Rational Strain Design for Value Added Products: A Systems Metabolic Engineering and Synthetic Biology Approach

Thesis submitted to AcSIR
For the Award of the Degree of

DOCTOR OF PHILOSOPHY
In
Biological Sciences

By
Mayooreshwar P. Rajankar
10BB12J26032

Under the guidance of
Dr. Anu Raghunathan
(Research Supervisor)

Chemical Engineering Division
CSIR-National Chemical Laboratory
Pune-411008, India

July 2018

सीएसआयआर–राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग, पुणे – 411 008. भारत
CSIR-NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411008. India

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled **Rational Strain Design for Value Added Products: A Systems Metabolic Engineering and Synthetic Biology Approach** submitted by **Mr. Mayooreshwar P. Rajankar** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor Of Philosophy**, embodies original research work under my supervision. I further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

Mayoorehwar P. Rajankar
(Student)

Dr. Anu Raghunathan
(PhD Supervisor)
Principal Scientist
Chemical Engineering Division
CSIR- National Chemical Laboratory
Pune-411008

Place: Pune
Date: 24/01/2019

Communication Channels
☎
NCL Level DID : 2590
NCL Board No. : +91-20-25902000
EPABX : +91-20-25893300
: +91-20-25893400
FAX
Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
COS&P's Office : +91-20-25902664
WEBSITE
www.ncl-india.org

Declaration of Authorship

I hereby declare that the PhD thesis entitled **Rational Strain Design for Value Added Product: a Systems Metabolic Engineering and Synthetic Biology Approach** was completely carried out by me and for the degree of **Doctor of Philosophy** in **Biological Sciences** under the guidance and the supervision of **Dr. Anu Raghunathan**, Principal Scientist, CSIR-National Chemical Laboratory, Pune, India.

I confirm that this thesis research is my own work while in candidature for the research degree at this institution and the content of this thesis is original. I also affirm that no part of this thesis is previously been submitted for a degree or any other qualification at this institution or any other institution. Also the interpretation put forth are based solely on my knowledge and understanding of original research articles and all resources have been duly cited and acknowledged as and when appropriate.

Mayooreshwar P. Rajankar
(PhD Student)
Metabolic Inquiry and Cellular Engineering (MICE) Lab
CEPD Division CSIR-National Chemical Laboratory
Pune-411008

Place: Pune
Date: 24/01/2019

Dedicated to

Family and Friends

Acknowledgement

The work in this dissertation was carried out in Chemical Engineering division (CEPD), CSIR-National Chemical Laboratory (NCL), Pune, India. Research funding from CSIR-NCL and fellowship from University Grant Commission (UGC) is gratefully acknowledged. I am also thankful to Department of Biotechnology (DBT), Department of Science and Technology (DST) for providing fund through research grant to Metabolic Inquiry and Cellular Engineering (MICE) Lab. I am also thankful to University of Toronto Impact centre for funding to attend Connaught Summer Institute on Synthetic Biology 2014. I also owe a debt of gratitude to Director of NCL and CEPD head of division, for allowing me to complete my research work in the premises and avail facilities offered by NCL.

Completion of this dissertation was possible with the support of several people. I would like to express my sincere gratitude to all of them. First and foremost, my sincerest thanks are extended to my research guide Dr. Anu Raghunathan for her valuable guidance, scholarly inputs and consistent encouragement received throughout the research work.

I would also like to thank the Doctoral Advisory Committee members, Dr. G J Sanjayan, Dr. Venkat Panchagnula, and Dr. Mugdha Gadgil who contributed immensely during my journey throughout this research by addressing the shortcomings. Similar, profound gratitude goes to the collaborators Dr. P R Rajmohanan and Dr. Sapna Ravindranathan.

I extend my gratitude towards my fellow colleagues Rupa, Deepanwita, Vishwanath, Jyoti Rawat and Nimisha Parekh for being very kind and patient and always being supportive. I gained a lot from them through their personal and scholarly interactions, their suggestions at various points of my research program. Rupa is a constant positive support and extremely patient helping hand. I appreciate the interest shown by Deepa, in updating me with the latest publications and troubleshooting while data analysis. I am also grateful to the members from Biochemical Engineering Lab Dr. Ramachandra Gadre, Dr. Rahul Bhambure, Deepa, Kayanat, Aatir, Sayli, Prashant and Manoj for helping with experimental work and allowing me to access their facility every now and then.

I am also hugely appreciative to Dr. Deepak Saini, Dr.Ruchi Agarwal and Dr. Ruchi Jain from IISc Bangalore for my initial training and introduction to field of scientific research.

I would like to thank all administrative staff from NCL, especially from C&P and SAO section and Mr. Raheja from CEPD section, who helped me with all diligence, for promptly completing any official formalities required, throughout my PhD.

I would also like to take this special opportunity to thank Dr. Jyoti Kawalekar, and Prof. R S Maben RLS Institute Belgavi, during my graduation. I also extend my sincere thanks towards Dr Pramod B. Gai, Vice Chancellor Karanataka University, Dharwad. Their encouragement and passionate teaching allowed me to keep open mind for all these scientific endeavors so far, since that young age.

It is my pleasure to thank my friends from NCL Avinash Ghanate, Bharat, Dharmesh, Meghana, Sandeep, Ranjeet, Shraddha, Pooja Sucheta, Anushua, Gayatri and Nivedita for the wonderful times we shared especially during the evening snacks and gave me necessary distractions from my research and made this journey a memorable one.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. I am grateful to my wife (Seema) and sister (Bhakti) for always being there for me as a friend, and my friends Priyanka Gangale, Rekha, Ravi Mengane, Pavan, Kamaal Hassan and Suresh Khot for encouraging and supporting me throughout this journey. I am forever indebted to my parents for giving me the opportunities and experiences that have made me who I am. They selflessly encouraged me to explore new directions in life and seek my own destiny. This journey would not have been possible if not for them, and I dedicate this milestone to them.

Table of Content

Certificate……………………………………………………………………………...iii

Declaration…………………………………………………………………………....v

Acknowledgement…………………………………………………………………..ix

Table of Content……………………………………………………………….....xiii

Abbreviations…………………………………………………………….........xvi

List of figures…………………………………………………………………....xvii

List of Tables…………………………………………………………………….xix

Thesis Abstract……………………………………………………………………….1

Chapter 1 Introduction **... 9**

1.1 Motivation**.. 10**

1.2 Overall objective and specific aims **... 11**

1.3 Literature Review **.. 11**

1.4 Synthetic biology **.. 18**

1.5 Bioinformatics and Abstraction**... 21**

1.6 Systems Biology for Metabolic Engineering**.. 22**

1.7 Developing experimental hosts for metabolic engineering **....................... 26**

1.8 Experimental technologies for manipulation **.. 27**

1.9 Violacaein – a pigment and a potential drug molecule **............................ 28**

1.10 The Polyhydroxyalkanoates and Bioplastics **... 29**

1.11 Examples of Synthetic Biology and Metabolic Engineering**.................... 30**

1.12 Challenges for Synthetic Biology and Metabolic Engineering**................ 31**

1.13 Security and Risk **... 32**

1.14 Overview of the Thesis **... 32**

1.15 Thesis Organization 33

Chapter 2 Methodology: Design, Build, Test, and Analyze 35

2.1 Design: Operon, Genetic Circuit and Pathway Design 36

2.2 Pathway Design for Violacein and PHB production 42

2.3 Build: Gene Synthesis, Cloning and Strain design 43

2.4 Test: Qualitative and Quantitative Estimation 47

2.5 Analyze: Genome scale model driven analysis of engineered strains 65

Chapter 3 Synthetic and Systems Metabolic Engineering for Violacein production 71

3.1 Introduction 72

3.2 Growth optimality analysis of Violacein pathway by Flux Balance Modeling 73

3.3 Synthetic operon optimization and construction 80

3.4 Violacein production increased through metabolic engineering of E. coli 84

3.5 Conclusion 86

Chapter 4 Rational Strain Design for Poly(R)-3-Hydroxybutyrate biosynthesis 89

4.1 Introduction 90

4.2 Genetic circuit design 91

4.3 PHB production via metabolic engineering of E. coli 93

4.4 Substrate utilization and PHB production 94

4.5 Confirmation of PHA producing cells and Carbonosomes using microscopy 97

4.6 Confirmation of PHB by NMR spectroscopy 98

4.7 Genome Integration of the PHB circuit into E. coli 100

4.8 Confirmation of PHB produced in the genome integrated E. coli strain 102

4.9 Growth and PHB production in the genome integrated PHB producing E. coli strain 102

4.10 Insilico analysis of PHB producing strain 104

4.11 Conclusions *106*

Chapter 5 Conclusion and Future scope *109*

5.1 Recapitulation *110*

5.2 Future of Metabolic Engineering *111*

Appendix I *113*

Bibliography *119*

List of publication *135*

Abbreviations

ΔG: Free energy;

Δpgi: pgi E. coli knockout strain

ΔpheA: pheA E. coli knockout strain

ΔtrpR: trpR E. coli knockout strain

CAI: Codon Adaptive Index

COBRA: Constraints Based Reconstruction And Analysis

CYTBO3_4pp: Periplasmic cytochrome oxidase bo3

EDD: 6-phosphogluconate dehydratase

ENO: Enolase

FBA: Flux Balance Analysis

GAPD: Glyceraldehyde-3-Phosphate Dehydrogenase

GUR: Glucose Uptake Rate

NADPH: Nicotinamide Adenine Dinucleotide Phosphate

Nc: Effective Codon Usage

PCR: Polymerase Chain Reaction

PhPPs: Phenotypic Phase Planes

PPP: Pentose Phosphate Pathway

RBS: Ribosome Binding Site

SYNO: Synthetic Operon

SP: Shadow Price

TC: Translation Coupling Factor

TCA: Citric Acid Cycle

TIR: Translation Initiation Rates

WTO: Wild-Type Operon

PHA: Polyhydroxyalkanoate

PHB: Polyhydroxybutyrate

NMR: Nuclear Magnetic Reasonance

COSY: Correlation Spectroscopy

List of Figures

Figure 1.1 Synthesis of value added products using Systems and Synthetic Biology *9*

Figure 1.2 Chemical sysnthesis of oligonucleotide using phosphorhamadite method *14*

Figure 1.3 Assembly methods for gene synthesis *16*

Figure 2.1 Synthetic Biology Paradigms *35*

Figure 2.2 Design of synthetic vioABCDE operon (SYNO) and gene sequence analysis . *41*

Figure 2.3 hairpin loop around the mRNA initiation codon (AUG) of the synthetic operon SYNO *42*

Figure2.4 Pathway design for (A) Violacein and (B) PHB biosynthetic pathway *43*

Figure 2.5 Establishing standard curve for estimation of intracellular violacein *48*

Figure 2.6 Pictorial representations of workflows for improved accuracy of quantitation of PHB *51*

Figure 2.7 Excitation and Emission scan for standards *53*

Figure 2.8 Development of Analytical Method for PHB *57*

Figure 2.9 PHB Quantitation from recombinant E. coli *63*

Figure 2.10 : Genome Scale model and flux balance analysis *67*

Figure 3.1 Synergies among synthetic biology, systems biology and metabolic engineering in the drug Violacein production *71*

Figure 3.2 Overall work flow for integration of constraint based modeling for experimental rational strain design. *73*

Figure 3.3 Phenotypic Phase Plane (PhPP) analysis for violacein biosynthesis *75*

Figure 3.4 Shadow price from phenotypic phase plane (PhPP) using in silico strain for NADPH *76*

Figure 3.5 Robustness analysis for in silico strain for sensitivity *77*

Figure 3.6 Structure of wildtype (pCVW) and synthetic (pSYN) plasmid *80*

Figure3.7 Tryptophan levels from MALDI-Analysis 83

Figure 3.8 Growth and Violacein analysis of genome engineered E. coli strains 84

Figure 3.9 Violacein yield variation, Comparison and Range 85

Figure 4.1 Genetic circuit design 91

Figure 4.2 Sequence optimization for PHB_Wt and PHB_Syn genetic circuit 92

Figure 4.3 Glucose titer for to determine the optimum glucose concentration in LB broth 93

Figure 4.4 Arabinose titration in LB media supplemented with 2 mg/ml glucose 94

Figure 4.5 Growth and Uptake rate for pPHB_Syn and pPHB_Wt transformed E. coli strains 95

Figure 4.6 PHB yields in E. coli strains transformed with plasmid A) pPHB_Wt and pPHB_Syn 96

Figure 4.7 Fluorescence microscopy of PHB producing E.coli transformed with pPHB_Syn 97

Figure 4.8 1H NMR spectrum for polymer extracted from PHB Producing E.coli 98

Figure 4.9 NMR Spectroscopy for structural confirmation 99

Figure 4.10 Genome Integration of PHB_Syn operon in E.coli genome by Tn7 Transposon 100

Figure 4.11 Fluorescence microscopy for PHB producing genome integrated (PHB_GISyn) 101

Figure 4.12 Genome Integrated strain SDS-PAGE Expression of Pctap and PhaC in induced cel 101

Figure 4.13 ^{1}*H NMR spectrum for polymer extracted from E.coli transformed 102*

Figure 4.14 Growth and substrate utilization for PHB_GISyn on LB media supplemented with glucose and glycerol 103

Figure5.1 Synthetic biology paradigm culminating into cell factories; End of the loop 109

List of Tables

Table 2.1 Codons preferred in E. coli for high expression of genes 38

Table 2.2 Performance characteristic of Nile red based PHB estimation methods using fluorescence spectroscopy 59

Table 2.3 Performance characteristics of Nile red based PHB quantitation in PHB producing E.coli cells compared to U V Spectrophotometery... 60

Table 2.4 Non-sonicated granules physical properties 60

Table 2.5 Sonicated granules physical properties 61

Table 2.6 Nile red stained cells in glycine-HCl buffer physical properties....................... 61

Table 2.7 ANOVA sonicated, ethanolic extract,spectrofluorometry................................ 62

Table 2.8 ANOVA non-sonicated, ethanolic extract, spectrofluorometry 62

Table 2.9 ANOVA sonicated spectrofluorometry 62

Table 2.10 ANOVA UV Spectrophotometry.. 62

Table 3.1 Predicting yields of biomass and violacein using iAF1260vio.......................... 74

Table 3.2 Shadow price (SP) analysis for optimal growth and Violacein in E. coli 78

Table 3.3 Predictions using biomass as the objective function and constraining Violacein secretion 79

Table 3.4 Reduced cost analysis for the optimal synthesis of violacein using in silico strains.. 81

Table 3.5 Strains used in this study 82

Table 3.6 Growth rates of engineered E. coli strains 82

Table 3.7 Growth rates of engineered strains on M9 media 83

Table 3.8 Summary of maximum violacein yields by engineered strains 85

Table 4.1 Growth rate, Substrate uptake rate and PHB yields for plasmid transformed E. coli Strain on Glycerol.. 96

Table 4.2 Growth rate and Substrate uptake rate for plasmid transformed E. coli Strain on Glucose 96

Table 4.3 Growth rate, Substrate uptake rate and PHB yields for PHB_GISyn 103

Table 4.4 Reactions added in iJO1366 GSM for simulating for PHB production for experimental growth and substrate uptake rate..104

Table 4.5 Constraints to simulate LB media conditions to derived from reduced cost analysis 105

Table 4.6 Constraints used to simulate Induced and Un-induced condition..................105

Table 4.7 FBA prediction after applying experimental constraints 106

Thesis Abstract

The grand challenge of metabolic engineering lies in the complexity and redundancy of cellular pathways and the evolutionary drive of a cell to maximize growth rather than a forced bioengineering objective. Engineering microorganisms to thus produce value added products from bulk chemicals as carbon source is now greatly accelerated by use of Synthetic Biology. The fast forwarding evolution has thus uncapped the limits of engineering biological systems. Rational strain design for production of value added products requires channeling of basic substrate molecules towards a desirable metabolic output to make products of interest. When complex pathways are introduced inside the cell, limitations including intermediate toxicity, low enzyme activity, metabolic burden (cofactor imbalance etc.) need to be overcome for high performance. Such bottlenecks can be addressed using pathway engineering that exploits the synergies of synthetic biology, metabolic engineering and systems biology. Successful metabolic engineering for platform cell factories to produce a wide range of fuels and chemicals necessitates identifying the sensitivity of product/process to nutrient precursors and cofactors by coupling of cellular objectives of growth and energy to desired bioengineering objectives. This thesis explores the application of these principles to develop scalable systems to make a drug molecule violacein and a biopolymer Polyhydroxyalkanoates (PHAs). Violacein is a bacterial bis-indole pigment of commercial interest having antibacterial, antitumoral, antiviral, trypanocidal and antiprotozoan properties. PHAs form a class of natural polyesters, commonly referred to as bioplastics, that many organisms accumulate as intracellular granules to store carbon and reducing equivalents in response to specific environmental conditions. This thesis discusses the rational strain design and development in the context of systems metabolic engineering and synthetic biology for violacein and polyhydroxybutyrate.

Chapter 1: Introduction

The strategy for strain design depends on chemical nature of the molecule of interest and its pathway design(Chubukov et al., 2016). Unlike recombinant expression of proteins, in metabolite production more than one enzyme; coenzyme, ATP and cofactors are involved. Moreover the carbon from energy metabolism has to be diverted towards product biosynthesis. The pathway involved may not be linear and may require augmenting the missing nodes in host metabolism. To understand the intricacies of strain designing, *Escherichia coli* were engineered to produce two molecules Violacein and Poly(R)-3-Hydroxybutyrate (PHB). Pathway for Violacein biosynthesis is a typical example of linear pathway which requires two tryptophan molecules as precursor. The pathway involves five genes *vioA, vioB, vioC, vioD* and *vioE,* where two of reaction requires reducing capabilities of NADPH. PHB biosynthesis is an example of a non linear pathway where the precursor, (R)-3-Hydroxybutyrate (3HB) is generated from many different routes and an incomplete pathway for polymer biosynthesis. Therefore it required a careful augmentation of two genes Propionyl-CoA Transferase (pct_{ap}) from *Acetobacter pasteurianus* that transfers CoA group to acetate generated due to carbtree effect to give acetyl-CoA. Acetyl-CoA is the precursor for the incomplete PHB biosysnthesis pathway and PHA polymerase ($phaC_{cv}$) from *Chromobacterium violaceium* that polymerizes 3HB into PHB. The gene sequences (coding for synthetic operons) were optimized for stability and optimum expression in *E. coli* host. Various molecular biology techniques such as gene knockouts, molecular cloning and transposon based genome integration techniques were used. Production of violacein was confirmed by appearance of violet colour in *E. coli* strains transformed with plasmid containing violacein producing operon. Further quantitative and qualitative analysis was done using absorption of ethanol extract from violacein producing cells at 550 nm. Qualitative estimation of PHB from *E. coli* transformed with plasmid containing PHB producing operon was confirmed with ^{1}H and ^{13}C NMR combined with fluorescent microscopy. A scalable method was developed for the quantitative estimation of PHB using nile red based spectrofluorometry. Further to eliminate the use of antibiotics and supplementing growth media with auxotrophic requirements, the PHB producing synthetic operon was genome integrated in *E. coli* K12 MG1655 strain using Tn7 transposon based genome editing technique.

Chapter 2: Methodology: Design, Build, Test and Analyze

Design: Heterologous genes may not use synonymous codons with same frequency which results into poor expression. Synthetic biology plays a crucial role of alleviating the effect of codon bias, while designing the gene for heterologous expression; rare codons are replaced with more frequent synonymous codons, without affecting the structure of translated protein. Several gene optimization tools are available for gene optimization according to the specific host all the approaches used mathematical algorithm to optimize genes. Apart from codon bias, GC content, tandem repeats and inverted repeat, DNA-protein interaction motifs; cryptic splice sites, restriction sites, which can negatively affect the transcription, stability of RNA and translation, can be eliminated. Nucleotides sequence with specific function are called standard parts that are categorized as Gene, Promoter, RBS, riboswitches and Terminator, standard parts put together into devices; Protein or RNA, and devices are put together with a System; Metabolic Pathway. This method of abstraction allows us to make replica of natural circuit using synthetic parts, replacing or adding these synthetic circuit that has a tight control over its functions, give us clues how genetic circuit works and allows to design complex genetic circuits like Toggle Switch, oscillators, repressilators(Elowitz and Leibler, 2000). The characteristic of the standard parts can also be predicted using pathway design tools based on thermodynamics of promoter-transcription factor interaction and coupling of genes within the operon(Salis et al., 2009).

Build: Methodology for gene synthesis has not changed since its invention, the monomers used for gene synthesis is nucleoside phosphoramidite. It is deoxyribonucleoside protected at exocyclic reactive group (amine, hydroxyl and phosphate) to avoid any nonspecific reaction while gene synthesis. The gene synthesis is carried out from 3' to 5' direction by adding each monomer at a time into oligonucleotide. Final step in gene synthesis is assembly of oligonucleotides into full length gene. Generally assembly is done by using cocktail of enzymes; ligase, DNA polymerase and polynucleotide kinase, this is known as Polymerase Cycling Assembly (PCA). With the discovery of Thermostable ligase, an alternative method involving Thermostable ligase and Polynucleotide kinase without any polymerase in the assembly reaction was developed, this method requires completely overlapping oligonucleotides, which can be phosporylated using Polynucleotide kinase and assembled using Ligation Chain Reaction (LCR). Many restriction enzymes-free cloning technique has been developed which enables not only the assembly of standard parts into pathway but entire genome using yeast as a host(Gibson et al., 2010).

Test: This is an ad-hoc part of any strain designing process, because it depends on the phenotype of the engineered host. Estimation of product yield is not the only sufficient criteria to test any strain. Strain engineering goes through many rounds of improvement and also should support future scale up process. Apart from production, growth rate, growth yield and substrate uptake rates are some of the important phenotypic characteristics of strain. Various methods are available for measurement of substrate uptake for all common substrate and intermediary metabolites. Both quantitative and qualitative characterization of molecules is necessary. Since Violacein is a coloured compound and soluble in ethanol, it can be characterized using colorimetric assay. The PHB characterization requires spectrofluorometric method to determine the product yield. The method was developed based on the fact that Nile red, a fluorochrome, quantitatively binds with intracellular PHB. The Nile red bound quantitatively to PHB can be extracted with ethanol and fluorescence can be measured. A calibration with standard PHB was established for absolute quantitation. For qualitative analysis NMR and mass spectrometry was used for PHB and Violacein respectively.

Analyze: The genome-scale metabolic model of *E. coli* iAF1260 was extended by adding curated violacein biosynthesis reaction, including gene-protein associations, reaction stoichiometry, and reversibility. Similarly CoA transferase and PHB polymerization reaction was added to genome scale model of *E. coli*iJO1366. The resulting model *i*AF1260vio and iJO1366PHB was able to synthesize Violacein and PHB. The set of constraints used in different simulations included (i) Substrate (Glucose) uptake rates (GUR) (ii) Growth yields (iii) product secretion associated with molar growth yields of each strain. A basis of 1 g biomass was used to calculate specific growth rates. Implementation of the genome scale metabolic model with experimental and legacy data based constraints was done using Constraints Based Reconstruction and Analysis (COBRA) Toolbox MOMA (Minimization of metabolic adjustment) Analysis was performed in COBRA toolbox using the function MOMA() to study the effect of gene deletions *in silico*, that allows for selecting more appropriate optimal solutions. Shadow price and reduced costs analysis;two sensitivity parameters – shadow prices and reduced costs were assessed at maximal yields and biomass in order to understand the effects of changing biomass, metabolites and reactions in the network. Shadow price corresponds to the sensitivity of the growth rate as an objective function in response to a change in the availability of a metabolite, and indicates how much an increment in that metabolite will increase or decrease the growth rate. Analogous to shadow price, reduced cost is the sensitivity of the objective function in response to change in fluxes of a particular reaction and its effect on the objective. In addition to the primal solution (optimal fluxes), the LP solver provides the corresponding dual solution i.e., shadow price and reduced cost for the FBA problem. Flux variability analysis (FVA) can be set up in COBRA toolbox using the function fluxVariability() to study the resulting space of feasible flux distributions (Mahadevan and Schilling, 2003). Nine categories can be mapped onto the flux variability distribution based on the magnitude and direction of the flux. Robustness analysis set up in COBRA toolbox using the function robustnessAnalysis() is used to better understand the phenotype of an organism under different environmental perturbations. Phenotypic phase planes (PhPP) analysis was performed characterizing all optimal flux distributions as a function of NADPH and Tryptophan demands at experimental biomass. Simultaneous sensitivity/shadow price analysis allows one to delineate the changing shadow prices in each phase of the plane.

Chapter 3: Synthetic and Systems Metabolic Engineering for Violacein production

Violacein is a bacterial bis-indole pigment of commercial interest having antibacterial, antitumoral, antiviral, trypanocidal and antiprotozoan properties(Durán and Menck, 2001). The violacein biosynthetic pathway is complex due to a coordinated five gene operon structure in *Chromobacterium violaceum*. The rate limiting step involves the condensation of two molecules of tryptophan to 2-imino-3-(Indole-3-yl) propionate (*vioA).* Further steps include conversion to protodeoxyviolaceinate by *vioB* and *vioE*, followed by conversion to violacein and deoxyviolacein by catalytic activity of *vioD* and *vioC*. The dynamics of violacein biosynthesis are thus dependent on the coordinated levels of transcripts, proteins, action of promoter and ribosome binding sites (RBS). This necessitates minimization of transcriptional noise (relative stability of transcripts) caused by synthesis and degradation of mRNA molecules and increasing the efficiency of translation of mRNAs into proteins by modulating translation initiation rate (dictated in part by mRNA secondary structures). Translational coupling (TC) significantly modulates translation efficiency of individual genes in a multi-gene operon structure within *E. coli*. In this study, an integrative computational and experimental strategy for strain design and genome engineering of *E. coli* to produce violacein was developed. Predicted gene deletions Δ*trpR*, Δ*pheA* and Δ*pgi* that increase the yields of the precursor tryptophan were constructed. Plasmid constructs pCVW and pSYN based on the wild type operon (WTO) and synthetic operon (SYNO) designs of the vioABCDE operon were expressed in *E. coli* K12 and its derivatives with varying genetic backgrounds. The impact of exploring the genetic design space on violacein production at molar growth yields was delineated. The grand challenge of systems metabolic engineering lies in the complexity and redundancy of cellular pathways and the evolutionarily drive to maximize growth/fitness rather than a forced bioengineering objective. Constraints based flux balance analysis (FBA) of metabolic models has been used to design strains *in silico* that simultaneously maximize fitness and the desired product. These models predict intracellular reaction fluxes and identify strategies for substrate uptake, energy and cofactor balance.

Chapter 4: Rational Strain Design for Poly(R)-3-Hydroxybutyrate biosynthesis

Naturally PHB biosynthesis operon consists of three genes, PhaA, PhaB and PhaC. These three enzymes catalyze the conversion of acetyl-CoA into PHB. Several attempts have been made to clone PHB biosynthetic pathway in *E. coli* to produce PHB and its co-polymers(Lee et al., 1994). There are several reports where an unnatural monomer is co-polymerized along with PHB(Choi et al., 2016). There are two routes in *E. coli* metabolism to generate (R)-3-hydroxybutyrate-CoA (3HB), which is the monomer for PHB. One of the routes is from central metabolism Glycolysis -> Acetyl-CoA-> 3HB and another route is to generate 3HB from fatty acid degradation. Since *E. coli* can produce 3HB, instead of introducing entire PHB producing operon, in this study, enzymes augmenting the incomplete PHB producing pathway were introduced in *E. coli* . Here, we have used the synthetic genetic circuit constructed using two enzymes, including Propionate-CoA Transferase (pct_{ap}) from *Acetobacter pasteurianus* which transfers CoA group from acyl-CoA (mostly Butanoyl-CoA) to acetate and Polyhydroxyalkonate polymerase ($phaC_{cv}$) from *Chromobacterium violaceium*, which polymerizes the monomer into PHB. The incorporation of act_{ap} is to increase the pool of acetyl-CoA by utilizing acetate produced during growth on *E. coli* culture on glucose and glycerol due to overflow metabolism. This was in agreement with our experimental results were *E. coli* transformed with wild-type and synthetic gene circuit produced PHB with glycerol supplemented media and did not produce any PHB with glucose and LB media. The PHB was not produced with glucose because acyl-CoA is not produced with glucose where as growth of glycerol accumulates acyl-CoA which transfers it CoA to acetate to produce acetyl-CoA. Both the genes are expressed through pBAD promoter which is induced by L-arabinose. Unlike natural operon structure, synthetic biology approach was used to create the genetic circuit. Both monocistronic design and polycistronic design were investigated using RBS calculator (https://salislab.net/software/) and monocistronic design was adopted which showed 1:1 transcription coupling for both the gene and controlled Translation initiation rate. Two gene pathways for biosynthesis of PHB in *E. coli* i.e each gene is controlled by separate pBAD promoter and terminator and placed in tandem. Glycerol substrate resulted in yields of 485.08 µg/mgDCW and 395.41 µg/mgDCW of PHB in the plasmid strains and the Genome integrated strains.

