New Methods of DNA Assembly, Gene Regulation with a Synthetic sRNA, and Cyanobacterium Phenotype Monitoring with Raman Spectroscopy

Imen Tanniche

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Biological Systems Engineering

Ryan S. Senger, Chair

Eva Collakova
Chenming (Mike) Zhang
Justin R. Barone

April 26, 2019
Blacksburg, Virginia

Keywords: Lambda-PCR, DNA assembly, synthetic small RNA, cell-free-system, Raman spectroscopy, Rametrix^{TM}, phenotypic response, external stimuli, metagenomic library

New Methods of DNA Assembly, Gene Regulation with a Synthetic sRNA, and Cyanobacterium Phenotype Monitoring with Raman Spectroscopy

Imen Tanniche

Abstract

Metabolic engineering has enabled studying microorganisms by the modification of their genetic material and analysis of their metabolism for the isolation of microbial strains capable of producing high yields of high value chemicals and biofuels. In this research, novel tools were developed to improve genetic engineering of microbial cells. In this matter, λ-PCR (lambda-PCR) was developed enabling the construction of plasmid DNA. This technique allows DNA assembly and manipulation (insertion, substitution and/or deletion) at any location of a vector. λ-PCR addresses the need for an easy, highly-efficient, rapid and inexpensive tool for genetic engineering and overcoming limitations encountered with traditional techniques. Then, novel synthetic small RNA (sRNA) regulators were designed in a cell-free-system (*in vitro*) in order to modulate protein expression in biosynthetic pathways. The ability of the sRNAs to regulate mRNA expression with statistical significance was demonstrated. Up to 70% decrease in protein expression level was achieved by targeting specific secondary structures of the mRNA with antisense binding regions of the sRNA. Most importantly, a sRNA was identified capable of protein overexpression by up to 65%. An understanding of its mechanism showed that its mRNA target region(s) likely lead to occlusion of RNase E binding. This mechanism was translated for expression of a diaphorase enzyme, which has relevance to synthetic biology and metabolic engineering in *in vitro* systems.

Results were successful, showing a greater than 75% increase in diaphorase expression in a cell-free protein synthesis reaction.

Next, Raman spectroscopy was employed as a near real-time method for microbial phenotyping. Here, Raman spectroscopy was used in combination with chemometric analysis methods through Rametrix™ Toolboxes to study the effects of environmental conditions (i.e. illumination, glucose, nitrate deprivation, acetate, sodium chloride and magnesium sulfate) on the phenotypic response of the cyanobacterium *Synechocystis* sp. PCC6803. The Rametrix™ LITE Toolbox for MATLAB® enabled processing of Raman spectra and application of principal component analysis (PCA) and discriminant analysis of principal components (DAPC). Two studies were performed. PCA and DAPC produces distinct clustering of Raman spectra, representing multiple *Synechocystis* phenotypes, based on the (i) presence of glucose in the growth medium, (ii) illumination, (iii) nitrate limitation, and (iv) throughout a circadian rhythm growth cycle, in the first study. The second study focused on the phenotypic response based on (i) growth in presence of acetate, (ii) presence of high concentrations of sodium chloride and (iii) magnesium sulfate starvation. Rametrix™ PRO was applied for the validation of the DAPC models through leave-one-out method that allowed calculation of prediction accuracy, sensitivity and selectivity for an unkown Raman spectrum. Statistical tests (ANOVA and pairwise comparison) were performed on Raman spectra to identify statistically relevant changes in *Synechocystis* phenotypes. Next, comparison between Raman data and standardized analytical methods (GF-FID, UPLC, spectrometric assays) was established. Overall, good correlation were obtained ($R > 0.7$).

Finally, genomic DNA libraries were enriched to isolate a deoxynivalenol detoxifying enzyme. To do this, library fragments from microorganisms was generated through oligonucleotide primed polymerase chain reaction (DOP-PCR) and transformed in a DON-

sensitive yeast strain. Rounds of subculture were performed in the presence of DON and ferulic acid in order to isolate a strain capable of enzymatic degradation of DON.

New Methods of DNA Assembly, Gene Regulation with a Synthetic sRNA, and Cyanobacterium Phenotype Monitoring with Raman Spectroscopy

Imen Tanniche

General Audience Abstract

Metabolic engineering is the use of genetic engineering to modify microorganisms in order to produce high yields of valuable commodity chemicals. The goal of this research is to develop new methods to improve genetic modification and selection of microbial cells. The specific objectives were to: (i) develop new tools for DNA assembly and manipulation, (ii) utilize small synthetic RNA to control protein expression level, (iii) use Raman spectroscopy to study phenotypic responses to environmental changes and (iv) enrich for microoganisms that detoxify dangerous toxins. First, a new technique for DNA assembly, named λ-PCR (lambda-PCR), was developed. This method allows the easy manipulation of plasmid DNA with high-efficiency and low-cost compared to traditional techniques. Second, novel synthetic small RNA (sRNA) regulators were designed in a cell-free-system in order to modulate (downregulate or overexpress) fluorescent protein expression. Next, Raman spectroscopy was used to assess phenotypic response of cyanobacterial cells to different environmental modifications (light settings, salts, sugar, etc…). Finally, genomic library was used to discover and characterize enzymes capable of degrading a mycotoxin.

Dedication

*In the loving memory of my son **Adam**.*

Dedicated to

My husband Yassine Mahjoubi

My parents Zohra and Zineddine

My siblings Hanen, Sondes and Ali

My mother and father in law, Saida and Ahmed

My sister and brother in law, Chaima and Aymen

My family and friends

Acknowledgments

I would like to express my sincere gratitude to all those who contributed to this dissertation. My sincerest appreciation is extended to my advisor **Dr. Ryan Senge**r for his continuous support and guidance during my research work. His valuable advices and feedback shaped this project into an achievement. Beyond that, his mentorship on professional and personal aspects added considerably to my Ph.D. journey. I would like to express my gratitude for his help and understanding during the toughest moments.

I was honored to work with my committee member **Dr. Eva Collakova** who dedicated hours helping and training me for the use of Gas chromatography and UPLC in her laboratory. I am deeply thankful for her patience, guidance and thoughtful insights throughout this Ph.D.

I would like also to express my regards to my other committee members, **Dr. Chenming (Mike) Zhang**, and **Dr. Justin R. Barone**. I am grateful for their guidance and useful feedback about my research.

Special thanks to my research group in the Senger lab, Dr. Hadi Nazeem-Bokaee, Dr. Theresa Zu, Dr. Benjamin Freedman, Dr. Jiun Yen, Dr. Pedro Guimaraes, Emily Berg, David Scherr, Hunter Flick and Herbert Huttanus. I would like to thank also Sara Schlemmer, Kendall Durham and Manuel Martinez Saldgado for their helps.

I would like to thank my friends for their considerable help and support. I would like to extend my thanks particularly to Dr. Saher Lahouar and his wife Afek Jaballah, Dr. Mohamed Jrad and his wife Maha Elouni, Prof. Abdessattar Abdelkefi and his wife Dr. Samah Ben Ayed, for their

generous care and the home feeling atmosphere they created. My thanks goes to my friends here in Blacksburg Dr. Youssef Bichiou, Dr. Bilel Aidi and Karim Fadhloun and my best friends Amel Boukthir, Sonia Gouider and Sara Ghariani.

My deep thanks go to my parents Zohra and Zineddine who encouraged me over the years. I would not thank them enough for everything they have done to me. They raised me with a love for science and the eagerness to succeed, and supported me in all my life goals. My utmost sincere thanks go to my sisters Hanen and Sondes for their continuous support.

Lastly, I thank my loving husband Yassine for his encouragement, advice and patience and most importantly for his sacrifice to allow me fulfill my dreams. His deep faith in my abilities, even when I doubted myself, made this journey possible.

Attribution

Several colleagues and coworkers contributed to the research in this dissertation. A brief description of their background and contributions is described below.

Chapter 2: λ-PCR for precise DNA assembly and modification

Amanda K. Fisher, Ph.D. was a researcher in Dr. Senger's lab and Dr. Bevan's lab and is currently research scientist at BioHybrid Solutions LLC in Monroeville, Pennsylvania. Dr. Fisher is a contributing author to this paper and performed experiments related to a case study with the λ-PCR technique.

Frank Gillam, Ph.D. was a researcher in Dr. Zhang's lab and is currently process development scientist in Raleigh-Durham, North Carolina Area. Dr. Gillam is a contributing author to this paper and performed experiments related to a case study with the λ-PCR technique.

Eva Collakova, Ph.D. is currently an associate professor in the School of Plant & Environmental Sciences at Virginia Tech. Dr. Collakova is a contributing author to the manuscript.

Chenming Zhang, Ph.D. is currently a professor in the department of Biological Systems Engineering at Virginia Tech. Dr. Zhang is a contributing author to the manuscript.

David R. Bevan, Ph.D. is currently a professor in the department of Biochemistry at Virginia Tech. Dr. Bevan is a contributing author to the manuscript.

Ryan S. Senger, Ph. D. is currently an associate professor in the department of Biological Systems Engineering at Virginia Tech and served as the principal investigator for the project. Dr. Senger is a contributing author to the manuscript.

Chapter 3: A novel synthetic sRNA promoting protein overexpression in cell-free systems

Hadi Nazem-Bokaee, Ph.D. was a researcher in Dr. Senger's lab and is currently a research scientist at CSIRO in Australia. Dr. Nazem-Bokaee contributed to the design of the stem-loop system of the antisense-RNA and the design of the plasmids.

David Scherr, is currently a Ph.D. student in Dr. Senger's lab. David performed the Bradford assays for the diaphorase case study.

Sara Schlemmer, is an undergraduate student in the Chemical Engineering department. Sara helped with the optimization of the cell-free-system parameters.

Ryan S. Senger, Ph. D. is currently an associate professor in the department of Biological Systems Engineering at Virginia Tech and served as the principal investigator for the project. Dr. Senger is a contributing author to the manuscript.

Chapter 4: Characterizing *Synechocystis* sp. PCC 6803 with Rametrix™: Glucose, illumination, and nitrogen deprivation-induced phenotypes

Cynthia Denbow, Ph.D., is currently the laboratory and research manager in the Collakova's lab. Dr. Denbow helped with all GC-MS/FID and UPLC experiments.

Eva Collakova, Ph.D. is currently an associate professor in the School of Plant & Environmental Sciences at Virginia Tech. Dr. Collakova assisted all GC-MS/FID and UPLC experiments and analysis and is also a contributing author to the manuscript.

Ryan S. Senger, Ph. D. is currently an associate professor in the department of Biological Systems Engineering at Virginia Tech and served as the principal investigator for the project. Dr. Senger is a contributing author to the manuscript.

Chapter 5: Characterizing *Synechocystis* sp. PCC 6803 with Rametrix^{TM}: Acetate, NaCl, and $MgSO_4$ deprivation-induced phenotypes

Cynthia Denbow, Ph.D., is currently the laboratory and research manager in the Collakova's lab. Dr. Denbow helped with all GC-MS/FID and UPLC experiments.

Eva Collakova, Ph.D. is currently an associate professor in the School of Plant & Environmental Sciences at Virginia Tech. Dr. Collakova assisted all GC-MS/FID and UPLC experiments and analysis and is also a contributing author to the manuscript.

Ryan S. Senger, Ph.D. is currently an associate professor in the department of Biological Systems Engineering at Virginia Tech and served as the principal investigator for the project. Dr. Senger is a contributing author to the manuscript.

Chapter 6: Screening of metagenomic DNA libraries to identify enzymes able to detoxify deoxynivalenol

Nina Wilson, Ph.D. was a Ph.D. student in the Schmale's lab. Nina prepared the cultures from the environmental samples and generated the genomic DNA library.

Nicole McMaster, is currently a research associate in the Schmale's lab. Niki helped with all GC-MS analysis.

David G. Schmale, Ph.D. is currently an associate professor at Virginia Tech. Dr. Schmale is a contributing author to the manuscript.

Ryan S. Senger, Ph. D. is currently an associate professor in the department of Biological Systems Engineering at Virginia Tech and served as the principal investigator for the project. Dr. Senger is a contributing author to the manuscript.

Table of contents

List of Figures

List of Tables

Chapter 1: Introduction

METABOLIC ENGINEERING

Metabolic engineering is the improvement of cellular activities by the manipulation of the metabolic network with the use of recombinant DNA technologies for objectives important in biotechnological applications (1). It has become a key approach for engineering microbial cells factories for the production of valuable chemicals, biofuels, and pharmaceuticals among many others. However, the production of these metabolites from microorganisms has a limited productivity (1,2). Therefore, several high throughput strategies have been developed to improve product formation (3–5). These strategies involve genome modifications (6,7), transcription and translation regulation (5,8,9), protein overexpression, and integration of replicating vectors. Nonetheless, most of these techniques require prior knowledge of bacterial genome, the identification of target gene(s), and an extensive rewiring of cellular metabolism to obtain performant cell factories (10). Despite the advanced systems and synthetic biology, the implementation of these conventional techniques is still laborious, takes time, and is costly. As a result, there is a need for the development of new, rapid, and less expensive tools. These tools would allow for the genetic engineering of microorganisms and the fine-tuning of gene(s) expression levels to achieve industrially attractive yields. These challenges of metabolic engineering of microbial cell factories were among the main motivations of the research described in this dissertation.

Genetic engineering

One of the key requirements of metabolic engineering is the availability of genetic tools for genetic engineering of the host cell (11,12). As emphasized previously, manipulation of the cellular metabolism involves gene knockouts or knock-ins, overexpression, and modification of regulating elements (e.g. promoter, ribosome binding sites, terminators). Most of these modifications use plasmids to introduce genes or to disrupt them, or to modify regulators. In all of these cases, it is desirable to have a tool to manipulate DNA and build plasmids quicky, easily, and reliably. Currently, a wide-range of DNA assembly methods is available, from the conventional restriction digestion and ligation to commercialized kits (13–16). However, the cost of some of these cloning methods, their time-consuming aspect, and irreproducibility limit their use. One of the goals of the present research was to develop an easy construction of DNA vector with a low cost and in few steps. This will direct the metabolic engineering to faster, robust, and reliable designs in a short time.

Small RNAs and Antisense RNAs

Genome editing represents one of the major challenges for metabolic engineering. Chromosomal DNA modification involving gene knockouts, disruption or introduction is limited to predefined locations across the genome and in most cases is lethal when the gene is crucial for growth (10,17–19). Therefore, post-transcriptional mediated regulation is becoming more attractive and synthetic regulatory small RNAs (sRNAs) and antisense RNAs (asRNAs) are gaining more attention (17,18,20,21). asRNAs is a well-studied category of RNA regulators that can regulate conditionally gene expression (22,23). They bind by complementarity to the target mRNA sequence and inhibit gene expression (24–27). Synthetic asRNA regulation has been used

in many metabolic engineering applications such as the optimization of expression level within a target pathway or to down-regulate competing pathways (5,28–30). There are many examples of the application of sRNAs and asRNAs in diverse bacteria and they have great implications for advancement in metabolic engineering of cells.

Motivation for Raman Spectroscopy

While genetic engineering is performed on cells to optimize target product formation, continuous analysis of cells and products is of great importance to assess the performance of the engineered cells. Having a rapid near-real time method to analyze cells and products and that can detect any phenotype change of microbes is of great importance. Initially discovered in 1928, Raman spectroscopy is an optical technique based on inelastic scattering of monochromatic light photons (31) upon interaction with a sample. The acquired Raman signal provides information regarding the structure, bond symmetry and electronic environment of the molecule (32). These unique Raman signal of biological molecules such as nucleic acids, proteins, polysaccharides and lipids generate unique and specific Raman spectra of the whole cells (33,34). Raman Spectroscopy requires minimal or no sample preparation, it is non-destructive, and allows further analysis on the same sample, and allows data acquisition in near-real time (35). Because of these advantages, this technique could be further utilized to detect phenotype changes of cells upon change of their culturing conditions to optimize either target product formation or to differentiate the response mechanism of cells upon exposure to different substrate and screen for desirable phenotypes.

Cell-Free System

Even though the main focus was on microbial cells, a major hindrance for advancement in this field is the extensive robustness of cell metabolism which is due to tight regulation and

interaction (2). In fact, metabolism has evolved to support cell growth and maintenance and not to produce specific metabolites at high rates (2). In this matter, a cell-free system (CFS) offers the possibility to bypass cellular barriers. CFS has emerged as a powerful platform enabling the activation of cellular machinery without the use of living cells. It allows direct control of transcription, translation, and metabolism in an open environment (36,37). Hence, the metabolism is specifically directed to the production of the target product and the requirement of cellular viability is eliminated (38,39). This strategy offers, ultimately, a way to circumvent the robustness of cell metabolism direct the cellular resources toward the desired objective.

CHAPTERS AND ORGANISATION OF THE DISSERTATION

Chapter 2 of this dissertation describes a new method of gene cloning, λ-PCR (lambda-PCR). This technique enables DNA assembly and construction of cloning vectors and was used for the construction of the major part of plasmids used in this research. λ-PCR was proven highly-efficient, rapid and inexpensive. It can be used to assemble DNA fragments or to insert double-stranded DNA at nearly any location of a plasmid.

Chapter 3 investigates the effect of a novel synthetic sRNA-based control system on the regulation of mRNA. A fluorescent gene reporter system was constructed for studying the changes in protein expression level in the presence of the sRNAs, which contain antisense binding sequences. Challenges with living cells were encountered and inconsistency of the results was revealed with the testing of our designs and confirmed by statistical analysis. Therefore, the design and implementation was switched into an *in vitro* CFS. Interestingly, results improved and consistent expression was maintained. The ultimate application of such system will be in fine-

tunable metabolic pathways engineering where the goal is specific gene down-regulation or over-expression in order to produce target metabolites.

Chapters 4 and 5 focus on phenotypic changes of *Synechocystis* sp. PCC6803 when exposed to different environmental changes (i.e. light, glucose concentration, nitrate deprivation, acetate, sodium chloride and magnesium sulfate). Rametrix^{TM} Toolboxes enabled the chemometric analysis of Raman spectra and the validation of the models through the calculation of prediction accuracy, sensitivity and selectivity. Statistical tests were also performed to demonstrate statistically relevant results. Finally, a comparative study was established between specific Raman bands and standardized analytical techniques. The purpose of these chapters is to develop and demonstrate the use of Raman spectroscopy as a near real-time analysis of cellular phenotypic response to changing environmental conditions (e.g. bioprocessing applications) and to identify the major biomolecules modification contributing to a specific phenotype.

Chapter 6 is devoted to the screening of enzymes capable of the degradation of the mycotoxin deoxynivalenol (DON) using enrichment of metagenomic DNA libraries. DNA library fragments were expressed in DON sensitive strain of *Saccharomyces cerevisiae*. For these experiments, DON was used as the sole carbon source and ferulic acid was added to prevent efflux pump enrichment. The goal of this study is to develop an enzymatic strategy for the decontamination of DON. The method is applicable for detoxification for the food and feed industry and even genetic engineering of crops for optimized mycotoxin degradation.

References

1. Bailey JE. Toward a Science of Metabolic Engineering. Science. 1991;252(5013):1668–75.
2. Nielsen J, Keasling JD. Engineering Cellular Metabolism. Cell. 2016;164(6):1185–97.
3. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, et al. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. Science. 2010;330(6000):70.
4. Chaudhary AK, Dhakal D, Sohng JK. An Insight into the "-Omics" Based Engineering of Streptomycetes for Secondary Metabolite Overproduction. BioMed Res Int. 2013;2013:15.
5. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotechnol. 2013;31(2):170–4.
6. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature. 2009;460:894.
7. Sandoval NR, Kim JYH, Glebes TY, Reeder PJ, Aucoin HR, Warner JR, et al. Strategy for directing combinatorial genome engineering in *Escherichia coli*. Proc Natl Acad Sci. 2012;109(26):10540.
8. Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol. 2006;24(8):1027–32.
9. Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol. 2012;30:521.
10. Chaudhary AK, Na D, Lee EY. Rapid and high-throughput construction of microbial cell-factories with regulatory noncoding RNAs. Biotechnol Adv. 2015;33(6 Pt 1):914–30.
11. Redden H, Morse N, Alper HS. The synthetic biology toolbox for tuning gene expression in yeast. FEMS Yeast Res. 2015;15(1):1–10.
12. Jensen MK, Keasling JD. Recent applications of synthetic biology tools for yeast metabolic engineering. FEMS Yeast Res. 2015;15(1):1–10.
13. Hartley JL, Temple GF, Brasch MA. DNA cloning using in vitro site-specific recombination. Genome Res. 2000;10(11):1788–95.
14. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison Iii CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 online;6:343.
15. Engler C, Marillonnet S. Combinatorial DNA assembly using Golden Gate cloning. Methods Mol Biol. 2013;1073:141–56.
16. Engler C, Marillonnet S. Golden Gate cloning. Methods Mol Biol. 2014;1116:119–31.
17. Nakashima N, Tamura T, Good L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in *Escherichia coli*. Nucleic Acids Res. 2006;34(20):e138.
18. Nakashima N, Tamura T. Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. Nucleic Acids Res. 2009;37(15):e103.
19. Nakashima N, Goh S, Good L, Tamura T. Multiple-gene silencing using antisense RNAs in *Escherichia coli*. Methods Mol Biol. 2012;815:307–19.
20. Grosshans H, Filipowicz W. Molecular biology: the expanding world of small RNAs. Nature. 2008 Jan 24;451(7177):414–6.
21. Goh S, Boberek JM, Nakashima N, Stach J, Good L. Concurrent growth rate and transcript analyses reveal essential gene stringency in *Escherichia coli*. PLoS One. 2009;4(6):e6061.
22. Engdahl HM, Hjalt TA, Wagner EG. A two unit antisense RNA cassette test system for silencing of target genes. Nucleic Acids Res. 1997;25(16):3218–27.

23. Yin D, Ji Y. Genomic analysis using conditional phenotypes generated by antisense RNA. Curr Opin Microbiol. 2002;5(3):330–3.
24. Qi LS, Arkin AP. A versatile framework for microbial engineering using synthetic non-coding RNAs. Nat Rev Microbiol. 2014;12(5):341–54.
25. Hussein R, Lim HN. Disruption of small RNA signaling caused by competition for Hfq. Proc Natl Acad Sci U A. 2011;108(3):1110–5.
26. Breaker RR. Prospects for riboswitch discovery and analysis. Mol Cell. 2011;43(6):867–79.
27. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell. 2011;43(6):880–91.
28. Solomon KV, Sanders TM, Prather KL. A dynamic metabolite valve for the control of central carbon metabolism. Metab Eng. 2012;14(6):661–71.
29. Tummala SB, Welker NE, Papoutsakis ET. Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. J Bacteriol. 2003;185(6):1923–34.
30. Yang Y, Lin Y, Li L, Linhardt RJ, Yan Y. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. Metab Eng. 2015;29:217–26.
31. Bocklitz T, Walter A, Hartmann K, Rosch P, Popp J. How to pre-process Raman spectra for reliable and stable models? Anal Chim Acta. 2011;704(1–2):47–56.
32. Das RS, Agrawal YK. Raman spectroscopy: Recent advancements, techniques and applications. Vib Spectrosc. 2011;57(2):163–76.
33. She CY, Dinh ND, Tu AT. Laser raman scattering of glucosamine N-acetylglucosamine, and glucuronic acid. Biochim Biophys Acta BBA - Gen Subj. 1974;372(2):345–57.
34. K. Dutta R, Sharma P, Pandey A. Surface enhanced Raman spectra of Escherichia coli cells using Zno nanoparticles. Vol. 4. 2009. 83–87 p.
35. Zu TNK, Athamneh AIM, Wallace RS, Collakova E, Senger RS. Near-Real-Time Analysis of the Phenotypic Responses of *Escherichia coli* to 1-Butanol Exposure Using Raman Spectroscopy. 2014;
36. Lu Y. Cell-free synthetic biology: Engineering in an open world. Synth Syst Biotechnol. 2017;2(1):23–7.
37. Karim AS, Heggestad JT, Crowe SA, Jewett MC. Controlling cell-free metabolism through physiochemical perturbations. Metab Eng. 2018;45:86–94.
38. Harris DC, Jewett MC. Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. Curr Opin Biotechnol. 2012;23(5):672–8.
39. Hodgman CE, Jewett MC. Cell-free synthetic biology: Thinking outside the cell. Metab Eng. 2012;14(3):261–9.

Chapter 2: λ-PCR for precise DNA assembly and modification

Imen Tanniche[1], Amanda K. Fisher[2], Frank Gillam[1], Eva Collakova[3], Chenming Zhang[1], David R. Bevan[2,4], Ryan S. Senger*[1,5]

[1] Department of Biological Systems Engineering; Virginia Tech; Blacksburg, VA

[2] Interdisciplinary PhD Program in Genetics, Bioinformatics, and Computational Biology; Virginia Tech; Blacksburg, VA

[3] School of Plant & Environmental Sciences; Virginia Tech; Blacksburg, VA

[4] Department of Biochemistry; Virginia Tech; Blacksburg, VA

[5] Department of Chemical Engineering; Virginia Tech; Blacksburg, VA

* Corresponding Author
1230 Washington St.
301C HABB1
Blacksburg, VA 24061
Email: senger@vt.edu
Phone: 540-231-9501

λ-PCR for precise DNA assembly and modification

Abstract

λ-PCR (lambda-PCR) is a novel and open-source method for DNA assembly and cloning projects that has been successful where conventional methods and commercial kits have failed. λ-PCR uses overlap extension to ultimately assemble linear and circular DNA fragments, but it allows the single-stranded DNA (ssDNA) primers of the PCR reaction to first exist as double-stranded DNA (dsDNA). Having dsDNA is advantageous to avoid inhibitory secondary structures during direct synthesis and to reduce costs. Three variations of λ-PCR were created to convert an initial dsDNA product into a ssDNA "megaprimer" to be used in overlap extension: (i) complete digestion by λ-exonuclease, (ii) asymmetric PCR, and (iii) partial digestion by λ-exonuclease. Four case studies are presented that demonstrate the use of λ-PCR in simple gene cloning, simultaneous multi-part assemblies, gene cloning not achievable with commercial kits, and the use of thermodynamic simulations to guide λ-PCR assembly strategies. High DNA assembly and cloning efficiencies have been achieved with λ-PCR for a fraction of the cost and time associated with conventional methods and other commercial kits.

2.1.Introduction

DNA assembly, cloning, and sequence modifications are critical to the biological sciences and engineering, yet these can be error prone and time consuming when performed by conventional methods. In addition, custom DNA synthesis and some commercial cloning kits can still be cost-prohibitive for large projects in many labs. Here, we introduce a new open-source protocol, "λ-PCR" (lambda-PCR), that can be used for DNA manipulations, including routine cloning and large assemblies. It is cost-effective and reliable, accommodates large DNA fragments, and allows assembly of many DNA fragments. In addition, it requires no restriction digestions of any kind and can be designed for nearly any location in a DNA vector (single base-pair resolution). Variations of λ-PCR have been developed to allow different types of genetic manipulations including DNA insertion, substitution, deletion, and point mutations.

PCR-based DNA cloning is used routinely, and most approaches require at least one primer to contain the DNA to be cloned. In some cases, this primer set cannot be synthesized, often due to secondary structure formation or prohibitive synthesis costs. λ-PCR is a PCR-based cloning method, but the DNA to be cloned exists as double-stranded DNA (dsDNA). This approach allows the DNA to be cloned to be PCR amplified (if a template exists), rather than synthesized. In λ-PCR, a dsDNA to be cloned is converted into a single-stranded DNA (ssDNA) "megaprimer" that is used in a second PCR to amplify the vector (e.g., plasmid) DNA while incorporating the megaprimer DNA. Different methods of converting the dsDNA to be cloned into the the ssDNA megaprimer have been explored and have uses in different scenarios. These methods include: (i) complete λ-exonuclease digestion, (ii) asymmetric PCR, and (iii) partial λ-exonuclease digestion. These represent the three variations of λ-PCR that are shown in Figures 2.1-2.3.

Conventional methods of DNA assembly and plasmid vector construction involve restriction digestion and ligation. These strategies can be limiting due to the lack of restriction sites and inherent inefficiencies in cloning certain genes or DNA fragments. As a result, a panoply of DNA plasmid assembly and genetic manipulation techniques has emerged, and many have been commercialized. These techniques are commonly ligase-free, and several involve recombination reactions (1,2). They include the overlap extension PCR (3–6), *in vitro* site-specific recombinational cloning (Gateway® cloning) (7), In-Fusion™ assembly (8), Gibson Assembly® (9), Golden Gate Cloning (10,11), and others. However, the cost of cloning kits, supplies, and DNA synthesis limits the use of some of these technologies in academic labs, and others are limited by the size and number of DNA fragments that can be assembled.

The Omega-PCR (Ω-PCR) method (12) was developed as another alternative to digestion-ligation methods and commercial cloning kits. The underlying basis of Ω-PCR is the use of chimeric PCR primers, where the 5' ends of the forward and reverse primers contain sequences homologous to the flanking ends of the cloning vector, while the 3' ends of the primers are used to amplify the DNA to be cloned or assembled. A first PCR amplifies the DNA to be cloned. This PCR product is denatured at 95°C to form dual ssDNA megaprimers. A second PCR uses these megaprimers, in the presence of a plasmid, to complete the cloning reaction through overlap extension. Since the megaprimers bind to the plasmid backbone in two different and adjacent locations, an 'Ω-shaped' secondary structure is produced, which inspired the Ω-PCR name (12). However, the Ω-PCR protocol failed in our laboratory with our particular gene cloning tasks. Thermodynamic calculations using NUPACK (http://www.nupack.org) (13) revealed a much greater free energy incentive for the megaprimers to re-anneal into a dsDNA product rather than anneal to the plasmid. However, the Ω-PCR protocol inspired the design of λ-PCR and its

variations. The λ-PCR protocol ensures that the long ssDNA megaprimers are favored thermodynamically to bind to the DNA vector and facilitate assembly rather than re-anneal with themselves. There are different methods to produce ssDNA from dsDNA (14,15) including (i) λ-exonuclease enzymatic digestion (14,16–18), (ii) asymmetric PCR (19–21), (iii) streptavidin biotin separation (22), and (iv) denaturing-urea PAGE separation following PCR amplification (23). The λ-exonuclease digestion and asymmetric PCR (Fig.s 2.1-2.3) were used to develop the λ-PCR variations.

λ-Exonuclease is able to generate ssDNA fragments by digesting phosphorylated strand(s) of blunted dsDNA in the 5' to 3' direction (14, 24). It has much higher affinity for a phosphorylated than a hydroxylated 5' end. Additionally, it has reduced activity for nicked, gapped, and ssDNA (25). Therefore, λ-exonuclease was selected to generate ssDNA from the PCR product to be cloned in variation 1 (v1) of λ-PCR (Fig. 2.1). Target dsDNA (i.e, dsDNA to be cloned, inserted, or assembled) is first amplified through PCR, which is done using chimeric primers, where the 3' region is used to amplify the target DNA, and the 5' region has homology to the insertion site on the vector. In addition, one of the primers is phosphorylated at the 5' end; we have a preference for phosphorylating the reverse primer. The amplified dsDNA is then digested by λ-exonuclease overnight, leaving only the complementary strand, which serves as the megaprimer. The megaprimer is involved in a second PCR to insert DNA into the cloning vector. This method of generating the megaprimer using the λ-exonuclease is the preferred variation of λ-PCR. While this has been successful nearly universally in our lab (tested up to inserting 2 kb DNA fragments), two other λ-PCR variations were developed and involve modified protocols for generating the ssDNA megaprimer. The second λ-PCR variation (v2) uses asymmetric PCR (Fig. 2.2) and the third λ-PCR variation (v3) allows non-terminal portions of the megaprimer to remain as dsDNA (Fig.

2.3). This provides an advantage in cloning or assembling very large DNA fragments or sequences prone to forming inhibitory secondary structures as ssDNA.

Asymmetric PCR (or single-stranded PCR) uses unequal molar amounts of PCR primers to generate ssDNA for sequencing and to improve detection signals in real-time PCR (19), and recent studies have focused on its optimization (20,21). Asymmetric PCR begins with an exponential amplification phase, where both primers are used. Once the limiting primer is depleted, the reaction switches to linear amplification, and the ssDNA product is accumulated for the remaining thermal cycles (26). The ssDNA product serves as the megaprimer in λ-PCR v2 (Fig. 2.2).

λ-PCR v3 is related to the original Ω-PCR protocol (12). In our trials with Ω-PCR, the dual megaprimers were thermodynamically favored (via NUPACK calculations) to re-anneal rather than bind the DNA vector and facilitate assembly. In λ-PCR v3, both primers of the initial PCR are phosphorylated, and λ-exonuclease is applied for 1-hour to partially digest the dsDNA and create single-stranded overhangs with homology to the vector (Fig. 2.3). Based on thermodynamics, this dsDNA may anneal to the vector or re-anneal to itself during the assembly/cloning reaction. In Fig. 2.3, it is shown annealing to the vector as ssDNA. If it re-anneals to itself, the ssDNA overhangs will still anneal to the vector (Fig. 2.3). In our experience, λ-PCR v3 is less efficient than v1 and v2 described above. However, v3 may be of value when a ssDNA megaprimer forms inhibitory secondary structures or leads to undesirable reactions.

To demonstrate the different applications and capabilities of λ-PCR, four case studies are presented. Case Study 1 involves simplified cloning procedures involving *green fluorescent protein* (*gfp*) gene, *cyan fluorescent protein* (*amcyan*) gene, and the pUC19 plasmid. Three distinct

genetic manipulations are performed using λ-PCR in this case study: (i) direct insertion of the *gfp* gene behind *lacZα* in pUC19, (ii) simultaneous insertion of both *gfp* and *amcyan* genes in pUC19, and (iii) substitution of the *lacZα* gene in pUC19 with *gfp*. λ-PCR variations v1 and v2 were both used in these applications. Case Study 2 involves the creation of plasmids for gene knockout by homologous recombination in the cyanobacterium *Synechocystis* PCC 6803. λ-PCR was used to construct plasmids containing a kanamycin resistance (*kan*^{r}) gene flanked by more than 100 bp sequences with chromosomal homology. The expansion of the homologous regions beyond 100 bp increased recombination efficiency, and λ-PCR enabled simple construction of this plasmid in four steps. Case Study 3 demonstrates λ-PCR v3 to clone "hard to clone" genes used in metabolic and enzyme engineering research. Finally, Case Study 4 demonstrates the use of thermodynamic calculations using NUPACK (13) to troubleshoot λ-PCR, design effective primers, and simulate annealing of the megaprimer.

2.2. Materials and methods

2.2.1. Plasmids, genes, primers, and strains

The case studies made use of pUC19, pBAD/HisA, pACYC177, and pET28a plasmids as cloning vectors. All plasmids used and created are listed in Table A1. The reporter genes for green fluorescent protein (*gfp*) and cyan fluorescent protein (*amcyan*) were obtained from plasmids pET-GFP and pAmCyan (Takara Bio USA; Mountain View, CA). Chimeric primers for each DNA insertion product and one reverse primer complementary to the pUC19 backbone were designed and are listed in Table A2 in the Supplementary Appendix A. Chimeric primers contained two regions: (i) for amplifying the target DNA and (ii) homologous to the plasmid vector backbone to complete the cloning reaction, as shown in Fig.s 2.1-2.3. All primers were synthesized by

Integrated DNA Technologies (Coralville, IA). *E. coli* 10-beta, *E. coli* BL21DE3, and T7 Express® cells (New England Biolabs; Ipswitch, MA) were used for transformations.

2.2.2. Megaprimer generation by λ-exonuclease digestion

To generate the megaprimer for λ-PCR v1, the reverse primer in the first PCR was phosphorylated (Fig. 2.1) as follows: (i) 0.5 µL of reverse primer (100 µM), (ii) 0.5 µL of 10x T4 DNA ligase buffer (New England Biolabs), (iii) 0.5 µL T4 polynucleotide kinase (New England Biolabs), and (iv) molecular biology grade water adjusted to 5.0 µL. The reaction was incubated at 37°C for 1 hour. The phosphorylated reverse primer and un-phosphorylated forward primer were used to amplify the target DNA sequence to be cloned. Final PCR volumes were 25 µL each and consisted of: (i) 0.5 µL reverse primer, (ii) 0.5 µL forward primer, (iii) template DNA (~50 ng/µL), (iv) 12.5 µL Q5® High-Fidelity 2X Master Mix (New England Biolabs), and (v) molecular biology grade water adjusted to 25 µL. After 30 cycles of standard PCR, the amplified DNA was purified using the GeneJET™ PCR purification kit (ThermoFisher Scientific; Waltham, MA). The recovered PCR fragments were then digested with λ-exonuclease enzyme (New England Biolabs). The digestion reaction consisted of: (i) 0.5 µL 10x reaction buffer, (ii) 5.0 µL of purified PCR product, (iii) 0.5 µL λ-exonuclease enzyme, and (iv) volume adjusted to 10 µL with molecular biology grade water. The reaction was incubated overnight, and then the λ-exonuclease was deactivated by incubation at 75°C for 20 min.

2.2.3. Megaprimer generation by asymmetric PCR

A previously published asymmetric PCR protocol (21) was followed: (i) 1 pmol limit primer (reverse primer), (ii) 15 pmol excess primer (forward primer), (iii) template DNA (~50 ng/µL), (iv) 20.0 µL Q5® High-Fidelity 2X Master Mix (New England Biolabs), and (v) molecular

biology grade water adjusted to 40 µL. The PCR cycling conditions included an initial denaturing at 98°C for 2 min followed by 30 cycles of: (i) denaturing at 98°C for 30 seconds, (ii) annealing for 20 seconds, and (iii) elongating at 72°C (allowing 30 s per kb). A final extension step at 72°C for 3 min was then applied. PCR products were visualized by 1% agarose gel. To verify that the PCR yielded ssDNA, 5 µL of the mixture containing PCR products were treated with 20 U of S1 nuclease (ThermoFisher Scientific) to degrade ssDNA. The reaction was performed at room temperature for 1 h, and the DNA mixture was re-visualized.

2.2.4. DNA cloning and assembly reactions

The ssDNA megaprimer, generated from a method described above, was used in a second PCR to complete the cloning reaction. The second PCR consisted of: (i) 0.5 µL reverse primer (100 µM), (ii) 2.0 µL of the first PCR product containing the megaprimer, (ii) cloning plasmid (~50 ng/µL), (v) 12.5 µL Q5® High-Fidelity 2X Master Mix (New England Biolabs), and (v) molecular biology grade water to 25 µL. A standard PCR program was run for 30 cycles. The newly amplified plasmid was recovered using a GeneJET™ PCR purification kit (ThermoFisher Scientific). The final product is a dsDNA plasmid with two nicks (one on each strand). It was then transformed into competent *E. coli*, which seal the nicks with DNA repair enzymes. In some instances, where only one of the chimeric primers has an overlap with the plasmid (Case Study 2), the final product is linear. An optional ligation step with T4 DNA ligase (New England Biolabs) was used here. DpnI treatment (New England Biolabs), according to the manufacturer's protocol, was used to digest original plasmid templates and improve cloning efficiency.

2.2.5. Transformation, colony screening, and fluorescence detection

Competent *E. coli* 10-beta cells were transformed by heat-shock according to the manufacturer's protocol and incubated overnight in LB plates supplemented with 100 µg/mL ampicillin and 0.05 µM XGal/IPTG. Successful clones were identified by colony PCR using the One*Taq*® 2X master mix with standard buffer (New England Biolabs). PCR products were visualized by 1% agarose gel, and positive colonies were grown in liquid LB at 30°C in 96-well microplates in an incubating Synergy H4 microplate reader (BioTek; Winooski, VT). GFP and AmCyan fluorescence intensity was monitored using excitation/emission wavelengths of 480/515 nm ($I_{480/515}$) and 458/489 nm ($I_{458/489}$), respectively. Culture growth was monitored by optical density at 600 nm (OD_{600}).

2.3. Results

2.3.1. Case Study 1: Fluorescent reporter gene insertions into pUC19

The λ-PCR protocol was demonstrated using the fluorescent reporter genes (*gfp* and *amcyan*) and the pUC19 plasmid. The following manipulations were performed: (i) insertion of *gfp* behind the native *lacZα* of pUC19, (ii) simultaneous insertion of both *gfp* and *amcyan* genes behind *lacZα*, and (iii) substitution of *lacZα* by *gfp*. For this case study, two versions of λ-PCR were used: v1 (λ-exonuclease) and v2 (asymmetric PCR). All plasmids used and constructed are given in Table A1, and all DNA primers used in this case study are given in Table A2. A schematic of the λ-PCR cloning strategies are given in Fig. A1 of the Supplementary Appendix A.

Successful clones were obtained for all three manipulations described above with all three variations of λ-PCR (DNA gel images provided as Fig. A2). The λ-PCR v1 protocol led to the

highest cloning efficiency (thousands of positive colonies per µg of DNA transformed). Fluorescence data for all three constructed plasmids with λ-PCR v1 and v2 are given in Fig. 2.4 and demonstrate successful gene cloning and expression. Cultures transformed with the native pUC19 plasmid served as the negative control, and cultures with the pET-GFP and pAmCyan plasmids served as positive controls for *gfp* and *amcyan* expression, respectively.

2.3.2. Case Study 2: Plasmid construction for Synechocystis PCC 6803 gene knockout

Gene knockout in *Synechocystis* PCC 6803 is performed by homologous recombination (27). In this case study, a *kan*^{r} gene was cloned into pUC19 and was flanked by regions (greater than 100 bp) with chromosomal homology. With conventional plasmid construction methods (i.e., restriction digestion and ligation), building this plasmid involves the ligation of 3 DNA fragments (2 homologous region sequences flanking the antibiotic resistance gene) and insertion of this product into a multiple cloning site in the plasmid. In our experience, this has proven difficult, as the homologous region sequences can approach hundreds of base pairs for effective homologous recombination in *Synechocystis.* This has led to very low cloning efficiency and reliance on the availability of adequate restriction sites.

With λ-PCR, this construction was simplified considerably and was done with high-efficiency. The photosystem II (*psba2*) gene was chosen as the site of homologous recombination in *Synechocystis* because this cyanobacterium has two copies of this gene and knocking out of them does not result in lethality. Four steps were required for the final assembly of the plasmid. First, the target *psba2* gene (1083 bp) was amplified from the *Synechocystis* genome with phosphorylated forward and reverse primers. Next, it was partially digested with λ-exonuclease for 30 min (λ-PCR v3). The resulting product, with ssDNA 3’ ends, was used as the megaprimer

in the cloning reaction with pUC19 to replace the *lacZα* gene. Third, another PCR (with a phosphorylated reverse primer only) amplified the *kan*^{r} gene from the pACYC177 plasmid, and it was digested with λ-exonuclease overnight (λ-PCR v1). This served as the ssDNA megaprimer to insert the *kan*^{r} gene into the middle of the already cloned *psba2* gene. The resulting plasmid, pUC19-psba2-kan, was transformed into *Synechocystis*, and cultures were grown in BG11 medium (28) under increasing levels of kanamycin, as described previously (29, 30). After approximately 1 month, a culture capable of growing in 15 μg/mL of kanamycin was obtained, and genetic screening indicated the presence of the *kan*^{r} gene inserted into the genome.

2.3.3. Case Study 3: Cloning heterologous genes for enzyme engineering

To build a *de novo* biosynthetic pathway for *in vivo* production of 3-hydroxy-2-pyrrolidinone in *E. coli*, two heterologous genes are needed. The first is malyl-CoA lyase (*mcla*), which was isolated from *Chloroflexus auriantiacus* (31,32), and the second is malonyl-CoA reductase (*mcrt*), which was obtained from *Sulfolobus tokodaii* (33,34). For a functional pathway, both enzymes must be engineered to accept non-native substrates, and that research is ongoing. Here, we describe how to clone these genes using λ-PCR v3, as cloning these genes was unsuccessful (for reasons unknown) after months of using conventional methods and several commercial cloning kits. The pBAD/HisA plasmid of the pBAD/His kit (ThermoFisher Scientific) was selected as the cloning vector due to its histidine tag, reliable arabinose-inducible promoter, and ampicillin resistance (*amp*^{r}) gene. λ-PCR v3 was used to insert the 1047 bp *mcla* and 1071 bp *mcrt* genes into separate pBAD/HisA plasmids. This case study showed that λ-PCR v3 can accommodate full gene-sized inserts (~1 kb). It has now been shown that the method can be used to clone DNA fragments in excess of 2 kb (data not shown). With λ-PCR v3, *mcla* cloning was

successful on the first transformation, after months of failures with other protocols. The *mcrt* gene required a few transformations to obtain positive colonies, but significantly improved cloning efficiency was observed for *mcrt* overall using λ-PCR v3. This suggested that a full ssDNA megaprimer containing the *mcrt* gene may have been prone to forming secondary structures inhibitory to the cloning reaction. The improved efficiency using v3 (with a dsDNA megaprimer) illustrates the value of this variation for cases where the usually more efficient v1 fails. Additional details of this case study are given in the Supplementary Appendix A.

2.3.4. Case Study 4: Troubleshooting using thermodynamic calculations

After many successful trials with λ-PCR, we identified a case where gene cloning was unsuccessful. This was investigated by thermodynamic calculations using NUPACK (13), and a solution was found. The purpose of this case study was to illustrate this potential problem, describe the solution, and demonstrate how to identify this problem *a priori*. Hepatitis b core antigen (*HBcAg*) gene (509 bp) was cloned previously into pET-28a for expression and assembly into a virus-like particle. To use the protein as a backbone for vaccines to other viruses, foreign epitopes (128 bp) were to be inserted into the major immunodominant region (MIR) of the *HBcAg* gene using λ-PCR. The only change to the λ-PCR v1 protocol described previously is that plasmids were transformed into freshly thawed T7 Express® cells. Further details and Figures related to this case study are given in the Supplementary Appendix A. Plasmids used are given in Table A1, and all primers are available in Table A2.

Following the attempted insertion of the foreign epitope by λ-PCR v1, the cloning products were visualized by gel electrophoresis. DNA fragment "streak," consisting of multiple indistinguishable bands of varying intensity was evident, as shown in Fig. A8A in the

Supplementary Appendix A. This phenomenon is indicative of mis-priming (i.e., binding of a primer or megaprimer in multiple or undesired locations) due to secondary structure formation in the binding region of a primer (or megaprimer). Risk of secondary structure formation increases in GC-rich regions. In this case study, the foreign epitope was designed to encode glycine-rich terminal ends by including three glycine residues upstream and downstream of the inserted peptides to improve flexibility of the epitope peptide insert. The only codons available for glycine in *E. coli* are $GG(X)_3$. Following initial λ-PCR v1 cloning, 112 colonies were screened to locate a positive clone, but none was ever found.

To solve this problem, thermodynamic calculations with NUPACK (13) were used to simulate secondary structure formation of the ssDNA megaprimer during λ-PCR v1. The cloning reaction of λ-PCR was run with an annealing temperature of 64°C. The 3' binding portion of the megaprimer must have affinity to the plasmid at this temperature, rather than a secondary structure within itself. NUPACK simulations of the megaprimer at 95°C, 80°C, 65°C, and 64°C are shown in Figs. A9A-D in the Supplementary Appendix A. Here, it is apparent that a large hairpin secondary structure is predicted to form near the 3' end at 65°C, with an additional hairpin structure formation at the 5' end at 64°C (although this should not impact gene insertion by λ-PCR). With much of the 3' end of the megaprimer (with complementarity to the pET-28a plasmid) involved in a secondary structure, proper plasmid annealing in λ-PCR does not occur, and the gene cloning reaction fails. Alternatively, the presence of the secondary structure allows other portions of the megaprimer to bind elsewhere to the plasmid, giving rise to several incomplete gene insertion products, as shown in Fig. A8A in the Supplementary Appendix A.

To implicate the $GG(X)_3$ nucleotide sequences in the failed cloning reaction, they were removed from the initial primers, and λ-PCR v1 was repeated. This time, the cloning reaction was successful (see Fig. A8B in the Supplementary Appendix A). NUPACK simulations (shown in Fig. A9E in the Supplementary Appendix A) indicated there were no longer inhibitory structures in the ssDNA megaprimer at the annealing temperature of the cloning reaction. Thus, the question remained as to how this cloning reaction can be performed if inhibitory secondary structure forms in the megaprimer? λ-PCR provides a unique solution to this "hard to clone" problem, which is common among all methods that involve ssDNA (even in shorter overhangs). New primers were designed that create a megaprimer to contain the foreign epitope DNA, but the 3' end was extended further past the $GG(X)_3$ hairpin forming sequences, into the *HBcAg* gene itself. The goal was to create a megaprimer for the cloning reaction that did not form secondary structures at the 3' annealing portion at the desired annealing temperature (64°C in this case). Because λ-PCR allows DNA cloning/assembly to occur anywhere (down to single nucleotide resolution), the megaprimer can be extended as needed to minimize inhibitory secondary structures at the annealing temperature of the cloning/assembly reaction. This process is illustrated in Fig. A10 in the Supplementary Appendix A and demonstrates how thermodynamic calculations can be used to optimize λ-PCR. The NUPACK simulation of this final design is shown in Fig. A9F, and the successful cloning result is shown in Fig. A8C. In our experience, preventing the formation of secondary structures in the ssDNA binding portion of the megaprimer at the annealing temperature is key to success.

2.4. Discussion

λ-PCR is a time- and resource-saving open-source method for DNA assembly and manipulation. It can be applied broadly and used for routine DNA cloning as well as for the "hard to clone" cases. The trend in DNA cloning technologies has moved away from restriction digest and ligation methods, and many versions of these methods are now available as commercial kits. However, their cost can be prohibitive to small and academic labs, especially when the cloning reaction efficiency is low for a particular DNA fragment. λ-PCR was developed in response to our failures to clone particular "hard to clone" DNA genes/fragments using traditional methods and several commercial kits. With λ-PCR, we succeeded cloning the "hard to clone" DNA fragments and genes, and generally higher cloning efficiencies have been observed. In addition to being restriction enzyme-free, the λ-PCR method also allows direct cloning of nearly any DNA fragment into nearly any vector at nearly any location (with single base pair resolution), which is another advantage not available from traditional methods or commercial kits that rely on restriction digestion.

The λ-PCR method (including the name) was inspired by the published Ω-PCR method (12) and resembles the overlap extension PCR technique (3–6, 35–37). The major difference is that λ-PCR allows the incorporation of large DNA fragments without the custom synthesis of long primers, which is accomplished through the generation of a megaprimer by PCR. The three variations of λ-PCR deal with how to convert this dsDNA megaprimer product into a form with ssDNA ends that can then complete the overlap extension by PCR. In our experience, this added step in λ-PCR is crucial because secondary structures can inhibit ssDNA custom synthesis and cloning/assembly. The generation of the megaprimer may provide the flexibility to accommodate

many more DNA fragments prone to secondary structure formation. While this scenario is unavoidable with some custom DNA projects, λ-PCR provides a way to mitigate this. Finally, as with overlap extension PCR, λ-PCR can be used to add DNA fragments serially to a plasmid or linear DNA strand. This makes DNA assembly limited by the molecular stability of the dsDNA strand itself, not by the assembly technique. In our practice, each new DNA fragment can be added in 1-2 days (PCR generation of the megaprimer, conversion of dsDNA to ssDNA, and cloning/assembly by PCR). This time can be reduced dramatically if the generation of multiple megaprimers and conversion to ssDNA is done in parallel. While the importance of thermodynamic calculations has been illustrated in Case Study 4, the adoption of λ-PCR more broadly will be aided by megaprimer design software that will automatically choose primers to minimize secondary structure formation in the binding region. This software is in development and will be offered as a freely-available platform to design λ-PCR primers.

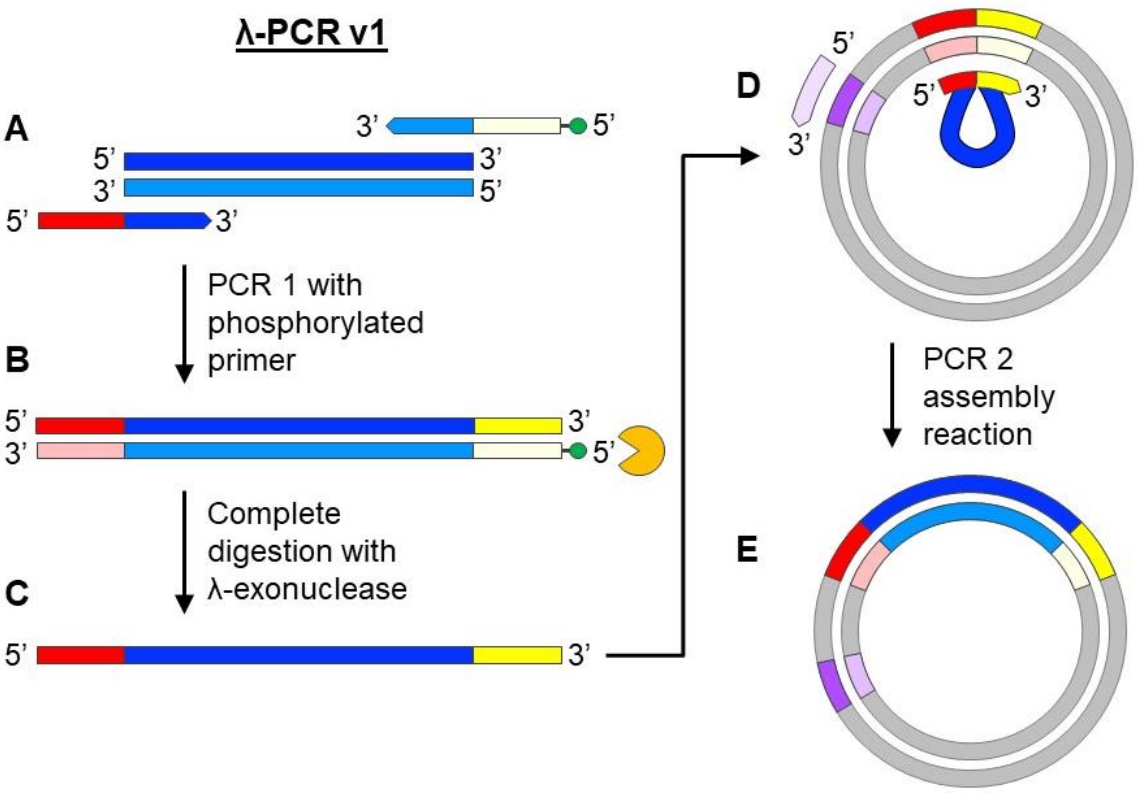


Figure 2.1. Schematic of the λ-PCR v1 protocol, which uses complete digestion with λ-exonuclease to generate the ssDNA megaprimer.

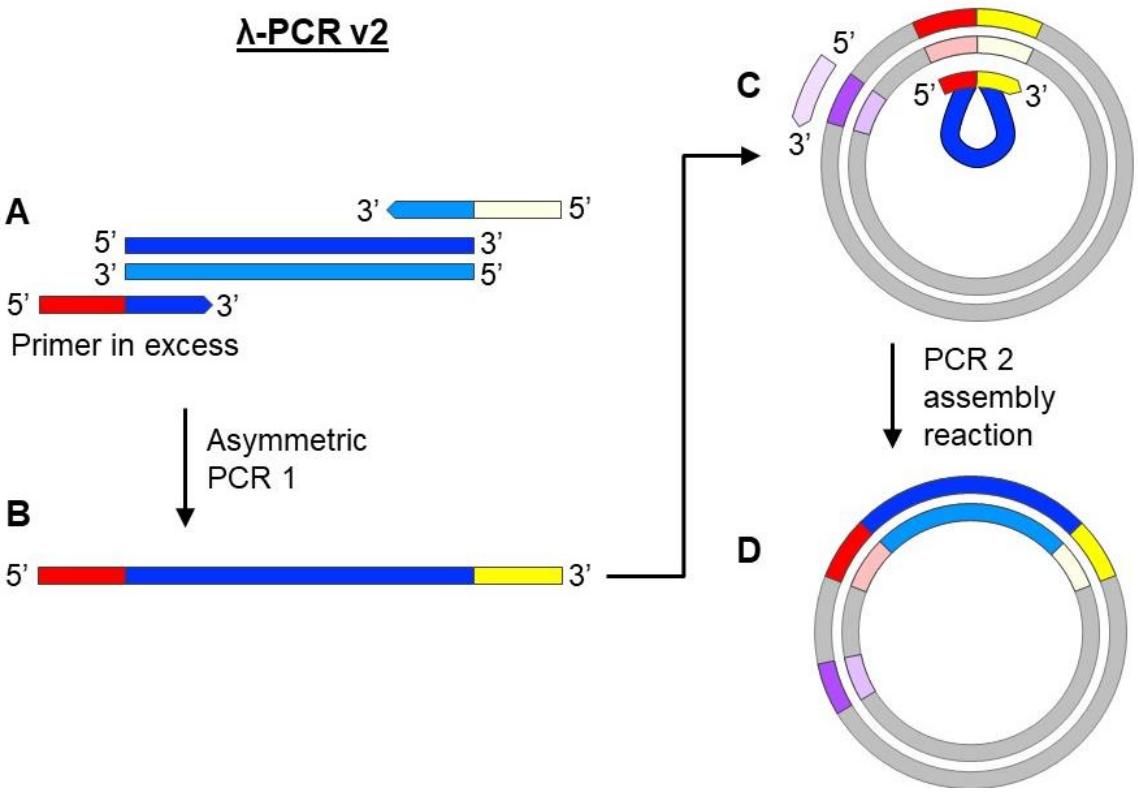


Figure 2.2. Schematic of the λ-PCR v2 protocol, which uses asymmetric PCR to generate the ssDNA megaprimer.

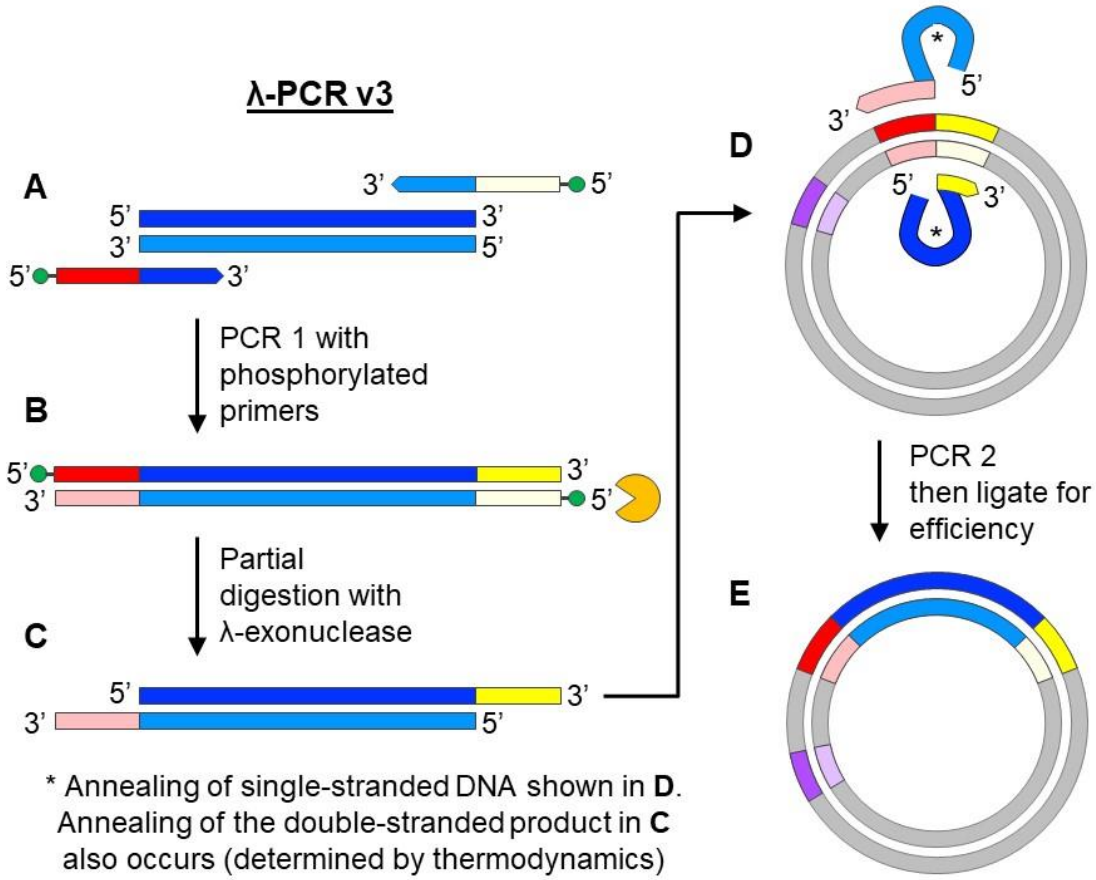


Figure 2.3. Schematic of the λ-PCR v3 protocol, which uses partial digestion with λ-exonuclease to generate the dsDNA megaprimer with ssDNA overhangs.

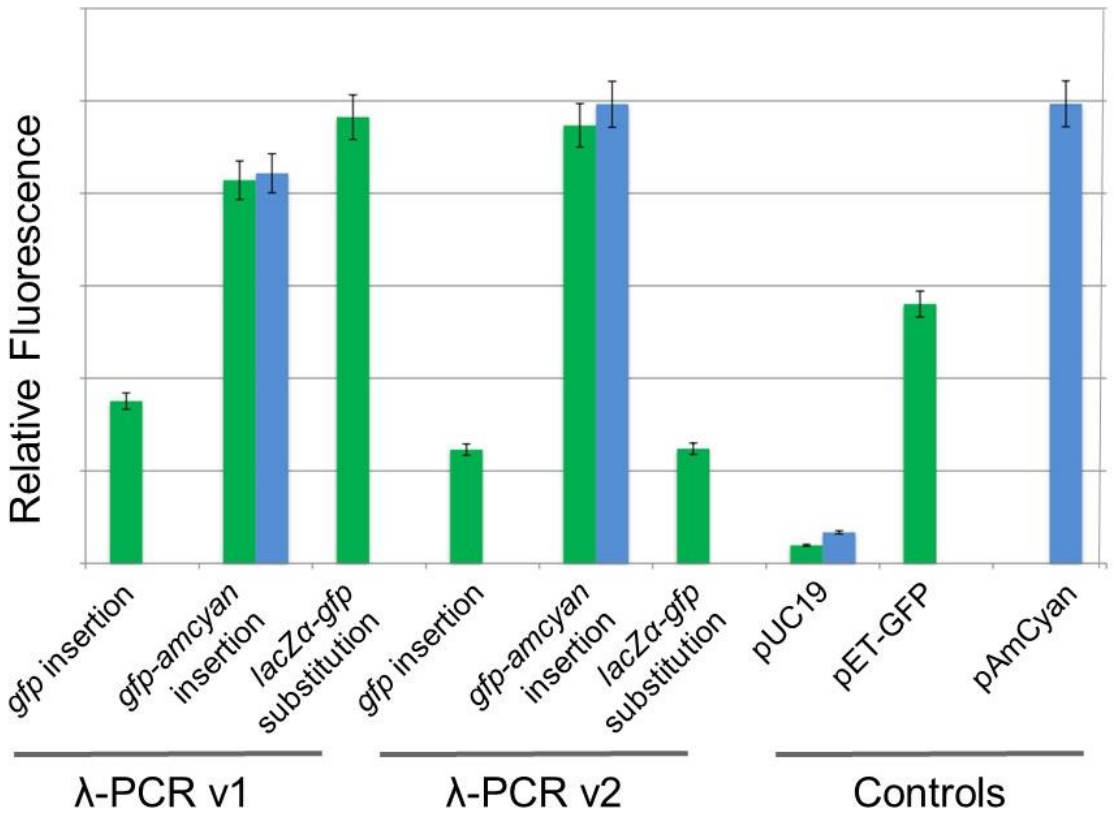


Figure 2.4. GFP (Green) and AmCyan (Blue) fluorescence levels of transformed cultures.

References

1. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 1990 Oct 25;18(20):6069–74.
2. Li MZ, Elledge SJ. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat Methods. 2007 Mar;4(3):251–6.
3. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene. 1989 Apr;77(1):51–9.
4. Heckman KL, Pease LR. Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc. 2007;2(4):924.
5. Bryksin A, Matsumura I. Overlap extension PCR cloning. Methods Mol Biol Clifton NJ. 2013;1073:31–42.
6. Bryksin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques. 2010;48(6):463–465.
7. Hartley JL, Temple GF, Brasch MA. DNA Cloning Using In Vitro Site-Specific Recombination. Genome Res. 2000 Nov 1;10(11):1788–95.
8. Zhu B, Cai G, Hall EO, Freeman GJ. In-Fusion^{TM} assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. Biotechniques. 2007;43(3):354–359.
9. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison III CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343.
10. Engler C, Kandzia R, Marillonnet S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLOS ONE. 2008 Nov 5;3(11):e3647.
11. Engler C, Marillonnet S. Golden Gate Cloning. In: Valla S, Lale R, editors. DNA Cloning and Assembly Methods [Internet]. Totowa, NJ: Humana Press; 2014 [cited 2019 Jan 22]. p. 119–31. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-62703-764-8_9
12. Chen L, Wang F, Wang X, Liu Y-G. Robust one-tube $Ømega$-PCR strategy accelerates precise sequence modification of plasmids for functional genomics. Plant Cell Physiol. 2013;54(4):634–642.
13. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, et al. NUPACK: analysis and design of nucleic acid systems. J Comput Chem. 2011;32(1):170–173.
14. Citartan M, Tang T-H, Tan S-C, Gopinath SC. Conditions optimized for the preparation of single-stranded DNA (ssDNA) employing lambda exonuclease digestion in generating DNA aptamer. World J Microbiol Biotechnol. 2011;27(5):1167–1173.
15. Citartan1 M, Tang1 T-H, Tan S-C, Hoe1 C-H, Saini R, Tominaga4 J, et al. Asymmetric PCR for good quality ssDNA generation towards DNA aptamer production. Sonklanakarin J Sci Technol. 2012;34(2):125.
16. Mitsis PG, Kwagh JG. Characterization of the interaction of lambda exonuclease with the ends of DNA. Nucleic Acids Res. 1999;27(15):3057–3063.
17. Null AP, Hannis JC, Muddiman DC. Preparation of single-stranded PCR products for electrospray ionization mass spectrometry using the DNA repair enzyme lambda exonuclease. Analyst. 2000;125(4):619–626.

18. Higuchi RG, Ochman H. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. 1989;17(14):5865–5865.
19. Gyllensten UB, Erlich HA. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc Natl Acad Sci. 1988;85(20):7652–7656.
20. Sanchez JA, Pierce KE, Rice JE, Wangh LJ. Linear-After-The-Exponential (LATE)–PCR: An advanced method of asymmetric PCR and its uses in quantitative real-time analysis. Proc Natl Acad Sci. 2004;101(7):1933–1938.
21. Tang X, Morris SL, Langone JJ, Bockstahler LE. Simple and effective method for generating single-stranded DNA targets and probes. BioTechniques. 2006;40(6):759–763.
22. Hultman T, StAhl S, Homes E, Uhlen M. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. Nucleic Acids Res. 1989;17(13):4937–4946.
23. Williams KP, Bartel DP. PCR product with strands of unequal length. Nucleic Acids Res. 1995 Oct 25;23(20):4220–1.
24. Subramanian K, Rutvisuttinunt W, Scott W, Myers RS. The enzymatic basis of processivity in λ exonuclease. Nucleic Acids Res. 2003 Mar 15;31(6):1585–96.
25. Little JW, Lehman IR, Kaiser AD. An exonuclease induced by bacteriophage λ I. Preparation of the crystalline enzyme. J Biol Chem. 1967;242(4):672–678.
26. Rice JE, Sanchez JA, Pierce KE, Reis Jr AH, Osborne A, Wangh LJ. Monoplex/multiplex linear-after-the-exponential-PCR assays combined with PrimeSafe and Dilute-'N'-Go sequencing. Nat Protoc. 2007;2(10):2429.
27. Heidorn T, Camsund D, Huang H-H, Lindberg P, Oliveira P, Stensjö K, et al. Synthetic biology in cyanobacteria: engineering and analyzing novel functions. In: Methods in enzymology. Elsevier; 2011. p. 539–579.
28. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology. 1979;111(1):1–61.
29. Grigorieva G, Shestakov S. Transformation in the cyanobacterium Synechocystis sp. 6803. FEMS Microbiol Lett. 1982;13(4):367–70.
30. Schwarzkopf M, Yoo YC, Huckelhoven R, Park YM, Proels RK. Cyanobacterial Phytochrome2 Regulates the Heterotrophic Metabolism and Has a Function in the Heat and High-Light Stress Response. PLANT Physiol. 2014 Apr 1;164(4):2157–66.
31. Herter S, Farfsing J, Gad'On N, Rieder C, Eisenreich W, Bacher A, et al. Autotrophic CO2 fixation by Chloroflexus aurantiacus: study of glyoxylate formation and assimilation via the 3-hydroxypropionate cycle. J Bacteriol. 2001;183(14):4305–4316.
32. Herter S, Busch A, Fuchs G. L-Malyl-coenzyme A lyase/β-methylmalyl-coenzyme A lyase from Chloroflexus aurantiacus, a bifunctional enzyme involved in autotrophic CO2 fixation. J Bacteriol. 2002;184(21):5999–6006.
33. Alber B, Olinger M, Rieder A, Kockelkorn D, Jobst B, Hügler M, et al. Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal Metallosphaera and Sulfolobus spp. J Bacteriol. 2006;188(24):8551–8559.
34. Demmer U, Warkentin E, Srivastava A, Kockelkorn D, Pötter M, Marx A, et al. Structural basis for a bispecific NADP+ and CoA binding site in an archaeal malonyl-coenzyme A reductase. J Biol Chem. 2013;jbc–M112.

35. Gaugué I, Oberto J, Putzer H, Plumbridge J. The use of amino sugars by Bacillus subtilis: presence of a unique operon for the catabolism of glucosamine. PLoS One. 2013;8(5):e63025.
36. Jiang X, Yang J, Zhang H, Zou H, Wang C, Xian M. In vitro assembly of multiple DNA fragments using successive hybridization. PLoS One. 2012;7(1):e30267.
37. Quan J, Tian J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc. 2011;6(2):242.

Chapter 3: A novel synthetic sRNA promoting protein overexpression in cell-free systems

Imen Tanniche[1], Hadi Nazem-Bokaee[1], David M. Scherr[1], Sara Schlemmer[2], Ryan S. Senger*[1,2]

[1]Department of Biological Systems Engineering; Virginia Tech; Blacksburg, VA

[2]Department of Chemical Engineering; Virginia Tech; Blacksburg, VA

* Corresponding Author
1230 Washington St.
301C HABB1
Blacksburg, VA 24061
Email: senger@vt.edu
Phone: 540-231-9501

A novel synthetic sRNA promoting protein overexpression in cell-free systems

Abstract

Bacterial small RNAs (sRNAs) that regulate gene expression have been engineered for uses in synthetic biology and metabolic engineering. Here, we designed a novel non-Hfq-dependent sRNA scaffold that makes use of a 20 nt antisense binding region to selectively target mRNAs and influence expression. The system was developed for regulation of a fluorescent reporter *in vivo* using *E. coli*, but the system was found to be more responsive and produce statistically significant results when applied to protein synthesis in *in vitro* cell-free systems (CFS). Antisense binding sequences were designed to target not only translation initiation regions but various secondary structures in the reporter mRNA. Targeting a high energy stem loop structure and the 3' end of the mRNA produced expression knock-downs that approached 70%. Notably, targeting a low energy stem structure near a potential RNase E binding site lead to a statistically significant 65% increase in gene expression. These results were not obtainable *in vivo*, and the underlying mechanism was translated from the reporter system to achieve better than 75% increase in recombinant diaphorase expression in an *in vitro* CFS. It is possible the designs developed here can be applied to improve expression of other proteins in a CFS.

3.1. Introduction

Non-coding bacterial RNAs such as antisense RNAs (asRNAs) and small RNAs (sRNAs) have become important tools in synthetic biology (3,20,27–32) and metabolic engineering (33,34). They have emerged as important regulators in both prokaryotes and eukaryotes (1,2) and have been found to play a central role in regulating gene expression (3) at post-transcriptional (3–5) and translational levels (6,7). Most act in *trans* (8) by annealing to target mRNAs, typically at or near the ribosome-binding site (RBS) sequence (9,10). sRNAs can have antisense sequences (a short asRNA portion) to bind mRNA targets and/or act with catalytic function. The asRNA-mRNA duplex formation has been documented to decrease translation (9,11–18) and/or increase the degradation of target mRNA (9,19–22). This resulting decrease in target gene expression has been found to be tunable through calculation and design of asRNA-mRNA binding free energy (23). Synthetic regulatory asRNAs have been shown to be suitable for conditional gene silencing (24,25) and can serve as a convenient tool to engineer metabolism due to their ability to regulate gene expression without intensive genetic manipulations such as gene knockouts or disruption (10,17); they can also be expressed from inducible promoters, providing additional metabolic controls. Synthetic sRNAs and asRNAs have already been used in a number of applications ranging from detecting metabolic state (26), balancing metabolic pathway expression (22), and tightly regulating toxin genes (27). Promising applications also include multiplexing and combinatorial applications to screen for enhanced target chemical production (23) and optimization of target pathway gene expression levels and down-regulation of competing pathways (23,28–30). Finally, asRNA technology has been used in industrial processes to improve yields and reduce byproducts (13,30–32), to optimize protein production (11), and it has been combined with CRISPR/Cas9 systems for derepression (5,33,34).

The majority of the asRNAs applications mentioned above and strategies in synthetic biology, in general, make use of living cells. However, the complexities of living cells, including selective induction, natural variability, and difficulties in standardization can make these systems difficult to work with and sometimes irreproducible (35,36). To address these challenges, cell-free systems (CFS) were developed (35–38). Cell-free biology systems harnesses cellular machinery *in vitro* (35,39,40) and can be categorized into three main classes. The first is a simplified CFS composed of crude extract prepared simply by lysing bacterial, plant, or animal cells (35,36,41). The second contains purified *E. coli* translational components and required buffers, amino acids, and energy/cofactor components (37,42–45), and the third class of CFS contains these with purified enzymes and complexes from different sources (46–51). The ability CFSs to bypass the complexities of living cells in an *in vitro* environment offers engineering flexibility and other advantages such as controllable transcription/translation modifications, high synthesis rates and product yields, high tolerance to toxic substrates/products/intermediates, and easy product purification (36,38,40,41,46,52).

Here, we report the development of a novel synthetic sRNA for modulating gene expression that expands the existing toolset by enabling gene overexpression in addition to expression knock-down. Our synthetic sRNA construct is composed of three parts: (i) a stem-loop stabilizer sequence, (ii) an antisense target binding sequence, and (iii) a terminator. The sRNA operates independently of Hfq, and, to our knowledge, its small size (48 nt total) allows multiple sRNAs to be contained on a single plasmid or genomic insert, and the antisense sequence is easily swappable by PCR. To demonstrate its function, the antisense portation of the sRNA was designed to target several predicted secondary structures in a fluorescent reporter, in addition to the RBS and 3’ end. The system was initially built and evaluated in *E. coli*, but here, we also demonstrate

that due to cell induction inconsistency, a simplified CFS provided more statistically significant results when testing the system. Most importantly, we were able to increase the expression level of the fluorescent protein reporter used in this study by more than 50% by occluding the target region of RNAse E. Mechanisms were learned and applied to diaphorase expression *in vitro* to show the system can be modified for nearly any target gene expression in a CFS or cell-free protein synthesis (CFPS) reaction.

3.2. Materials and methods

3.2.1. Bacterial strains and culture conditions

Escherichia coli 10-β (New England Biolabs; Ipswitch, MA) were used as the host expression system in all experiments and were cultured in LB growth medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride) supplemented with ampicillin to the final concentration of 100 μg/ml. Appropriate concentrations arabinose "Ara" and anhydrotetracycline "ATc" were used to induce protein and asRNA expression. *E coli* BL21 Star (DE3) (New England Biolabs) were used for diaphorase expression in CFS experiments. A description of all strains and plasmids used in this study are given in Table B1, and all PCR primer sequences are given in Table B2 of Supplementary Appendix B.

3.2.2. Plasmid construction and sRNA design

Three different plasmids were constructed for this study and are shown in Fig. B1 (and described in Table B1) in Supplementary Appendix B. The first plasmid, pART15-C1, was used as a reporter and contains *AmCyan* (cyan) fluorescent gene under the control of Ara-inducible P_{BAD} promoter. pART15-C2 was also used as a reporter to determine the expression levels of *AmCyan*

under the control of ATc-inducible $P_{LtetO-1}$. This was used to determine relative expression levels of sRNAs. pART15-C1-LacZ was used to substitute sRNA constructs for *lacZ* by inverse PCR. A blue/white screen enabled the identification of positive colonies after transformation. The final plasmids, versions of pART15-AS, contained the *AmCyan* gene under the control of P_{BAD} promoter and various sRNAs under the control of $P_{LtetO-1}$. This synthetic circuit is shown in Fig. B2. The construction of all necessary parts (origin of replication, promoters, terminators, RBSs, reporter gene) was performed by Gibson assembly® cloning (53) and λ-PCR (54).

The synthetic sRNA scaffold is shown in Fig. 3.1 and was designed to have three parts, described from 5' to 3': (i) a 14 nt synthetic stem-loop stabilizer, (ii) a 20 nt antisense sequence complementary to the target mRNA, and (iii) a 14 nt stem-loop transcriptional terminator. sRNA antisense sequences (Table B3) were designed to target the following regions of the *AmCyan* mRNA: (i) the ribosome binding site (RBS) of (antisense sequence referred to as "asRBS"), (ii) the first 20 bp of *AmCyan* from the 5' end (asB1), (iii) a potential high energy loop (asB30), (iv-v) potential low energy loops (asB57, asB180), (vi-viii) potential high energy stem structures (asB407, asB430, asB539), (ix-x) potential low energy stem structures (asB606, asB615), (xi-xii) near the 3' end (asB693, asB701), and (xiii) a control without a binding region, containing only the stabilizing region and terminator (asCTRL). Predictions of RNA secondary structures and energy regions were computed using NUPACK software (http://www.nupack.org) (55). All of the target binding sequences were blasted against the *E. coli* genome and pART15 plasmid to ensure no other interactions were anticipated.

