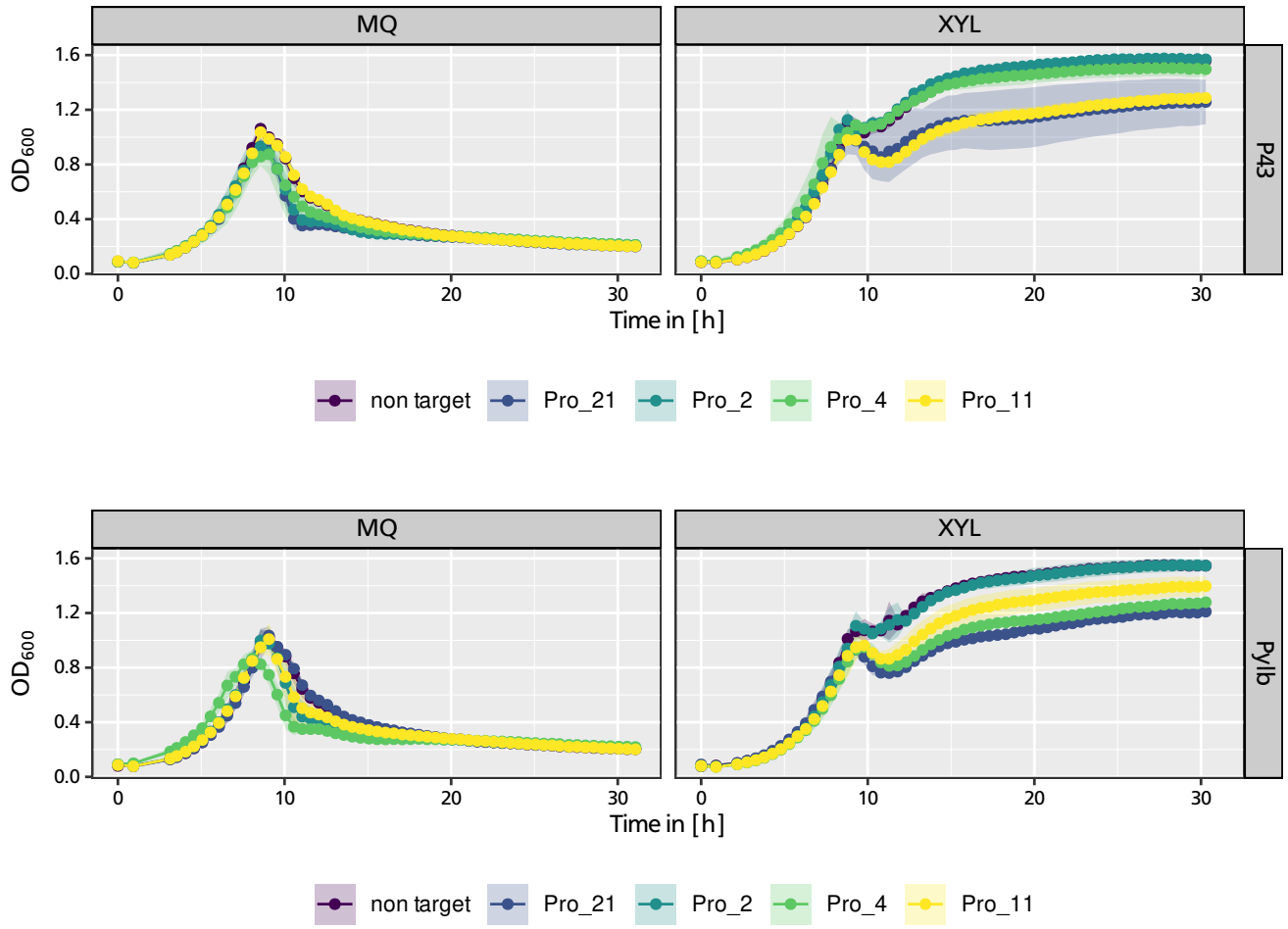


Figure A.23.: Growth of *B. subtilis* PY79 with CPad, sfGFP and sgRNAs (indicated by different colors) which were constructed by Georg Schmidt. Strains were cultivated in the robotics platform in defined minimal media (MSM) containing 0.3 % D-glucose as carbon source. Inducer: MQ - ddH_2O, XYL - 1 % D-xylose. The ribbon indicates the standard deviation of n = 8 samples. Data derived from the Bachelor thesis of Georg Schmidt and plotted again for this thesis.