Chapter 5: Conclusions and Future scope

The premise of this thesis is to use the synthetic biology paradigm combined with systems metabolic engineering for rational strain design and development. Two diverse examples Violacein and Poly(R)-3-Hydroxybutyrate demonstrated various ways in which such goals can be achieved. The common strategy involved here was designing operons via gene sequence optimization that had the significant impact on the product yields. Synthetic designs were more robust. Violacein operon designs were more conservative and retained the native promoter and gene order.Genetic circuit design targeting native precursors made in the E. coli host that can be converted into the monomer for polymerization to PHB was adopted. Herein two genes were expressed under independent pBAD L-arabinose inducible promoter control and placed in tandem in a plasmid. Further to address the issues with antibiotic resistance markers, plasmid stability, the designed synthetic genetic circuit was integrated into the genome of *E. coli K12* genome using Tn7 transposon. Constrained based analysis of an extended genome scale model with a module for the desired product can predict genome engineering strategies for the host cell for better redox and energy balance and media optimization to increase the product yields. The standard parts established here are now well characterized. These kinds of operon and genetic circuit designs are scalable and can be used in future strain design projects.

Chapter 1 Introduction

Come gather Metabolic Engineers 'cross the land
At ME III we'll take command
Of cells that are too slow to produce or grow.
If it's higher fluxes you're needing'
Then we'll shift the controls, and block bad outflows.
The Metabolic Engineer has the answers for you.
We'll import new pathways, and shuffle them too.
For the times, they are a changing.
-Jay Bailey, 2000

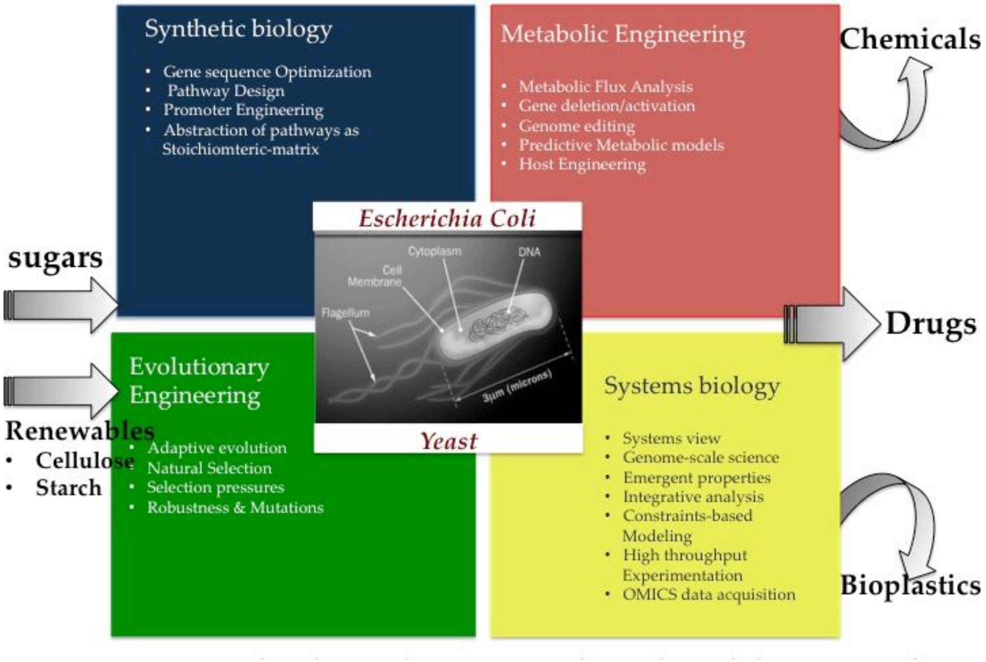

Synthetic biology
Gene sequence Optimization
Pathway Design
Promoter Engineering
Abstraction of pathways as Stoichiomteric-matrix
Metabolic Engineering
Metabolic Flux Analysis
Gene deletion/activation
Genome editing
Predictive Metabolic models
Host Engineering
Chemicals
Escherichia Coli
Cytoplasm
DNA
Cell Membrane
Flagellum
3μm (microns)
Yeast
sugars
Drugs
Evolutionary Engineering
Adaptive evolution
Natural Selection
Selection pressures
Robustness & Mutations
Renewables
Cellulose
Starch
Systems biology
Systems view
Genome-scale science
Emergent properties
Integrative analysis
Constraints-based Modeling
High throughput Experimentation
OMICS data acquisition
Bioplastics

Figure 1.1 Synthesis of value added products using Systems and Synthetic Biology

Nature produces a wide assortment of compounds that are highly diverse in structure and function. Many of these compounds are therapeutic and play an important role in maintaining human health. A large percentage of these are produced as secondary metabolites by their native hosts in miniscule amounts as needed for their specific function. Another big class of compounds is biopolymers. They are formed by repetitive covalent bonding of monomers for various purposes such as storing and transferring information (DNA), structural and functional units (proteins, lignocelluloses) and energy storage (Polyhydroxyalkanoate, starch, glycogen). Limitation in fossil resources and an increased awareness about environmental issues, has anticipated that renewable biopolymers will replace a substantial fraction of the market for synthetic polymers that exists today (Rehm, 2010).

Metabolic engineering, defined as the purposeful alteration of the genetic code of an organism to redirect its metabolic flux, is important in production of these value added products for many reasons:

(i) To simplify the process of synthesis

(ii) To optimize the yield of a natural compound or its intermediate and

(iii) To address the problem of insufficient supply vs market demand

1.1 Motivation

Rational construction of desired metabolic traits in a heterologous host began several decades ago and has now progressed to the assembly of very complex pathways. Genetically well characterized hosts like *E. coli* and *S. cerevisiae*, with favorable growth rates and simple nutrient requirements are generally attractive biosynthetic hosts. With the increased insight into the design principles it is now possible to extend the spectrum of products synthesized to drugs and biopolymers. However, the complexity of metabolic pathways, along with regulation often results in unexpected outcomes for metabolic engineering and the resulting strains may require extensive fine tuning to be economically viable. In recent years, metabolic engineering has shifted its paradigm towards a systemic approach that relies on large scale experimentation and computational analysis of metabolic and regulatory networks. High throughput technologies have resulted in databases that contain biochemical, molecular and genomic information that enable more systemic metabolic engineering approaches. The reconstruction of *in silico* genome scale stoichiometric models of metabolic networks that compute the functional state of the cell accurately facilitate *a priori* determination of effects of gene deletions and insertions on metabolic state, growth and redox state for improved product synthesis. The design principles and engineering tools to build and optimize cellular processes are state of the art today and encompass the field of synthetic biology. This process allows access to 3-tiers of optimization: the central dogma, intrinsic regulatory interactions and extrinsic environmental interactions leading to their seamless integration. Based on this background one can visualize the synergy between synthetic biology, systems biology and metabolic engineering which leads to inevitable microbial cell factories.

1.2 Overall objective and specific aims

The main objective of this thesis is to investigate rational strain designs using synthetic and systems biology approaches for metabolic engineering of a drug molecules and biopolymers. To satisfy this objective, certain specific aims were put forth and included:

- Design of a synthetic operon for the violacein pathway expression in *E. coli*
- Design of a synthetic operon for PHB pathway expression in *E. coli*
- Metabolic engineering of the designed operons/circuits in *E. coli*
- Development of reactions modules for metabolic reconstruction and Constraints-based flux balance analysis for product biosynthesis in E. coli
- Genome engineering of *E. coli*
    - to compensate for precursor/cofactor imbalance in violacein production
    - to remove antibiotic selection markers in PHB production

The thesis aims to understand synergies between synthetic and metabolic systems biology for metabolic engineering of violacein and PHB. The underlying hypothesis is that development of rational strain designs through the Design-Build-Test-Analyze paradigm of synthetic biology would fast forward evolution and create scalable workflows for optimization of value added products.

1.3 Literature Review

Since the advance of modern biology, humans have gained the comprehensive knowledge by understanding basic molecules of life. These molecules are mainly built from six elements: Carbon, Hydrogen, Oxygen, Nitrogen, Sulfur, Phosphorus, some macro-molecules and few trace elements. Together these elements can result into a highly complex system. Strain development for production of value added products requires channeling these basic molecules towards the desirable metabolic output to make product of interest. Engineering microorganism to produce value added products such as drugs, biofuel, chemicals, and polymers, a field popularly known as Metabolic Engineering, is now greatly accelerated by use of Synthetic Biology uncapping the limits to engineer biological system. Due to the scale at which the synthetic biology operates, instead of applying a reductionist approach a systems approach is preferred. Therefore the comprehensive approach is, to apply synthetic biology in combination with systems metabolic engineering. This thesis discusses the goals achieved by the systems metabolic engineering and synthetic biology approach towards the strain development

1.3.1 Historical Overview

The first evidence explaining that the genome of an organism can be perturbed comes from the Griffiths Experiment where in transformation principle was found in Pneumonococcus (Griffith, 1928), since back then DNA was not established as a genetic material, it was only after Avery, Macleod and McCarty work on Pneumococcus type III cell fractions , it was proved that transforming entity is DNA (Avery, 1944) Oswald Avery’s discovery was under lot of controversy until Lederberg, Tantum and Beadle published their findings on conjugation in 1946 (Tatum and Lederberg, 1947). This ignited the minds of many researchers to find the structure of DNA and it happened very soon, when Watson and Crick discovered the structures of DNA(H C Crick, 1953) . Immediately after the discovery of Structure of DNA, the quest for deciphering genetic code became the priority and competition. It was a collective effort of three different Scientist who decoded the secret of life, H G Khorana(“The Nobel Prize in Physiology or Medicine 1968,” n.d.), Marshall Nirenberg(“The Nobel Prize in Physiology or Medicine 1968,” n.d.), William Robert Holley(Holley et al., 1965). Following his work even further H G Khorana was able to chemically synthesize Alanine-tRNA gene from yeast and published 17 manuscripts in single edition of Journal of Molecular Biology, where he and his team has established the basic principles for gene synthesis and polymerase chain reaction . This is about the journey from the discovery of DNA to its chemical synthesis. Simultaneous with these landmark discoveries, Arthur Kornberg et al(Lehman et al., 1958) discovered enzymatic synthesis and discovered DNA polymerase that was used in gene synthesis methodology. Some other enzymes were being discovered at the same time, which acts on the DNA such as Restriction Enzymes(Kelly Jr and Smith, 1970; Meselson and Yuan, 1968; Smith and Welcox, 1970),T4 ligase(Weiss and Richardson, 1967), T4 Polynucleotide kinase (Novogrodsky et al., 1966), Alkaline Phosphatase (Maxam and Gilbert, 1980). All these discoveries lead to a new technology called Recombinant DNA Technology, which contributed to most of the breakthroughs in the field of molecular biology and basic biological sciences. Development of DNA sequencing by Frederick Sanger(Sanger et al., 1973) and Polymerase Chain Reaction (PCR) by Kary Mullis. There is no change in the basic methodology of gene synthesis other than the throughput and size of DNA fragments that can be synthesized. The cloning methodology has advanced with a large extent which uses not only restriction digestion based cloning but also the recombination based cloning(Gibson et al., 2010).

1.3.2 Gene synthesis

Methodology for gene synthesis has not changed since its invention [6], the monomers used for gene synthesis is nucleoside phosphoramidite, it is deoxyribonucleosides protected at exocyclic reactive group (amine, hydroxyl and phosphate) to avoid any nonspecific reaction while gene synthesis. The gene synthesis is carried out from 3’ to 5’ direction by adding each monomer at a time. The process is carried out in four steps i) Deprotection, ii) Coupling, iii) Capping iv) Oxidation

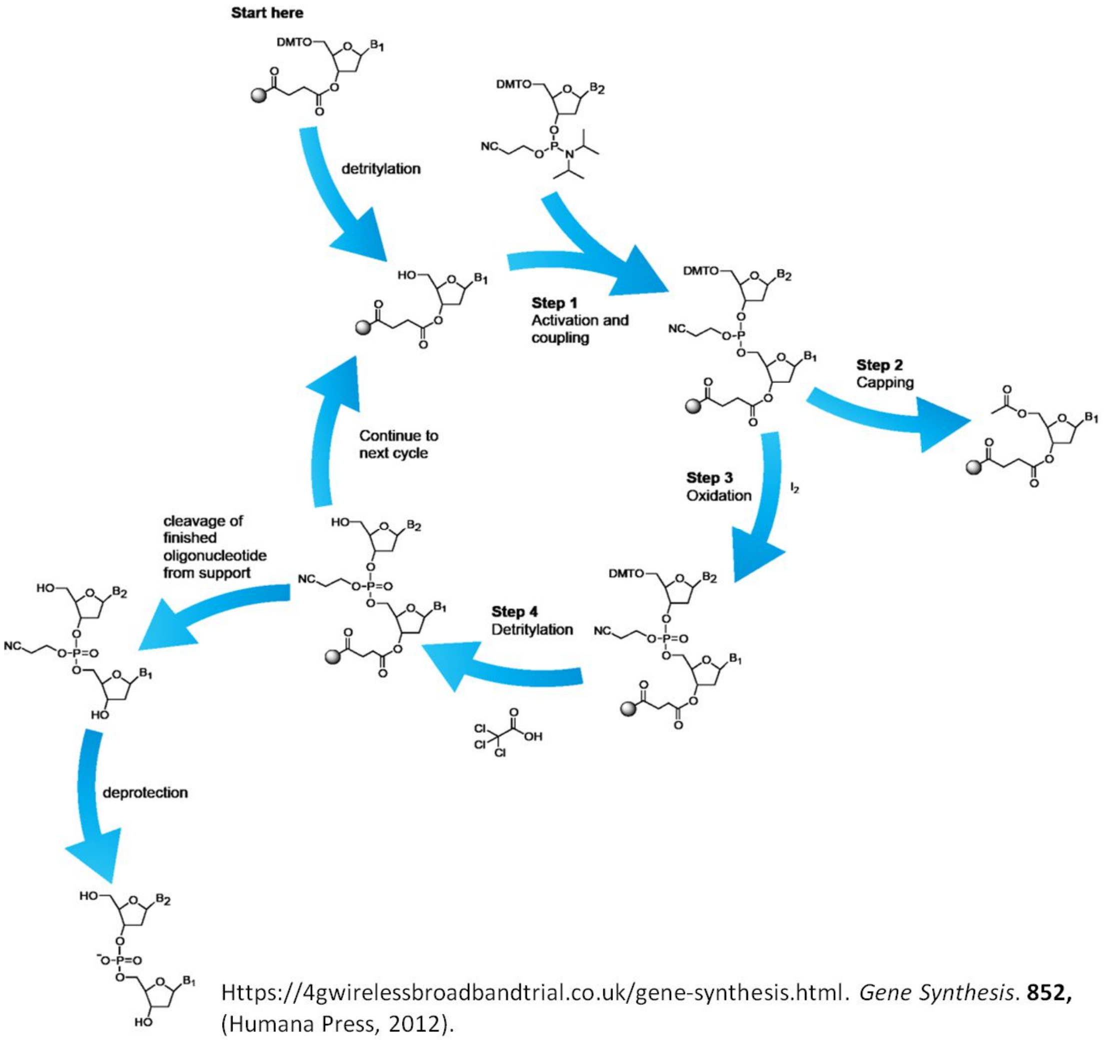
i) **Deprotection:** Removal of 5’ protecting group; Di-Methoxytrityl (DMT), by adding a weak acid.

ii) **Coupling:** Addition of monomer along with coupling agent or activator (tetrazol)

iii) **Capping:** Capping uncoupled growing chains in the previous reaction by acylation,

iv) **Oxidation:** Unstable phosphite group is oxidised to more stable phosphate triester.

v) **Detritylation:** The 5'-DMT protecting group is removed by TCA (trichloroacetic acid) in the solvent dichloromethane


Start here
DMTO
B_1
detritylation
HO
Step 1
Activation and coupling
DMTO
B_2
NC
O-P
N
Step 2
Capping
Step 3
Oxidation
I_2
Step 4
Detritylation
Continue to next cycle
cleavage of finished oligonucleotide from support
O-P=O
deprotection
Cl
OH
HO

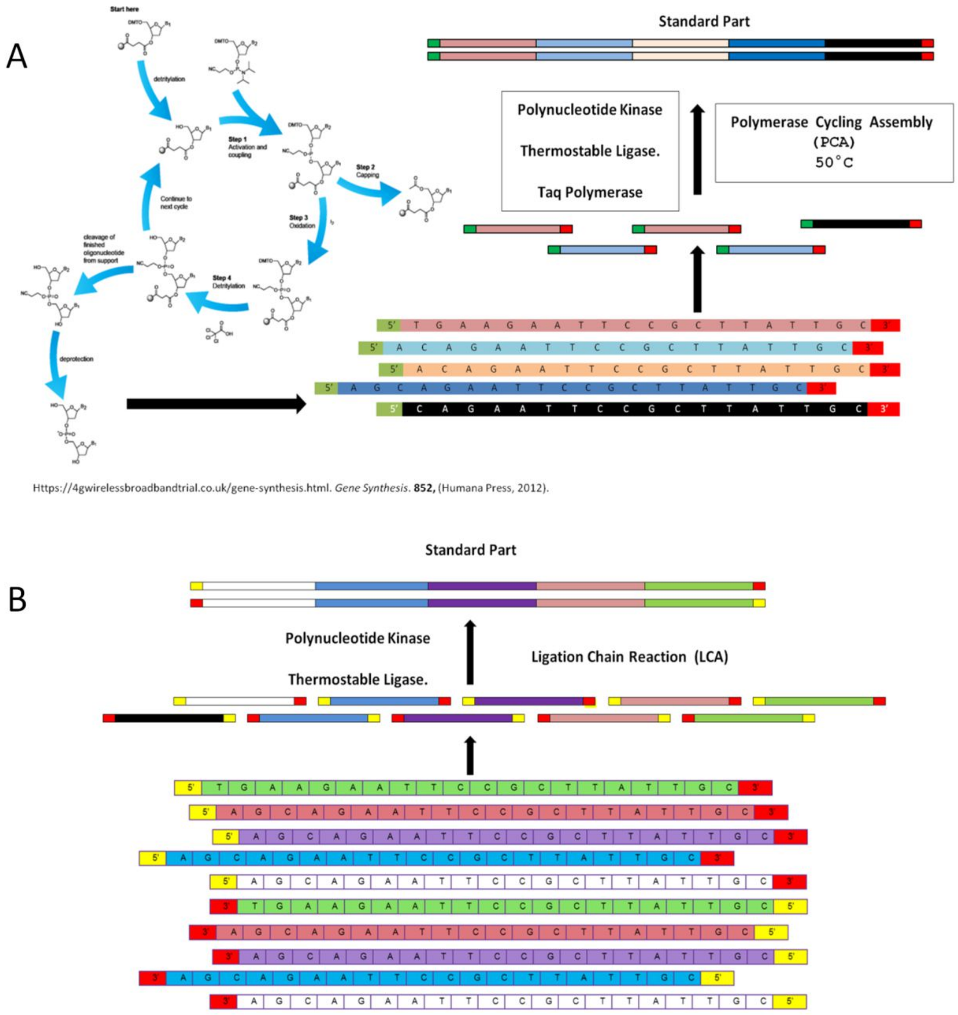
Https://4gwirelessbroadbandtrial.co.uk/gene-synthesis.html. *Gene Synthesis*. **852,** (Humana Press, 2012).

Figure 1.2 Chemical sysnthesis of oligonucleotide using phosphorhamadite method

Gene synthesis using protected Phosphorhamadite on Controlled Pore Glass (CPG) as a solid support. The change in the protocol in last few decades had occurred to reduce the number of steps involved, improving the throughput of synthesis and reducing the cost of sequencing. To reduce the number of steps, deprotection can be coupled with Oxidation, which can be achieved by deprotection by oxidation, if 5'-DMT is replaced by aryloxycarbonyl, oxidation with peroxy anion will simultaneously remove carbonate protecting group and oxidise phosphite linkage to phoshotriester bond. To improve the throughput of synthesis process various methods have been developed that are mostly based microarray chip(Richmond, 2004; Tian et al., 2004). Such technology exploits expertise from electronics and microchip industry. Photolithigraphy is one such technique where nucleoside phosphoramidites are protected by photolabile protecting groups and area where incoming monomer couples is exposed to light masking the rest of the surface. Electrochemical array uses electricity to change the local pH on the slide; this requires special slides itched with microscopic microelectrode using photolithography. Other techniques like inkjet printing, microfluidics and LED-controlled capillary synthesis have been adopted. These method differ mainly in protecting group, method of deprotection and solvent systems(Tian et al., 2009). Currently polymerase-nucleotide conjugate based de-novo methods is being developed which can bring the gene synthesis to bench top(Sebastian Palluk et al, 2018).

Gene Assembly: Final step in gene synthesis is assembly of oligonucleotides into full length gene; generally assembly is done by using cocktail of enzymes containing DNA ligase, DNA polymerase and polynucleotide kinase, this is known as Polymerase Cycling Assembly (PCA)(Willem P.C. Stemmer et al., 1995). With the discovery of Thermostable ligase, an alternative method involving Thermostable ligase and Polynucleotide kinase without any polymerase in the assembly reaction was developed (Au et al., 1998), for this method completely overlapping oliogonucleiotides, which can be phosporylated using Polynucleotide kinase and assembled using Ligation Chain Reaction (LCR) (Au et al., 1998). Both the methods require a PCR reaction, after removal of incorrect sequence, by end primers to amplify full gene. The PCA can generate more errors than LCR, whereas LCR will require more no of oligonucleotides which makes it more expensive. By combining the above two methods, the whole process can become one step by using thermostable ligase.The oligonucleotide sequence is not error free because the efficiency of coupling is not 100% in each reaction in such cases there can be deletions or insertions in final sequence. Sequence with such errors can be rectified using HPLC or PAGE purification.

The real problem comes when error is an SNP, which can be rectified using mismatch binding protein such as MutS to separate the desired synthetic construct from mutated synthetic construct Second method is to treat the synthetic construct with mismatch specific endonuclease for Example T4 endonuclease or MutHLS endonucleases, which


A
Start here
detritylation
Step 1 Activation and coupling
Step 2 Capping
Step 3 Oxidation
Step 4 Detritylation
Continue to next cycle
cleavage of finished oligonucleotide from support
deprotection
Standard Part
Polynucleotide Kinase
Thermostable Ligase.
Taq Polymerase
Polymerase Cycling Assembly (PCA) 50°C
B
Standard Part
Polynucleotide Kinase
Thermostable Ligase.
Ligation Chain Reaction (LCA)
5'
3'

Https://4gwirelessbroadbandtrial.co.uk/gene-synthesis.html. *Gene Synthesis*. **852,** (Humana Press, 2012).

Figure 1.3 Assembly methods for gene synthesis

(A) Polymerase Cycling Assembly (PCA) uses thermostable ligase polynucleotide kinase and Taq polymerase to assemble oligonucleotide. (B) Ligation Chain Reaction uses only Thermostable ligase and polynucleotide kinase but requires more oligonucleotides.

will cut the sequence at the site of mismatch, which can be separated using size exclusion gel Electrophoresis(Carr et al., 2004).

1.3.3 Assembling Plasmids to Genomes

Traditional cloning techniques based on restriction enzymes are limited to 1 or 2 fragment assembly, synthetic biology requires large no of standard parts with varying size to be assembled. Several recombination based assembly methods are developed, both *in-vitro*(Gibson et al., 2009) and intracellular in yeast(Gibson, 2009). The first two genome synthesis of Polio Virus(Cello et al., 2002) and ϕX174(H O Smith et al., 2003) genome sequence did not involve any cell mediated assembly steps since their genome is small. To go beyond the size limit of above mentioned in-vitro method, it is necessary to involve living cells two usual choices would be *E. coli* and yeast. Even if larger sequence can be assembled using oligos the assembly will have lot of errors and desired sequence will be few in number in the heterogeneous mixture of erroneous fragments that are difficult to isolate from rest of the sequences. To overcome this, PCA is used to make smaller fragments called synthons (500–700bp)(Kodumal et al., 2004) these are cloned into *E. coli* and several clones are screened by sequencing to find out the error free sequence. Synthons can be further assembled using enzymatic assembly using cocktail of three enzymes; T5 exonuclease , DNA polymerase and Taq ligase (Gibson et al., 2009). At a time 4 to 10 stynthons can be assembled, desired no. of synthons are mixed with Assembly buffer, dNTP mix and Enzyme cocktail. The T5 exonuclease chews away 5' fragment leaving 3' single stranded overhang at 50°C, but at 50°C it is not stable and gets deactivated preventing longer 3' overhang. Synthons having end homology anneals and polymerization is carried out at 50°C, this is followed by ligation with thermostable Taq ligase at 50°C. Since all the reaction takes place at 50°C this method does not require any specialized instrument like thermo cycler. One of the synthons can be a linearized plasmid so that the entire construct can be transformed to get circular clones that can be authenticated using sequencing.

The scale of synthesis is improved to an extent that, first bacterial genome has been synthesized using above method. The choice of organism is *Mycoplasma genitalium* which has smallest genome among bacteria (580kb)(Gibson et al., 2010), and the assembly involving longer sequence were assembled using yeast as a host for cloning purposes, since larger sequences are unstable in the *E. coli* in the form of plasmid. Currently genome of *Saccharomyces cerevisiae* is being synthesized with a global mission in action to make synthetic yeast Sc2.0, until 2017 five chromosomes are already synthesized(Richardson et al., 2017)

1.4 Synthetic biology

Synthetic biology is defined in several ways by pioneers in this field leading way to modular biological systems that can be fabricated individually and used in a plug and play mode (based on the desired hierarchy) for basic or applied research. Some of these definitions are delineated below:

"Synthetic biology is a) the design and construction of new biological parts, devices and systems and b) the re-design of existing natural biological systems for useful purposes."

Source: Synthetic Biology.org

"Synthetic biology is an emerging area of research that can broadly be described as the design and construction of novel artificial biological pathways, organisms or devices, or the redesign of existing natural biological systems."

Source: UK Royal Society

"Synthetic biology is a maturing scientific discipline that combines science and engineering in order to design and build novel biological functions and systems. This includes the design and construction of new biological parts, devices, and systems , as well as the re-design of existing, natural biological systems for useful purposes"
Source: SynBERC

"Synthetic biology is the engineering of biology: the synthesis of complex, biologically based (or inspired) systems which display functions that do not exist in nature. This engineering perspective may be applied at all levels of the hierarchy of biological structures – from individual molecules to whole cells, tissues and organisms. In essence, synthetic biology will enable the design of 'biological systems' in a rational and systematic way."
Source: High-level Expert Group European Commission

All the above definitions are established with the intentions of broad application for synthetic biology. In the context of this study, Synthetic biology may be looked upon as an interdisciplinary branch at the interface of biology and engineering that allows the application of engineering principles, at every level of central dogma of molecular biology to achieve desired metabolic output. This output is generally a value added product and can range from small molecules like amino acids, organic acids, drug molecules to polymers. The three Salient features of Synthetic Biology include:

(i) Gene Synthesis.

(ii) Abstraction.

(iii) Standardization

1.4.1 Synthetic Biology an Engineering Perspective

Engineering already existing natural standard parts by tweaking its base sequence is a common approach to engineer standard parts for synthetic biology. Sequences for many natural as well as engineered parts are available in open access online repository http://parts.igem.org/Main_Page , and people across the world are adding new designs every day to this repository which includes Promoters, RBS, Genes, Terminators, antibiotics resistance cassettes, origin of replication and Polylinker. Every sequence is with its description and references suggesting its potential use. It is not necessary to use engineered standard part every time; sometimes a natural sequence can do, but combination of standard part which makes a module or a device should support the objective of using every part. Generally an artificial operon is constructed out of standard parts, and one can take assistance of genome scale metabolic models to predict the global impact of introducing synthetic operon into the host. Several iterative predictions can be performed to fine tune the final synthetic sequence before its chemical synthesis.