3.2.3. Optimizing induction with pART15-C1 and pART15-C2

Positive colonies transformed with pART15-C1 and pART15-C2 were grown overnight at 37°C and 200 rpm in 5 mL of LB medium supplemented with 100 µg/ml ampicillin. Cells were diluted 100X into 200 µL of the same medium supplemented with inducers (0-100 mM Ara and 0-100 ng/mL ATc) in a 96-well plate. Colonies were grown for 24 h at 30°C and OD_{600} and fluorescence readings were measured in a Synergy H4 Hybrid Microplate Reader (BioTek; Winooski, VT). AmCyan fluorescence was determined at ex. 458 nm and em. 489 nm.

3.2.4. Determining sRNA influence on gene expression

Cells harboring pART15-AS constructs, containing different synthetic sRNAs (differing by antisense sequence), were tested by growing transformed cells overnight in LB media with ampicillin and diluting 100X in 200 µl of LB media supplemented with ampicillin and ATc (0 ng/ml or 100 ng/ml). After 6 h of growth, Ara (0 mM or 0.05 mM) was added, and cells were incubated for 24 h at 30°C in a 96-well plate. Cell growth was monitored by OD_{600} and fluorescence levels were measured at ex. 458 nm, em. 489 nm. Normalized fluorescence (F_N) was calculated by dividing the fluorescence measurement by the OD_{600} reading. Fluorescence change (F_Δ) was calculated by comparing the fluorescence of cells grown supplemented with both ATc and Ara with the positive control fluorescence (pART15-C1), using Eq 1:

$$F_\Delta = \left(F_{N,pART15-AS} - F_{N,pART15-C1}\right)/F_{N,pART15-C1} \times 100 \quad \text{(Eq. 1)}$$

3.2.5. Flow cytometry

Cells were grown overnight in 5 mL LB media with 100 µg/mL ampicillin at 37°C and 200 rpm. 10 µL of overnight culture was transferred to fresh 600 µL LB media (OD600 ≈ 0.05)

supplemented with appropriate antibiotic and inducers. After 8 h of growth at 37°C and 200 rpm, cells were centrifuged and resuspended in phosphate-buffered saline (PBS) for fluorescence measurements. Fluorescence (ex. 458 nm; em. 489 nm) was measured via flow cytometry using a BD Accuri® C6 flow cytometer (BD Biosciences; San Jose, CA), by collecting 10,000 events per replicate.

3.2.6. In vitro CFS

A pre-culture was prepared by growing cells overnight in 5 mL LB media supplemented with ampicillin at 37°C and 200 rpm. The overnight culture was diluted in 100 mL LB media and incubated overnight at 37°C and 200 rpm and then centrifuged. The cell pellet was resuspended in the lysis buffer (20 mM Na_2HPO_4, 50 mM NaCl, 10 mM imidazole and 2% triton). Cells were subsequently disrupted by sonication using Branson 450 Digital Sonifier® (40% amplitude, 3 min 30, 7 s pulse and 5 s OFF). Cell lysate was supplemented with the appropriate inducers and incubated for 96 h in a 96-well plate. OD_{600} and fluorescence readings (ex. 458 nm and em. 489 nm) were measured in a Synergy H4 Hybrid Microplate Reader (BioTek; Winooski, VT).

3.2.7. Diaphorase over-expression

Four plasmids expressing sRNAs targeting potential binding sites of RNase E in diaphorase (DI, GenBank accession number JQ040550) were constructed. pBAD-LIC was used as the backbone plasmid to clone the four antisense fragments (Table B3 in the Supplementary Appendix B): (i) asDI410, (ii) asDI490, (iii) asDI510 and (iv) asDI528. The recombinant pBAD-AS plasmids (pBAD-asDI410, pBAD-asDI490, pBAD-asDI510 and pBAD-asDI528) were transformed into *E.coli* BL21 Star (DE3). Plasmids and primers are summarized in Tables B1 and B2.

Separate *E. coli* BL21 (DE3) cultures, one harboring the diaphorase (56) and the other the recombinant pBAD-AS plasmids, were used to generate the *in vitro* CFS. The procedures of the *in vitro* CFS were followed as described in the previous section. Cultures were mixed in a 1:1 ratio (diaphorase to pBAD-AS), and the cell mixture was disrupted by sonication. Aliquots of the crude extract (200 µL) were used in triplicate in a 96-well plate in the presence of 0.5 mM IPTG to induce diaphorase expression and 5 mM Ara to induce sRNAs. The empty plasmid pBAD-LIC was used as the control. Bradford protein assay (57) was performed at time 0 and 72 hours to determine protein content in the *in vitro* CFS.

3.3. Results

3.3.1. Optimizing induction with pART15-C1 and pART15-C2 (in vivo system)

pART15-AS was built to characterize the effects of different antisense sequences in our synthetic sRNA on the expression level of the AmCyan (cyan) fluorescent reporter. Two control plasmids, pART15-C1 and pART15-C2 (described in the Materials and Methods section and in Fig. B1 and Table B1) were used to assess the expression of the sRNA and the reporter protein. Both control plasmids were first characterized to determine the working concentrations of both inducers. We assessed the susceptibility to modulation of both promoters P_{BAD} and $P_{Ltet\text{-}O1}$ by measuring AmCyan fluorescence at different levels of induction by Ara and ATc. The expression level of AmCyan with induction from 0-100 mM Ara is shown in Fig. B3. A linear relationship was observed between Ara concentration and fluorescence. Maximum fluorescence was obtained at an Ara concentration of 10 mM, and the response plateaued beyond this point. Expression from $P_{Ltet\text{-}O1}$ is an indirect measure of sRNA levels. A plot of AmCyan fluorescence as a function of ATc concentration (0-100 ng/mL) is also shown in Fig. B3, and a linear relationship was observed

over the induction range to 1 ng/mL, where only slight increases were observed at higher levels. The maximum fluorescence level (indicative of asRNA levels) was observed at 100 ng/mL. These results indicated that both parts of our system (fluorescent reporter and sRNA expression) could be modulated over a wide range of inducer concentrations. At the individual cell level, fluorescence intensities have been shown to vary over a broad range within populations growing at sub-saturating inducer concentrations (58). Thus, a set of Ara and ATc inducer levels was sought that yields high expression levels of both asRNA and AmCyan, where all cells were induced uniformly. Therefore, flow cytometry was employed to determine this optimum operating point of induction.

3.3.2. Flow cytometry (in vivo system)

Flow cytometry experiments were conducted to examine the induction homogeneity of cell populations under different concentrations of inducers. First, AmCyan expression in pART15-C1, under different concentrations of Ara, was examined. Flow cytometry analysis showed that induction with 0.05 mM Ara was sufficient to induce the entire population. This is shown in Fig. B4. Similarly, cells harboring pART15-C2 were induced with different levels of ATc, and a concentration of 100 ng/mL was observed to yield full and uniform induction of all cells (shown in Fig. B5).

3.3.3. Plate-reader assays (in vivo system)

After characterizing both parts of the synthetic circuit, all versions of pART15-AS (with the different sRNAs) were assessed. Cells transformed with the pART15-AS vectors were incubated in 96-well plate for 6 hours with 0.05 mM arabinose, then 100 ng/ml ATc was added. Normalized fluorescence (F_N) intensity and the fluorescence change (F_Δ) (Eq. 1) were calculated.

A total of 12 different antisense binding sequences (Table B3) were tested. Replicates for the sRNAs variants did not produce consistent results, with some replicates showing decreases in reporter fluorescence, while others with the same sRNA construct showed increases. Statistical analysis of fluorescence was performed using ANOVA tests and pairwise comparisons using Tukey's honest significant difference (HSD) method at three different time points (6 h, 18 h, and 29 h) with the pART15-AS constructs and the control pART15-C1, which contained no asRNA. A total of 64 replicates were run for each sRNA, and data were collected at multiple time points. All fluorescence data and full ANOVA tests and pairwise comparison results are included in Supplementary Appendix B, a graphical representation of fluorescence data is given in Fig. B6, and ANOVA results are summarized in Tables 3.1 and 3.2. Results for the *in vivo* system (Table 3.1) showed that targeting the RBS and first 20 bp of the *AmCyan* mRNA lead to statistically significant reductions in expression of more than 10% at 29 h. The approach of targeting translation initiation has been documented in the literature; although, our sRNA is not Hfq dependent, to our knowledge. Results also showed an even greater statistically significant expression knock-down of more than 35% at 29 h by targeting a high energy stem in the mRNA secondary structure (asB430). This sRNA also was able to achieve about 25% statistically significant knock-down by 6 h in the culture. Others targeting the translation initiation region did not show statistical significance until late in the culture. The antisense sequence of the asB430 sRNA was designed to bind nt 430-450 of the reporter mRNA. This is far from the translation initiation region. Targeting the high energy stem further (asB407 and asB539) produced decreases in fluorescence, with some results being statistically significant (e.g., -43% for asB407 at 18 h). Notably, we observed a statistically significant increase in fluorescence of more than 35% from asB606 at 18 h, but a statistically significant increase was not observed at 29 h. Pairwise comparisons testing was

applied where fluorescence results from all sRNAs at every time point were compared with every other to identify statistically significant differences. Of all comparisons made, only 7.7% were statistically different. Thus, despite the significant findings described above, most of the sRNA constructs varied (some widely) in fluorescence readings, and it has been suggested in the literature that the extreme complexity of living cells may play a role (35). In addition, the objective of microbial cells is growth, not the production of target products, which is usually opposed to synthetic biology and metabolic engineering goals (59). These results were investigated further by observing non-uniform induction of the cultures through flow cytometry when using both the Ara and ATc inducers. This is shown in Fig. B7 and demonstrates induction heterogeneity that differed given different sRNAs and even between replicates. The mechanisms leading to this behavior remains unknown; however, a CFS was used to repeat these experiments, since cell-to-cell induction variations would be negligible.

3.3.4. Implementation of a CFS (in vitro system)

To facilitate the analysis of the synthetic asRNAs and to overcome the induction heterogeneity of implementation *in vivo*, we constructed a very simple CFS system. It consisted of cells harboring the pART15 plasmids, which were harvested and sonicated to disrupt cell membranes and liberate the cell contents. Optimization of the system parameters (activity, sonication time, and volume) are shown in Figs. B8-B12. In order to test the induction linearity in the the CFS, the fluorescence response to Ara and ATc inducers was measured as before with pART15-C1 and pART15-C2 and the same inducer levels (0-100 mM for Ara; and 0-100 ng/mL for ATc). Again, linear relationships were obtained for the CFS (Fig. B13), and the same optimal concentrations of of Ara (0.05 mM) and ATc (100 ng/mL) were identified for pART15-AS and the CFS. The OD_{600} of the lysate following sonication remained constant, as shown in Fig. B14.

3.3.5. Plate-reader assays (in vitro system)

Sonicated cells previously transformed with pART15-AS constructs were induced *in vitro* with 0.05 mM Ara and 100 ng/ml ATc and incubated for 96 h until the fluorescence reached a constant level. With this simplified CFS, the system required 24 h for induction (the point where fluorescence appeared), and the maximum was achieved by 96 h. All fluorescence readings were averaged and normalized by the OD_{600}. Unlike the *in vivo* system, fluorescence data with pART15-AS constructs were more consistent, and these data are shown in Fig. B15 and are given with statistical test results in Supplementary Appendix B. Three time points were chosen to analyze the data (48 h, 72 h, and 96 h). The CFS produced many more statistically significant results than did the *in vivo* system. In fact, 44.6% of all pairwise comparisons were found statistically significant ($p < 0.05$). This was a stark increase from the 7.7% of the *in vivo* system, and fluorescence results seemed far more consistent. ANOVA test results are given in Table 3.2. Targeting the RBS did not produce a statistically significant decrease in fluorescence in the CFS, but this was observed when targeting the first 20 nt and repression levels were higher than the *in vivo* system. However, targeting other secondary structures produced far superior repression. For example, targeting the high energy stem produced greater than 65% (statistically significant) fluorescence reduction, and targeting the 3' end of the mRNA (asB701) lead to more than 70% fluorescence reduction, compared to about 19% reduction in the *in vivo* system (both statistically significant). However, we found that selectively targeting the low energy stem (asB606) lead to a statistically significant increase in fluorescence of more than 66% at 48 and 96 h (it was more than 50% at 48 h). This is the first time, to our knowledge, that a synthetic sRNA produced an increase in target protein expression level by targeting an mRNA secondary structure, in this case a low energy stem near the 3'end of the mRNA. We attempted to target another portion of this low energy stem loop

(asB615) in the CFS and observed statistically significant decreases in fluorescence ranging from 30-50%. In addition, it was sought whether existing sRNA scaffolds that recruit Hfq would produce similar results. We chose the MicC scaffold from previous published studies (23,60–62). MicC is a naturally occurring small RNA involved in the repression of the *E.coli* outer membrane protein, OmpC (63). MicC requires the Hfq RNA chaperone for its function (63,64), and studies have shown its superior repression capabilities when compared to other scaffolds (23). In fact, there is a high frequency of RNase E mediated cleavages in Hfq-dependent asRNAs (65–68). Using the MicC scaffold, we added the same antisense binding regions used in asRBS (targets the mRNA RBS), asB1 (targets the first 20 nt of the mRNA) and asB606 (targets the low energy stem loop of the reporter mRNA starting at nt 606). Results are also given in Table 3.2. Targeting the RBS with our MicC-asRBS sRNA lead to an increase in fluorescence expression, which is contrary to results published for other *in vivo* systems. A possible mechanism for this is discussed later. However, targeting the first 20 nt with the MicC-asB1 lead to a statistically significant reduction in fluorescence of better than 40%. This was greater repression than the 20% (and not statistically significant at 96 h) achieved when using our non-Hfq-dependent synthetic scaffold. Using the MicC scaffold with asB606 (MicC-asB606) resulted in virtually no change in fluorescence at 72 and 96 h (with significant increase observed at 48 h) relative to the control; whereas, our synthetic scaffold produced an increase in fluorescence of over 65%. Thus, a mechanism was sought to explain the overexpression observation by our asB606 sRNA.

3.3.6. Probing the mechanism of asB606 with NUPACK and catRAPID

To determine the mechanism behind asB606 sRNA mediated overexpression, analysis with the NUPACK thermodynamic software (55) was performed. Comparison of all sRNAs aligning and binding with *AmCyan* mRNA (shown in Fig. B16 and Fig. B17 for MicC constructs) showed

a non-significant change in free energy of binding. NUPACK analyses allowed the analysis of binding probability when an asRNA is bound to its target region. *AmCyan* mRNA served as a template to compare changes of synthetic sRNA antisense-mRNA pairing probability. Only the last 200 nt of the *AmCyan* mRNA were considered for this analysis since there were negligible changes in pairing when also including the first 400 nt. Comparison of pair probability (Fig. B18 and Fig. B19) revealed that asB606 binds to base pairs 606-615 with the sequence "CCAAACCAUG." Further analysis revealed this site contains a potential consensus binding site for RNase E (65), which is the most important endonuclease initiating mRNA decay in *E. coli* (66,69–72). RNase E is known to play a role in regulation of gene expression through transcript degradation (65–67,71), and in its first step, RNase E recognizes a cleavage site frequently found upstream of a stem-loop protected 3'end. Following recognition, an endonucleolytic cut of the mRNA is performed (67,69,73,74). Here, the particular site asB606 targets in the reporter mRNA is a region near a stem-loop and, when bound, it likely obscurs RNase E recognition. This would extend the half-life of *AmCyan* mRNA in the CFS, resulting in increased translation and expression level of the fluorescent reporter.

To probe this mechanism further, more analyses were conducted with catRAPID software (75–77), an accurate prediction tool for protein-RNA interaction propensities. *AmCyan* mRNA and the catalytic and binding regions of RNase E were interacted in catRAPID, and results showed strong binding potential between RNAse E and *AmCyan* mRNA in the 500-650 bp region (shown in Fig. B20), providing more evidence sRNA asB606 occludes a potential recognition site of RNase E.

With NUPACK analyses, we also probed why MicC-asRBS lead to statistically significant increases in reporter fluorescence, when a knock-down was hypothesized based on observations

in the literature with other reporter genes. The NUPACK predicted *AmCyan* mRNA secondary structure is included as Fig. B16a, and the structure of this bound with the sRNA MicC-asRNA is shown in Fig. B17d. A comparison of these structures suggests the native *AmCyan* mRNA 5' end is sequestered in a stem configuration, and this is alleviated by binding the MicC-asRNA sRNA. This also suggests MicC-asRNA does not occlude the RBS or an alternative RBS is liberated by unfolding the stem loop structure.

3.3.7. Implementation for diaphorase production in vitro

Diaphorase is a soluble NAD(P)H dehydrogenase (EC 1.6.99.1 or EC 1.6.99.3). It is widely applied in synthetic enzymatic pathways for the production of high-value chemicals (48,56,78,79). Therefore, overexpression of this enzyme as part of a synthetic pathway is of interest. Previous results with the sRNA asB606 showed that occluding the RNase E binding site increased the expression of a fluorescent reporter by 50-80% in a CFS. Here, this design was adapted and implemented for diaphorase expression in a CFS. Potential binding sites for the RNase E were predicted by catRAPID (75–77) (shown in Fig. B21), and four targets were identified (20 nt sequences starting at 410, 490, 510, and 528). These are close to stem loop structures and are in the predicted interaction region(s) for diaphorase mRNA and RNase E (Fig. B21). Plasmids (pBAD-AS) harboring asRNAs targeting those regions were constructed (pBAD-asDI410, pBAD-asDI490, pBAD-asDI510, and pBAD-asDI528). Descriptions, primers, and antisense sequences are given in Tables B1-B3. Next, an *in vitro* diaphorase-CFS was developed by combining cells expressing diaphorase and cells transformed with one of the pBAD-AS plasmids in the same sonication reaction. Diaphorase overexpression was observed in the presence of multiple sRNAs, as shown in Fig. 3.2. About 30% increase in expression was observed for asDI410 replicate 1 and about 50% for the replicate 2 (separate colonies). The induction of asDI490 replicate 2 and

asDI510 improved the expression level of diaphorase by about 66% and 75%, respectively. However, no increase was observed with asDI490 replicate 1, and asDI510 showed a slight increase (3.5%). This result might be due to many factors including: (i) the presence of multiple sites for RNase E binding (unlike *AmCyan* mRNA which contained a single site) and (ii) the diapohorasae-expressing strain was *E. coli* BL21, which can metabolize Ara, lowering the induction of sRNAs. However, results suggest the sRNAs described here can be adapted over expression of multiple mRNAs in CFSs and CFPSs.

3.4. Discussion

Natural and synthetic sRNAs have many potential applications in synthetic biology and metabolic engineering. Synthetic sRNAs are small (allowing several to be expressed from a single plasmid or genomic insertion), their expression can be inducible if desired, they can be "fine tuned" according to binding thermodynamics, and they can modulate gene expression where gene knockout is lethal and conditional gene silencing is required. Here, we have shown, for the first time to our knowledge, a synthetic sRNA can also be used to selectively enhance target protein expression. In this study, a new synthetic scaffold was designed to support 20 nt antisense sequences that bind target mRNAs, and several sequences were designed to target different regions and secondary structures of an *AmCyan* fluorescent reporter mRNA. An unexpected complication was encountered where few of our synthetic sRNAs gave consistent (or statistically significant) results when tested *in vivo*. These experiments highlighted the inherent variability and heterogeneity of living cultures, particularly in response to multiple inducers. Furthermore, for cells to thrive to grow and divide is contrary to our objective to express a foreign protein or execute a synthetic circuit (52,59). This is supported further, as the burden of expressing a heterologous gene is imposed by competition between the target and host cell genes for limited cellular resources

(e.g., transcriptional/translational machinery, metabolites, and co-factors) (80,81). Thus, it remains unclear to us whether this phenomenon is primarily confined to our system or is present in other synthetic circuit designs executed *in vivo*. Nevertheless, we presented methods on how this can be checked using flow cytometry and statistical tests, and have implemented our synthetic sRNA circuit in a simple and inexpensive CFS.

Executions of experiments in our simple CFS came at the expense of time (96 h for expression *in vitro* compared to 18 h for *in vivo*). With addition of transcriptional/translational machinery, nucleotides, and energy buffers (among others), researchers have been able to improve yields and reduce protein synthesis times in a CFS from days to hours (40,82,83). This, however, comes at the expense of reagents, and our simplified CFS that used cell lysates supplemented with LB medium served as a viable option for evaluating our synthetic sRNAs and produced statistically significant results. After characterizing this new system, the ability of sRNAs to regulate the fluorescence level was demonstrated. Downregulation of the fluorescence level was observed with multiple sRNAs that targeted different secondary structures of the mRNA, the RBS, the first 20 bp, and the 3' end. It was observed that the modulation of expression can be performed either by changing the target region and/or the incubation time. Most importantly, AmCyan overexpression was observed with asB606. This result was observed but not statistically significant with *in vivo* expression, but the effect was amplified and found to be statistically significant when using our simplified CFS. Furthermore, this effect disappeared when switching to a Hfq-dependent scaffold.

An understanding of the asB606 mechanism showed that this particular asRNA likely promoted overexpression by occluding RNase E, which serves to inhibit mRNA degradation. This effectively extends its half-life, enabling increased translation. Importantly, this mechanism is likely predictable using NUPACK and catRAPID software. We tested this with the expression of

diaphorase *in vitro* and found increased expression relative to the wild-type. We anticipate this mechanism can be used in CFPS reactions to further increase yields. There is strong interest in utilizing these systems to produce enzymes for catalysis and *in vitro* metabolic engineering (46,56,78,79). The translation of our sRNAs from improving fluorescent reporter expression to diaphorase shows promise for this system to be applied universally in CFSs. Further improvements are required for reliable function of our synthetic sRNAs *in vivo*, yet it remains unknown how the system is leading to heterogeneous, and somewhat randomized, induction. It is recognized that this phenomenon can be inherent of analyzing some synthetic circuit *in vivo*, making the simplified CFS used here a reliable alternative.

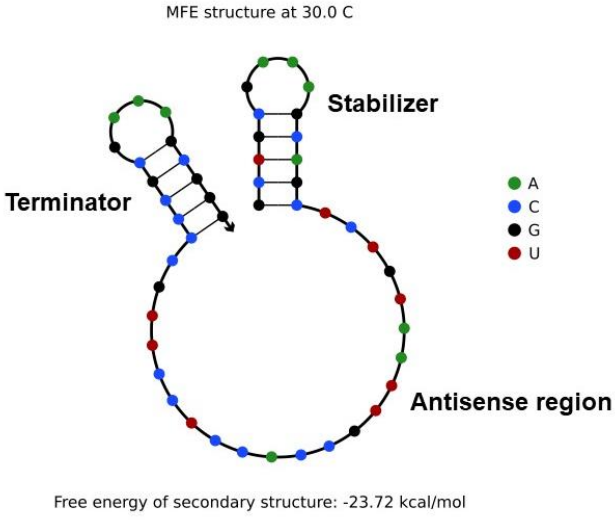


Figure 3.1. Synthetic sRNA scaffold structure containing a 14 nt synthetic stem-loop as stabilizing sequence, a 20 nt binding sequence complementary to the target region, and a 14 nt stem-loop transcriptional terminator. The structure shown was exported from NUPACK.

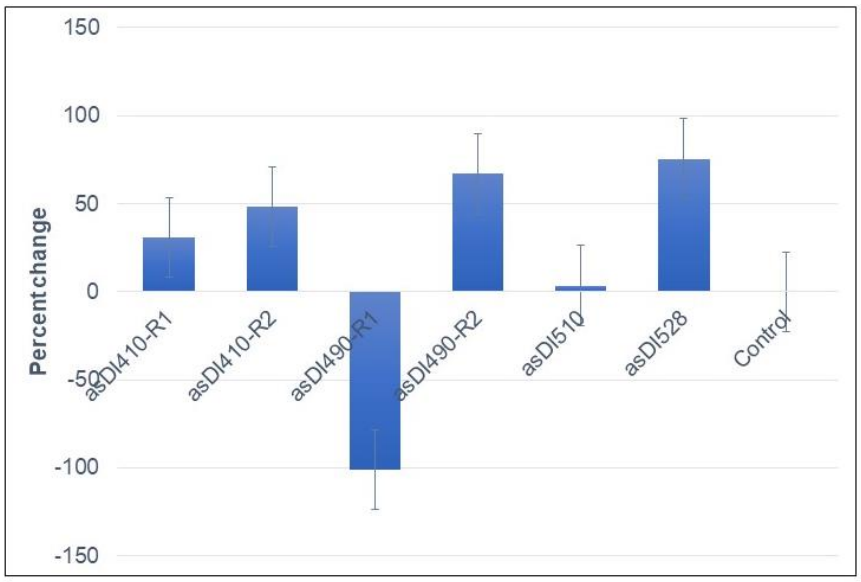


Figure 3.2. Percent changes in diaphorase expression in the presence of sRNAs containing the labeled antisense fragments. Each data point is the average of three biological replicates.

Table 3.1. ANOVA test results for *in vivo* experiments with pART15-AS.

		6 h (64 replicates)		18 h (64 replicates)		29 h (64 replicates)	
Antisense Sequence	Target Description	ANOVA Result ($p < 0.05$)	Average Response	ANOVA Result ($p < 0.05$)	Average Response	ANOVA Result ($p < 0.05$)	Average Response
asCTRL	sRNA without antisense region	Not Significant	-16.7%	Not Significant	-11.8%	Not Significant	-4.6%
asRBS	RBS	Not Significant	-6.3%	Not Significant	+4.8%	Significant	-12.3%
asB1	First 20 bp	Not Significant	-17.9%	Not Significant	+5.0%	Significant	-15.6%
asB30	High energy loop	Not Significant	-8.8%	Not Significant	-12.6%	Not Significant	-6.3%
asB57	Low energy loop	Not Significant	+5.3%	Not Significant	-23.3%	Significant	-25.9%
asB180	Low energy loop	Significant	-22.4%	Not Significant	+7.4%	Not Significant	-7.4%
asB407	High energy stem	Not Significant	+2.1%	Significant	-43.1%	Not Significant	-16.6%
asB430	High energy stem	Significant	-25.5%	Not Significant	-16.9%	Significant	-35.5%
asB539	High energy stem	Not Significant	+0.8%	Not Significant	-11.7%	Significant	-18.3%
asB606	Low energy stem	Significant	-22.2%	Significant	+36.7%	Not Significant	+ 7.4%
asB615	Low energy stem	Not Significant	+3.0%	Significant	-19.8%	Not Significant	-17.2%
asB693	3' end	Significant	-31.0%	Not Significant	+14.2%	Not Significant	-9.1%
asB701	3’ end	Not Significant	+7.0%	Not Significant	-20.5%	Significant	-19.4%

* The average response is the percent increase or decrease in fluorescence observed relative to pART15-C1.

** The asCTRL contains the stabilizer stem loop and terminator but no antisense sequence.

Table 3.2. ANOVA test results for *in vitro* CFS experiments with pART15-AS.

		48 h (30 replicates)		**72 h (30 replicates)**		**96 h (30 replicates)**	
Antisense Sequence	Target Description	ANOVA Result ($p < 0.05$)	Average Response	ANOVA Result ($p < 0.05$)	Average Response	ANOVA Result ($p < 0.05$)	Average Response
asCTRL	sRNA without antisense region	Not Significant	-8.7%	Not Significant	+0.8%	Not Significant	-2.8%
asRBS	RBS	Not Significant	-8.5%	Not Significant	+1.1%	Not Significant	+8.6%
asB1	First 20 bp	Significant	-22.1%	Significant	-17.1%	Not Significant	-11.3%
asB57	Low energy loop	Significant	-32.3%	Significant	-48.0%	Significant	-41.9%
asB30	High energy loop	Not Significant	-2.1%	Not Significant	+6.0%	Not Significant	+5.1%
asB180	Low energy loop	Not Significant	-10.1%	Not Significant	-9.3%	Not Significant	-11.0%
asB407	High energy stem	Significant	-39.1%	Significant	-39.9%	Significant	-38.2%
asB430	High energy stem	Not Significant	+19.7%	Not Significant	+14.9%	Not Significant	+9.3%
asB539	High energy stem	Significant	-49.0%	Significant	-65.1%	Significant	-65.8%
asB606	Low energy stem	Significant	+52.4%	Significant	+66.0%	Significant	+66.1%
asB615	Low energy stem	Significant	-32.7%	Significant	-58.1%	Significant	-34.0%
asB693	3' end	Not Significant	-10.5%	Not Significant	-6.4%	Not Significant	-3.4%
asB701	3’ end	Significant	-38.5%	Significant	-72.0%	Significant	-71.0%
MicC-asRBS	RBS with Hfq	Significant	+35.2%	Significant	+23.2%	Significant	+17.5%
MicC-asB1	First 20 bp with Hfq	Not Significant	-0.2%	Significant	-34.6%	Significant	-42.3%
MicC-asB606	Low energy stem with Hfq	Significant	+29.5%	Not Significant	0.0%	Not Significant	+3.7%

* The average response is the percent increase or decrease in fluorescence observed relative to pART15-C1.

** The asCTRL contains the stabilizer stem loop and terminator but no antisense sequence.

References

1. Hussein R, Lim HN. Disruption of small RNA signaling caused by competition for Hfq. Proc Natl Acad Sci U A. 2011;108(3):1110–5.
2. Altuvia S. Identification of bacterial small non-coding RNAs: experimental approaches. Curr Opin Microbiol. 2007 Jun;10(3):257–61.
3. Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. Proc Natl Acad Sci U A. 2011;108(21):8617–22.
4. Chappell J, Takahashi MK, Lucks JB. Creating small transcription activating RNAs. Nat Chem Biol. 2015;11(3):214–20.
5. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013;152(5):1173–83.
6. Green AA, Silver PA, Collins JJ, Yin P. Toehold switches: de-novo-designed regulators of gene expression. Cell. 2014;159(4):925–39.
7. Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW, et al. Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. Cell. 2016;165(5):1255–66.
8. Gottesman S. The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annu Rev Microbiol. 2004;58:303–28.
9. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell. 2011;43(6):880–91.
10. Nakashima N, Tamura T. Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. Nucleic Acids Res. 2009;37(15):e103.
11. Kim JY, Cha HJ. Down-regulation of acetate pathway through antisense strategy in *Escherichia coli*: improved foreign protein production. Biotechnol Bioeng. 2003;83(7):841–53.
12. Chen H, Ferbeyre G, Cedergren R. Efficient hammerhead ribozyme and antisense RNA targeting in a slow ribosome *Escherichia coli* mutant. Nat Biotechnol. 1997;15(5):432–5.
13. Desai RP, Papoutsakis ET. Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. Appl Env Microbiol. 1999;65(3):936–45.
14. Engdahl HM, Lindell M, Wagner EG. Introduction of an RNA stability element at the 5'-end of an antisense RNA cassette increases the inhibition of target RNA translation. Antisense Nucleic Acid Drug Dev. 2001;11(1):29–40.
15. Ji Y, Yin D, Fox B, Holmes DJ, Payne D, Rosenberg M. Validation of antibacterial mechanism of action using regulated antisense RNA expression in Staphylococcus aureus. FEMS Microbiol Lett. 2004;231(2):177–84.
16. Tchurikov NA, Chistyakova LG, Zavilgelsky GB, Manukhov IV, Chernov BK, Golova YB. Gene-specific silencing by expression of parallel complementary RNA in *Escherichia coli*. J Biol Chem. 2000;275(34):26523–9.
17. Nakashima N, Tamura T, Good L. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in *Escherichia coli*. Nucleic Acids Res. 2006;34(20):e138.
18. Dryselius R, Aswasti SK, Rajarao GK, Nielsen PE, Good L. The translation start codon region is sensitive to antisense PNA inhibition in Escherichia coli. Oligonucleotides. 2003;13(6):427–33.

19. Breaker RR. Prospects for riboswitch discovery and analysis. Mol Cell. 2011;43(6):867–79.
20. Carothers JM, Goler JA, Juminaga D, Keasling JD. Model-driven engineering of RNA devices to quantitatively program gene expression. Science. 2011;334(6063):1716–9.
21. Carrier TA, Keasling JD. Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. Biotechnol Prog. 1999;15(1):58–64.
22. Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol. 2006;24(8):1027–32.
23. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotechnol. 2013;31(2):170–4.
24. Engdahl HM, Hjalt TA, Wagner EG. A two unit antisense RNA cassette test system for silencing of target genes. Nucleic Acids Res. 1997;25(16):3218–27.
25. Yin D, Ji Y. Genomic analysis using conditional phenotypes generated by antisense RNA. Curr Opin Microbiol. 2002;5(3):330–3.
26. Win MN, Smolke CD. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U A. 2007;104(36):14283–8.
27. Callura JM, Dwyer DJ, Isaacs FJ, Cantor CR, Collins JJ. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. Proc Natl Acad Sci U A. 2010;107(36):15898–903.
28. Yang Y, Lin Y, Li L, Linhardt RJ, Yan Y. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. Metab Eng. 2015;29:217–26.
29. Solomon KV, Sanders TM, Prather KL. A dynamic metabolite valve for the control of central carbon metabolism. Metab Eng. 2012;14(6):661–71.
30. Tummala SB, Welker NE, Papoutsakis ET. Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. J Bacteriol. 2003;185(6):1923–34.
31. Sillers R, Al-Hinai MA, Papoutsakis ET. Aldehyde-alcohol dehydrogenase and/or thiolase overexpression coupled with CoA transferase downregulation lead to higher alcohol titers and selectivity in *Clostridium acetobutylicum* fermentations. Biotechnol Bioeng. 2009;102(1):38–49.
32. Tummala SB, Junne SG, Papoutsakis ET. Antisense RNA downregulation of coenzyme A transferase combined with alcohol-aldehyde dehydrogenase overexpression leads to predominantly alcohologenic *Clostridium acetobutylicum* fermentations. J Bacteriol. 2003;185(12):3644–53.
33. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. 2013;41(15):7429–37.
34. Lee YJ, Hoynes-O'Connor A, Leong MC, Moon TS. Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. Nucleic Acids Res. 2016;44(5):2462–73.
35. Lu Y. Cell-free synthetic biology: Engineering in an open world. Synth Syst Biotechnol. 2017;2(1):23–7.
36. Dudley QM, Karim AS, Jewett MC. Cell-free metabolic engineering: biomanufacturing beyond the cell. Biotechnol J. 2015;10(1):69–82.
37. Lentini R, Forlin M, Martini L, Del Bianco C, Spencer AC, Torino D, et al. Fluorescent Proteins and in Vitro Genetic Organization for Cell-Free Synthetic Biology. ACS Synth Biol. 2013;2(9):482–9.

38. Harris DC, Jewett MC. Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. Curr Opin Biotechnol. 2012;23(5):672–8.
39. Bujara M, Panke S. Engineering in complex systems. Curr Opin Biotechnol. 2010;21(5):586–91.
40. Swartz JR. Transforming biochemical engineering with cell-free biology. AIChE J. 2012;58(1):5–13.
41. Carlson ED, Gan R, Hodgman CE, Jewett MC. Cell-free protein synthesis: Applications come of age. Biotechnol Adv. 2012;30(5):1185–94.
42. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, et al. Cell-free translation reconstituted with purified components. Nat Biotechnol. 2001;19:751.
43. Marshall R, Maxwell CS, Collins SP, Jacobsen T, Luo ML, Begemann MB, et al. Rapid and Scalable Characterization of CRISPR Technologies Using an *E. coli* Cell-Free Transcription-Translation System. Mol Cell. 2018;69(1):146-157.e3.
44. Garamella J, Marshall R, Rustad M, Noireaux V. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. ACS Synth Biol. 2016;5(4):344–55.
45. Lavickova B, Maerkl SJ. A Simple, Robust, and Low-Cost Method To Produce the PURE Cell-Free System. ACS Synth Biol. 2019;
46. Zhang Y-HP. Production of biofuels and biochemicals by in vitro synthetic biosystems: Opportunities and challenges. Biotechnol Adv. 2015;33(7):1467–83.
47. Guterl JK, Garbe D, Carsten J, Steffler F, Sommer B, Reisse S, et al. Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. ChemSusChem. 2012;5(11):2165–72.
48. Zhang YH, Evans BR, Mielenz JR, Hopkins RC, Adams MW. High-yield hydrogen production from starch and water by a synthetic enzymatic pathway. PLoS One. 2007;2(5):e456.
49. Bujara M, Schumperli M, Billerbeck S, Heinemann M, Panke S. Exploiting cell-free systems: Implementation and debugging of a system of biotransformations. Biotechnol Bioeng. 2010;106(3):376–89.
50. Karim AS, Jewett MC. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. Metab Eng. 2016;36:116–26.
51. Karim AS, Heggestad JT, Crowe SA, Jewett MC. Controlling cell-free metabolism through physiochemical perturbations. Metab Eng. 2018;45:86–94.
52. Hodgman CE, Jewett MC. Cell-free synthetic biology: Thinking outside the cell. Metab Eng. 2012;14(3):261–9.
53. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison Iii CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 online;6:343.
54. Tanniche I, Fisher AK, Gillam F, Collakova E, Zhang C, Bevan DR, et al. λ-PCR for precise DNA assembly and modification. 2019;(Under Review).
55. Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, et al. NUPACK: Analysis and design of nucleic acid systems. J Comput Chem. 2011;32(1):170–3.
56. Collins J, Zhang T, Huston S, Sun F, Zhang Y-HP, Fu J. A Hidden Transhydrogen Activity of a FMN-Bound Diaphorase under Anaerobic Conditions. PLOS ONE. 2016;11(5):e0154865.
57. Kruger NJ. The Bradford method for protein quantitation. In: The protein protocols handbook. Springer; 2009. p. 17–24.

58. Siegele DA, Hu JC. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. Proc Natl Acad Sci U S A. 1997;94(15):8168–72.
59. Nielsen J, Keasling JD. Engineering Cellular Metabolism. Cell. 2016;164(6):1185–97.
60. Yoo SM, Na D, Lee SY. Design and use of synthetic regulatory small RNAs to control gene expression in *Escherichia coli*. Nat Protoc. 2013 Sep;8(9):1694–707.
61. Cho C, Lee SY. Efficient gene knockdown in *Clostridium acetobutylicum* by synthetic small regulatory RNAs. Biotechnol Bioeng. 2017;114(2):374–83.
62. Sun D, Chen J, Wang Y, Li M, Rao D, Guo Y, et al. Metabolic engineering of *Corynebacterium glutamicum* by synthetic small regulatory RNAs. J Ind Microbiol Biotechnol. 2019;46(2):203–8.
63. Chen S, Zhang A, Blyn LB, Storz G. MicC, a Second Small-RNA Regulator of Omp Protein Expression in *Escherichia coli*. J Bacteriol. 2004 Oct 15;186(20):6689–97.
64. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol. 2003;50(4):1111–24.
65. Chao Y, Li L, Girodat D, Förstner KU, Said N, Corcoran C, et al. In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. Mol Cell. 2017;65(1):39–51.
66. Mackie GA. RNase E: at the interface of bacterial RNA processing and decay. Nat Rev Microbiol. 2012;11:45.
67. Morita T, Aiba H. RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. Genes Dev. 2011;25(4):294–8.
68. Bandyra KJ, Said N, Pfeiffer V, Górna MW, Vogel J, Luisi BF. The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. Mol Cell. 2012;47(6):943–53.
69. Klug G, Rauhut R. mRNA degradation in bacteria. FEMS Microbiol Rev. 1999;23(3):353–70.
70. Steege DA. Emerging features of mRNA decay in bacteria. Rna. 2000 Aug;6(8):1079–90.
71. Bandyra KJ, Wandzik JM, Luisi BF. Substrate Recognition and Autoinhibition in the Central Ribonuclease RNase E. Mol Cell. 2018;72(2):275-285.e4.
72. Laalami S, Putzer H. mRNA degradation and maturation in prokaryotes: the global players. Biomol Concepts. 2011;2(6):491–506.
73. McDowall KJ, Kaberdin VR, Wu S-W, Cohen SN, Lin-Chao S. Site-specific RNase E cleavage of oligonucleotides and inhibition by stem–loops. Nature. 1995;374(6519):287–90.
74. Jiang X, Diwa A, Belasco JG. Regions of RNase E important for 5'-end-dependent RNA cleavage and autoregulated synthesis. J Bacteriol. 2000;182(9):2468–75.
75. Bellucci M, Agostini F, Masin M, Tartaglia GG. Predicting protein associations with long noncoding RNAs. Nat Methods. 2011;8(6):444–5.
76. Agostini F, Zanzoni A, Klus P, Marchese D, Cirillo D, Tartaglia GG. catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. Bioinformatics. 2013;29(22):2928–30.
77. Livi CM, Klus P, Delli Ponti R, Tartaglia GG. catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. Bioinformatics. 2016;32(5):773–5.
78. Zhu Z, Zhang Y-HP. Use of nonimmobilized enzymes and mediators achieved high power densities in closed biobatteries. Energy Sci Eng. 2015;3(5):490–7.

79. Zhu Z, Sun F, Zhang X, Zhang Y-HP. Deep oxidation of glucose in enzymatic fuel cells through a synthetic enzymatic pathway containing a cascade of two thermostable dehydrogenases. Biosens Bioelectron. 2012;36(1):110–5.
80. Borkowski O, Bricio C, Murgiano M, Rothschild-Mancinelli B, Stan G-B, Ellis T. Cell-free prediction of protein expression costs for growing cells. Nat Commun. 2018;9(1):1457.
81. Garcia-Ramallo E, Carbonell-Ballestero M, Montañez R, Rodriguez-Caso C, Macía J. Dealing with the genetic load in bacterial synthetic biology circuits: convergences with the Ohm's law. Nucleic Acids Res. 2015;44(1):496–507.
82. Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR. An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol. 2008;4:220.
83. Swartz JR, Jewett MC, Woodrow KA. Cell-Free Protein Synthesis With Prokaryotic Combined Transcription-Translation. In: Balbás P, Lorence A, editors. Recombinant Gene Expression: Reviews and Protocols. Totowa, NJ: Humana Press; 2004. p. 169–82. (Methods in Molecular Biology).

Chapter 4: Characterizing *Synechocystis* sp. PCC 6803 with Rametrix™: Glucose, illumination, and nitrogen deprivation-induced phenotypes

Imen Tanniche[1], Eva Collakova[2], Cynthia Denbow[2], Ryan S. Senger*[1,3]

[1]Department of Biological Systems Engineering; Virginia Tech; Blacksburg, VA

[2]School of Plant & Environmental Sciences; Virginia Tech; Blacksburg, VA

[3]Department of Chemical Engineering; Virginia Tech; Blacksburg, VA

* Corresponding Author
1230 Washington St.
301C HABB1
Blacksburg, VA 24061
Email: senger@vt.edu
Phone: 540-231-9501

Characterizing *Synechocystis* sp. PCC 6803 with Rametrix™: Glucose, illumination, and nitrogen deprivation-induced phenotypes

Abstract

Background: *Synechocystis sp.* PCC 6803 is a model cyanobacterium that has been studied widely and is considered for metabolic engineering applications. Here, Raman spectroscopy and Raman chemometrics (Rametrix™) were used to (i) study broad phenotypic changes in response to growth conditions, (ii) identify phenotypic changes associated with its circadian rhythm, and (iii) correlate individual Raman bands with biomolecules and verify these with more accepted analytical methods.

Methods: *Synechocystis* cultures were grown under the following conditions: (i) autotrophic, (ii) heterotrophic (with 5 mM glucose), (iii) photoautotrophic, (iv) photoheterotrophic, (v) nitrogen deprivation, (vi) dark heterotrophic, and (vii) dark autotrophic (control). The Rametrix™ LITE Toolbox for MATLAB® was used to process Raman spectra and perform principal component analysis (PCA) and discriminant analysis of principal components (DAPC). The Rametrix™ PRO Toolbox was used to validate these models through leave-one-out routines that allowed calculation of prediction accuracy, sensitivity, and selectivity for an unkown Raman spectrum. Raman spectra were also subjected to statistical tests (ANOVA and pairwise comparisons) to identify statistically relevant changes in *Synechocystis* phenotypes. Finally, experimental methods, including widely used analytical and spectroscopic assays were used to quantify glycogen, fatty acids, amino acids, and chlorophyll a for correlations with Raman data.

Results: PCA and DAPC models produced distinct clustering of Raman spectra, representing multiple *Synechocystis* phenotypes, based on (i) growth in the presence of glucose, (ii) illumination (particularly continuous light), (iii) nitrogen deprivation, and (iv) throughout a circadian rhythm growth cycle. Rametrix^{TM} PRO was successful in identifying glucose-induced phenotypes in unkown Raman spectra, with 95.3% accuracy, 93.4% sensitivity, and 96.9% selectivity. Prediction accuracy was above random chance values for all other studies. Circadian rhythm analysis showed a return to the initial phenotype after 24 hours for cultures grown in light/dark (12h/12h) cycles; this did not occur for cultures grown in the dark. Finally, correlation coefficients ($R > 0.7$) were found for glycogen, all amino acids, and chlorophyll a when comparing specific Raman bands to other experimental results.

4.1. Introduction

Cyanobacteria constitute a widespread group of Gram-negative bacteria (1,2), and they are the only prokaryotes capable of plant-like oxygenic photosynthesis (3). Among the diverse cyanobacterial strains, *Synechocystis sp.* PCC6803 is the most extensively studied model organism (4). It has been genetically well characterized and its genome fully sequenced (5), annotated, and available at CyanoBase (6). In addition, biochemical and physiological data are well established (7), and a "genome-scale" metabolic model has been produced (8). The ability of *Synechocystis* to grow photoautotrophically (fast growth compared to plants), and the ease and high efficiency of genetic engineering make this blue-green algae an attractive candidate for biotechnological and metabolic engineering applications (3,7). The most interesting are the sustainable biosynthesis of commodity chemicals (e.g. isoprene, ethylene), biofuels (e.g. ethanol, butane, fatty acids and fatty alcohols), biomaterials, and health-related compounds (3,9,10). *Synechocystis sp.* PCC 6803 has no specific nutritional demands as it has the ability to adapt different growth modes, going from fully photoautotrophic, in the absence of an added fixed-carbon source, to fully heterotrophic (11), and even chemoheterotrophic when exposed to short periods of blue-light (12). The oxygenic photosynthesis is achieved mainly by two multi-subunit complexes, photosystem I and photosystem II, which are embedded in the thylakoid membrane (13) and allow adaptation to daily fluctuations in nutrients and light. As a result, a circadian rhythm is observed in *Synechocystis sp.* PCC6803 with an approximate period length of 24 h in response to daily environmental changes (14). This oscillatory behavior has been well-studied and showed that this cyanobacterium carries out photosynthesis and glycogen synthesis in the light and respiration and glycogen degradation in the dark (15–18).

Although acclimation of *Synechocystis* to the nutrient availability in the environment has been evaluated at the transcriptional and physiological levels (16,19–23), the approaches used were time and resource-demanding. In the effort to develop a method, by which *Synechocystis* (or any microorganism) can be rapidly and inexpensively identified/characterized, we chose to evaluate *Synechocystis* phenotypes of the circadian rhythm with Raman spectroscopy and Rametrix^{TM} (24). This powerful analytical technique has been applied to a wide range of biological samples (25) including whole-cell bacteria (26–28). Raman spectroscopy is a non-destructive method and requires minimal or no sample preparation. The biological sample is excited by a monochromatic laser and the obtained spectrum shows the intensity of the Raman scattered radiation as a function of wavenumber, and spectra are extremely sample-specific (29–31) and represents a "snapshot" of molecular composition. A major challenge with Raman spectroscopy is to deconvolute the complex Raman signal to extract molecular composition information. Nonetheless, Raman has been proven a rapid and reliable method for the detection and identification of microorganisms (28,32). It has also enabled the detection of phenotype changes of *E. coli* upon exposure to different alcohols (33,34) and antibiotics (35). Furthermore, it has been used in conjunction with peptide-guided Surface-Enhanced Raman Scattering (pgSERS) probes to locally characterize sub-cellular compartment composition (36).

In the present study, we used Raman spectroscopy to explore the phenotypic changes in the *Synechocystis sp.* PCC6803 circadian rhythm under different growth conditions: (i) autotrophic (i.e., photosynthesis in light/dark cycles), (ii) photoautotrophic (i.e., continuous light), (iii) heterotrophic (i.e., growth on glucose in light/dark cycles), (iv) photoheterotrophic (i.e., growth on glucose in continuous light), and (v) under nitrogen limitation in continuous light. We also included (vi) an autotrophic culture grown in the dark (i.e., "dark autotrophic") and (vii) a

heterotrophic culture grown in the dark (i.e., "dark heterotrophic") as a control. Multivariate statistical analyses were performed using the Rametrix^{TM} LITE (24) and PRO (37) Toolboxes for MATLAB®. Here, whole Raman spectra (over the biological range), were used in chemometric analyses to distinguish among whole-cell phenotypes in response to metabolic changes and light-to-dark transitions. Additionally, levels of some metabolites and macromolecules (glycogen, amino acids, fatty acids and chlorophyll) were determined using accepted analytical methods and compared with specific Raman bands to validate results.

4.2. Material and Methods

4.2.1. Cyanobacterial strain

Kanamycin resistant mutants of *Synechocystis sp.* PCC 6803 (ATCC® 27184™, a glucose-tolerant strain) were generated in previous research (38) and used in all experiments.

4.2.2. Culture media and growth conditions

Synechocystis cultures were grown in BG-11 medium (39) containing different amounts of nitrogen and glucose, and cultures were grown under different and light settings. "Autotrophic" conditions were achieved by growing cells in the minimal BG-11 medium and light to alternate light/dark (or day/night) cycles (12h/12h). Photoautotrophic conditions were obtained by growing cells in the minimal BG-11 media under continuous artificial light (20 μE m^{-2} s^{-1}). Heterotrophic conditions consisted of growth under light/dark cycles (12h/12h) in BG-11 supplemented with 5 mM glucose. "Photoheterotrophic" conditions consisted of BG-11 medium supplemented with 5 mM glucose and growth under continuous artificial light (20 μE m^{-2} s^{-1}). Nitrogen starvation conditions consisted of BG-11^0 medium (BG-11 medium without nitrate) supplemented with

different levels of $NaNO_3$ (0.88 mM, 1.76 mM, 3.52 mM, 8.8 mM, 13.2 mM, 16.72 mM and 17.6 mM corresponding to 5%, 10%, 20%, 50%, 75%, 95%, and 100% of nitrate in the BG-11 medium), where 17.6 mM corresponds to an unaltered BG-11 medium. Dark autotrophic and dark heterotrophic conditions were identical to autotrophic and Heterotrophic conditions, respectively, except the cultures were kept in the dark through the duration of the experiments. Culture growth was monitored by optical density at 730 nm (OD_{730}). All cells were grown to the exponential phase ($OD_{730} \approx 1.0$) and harvested by centrifugation. For the circadian rhythm phenotype dynamics experiments, *Synechocystis* cells were grown to the log phase in BG-11 media, with or without glucose. An aliquot of cells (5-10 µl) was centrifuged every 2 hours for 24 hours prior to analysis.

4.2.3. Raman spectroscopy

Cells (5 µl of culture; three biological replicates) were air dried on aluminum foil at room temperature prior to analysis. Raman spectroscopy was performed using an Agiltron PeakSeeker PRO-785 (Agiltron, Woburn, MA) Raman microscope. The spots of dried cells were imaged using a 10x objective, and the following Raman settings were used: 785 nm laser excitation for 5 seconds with spectral resolution of 13 cm^{-1}. Twenty spectra were acquired per sample by focusing on different regions of the dried spot of cells.

4.2.4. Rametrix™ and statistical analyses

Raman data were collected/visualized using RSIQ™ software (Agiltron) and analyzed further with the Rametrix™ LITE (24) and PRO (37) Toolboxes for MATLAB®. All calculations were performed using MATLAB® r2018a (Mathworks; Natick, MA) and required the Statistics and Machine Learning Toolbox.

The Rametrix™ LITE Toolbox was used to process all spectra, perform principal component analysis (PCA), and discriminant analysis of principal components (DAPC). Processing Raman spectra consisted of: (i) averaging the 20 spectra replicates; (ii) truncating spectra to the Raman shift biological range (400 - 1,800 cm^{-1}); (iii) baselining using the Goldindec algorithm (40); and (iv) vector normalization. Goldingdec algorithm parameters used were: (i) baseline polynomial order = 3; (ii) estimated peak ratio = 0.5; and (iii) smoothing window size = 5.

The Rametrix™ PRO Toolbox was used to further evaluate DAPC models generated by the LITE Toolbox. Here, leave-one-out DAPC model validations were performed where (i) one sample was held out of DAPC model construction, (ii) the classification of that sample was predicted by the model, and (iii) the model prediction was compared to the actual classification. This procedure was repeated so that every sample was held out of model construction and used for prediction accuracy. The following were used to convey results: (i) accuracy (the percentage of samples predicted correctly by the model), (ii) sensitivity (the true-positive prediction accuracy), and (iii) selectivity (the true-negative prediction accuracy). The Rametrix™ PRO Toolbox was also used to determine the "random chance" values of accuracy, sensitivity, and selectivity. These are calculated with random classification predictions and can vary widely based on the number of classification factors in a dataset and their distribution. In short, Rametrix™ PRO predictions must exceed the random chance values to show effectiveness of Raman spectroscopy and Rametrix™ to classify unknown samples correctly.

Multi-way ANOVA and pairwise comparisons using using Tukey's honestly significant difference (HSD) method, with 99% confidence interval were used to compare Raman spectra directly. However Raman spectra consist of intensity data points over a wide range of Raman shifts

(400-1,800 cm^{-1}). Thus, these data-rich spectra need to be reduced to a single value if several are to be compared using the statistical tools above. This is done by calculating the multi-dimensional distance between a baselined and vector normalized Raman spectrum and a reference spectrum. This is referred to as the Total Spectral Distance (TSD) and is calculated according to Eq. 1,

$$TSD = \sum_{i=400}^{1,800} \sqrt{\left(S_{x,i} - S_{reference,i}\right)^2} \quad (Eq. 1)$$

where $S_{x,i}$is the Raman intensity value of sample *x* at Raman shift *i* (where *i* goes from 400-1,800 cm^{-1}). The term $S_{reference,i}$represents the Raman intensity value of the reference (or control) spectrum over the Raman shift range. The distance calculation was also applied to principal components (PCs) from PCA analysis by the Rametrix™ LITE Toolbox according to Eq. 2, where the first five PCs (usually representing over 95% of the dataset variations) were used. This is referred to as the Total Principal Component Distance (TPD).

$$TPD = \sum_{i=1}^{5} \sqrt{\left(P_{x,i} - P_{reference,i}\right)^2} \quad (Eq. 2)$$

Finally, this formula was applied to canonical values resulting from DAPC models built using the Rametrix™ LITE Toolbox. This is referred to as the Total Canonical Distance (TCD) and is given in Eq. 3. The TCD is calculated from the top five canonicals in the DAPC model. The TSD, TPD, and TCD distance metrics have been developed and used in previous research (41).

$$TCD = \sum_{i=1}^{5} \sqrt{\left(C_{x,i} - C_{reference,i}\right)^2} \quad (Eq. 3)$$

4.2.5. Estimation of total glucose (glycogen and free glucose) levels

Glycogen levels were determined with a previously published protocol (42). Briefly, lyophilized cells (0.5 mg) were suspended in 100 µl of 3.5% (v/v) sulfuric acid solution and boiled for 40 min. Glucose produced by acid hydrolysis was assayed spectrophotometrically at 635 nm using 6% o-toluidine solution.

4.2.6. Amino acid and fatty acid analyses

Amino acids and fatty acids were analyzed as AccQ-Tag™ derivatives and fatty acid methyl esters (FAME) using Waters ultra-performance liquid chromatography (UPLC) and gas chromatography coupled with flame ionization detection (GC-FID), respectively, as described (43). Briefly, 1 mg of lyophilized *Synechocystis* cells was subjected to biphasic extractions using 10 µg heptadecanoic acid (C17:0) and 50 mM norvaline as internal standards, followed by protein and lipid hydrolysis and derivatizations prior to UPLC and GC-FID analyses. Amino acids were analyzed on a Waters Acquity™ H-class UPLC system (Waters Corporation, Milford, MA) using fluorescent detection, while FAME on an Agilent 7890A series GC and 5975C series single quadrupole MS (Agilent Technologies, Santa Clara, CA).

4.2.7. Chlorophyll a measurement

Lyophilized cells were suspended in 100% methanol and disrupted for 5 minutes with bead beater. Chlorophyll a was determined spectrophotometrically at 665 nm and 720 nm after 20 minutes of incubation in the dark at 4ºC (44).

4.2.8. Public availability

All Raman spectra and raw data measurements are available publicly at VTechData (https://doi.org/10.7294/PKC0-5K96). The Rametrix™ LITE Toolbox is available through GitHub (https://github.com/SengerLab/RametrixLITEToolbox). The Rametrix™ PRO Toolbox is also shared through GitHub (https://github.com/SengerLab/RametrixPROToolbox). Both are shared under licence agreement.

4.3. Results

4.3.1. Raman spectroscopy of Synechocystis grown under different conditions

Raman spectroscopy was used to detect specific phenotypic changes in *Synechocystis sp.* PCC6803 under the following growth conditions: (i) autotrophic, (ii) heterotrophic, (iii) photoautotrophic, (iv) photoheterotrophic, (v) nitrogen deprivation, (vi) dark autotrophic (control), and (vii) dark heterotrophic (control). Cells were analyzed by Raman microscopy, and the resulting spectra were processed using the Rametrix™ LITE (24) and PRO (37) Toolboxes in MATLAB®. Averaged, truncated (400 - 1800 cm^{-1}), baselined, and vector normalized Raman spectra are shown in Fig. 4.1 for the different growth conditions defined above. While the overall spectral signatures remained consistent with each other, Raman spectroscopy was able to detect significant differences in cell phenotypes that were induced by the different growth conditions. These differences were analyzed further using PCA and DAPC in the Rametrix™ LITE Toolbox, and the Rametrix™ PRO Toolbox was used to determine if these phenotypes (i.e., growth conditions) were predictable from a Raman spectrum alone (when no other information is given). Furthermore, the phenotype changes were further validated through more traditional measurements (e.g., UPLC, GC-FID, and spectroscopy), as described in the Materials and Methods section. Thus, the remainder of the

Results section details two different types of analyses: (i) chemometric phenotype analyses by Rametrix™ and (ii) correlation of individual Raman bands with metabolites and their levels analyzed by well-established analytical approaches.

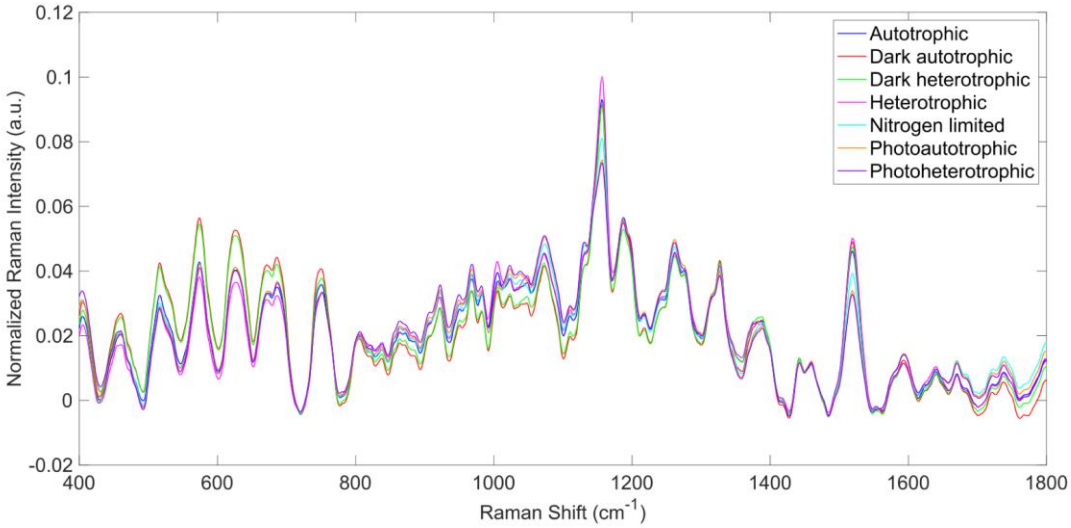


Figure 4.1. Averaged, truncated (400-1,800 cm^{-1}), baselined, and vector normalized Raman spectra from all culture conditions studied.

4.3.2. Chemometric phenotype analyses by Rametrix™

4.3.2.1. Glucose-induced phenotypes

To investigate the effect of glucose addition to BG-11 medium on *Synechocystis* phenotypes, Rametrix™ was used to analyze the autotrophic, heterotrophic, photoautotrophic, and photoheterotrophic growth. These different growth modes were assigned "BG-11 medium" and "BG-11 (with 5 mM glucose)" classifications. PCA, DAPC, and principal component contribution results generated by the Rametrix™ LITE Toolbox are given in Fig. 4.2. PCA of Raman spectra (Fig. 4.2A) demonstrated no apparent clustering in the first two PCs, which comprised over 68%

of the dataset variance. Application of DAPC (Fig. 4.2B), however, revealed two marked clusters, based on the presence of glucose in the culture medium, when using 40 PCs (representing over 99.9% of the dataset variance). These two clusters suggest cellular phenotype differences could exist as a result of adding glucose to BG-11 medium. Next, the contributions of different Raman shifts to the PCA and DAPC results were examined. This analysis lends to discovering molecular differences between clusters of Raman scans. In this case, it provided information about the molecular differences between cultures grown in the presence and absence of glucose. The Raman shift contributions between the two groups in PCA are shown in Fig. 4.2C, and the contributions leading to the separations in DAPC are shown in Supplementary Appendix C, Fig. C1. Raman band assignments were assigned based on a popular published library (31). Specific lists of Raman band contributions and molecular assignments (for both PCA and DAPC) are given in Supplementary Appendix C, and highlights are given in Table 4.1.

Next, from spectral, PCA, and DAPC data, the TSD, TPD, and TCD distance values (Eqs. 1-3) were calculated and used in ANOVA and pairwise comparison tests. As shown previously (41), statistical calculations with TSD and TPD are very similar, so only TPD and TCD are reported here. All distance data (both raw and mean values) are given in Supplementary Appendix C. With TPD data, ANOVA returned an insignificant *p*-value ($p = 0.13$) for the overall change in phenotypes between cultures grown in BG-11 medium and cultures grown in BG-11 (with 5 mM glucose). TCD calculations (when using 40 PCs in DAPC), however, did find significant differences ($p < 0.001$), and these findings are consistent with the clustering shown in Figs. 4.2AB. With only 2 possible classifications (i.e., BG-11 with and without glucose), pairwise comparisons were identical to ANOVA results here.

Finally, Rametrix™ PRO was used to apply leave-one-out analysis to the DAPC clustering results in Fig. 4.2B. Essentially, this analysis determines the ability of the DAPC model to correctly predict the classification (i.e., BG-11 medium with or without glucose) when presented with an unknown Raman spectrum of *Synechocystis* cells. All Rametrix™ PRO results, for DAPC models constructed with different numbers of PCs, are given in Supplementary Appendix C, and the best-performer DAPC model is summarized in Table 4.2. Given that nearly equal numbers of positive and negative scans were used, the random chance accuracy, sensitivity, and selectivity for this case are all 50%. Thus, model performance far exceeded random chance. DAPC models with fewer PCs (i.e., including less dataset variability) performed worse, as did models including more than 40 PCs. Models with more than 40 PCs likely suffered from over-fitting.

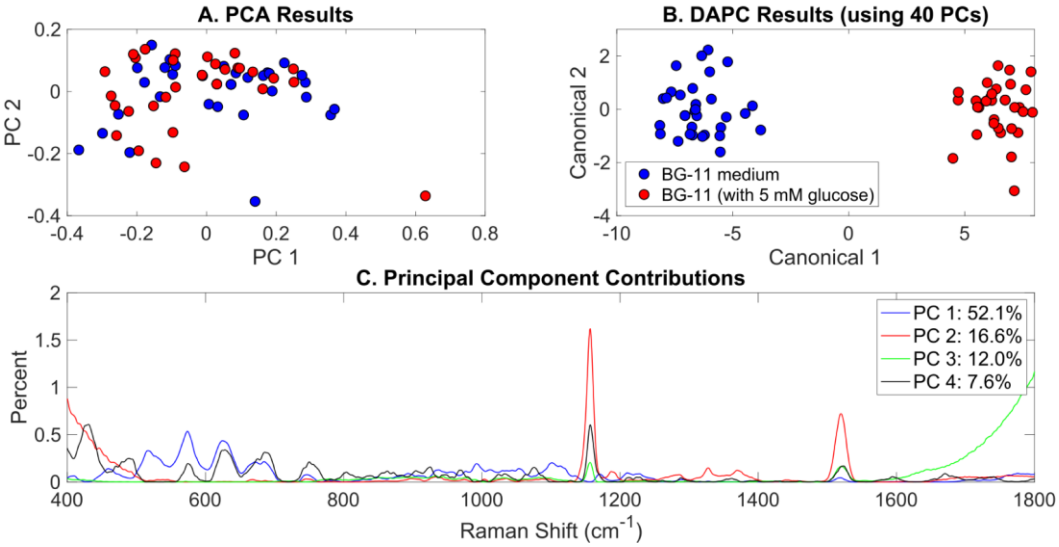


Figure 4.2. Rametrix™ models for the glucose-induced phenotypes study. (A) PCA results, (B) DAPC results when using 40 PCs, and (C) Raman shift contributions to the differences between the groups in PCA.

4.3.2.2. Illumination phenotypes

Similar to the previous section, Raman spectra of cultures grown in autotrophic, heterotrophic, photoautotrophic, photoheterotrophic, dark autotrophic, and dark heterotrophic conditions were analyzed according to the different light settings: (i) continuous artificial light, (ii) dark/light cycle (12h/12h), and (iii) dark. PCA, DAPC, and PC contributions generated by the Rametrix^{TM} LITE Toolbox are shown in Fig. 4.3. Some clustering was apparent in PCA, and this was improved significantly for the DAPC model constructed with 9 PCs, representing approximately 99% of the dataset variance. The full lists of Raman shift contributions for PCA and DAPC are given in Supplementary Appendix C, and these results for the DAPC model are shown in Fig. C2. A summary is given in Table 4.1. For Raman shift contributions to PCA, similar results were observed as reported for the glucose-induced phenotypes.

TSD, TPD, and TCD distances (Eqs. 1-3) were calculated based on spectral, PCA, and DAPC data and are given in Supplementary Appendix C (both raw and mean values). An ANOVA test based on the type of illumination (options listed above) revealed statistically significant differences in TPD data ($p < 0.01$) and TCD data ($p < 0.001$). Pairwise comparisons showed insignificant differences in TPD values between light/dark and dark growth conditions ($p = 0.81$), somewhat significant differences between light/dark and continuous light ($p = 0.022$), and significant differences between dark and continuous light ($p < 0.01$). With TCD data (when using 9 PCs in DAPC), both pairwise comparisons involving continuous light were statistically significant ($p < 0.001$). However, the pairwise comparison for dark and light/dark was not statistically significant ($p = 0.55$). These results agree with the clustering results in Fig. 4.3AB, suggesting that phenotypes resulting from growth in continuous light are more easily distinguished (and separated by clustering) than those arising from growth in the dark or light/dark.

Rametrix™ PRO was used to apply leave-one-out analysis to the DAPC results in Fig. 4.3B. Here, Rametrix™ PRO had three classification options, dark, light/dark, or continuous light. The following example shows how the positive and negative conditions were assigned when using more than two classifications. To test the condition dark, this was assigned as the positive condition. Both light/dark and continuous light comprised the negative condition. All three conditions were tested as the positive condition in this analysis. Here, the prediction accuracy metric remains the same, the ability to classify correctly the positive and negative conditions. The sensitivity (true positive rate) holds more importance, where the selectivity (true negative rate) must also be high (i.e., above random chance). With our dataset, the random chance selectivity is 33% for each classification, and the random chance selectivity is 67%. Thus, model performance sensitivity and selectivity must be higher than these values if the model is truly able to classify *Synechocystis* phenotypes based on Raman spectra.

Using a DAPC model built with 9 PCs (99% of dataset variability), all conditions far exceeded the random chance sensitivity and selectivity values, as shown in Table 4.2. In fact, the phenotypes resulting from continuous light were highly distinguishable, resulting in 100% accuracy, sensitivity, and selectivity. Results for other DAPC models, constructed with different numbers of PCs, are given in Supplementary Appendix C.

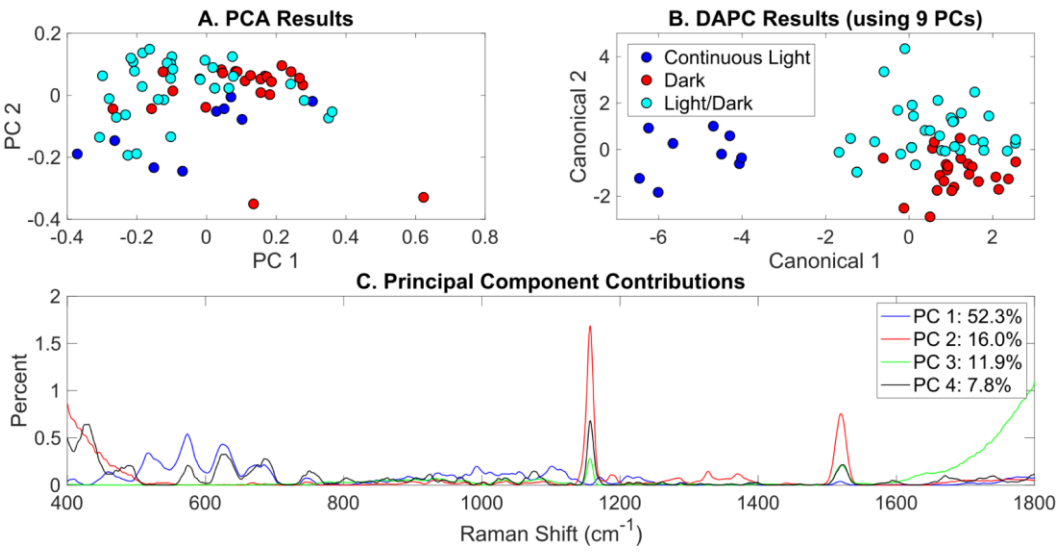


Figure 4.3. Rametrix™ models for the illumination-induced phenotypes study. (A) PCA results, (B) DAPC results when using 40 PCs, and (C) Raman shift contributions to the differences between the groups in PCA.

4.3.2.3. Nitrogen limitation phenotypes

Nitrate is the most common nitrogen source used by *Synechocystis* (45,46), and we hypothesized that growth with different nitrate levels would result in significant phenotype changes observable by Rametrix™. Thus, *Synechocystis* cultures were grown in BG-11 medium with different nitrate levels (0-100% of native BG-11) and continuous light. Raman spectra were obtained of cultures at each nitrate level and were truncated (400-1,800 cm^{-1}), baseline corrected, vector normalized, and averaged. These results are shown in Fig. C5 in Supplementary Appendix C. Clear differences were observed in Raman signal intensity over certain regions of the Raman shift, suggesting that there are specific changes in culture phenotypes related to nitrate deprivation.

PCA and DAPC model results are presented in Fig. 4.4. Again, little clustering was observed with PCA (Fig. 4.4A). DAPC models were built with several numbers of PCs, and

analyses of these are given in Supplementary Appendix C. DAPC models built with 3, 5, and 10 PCs (representing 83.2%, 95.6%, and 99.4% of dataset variance, respectively) are shown in Figs. 4.4B-D. As more PCs were included in the DAPC model, the separation of the more nitrate deprived cultures (5% and 10%) separated and cultures receiving 50-100% of BG-11 levels of nitrate clustered together, suggesting significantly altered phenotypes may only occur at low nitrate levels. PCA and canonical contributions are given in Supplementary Appendix C and are summzarized in Table 4.1.

TSD, TPD, and TCD data were generated (Eqs. 1-3) and are available in Supplementary Appendix C. ANOVA and pairwise comparison tests were applied, and nitrate concentrations were found to be statistically significant in ANOVA ($p < 0.001$) for all TSD, TPD, and TCD data. Some pairwise comparisons showed statistical significance when considering TCD data (for DAPC model built with 3 PCs). In particular, no statistical differences were found between 100%, 95%, and 75% of nitrate present in the medium. Comparing 100% and 50% gave a *p*-value of 0.02, and comparisons with 10% and 5% gave *p*-values ≤ 0.01. Unique to this analysis, regression analysis was performed to determine the correlation between the percentage of nitrate included and calculated distances (TSD, TPD, and TCD). These results are also given in Supplementary Appendix C, and are summarized here. In particular, a correlation coefficient (R) of 0.87 was found between TSD values and the percentage of nitrate in BG-11 medium (0-100%). With TPD, a value of R = 0.84 was calculated. The calculation was repeated for TCD data from DAPC models built with different numbers of PCs. For a model built with 3 PCs, R = 0.81. For 5 PCs, R = 0.88, but for 10 PCs, R = 0.59. This further shows the incidence of "overfitting" as too many PCs are used to build a DAPC model.

Finally, leave-one-out analysis was applied using Rametrix^{TM} PRO. For the nitrate limitation dataset, sensitivity must exceed the random chance sensitivity and selectivity values of 14% and 86%, respectively, demonstrate some prediction success. The best-performer DAPC model results are given in Table 4.2, and results for all models tested are in Supplementary Appendix 1. The DAPC model built with 3 PCs (Table 4.2) generated an overall sensitivity of 44.4% and selectivity of 86.4% across all nitrate levels, exceeding the random chance values. It is noted that some individual sensitivity/selectivity values for this study (Table 4.2) fell below the random chance values. This would likely be remedied by expanding the number of measurements in the dataset. Models built with more PCs showed greater overall sensitivity but had inadequate selectivity. For example, 5 PCs yielded sensitivity = 77.8% but selectivity = 56.5%, and 10 PCs yielded selectivity = 77.8% but selectivity = 22.2% (Supplementary Appendix C).

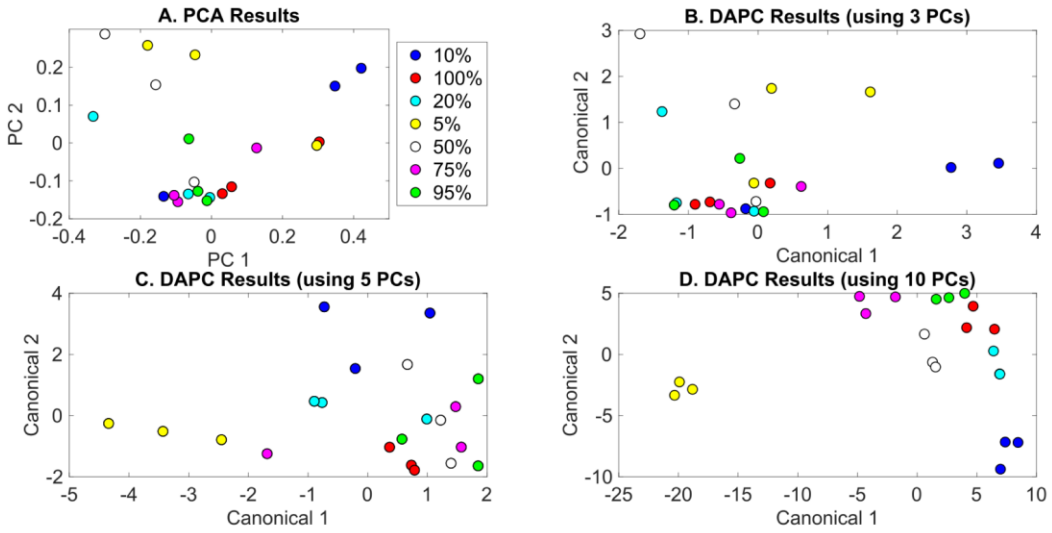


Figure 4.4. Rametrix™ models for the nitrate limitation phenotypes study. (A) PCA results, (B) DAPC results when using 3 PCs, and (C) Raman shift contributions to the differences between the groups in PCA. Percentages of nitrate in BG-11 medium are given (with 100% being native BG-11). The following are conversions between percentages and mM: 5% = 0.88 mM; 10% = 1.76 mM; 20% = 3.52 mM; 50% = 8.8 mM; 75% = 13.2 mM; 95% = 16.72 mM; 100% = 17.6 mM.

4.3.2.4. Classification of all phenotypes

The production of PCA and DAPC models with Rametrix™ LITE and validation with leave-one-out analysis with Rametrix™ PRO was repeated for a dataset consisting of all Raman scans used in this study. This included the growth conditions: (i) autotrophic, (ii) heterotrophic, (iii) photoautotrophic, (iv) photoheterotrophic, (v) nitrate limited, (vi) dark autotrophic, and (vii) dark heterotrophic. The autotrophic and heterotrophic conditions used light/dark illumination, and the photoautotrophic, photoheterotrophic, and nitrate-limited conditions used continuous light. The dark heterotrophic and dark autotrophic (control) cultures were grown in the dark. PCA results are shown in Fig. 4.5A and, again, reveal little clustering. DAPC models were built using several

numbers of PCs (complete list in Supplemental Appendix C). Models constructed with 8 and 50 PCs are shown in Figs. 4.5B and 4.5C, respectively. The clustering in Fig. 4.5C largely reveals clustering of phenotypes according to the presence of glucose and illumination. The following clusters were observed: (i) autotrophic and dark autotrophic conditions (no glucose; light/dark and dark), (ii) nitrate limited and photoautotrophic (no glucose; continuous light), and (iii) heterotrophic and dark heterotrophic (no glucose; light/dark and dark). The only condition to not cluster was photoheterotrophic, which appeared closer to the continuous light cluster than the heterotrophic cluster in Fig. 4.5C.

When considering all phenotypes (listed above), statistically significant ANOVA test results ($p < 0.001$) were obtained from both TPD and TCD data. All TSD, TPD, and TCD raw and mean distance values are given in Supplementary Appendix C, along with ANOVA and pairwise comparison test results. For TPD data, pairwise comparisons showed statistical significance ($p \leq 0.001$) between the pairs: autotrophic/heterotrophic, autotrophic/photoautotrophic, and autotrophic/photoheterotrophic. For TCD data, all comparisons were statistically significant ($p < 0.001$), meaning differences were found in the cell phenotypes.