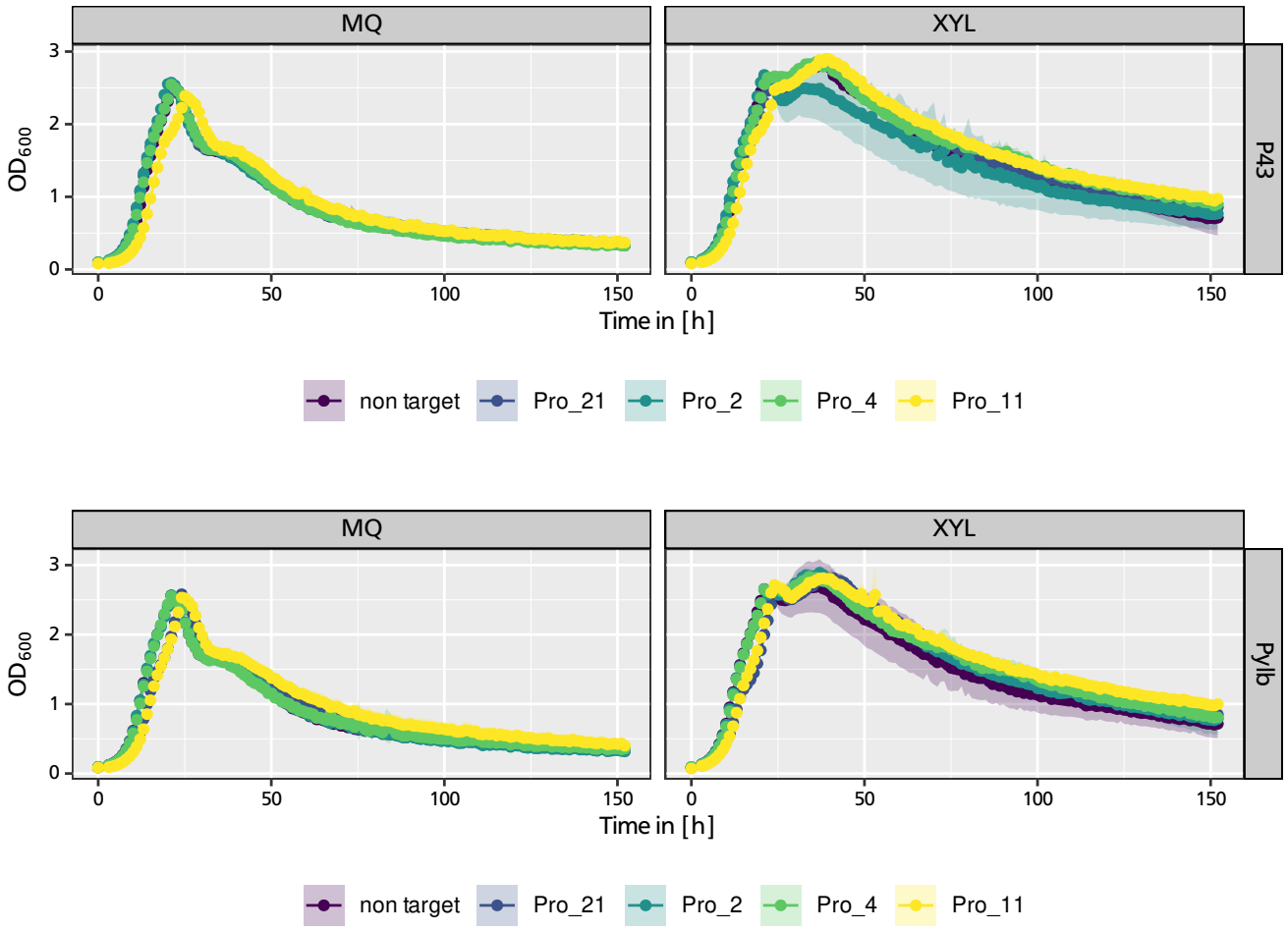


Figure A.24.: Growth of *B. subtilis* PY79 with CPad, sfGFP and sgRNAs (indicated by different colors) which were constructed by Georg Schmidt. Strains were cultivated in the robotics platform in defined minimal media (MSM) containing 1 % D-fructose as carbon source. Inducer: MQ - ddH_2O, XYL - 1 % D-xylose. The ribbon indicates the standard deviation of n = 8 samples.