1.4.2 Potential applications of Synthetic biology

One cannot imagine human life without chemicals, the major source of these chemicals is crude oil which is depleting at a fast rate. Currently seven billion humans inhabit earth and it is estimated that till 2025 we will be 12 billion, which is almost double and that to with the improved life style. The demand is going to increase every day and impact on the environment is another problem. Clearly we need to switch to the renewable resources as soon as possible. Energy is central to the world's major demand and transport fuel is the major target of synthetic Biology. There are two types of biofuels currently into development BioEthanol and BioDiesel. BioEthanol is produced by fermentation of sugar by Yeast (*Saccharomyces cereviciae)* the sugar can be derived from enzymatic degradation of cellulose, the most abundant polysaccharide in nature. Many attempts in cloning several cellulose degrading enzymes have failed to produce significant amount of ethanol. Biodiesel is produced by etherification of fatty acids triglycerides to produce methyl esters of saturated or monounsaturated fatty acids and glycerol as a by-product several algae produces high amount of fatty acids but not enough to make BioDiesel commercially. Not much has been done in synthetic biology to improve the strains of yeast and algae to produce biofuel and much work needs to be done in this area with respect to synthetic biology. Many monomers are produced by microorganisms which can be polymerized to produced polymers PolyHydroxyButyrate (PHB) (Rehm et al., 2002) and Poly Lactic Acid (PLA) are candidates for the production of Bio Degradable Plastic out of which PHB is naturally synthesized by it monomer (Rehm et al., 2002), but PLA was produced using evolved heterologous enzymes in *E. coli* (Yang et al., 2010). Vaccines are attenuated virus which is generated by selecting non-virulent viral strain; this process takes long time, from months to year and in case of epidemic several lives can get affected in this time period. One of the ways to accelerate this process is to synthesize a non-virulent viral strain based on the knowledge of its genome; this might be a month job. In fact one of the first genome to be synthesizing using artificial genome sequence is a virus ϕ174. Synthetic biology has several applications for the future world and biologists across the world are adapting synthetic biology to solve many real world problems.

The first two genome synthesis of Polio Virus and ϕX174 genome sequence did not involve any cell mediated assembly steps since their genome is small (H. O. Smith et al., 2003), in the previous case people did use conventional cloning using restriction enzymes. To go beyond the size limit of above mentioned in-vitro method, it is necessary to involve living cells two usual choices would be *E. coli* and yeast. Even if larger sequence can be assembled using oligos the assembly will have lot of errors and desired sequence will be few in number that are difficult to isolate from rest of the sequences. To overcome this, PCA is used to make smaller fragments called synthons (500–700bp)

(Kodumal et al., 2004); these are cloned into *E. coli.* Several clones are subsequently screened by sequencing to find the error free sequence. Synthons can be further assembled using enzymatic assembly using cocktail of three enzymes; T5 exonuclease , DNA polymerase and Taq ligase (Au et al., 1998). At a time 4 to 10 synthons can be assembled, desired no. of synthons are mixed with Assembly buffer, dNTP mix and Enzyme cocktail. The T5 exonuclease chews away 5' fragment leaving 3' single stranded overhang at 50°C, but at 50°C it is not stable and gets deactivated preventing longer 3' overhang, synthons having end homology anneals and polymerization is carried out at 50°C, this is followed by ligation with Taq ligase also at 50°C. Since all the reaction takes place at 50°C this method does not require any specialized instrument like thermo cycler. One of the synthones can be a linearized plasmid for *E. coli* so that the entire construct can be transformed to get circular clones which can be authenticated using sequencing. The scale of synthesis is improved to an extent that, first bacterial genome has synthesized using above method, the choice of organism is *Mycoplasma genitalium* which has smallest genome among bacteria (580kb) (Gibson et al., 2008), and the assembly involving longer sequence were assembled using yeast as a host for cloning purposes, since larger sequences are unstable in the *E. coli* in the form of plasmid. Functionality of the synthetic genome was authenticated by transplanting the genome into another species of Mycoplasma (Gibson et al., 2010), only this time the genome of *M. mycoides* was synthesized and transplanted into *M. capricolum.*

1.5 Bioinformatics and Abstraction

Clearly lot of information is generated across the world and this information is complex, there is an urgent need of means to communicate this kind of information between synthetic biologists across the world. Abstraction provides means to communicate such information, it associates complex information to a single function and it works at every level of complexity. In case of synthetic biology, it associates nucleotides sequence with Standard parts; Gene, Promoter, RBS, or Terminator, it associates Standard parts put together with Device; Protein or RNA, it associates Devices put together with a System; Metabolic Pathway. In general abstraction is annotation at every level of complexity. The abstraction allows us to make replica of natural circuit using synthetic parts, replacing or adding these synthetic circuit that has a tight control over its functions. This will give us clues how genetic circuit works and allow us to design complex circuits like Toggle Switch, oscillators, repressilators. The toggle switch is generated by placing two constitutive promoters P1 and P2 with two repressors R1 and R2 respectively, where R1 represses P2 and R2 represses P, this kind of Toggle switch will give bistable toggle network (Heinemann and Panke, 2006). Oscillators are combinations of three pairs of promoters and repressors in loop these are also known as Repressilators (Appleton et al., 2014). Another new method of constructing genetic logic gates using recombineering techniques based λ phage serine integrase have been successfully attempted. These genetic logic gates were used to apply Boolean logic in genetic circuits (Bonnet et al., 2013).

1.6 Systems Biology for Metabolic Engineering

The central dogma of molecular biology, i.e, information stored in DNA is transcribed to mRNA and then translated into a protein that carries out catalytic, regulatory or structural function in a cell, governs metabolism and gives it a hierarchical multi-scale structure. This multi-scale nature is seen in the different layers of metabolism being integrated via explicit dependencies. Thus, integrating transcript, protein and metabolite information is essential to understand the phenotype under specific conditions. Delineating the basis of the fundamental genotype-phenotype relationship in biology thus needs a systems approach. This systems approach has flourished with the advent of several high throughput genome-scale technologies and has been used to understand the molecular composition (genes, transcripts, protein and metabolites) and their integrated functioning in a cell. More recently, these techniques have been exploited for design purposes and engineering a cell for a desired trait or function. Metabolic engineering has benefited immensely by large scale screening and implementation, due to the global effects of local changes in pathways to maximize a desired product. Further, building *in silico* models has allowed us to analyze complex network structure to ensure stability, robustness to perturbations and display emergent properties for manipulation experimentally. This review summarizes some of the popular systems approaches *in silico* and high through put biology that have been used to drive metabolic engineering outcomes. This is followed by a section on the current status of alkaloid and vinblastine synthesis.

1.6.1 *In silico* methods for metabolic engineering

There are several computational and mathematical approaches used to model metabolic network behavior. These include kinetic methodologies such as enzyme kinetics (Varma and Palsson, 1993) and metabolic control analysis (MCA) (Schuster et al., 2008), stoichiometric approaches, such as metabolic flux analysis (MFA) (He et al., 2017) and flux balance analysis (FBA) (Yadav et al., 2012), and power law formalisms such as biochemical systems theory (BST) (Rehm and Steinbu, 1999).In kinetic methodologies, mathematical models of metabolic networks are created with details of enzyme kinetic equations. Success depends on a complete knowledge of system parameters – rate constants *a priori*. This results in poor prediction accuracies of cellular response and requires iterative experimental work. Thus, these methods are generally suited to small systems and used to determine regulatory steps in specific metabolic pathways.

The philosophy in MCA focuses on discovering collective control of a series of interconnected reactions. MCA proposes the idea that the regulation of the cell requires coordinated changes of multiple enzyme activities. It analyzes how control of flux and concentrations of intermediates in a pathway are distributed. Application of MCA in metabolic engineering has achieved notable success and involves detailed flux calculations with rationalized strain improvements (Stafford and Stephanopoulos, 2001).

Stoichiometric methodologies such as MFA and FBA are used when detailed kinetic information of metabolic interactions are not available. These models rely on mass action constraints to mathematically represent the direction of metabolic modulation. There are several algorithms developed to explore the optimal use of proteins and pathways for a given metabolic phenotype. There are several techniques like MOMA, OPTKNOCK, OPTGENE that have been developed specifically with metabolic engineering in mind (Kim et al., 2008). Robustness analysis and phase-plane analysis have allowed narrowing down choices when alternate feasible solutions exist. Since these models do not depend on detailed kinetic information (Thiele and Palsson, 2010), they are the only kind of models that have been built on a genome scale till date (Feist et al., 2010, 2009; Schellenberger et al., 2010). Their prediction of metabolic phenotype has been successful to apply to many areas of research including metabolic engineering (Stephanopoulos, 2012; Thiele and Palsson, 2010).

1.6.2 Genome Scale Models and Flux balance analysis in metabolic engineering

One of the major challenges in efforts to engineer the over production of desired products is identifying genes to be manipulated to generate the desired phenotype. Choosing the best genome manipulation strategies requires understanding not only the main pathway for alteration, but how the altered metabolic pathway will function in the context of the whole cell/organism. Although, high throughput data-based techniques identify potential bottlenecks post engineering, they are less useful when used in a prospective fashion to select genes for manipulation. Genome-scale models of metabolic networks hold great promise and have delivered results in this prospective design phase of metabolic engineering. Constraint-based modeling techniques like flux balance analysis and other novel methods can be applied to these models to understand the interplay of internetworked pathways with respect to the desired product and allow rapid evaluation of potential engineering strategies. Also, the ability to integrate heterogeneous data types within this framework is a great asset that helps multi-scale analysis and prediction at the gene, protein and metabolite level. There is a sudden increase in the application of such models to engineering micro-organisms to overproduce commercially desirable metabolic products (Jung et al., 2010; Lee et al., 2005; Willem P C Stemmer et al., 1995; Yang et al., 2010). Lee et al. provide a typical example where optimal gene knock-out targets improve production of succinic acid in a strain of *E. coli* that naturally produces only miminal quantities. Using comparative genomics with *M. succiniproducens* in combination with gene knock out predictions using FBA on a genome-scale model of *E. coli* resulted in an optimal design that over produced succinic acid.

The OptKnock method (Burgard et al., 2003) is a computational tool that was developed for genome-scale metabolic models, specifically with the metabolic engineering application in mind. The underlying idea is to identify gene knockouts that couple the production of desired metabolite with cellular growth. The optknock approach relies on a bi-level mixed integer optimization that finds the optimal combination of gene deletions that allow maximizing the bioproduct simultaneously with biomass formation rate. This framework in conjunction with evolutionary engineering has been successfully applied for growth coupled over production of succinate, lactate, 1,3 propane diol and amino acids like threonine and valine (Burgard et al., 2003; Kim et al., 2008; Pharkya et al., 2004, 2003). *In silico* experiments would play a key role in metabolic engineering applications beyond over production of metabolites native to the host. Introducing heterologous pathways *in silico* would be a valuable to tool to investigate the optimal strain design for overproduction of violacein and also PHA.

1.6.3 High through-put experimental methods for metabolic engineering utilizing single high throughput technologies

High throughput and comparative genome sequencing methods have allowed creating databases of gene and gene sequence repertoire for both prokaryotes and eukaryotes (Burge and Karlin, 1997). Analyzing this digital information forms a basis for metabolic engineering of various compounds of interest. The comparative genomics approach has demonstrated the over production of lysine in *Corynebacterium glutamicum* (Ohnishi et al., 2002). The insertion and deletion of genes for metabolic engineering of carotenoids in *E. coli* and yeast has been successful and drove research towards overproducing the precursur of artemisinin (Tsuruta et al., 2009), an antimalarial from plants, in yeast. *E. coli* has been engineered with several genes for cellulose and hemicellulose degradation to extend the substrate range to renewable lignocellulosic resources (Ingram et al., 2010). These strains have been used for green production of fuels like ethanol and other building blocks.

Transcriptomics, using high density DNA microarrays allows for parallel study of mRNAs under different environmental and genetic conditions. These genome wide profiles at different times, give insight into the metabolic and regulatory structure and can explain dynamic physiological changes like growth rates and alcohol tolerance in the organism (Gonzalez et al., 2003; Ingram et al., 2010). Transcriptomic profiles were successful in improving the production of the human insulin-like growth factor I fusion protein in *E. coli* (Choi et al., 2003).

Proteomics allows the identification and quantification of proteins present in an organism in a given condition. Stable isotope coding, 2D electorphoresis or other technologies are combined with mass spectroscopy for this method (Patterson and Aebersold, 2003). Although proteome analysis has not been exploited in metabolic engineering like its counterparts – genomics and transcriptomics, it can provided unparallel insight on the activity of metabolic pathways in engineered strains. An *E. coli* strain for producing the human hormone leptin (Choi et al., 2003) was enhanced by identifying the bottle-neck in the process -the serine biosynthetic pathway through proteomics and then over expressing a key enzyme to improve production of leptin.

Metabolomics is devoted to identification and quantification of intracellular metabolites in a system using MS, NMR and other analytical techniques (Dunn et al., 2005). Although, this is by far the most important technique that can propel metabolic engineering, it has not been used widely in metabolic engineering, because of technological limitations in detection of small molecules.

1.6.4 Integrating heterogeneous data types

Due to the multi-scale nature of metabolism, it is not enough to have a single high-throughput data to fully characterize the behavior of an organism. Successful integration of multiple high-throughput data types has been achieved in the case of HCDC of *E. coli* for bio product formation (Han et al., 2003). The specific production rate was inversely related to cell density. The transcriptional and proteomic analysis revealed hampered amino acid biosynthesis, which allowed development of new strategies for increasing production. Having an *in silico* framework to integrate these data types will not only allow ease of analysis, but will be helpful in prospective design.

1.7 Developing experimental hosts for metabolic engineering

E. coli has been a model organism for studying many cellular processes. It maintains a strong position as a host for protein production, despite limitations due to its ease of cultivation, rapid growth rate, simple nutritional requirements and ample availability of genetic tools. It was the first bacterium to be engineered using molecular biology techniques to produce human insulin. To enable targeted metabolic engineering, a detailed knowledge of metabolic pathways and control elements for gene expression is necessary. In order to obtain a desired product in *E. coli*, one may have to modify substrate uptake, increase yield of precursor molecules, insert genes for synthesis of desired compound or modify existing genes.

Whole genome sequencing has led to the explosion of annotated genome information and advent of high-throughput technologies has enabled a systematic improvement of industrial microbes through genome engineering. Restructuring of microbial genomes has been shown to have several advantages over conventional approaches for strain improvement. Restructured genomes with the desired functionalities have served as customized industrial strains that display (i) streamlined metabolic pathways for the production of selected biomaterials, (ii) a reduced production of unwanted by-products, and (iii) increased genome stability (Lee et al., 2009).Even a simple biocatalyst, such as the laboratory work- horse *E. coli*, is a complex system of an estimated 4603 genes, 2077 reactions, and 1039 unique metabolites, and while the steps outlined above are relatively straightforward, it is still difficult to quickly and reliably engineer a biocatalyst to perform desired behaviors.

1.8 Experimental technologies for manipulation

There is plethora of techniques available for the perturbing genome and introducing heterologous gene in host for metabolic engineering. Based on heuristic approach, very good biochemists can easily find out, which gene should be knocked out and which gene should be over expressed. Another approach would be to use mathematical tools for decision making. Whatever may be the method, controlling the complexity of biological system while engineering an organism requires rigorous decision making at every stage of central dogma. Including any standard part would require proper justification, a notorious standard part can collapse the project and any standard part falsely predicted may be a burden. Before adding any heterologous expression system into the cell, one needs to knock out any alternate pathway that consumes the precursor for the pathway if it does not affect the growth significantly. If single gene has to be knocked down, a recombineering approach can be used that uses bacteriophage λ encoded Red recombinase to flip the target gene with Antibiotic resistance gene (Murphy, 1998; Muyrers et al., 1999; Zhang et al., 1998). In case more than one knock out is needed high throughput methods like Multiplex Automated Genome Engineering (MAGE) can be used, MAGE incorporates linear DNA into *E. coli* by electroporation, with the λ-Red β protein integrating the oligonucleotides into the genome. Oligonucleotides can be synthesized to introduce mutations at precise genomic loci. Iterated rounds of MAGE introduce increasing amounts of diversity at these loci, although many cells are killed at the electroporation step. The basic Knock out strain developed is now ready to incorporated, the incorporation can be in the form of plasmid of can be inserted into the genome using above mentioned recombineering techniques. MAGE can also be used to replace the RBS sequence of specific property to manipulate the expression of housekeeping genes that cannot be knocked out easily without losing the viability or growth of cells. The riboswitches (Mandal and Breaker, 2004) are another group of RNA mediated control switches, which in the presence on some particular metabolite switches on or off the gene, this leads to steady expression of desired gene, these riboswitches should be included into the 3’ UTR region of heterologous gene and controlling can be done by adding specific metabolite. Gene expression can be controlled at various levels and sub levels and this makes possible the highly controlled metabolic engineering assuring the success of complete task. Genome editing with engineered nucleases (GEEN) encompasses new sets of tools to engineer genome, these include Zinc Finger Nuclease (ZFN), Transcription Activator like Effector Nuclease (TALENs), CRISPER/Cas system and Engineered Meganuclease

Thus synergy of Synthetic Biology with Metabolic engineering brings a paradigm shift in rational strain design. With the reduction in the cost of gene synthesis technology it is easier to make hundreds of standard parts in weeks’ time. It is obvious that such variations do not exist in nature because microorganisms are programmed to make only small amount of chemicals for their own use, to convert these microorganisms into bio-factories, copying these parts from nature will not help to address the shortcoming in metabolic engineering.

1.9 Violacaein – a pigment and a potential drug molecule

(*text in this section partly adapted from Immanuel et al, 2018* (Immanuel et al., 2018))

Violacein is a bacterial bis-indole pigment of commercial interest having antibacterial, antitumoral, antiviral, trypanocidal and antiprotozoan properties (Durán et al., 2007; Durán and Menck, 2001; Ferreira et al., 2004; Queiroz et al., 2012). It is formed by the condensation of two L-tryptophan molecules controlled by the enzymes of a complex biosynthetic pathway (Fig. 1). The impact of the double bonds, conjugation and hydroxyl groups potentially attribute chromophoric properties (Fig. 1B) to final violet colored product violacein of the pathway (Hoshino, 2011). Violacein has been tested to show anti-bacterial (gram positive), antineoplastic and antifungal properties (Durán et al., 2016). Other tryptophan based small molecule therapeutics like rebeccamycin and staurosporine have been reported as important antitumor molecules (Howard-Jones and Walsh, 2006).

The violacein biosynthetic pathway is complex due to a coordinated five gene operon structure in *Chromobacterium violaceum* (Balibar and Walsh, 2006; Hoshino, 2011). The rate limiting step involves the condensation of two molecules of tryptophan to 2-imino-3-(Indole-3-yl) propionate (*vioA)*. Further steps include conversion to protodeoxy violaceinate by *vioB* and *vioE*, followed by conversion to violacein and deoxyviolacein by catalytic activity of *vioD* and *vioC*. The dynamics of violacein biosynthesis are thus dependent on the coordinated levels of transcripts, proteins, action of promoter and ribosome binding sites (RBS) (Lee et al., 2012; Salis et al., 2009). This necessitates minimization of transcriptional noise (relative stability of transcripts) caused by synthesis and degradation of mRNA molecules and increasing the efficiency of translation of mRNAs into proteins by modulating translation initiation rate (dictated in part by mRNA secondary structures). Translational coupling (TC) significantly modulates translation efficiency of individual genes in a multi-gene operon structure within *E. coli* (Lim et al., 2011; Tian and Salis, 2015).

There have been legacy efforts for violacein production using native producers including *Chromobacterium*, *Duganella* spp. and recombinant *Corynebacterium*, *Citrobacter* and *E. coli* hosts (Table S10). The highest level of violacein attained in native producer *Duganella* spp. B2 is 1.62 g/l. Current efforts on producing violacein in non-native hosts include the extensive testing of the potential of *E. coli* (up to 1.92 g/l with genetic variations) as a platform for production of tryptophan-based therapeutics. The strain designs with *vioABCE* operon from *Chromobacterium* produced violacein up to 70 mg/l in batch processes using arabinose as substrate (Rodrigues et al., 2013). A fed-batch process increased violacein by 10 fold. *Corynebacterium glutamicum* spp. (Sun et al., 2016) and *Citrobacter freundii* (Jiang et al., 2010), classical amino acid producing strains have also been engineered to produced violacein.

1.10 The Polyhydroxyalkanoates and Bioplastics

Polyhydroxyalkanoates (PHAs) form a class of natural polyesters, commonly refered to as bioplastics, that many organisms accumulate as intracellular granules to store carbon and reducing equivalents in response to specific environmental conditions. Traditionally there are over 140 possible constituent monomers of PHAs, classified as short chain length (C4 and C5) and medium chain length (C6) hydroxyalkanoates (Berlanga et al., 2006). Different monomer units result in varying polymer composition providing PHAs with diverse material properties. The most common is a polymer of 3--hydroxybutyrate, polyhydroxybutyrate (PHB). Recombinant E. coli for PHA and PHB production (Chen, 2009) has been commercialized also producing a tailored PHA composed of (R)--3--hydroxybutyrate and (R)--3--hydroxyvalerate and/or (R)--3--hydroxyhexanoate for increased elasticity and decreased brittleness (Appleton et al., 2014). Although, the biopolymer synthesis process is complex, the dependency of monomeric composition on the host cell PHA synthase and global metabolic phenotype of the host cell facilitates production of designer PHA polymers. The use of lactate (C3) as a monomer though difficult, can be overcome by increasing substrate specificity of the existing PHA synthases, to even synthesize a PHA with a C3 monomer (polylactic acid). Cellular biosynthetic routes to PHA monomers compete with and/or rely on important pathways such as the tricarboxylic acid (TCA) cycle, fatty acid degradation or ß--oxidation and fatty acid biosynthesis for precursors, involving important central metabolites like AcCoA and cofactors like NADH/NADPH. This highlights the need and makes PHA production a model system for metabolic engineering, allowing control over polymer properties like monomeric composition, chain length, copolymer microstructure, as well as optimizing yield. There are only a few reports on metabolic engineering to produce poly lactic acid (Jung et al., 2010; Yang et al., 2010).

1.11 Examples of Synthetic Biology and Metabolic Engineering

There is much evidence where synthetic biology is applied in order to understand the scope of synthetic biology for rational strain design. Synthetic biology approach is used to engineer *E. coli* to produce n-Butanol. n-Butanol is biosynthetically synthesized by condensing two acetyl-CoA monomer followed by two reductions and dehydration to produce butyryl-CoA. To do this job, PhaAB genes were derived from Polyhydroxyalkanoate biosysnthesis pathway of *Ralstonia eutropha*, Crotonase (Crt) from *C.acetobutylicum*, crotonyl-CoA reductase (Ccr) from Streptomycis Collinus and bifunctional butyraldehyde and butanol dehydrogenase (AdhE2) from *C. Acetobutylicum*. All the genes were optimized for codon usage and synthesized using the methods meentioned previously. By further optimizing the promoters and hosts 30 fold increase in the titers of n-butanol was achieved. One of the most important achievement in the field of synthetic biology is Production of Amorpha-4, 11-diene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin (Tsuruta et al., 2009), the technology is in the bioprocess development stage Amyris Inc. In this effort of synthetic biology, a modified mevalonate pathway was used, and the yeast cells were engineered to express the enzyme amorphadiene synthase and a cytochrome P450 monooxygenase (CYP71AV1), both from *A. annua*. A three-step oxidation of amorpha-4,11-diene gives the resulting artemisinic acid. The other two products 1,4-Butanediol and 1,3 Propane Diol have been taken to commercial realization by Genomatica Inc., and Dupont, Inc.

1.12 Challenges for Synthetic Biology and Metabolic Engineering

To begin with the first challenge is to make an ideal strain that can be starting point of every process. Many researchers predict that the cells with minimal genome and faster growth rate is the key to the success of any strategy, so the key is to find the minimum set of genes required to grow a healthy cell(Hutchison et al., 2016). The assembly of standard parts is another area that needs a constant improvement more recombination compatible systems are necessary to make strategies to clone genes. Another difficulty is plasmid instability in the cell because of metabolic burden to maintain and replicate plasmids, solution for this problem is the genome integration of desired construct that has high stability. There are thousands of standard parts in the labs across the world, to avoid any redundancy in the synthesis of standard parts there should be a common portal to acquire from any part of the world with minimum possible compensation this will accelerate the development of field altogether.

When complex pathways are introduced inside the cell, limitations including intermediate toxicity, low enzyme activity, metabolic burden (cofactor imbalance etc.) need to be overcome for high performance. Such bottlenecks can be addressed using pathway engineering that exploits the synergies of synthetic biology, metabolic engineering and systems biology (Nielsen and Keasling, 2016, 2011; Stephanopoulos, 2012; Wu et al., 2016; Yadav et al., 2012). Successful metabolic engineering for platform cell factories to produce a wide range of fuels and chemicals necessitates identifying the sensitivity of product/process to nutrient precursors and cofactors *a priori.* Such complementation supports coupling of cellular objectives of growth and energy to desired bioengineering objectives. Comprehensive computational strain designs for stoichiometric growth-coupling of desired products of central metabolism have been identified through pathway analysis (Klamt and Mahadevan, 2015; Von Kamp and Klamt, 2017).

1.13 Security and Risk

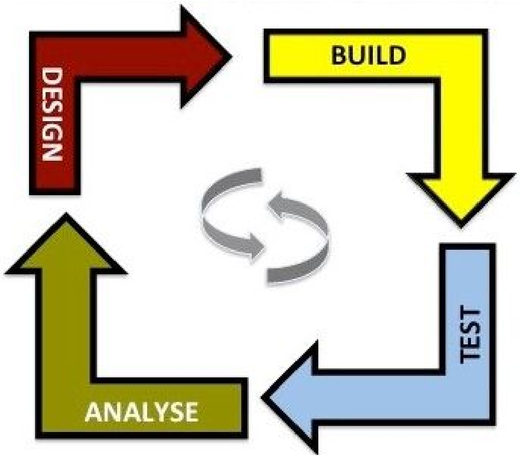
Clearly such a technology in the wrong hands can be a disaster. The deadly pathogen can be created in matter of few days and can lead to bioterrorism. Understanding the risk involved in this field, it is under the tight regulation by security agencies across the world, it is important to share the details of every gene that is synthesized in any corner of the world to stakeholders of world security(IGSG, 2013). It is necessary to practice synthetic biology by the scientists across the world while understanding its potential hazards. Most of the synthetic biology projects use laboratory strains that cannot survive in the wild which prevent the potential environmental hazards from engineered strains. The artificial gene circuit can escape the lab, therefore it is highly recommendable not to practice synthetic biology on pathogenic organism and with medical procedure unless the approved as safe. It should be taken care that synthetic genes should not escape the boundaries of laboratory; it is a responsibility of every synthetic biology labs to practice proper waste disposal methods.

1.14 Overview of the Thesis

The transition from traditional processes that use petroleum feedstock to bioprocesses and microbial cell factories is imminent. The main challenges include (i) development of new biological strain designs to produce commercial products, and (ii) the optimization of processes to make them commercially viable. The work described in this thesis focuses on the first challenge by developing rational strain designs for metabolic engineering of value added products using synthetic and systems biology approaches. The strategy for strain design depends on chemical nature of the molecule of interest and its pathway design The value added products discussed in this thesis are (i) a violet pigmented drug molecule violacein and (ii) a bioplastic, Poly Hydroxy Butyrate (PHB).

The main highlights of the work executed on developing strategies using synthetic biology and systems biology for metabolic engineering include:

- Design and development of a synthetic construct producing violacein
- ΔtrpR, Δ*pheA* and Δ*pgi* gene deletions in *E. coli* increased tryptophan levels as predicted *in silico*
- Constraints based modeling identified precursor tryptophan and cofactor NADPH as limiting
- A synthetic construct designed by vioABCDE operon sequence optimization was engineered in *E. coli*
- Genome engineering of deletions ΔtrpR, Δ*pheA* and Δ*pgi* in *E. coli* for increased tryptophan levels
- Constraints based modeling identified precursor tryptophan and cofactor NADPH as limiting



- Development of a novel scalable absolute quantitation method for estimation of PHB
- Design and development of a synthetic construct producing PHB
- Development of a genome integrated PHB circuit in host strain of *E.coli*
- Integrative design strategy and systems analysis increasing violacein and PHB yields
- A Violacein pathway module and PHB module was reconstructed in genome scale model of *E. coli*

1.15 Thesis Organization

The thesis is organized into 5 Chapters.

Chapter 1 is the Introduction that discusses the motivation and specific aims and objectives of the thesis along with a brief literature review. The literature review will cover basic aspects of synthetic biology that will make a case for its use in metabolic engineering with the help of metabolic systems biology. In addition, the work till date on violacein and PHA production using biological processes will be discussed.