Rametrix^{TM} PRO was used to determine if these phenotypes could be predicted from Raman spectra. With seven different classifications, the random sensitivity was calculated at 14% for each, and random selectivity was calculated at 86% for each. The DAPC model built with 8 PCs (98% of dataset variance) performed the best. Results were better than random chance (except for one selectivity value) and included sensitivity and selectivity values (respectively) of: (i) 25% and 91% for autotrophic, (ii) 38% and 94% for autotrophic control, (iii) 38% and 93% for dark heterotrophic, (iv) 50% and 87% for heterotrophic, (v) 39% and 91% for nitrate limited, (vi) 50% and 91% for photoautotrophic, and (vii) 67% and 76% for photoheterotrophic. Ultimately,

prediction metrics would likely be improved with more data points given seven different possible classifications.

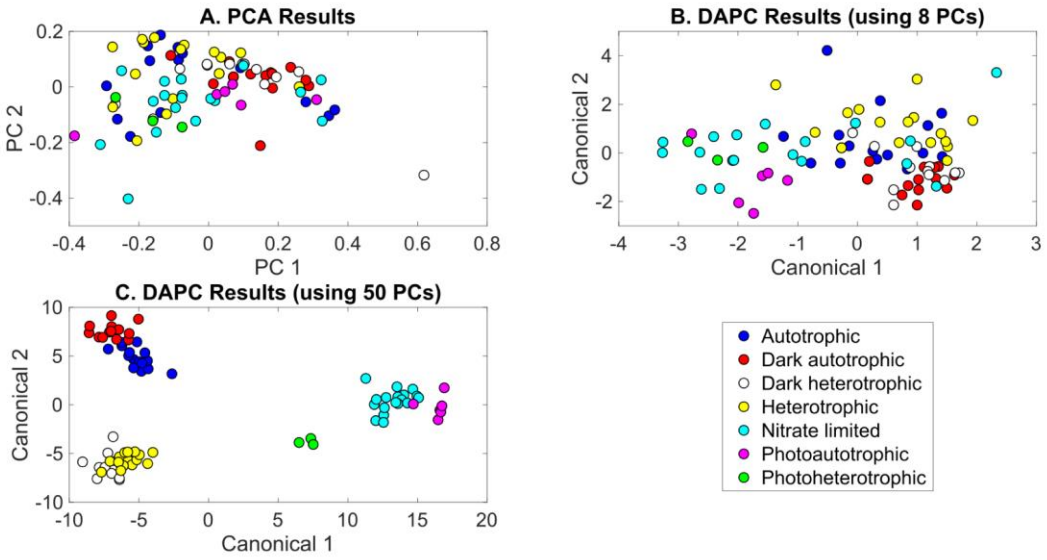


Figure 4.5. Rametrix™ models for the all phenotypes study. (A) PCA, (B) DAPC model built with 8 PCs, (C) DAPC model built with 50 PCs.

4.3.2.5. Circadian rhythm growth phenotypes

Synechocystis have a circadian rhythm to cope with daily fluctuations in available nutrients and light intensities (14). Studies have shown that photosynthesis and glycogen synthesis is carried-out in the light, and respiration and glycogen degradation is adopted in the dark (15–18). Here, Rametrix™ was employed to detect phenotypic changes related to the circadian rhythm in *Synechocystis sp.* PCC 6803 grown under four different conditions: (i) autotrophic (with 12-hour light/dark cycles), (ii) heterotrophic (with 12-hour light/dark cycles), (iii) autotrophic in the dark, and (iv) heterotrophic in the dark. It was hypothesized that a circadian rhythm would be observed in cultures grown under light/dark cycles and absent from cultures grown in the dark only. For a

circadian rhythm, phenotypes should deviate from its initial point during the course of the 24-hour light/dark cycle and eventually return to the starting point at the end of the cycle. Rametrix™ is well-suited to probe this cycle, and the TSD was chosen to represent phenotype changes. The TPD calculations would also work here. As a phenotype deviates from its initial phenotype, the TSD should become larger in value, and as a culture returns to its initial phenotype, the TSD value will become smaller (or very close to zero). Results for the four culturing conditions are shown in Fig. 4.6. Raw TSD data (available in Supplementary Appendix C) were first filtered in MATLAB to remove outliers according to the median method. Next, TSD values exceeding 0.15 were excluded as outliers from the time-course data in Fig. 4.6. Cultures grown in light/dark cycles (both autotrophic and heterotrophic) exhibited bimodal deviations from the initial phenotype and including the time-course outliers still produced the bimodal shape. The highest values of TSD occurred at 6 and 18 hours, with a near return to the initial phenotype at 12 hours and a full return to the initial phenotype at 24 hours. Fourth-order polynomials were then fitted through the time-course data for each growth condition in Fig. 4.6. These high-order polynomials were chosen because of their ability to detect the bimodal deviation, should it exist. As shown in Fig. 4.6CD, the bimodal deviation did not occur for the autotrophic and heterotrophic cultures grown in the dark and the final phenotypes (after 24 hours) did not return to the initial phenotypes.

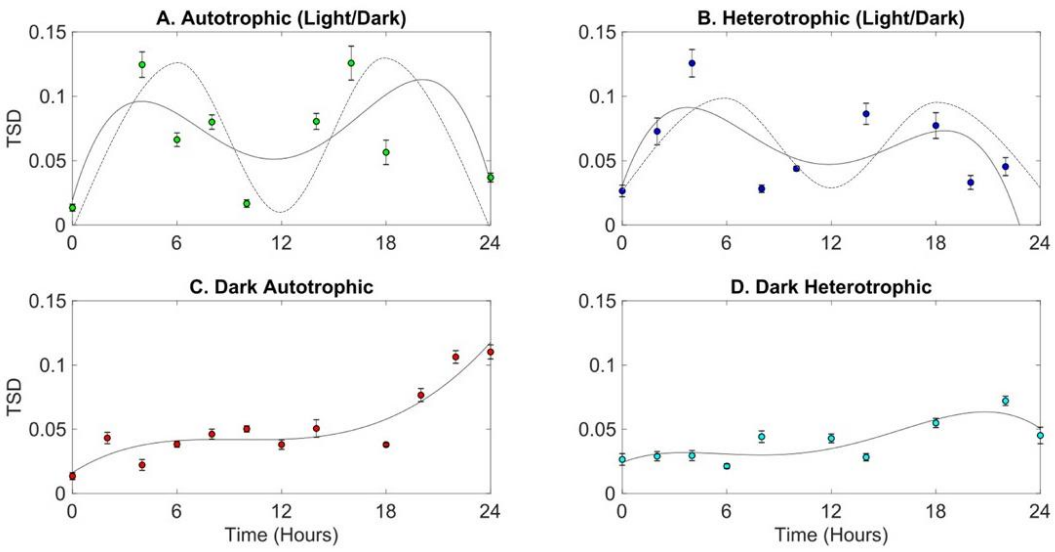


Figure 4.6. Identification of circadian rhythm phenotype changes based on TSD data for cultures grown in the following conditions: (A) autotrophic (light/dark), (B) heterotrophic (light/dark), (C) dark autotrophic, and (D) dark heterotrophic.

4.3.3. Correlation of individual Raman bands with metabolites and their levels analyzed by well-established analytical approaches

The nitrogen-limited cultures (0-100% of BG-11 medium nitrate; continuous light), described in earlier chemometric studies, were used in further phenotype validation experiments using analytical methods described in the Materials and Methods section. The purpose was to correlate individual Raman band intensities with specific biomolecules and results from other, widely accepted analytical methods (e.g., GC-FID, UPLC, and spectroscopy).

4.3.3.1. Glycogen

Previous studies reported an activation of glycogen synthesis during nitrogen-limited conditions (47–51). Therefore, glycogen levels were studied by comparing glycogen assigned

bands in Raman spectroscopy (31,52) and an established spectrophotometric method (42). As expected, glycogen levels increased in response to nitrogen deprivation by using both methods (Fig. C6). Cultures given 0.88 mM nitrate (5% of BG-11) showed about 2.5-fold increase in glycogen levels compared to cells grown at 17.6 mM (100% of BG-11) nitrate. Cultures grown in nitrate concentrations of 1.76 mM (10%), 13.2 mM (75%) and 16.72 mM (95%) showed increased glycogen levels by about 1.5-fold. However, growth in 3.52 mM (20%) and 8.8 mM (50%) nitrate showed no additional accumulation of glycogen.

Raman band assignments for glycogen (31,52) were correlated with the spectrophotometric measurements (Fig. C6). A strong correlation was obtained with the glycogen-related bands at 1150 cm^{-1} (R = 0.82) and at 1155 cm^{-1} (R = 0.80), suggesting that Raman band analysis may be used for determining glycogen levels. Although Raman spectroscopy cannot be used to obtain absolute concentrations of molecules, the trends reflecting changes in glycogen levels in response to varying nitrate availability were comparable between the two methods.

4.3.3.2. Amino acids

Proteins and amino acids represent the major nitrogen-containing compounds in a cell. Previous studies have demonstrated that protein and amino acid levels decrease dramatically with prolonged nitrate limitation (48–51,53). The levels of total amino acids (protein-derived and free amino acids) were determined by both Raman spectroscopy (54,55) and UPLC coupled with fluorescence detection (Fig. C7). Individual Raman bands used for each amino acid are given in Table C11 in Supplementary Appendix 2. Strong correlations that varied between R = 0.6 - 0.91 were observed between Raman spectroscopy and UPLC data (Fig. C8) for the majority of amino acids. Only Gly and Phe had moderate correlations, where the correlation coefficient varied from

R = 0.44 - 0.54. It is important to note that some amino acids such as Cys and Trp are degraded during acidic hydrolysis and are not detected by UPLC (56–58). Raman band data enabled the prediction of expression trend of these two amino acids (Fig. C9), suggesting also a decrease in Cys and Trp levels during prolonged nitrate starvation. In addition, Asn and Gln are converted to Asp and Glu during acidic protein hydrolysis and, as such, Asp and Glu levels reflect Asp/Asn and Glu/Gln levels, respectively (59,60). Our Raman spectroscopy and UPLC measurements are in good agreement, and the results are consistent with previously published metabolic phenotypes reflecting metabolic adaptations to nitrogen deprivation (51,53).

4.3.3.3. Fatty acids

Nitrogen starvation leads to oxidative stress, which causes damage to membrane lipids (61–64). The levels of total saturated and unsaturated fatty acids were determined by both GC-FID (Fig. C10A) and Raman spectroscopy. Overall, decreases in fatty acids levels were observed at low nitrate concentrations (0.88 mM and 1.76 mM; 5% and 10% of BG-11 medium, respectively). Moderate correlations were obtained between Raman assigned bands to fatty acids (31) and GC-FID data. While the correlation coefficients were somewhat low (R ~ 0.5), this may be due to GC-FID-measured fatty acids at 13.2 mM (75%) nitrate that do not appear to fit with time-course data, suggesting a problem with this particular GC-FID data point (Fig. C10). For example, with measurements at 13.2 mM nitrate included in the dataset, the correlation with the Raman band intensity at 1,078 cm^{-1} (I_{1078}) was R < 0.5. With the data at 13.2 mM nitrate removed, the correlation improved, R > 0.85.

4.3.3.4. Chlorophyll a

Nitrogen deprivation stress leads to damage of photosynthetic components, including chlorophyll degradation, resulting in chlorosis (48,65). Levels of photosynthetic pigments can be measured by both Raman spectroscopy and using spectrophotometric methods (44). As such, changes in chlorophyll levels were evaluated by these two approaches (Fig. C11 in Supplementary Appendix C). Raman signals corresponding to chlorophyll a (66,67) have a strong correlation with the spectrophotometric measurements (Fig. C11). The Raman band assignment 1,239 cm^{-1} had the highest correlation coefficient (R = 0.92) with spectroscopic data. As expected, chlorophyll a levels decreased with the severity of nitrogen deprivation. Based on these results, Raman spectroscopy may also be used to quantify the levels of pigments in *Synechocystis*.

4.4. Discussion

Raman spectroscopy with Rametrix^{TM} was used as an analysis tool to study metabolic changes related to different growth conditions in *Synechocystis sp.* PCC6803. Two methods were explored here: (i) broad phenotype changes to external stimuli using chemometric approaches and (ii) the correlation of individual Raman bands with specific biomolecules, with validation by accepted analytical methods. Raman spectroscopy is rapid, non-destructive, and reliable for the characterization of changes in metabolic phenotypes as a response to changes in growth conditions. Clear distinctions were observed between cells grown in the absence or presence of glucose and in various light conditions. In most cases, Raman spectral data were found to represent changes in the biomolecular composition of cells well and enabled the differentiation of phenotypes. For example, clear statistically relevant differences were found between phenotypes of autotrophic cultures growing in BG-11 medium and heterotrophic cultures growing in BG-11 medium with 5

mM glucose. Given an unkown Raman spectrum of *Synechocystis* growing in one of these environments, Rametrix™ PRO could determine that environment with 95.3% accuracy (sensitivity = 93.4%; selectivity = 96.9%). The observation of glucose-induced phenotypes agrees with previous studies (10,11,45,47,68). It was reported previously that glucose has an effect on the *Synechocystis* growth rate and it activates other sets of metabolic pathways (10,45,47,68); these should lead to phenotype changes detectable by Raman spectroscopy.

Phenotype changes due to different illumination were also found through Rametrix™ LITE clustering of Raman spectral data, confirmed for the continuous light phenotypes through statistical analyses (ANOVA and pairwise comparisons), and predictable over random chance using Rametrix™ PRO (including 100% sensitivity and selectivity for the continuous light phenotype). Previous studies have revealed that light is one of the most important environmental factors relevant to growth of photosynthetic organisms (11,12,45,69,70) . Light plays fundamental roles including, but not limited to, altering gene expression (70,71), initiating phototaxis (72), and resetting circadian rhythm (15,72). In addition, excess of light (11,73–75) and UV (76) might induce damage to the photosystem II. These light excess phenotypic changes appear to be the most easily detected by Raman spectroscopy and Rametrix™.

While clear cluster separations were not observed with Rametrix™ LITE models for nitrogen-limited phenotypes, ANOVA and pairwise comparisons confirmed phenotype changes. In addition, Rametrix™ PRO was able to predict the level of nitrate in the culture medium with accuracy, sensitivity, and selectivity greater than by random chance. Further, analyses by accepted analytical methods (e.g., GC-FID, UPLC, spectroscopy, etc.) confirmed changes in glycogen, fatty acids, amino acids, and chlorophyll a, which were then correlated with specific Raman bands in good agreement. These results agree with several literature reports suggesting nitrogen-limitation

induced phenotype changes. In particular, *Synechocystis* cells respond to nitrate starvation by reprogramming central carbon and nitrogen metabolism, which is reflected by changes in the levels of various high abundance macromolecules and metabolites, including glycogen (47–50), proteins, amino acids (48–50,53), lipids (61–64), and photosynthetic pigments (43,60). All of these changes and phenotypes were confirmed here as the metabolites and macromolecules could be reliably detected from Raman spectra. Previous studies also revealed that nitrogen limitation induces specific cellular responses including nitrogen assimilation up-regulation (60) as well as degradation of chlorophyll, phycobilisomes, and light harvesting complexes, leading to chlorosis (48,65,77), which is consistent with our results (Fig. C5B in Supplementary Appendix C). In addition, cyanobacteria placed under low nitrogen stress conditions degrade intracellular proteins to amino acids to mobilize nitrogen and to catabolize amino acids to provide carbon precursors for sugar and glycogen synthesis (48–51).

Regarding the circadian rhythm results, our observations confirmed our hypothesis that TSD values (representative of phenotypes) would return to their initial values after 24 hours for cultures grown in light/dark (12h/12h) cycles. This was true of both autotrophic and heterotrophic cultures. The bimodal nature of this 24-hour cycle was unexpected and warrants further investigations. We also noted more extreme variances for light/dark grown cultures compared to cultures grown in the dark. This could be indicative of a higher degree of heterogeneity for light/dark grown cultures, based on whether analyzed cells were closer to the flask exterior or center when sampled. This has been known in the literature to play a role in metabolic activity (78).

Overall, there was a strong correlation between Raman band intensities and the data obtained by well-established analytical methods. As such, Raman spectroscopy is suitable for

comparative studies when relative levels of metabolites and macromolecules need to be determined. However, Raman spectra are highly convoluted and finding correlating bands may not always be possible. In addition, preliminary experiments such as the ones described in this study, should be performed for new cell types and culturing environments. Specific bands may become "hidden" due to potentially interfering bands belonging to other biomolecules. The presence of interfering compounds will be highly dependent on the type of the sample. Once the correlating bands are established, Raman spectroscopy can be used for nearly any type of high-throughput screening of metabolic phenotypes.

Table 4.1. Raman shift contributions to PCA and DAPC models for selected studies.

Study	**Model**	**Biomolecules and Raman bands (cm^{-1})**
Glucose-induced	PCA	Sterols and phosphatidylinositol (430 and 575 cm^{-1}, respectively); protein (750, 1102, and 1156 cm^{-1}); and carotenoids (1156, 1520 cm^{-1})
Glucose-induced	DAPC	Sterols and phosphatidylinositol (430, 704, 778 cm^{-1}); protein and amino acids (509, 632, 1052, 1245, 1403 cm^{-1}); DNA/RNA (664, 747 cm^{-1}), and lipids (1305 cm^{-1})
Illumination-induced	PCA	Sterols and phosphatidylinositol (430, 516, 574 cm^{-1}) and protein/carotenoids (1156, 1520 cm^{-1})
Illumination-induced	DAPC	Phosphatidylinositol (509, 776 cm^{-1}), phospholipids (719, 1328 cm^{-1}), lipids (1063, 1074 cm^{-1}), glycogen (1049, 1155 cm^{-1}), carotenoids (1155 cm^{-1}), porphyrin (1620 cm^{-1}), and several protein/polypeptide/amide bands (1005, 1155, 1628, 1633, 1638 cm^{-1})
Nitrogen limitation-induced	PCA, DAPC	Protein-related bands (571, 623, 639, 1156, 1359 cm^{-1}), carotenoids and porphyrin (1156, 1518, 1520 cm^{-1})

Table 4.2. Summary of Rametrix™ PRO results for best-performing DAPC models.

Study	Number of PCs*	Classification Predicted	Accuracy	Sensitivity	Selectivity
Glucose-induced	Random chance**		50%	50%	50%
Glucose-induced	40 (99.9%)	BG-11 with 5 mM Glucose	95%	93%	97%
Illumination-induced	Random chance**		56%***	33%	67%
Illumination-induced	9 (99%)	Dark	89.5%	69%	90%%
Illumination-induced	9 (99%)	Light/Dark	78%	81%	75%
Illumination-induced	9 (99%)	Continuous Light	100%	100%	100%
Nitrate limitation-induced	Random chance**		76%	14%	86%
Nitrate limitation-induced	3 (83%)	5% (0.88 mM)	76%	67%	78%
Nitrate limitation-induced	3 (83%)	10% (1.76 mM)	95%	67%	100%
Nitrate limitation-induced	3 (83%)	20% (3.52 mM)	62%	0%	72%
Nitrate limitation-induced	3 (83%)	50% (8.8 mM)	81%	33%	89%
Nitrate limitation-induced	3 (83%)	75% (13.2 mM)	95%	67%	100%
Nitrate limitation-induced	3 (83%)	95% (16.72 mM)	52%	33%	80%
All phenotypes	Random chance**		76%***	14%	86%
All phenotypes	8 (98%)	Autotrophic	79%	25%	91%

All phenotypes	8 (98%)	Autotrophic control	86%	38%	94%
All phenotypes	8 (98%)	Dark heterotrophic	85%	38%	93%
All phenotypes	8 (98%)	Heterotrophic	80%	50%	87%
All phenotypes	8 (98%)	Nitrate limited	80%	39%	91%
All phenotypes	8 (98%)	Photoautotrophic	88%	50%	91%
All phenotypes	8 (98%)	Photoheterotrophic	75%	67%	76%

* Number of PCs used to build the DAPC model. The percent variance of the dataset represented is given in parentheses.

** The random chance sensitivity and selectivity values were calculated statistically, not by Rametrix^{TM} PRO. They were the same for all studies here.

*** These values are classification-dependent. The average value is given here, and individual values are given in Supplementary Appendix C.

References

1. Stanier RY, Cohen-Bazire G. Phototrophic prokaryotes: the cyanobacteria. Annu Rev Microbiol. 1977;31:225–74.
2. Waterbury JB. The cyanobacteria-Isolation, purification and identification. " Prokaryotes" M Dworkin Al Eds. 2006;1053–1073.
3. Heidorn T, Camsund D, Huang H-H, Lindberg P, Oliveira P, Stensjö K, et al. Synthetic biology in cyanobacteria engineering and analyzing novel functions. Methods Enzym. 2011;497:539–79.
4. Knoop H, Zilliges Y, Lockau W, Steuer R. The Metabolic Network of *Synechocystis* sp. PCC 6803: Systemic Properties of Autotrophic Growth. Plant Physiol. 2010;154(1):410–22.
5. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, et al. Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-coding Regions. DNA Res. 1996;3(3):109–36.
6. Nakamura Y, Kaneko T, Tabata S. CyanoBase, the genome database for *Synechocystis* sp. strain PCC6803: status for the year 2000. Nucleic Acids Res. 2000;28(1):72.
7. Angermayr SA, Hellingwerf KJ, Lindblad P, Teixeira de Mattos MJ. Energy biotechnology with cyanobacteria. Curr Opin Biotechnol. 2009;20(3):257–63.
8. Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proc Natl Acad Sci U A. 2012;109(7):2678–83.
9. Ducat DC, Way JC, Silver PA. Engineering cyanobacteria to generate high-value products. Trends Biotechnol. 2011;29(2):95–103.
10. Yu Y, You L, Liu D, Hollinshead W, Tang YJ, Zhang F. Development of *Synechocystis* sp. PCC 6803 as a phototrophic cell factory. Mar Drugs. 2013;11(8):2894–916.
11. Vermaas W. Molecular genetics of the cyanobacterium *Synechocystis* sp. PCC 6803: Principles and possible biotechnology applications. J Appl Phycol. 1996;8:263–73.
12. Anderson SL, McIntosh L. Light-activated heterotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803: a blue-light-requiring process. J Bacteriol. 1991;173(9):2761–7.
13. Yao DCI, Brune DC, Vermaas WFJ. Lifetimes of photosystem I and II proteins in the cyanobacterium *Synechocystis* sp. PCC 6803. FEBS Lett. 2012;586(2):169–73.
14. Tu BP, McKnight SL. Metabolic cycles as an underlying basis of biological oscillations. Nat Rev Mol Cell Biol. 2006;7(9):696–701.
15. Kucho K, Okamoto K, Tsuchiya Y, Nomura S, Nango M, Kanehisa M, et al. Global Analysis of Circadian Expression in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. J Bacteriol. 2005;187(6):2190–9.
16. Saha R, Liu D, Hoynes-O'Connor A, Liberton M, Yu J, Bhattacharyya-Pakrasi M, et al. Diurnal regulation of cellular processes in the cyanobacterium *Synechocystis* sp. Strain PCC 6803: Insights from transcriptomic, fluxomic, and physiological analyses. mBio. 2016;7(3):e00464.
17. van Alphen P, Hellingwerf KJ. Sustained Circadian Rhythms in Continuous Light in *Synechocystis* sp. PCC6803 Growing in a Well-Controlled Photobioreactor. PLOS ONE. 2015;10(6):e0127715.
18. Whitton BA. Diversity, ecology, and taxonomy of the cyanobacteria. In 1992.

19. Beck C, Hertel S, Rediger A, Lehmann R, Wiegard A, Kolsch A, et al. Daily Expression Pattern of Protein-Encoding Genes and Small Noncoding RNAs in *Synechocystis* sp. Strain PCC 6803. Appl Environ Microbiol. 2014;80(17):5195–206.
20. Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchi M. DNA Microarray Analysis of Cyanobacterial Gene Expression during Acclimation to High Light. Plant Cell. 2001;13(4):793–806.
21. Imamura S, Tanaka K, Shirai M, Asayama M. Growth Phase-dependent Activation of Nitrogen-related Genes by a Control Network of Group 1 and Group 2 sigma Factors in a Cyanobacterium. J Biol Chem. 2006;281(5):2668–75.
22. Takahashi H, Uchimiya H, Hihara Y. Difference in metabolite levels between photoautotrophic and photomixotrophic cultures of *Synechocystis* sp. PCC 6803 examined by capillary electrophoresis electrospray ionization mass spectrometry. J Exp Bot. 2008;59(11):3009–18.
23. Yang C, Hua Q, Shimizu K. Metabolic Flux Analysis in *Synechocystis* Using Isotope Distribution from ^{13}C-Labeled Glucose. Metab Eng. 2002;4(3):202–16.
24. Fisher AK, Carswell WF, Athamneh AIM, Sullivan MC, Robertson JL, Bevan DR, et al. The Rametrix^{TM} LITE Toolbox v1.0 for MATLAB®. J Raman Spectrosc. 2018;49:885–896.
25. Das RS, Agrawal YK. Raman spectroscopy: Recent advancements, techniques and applications. Vib Spectrosc. 2011;57(2):163–76.
26. Gaus K, Rosch P, Petry R, Peschke KD, Ronneberger O, Burkhardt H, et al. Classification of lactic acid bacteria with UV-resonance Raman spectroscopy. Biopolymers. 2006;82(4):286–90.
27. Jarvis RM, Goodacre R. Ultra-violet resonance Raman spectroscopy for the rapid discrimination of urinary tract infection bacteria. FEMS Microbiol Lett. 2004;232(2):127–32.
28. Pahlow S, Meisel S, Cialla-May D, Weber K, Rösch P, Popp J. Isolation and identification of bacteria by means of Raman spectroscopy. Adv Drug Deliv Rev. 2015;89:105–20.
29. Crow P, Stone N, Kendall CA, Uff JS, Farmer JA, Barr H, et al. The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma in vitro. Br J Cancer. 2003;89(1):106–8.
30. Mahadevan-Jansen A, Mitchell MF, Ramanujam N, Malpica A, Thomsen S, Utzinger U, et al. Near-infrared Raman spectroscopy for in vitro detection of cervical precancers. Photochem Photobiol. 1998;68(1):123–32.
31. Movasaghi Z, Rehman S, Rehman IU. Raman Spectroscopy of Biological Tissues. Appl Spectrosc Rev. 2007;42(5):493–541.
32. Nelson WH, Manoharan R, Sperry JF. UV Resonance Raman Studies of Bacteria. Appl Spectrosc Rev. 1992;27(1):67–124.
33. Zu TNK, Athamneh AIM, Wallace RS, Collakova E, Senger RS. Near-Real-Time Analysis of the Phenotypic Responses of *Escherichia coli* to 1-Butanol Exposure Using Raman Spectroscopy. 2014;
34. Freedman BG, Zu TNK, Wallace RS, Senger RS. Raman spectroscopy detects phenotypic differences among *Escherichia coli* enriched for 1-butanol tolerance using a metagenomic DNA library. Biotechnol J. 2016;11(7):877–89.
35. Athamneh AI, Alajlouni RA, Wallace RS, Seleem MN, Senger RS. Phenotypic profiling of antibiotic response signatures in *Escherichia coli* using Raman spectroscopy. Antimicrob Agents Chemother. 2014;58(3):1302–14.

36. Athamneh AI, Senger RS. Peptide-guided surface-enhanced Raman scattering probes for localized cell composition analysis. Appl Env Microbiol. 2012;78(21):7805–8.
37. Senger RS, Robertson JL. The Rametrix™ PRO Toolbox for MATLAB. 2019;(Under Review).
38. Tanniche I, Fisher AK, Gillam F, Collakova E, Zhang C, Bevan DR, et al. λ-PCR for precise DNA assembly and modification. 2019;(Under Review).
39. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. J Gen Microbiol. 1978;111:1–61.
40. Liu J, Sun J, Huang X, Li G, Liu B. Goldindec: A Novel Algorithm for Raman Spectrum Baseline Correction. Appl Spectrosc. 2015;69(7):834–42.
41. Senger RS, Du P, DeLaTorre Campos D, Carswell WF, Webster K, Sullivan MC, et al. Assessment of urine specimen storage conditions using Raman Chemometric Urinalysis™ (Rametrix™). 2019;(Under Review).
42. Osanai T, Oikawa A, Azuma M, Tanaka K, Saito K, Hirai MY, et al. Genetic engineering of group 2 sigma factor SigE widely activates expressions of sugar catabolic genes in *Synechocystis* species PCC 6803. J Biol Chem. 2011;286(35):30962–71.
44. Sinetova MA, Cerveny J, Zavrel T, Nedbal L. On the dynamics and constraints of batch culture growth of the cyanobacterium *Cyanothece* sp. ATCC 51142. J Biotechnol. 2012;162(1):148–55.
45. Lopo M, Montagud A, Navarro E, Cunha I, Zille A, de Cordoba PF, et al. Experimental and modeling analysis of *Synechocystis* sp. PCC 6803 growth. J Mol Microbiol Biotechnol. 2012;22(2):71–82.
46. Flores E, Herrero A. Nitrogen assimilation and nitrogen control in cyanobacteria. Biochem Soc Trans. 2005;33(1):164–7.
47. Yoo S-H, Keppel C, Spalding M, Jane J. Effects of growth condition on the structure of glycogen produced in cyanobacterium *Synechocystis* sp. PCC6803. Int J Biol Macromol. 2007;40(5):498–504.
48. Hasunuma T, Kikuyama F, Matsuda M, Aikawa S, Izumi Y, Kondo A. Dynamic metabolic profiling of cyanobacterial glycogen biosynthesis under conditions of nitrate depletion. J Exp Bot. 2013;64(10):2943–54.
49. Adebiyi AO, Jazmin LJ, Young JD. ^{13}C flux analysis of cyanobacterial metabolism. Photosynth Res. 2015;126(1):19–32.
50. Osanai T, Imamura S, Asayama M, Shirai M, Suzuki I, Murata N, et al. Nitrogen induction of sugar catabolic gene expression in *Synechocystis* sp. PCC 6803. DNA Res. 2006;13(5):185–95.
51. Osanai T, Oikawa A, Shirai T, Kuwahara A, Iijima H, Tanaka K, et al. Capillary electrophoresis-mass spectrometry reveals the distribution of carbon metabolites during nitrogen starvation in *Synechocystis* sp. PCC 6803. Env Microbiol. 2014;16(2):512–24.
52. Kamemoto LE, Misra AK, Sharma SK, Goodman MT, Luk H, Dykes AC, et al. Near-infrared micro-Raman spectroscopy for in vitro detection of cervical cancer. Appl Spectrosc. 2010;64(3):255–61.
53. Hauf W, Schlebusch M, Hüge J, Kopka J, Hagemann M, Forchhammer K. Metabolic Changes in *Synechocystis* PCC6803 upon Nitrogen-Starvation: Excess NADPH Sustains Polyhydroxybutyrate Accumulation. Metabolites. 2013;3(1):101–18.

54. De Gelder J, De Gussem K, Vandenabeele P, Moens L. Reference database of Raman spectra of biological molecules. J Raman Spectrosc. 2007;38(9):1133–47.
55. Zhu G, Zhu X, Fan Q, Wan X. Raman spectra of amino acids and their aqueous solutions. Spectrochim Acta Mol Biomol Spectrosc. 2011;78(3):1187–95.
56. Hugli TE, Moore S. Determination of the tryptophan content of proteins by ion exchange chromatography of alkaline hydrolysates. J Biol Chem. 1972;247(9):2828–34.
57. Olcott HS, Fraenkel-Conrat H. Formation and loss of cysteine during acid hydrolysis of proteins. Role of tryptophan. J Biol Chem. 1947;171:583–94.
58. Yamada H, Moriya H, Tsugita A. Development of an acid hydrolysis method with high recoveries of tryptophan and cysteine for microquantities of protein. Anal Biochem. 1991;198(1):1–5.
59. Hesse A, Weller MG. Protein Quantification by Derivatization-Free High-Performance Liquid Chromatography of Aromatic Amino Acids. J Amino Acids. 2016;2016:8.
60. Holt LA, Milligan B, Roxburgh CM. Aspartic acid, asparagine, glutamic acid, and glutamine contents of wool and two derived protein fractions. Aust J Biol Sci. 1971;24(3):509–14.
61. Allen MM, Smith AJ. Nitrogen chlorosis in blue-green algae. Arch Für Mikrobiol. 1969;69(2):114–20.
62. Kumar Saha S, Uma L, Subramanian G. Nitrogen stress induced changes in the marine cyanobacterium *Oscillatoria willei* BDU 130511. FEMS Microbiol Ecol. 2003;45(3):263–72.
63. Stevens SE, Balkwill DL, Paone DAM. The effects of nitrogen limitation on the ultrastructure of the cyanobacterium *Agmenellum quadruplicatum*. Arch Microbiol. 1981;130(3):204–12.
64. Wanner G, Henkelmann G, Schmidt A, Köst HP. Nitrogen and Sulfur Starvation of the Cyanobacterium *Synechococcus* 6301 An Ultrastructural, Morphometrical, and Biochemical Comparison. Z Für Naturforschung C. 1986;41(7–8):741.
65. Krasikov V, Aguirre von Wobeser E, Dekker HL, Huisman J, Matthijs HCP. Time-series resolution of gradual nitrogen starvation and its impact on photosynthesis in the cyanobacterium *Synechocystis* PCC 6803. Physiol Plant. 2012;145(3):426–39.
66. Wood BR, Heraud P, Stojkovic S, Morrison D, Beardall J, McNaughton D. A portable Raman acoustic levitation spectroscopic system for the identification and environmental monitoring of algal cells. Anal Chem. 2005;77(15):4955–61.
67. Jehlicka J, Edwards HG, Oren A. Raman spectroscopy of microbial pigments. Appl Env Microbiol. 2014;80(11):3286–95.
68. Williams JGK. [85] Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. In: Methods in Enzymology. Academic Press; 1988. p. 766–78.
69. Dechatiwongse P, Srisamai S, Maitland G, Hellgardt K. Effects of light and temperature on the photoautotrophic growth and photoinhibition of nitrogen-fixing cyanobacterium *Cyanothece* sp. ATCC 51142. Algal Res. 2014;5(Supplement C):103–11.
70. Mohamed A, Jansson C. Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium Synechocystis 6803. Plant Mol Biol. 1989;13(6):693–700.
71. Gill RT, Katsoulakis E, Schmitt W, Taroncher-Oldenburg G, Misra J, Stephanopoulos G. Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis* sp. strain PCC 6803. J Bacteriol. 2002;184(13):3671–81.
72. Mullineaux CW. How do cyanobacteria sense and respond to light? Mol Microbiol. 2001;41(5):965–71.

73. Allakhverdiev SI, Murata N. Environmental stress inhibits the synthesis de novo of proteins involved in the photodamage-repair cycle of Photosystem II in *Synechocystis* sp. PCC 6803. Biochim Biophys Acta. 2004;1657(1):23–32.
74. Aro E-M, Virgin I, Andersson B. Photoinhibition of Photosystem II. Inactivation, protein damage and turnover. Biochim Biophys Acta BBA - Bioenerg. 1993;1143(2):113–34.
75. Zer H, Ohad I. Photoinactivation of photosystem II induces changes in the photochemical reaction center II abolishing the regulatory role of the QB site in the D1 protein degradation. Eur J Biochem. 1995;231(2):448–53.
76. Vass I, Kirilovsky D, Etienne AL. UV-B radiation-induced donor- and acceptor-side modifications of photosystem II in the cyanobacterium *Synechocystis* sp. PCC 6803. Biochemistry. 1999;38(39):12786–94.
77. Grossman AR, Schaefer MR, Chiang GG, Collier JL. The Responses of Cyanobacteria to Environmental Conditions: Light and Nutrients. In: Bryant DA, editor. The Molecular Biology of Cyanobacteria. Dordrecht: Springer Netherlands; 1994. p. 641–75.
78. Kumar K, Dasgupta CN, Nayak B, Lindblad P, Das D. Development of suitable photobioreactors for CO_2 sequestration addressing global warming using green algae and cyanobacteria. Bioresour Technol. 2011;102(8):4945–53.

Chapter 5: Characterizing *Synechocystis* sp. PCC 6803 with Rametrix™: Acetate, NaCl, and $MgSO_4$ deprivation-induced phenotypes

Imen Tanniche[1], Eva Collakova[2], Cynthia Denbow[2], Ryan S. Senger*[1,3]

[1]Department of Biological Systems Engineering; Virginia Tech; Blacksburg, VA

[2]School of Plant & Environmental Sciences; Virginia Tech; Blacksburg, VA

[3]Department of Chemical Engineering; Virginia Tech; Blacksburg, VA

* Corresponding Author
1230 Washington St.
301C HABB1
Blacksburg, VA 24061
Email: senger@vt.edu
Phone: 540-231-9501

Characterizing *Synechocystis* sp. PCC 6803 with Rametrix™: Acetate, NaCl, and $MgSO_4$ deprivation-induced phenotypes

Abstract

Background: During their long evolution, *Synechocystis* sp. PCC6803 developed a remarkable capacity to acclimate to diverse environmental conditions. In this study, Raman spectroscopy and Raman chemometrics tools (Rametrix™) were employed to investigate the phenotypic changes in response to external stressors and correlate specific Raman bands with their corresponding biomolecules determined with widely used analytical methods.

Methods: *Synechocystis* cells were grown in the presence of (i) acetate, (ii) NaCl and (iii) limiting levels of $MgSO_4$. Principal component analysis (PCA) and discriminant analysis of PCs (DAPC) were performed through Rametrix™ LITE Toolbox for MATLAB®. Next, validation of these models was realized via Rametrix™ PRO Toolbox where prediction of accuracy, sensitivity, and selectivity for an unknown Raman spectrum was calculated. These analyses were coupled with statistical tests (ANOVA and pairwise comparison) to determine statistically significant changes in the phenotypic responses. Finally, amino acid and fatty acid levels were measured with UPLC and GC-FID, respectively. The obtained data were correlated with previously established Raman bands assigned to these biomolecules.

Results: Distinguishable clusterings representative of phenotypic responses were observed based on the external stimuli (i.e. acetate, NaCl, $MgSO_4$, and controls grown on BG-11 medium) or its concentration when analyzing separately. For all these cases, Rametrix™ PRO was able to predict efficiently the corresponding concentration in the culture media for an unknown Raman spectra

with accuracy, sensitivity and selectivity exceeding random chance. Finally, overall strong correlations ($R > 0.7$) were obtained for all amino acids and fatty acids between well-established analytical methods and Raman bands.

5.1. Introduction

Microbial cells undergo phenotypic variations in response to changing environmental conditions. These changes directly affect the chemical composition of the cell, allowing it to maintain integrity, and the study of these changes is referred to as "microbial phenotyping" (1–4). For example, exposure to high salt concentrations (e.g., NaCl, $MgSO_4$) results in ionic and osmotic imbalances (5–10). This leads to the accumulation of carboxylic acids (e.g. acetate), which can be toxic and require a reorganization of the cell membrane for survival (11–14). Thus, cells have the ability to respond to such chemical stresses, adapt for survival, and even acquire tolerance (1). These characteristics are necessary in biotechnology, where cultures are engineered to produce products that are toxic in high concentrations (e.g., biofuels), operate in stressful environments (e.g., low pH), or perform bioremediation. Here, Raman spectroscopy was further developed as a microbial phenotyping methodology, with the focus on analyzing responses of cyanobacteria to salt and acid stressors.

In this context, the cyanobacterium *Synechocystis* sp. PCC6803 (hereafter *Synechocystis*) is of interest to biotechnology because of its ability to grow on low-cost resources (i.e. CO_2 and sunlight) and produce high value chemicals (15). This species demonstrates a highly versatile carbon metabolism, growing in autotrophic as well as under heterotrophic and mixotrophic conditions (16). It has been observed that acclimation of *Synechocystis* to high salt concentrations begins with adjustment of osmotic equilibrium with the outside medium (6,9,17), followed by the accumulation of the osmoprotectant glycosylglycerol (6,17–19). This osmoprotectant enhances the internal osmotic potential and protects proteins and membranes (20,21). Mechanisms of acetate toxicity have been studied mainly in Gram-negative bacteria and include alterations in cell membranes, active secretion, and activation of stress responses (22,23). However, relatively little

is known about the phenotypic responses of *Synechocystis* to carboxylic acids (e.g., acetate) and salts (e.g., NaCl and $MgSO_4$). Commonly, biological studies of physiological responses to chemical stressors have have taken place using genomic methods (9,17,24–26) to study gene expression and biochemical approaches (6,27) to characterize biomolecules. These require extensive sample preparation and long analysis times. Near real-time methods of analysis are needed alongside standardized procedures to provide rapid information to further understand microbial phenotypic response to environmental stress or for phenotype screening purposes. With such methods, cultures could be monitored in near real-time in biotechnological applications so that the appearance of unhealthy or unproductive phenotypes could be remedied quickly.

Raman spectroscopy has been proven a powerful analytical technique for the analysis of biological materials. It uses monochromatic light to provide a qualitative measurement of the biochemical composition of the biological specimen (28) and gives distinctive signals from macromolecules (i.e. proteins, lipids, carbohydrates, and nucleic acids) (29). Therefore, a characteristic fingerprint is assigned to each biological sample and allows a better understanding of its chemical composition. Raman spectroscopy allows near real-time and noninvasive acquisition of phenotype data from living systems (28–33). In addition, the analysis is label free, requires minimal or no sample preparation, and there is no spectral interference from water (30–32). The application of Raman spectroscopy is an expanding field and it includes fermentation monitoring (34–37), detection and identification of microorganisms (38–40), monitoring the kinetics of germination of individual *Clostridium difficile* spores (41) and detection of its toxins (42). It has been also demonstrated that Raman spectroscopy can be used in near-real time phenotyping of *E. coli* exposed to alcohol (31,32) and antibiotics (43), single cell phenotyping (33,44), and characterizing phenotypic differences among *E. coli* enriched for 1-butanol tolerance

(28). Thus, Raman spectroscopy was selected as an effective method for probing changes in *Synechocystis* phenotypes in near real-time when exposed to external stimuli.

In this research, Raman spectroscopy was used to perform microbial phenotyping of *Synechocystis* cultures growing in different concentrations of (i) acetic acid, (ii) NaCl, and (iii) $MgSO_4$. Two types of analyses were performed with the resulting Raman spectra: (i) Raman chemometrics (Rametrix™) using Raman spectra processed with principal component analysis (PCA) and linear discriminant analysis of principal components (DAPC) and (ii) Raman band intensity analysis (31). The Rametrix™ LITE Toolbox for MATLAB® (45) was used to build PCA and DAPC models, and the Rametrix™ PRO Toolbox (46) was used to validate these models and determine what characteristics of a growing *Synechocystis* culture could be determined given only a Raman scan. Routine analytical methodology was used to measure the levels of several biomolecules (e.g., fatty acids and amino acids) so correlations could be made with individual Raman bands.

5.2. Material and Methods

5.2.1. Bacterial strain and culture conditions

The wild-type and glucose tolerant strain *Synechocystis* sp. PCC6803 (ATCC®27184™) were used to create kanamycin resistant mutants for the comparative growth experiments (47). This strain was used with kanamycin to help prevent contamination. Cells were cultivated in the BG-11 medium (48) supplemented with 5 mM glucose. Kanamycin at a final concentration of 15 μg/mL was added to the medium. BG-11 plates containing additional 1.5% (w/v) phyto-agar, 1% (v/v) TES/NaOH buffer pH 8.2 and 3 g/L sodium thiosulfate were used for transformation and cell selection.

The kanamycin resistant strains were grown in three different conditions: (i) NaCl (50, 100, and 150 mM), (ii) acetate (7.5, 15, and 30 mM) and (iii) $MgSO_4$ (0, 15.6, 31.2, and 62.5 mM; representing 0%, 25%, 50% and 100% of BG-11). Cultures grown in BG-11 were used as controls. Each culture type condition experiment was performed in biological triplicate. All flask cultures were incubated at 25°C under continuous light, 140 rpm, and ambient CO_2. Cells were harvested at mid-exponential phase and washed twice with ice-cold purified water. 2 μL of cells were kept for Raman scans and the remaining pellet was freeze-dried and prepared for analyses by UPLC and GC-FID.

5.2.2. Raman Spectroscopy

To prepare samples for Raman analysis, 2 μL of washed cells were dried at room temperature on aluminum foil (three samples per biological replicate). Dried cells were analyzed using a PeakSeeker PRO-785 Raman microscope (Agiltron; Woburn ,MA) with 10X objective. Measurements were carried out using the following settings: laser excitation of 785 nm for 5 seconds with spectral resolution of 13 cm^{-1}. Twenty individual scans were taken per sample by focusing on different areas of the dried cells. Raman data were collected using RSIQ™ software.

5.2.3. Computational methods

Acquired Raman scans were processed and analyzed using the Rametrix™ LITE Toolbox (45) and Rametrix™ PRO Toolbox (46) for MATLAB®. MATLAB® R1018A with the Statistics and Machine Learning Toolbox was used for all calculations. Data analysis with Rametrix™ LITE Toolbox (45) consisted of averaging the 20 spectra replicates and then baselining with the Goldindec algorithm (53) (baseline polynomial order = 3, estimated peak ratio = 0.5, and smoothing window size = 5) and vector normalizing over the biological range (400 cm^{-1}-1800 cm^{-}

[1]). PCA was performed, and outlier scans were identified and excluded from analysis. Next, DAPC was applied, and several DAPC models were constructed by varying the number of principal components (PCs) used to build each one. DAPC models enabled the separation of Raman spectra according to a preset experimental factor. Here, the culture medium composition and environmental stimuli (acetate, NaCl, and $MgSO_4$) concentrations were used. The Rametrix™ PRO Toolbox (46) was used to apply leave-one-out analysis to DAPC clustering. This analysis determines the ability of DAPC model to predict correctly the classification when presented with an unknown Raman spectrum. The performance of DAPC models are quantitated by prediction accuracy (percent of unknown spectra classified correctly), sensitivity (true positive percentage), and selectivity (true negative percentage).

Other statistical analyses required the representation of the Rametrix™ LITE outputs (DAPC, PCA, and spectra) as a single value. From here, the distance formula is applied to determine the similarity between two spectra, where one is a reference. When using PCA data, the Total Principal Component Distance (TPD) is calculated according to Eq.1, where the first five PCs (representing more than 95% of the dataset variance) were used and

$$TPD = \sum_{i=1}^{5} \sqrt{\left(P_{x,i} - P_{reference,i}\right)^2} \qquad (Eq.1)$$

where $P_{x,i}$ is the i^{th} PC of sample x and $P_{reference,i}$ represents the i^{th} PC of the reference. Similarly, the distance calculation was applied to calculate the Total Spectral Distance (TSD) and the Total Canonical Distance (TCD) based on the first five canonicals in the DAPC model. More detailed descriptions of these calculations have been published (54). Reduction of the data to a single value allowed statistical analysis including analysis of variance (ANOVA), pairwise

comparisons using Tukey's honest significant difference (HSD), and regression. These analyses were performed to assess whether significant differences exist among the separated groups of spectra (ANOVA), to determine which group differences are statistically significant (pairwise comparisons), and establish a linear fit among data.

5.2.4. Biomass measurements

Fatty acid content was determined by GC-FID after direct hydrolysis and conversion to fatty acid methyl esters (FAME). Lyophilized cells were incubated with methanolic HCl and 10 µg heptadecanoic acid (C17:0) as internal standard at 75°C for 2 h to form FAME. GC-FID was done on an Agilent 7890A series GC equipped with FID. For fatty acid identification purposes, the same GC was coupled to a 5975C series single quadrupole MS and fatty acids were identified by comparing the spectra with the corresponding spectra in the NIST library (Agilent Technologies, Santa Clara, CA).

Protein hydrolysis was carried out in a custom-made teflon hydrolysis chamber (see photo in Supplementary Appendix D) under vacuum at 110°C in the presence of 6N HCl. 50 mM norvaline was used as an internal standard. Amino acid levels were determined after derivatization with AccQ-Tag™ reagent using a Waters Acquity™ H-class UPLC system equipped with a fluorescent detector (Waters Corporation, Mildford, MA).

5.2.5. Public access

The Rametrix™ LITE Toolbox is shared through GitHub with license agreement (https://github.com/SengerLab/RametrixLITEToolbox), and the Rametrix™ PRO Toolbox is also

accessible through GitHub with licence agreement (https://github.com/SengerLab/RametrixPROToolbox).

5.3. Results

5.3.1. Raman spectroscopy of Synechocystis grown in the presence of acetate and salts

Raman spectroscopy was used to identify phenotypic changes in *Synechocystis* induced by the presence of: (i) acetate, (ii) NaCl, and (iii) $MgSO_4$. Cells were scanned with Raman microscopy and the resulting spectra were analyzed using the Rametrix™ LITE (48) and PRO (46) Toolboxes. The obtained spectra were averaged, truncated (400 – 1800 cm^{-1}), baselined, and vector normalized prior to further processing using PCA and DAPC in the Rametrix™ LITE (48). Rametrix™ PRO (46) was used to determine the ability to predict a particular metabolic phenotype given a Raman spectrum of cells but unknown growth conditions. Next, statistical analyses (i.e., ANOVA and pairwise comparisons) were performed on TCD and TPD results. It was noticed previously that TPD and TSD produced similar results (54); thus, analysis of TSD was omitted. Finally, specific Raman bands assigned to certain metabolites (i.e. amino acids and fatty acids) were compared to standardized analytical methods (e.g., UPLC, GC-FID). The remainder of the results section is divided according to analysis type: (i) Rametrix™ analyses and (ii) correlation of individual Raman bands with well-established analytical approaches.

5.3.2. Rametrix™ analyses

5.3.2.1. Acetate-induced phenotypes

Raman spectra of cultures growing in three concentrations of acetate (7.5 mM, 15 mM and 30 mM) and the control (BG-11 without acetate) were analyzed by Raman spectroscopy and the

Rametrix™ LITE and Rametrix™ PRO Toolboxes. PCA, DAPC, and PC contributions results are shown in Fig. 5.1. PCA of Raman spectra (Fig. 5.1A) showed some separation of clusters in the two first PCs, which contained more than 68% of the dataset variance. This clustering is more noticeable with the DAPC model (Fig. 5.1B) constructed with 12 PCs (representing more than 99% of the dataset variance). Although four marked clusters were revealed, based on the concentration of acetate in the BG-11 medium, only three major groups were observed. These clusters suggest that phenotypic differences could be generated as a result of acetate addition to BG-11 medium and that cells grown in the presence of 7.5 mM and 15 mM could have similar phenotypic responses. Next, Raman shift contributions for both PCA and DAPC were studied to point to molecular differences between clusters of Raman scans. In this particular case, it provided information about the molecular differences between cultures grown in the presence and absence of acetate. Raman shift contributions between groups in PCA are shown in Fig. 5.1C and the contributions associated with DAPC model are represented in Fig. D1. Raman band assignments were selected based on published libraries (55–57). The full list of assignments is given in Supplementary Appendix D, and highlights are given in Table 5.1.

Next, TCD, TPD and TSD (both raw and mean values) were calculated using PCA and DAPC data and used in ANOVA and pairwise comparisons. Full results are given in the Supplementary Appendix D. In general, as the phenotype deviates from the reference (represented here by phenotypes of cells grown on BG-11 medium), the TCD (or the TPD) mean value should become larger. Similar phenotypes should have comparable values. In the acetate-induced phenotypes, TCD and TPD mean values increased proportionally to the increase of acetate concentration. Based on TCD data (using 12 PCs in DAPC model) ANOVA results indicated that differences between the groups are significant ($p < 0.001$). With TPD data, ANOVA returned a

significant *p*-value ($p < 0.001$). This result indicates that even though no clear clustering was observed in PCA (Fig. 5.1A), there are statistically significant differences between the groups. Pairwise comparisons identified all phenotypes (groups) to be significantly different based on the TCD data ($p < 0.001$). However, using TPD data, 7.5 mM and 15 mM induced phenotypes were found not significantly different ($p = 0.997$). The small difference in mean TPD and the near clustering observed in Fig. 5.1B are consistent with this result. Next, regression (Supplementary Appendix D) was computed between acetate concentrations (0 mM, 7.5 mM, 15 mM, 30 mM) and the TCD and TPD mean values, respectively. Results show a good fit and linear relationship among the data with a coefficient of determination (R^2) of 0.97 and 0.84 for TPD and TCD, respectively.

Rametrix^{TM} PRO was used to apply leave-one-out analysis to the DAPC clustering in Fig. 5.1B. The purpose of this analysis is to determine the ability of the DAPC model to predict precisely the classification (i.e. 0 mM, 7.5 mM, 15 mM, 30 mM of acetate) when presented with a Raman spectrum of *Synechocystis* cells grown in an unknown condition. Rametrix^{TM} PRO results are also given in Supplementary Appendix D. Here, four classification options, 0 mM, 7.5 mM, 15 mM and 30 mM, were assessed. To test the first condition, 0 mM was assigned as the positive condition. The remaining conditions were assigned as negative. All four conditions were tested as positive condition in this analysis. Here, the prediction accuracy reports the ability to classify correctly the positive and negative conditions. The sensitivity (true positive rate) is more important, where the selectivity (true negative rate) must be also high (i.e. above the random chance). With four different classifications, the random chance selectivity was calculated as 25% and the random chance selectivity as 75% for this dataset. Therefore, model performance sensitivity and selectivity must be higher than these values if the model is truly able to predict *Synechocystis* phenotypes based on Raman spectra. Highlighted Rametrix^{TM} PRO results are given

in Table 5.2, and all sensitivity and selectivity values exceeded random chance values, with selectivity and sensitivity values ranging between 56-100% and 81-100%, respectively. Thus, the approach showed some success at identifying acetate-induced metabolic phenotypes in *Synechocystis*.

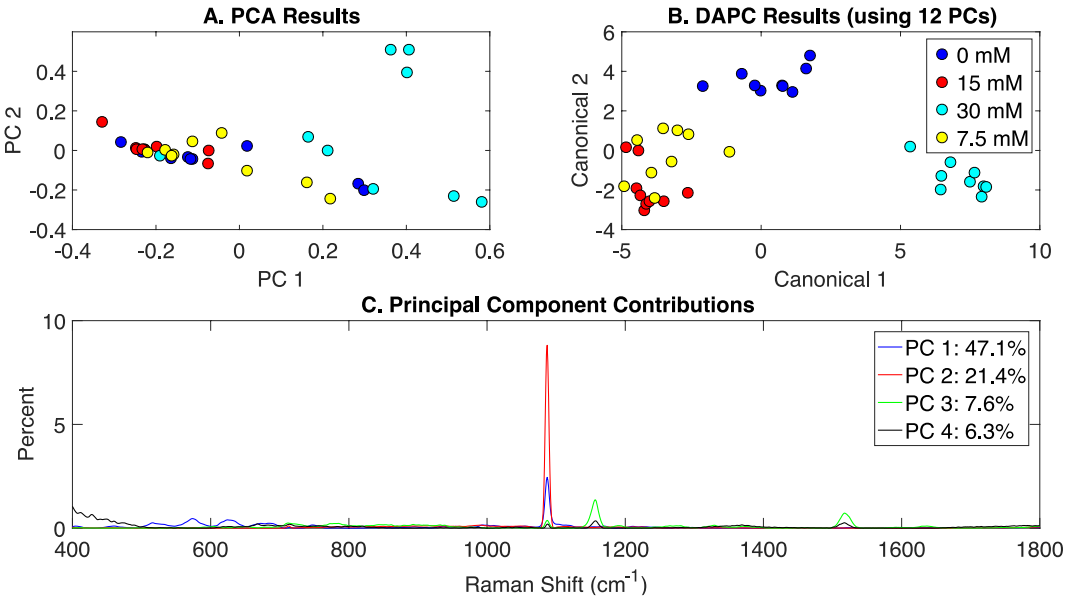


Figure 5.1. Rametrix™ LITE analysis of acetate induced phenotypes. A. PCA, B. DAPC and C. Raman shifts contributions between groups in PCA. (Acetate concentration in mM, 0 mM represents BG-11 medium).

5.3.2.2. NaCl-induced metabolic phenotypes

In order to determine the effect of NaCl addition to BG-11 medium on *Synechocystis* metabolic phenotypes, cells were grown in BG-11 at three salt concentrations (50 mM, 100 mM, and 150 mM) and compared to a control (cells grown in BG-11 only, 0 mM NaCl). PCA, DAPC and PC contribution results generated by the Rametrix™ LITE are represented in Fig. 5.2. PCA of Raman spectra (Fig. 5.2A), again, showed no apparent clustering in the first two PCs, which

comprised about 69% of the dataset variance. However, the application of DAPC (Fig. 5.2B) demonstrated four clusters representative of the three different concentrations of NaCl and the control. These results were obtained using 12 PCs (representing over 99% of the dataset variance). Based on the presence of NaCl, two major groups were noticed (Fig. 5.2B), suggesting significant cellular phenotype differences could occur in response to as little as 50 mM NaCl in BG-11 medium. Next, Raman shift contributions to the PCA and DAPC results were examined. Raman shift contributions between the groups in PCA are shown in Fig. 5.2C, and those related to DAPC contributions are represented in Fig. D2. Specific lists of Raman band contributions and molecular assignments (for both PCA and DAPC) are given in Supplementary Appendix D, and highlights are given in Table 5.1.

DAPC model and PCA were used to compute TCD, TPD, TSD, and their mean values. Multi-way ANOVA and pairwise comparisons were used to assess these parameters, and full results are given in Supplementary Appendix D. As mentioned previously, the TCD and TPD mean values increase with the increased difference from the reference phenotype. Using the TPD mean values, as the concentration of NaCl increased, the distance also increased. However, a small difference was observed between the 50 mM and the 100 mM TPD means, suggesting comparable phenotypes. The TCD mean values and the clustering in the Fig. 5.2B confirm this result. ANOVA revealed that molecular phenotypes induced at different NaCl concentrations were significantly different ($p < 0.001$) for TCD and TPD data. Pairwise comparison allowed the identification of significantly different phenotypes. Based on TPD data, the phenotypic responses associated with 100 mM NaCl were not significantly different from the BG-11 medium ($p = 0.165$) and 50 mM ($p = 0.091$). Pairwise comparisons also showed that the phenotypes observed at 150 mM NaCl were not statistically different from the phenotypes at 50 mM ($p = 0.04$). On the other hand, based on

the TCD data, the phenotypes resulting from 50 mM, 100 mM, and 150 mM NaCl treatments were all statistically different from the BG-11 medium control (all *p*-values < 0.001), but they were not different from one another (all *p*-values > 0.027). Finally, Regression (Supplementary Appendix D) was applied between NaCl concentrations and the TCD and TPD mean values. A coefficient of determination (R^2) of 0.64 and 0.51 were obtained for TPD and TCD, respectively.

Rametrix™ PRO was used to apply leave-one-out analysis to DAPC in Fig. 5.2B. Similar to the previous analysis, four classification options were included: 0 mM, 50 mM, 100 mM, and 150 mM NaCl. Each of these conditions was selected as the positive condition to generate accuracy, sensitivity, and selectivity results. As mentioned previously, sensitivity and selectivity values should exceed those of the random chance values to confirm at least some function of the model in correctly identifying whether a *Synechocystis* culture was exposed to NaCl given only a Raman spectrum of cells. Highlighted results for a DAPC model built with 12 PCs (representing more than 99% of dataset variability) are given in Table 5.2. Results for additional DAPC models are given in Supplementary Appendix D. The best-performing model (Table 5.2) returned sensitivity and selectivity values that exceeded the random chance values. In addition, 0 mM classification showed the highest sensitivity and selectivity rates (100%). This suggests that NaCl-induced phenotypes may occur below 50 mM and remain somewhat similar through 150 mM. The BG-11 control phenotypes were significantly distinguishable from the ones observed for the other three conditions, as shown in Fig. 5.2B.

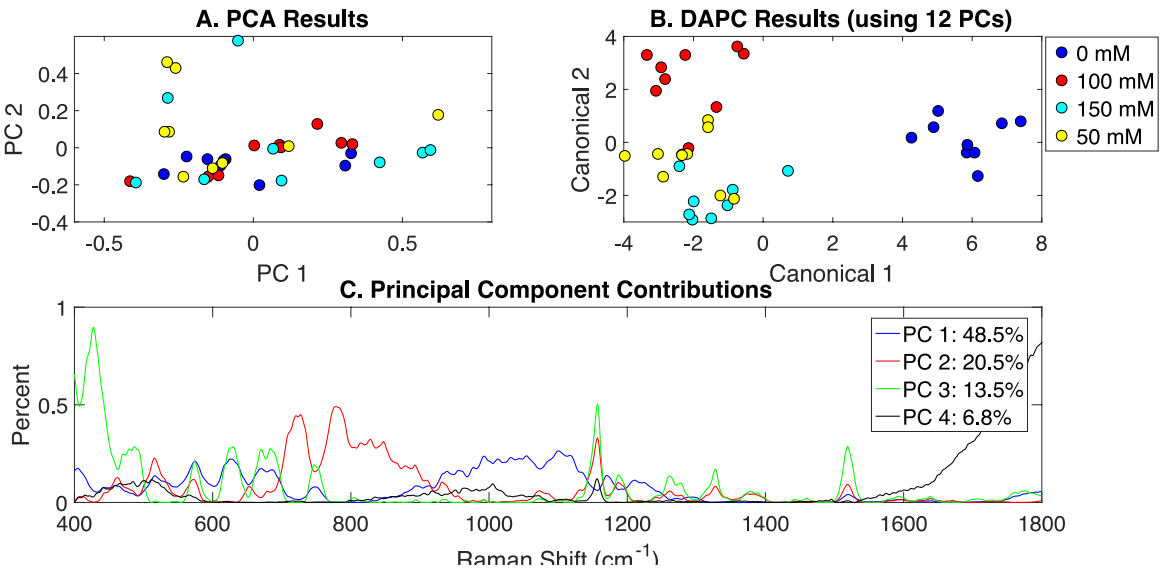


Figure 5.2. Rametrix™ LITE analysis of NaCl induced phenotypes. A. PCA, B. DAPC and C. Raman shifts contributions between groups in PCA. (Sodium chloride concentration in mM, 0 mM represents BG-11 medium).

5.3.2.3. $MgSO_4$-induced phenotypes

Synechocystis cells growing under different levels of $MgSO_4$ present in BG-11 medium were analyzed using Raman spectroscopy. PCA, DAPC and PCA contributions generated by the Rametrix™ LITE Toolbox are presented in Fig. 5.3. Again, no clustering was observed with PCA of Raman spectra (Fig. 5.3A) in the first two PCs (representing over 75% of the dataset variance). Application of DAPC using 12 PCs (representative of more than 99% of the dataset variance) revealed four distinguishable clusters (Fig. 5.3B), based on the concentration of $MgSO_4$ in BG-11 medium. These clusters formed two separate groups representing cells grown under limiting concentrations of $MgSO_4$ and BG-11 medium, confirming that limiting $MgSO_4$ in the medium is linked with altering cellular phenotypes. Analysis of Raman shift contributions to the PCA and

DAPC groups was also performed. Full results are given in Supplementary Appendix D, and highlights are given in Table 5.1. DAPC contributions are also identified in Fig. D3.

TSD, TCD and TPD values were calculated based on spectra, DAPC, and PCA results. These were analyzed further by ANOVA and pairwise comparisons. All results are available in the Supplementary Appendix D. In general, TCD and TPD increased with the decrease of $MgSO_4$ concentration, with the 15.6 mM classification mean values being closer to the reference when using the TPD means, compared to TCD mean values. Next, ANOVA and pairwise comparison tests were performed. It was found that differences between the groups (in both DAPC and PCA) are statistically significant ($p < 0.001$). Pairwise comparisons allowed the identification of significantly different phenotypes. With both TPD and TCD data, all phenotypes were different from those observed when cells were grown in the original BG-11 medium containing 62.5 mM $MgSO_4$. These results confirm that even at 50% reduction in $MgSO_4$ from the original BG-11 medium gives rise to various phenotypes. Results also show that phenotypes associated with the reduced $MgSO_4$ content (31 mM) are not significantly different from 0 mM ($p = 0.096$) and phenotypes with even more reduced content (15 mM) ($p = 0.032$), when using TPD data. Similar results were found when using TCD values. Regression analysis (Supplementary Appendix D) between $MgSO_4$ concentrations and both TCD and TPD mean values was performed. Results show a good fit and linear relationship between the data for both cases with a coefficient of determination (R^2) of 0.81 and 0.76 for TPD and TCD, respectively.

For Rametrix^{TM} PRO leave-one-out analysis on the DAPC model (Fig. 5.3B), four classification options were included for the four tested concentrations of $MgSO_4$. Each one of these conditions was treated as a positive condition to generate accuracy, sensitivity, and selectivity results. Again, results for several DAPC models are given in Supplementary Appendix D, and

results for the DAPC model generated with 12 PCs (the best performer) are given in Table 5.2. Results confirmed the BG-11 cluster (62.5 mM $MgSO_4$) (100% sensitivity and selectivity) is significantly different from the other clusters. Again, all sensitivity and selectivity values exceeded random chance, meaning the DAPC model was capable of applying these classifications for an unknown Raman scan of *Synechocystis* with accuracy greater than random chance.

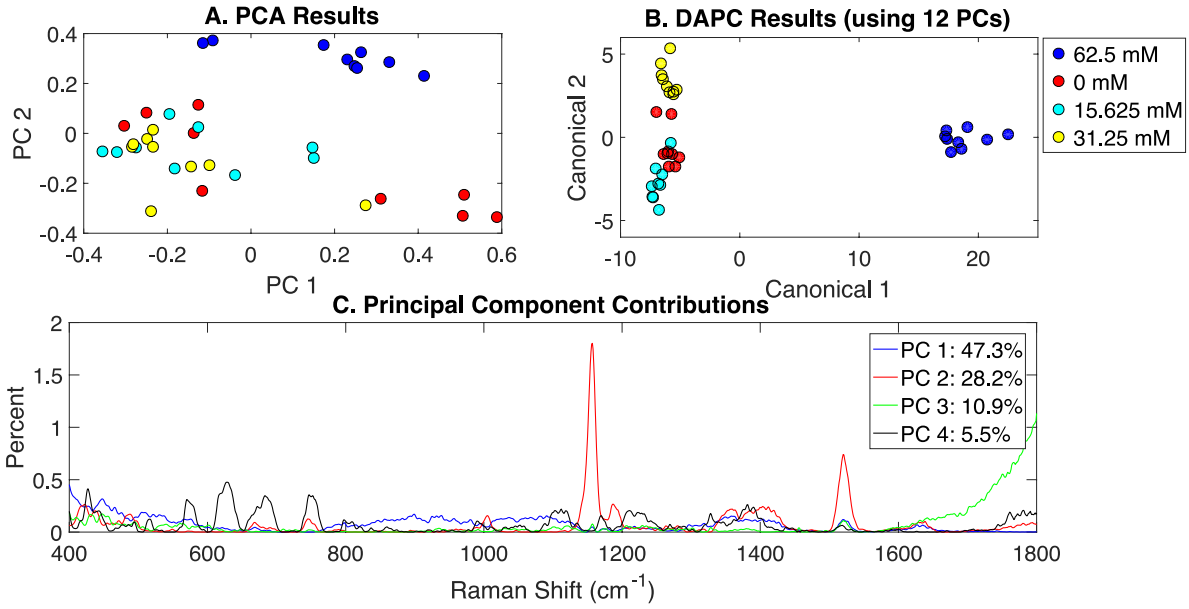


Figure 5.3. Rametrix™ LITE analysis of $MgSO_4$ induced phenotypes. A. PCA, B. DAPC and C. Raman shifts contributions between groups in PCA. (Magnesium sulfate concentration in mM, 62.5 mM represents BG-11 medium).

5.3.2.4. Classification of all phenotypes

All Raman scans performed in this study were compiled in a single dataset and included the growth conditions in the presence of the following environmental stimuli: (i) acetate, (ii) NaCl, (iii) $MgSO_4$ and (iv) BG-11 medium (set as the control). Data analysis included the generation of PCA and DAPC models and validation with the leave-one-out analysis using Rametrix™ LITE

and Rametrix^{TM} PRO Toolboxes, respectively. PCA results are given in Fig. 5.4A and show some clustering, based on the two first two PCs which covered over 59% of the dataset variance. A DAPC model was constructed based on 15 PCs (representing over 99% of the dataset variance) and indicates four distinct clusters based on the presence or absence of environmental stimuli in the BG-11 medium (Fig. 5.4B). These clusters could be the result of potential differences in the cellular phenotypes obtained with every growth condition. Most importantly, a major segregation was observed for cells grown under $MgSO_4$, clearly separating from the other conditions. Raman shift contributions to PCA (Fig. 5.4C) and DAPC groups (Fig. D4) are listed in the Supplementary Appendix D.

The analysis of TCD and TPD, using the BG-11 medium as a reference, allowed the classification of phenotypes. Both TCD and TPD mean values indicated the distance from the control (indicating more different phenotypes): (i) acetate-induced phenotypes, (ii) NaCl-induced phenotypes and (iii) $MgSO_4$-induced phenotypes. It appears that $MgSO_4$ limitation generated the most significant changes in the cellular phenotypic response (also shown in Fig. 5.4B). In addition, multi-way ANOVA indicated statistically significant differences between the groups representing each phenotype ($p < 0.001$). Using TPD data, pairwise comparisons indicated that all groups are significantly different from each other, only acetate (over all concentrations together) was not statistically different from BG-11 ($p = 0.021$). Based on TCD data (using 15 PCs), all phenotypes are significantly different ($p < 0.001$).

Next, Rametrix^{TM} PRO was applied to determine if the phenotypes generated by the DAPC model could be predicted from Raman spectra (Supplementary Appendix D). Results are highlighted in Table 5.2. Again, all results exceeded the random chance values, indicating ability

by the DAPC model to identify environmental stimuli exposure with accuracy much better than random chance.

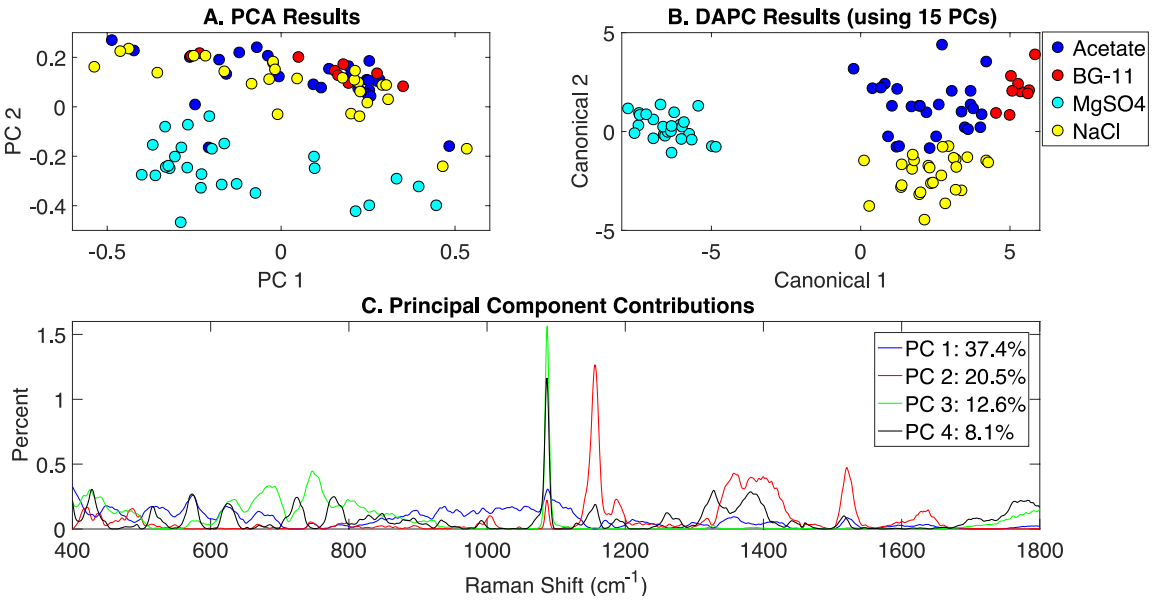


Figure 5.4. Rametrix™ LITE analysis of all phenotypes. A. PCA, B. DAPC and C. Raman shifts contributions between groups in PCA.

5.3.3. Correlation of individual Raman bands with well-established analytical approaches

Another method of comparing phenotypes is analysis of Raman bands assigned to functional groups of biomolecules. These signals are then correlated with results from accepted analytical methods such as GC-FID and UPLC for fatty acids and amino acids, respectively. The analysis of individual Raman band intensities was conducted based on the sets of Raman bands identified in previous research (55–57) and all band assignment are available in Supplementary Appendix D (Table D13).

5.3.3.1. Acetate study

Synechocystis has the ability to grow mixotrophically in the presence of glucose and acetate (58); however, its utilization as the sole carbon source is limited as it lacks major enzymes required for acetate uptake (59–61). Acetate could serve as a supplementary precursor for the synthesis of certain biomass building blocks such as amino acids (59,61) and lipids (59). The levels of amino acids (protein-derived and free amino acids) were determined by both Raman spectroscopy and UPLC analysis (Fig. D5). The general trend is an increase in amino acids levels for the three tested concentration of acetate (7.5 mM, 15 mM, 30 mM) compared to the control with an over-production of amino acids at 15 mM acetate and comparable levels at 7.5 mM and 30 mM acetate.

Raman assigned bands for the different amino acids (56,57) were correlated with the UPLC data (Fig. D6). Strong to very strong correlations were obtained for most amino acids with the correlation coefficients (R) varying between 0.61 and 0.97. Ala, Leu, Lys, Met, and Ser had moderate correlations with values of R ranging from 0.43 to 0.59. Since the acidic protein hydrolysis method used to release amino acid prevents from determining Cys and Trp levels (62–64), these amino acids were predicted using Raman band assignment alone (Fig. D7). The same trend was observed where Cys and Trp levels achieved a higher expression at 15 mM of acetate.

Similarly, fatty acid content was evaluated and then compared to Raman bands (57,65). Overall, there was an increase in fatty acid levels, mainly for C16:0, C18:2 and C18:3, where 15 mM showed the highest values (Fig. D8). However, minor changes were observed for C16:1 and C18:1. Next, Raman signals corresponding those fatty acids were compared with GC-FID results (Fig. S9). Very strong correlations were observed for C16:0 (R = 0.92) and C18:2 (R = 0.82),

strong correlations were found for C18:1 and C18:3 with the correlation coefficients of 0.67 and 0.68, respectively. C16:1 had a moderate correlation between both data (R = 0.46).

5.3.3.2. NaCl study

Cyanobacteria is known to be able to tolerate high salt concentrations (7). It has been shown previously that long-term salt stress induces accumulation of specific proteins (25,26,66,67) and alteration in fatty acids composition (68–71). Cells growing in different concentrations of NaCl were examined for their amino acid and fatty acid levels with UPLC and GC-FID, respectively. Raman signals corresponding to these biomolecules were used for comparison. Amino acid analysis revealed an increase in their levels proportionally to the amount of NaCl supplied to the media with a distinguishable over production of amino acids in the cells grown in the presence of 150 mM NaCl (Fig. D10). Comparison with the associated Raman bands showed a very strong correlation between the UPLC and Raman microscopy data (Fig. D11), where correlation values were comprised between 0.7-0.99 and a strong correlation obtained with Val (R = 0.64). As stated previously, Cys and Trp are only predicted using Raman spectroscopy since they degrade during acidic hydrolysis (62–64). Again, Raman assignments were able to resolve the patterns of these two amino acids showing an increase of their levels mainly at 150 mM of acetate (Fig. D12).

Finally,GC-FID analysis of the levels of free and lipid-derived fatty acids (Fig. D13) revealed an increase in the levels of C16:0, C18:2, and C18:3 in cells subjected to 50 mM of NaCl. In addition, no changes were observed for C16:1 and C18:1 levels. These results were compared with specific Raman bands assigned to each fatty acid (Fig. D14). With the exception of C18:2, very strong correlations ($0.7 < R < 0.95$) were obtained. A weak correlation was observed for the

C18:2 between the Raman and GC-FID data (R = 0.28), which improved to 0.7 when the control (0 mM NaCl) was removed.

5.3.3.3. $MgSO_4$ study

The correlation of the individual Raman bands with their associated biomolecules could not be achieved for this study. $MgSO_4$ is the major source of both elements in the BG-11 medium and sufficient amounts of the biomass could not be obtained for cells deprived of this salt, which prevented UPLC and GC-FID analyses. Magnesium is an important cofactor for activities of many enzymes, while sulfur represents one of the major macronutrients (0.6% of the biomass DW) (72,73) as it is incorporated in proteins, lipids, and other vital compounds. In addition, sulfur is needed for the thioredoxin-mediated regulation of involved in photosynthesis and respiratory activities (24,74). Therefore, obtained biomass after growth in limiting $MgSO_4$, which is the major source of both Mg and S in the BG-11 medium, was very little even at high volume of BG-11 and prevented the analysis with UPLC or GC-FID.

5.4. Discussion

Newly developed methods of analysis by Raman spectroscopy using Rametrix^{TM} enabled the assessment of phenotypic variations among *Synechocystis* cells growing in the presence of external substrates or stressors (i.e. acetate, NaCl, and $MgSO_4$). Two different approaches of analysis were undertaken: (i) overall metabolic phenotype changes using Rametrix^{TM} chemometric methods and (ii) correlation of specific Raman bands with their corresponding biomolecules and comparisons with data obtained by well-established analytical methods (i.e. UPLC and GC-FID).

Synechocystis can grow in the presence of acetate in heterotrophic conditions (58). However, this microorganism cannot utilize acetate efficiently as it does not express necessary transporters (59–61). Phenotype changes due to different concentrations of acetate were observed through RametrixTM LITE clustering of Raman spectral data and validated through statistical analyses (ANOVA and pairwise comparisons). Rametrix^{TM} PRO displayed higher values of sensitivity and selectivity (100% and 89%, respectively) over random chance values, supporting that our model can predict effectively the level of acetate in the culture media. Previous studies revealed that acetate is used for the biosynthesis of amino acids (59,61) and lipids (59) which is in agreement with our results (both with Raman spectroscopy and standard analytical methods). An over-accumulation of amino acids and fatty acids was observed mainly for acetate concentration of 15 mM. However, addition of 30 mM acetate into the BG-11 medium did not have a stronger effect compared to the 15 mM acetate. It is obvious that the amount of acetate supplied is limited by diffusion through the cell membrane (59–61) and that 15 mM might be the maximum concentration that could induce changes in the levels of the tested biomolecules. This result does not exclude the changes in the phenotypic responses observed with the clustering in Rametrix^{TM} LITE and suggest that other changes might happen at 30 mM acetate. According to DAPC contributions, these changes include altered polysaccharide composition, and the levels and types of nucleotides and carotenoids.