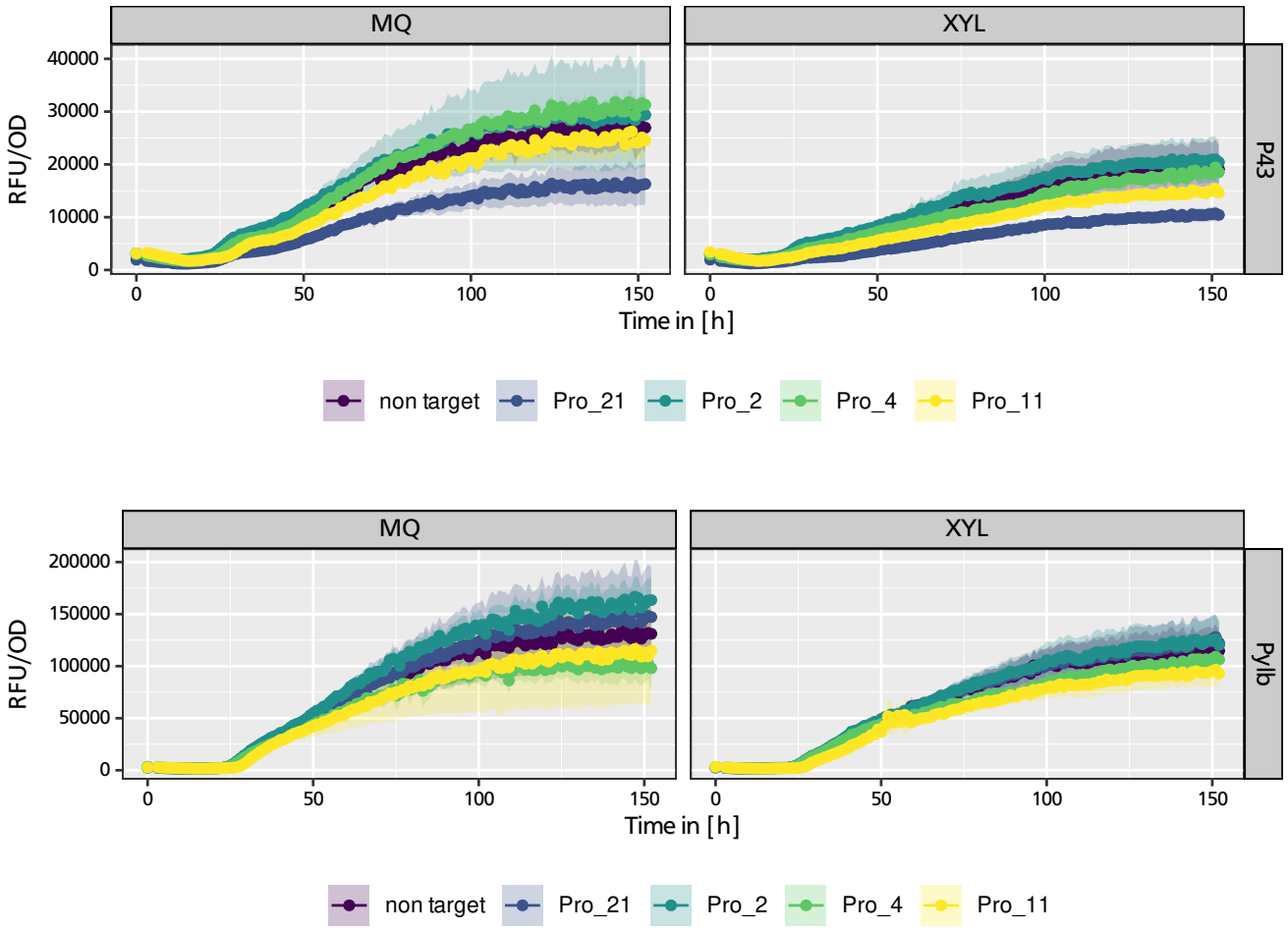


Figure A.25.: Normalized FI of *B. subtilis* PY79 with CPad, sfGFP and sgRNAs (indicated by different colors) which were constructed by Georg Schmidt. Strains were cultivated in the robotics platform in defined minimal media (MSM) containing 1 % D-fructose as carbon source. Inducer: MQ - ddH_2O, XYL - 1 % D-xylose. The ribbon indicates the standard deviation of n = 8 samples.