Chapter 2 on Methodology discusses the methods used in this thesis. The work discussed in the thesis is based on the Design-Test-Build-Analyze paradigm. All methods pertaining to each step in this paradigm are elaborated in this chapter. A novel scalable method for quantitation of PHB developed in this thesis is also described here.

Chapter 3 is based on the application of the methods discussed to rational strain designs and engineering of a violacein pathway in *Escherichia Coli.* This chapter highights the synergistic use of algorithms for operon sequence optimization and mechanistic modeling in metabolic engineering of violacein. It underscores the importance of metabolic modeling in identifying rate limiting precursors and cofactors.

Chapter 4 is based on circuit development to augment's *E. coli* native metabolism and make PHB. The chapter discusses not only how glycerol is an optimal substrate for PHB synthesis but also the integration of the operon into the genome of E. coli. This eliminates the need for antibiotic selection markers and make these strains GRAS for industry.

Chapter 5 recapitulates the work done and discusses some of the future prospects of microbial cell factories.

Chapter 2

Methodology: Design, Build, Test, and Analyze

The technology of synthetic biology is currently accelerating at four times the rate of Moore's law.
Steward Brand

There nearly always is a method in madness, its what drives men mad, being methodical.
G.K. Chesterton

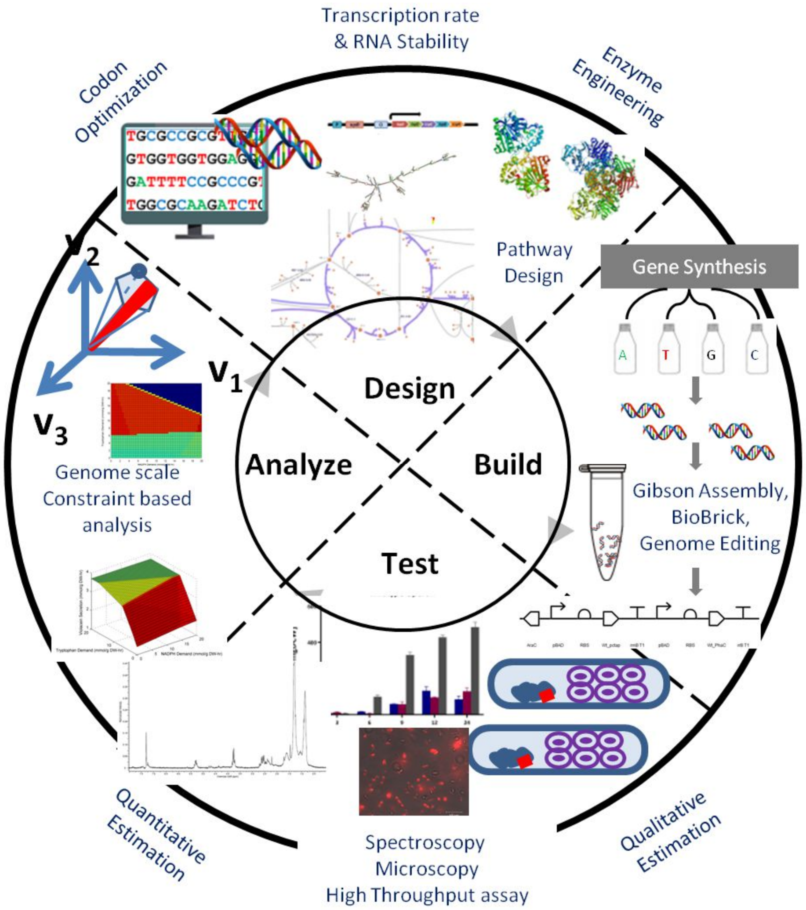

Transcription rate
& RNA Stability
Codon
Optimization
Enzyme
Engineering
TGCGCCGCGT
GTGGTGGTGGAG
GATTTTCCGCCCGT
TGGCGCAAGATCTC
Pathway
Design
Gene Synthesis
A
T
G
C
Gibson Assembly,
BioBrick,
Genome Editing
V2
V1
V3
Design
Analyze
Build
Test
Genome scale
Constraint based
analysis
Quantitative
Estimation
Spectroscopy
Microscopy
High Throughput assay
Qualitative
Estimation

Figure 2.1 Synthetic Biology Paradigms

This chapter discusses the methodology developed and used in this thesis for rational strain design using a synthetic biology approach. The synthetic biology paradigm discussed in Chapter 1 relates to the Design-Build-Test-Analyze iterative sequence (Figure 2.1). Methods corresponding to each of these steps as used in metabolic engineering of violacein and PHB are discussed here.

2.1 Design: Operon, Genetic Circuit and Pathway Design

The first step of the synthetic biology paradigm is “Design”. Design in here describes the re-design of gene sequences that are part of circuits and pathways that are desired. Heterologous expression of genes for metabolite production is an extensive engineering process. Engineering already existing natural standard parts by optimizing its DNA sequence is a common approach to engineer standard parts for synthetic biology. Sequences for many natural as well as engineered parts are available in online repositories like http://parts.igem.org/Main_Page and https://www.addgene.org. Laboratories across the world are adding new designs every day to such repositories which includes promoters, ribosomal binding site (RBS), genes, terminators, antibiotics resistance cassettes, selection markers and origin of replication (Ori). Every sequence is with its description and references suggesting its potential use. It is not necessary to use engineered standard part every time, sometimes a natural sequence can serve the purpose, but combination of standard parts which makes a module should support the objective of using every part.

A synthetic operon or genetic circuit is constructed out of standard parts, and the global impact of synthetic and wild type genes on cell metabolism can be predicted using genome scale metabolic models (GSM) (Chubukov et al., 2016; Immanuel et al., 2018; Jung et al., 2010). For Example, the production of value added product requires reducing power that can be generated in excess and used exclusively for product biosynthesis (Charusanti et al., 2010; Lee et al., 1996b; Spaans et al., 2015). Thus GSM is used to perform the systems level predictions for the optimal balance between growth and production of metabolite. Several iterative predictions can be performed to fine tune the final synthetic sequence and design before its chemical synthesis and pathway design. Adding an appropriate promoter can only provide high mRNA copy number but the codon usage has a profound effect on the transcription and translation independently

(Zhou et al., 2016) . Heterologous genes inside its host may not use synonymous codons with same frequency (Grantham et al., 1980) which results into poor expression; in fact there is direct correlation between a codon preference and levels of its cognate t-RNA in cells (Ikemura, 1981; Nakamura et al., 1999). At this point synthetic biology plays a crucial role of alleviating the effect of codon bias, since while designing the gene for heterologous expression, rare codons can be replaced with the more frequent codons. The degeneracy of codons allows almost one third of the codons replace with the synonymous codons without affecting the primary structure of translated protein. The minimum number of codons that can be used to express a protein containing all 20 amino acid is 20 and maximum is 61, this is known as effective number of codons (Nc) (Wright, 1990). Several software and online tools are available to optimize sequence according to the specific host (Grote et al., 2005; Jung and Mcdonald, 2011; Raab et al., 2010; Villalobos et al., 2006)' all the approaches used various algorithms and software platform to optimize genes, but the basic biological objective remains the same. Apart from codon bias, GC content, tandem repeats and inverted repeat, DNA-protein interaction motifs, cryptic splice sites, restriction sites, promoters and ribosome binding sites (RBS) which can negatively affect the transcription, stability of RNA and translation, should be removed. Although not all the parameter can be scored to 1.0 (100%) in a given sequence but priority can be provided to the parameter which is most important.

2.1.1 Sequence Optimization

The main parameters that need to be optimized include codon bias, GC content and free energy of folding. The other issues are related to restriction site and ribonuclease site minimization and elimination as desired. The following sections discuss these parameters in detail.

Codon Bias: To decide the most appropriate codon(s) for each amino acid, codon frequency in the highly expressed gene such as genes for ribosomal protein and housekeeping genes can utilized .The most popular parameter in scoring the alternative triplet codon throughout the sequence is Codon Adaptive Index (CAI), it is geometric mean of relative adaptiveness of codons. Relative adaptiveness is the ratio of frequency of a codon for an amino acid to frequency of most frequent codon for an amino acid in highly expressed genes(Raab et al., 2010). CAI for a gene sequence approaching 1.0

suggests the similarity with highly expressed genes with respect to relative codon frequency.

Table 2.1 Codons preferred in *E. coli* for high expression of genes

UUU	F	0.57	UCU	S	0.11	UAU	Y	0.53	UGU	C	0.42
UUC	F	0.43	UCC	S	0.11	UAC	Y	0.47	UGC	C	0.58
UUA	L	0.15	UCA	S	0.15	UAA	*	0.64	UGA	*	0.36
UUG	L	0.12	UCG	S	0.16	UAG	*	0.0	UGG	W	1.00
CUU	L	0.12	CCU	P	0.17	CAU	H	0.55	CCU	R	0.36
CUC	L	0.10	CCC	P	0.13	CAC	H	0.45	CCC	R	0.44
CUA	L	0.05	CCA	P	0.14	CAA	Q	0.30	CCA	R	0.07
CUG	L	0.46	CCG	P	0.55	CAG	Q	0.70	CCG	R	0.07
AUU	I	0.58	ACU	T	0.16	AAU	N	0.47	ACU	S	0.14
AUC	I	0.35	ACC	T	0.47	AAC	N	0.53	ACC	S	0.33
AUA	I	0.07	ACA	T	0.13	AAA	K	0.73	ACA	R	0.02
AUG	I	1.0	ACG	T	0.24	AAG	K	0.27	ACG	R	0.03
GUU	V	0.25	GCU	A	0.11	GAU	D	0.65	GCU	G	0.29
GUC	V	0.18	GCC	A	0.31	GAC	D	0.35	GCC	G	0.46
GUA	V	0.17	GCA	A	0.21	GAA	E	0.70	GCA	G	0.13
GUG	V	0.40	GCG	A	0.38	GAG	E	0.30	GCG	G	0.12

In a different approach Individual Codon Usage (ICU) and Codon Context (CC) is used (Chung and Lee, 2012) , ICU distribution is calculated for each amino acid in highly expressed genes using Pearson Chi-Square test, using a p-value cut off of 0.05 to determine the amino acid with biased ICU. For optimizing a gene for the expression, its ICU distribution has to be adjusted to the host ICU distribution. Similar Approach is used for optimization using Codon Context CC distribution but in this case codon pair is considered in place of single codon, codon context is an effect of adjacent codon on the rate of translation due to tRNA-tRNA stearic interaction within the ribosome, in many cases it helps in the translation of rare codon when placed with a particular partner codon.

GC Content: GC content is the frequency of Guanine and Cytosine (GC) nucleotide in gene or a genome, optimal GC content (40-60%) is necessary for number of reasons in gene synthesis. Firstly the gene synthesis method requires short oligos (200 Nt to 300 Nt) to be assembled and it is necessary that oligos should have similar melting temperature for assembly reaction, for this reason optimized gene should not have spikes of high GC and AT rich region for error free assembly reaction. Secondly high GC content can lead to very stable RNA secondary structure affecting translation rate. Deviation from equal GC to AT content often leads to genetic instability. Unlike codon optimization GC optimization is done for longer stretch of sequence considering window of 40 nucleotides. Every time the codon is added while optimizing codons a score can be calculated for the GC content of 40 nucleotides keeping the added codon at the centre, this way by maintaining consistent GC content throughout the codon optimization will always keep GC content in check.

Tandem repeats and Inverted repeats: Tandem repeats are long been known to be associated with recombination events leading to the genetic instability of genes whereas, inverted repeats leads to stable mRNA secondary structure (hairpin loops and loop stem secondary structure). To avoid any tandem repeats each window with 40 nucleotides can be aligned and to itself and its subsequent windows and one with the higher alignment score can be altered to minimize the alignment score. Similarly inverter repeats can be scored but every window should be converted into reverse complementary before performing alignment.

DNA-Protein interaction motifs: Some motifs can be recognized by certain protein, normally it has a specific purpose for example restriction sites helps in eliminating any viral DNA by cleaving at specific recognition site; Cryptic splice site in case of eukaryotic cells are involved in splicing mRNA; Promoters and RBS are recognized by transcription and translation complex respectively. Certainly, such sites in synthetic gene will have a deleterious effect on the expression of proteins. Simple search tool can be used to find and eliminate library of known motifs for the host systems.

2.1.2 Gene optimization and Pathway design for Violacein and Poly(R)-3-Hydroxybutyrate

Sequence optimization for genes *vioA, vioB, vioC, vioD and vioE* that form the operon was performed to design Synthetic Violacein Operon (SYNO) and stitched into a plasmid vector for experimental testing. The SYNO design increased Codon Adaptive Index (CAI) values (Figure 2.2 C), optimum GC content (Figure 2.2 D) and lowered effective codon usage (Nc) (Figure 2.2 E) as compared to Wild Type Operon (WTO) derived from *C. violaceum.* The change in CAI was concomitant with optimum GC content (CAI: 0.9, GC content: 50%) to match *E. coli* genome characteristics. The design space explored did not deviate in operon structure, gene order, intergenic sequence and retained *vioA* (CviR Box) promoter control for mRNA transcription (**Error! Reference source not found.** A,B) The design used in this study was compared to the iGEM design using biobricks (Part:BBa_K274002) modified versions of which have been used for violacein production (He et al., 2017) in other studies. RBS Calculator (Espah Borujeni et al., 2014; Salis et al., 2009) (available at http://voigtlab.ucsf.edu/software) estimated increased (minimally by 5 fold) translation initiation rates (TIR) (Figure 2.2 F) and low translational coupling factor (Figure 2.2 G) for SYNO that allows for higher protein abundances and activities as compared to WTO. WTO potentially could couple the translation of three transcripts *vioB*, *vioC* and *vioD* using one single ribosome. This was based on the 3-fold lower free energies (ΔG) of transcription (Figure 2.2 H) and 20-fold higher TIRs for the tricistronic mRNA (*vioBCD*) *vis a vis* single mRNA structures. This is in contrast to critical mRNA features of SYNO using multiple ribosomes for each coding region. The mRNA secondary structure around the initiation region in the synthetic operon including the high- and low-folding energy hairpin structures can shape translation efficiency (Figure 2.3). These are reflected in the magnitude of folding free energies as represented (Figure 2.2 H). The analysis shows that *vioA* (that synthesizes the first intermediate of pathway from tryptophan) is rate limiting for the WTO structure, while *vioD* (that commits the cell to violacein biosynthesis using NADPH) is limiting in the SYNO structure.

For synthetic plasmid construction for PHB production, genes were optimized using GeneOptimizer (Thermo Fisher Scientific) that uses sliding window algorithm (Raab et al., 2010). Genes were optimized for maximum CAI, optimum GC%. Sequence optimization also involved removal of tandem and inverted repeats along with removal of common restriction sites. Genes were re-examined using Visual Gene Developer to compare wild-type genes with synthetic genes for CAI value, GC% and additionally for effective number of codons (Nc) (Jung and Mcdonald, 2011). The promoter sequence was

For synthetic plasmid construction for PHB production, genes were optimized using GeneOptimizer (Thermo Fisher Scientific) that uses sliding window algorithm (Raab et al., 2010). Genes were optimized for maximum CAI, optimum GC %. Sequence optimization also involved removal of tandem and inverted repeats along with removal of common restriction sites. Genes were re-examined using Visual Gene Developer to

compare wild-type genes with synthetic genes for CAI value, GC% and additionally for effective number of codons (Nc) (Jung and Mcdonald, 2011). The promoter sequence was analyzed for Translation Initiation Rate (TIR) using RBS calculator (https://salislab.net/software/) and Minimum Free Energy (secondary structure) secondary structure was predicted using mfold (http://unafold.rna.albany.edu/). Both the genes are expressed through pBAD promoter which is induced by L-arabinose. Unlike natural operon structure, synthetic biology approach was used to create the genetic circuit. Both monocistronic design and polycistronic design were investigated using RBS calculator (https://salislab.net/software/) and monocistronic design was adopted which showed 1:1 transcription coupling for both the gene and controlled Translation initiation rate for both the genes

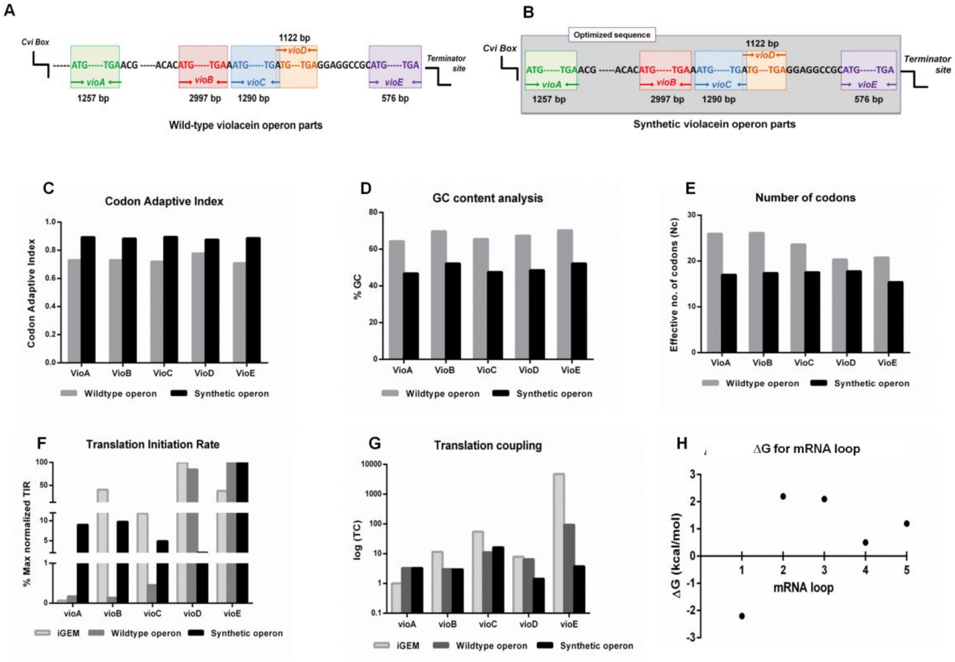


Figure 2.2 Design of synthetic vioABCDE operon (SYNO) and gene sequence analysis

The native operon sequence from *C. violaceum* was used as WTO (A). The optimized synthetic operon SYNO (B) with DNA sequence indicators. The design included optimized Codon Adaptive Indices (C), GC content (D) and Effective number of codons (E). Translation initiation rates (TIR) for wild-type and synthetic operon (F). All normalized values are shown. (G) Translation coupling (TC) for wild-type and synthetic violacein operon. Translation efficiency represented by folding energies for stem loops in mRNA (H) and hypothetical hairpin structures

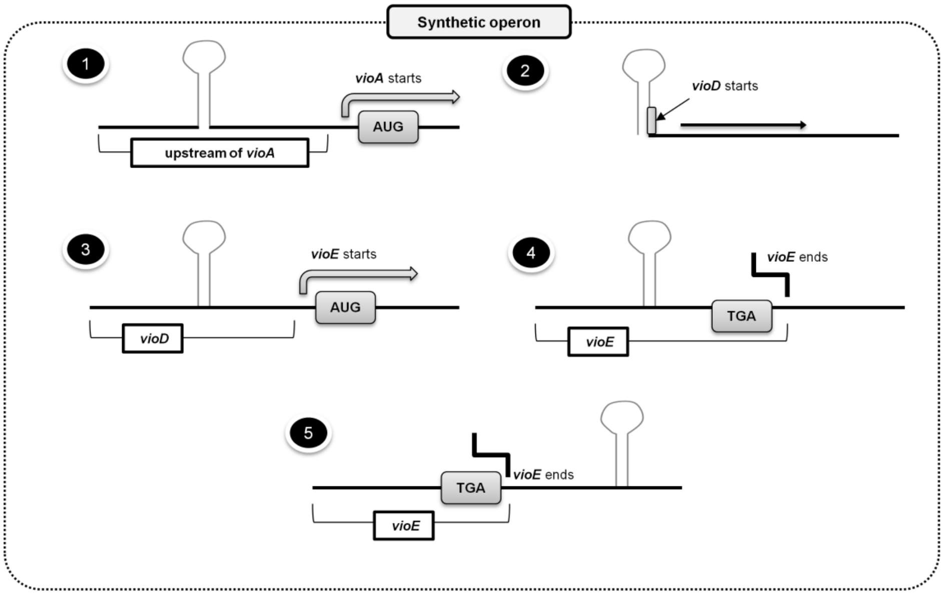


Figure 2.3 hairpin loop around the mRNA initiation codon (AUG) of the synthetic operon SYNO

2.2 Pathway Design for Violacein and PHB production

Two separate strategies were used for the Violacein and PHB production, Violacein being the linear pathway starting from two molecule of Tryptophan is converted into violacein. The metabolic requirement for the violacein synthesis is to generate precursor and reducing power. Later objective was achieved by knocking out *pgi* (Phosphoglucose isomerase) in *E.coli K* 12 .This diverts the glycolytic flux through pentose phosphate pathway (PPP) and marginally through Entner-Doudoroff pathway for glucose catabolism. PPP pathway generates 2 molecules of NADPH per glucose molecule, while growth is reduced significantly in *E.coli K 12:Δpgi* due to redox imbalance and reduced glucose uptake rate. *Δpgi* bypasses redox imbalance by activating glyoxylate shunt instead of TCA cycle, towards gluconeogenesis and partially re-establishes the redox and energy balance. Other two knockouts i.e *ΔpheA* (prephenate dehydratase) and *ΔtrpR* (Tryptophan operon feedback regulatory operator) increases precursor yield. *ΔpheA* eliminates competing reaction for tryptophan which generates two other aromatic amino acid, phenyl alanine and tyrosine. *ΔtrpR* removes the feedback inhibition from tryptophan operon expression (Figure2.4 A).

Pathway design for PHB production has a different strategy, the precursor for PHB synthesis is Acetyl-CoA and *E .coli* generates Acetyl-CoA through carbohydrate metabolism, TCA cycle and fatty acid metabolism. Apart from this *E. coli* has a partial pathway PHB synthesis making (R) 3-hydroxy butyrate (3HB) which is the immediate precursor for Polyhydroxyalkanoate Synthase (PhaC). Thus unlike violacein pathway the entire PHB synthesis operon is not necessary to be incorporated into the pathway design. Therefore pathway design for PHB synthesis requires two genes, pct_{ap} and $phaC_{cv}$ to be incorporated into *E.coli* metabolism to augment the incomplete PHB synthesis pathway. Pct_{ap} diverts the flux to towards acetyl-CoA generation while PhaC polymerizes the 3HB (Figure2.4 B).

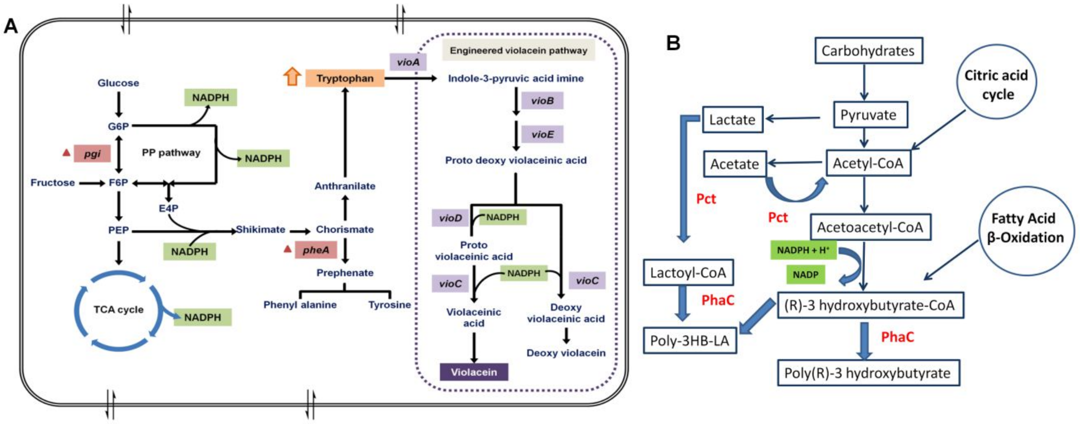

A
Glucose
NADPH
G6P
pgi
PP pathway
NADPH
Fructose
F6P
E4P
PEP
Shikimate
Chorismate
NADPH
pheA
Prephenate
Phenyl alanine
Tyrosine
TCA cycle
NADPH
Tryptophan
Anthranilate
vioA
Engineered violacein pathway
Indole-3-pyruvic acid imine
vioB
vioE
Proto deoxy violaceinic acid
vioD
NADPH
Proto violaceinic acid
vioC
NADPH
vioC
Violaceinic acid
Deoxy violaceinic acid
Deoxy violacein
Violacein
B
Carbohydrates
Citric acid cycle
Pyruvate
Lactate
Acetate
Acetyl-CoA
Pct
Pct
Acetoacetyl-CoA
Fatty Acid β-Oxidation
NADPH + H⁺
NADP
Lactoyl-CoA
PhaC
(R)-3 hydroxybutyrate-CoA
Poly-3HB-LA
PhaC
Poly(R)-3 hydroxybutyrate

Figure2.4 Pathway design for (A) Violacein and (B) PHB biosynthetic pathway

2.3 Build: Gene Synthesis, Cloning and Strain design

The build phase of the synthetic biology paradigm involves using the DNA parts designed above and stitching them together to create circuits and pathways for expression. There are many traditional molecular biology methods that have been established that will be discussed in this section in the context of building systems to make violacein and PHB. Some more recent protocols that have been developed to build larger pathways (violacein) will also be discussed. In applications of metabolic engineering, antibiotic markers are not considered safe for strain designs to be used in industry. We also discuss a protocol of genome integration for PHB strain design. The cloning methodology has advanced with to a large extent which uses not only restriction digestion based cloning but also the recombination based cloning(Gibson et al., 2010).

2.3.1 Cloning and expressing pathways and circuits

Traditional cloning techniques based on restriction enzymes are limited to 1 or 2 fragment assembly, synthetic biology requires large no of standard parts with varying size to be assembled. Several recombination based assembly methods are developed, both *in-vitro*(Gibson et al., 2009) and intracellular in yeast(Gibson, 2009). The first two genome synthesis of Polio Virus(Cello et al., 2002) and ϕX174(H O Smith et al., 2003) genome sequence did not involve any cell mediated assembly steps since their genome is small. To go beyond the size limit of above mentioned in-vitro method, it is necessary to involve living cells two usual choices would be *E. coli* and yeast. Even if larger sequence can be assembled using oligos the assembly will have lot of errors and desired sequence will be few in number in the heterogeneous mixture of erroneous fragments that are difficult to isolate from rest of the sequences. To overcome this, PCA is used to make smaller fragments called synthons (500–700bp)(Kodumal et al., 2004) these are cloned into *E. coli* and several clones are screened by sequencing to find out the error free sequence. Synthons can be further assembled using enzymatic assembly using cocktail of three enzymes; T5 exonuclease , DNA polymerase and Taq ligase(Gibson et al., 2009). At a time 4 to 10 stynthons can be assembled, desired no. of synthons are mixed with Assembly buffer, dNTP mix and Enzyme cocktail. The T5 exonuclease chews away 5' fragment leaving 3' single stranded overhang at 50°C, but at 50°C it is not stable and gets deactivated preventing longer 3' overhang. Synthons having end homology anneals and polymerisation is carried out at 50°C, this is followed by ligation with thermostable Taq ligase at 50°C. Since all the reaction takes place at 50°C this method does not require any specialized instrument like thermo cycler. One of the synthons can be a linearized plasmid so that the entire construct can be transformed to get circular clones that can be authenticated using sequencing. The scale of synthesis is improved to an extent that, first bacterial genome has been synthesized using above method. The choice of organism is *Mycoplasma genitalium* which has smallest genome among bacteria (580kb)(Gibson et al., 2010), and the assembly involving longer sequence were assembled using yeast as a host for cloning purposes, since larger sequences are unstable in the *E. coli* in the form of plasmid. Currently genome of *Saccharomyces cerevisiae* is being synthesized with a global mission in action to make synthetic yeast Sc2.0, until 2017 five chromosomes are already synthesized(Richardson et al., 2017)

2.3.2 Synthetic and wild-type plasmid construction for expression of violacein operon

Single *C.violaceium* ATCC 12472 (Sequence ID: AE016825.1) (Haselkorn et al., 2003) colony was inoculated in 5ml Luria Bertani Broth, Miller (HiMedia Cat no. M1245) and incubated overnight at 30°C at 200 rpm shaking. Genomic DNA was isolated from *C.volaceium* using Qiagen DNeasy Blood & Tissue Kits (Cat. no. 69506).Violacein biosynthesis wild type gene cluster was PCR amplified with Q5 High-Fidelity DNA polymerase (Cat. no. NEB M0491) on Applied Biosystems Veriti® Thermo Cycler using VA_FW, VA_RV,VB_FW,VB_RV,VCD_FW and VCD_RV(*Appendix I* Table II) primers, into three standard parts VioA, VioB and VioCDE of 3231bp, 3010bp and 3534bp repectively, comprising 5 violacein biosynthetic genes. Final 50µl of PCR mix contained 0.5µM of each primer, 200µM of dNTP, 50 ng of *C.volaceium* genomic DNA. Q5 5X Buffer and GC enhancer was added as recommended by supplier. Gene synthesis service was provided by GeneArt gene synthesis service from Thermo Fisher Scientific. Four plasmids carrying Violacein biosysnthesis genes cloned in pMK-RQ(KanR) at sfiI/sfiI site were received in the lyophilized form. Plasmids were reconstituted in T.E buffer and used as template to PCR amplifiy synthetic gene fragmants for Gibson Assembly. Synthetic gene fragments were amplified with Accuprime™ Pfx Supermix Thermo Fisher Scientific(Cat. no.12344040) using VA_FWSyn, VA_ RVSyn, VB_FWSyn, VB_RVSyn, VC_FWSyn, VC_RVSyn, VDE_FWSyn and VCD_RV (*Appendix II* Table II) primers into 4 Synthetic standard part, VioASyn, VioBSyn, VioCSyn and VioDESyn of 2659bp 1751bp, 2917bp,1323bp and 2204bp respectively . Final 50µl of PCR mix contained 0.5µM primers, 50 ng of *C.volaceium* genomic DNA and PCR mix as per the recommendation from supplier. pUC19 vector(Thermo Fisher scientific Cat. no. SD0061**)** was used for the construction of both recombinant and synthetic construct. pUC19 was double digested with KpnI/PstI for one hour at 37°C(pUC19L). Digested vector and gene fragments were gel purified using Qiagen MiniElute® Gel Extraction kit (Cat no. 28604) and purified fragments were used for Gibson assembly to construct pCVW and pSYN (Figure 3.6) pwith vioABCDE wild type operon (WTO) and synthetic operon (SYNO) respectively. Gibson assembly(Gibson et al., 2009) was carried in 10ul reaction mixture containing 40 ng pUC19L and 2 fold excess of gene fragments, the assembly mix was incubated at 50°C for 60 min and stored at -20 for subsequent transformation. Chemically competent E.coli DH10β were transformed with assembly mix and cells were platted on Luria Bertani Agar, Miller (HiMedia Cat no. M1151) with ampicillin (100µg/ml) selection.