NaCl-induced phenotypes were clearly different among the cells grown in the presence of different concentrations of this salt. ANOVA confirmed these results and pairwise comparisons revealed which phenotypes are significantly different from each other. In addition, Rametrix^{TM} PRO was able to predict with accuracy, sensitivity (100%) and selectivity (100%) greater than random chance the concentration of NaCl in an unknown Raman spectrum of *Synechocystis* cells.

Data from both Raman and the chromatography-based methods showed strong correlations and were in an agreement with previously published data. In fact, salt stress triggers specific cellular responses including synthesis of stress related proteins (25,26,66,67) and alterations in fatty acid composition (68–71). In particular, 150 mM NaCl induces an over-production of amino acids (exceeding a 15-fold increase compared to the control) and significant changes in fatty acid composition.

Clustering of Raman spectral data via Rametrix^{TM} LITE allowed the differentiation of metabolic phenotypes among *Synechocystis* cells in the presence of $MgSO_4$ limiting concentrations. These results were validated by ANOVA and pairwise comparison. In addition, the DAPC model could predict the $MgSO_4$ levels in the culture media with 100% selectivity and sensitivity. The importance of Raman spectroscopy is well-demonstrated also, as it provided valuable information with a limited biomass sample and without upstream preparation when the traditional analytical approaches could not be implemented. In addition, Raman shift contributions to the PCA and DAPC allowed the identification of specific biomolecules that contributed to the phenotypic changes observed with Rametrix^{TM} LITE. These biomolecules included carotenoids, amino acids, lipids, phospholipids, and saccharides. Specific responses observed in photosynthetic organisms during acclimation to Mg and S deprivation include glycogen accumulation (72), changes in amino acids composition (74) as well as lipids (75) and degradation of the photosynthetic system (i.e. chlorophyll a, PBS and carotenoids) (76–78). Interestingly, most of these biomolecules were identified with Raman shift contributions, suggesting that our findings are in an agreement with previously established results.

In conclusion, Raman spectroscopy enabled the identification of phenotypic changes in cells growing in the presence of different concentrations of external stressor or nutrient. In general,

a strong correlation was observed between Raman band intensities and measurements obtained with well-established analytical methods. Most importantly, Raman spectroscopy enabled the identification of the major biomolecules contributing to the phenotypic changes in samples with limited biomass, demonstrating its usefulness.

Table 5.1. Highlighted Raman shift contributions to PCA and DAPC models.

Study	**Model**	**Biomolecules and Raman bands (cm^{-1})**
Acetate-induced	PCA	Glycerol (630 cm^{-1}), amino acid (573 cm^{-1}) protein (1156 cm^{-1}), and carotenoids (1156, 1518 cm^{-1})
Acetate-induced	DAPC	Proteins (524, 618, 1005, 1155 cm^{-1}), carotenoids (1155, 1528 cm^{-1}), amino acids (532, 638, 715, 920, 988, 1005, 1196 1327, 1495, 1550, 1558, 1615 cm^{-1}), phosphatidylinositol (415 cm^{-1}), polysaccharides (477 cm^{-1}), DNA/RNA (746, 1510 cm^{-1}), and lipids (1086 cm^{-1})
NaCl-induced	PCA	Glycerol (630 cm^{-1}); proteins, amides and lipids (1260 cm^{-1}); carotenoids and β-carotene (1157, 1520 cm^{-1}); and amino acids (515, 573, 669, 684, 1188 cm^{-1})
NaCl-induced	DAPC	Sterols (430 cm^{-1}), phosphatidylinositol (594 cm^{-1}), DNA/RNA (746 cm^{-1}), DNA (1090, 1329 cm^{-1}), amino acids (730, 854, 1040, 1495 cm^{-1}), carotenoids and β-carotene (1157, 1517 cm^{-1}), nucleic acids (650, 1200, 1329 cm^{-1}), and lipids (1255 cm^{-1})
$MgSO_4$-induced	PCA	Sterols (429 cm^{-1}), glycerol (630 cm^{-1}), carotenoids and carotene (1157, 1520 cm^{-1}), amino acids (573, 684, 1188 cm^{-1}), lipids (1380 cm^{-1}), and phospholipids (1330 cm^{-1})
$MgSO_4$-induced	DAPC	Carotenoids and carotene (1157, 1520 cm^{-1}) and amino acids (850, 992 cm^{-1}) in addition to phosphatidylinositol (412 cm^{-1}) and saccharides (850, 868, 1456 cm^{-1})

Table 5.2. Highlighted Rametrix™ PRO results.

Study	Classification Predicted	Sensitivity	Selectivity
All Studies	Random chance**	25%	75%
Acetate-induced	0 mM	100%	89%
Acetate-induced	7.5 mM	67%	81%
Acetate-induced	15 mM	56%	100%
Acetate-induced	30 mM	56%	96%
NaCl-induced	0 mM	100%	100%
NaCl-induced	50 mM	33%	96%
NaCl-induced	100 mM	44%	100%
NaCl-induced	150 mM	44%	89%
$MgSO_4$-induced	62.5 mM	100%	100%
$MgSO_4$-induced	31.2 mM	89%	96%
$MgSO_4$-induced	15.6 mM	100%	89%
$MgSO_4$-induced	0 mM	56%	78%
All phenotypes	BG-11 control	100%	100%
All phenotypes	Acetate-induced	56%	100%
All phenotypes	NaCl-induced	71%	95%
All phenotypes	$MgSO_4$-induced	93%	100%

* All DAPC models were built with 12 PCs representing 99% of the dataset variance.
** The random chance sensitivity and selectivity values were calculated statistically, not by Rametrix™ PRO. They were the same for all studies here.

References

1. Nicolaou SA, Gaida SM, Papoutsakis ET. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. Metab Eng. 2010;12(4):307–31.
2. Alsaker KV, Paredes C, Papoutsakis ET. Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*. Biotechnol Bioeng. 2010;105(6):1131–47.
3. Sellick CA, Reece RJ. Eukaryotic transcription factors as direct nutrient sensors. Trends Biochem Sci. 2005;30(7):405–12.
4. Tummala SB, Junne SG, Paredes CJ, Papoutsakis ET. Transcriptional analysis of product-concentration driven changes in cellular programs of recombinant Clostridium acetobutylicum strains. Biotechnol Bioeng. 2003;84(7):842–54.
5. Lusk JE, Williams RJ, Kennedy EP. Magnesium and the growth of *Escherichia coli*. J Biol Chem. 1968;243(10):2618–24.
6. Marin K, Stirnberg M, Eisenhut M, Kramer R, Hagemann M. Osmotic stress in *Synechocystis* sp. PCC 6803: low tolerance towards nonionic osmotic stress results from lacking activation of glucosylglycerol accumulation. Microbiology. 2006;152(Pt 7):2023–30.
7. Reed RH, Warr SRC, Richardson DL, Moore DJ, Stewart WDP. Multiphasic osmotic adjustment in a euryhaline cyanobacterium. FEMS Microbiol Lett. 1985;28(3):225–9.
8. Mikkat S, Galinski EA, Berg G, Minkwitz A, Schoor A. Salt Adaptation in Pseudomonads: Characterization of Glucosylglycerol-Synthesizing Isolates from Brackish Coastal Waters and the Rhizosphere. Syst Appl Microbiol. 2000;23(1):31–40.
9. Karandashova IV, Elanskaya IV. Genetic Control and Mechanisms of Salt and Hyperosmotic Stress Resistance in Cyanobacteria. Russ J Genet. 2005;41(12):1311–21.
10. Markovitz A, Sylvan S. Effect of sodium sulfate and magnesium sulfate on heteropolysaccharide synthesis in gram-negative soil bacteria. J Bacteriol. 1962;83(3):483–9.
11. Papoutsakis ET, Bussineau CM, Chu IM, Diwan AR, Huesemann M. Transport of substrates and metabolites and their effect on cell metabolism (in butyric-acid and methylotrophic fermentations). Ann N Acad Sci. 1987;506:24–50.
12. Russell JB. Another explanation for the toxicity of fermentation acids at low pH: anion accumulation versus uncoupling. J Appl Bacteriol. 1992;73(5):363–70.
13. Lasko DR, Schwerdel C, Bailey JE, Sauer U. Acetate-specific stress response in acetate-resistant bacteria: an analysis of protein patterns. Biotechnol Prog. 1997;13(5):519–23.
14. Roe AJ, O'Byrne C, McLaggan D, Booth IR. Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. Microbiology. 2002;148(Pt 7):2215–22.
15. Yu Yi, You Le, Liu Dianyi, Hollinshead Whitney, Tang YinjieJ, Zhang Fuzhong. Development of *Synechocystis* sp. PCC 6803 as a phototrophic cell factory. Mar Drugs. 2013;11(8):2894–916.
16. Vermaas W. Molecular genetics of the cyanobacterium *Synechocystis* sp. PCC 6803: Principles and possible biotechnology applications. J Appl Phycol. 1996;8:263–73.
17. Marin K, Kanesaki Y, Los DA, Murata N, Suzuki I, Hagemann M. Gene expression profiling reflects physiological processes in salt acclimation of *Synechocystis* sp. strain PCC 6803. Plant Physiol. 2004;136(2):3290–300.

18. Reed RH, Borowitzka LJ, Mackay MA, Chudek JA, Foster R, Warr SRC, et al. Organic solute accumulation in osmotically stressed cyanobacteria. FEMS Microbiol Lett. 1986;39(1):51–6.
19. Hagemann M. Chapter Two - Genomics of Salt Acclimation: Synthesis of Compatible Solutes among Cyanobacteria. In: Chauvat F, Cassier-Chauvat C, editors. Advances in Botanical Research. Academic Press; 2013. p. 27–55.
20. Borges N, Ramos A, Raven ND, Sharp RJ, Santos H. Comparative study of the thermostabilizing properties of mannosylglycerate and other compatible solutes on model enzymes. Extremophiles. 2002;6(3):209–16.
21. Hincha DK, Hagemann M. Stabilization of model membranes during drying by compatible solutes involved in the stress tolerance of plants and microorganisms. Biochem J. 2004;383(Pt 2):277–83.
22. Isken S, de Bont JA. Bacteria tolerant to organic solvents. Extremophiles. 1998;2(3):229–38.
23. Sardessai Y, Bhosle S. Tolerance of bacteria to organic solvents. Res Microbiol. 2002;153(5):263–8.
24. Zhang Z, Pendse ND, Phillips KN, Cotner JB, Khodursky A. Gene expression patterns of sulfur starvation in *Synechocystis* sp. PCC 6803. BMC Genomics. 2008;9(1):344.
25. Huang F, Fulda S, Hagemann M, Norling B. Proteomic screening of salt-stress-induced changes in plasma membranes of *Synechocystis* sp. strain PCC 6803. PROTEOMICS. 2006;6(3):910–20.
26. Fulda S, Huang F, Nilsson F, Hagemann M, Norling B. Proteomics of *Synechocystis* sp. strain PCC 6803. Eur J Biochem. 2000;267(19):5900–7.
27. Zavřel T, Očenášová P, Červený J. Phenotypic characterization of *Synechocystis* sp. PCC 6803 substrains reveals differences in sensitivity to abiotic stress. PloS One. 2017;12(12):e0189130–e0189130.
28. Butler HJ, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, et al. Using Raman spectroscopy to characterize biological materials. Nat Protoc. 2016;11(4):664–87.
29. Das RS, Agrawal YK. Raman spectroscopy: Recent advancements, techniques and applications. Vib Spectrosc. 2011;57(2):163–76.
30. Freedman BG, Zu TNK, Wallace RS, Senger RS. Raman spectroscopy detects phenotypic differences among *Escherichia coli* enriched for 1-butanol tolerance using a metagenomic DNA library. Biotechnol J. 2016;11(7):877–89.
31. Zu TNK, Athamneh AIM, Wallace RS, Collakova E, Senger RS. Near-Real-Time Analysis of the Phenotypic Responses of *Escherichia coli* to 1-Butanol Exposure Using Raman Spectroscopy. 2014;
32. Zu NT, Athamneh IA, Senger SR. Characterizing the Phenotypic Responses of *Escherichia coli* to Multiple 4-Carbon Alcohols with Raman Spectroscopy. Fermentation. 2016;2(1).
33. García-Timermans C, Rubbens P, Heyse J, Props R, Kerckhof F-M, Skirtach A, et al. Characterizing phenotypic heterogeneity in isogenic bacterial populations using flow cytometry and Raman spectroscopy. bioRxiv. 2019;545681.
34. Ewanick SM, Thompson WJ, Marquardt BJ, Bura R. Real-time understanding of lignocellulosic bioethanol fermentation by Raman spectroscopy. Biotechnol Biofuels. 2013;6(1):28.
35. Sivakesava S, Irudayaraj J, Demirci A. Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. J Ind Microbiol Biotechnol. 2001;26(4):185–90.

36. Zu TNK, Liu S, Germane KL, Servinsky MD, Gerlach ES, Mackie DM, et al. Predictive modeling in *Clostridium acetobutylicum* fermentations employing Raman spectroscopy and multivariate data analysis for real-time culture monitoring. Vol. 9863. SPIE; 2016.
37. Olson ML, Johnson J, Carswell WF, Reyes LH, Senger RS, Kao KC. Characterization of an evolved carotenoids hyper-producer of *Saccharomyces cerevisiae* through bioreactor parameter optimization and Raman spectroscopy. J Ind Microbiol Biotechnol. 2016;43(10):1355–63.
38. Nelson WH, Manoharan R, Sperry JF. UV Resonance Raman Studies of Bacteria. Appl Spectrosc Rev. 1992;27(1):67–124.
39. Pahlow S, Meisel S, Cialla-May D, Weber K, Rösch P, Popp J. Isolation and identification of bacteria by means of Raman spectroscopy. Adv Drug Deliv Rev. 2015;89:105–20.
40. Kirschner C, Maquelin K, Pina P, Ngo Thi NA, Choo-Smith LP, Sockalingum GD, et al. Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. J Clin Microbiol. 2001;39(5):1763–70.
41. Wang S, Shen A, Setlow P, Li Y. Characterization of the Dynamic Germination of Individual *Clostridium difficile* Spores Using Raman Spectroscopy and Differential Interference Contrast Microscopy. J Bacteriol. 2015;197(14):2361.
42. Koya SK, Brusatori M, Martin JV, Yurgelevic S, Huang C, Liberati DM, et al. Rapid Detection of *Clostridium difficile* Toxins in Serum by Raman Spectroscopy. J Surg Res. 2018;232:195–201.
43. Athamneh AI, Alajlouni RA, Wallace RS, Seleem MN, Senger RS. Phenotypic profiling of antibiotic response signatures in *Escherichia coli* using Raman spectroscopy. Antimicrob Agents Chemother. 2014;58(3):1302–14.
44. Sun S, Wang X, Gao X, Ren L, Su X, Bu D, et al. Condensing Raman spectrum for single-cell phenotype analysis. BMC Bioinformatics. 2015;16 Suppl 18:S15.
45. Fisher AK, Carswell WF, Athamneh AIM, Sullivan MC, Robertson JL, Bevan DR, et al. The Rametrix™ LITE Toolbox v1.0 for MATLAB®. J Raman Spectrosc. 2018;49:885–896.
46. Senger RS, Robertson JL. The Rametrix™ PRO Toolbox for MATLAB. 2019;(Under Review).
47. Tanniche I, Fisher AK, Gillam F, Collakova E, Zhang C, Bevan DR, et al. λ-PCR for precise DNA assembly and modification. 2019;(Under Review).
48. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. J Gen Microbiol. 1978;111:1–61.
49. Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G. Purification and properties of unicellular blue-green algae (order Chroococcales). Bacteriol Rev. 1971;35(2):171–205.
50. Grigorieva G, Shestakov S. Transformation in the cyanobacterium *Synechocystis* sp. 6803. FEMS Microbiol Lett. 1982;13(4):367–70.
51. Desai SH, Atsumi S. Photosynthetic approaches to chemical biotechnology. Curr Opin Biotechnol. 2013;24(6):1031–6.
52. Schwarzkopf M, Yoo YC, Huckelhoven R, Park YM, Proels RK. Cyanobacterial phytochrome2 regulates the heterotrophic metabolism and has a function in the heat and high-light stress response. Plant Physiol. 2014;164(4):2157–66.
53. Liu J, Sun J, Huang X, Li G, Liu B. Goldindec: A Novel Algorithm for Raman Spectrum Baseline Correction. Appl Spectrosc. 2015 Jul;69(7):834–42.

54. Senger RS, Du P, DeLaTorre Campos D, Carswell WF, Webster K, Sullivan MC, et al. Assessment of urine specimen storage conditions using Raman Chemometric Urinalysis™ (Rametrix™). 2019;(Under Review).
55. Movasaghi Z, Rehman S, Rehman IU. Raman Spectroscopy of Biological Tissues. Appl Spectrosc Rev. 2007;42(5):493–541.
56. Zhu G, Zhu X, Fan Q, Wan X. Raman spectra of amino acids and their aqueous solutions. Spectrochim Acta Mol Biomol Spectrosc. 2011;78(3):1187–95.
57. De Gelder J, De Gussem K, Vandenabeele P, Moens L. Reference database of Raman spectra of biological molecules. J Raman Spectrosc. 2007;38(9):1133–47.
58. Wu GF, Shen ZY, Wu QY. Modification of carbon partitioning to enhance PHB production in *Synechocystis* sp. PCC6803. Enzyme Microb Technol. 2002;30(6):710–5.
59. Thiel K, Vuorio E, Aro E-M, Kallio PT. The effect of enhanced acetate influx on *Synechocystis* sp. PCC 6803 metabolism. Microb Cell Factories. 2017;16(1):21.
60. Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, et al. Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803. PLOS Comput Biol. 2013;9(6):e1003081.
61. Varman AM, Yu Y, You L, Tang YJ. Photoautotrophic production of D-lactic acid in an engineered cyanobacterium. Microb Cell Factories. 2013;12(1):117.
62. Hugli TE, Moore S. Determination of the tryptophan content of proteins by ion exchange chromatography of alkaline hydrolysates. J Biol Chem. 1972;247(9):2828–34.
63. Olcott HS, Fraenkel-Conrat H. Formation and loss of cysteine during acid hydrolysis of proteins. Role of tryptophan. J Biol Chem. 1947;171:583–94.
64. Yamada H, Moriya H, Tsugita A. Development of an acid hydrolysis method with high recoveries of tryptophan and cysteine for microquantities of protein. Anal Biochem. 1991;198(1):1–5.
65. Czamara K, Majzner K, Pacia MZ, Kochan K, Kaczor A, Baranska M. Raman spectroscopy of lipids: a review. J Raman Spectrosc. 2015;46(1):4–20.
66. Hagemann M, Techel D, Rensing L. Comparison of salt- and heat-induced alterations of protein synthesis in the cyanobacterium *Synechocystis* sp. PCC 6803. Arch Microbiol. 1991;155(6):587–92.
67. Hagemann M, Erdmann N. Activation and pathway of glucosylglycerol synthesis in the cyanobacterium *Synechocystis* sp. PCC 6803. Microbiology. 1994;140(6):1427–31.
68. Huflejt ME, Tremolieres A, Pineau B, Lang JK, Hatheway J, Packer L. Changes in membrane lipid composition during saline growth of the fresh water cyanobacterium Synechococcus 6311. Plant Physiol. 1990;94(4):1512–21.
69. Singh S, Sinha RP, Häder D-P. Role of Lipids and Fatty Acids in Stress Tolerance in Cyanobacteria. In 2002.
70. Khomutov G, Fry IV, Huflejt ME, Packer L. Membrane lipid composition, fluidity, and surface charge changes in response to growth of the fresh water cyanobacterium *Synechococcus* 6311 under high salinity. Arch Biochem Biophys. 1990;277(2):263–7.
71. Allakhverdiev SI, Nishiyama Y, Suzuki I, Tasaka Y, Murata N. Genetic engineering of the unsaturation of fatty acids in membrane lipids alters the tolerance of *Synechocystis* to salt stress. Proc Natl Acad Sci. 1999;96(10):5862–7.
72. Schwarz R, Forchhammer K. Acclimation of unicellular cyanobacteria to macronutrient deficiency: emergence of a complex network of cellular responses. Microbiology. 2005;151(8):2503–14.

73. Kim HW, Vannela R, Zhou C, Rittmann BE. Nutrient acquisition and limitation for the photoautotrophic growth of <i>Synechocystis<i> sp. PCC6803 as a renewable biomass source. Biotechnol Bioeng. 2011;108(2):277–85.
74. Kiyota H, Ikeuchi M, Hirai MY. Response of Amino Acid Metabolism to Sulfur Starvation in *Synechocystis* sp. PCC 6803. In: De Kok LJ, Tabe L, Tausz M, Hawkesford MJ, Hoefgen R, McManus MT, et al., editors. Sulfur Metabolism in Plants. Springer Netherlands; 2012. p. 53–9. (Proceedings of the International Plant Sulfur Workshop).
75. Sato N, Kamimura R, Kaneta K, Yoshikawa M, Tsuzuki M. Species-specific roles of sulfolipid metabolism in acclimation of photosynthetic microbes to sulfur-starvation stress. PLOS ONE. 2017;12(10):e0186154.
76. Richaud C, Zabulon G, Joder A, Thomas J-C. Nitrogen or Sulfur Starvation Differentially Affects Phycobilisome Degradation and Expression of the nblA Gene in *Synechocystis* Strain PCC 6803. J Bacteriol. 2001;183(10):2989–94.
77. Schmidt A, Erdle I, Köst H-P. Changes of C-Phycocyanin in *Synechococcus* 6301 in Relation to Growth on various Sulfur Compounds Materials and Methods. Z Für Naturforschung C. 2014;37(10):870–876.
78. Jensen TE, Rachin JW. Effect of varying sulphur deficiency on structural components of a cyanobacterium *Synechococcus leopoliensis*: a morphometric study. Cytobios. 1984;

Chapter 6: Screening metagenomic DNA libraries for enzymes able to detoxify deoxynivalenol

Imen Tanniche[1], Nina Wilson[2], Nicole McMaster[2], David Schmale[2], Ryan S. Senger*[1,3]

[1] Department of Biological Systems Engineering; Virginia Tech; Blacksburg, VA

[2] School of Plant & Environmental Sciences; Virginia Tech; Blacksburg, VA

[3] Department of Chemical Engineering; Virginia Tech; Blacksburg, VA

* Corresponding Author
1230 Washington St.
301C HABB1
Blacksburg, VA 24061
Email: senger@vt.edu
Phone: 540-231-9501

Screening metagenomic DNA libraries for enzymes able to detoxify deoxynivalenol

Abstract

Enzymatic inactivation of fungal toxins is an attractive strategy for the detoxification of crops and their food derivatives. The trichothecene mycotoxin deoxynivalenol (DON) represents one of the common contaminant of grain. DON-degrading bacteria and fungi are likely to be found in metagenomic environments in contact with this mycotoxin. In this study, we used metagenomic DNA library preparation and enrichment for the selection of cells capable of enzymatic degradation of DON. Different library fragments were cultured in the presence of DON as the unique carbon source and in the presence of ferulic acid, an efflux pump inhibitor. Five cultures showed capabilities to reduce DON concentrations by up to 30% in cultures. However, these strains failed to out-perform the DON-sensitive control culture with no DNA library. On the other hand, with the control culture also showing 30% DON degradation, it is possible that DON degradation mechanism(s) have evolved in DON-sensitive control culture after several rounds of sub-culturing in the presence of DON.

6.1. Introduction

Mycotoxins are highly toxic secondary metabolites produced by fungi and compromises the health of human and animals (1–3). The trichothecenes are a major class of mycotoxins and comprise over 180 structurally related toxic compounds (3). These mycotoxins bind to eukaryotic ribosomes and prevent protein and polypeptide synthesis (1,3–6). The trichothecene mycotoxin deoxynivalenol (DON) is produced by the fungus *Fusarium graminearum* in cereal crops in the field or during storage (1–3,7). Exposure to low doses of DON causes diarrhea, vomiting, and gastroenteritis (1,3), and higher doses result in severe health issues such as immune system dysregulation (1,2,8) and reproductive and teratogenic disorders (1,2,7,9,10). One of the most important physicochemical properties of DON is its heat-stability (2,11), which contributes to the risk of its occurrence in food. Therefore, degradation of DON, by either chemical, physical, or biological means (12–14) is of great importance.

A large number of mycotoxin-degrading microorganisms are found in rumen fluids (15–17). Soil is also a promising source of DON-degrading bacteria and fungi. Previous studies focused on the isolation of these organisms from metagenomic environmental samples (18–23). In general, biological degradation of DON includes enzymatic degradation or transformation to a less toxic products (12,24–27).

A genomic DNA library is a collection of DNA fragments representing nearly the entire genome of the organism(s) from which the DNA was extracted. These fragments are then transformed into expression vectors (28,29). DNA library enrichment is a combinatorial approach that involves exposing the culture to a selective pressure and screening for the targeted trait (19,28). The DNA fragment(s) conferring that trait are then enriched in the culture. In the current

study, metagenomic DNA libraries isolated from natural soil environments and plants (18) were transformed into a DON sensitive yeast strain (18) are and enriched in the presence of toxic levels of DON. In particular, these strains were incubated in the presence of 100 µg/mL DON as the sole carbon source and different concentrations of ferulic acid. Here, ferulic acid is used to prevent the enrichment of efflux pumps (30), which was seen previously, since those membrane proteins are responsible for exporting trichothecenes and not their detoxification (26,31). In this case, DNA library fragments that confer higher growth rates and DON detoxification will ultimately dominate the culture to where a DON-detoxifying enzyme can be identified and isolated for overexpression in pure cultures.

6.2. Material and Methods

6.2.1. Strain and culture media

Wild-type *Saccharomyces cerevisiae* strains RW2802 (PDR5 leu2 ura3-52 met5) and JG436 (pdr5::Tn5 leu2 ura3-35 met5) were used to generate the DON sensitive strain (18). Genes encoding for transporters under stress conditions *PDR5* and *PDR10* or *PDR10* and acetyltransferase (*AYTI*) were knocked out (Table 6.1).

The DON sensitive strain containing a metagenomic DNA library was grown for enrichment on synthetic drop-out medium without uracil (SD-U) containing all standard amino acids (76 mg/liter final) plus leucine (380 mg/liter final) and lacked uracil (Sigma, St. Louis, MO). Yeast extract-peptone-dextrose (YPD) medium (1% yeast extract, 2% peptone, and 2% glucose) was used to grow *S. cerevisiae* before the enrichment.

6.2.2. Microbial DNA library fragments

Microbial DNA library fragments were generated from one pure culture of bacteria (*Achromobacter*), and three mixed cultures obtained from plant and soil environmental samples (18). DOP-PCR (28) was used to construct the metagenomic DNA library from DNA extracted from soil microbes. Next, the library DNA was cloned into pYesDEST-52 using the Invitrogen Gateway LR Clonase kit (Thermo Fisher Scientific, Waltham, MA) and then transformed into both DON sensitive yeast strains.

6.2.3. Enrichment and isolation

For culture enrichment, samples of the different genomic library (yeast strain transformed with different DNA fragments) were suspended in YPD media and grown for 2 days at 28℃. Aliquots of the cultures (200 µL) were used to inoculate 2 mL of SD-U media and were incubated at 28℃ for two days. Assays were then prepared in a 96-well plate. 50 µL of the culture was used to inoculate a 200 µL SD-U media supplemented with 100 µg/mL DON and ferulic acid at the following concentration: 0.5 mM and 1.5 mM. A control without ferulic acid was also included. The cultures were incubated at 28℃ and 140 rpm for 3 days. Afterward, 50 µL of the culture was subcultured into a freshly prepared 200 µL SD-U media supplemented with 100 µg/mL DON and the different concentration of ferulic acid.

6.2.4. Determination of DON level with GC/MS

An aliquot of the resulting cultures (50 µL) was used to screen for DON degradation with GC/MS following standard protocols (32). Each sample was dried under a nitrogen flow at 55℃ and then derivatized at room temperature with 99 µL of N-trimethylsilylimidazole (TMSI) and 1

µL of trimethylchlorosilane (TMCS) for 20 min. Then, 500 µL of isooctane containing 0.5 µg/g mirex were added to the samples and vortexed. Finally, 500 µL of water was added to quench the reaction and a volume of 150 µL was used for analysis. An Agilent 6890/5975 system was used for GC/MS analysis operating in selected ion monitoring (SIM) mode.

6.3. Results

An initial screening of nine DNA libraries was performed using DON (100 µg/mL) as the sole carbon source and in the presence of 0.5 mM and 1 mM of ferulic acid, respectively. The DON sensitive strain transformed with the empty vector pYesDEST52 (Table 6.1) was used as the negative control. Five cultures, which sustained growth, were identified where DON relative concentration (compared to the empty vector) decreased up to 30% in presence of 0.5 and 1.5 mM ferulic acid (Fig. 6.1AB). About 40% decrease in DON concentration was observed in the absence of ferulic acid (Fig. 6.1C). These five cultures were subjected to more subculturing. Results indicated that there is no additional degradation of DON compared to the control. However, as shown in Fig. 6.2, about a 30% decrease in DON was observed for the DON-sensitive control culture. This suggests that this strain may have evolved a natural mechanism(s) for degrading DON. Known DON degradation products have not been detected by GC/MS analysis.

6.4. Discussion

Enzymes responsible for the detoxification of DON may be used for the development of inactivation procedures at agricultural level, animal feeds and food production (12). In this study, we were able to isolate five culture that reduced concentration of DON by about 30%, when grown on DON as their sole carbon source and in the presence of efflux pump inhibitor ferulic acid. These cultures provide a first step toward the identification of an enzyme able of the detoxification of

DON. Even though further enrichment of these cultures did not yield lower levels of this mycotoxin, the DON-sensitive yeast strain control showed decreased concentrations of DON. We hypothesize that the DON-sensitive strain is acquiring DON-degradation capabilities after several rounds of sub-culturing in the presence of DON. This should be investigated further by additional sub-culturing and possibly through the addition of chemical mutagens to increase the culture evolution rate. This observed inconsistency might also be due to the storage of the enriched libraries in glycerol at -80°C, which could affect their ability to modify DON. In addition, after the long-term storage, cells are usually revived in YPD and SD-U without DON, which might allow for the growth of all the organisms in the mixed culture.

6.5. Future development

In addition to the subculturing mentioned above, the adjustment of the enrichment technique in this study is required in order to maintain the acquired ability of cells to degrade DON. It is suggested to do more round of subculturing (19) in order to isolate a strain capable of detoxifying DON. The next step should be the identification by sequencing of the gene responsible of DON disappearance and its mechanism. Another alternative is the use of cell-free-system (12) alongside with the genomic DNA library to identify enzymes capable of degrading DON. In fact, based on preliminary studies, where *Saccharomyces cerevisiae* cellular membrane was disrupted by sonication, the crude extract sustained the production of two reporter fluorescent proteins (GFP and mCherry) for about 96 h (Fig. 6.3 AB). This represents a promising result toward the use of this technique in future work.

Table 6.1. *Saccharomyces cerevisiae* strains description

Strain	Culture ID	Description	Reference
RW2802	NA	Parent strain to DON sensitive yeast strain 5 (DSY5)	invitrogen
JG436	NA	Parent strain to DON sensitive yeast strain 8 (DSY8)	invitrogen
DSY5	NA	DON sensitive yeast strain 5 - has two efflux pumps knocked out, PDR5 and PDR10	(18)
DSY8	NA	DON sensitive yeast strain 8 - has one efflux pump knocked out, PDR10 and an acetyltransferase gene knocked out, AYT1	(18)
pYesDEST52	NA	DON sensitive yeast strain transformed with Empty vector pYesDEST52	(18)
S3B YG5 5 PYO	mixed culture 1	pYesDEST vector without the TEF1	(18)
S3B YG3 5 PYW	mixed culture 1	pYesDEST vector containing the TEF1 promoter	(18)
S3B YG5 5 PYW	mixed culture 1	pYesDEST vector containing the TEF1 promoter	(18)
S3I YG5 5 PYO	mixed culture 3	pYesDEST vector without the TEF1	(18)
S3I YG5 5 PYW	mixed culture 3	pYesDEST vector containing the TEF1 promoter	(18)
S3H YBAC YG5 PYO	pure culture 1	pYesDEST vector without the TEF1	(18)
S3H YBAC YG5 5 PYW	mixed culture 2	pYesDEST vector containing the TEF1 promoter	(18)
S3H YG5 5 PYW	mixed culture 2	pYesDEST vector containing the TEF1 promoter	(18)
S3H YG3 5 PYO	mixed culture 2	pYesDEST vector without the TEF1	(18)

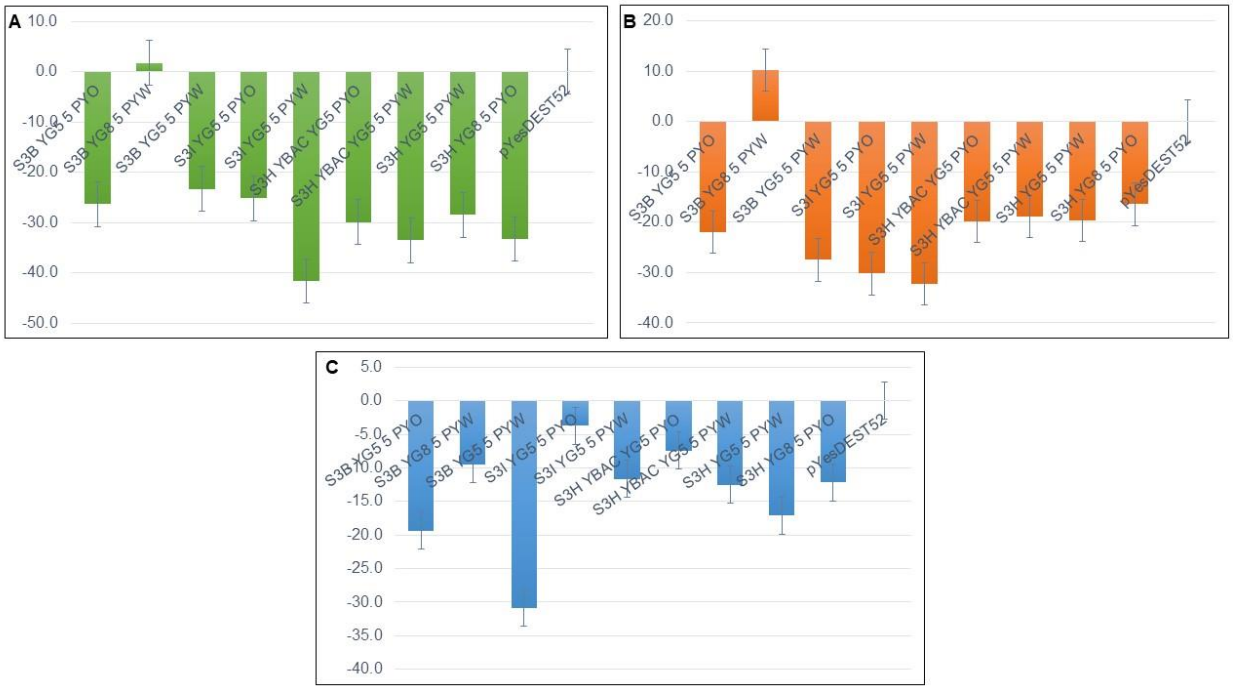


Figure 6.1. DON modification by metagenomic DNA library in *Saccharomyces cerevisiae*. (A) Percent change in the presence of 0 mM ferulic acid, (B), Percent change in the presence of 0.5 mM ferulic acid and (C) Percent change in the presence of 1.5 mM ferulic acid. Results are averages of two biological replicates.

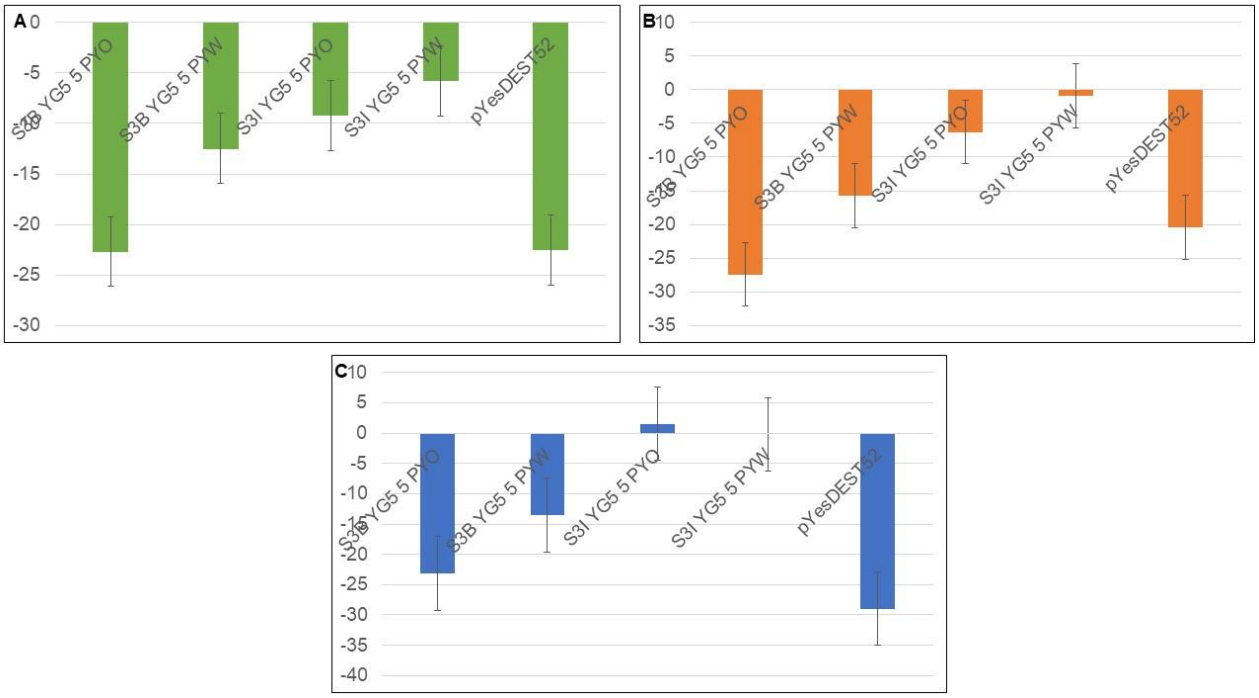


Figure 6.2. DON modification by enriched metagenomics library. (A) Percent change in the presence of 0 mM ferulic acid, (B), Percent change in the presence of 0.5 mM ferulic acid and (C) Percent change in the presence of 1.5 mM ferulic acid. Results are averages of three biological replicates.

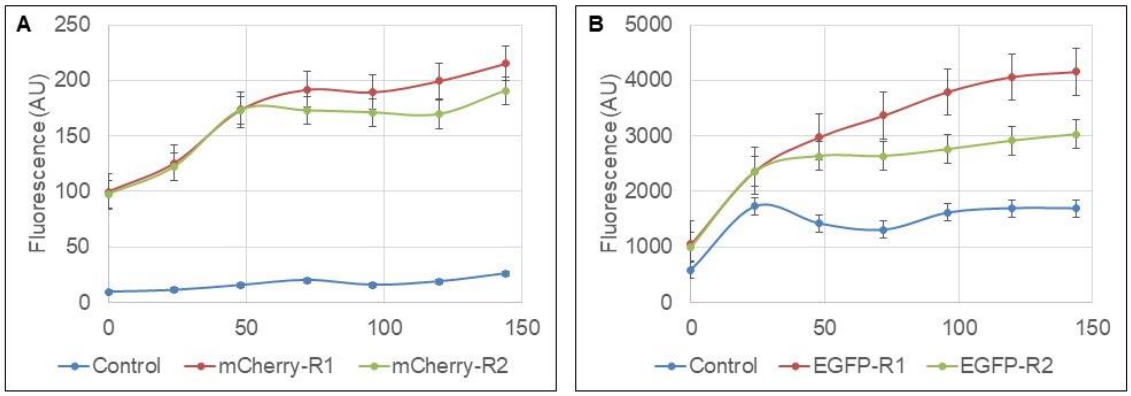


Figure 6.3. Expression of fluorescent proteins in *Saccharomyces cerevisiae* cell-free-system. (A) mCherry fluorescence level, (B) EGFP fluorescence level. Results are averages of 12 biological replicates.

References

1. Pestka JJ. Deoxynivalenol: mechanisms of action, human exposure, and toxicological relevance. Arch Toxicol. 2010;84(9):663–79.
2. Sobrova P, Adam V, Vasatkova A, Beklova M, Zeman L, Kizek R. Deoxynivalenol and its toxicity. Interdiscip Toxicol. 2010;3(3):94–9.
3. Pestka JJ. Deoxynivalenol: Toxicity, mechanisms and animal health risks. Anim Feed Sci Technol. 2007;137(3):283–98.
4. Ueno Y. Toxicological features of T-2 toxin and related trichothecenes. Toxicol Sci. 1984;4(2part2):124–32.
5. Carter CJ, Cannon M. Structural requirements for the inhibitory action of 12, 13-epoxytrichothecenes on protein synthesis in eukaryotes. Biochem J. 1977;166(3):399–409.
6. Ueno Y, Nakajima M, Sakai K, Ishii K, Sato N, Shimada N. Comparative Toxicology of Trichothec Mycotoxins: Inhibition of Protein Synthesis in Animal Cells. J Biochem (Tokyo). 1973 Aug 25;74(2):285–96.
7. Döll S, Dänicke S. The Fusarium toxins deoxynivalenol (DON) and zearalenone (ZON) in animal feeding. Prev Vet Med. 2011;102(2):132–45.
8. Pestka JJ. Mechanisms of Deoxynivalenol-Induced Gene Expression and Apoptosis. Food Addit Contam Part Chem Anal Control Expo Risk Assess. 2008;25(9):1128–40.
9. Döll S, Dänicke S, Ueberschär K-H, Valenta H, Schnurrbusch U, Ganter M, et al. Effects of graded levels of Fusarium toxin contaminated maize in diets for female weaned piglets. Arch Anim Nutr. 2003;57(5):311–34.
10. EFSA J. Opinion of the scientific panel on contaminants in the food chain on a request from the commission related to zearalenone as undesirable substance in animal feed. EFSA J. 2004;89:1–35.
11. Hughes DM, Gahl MJ, Graham CH, Grieb SL. Overt signs of toxicity to dogs and cats of dietary deoxynivalenol. J Anim Sci. 1999;77(3):693–700.
12. Karlovsky P. Biological detoxification of fungal toxins and its use in plant breeding, feed and food production. Nat Toxins. 1999;7(1):1–23.
13. Karlovsky P. Biological detoxification of the mycotoxin deoxynivalenol and its use in genetically engineered crops and feed additives. Appl Microbiol Biotechnol. 2011;91(3):491–504.
14. McCormick SP. Microbial Detoxification of Mycotoxins. J Chem Ecol. 2013;39(7):907–18.
15. Yu H, Zhou T, Gong J, Young C, Su X, Li X-Z, et al. Isolation of deoxynivalenol-transforming bacteria from the chicken intestines using the approach of PCR-DGGE guided microbial selection. BMC Microbiol. 2010;10(1):182.
16. Kollarczik B, Gareis M, Hanelt M. In vitro transformation of the Fusarium mycotoxins deoxynivalenol and zearalenone by the normal gut microflora of pigs. Nat Toxins. 1994;2(3):105–10.
17. Young JC, Zhou T, Yu H, Zhu H, Gong J. Degradation of trichothecene mycotoxins by chicken intestinal microbes. Food Chem Toxicol. 2007 Jan 1;45(1):136–43.
18. Wilson NM, McMaster N, Gantulga D, Soyars C, McCormick SP, Knott K, et al. Modification of the Mycotoxin Deoxynivalenol Using Microorganisms Isolated from Environmental Samples. Toxins. 2017;9(4).

19. Shima J, Takase S, Takahashi Y, Iwai Y, Fujimoto H, Yamazaki M, et al. Novel detoxification of the trichothecene mycotoxin deoxynivalenol by a soil bacterium isolated by enrichment culture. Appl Environ Microbiol. 1997;63(10):3825–30.
20. Fuchs E, Binder EM, Heidler D, Krska R. Structural characterization of metabolites after the microbial degradation of type A trichothecenes by the bacterial strain BBSH 797. Food Addit Contam. 2002;19(4):379–86.
21. Völkl A, Vogler B, Schollenberger M, Karlovsky P. Microbial detoxification of mycotoxin deoxynivalenol. J Basic Microbiol. 2004;44(2):147–56.
22. Sato I, Ito M, Ishizaka M, Ikunaga Y, Sato Y, Yoshida S, et al. Thirteen novel deoxynivalenol-degrading bacteria are classified within two genera with distinct degradation mechanisms. FEMS Microbiol Lett. 2012;327(2):110–7.
23. Ikunaga Y, Sato I, Grond S, Numaziri N, Yoshida S, Yamaya H, et al. Nocardioides sp. strain WSN05-2, isolated from a wheat field, degrades deoxynivalenol, producing the novel intermediate 3-epi-deoxynivalenol. Appl Microbiol Biotechnol. 2011;89(2):419–27.
24. Awad WA, Ghareeb K, Bohm J, Zentek J. Decontamination and detoxification strategies for the *Fusarium* mycotoxin deoxynivalenol in animal feed and the effectiveness of microbial biodegradation. Food Addit Contam Part -- Chem Anal Control Expo Risk Assess. 2010;27(4):510–20.
25. Carere J, Hassan YI, Lepp D, Zhou T. The enzymatic detoxification of the mycotoxin deoxynivalenol: identification of DepA from the DON epimerization pathway. Microb Biotechnol. 2017;11(6):1106–11.
26. Khatibi PA, Newmister SA, Rayment I, McCormick SP, Alexander NJ, Schmale DG. Bioprospecting for trichothecene 3-O-acetyltransferases in the fungal genus *Fusarium* yields functional enzymes with different abilities to modify the mycotoxin deoxynivalenol. Appl Environ Microbiol. 2011;77(4):1162–70.
27. Vanhoutte I, De Mets L, De Boevre M, Uka V, Di Mavungu JD, De Saeger S, et al. Microbial Detoxification of Deoxynivalenol (DON), Assessed via a Lemna minor L. Bioassay, through Biotransformation to 3-epi-DON and 3-epi-DOM-1. Toxins. 2017;9(2).
28. Freedman BG, Zu TNK, Wallace RS, Senger RS. Raman spectroscopy detects phenotypic differences among *Escherichia coli* enriched for 1-butanol tolerance using a metagenomic DNA library. Biotechnol J. 2016;11(7):877–89.
29. Theophilus BDM. Genomic DNA Libraries. In: Rapley R, Walker JM, editors. Molecular Biomethods Handbook [Internet]. Totowa, NJ: Humana Press; 1998 [cited 2019 Apr 7]. p. 145–52. Available from: https://doi.org/10.1007/978-1-59259-642-3_14
30. Sundaramoorthy NS, Mitra K, Ganesh JS, Makala H, Lotha R, Bhanuvalli SR, et al. Ferulic acid derivative inhibits NorA efflux and in combination with ciprofloxacin curtails growth of MRSA in vitro and in vivo. Microb Pathog. 2018;124:54–62.
31. Alexander NJ, McCormick SP, Hohn TM. TRI12, a trichothecene efflux pump from *Fusarium sporotrichioides*: gene isolation and expression in yeast. Mol Gen Genet MGG. 1999;261(6):977–84.
32. Khatibi PA, McMaster NJ, Musser R, Schmale DG. Survey of Mycotoxins in Corn Distillers' Dried Grains with Solubles from Seventy-Eight Ethanol Plants in Twelve States in the U.S. in 2011. Toxins. 2014;6(4):1155–68.

Chapter 7: Conclusions

Metabolic engineering of microbes for the production of high value biochemicals and biofuels needs new tools for (i) genetic engineering, (ii) synthetic or metabolic pathways "fine-tuning", (iii) analytical methods that can be applied in near real-time, and (iv) techniques for the isolation of microbial strains with desired features. In this study, λ-PCR, a novel method for plasmid DNA construction was developed (Chapter 2). λ-PCR is a time- and resource-saving open-source method for DNA manipulation. This technique allows different types of genetic manipulations including DNA insertion, deletion, substitution and point mutation for nearly any location in a DNA vector. High DNA assembly and cloning efficiencies have been achieved for a fraction of cost and time associated with conventional methods and commercial kits.

In the next study, novel synthetic small RNAs (sRNAs) with antisense binding regions were developed. The technology was of interest as it offers protein synthesis modulations, applicable to synthetic biology and metabolic engineering. A new synthetic scaffold was designed to support 20 bp antisense sequences that bind specific mRNAs regions (Chapter 3). This sRNA was part of an inducible synthetic-circuit expressed in cell-free-system to assess the different sRNA fragments of a fluorescent reporter protein. Downregulation of the fluorescence level was observed either by changing the target region or incubation time. Interestingly, we were able to characterize a synthetic sRNA that can increase target protein expression. This mechanism can be used in cell-free protein synthesis reactions to improve yields.

Analysis of phenotypic response of the cyanobacterium *Synechocystis* was performed using Raman spectroscopy (Chapters 4 and 5). Two methods were explored: (i) broad phenotype

changes to environmental conditions using chemometric methods and (ii) the correlation of individual Raman bands with their corresponding biomolecules with validation by accepted analytical methods. Rametrix™ Toolboxes enabled the characterization of changes in metabolic phenotypes as a response to changes in growth conditions. Clear distinctions were observed between cells growing in different environments. These results were validated through Rametrix™ with high accuracy, sensitivity, and selectivity. In addition, ANOVA and pairwise comparison confirmed our findings and determined which clusterings are statistically relevant. Finally, Raman data were correlated with standardized analytical methods (UPLC, GC-FID and spectrometric analysis), which showed strong correlation.

Finally, detoxification of the mycotoxin deoxynivalenol (DON) was investigated. For this, metagenomic DNA library fragments from microorganisms were enriched in a DON sensitive yeast strain (*Saccharomyces cerevisiae*) for the isolation of an enzyme(s) capable of DON degradation or modification. Although the enrichment for an enzymatic activity was not fully achieved, our results showed optimistic future possibilities in this regard. Low level of DON were observed after rounds of sub-culturing even for the sensitive strain, suggesting a possible acquisition of resistance or degradation mechanism(s). The path forward for this research is (i) increase of enrichment cycles, (ii) characterization and purification of detoxification enzymes. Cloning of these genes is easily achieved through λ-PCR and overexpression is accomplished via our asRNA system expressed *in vitro*. Detoxification enzymes can be cloned and expressed in crops to confer resistance to DON contamination.

Supplementary Appendices

Supplementary Appendix A: λ-PCR for precise DNA assembly and modification

Table A. 1. Plasmids used and constructed in this study.

Plasmid	Description	Origin
pUC19	ColE1 ori, Amp^r, P_{lac} promoter	New England Biolabs
pET-GFP	pBR322 ori, Amp^r, T7 promoter	Contributed by collaborator
pAmCyan	pUC ori, Amp^r, P_{lac} promoter, *amcyan* bp 289 - 978	Takara Bio USA
pACYC177	p15A ori, Amp^r and Kan^r	Life Technologies
pBAD-B	pBR322 ori, Amp^r, P_{BAD} promoter, His-Tag coding sequence 331-348	ThermoFisher Scientific
pET-28a	pBR322 ori, Kan^r, T7 promoter, His-Tag coding sequence 270-287	Invitrogen
pET-HBcAg	pET-28a with Hepatitis B Core Antigen (HBcAg) Virus Like Particle (VLP) gene in the MCS corresponding to the NheI and HindIII restriction enzyme cut sites	This study
pUC19-lacZα-GFP	pUC19 backbone, *gfp* bp 2610-3329	This study
pUC19-lacZα-GFP-AmCyan	pUC19 backbone, *gfp-cfp* bp 2683-4109	This study
pUC19-GFP	pUC19 backbone, *gfp* bp 45-764	This study

pUC19-psba2	pUC19 backbone, constructed by substitution of l*acZα* gene by *psba2* gene	This study
pUC19-psba2-kan	pUC19 backbone, constructed by insertion of kanamycin resistance gene in the middle of *psba2* gene	This study
pBAD-B-MCLA	pBAD-B harboring *mcla* gene downstream of PBAD promoter	This study
pBAD-MCRT	pBAD-B harboring *mcrt* gene downstream of PBAD promoter	This study
pET-HBcAgEps	pET-28a with HBcAg inserted into the MCS of the plasmid, with the addition of the gene for foreign epitopes of another virus inserted into the MIR	This study

Table A. 2. Primers used in λ-PCR case studies.

Primer Sequence (5' → 3')	Primer Name	Description
TAATAATGGTTTCTTAGACG_AGGAGGAACAGCT_ATGGTGAGCAAGGGCGA	Om_GFP_F	Case Study 1; Insertion of *gfp* into pUC19; Forward primer
ACATTTCCCCGAAAAGT_TTACTTGTACAGCTCGTCCATGC	Om_GFP_R	Case Study 1: Insertion of *gfp* into pUC19; Reverse primer
AACTTGTTTGAAAGAGCCAT_AGCTGTTCCTCCTTTTT_TTACTTGTACAGCTCGTCCATGC	Om_GFP_CFP_R	Case Study 1; Insertion of *gfp* and am*cyan* into pUC19; Reverse primer 1
ACATTTCCCCGAAAAGT_TCAGAAAGGGACAACAGAG	Om_CFP_R	Case Study 1; Insertion of *gfp* and am*cyan* into pUC19; Reverse primer 2
TTCACACAGGAAACAGCT_ATGGTGAGCAAGGGCGA	Om_Sub_GFP_F	Case Study 1; Substitution of *lacZα* with *gfp* in pUC19; Forward primer
AACCTCTGACACATGCAG_TTACTTGTACAGCTCGTCCATGC	Om_Sub_GFP_R	Case Study 1; Substitution of *lacZα* with *gfp* in pUC19; Reverse primer
TTGTAAAACGACGGCCAGTG_ATGACAACGACTCTCCAACAG	PSBA2_Om_Fw	Case Study 2; Substitution of *lacZα* with *psbas2* in pUC19; Forward primer
GCTATGACCATGATTACGCCA_TTAACCGTTGACAGCAGGAGCGGTC	PSBA2_Om_Rev	Case Study 2; Substitution of *lacZα* with *psba2* in pUC19; Reverse primer
CTACCGCTTAGGTATGCGTC_TTCAACAAAGCCACGTTG	KanR_Om_Fw	Case Study 2; Insertion of *kan*^{r} in the middle of *psbas2* in pUC19; Forward primer
AGTTCTGGGATTCAACTTCGG_CCAAATTTGATTCCTTTCAGCTTTGCTTCCT	KanR_Om_Rev	Case Study 2; Insertion of *kan*^{r} in the middle of *psbas2* in pUC19; Reverse primer
AACCCCTATTTGTTTATTTTTCTAAATAC	p19-*λ*-R	Case Study 2; Reverse primer

GTATGTTGTGTGGAATTGT	Check_F	Case Study 1 and 2; Forward Check primer
GTCACAGCTTGTCTGTAAG	Check_R	Case Study 1 and 2; Reverse Check primer
CGCAAGCTAGCTCACAAC	Lambda1_mcla_F	Case Study 3; Insertion of *mcla* into pBAD24; Forward primer
CAGGCGTATGGTCTGTGA_GCAG CTGGTACCATATGGG	Lambda1_mcla_R	Case Study 3: Insertion of *mcla* into pBAD24; Reverse primer
GATCCGAGCTCGAGATCT	Lambda2_mcla_R	Case Study 3: Insertion of *mcla* into pBAD24; Reverse primer
CGCCGTACCTTAAAA	Lambda1_mcrt_F	Case Study 3; Insertion of *mcrt* into pBAD24; Forward primer
AAAAAGGTTACATTGAAAAATAA _GCAGCTGGTACCATATGGGAATT CGAAGCTT	Lambda1_mcrt_R	Case Study 3: Insertion of *mcrt* into pBAD24; Reverse primer
CGATGACGATAAGGATCCGAGCT CGAGATCT	Lambda2_mcrt_R	Case Study 3: Insertion of *mcrt* into pBAD24; Reverse primer
ACTTCTCGTGACCTGGTTG	HBcAgMIR_F	Case Study 3: Insertion of *HBcAg* into pET28a; Forward primer
GTCTTCCAGGTTACCACC	HBcAgMIR_R	Case Study 3: Insertion of *HBcAg* into pET28a; Reverse primer

*Different chimeric sections of the primer are separated by underscore symbol(s).

*The red section of the primer anneals to the pUC19 plasmid.

*The green section of the primer anneals to the pBAD24 plasmid.

*The yellow section of the primer anneals to the *psb2* gene.

CASE STUDY 1: Fluorescent reporter gene insertions into pUC19

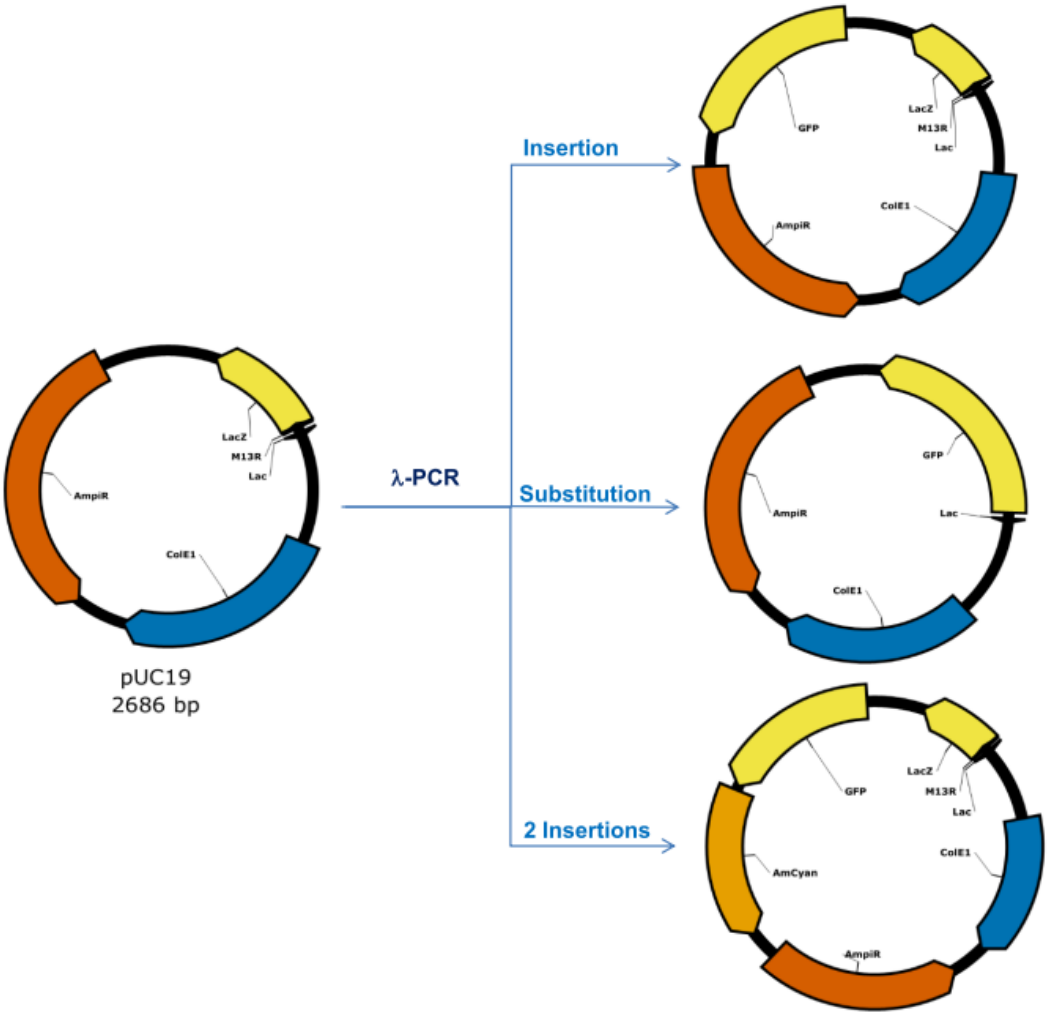


Figure A. 1. Plasmid constructions using the *gfp* and *amcyan* genes to validate the λ-PCR method. pUC19 was used as the backbone in all plasmid constructions.

(top – single insertion experiment) The *gfp* gene inserted downstream of the *lacZα* gene.
(middle – substitution experiment) Replacing the *lacZα* gene by the *gfp* gene.
(bottom – double insertion experiment) The *gfp* and *amcyan* genes were inserted simultaneously downstream of the *lacZα* gene.

Gene map:
colE1: Blue
ampR: Red-orange
lacZα: Yellow (short)
gfp: Yellow (long)
amcyan: Light orange

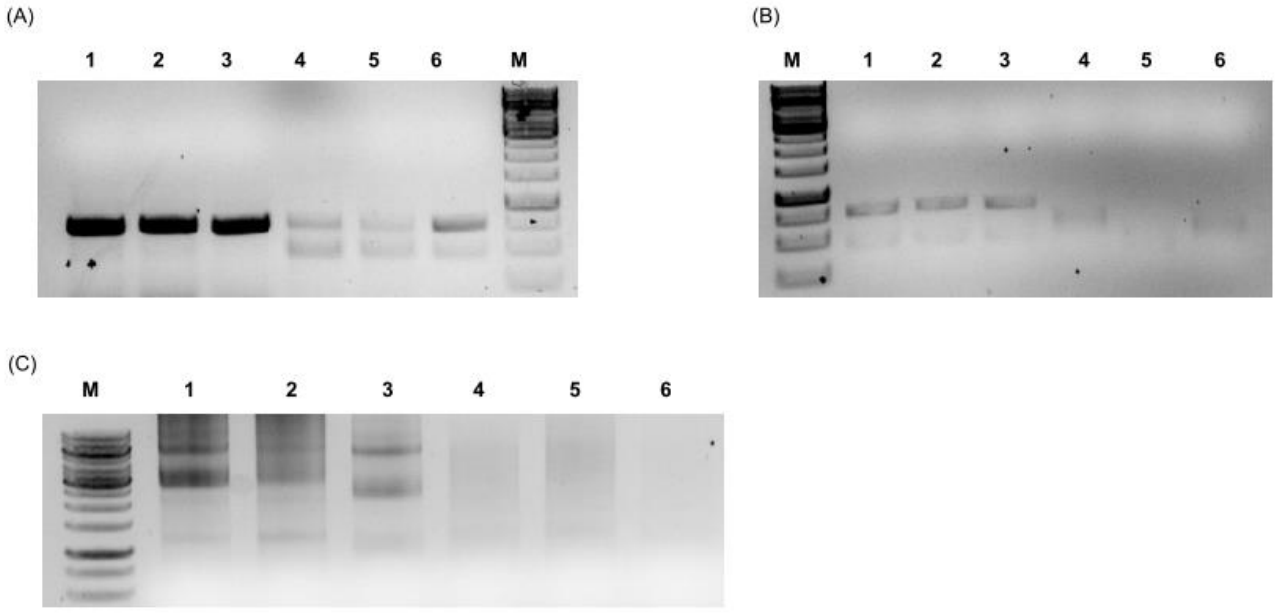


Figure A. 2. Validation of fluorescent reporter genes insertion into pUC19.

(A) PCR amplification of the *gfp* gene (720 bp) (lanes 1-3), λ-exonuclease digestion of the *gfp* PCR products (lanes 4-6), 1 kb DNA marker (lane M);
(B) Asymmetric PCR of the *gfp* gene (lanes 1-3), S1 nuclease digestion of asymmetric PCR products (lanes 4-6). 1 kb DNA marker (lane M);
(C) PCR amplification of pUC19 with the ssDNA megaprimer generated through complete λ-exonuclease digestion (lanes 1-3), and generated with asymmetric PCR (lanes 4-6). 1 kb DNA marker (lane M).

In all gels, lanes 1 and 4 were used in single *gfp* insertion experiment; lanes 2 and 5 were used in the double *gfp-amcyan* insertion experiment, and lanes 3 and 6 were used in the *gfp* for *lacZα* substitution experiment.

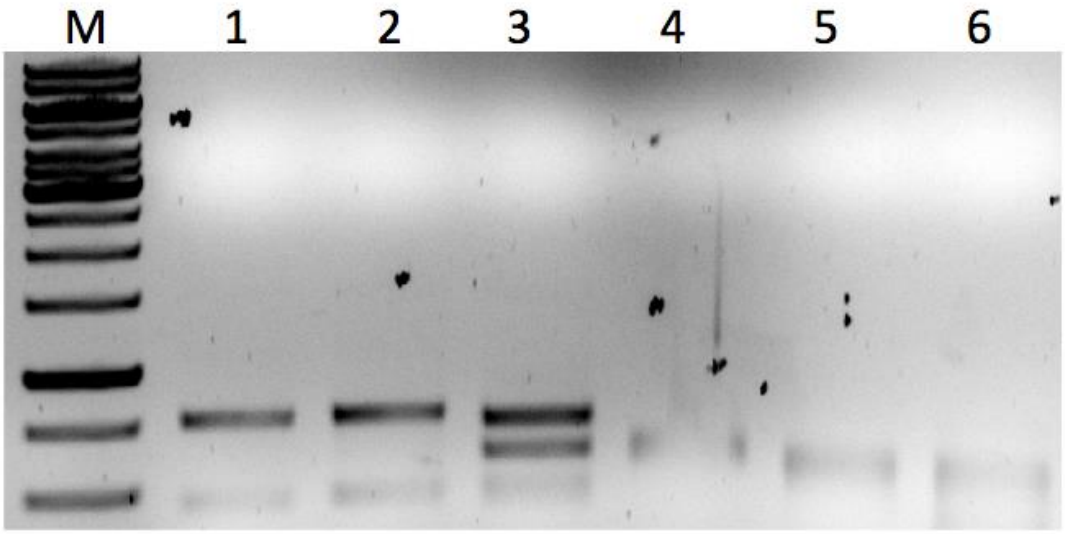


Figure A. 3. Additional asymmetric PCR results.

Lane M: 1 Kb DNA ladder,
Lanes 1-3: before S1 nuclease digestion;
Lanes 4-6: after S1 nuclease digestion;
Lanes 1 and 4: *gfp* insertion;
Lanes 2 and 5: *gfp-cfp* simultaneous insertion;
Lanes 3 and 6: *gfp* substitution.

CASE STUDY 2: Plasmid construction for Synechocystis PCC 6803 gene knockout

Step 1. Amplification and insertion of psba2

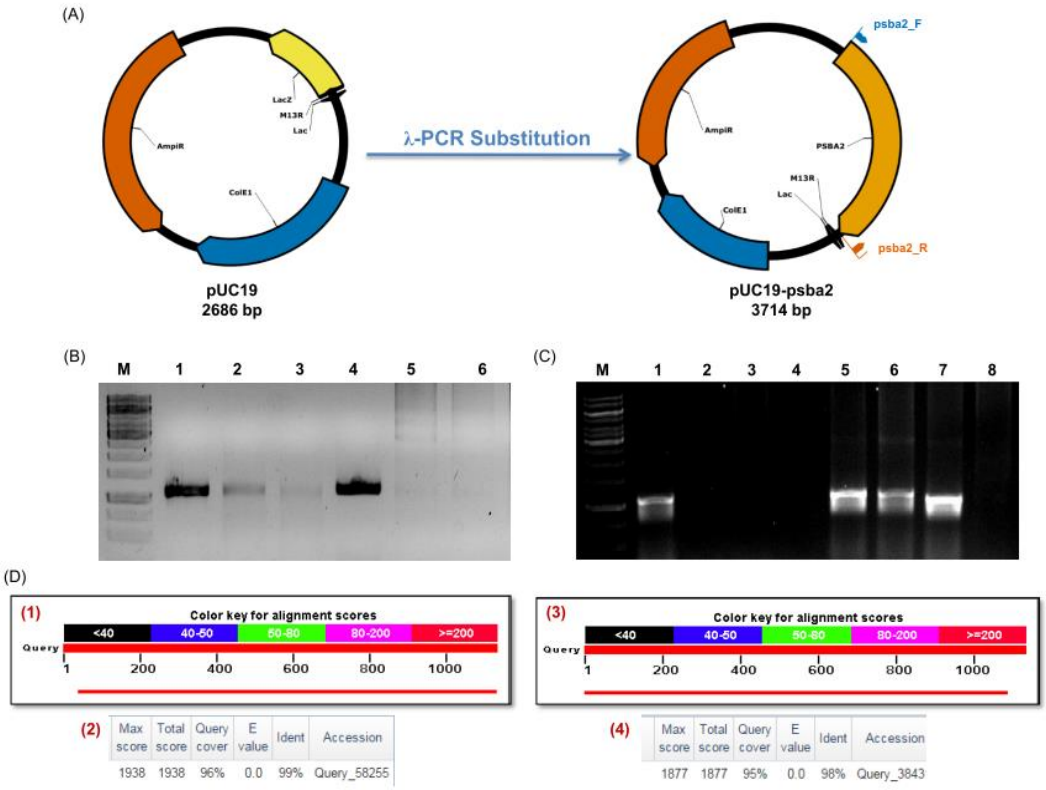


Figure A. 4. Plasmid construction to insert the *psba2* gene (step 1).

(A) Schematic of the pUC19-psba2 construction by λ-PCR.

(B) λ-Exonuclease partial digestion of *psba2* gene (1083 bp) (lanes 1-3). 1 kb DNA marker (lane M), the PCR-amplified and undigested *psba2* gene (lanes 2 and 4), after 15 min of λ-exonuclease digestion (lane 2), after 30 min of λ-exonuclease digestion (lane 3), PCR amplification of pUC19 with the dsDNA megaprimer after 15 min of λ-exonuclease digestion (lane 5), and PCR amplification of pUC19 with the dsDNA megaprimer after 30 min of λ-exonuclease digestion (lane 6).

(C) Colony PCR following pUC19-psba2 plasmid transformation. 1 kb DNA marker (lane M), pUC19-psba2 constructed with dsDNA megaprimer after 15 min of λ-exonuclease digestion (lanes 1-4). pUC19-psba2 constructed with dsDNA megaprimer after 30 min of λ-exonuclease digestion (lanes 5-7).

(D) Sequencing results to confirm insertion of *psba2* in the pUC19 plasmid (when using the dsDNA megaprimer after 30 min of λ-exonuclease digestion). (1) and (2) sequencing with the forward primer, (3) and (4) results of sequencing with reverse primer.

Gene map:
colE1: Blue
ampR: Red-orange
lacZα: Yellow (short)
psba2: Orange

Step 2. Insertion of the kanamycin resistance gene

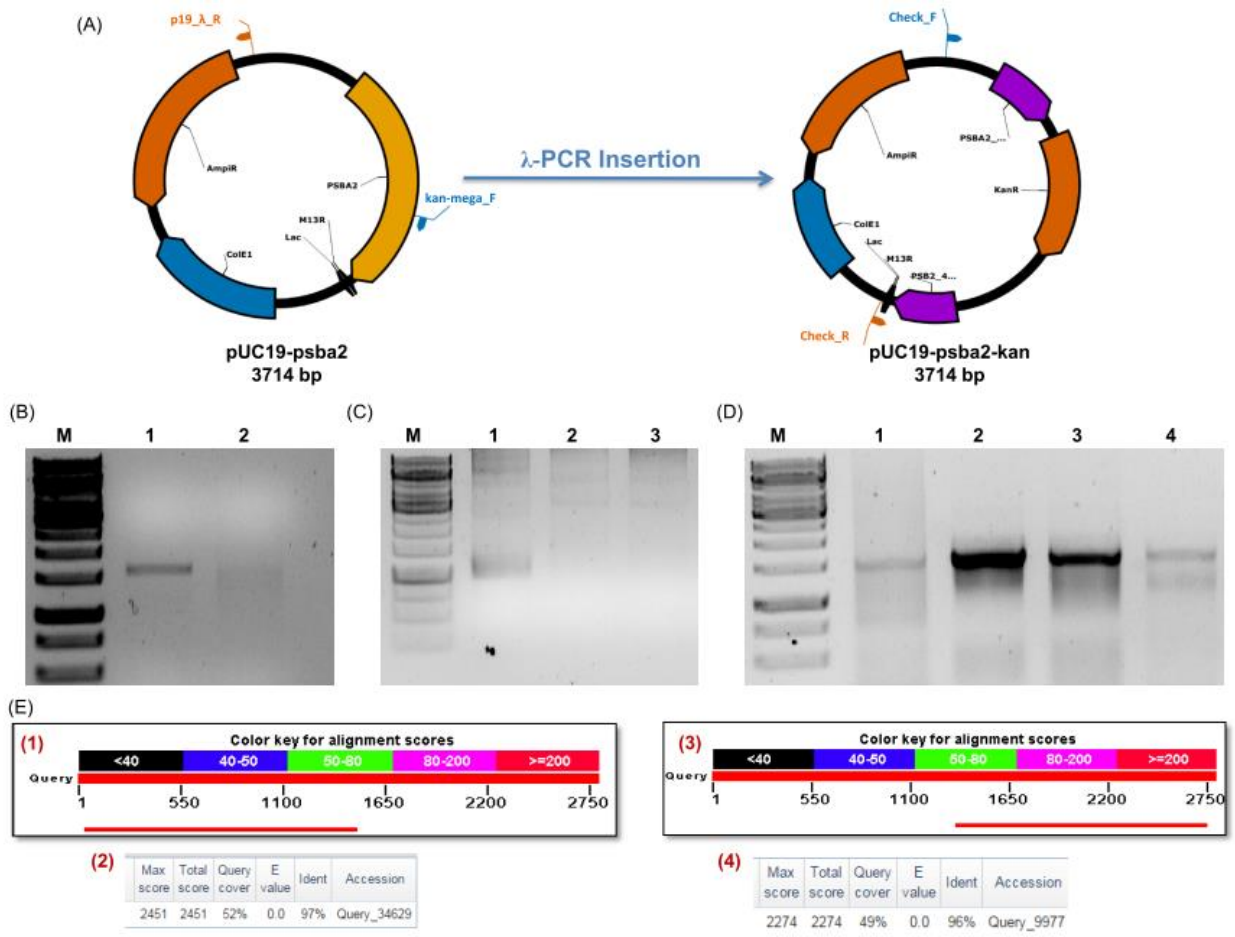


Figure A. 5. Plasmid construction to insert the kanamycin resistance gene into the *psba2* gene.

(A) Schematic of the pUC19-psba2-kanr construction by λ-PCR.
(B) PCR amplification of the kanamycin resistance gene (lane 1), λ-exonuclease digestion of the kanamycin resistance gene (lane 2).
(C) PCR amplification of the pUC19-psba2-kanr plasmid.
(D) Colony PCR of the pUC19-psba2-kanr transformation (lanes 1-4). 1 kb DNA marker (lane M).
(E) Sequencing results to confirm insertion of the kanamycin resistance gene into the pUC19-psba2 plasmid. (1) and (2) sequencing with the forward primer, (3) and (4) results of sequencing with reverse primer.

Gene map:
colE1: Blue
ampR: Red-orange
kanR: Red-orange
psba2 homology regions: Purple

CASE STUDY 3: Cloning heterologous genes for enzyme engineering

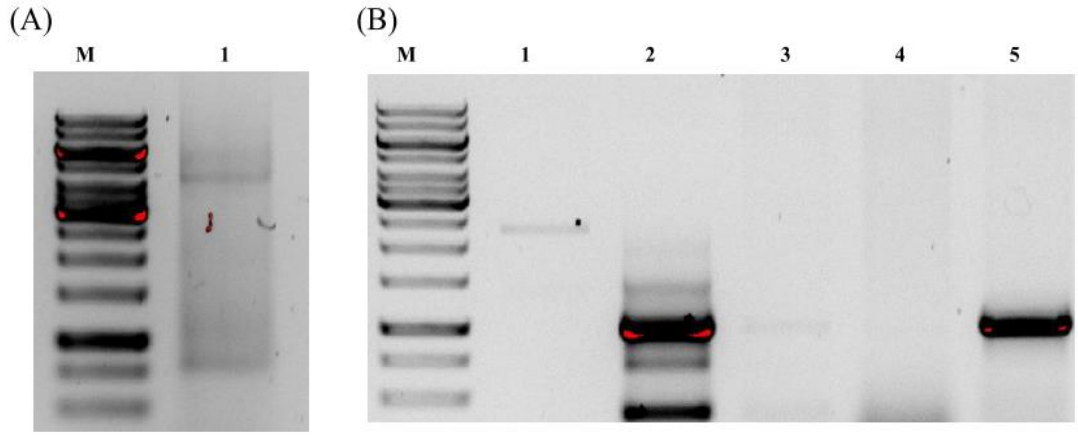


Figure A. 6. *Mcla* gene cloning in pBAD-A by λ-PCR v3.
(A) pBAD-mcla amplification with the *mcla* ssDNA megaprimer and the reverse primer (lane 1);
(B) Colony PCR of pBAD-mcla. Negative colonies (lanes 1, 3 and 4) and positive colonies (lanes 2 and 5). 1 kb gene marker (lane M).

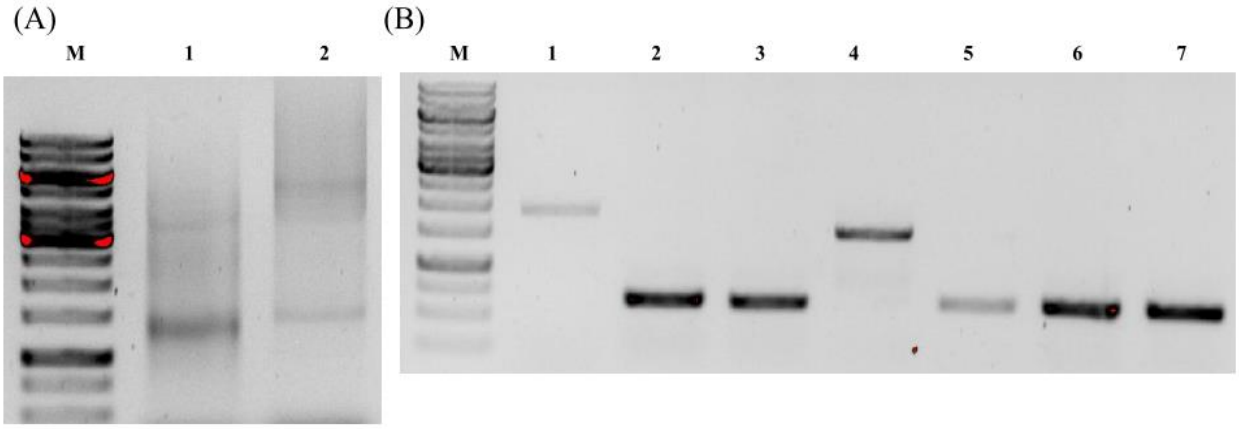


Figure A. 7. *Mcrt* gene cloning in pBAD-A by λ-PCR v3.