A.4. Automation: Additional figures

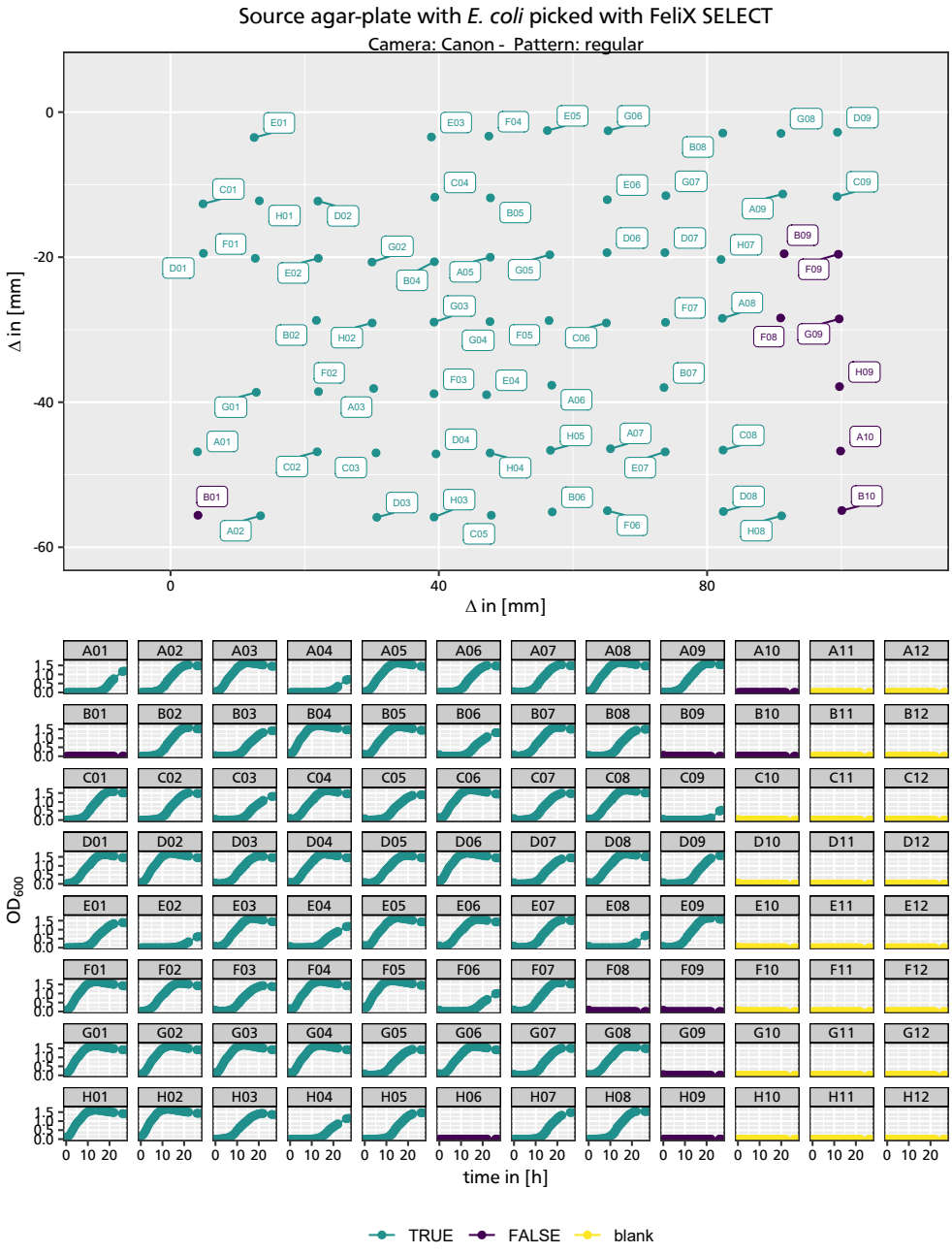


Figure A.26.: Colony Picking using 3DpickO OpenCV. Different cameras and patterns were tested and picked in LB Medium. Camera: Canon, Pattern: 96 Well. Picking success was classified as growth (TRUE) and no growth (FALSE). Wells that did not receive a sample were classified as blank. Growth was determined by the mean OD. Mean OD per well > 0.05 was classified as growth.

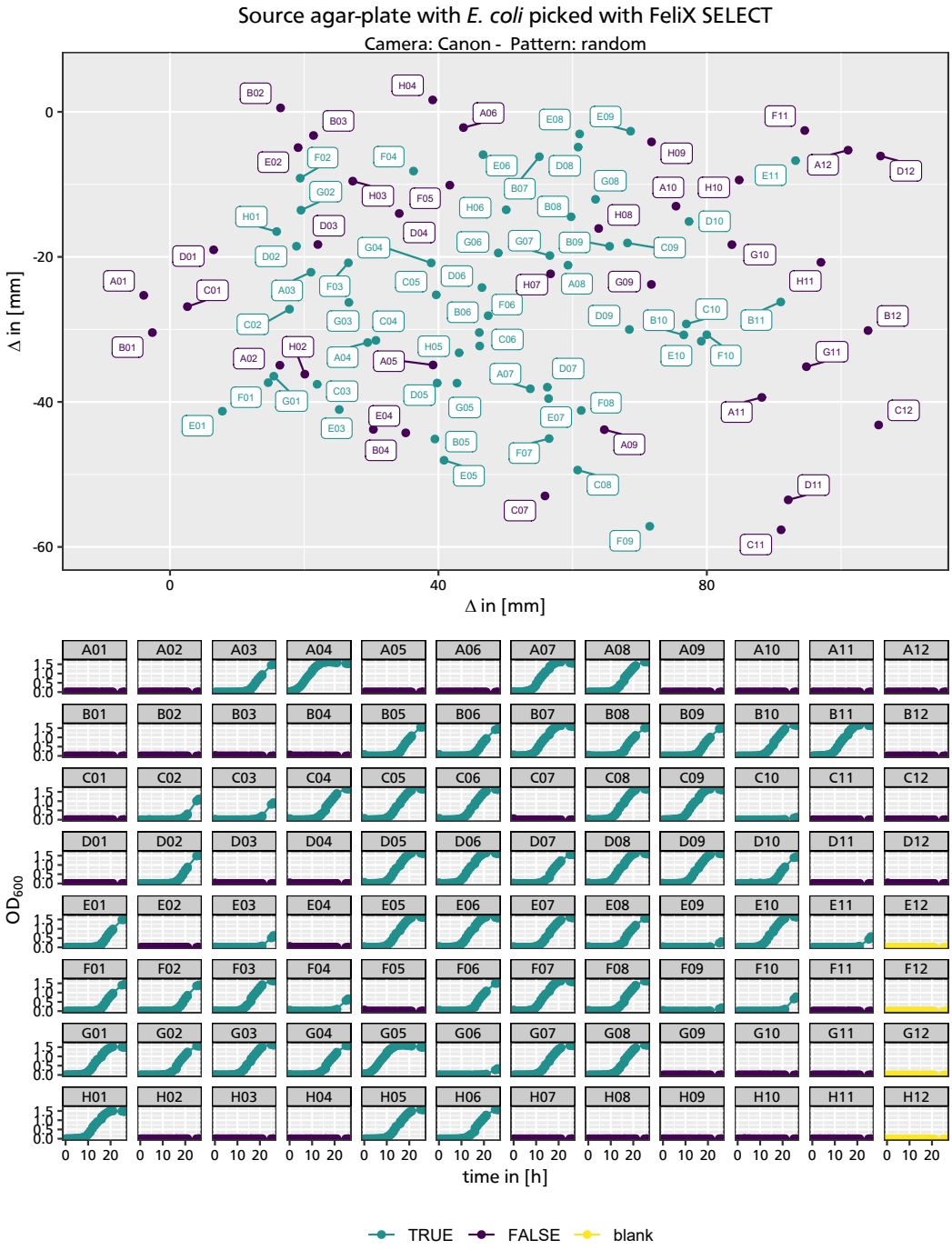


Figure A.27.: Colony Picking using 3DpickO OpenCV. Different cameras and patterns were tested and picked in LB Medium. Camera: Canon, Pattern: Random. Picking success was classified as growth (TRUE) and no growth (FALSE). Wells that did not receive a sample were classified as blank. Growth was determined by the mean OD. Mean OD per well > 0.05 was classified as growth.