2.3.3 Construction of *E. coli* knockout strains

The knockout strains were created using the standard method using λ Red recombinase system. It is a recombineering technique that allows the precise generation of insertions, deletions, and point mutations at loci specified by flanking homology regions of as few as 35 bp(Sharan et al., 2009) . Lambda Red, as well as the similar RecET system , is capable of modifying *Escherichia coli* genome, plasmid(Thomason et al., 2007; Yu et al., 2000; Zhang et al., 1998)and BAC(Muyrers et al., 1999) targets using either dsDNA cassettes or ssDNA oligonucleotides. The Quick and Easy *E. coli* Gene Deletion Kit (Gene Bridges) was used to generate Δ*pgi*, Δ*trpR* and Δ*pheA* strains as per manufacturer's protocol. List of primers used are listed *(Appendix II Table I)*.

2.3.4 Construction of wild type and synthetic genetic circuit for PHB production

A synthetic genetic circuit (SynPHB) was designed with two sequence optimized gene for biosynthesis of PHB in *E.coli* where each genes are controlled by separate pBAD promoter and rrnB T1 terminator and placed in tandem in vector backbone with ampicillin resistance cassette (AmpR), AraC ORF and pBR322 Ori. The The synPHB was chemically synthesized and cloned into pBADHis_C_A238 *(Appedndix II Table 2*) vector double digested with NcoI/hindIII to construct pPHB_Syn (Figure. 4A). Apart than vector backbone pPHB_Syn is composed 3 standard parts i)sequence optimized Propionyl-CoA Transferase (Syn_pct$_{ap}$) from *Acetobacter pasteurianus* ii) sequence optimized Polyhydroxyalkanoate synthase (Syn_phaCcv) from *Chromobacterium violaceum* iii) rrnB T1 terminator-araBAD Promoter (Ter-Pro) . The wild type genetic circuit (WtPHB) was created using Gibson assembly (Gibson et al., 2010, 2009)which is a recombination based cloning. Wild type Propionyl-CoA Transferase (Wt_pct$_{ap}$) from

Acetobacter pasteurianus 386B (Sequence ID: HF677570.1) was chemically synthesized and cloned into pBADHis_C_A238 vector double digested with NcoI/hindIII to construct pBAD_Pct (*Appendix II Table 2*). Satandard part rrnB terminator-araBAD Promoter (Ter-Pro) and Polyhydroxyalkanoate synthase (WT_phaCcv) from *Chromobacterium violaceum* ATCC 12472 (Sequence ID: AE016825.1) was PCR amplified with overhang for Gibson assembly from *Chromobacterium violaceum* genomic DNA and pPHB_Syn with primers listed in (*Appendix II Table 3*) and cloned into pBAD_Pct vector double digested with KpnI/HindIII to construct wild type plasmid (pPHB_Wt) Gene synthesis services were provided by GeneArt Gene Synthesis services and DNA was received in lyophilized form and handled as per service provider instructions . Gibson assembly was carried in 10ul reaction mixture containing 40 ng pBAD_Pct digested with KpnI and HindIII and 2 fold excess of Wt) phaCcv and Ter-Pro , the assembly mix was incubated at 50°C for 60 min and stored at -20 for subsequent transformation. Chemically competent E.coli DH10β were transformed with assembly mix and cells were platted on Luria Bertani Agar, Miller (HiMedia Cat no. M1151)with ampicillin (100µg/ml) selection. pPHB_Wt (Figure 2.4 G) was confirmed with restriction digestions with ScaI followed by sequencing.

2.3.5 Genome Integration

The WtPHB and SynPHB was cloned in pGRG36 plasmid to construct pGRGSyn, pGRG36 was a gift from Nancy Craig (Addgene plasmid # 16666). pGRGSyn was transformed in *E.coli* K12 MG1655 and single colony was inoculated in Luria Bertani (LB) Broth, Miller (HiMedia cat no. M1245) with 100 µg/ml ampicillin and inoculated at 30°C; 200 rpm shaking. Three colonies were selected and streaked out on LB agar plates with ampicillin. Isolated colonies were inoculated in LB Broth and allowed to grow for overnight at 30°C; 200 rpm shaking, without adding ampicillin. 50 µl of Culture diluted to 10^{-7} dilution factor was spread on the LB agar plate ampicillin sensitive colonies were selected by replica platting and the genome integration was confirmed by colony PCR with pctRT_Fw and phaCRT_Rv primer (Appendix II Table. II) and sequencing

2.4 Test: Qualitative and Quantitative Estimation

Testing the systems built is critical to gauge the success of the first two steps of the paradigm. Testing can be accomplished through simple screening assays based on simple visual inspection or microscopy that are qualitative measurements to more sophisticated and quantitative measurements using NMR Spectroscopy. This section covers the methods used in this thesis for testing. Previously published methods for qualitative and quantitative assessment of violacein are discussed. Although qualitative assessment of PHB is possible, there were no scalable quick and accurate methods for quantitation. Such a method developed for PHB quantitation using spectrofluorometry.

2.4.1 Qualitative and Quantitative estimation of intracellular Violacein

Absorbance scan for standard violacein (Sigma V9389-1 mg) in ethanol was measured for 0.2 mg/ml in ethanol using UV-Vis spectrophotometer Pharo 300 and absorption with absorbtion maximum at 550 nm (Figure 2.5). Standard curve was developed by measuring the absorbance of Violacein in ethanol in conc. Range of 0.01-0.05 above this range(Figure 2.5).1 mL of culture was centrifuged at 14,000g for 20 min and the pellet was re-suspended in equal volumes of absolute ethanol to extract violacein. The supernatant was collected after centrifugation at 12,000 g for 10 min. The concentration of violacein was determined by monitoring its absorbance at 550 nm. Quantitation was done using calibration curves of a violacein standard.

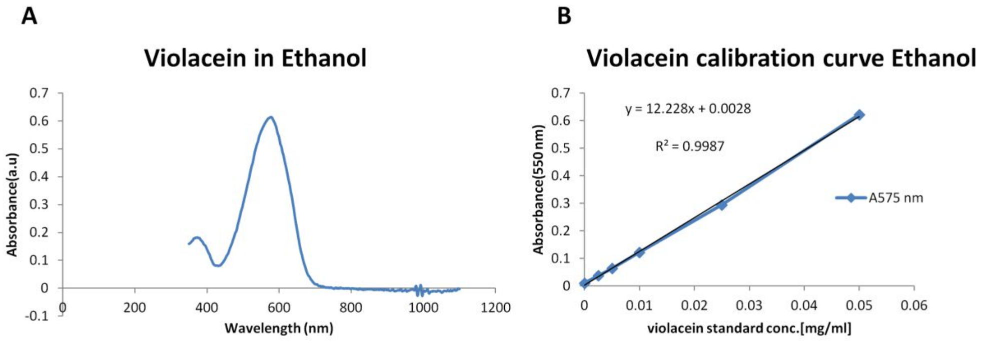


Figure 2.5 Establishing standard curve for estimation of intracellular violacein

A) Absorbance scan B) Standard curve

2.4.2

2.4.3 Qualitative estimation of intracellular Poly (R) 3-hydroxybutyrate

2.4.3.1 Expression of PHB producing operon

Single colony of a *E.coli* DH10β transformed with pPHB_Syn and pPHB_Wt was inoculated in 10 mL LB media and incubated overnight at 37°C with shaking at 200 rpm. Overnight culture was sub-cultured in 10 mL LB media supplemented with 2 mg/mL glycerol. Culture was induced with 0.025% L-arabinose to express PHB producing genes after 3 hours of growth. 1 mL culture was harvested by centrifugation at 10000 rpm for 5 min, 9 hours after induction and re-suspended in 1 mL of PBS while adjusting cell density (OD_{600}) to 1.0. Cells were washed by centrifugation at 10000 rpm for 5 min with 20 mM Tris-HCl (pH: 8.0) and re-suspended in 100 µL lysis buffer (20 mM Tris-HCl and 2 mM EDTA containing 1% Triton X-100). Lysozyme (1 mg/mL) was added to the cells in lysis buffer and incubated with shaking (500 rpm) at 37°C for 1 hour. The soluble fraction of cell lysate was analyzed on an Agilent Bioanalyzer protein 80 chip. Soluble fraction did not show the over expression of $phaC_{cv,}$ since it is an insoluble protein in its active form. Therefore complete cell lysate was prepared by heating cell pellet in 100 µL 1X SDS-PAGE sample buffer (80 mM Tris-Cl pH: 6.8, 2% SDS, 10% glycerol, 5 % β-Mercaptoethanol and 0.002% Bromophenol blue) for 5 min at 95°C. Cell lysate was centrifuged at 8000 rpm for 5 min and 10 µL loaded on 10 % SDS-PAGE gel in Tris-glycine running buffer pH: 8.3 (25 mM Tris-Cl, 250 mM Glycine, 0.1 % SDS).

2.4.3.2 NMR Spectroscopy

Transformed *E.coli* DH10β cells were cultured in 800 mL LB media supplemented with glycerol and glucose as substrate. The cells were harvested after 24 hrs for extraction of PHB using combination of Trypsin-sodium hypochlorite method and solvent extraction. The solvent mixture was evaporated in a Rotavapor® R-100 and the extracted polymer was dissolved in $CDCl_3$. NMR experiments were carried out at 25 °C on Bruker Avance 200 and Bruker Avance 500 spectrometers, using BBO probes with z-gradients. ^{1}H, ^{13}C and ^{13}C DEPT spectra were recorded with 32, 10000 and 3200 scans respectively. Flip angle of 30° and interscan delay of 3 s was employed for 1D ^{1}H and ^{13}C spectra. The diffusion filtered NMR spectrum was recorded using pulsed field gradient stimulated echo experiments, setting diffusion time to 80 ms; encode/decode gradient pulses to 2300 µs and a gradient strength of 48 G/cm. The high gradient strength is sufficient to filter out signals from all low molecular weight species in the solution, however the signals from

PHB is observed in the spectrum due to the slow diffusion of the polymer. The diffusion coefficient determined for PHB is ~ 2.35 × 10^{-11} m^2/s which is typical for polymeric species. The 2D correlation spectroscopy (COSY) experiment was recorded with 2K × 128 data points and 16 scans. The connectivity pattern between the different types of protons clearly shows that the polymeric species is PHB.

2.4.3.3 Cell Microscopy

Overnight culture was sub-cultured in 10 mL LB media supplemented with 2 mg/mL glycerol, with and without Nile red. Culture was induced with 0.025% L-arabinose to express PHB producing genes after 3 hours of growth. To visualize intracellular carbonosome, 100 µL culture without Nile red was harvested after 12 hrs (9 hrs after induction) and cells were washed by centrifugation at 10000 rpm with 100 µL PBS buffer and re-suspended in 1 mL distilled water. 20µL cells suspended in water were smeared on a microscopy glass slide and heat fixed. Smear was flooded with 1% aqueous solution of Nile red and incubated for 10 min at 55°C. Excess staining solution was removed and smear was washed gently with water. Smear was flooded with 8% (v/v) aqueous acetic acid and incubated for 1 min at room temperature. The excess was removed and smear was washed gently with water. Cover slip was placed and the stained cells were examined with Axio Observer Z1 inverted fluorescence microscope under 63x/1.4 oil immersion with DIC and 20 Rhodamine channel.

2.4.4 Quantitative estimation of intracellular Poly (R) 3-hydroxybutyrate

Our study attempts to develop an improved method (Figure 2.6) overcome all the limitations discussed in existing methods of PHB quantitation (Berlanga et al., 2006) by including two additional steps in the Nile red based fluorescence method. The steps of sonication and ethanol extraction improve the recovery and thus accuracy and sensitivity of PHB quantitation. An additional protein digestion step improved the accuracy of quantitation significantly in PHB producing recombinant *E.coli* cells.

2.4.4.1 Calibration curves for PHB Standard using Spectrofluorometry and UV Spectrophotometry

A spectrofluorometric method for PHB quantitation was developed using Nile red fluorophore on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific; Cat no. Q32866) with LED light source. The blue LED light source in Qubit 2.0 was used in a raw mode for excitation; it has the maximum intensity at 470 nm with excitation filter ranging from 430 nm to 495 nm and emission filter ranging from 665 to 720.

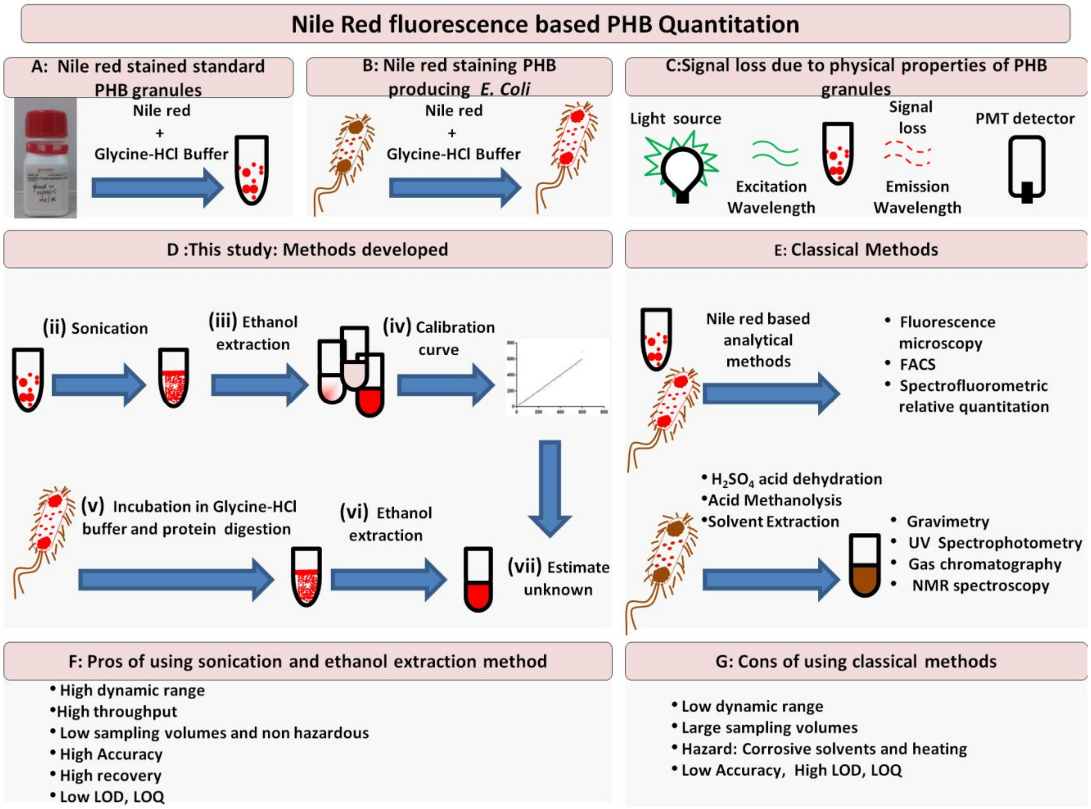

Nile Red fluorescence based PHB Quantitation
A: Nile red stained standard PHB granules
Nile red
+
Glycine-HCl Buffer
B: Nile red staining PHB producing *E. Coli*
Nile red
+
Glycine-HCl Buffer
C:Signal loss due to physical properties of PHB granules
Light source
Excitation Wavelength
Signal loss
Emission Wavelength
PMT detector
D :This study: Methods developed
(ii) Sonication
(iii) Ethanol extraction
(iv) Calibration curve
(v) Incubation in Glycine-HCl buffer and protein digestion
(vi) Ethanol extraction
(vii) Estimate unknown
E: Classical Methods
Nile red based analytical methods
• Fluorescence microscopy
• FACS
• Spectrofluorometric relative quantitation
• H_2SO_4 acid dehydration
• Acid Methanolysis
• Solvent Extraction
• Gravimetry
• UV Spectrophotometry
• Gas chromatography
• NMR spectroscopy
F: Pros of using sonication and ethanol extraction method
• High dynamic range
• High throughput
• Low sampling volumes and non hazardous
• High Accuracy
• High recovery
• Low LOD, LOQ
G: Cons of using classical methods
• Low dynamic range
• Large sampling volumes
• Hazard: Corrosive solvents and heating
• Low Accuracy, High LOD, LOQ

Figure 2.6 Pictorial representations of workflows for improved accuracy of quantitation of PHB The principle of the assay is depicted in Panels A-C. Standard PHB suspensions in aqueous glycine-HCl buffer bind quantitatively to Nile red and have specific fluorescence spectral characteristics Panel B shows how cells producing PHA can be, grown in the presence of Nile red that can bind to intracellular carbonosome, an intracellular PHA producing inclusion body. Panel C shows loss in emission signal of fluorescence dependent on physical properties of granules Panel D represents the methodology developed in this study while panel E represents the classical method for quantitation using UV spectrophotometry. Panel F-G discuss the Pros and Cons of the method

Suspensions of 1000 µg/mL standard PHB in triplicate were prepared by suspending 10 mg of PHB (Sigma: cat no. 363502) in 10 mL of 0.1 M glycine-HCl buffer pH 3.0 in 30 mL Borosil flat bottom culture tube (Cat no. 9910010). Suspension was kept in ice bath for 20 min and moved to a sonication chamber with an ice bath. The suspension was probe sonicated at 80% amplitude for 10 min at a 10/20 cycle on Sonic Vibracell VCX130 (130 wattz, 20 kHz) equipped with 6 mm stepped tip sonication probe. Suspensions with the same volume and concentration were also prepared and used without sonication. All experiments were performed in triplicate. Non-sonicated and sonicated standard suspensions were added to 1µg/mL Nile red (Sigma: cat no. N3013) (in DMSO 0.5 mg/mL) and incubated for 2 h (Figure 2.7 C) at room temperature (25 °C). To establish that Qubit 2.0 can be used for Nile red based quantitation we measured the excitation and emission fluorescence spectrum of 1 mL (1000 µg/mL) of Nile red stained standard PHB suspension (Figure 2.7 A) and its ethanolic extracts (Figure 2.7 B) on Photon Technology Fluorescence QM40 Spectrophotometer in Quartz high precision cell (Hellma analytics Art. No.101-10-40). Since blue LED in Qubit 2.0 has maximum exitaton at 470 nm, emission fluorescence spectrum was measured with excitation at 470 nm. For calibration curves using standards, Nile red stained sonicated and non-sonicated PHB suspensions were diluted with 0.1 M glycine-HCl buffer to 1 mL using glass pipettes. 200 µL suspensions were aliquoted at various concentrations (in the range of 20 to 1000 µg/mL) and washed twice by vortexing in 200 µL of 0.1 M glycine-HCl buffer followed by centrifugation at 4000 rpm for 5 min in Eppendorf® 5430 R centrifuge. The pellet was re-suspended in 200 µL glycine-HCl buffer and fluorescence measured. For ethanol extraction, 200 µL Nile red stained sonicated and non-sonicated granules were washed twice by vortexing in 200 µL 0.1 M glycine-HCl buffer followed by centrifugation at 12000 rpm for 5 min. The Nile red bound to the PHB standard was released by re-suspending stained granules in 200 µL ethanol. The fluorescence was measured after excitation at the determined wavelengths.

For UV spectrophotometric estimation, the suspension of PHB granules (20 to100 µg/mL) in triplicate were centrifuged and re-suspended in 100 µL ethanol and transferred into glass tubes. This procedure was repeated four times to ensure complete transfer of all granules. Ethanol was evaporated by heating at 60°C and drying under vacuum. Concentrated H_2SO_4 (1 mL) was added to the dried pellet, and capped tubes were heated for 10 min in a boiling water bath. The solution was cooled to room temperature (25 °C) and transferred to a high precision quartz cuvette to measure absorbance at 230 nm on a Spectrophotometer (BioPhotometer Plus, Eppendorf).

2.4.4.2 Quantitation of PHB using Spectrofluorometry from recombinant E. coli

Single colony of an *E.coli* DH10β transformed with ppct_phaC_Wt was inoculated in 10 mL LB media and incubated overnight at 37°C with shaking at 200 rpm.

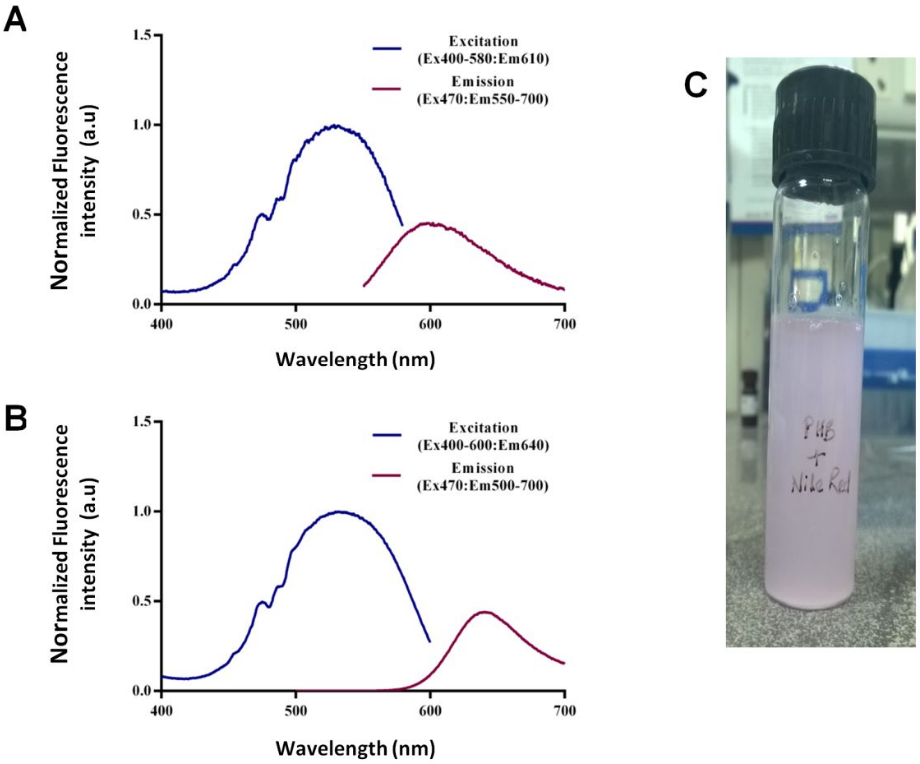


Figure 2.7 Excitation and Emission scan for standards

A) Nile red extracted from standard PHB granule B) Nile red stained standard PHB granule in glycine-HCl Buffer C) Sonicated Standard PHB suspension in glycine-HCl buffer with 1µg/ml Nile red

2.4.4.3 Quantitation of PHB using Spectrofluorometry from recombinant E. coli

Overnight culture was sub-cultured in 10 mL LB media supplemented with no extra carbon source, glycerol and glucose. 1 µg/mL Nile red was added to LB media while sub-culturing for staining PHB producing cells.Single colony of a *E.coli* DH10β transformed with ppct_phaC_Wt was inoculated in 10 mL LB media and incubated overnight at 37°C with shaking at 200 rpm. Overnight culture was sub-cultured in 10 mL LB media supplemented with no extra carbon source, glycerol and glucose. 1 µg/mL Nile red was added to LB media while sub-culturing for staining PHB producing cells. Culture was induced with 0.025% L-arabinose to express PHB producing genes after 3 hours of growth. Un-induced culture was used as control, grown in similar conditions with Nile red. 200 µL culture was harvested at regular intervals for 24 hours by centrifugation at 10000 rpm for 5 min. Cells were washed with PBS pH 7.4 by vortexing and followed by centrifugation at 10000 rpm for 5 min Cells can be stored in PBS at 4°C if assay is not carried out immediately after harvesting. Cells were washed twice with 200 µL of 0.1 M glycine-HCl buffer by vortexing followed by centrifugation at 4000 rpm. Pellet was re-suspended in 200 µL of 0.1 M Glycine-HCl buffer at pH 3.0 and incubated for 2 hours. Fluorescence was measured in a Qubit 2.0 fluorometer (Berlanga et al., 2006; Gorenflo et al., 1999; Spiekermann et al., 1999). For ethanol extract, the cell suspension in glycine-HCl buffer was centrifuged at 10000 rpm for 5 min and supernatant was discarded. 200 µL ethanol was added to the pellet to extract PHB bound Nile red. The fluorescence of the ethanolic extract was measured in a Qubit 2.0 fluorometer.

2.4.4.4 Microscopy Standard

Two aliquots of 1000 µg/mL standard PHB suspension were prepared with 10 mg of PHB in 10 mL of 0.1 M glycine-HCl buffer pH 3.0 in 30 mL Borosil flat bottom culture tube. Suspensions were kept in an ice bath for 20 min and moved to a sonicator chamber also with an ice bath. Suspension was probe sonicated at 80% amplitude for 10 min at a 10/20 cycle on Sonic Vibracell VCX130 (130 wattz, 20 kHz) equipped with 6 mm stepped tip sonication probe. 1 µg/mL Nile red in DMSO (0.5mg/mL) was added to both non-sonicated and sonicated suspension and incubated for 2 hours. 20 µL was spread over microscopy glass slide, covered with a cover slip and examined on an Axio Observer.Z1 inverted fluorescence microscope under 63x/1.4 oil immersion with DIC and 20 Rhodamine channel. Microscopy images were analyzed for granule surface area and volume using ZEN lite (Blue version) software from ZEISS.

200 µL of growing culture sub-cultured with Nile red was harvested, washed twice with 200 µL of 0.1 M glycine-HCl buffer by vortexing followed by centrifugation at 4000 rpm, re-suspended in 200 µL 0.1 M glycine-HCl buffer pH 3.0 and incubated for 2 hours at room temperature. 20 µL suspensions was spread over microscopy glass slide, covered with cover slip and examined with Axio Observer Z1 inverted fluorescence microscope

under 63x/1.4 oil immersion with DIC and 20 Rhodamine channel. Microscopy images were analyzed for granule surface area and volume using ZEN lite (Blue version) software from ZEISS.