(A) pBAD-mcrt amplification with the *mcrt* ssDNA megaprimer and the reverse primer (lanes 1 and 2);
(B) Colony PCR of pBAD-mcrt. Negative colonies (lanes 2, 3, 5-7) and positive colonies (lanes 1 and 4). 1 kb gene marker (lane M).

CASE STUDY 4: Troubleshooting using thermodynamic calculations

DNA gel images of the three attempts at inserting an immunodominant region (MIR) into the Hepatitis B Core Antigen (*HBcAg*) gene, which was previously inserted into the pET-28a plasmid, are shown in Fig. A8. In each case, an additional MIR sequence was inserted into the existing MIR of the *HBcAg*, at approximately the mid-point of the gene. The first image (Fig. A8A) shows an attempt where the result was a "smear" of DNA without any dominant products, indicating a failure of single product amplification and the possible presence of a heterogeneous mixture of PCR products. It was hypothesized that $GG(X)_3$ regions (encoding glycine residues) of the ssDNA portion of the megaprimer were forming inhibitory secondary structures. These were validated by thermodynamic calculations, which are described in detail below. The second attempt (Fig. A8B), with the $GG(X)_3$ rich portion of the inserted MIR DNA fragment removed, led to successful single-product amplification. Additionally, the extension of the annealing portion of the ssDNA megaprimer past the hairpin loop indicated by thermodynamic calculations allowed for successful amplification (Fig. A8C) of the plasmid with the glycine residues and the $GG(X)_3$ portions of DNA to occur. The two successful plasmids were re-circularized (using ligase in this case) and transformed into *E. coli* BL21-DE3. The transformed plasmids were then sequenced, each showing successful cloning took place (Fig. A8D-E).

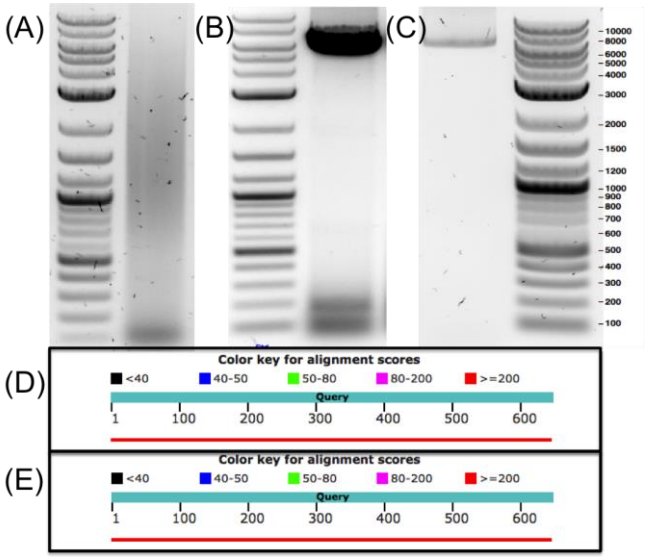


Figure A. 8. Attempts at inserting the MIR into the *HBcAg* gene in the pET-28a plasmid.
(A) pET-28a-MIR1 amplification using the native MIR, suspected of forming inhibitory secondary structures;
(B) pET-28a-MIR2 amplification using the MIR with $GG(X)_3$ sequences removed;
(C) pET-28a-MIR3 amplification using the native MIR with an extended annealing sequence that avoids secondary structure formation.
(D,E) Sequencing results confirming the correct identity of cloned gene products in (B) and (C).

It was hypothesized that the reason for the success and failure of these amplification attempts was related to inhibitory secondary structures in the annealing ssDNA portion of the megaprimer during PCR. This hypothesis was tested by simulating ssDNA folding using the NUPACK web application (http://www.nupack.org). NUPACK simulations of the ssDNA megaprimer at various temperatures are shown in Figs. A9A-F. The first two simulations (Fig. A9A,B) demonstrate a linear ssDNA strand with a hairpin secondary structure toward the middle of the megaprimer at 95°C and 80°C, respectively. The second two simulations (Fig. A9C,D) indicate that second and third hairpin structures form at 65°C and 64°C, respectively. The larger of these structures is located only 6 bp away from the 3' annealing end, as indicated by the arrows. Because this secondary structure was within the annealing portion of the ssDNA megaprimer to the plasmid, it was hypothesized that there was disruption of the original design, and correct annealing did not occur. Instead, this phenomenon resulted in somewhat random annealing, leading to the resulting "smear" in the gel image (Fig. A8A). The removal of the $GG(X)_3$ regions from the megaprimer eliminated the second and third hairpin structures at the annealing temperature (Fig. AE), which allowed for correct annealing and the plasmid amplification to take place (Fig. A8B). Finally, simple extension of the annealing portion, shown in Fig. A9F, moved the hairpin structure away from the terminal 3' end, allowing proper annealing to occur with the plasmid. There was, therefore, a new annealing sequence, which remained as unbound ssDNA at the 64°C annealing temperature in the cloning reaction.

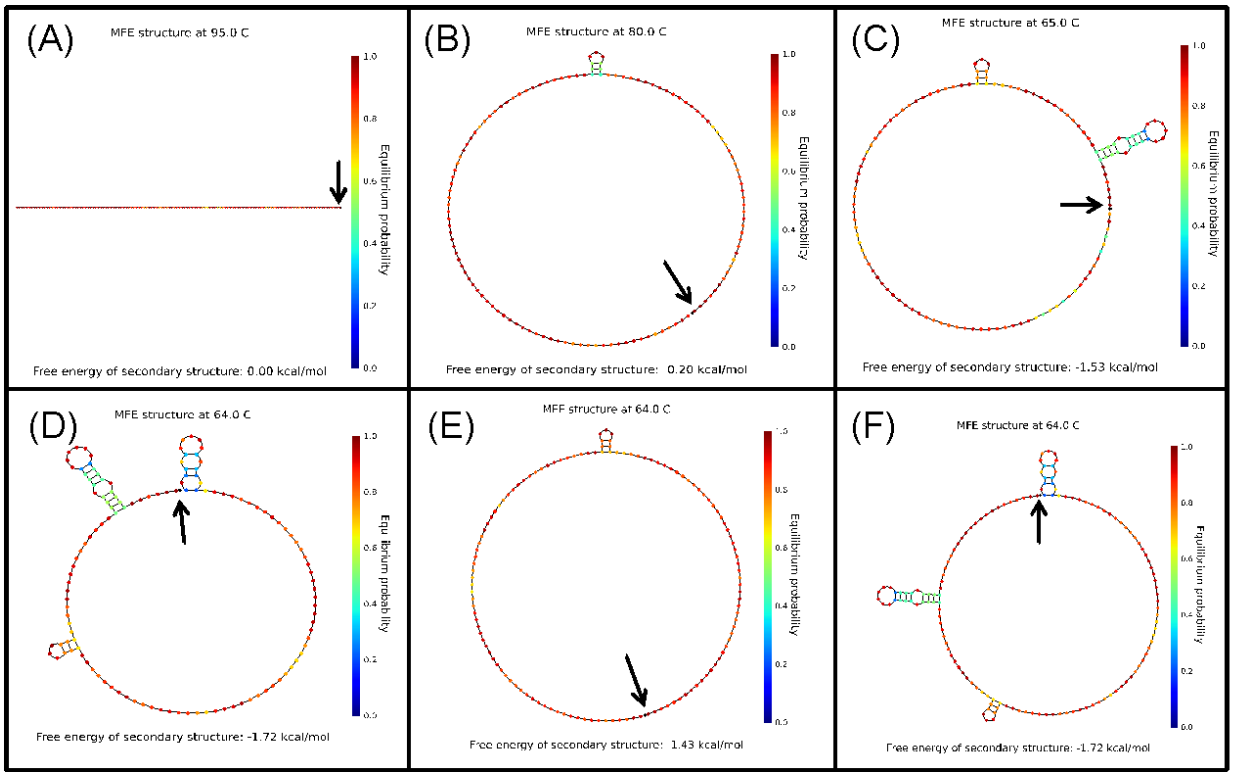


Figure A. 9. NUPACK folding simulation screenshots of the ssDNA megaprimer containing the MIR to be cloned at different temperatures. The 3' terminal end is identified by an arrow in each screenshot.
(A) The original megaprimer at 95°C;
(B) the original megaprimer at 80°C;
(C) the original megaprimer at 65°C;
(D) the original megaprimer at 64°C;
(E) the megaprimer with $GG(X)_3$ regions removed at 64°C; and
(F) the extended megaprimer at 64°C. The sequences upstream of the 3' terminal end anneal to the plasmid. It was hypothesized that secondary structures interfered with this annealing process.

The ssDNA megaprimer containing the foreign epitope gene with the $GG(X)_3$ terminal ends encoding for glycine residues is shown below in Fig. A10. The grey highlighted portion represents the original binding region to the plasmid. The bold region 3' downstream of the highlighted region represents the extended annealing portion used to successfully clone in the epitopes with the $GG(X)_3$ codons. Interior bases not involved in secondary structure formation have been changed to "X" to protect sequence identity for other ongoing research.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
ATCTCTCGTGACCTGGTTGT**TTCTTACGTTAACACC**

Figure A. 10. The final design for extending the ssDNA megaprimer to clone the MIR.

Supplementary Appendix B: A novel synthetic sRNA promoting protein overexpression in cell-free systems

Table B. 1. Strains and plasmids

Strain/plasmid	**Relevant Characteristic**	**Reference**
Escherichia coli 10-β	*Δ*(ara-leu) 7697 araD139 fhuA *Δ*lacX74 galK16 galE15 e14- *ϕ*80*d*lacZ*Δ*M15 recA1 relA1 endA1 nupG rpsL *(StrR)* rph spoT1 *Δ*(mrr-hsdRMS-mcrBC)	New England Biolabs
pACYC177	Amp^r; Kan^r; p15 origin	Life Technologies
pBAD24	Amp^r; ColE1 origin; *araC*; *rrnB*; P_{BAD}	(1)
pKLiO20	Amp^r; ColE1 origin; *tetR*	(2)
pKLiO30	Amp^r; ColE1 origin; $P_{LtetO-1}$	(2)
pUC19	Amp^r; ColE1 origin; P_{lac}; *lacZ*	New England Biolabs
pAmCyan	Amp^r; pUC origin; *AmCyan*	Takara Bio USA, Mountain View, CA
pART15	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; P_{BAD}; $P_{LtetO-1}$	This study
pART15-C1	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$	This study
pART15-C1-LacZ	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; *lacZ*; P_{BAD}; $P_{LtetO-1}$	This study
pART15-C2	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$	This study
pART15-asCTL	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; Control	This study
pART15-asRBS	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-RBS	This study
pART15-asB1	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 1-20	This study
pART15-asB30	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 30-50	This study
pART15-asB57	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 57-77	This study

pART15-asB180	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 180-200	This study
pART15-asB407	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 407-427	This study
pART15-asB430	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 430-450	This study
pART15-asB539	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 539-559	This study
pART15-asB606	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 606-626	This study
pART15-asB615	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 615-635	This study
pART15-asB693	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 693-713	This study
pART15-asB701	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; asRNA anti-bp 701-721	This study
pART15-MicC	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; MicC	This study
pART15-MicC-asRBS	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; MicC-asRNA anti-RBS	This study
pART15-MicC-aB1	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; MicC-asRNA anti-bp 1-20	This study
pART15-MicC-aB606	Amp^r; p15 origin; *araC*; *rrnB*; *tetR*; *AmCyan*; P_{BAD}; $P_{LtetO-1}$; MicC-asRNA anti-bp 606-626	This study
pET20b-Gsdi	Amp^r; pBR322 origin; *diaphorase*; T7; T7 *terminator*	(3)
pBAD-LIC (8A)	Amp^r; ColE1 origin; *araC*; *rrnB*; P_{BAD}; asRNA anti-bp 701-721	Addgene
pBAD-asDI420	Amp^r; ColE1 origin; *araC*; *rrnB*; P_{BAD}; asRNA anti-bp 420-440	This study
pBAD-asDI490	Amp^r; ColE1 origin; *araC*; *rrnB*; P_{BAD}; asRNA anti-bp 490-510	This study
pBAD-asDI510	Amp^r; ColE1 origin; *araC*; *rrnB*; P_{BAD}; asRNA anti-bp 510-530	This study

pBAD-asDI528	Amp^r; ColE1 origin; *araC*; *rrnB*; P_{BAD}; asRNA anti-bp 528-548	This study

Table B. 2. Primer sequences

Primer Name	**Primer Sequence (5'→3')**	**Description**
pART15-177-F	GAGTCAGCAACACCTTCTTC	Amplification of p15 origin of replication and *amp*^{r} gene from pACYC177
pART15-177-R	AACCCCTATTTGTTTATTTTTCTAAATAC	Amplification of p15 origin of replication and *amp*^{r} gene from pACYC177
pART15-24-F	aaataaacaaataggggttGCTCATGAGCCCGAAGTG	Amplification of *araC* gene with P_{BAD} and T1T2 segment from pBAD24
pART15-24-R	aggcgcccAAACAAAAGAGTTTGTAGAAACGC	Amplification of *araC* gene with P_{BAD} and T1T2 segment from pBAD24
pART15-20-F	tcttttgtttGGGCGCCTGTCACTTTGC	Amplification of *tetR* gene from pKLiO20
pART15-20-R	aaggtgttgctgactcCCAGGCATCAAATAAAACGAAAGG	Amplification of *tetR* gene from pKLiO20
pART15-30-F	atgagaattatttaaccttaCTCGAGATCTGATCTTCAGCATCTTTTACTTTCAC	Amplification of $P_{Lteto-1}$ promoter from pKLiO30
pART15-30-R	cgttgctttttctcatttaGGATCCCAGCTGTCCCTA	Amplification of $P_{Lteto-1}$ promoter from pKLiO30
pART15-C1-F	cgtttttttgggctagcaggaggaaagctATGGCTCTTTCAAACAAGTTTATC	Insertion of *amcyan gene* downstream of P_{BAD} promoter
pART15-C1-R	gatccccgggtaccatggtgTCAGAAAGGGACAACAGAG	Insertion of *amcyan gene* downstream of P_{BAD} promoter
pART15-C2-F	tgagaattatttaaccttacTCAGAAAGGGACAACAGAG	Insertion of *amcyan gene* downstream of $P_{Lteto-1}$ promoter
pART15-C2-R	gagcacatcagcaggagatcATGGCTCTTTCAAACAAGTTTATC	Insertion of *amcyan gene* downstream of $P_{Lteto-1}$ promoter

Pc-RBS-F	cagcatagcatagacaaccaAGGAGG AACAGCTATGTCTCGTTTA GATAAAAGTAAAGTG	Insertion of RBS region upstream of *TetR* gene
Pc-RBS-R	atctacaactggcatgcCGGCAGAA AAGTCCACATTGATCAGCT GTTCAGCGACTAAG	Insertion of RBS region upstream of *TetR* gene
Inverse-RBS2-F	caaaaaaacgggtatggagaaTCTCT ATCACTGATAGGGATGTC	Insertion of RBS region upstream of *amcyan* gene in pART15-C2
AmCyan2-RBS-R	ggctagcaggaggaaagctATGGCT CTTTCAAACAAGTTTATC	Insertion of RBS region upstream of *amcyan* gene in pART15-C2
pART15-LacZ-F	gagaattatttaaccttacCTATGCGG CATCAGAGCAG	Insertion of *lacz gene* downstream of $P_{Lteto-1}$ promoter
pART15-LacZ-R	gagcacatcagcaggagatcaggaggaa agctATGACCATGATTACGC CAAG	Insertion of *lacz gene* downstream of $P_{Lteto-1}$ promoter
p15-C1-V2-F	TCCTGCTGATGTGCTCAG	Modification of $P_{Lteto-1}$ promoter downstream region
p15-LacZ-V2-R	CGCACTGACCGATCAGGA GGAAAGCTATGACC	Modification of $P_{Lteto-1}$ promoter downstream region
ST-V2-RBS-F	GGGCTAGCAGGAGGAAAG CTGCTGCTTTCGCAGCGGT CAGTGCGTCCTG	Insertion of antisense RNA anti-RBS downstream of $P_{Lteto-1}$ promoter
ST-V2-B1-F	ATGGCTCTTTCAAACAAGT TGCTGCTTTCGCAGCGGTC AGTGCGTCCTG	Insertion of antisense RNA anti-B1 downstream of $P_{Lteto-1}$ promoter
ST-V2-B30-F	CTATGGCTCTTTCAAACAA GGCTGCTTTCGCAGCGGTC AGTGCGTCCTG	Insertion of antisense RNA anti-B30 downstream of $P_{Lteto-1}$ promoter
ST-V2-B57-F	AGATGACATGAAAATGAC CTGCTGCTTTCGCAGCGGT CAGTGCGTCCTG	Insertion of antisense RNA anti-B57 downstream of $P_{Lteto-1}$ promoter
ST-V2-B180-F	TGGCCAACGGTGGGCCCCT TGCTGCTTTCGCAGCGGTC AGTGCGTCCTG	Insertion of antisense RNA anti-B180 downstream of $P_{Lteto-1}$ promoter
ST-V2-B407-F	TTTCATGGAGTGAACTTTC CGCTGCTTTCGCAGCGGTC AGTGCGTCCTG	Insertion of antisense RNA anti-B407 downstream of $P_{Lteto-1}$ promoter

ST-V2-B430-F	TGATGGACCTGTGATGGCGAGCTGCTTTCGCAGCGGTCAGTGCGTCCTG	Insertion of antisense RNA anti-B430 downstream of $P_{Lteto-1}$ promoter
ST-V2-B539-F	AAGGAGGTGGCAATTACAGAGCTGCTTTCGCAGCGGTCAGTGCGTCCTG	Insertion of antisense RNA anti-B539 downstream of $P_{Lteto-1}$ promoter
ST-V2-B606-F	CACCAAACCATGCGGTGGAAGCTGCTTTCGCAGCGGTCAGTGCGTCCTG	Insertion of antisense RNA anti-B606 downstream of $P_{Lteto-1}$ promoter
ST-V2-B615-F	ATGCGGTGGAACATCGCATTGCTGCTTTCGCAGCGGTCAGTGCGTCCTG	Insertion of antisense RNA anti-B615 downstream of $P_{Lteto-1}$ promoter
ST-V2-B693-F	CACATATAACCTCTGTTGTCGCTGCTTTCGCAGCGGTCAGTGCGTCCTG	Insertion of antisense RNA anti-B693 downstream of $P_{Lteto-1}$ promoter
ST-V2-B701-F	CCTCTGTTGTCCCTTTCTGAGCTGCTTTCGCAGCGGTCAGTGCGTCCTG	Insertion of antisense RNA anti-B701 downstream of $P_{Lteto-1}$ promoter
C-V2-control-F	GCTGCTTTCGCAGCGGTCAGTGCGTCCTG	Insertion of antisense RNA control downstream of $P_{Lteto-1}$ promoter
asRNA-UniR-2	CCCGCGAAAGCGGGAAGGTTAAATAATTCTCATATATCAAGCAAAG	Reverse primer for the insertion of all antisense RNAs
pART15-MicC-F	TGTTGGAAAATCAGTGGCAATGCAATGGCCCAACAGAAACCTATGCGGCATCAGAGCAG	Insertion of MicC scaffold downstream of $P_{Lteto-1}$ promoter
pART15-MicC-R	TATAAAAAGACAAGCCCGAACAGTCGTCCGGGCTTTTTTTAAGGTTAAATAATTCTCATATATCAAGCAAAGTGA	Insertion of MicC scaffold downstream of $P_{Lteto-1}$ promoter
MicC-asRBS-F	GGGCTAGCAGGAGGAAAGCTCCTGCTGATGTGCTCAGTATC	Insertion of antisense RNA control MicC-asRBS downstream of $P_{Lteto-1}$ promoter

MicC-asB1-F	ATGGCTCTTTCAAACAAGTTCCTGCTGATGTGCTCAGTATC	Insertion of antisense RNA control MicC-asB1 downstream of $P_{Lteto-1}$ promoter
MicC-asB606-F	CACCAAACCATGCGGTGGAACCTGCTGATGTGCTCAGTATC	Insertion of antisense RNA control MicC-asB606 downstream of $P_{Lteto-1}$ promoter
MicC-Uni-R	TTCTGTTGGGCCATTGCATTG	Reverse primer for the insertion of MicC antisense RNAs
DI410-F	GTGGAGCGCTTTTTTATCGGCCCGCGAAAGCGGGCCCGTTTTTTTGGGCTAG	Insertion of antisense asDI410 downstream of P_{BAD} promoter in pBAD-LIC plasmid
DI490-F	ATAATGACGCTTAAGTACCGCCCGCGAAAGCGGGCCCGTTTTTTTGGGCTAG	Insertion of antisense asDI490 downstream of P_{BAD} promoter in pBAD-LIC plasmid
DI510-F	ACGGAACACCGAAAAATTGCCCCGCGAAAGCGGGCCCGTTTTTTTGGGCTAG	Insertion of antisense asDI510 downstream of P_{BAD} promoter in pBAD-LIC plasmid
DI528-F	CAAACAGTCCTTCAAACGACCCCGCGAAAGCGGGCCCGTTTTTTTGGGCTAG	Insertion of antisense asDI528 downstream of P_{BAD} promoter in pBAD-LIC plasmid
DI-Uni-R	GCTGCTTTCGCAGCTATGGAGAAACAGTAGAGAG	Reverse primer for the insertion of antisense RNAs in PBAD-LIC

*Lowercase are overlaps for Gibson assembly® cloning.
**RBS sequence is in red.
*** The underlined sequence in pART15-30-F is cut sites for XhoI and BglII.
****Antisense RNA binding region is blue.
*****Stabilizer region is in green.
******Terminator is orange.
*******MicC scaffold is in purple.

Table B. 3. Antisense sequences added to the sRNA

Antisense Name	**Sequence (5' → 3')**
asRBS	GCTGCGAAAGCAGCAGCTTTCCTCCTGCTAGCCCCCCGCGAAAGCGGG

asB1	GCTGCGAAAGCAGCAACTTGTTTGAAAGAGCCATCCCG CGAAAGCGGG
asB30	GCTGCGAAAGCAGCCTTGTTTGAAAGAGCCATAGCCCG CGAAAGCGGG
asB57	GCTGCGAAAGCAGCAGGTCATTTTCATGTCATCTCCCCG CGAAAGCGGG
asB180	GCTGCGAAAGCAGCAAGGGGCCCACCGTTGGCCACCCG CGAAAGCGGG
asB407	GCTGCGAAAGCAGCGGAAAGTTCACTCCATGAAACCCC GCGAAAGCGGG
asB430	GCTGCGAAAGCAGCTCGCCATCACAGGTCCATCACCCG CGAAAGCGGG
asB539	GCTGCGAAAGCAGCTCTGTAATTGCCACCTCCTTGCCCC GCGAAAGCGGG
asB606	GCTGCGAAAGCAGCTTCCACCGCATGGTTTGGTGCCCG CGAAAGCGGG
asB615	GCTGCGAAAGCAGCAATGCGATGTTCCACCGCATCCCC GCGAAAGCGGG
asB693	GCTGCGAAAGCAGCGACAACAGAGGTTATATGTGCCCG CGAAAGCGGG
asB701	GCTGCGAAAGCAGCTCAGAAAGGGACAACAGAGGCCC CGCGAAAGCGGG
asCTRL	GCTGCGAAAGCAGCCCCGCGAAAGCGGG
MicC-asRBS	AGCUUUCCUCCUGCUAGCCCUUCUGUUGGGCCAUUGC AUUGCCACUGAUUUUCCAACAUAUAAAAAGACAAGCC CGAACAGUCGUCCGGGCUUUUUUU
MicC-asB1	AACUUGUUUGAAAGAGCCAUUUCUGUUGGGCCAUUGC AUUGCCACUGAUUUUCCAACAUAUAAAAAGACAAGCC CGAACAGUCGUCCGGGCUUUUUUU
MicC-asB606	TTCCACCGCATGGTTTGGTGUUCUGUUGGGCCAUUGCA UUGCCACUGAUUUUCCAACAUAUAAAAAGACAAGCCC GAACAGUCGUCCGGGCUUUUUUU
asDI410	GCTGCGAAAGCAGCGTGGAGCGCTTTTTTATCGGCCCG CGAAAGCGGG
asDI490	GCTGCGAAAGCAGCATAATGACGCTTAAGTACCGCCCG CGAAAGCGGG
asDI510	GCTGCGAAAGCAGCACGGAACACCGAAAAATTGCCCCG CGAAAGCGGG
asDI528	GCTGCGAAAGCAGCCAAACAGTCCTTCAAACGACCCCG CGAAAGCGGG

*Stabilizer region is highlighted in red.
** 20 nt antisense region is in black.
***Terminator region is highlighted in blue.
****MicC scaffold is highlighted in green.

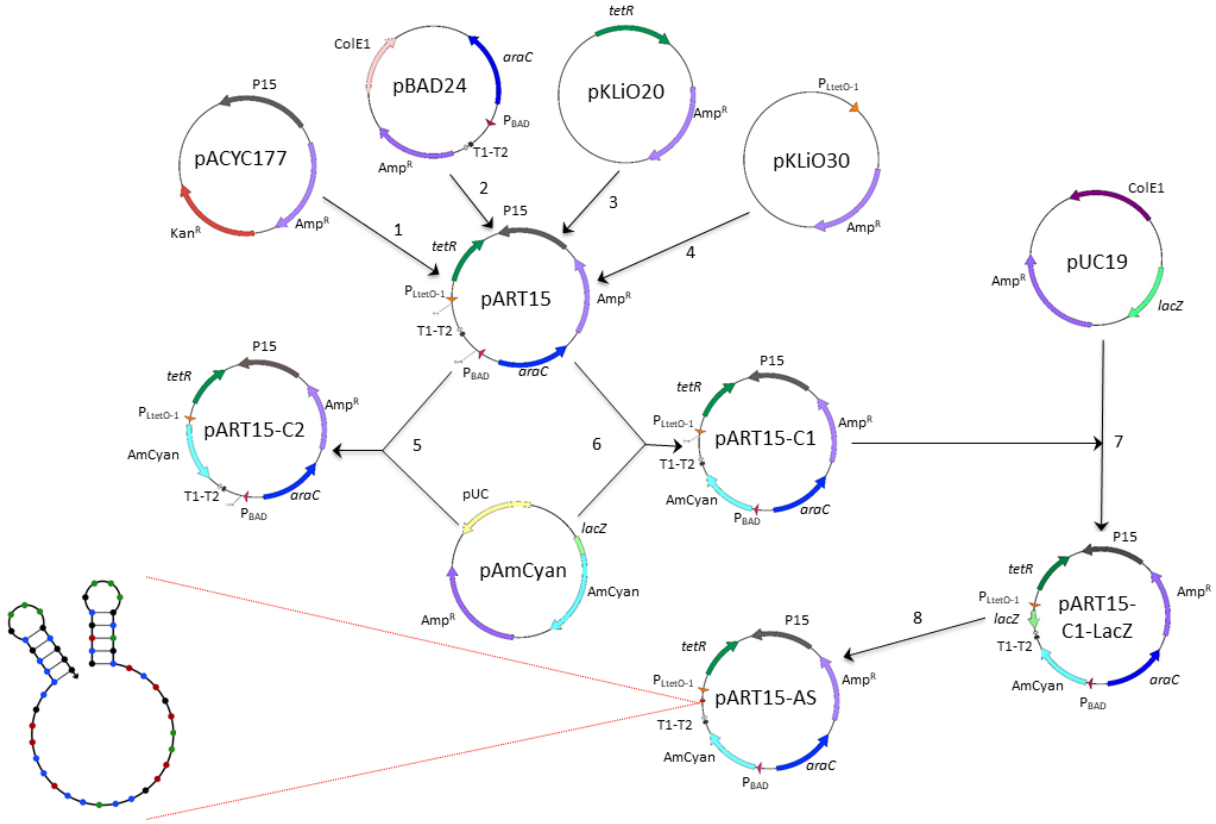
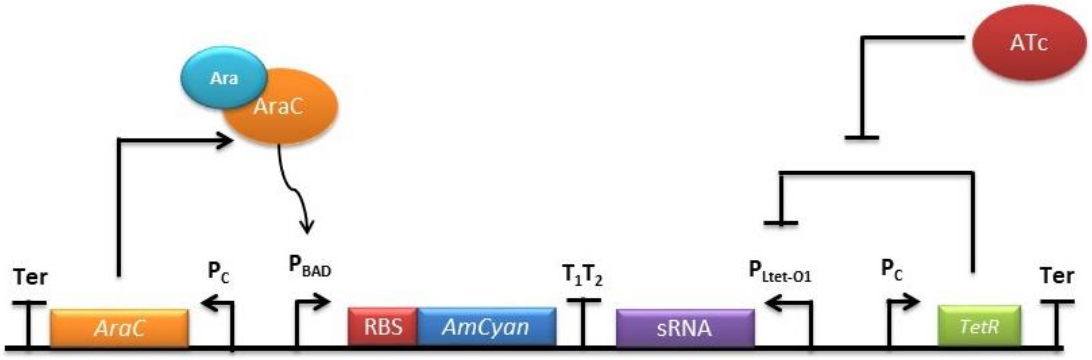


Figure B. 1. Construction of the single-plasmid dual-inducible pART15 reporter system. (1) Amplification of p15 origin of replication and ampicillin resistant gene from pACYC177, (2) amplification of *araC* gene, P_{BAD} promoter, and T1T2 terminator from pBAD24, (3) amplification of *tetR* gene from pKLiO20, (4) amplification of $P_{LtetO-1}$ promoter from pKLiO30 and cloning to construct pART15. (5) Construction of control plasmids pART15-C1 (6) and pART15-C2 (5) by cloning *AmCyan* gene from pAmCyan plasmid under the control of P_{BAD} promoter and $P_{LtetO-1}$ promoter for pART15-C1 and pART15-C2, respectively. (7) Construction of pART15-C1-LacZ by cloning *lacZ* gene downstream of $P_{LtetO-1}$ promoter. (8) Construction of pART15-AS from pART15-C1-LacZ by reverse PCR and blue/white screening results in generation of various reporter systems each contain a different asRNA sequence. The dual stem-loop structure (bottom-left) shows the secondary structure of a sample asRNA designed and used in this research.



Inducer	Output
Nothing	*AmCyan* OFF, sRNA OFF
Arabinose	*AmCyan* ON, sRNA OFF
ATc	*AmCyan* OFF, sRNA ON
Arabinose & ATc	*AmCyan* ON, sRNA ON

Figure B. 2. The synthetic circuit for sRNA and AmCyan expression used in pART15-AS.

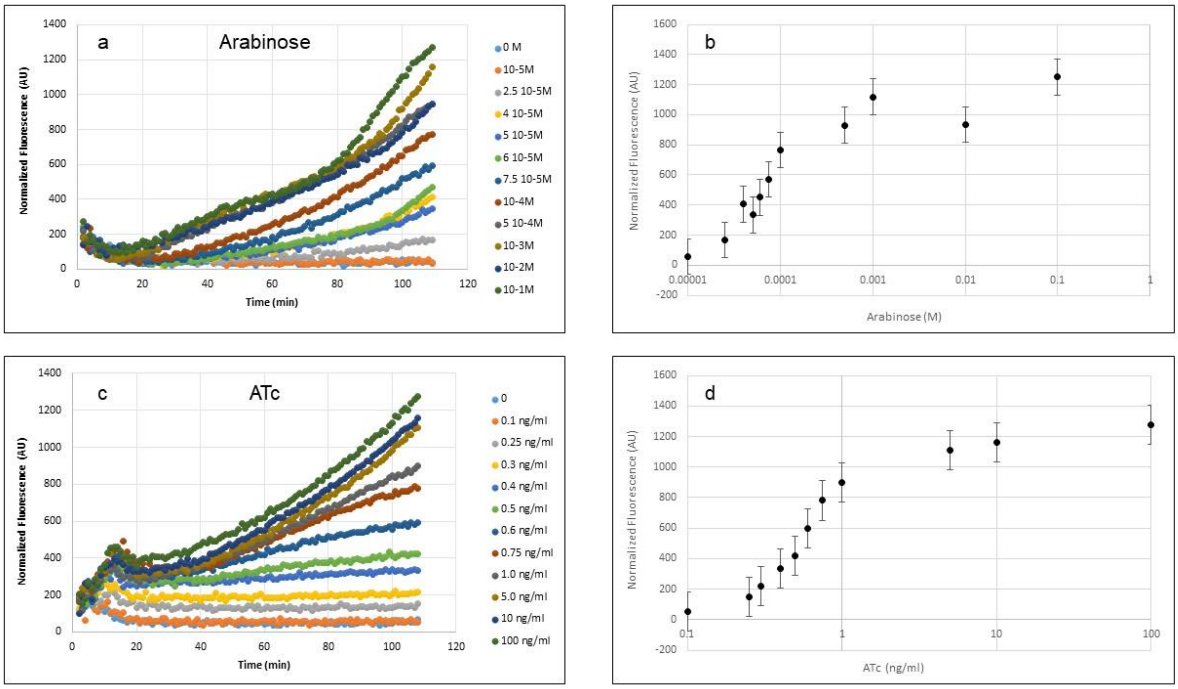


Figure B. 3. Modulation of expression from the araC-P_{BAD} (pART15-C1) and tetR-$P_{LtetO-1}$ (pART15-C2) systems. (a) Induction of pART15-C1 activity under the control of P_{BAD} at arabinose concentrations indicated. (b) Dependence of AmCyan levels on arabinose concentration. Data shown are the average values for three independent experiments. Error bars indicate 1 standard deviation at each arabinose concentration. (C) Induction of pART15-C2 activity under the control of $P_{LtetO-1}$ at ATc concentrations indicated. (b) Dependence of AmCyan levels on ATc concentration. Data shown are the average values for three independent experiments. Error bars indicate 1 standard deviation at each ATc concentration.

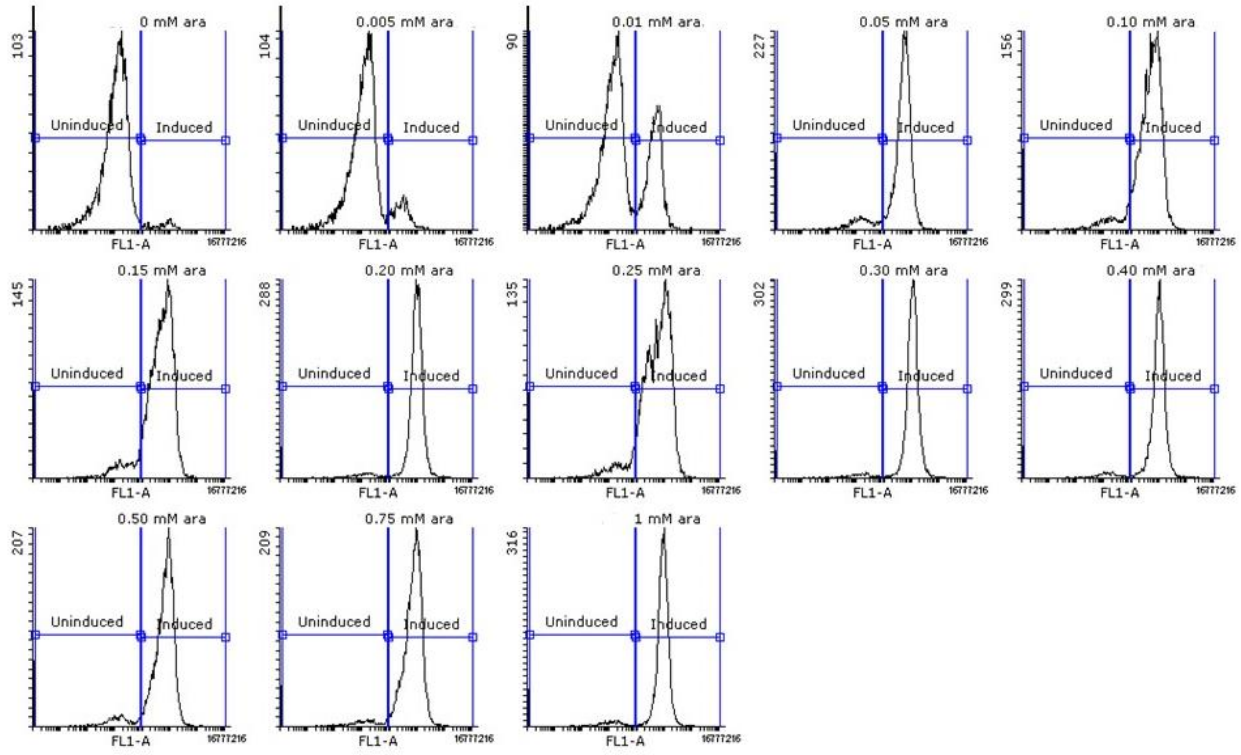


Figure B. 4. Flow cytometry analysis of induction of AmCyan expression with different concentrations of arabinose (pART15-C1). Each histogram represents 100,000 cells.

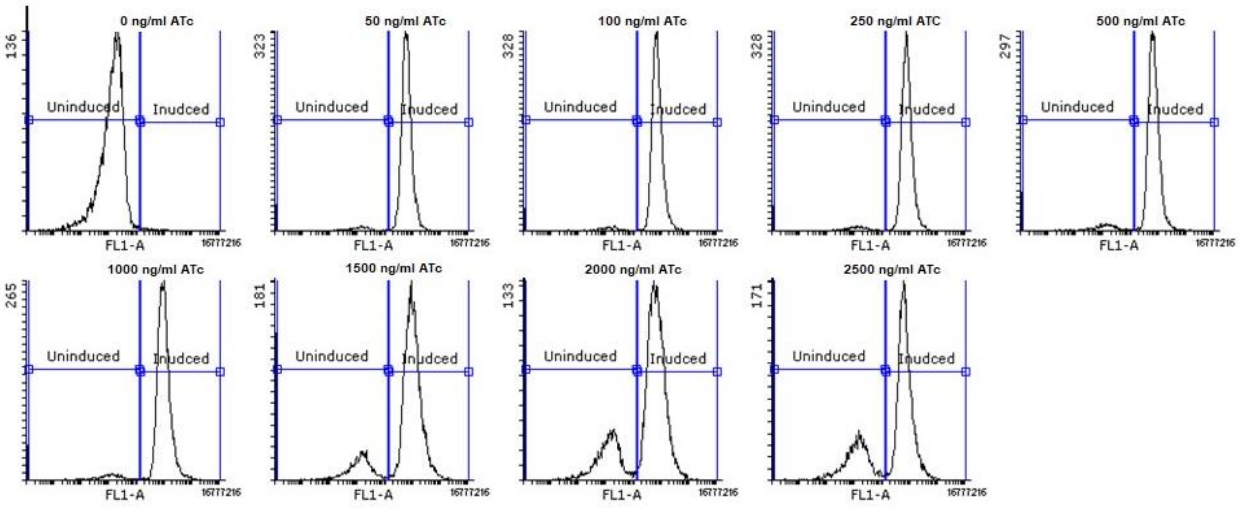


Figure B. 5. Flow cytometry analysis of induction of AmCyan expression with different concentrations of ATc (pART15-C2). Each histogram represents 100,000 cells.

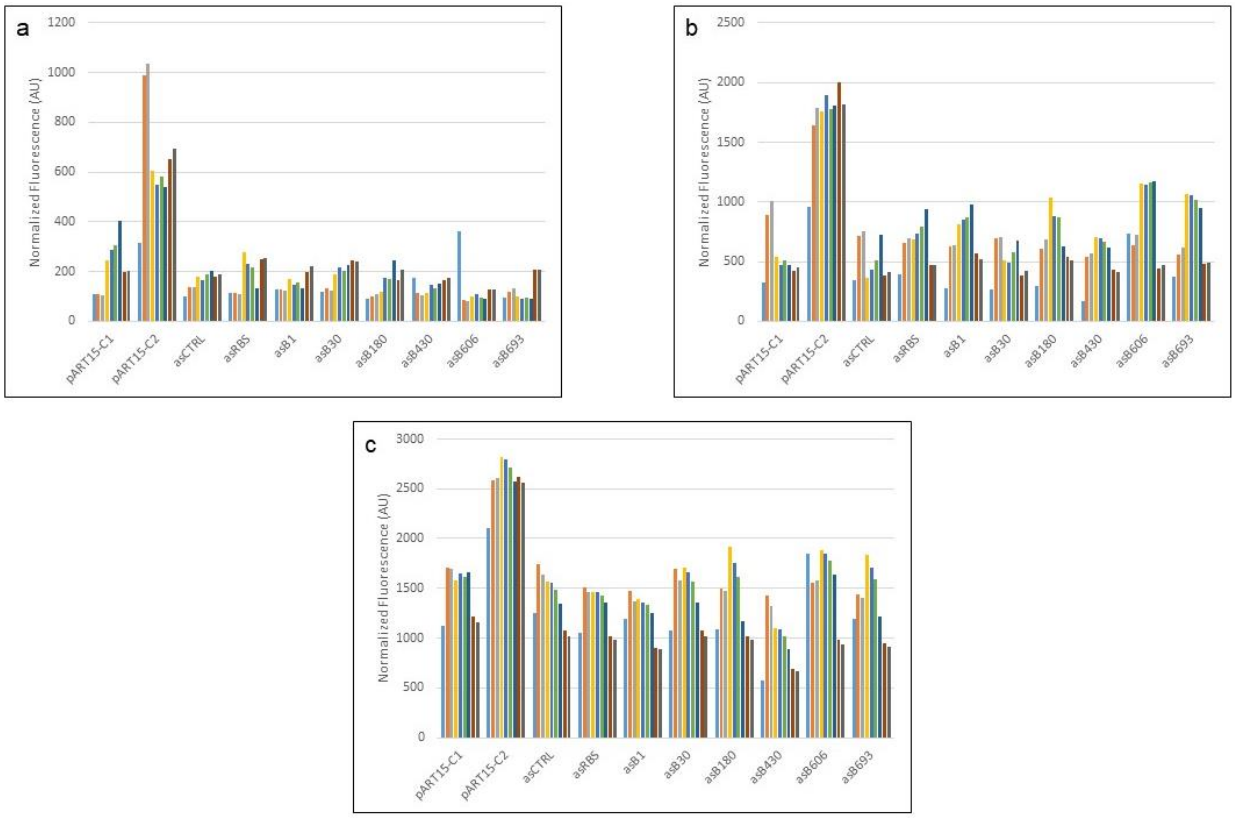


Figure B. 6. Normalized Fluorescence levels for the different sRNA when induced *in vivo* with both Ara and ATc. The following time points are represented: (a) 6hrs, (b) 18hrs and (c) 29 hrs.

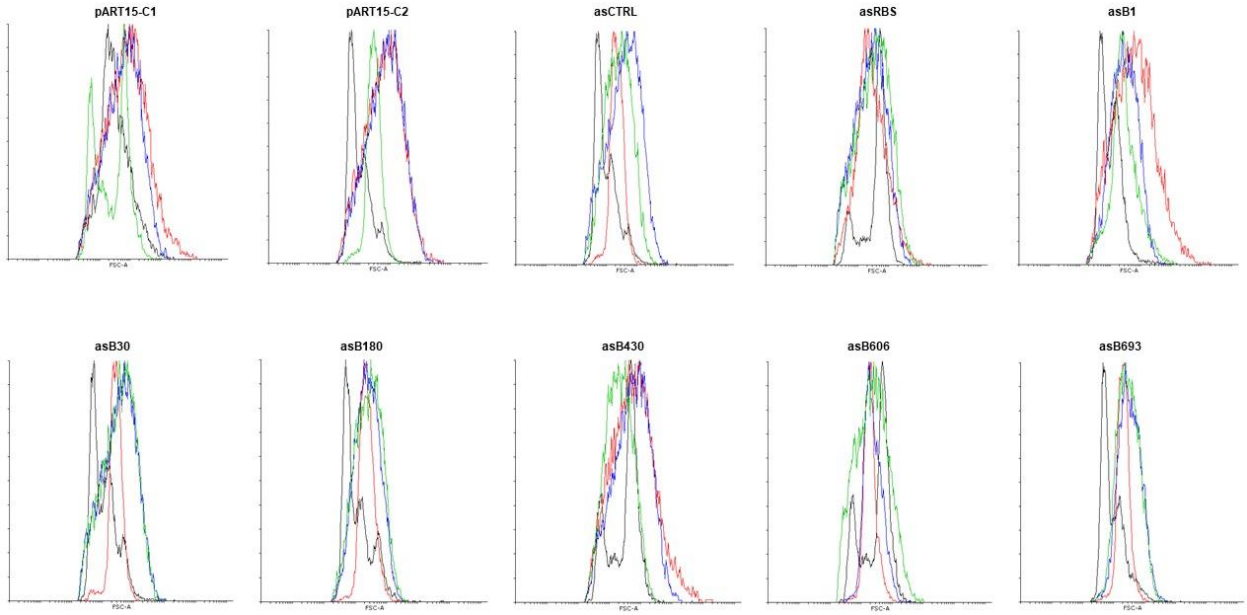


Figure B. 7. Flow cytometry results for induced cultures transformed with pART15-AS plasmids constructs. The antisense sequences are labeled above. Each histogram contains four biological replicates, each representing 100,000 cells.

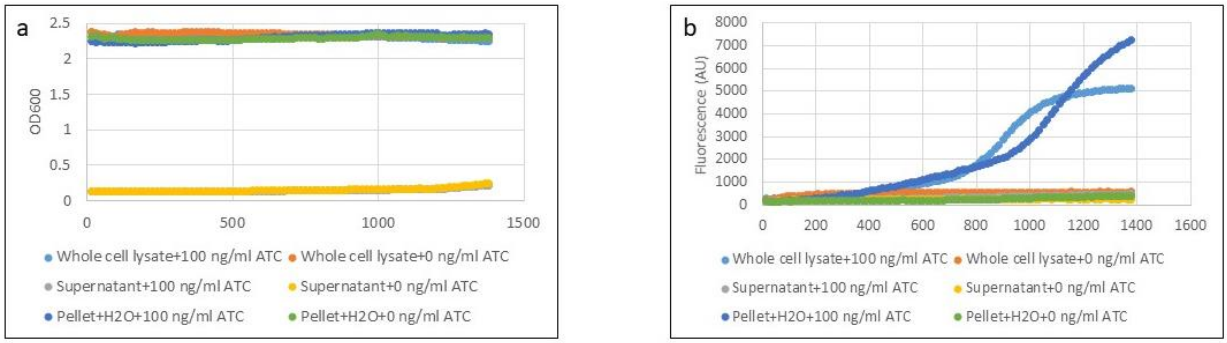


Figure B. 8. Optimization of sonication parameters: localization of the activity. (a) Optical density (OD_{600}). (b) Fluorescence level. Cell were sonicated as described in the manuscript Materials and Methods section. Fluorescence was measured on the whole cell lysate. Then the cell lysate was pelleted and the fluorescence was measured on both the cell pellet supplemented with water and the supernatant. All experiments were performed on pART15-C2 with and without ATc. Data shown are the average values for three independent experiments.

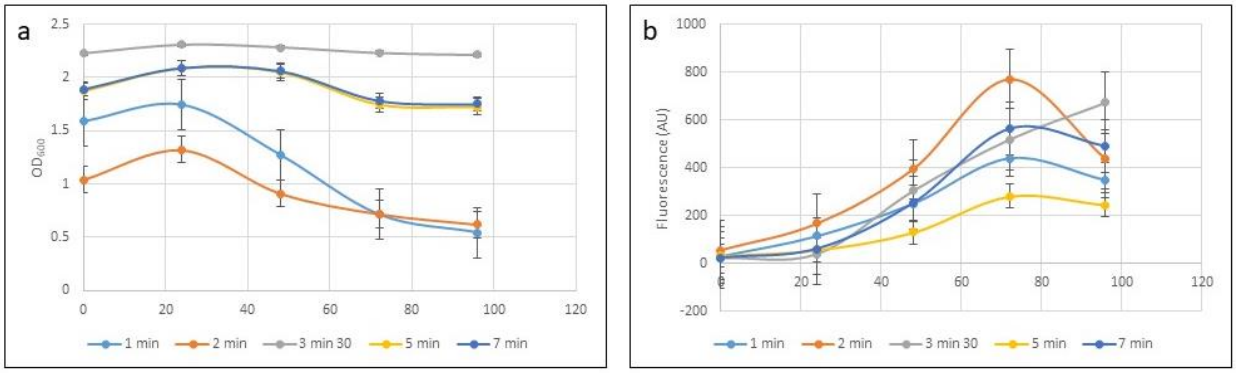


Figure B. 9. Optimization of sonication parameters: sonication time. (a) Optical density (OD_{600}). (b) Fluorescence level. A set volume of cells was sonicated at 5 different times: 1 min, 2 min, 3 min30, 5 min and 7 min.

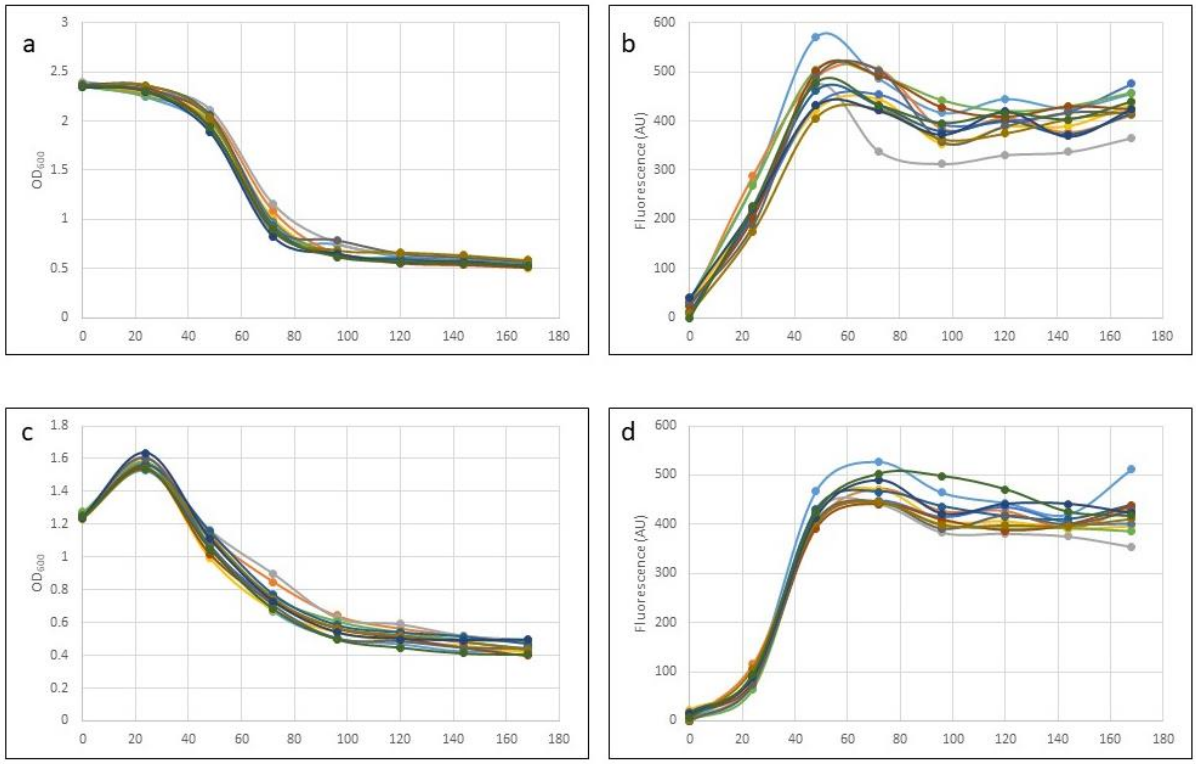


Figure B. 10. Optimization of sonication parameters: sonication volume (5 mL). (a) and (c) Optical density (OD600). (b) and (d) Fluorescence level. Each figure represents 12 biological replicates.

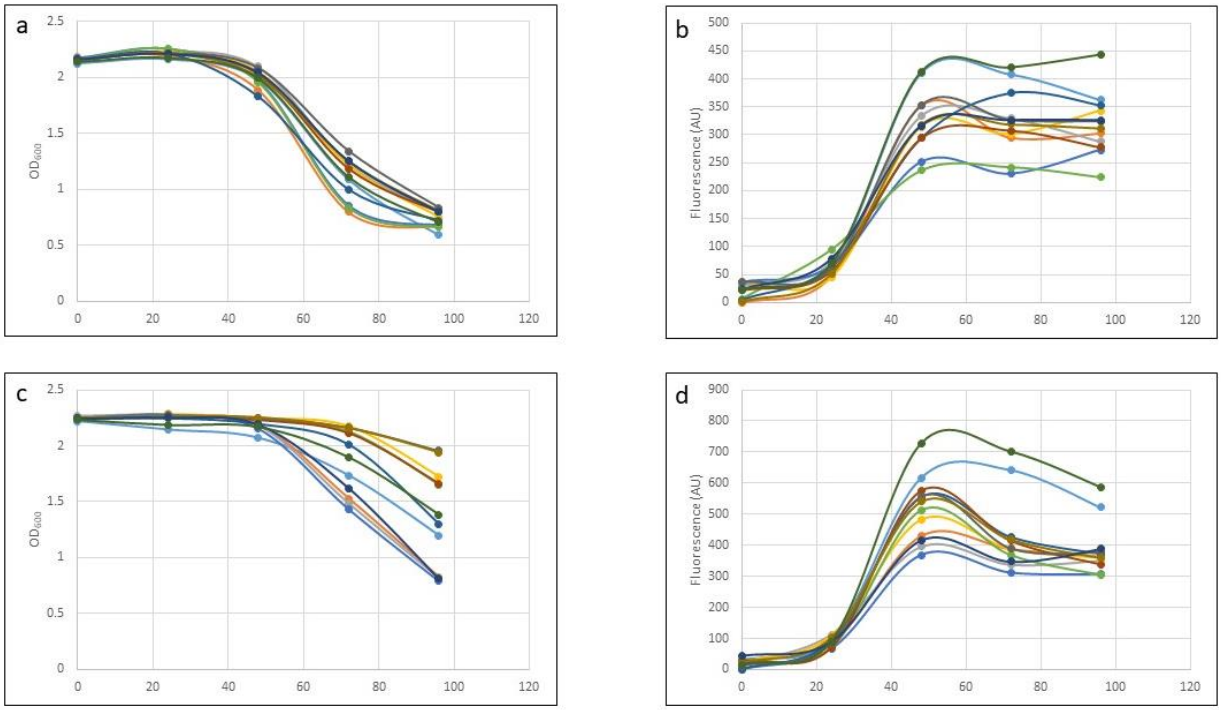


Figure B. 11. Optimization of sonication parameters: sonication volume (5.5 mL). (a) and (c) Optical density (OD600). (b) and (d) Fluorescence level. Each figure represents 12 biological replicates.

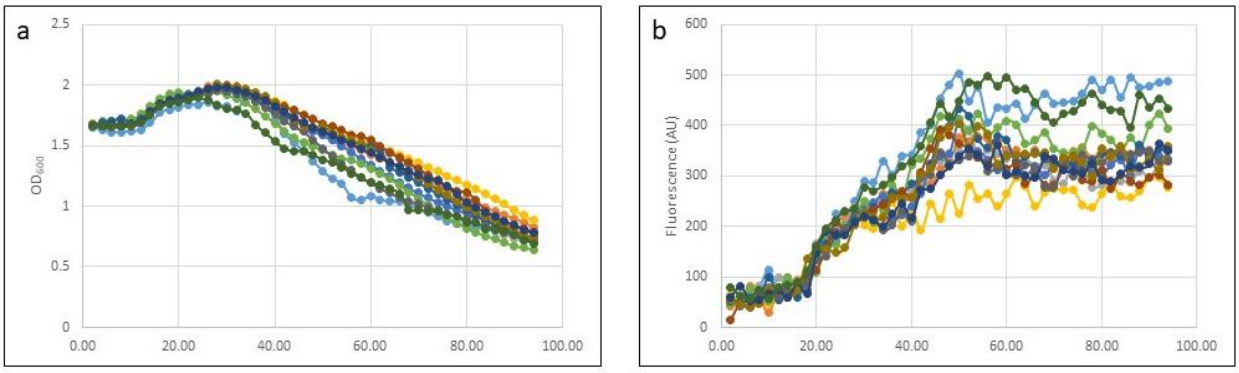


Figure B. 12. Optimization of sonication parameters: sonication volume (6.5 mL). (a) and (c) Optical density (OD600). (b) and (d) Fluorescence level. Each figure represents 12 biological replicates.

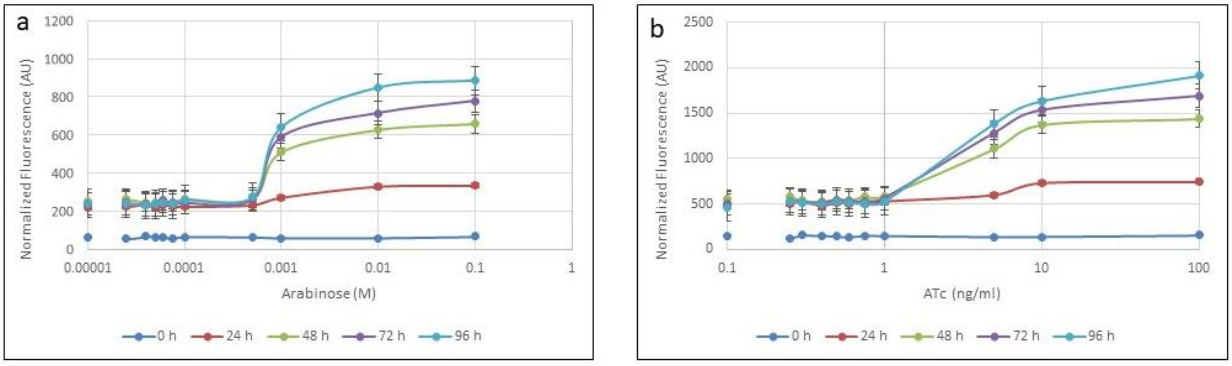


Figure B. 13. Modulation of expression from the araC-P_{BAD} (pART15-C1) and tetR-$P_{LtetO-1}$ (pART15-C2) cell-free-systems. (a) Induction of pART15-C1 activity under the control of P_{BAD} at arabinose concentrations indicated. (b) Induction of pART15-C2 activity under the control of $P_{LtetO-1}$ at ATc concentrations indicated. Data represent the average values for four independent experiments. Error bars indicate 1 standard deviation.

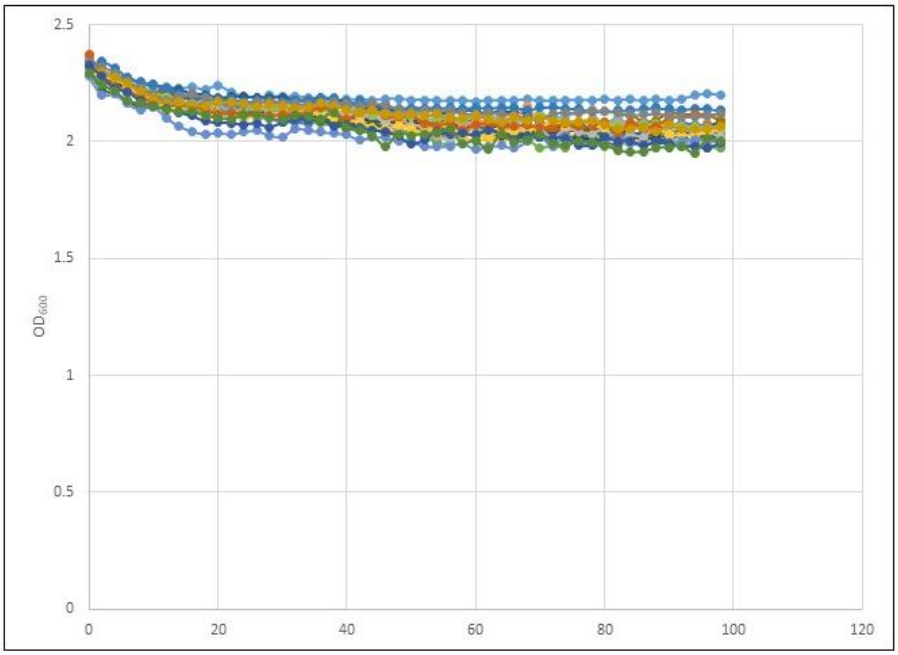


Figure B. 14. Stability of OD_{600} for different replicates after sonication. Data represent 24 biological replicates.

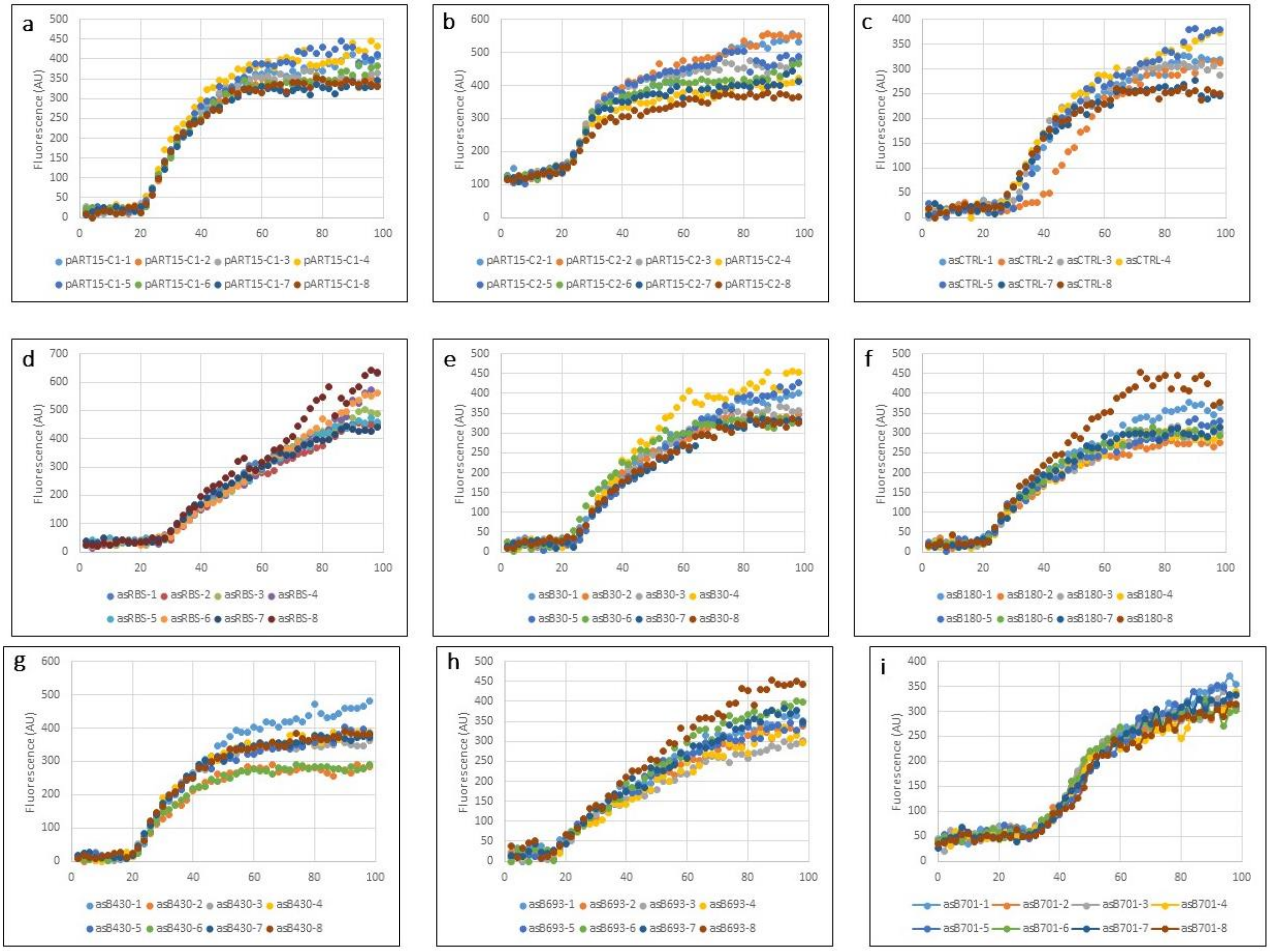


Figure B. 15. Normalized Fluorescence levels for the different pART15-AS plasmids after sonication and induction with 0.05 mM Ara and 100 ng/ml ATc. (a) pART15-C1, (b) pART15-C2, (c) pART15-asCTRL, (d) pART15-asRBS, (e) pART15-asB30, (f) pART15-asB180, (i) pART15-asB430, (j) pART15-asB693, (k) pART15-asB701. Each figure represents eight biological replicates.

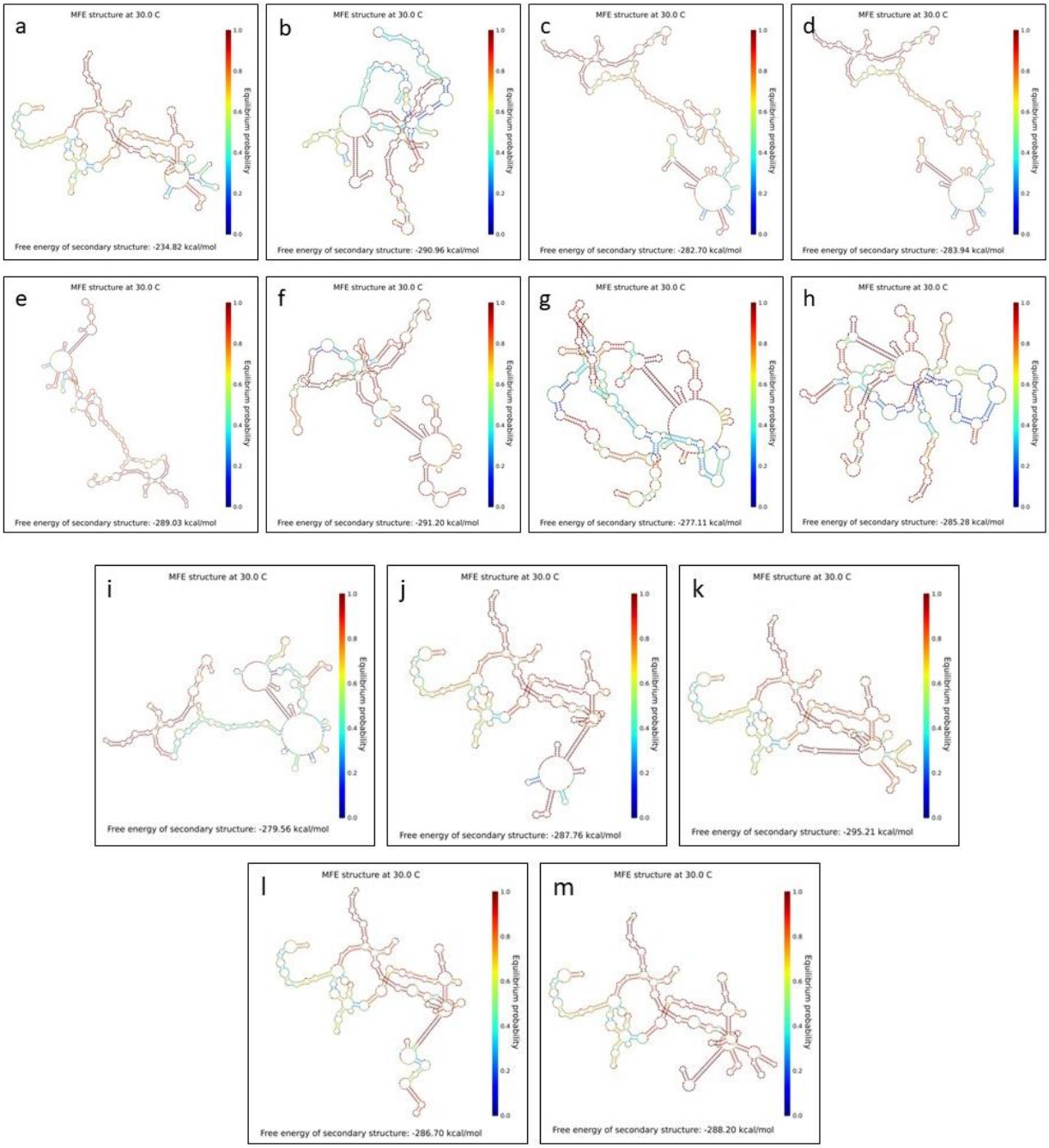


Figure B. 16. NUPACK predictions of the secondary structure interactions of the *AmCyan* mRNA with sRNA constructs. (a) *AmCyan* mRNA, (b) asRBS, (c) asB1, (d) asB30, (e) asB57, (f) asB180, (g) asB407, (h) asB430, (i) asB539, (j) asB606, (k) asB615, (l) asB693, and (m) asB701. Images shown here were exported from NUPACK directly.

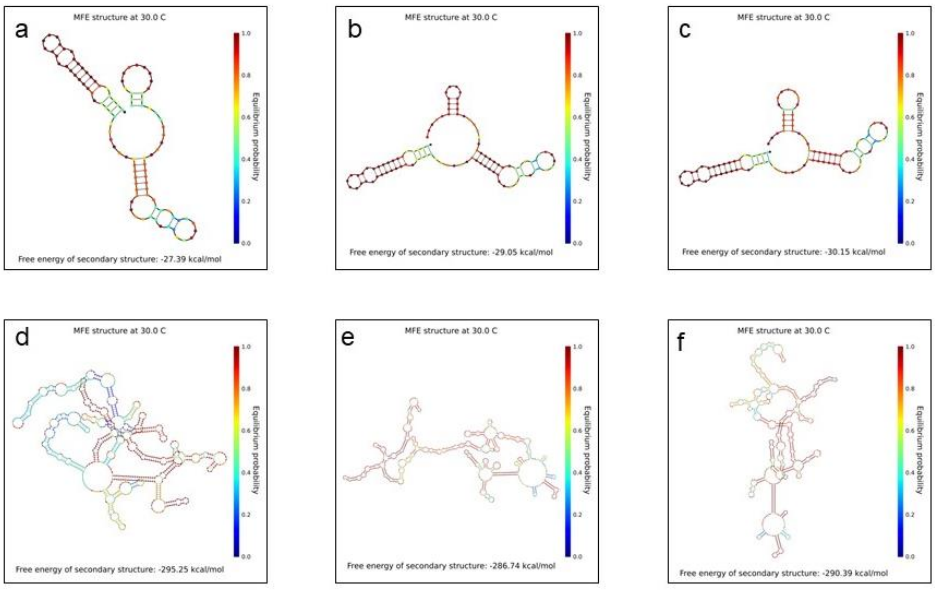


Figure B. 17. NUPACK predictions of secondary structures. (a) MicC-asRBS, (b) MicC-asB1, (c) MicC-asB606, (d) interaction of *AmCyan* mRNA with MicC-asRBS, (e) interaction of *AmCyan* mRNA with MicC-asB1, and (f) Interaction of *AmCyan* mRNA with MicC-asB606. Images shown here were exported from NUPACK directly.

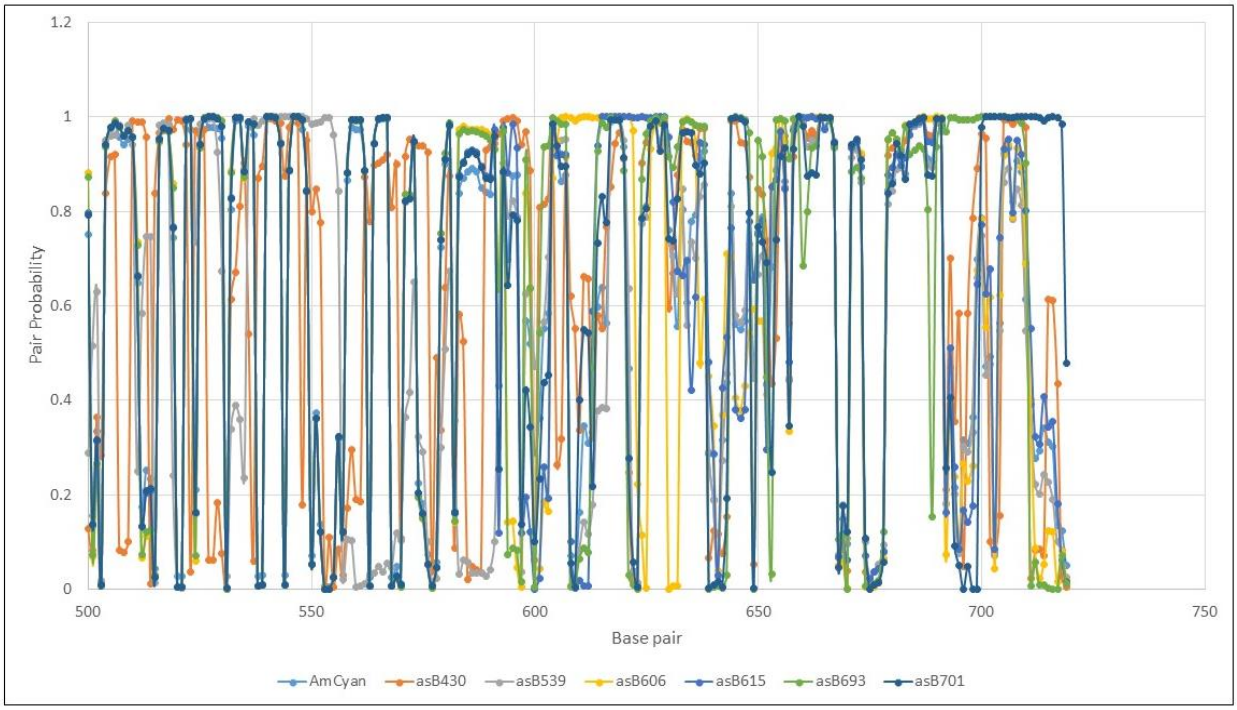


Figure B. 18. NUPACK predictions for the pairing probability of antisense binding sequences to the 3' end of *AmCyan* mRNA.

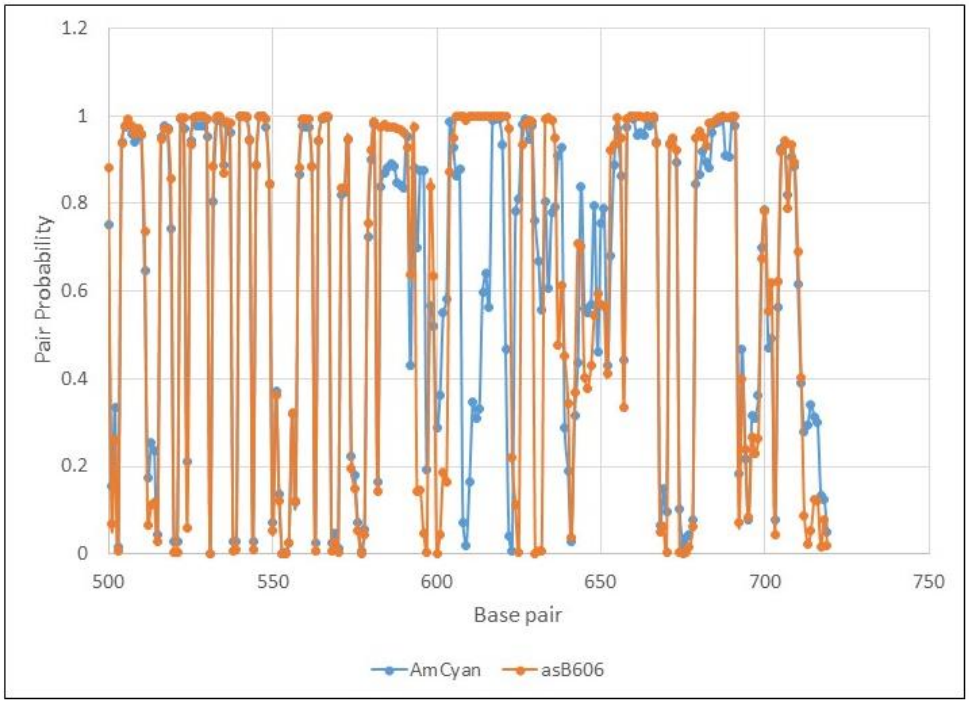


Figure B. 19. NUPACK predictions for the pairing probability of asB606 compared to *AmCyan* mRNA.

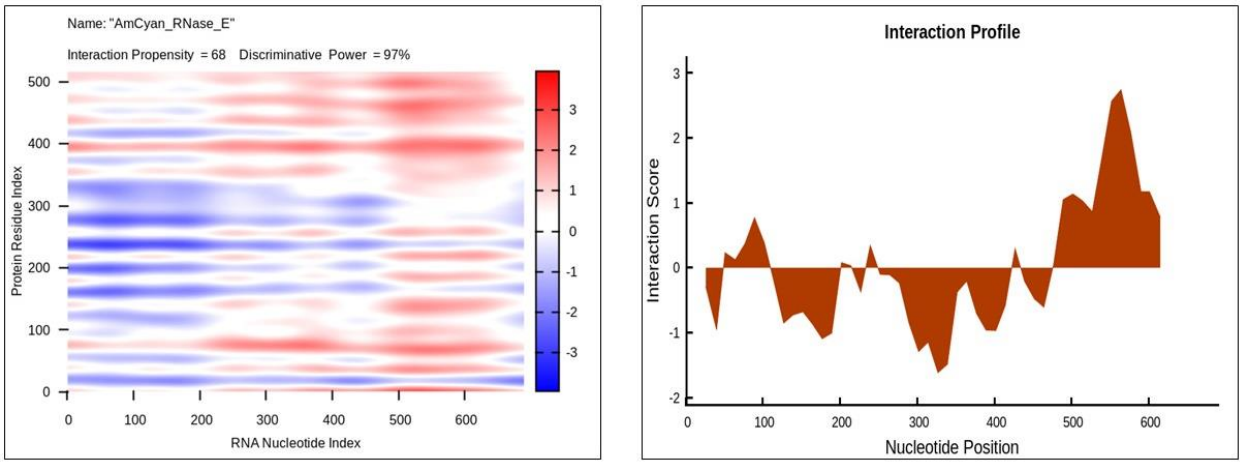


Figure B. 20. catRAPID prediction of the interactions between *AmCyan* mRNA and RNase E. (a) Heat-map of the interaction between *AmCyan* mRNA and RNase E. (b) Interaction profile of *AmCyan* mRNA and RNase E. Strong binding potential is shown between nucleotides 500-650. Images shown here were exported from catRAPID directly.