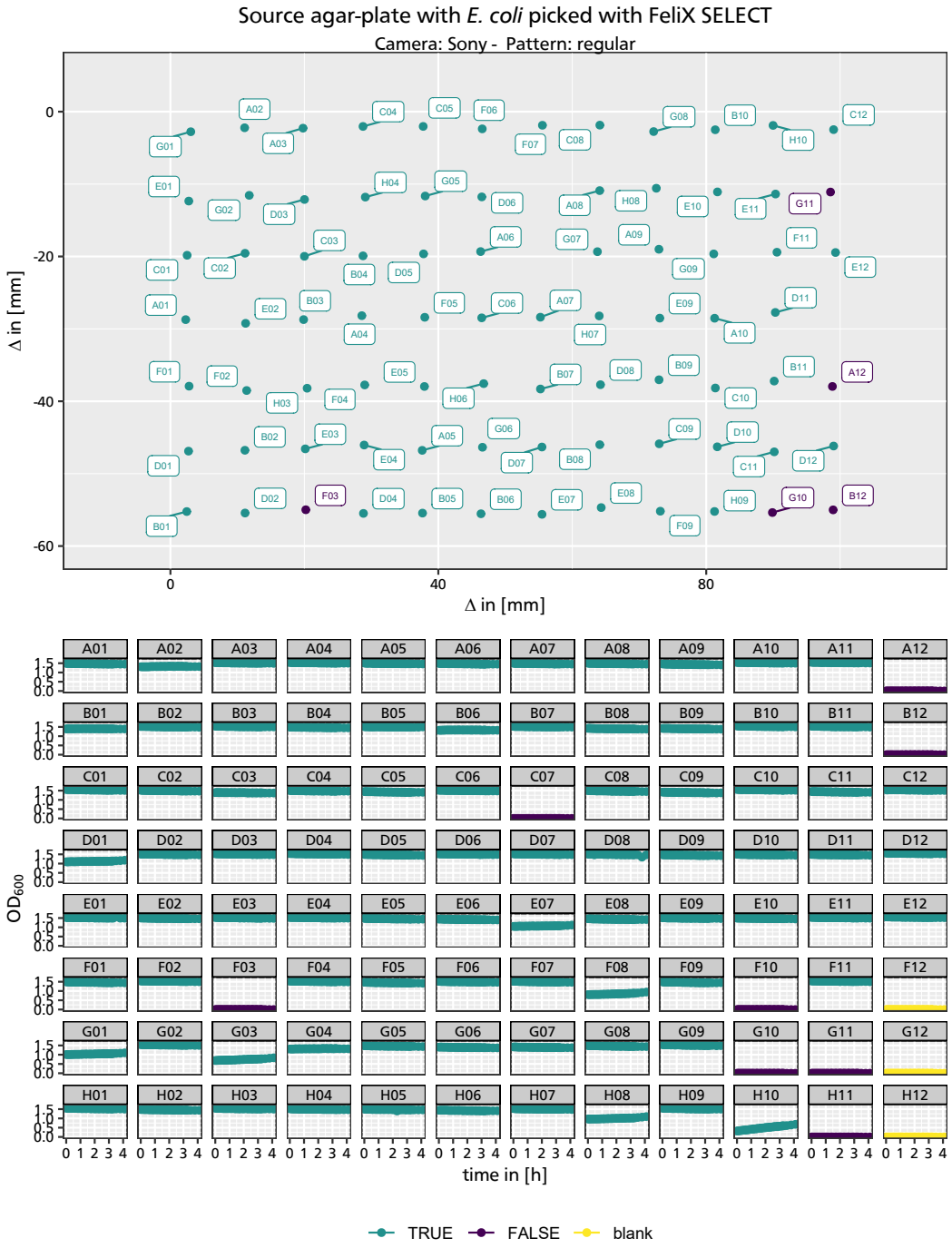


Figure A.28.: Colony Picking using 3DpickO OpenCV. Different cameras and patterns were tested and picked in LB Medium. Camera: Sony, Pattern: 96 Well. Picking success was classified as growth (TRUE) and no growth (FALSE). Wells that did not receive a sample were classified as blank. Growth was determined by the mean OD. Mean OD per well > 0.05 was classified as growth.