2.4.4.5 Quantitation of PHB using UV Spectrophotometry from recombinant E. coli

For UV Spectrophotometric estimation, 100 µl of culture was harvested at regular time intervals. Cells were washed with PBS pH 7.4 followed by centrifugation at 10000 rpm for 5 min. This was followed by washing with glycine-HCl buffer. Pellet was re-suspended in glycine-HCl buffer and incubated for 2 hours at room temperature (25 °C). Cell suspension was centrifuged at 12000 rpm at room temperature for 10 min and re-suspended in 1mL chloroform. Separated chloroform extract was transferred to clean glass tubes and chloroform was allowed to evaporate at 50 °C and complete drying was achieved under vacuum. To the dried pellet, 1 mL concentrated H_2SO_4 was added and the capped tubes were heated for 10 min in a boiling water bath. The solution was cooled and transferred to Quartz high precision cell and absorbance was measured at 230 nm on a Spectrophotometer (BioPhotometer Plus, Eppendorf).

2.4.4.6 Estimation of fluorescence contributions from expressed proteins in growing cells

Nile red (1 µg/mL) in DMSO (0.5 mg/mL) was added to the LB media while sub culturing cells. Culture was induced with 0.025% L-arabinose to express PHB producing genes at 3 hours. Cells were harvested by centrifuging at 10000 rpm at 12 and 24 hrs after inoculation. Cell pellets were re-suspended in 1 mL of PBS and the cell density was adjusted (OD_{600}=1.0). Cells were washed with 20 mM Tris-HCl (pH: 8.0) and re-suspended in 100 µl lysis buffer (20 mM Tris-HCl and 2 mM EDTA containing 1% Triton X-100). Lysozyme (1 mg/mL) and trypsin (1 mg/mL) was added followed by incubation with shaking (500 rpm) at 37°C for 1 hour. The soluble fraction of cell lysate was analyzed on the Agilent Bioanalyzer protein 80 chip. Insoluble fraction was mixed with 200 µL 0.1 M glycine-HCl buffer and incubated for 2 hrs. Treated insoluble fraction was centrifuged and supernatant was discarded. To the treated insoluble fraction, 1 mL of ethanol was added and fluorescence was measured from an ethanolic extract. Replicates of the same culture without protein digestion were quantitated for PHB using the established spectro-fluorometric method.

2.4.4.7 Developing Calibration Curves

The analytical calibration curves (n=14) of fluorescence intensity measurements of PHB granules stained with Nile red, sonicated and extracted into ethanol were linear in the broad range of 125 to 1000 µg/mL (Figure 2.8 A). Fluorescence intensity measurements were made on a Qubit 2.0 with appropriate excitation (470 nm, slits fixed at 20 nm) and emission (600-640 nm) wavelengths (Figure 2.7 A, B). The standard error of mean (SEM) values is indicated in (Table 2.2,Table 2.3) the linearity was assessed by calculating the regression equation (y = a$\underline{x}$ ± b) and the correlation coefficient (r^2) by the least squares method. The linear regression equation was described by equation (1) and r^2 = 0.9894

$$y = 360.1x + 1366 \tag{1}$$

and non-sonicated granule suspensions with ethanol extraction (Figure 2.8 B). The methods excluding either of the added steps, sonication or ethanol extraction show higher % relative error (Table 2.3) and estimate higher concentrations than actual PHB standard. Sonication prevents aggregates (Figure 2.8 E, F, G, H) and controls size distribution and surface area to volume ratios (Table 2.4,Table 2.5,Table 2.6) . This results in uniform staining of granules with optimally sized granules with Nile red (Figure 2.8 I). The result of the method with both the added steps of sonication and extraction was comparable to the traditional quantitation method, based on degradation of PHB to crotonic acid using concentrated sulfuric acid followed by absorbance measurements at 230 nm (Fig: 4a). However, the maximum measurable value was an order of magnitude higher (dynamic range of UV-vis spectroscopy: 20 to 100 µg/mL). The spectrophotometric calibration curve (Fig: 2d) was described by the regression equation

$$y \;=\; 0.016 \times x \;-\; 0.026 \tag{2}$$

where ‘y’ is the absorbance at 230 nm and ‘x’ is the concentration of the standard in µg. The curve was highly linear with the regression coefficient r^2 being 0.9666.

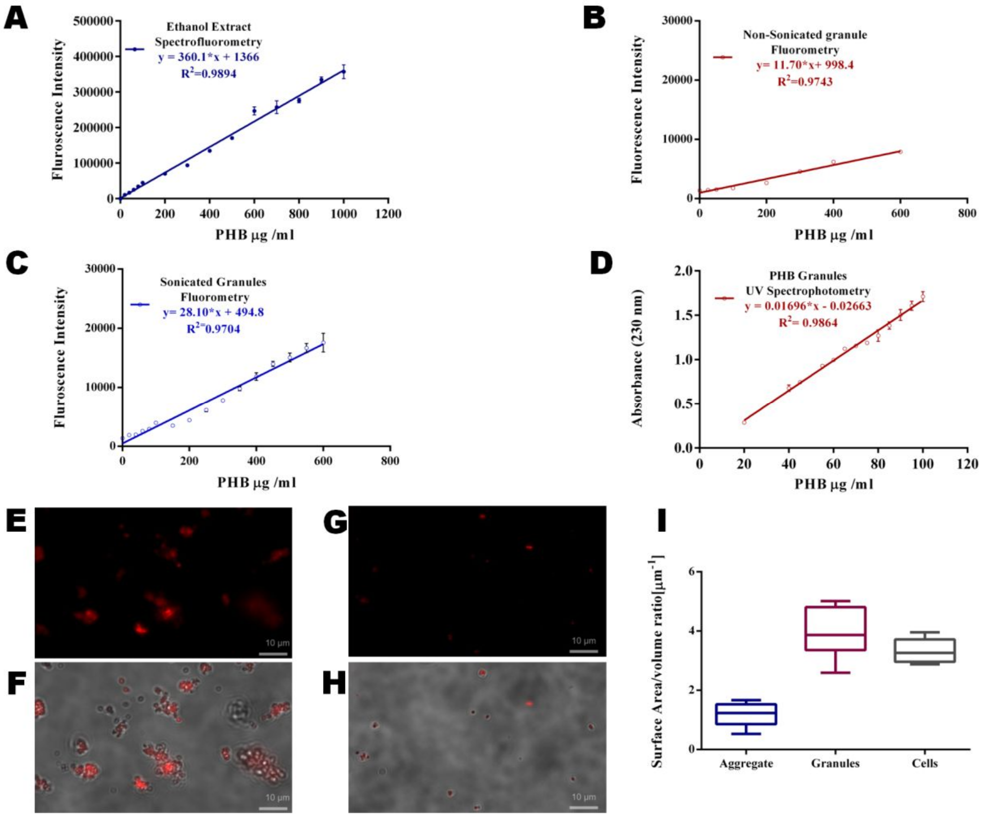


Figure 2.8 Development of Analytical Method for PHB

Calibration curves for PHB standard and physical parameters of standard PHB granules A) Spectrofluorometric calibration curve of ethanol extract from sonicated Nile red stained PHB standard granules in glycine-HCl buffer (excitation wavelength:470nm; emission wavelength:640 nm) in the concentration range of 20 to 1000 µg/mL B) Spectrofluorometric calibration curve of ethanol extract from Non-sonicated Nile red stained PHB standard granules in buffer (excitation wavelength:470nm;emission wavelength:640 nm) in the concentration range of 25 to 600 µg/mL C) Spectrofluorometric Calibration curve for sonicated PHB granules stained with Nile red in Glycine HCl buffer using (excitation wavelength:470 nm; emission wavelength:600 nm) in the concentration range of 20 to 600 µg/mL D) calibration curve for PHB standards using crotonic acid based UV spectrophotometry (absorbance wavelenght:230 nm) in concentration range of 20 to 100 µg/mL. Image panels depict Fluorescence microscopy and DIC images of Nile red stained granules incubated in glycine-HCl buffer for E-F) unsonicated PHB granules stained G-H) sonicated PHB granules I) Differential surface area to volume ratios for non-sonicated aggregates, sonicated PHB granules and cells.

The inclusion of the sonication and ethanol extraction steps had a statistically significant effect on precision, accuracy and sensitivity of measurement of true values. The standard measurement errors, RSD and standard deviations were lower. The method also had the lowest limits of detection and quantitation (LOD= 41.12 µg/mL; LOQ= 124.6 µg/mL). The dynamic range was high (124.6 µg/mL to 1000 µg/mL). The accuracy of the finalized method was the best with lowest relative error (9.59%). This underscores the impact based on systemic errors *vis a vis* random errors. Further, based on variance analysis using ANOVA, the effect of sonication had the highest statistical significance as reflected in the F-statistic ($F_{1,13}$ =1657.83, p= 4.26252e^{-15}). The ANOVA output of the linear regression model described (Table 2.7, Table 2.8, Table 2.9, Table 2.10) also confirmed linearity. Variance analysis resulted in correlation coefficient with a P-value < 0.05. r^2 values greater than 0.999 indicate a good correlation of linearity through all the concentrations used and a homoscedastic distribution of replicates in the calibration curve assembly PHB quantitation was comparable (standard deviation very low) using both the methods over a time frame of 24 hours from a growing culture of recombinant *E. coli* cells. Using glycerol as primary carbon source, the cells were seen to produce a maximum of 139.73 ± 9.02 µg/mL using our method as opposed to 154.02 ± 6.87 µg/mL with the crotonic acid method (Figure 2.9 A). All fluorescence measurements of induced cultures were normalized to that of un-induced control culture. When LB media with and without glucose was used, no PHA signals were identified in the NMR data indicating no PHB formation. An absorbance corresponding to 38 µg/mL of PHB over estimated PHB in the crotonic acid based method (Figure 2.9 B). Fluorescence intensities of 14633 a.u and 12579 a.u were measured respectively estimating falsely an amount of PHB equivalent to 36.84±4.32 µg/mL and 31.14±13.01 µg/mL after 24 hours of growth (Figure 2.9 C). NMR experiments confirmed that recombinant *E. coli* growing on glycerol produced PH. Since the NMR results did not detect any PHA in the spectrum of *E. coli* grown on glucose, it was critical to identify the source of contribution of fluorescent intensity in these samples. Since PHB is known to form carbonosomes around themselves with PHA producing proteins, we hypothesized that the expressed proteins were contributing to the fluorescence.

Table 2.2 Performance characteristic of Nile red based PHB estimation methods using fluorescence spectroscopy

	Accuracy	Precision						
Method Detail	%RE[a]	ANOVA[b] F(df)	% RSD[c]	SEM[d] (µg/µL)	Sensitivity (Slope)	LOD[e] (µg/mL)	LOQ[f] (µg/mL)	DynamicRange (µg/mL)
Sonication ethanol extraction spectrofluorometry	9.59	1657.4, (1,13) p=3E-15	2.27	7.95	360.12	41.12	124.6	124.6 to 1000
Ethanol extraction Spectrofluorometry	14.53	541.6, (1,13) p=5.5E-12	4.99	8.58	28.54	46.80	141.83	141.83 to 600
Sonication Spectroflurometry	19.85	229.3, (1,6) p=5.2E-06	2.10	2.00	11.70	62.73	190.09	190.09 to 600
UV Spectrophotometery	24.96	1463.6,(1,11) p=4.7E-13	2.11	1.01	0.0169	6.15	18.64	18.64 to 100

a: Percentage relative error; b: Analysis of Variance ; c:Relative standard deviation ; d:standard error of mean; e: Limit of detection ; f: Limit of quantitation

Table 2.3 Performance characteristics of Nile red based PHB quantitation in PHB producing E.coli cells compared to U V Spectrophotometery

	Accuracy		Precision	
Method Detail	%RE[a]	ANOVA[b] F(df)	% RSD[c]	SEM[d] µg/µl
Ethanol extraction spectrofluorometry	39.43	21.02, (1,4) p=0.019	12.36	4.52
Protein digestion, Ethanol extraction Spectrofluorometry	17.22	0.12, (1,4) p=0.73	5.531	3.47

Table 2.4 Non-sonicated granules physical properties

Aggregate S.no	**Diameter µm**	**Surface Area (μm^2)**	**Volume (μm^3)**	**SA/V (μm^{-1})**
A1	2.06	252.08	375.50	0.67
A2	1.66	73.25	58.82	1.25
A3	1.52	77.87	64.47	1.21
A4	2.09	133.89	145.36	0.92
A5	1.83	73.25	58.82	1.25
A6	1.86	127.01	134.30	0.95
A7	2.02	50.74	33.91	1.50
A8	1.63	43.69	27.09	1.61
A9	1.94	410.94	781.57	0.53
A10	1.58	40.92	24.56	1.67

Table 2.5 Sonicated granules physical properties

Aggregate S.no	Diameter µm	Surface Area (μm^2)	Volume (μm^3)	SA/V (μm^{-1})
G1	2.06	7.350	1.870	3.931
G2	1.66	7.642	1.982	3.856
G3	1.52	7.544	1.944	3.881
G4	2.09	8.343	2.261	3.690
G5	1.83	4.522	0.902	5.013
G6	1.86	16.901	6.519	2.593
G7	2.02	9.182	2.610	3.518
G8	1.63	13.716	4.766	2.878
G9	1.94	4.522	0.902	5.013
G10	1.58	5.065	1.069	4.736

Table 2.6 Nile red stained cells in glycine-HCl buffer physical properties

Aggregate S.no	Diameter µm	Surface Area (μm^2)	Volume (μm^3)	SA/V (μm^{-1})
C1	2.06	13.325	4.563	2.920
C2	1.66	8.653	2.388	3.624
C3	1.52	7.255	1.833	3.957
C4	2.09	13.716	4.766	2.878
C5	1.83	10.516	3.199	3.287
C6	1.86	10.863	3.359	3.234
C7	2.02	12.812	4.303	2.978
C8	1.63	8.343	2.261	3.690
C9	1.94	11.818	3.812	3.101
C10	1.58	7.839	2.059	3.807

Table 2.7 ANOVA sonicated, ethanolic extract,spectrofluorometry

Variation	Degree of Freedom (df)	Sum of squares (SS)	Mean square	F	Significance (p value)
Regression	1	2.1996E+11	2.2E+11	1657.83	4.26E-15
Residual	13	1724832472	1.33E+08		
Total	14	2.21685E+11			

Table 2.8 ANOVA non-sonicated, ethanolic extract, spectrofluorometry

Variation	Degree of Freedom (df)	Sum of squares (SS)	Mean square	F	Significance (p value)
Regression	1	42784751	42784751	229.3886	5.22E-06
Residual	6	1119099	186516.4		
Total	7	43903849			

Table 2.9 ANOVA sonicated spectrofluorometry

Variation	Degree of Freedom (df)	Sum of squares (SS)	Mean square	F	Significance (p value)
Regression	1	440881190.7	4.41E+08	541.6811	5.56E-12
Residual	13	10580866.69	813912.8		
Total	14	451462057.4			

Table 2.10 ANOVA UV Spectrophotometry

Variation	Degree of Freedom (df)	Sum of squares (SS)	Mean square	F	Significance (p value)
Regression	1	1.893948	1.893948	1463.695	4.71E-13
Residual	11	0.014233	0.00129 4		
Total	12	1.908181			

The relative over expressed protein contributed to 8.9%, 38.9% and 27.8% of the total cell protein when cells were grown on glucose, glycerol and LB media respectively (Figure 2.9 D). Quantitation of expressed protein suggests an eleven fold (Figure 2.9 E) ratio to PHB produced when grown on glycerol. Even in the absence of PHB biosynthesis, when cells are grown on LB with and without glucose, 7.75 and 9.41 µg of protein are produced per mL culture. It was evident on lysozyme digestion that in cells grown on glycerol, the over expressed protein contributed to 24% of the total fluorescence.

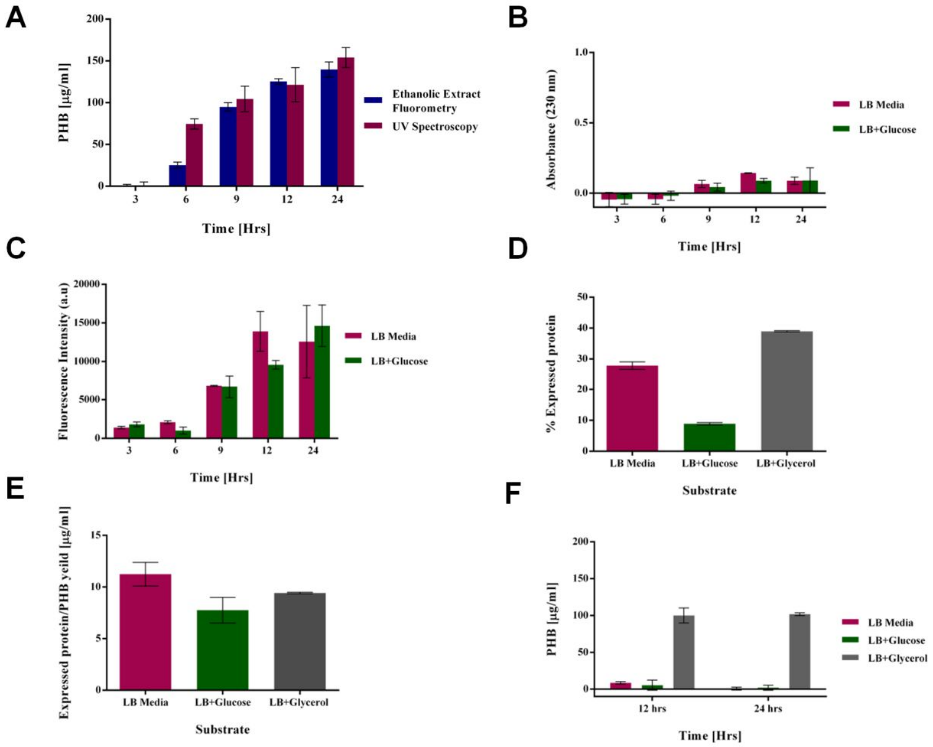


Figure 2.9 PHB Quantitation from recombinant *E. coli*

(A) Comparison of PHB quantitation from recombinant E.coli grown on 2mg/mL glycerol for 24 hours using the classical UV spectrophotometric method and the method developed in our study (B) Over estimation of PHB grown on LB media and LB Media supplemented with glucose using the UV spectrophotometric method (C) Contributing fluorescence intensities of PHB producing cells grown on LB media without any supplement and LB media supplemented with glucose over estimating PHB (D) Percentage of over expressed protein to the total protein after 24 hours of growth (E) Ratio of over expressed protein to the total PHB produced after 24 hours of growth (F) PHB quantitation in cell samples after protein digestion

All the fluorescence intensity (100%) emitted by the cells grown in LB with and without glucose supplementation was due to expressed protein (Figure 2.9 F) accounting for lack of PHA signals in the NMR spectrum. This indicates potential carbonosome formation and entrapment of some amounts of Nile red that contribute to incorrect or over estimation of PHB. The performance characteristic of the cell samples using the Nile red based method with an additional step of protein digestion was higher in accuracy and precision as compared to the classical UV-spectrophotometric method.

Nile red based spectrofluorometry is a widely used technique for relative quantitation of PHB. Limitations mainly arise out of the suspension assay either due to surface to volume ratio in standard PHB granules or background fluorescence of hydrophobic moieties in cells. In this study using Nile red based fluorescence analytics, an analytical method for absolute quantitation of PHB using standards has been developed that includes two additional steps to improve accuracy. The method is, has a higher dynamic range, lower relative error, LOD, LOQ as compared to the classical method. The crotonic acid based assay is also unsafe due to use of corrosive agents and heating at high temperatures. The inaccuracies in suspension assay, mainly attributable to the aggregation of PHB granules and microparticle size and distribution are corrected for through controlled sonication protocols. Fluorescence measurements after sonication cycles break up the clumped granules and increase the exposed area of binding Nile red (Figure 2.8 E, F, G, H & I). The recovery is improved through the extraction into ethanol.

The bigger challenge of transcending from standards to PHB measurements in cells was overcome by correcting for noise due to fluorescing protein moieties in the cell with appropriate controls before releasing the PHB-bound Nile red through ethanol extraction. To correct for the lipid background fluorescence from PHB producing cells, induced cells was normalized to un-induced cells. The noise related to protein over-expression that over estimated PHB was lowered by a simple lysozyme digestion step. Ethanolic extracts of hydrophobic granules and Nile red also allow the use of inexpensive polypropylene tubes instead of quartz cuvettes, in a compact LED based fluorometer. This is further suitable for high throughput scale up if microtiter plate readers are used. The method is also scalable to other PHAs.

2.5 Analyze: Genome scale model driven analysis of engineered strains

The last and final step of iteration of the synthetic biology paradigm is “Analyze”. Analyzing goes beyond looking at direct correlations and emergent properties are critical for systems approaches in metabolic engineering. Computational models have been used successfully to predict behavior of rationally designed strains and have been used similarly in engineering strains for violacein and PHA synthesis. The modeling methodology employed was constraints-based flux balance analysis. The details are discussed below.

Flux balance analysis (FBA) (Edwards and Palsson, 1999; Orth et al., 2010; Varma et al., 1993) is a widely used approach for analyzing biochemical reaction networks, particularly for genome-scale metabolic network reconstructions, that contain information about all the metabolic reactions and genes associated with them. FBA can predict the flow of metabolites in this metabolic network, resulting in to the prediction of growth rate of an organism or the rate of production of a biotechnologically important metabolite. With metabolic models for different organisms already available and high-throughput technologies enabling the construction of many more each year, FBA is an important tool for harnessing the knowledge encoded in these models.

Again start with the dynamic mass balance equation

$$\frac{\mathbf{dX}}{\mathbf{dt}} = \mathbf{S}.\mathbf{v}$$

For a given metabolic network, **S** is the stoichiometric matrix, $\frac{\mathbf{dX}}{\mathbf{dt}}$ is vector, consisting the rate of change of concentration of all the metabolites (internal and external both) and v is the flux vector carrying rate of all reactions. Now, integration of this equation over time will give us the time course of concentrations x(t). Now because the rates of reaction (fluxes) v are a function of the concentrations x and enzyme kinetic constants and other parameters as well and it is hard to measure all the kinetic constants required for dynamic simulation of the reaction network, the steady states of the system are studied (assuming the rate of production and consumption of all metabolites are equal). Steady state suggests, $\frac{\mathbf{dX}}{\mathbf{dt}} = \mathbf{0}$, thus$\mathbf{S.v} = \mathbf{0}$. In other words the (right) null space of matrix **S** contains the steady state flux distributions.

Thus the equation of concern is,

$$\mathbf{S.v} = \mathbf{0}$$

As major metabolic networks have lesser number of metabolites then the number of reactions, thus above equation represents a highly underdetermined system of equations, which can have infinite number of solutions.

To further constrain the solution space and take it closer to the exact physiological flux values by putting upper and lower limit to reaction fluxes. Different biological functional targets can be achieved under these constraints, and organisms willingly accept theses constraints by implication of regulatory mechanisms to distinguish useful functional states from all possible allowable states. These constraints give rise to a constraint-based approach which makes the simultaneous analysis of physicochemical factors and biological properties plausible

2.5.1 Mathematical representation of constraints: balances and bounds

In steady state, there is no accumulation or depletion of compounds; thus, the rate of production equals the rate of consumption for each compound in the network. This balance is represented mathematically as $\mathbf{S.v} = \mathbf{0}$.

A bound on flux values can be defined as an inequality, which confines the numerical range of fluxes. Upper and lower limits can be applied to individual fluxes ($\mathbf{v_{min}} \leq \mathbf{v} \leq \mathbf{v_{max}}$). In such analysis, flux bounds for irreversible reactions $\mathrm{v_{min}} = 0$ (representing thermodynamic constraint) and $\mathrm{v_{max}} = 1000$ (representing kinetic constraint) and similarly for reversible reactions $\mathrm{v_{min}} = -1000$ and $\mathrm{v_{max}} = 1000$ leading to free flow of metabolites are used. Thus application of constraints (**Error! Reference source not found.** D) on flux values leads to a constrained solution space carrying all possible physiological flux distributions **v**.

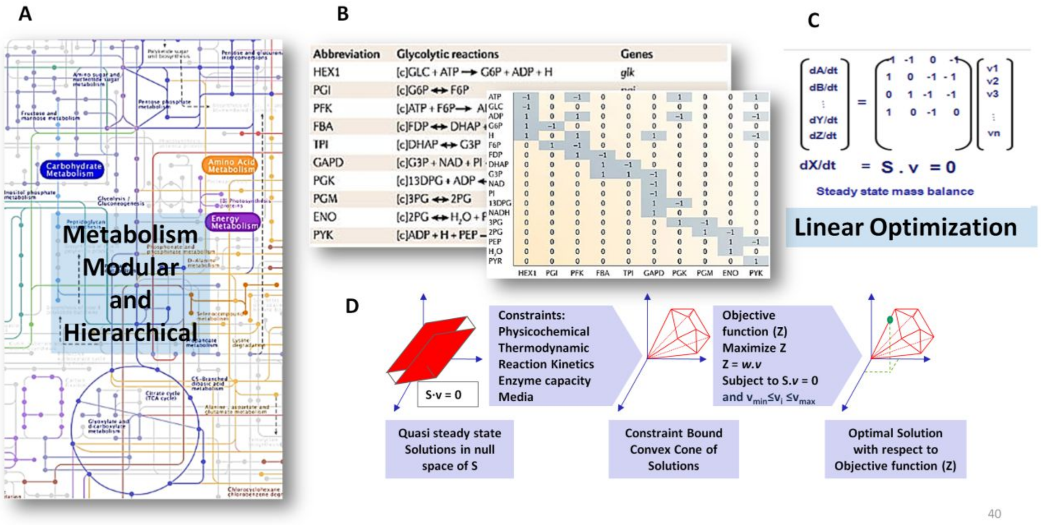

A
B
C
D
Metabolism
Modular
and
Hierarchical
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Abbreviation
Glycolytic reactions
Genes
HEX1
[c]GLC + ATP → G6P + ADP + H
glk
PGI
[c]G6P ↔ F6P
PFK
[c]ATP + F6P →
FBA
[c]FDP ↔ DHAP
TPI
[c]DHAP ↔ G3P
GAPD
[c]G3P + NAD + Pi
PGK
[c]13DPG + ADP
PGM
[c]3PG ↔ 2PG
ENO
[c]2PG ↔ H2O +
PYK
[c]ADP + H + PEP
dA/dt
dB/dt
dY/dt
dZ/dt
v1
v2
v3
vn
dX/dt = S . v = 0
Steady state mass balance
Linear Optimization
S·v = 0
Constraints:
Physicochemical
Thermodynamic
Reaction Kinetics
Enzyme capacity
Media
Objective function (Z)
Maximize Z
Z = w.v
Subject to S.v = 0
and vmin≤vi ≤vmax
Quasi steady state Solutions in null space of S
Constraint Bound Convex Cone of Solutions
Optimal Solution with respect to Objective function (Z)
40

Figure 2.10 : Genome Scale model and flux balance analysis

(A) Modular nature of metabolism (B) Reactions to Stoichiometric Matrix (C) Metabolic reconstruction and flux balance analysis (D) Applying Steady State and Flux Constraints

2.5.2 Metabolic reconstruction and flux balance analysis (FBA)

The genome-scale metabolic model of *E. coli i*AF1260 (Feist et al., 2007) was used for the present study. The curated violacein biosynthesis reaction list including gene-protein associations, reaction stoichiometry, and reversibility was incorporated into this model. The resulting model *i*AF1260vio was able to synthesize violacein. The detailed description of the model can be found in Supplementary File 2. The set of constraints used in different simulations included (i) Substrate (Glucose) uptake rates (GUR) (ii) Growth yields (iii) Violacein secretion associated with molar growth yields of each strain as discussed (*Appendix II* Sheet 4: Constraints). A basis of 1 g biomass was used to calculate specific growth rates. Implementation of the genome scale metabolic model *i*AF1260 with experimental and legacy data based constraints was done using Constraints Based Reconstruction and Analysis (COBRA) Toolbox 2.0.2 (Schellenberger et al., 2011) with MATLAB v 7.11, (R2010b) and TOMLAB/CPLEX v7.7 optimizer. MATLAB codes for all referenced COBRA functions are available at the COBRA's website (https://opencobra.github.io/). The function optimizeCbModel (), in COBRA toolbox was used for maximization of growth by fixing the objective as the biomass equation in the model. The following constraints based methods are used to probe the network space of *i*AF1260vio strains *in silico.*

MOMA (Minimization of metabolic adjustment) Analysis

MOMA was performed in COBRA toolbox using the function MOMA() to study the effect of gene deletions *in silico*, that allows for selecting more appropriate optimal solutions (Alper et al., 2005; Segre et al., 2002).