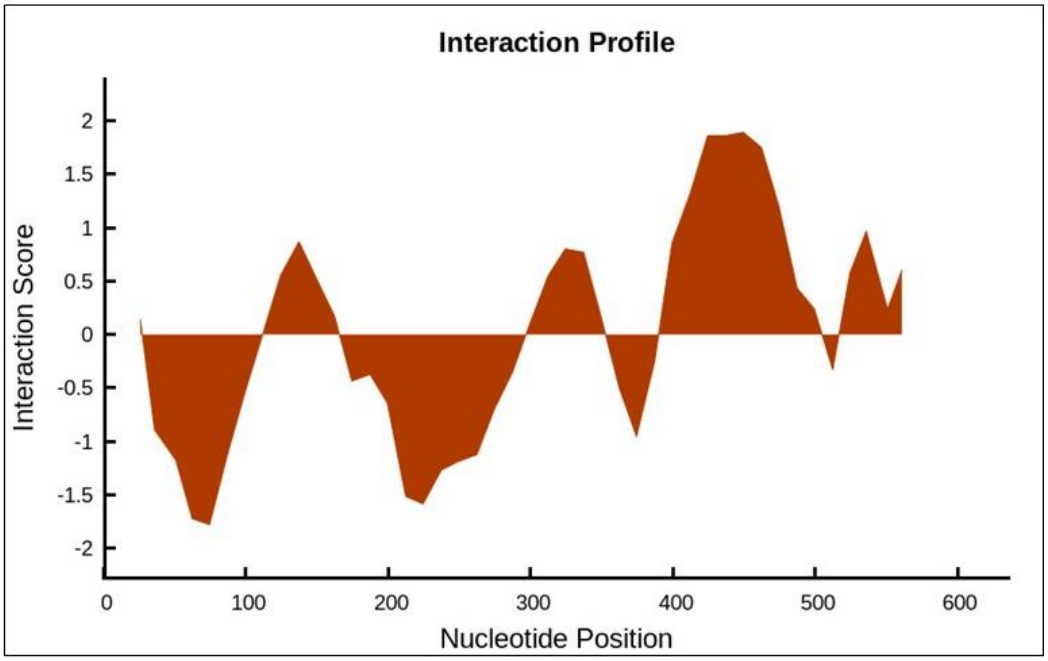


Figure B. 21. catRAPID predicted interaction profile between diaphorase mRNA and RNase E. Image shown here was exported from catRAPID directly.

Table B. 4. Statistical analysis of cells at 6 hours

Analysis at 6 hours		12 Replicates			12 Replicates			16 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-7.663202	Not Significant	Increase	25.05724	Significant	Increase	35.292322
asRBS	anti-RBS	Not Significant	Increase	2.5849878	Not Significant	Increase	6.2377387	Not Significant	Increase	7.8641634
asB1	anti B1-B20	Not Significant	Increase	14.218705	Not Significant	Increase	19.313995	Not Significant	Increase	21.543783
asB30	High energy loop	Not Significant	Increase	7.6869364	Not Significant	Increase	22.750795	Not Significant	Increase	18.882914
asB180	Low energy loop	Not Significant	Decrease	-17.33571	Not Significant	Decrease	-8.829779	Not Significant	Increase	8.3549574
asB430	High energy stem	Significant	Increase	58.664968	Not Significant	Increase	5.0890216	Not Significant	Increase	0.4277631
asB606	Low energy stem	Significant	Increase	225.34136	Not Significant	Decrease	-21.26586	Significant	Decrease	-22.67322
asB693	3' end	Not Significant	Decrease	-12.6226	Not Significant	Increase	9.0216175	Significant	Increase	27.522785

Analysis at 6 hours		12 Replicates			16 Replicates			20 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-25.28448	Significant	Decrease	-41.98634	Significant	Decrease	-38.04358
asRBS	anti-RBS	Not Significant	Increase	14.215538	Not Significant	Decrease	-19.86572	Significant	Decrease	-28.62801
asB1	anti B1-B20	Significant	Decrease	-29.34372	Significant	Decrease	-49.44783	Significant	Decrease	-49.28037
asB30	High energy loop	Not Significant	Decrease	-23.03424	Not Significant	Decrease	-23.71095	Significant	Decrease	-34.14256
asB180	Low energy loop	Significant	Decrease	-51.87402	Significant	Decrease	-38.80624	Significant	Decrease	-44.91313
asB430	High energy stem	Significant	Decrease	-52.46481	Significant	Decrease	-48.88103	Significant	Decrease	-57.35283
asB606	Low energy stem	Significant	Decrease	-58.31373	Significant	Decrease	-62.18697	Significant	Decrease	-68.28539
asB693	3' end	Significant	Decrease	-58.90158	Significant	Decrease	-68.95734	Significant	Decrease	-68.69067

Analysis at 6 hours		8 Replicates			12 Replicates			16 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Decrease	-49.60996	Not Significant	Decrease	-9.289039	Not Significant	Decrease	-6.32448
asRBS	anti-RBS	Significant	Decrease	-67.46655	Not Significant	Increase	24.240042	Not Significant	Increase	25.19224
asB1	anti B1-B20	Significant	Decrease	-67.35334	Not Significant	Decrease	-1.329196	Not Significant	Increase	8.1090045
asB30	High energy loop	Significant	Decrease	-44.21247	Not Significant	Increase	21.445525	Not Significant	Increase	18.33155
asB180	Low energy loop	Significant	Decrease	-38.60295	Not Significant	Decrease	-16.66149	Not Significant	Increase	2.5934136
asB430	High energy stem	Significant	Decrease	-61.78392	Not Significant	Decrease	-18.00572	Not Significant	Decrease	-13.24962
asB606	Low energy stem	Significant	Decrease	-77.32491	Significant	Decrease	-35.88656	Significant	Decrease	-36.3959
asB693	3' end	Significant	Decrease	-77.56468	Not Significant	Increase	4.6835431	Not Significant	Increase	3.0668937

Analysis at 6 hours		12 Replicates			64 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-5.619308	Not Significant	Decrease	-16.65345
asRBS	anti-RBS	Significant	Increase	51.630093	Not Significant	Decrease	-6.255507
asB1	anti B1-B20	Not Significant	Increase	20.191692	Not Significant	Decrease	-17.91576
asB30	High energy loop	Significant	Increase	37.028743	Not Significant	Decrease	-8.775777
asB180	Low energy loop	Not Significant	Increase	21.255962	Significant	Decrease	-22.35274
asB430	High energy stem	Not Significant	Decrease	-8.460255	Significant	Decrease	-25.45119
asB606	Low energy stem	Significant	Decrease	-28.26426	Significant	Decrease	-22.23165
asB693	3' end	Not Significant	Decrease	-24.57304	Significant	Decrease	-30.96131

Analysis at 6 hours		16 Replicates			12 Replicates			12 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Increase	40.761202	Not Significant	Increase	28.235969	Not Significant	Decrease	-16.17517
asB57	Low energy loop	Significant	Increase	47.706932	Significant	Increase	9.3540536	Not Significant	Decrease	-25.8866
asB407	High energy stem	Significant	Increase	39.094567	Significant	Increase	21.987944	Significant	Decrease	-36.67383
asB539	High energy stem	Significant	Increase	33.271043	Not Significant	Increase	15.344119	Significant	Decrease	-31.07188
asB615	Low energy stem	Not Significant	Increase	25.493485	Not Significant	Increase	8.0091017	Not Significant	Decrease	-15.50954
asB701	3' end	Significant	Increase	37.829949	Not Significant	Increase	7.4005435	Not Significant	Decrease	-13.89445

Analysis at 6 hours		40 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Increase	13.033286
asB57	Low energy loop	Not Significant	Increase	5.3109544
asB407	High energy stem	Not Significant	Increase	2.0657457
asB539	High energy stem	Not Significant	Increase	0.8202616
asB615	Low energy stem	Not Significant	Increase	3.0212919
asB701	3' end	Not Significant	Increase	7.0049184

*****Statistical analysis was performed using N-way analysis of variance (anovan) in comparison with control pART15-C1.

Table B. 5. Statistical analysis of cells at 18 hours

Analysis at 18 hours		12 Replicates			12 Replicates			16 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Increase	4.3397063	Not Significant	Decrease	-19.92991	Significant	Decrease	-25.45696
asRBS	anti-RBS	Not Significant	Increase	20.388573	Not Significant	Decrease	-26.77061	Significant	Decrease	-30.66551
asB1	anti B1-B20	Not Significant	Decrease	-17.41496	Not Significant	Decrease	-29.06459	Significant	Decrease	-36.33874
asB30	High energy loop	Not Significant	Decrease	-18.28792	Not Significant	Decrease	-21.8097	Significant	Decrease	-30.15082
asB180	Low energy loop	Not Significant	Decrease	-8.891648	Significant	Decrease	-31.19874	Significant	Decrease	-32.0784
asB430	High energy stem	Significant	Decrease	-48.41497	Significant	Decrease	-39.05451	Significant	Decrease	-43.41052
asB606	Low energy stem	Significant	Increase	122.6962	Significant	Decrease	-28.01504	Significant	Decrease	-27.81517
asB693	3' end	Not Significant	Increase	14.632378	Significant	Decrease	-36.70026	Significant	Decrease	-38.22331

Analysis at 18 hours		12 Replicates			16 Replicates			20 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-32.25984	Not Significant	Decrease	-7.536022	Not Significant	Decrease	-0.140804
asRBS	anti-RBS	Not Significant	Increase	26.932489	Significant	Increase	55.594663	Significant	Increase	54.047143
asB1	anti B1-B20	Significant	Increase	49.334714	Significant	Increase	81.249425	Significant	Increase	70.854608
asB30	High energy loop	Not Significant	Decrease	-5.72862	Not Significant	Increase	3.6204659	Not Significant	Increase	12.403016
asB180	Low energy loop	Significant	Increase	90.998953	Significant	Increase	85.454474	Significant	Increase	70.524074
asB430	High energy stem	Not Significant	Increase	29.517837	Significant	Increase	46.868527	Significant	Increase	30.211988
asB606	Low energy stem	Significant	Increase	113.68292	Significant	Increase	142.80968	Significant	Increase	127.54198
asB693	3' end	Significant	Increase	96.947623	Significant	Increase	123.12852	Significant	Increase	98.862198

Analysis at 18 hours		8 Replicates			12 Replicates			16 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Increase	55.626948	Not Significant	Decrease	-9.70834	Not Significant	Decrease	-6.943725
asRBS	anti-RBS	Significant	Increase	101.12587	Not Significant	Increase	10.816202	Not Significant	Increase	5.2635583
asB1	anti B1-B20	Significant	Increase	108.21925	Significant	Increase	33.341433	Not Significant	Increase	15.9821
asB30	High energy loop	Not Significant	Increase	43.884678	Not Significant	Decrease	-8.232026	Not Significant	Decrease	-5.917549
asB180	Low energy loop	Not Significant	Increase	34.973877	Not Significant	Increase	26.518942	Not Significant	Increase	14.829511
asB430	High energy stem	Not Significant	Increase	31.41723	Not Significant	Increase	2.8598941	Not Significant	Decrease	-7.153498
asB606	Low energy stem	Significant	Increase	151.60523	Not Significant	Increase	5.2939669	Not Significant	Increase	4.4153944
asB693	3' end	Significant	Increase	102.18644	Not Significant	Increase	14.725811	Not Significant	Increase	10.550582

Analysis at 18 hours		12 Replicates			64 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-192.2916	Not Significant	Decrease	-11.84918
asRBS	anti-RBS	Not Significant	Increase	-206.1378	Not Significant	Increase	4.7540968
asB1	anti B1-B20	Not Significant	Increase	-208.0914	Not Significant	Increase	4.9729289
asB30	High energy loop	Not Significant	Decrease	-196.3262	Not Significant	Decrease	-12.62888
asB180	Low energy loop	Not Significant	Increase	-215.1794	Not Significant	Increase	7.3920946
asB430	High energy stem	Not Significant	Decrease	-185.4294	Not Significant	Decrease	-16.87476
asB606	Low energy stem	Not Significant	Increase	-204.1127	Significant	Increase	36.700554
asB693	3' end	Not Significant	Increase	-205.1565	Not Significant	Increase	14.160336

Analysis at 18 hours		16 Replicates			12 Replicates			12 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Decrease	-35.71189	Significant	Decrease	-38.41668	Not Significant	Decrease	-6.129228
asB57	Low energy loop	Significant	Decrease	-32.75344	Not Significant	Decrease	-9.682016	Not Significant	Increase	7.5371203
asB407	High energy stem	Significant	Decrease	-54.05706	Not Significant	Decrease	-23.09818	Not Significant	Decrease	-13.06164
asB539	High energy stem	Not Significant	Decrease	-17.5194	Not Significant	Decrease	-2.863929	Not Significant	Increase	6.7126095
asB615	Low energy stem	Significant	Decrease	-28.15991	Not Significant	Increase	1.4701739	Not Significant	Decrease	-3.711088
asB701	3' end	Significant	Decrease	-23.43433	Not Significant	Decrease	-13.80121	Not Significant	Decrease	-14.21201

Analysis at 18 hours		40 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Decrease	-32.09605
asB57	Low energy loop	Not Significant	Decrease	-23.28025
asB407	High energy stem	Significant	Decrease	-43.14586
asB539	High energy stem	Not Significant	Decrease	-11.68939
asB615	Low energy stem	Significant	Decrease	-19.75427
asB701	3' end	Not Significant	Decrease	-20.52607

*****Statistical analysis was performed using N-way analysis of variance (anovan) in comparison with control pART15-C1.

Table B. 6. Statistical analysis of cells at 29 hours

Analysis at 29 hours		12 Replicates			12 Replicates			16 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not significant	Decrease	-5.683361	Not significant	Decrease	-15.51244	Significant	Decrease	-17.04073
asRBS	anti-RBS	Significant	Decrease	-20.01115	Not significant	Decrease	-7.748991	Significant	Decrease	-9.709909
asB1	anti B1-B20	Significant	Decrease	-20.74451	Significant	Decrease	-14.54478	Significant	Decrease	-18.56455
asB30	High energy loop	Not significant	Decrease	-10.97003	Not significant	Decrease	-10.60853	Significant	Decrease	-16.73551
asB180	Low energy loop	Significant	Decrease	-18.21651	Not significant	No effect	-0.207514	Not significant	No effect	-2.086171
asB430	High energy stem	Significant	Decrease	-21.36917	Significant	Increase	19.274308	Not significant	Increase	5.6307373
asB606	Low energy stem	Significant	Decrease	-13.54864	Not significant	Increase	10.468811	Significant	Increase	13.30714
asB693	3' end	Not significant	Increase	8.0104533	Not significant	Decrease	-4.274197	Not significant	Decrease	-7.862256

Analysis at 29 hours		12 Replicates			16 Replicates			20 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not significant	No effect	-0.670305	Not significant	Decrease	-6.05996	Not significant	Decrease	-8.154373
asRBS	anti-RBS	Not significant	Decrease	-7.509871	Not significant	Decrease	-11.7324	Significant	Decrease	-11.8513
asB1	anti B1-B20	Significant	Decrease	-12.08383	Significant	Decrease	-18.1036	Significant	Decrease	-17.15353
asB30	High energy loop	Not significant	Increase	7.8039586	Not significant	No effect	0.7477681	Not significant	Decrease	-2.784887
asB180	Low energy loop	Significant	Increase	20.988624	Not significant	Increase	6.2110524	Not significant	No effect	0.3505881
asB430	High energy stem	Significant	Decrease	-30.49322	Significant	Decrease	-34.02691	Significant	Decrease	-36.91787
asB606	Low energy stem	Significant	Increase	18.906544	Not significant	Increase	11.984707	Not significant	Increase	10.503254
asB693	3' end	Significant	Increase	16.371877	Not significant	Increase	3.0871631	Not significant	Decrease	-1.379601

Analysis at 29 hours		8 Replicates			12 Replicates			16 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Decrease	-18.85576	Not significant	Decrease	-11.64823	Not significant	Decrease	-11.65833
asRBS	anti-RBS	Significant	Decrease	-18.05907	Significant	Decrease	-16.05889	Significant	Decrease	-14.89099
asB1	anti B1-B20	Significant	Decrease	-24.40265	Significant	Decrease	-26.28668	Significant	Decrease	-23.54894
asB30	High energy loop	Significant	Decrease	-17.92576	Not significant	Decrease	-11.14793	Significant	Decrease	-12.24739
asB180	Low energy loop	Significant	Decrease	-29.15951	Significant	Decrease	-16.64868	Significant	Decrease	-15.46537
asB430	High energy stem	Significant	Decrease	-46.1044	Significant	Decrease	-43.43012	Significant	Decrease	-42.01807
asB606	Low energy stem	Not significant	Decrease	-1.512516	Significant	Decrease	-18.90089	Significant	Decrease	-18.64338
asB693	3' end	Significant	Decrease	-26.76224	Significant	Decrease	-21.99318	Significant	Decrease	-21.40519

Analysis at 29 hours		12 Replicates			64 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not significant	Decrease	-9.535941	Not significant	Decrease	-4.554005
asRBS	anti-RBS	Significant	Decrease	-16.18159	Significant	Decrease	-12.29715
asB1	anti B1-B20	Significant	Decrease	-17.44661	Significant	Decrease	-15.63684
asB30	High energy loop	Not significant	Decrease	-10.47858	Not significant	Decrease	-6.344419
asB180	Low energy loop	Significant	Decrease	-13.47242	Not significant	Decrease	-7.436986
asB430	High energy stem	Significant	Decrease	-39.59259	Significant	Decrease	-35.45169
asB606	Low energy stem	Significant	Decrease	-18.02169	Not significant	Increase	7.4302766
asB693	3' end	Significant	Decrease	-18.27027	Not significant	Decrease	-9.112476

Analysis at 29 hours		16 Replicates			12 Replicates			12 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-3.497204	Significant	Decrease	-30.51684	Singificant	Decrease	-29.70957
asB57	Low energy loop	Significant	Decrease	-8.332703	Significant	Decrease	-45.65555	Singificant	Decrease	-46.39952
asB407	High energy stem	Not Significant	Increase	1.1458683	Significant	Decrease	-37.09147	Singificant	Decrease	-36.76632
asB539	High energy stem	Not Significant	Decrease	-2.826996	Significant	Decrease	-29.72585	Singificant	Decrease	-43.37282
asB615	Low energy stem	Not Significant	Increase	5.0977567	Significant	Decrease	-38.13665	Singificant	Decrease	-48.32856
asB701	3' end	Not Significant	Decrease	-3.892298	Significant	Decrease	-32.81387	Singificant	Decrease	-42.3807

Analysis at 29 hours		40 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Decrease	-15.91455
asB57	Low energy loop	Significant	Decrease	-25.88216
asB407	High energy stem	Not Significant	Decrease	-16.60134
asB539	High energy stem	Significant	Decrease	-18.27546
asB615	Low energy stem	Not Significant	Decrease	-17.22425
asB701	3' end	Significant	Decrease	-19.41173

*****Statistical analysis was performed using N-way analysis of variance (anovan) in comparison with control pART15-C1.

Table B. 7. Statistical analysis of Cell-free-system at 48 hours

Analysis at 48 hours		6 Replicates			8 Replicates			8 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Increase	-11.20146925	Significant	Increase	23.128136	Not Significant	Decrease	-3.953029
asRBS	anti-RBS	Not Significant	Increase	-1.372037921	Significant	Increase	16.279269	Not Significant	Decrease	-11.07189
asB1	anti B1-B20	Not Significant	Decrease	-18.2314361	Not Significant	Decrease	-7.451443	Significant	Decrease	-33.57258
asB30	High energy loop	Not Significant	Decrease	-18.09767613	Significant	Increase	26.732461	Not Significant	Decrease	13.893922
asB180	Low energy loop	Not Significant	Decrease	-9.801968869	Not Significant	Decrease	-0.709013	Not Significant	Decrease	1.8232121
asB430	High energy stem	Not Significant	Increase	26.60149975	Significant	Increase	84.959211	Not Significant	Decrease	-19.40422
asB606	Low energy stem	Significant	Increase	44.4266578	Significant	Increase	78.516569	Significant	Increase	62.577262
asB693	3' end	Not Significant	Decrease	-13.57524021	Significant	Increase	16.279269	Not Significant	Decrease	-5.014126

Analysis at 48 hours		8 Replicates			30 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Decrease	-34.99905	Not Significant	Decrease	-8.678047
asRBS	anti-RBS	Significant	Decrease	-30.80636	Not Significant	Decrease	-8.489924
asB1	anti B1-B20	Significant	Decrease	-27.39468	Significant	Decrease	-22.12331
asB30	High energy loop	Significant	Decrease	-24.88976	Not Significant	Decrease	-2.075451
asB180	Low energy loop	Significant	Decrease	-26.96888	Not Significant	Decrease	-10.14512
asB430	High energy stem	Not Significant	Decrease	-5.238043	Not Significant	Increase	19.71038
asB606	Low energy stem	Significant	Increase	30.134399	Significant	Increase	52.368835
asB693	3' end	Significant	Decrease	-33.18081	Not Significant	Decrease	-10.51457

Analysis at 48 hours		16 Replicates			16 Replicates			32 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asB57	Low energy loop	Significant	Decrease	-47.22086	Significant	Decrease	-14.74997	Significant	Decrease	-32.29603
asB407	High energy stem	Significant	Decrease	-48.82772	Significant	Decrease	-27.63458	Significant	Decrease	-39.08656
asB539	High energy stem	Significant	Decrease	-57.46463	Significant	Decrease	-39.11799	Significant	Decrease	-49.03183
asB615	Low energy stem	Significant	Decrease	-38.33697	Significant	Decrease	-26.05452	Significant	Decrease	-32.6915
asB701	3' end	Significant	Decrease	-40.20867	Significant	Decrease	-36.38779	Significant	Decrease	-38.45245

*****Statistical analysis was performed using N-way analysis of variance (anovan) in comparison with control pART15-C1.

Table B. 8. Statistical analysis of Cell-free-system at 72 hours

Analysis at 72 hours		6 Replicates			8 Replicates			8 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-4.280866598	Not Significant	Increase	10.640972	Not Significant	Decrease	16.942851
asRBS	anti-RBS	Not Significant	Increase	7.771485989	Not Significant	Increase	1.7493727	Not Significant	Decrease	-11.02253
asB1	anti B1-B20	Not Significant	Decrease	-7.992151962	Significant	Decrease	-23.26966	Significant	Decrease	-32.23958
asB30	High energy loop	Not Significant	Decrease	-4.326132153	Significant	Increase	23.940376	Not Significant	Decrease	10.124491
asB180	Low energy loop	Not Significant	Decrease	-18.49752523	Not Significant	Increase	1.3422884	Not Significant	Decrease	-8.598146
asB430	High energy stem	Not Significant	Increase	13.75482238	Significant	Increase	74.495343	Not Significant	Decrease	-19.92038
asB606	Low energy stem	Significant	Increase	52.49089518	Significant	Increase	105.64317	Significant	Increase	57.081073
asB693	3' end	Not Significant	Decrease	-6.819328562	Not Significant	Increase	1.7493727	Not Significant	Decrease	-5.790364

Analysis at 72 hours		8 Replicates			30 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Significant	Decrease	-19.14528	Not Significant	Increase	0.8209287
asRBS	anti-RBS	Not Significant	Increase	6.4414176	Not Significant	Increase	1.1433757
asB1	anti B1-B20	Not Significant	Increase	-4.948475	Significant	Decrease	-17.0942
asB30	High energy loop	Not Significant	Decrease	-5.116547	Not Significant	Increase	6.0012479
asB180	Low energy loop	Not Significant	Decrease	-11.34196	Not Significant	Decrease	-9.263868
asB430	High energy stem	Significant	Decrease	-4.605267	Not Significant	Increase	14.924015
asB606	Low energy stem	Significant	Increase	53.796714	Significant	Increase	66.860869
asB693	3' end	Significant	Decrease	-14.10125	Not Significant	Decrease	-6.418191

Analysis at 72 hours		16 Replicates			16 Replicates			32 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asB57	Low energy loop	Significant	Decrease	-74.71092	Not Significant	Decrease	-9.581926	Significant	Decrease	-47.96216
asB407	High energy stem	Significant	Decrease	-71.93235	Not Significant	Increase	6.0039617	Significant	Decrease	-39.92357
asB539	High energy stem	Significant	Decrease	-61.09395	Significant	Decrease	-70.77354	Significant	Decrease	-65.0694
asB615	Low energy stem	Significant	Decrease	-61.60226	Significant	Decrease	-53.10634	Significant	Decrease	-58.11295
asB701	3' end	Significant	Decrease	-75.39302	Significant	Decrease	-67.11225	Significant	Decrease	-71.99207

*****Statistical analysis was performed using N-way analysis of variance (anovan) in comparison with control pART15-C1.

Table B. 9. Statistical analysis of Cell-free-system at 96 hours

Analysis at 96 hours		6 Replicates			8 Replicates			8 Replicates		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Not Significant	Decrease	-7.739444766	Not Significant	Increase	5.5820883	Not Significant	Decrease	6.7512092
asRBS	anti-RBS	Not Significant	Decrease	-1.843620524	Not Significant	Increase	9.3804799	Not Significant	Decrease	-8.453662
asB1	anti B1-B20	Not Significant	Increase	3.609065193	Significant	Decrease	-27.53853	Significant	Decrease	-30.78369
asB30	High energy loop	Not Significant	Decrease	-10.5513797	Significant	Increase	26.765778	Not Significant	Decrease	7.0489649
asB180	Low energy loop	Not Significant	Decrease	-16.28182846	Not Significant	Increase	11.18673	Significant	Decrease	-20.42169
asB430	High energy stem	Not Significant	Increase	10.00441857	Significant	Increase	62.130237	Significant	Decrease	-26.08113
asB606	Low energy stem	Not Significant	Increase	43.28356403	Significant	Increase	120.96215	Significant	Decrease	46.372422
asB693	3' end	Not Significant	Decrease	-14.71706127	Not Significant	Increase	9.3804799	Not Significant	Decrease	-2.749724

Analysis at 96 hours		8 Replicates			30 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage
asCTRL	asRNA without binding region	Siginificant	Decrease	-15.46696	Not Significant	Decrease	-2.810698
asRBS	anti-RBS	Siginificant	Increase	35.930829	Not Significant	Increase	8.649579
asB1	anti B1-B20	Not Significant	Increase	9.142867	Not Significant	Decrease	-11.25455
asB30	High energy loop	Not Significant	Decrease	-1.352749	Not Significant	Increase	5.0559446
asB180	Low energy loop	Siginificant	Decrease	-16.03485	Not Significant	Decrease	-10.98527
asB430	High energy stem	Not Significant	Decrease	-2.355732	Not Significant	Increase	9.3054399
asB606	Low energy stem	Siginificant	Increase	59.510229	Significant	Increase	66.146807
asB693	3' end	Not Significant	Decrease	-4.404644	Not Significant	Decrease	-3.361623

Analysis at 96 hours		16 Replicates			16 Replicates			32 Replicates (all data)		
Antisense fragment	**Description**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
asB57	Low energy loop	Significant	Decrease	-62.77153	Significant	Decrease	-24.66437	Significant	Decrease	-41.90937
asB407	High energy stem	Significant	Decrease	-48.30012	Significant	Decrease	-29.79405	Significant	Decrease	-38.16878
asB539	High energy stem	Significant	Decrease	-57.33846	Significant	Decrease	-72.75598	Significant	Decrease	-65.77894
asB615	Low energy stem	Significant	Decrease	-15.50988	Significant	Decrease	-49.31215	Significant	Decrease	-34.01529
asB701	3' end	Significant	Decrease	-65.94318	Significant	Decrease	-75.21886	Significant	Decrease	-71.02125

*****Statistical analysis was performed using N-way analysis of variance (anovan) in comparison with control pART15-C1.

Table B. 10. Statistical analysis of Cell-free-system with MicC scaffold

			48 hours			72 hours			96 hours		
Antisense fragment	**Description**	**Replicates**	Test	Effect	Percentage	Test	Effect	Percentage	Test	Effect	Percentage
MicC-asRBS	anti-RBS	23	Significant	Increase	35.23607	Significant	Increase	23.19246	Significant	Increase	17.50201
MicC-asB1	anti B1-B20	24	Not Significant	Decrease	-0.23998	Significant	Decrease	-34.6396	Significant	Decrease	-42.2982
MiCC-asB606	Low energy stem	14	Significant	Increase	29.46057	Not Significant	Decrease	-0.04392	Not Significant	Increase	3.704833

*****Statistical analysis was performed using N-way analysis of variance (anovan) in comparison with control pART15-C1

Table B. 11. Pairwise comparison test of pART15 plasmids for cells

		6 hrs	18 hrs	29 hrs
pART15-C1	asCTRL	0.999871	0.999989	0.987148
pART15-C1	asRBS	0.999999	0.962367	0.999779
pART15-C1	asB1	0.99947	0.957866	0.934912
pART15-C1	asB30	1	0.999957	0.999628
pART15-C1	asB180	0.984754	0.880324	0.999991
pART15-C1	asB430	0.901272	0.994289	5.04E-07
pART15-C1	asB606	0.985762	0.000243	0.00752
pART15-C1	asB693	0.695221	0.43938	1
asCTRL	asRBS	0.995919	0.793383	0.789251
asCTRL	asB1	1	0.780359	0.296221
asCTRL	asB30	0.999542	1	0.999996
asCTRL	asB180	0.99997	0.616666	0.999779
asCTRL	asB430	0.997054	0.999963	1.27E-07
asCTRL	asB606	0.999975	2.75E-05	0.194963
asCTRL	asB693	0.961413	0.185912	0.991899
asRBS	asB1	0.990477	1	0.999263
asRBS	asB30	1	0.745343	0.950333
asRBS	asB180	0.92058	1	0.987162
asRBS	asB430	0.736048	0.442607	1.37E-05
asRBS	asB606	0.924007	0.034954	0.000548
asRBS	asB693	0.468466	0.994079	0.999499
asB1	asB30	0.998493	0.731098	0.560365
asB1	asB180	0.999997	1	0.727468
asB1	asB430	0.99898	0.427198	0.000502
asB1	asB606	0.999997	0.037523	1.52E-05
asB1	asB693	0.979056	0.995032	0.913984
asB30	asB180	0.972602	0.559518	1
asB30	asB430	0.859811	0.999991	1.32E-07
asB30	asB606	0.97419	1.82E-05	0.070849
asB30	asB693	0.627113	0.154009	0.999842
asB180	asB430	0.999998	0.273062	1.52E-07
asB180	asB606	1	0.078473	0.034195
asB180	asB693	0.999063	0.999549	0.999998
asB430	asB606	0.999997	1.80E-06	1.27E-07
asB430	asB693	0.999997	0.046753	3.79E-07
asB606	asB693	0.998953	0.379766	0.009628

		6 hrs	18 hrs	29 hrs
pART15-C1	asCTRL	0.633809	0.996906	1
pART15-C1	asB57	0.98974	1	0.980949
pART15-C1	asB407	0.999787	0.717253	1
pART15-C1	asB539	0.999984	0.966707	1
pART15-C1	asB615	0.999093	0.999966	1
pART15-C1	asB701	0.965147	0.999992	0.999997
asCTRL	asB57	0.983804	0.99511	0.960066
asCTRL	asB407	0.894361	0.98107	1
asCTRL	asB539	0.827737	0.651625	0.999996
asCTRL	asB615	0.932652	0.964738	1
asCTRL	asB701	0.99639	0.975386	0.999946
asB57	asB407	0.999938	0.682637	0.973062
asB57	asB539	0.999458	0.975134	0.99157
asB57	asB615	0.999994	0.999989	0.981867
asB57	asB701	0.999999	0.999998	0.996898
asB407	asB539	1	0.121442	1
asB407	asB615	1	0.474807	1
asB407	asB701	0.99899	0.520406	0.999988
asB539	asB615	0.999996	0.997199	1
asB539	asB701	0.995764	0.995038	1
asB615	asB701	0.999755	1	0.999998

*Pairwise comparisons was performed using Tukey's honestly significant difference (HSD) method, with 99% confidence interval.

*The mean responses that are significantly different are highlighted in green.

Table B. 12. Multiple comparison test of pART15 plasmids for cell-free-system

		48 hrs	72 hrs	96 hrs
pART15-C1	asCTRL	0.966494	1	0.999999
pART15-C1	asRBS	0.970981	1	0.985356
pART15-C1	asB1	0.05156	0.425285	0.919986
pART15-C1	asB30	1	0.998755	0.999771
pART15-C1	asB180	0.912633	0.969894	0.930604
pART15-C1	asB430	1.33E-01	0.627066	0.975808
pART15-C1	asB606	1.27E-07	1.27E-07	1.27E-07
pART15-C1	asB693	0.893149	0.99789	0.999993
asCTRL	asRBS	1	1	0.91118
asCTRL	asB1	0.657794	0.354398	0.987642
asCTRL	asB30	0.99511	0.99962	0.992585
asCTRL	asB180	1	0.948198	0.990198
asCTRL	asB430	2.11E-03	7.01E-01	0.879011
asCTRL	asB606	1.27E-07	1.27E-07	1.27E-07
asCTRL	asB693	1	0.994715	1
asRBS	asB1	0.639118	0.328168	0.258116
asRBS	asB30	0.996067	0.999776	0.999987
asRBS	asB180	1	0.937211	0.27654
asRBS	asB430	2.36E-03	0.728926	1
asRBS	asB606	1.27E-07	1.27E-07	1.27E-07
asRBS	asB693	1	0.992704	0.884579
asB1	asB30	0.117733	0.073083	0.553179
asB1	asB180	0.791664	0.990591	1
asB1	asB430	2.33E-07	1.08E-03	0.216522
asB1	asB606	1.27E-07	1.27E-07	1.27E-07
asB1	asB693	0.820787	0.926898	0.992401
asB30	asB180	0.979363	0.595269	0.577607
asB30	asB430	0.05947	0.976553	0.999947
asB30	asB606	1.27E-07	1.27E-07	1.27E-07
asB30	asB693	0.972112	0.832934	0.987912
asB180	asB430	8.71E-04	4.80E-02	0.23303
asB180	asB606	1.27E-07	1.27E-07	1.27E-07
asB180	asB693	1	0.999998	0.994117
asB430	asB606	0.00014	1.27E-07	1.27E-07
asB430	asB693	6.91E-04	0.135032	0.847169
asB606	asB693	1.27E-07	1.27E-07	1.27E-07

		48 hrs	72 hrs	96 hrs
pART15-C1	asCTRL	0.003739	3.71E-08	3.71E-08
pART15-C1	asB57	3.71E-08	3.71E-08	3.71E-08
pART15-C1	asB407	3.71E-08	3.71E-08	3.71E-08
pART15-C1	asB539	3.71E-08	3.71E-08	3.71E-08
pART15-C1	asB615	3.71E-08	3.71E-08	4.07E-08
pART15-C1	asB701	3.71E-08	3.71E-08	3.71E-08
asCTRL	asB57	8.99E-06	0.204903	0.001002
asCTRL	asB407	3.72E-08	0.002906	4.38E-05
asCTRL	asB539	3.71E-08	0.996857	0.999861
asCTRL	asB615	5.11E-06	0.996394	7.98E-07
asCTRL	asB701	3.74E-08	0.543476	0.845843
asB57	asB407	0.496639	0.78828	0.992595
asB57	asB539	7.75E-05	0.038975	0.000158
asB57	asB615	1	0.548141	0.757638
asB57	asB701	0.615756	0.000408	1.06E-06
asB407	asB539	0.086952	0.000167	4.83E-06
asB407	asB615	0.570965	0.021468	0.987069
asB407	asB701	0.999998	3.06E-07	5.22E-08
asB539	asB615	0.00013	0.881253	9.17E-08
asB539	asB701	0.053897	0.883684	0.957993
asB615	asB701	0.688068	0.173747	3.71E-08

*****Statistical analysis was performed using multiple comparison test in Matlab (multcompare).

*The mean responses that are significantly different are highlighted in green.

References

1. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995;177(14):4121–30.
2. Litcofsky KD, Afeyan RB, Krom RJ, Khalil AS, Collins JJ. Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. Nat Methods. 2012;9(11):1077–80.
3. Collins J, Zhang T, Huston S, Sun F, Zhang Y-HP, Fu J. A Hidden Transhydrogen Activity of a FMN-Bound Diaphorase under Anaerobic Conditions. PLOS ONE. 2016;11(5):e0154865.

Supplementary Appendix C: Characterizing *Synechocystis* sp. PCC 6803 with Rametrix™: Glucose, illumination, and nitrogen deprivation-induced phenotypes

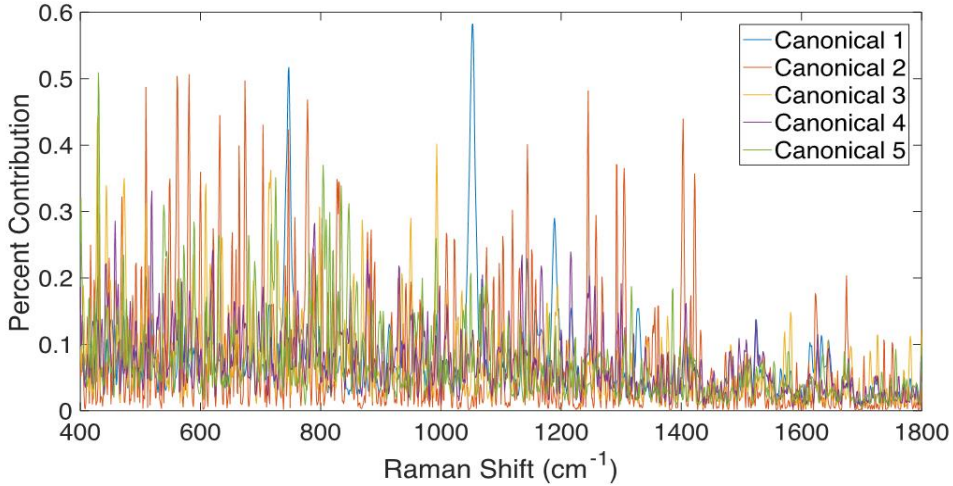


Figure C. 1. Raman shift contributions to DAPC clustering in Figure 4.2B, separation by the presence of glucose in BG-11 medium.

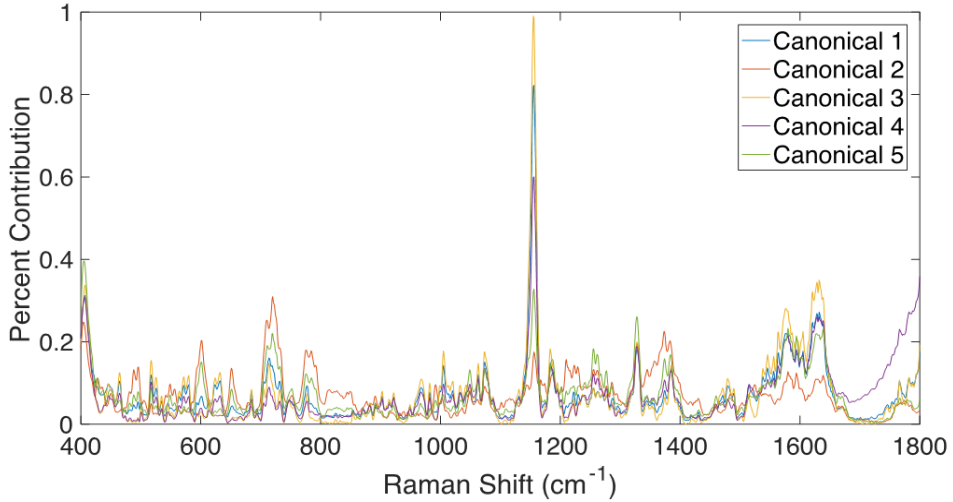


Figure C. 2. Raman shift contributions to DAPC clustering in Figure 4.3B, separation by differences in illumination.

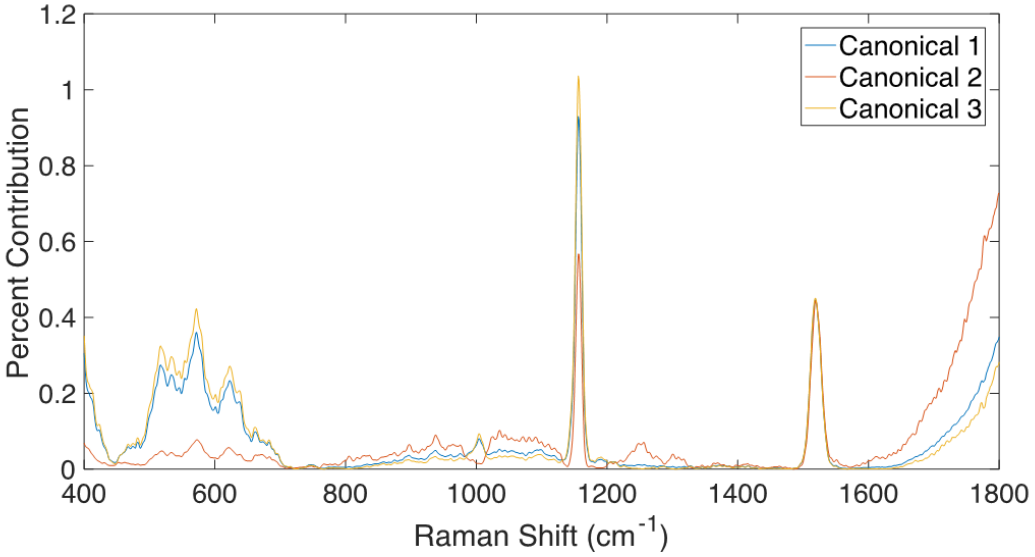


Figure C. 3. Raman shift contributions to DAPC clustering in Figure 4.4B, separation by nitrate limitations.

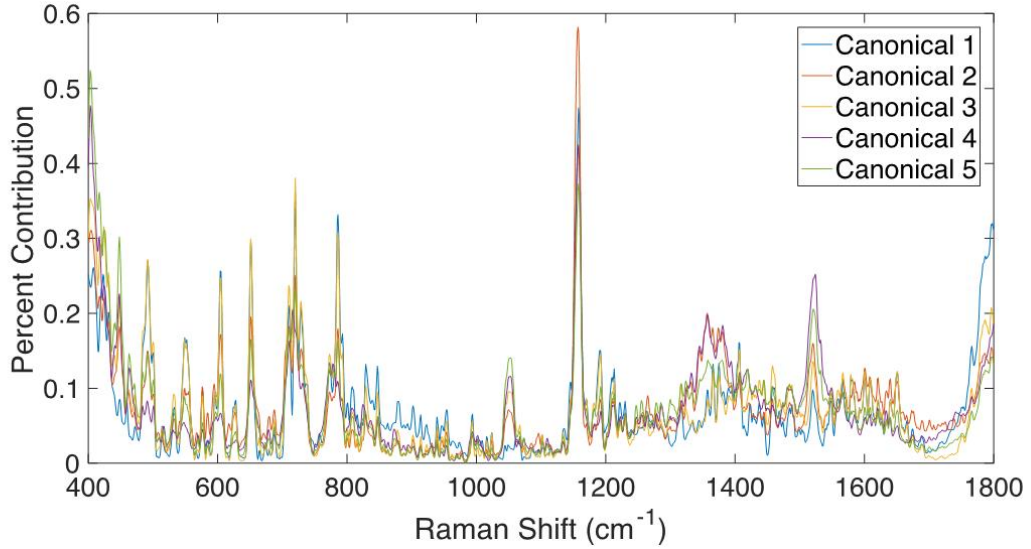


Figure C. 4. Raman shift contributions to DAPC clustering in Figure 4.5B, separation by differences in all growth conditions.

Table C. 1. PCA contributions

Glucose Study (Figure 4.2C)		
Major Band	**Designation**	**Reference (cited in Ref (1))**
430	Cholesterol, cholesterol ester	(2)
575	Phosphatidylinositol	(2)
625	Unknown; Possibly glycerol	(2)
690	Unknown	
750	Protein assignment	(3–5)
992	Unknown	
1102	Phenylalanine; Protein assignment	(6)
1156	Protein assignment; Carotenoids	(7,8)
1520	Carotenoids	(9,10)
Illumination Study (Figure 4.3C)		
Major Band	**Designation**	**Reference (cited in Ref 1)**
430	Cholesterol, cholesterol ester	(2)
516	Phosphatidylinositol	(2)
574	Phosphatidylinositol	(2)
625	Unknown	
685	Unknown	
1156	Protein assignment; Carotenoids	(7,8)
1520	Carotenoids	(9,10)
Nitrate Limitation Study (Figure 4.4C)		
Major Band	**Designation**	**Reference (cited in Ref 1)**
518	Phosphatidylinositol	(6)
534	Unknown	

571	Tryptophan; cytosine, guanine	(10)
623	Protein	(4,10,11)
1156	Protein assignment; Carotenoids	(7,8)
1359	Tryptophan	(5)
1518	Porphyrin; Carotenoids	(3)
1520	Carotene	(9)
All Phenotypes Study (Figure 4.5C)		
Major Band	**Designation**	**Reference (cited in Ref 1)**
516	Phosphatidylinositol	(2)
574	Phosphatidylinositol	(2)
624	Unknown	
630	Glycerol	(2)
667	DNA/RNA	(11)
675	DNA	(11)
685	Unknown	
938	C-C stretch backbone	(10)
991	Unknown	
1101	Proteins or fatty acids	(6,12)
1114	Unknown	
1156	Protein assignment; Carotenoids	(7,8)
1158	Protein	(11)
1265	Protein	(13)
1333	Protein; Nucleic acids	(3,14)
1345	Unknown	
1357	Unknown	
1374	Unknown	
1381	Unknown	
1395	Unknown	

1520	Carotenoids	(9,10)
1522	Carotenoids	(10)
1525	Carotenoids	(10)

Table C. 2. DAPC contributions

Glucose Study (Fig. C1)		
Major Band	**Designation**	**Reference (cited in Ref 1)**
430	Cholesterol	(2)
443	Unknown	
509	S-S disulfide stretching	(12)
561	Unknown	
581	Unknown	
632	Methionine or glycerol	(2,12)
664	RNA	(7)
674	Unknown	
704	Cholesterol	(2)
747	DNA/RNA	(7,15)
778	Phosphatidylinositol	(2)
993	C-O Ribose	(11)
1052	Protein	(7)
1144	Unknown	
1245	Amide III	(12)
1292	Cytocine	(14)
1305	Lipids	(10)
1403	Amino acids	(12)

Illumination Study (Fig. C2)		
Major Band	**Designation**	**Reference (cited in Ref 1)**
405	Unknown	
599	Phosphatidylinositol	(2)
710	Unknown	
719	Phospholipids	(2)
776	Phosphatidylinositol	(2)
785	DNA/RNA	(7)
792	Unknown	
1005	Phenylalanine (protein)	(11)
1049	Glycogen	(15)
1063	Lipids	(10)
1074	Lipids	(9)
1155	Proteins, Glycogen, Carotenoids	(4,10,16)
1328	Phospholipids	(9)
1374	DNA/RNA	(6)
1385	CH3 band	(10)
1576	Nucleic acids	(17)
1620	Porphyrin	(13)
1628	Beta-form polypeptide films	(18)
1633	Amide I	(16)
1638	Amide I	(19)
Nitrate Limitation Study (Fig. C3)		
Major Band	**Designation**	**Reference (cited in Ref 1)**
518	Phosphatidylinositol	(6)
534	Unknown	
571	Tryptophan; cytosine, guanine	(10)

623	Protein	(4,10,11)
639	Protein	(7)
1156	Protein assignment; Carotenoids	(7,8)
1518	Porphyrin; Carotenoids	(3)
1520	Carotene	(9)
All Phenotypes Study (Fig. C4)		
Major Band	**Designation**	**Reference (cited in Ref 1)**
447	Phenyl	(20)
492	Unknown	
605	Glycerol	(2)
651	Unknown	
720	DNA	(15)
786	Nucleic acids	(21,22)
1054	Nucleic acids	(16)
1156	Protein assignment; Carotenoids	(7,8)
1357	Nucleic acids	(14)
1524	Carotenoids	(4,8)

Table C. 3. Mean values of Total Spectral Distance (TSD), Total Principal canonical Distance (TPD) and Total Canonical Distance (TCD)

Glucose Study (Reference = BG-11; 40 PCs (99.94%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
BG11	0.142	0.121	0.124	0.118	75.900	24.722
BG11 (5 mM Glucose)	0.150	0.137	0.132	0.135	242.946	44.719
Illumination Study (Reference = Continuous Light; 9 PCs (98.95%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**

Continuous Light	0.122	0.118	0.107	0.114	15.637	10.460
Dark	0.172	0.154	0.155	0.154	53.019	18.777
Light/Dark	0.171	0.096	0.148	0.094	51.402	18.044
Nitrate Limitation Study (Reference = 100%; 3 PCs (83.23%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%	0.2087	0.1201	0.18	0.132972682	7.599177606	6.198677987
10%	0.210	0.094	0.198	0.095	11.646	5.537
20%	0.155	0.135	0.145	0.134	4.718	3.917
50%	0.219	0.152	0.212	0.152	8.290	5.311
75%	0.095	0.081	0.084	0.081	2.207	0.460
95%	0.081	0.069	0.073	0.068	2.926	2.462
100%	0.045	0.052	0.040	0.047	0.020	0.028
Nitrate Limitation Study (Reference = 100%; 4 PCs (92.15%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%					9.473808032	7.60474009
10%					22.170	6.957
20%					9.234	6.174
50%					12.048	7.737
75%					5.055	3.109
95%					5.508	4.995
100%					1.460	1.670
Nitrate Limitation Study (Reference = 100%; 5 PCs (95.58%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%					24.13960598	10.07295551
10%					24.440	5.583
20%					9.875	6.596
50%					13.719	8.543

75%					6.648	2.787
95%					5.883	4.796
100%					1.483	1.666
Nitrate Limitation Study (Reference = 100%; 6 PCs (97.74%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%					264.5835719	42.23825643
10%					30.459	5.943
20%					14.936	5.902
50%					27.298	10.542
75%					30.861	10.168
95%					9.188	4.638
100%					2.971	3.085
Nitrate Limitation Study (Reference = 100%; 7 PCs (98.38%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%					333.9629081	29.96097684
10%					41.552	12.228
20%					21.583	8.264
50%					29.096	12.229
75%					34.377	9.876
95%					11.750	5.902
100%					5.460	5.888
Nitrate Limitation Study (Reference = 100%; 8 PCs (98.80%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%					379.6526228	42.07181949
10%					55.619	22.721
20%					23.258	8.486
50%					33.620	16.830

75%					42.642	15.430
95%					13.368	6.024
100%					8.140	9.050
Nitrate Limitation Study (Reference = 100%; 9 PCs (99.19%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%					662.2821572	62.77970646
10%					58.397	24.074
20%					37.808	14.047
50%					41.089	18.996
75%					92.971	34.650
95%					24.271	9.977
100%					10.502	9.525
Nitrate Limitation Study (Reference = 100%; 10 PCs (99.41%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
5%					664.980791	62.82454032
10%					138.014	36.285
20%					41.802	15.561
50%					51.258	18.936
75%					101.278	36.010
95%					33.321	9.734
100%					10.850	9.868
All Phenotypes Study (Reference = Autotrophic, 8 PCs (99.79%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
Autotrophic	0.018	0.015	0.007	0.005	10.667	8.000
Heterotrophic	0.220	0.116	0.209	0.114	283.131	22.425
Photoautotrophic	0.275	0.208	0.266	0.210	192.109	42.165
Photoheterotrophic	0.337	0.086	0.327	0.088	660.015	72.151

Circadian Rhythm Study (Reference = Autotrophic- 00h, 32 PCs (95%))						
Incubation Time	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
autotrophic - 00h	0.0135	0.0027	0.00	0.002415486	47.575384	11.86614368
autotrophic - 02h	0.3223	0.0094	0.19	0.01003466	403.0254163	26.10595162
autotrophic - 04h	0.1246	0.0099	0.08	0.010324777	252.5189898	32.18651665
autotrophic - 06h	0.0663	0.0053	0.06	0.005019474	243.4495726	27.01559472
autotrophic - 08h	0.0800	0.0056	0.07	0.005416715	253.9657946	26.55647512
autotrophic - 10h	0.0167	0.0029	0.01	0.00233366	74.71590441	15.62123649
autotrophic - 12h	0.1946	0.0080	0.11	0.004781598	877.4240204	43.45062543
autotrophic - 14h	0.0805	0.0062	0.07	0.006767638	191.0140419	21.80076453
autotrophic - 16h	0.1258	0.0132	0.10	0.013689194	289.9684763	31.5691112
autotrophic - 18h	0.0565	0.0095	0.04	0.009746724	190.794724	19.56369613
autotrophic - 20h	0.2476	0.0125	0.24	0.012385247	473.2323463	30.86793238
autotrophic - 22h	0.1748	0.0161	0.14	0.016846251	295.4347274	22.6548513
autotrophic - 24h	0.0369	0.0034	0.01	0.002531542	147.591785	22.08116919
dark autotrophic - 00h	0.0135	0.0027	0.00	0.002415486	47.575384	11.86614368
dark autotrophic - 02h	0.0432	0.0044	0.01	0.002271001	209.9644047	24.23029326
dark autotrophic - 04h	0.0222	0.0042	0.01	0.00438403	89.65537575	21.78130077
dark autotrophic - 06h	0.0384	0.0023	0.02	0.003126053	116.1005494	15.78577056
dark autotrophic - 08h	0.0462	0.0039	0.03	0.003921029	112.8706363	16.40593689
dark autotrophic - 10h	0.0503	0.0024	0.03	0.003870265	145.4945777	11.80596462
dark autotrophic - 12h	0.0381	0.0037	0.02	0.003052838	116.3511313	15.09081365
dark autotrophic - 14h	0.0506	0.0068	0.03	0.007181666	243.6916609	24.75337642
dark autotrophic - 16h	0.1872	0.0116	0.17	0.011891275	492.8847031	29.13725649
dark autotrophic - 18h	0.0380	0.0015	0.02	0.002894577	147.1533002	11.17443674
dark autotrophic - 20h	0.0766	0.0051	0.06	0.005822875	160.0446129	11.58233326
dark autotrophic - 22h	0.1063	0.0049	0.09	0.00624874	161.8590657	15.45156905
dark autotrophic - 24h	0.1102	0.0055	0.09	0.006193981	161.3285846	13.36215565

dark heterotrophic - 00h	0.0265	0.0045	0.01	0.003758448	139.6676073	20.46607052
dark heterotrophic - 02h	0.0290	0.0036	0.01	0.00246918	202.7485922	33.24183082
dark heterotrophic - 04h	0.0295	0.0039	0.02	0.003942036	168.8237508	25.79845765
dark heterotrophic - 06h	0.0213	0.0018	0.01	0.002242661	129.1667795	19.28403605
dark heterotrophic - 08h	0.0442	0.0045	0.03	0.004761201	137.8248585	24.05929207
dark heterotrophic - 10h	0.2929	0.0190	0.26	0.01902287	389.2924598	24.23496546
dark heterotrophic - 12h	0.0429	0.0034	0.03	0.003772676	179.2845999	19.18904711
dark heterotrophic - 14h	0.0282	0.0028	0.01	0.003107979	177.5798673	22.9184634
dark heterotrophic - 16h	0.5818	0.0186	0.56	0.019344337	201.2620944	17.74284173
dark heterotrophic - 18h	0.0549	0.0036	0.04	0.004670591	151.6795451	21.64352722
dark heterotrophic - 20h	0.1553	0.0069	0.15	0.00716113	249.8973582	26.10548678
dark heterotrophic - 22h	0.0721	0.0036	0.05	0.004517682	233.3933103	25.6088299
dark heterotrophic - 24h	0.0452	0.0064	0.03	0.005816656	188.6156604	20.24447394
heterotrophic - 00h	0.0265	0.0045	0.01	0.003758448	139.6676073	20.46607052
heterotrophic - 02h	0.0728	0.0104	0.04	0.010103278	254.6633087	28.15601974
heterotrophic - 04h	0.1257	0.0107	0.11	0.010921809	159.2780898	21.87657878
heterotrophic - 06h	0.8565	0.0216	0.82	0.022806807	217.4271784	21.46930242
heterotrophic - 08h	0.0282	0.0029	0.01	0.003110624	137.5058908	18.60694622
heterotrophic - 10h	0.0439	0.0018	0.02	0.003194187	162.9886428	15.08617139
heterotrophic - 12h	0.1509	0.0079	0.09	0.00521718	770.2068283	48.12157467
heterotrophic - 14h	0.0864	0.0082	0.08	0.008284332	288.4140201	29.40515586
heterotrophic - 16h	0.1677	0.0123	0.16	0.012790431	236.4500369	22.68688328
heterotrophic - 18h	0.0773	0.0100	0.07	0.010502684	259.5813169	26.80400714
heterotrophic - 20h	0.0331	0.0054	0.02	0.005774465	161.0191639	20.32070082
heterotrophic - 22h	0.0454	0.0070	0.03	0.007556826	189.0266663	21.74808893
heterotrophic - 24h	0.1969	0.0120	0.186293251	0.012161165	333.2301309	31.34231638

Table C. 4. ANOVA test based on TPD data

Glucose Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	0.036660526	1	0.036660526	2.280251495	0.131185037
Error'	32.89437035	2046	0.016077405	[]	[]
Total'	32.93103088	2047	[]	[]	[]
Illumination Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	0.142388151	2	0.071194076	4.684500223	0.009577848
Error'	9.118677208	600	0.015197795	[]	[]
Total'	9.261065359	602	[]	[]	[]
Nitrate Limitation Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	0.248382292	6	0.041397049	3.560216339	0.004669223
Error'	0.651149961	56	0.011627678	[]	[]
Total'	0.899532253	62	[]	[]	[]
All Phenotypes Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	0.52029041	3	0.173430137	10.71671044	4.94E-05
Error'	0.517860812	32	0.01618315	[]	[]
Total'	1.038151222	35	[]	[]	[]

Table C. 5. ANOVA test based on TCD data

Glucose Study (40 PCs)					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	14287005.23	1	14287005.23	10943.84564	0
Error'	2671018.368	2046	1305.483073	[]	[]
Total'	16958023.6	2047	[]	[]	[]
Illumination Study (9 PCs)					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	93702.51994	2	46851.25997	152.4899413	2.84E-54
Error'	184344.9852	600	307.2416421	[]	[]
Total'	278047.5052	602	[]	[]	[]
Nitrate Limitation Study (2 PCs)					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	272.930289	6	45.4883815	3.835924704	0.002836752
Error'	664.0769986	56	11.85851783	[]	[]
Total'	937.0072875	62	[]	[]	[]
All Phenotypes Study (8 PCs)					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	2020665.298	3	673555.0994	356.8273897	1.16E-24
Error'	60403.89219	32	1887.621631	[]	[]
Total'	2081069.191	35	[]	[]	[]

Table C. 6. Pairwise Comparison based on TPD data

Glucose Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
BG-11	BG-11 (5 mM Glucose)	-0.019444729	-0.008461831	0.002521066	0.131030625
Illumination Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
Dark	Light/Dark	-0.018683878	0.006744784	0.032173445	0.808246756
Dark	Continuous Light	0.010597357	0.047844869	0.08509238	0.007355676
Light/Dark	Continuous Light	0.004761597	0.041100085	0.077438573	0.021889717
Nitrate Limitation Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
5%	10%	-0.170063392	-0.014617649	0.140828094	0.999949502
5%	20%	-0.11722692	0.038218823	0.193664566	0.988433404
5%	50%	-0.184117402	-0.028671659	0.126774084	0.997560062
5%	75%	-0.056344802	0.099100941	0.254546684	0.45747255
5%	95%	-0.04571264	0.109733103	0.265178846	0.334114954
5%	100%	-0.012616447	0.142829296	0.29827504	0.091397369
10%	20%	-0.102609271	0.052836472	0.208282215	0.942410892
10%	50%	-0.169499754	-0.01405401	0.141391733	0.999959939
10%	75%	-0.041727154	0.113718589	0.269164333	0.292949723
10%	95%	-0.031094991	0.124350752	0.279796495	0.199274486
10%	100%	0.002001202	0.157446945	0.312892688	0.045221317

20%	50%	-0.222336225	-0.066890482	0.088555261	0.84166019
20%	75%	-0.094563625	0.060882118	0.216327861	0.892114327
20%	95%	-0.083931463	0.07151428	0.226960023	0.795859917
20%	100%	-0.05083527	0.104610473	0.260056217	0.39136323
50%	75%	-0.027673143	0.1277726	0.283218343	0.174243328
50%	95%	-0.017040981	0.138404762	0.293850505	0.111462367
50%	100%	0.016055212	0.171500956	0.326946699	0.021595557
75%	95%	-0.144813581	0.010632162	0.166077906	0.999992314
75%	100%	-0.111717388	0.043728356	0.199174099	0.976906489
95%	100%	-0.12234955	0.033096193	0.188541936	0.994631908
All Phenotypes Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
Heterotrophic	Photoheterotrophic	-0.280157034	-0.117679896	0.044797242	0.223326845
Heterotrophic	Autotrophic	0.039786715	0.202263852	0.36474099	0.010054239
Heterotrophic	Photoautotrophic	-0.219414829	-0.056937691	0.105539447	0.778553807
Photoheterotrophic	Autotrophic	0.15746661	0.319943748	0.482420886	4.28E-05
Photoheterotrophic	Photoautotrophic	-0.101734933	0.060742205	0.223219342	0.743127231
Autotrophic	Photoautotrophic	-0.421678681	-0.259201544	-0.096724406	0.000774189
Heterotrophic	Photoheterotrophic	-0.280157034	-0.117679896	0.044797242	0.223326845

Table C. 7. Pairwise Comparison based on TCD data

Glucose Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
BG-11	BG-11 (5 mM Glucose)	-170.1754647	-167.0458233	-163.916182	1.06E-10
Illumination Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
Dark	Light/Dark	-1.999401613	1.616139018	5.231679648	0.546777093
Dark	Continuous Light	32.085659	37.38164717	42.67763534	9.56E-10
Light/Dark	Continuous Light	30.59876828	35.76550815	40.93224803	9.56E-10
Nitrate Limitation Study (2 PCs)					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
5%	10%	-4.149685014	0.814495245	5.778675503	0.998734295
5%	20%	-2.715951227	2.248229032	7.212409291	0.807409115
5%	50%	-5.813381598	-0.84920134	4.114978919	0.998398721
5%	75%	-0.872547214	4.091633045	9.055813303	0.171816473
5%	95%	-0.697287128	4.266893131	9.23107339	0.137074862
5%	100%	-0.231006392	4.733173866	9.697354125	0.071122768
10%	20%	-3.530446471	1.433733787	6.397914046	0.97365881
10%	50%	-6.627876843	-1.663696585	3.300483674	0.946056827
10%	75%	-1.687042459	3.2771378	8.241318058	0.414785997
10%	95%	-1.511782372	3.452397886	8.416578145	0.351791407
10%	100%	-1.045501637	3.918678622	8.88285888	0.212176469

20%	50%	-8.061610631	-3.097430372	1.866749887	0.483742006
20%	75%	-3.120776246	1.843404012	6.807584271	0.914230162
20%	95%	-2.94551616	2.018664099	6.982844358	0.873837676
20%	100%	-2.479235424	2.484944834	7.449125093	0.72535109
50%	75%	-0.023345874	4.940834384	9.905014643	0.051852602
50%	95%	0.151914212	5.116094471	10.08027473	0.039299547
50%	100%	0.618194948	5.582375206	10.54655546	0.018007587
75%	95%	-4.788920172	0.175260086	5.139440345	0.999999851
75%	100%	-4.322639437	0.641540822	5.605721081	0.999676933
95%	100%	-4.497899523	0.466280736	5.430460994	0.999949842
Nitrate Limitation Study (3 PCs)					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
5%	10%	-9.172647598	-2.750394948	3.671857701	0.844608303
5%	20%	-2.244789206	4.177463444	10.59971609	0.432873263
5%	50%	-5.816495025	0.605757624	7.028010274	0.999948596
5%	75%	-1.076679276	5.345573374	11.76782602	0.16328977
5%	95%	-0.45207014	5.97018251	12.39243516	0.084629029
5%	100%	1.103996823	7.526249472	13.94850212	0.011888565
10%	20%	0.505605742	6.927858392	13.35011104	0.026544746
10%	50%	-3.066100077	3.356152572	9.778405222	0.684039104
10%	75%	1.673715672	8.095968322	14.51822097	0.005277468
10%	95%	2.298324808	8.720577458	15.14283011	0.002069685
10%	100%	3.854391771	10.27664442	16.69889707	0.000170639
20%	50%	-9.993958469	-3.57170582	2.85054683	0.618442918
20%	75%	-5.25414272	1.16810993	7.590362579	0.997742937
20%	95%	-4.629533584	1.792719066	8.214971715	0.977787432
20%	100%	-3.073466621	3.348786028	9.771038678	0.686232809

50%	75%	-1.6824369	4.739815749	11.1620684	0.283072458
50%	95%	-1.057827764	5.364424885	11.78667753	0.160287609
50%	100%	0.498239198	6.920491848	13.3427445	0.026799174
75%	95%	-5.797643514	0.624609136	7.046861786	0.999938449
75%	100%	-4.241576551	2.180676099	8.602928748	0.942686573
95%	100%	-4.866185687	1.556066963	7.978319612	0.989288978
All Phenotypes Study (8 PCs)					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
Heterotrophic	Photoheterotrophic	-432.3740453	-376.8836016	-321.3931579	3.77E-09
Heterotrophic	Autotrophic	216.9741654	272.4646091	327.9550529	3.77E-09
Heterotrophic	Photoautotrophic	35.53165941	91.02210315	146.5125469	0.00054925
Photoheterotrophic	Autotrophic	593.857767	649.3482107	704.8386545	3.77E-09
Photoheterotrophic	Photoautotrophic	412.415261	467.9057048	523.3961485	3.77E-09
Autotrophic	Photoautotrophic	-236.9329497	-181.442506	-125.9520622	6.15E-09
Heterotrophic	Photoheterotrophic	-432.3740453	-376.8836016	-321.3931579	3.77E-09

Table C. 8. Nitrate Limitation Study: Regression of TSD, TPD and TCD data

			3 PCs (83.23%)	4 PCs (92.15%)	5 PCs (95.58%)	6 PCs (97.74%)	7 PCs (98.38%)	8 PCs (98.80%)	9 PCs (99.19%)	10 PCs (99.41%)
	TSD (mean)	**TPD (mean)**	**TCD (mean)**	**TCD (mean)**	**TCD (mean)**	**TCD (mean)**	**TCD (mean**	**TCD (mean**	**TCD (mean**	**TCD (mean)**
Slope	-0.002	-0.001	-0.082	-0.123	-0.198	-1.294	-1.656	-1.895	-3.101	-3.399
Intercept	0.223	0.206	9.477	15.515	22.363	119.931	152.248	175.587	289.752	321.141
R^2	0.752	0.708	0.656	0.548	0.782	0.310	0.318	0.325	0.280	0.347
R	0.867	0.841	0.810	0.7400	0.884	0.556	0.5643	0.5704	0.5296	0.5889

Table C. 9. Rametrix™ PRO analysis

Study	Factor	Number of PCs	% Variability Explained by PCs	Accuracy	Sensitivity	Selectivity	Scans Used
Glucose	Glucose	6	95.00%	50.00%	50.00%	50.00%	CYB1 - CYB64
Glucose	Glucose	10	99.00%	89.06%	87.50%	90.63%	CYB1 - CYB64
Glucose	Glucose	15	99.50%	90.63%	90.63%	90.63%	CYB1 - CYB64
Glucose	Glucose	25	99.78%	84.38%	81.25%	87.50%	CYB1 - CYB64
Glucose	Glucose	30	99.85%	90.63%	93.75%	87.50%	CYB1 - CYB64
Glucose	Glucose	40	99.94%	95.31%	93.75%	96.88%	CYB1 - CYB64
Glucose	Glucose	50	99.98%	92.19%	90.63%	93.75%	CYB1 - CYB64
Glucose	Glucose	60	100.00%	57.81%	62.50%	53.13%	CYB1 - CYB64
Illumination	Continuous Light	8	98.54%	98.28%	100.00%	98.44%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Continuous Light	9	98.98%	100.00%	100.00%	100.00%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Continuous Light	10	99.19%	93.75%	100.00%	93.10%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Dark	6	96.70%	84.21%	69.23%	78.13%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Dark	7	97.97%	78.95%	73.08%	76.56%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Dark	8	98.54%	79.69%	65.39%	89.47%	CYB1 - CYB64, CYB83 - CYB85

Illumination	Dark	9	98.98%	89.47%	69.23%	89.47%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Dark	10	99.19%	76.56%	65.39%	84.21%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Dark	20	99.68%	65.63%	42.31%	81.58%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Light/Dark	7	97.97%	76.56%	78.13%	75.00%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Light/Dark	8	98.54%	71.88%	81.25%	62.50%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Light/Dark	9	98.98%	78.13%	81.25%	75.00%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Light/Dark	10	99.19%	78.13%	84.38%	71.88%	CYB1 - CYB64, CYB83 - CYB85
Illumination	Light/Dark	11	99.28%	76.56%	81.25%	71.88%	CYB1 - CYB64, CYB83 - CYB85
Nitrate Limitation	5%	2	71.68%	80.95%	33.33%	88.89%	CYB65 - CYB85
Nitrate Limitation	5%	3	83.23%	76.19%	66.67%	77.78%	CYB65 - CYB85
Nitrate Limitation	5%	4	92.15%	85.71%	66.67%	88.89%	CYB65 - CYB85
Nitrate Limitation	5%	5	95.58%	71.43%	66.67%	72.22%	CYB65 - CYB85
Nitrate Limitation	5%	6	97.74%	90.48%	66.67%	94.44%	CYB65 - CYB85

Nitrate Limitation	5%	7	98.38%	85.71%	33.33%	94.44%	CYB65 - CYB85
Nitrate Limitation	5%	8	98.81%	80.95%	33.33%	88.89%	CYB65 - CYB85
Nitrate Limitation	5%	10	99.41%	47.62%	33.33%	50.00%	CYB65 - CYB85
Nitrate Limitation	10%	2	71.68%	95.24%	66.67%	100.00%	CYB65 - CYB85
Nitrate Limitation	10%	3	83.23%	95.24%	66.67%	100.00%	CYB65 - CYB85
Nitrate Limitation	10%	4	92.15%	85.71%	100.00%	83.33%	CYB65 - CYB85
Nitrate Limitation	10%	5	95.58%	80.95%	66.67%	83.33%	CYB65 - CYB85
Nitrate Limitation	10%	6	97.74%	71.43%	66.67%	72.22%	CYB65 - CYB85
Nitrate Limitation	10%	7	98.38%	61.90%	66.67%	61.11%	CYB65 - CYB85
Nitrate Limitation	10%	10	99.41%	23.81%	100.00%	11.11%	CYB65 - CYB85
Nitrate Limitation	20%	2	71.68%	76.19%	0.00%	88.89%	CYB65 - CYB85
Nitrate Limitation	20%	3	83.23%	61.90%	0.00%	72.22%	CYB65 - CYB85
Nitrate Limitation	20%	4	92.15%	66.67%	33.33%	72.22%	CYB65 - CYB85

Nitrate Limitation	20%	5	95.58%	52.38%	66.67%	50.00%	CYB65 - CYB85
Nitrate Limitation	20%	6	97.74%	23.81%	33.33%	22.22%	CYB65 - CYB85
Nitrate Limitation	50%	2	71.68%	85.71%	33.33%	94.44%	CYB65 - CYB85
Nitrate Limitation	50%	3	83.23%	80.95%	33.33%	88.89%	CYB65 - CYB85
Nitrate Limitation	50%	4	92.15%	80.95%	66.67%	83.33%	CYB65 - CYB85
Nitrate Limitation	50%	5	95.58%	61.90%	100.00%	55.56%	CYB65 - CYB85
Nitrate Limitation	50%	6	97.74%	52.38%	100.00%	44.44%	CYB65 - CYB85
Nitrate Limitation	75%	2	71.68%	71.43%	0.00%	83.33%	CYB65 - CYB85
Nitrate Limitation	75%	3	83.23%	95.24%	66.67%	100.00%	CYB65 - CYB85
Nitrate Limitation	75%	4	92.15%	66.67%	0.00%	77.78%	CYB65 - CYB85
Nitrate Limitation	75%	5	95.58%	42.86%	100.00%	33.33%	CYB65 - CYB85
Nitrate Limitation	75%	6	97.74%	33.33%	100.00%	22.22%	CYB65 - CYB85
Nitrate Limitation	95%	2	71.68%	80.95%	0.00%	94.44%	CYB65 - CYB85

Nitrate Limitation	95%	3	83.23%	52.38%	33.33%	79.56%	CYB65 - CYB85
Nitrate Limitation	95%	4	92.15%	38.10%	66.67%	33.33%	CYB65 - CYB85
Nitrate Limitation	95%	5	95.58%	47.62%	66.67%	44.44%	CYB65 - CYB85
Nitrate Limitation	95%	6	97.74%	38.10%	66.67%	33.33%	CYB65 - CYB85
Nitrate Limitation	95%	10	99.41%	19.05%	100.00%	5.56%	CYB65 - CYB85
All Phenotypes	Autotrophic	5	92.15%	71.77%	12.50%	85.51%	CYB1 - CYB85
All Phenotypes	Autotrophic	6	95.01%	69.41%	12.50%	82.61%	CYB1 - CYB85
All Phenotypes	Autotrophic	7	96.69%	74.12%	25.00%	85.51%	CYB1 - CYB85
All Phenotypes	Autotrophic	8	97.80%	78.82%	25.00%	91.30%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	2	63.15%	84.71%	0.00%	100.00%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	3	78.59%	69.41%	53.85%	72.22%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	4	87.11%	75.29%	61.54%	77.78%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	5	92.15%	75.29%	53.85%	79.17%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	6	95.01%	75.29%	46.15%	80.56%	CYB1 - CYB85

All Phenotypes	Autotrophic Control	7	96.69%	78.82%	53.85%	83.33%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	8	97.00%	85.88%	38.46%	94.44%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	10	98.85%	87.06%	30.77%	97.22%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	11	99.00%	84.71%	7.69%	98.61%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	40	99.86%	15.29%	100.00%	0.00%	CYB1 - CYB85
All Phenotypes	Autotrophic Control	50	99.92%	15.29%	100.00%	0.00%	CYB1 - CYB85
All Phenotypes	Dark Heterotrophic	4	87.11%	77.65%	0.00%	91.67%	CYB1 - CYB85
All Phenotypes	Dark Heterotrophic	5	92.15%	83.53%	30.77%	93.06%	CYB1 - CYB85
All Phenotypes	Dark Heterotrophic	6	95.01%	84.71%	38.46%	93.06%	CYB1 - CYB85
All Phenotypes	Dark Heterotrophic	7	96.69%	84.71%	30.77%	94.44%	CYB1 - CYB85
All Phenotypes	Dark Heterotrophic	8	97.80%	84.71%	38.46%	93.06%	CYB1 - CYB85
All Phenotypes	Dark Heterotrophic	9	98.40%	84.71%	23.08%	95.83%	CYB1 - CYB85
All Phenotypes	Dark Heterotrophic	10	98.85%	85.88%	15.39%	98.61%	CYB1 - CYB85

All Phenotypes	Heterotrophic	5	92.15%	74.12%	31.25%	84.06%	CYB1 - CYB85
All Phenotypes	Heterotrophic	6	95.01%	77.65%	50.00%	84.06%	CYB1 - CYB85
All Phenotypes	Heterotrophic	7	96.69%	71.77%	43.75%	78.26%	CYB1 - CYB85
All Phenotypes	Heterotrophic	8	97.80%	80.00%	50.00%	86.96%	CYB1 - CYB85
All Phenotypes	Heterotrophic	11	99.05%	80.00%	37.50%	89.86%	CYB1 - CYB85
All Phenotypes	Nitrate Limited	7	96.69%	74.12%	33.33%	85.08%	CYB1 - CYB85
All Phenotypes	Nitrate Limited	8	97.80%	80.00%	38.89%	91.05%	CYB1 - CYB85
All Phenotypes	Nitrate Limited	9	98.40%	91.05%	38.89%	92.54%	CYB1 - CYB85
All Phenotypes	Photoautotrophic	7	96.69%	91.77%	66.67%	93.67%	CYB1 - CYB85
All Phenotypes	Photoautotrophic	8	97.80%	88.24%	50.00%	91.14%	CYB1 - CYB85
All Phenotypes	Photoautotrophic	9	98.40%	87.06%	50.00%	89.87%	CYB1 - CYB85
All Phenotypes	Photoheterotrophic	6	95.01%	60.00%	100.00%	58.54%	CYB1 - CYB85
All Phenotypes	Photoheterotrophic	7	96.69%	32.94%	100.00%	30.49%	CYB1 - CYB85
All Phenotypes	Photoheterotrophic	8	97.80%	75.29%	66.67%	75.61%	CYB1 - CYB85
All Phenotypes	Photoheterotrophic	9	98.40%	43.53%	33.33%	43.90%	CYB1 - CYB85

Table C. 10. Random chance analysis

Study	Random Accuracy	Random Sensitivity	Random Selectivity
Glucose Phenotypes (n = 100,000)			
BG-11	50%	50%	50%
BG-11 (with 5 mM glucose)	50%	50%	50%
Illumination Phenotypes (n = 100,000)			

Continuous Light	63.6%	33.4%	66.7%
Dark	53.2%	33.4%	66.7%
Light/Dark	50.0%	33.4%	66.7%
Nitrate Limitation Phenotypes (n = 100,000)			
5%	75.5%	14.2%	85.7%
10%	75.5%	14.2%	85.7%
20%	75.6%	14.4%	85.7%
50%	75.5%	14.2%	85.7%
75%	75.5%	14.3%	85.7%
95%	75.5%	14.1%	85.7%
All Phenotypes (n = 100,000)			
Autotrophic	72.3%	14.3%	85.7%
Dark control	74.8%	14.3%	85.7%
Dark heterotrophic	74.8%	14.2%	85.7%
Heterotrophic	72.3%	14.2%	85.7%
Nitrate limited	70.6%	14.3%	85.7%
Photoautotrophic	80.7%	14.2%	85.7%
Photoheterotrophic	83.2%	14.2%	85.7%

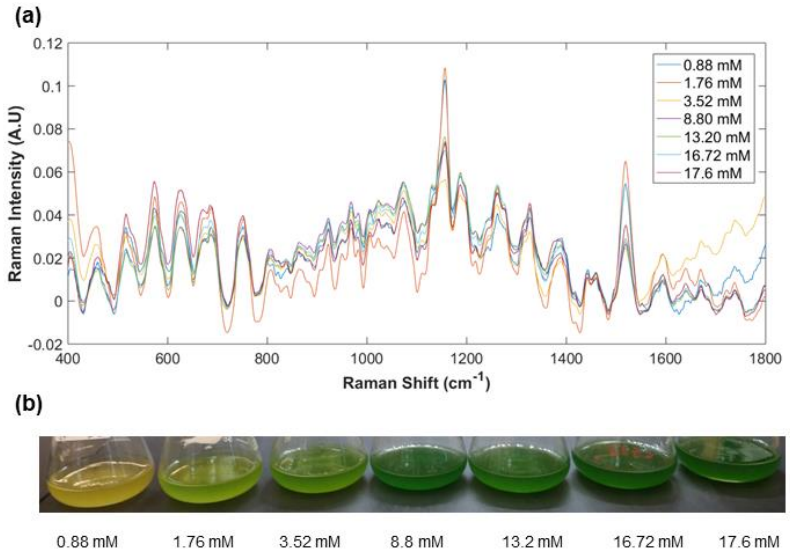


Figure C. 5. *Synechocystis sp*. PCC6803 culture growth under different concentrations of nitrate. (a) Averaged spectra baselined and vector normalized over the range 400-1800 cm^{-1}. (b) Visual comparison of bleaching of cells growing under low concentrations of nitrate, where 17.6 mM is nitrate concentration in unaltered BG-11 medium.