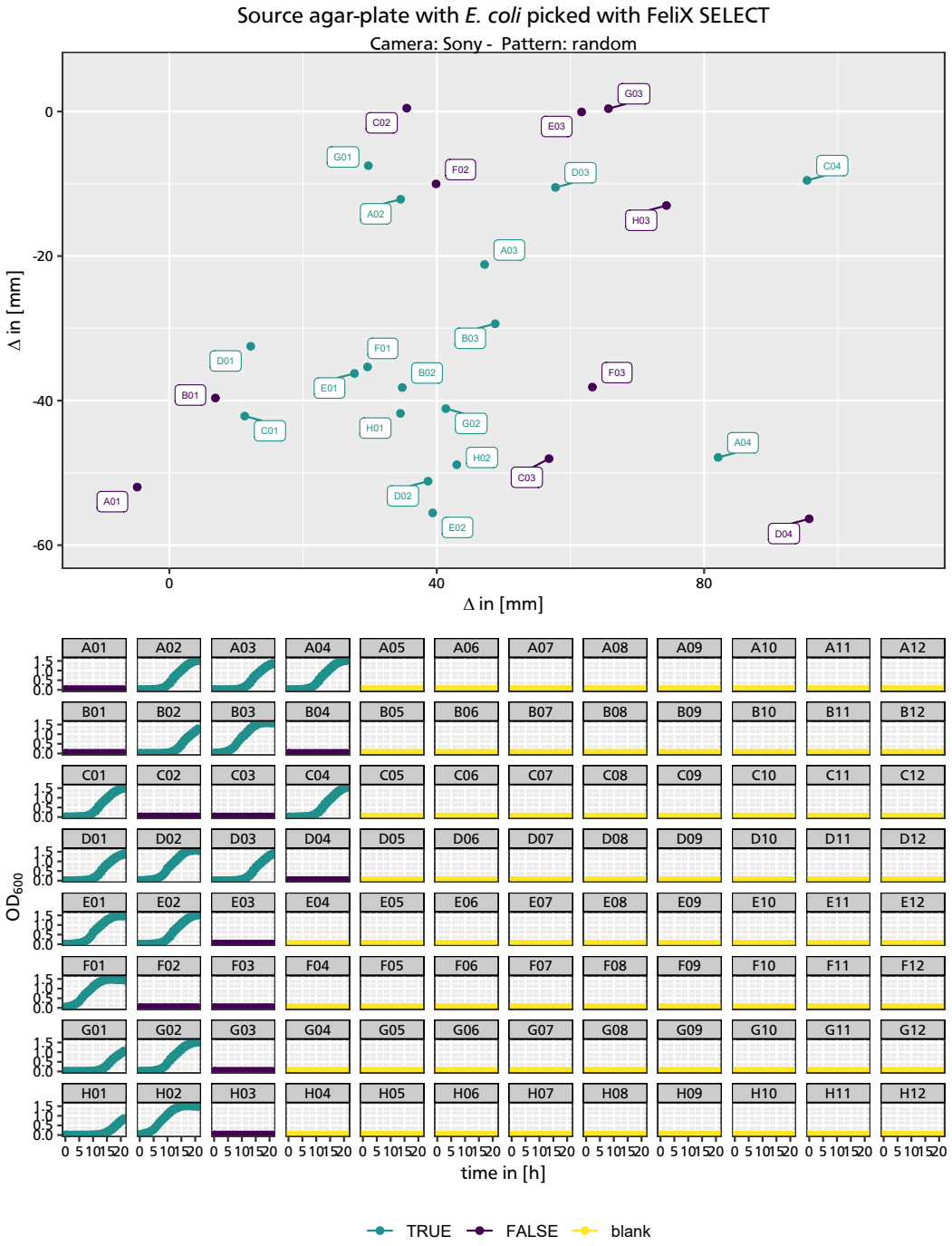


Figure A.29.: Colony Picking using 3DpickO OpenCV. Different cameras and patterns were tested and picked in LB Medium. Camera: Sony, Pattern: Random. Picking success was classified as growth (TRUE) and no growth (FALSE). Wells that did not receive a sample were classified as blank. Growth was determined by the mean OD. Mean OD per well > 0.05 was classified as growth.