Shadow price and reduced costs analysis

Two sensitivity parameters – shadow prices and reduced costs (Schellenberger et al., 2011) were assessed at maximal yields of violacein and biomass in order to understand the effects of changing biomass, metabolites and reactions in the network. Shadow price corresponds to the sensitivity of the growth rate as an objective function (Z) in response to a change in the availability of a metabolite (i), and indicates how much an increment in that metabolite will increase or decrease the growth rate. Analogous to shadow price, reduced cost is the sensitivity of the objective function in response to change in fluxes of a particular reaction and its effect on the objective. In addition to the primal solution (optimal fluxes), the LP solver provides the corresponding dual solution i.e., shadow price and reduced cost for the FBA problem (Acevedo et al., 2014; Orth et al., 2010; Reznik et al., 2013).

Flux variability analysis (FVA)

FVA can be set up in COBRA toolbox using the function fluxVariability() to study the resulting space of feasible flux distributions (Mahadevan and Schilling, 2003). Nine categories can be mapped onto the flux variability distribution based on the magnitude and direction of the flux as discussed previously (Banerjee et al., 2017) .

Robustness analysis

Robustness analysis set up in COBRA toolbox using the function robustnessAnalysis() is used to better understand the phenotype of an organism under different environmental perturbations. The carbon flux diversion towards aromatic amino acids for all *E. coli* strains while maximizing for violacein was studied.

Phenotypic phase planes (PhPP) analysis

PhPP analysis (Edwards et al., 2002) was performed characterizing all optimal flux distributions as a function of NADPH and Tryptophan demands at experimental biomass. Simultaneous sensitivity/shadow price analysis allows one to delineate the changing shadow prices in each phase of the plane.

Conclusions

The methodology used in the Design, Build, and Test and Analyze iterative cycle was discussed in this chapter. The pathway design was meticulously analyzed that will help in further improvement of design starting from the strains produced in this work. Most of the methods discussed are previously published legacy protocols that have been applied in this thesis for novel applications. These methods were implemented for violacein and PHB biosynthesis.

A new method was developed for absolute PHB quantitation in standards and cells. This method is rapid, accurate, and sensitive and a significant improvement over the already existing methods, accounts for interference from lipophilic cell constituent, carbonosome formation and protein. Further, it eliminates inconsistencies due to PHB granule aggregation and solvatochromism. This method can be scaled for higher throughput in systems biology and metabolic engineering approaches. The use of Nile-red based estimations are also scalable for PHAs, plastics (Maes et al., 2017)in the environment and multiple cell types.

Chapter 3

Synthetic and Systems Metabolic Engineering for Violacein production

How is phenotype controlled by the genes?
Nobody knows, least of all machines,
Medicine will thrive if we discover the means,
To merge our knowledge and information
And find genes' intent and control by environment
For the "metabolic" times they are a changing
Jay Bailey

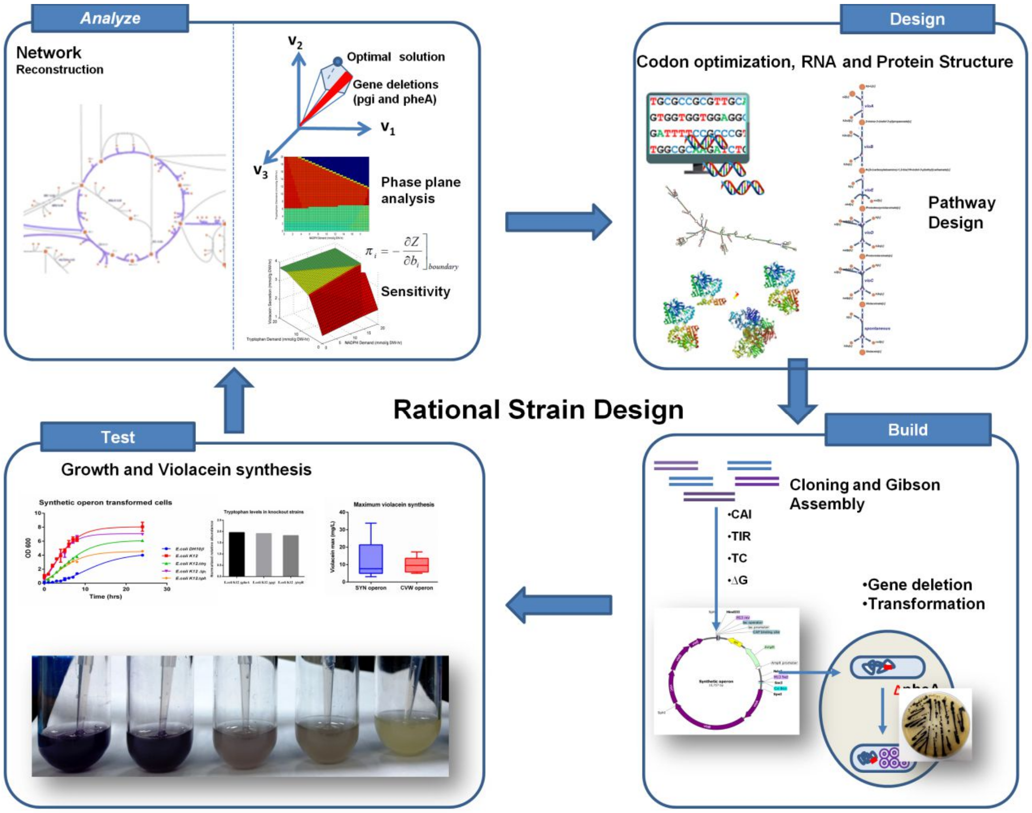

Analyze
Network
Reconstruction
v_2
Optimal solution
Gene deletions
(pgi and pheA)
v_1
v_3
Phase plane
analysis
$\pi_i = -\frac{\partial Z}{\partial b_i}\Big]_{boundary}$
Sensitivity
Design
Codon optimization, RNA and Protein Structure
TGCGCCGCGTTGC
GTGGTGGTGGAGGC
GATTTTCCGCCCGT
TGGCGC
Pathway
Design
Rational Strain Design
Build
Cloning and Gibson
Assembly
•CAI
•TIR
•TC
•ΔG
•Gene deletion
•Transformation
Synthetic operon
ΔpheA
Test
Growth and Violacein synthesis
Synthetic operon transformed cells
OD 600
Time (hrs)
0
10
20
30
Tryptophan levels in knockout strains
Maximum violacein synthesis
Violacein max (mg/L)
SYN operon
CVW operon

Figure 3.1 Synergies among synthetic biology, systems biology and metabolic engineering in the drug Violacein production

3.1 Introduction

L-Tryptophan is a unique amino acid for having an electron rich indole ring, it is the largest amino acid and least abundant as a part of protein structure and in its free form. The electron rich property of indole ring in tryptophan makes it a highly reactive molecule. It is also a very important amino acid for investigation since it is a precursor for many important molecules in the physiology of all categories of organism(Immanuel et al., 2018). In humans it is precursor for neurotransmitter serotonin, hormone melatonin and vitamin Nicotinic acid (Palego et al., 2016). In plants it is precursor for plant hormones like Indole Acetic Acid (IAA) phytoalexins, glucosinolates, and both indole- and anthranilate-derived alkaloids(Radwanski and Last, 1995). In bacteria there are hundreds of molecules produced starting from L-tryptophan, not only through enzymatic conversions but also spontaneous non-enzymatic conversion(Alkhalaf and Ryan, 2015). Some of the examples of bacterial secondary metabolite produced form tryptophan is Sturosporin, Rebeccamycin, chromopyrrolic acid and violacein (Durán et al., 2007; Rodrigues et al., 2013). Many such secondary metabolites because of their reactive and planar nature interact with protein as inhibitors(Ferreira et al., 2004; Hosokawa et al., 2016). This make such compounds the potential candidate for drug discovery.

Violacein is a bis-indole pigment produced by diverse form of bacteria inhabiting deep sea water (*Psedoalteromonas sp.*), Glaciers (*Janthinobacterium lividum, J.svalbardensis*), agricultural soil (Duganella sp.), river and waste water treatment plant (*Chromobacterium violaceum*) and as a commensal on sea sponge surface (*Psedoalteromonas luteviolacea*). The cosmopolitan presence of violacein biosynthetic operon constituting five gene pathway (VioABCDE) suggests towards the evolutionary benefit of Violacein and other L-Tryptophan derived molecules because of their highly reactive and protein inhibitory nature. Violacein is known to have anti-bacterial ,anti-viral, anti-fungal and anti-tumor activity(Andrighetti-Fröhner et al., 2003; Choi et al., 2015; Durán et al., 2016) and it is also a pigment appearing violet-blue in color, making its qualitative and quantitative estimation from cells easy. Therefore it can be used as a model molecule for all L-tryptophan derived metabolites with potential to be new drug. The violacein biosynthetic pathway also uses 2 molecules NADPH as reducing power which is also common among most secondary metabolite production and balancing the redox potential is the main obejective of the strain design for metabolite production. Violacein makes an excellent example for rational strain design because its linear pathway and simple estimation techniques for end product, this allows for the high throughput estimation and test the systems driven approach for rational strain design.

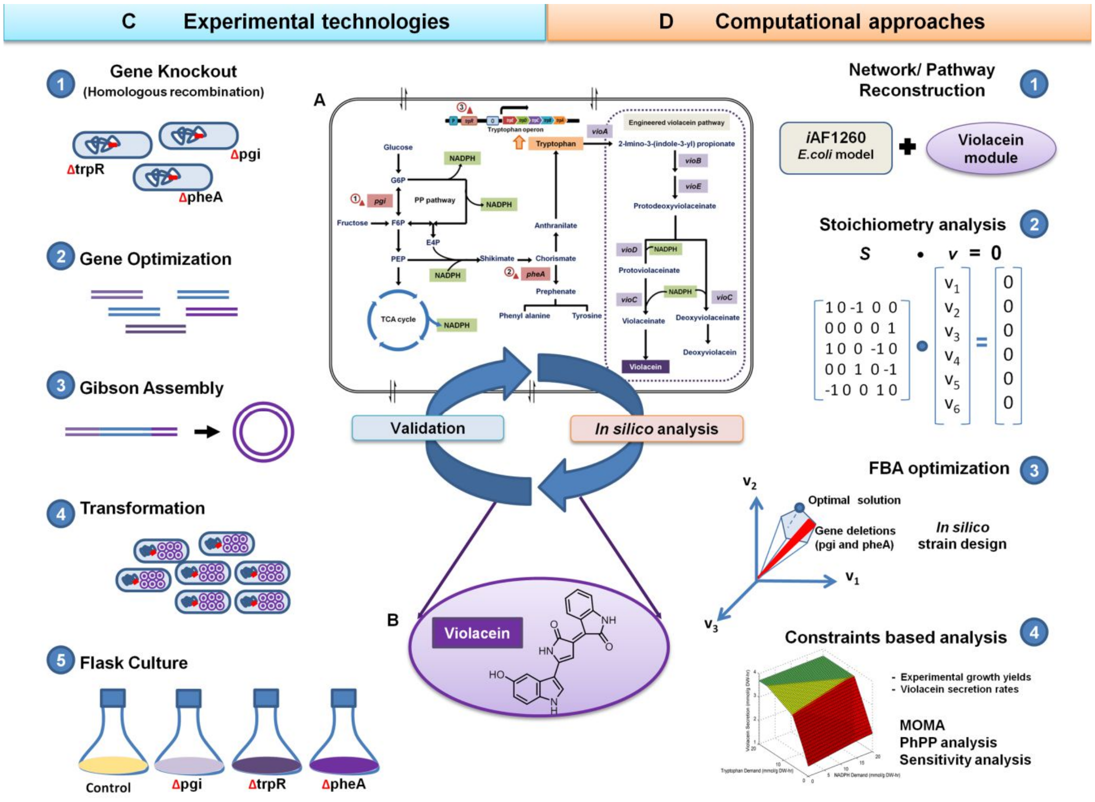

C Experimental technologies
D Computational approaches
1 Gene Knockout (Homologous recombination)
ΔtrpR
Δpgi
ΔpheA
2 Gene Optimization
3 Gibson Assembly
4 Transformation
5 Flask Culture
Control
Δpgi
ΔtrpR
ΔpheA
A
Tryptophan operon
Engineered violacein pathway
vioA
Tryptophan
2-Imino-3-(indole-3-yl) propionate
vioB
vioE
Protodeoxyviolaceinate
vioD
NADPH
Protoviolaceinate
vioC
vioC
Violaceinate
Deoxyviolaceinate
Deoxyviolacein
Violacein
Glucose
G6P
pgi
PP pathway
Fructose
F6P
E4P
PEP
Shikimate
Chorismate
pheA
Prephenate
Anthranilate
Phenyl alanine
Tyrosine
TCA cycle
NADPH
Validation
In silico analysis
B
Violacein
1 Network/ Pathway Reconstruction
iAF1260 E.coli model
Violacein module
2 Stoichiometry analysis
S • v = 0
10 -1 0 0
00 0 0 1
10 0 -10
00 1 0 -1
-10 0 10
v1 v2 v3 v4 v5 v6
0 0 0 0 0 0
3 FBA optimization
v2
v1
v3
Optimal solution
Gene deletions (pgi and pheA)
In silico strain design
4 Constraints based analysis
- Experimental growth yields
- Violacein secretion rates
MOMA
PhPP analysis
Sensitivity analysis
Violacein Secretion (mmol/g DW-hr)
Tryptophan Demand (mmol/g DW-hr)
NADPH Demand (mmol/g DW-hr)

Figure 3.2 Overall work flow for integration of constraint based modeling for experimental rational strain design.

3.2 Growth optimality analysis of Violacein pathway by Flux Balance Modeling

The overall workflow (Figure 3.2) describes the integration of constraints-based modeling, SYNO design and experimental testing to interrogate violacein biosynthesis in *E. coli*. A violacein (Figure 3.2 B) producing *E. coli* organism (*i*AF1260vio) was reconstructed *in silico* by adding a violacein biosynthesis reaction module (Figure 3.2 A) into a previously published *i*AF1260 genome scale model (Feist et al., 2007). The module consisted of 10 intracellular reactions and 10 metabolites describing the biochemistry and

5 Gene-Protein-Relation (GPR) to account for the genetics (August et al., 2000; Banerjee et al., 2017; Hoshino, 2011). Both experimentally determined (Figure 3.1 C) and legacy data were used as constraints for simulations (Figure 3.1 D and *Appendix II*). Deletions Δ*pgi* and Δ*pheA* were identified to increase yields of precursor tryptophan that eventually increased violacein yields. These genes essentially exploit the redundancy in metabolic pathways to increase NADPH and tryptophan, the two bottlenecks identified for increasing violacein biosynthesis. The *i*AF1260 was able to match the growth yields (molar basis) of experimentally grown *E. coli* K12 with glucose as limiting substrate. Accounting for additional constraints (*Appendix II*), the accuracy of growth yield (again on a molar basis) predicted for all genetic mutants of K12 and *i*AF1260*vio* was greater than 88%. These constraints include (i) violacein secretion (ii) excess tryptophan produced and (iii) varying glucose uptake rates. The model was validated with the experimental growth yield of *E. coli* K12 and deletion strains on glucose, predicted with 91.5% accuracy. The predicted and experimental growth and violacein yields are listed (Table 3.1).

Table 3.1 Predicting yields of biomass and violacein using iAF1260vio

Experimental values				**Prediction Accuracy (%)**	
Strains	**GUR**	**Molar Yield**		**Biomass**	**Violacein**
		Biomass	**Violacein**		
K12+VioABCDE	7.6	0.038	0.054	96	88
	8.2	0.063	0.007	97	41
K12Δ*pgi*+VioABCDE	7.6	0.041	0.041	90	98
	8.2	0.103	0.078	100	99
K12Δ*pheA*+VioABCDE	7.6	0.034	0.441	99	99
	8.2	0.043	0.132	67	-

The sensitivity of biomass and violacein synthesis flux to the substrate tryptophan and the cofactor NADPH flux was delineated (Figure 3.3) through PhPP analysis for *i*AF1260 derived strains *in silico*. Constraints were derived from experimental data on all CVW and SYN strains. PhPP analysis delineates the sensitivity of the objectives growth and violacein biosynthesis to various rate limiting metabolites simultaneously by analyzing shadow prices in each phase of the plane. The PhPP analysis herein mapped the entire spectrum of violacein biosynthesis/secretion conditions at optimal experimental growth yields as a function of two critical environmental variables. The variables represented on the x and y axes are precursor tryptophan and cofactor NADPH respectively. The third dimension (the violacein secretion rate at the optimal cellular growth yield) is the bioengineering objective.

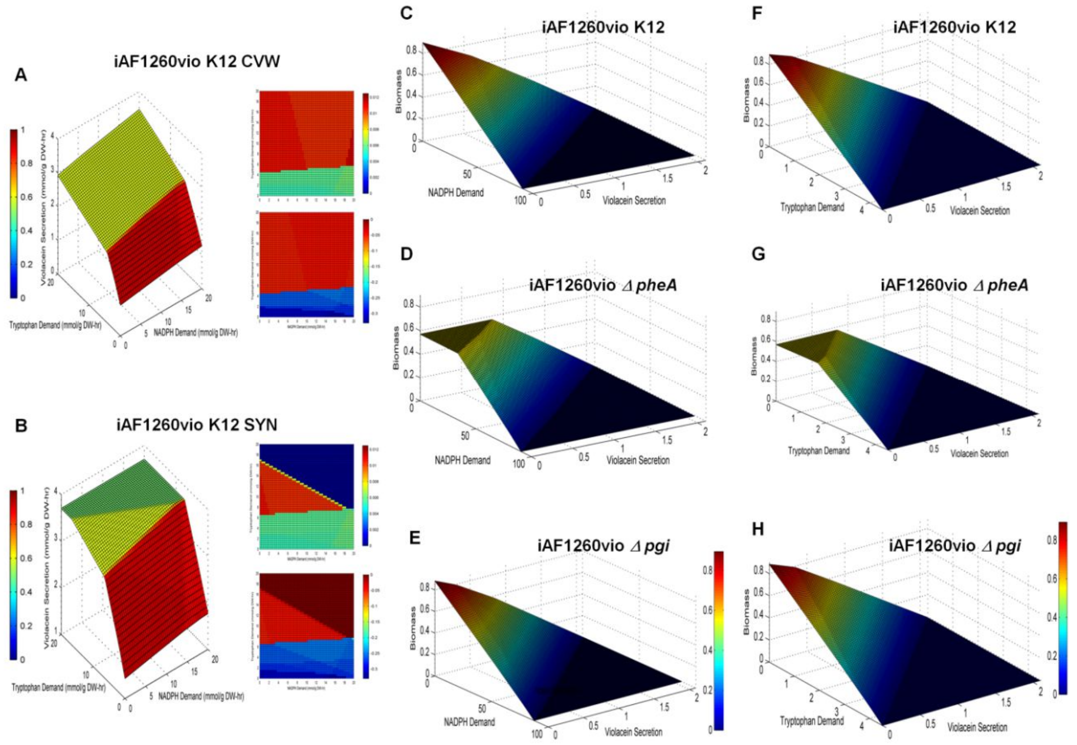


Figure 3.3 Phenotypic Phase Plane (PhPP) analysis for violacein biosynthesis

PhPP graphs as a function of Tryptophan and NADPH using iAF1260vio for (A) E. coli K12 containing WTO defined by CVW constraints and B. E. coli K12 containing SYNO defined by SYN constraints. Three distinct viable phenotypes were predicted feasible for *i*AF1260vio. Shadow prices for NADPH (inset top) and Tryptophan (inset bottom) are similar. (C – E) Double robustness analysis depicting sensitivity of biomass formation to NADPH and violacein for (C) *i*AF1260vio *in silico* strain, (D) *i*AF1260vio:*ΔpheA* and (E) *i*AF1260vio:*Δpgi*. F – H Sensitivity of biomass formation to Tryptophan and violacein for (F) *i*AF1260vio *in silico* strain, (G) *i*AF1260vio: *ΔpheA* and (H) *i*AF1260vio: *Δpgi*.

The shadow prices of these two variables projected onto a two-dimensional plane identified three different phases with distinct limiting patterns for violacein biosynthesis. iAF1260vio and iAF1260vio:Δpgi in silico behaved differently exhibiting only two phases. PhPP (where the shadow price is constant) represents a metabolic phenotype with specific pathway utilization (Figure 3.3 A, B). The shadow prices changed continuously at the boundary from one phase to the next. The operational space for *i*AF1260vio:Δ*pheA in silico* decreased for biomass while producing violacein at increased levels. Both NADPH and tryptophan have a threshold value determined through sensitivity/robustness analysis with no change in biomass in only *i*AF1260vio:Δ*pheA* (Figure 3.3 C-H).

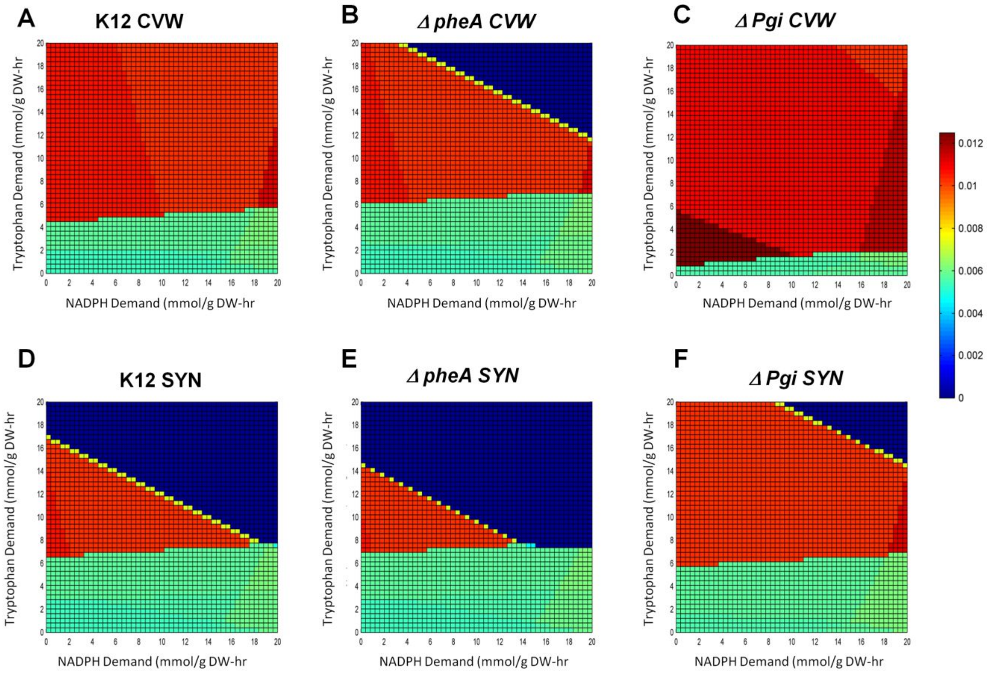


Figure 3.4 Shadow price from phenotypic phase plane (PhPP) using in silico strain for NADPH

Robustness analysis highlighted the sensitivity of anthranilate synthase (ANS), chorismate mutase (CHORM) to tryptophan and 6-phos- phogluconate dehydratase (EDD) to NADPH (Figure 3.5). The sensitivity of violacein secretion to changing the magnitude of NADPH and tryptophan can be further examined using shadow price analysis. While maximizing their fitness, iAF1260vio:ΔpheA is not limited in either tryptophan or NADPH, while in iAF1260vio:K12 and iAF1260vio:Δpgi they are limiting. However, with an added bioengineering objective, the cells in silico find both tryptophan and NADPH limiting. Tryptophan is the precursor for violacein as well as crucial for cell growth and biomass formation. NADPH is an essential reducing agent in anabolic pathways for biomass synthesis and is also crucial for the downstream reactions in the biosynthesis of violacein. A bioengineering objective (violacein formation) adds up a cost in addition to the survival and functioning of the cell (biomass formation) and as both tryptophan and NADPH are essential for biomass as well as violacein production they become rate limiting as represented by negative shadow prices in

Table 3.2. The analysis also highlights metabolites involved in two major pathways of aromatic amino acid synthesis (indole, phenyl pyruvate etc.) and oxidative phosphorylation (quinones) as rate limiting (

Table 3.2)

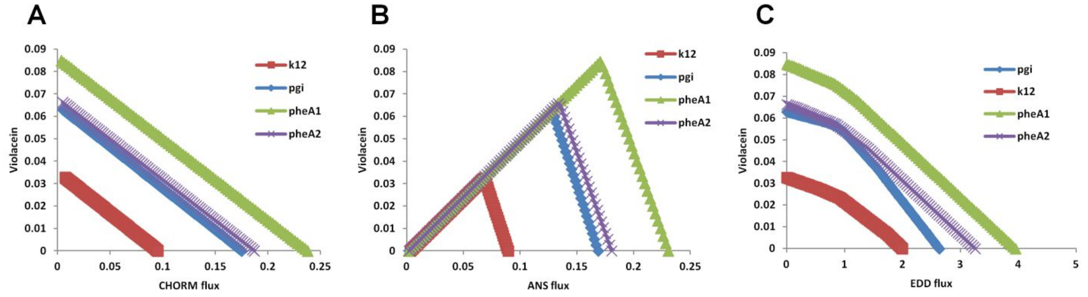


Figure 3.5 Robustness analysis for in silico strain for sensitivity (A)Chorismate mutase (CHORM) (B) Anthranilate synthase (ANS), to tryptophan and (C) 6-phosphogluconate dehydratase (EDD) to NADPH.

Flux variability analysis (FVA) identifies the redundancy of the network through identifying maximum and minimum flux distributions. At experimental growth yields flux distribution patterns were identified (Table 3.3) to establish rewiring in the wild-type and synthetic transformants of *E. coli.* This is necessary to compensate for metabolic burden of violacein biosynthesis. The flux variability analysis points to 3 major category changes between the Δ*pgi* and Δ*pheA* backgrounds that involve glyceraldehyde-3-phosphate dehydrogenase (GAPD), enolase (ENO) and periplasmic cytochrome oxidase bo3 (ubiquinol-8: 4 protons, CYTBO3_4pp). Reaction reduced costs are also found to be limiting for pathways leading to these metabolites. There are 75 and 85 category changes spanning glycolysis, PPP, TCA, Oxidative phosphorylation and alternate carbon metabolism when one compares the *i*AF1260vio to *i*AF1260vio:Δ*pheA* andΔ *i*AF1260vio: Δ*pgi* respectively. A third of these reactions need NADPH/ NADH as cofactor. Although only NADPH is used in the violacein pathway, the inclusion of a transhydrogenase reaction in the model treats NADH and NADPH as equivalent. The Flux variability analysis (FVA) (**Error! Reference source not found.**; *Appendix II*) identified acetate secretion as concomitant to violacein biosynthesis for all strains and genetic backgrounds.