Conversion to percentages: 5% = 0.88 mM; 10% = 1.76 mM; 20% = 3.52 mM; 50% = 8.8 mM; 75% = 13.2 mM; 95% = 16.72 mM; 100% = 17.6 mM.

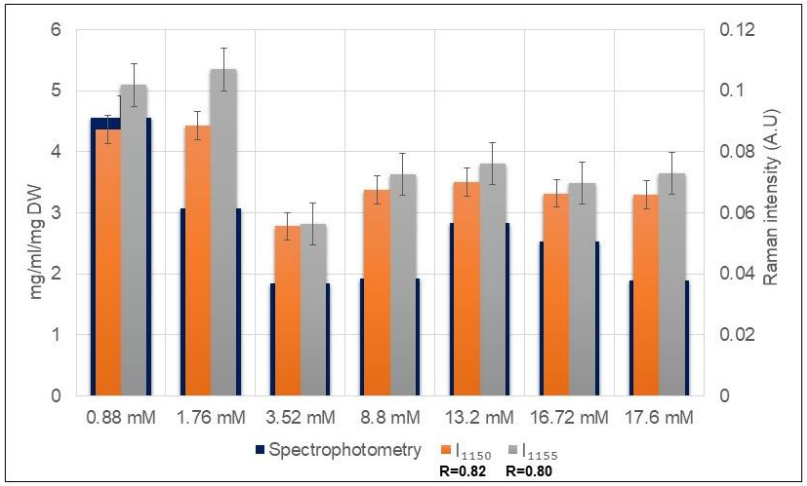


Figure C. 6. Glycogen analysis and Raman spectroscopy in in *Synechocystis sp*. PCC 6803 cells grown under different concentrations of nitrate. Correlation coefficients (R) between Raman bands and glycogen level are represented.

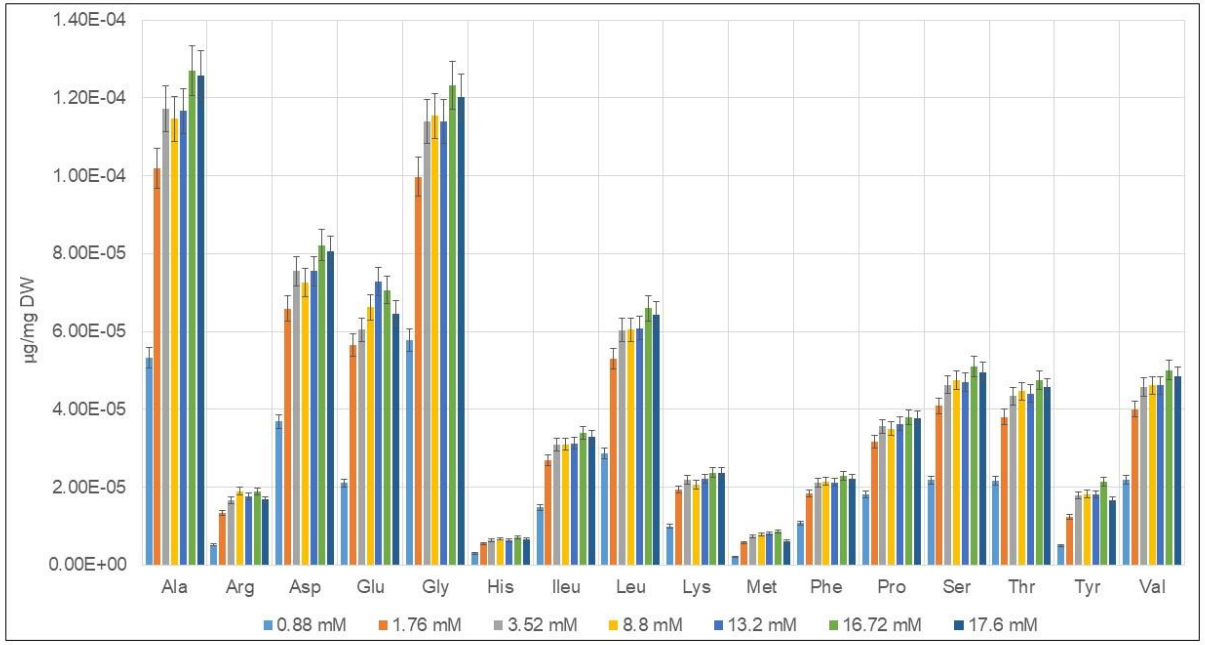
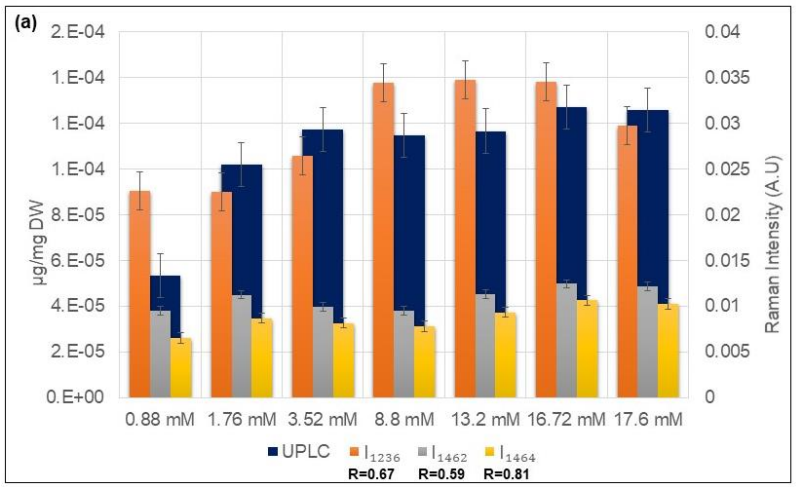
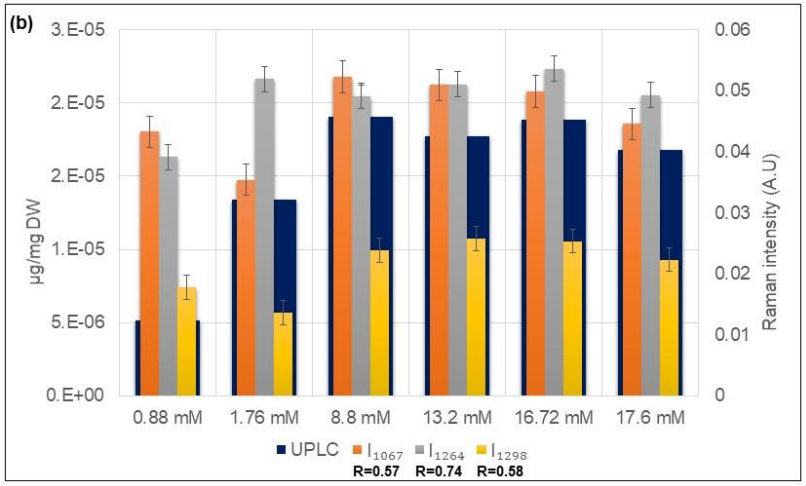
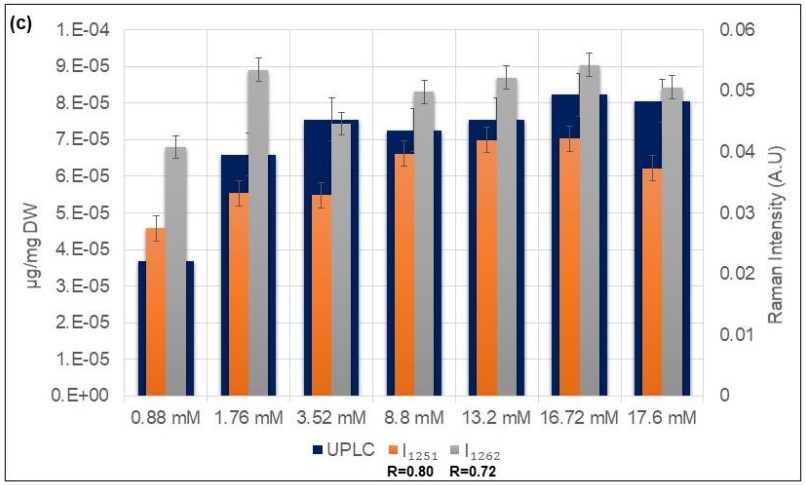
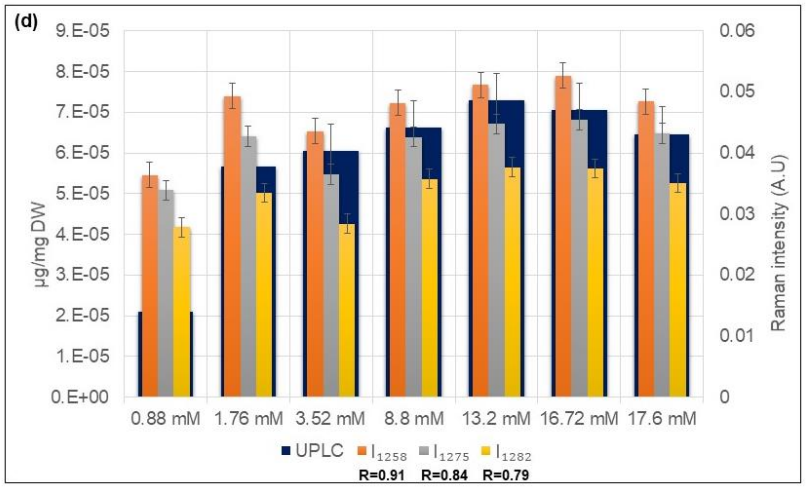


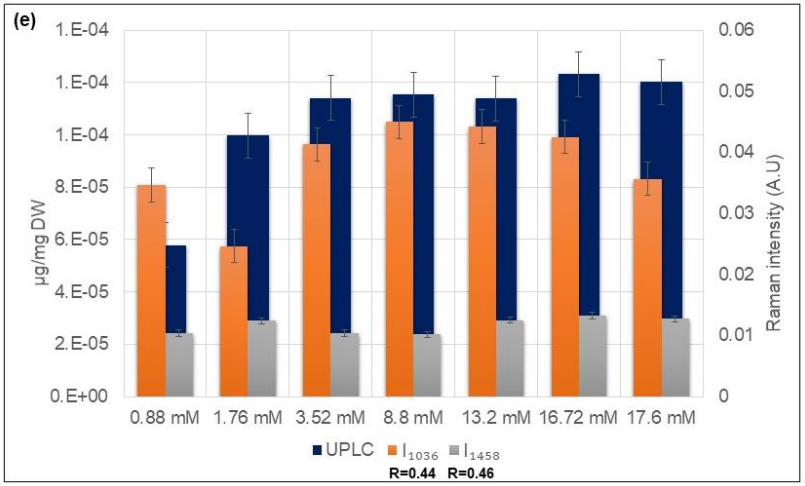
Figure C. 7. Levels of amino acids under nitrogen limitation conditions. Data represent means ± standard deviation (SD) of values from three independent experiments. *Synechocystis sp.* PCC6803 cells under 17.6 mM conditions represents the regular concentration of nitrate in BG-11 medium.

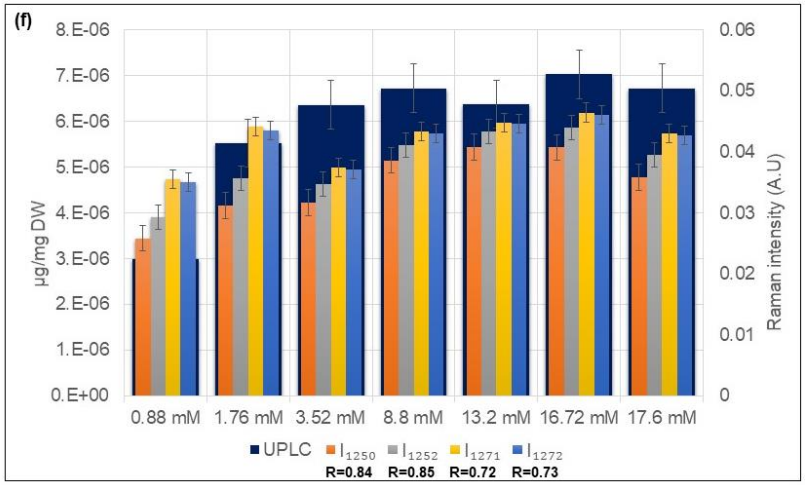

(a)
2.E-04
1.E-04
1.E-04
1.E-04
8.E-05
6.E-05
4.E-05
2.E-05
0.E+00
µg/mg DW
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1236
I1462
I1464
R=0.67
R=0.59
R=0.81

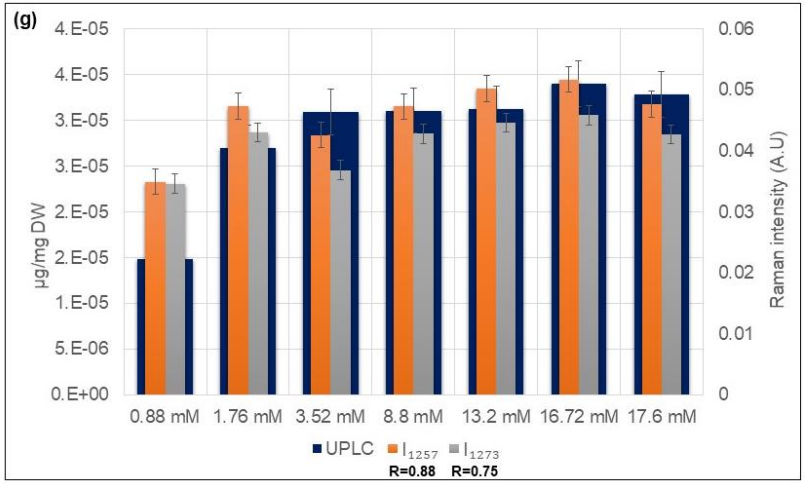

(b)
3.E-05
2.E-05
2.E-05
1.E-05
5.E-06
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1067
I1264
I1298
R=0.57
R=0.74
R=0.58

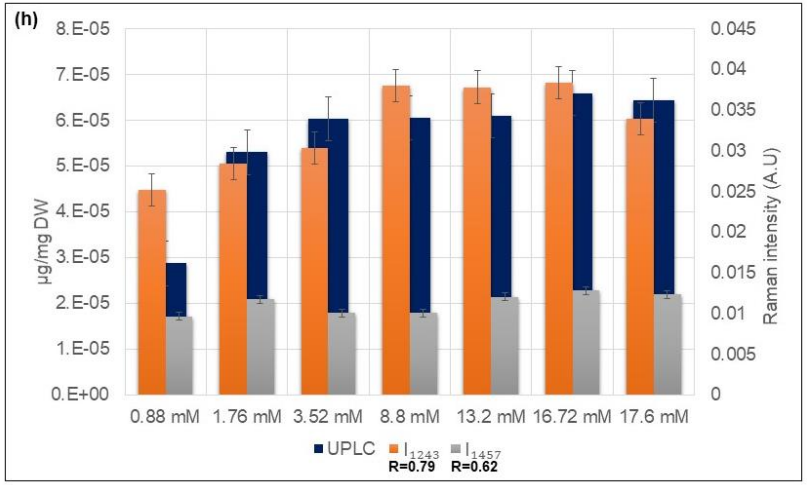

(c)
1.E-04
9.E-05
8.E-05
7.E-05
6.E-05
5.E-05
4.E-05
3.E-05
2.E-05
1.E-05
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman Intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1251
I1262
R=0.80
R=0.72

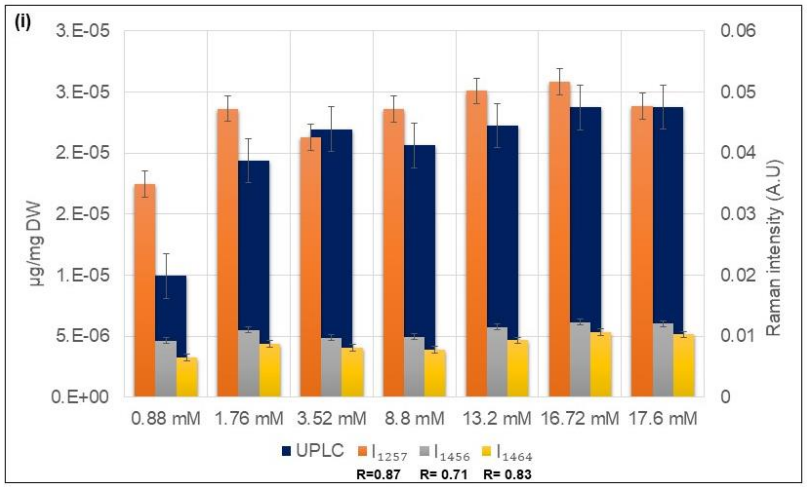

(d)
9.E-05
8.E-05
7.E-05
6.E-05
5.E-05
4.E-05
3.E-05
2.E-05
1.E-05
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1258
I1275
I1282
R=0.91
R=0.84
R=0.79

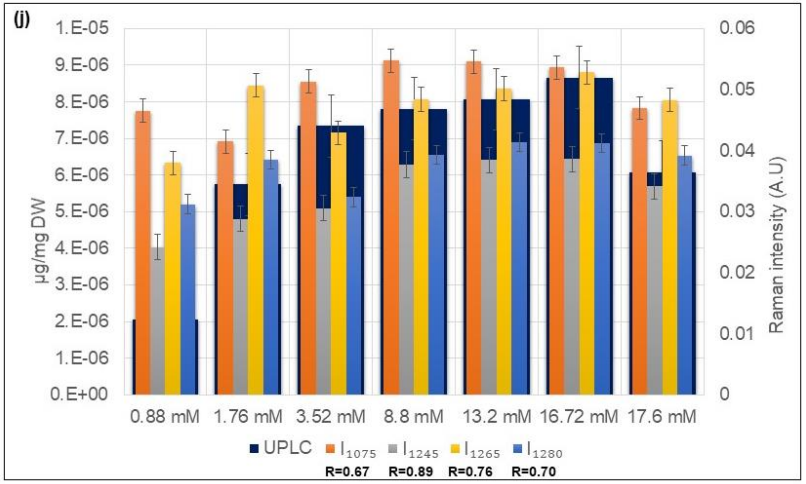

(e)
1.E-04
1.E-04
1.E-04
8.E-05
6.E-05
4.E-05
2.E-05
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1036
I1458
R=0.44
R=0.46

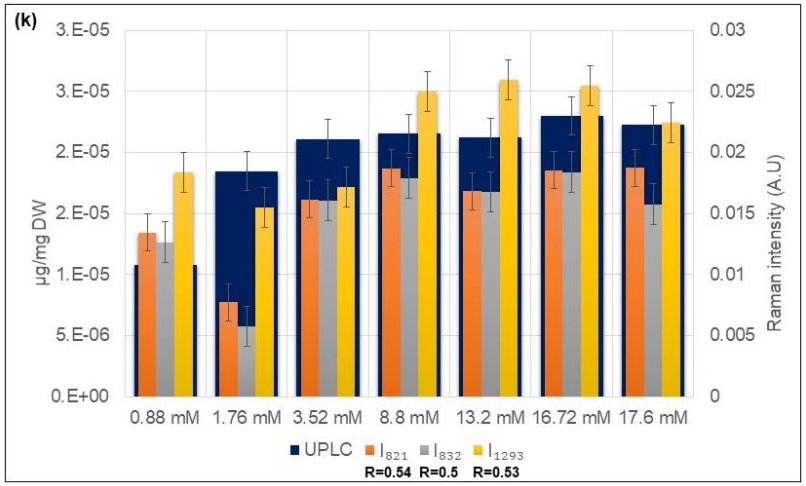

(f)
8.E-06
7.E-06
6.E-06
5.E-06
4.E-06
3.E-06
2.E-06
1.E-06
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1250
I1252
I1271
I1272
R=0.84
R=0.85
R=0.72
R=0.73

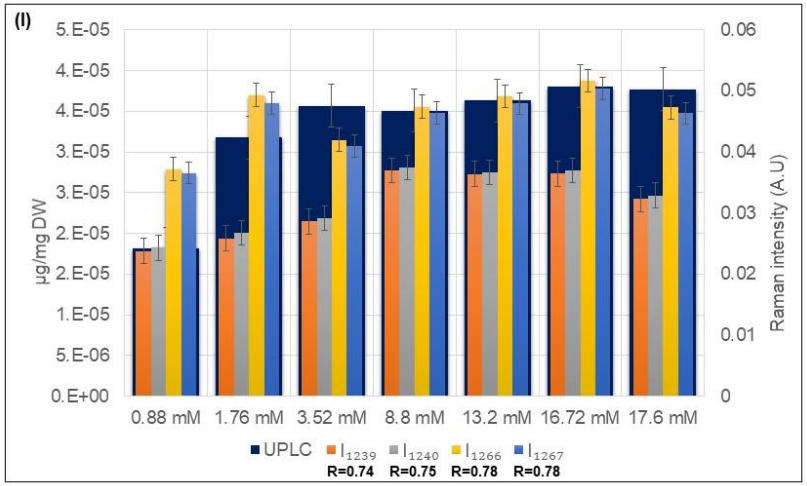

(g)
4.E-05
4.E-05
3.E-05
3.E-05
2.E-05
2.E-05
1.E-05
5.E-06
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1257
I1273
R=0.88
R=0.75


(h)
8.E-05
7.E-05
6.E-05
5.E-05
4.E-05
3.E-05
2.E-05
1.E-05
0.E+00
µg/mg DW
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1243
I1457
R=0.79
R=0.62


(i)
3.E-05
3.E-05
2.E-05
2.E-05
1.E-05
5.E-06
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1257
I1456
I1464
R=0.87
R= 0.71
R= 0.83


(j)
1.E-05
9.E-06
8.E-06
7.E-06
6.E-06
5.E-06
4.E-06
3.E-06
2.E-06
1.E-06
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1075
I1245
I1265
I1280
R=0.67
R=0.89
R=0.76
R=0.70


(k)
3.E-05
3.E-05
2.E-05
2.E-05
1.E-05
5.E-06
0.E+00
µg/mg DW
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I821
I832
I1293
R=0.54
R=0.5
R=0.53


(l)
5.E-05
4.E-05
4.E-05
3.E-05
3.E-05
2.E-05
2.E-05
1.E-05
5.E-06
0.E+00
µg/mg DW
0.06
0.05
0.04
0.03
0.02
0.01
0
Raman intensity (A.U)
0.88 mM
1.76 mM
3.52 mM
8.8 mM
13.2 mM
16.72 mM
17.6 mM
UPLC
I1239
I1240
I1266
I1267
R=0.74
R=0.75
R=0.78
R=0.78

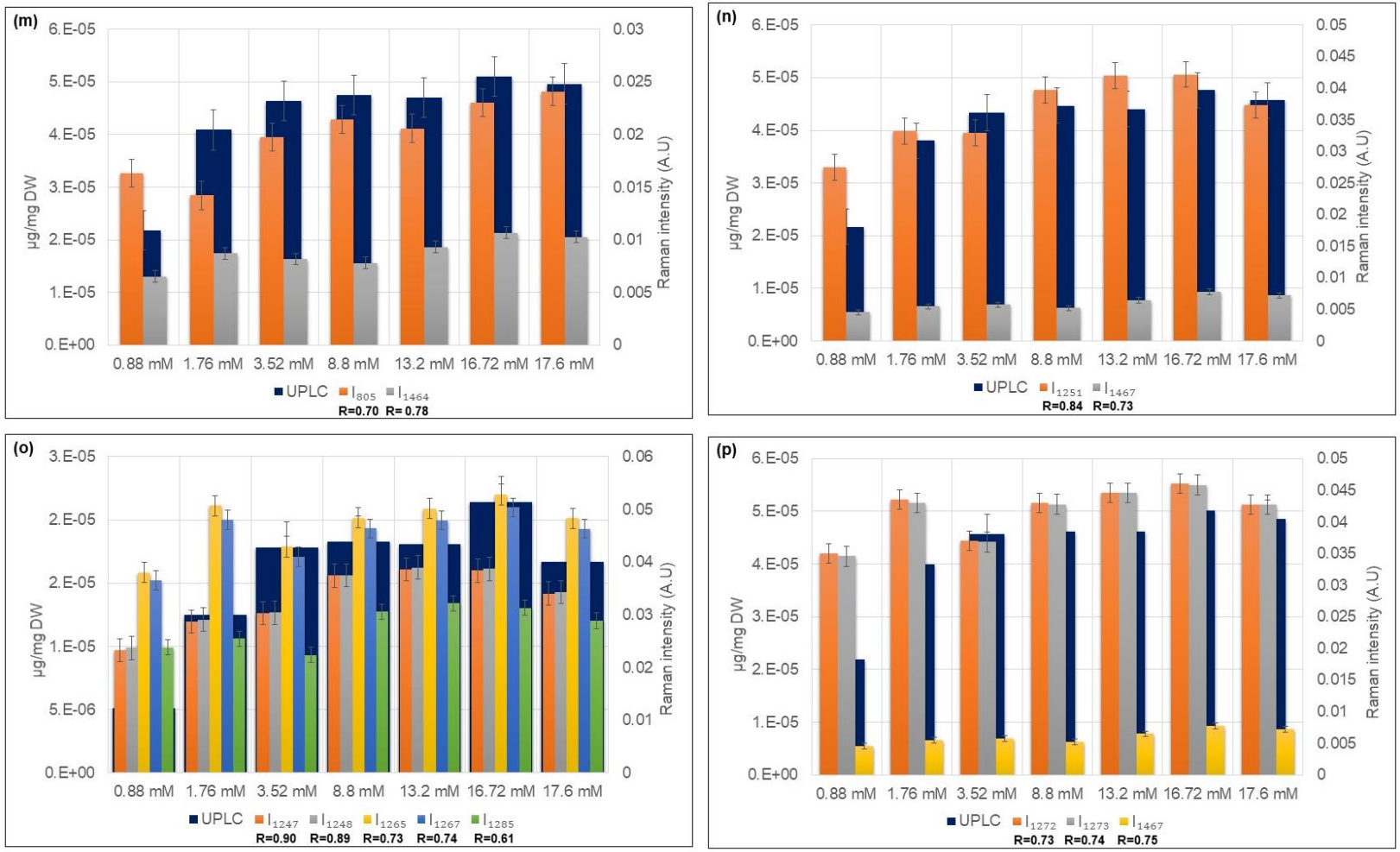


Figure C. 8. Correlation of Raman data with UPLC extraction of amino acids. (a) Alanine, (b) Arginine, (c) Aspartate/Asparagine, (d) Glutamate, (e) Glycine, (f) Histidine, (g) Isoleucine, (h) Leucine, (i) Lysine, (j) Methionine, (k) Phenylalanine, (l) Proline, (m) Serine, (n) Threonine, (o) Tyrosine, (p) Valine. Correlation coefficients (R) for each Raman band are represented.

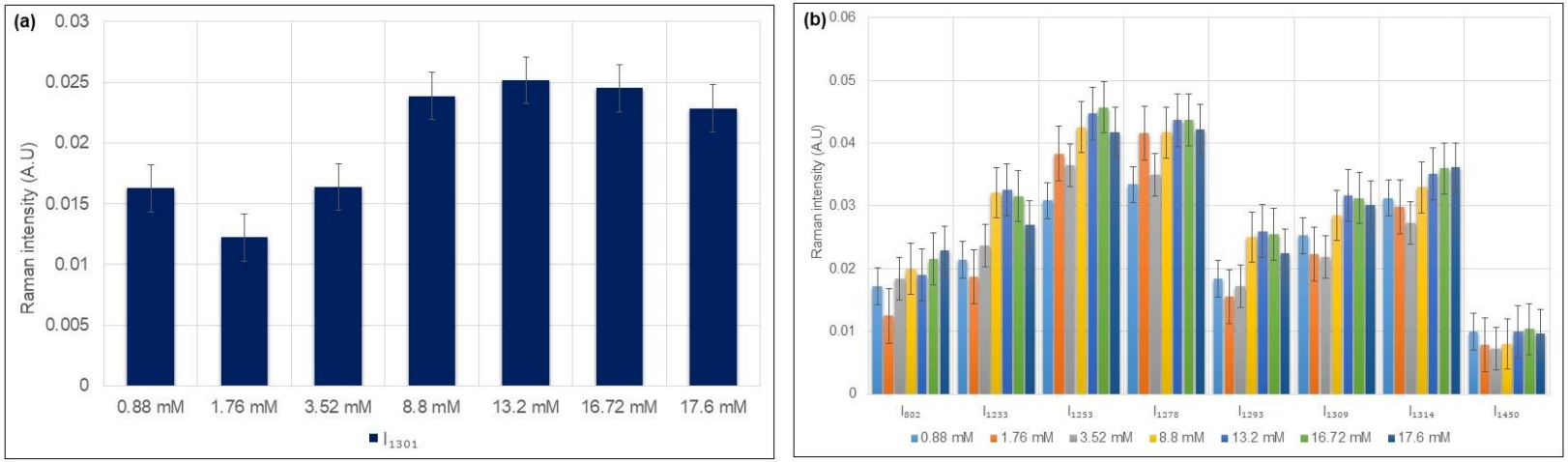


Figure C. 9. Raman bands predictions of amino acids non-resolved with UPLC. (a) Cysteine predictions, (b) Tryptophan predictions.

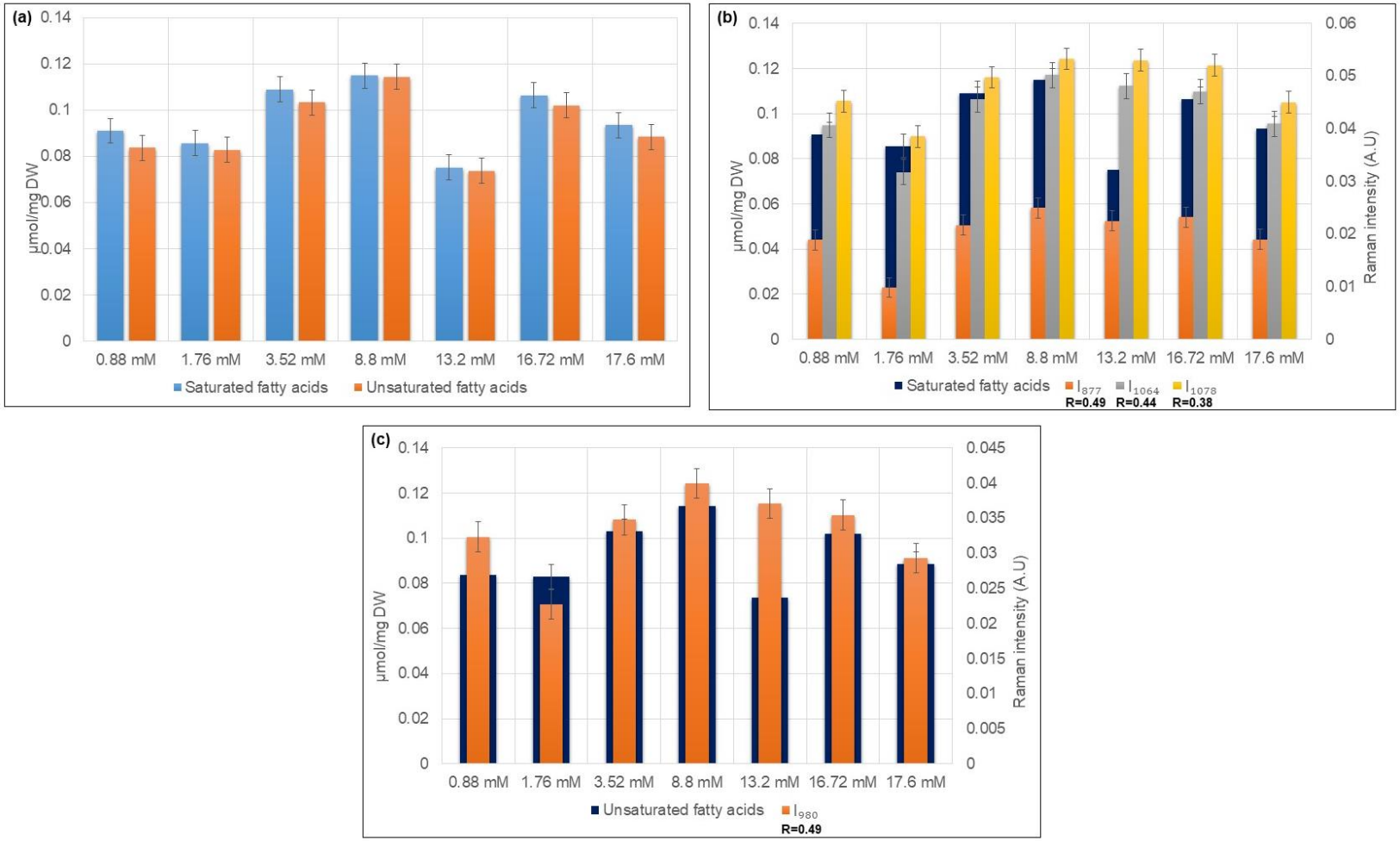


Figure C. 10. Fatty acid levels analysis with GC-FID and Raman spectroscopy in in *Synechocystis sp*. PCC 6803 cells grown under different concentrations of nitrate. (a) Total unsaturated fatty acid and total saturated fatty acids levels, (b) Correlation between Raman spectroscopy and GC-FID data of saturated fatty acids, (c) Correlation between Raman spectroscopy and GC-FID data of unsaturated fatty acids analysis. Correlation coefficients (R) for each Raman band are represented.

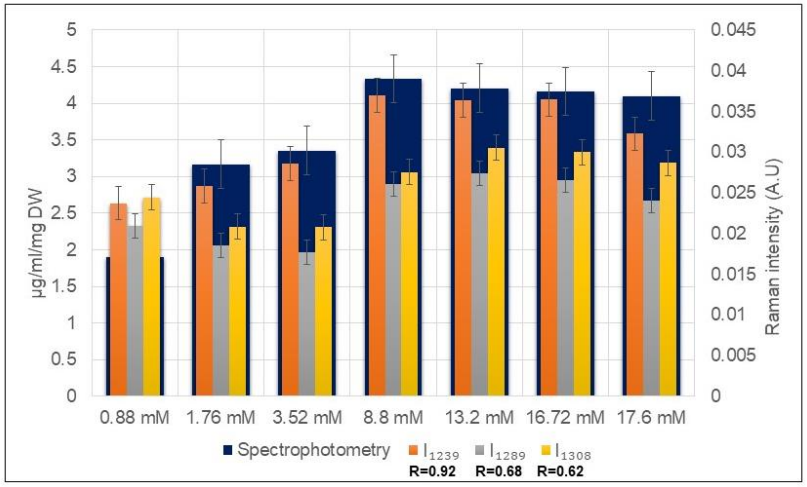


Figure C. 11. Chlorophyll a level and Raman spectroscopy in in *Synechocystis sp*. PCC 6803 cells grown under different concentrations of nitrate. Correlation coefficients (R) between Raman bands and chlorophyll a levels are represented.

Table C. 11. All Raman bands cited and tested for amino acids, chlorophyll a, glycogen and fatty acids.

Biomolecules	**Raman bands (cm^{-1}) and Correlation Coefficient (R)**	**References**
Ala	1236 (R = 0.67), 1462 (R = 0.59), 1464 (R = 0.81)	(23,24)
Arg	1067 (R = 0.57), 1264 (R = 0.74), 1298 (R = 0.58)	(23,24)
Asp/Asn	1251 (R = 0.80), 1262 (R = 0.72)	(24)
Cys*	1301	(24)
Glu/Gln	1258 (R = 0.91), 1275 (R = 0.84), 1282 (R = 0.79)	(23)
Gly	1036 (R = 0.44), 1458 (R = 0.46)	(24)
His	1250 (R = 0.84), 1252 (R = 0.85), 1271 (R = 0.72), 1272 (R = 0.73)	(23,24)
Ile	1257 (R = 0.88), 1273 (R = 0.75)	(24)
Leu	1243 (R = 0.79), 1457 (R = 0.62)	(24)
Lys	1257 (R = 0.87), 1456 (R = 0.71), 1464 (R = 0.83)	(24)

Met	1075 (R = 0.67), 1245 (R = 0.89), 1265 (R = 0.76), 1280 (R = 0.70)	(24)
Phe	821 (R = 0.54), 832 (R = 0.5), 1293 (R = 0.53)	(23,24)
Pro	1239 (R = 0.74), 1240 (R = 0.75), 1266 (R = 0.78), 1267 (R = 0.78)	(23,24)
Ser	805 (R = 0.70), 1464 (R = 0.78)	(24)
Thr	1251 (R = 0.84), 1467 (R = 0.73)	(24)
Trp*	802, 1233, 1253, 1278, 1293, 1309, 1314, 1450	(23,24)
Tyr	1247 (R = 0.90), 1248 (R = 0.89), 1265 (R = 0.73), 1267 (R = 0.74), 1285 (R = 0.61)	(23,24)
Val	1272 (R = 0.73), 1273 (R = 0.74), 1467 (R = 0.75)	(23,24)
Chlorophyll a	1239 (R = 0.92), 1289 (R = 0.68), 1308 (R = 0.62)	(25,26)
Glycogen	1150 (R = 0.82), 1155 (R = 0.80)	(27)
Unsaturated Fatty acids	980 (R = 0.49)	(1)
Saturated Fatty acids	877 (0.49), 1064 (0.44), 1078 (0.38)	(1)

* Values for Cys and Trp could not be obtained by the UPLC method used in this research. Only Raman predictions are provided.
** Possible overlapping bands between different biomolecules assigned Raman band

References

1. Movasaghi Z, Rehman S, Rehman IU. Raman Spectroscopy of Biological Tissues. Appl Spectrosc Rev. 2007;42(5):493–541.
2. Krafft C, Neudert L, Simat T, Salzer R. Near infrared Raman spectra of human brain lipids. Spectrochim Acta Mol Biomol Spectrosc. 2005;61(7):1529–35.
3. Huang Z, McWilliams A, Lui H, McLean DI, Lam S, Zeng H. Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. Int J Cancer. 2003;107(6):1047–52.
4. Stone N, Kendall C, Shepherd N, Crow P, Barr H. Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. J Raman Spectrosc. 2002;33(7):564–73.
5. Cheng WT, Liu MT, Liu HN, Lin SY. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. Microsc Res Tech. 2005;68(2):75–9.
6. Lakshmi RJ, Kartha VB, Murali Krishna C, JG RS, Ullas G, Uma Devi P. Tissue Raman spectroscopy for the study of radiation damage: brain irradiation of mice. Radiat Res. 2002;157(2):175–82.
7. Chan JW, Taylor DS, Zwerdling T, Lane SM, Ihara K, Huser T. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. Biophys J. 2006;90(2):648–56.
8. Mahadevan-Jansen A, Richards-Kortum R. Raman spectroscopy for cancer detection: a review. In 1997. p. 2722–8 vol.6.
9. Malini R, Venkatakrishna K, Kurien J, Pai KM, Rao L, Kartha VB, et al. Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study. Biopolymers. 2006;81(3):179–93.
10. Stone N, Kendall C, Smith J, Crow P, Barr H. Raman spectroscopy for identification of epithelial cancers. Faraday Discuss. 2004;126:141–57; discussion 169-83.
11. Notingher I, Green C, Dyer C, Perkins E, Hopkins N, Lindsay C, et al. Discrimination between ricin and sulphur mustard toxicity in vitro using Raman spectroscopy. J R Soc Interface. 2004;1(1):79–90.
12. Shetty G, Kendall C, Shepherd N, Stone N, Barr H. Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. Br J Cancer. 2006;94(10):1460–4.
13. Katainen E, Elomaa M, Laakkonen UM, Sippola E, Niemela P, Suhonen J, et al. Quantification of the amphetamine content in seized street samples by Raman spectroscopy. J Forensic Sci. 2007;52(1):88–92.
14. Ruiz-Chica AJ, Medina MA, Sánchez-Jiménez F, Ramírez FJ. Characterization by Raman spectroscopy of conformational changes on guanine–cytosine and adenine–thymine oligonucleotides induced by aminooxy analogues of spermidine. J Raman Spectrosc. 2004;35(2):93–100.
15. Binoy J, Abraham JP, Joe IH, Jayakumar VS, Pettit GR, Nielsen OF. NIR-FT Raman and FT-IR spectral studies and ab initio calculations of the anti-cancer drug combretastatin-A4. J Raman Spectrosc. 2004;35(11):939–46.
16. Dukor RK. Vibrational Spectroscopy in the Detection of Cancer. Biomed Appl. 2002;5:3335–3359.
17. Fung MFK, Senterman MK, Mikhael NZ, Lacelle S, Wong PTT. Pressure-tuning fourier transform infrared spectroscopic study of carcinogenesis in human endometrium. Biospectroscopy. 1996;2(3):155–65.
18. Shaw RA, Mantsch HH. Vibrational biospectroscopy: from plants to animals to humans. A

historical perspective. J Mol Struct. 1999;480–481:1–13.
19. Ó Faoláin E, Hunter MB, Byrne JM, Kelehan P, McNamara M, Byrne HJ, et al. A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. Vib Spectrosc. 2005;38(1):121–7.
20. Schulz H, Baranska M. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. Vib Spectrosc. 2007;43(1):13–25.
21. Naumann D. Infrared and NIR Raman spectroscopy in medical microbiology. Proc SPIE. 1998;3257:245–257.
22. Farquharson S, Shende C, Inscore FE, Maksymiuk P, Gift A. Analysis of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy. J Raman Spectrosc. 2005;36(3):208–12.
23. De Gelder J, De Gussem K, Vandenabeele P, Moens L. Reference database of Raman spectra of biological molecules. J Raman Spectrosc. 2007;38(9):1133–47.
24. Zhu G, Zhu X, Fan Q, Wan X. Raman spectra of amino acids and their aqueous solutions. Spectrochim Acta Mol Biomol Spectrosc. 2011;78(3):1187–95.
25. Jehlicka J, Edwards HG, Oren A. Raman spectroscopy of microbial pigments. Appl Env Microbiol. 2014;80(11):3286–95.
26. Wood BR, Heraud P, Stojkovic S, Morrison D, Beardall J, McNaughton D. A portable Raman acoustic levitation spectroscopic system for the identification and environmental monitoring of algal cells. Anal Chem. 2005;77(15):4955–61.
27. Kamemoto LE, Misra AK, Sharma SK, Goodman MT, Luk H, Dykes AC, et al. Near-infrared micro-Raman spectroscopy for in vitro detection of cervical cancer. Appl Spectrosc. 2010;64(3):255–61.

Supplementary Appendix D: Characterizing *Synechocystis sp.* PCC 6803 with Rametrix™: Acetate, NaCl, and $MgSO_4$ deprivation-induced phenotypes

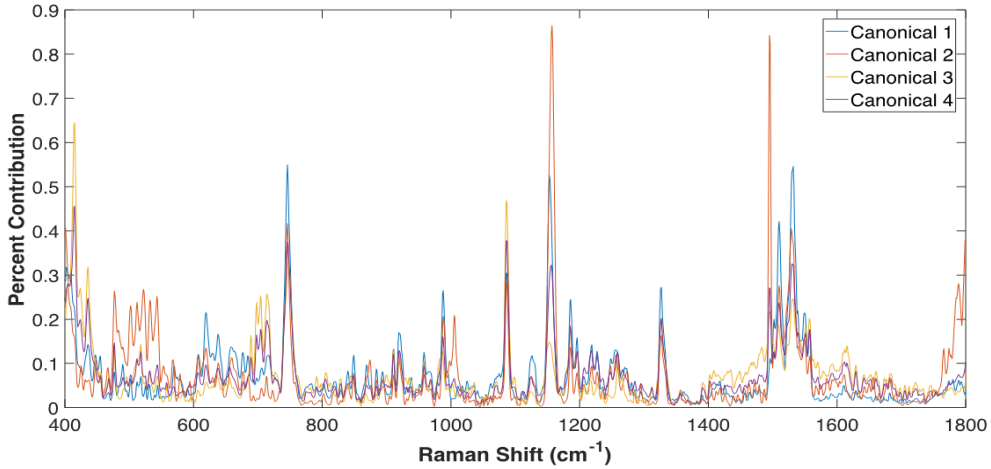


Figure D. 1. Raman shift contributions to DAPC clustering in Figure 5.1B, separation by the presence of acetate in BG-11 medium.

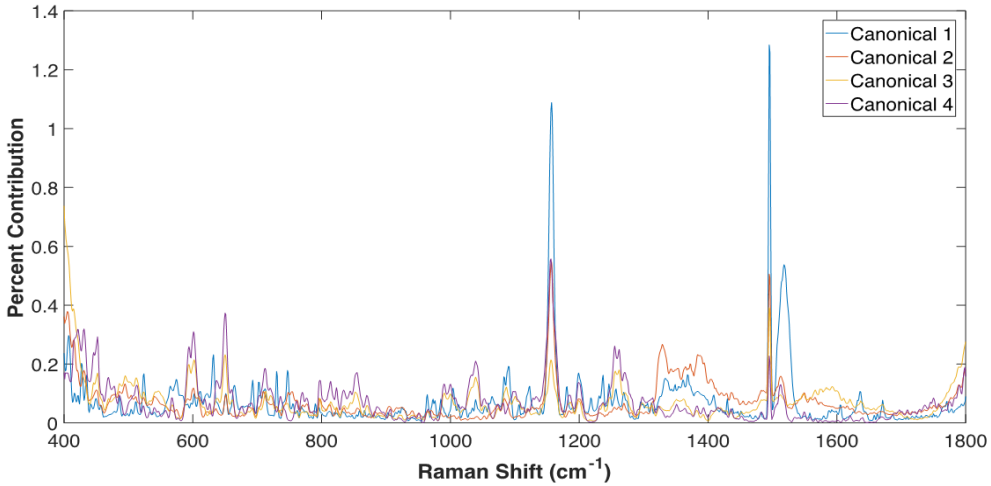


Figure D. 2. Raman shift contributions to DAPC clustering in Figure 5.2B, separation by the presence of NaCl in BG-11 medium.

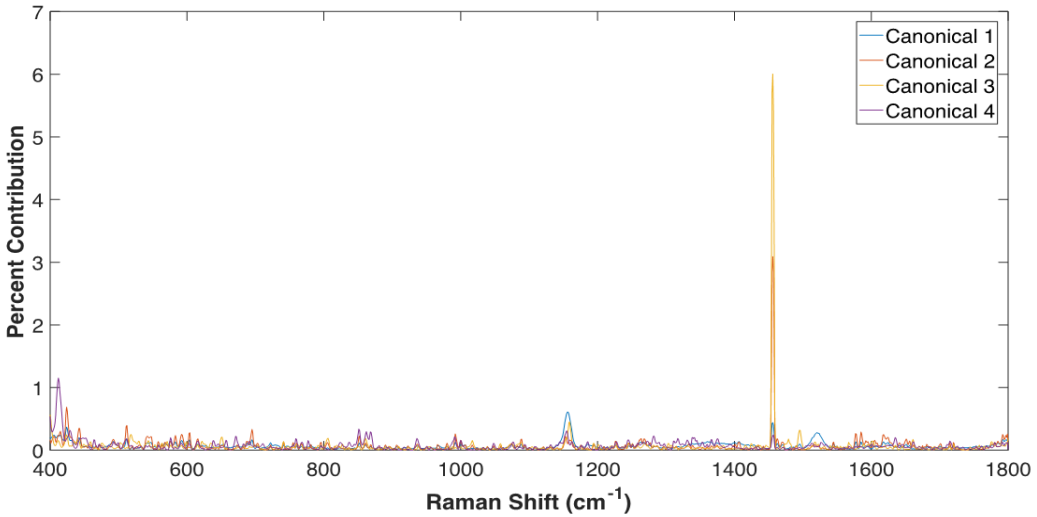


Figure D. 3. Raman shift contributions to DAPC clustering in Figure 5.3B, separation by the presence of $MgSO_4$ in BG-11 medium.

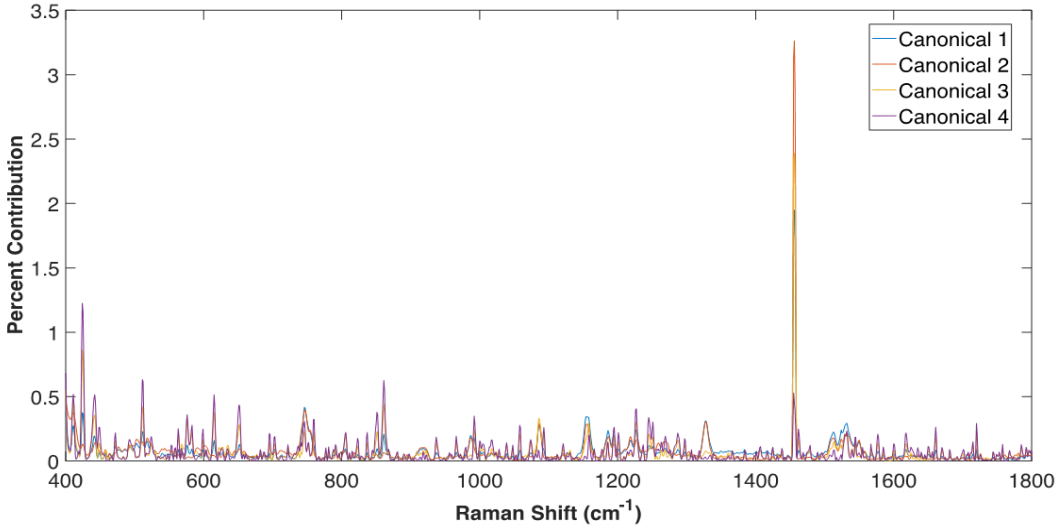


Figure D. 4. Raman shift contributions to DAPC clustering in Figure 5.4B, separation by all treatments in BG-11 medium.

Table D. 1. PCA contributions

Acetate Study (Figure 5.1C)		
Major Band	**Designation**	**Reference**
573	Tryptophan/cytosine, guanine	(1)
630	Glycerol	(2)
1088	C-C stretch	(1)
1156	Protein assignment; Carotenoids	(3–5)
1518	Carotenoids	(6)
NaCl Study (Figure 5.2C)		
Major Band	**Designation**	**Reference**
428	phosphate of HA	(7)
515	Serine	(8)
573	Tryptophan/cytosine, guanine	(1)
630	Glycerol	(2)
669	Cytosine	(1)
684	Phenylalanine	(9)
725	DNA/RNA	(10)
750	Lactic acid	(11)
780	Uracil	(12)
1157	Carotenoids, beta-carotene	(3–5)
1188	Phenylalanine	(9)
1260	Amide III, Second amide, lipids, proteins	(1,3,5,13–15)
1520	Carotenoid, carotene	(1,16)
$MgSO_4$ Study (Figure 5.3C)		
Major Band	**Designation**	**Reference**

429	Cholesterol, cholesterol ester	(2)
573	Tryptophan/cytosine, guanine	(1)
630	Glycerol	(2)
684	Phenylalanine	(9)
750	Lactic acid	(11)
1157	Carotenoids, beta-carotene	(3–5)
1188	Phenylalanine	(9)
1330	Phospholipids	(16)
1380	Lipid assignment	(17)
1520	Carotenoid, carotene	(1,16)
All Phenotypes Study (Figure 5.4C)		
Major Band	**Designation**	**Reference**
428	Phosphate of HA	(7)
575	Phosphatidylinositol	(2)
625	Lysine	(9)
685	Tryptophan	(9)
725	DNA/RNA	(10)
750	Lactic acid	(11)
780	Uracil	(12)
1087	acyl backbone in lipids	(7)
1157	Carotenoids, beta-carotene	(3–5)
1188	Phenylalanine	(9)
1328	Nucleic acids	(18)
1360	Tryptophan	(19)
1520	Carotenoid, carotene	(1,16)

Table D. 2. DAPC contributions

Acetate Study (Fig. D1)		
Major Band	**Designation**	**Reference**
415	Phosphatidylinositol	(2)
436	Unknown	
477	Polysaccharides, amylose, amylopectin	(19)
502	Riboflavin	(8)
512	Unknown	
524	Proteins, Phosphatidylserine	(1,2,20)
532	Alanine	(8)
543	Unknown	
618	Protein	(10)
638	Methionine	(19)
689	Unknown	
698	Arabinose, N-Acetyl-D-glucosamine	(8)
704	Galactosamine	(8)
715	Tyrosine, Valine	(9)
746	DNA/RNA	(10)
920	Proline ring, glucose, lactic acid	(1,20)
988	Glutamic acid, lysine, tryptophan	(9)
1005	Phenylalanine, proteins	(21)
1155	Protein assignment; Carotenoids	(1,20)
1086	acyl backbone in lipids	(7)
1196	Cysteine	(9)
1327	Glycine, serine	(9)
1495	Histidine	(9)

1510	DNA	(10)
1528	Carotenoids	(3)
1550	Tryptophan	(6)
1558	Tryptophan	(7)
1615	Tyrosine, tryptophan	(10)
1790	Unknown	
NaCl Study (Fig. D2)		
Major Band	**Designation**	**Reference**
405	Unknown	
422	Unknown	
430	Cholesterol, cholesterol ester	(2)
594	Phosphatidylinositol	
650	Guanine	(8)
712	Unknown	
730	Proline	(21)
746	DNA/RNA	(10)
798	Unknown	
854	Polysaccharides, pectin, tyrosine	(19,21)
1040	Carbohydrates, Proline	(13,22)
1090	Phosphate, DNA	(13,23)
1157	Carotenoids, beta-carotene	(3–5)
1200	Nucleic acids and phosphates	(24)
1255	Lipids	(25)
1329	DNA, Phospholipid collagen, Nucleic acids and phosphates	(16,24,26)
1495	Histidine	(9)
1517	Beta-carotene	(5)
$MgSO_4$ Study (Fig. D3)		

Major Band	Designation	Reference
412	Phosphatidylinositol	(2)
424	Unknown	
850	Amino acids, valine, polysaccharides, tyrosine	(19,27)
862	Unknown	
868	Monosaccharides	(19)
992	Aspartate, proline	(9)
1157	Carotenoids, beta-carotene	(3–5)
1456	Deoxyribose	(28)
1496	Riboflavin	(8)
1520	Carotenoid, carotene	(1,16)
1577	Guanine, Bound & free NADH	(16,28)
1585	C=C olefinic stretch	(29)
All Phenotypes Study (Fig. D4)		
Major Band	**Designation**	**Reference**
412	Phosphatidylinositol	(2)
424	Unknown	
442	Unknown	
509	Collagen, Cysteine	(7,19)
615	Cholesterol ester	(2)
652	Alanine, Arginine	(8,9)
746	DNA/RNA	(10)
850	Amino acids, valine, polysaccharides, tyrosine	(19,27)
861	Phosphate group, Phosphatidic acid	(2)
992	Aspartate, proline	(9)
1058	Lipids	(13)
1086	acyl backbone in lipids	(7)
1155	Proteins, carotenoid	(1,20)

1226	Lysine	(9)
1245	Amide III	(19)
1251	Guanine, cytosine	(28)
1327	Nucleic acids	(18)
1456	Deoxyribose	(28)
1462	Disaccharides, sucrose	(19)
1513	Cytosine	(28)
1577	Guanine, Bound & free NADH	(16,28)
1618	Tryptophan, Bound and free NADH	(6,16)
1661	Amide I, Lipids	(21,25)
1720	C=O	(2)

Table D. 3. Mean values of Total Spectral Distance (TSD), Total Principal canonical Distance (TPD) and Total Canonical Distance (TCD)

Acetate Study (Reference = 0; 12 PCs (99.2758%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
0'	0.1182644181	0.1242458332	0.09700696619	0.1137800413	13.55223002	7.993008539
15'	0.1564892439	0.163263582	0.1232850746	0.1487330242	64.58922464	19.12495605
30'	0.4191560852	0.2114457536	0.3537607939	0.2342432018	98.87762659	24.09609408
7.5'	0.145724401	0.1305455851	0.1181000853	0.1271310151	47.46829317	21.67648776
NaCl Study (Reference = 0; 12 PCs (99.5533%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
0'	0.118296135	0.1242557691	0.09551997633	0.1146655784	15.82404142	8.21365186
50'	0.2867847224	0.2354245872	0.2425456755	0.2470550451	86.10363268	25.1568395
100'	0.2088394906	0.1909190586	0.1645341464	0.192105271	88.13745897	24.36953821
150'	0.3689953028	0.2561297948	0.3309776715	0.2651606197	78.93040386	21.70275814

$MgSO_4$ Study (Reference = 7.5; 12 PCs (99.1432%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
62.5'	0.118296135	0.1242557691	0.09822638996	0.1222952466	20.25270236	11.98429652
0'	0.4844252476	0.2178491661	0.4618130925	0.2192188051	638.3105672	89.2194826
15.625 '	0.3701721025	0.1218309497	0.3388166134	0.1302162348	686.8762703	93.01281108
31.25'	0.4248986123	0.1343619981	0.4054577552	0.1293354092	643.5199691	88.54078455
All Phenotypes Study (Reference = 0, 15 PCs (99.2527%))						
Factor	**TSD (mean)**	**TSD (stdev)**	**TPD (mean)**	**TPD (stdev)**	**TCD (mean)**	**TCD (stdev)**
Acetate'	0.2404565767	0.2127848747	0.1572684831	0.1725368778	34.42091995	15.28187715
BG-11	0.1182644181	0.1242458332	0.09090765282	0.1088634695	11.16701488	7.362581579
MgSO4	0.4261322316	0.1698890353	0.3647929079	0.1576598915	162.700415	22.2394417
NaCl	0.2879949365	0.237425449	0.2169180631	0.2219105416	48.47501731	20.97899238

Table D. 4. ANOVA test based on TPD data

Acetate Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	3.558732924	3	1.186244308	44.72192164	3.66E-24
Error'	8.487966632	320	0.02652489572	[]	[]
Total'	12.04669956	323	[]	[]	[]
NaCl Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	2.499443359	3	0.8331477865	18.37160911	5.06E-11
Error'	14.51191837	320	0.04534974489	[]	[]
Total'	17.01136173	323	[]	[]	[]
$MgSO_4$ Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**

Groups'	6.221106936	3	2.073702312	85.78151801	9.47E-41
Error'	7.735754219	320	0.02417423193	[]	[]
Total'	13.95686116	323	[]	[]	[]
All Phenotypes Study					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	7.308706221	3	2.436235407	75.27830383	6.25E-43
Error'	26.08461719	806	0.03236304862	[]	[]
Total'	33.39332341	809	[]	[]	[]

Table D. 5. ANOVA test based on TCD data

Acetate Study (12 PCs)					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	306731.566	3	102243.8553	276.3078599	1.83E-88
Error'	118411.5201	320	370.0360003	[]	[]
Total'	425143.0861	323	[]	[]	[]
NaCl Study (12 PCs)					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	289397.3227	3	96465.77422	218.5927214	3.96E-77
Error'	141217.1803	320	441.3036885	[]	[]
Total'	430614.503	323	[]	[]	[]
$MgSO_4$ Study (12 PCs)					
Source'	**SS'**	**df'**	**MS'**	**F'**	**Prob>F'**
Groups'	24686980.04	3	8228993.348	1338.341862	1.11E-180
Error'	1967567.439	320	6148.648248	[]	[]
Total'	26654547.48	323	[]	[]	[]
All Phenotypes Study (15 PCs)					

Source'	SS'	df'	MS'	F'	Prob>F'
Groups'	2770121.423	3	923373.8077	2592.69579	0.00E+00
Error'	287052.2998	806	356.1442925	[]	[]
Total'	3057173.723	809	[]	[]	[]

Table D. 6. Pairwise Comparison based on TPD data

Acetate Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
15'	7.5'	-0.0605609306	0.005184989337	0.07093090928	0.997058447
15'	0'	-0.0394678115	0.02627810844	0.09202402838	7.34E-01
15'	30'	-0.2962216392	-0.2304757192	-0.1647297993	3.77E-09
7.5'	0'	-0.04465280084	0.0210931191	0.08683903904	8.43E-01
7.5'	30'	-0.3014066285	-0.2356607086	-0.1699147886	3.77E-09
0'	30'	-0.3224997476	-0.2567538277	-0.1910079077	3.77E-09
NaCl Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
0'	100'	-0.1549806983	-0.06901417005	0.01695235819	0.165547446
0'	150'	-0.3214242234	-0.2354576951	-0.1494911669	3.78E-09
0'	50'	-0.2329922274	-0.1470256992	-0.06105917092	6.58E-05
100'	150'	-0.2524100533	-0.1664435251	-0.08047699683	3.91E-06
100'	50'	-0.1639780574	-0.07801152911	0.007954999128	0.09100238259
150'	50'	0.002465467721	0.08843199596	0.1743985242	0.04099165105

$MgSO_4$ Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
62.5'	0'	-0.4263518142	-0.3635867025	-0.3008215908	3.77E-09
62.5'	15.625'	-0.3033553352	-0.2405902235	-0.1778251118	3.77E-09
62.5'	31.25'	-0.369996477	-0.3072313653	-0.2444662536	3.77E-09
0'	15.625'	0.06023136729	0.122996479	0.1857615907	2.86E-06
0'	31.25'	-0.006409774507	0.05635533721	0.1191204489	0.0964683347
15.625'	31.25'	-0.1294062535	-0.0666411418	-0.003876030081	0.03235263121
All Phenotypes Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
Acetate'	BG-11'	0.00706542426	0.06636083028	0.1256562363	0.02104970698
Acetate'	NaCl'	-0.1015777636	-0.05964957996	-0.01772139626	0.001468085923
Acetate'	MgSO4'	-0.2494526085	-0.2075244248	-0.1655962411	3.77E-09
BG-11'	NaCl'	-0.1853058163	-0.1260104102	-0.06671500422	2.89E-07
BG-11'	MgSO4'	-0.3331806611	-0.2738852551	-0.2145898491	3.77E-09
NaCl'	MgSO4'	-0.1898030286	-0.1478748449	-0.1059466612	3.77E-09

Table D. 7. Pairwise Comparison based on TCD data

Acetate Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
15'	7.5'	9.355529302	17.12093148	24.88633365	9.19E-08
15'	0'	43.27159245	51.03699462	58.8023968	3.77E-09
15'	30'	-42.05380412	-34.28840195	-26.52299978	3.77E-09
7.5'	0'	26.15066097	33.91606315	41.68146532	3.77E-09
7.5'	30'	-59.1747356	-51.40933342	-43.64393125	3.77E-09
0'	30'	-93.09079874	-85.32539657	-77.5599944	3.77E-09
NaCl Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
0'	100'	-80.7937083	-72.31341755	-63.8331268	3.77E-09
0'	150'	-71.58665319	-63.10636244	-54.62607169	3.77E-09
0'	50'	-78.75988201	-70.27959126	-61.79930051	3.77E-09
100'	150'	0.7267643553	9.207055108	17.68734586	0.02712138432
100'	50'	-6.446464461	2.033826291	10.51411704	0.9269796481
150'	50'	-15.65351957	-7.173228816	1.307061936	0.1307527458
$MgSO_4$ Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**

62.5'	0'	-649.7120974	-618.0578648	-586.4036323	3.77E-09
62.5'	15.625'	-698.2778005	-666.6235679	-634.9693354	3.77E-09
62.5'	31.25'	-654.9214993	-623.2672668	-591.6130342	3.77E-09
0'	15.625'	-80.21993564	-48.56570311	-16.91147059	0.0004699121492
0'	31.25'	-36.86363447	-5.209401942	26.44483058	0.974603151
15.625'	31.25'	11.70206865	43.35630117	75.01053369	0.00244751384
All Phenotypes Study					
Group 1	**Group 2**	**Lower Level 95% Confidence**	**Difference in Group Means**	**Upper Level 95% confidence**	***p*-value**
Acetate'	BG-11'	0.00706542426	0.06636083028	0.1256562363	0.02104970698
Acetate'	NaCl'	-0.1015777636	-0.05964957996	-0.01772139626	0.001468085923
Acetate'	$MgSO_4$'	-0.2494526085	-0.2075244248	-0.1655962411	3.77E-09
BG-11'	NaCl'	-0.1853058163	-0.1260104102	-0.06671500422	2.89E-07
BG-11'	$MgSO_4$'	-0.3331806611	-0.2738852551	-0.2145898491	3.77E-09
NaCl'	$MgSO_4$'	-0.1898030286	-0.1478748449	-0.1059466612	3.77E-09

Table D. 8. Regression of TSD, TPD and TCD data

Acetate Study			
	TSD fit	**TPD fit**	**TCD fit**
Slope	0.01014807574	0.008662031238	2.732256664
Intercept	0.07671504298	0.05934906999	20.26097489
R^2	0.8568330399	0.8405425314	0.9702867809

NaCl Study			
	TSD fit	**TPD fit**	**TCD fit**
Slope	0.001348304543	0.001256723113	0.3827058272
Intercept	0.144606072	0.114140134	38.54594719
R^2	0.6590482035	0.6397804086	0.5124254285
$MgSO_4$ Study			
	TSD fit	**TPD fit**	**TCD fit**
Slope	-0.04589606421	-0.04509937101	-86.90298653
Intercept	0.500044485	0.4740607739	782.3903018
R^2	0.8331753287	0.814645981	0.762249334

Table D. 9. Rametrix™ PRO analysis

Study	**Factor**	**Number of PCs**	**% Variability Explained by PCs**	**Accuracy**	**Sensitivity**	**Selectivity**	**Scans Used**
Acetate	0 mM	6	92.3537	77.78%	44.44%	88.89%	SYN1-SYN27, SYN53-SYN61
Acetate	0 mM	10	98.7642	77.78%	33.33%	92.59%	SYN1-SYN27, SYN53-SYN61
Acetate	0 mM	12	99.2758	91.67%	100.00%	88.89%	SYN1-SYN27, SYN53-SYN61
Acetate	0 mM	15	99.5727	83.33%	100.00%	77.78%	SYN1-SYN27, SYN53-SYN61
Acetate	0 mM	20	99.7728	66.67%	100.00%	55.56%	SYN1-SYN27, SYN53-SYN61
Acetate	0 mM	25	99.8883	30.56%	100.00%	7.41%	SYN1-SYN27, SYN53-SYN61
Acetate	0 mM	30	99.9604	75.00%	77.78%	74.07%	SYN1-SYN27, SYN53-SYN61

Acetate	0 mM	33	99.9889	66.67%	11.11%	85.19%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	6	92.3537	69.44%	22.22%	85.19%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	10	98.7642	77.78%	55.56%	85.19%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	12	99.2758	77.78%	66.67%	81.48%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	15	99.5727	80.56%	77.78%	81.48%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	20	99.7728	41.67%	100.00%	22.22%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	25	99.8883	27.78%	100.00%	3.70%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	30	99.9604	30.56%	100.00%	7.41%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	33	99.9889	63.89%	33.33%	74.07%	SYN1-SYN27, SYN53-SYN61
Acetate	7.5 mM	4	82.5808	91.67%	66.67%	100.00%	SYN1-SYN27, SYN53-SYN61
Acetate	15 mM	6	92.3537	83.33%	33.33%	100.00%	SYN1-SYN27, SYN53-SYN61
Acetate	15 mM	10	98.7642	88.89%	55.56%	100.00%	SYN1-SYN27, SYN53-SYN61
Acetate	15 mM	12	99.2758	88.89%	55.56%	100.00%	SYN1-SYN27, SYN53-SYN61
Acetate	15 mM	15	99.5727	86.11%	55.56%	96.30%	SYN1-SYN27, SYN53-SYN61

Acetate	15 mM	20	99.7728	69.44%	88.89%	62.96%	SYN1-SYN27, SYN53-SYN61
Acetate	15 mM	25	99.8883	27.78%	100.00%	3.70%	SYN1-SYN27, SYN53-SYN61
Acetate	15 mM	30	99.9604	30.56%	100.00%	7.41%	SYN1-SYN27, SYN53-SYN61
Acetate	15 mM	33	99.9889	72.22%	0.00%	96.30%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	6	92.3537	83.33%	55.56%	92.59%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	10	98.7642	86.11%	66.67%	92.59%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	12	99.2758	86.11%	55.56%	96.30%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	15	99.5727	86.11%	55.56%	96.30%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	20	99.7728	91.67%	100.00%	88.89%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	25	99.8883	55.56%	100.00%	40.74%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	30	99.9604	61.11%	100.00%	48.15%	SYN1-SYN27, SYN53-SYN61
Acetate	30 mM	33	99.9889	33.33%	88.89%	14.81%	SYN1-SYN27, SYN53-SYN61
NaCl	0 mM	6	95.9963	88.89%	55.56%	100.00%	SYN19-SYN52, SYN89-SYN91
NaCl	0 mM	8	98.453	97.22%	88.89%	100.00%	SYN19-SYN52, SYN89-SYN91

NaCl	0 mM	10	99.3168	100.00%	100.00%	100.00%	SYN19-SYN52, SYN89-SYN91
NaCl	0 mM	12	99.5533	100.00%	100.00%	100.00%	SYN19-SYN52, SYN89-SYN91
NaCl	0 mM	15	99.7577	97.22%	100.00%	96.30%	SYN19-SYN52, SYN89-SYN91
NaCl	0 mM	20	99.881	52.78%	100.00%	37.04%	SYN19-SYN52, SYN89-SYN91
NaCl	50 mM	6	95.9963	66.67%	33.33%	77.78%	SYN19-SYN52, SYN89-SYN91
NaCl	50 mM	8	98.453	77.78%	11.11%	100.00%	SYN19-SYN52, SYN89-SYN91
NaCl	50 mM	10	99.3168	69.44%	0.00%	92.59%	SYN19-SYN52, SYN89-SYN91
NaCl	50 mM	12	99.5533	80.56%	33.33%	96.30%	SYN19-SYN52, SYN89-SYN91
NaCl	50 mM	15	99.7577	63.89%	33.33%	74.07%	SYN19-SYN52, SYN89-SYN91
NaCl	50 mM	20	99.881	41.67%	100.00%	22.22%	SYN19-SYN52, SYN89-SYN91
NaCl	100 mM	6	95.9963	75.00%	44.44%	85.19%	SYN19-SYN52, SYN89-SYN91
NaCl	100 mM	8	98.453	83.33%	33.33%	100.00%	SYN19-SYN52, SYN89-SYN91
NaCl	100 mM	10	99.3168	91.67%	66.67%	100.00%	SYN19-SYN52, SYN89-SYN91
NaCl	100 mM	12	99.5533	86.11%	44.44%	100.00%	SYN19-SYN52, SYN89-SYN91

NaCl	100 mM	15	99.7577	83.33%	44.44%	96.30%	SYN19-SYN52, SYN89-SYN91
NaCl	100 mM	20	99.881	77.78%	66.67%	81.48%	SYN19-SYN52, SYN89-SYN91
NaCl	150 mM	6	95.9963	69.44%	22.22%	85.19%	SYN19-SYN52, SYN89-SYN91
NaCl	150 mM	8	98.453	80.56%	22.22%	100.00%	SYN19-SYN52, SYN89-SYN91
NaCl	150 mM	10	99.3168	72.22%	22.22%	88.89%	SYN19-SYN52, SYN89-SYN91
NaCl	150 mM	12	99.5533	77.78%	44.44%	88.89%	SYN19-SYN52, SYN89-SYN91
NaCl	150 mM	15	99.7577	66.67%	33.33%	77.78%	SYN19-SYN52, SYN89-SYN91
NaCl	150 mM	20	99.881	41.67%	88.89%	25.93%	SYN19-SYN52, SYN89-SYN91
$MgSO_4$	62.5 mM	2	75.5401	100.00%	100.00%	100.00%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	62.5 mM	4	92.0435	100.00%	100.00%	100.00%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	62.5 mM	6	97.0328	100.00%	100.00%	100.00%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	62.5 mM	8	98.2137	100.00%	100.00%	100.00%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	62.5 mM	10	98.8121	100.00%	100.00%	100.00%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	62.5 mM	12	99.1432	100.00%	100.00%	100.00%	SYN19-SYN27, SYN62-SYN88

$MgSO_4$	62.5 mM	20	99.7034	97.22%	100.00%	96.30%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	62.5 mM	33	99.981	50.00%	88.89%	37.04%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	2	75.5401	80.56%	44.44%	92.59%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	4	92.0435	80.56%	44.44%	92.59%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	6	97.0328	77.78%	44.44%	88.89%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	8	98.2137	77.78%	33.33%	92.59%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	10	98.8121	77.78%	22.22%	96.30%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	12	99.1432	72.22%	55.56%	77.78%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	14	99.3787	77.78%	77.78%	77.78%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	16	99.5225	69.44%	88.89%	62.96%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	18	99.6217	52.78%	100.00%	37.04%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	20	99.7034	47.22%	100.00%	29.63%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	0 mM	33	99.981	72.22%	22.22%	88.89%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	2	75.5401	61.11%	44.44%	66.67%	SYN19-SYN27, SYN62-SYN88

$MgSO_4$	15.625 mM	4	92.0435	66.67%	66.67%	66.67%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	6	97.0328	69.44%	22.22%	85.19%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	8	98.2137	86.11%	55.56%	96.30%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	10	98.8121	86.11%	66.67%	92.59%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	12	99.1432	91.67%	100.00%	88.89%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	14	99.3787	88.89%	100.00%	85.19%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	16	99.5225	83.33%	100.00%	77.78%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	18	99.6217	77.78%	100.00%	70.37%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	15.625 mM	20	99.7034	58.33%	100.00%	44.44%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	2	75.5401	72.22%	66.67%	74.07%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	4	92.0435	77.78%	66.67%	81.48%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	6	97.0328	83.33%	44.44%	96.30%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	8	98.2137	80.56%	22.22%	100.00%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	10	98.8121	88.89%	77.78%	92.59%	SYN19-SYN27, SYN62-SYN88

$MgSO_4$	31.25 mM	12	99.1432	94.44%	88.89%	96.30%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	14	99.3787	88.89%	100.00%	85.19%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	16	99.5225	88.89%	100.00%	85.19%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	18	99.6217	91.67%	100.00%	88.89%	SYN19-SYN27, SYN62-SYN88
$MgSO_4$	31.25 mM	20	99.7034	72.22%	100.00%	62.96%	SYN19-SYN27, SYN62-SYN88
All Phenotypes	BG-11	10	98.324	92.22%	22.22%	100.00%	SYN1-SYN91
All Phenotypes	BG-11	15	99.2527	100.00%	100.00%	100.00%	SYN1-SYN91
All Phenotypes	BG-11	20	99.5184	86.67%	100.00%	85.19%	SYN1-SYN91
All Phenotypes	BG-11	30	99.7286	40.00%	100.00%	33.33%	SYN1-SYN91
All Phenotypes	BG-11	40	99.8347	10.00%	100.00%	0.00%	SYN1-SYN91
All Phenotypes	BG-11	50	99.9036	10.00%	100.00%	0.00%	SYN1-SYN91
All Phenotypes	Acetate	10	98.324	81.11%	62.96%	88.89%	SYN1-SYN91
All Phenotypes	Acetate	15	99.2527	86.67%	55.56%	100.00%	SYN1-SYN91
All Phenotypes	Acetate	20	99.5184	84.44%	48.15%	100.00%	SYN1-SYN91
All Phenotypes	Acetate	30	99.7286	84.44%	48.15%	100.00%	SYN1-SYN91
All Phenotypes	Acetate	40	99.8347	88.89%	77.78%	93.65%	SYN1-SYN91
All Phenotypes	Acetate	50	99.9036	48.89%	100.00%	26.98%	SYN1-SYN91
All Phenotypes	NaCl	10	98.324	83.33%	74.07%	87.30%	SYN1-SYN91
All Phenotypes	NaCl	15	99.2527	87.78%	70.37%	95.24%	SYN1-SYN91
All Phenotypes	NaCl	20	99.5184	84.44%	48.15%	100.00%	SYN1-SYN91
All Phenotypes	NaCl	30	99.7286	80.00%	33.33%	100.00%	SYN1-SYN91