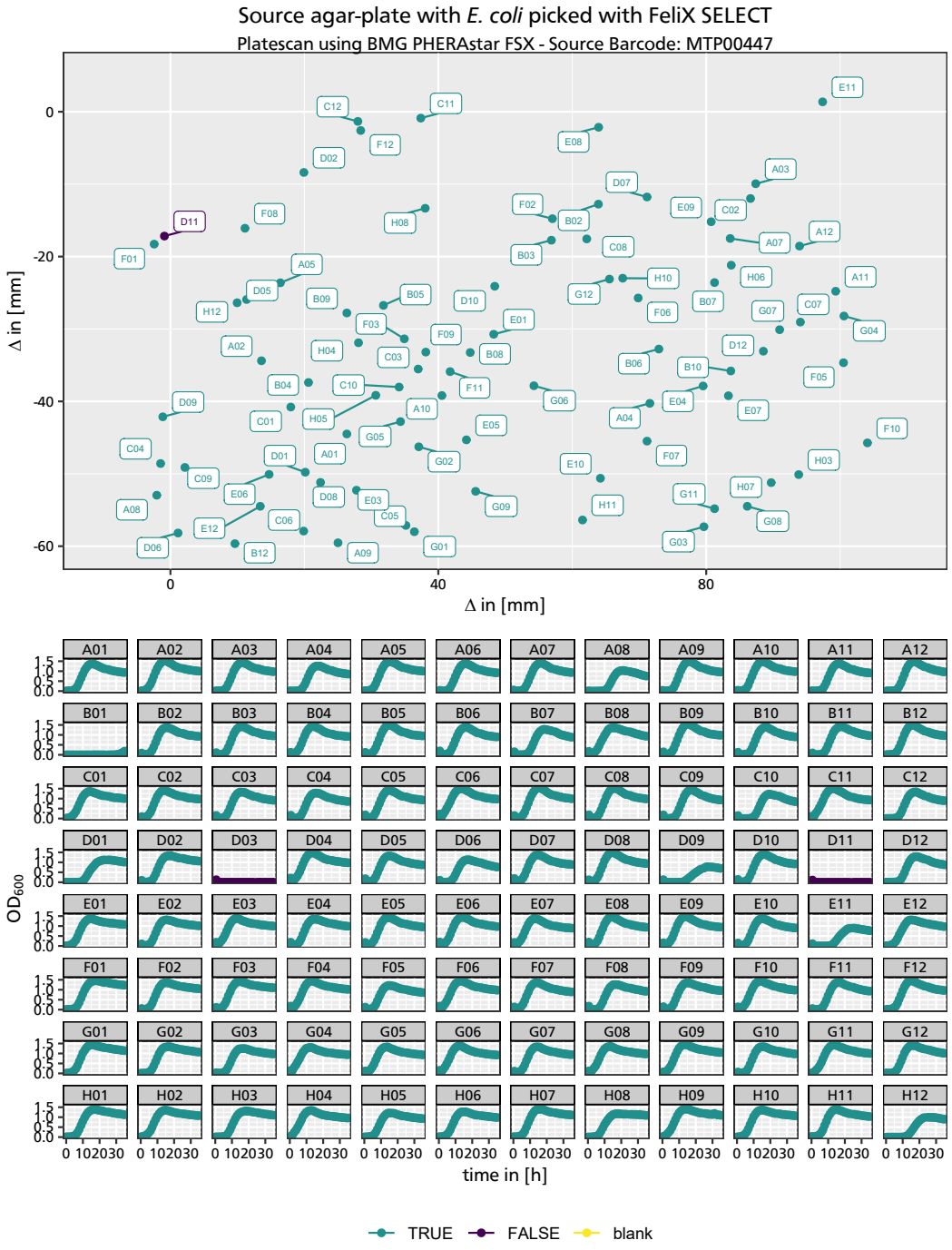


Figure A.30.: Plates were measured in the PHERAstar FSX and analyzed by Platescan. Example with multiple target plates. Target: 1/2. Picking success was classified as growth (TRUE) and no growth (FALSE). Wells that did not receive a sample were classified as blank. Growth was determined by the mean OD. Mean OD per well > 0.05 was classified as growth.

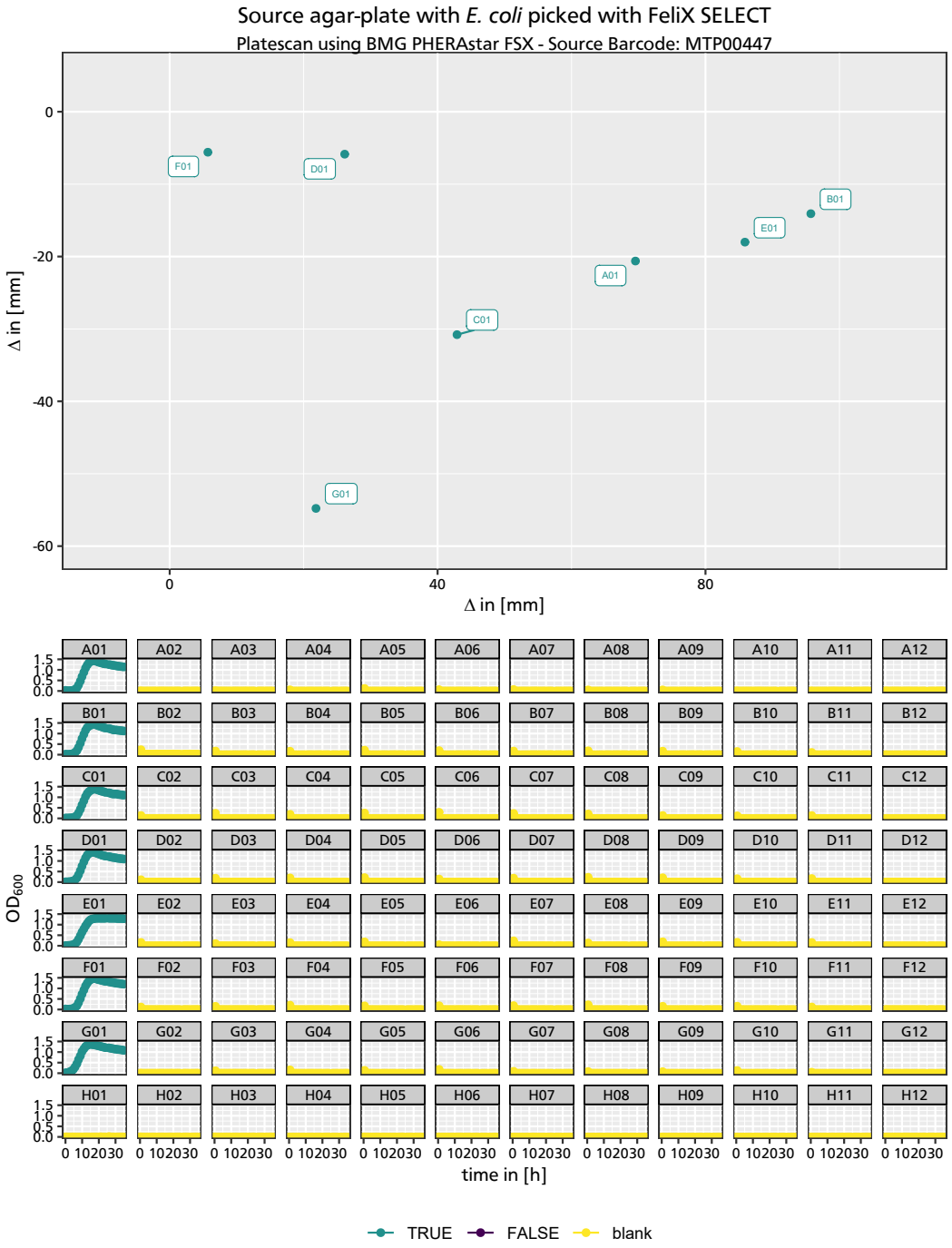


Figure A.31.: Plates were measured in the PHERAstar FSX and analyzed by Platescan. Example with multiple target plates. Target: 2/2. Picking success was classified as growth (TRUE) and no growth (FALSE). Wells that did not receive a sample were classified as blank. Growth was determined by the mean OD. Mean OD per well > 0.05 was classified as growth.

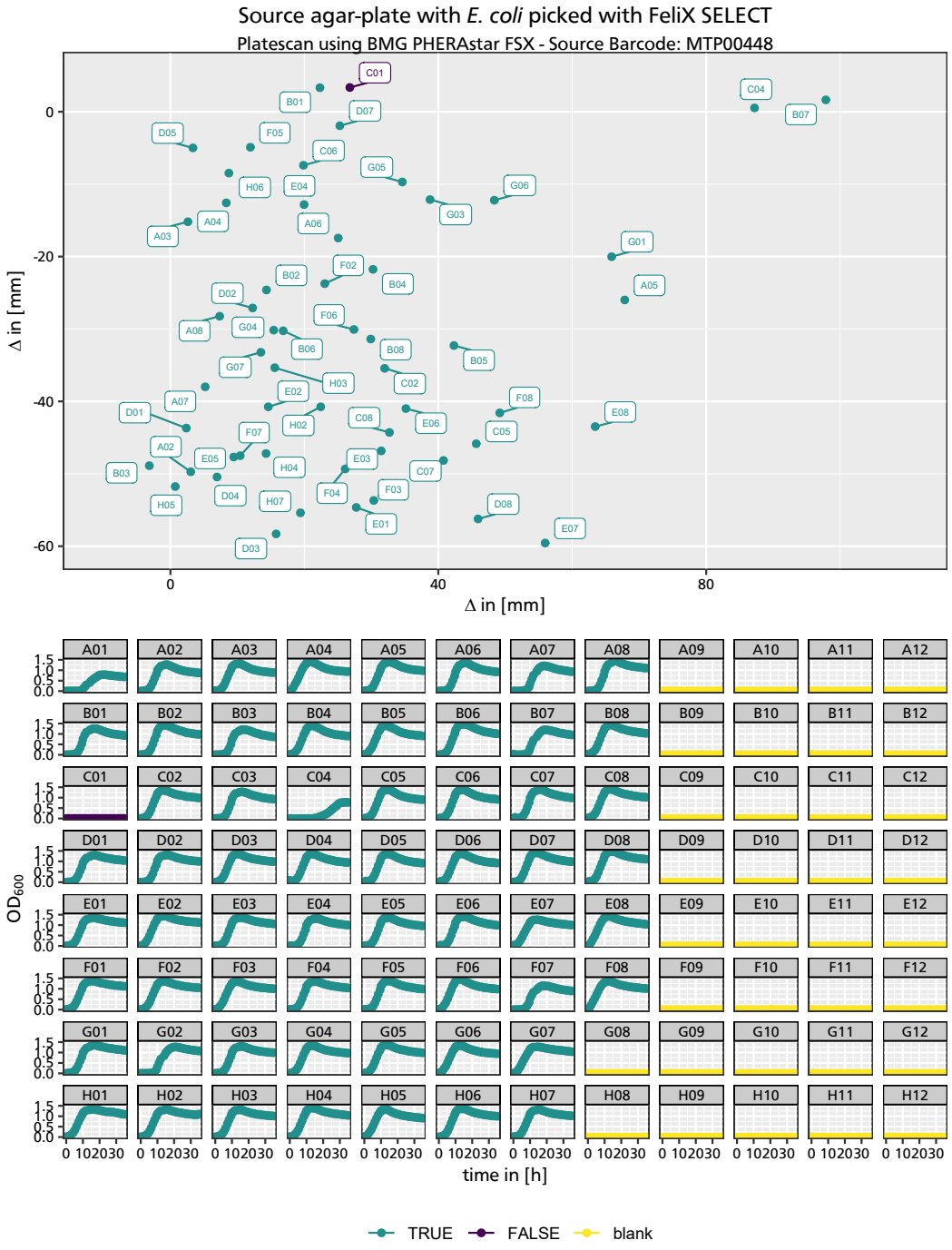


Figure A.32.: Plates were measured in the PHERAstar FSX and analyzed by Platescan. Example with one target plate. Picking success was classified as growth (TRUE) and no growth (FALSE). Wells that did not receive a sample were classified as blank. Growth was determined by the mean OD. Mean OD per well > 0.05 was classified as growth.