All Phenotypes	NaCl	40	99.8347	83.33%	51.85%	96.83%	SYN1-SYN91
All Phenotypes	NaCl	50	99.9036	86.67%	96.30%	82.54%	SYN1-SYN91
All Phenotypes	$MgSO_4$	10	98.324	96.67%	88.89%	100.00%	SYN1-SYN91
All Phenotypes	$MgSO_4$	15	99.2527	97.78%	92.59%	100.00%	SYN1-SYN91
All Phenotypes	$MgSO_4$	20	99.5184	95.56%	85.19%	100.00%	SYN1-SYN91
All Phenotypes	$MgSO_4$	30	99.7286	97.78%	92.59%	100.00%	SYN1-SYN91
All Phenotypes	$MgSO_4$	40	99.8347	97.78%	100.00%	96.83%	SYN1-SYN91
All Phenotypes	$MgSO_4$	50	99.9036	47.78%	100.00%	25.40%	SYN1-SYN91

Table D. 10. Random chance analysis

Study	Random Accuracy	Random Sensitivity	Random Selectivity
Acetate Phenotypes (n = 100,000)			
0 mM	62.48%	24.93%	75.00%
15 mM	62.48%	24.96%	74.99%
30 mM	62.48%	25.01%	74.97%
7.5 mM	62.47%	24.96%	74.97%
NaCl Phenotypes (n = 100,000)			
0 mM	62.50%	24.98%	75.01%
100 mM	62.51%	24.96%	75.02%
150 mM	62.49%	24.96%	74.99%
50 mM	62.50%	25.01%	75.00%
$MgSO_4$ Phenotypes (n = 100,000)			
62.5 mM	62.49%	24.94%	75.00%

0 mM	62.49%	24.99%	75.00%
15.625 mM	62.52%	25.07%	75.01%
31.25 mM	62.53%	25.06%	75.02%
All Phenotypes (n = 100,000)			
Acetate	59.99%	25.03%	74.98%
BG-11	70.03%	25.08%	75.03%
MgSO4	59.99%	25.02%	74.98%
NaCl	60.00%	24.99%	75.01%

Table D. 11. Strains and plasmids

Strain/plasmid	**Relevant Characteristic**	**Reference**
Synechocystis sp.	PCC 6803 [N-1]	American Type Culture Coll ection ; Manassas,
Synechocystis psba3-ko	Δ*psba3.km*^{R}	This study
Escherichia coli 10-β	Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- *ϕ*80*d*lacZΔM15 recA1 relA1 endA1 nupG rpsL *(StrR)* rph spoT1 Δ(mrr-hsdRMS-mcrBC)	New England Biolabs
pACYC177	Amp^r; Kan^r; p15 origin	Life Technologies
pUC19	Amp^r; ColE1 origin; P$_{lac}$; *lacZ*	New England Biolabs
pUC19-psba3	Amp^r; ColE1 origin; P$_{lac}$; *psba3*	This study
pPSBA3-K	Amp^r; Kan^r; ColE1 origin; P$_{lac}$; *psba3,*	This study

Table D. 12. Primers

Primer name	**Primer sequence (5'→ 3')**	**Description**
PSBA3_ F	TTGTAAAACGACGGCCAGTG_ATGACAACGACTCTCCAACAG	Insertion of *psab2* into pUC19; Forward primer
PSBA3_ R	GCTATGACCATGATTACGCCA_TTAACCGTTGACAGCAGGAGCGGTC	Insertion of *psab2* into pUC19; Reverse primer
KanR_ PSBA3_F	CTACCGCTTAGGTATGCGTC_TTCAACAAAGCCACGTTG	Insertion of *KanR* into pUC19-psba3; Forward primer
KanR_ PSBA3_R	AGTTCTGGGATTCAACTTCGG_CCAAATTTGATTCCTTTCAGCTTTGCTTCCT	Insertion of *KanR* into pUC19-psba3; Reverse primer
p19-λ-R	AACCCCTATTTGTTTATTTTTCTAAATAC	Reverse lambda primer
p19_Check_F	GTATGTTGTGTGGAATTGT	pUC19 forward check primer
p19_Check_R	GTCACAGCTTGTCTGTAAG	pUC19 reverse check primer
PSBA3_Check_F	GACAGATTTTTTTCAGCAAGC	*psba3* knockout forward check primer
PSBA3_Check_R	CCTACTCTCCTTCTAAGAGT	*psba3* knockout reverse check primer

* λ-PCR annealing part of the primers are separated by an underscore symbol

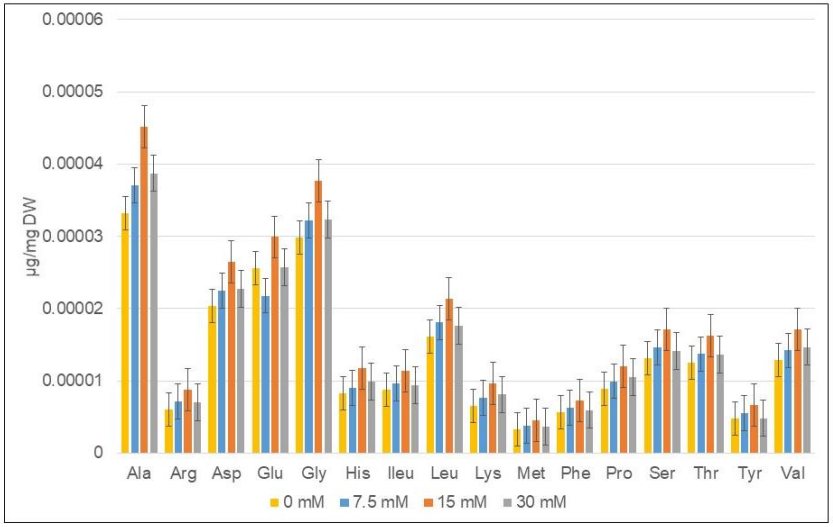
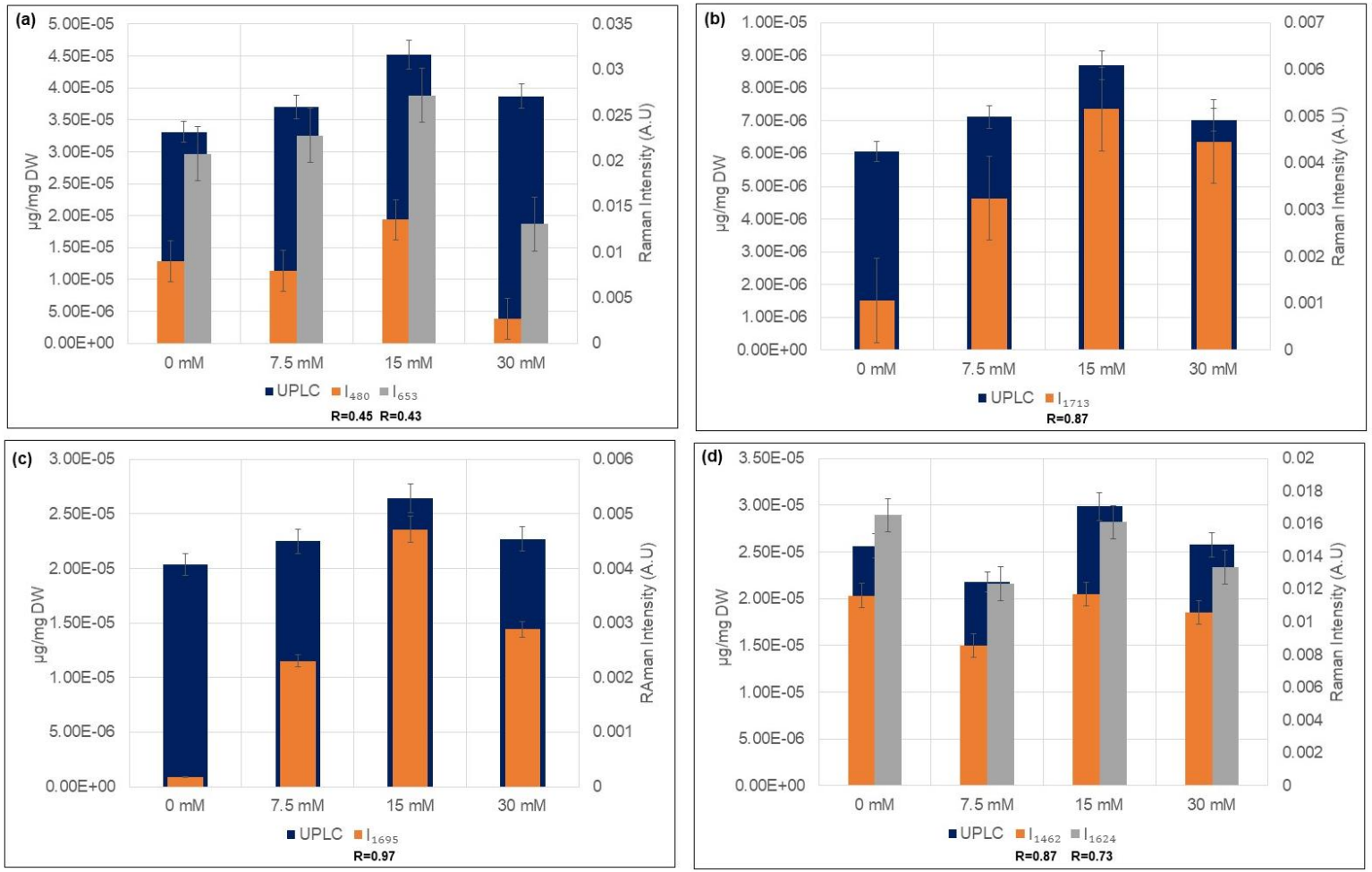
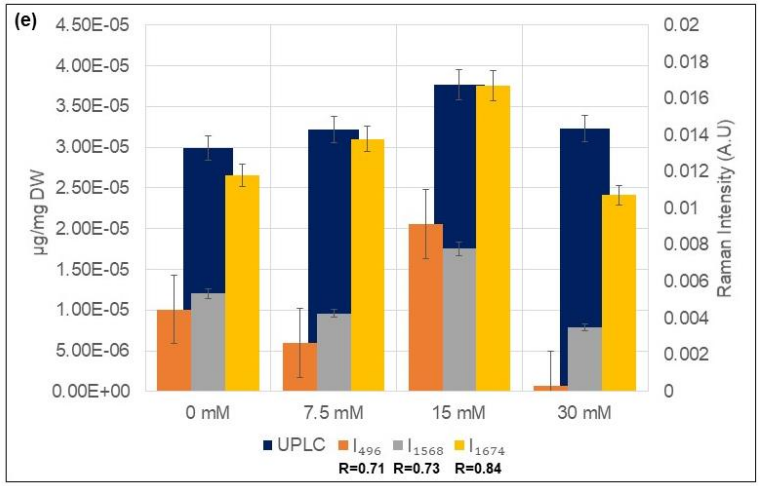
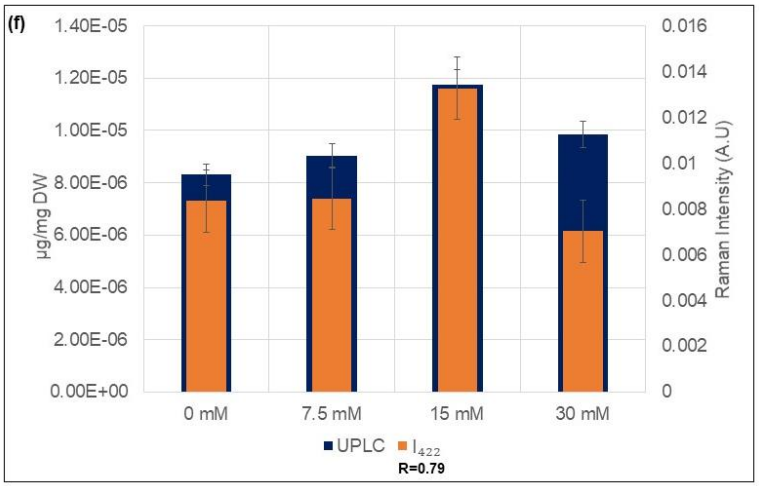
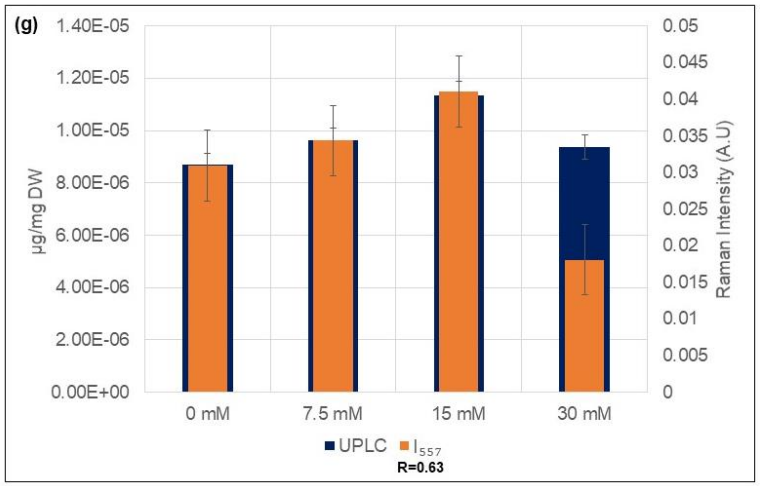


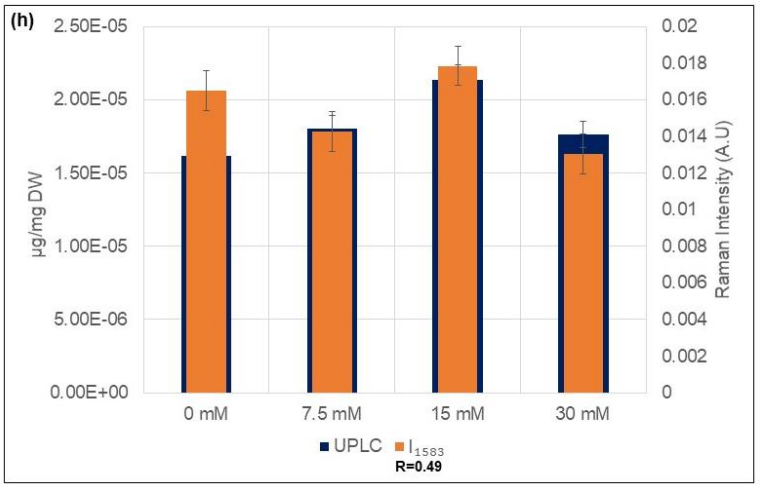
Figure D. 5. Amino acids analysis for acetate study. Data represent means ± standard deviation (SD) of values from three independent experiments.

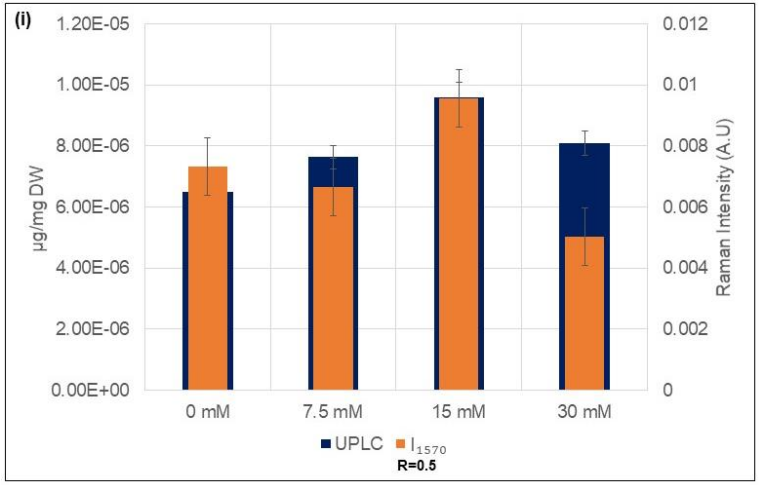

(a)
5.00E-05
4.50E-05
4.00E-05
3.50E-05
3.00E-05
2.50E-05
2.00E-05
1.50E-05
1.00E-05
5.00E-06
0.00E+00
µg/mg DW
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I480
I653
R=0.45 R=0.43
(b)
1.00E-05
9.00E-06
8.00E-06
7.00E-06
6.00E-06
5.00E-06
4.00E-06
3.00E-06
2.00E-06
1.00E-06
0.00E+00
µg/mg DW
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I1713
R=0.87
(c)
3.00E-05
2.50E-05
2.00E-05
1.50E-05
1.00E-05
5.00E-06
0.00E+00
µg/mg DW
0.006
0.005
0.004
0.003
0.002
0.001
0
RAman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I1695
R=0.97
(d)
3.50E-05
3.00E-05
2.50E-05
2.00E-05
1.50E-05
1.00E-05
5.00E-06
0.00E+00
µg/mg DW
0.02
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I1462
I1624
R=0.87 R=0.73

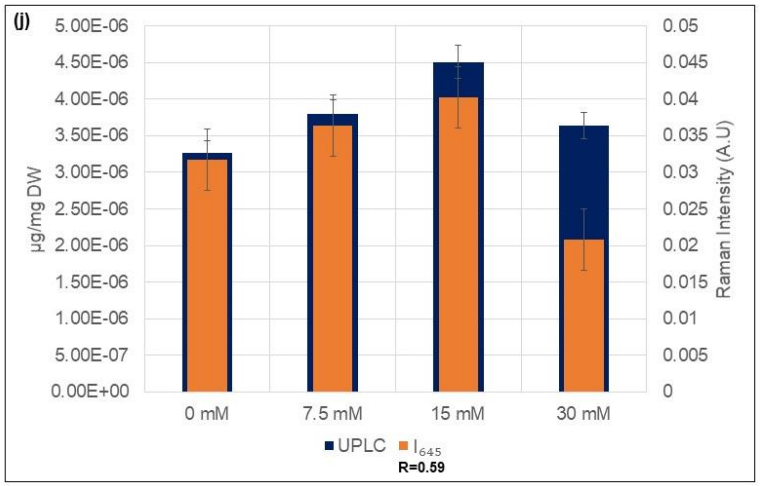

(e)
µg/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I496
I1568
I1674
R=0.71 R=0.73 R=0.84

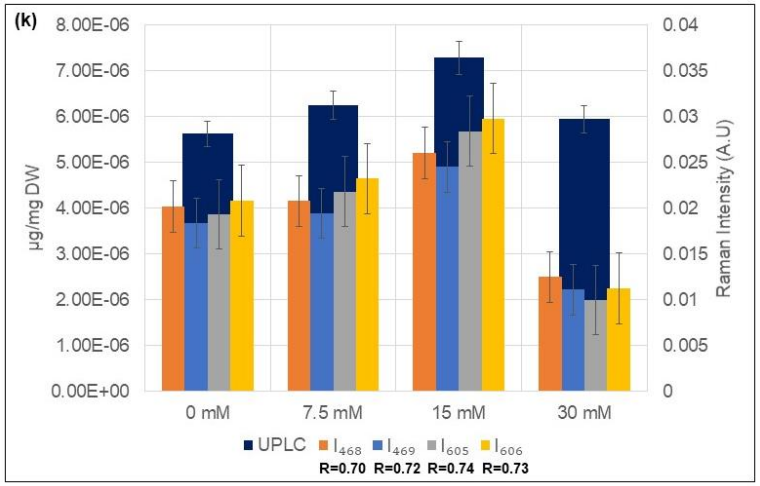

(f)
µg/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I422
R=0.79

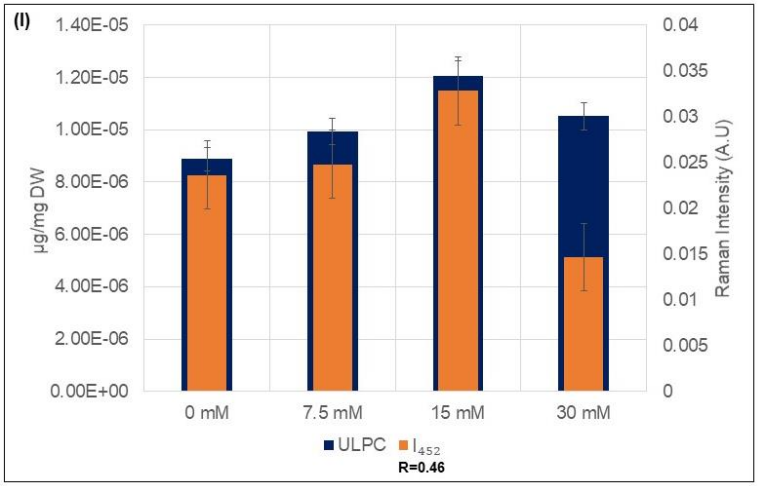

(g)
µg/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I557
R=0.63


(h)
µg/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I1583
R=0.49


(i)
1.20E-05
1.00E-05
8.00E-06
6.00E-06
4.00E-06
2.00E-06
0.00E+00
µg/mg DW
0.012
0.01
0.008
0.006
0.004
0.002
0
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I_{1570}
R=0.5


(j)
5.00E-06
4.50E-06
4.00E-06
3.50E-06
3.00E-06
2.50E-06
2.00E-06
1.50E-06
1.00E-06
5.00E-07
0.00E+00
µg/mg DW
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I_{645}
R=0.59


(k)
8.00E-06
7.00E-06
6.00E-06
5.00E-06
4.00E-06
3.00E-06
2.00E-06
1.00E-06
0.00E+00
µg/mg DW
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
UPLC
I_{468}
I_{469}
I_{605}
I_{606}
R=0.70 R=0.72 R=0.74 R=0.73


(l)
1.40E-05
1.20E-05
1.00E-05
8.00E-06
6.00E-06
4.00E-06
2.00E-06
0.00E+00
µg/mg DW
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
ULPC
I_{452}
R=0.46

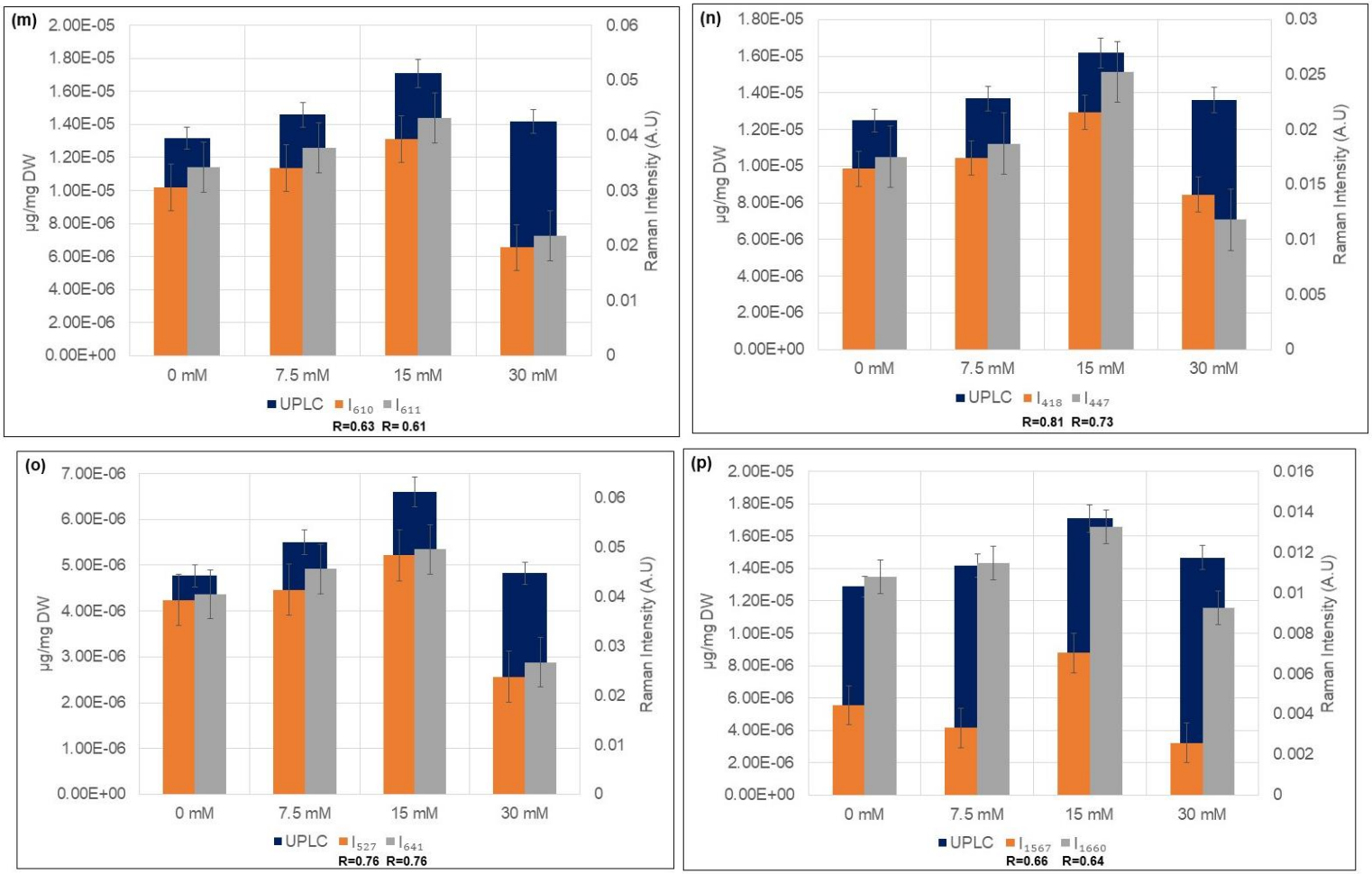


Figure D. 6. Correlation of Raman data with UPLC analysis of amino acids in the acetate study. (a) Alanine, (b) Arginine, (c) Aspartate/Asparagine, (d) Glutamate, (e) Glycine, (f) Histidine, (g) Isoleucine, (h) Leucine, (i) Lysine, (j) Methionine, (k) Phenylalanine, (l) Proline, (m) Serine, (n) Threonine, (o) Tyrosine, (p) Valine. Correlation coefficients (R) for each Raman band are represented.

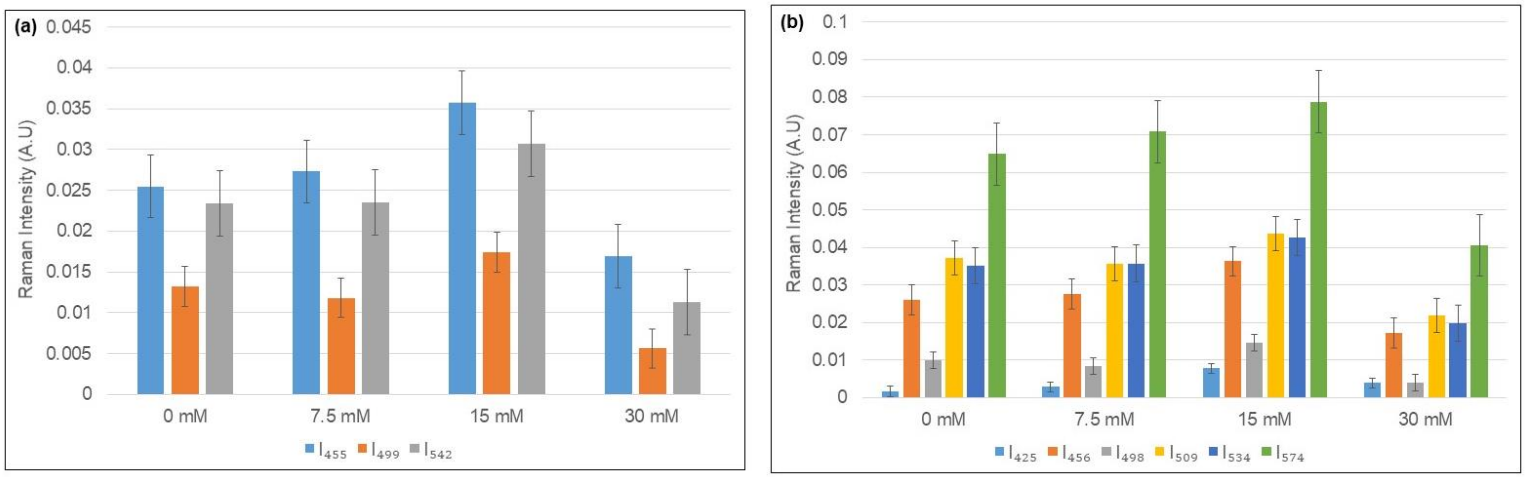


Figure D. 7. Acetate study: Raman bands predictions of amino acids non-resolved with UPLC. (a) Cysteine predictions, (b) Tryptophan predictions.

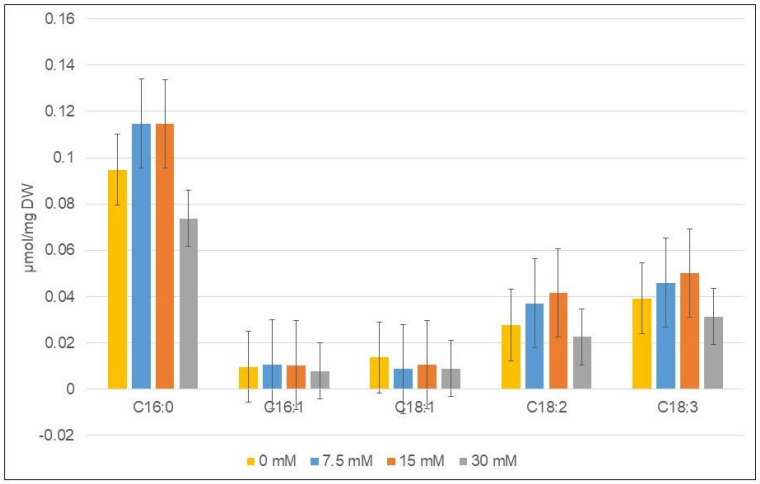
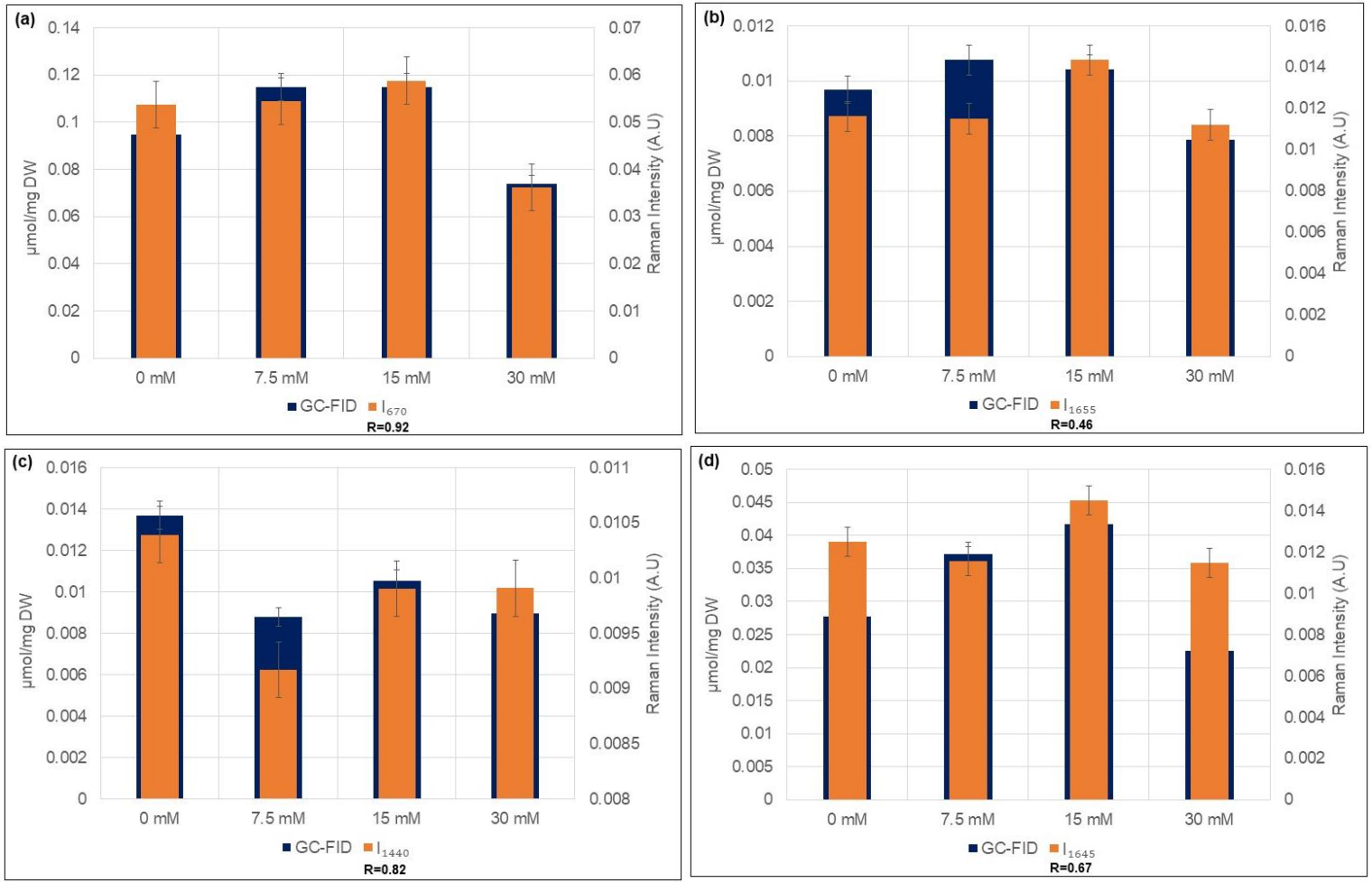


Figure D. 8. Levels of fatty acids for acetate study. Data represent means ± standard deviation (SD) of values from three independent experiments.


(a)
µmol/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
GC-FID
I670
R=0.92
(b)
µmol/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
GC-FID
I1655
R=0.46
(c)
µmol/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
GC-FID
I1440
R=0.82
(d)
µmol/mg DW
Raman Intensity (A.U)
0 mM
7.5 mM
15 mM
30 mM
GC-FID
I1645
R=0.67

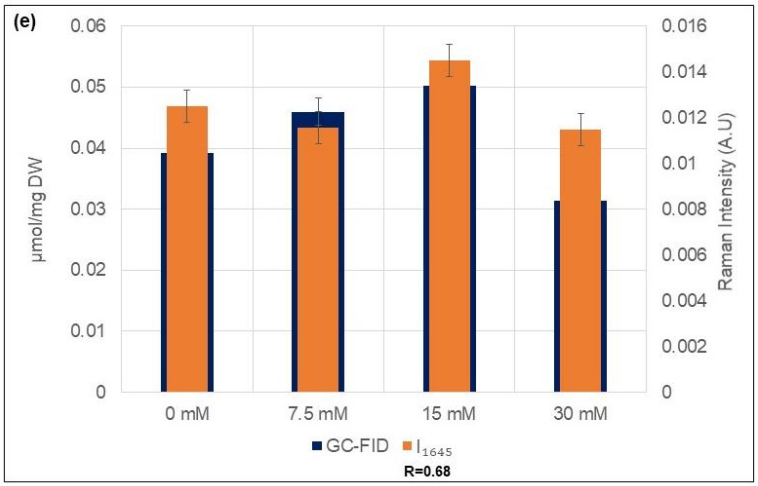


Figure D. 9. Correlation between Raman spectroscopy and GC-FID data of fatty acids for acetate study. (a) Palmitic acid, (b) Palmitoleic acid, (c) Oleic acid (d) Linoleic acid, (e) Linolenic acid. Correlation coefficients (R) for each Raman band are represented.

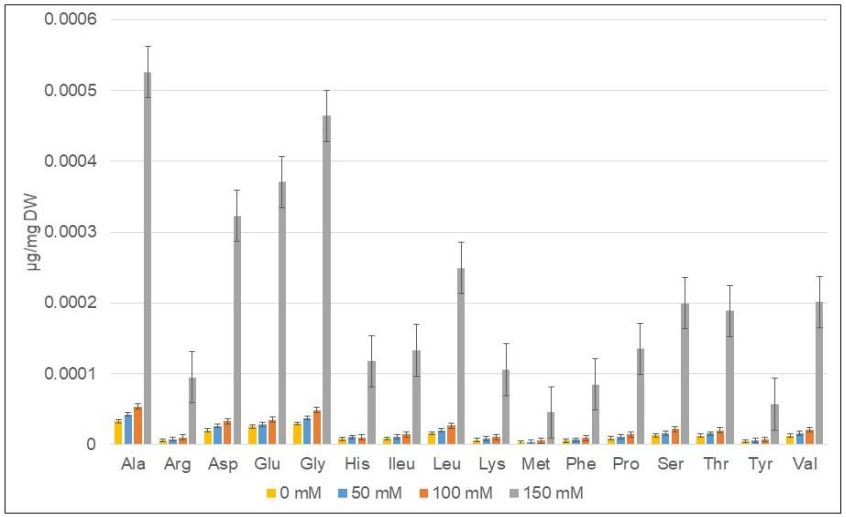
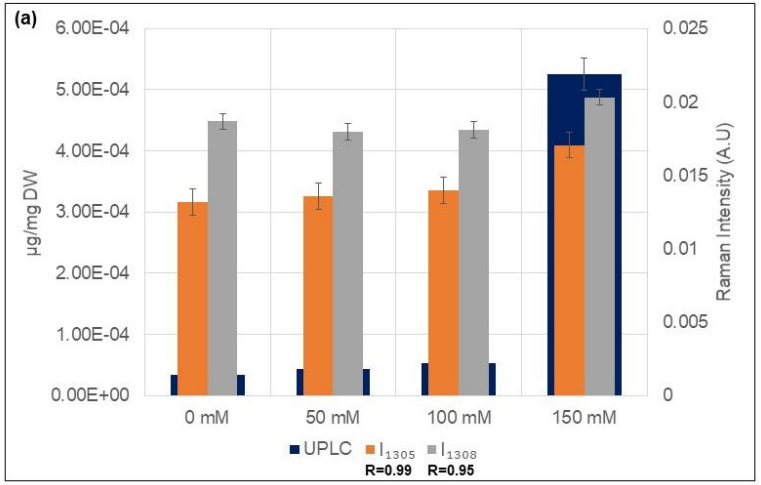
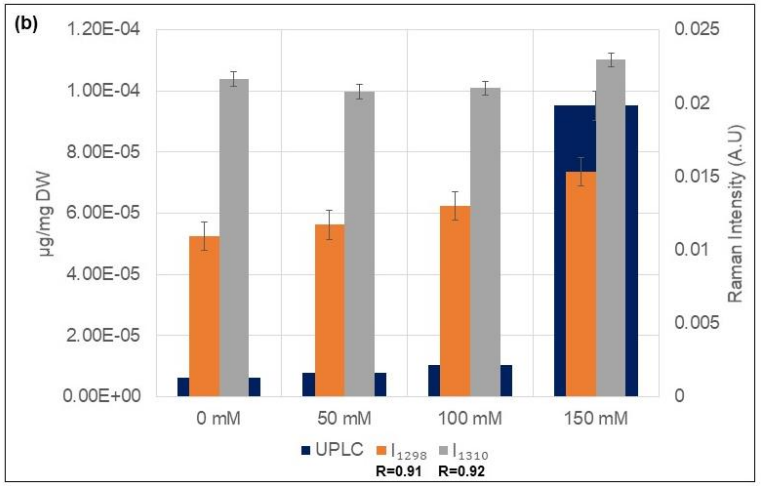
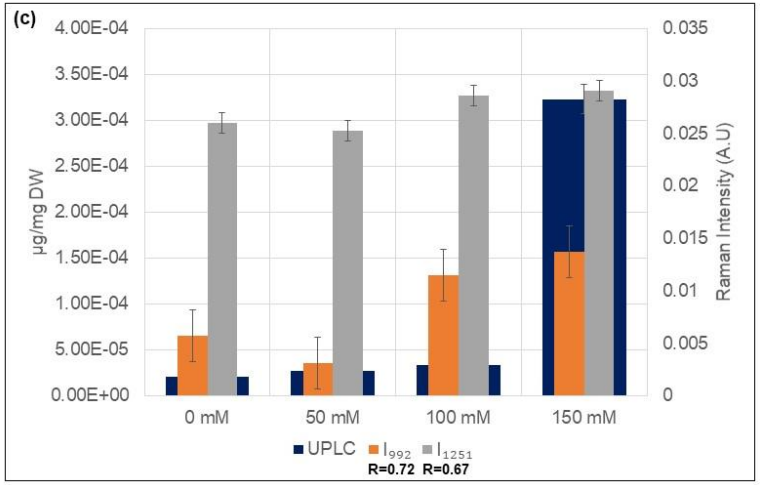
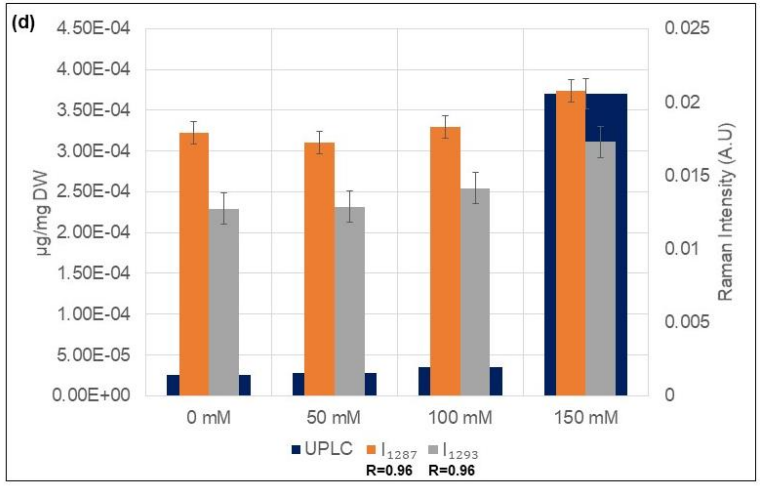


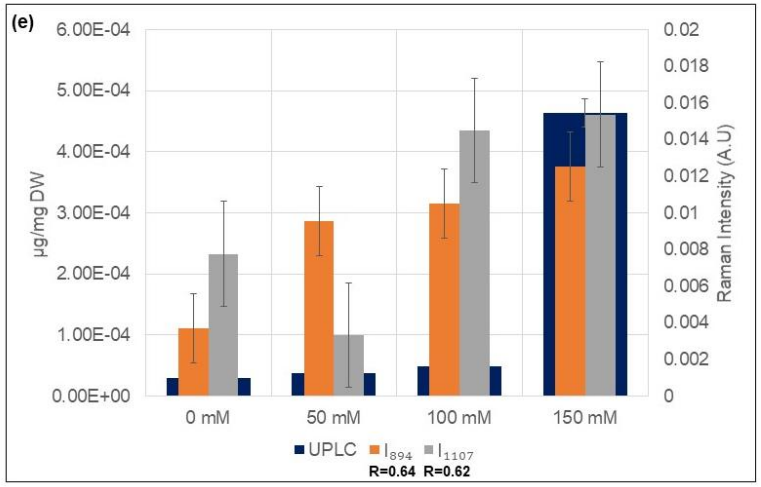
Figure D. 10. Amino acids analysis for NaCl study. Data represent means ± standard deviation (SD) of values from three independent experiments.

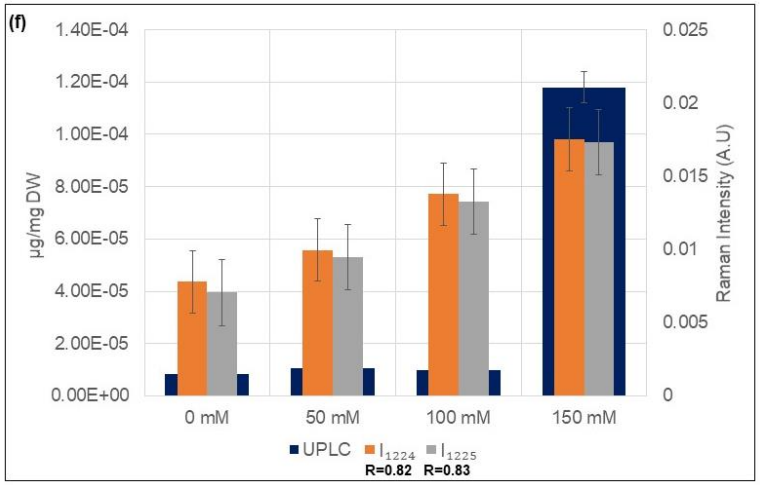

(a)
6.00E-04
5.00E-04
4.00E-04
3.00E-04
2.00E-04
1.00E-04
0.00E+00
µg/mg DW
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1305
I1308
R=0.99
R=0.95

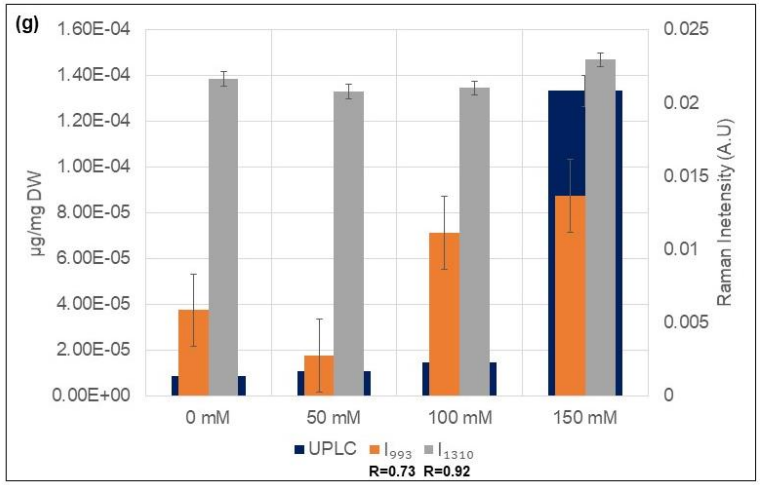

(b)
1.20E-04
1.00E-04
8.00E-05
6.00E-05
4.00E-05
2.00E-05
0.00E+00
µg/mg DW
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1298
I1310
R=0.91
R=0.92

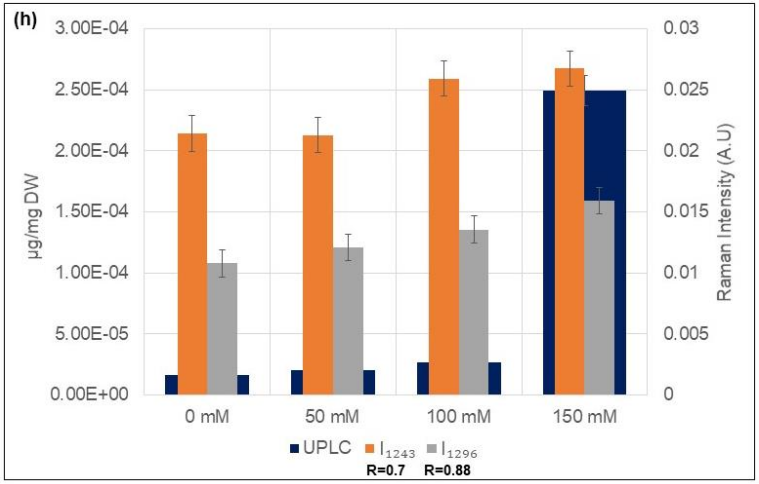

(c)
4.00E-04
3.50E-04
3.00E-04
2.50E-04
2.00E-04
1.50E-04
1.00E-04
5.00E-05
0.00E+00
µg/mg DW
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I992
I1251
R=0.72
R=0.67

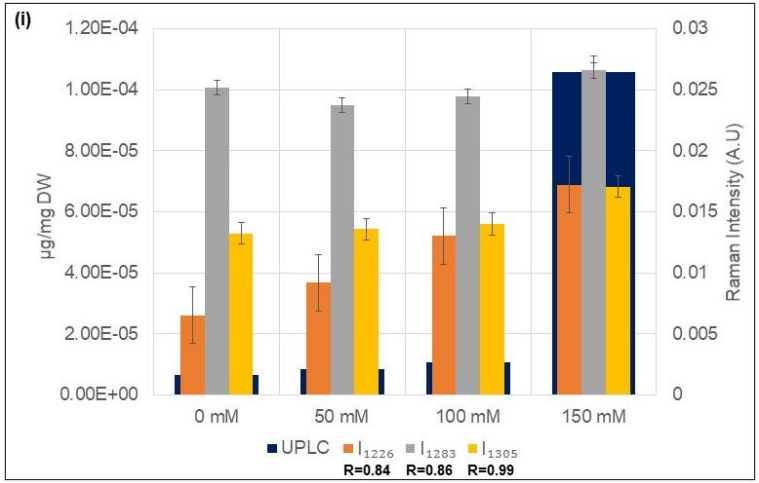

(d)
4.50E-04
4.00E-04
3.50E-04
3.00E-04
2.50E-04
2.00E-04
1.50E-04
1.00E-04
5.00E-05
0.00E+00
µg/mg DW
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1287
I1293
R=0.96
R=0.96

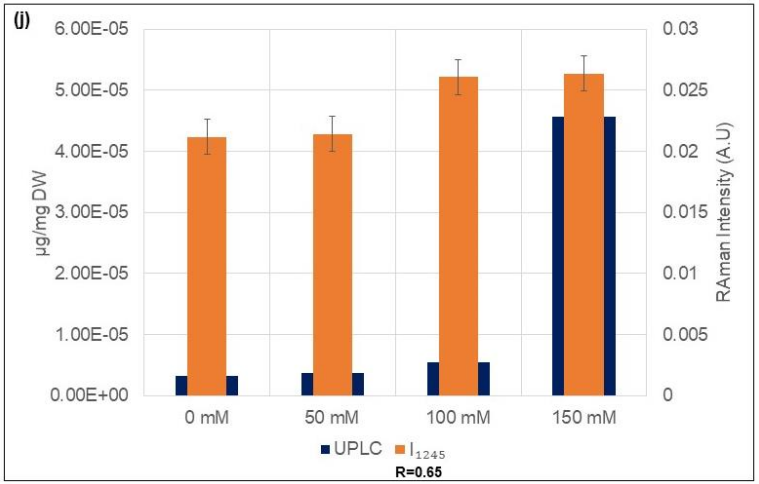

(e)
6.00E-04
5.00E-04
4.00E-04
3.00E-04
2.00E-04
1.00E-04
0.00E+00
µg/mg DW
0.02
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I894
I1107
R=0.64 R=0.62

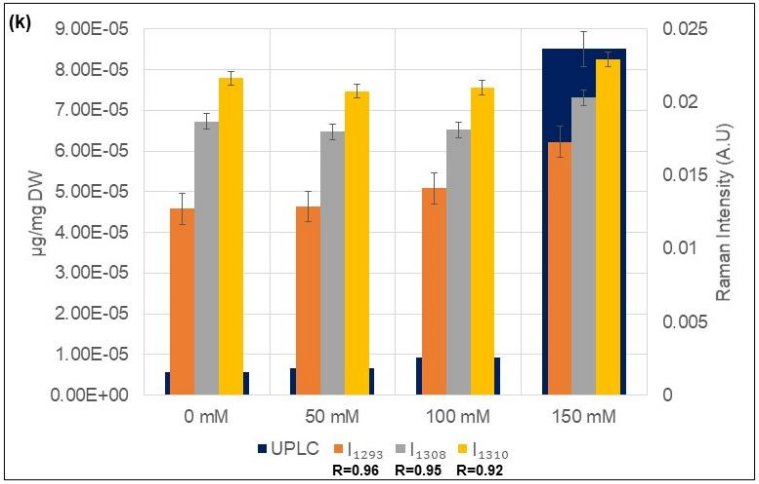

(f)
1.40E-04
1.20E-04
1.00E-04
8.00E-05
6.00E-05
4.00E-05
2.00E-05
0.00E+00
µg/mg DW
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1224
I1225
R=0.82 R=0.83

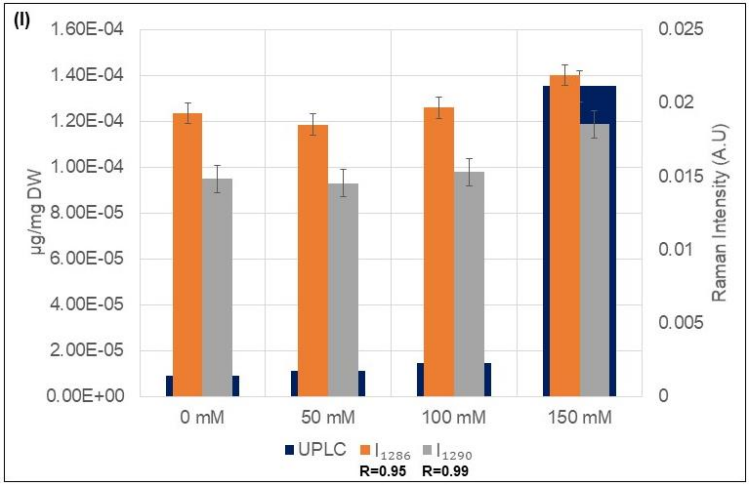

(g)
1.60E-04
1.40E-04
1.20E-04
1.00E-04
8.00E-05
6.00E-05
4.00E-05
2.00E-05
0.00E+00
µg/mg DW
0.025
0.02
0.015
0.01
0.005
0
Raman Inetensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I993
I1310
R=0.73 R=0.92


(h)
3.00E-04
2.50E-04
2.00E-04
1.50E-04
1.00E-04
5.00E-05
0.00E+00
µg/mg DW
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1243
I1296
R=0.7 R=0.88


(i)
1.20E-04
1.00E-04
8.00E-05
6.00E-05
4.00E-05
2.00E-05
0.00E+00
µg/mg DW
0.03
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1226
I1283
I1305
R=0.84
R=0.86
R=0.99


(j)
6.00E-05
5.00E-05
4.00E-05
3.00E-05
2.00E-05
1.00E-05
0.00E+00
µg/mg DW
0.03
0.025
0.02
0.015
0.01
0.005
0
RAman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1245
R=0.65


(k)
9.00E-05
8.00E-05
7.00E-05
6.00E-05
5.00E-05
4.00E-05
3.00E-05
2.00E-05
1.00E-05
0.00E+00
µg/mg DW
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1293
I1308
I1310
R=0.96
R=0.95
R=0.92


(l)
1.60E-04
1.40E-04
1.20E-04
1.00E-04
8.00E-05
6.00E-05
4.00E-05
2.00E-05
0.00E+00
µg/mg DW
0.025
0.02
0.015
0.01
0.005
0
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
UPLC
I1286
I1290
R=0.95
R=0.99

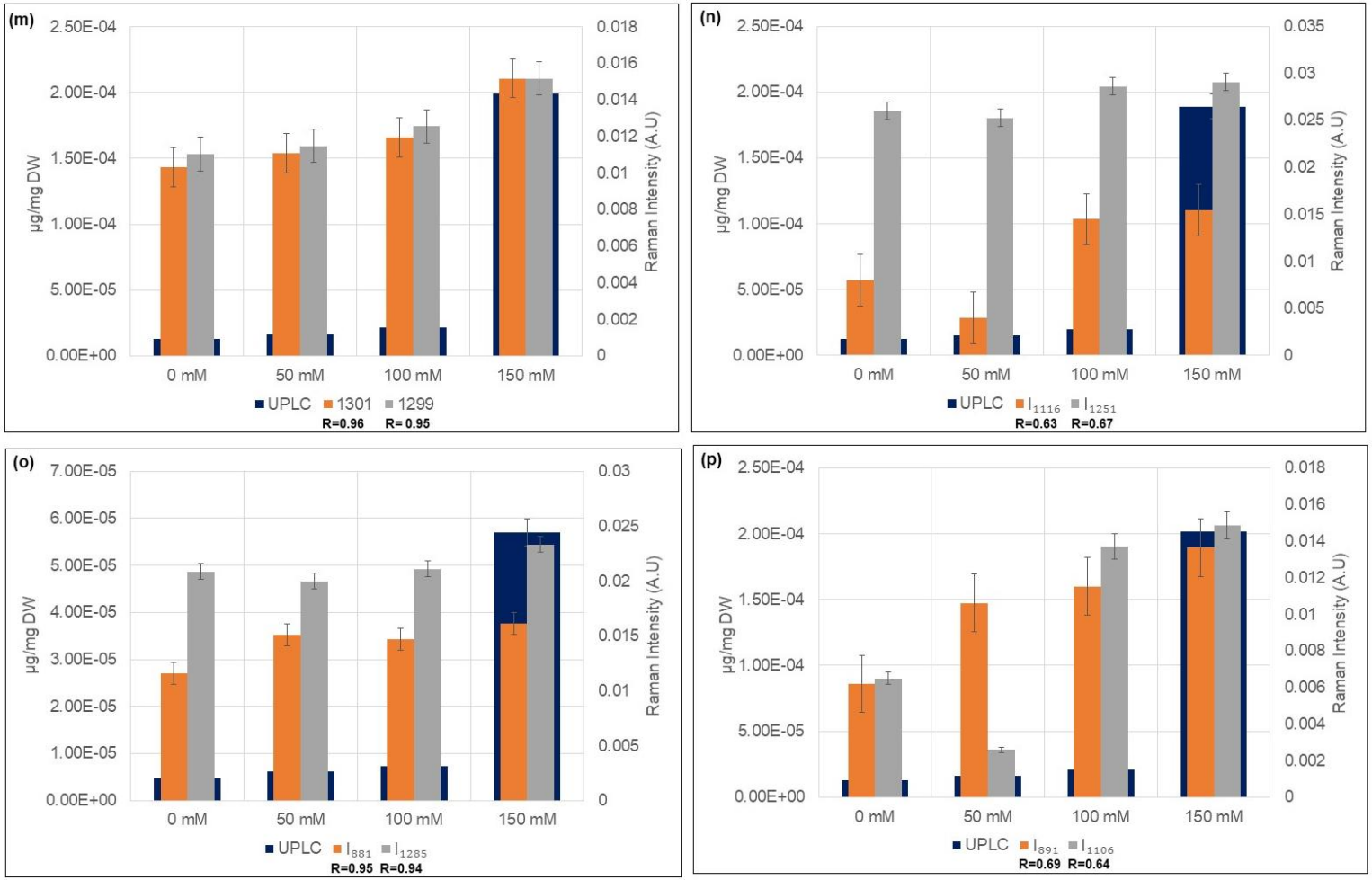


Figure D. 11. Correlation of Raman data with UPLC analysis of amino acids in the NaCl study. (a) Alanine, (b) Arginine, (c) Aspartate/Asparagine, (d) Glutamate, (e) Glycine, (f) Histidine, (g) Isoleucine, (h) Leucine, (i) Lysine, (j) Methionine, (k) Phenylalanine, (l) Proline, (m) Serine, (n) Threonine, (o) Tyrosine, (p) Valine. Correlation coefficients (R) for each Raman band are represented.

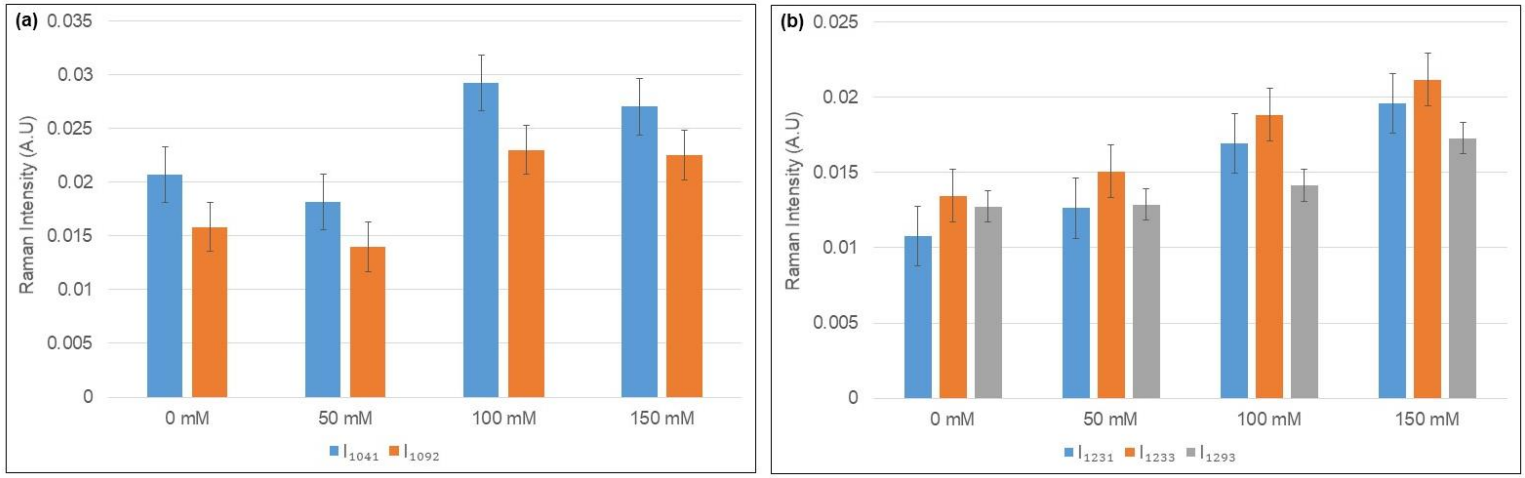


Figure D. 12. Acetate study: Raman bands predictions of amino acids non-resolved with UPLC. (a) Cysteine predictions, (b) Tryptophan predictions.

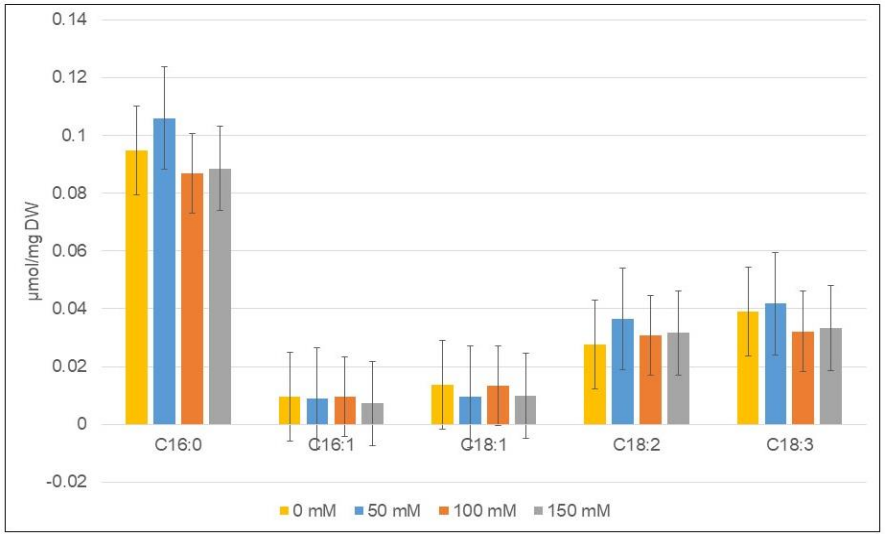
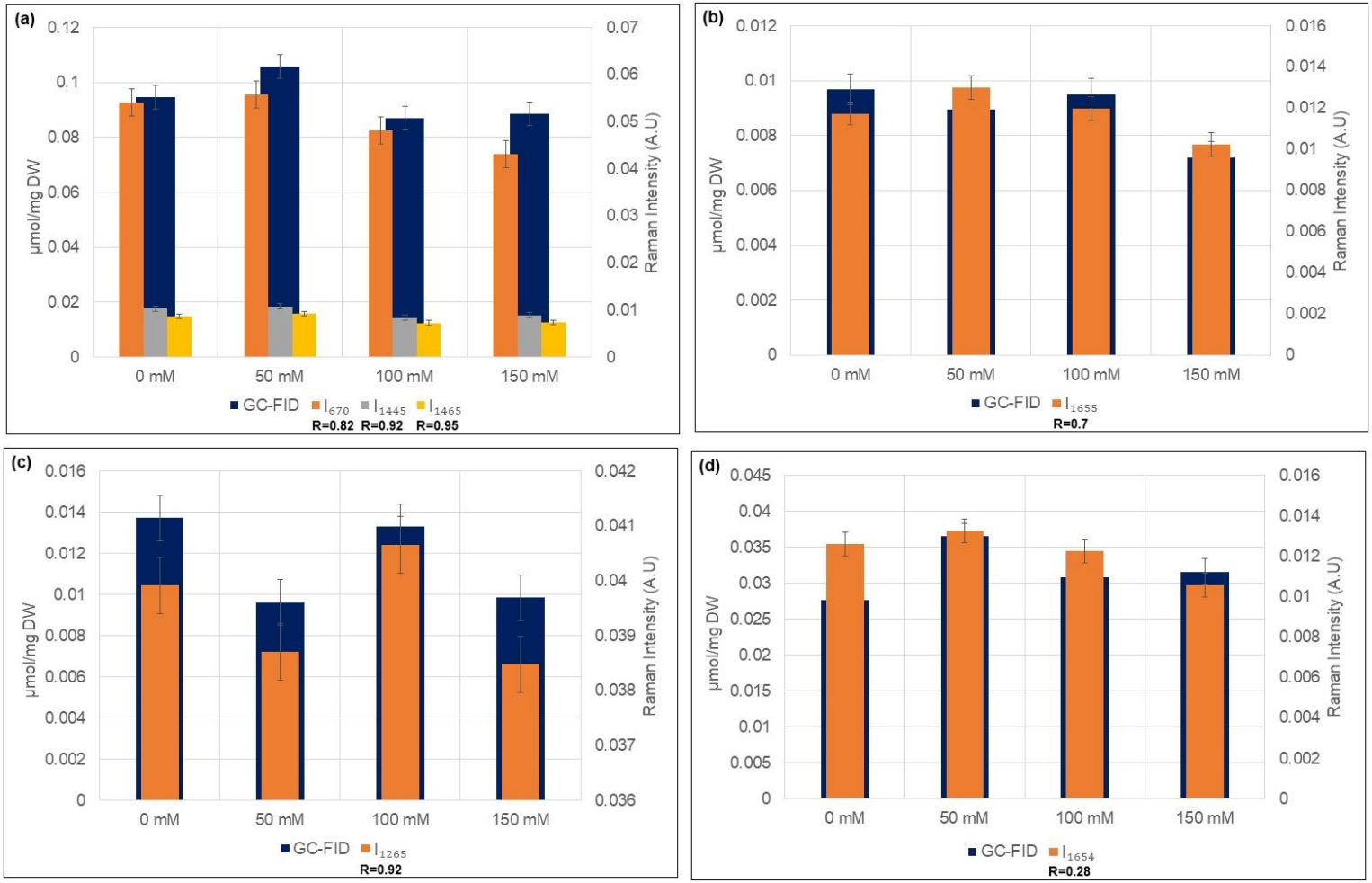


Figure D. 13. Levels of fatty acids for NaCl study. Data represent means ± standard deviation (SD) of values from three independent experiments.


(a)
µmol/mg DW
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
GC-FID
I670
I1445
I1465
R=0.82 R=0.92 R=0.95
(b)
µmol/mg DW
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
GC-FID
I1655
R=0.7
(c)
µmol/mg DW
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
GC-FID
I1265
R=0.92
(d)
µmol/mg DW
Raman Intensity (A.U)
0 mM
50 mM
100 mM
150 mM
GC-FID
I1654
R=0.28

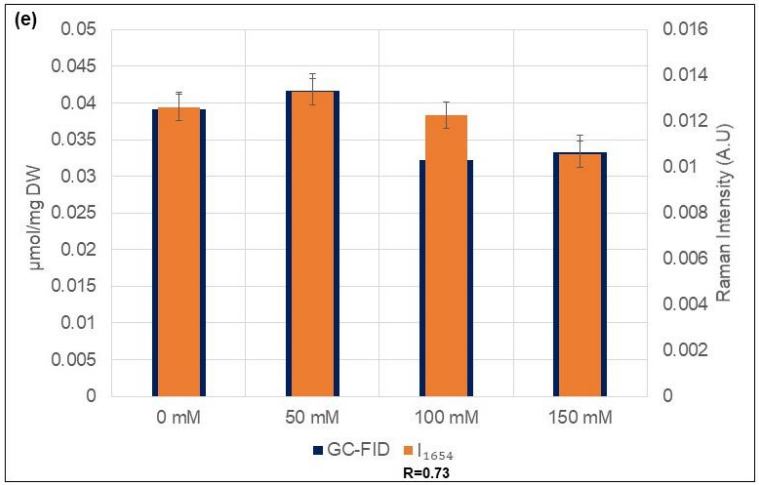


Figure D. 14. Correlation between Raman spectroscopy and GC-FID data of fatty acids for NaCl study. (a) Palmitic acid, (b) Palmitoleic acid, (c) Oleic acid (d) Linoleic acid, (e) Linolenic acid. Correlation coefficients (R) for each Raman band are represented.

Figure D. 15. The fabricated Teflon hydrolysis chamber.

Table D. 13. All Raman bands cited and tested for amino acids and fatty acids.

Acetate Study		
Biomolecules	**Raman bands (cm-1) and Correlation Coefficient**	**References**
Ala	480 (R = 0.45), 653 (R = 0.43)	(8,9)
Arg	1713 (R = 0.87)	(8)
Asp/Asn	1695 (R = 0.97)	(9)
Cys*	455, 499, 542	(9)
Glu/Gln	1462 (R = 0.87), 1624 (R = 0.73)	(8,9)
Gly	496 (R = 0.71), 1568 (R = 0.73), 1674 (R = 0.84)	(9)
His	422 (R = 0.79)	(8)
Ile	557 (R = 0.63)	(9)
Leu	1583 (R = 0.49)	(9)
Lys	1570 (R = 0.5)	(9)
Met	645 (R = 0.59)	(9)
Phe	468 (R = 0.70), 469 (R = 0.72), 0.74 (R = 605), 606 (R = 0.73)	(8,9)
Pro	452 (R = 0.46)	(8)
Ser	610 (R = 0.63), 611 (R = 0.61)	(8,9)
Thr	418 (R = 0.81), 447 (R = 0.73)	(9)
Trp*	425, 456, 498, 509, 509, 574	(8)
Tyr	527 (R = 0.76), 641 (R = 0.76)	(8,9)
Val	1567 (R = 0.66), 1660 (R = 0.64)	(8)
Palmitic acid (C16:0)	670 (R = 0.92)	(8)
Palmitoleic acid (C16:1)	1655 (R = 0.46)	(30)
Oleic acid (C18:1)	1440 (R = 0.82)	(8)
Linoleic acid (C18:2)	1645 (R = 0.67)	(30)
Linolenic acid (C18:3)	1645 (R = 0.68)	(30)
NaCl Study		
Biomolecules	**Raman bands (cm-1) and Correlation Coefficient**	**References**
Ala	1305 (R = 0.99), 1308 (R = 0.95)	(8,9)
Arg	1298 (R = 0.91), 1310 (R = 0.92)	(8,9)

Asp/Asn	992 (R = 0.72), 1251 (R = 0.67)	(9)
Cys*	1041, 1092	(9)
Glu/Gln	1287 (R = 0.96), 1293 (R = 0.96)	(8,9)
Gly	894 (R =0.64), 1107 (R = 0.62)	(8,9)
His	1224 (R = 0.82), 1225 (R = 0.83)	(8,9)
Ile	993 (R = 0.73), 1310 (R = 0.92)	(9)
Leu	1243 (R = 0.7), 1296 (R = 0.88)	(9)
Lys	1226 (R = 0.84), 1283 (R = 0.86), 1305 (R = 0.99)	(9)
Met	1245 (R = 0.65)	(9)
Phe	1293 (R = 0.96), 1308 (R = 0.95), 1310 (R = 0.92)	(8,9)
Pro	1286 (R = 0.95), 1290 (R = 0.99)	(8)
Ser	1299 (R = 0.95), 1301 (R = 0.96)	(8,9)
Thr	1116 (R = 0.63), 1251 (R = 0.67)	(9)
Trp*	1231, 1233, 1293	(8,9)
Tyr	881 (R = 0.95), 1285 (R = 0.94)	(9)
Val	891 (R = 0.69), 1106 (R = 0.64)	(8)
Palmitic acid (C16:0)	670 (R = 0.82), 1445 (R = 0.92), 1465 (R = 0.95)	(8)
Palmitoleic acid (C16:1)	1655 (R = 0.7)	(30)
Oleic acid (C18:1)	1265 (R = 0.92)	(8)
Linoleic acid (C18:2)	1654 (R = 0.28)	(30)
Linolenic acid (C18:3)	1654 (R = 0.73)	(30)

* Values for Cys and Trp could not be obtained by the UPLC method used in this research. Only Raman predictions are provided.
** Possible overlapping bands between different biomolecules assigned Raman bands

References

Note: All references are reviewed and band assignments summarized in reference (31), except references 8, 9 and 30.

1. Stone N, Kendall C, Smith J, Crow P, Barr H. Raman spectroscopy for identification of epithelial cancers. Faraday Discuss. 2004;126:141–57; discussion 169-83.
2. Krafft C, Neudert L, Simat T, Salzer R. Near infrared Raman spectra of human brain lipids. Spectrochim Acta Mol Biomol Spectrosc. 2005;61(7):1529–35.
3. Mahadevan-Jansen A, Richards-Kortum R. Raman spectroscopy for cancer detection: a review. In 1997. p. 2722–8 vol.6.
4. Puppels GJ, Garritsen HSP, Kummer JA, Greve J. Carotenoids located in human lymphocyte subpopulations and natural-killer-cells by Raman microspectroscopy. Cytometry. 1993;14(3):251–6.
5. Silveira L Jr, Sathaiah S, Zangaro RA, Pacheco MT, Chavantes MC, Pasqualucci CA. Correlation between near-infrared Raman spectroscopy and the histopathological analysis of atherosclerosis in human coronary arteries. Lasers Surg Med. 2002;30(4):290–7.
6. Huang Z, McWilliams A, Lui H, McLean DI, Lam S, Zeng H. Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. Int J Cancer. 2003;107(6):1047–52.
7. Cheng WT, Liu MT, Liu HN, Lin SY. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. Microsc Res Tech. 2005;68(2):75–9.
8. De Gelder J, De Gussem K, Vandenabeele P, Moens L. Reference database of Raman spectra of biological molecules. J Raman Spectrosc. 2007;38(9):1133–47.
9. Zhu G, Zhu X, Fan Q, Wan X. Raman spectra of amino acids and their aqueous solutions. Spectrochim Acta Mol Biomol Spectrosc. 2011;78(3):1187–95.
10. Chan JW, Taylor DS, Zwerdling T, Lane SM, Ihara K, Huser T. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. Biophys J. 2006;90(2):648–56.
11. Schulz H, Baranska M. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. Vib Spectrosc. 2007;43(1):13–25.
12. Farquharson S, Shende C, Inscore FE, Maksymiuk P, Gift A. Analysis of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy. J Raman Spectrosc. 2005;36(3):208–12.
13. Dukor RK. Vibrational Spectroscopy in the Detection of Cancer. Biomed Appl. 2002;5:3335–3359.
14. Hanlon EB, Manoharan R, Koo TW, Shafer KE, Motz JT, Fitzmaurice M, et al. Prospects for in vivo Raman spectroscopy. Phys Med Biol. 2000;45(2):R1-59.
15. Koljenović S, Schut TB, Vincent A, Kros JM, Puppels GJ. Detection of Meningioma in Dura Mater by Raman Spectroscopy. Anal Chem. 2005;77(24):7958–65.
16. Malini R, Venkatakrishna K, Kurien J, Pai KM, Rao L, Kartha VB, et al. Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study. Biopolymers. 2006;81(3):179–93.
17. Katainen E, Elomaa M, Laakkonen UM, Sippola E, Niemela P, Suhonen J, et al. Quantification of the amphetamine content in seized street samples by Raman spectroscopy. J Forensic Sci. 2007;52(1):88–92.
18. Viehoever AR, Anderson D, Jansen D, Mahadevan-Jansen A. Organotypic raft cultures as an effective in vitro tool for understanding Raman spectral analysis of tissue. Photochem

Photobiol. 2003;78(5):517–24.
19. Shetty G, Kendall C, Shepherd N, Stone N, Barr H. Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. Br J Cancer. 2006;94(10):1460–4.
20. Stone N, Kendall C, Shepherd N, Crow P, Barr H. Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. J Raman Spectrosc. 2002;33(7):564–73.
21. Notingher I, Green C, Dyer C, Perkins E, Hopkins N, Lindsay C, et al. Discrimination between ricin and sulphur mustard toxicity in vitro using Raman spectroscopy. J R Soc Interface. 2004;1(1):79–90.
22. Frank CJ, McCreery RL, Redd DC. Raman spectroscopy of normal and diseased human breast tissues. Anal Chem. 1995;67(5):777–83.
23. Chiriboga L, Xie P, Yee H, Vigorita V, Zarou D, Zakim D, et al. Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in the human cervix. Biospectroscopy. 1998;4(1):47–53.
24. Andrus PG, Strickland RD. Cancer grading by Fourier transform infrared spectroscopy. Biospectroscopy. 1998;4(1):37–46.
25. Naumann D. Infrared and NIR Raman spectroscopy in medical microbiology. Proc SPIE. 1998;3257:245–257.
26. Utzinger U, Heintzelman DL, Mahadevan-Jansen A, Malpica A, Follen M, Richards-Kortum R. Near-Infrared Raman Spectroscopy for in vivo Detection of Cervical Precancers. Appl Spectrosc. 2001;55(8):955–9.
27. Gniadecka M, Wulf HC, Nymark Mortensen N, Faurskov Nielsen O, Christensen DH. Diagnosis of Basal Cell Carcinoma by Raman Spectroscopy. J Raman Spectrosc. 1997;28(2-3):125–9.
28. Ruiz-Chica AJ, Medina MA, Sánchez-Jiménez F, Ramírez FJ. Characterization by Raman spectroscopy of conformational changes on guanine–cytosine and adenine–thymine oligonucleotides induced by aminooxy analogues of spermidine. J Raman Spectrosc. 2004;35(2):93–100.
29. Ó Faoláin E, Hunter MB, Byrne JM, Kelehan P, McNamara M, Byrne HJ, et al. A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. Vib Spectrosc. 2005;38(1):121–7.
30. Czamara K, Majzner K, Pacia MZ, Kochan K, Kaczor A, Baranska M. Raman spectroscopy of lipids: a review. J Raman Spectrosc. 2015;46(1):4–20.
31. Movasaghi Z, Rehman S, Rehman IU. Raman Spectroscopy of Biological Tissues. Appl Spectrosc Rev. 2007;42(5):493–541.