Table 3.2 Shadow price (SP) analysis for optimal growth and Violacein in E. coli

For maximum molar biomass yields

Strains	**Maximum biomass**	**SP NADPH**	**SP Tryptophan**
***i*AF1260**	0.0642	-0.3806	-0.2057
i*AF1260:Δ*pgi	0.0632	-0.3806	-0.2057
i*AF1260:Δ*pheA	0.0569	-7.51E-28	-1.58E-28

For maximum molar violacein yields

Strains	**Maximum violacein**	**SP NADPH**	**SPTryptophan**
***i*AF1260vio**	0.1416	-0.0207	-0.4677
***i*AF1260vio**:Δ*pgi*	0.1395	-0.7034	-0.4671
***i*AF1260vio**:Δ*pheA*	0.1416	-0.8433	-0.4677

For maximum molar violacein yields at 10% biomasss

Strains	**Maximum violacein**	**SP NADPH**	**SP Tryptophan**
***i*AF1260vio**	0.1290	-0.8618	-0.4677
***i*AF1260vio**:Δ*pgi*	0.1268	-0.8642	-0.4671
***i*AF1260vio**:Δ*pheA*	0.1290	-0.8618	-0.4677

For maximum violacein and experimental biomass

Metabolite Names	**Metabolite ID**	***i*AF1260**	***i*AF1260vio**		***i*AF1260vio:Δpgi**	***i*AF1260vio:Δpgi SYN**	***i*AF1260:ΔpheA**	***i*AF1260vio:ΔpheA CVW**
			SYN	**CVW**				
Aromatic amino acid metabolism								
L-Phenylalanine	phe-L[c]	-0.17	0	-0.39	-0.17	0	-5.69	0
Phenylpyruvate	phpyr[c]	-0.16	0.01	-0.37	-0.16	0.01	-5.69	0.06
Chorismate	chor[c]	-0.17	-0.37	-0.37	-0.17	-0.37	7.8E-29	-1.49
Isochorismate	ichor[c]	-0.17	-0.37	-0.37	-0.17	-0.37	7.8E-29	-1.49
Indole	indole[c]	-0.17	-0.38	-0.38	-0.17	-0.38	1.2E-28	-0.01
Central carbon metabolism								
D-Glucose	glc-D[c]	-0.1	-0.22	-0.22	-0.1	-0.22	6.8E-29	-0.88
Acetate	ac[c]	-0.03	-0.06	-0.06	-0.03	-0.06	3.2E-29	-0.25
D-Gluconate	glcn[c]	-0.09	-0.21	-0.21	-0.09	-0.21	5.9E-29	-0.85
1,4-Dihydroxy-2-naphthoate	dhna[c]	1.04	-0.44	-0.44	0.92	-0.44	-2.8E-27	-1.75
3-hydroxycinnamic acid	3hcinnm[c]	-0.12	-0.27	-0.27	-0.12	-0.27	0	-1.05
2,3-dihydroxicinnamic acid	dhcinnm[c]	-0.12	-0.28	-0.28	-0.12	-0.28	1.6E-28	-1.12
Oxidative phosphorylation								
2-Demethylmenaquinone 8	2dmmq8[c]	0.12	-2.53	0	0	-2.53	0	0
2-Demethylmenaquinol 8	2dmmql8[c]	0.12	-2.54	-0.01	0	-2.54	9.21E-30	-0.04
Menaquinol 8	mql8[c]	0	-2.81	0	0	-2.81	0	0
Menaquinone 8	mqn8[c]	0	-2.8	0.01	0	-2.8	-9.2E-30	0.04

* Green shaded values denote no rate limiting by the corresponding metabolites

Table 3.3 Predictions using biomass as the objective function and constraining Violacein secretion

Model	**Violacein**	**Other Constraints**[#]	**FBA**	**MOMA**[$]	**Total flux Difference**[*]
K12	-	Molar	0.0642	0.0642	-
Δ*pgi*	-	Molar	0.0632	0.0415	2.5105
Δ*pheA*	-	Molar	0.0569	0.0569	0.0041
K12 SYN	0.054	Molar	0.0399	0.0399	-
K12 Δ*pgi* SYN	0.041	Molar	0.0447	0.0196	2.8587
K12 Δ*pheA* SYN	0.441	Molar[2]	0.0336	0	19.9311
K12 CVW	0.007	Molar	0.0611	0.0611	-
K12 Δ*pgi* CVW	0.078	Molar[1]	0.1031	0.0328	3.0575
K12 Δ*pheA* CVW	0.132	Molar[2]	0.0569	0	2.1200

Predictions using violacein secretion as the objective function and constraining biomass

Model	**Biomass**	**Other Constraints**[#]	**FBA**	**MOMA**[$]	**Total flux Difference**[*]
K12 SYN	0.038	Molar	0.0582	0.0582	-
K12 Δ*pgi* SYN	0.041	Molar	0.0492	0.027	3.0956
K12 Δ*pheA* SYN	0.034	Molar[2]	0.4371	0.0581	0.0001
K12 CVW	0.063	Molar	0.0027	0.0027	-
K12 Δ*pgi* CVW	0.103	Molar[1]	0.0788	0	37.8334
K12 Δ*pheA* CVW	0.043	Molar[2]	0.3587	0	0.5663

[#] Constraints as mentioned in Appendix II

[*] Total flux difference with respect to WT in case of MOMA simulations

[$] MOMA (Segre et al., 2002) was implemented using MOMA() command in COBRA Toolbox and Tomlab CPLEX was used as the LP and QP solver.

1and 2 are additional constraints as mentioned in *Appendix II*

Flux variability analysis (FVA) identifies the redundancy of the network through identifying maximum and minimum flux distributions. At experimental growth yields flux distribution patterns were identified to establish rewiring in the wild-type and synthetic transformants of *E. coli.* This is necessary to compensate for metabolic burden of violacein biosynthesis. The flux variability analysis points to 3 major category changes between the Δ*pgi* and Δ*pheA* backgrounds that involve glyceraldehyde-3-phosphate dehydrogenase (GAPD), enolase (ENO) and periplasmic cytochrome oxidase bo3 (ubiquinol-8: 4 protons, CYTBO3_4pp). Negative reduced costs (Table 3.4) for reactions in pathways leading to these metabolites also showed them to be limiting. There are 75 and 85 category changes spanning glycolysis, PPP, TCA, Oxidative phosphorylation and alternate carbon metabolism when one compares the *i*AF1260vio to *i*AF1260vio:Δ*pheA* and *i*AF1260vio: Δ*pgi* respectively. A third of these reactions need NADPH/ NADH as cofactor. Although only NADPH is used in the violacein pathway, the inclusion of a transhydrogenase reaction in the model treats NADH and NADPH as equivalent. The Flux variability analysis (FVA) (Table 3.4; *Appendix II*) identified acetate secretion as concomitant to violacein biosynthesis for all strains and genetic backgrounds.

3.3 Synthetic operon optimization and construction

The Gibson assembly method successfully stitched the WTO and SYNO vioABCDE gene designs into a vector (Figure 3.6) that allows coordinated transcription of vioABCDE mRNA and translation into functional proteins and violacein producing violet colonies of *E. coli* (Figure 3.8 A). The growth and product characterization of these operons expressed in *E. coli* K12 and varying genetic backgrounds

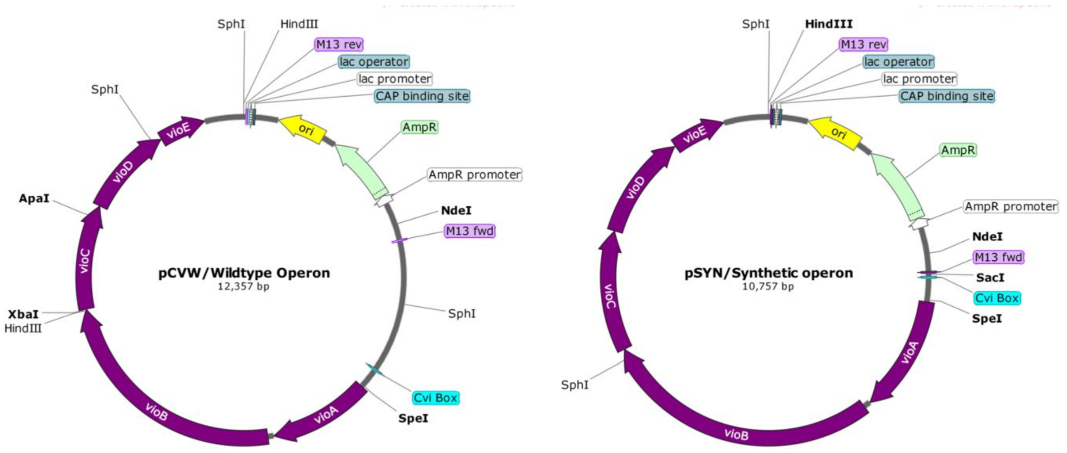

SphI
HindIII
M13 rev
lac operator
lac promoter
CAP binding site
SphI
vioE
ori
AmpR
AmpR promoter
vioD
ApaI
NdeI
M13 fwd
vioC
pCVW/Wildtype Operon
12,357 bp
SphI
XbaI
HindIII
Cvi Box
SpeI
vioA
vioB
SphI
HindIII
M13 rev
lac operator
lac promoter
CAP binding site
vioE
ori
AmpR
vioD
AmpR promoter
NdeI
M13 fwd
SacI
vioC
pSYN/Synthetic operon
10,757 bp
Cvi Box
SpeI
vioA
SphI
vioB

Figure 3.6 Structure of wildtype (pCVW) and synthetic (pSYN) plasmid

Table 3.4 Reduced cost analysis for the optimal synthesis of violacein using in silico strains

Rxns	**Rxn Names**	**Subsystem**	**Reaction Formula**	***iAF1260***	***iAF1260vio***	**iAF1260: Δ*pgi***	**iAF1260vio: Δ*pgi* SYNO**	**iAF1260: Δ*pheA***	**iAF1260vio: Δ*pheA* WTO**
ALDD2x	aldehyde dehydrogenase (acetaldehyde, NAD)	Alternate Carbon Metabolism	acald[c] + h2o[c] + nad[c] -> ac[c] + 2 h[c] + nadh[c]	-0.00414	-0.00939	-0.00414	-0.00939	2.13E-29	-0.03636
ALDD2y	aldehyde dehydrogenase (acetaldehyde, NADP)		acald[c] + h2o[c] + nadp[c] -> ac[c] + 2 h[c] + nadph[c]	-0.00414	-0.00939	-0.00414	-0.00939	2.52E-29	-0.02727
G6PP	glucose-6-phosphate phosphatase	Glycolysis/ Gluconeogenesis	g6p[c] + h2o[c] -> glc-D[c] + pi[c]	-0.0031	-0.00939	-0.00241	-0.00548	-1.55E-29	-0.02727
FBP	fructose-bisphosphatase		fdp[c] + h2o[c] -> f6p[c] + pi[c]	-0.00414	-0.00939	-0.00414	-0.00939	0.00E+00	-0.03636
PPS	phosphoenolpyruvate synthase		atp[c] + h2o[c] + pyr[c] -> amp[c] + 2 h[c] + pep[c] + pi[c]	-0.00414	-0.00939	-0.00414	-0.00939	1.51E-29	-0.03636
HEX1	hexokinase (D-glucose:ATP)		atp[c] + glc-D[c] -> adp[c] + g6p[c] + h[c]	-0.00103	5.55E-17	-0.00172	-0.00391	1.54E-29	-0.00909
EDD	6-phosphogluconate dehydratase	Pentose Phosphate Pathway	6pgc[c] -> 2ddg6p[c] + h2o[c]	-0.00241	-0.00939	-0.00241	0	0.00E+00	-0.02727
CITL	Citrate lyase	Citric Acid Cycle	cit[c] -> ac[c] + oaa[c]	-0.00414	-0.00939	-0.00414	-0.00939	-2.87E-30	-0.03636
MDH2	Malate dehydrogenase (ubiquinone 8 as acceptor)		mal-L[c] + q8[c] -> oaa[c] + q8h2[c]	-0.0031	-0.00704	-0.0031	-0.00704	1.52E-29	-0.02727
MDH3	Malate dehydrogenase (menaquinone 8 as acceptor)		mal-L[c] + mqn8[c] -> mql8[c] + oaa[c]	-0.0031	-0.00704	-0.0031	-0.00704	1.52E-29	-0.02727
TRPS1	tryptophan synthase (indoleglycerol phosphate)	Tyrosine, Tryptophan, and Phenylalanine Metabolism	3ig3p[c] + ser-L[c] -> g3p[c] + h2o[c] + trp-L[c]	-0.00414	-0.00939	-0.00414	-0.00939	-1.31E-27	-0.04545
TRPS2	tryptophan synthase (indole)		indole[c] + ser-L[c] -> h2o[c] + trp-L[c]	-0.00414	-0.00939	-0.00414	-0.00939	-1.31E-27	-0.04545
GLYCDx	Glycerol dehydrogenase	Alternate Carbon Metabolism	glyc[c] + nad[c] -> dha[c] + h[c] + nadh[c]	-0.00414	1.39E-17	-0.00414	8.33E-17	1.33E-27	-0.00909

$^{\#}$Constraints as mentioned in supplementary table * Total flux difference with respect to WT in case of MOMA simulations $^{\$}$MOMA (Segre et al., 2002) was implemented using MOMA() command in COBRA Toolbox and Tomlab CPLEX was used as the LP and QP solver. 1and 2 are additional constraints as mentioned in Supplementary Information

Table 3.5 Strains used in this study

Strains used	**Genotype**	**Antibiotic Resistance**
E. coli DH10β™	F– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara leu) 7697 galU galK rpsL nupG λ–	Streptomycin
E. coli K12 *MG1655 (ATCC47076)*	F- lambda- ilvG- rfb-50 rph-1	Nil
E. coli K12 *MG1655: ΔtrpR*	F- lambda- ilvG- rfb-50 rph-1Δ*trpR*	Nil
E. coli K12 *MG1655: ΔpheA*	F- lambda- ilvG- rfb-50 rph-1Δ*pheA*	Nil
E. coli K12 *MG1655: Δpgi*	F- lambda- ilvG- rfb-50 rph-1Δ*pgi*	Nil

Metabolic pathway engineering of *E. coli* K12 through gene deletions identified genes, *pgi* and *pheA* that potentially increase tryptophan levels. The Δ*pgi* strain diverts flux through the pentose phosphate shunt, increasing precursor for aromatic amino acid (AA) synthesis and cofactor NADPH levels; and Δ*pheA* strain diverts flux from the other competing aromatic AA phenylalanine and tyrosine into tryptophan synthesis. Additionally, Δ*trpR* strain de-regulates feedback inhibition of tryptophan synthesis.

Table 3.6 Growth rates of engineered E. coli strains

***E. coli* Strain**	**Growth rate in LB (hr^{-1}**
DH10B	0.54
K12	0.52
K12 :ΔtrpR	0.48
K12 :Δpgi	0.50
K12 :ΔpheA	0.48
DH10B: pCVW	0.55
K12 pCVW	0.51
K12 ΔtrpR pCVW	0.21
K12 Δpgi pCVW	0.84
K12 ΔpheA pCVW	0.35
DH10B pSYN	0.17
K12 pSYN	0.29
K12 :ΔtrpR pSYN	0.23
K12 :Δpgi pSYN	0.36
K12 :ΔpheA pSYN	0.26

Growth rates and yields of the deletion strains Δ*pgi,* Δ*trpR* and Δ*pheA* (Figure 3.8 B) in *E. coli* on Luria-Bertani (LB) media were similar with a two-fold increase in tryptophan levels (Figure3.7). The growth rates are listed (Table 3.6). The Δ*pheA* had slightly lower biomass yields while Δ*pgi* and Δ*trpR* had higher yields as compared to wild-type. On minimal media with glucose as growth limiting substrate, however, the growth rates were lower for the deletion strain Δ*pgi* while Δ*trpR* and Δ*pheA* were unaffected (Figure 3.8 D). The rates (Table 3.7) varied significantly for the Δ*pheA* and Δ*pgi* deletion strains, with growth profiles showing a longer lag phase. *E. coli* Δ*trpR* grows very similarly to wild-type *E. coli* K12 (Figure 3.8 C).

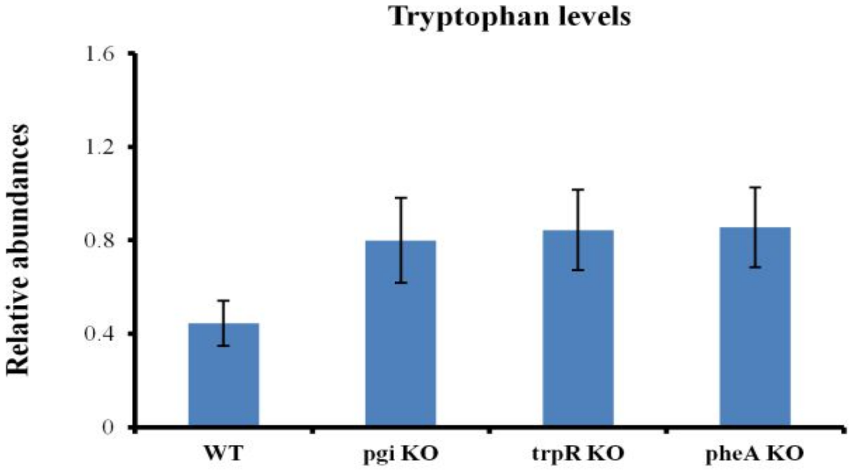


Figure3.7 Tryptophan levels from MALDI-Analysis

Table 3.7 Growth rates of engineered strains on M9 media

Growth on M9 media$^{\#}$						
Strain type	**K12 control**	**Δ *trpR* control**	**Δ*pgi* control**	**Δ*pheA* control**	**K12+ Vio operon**	**Δ*trpR*+Vio operon**
Growth rate	0.6426	0.6796	0.1766	0.5407	0.4704	0.6178

The growth rates of the WTO transformants (Figure 3.8 E) on LB media are similar to that of *E.coli* K12 with similar biomass yields as predicted *in silico* for *E.coli iAF1260vio* (Table 3.6). This indicates that inclusion of the secondary metabolite pathway allowed for removal of excess pathway intermediates tryptophan and NADPH that was being formed in the knockout deletion strains.

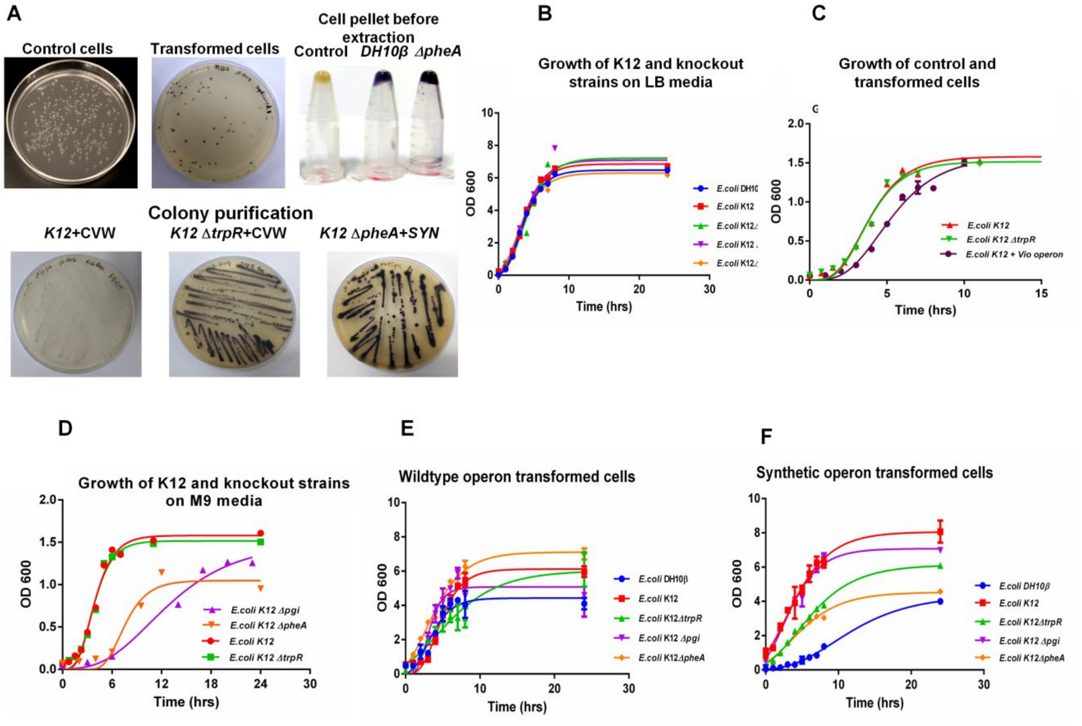


Figure 3.8 Growth and Violacein analysis of genome engineered E. coli strains

(A) Violacein producing *E. coli* transformants (violet colonies) on LB agar and their cell pellets. (B) Growth profiles of *E. coli* K12 and deletion strains on LB media. (C) Growth profiles of *E. coli* K12 and violacein operon transformants on M9 minimal media. (D) Growth profiles of *E. coli* K12 and deletion strains on minimal media. (E) Growth profiles of *E. coli* K12 and wild-type operon transformants (pCVW) on LB media. (F) Growth profiles of *E. coli* K12 and synthetic operon transformants (pSYN) on LB media.

3.4 Violacein production increased through metabolic engineering of *E. coli*

The metabolically engineered strains increased tryptophan levels in *E. coli* K12, had similar growth rates but varying growth yields (Figure 3.8 E). The variation of growth yields was higher among all SYNO transformants (Figure 3.8 F). The lowest growth yields was for the recombinant host *E.coli* DH10*β*. The SYNO *E. coli* transformants of Δ*trpR* and Δ*pheA* showed similar growth rates but higher growth yields (*trpR*).

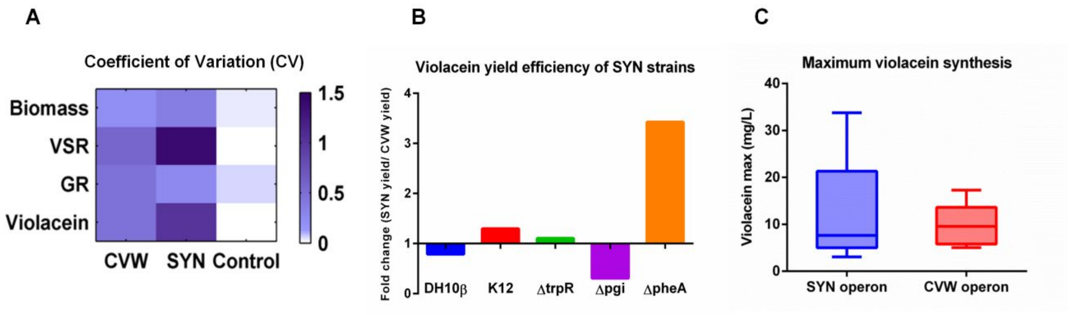


Figure 3.9 Violacein yield variation, Comparison and Range

(G) Coefficient of Variation (CV) for biomass, violacein secretion rates (VSR), growth rates (GR) and violacein yields across controls, wildtype and synthetic operon transformants. (H) Differential violacein yields across synthetic and wildtype transformants (I) Maximum violacein synthesis from synthetic and wildtype transformants (pCVW and pSYN).

Table 3.8 Summary of maximum violacein yields by engineered strains

Maximum violacein synthesis	
***E.coli* Strain**	**Violacein max (mg/L)**
DH10B CVW	9.9410
K12 CVW	5.0504
K12 ΔtrpR CVW	9.5464
K12 Δpgi CVW	6.5687
K12 ΔpheA CVW	17.2953
DH10B SYN	7.6420
K12 SYN	8.8647
K12 ΔtrpR SYN	6.9757
K12 Δpgi SYN	3.0604
K12 ΔpheA SYN	33.7566

The deletion Δ*pgi* in *E. coli* did not change growth rates. The final growth yields were however lower when the diversion through Pentose phosphate pathway (PPP) was utilized. The highest growth rates (with a 40% increase as compared to the wild-type strains) were for WTO Δ*pgi E. coli* transformants. All other knock-outs showed 20% lower growth rates (Table 3.6). Maximum violacein was produced by SYNO Δ*pheA E. coli* transformant amongst all strains (Figure 3.9 G-I; Table 3.8). All SYNO *E. coli* transformants produced higher violacein yield, with a three-fold increase relative to the control and a maximum six-fold increase in yields as compared to the WTO in *E. coli*. The highest titer of violacein observed was 33.8 mg/L for the SYNO transformant of the Δ*pheA E. coli* strain (Table 3.8). The wild type K12 transformant had the lowest yield (5.1mg/L). The variation in violacein production yields and rates were lower for all WTO transformants and showed.

3.5 Conclusion

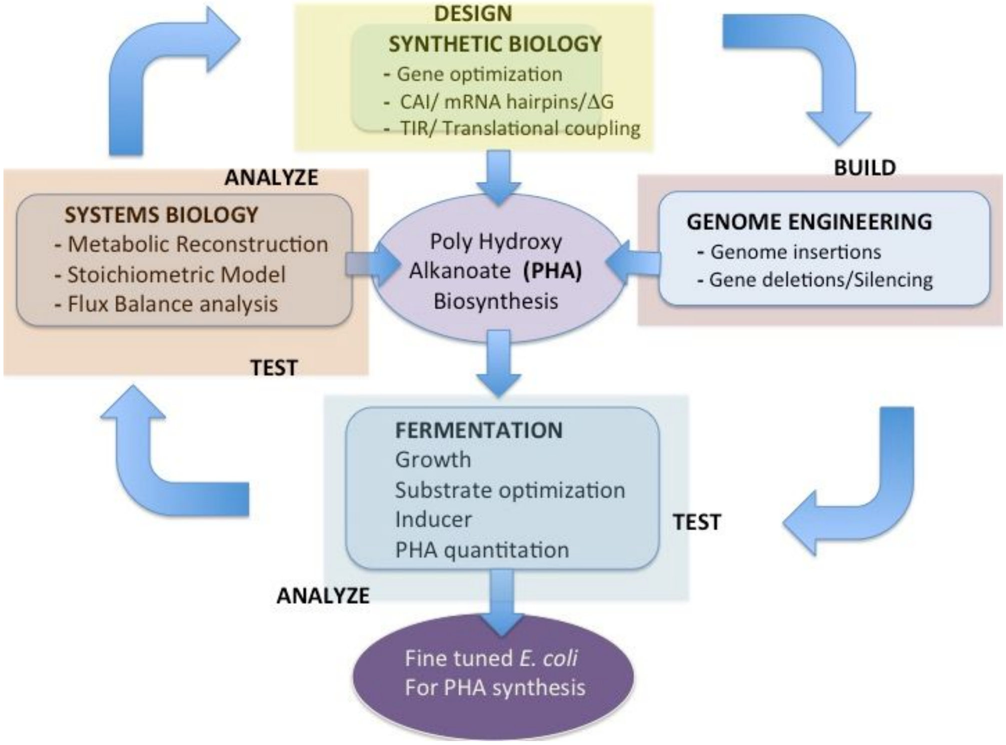
This chapter discussed an integrated approach for rational strain design and development for violacein biosynthesis. The maximum increase in violacein levels (six-fold change) observed in the engineered synthetic strain designs were comparable to that found in literature (He et al., 2017) .Four-fold increase in shake flask for a multi deletion operon (Δ*trpR* Δ*tnaA* Δ*pheA*) with simultaneous overexpression of *trpE*, *trpD*, five-fold increase by combinatorial T7 promoter engineering (He et al., 2017) and three-fold increase using multivariate statistical regulatory model with T7 promoter. *E.coli:*Δ*pheA*SYN transformant outperformed all other strains in terms of yields and rates of violacein production. The higher accuracy of prediction of experimental growth yields in al the SYNO transformants is also justified as FBA models are evolutionary optimality models and are known to predict the growth yields better for evolved/adapted populations (Ibarra et al., 2002).

Optimal violacein biosynthesis through minimal changes in operon gene sequences and moving through cell hierarchies via critical gene knockouts in hosts and varying environmental parameters of the design space was illustrated. The resultant design allowed for a magnitude increase in production yields and rates that could be explained through genome scale metabolic modeling. Such integrated computational and experimental strategies are invaluable for creating suites of strains for value added products and are critical to transcending from fundamental synthetic biology to metabolic engineering applications for industry.

Chapter 4

Rational Strain Design for Poly(R)-3-Hydroxybutyrate biosynthesis

> *Design is the transmission of ideas through things.*
>
> *Daisy Ginsberg*


DESIGN
SYNTHETIC BIOLOGY
- Gene optimization
- CAI/ mRNA hairpins/ΔG
- TIR/ Translational coupling
ANALYZE
SYSTEMS BIOLOGY
- Metabolic Reconstruction
- Stoichiometric Model
- Flux Balance analysis
TEST
Poly Hydroxy Alkanoate (PHA) Biosynthesis
BUILD
GENOME ENGINEERING
- Genome insertions
- Gene deletions/Silencing
FERMENTATION
Growth
Substrate optimization
Inducer
PHA quantitation
TEST
ANALYZE
Fine tuned E. coli For PHA synthesis

4.1 Introduction

In the light of the economic and environmental importance of bioplastics, it is indeed imperative to metabolically engineer a host strain for its over production. The main objective of this chapter was to design and develop a framework for engineering of *E. coli* to produce polyhydroxyalkanoates (PHA) using the systems/synthetic biology paradigm. The design principles and engineering tools to build and optimize cellular processes are state of the art today and the synergies allow 3--tiers of optimization: the central dogma, intrinsic regulatory interactions and extrinsic environmental interactions leading to their seamless integration.

The Poly(R)-3-Hydroxybutyrate (PHB) biosynthesis operon consists of three genes:*phaA,phaB and phaC*. The genes code for three enzymes catalyzes the conversion of acetyl-CoA into PHB. The reaction cascade includes 3-ketothiolase (PhaA**;** EC 2.3.1.9) that converts Acetyl-CoA to Acetoacetyl-CoA thatis furtherconverted to (R)-3-hydroxybutyryl-CoA (3HB-CoA) by Acetoacetyl-CoA reductase (PhaB**;** EC 1.1.1.36) using the reducing power of $NADPH/H^+$. 3HB-CoA is polymerized into PHB by a PHA synthase (PhaC; EC 2.3.1.-). Several successful attempts have been made to clone PHB biosynthetic pathway into *E.coli* to produce PHB and its co-polymers (S. C. Lee et al., 1994; S. Y. Lee et al., 1994; Mahishi et al., 2003; Mahishi and Rawal, 2002; Schubert et al., 1988; Steinbüchel et al., 1993) by introducing the entire operon into the host.There are several reports where an unnatural monomer is co-polymerized along with PHB(Choi et al., 2016; Lee et al., 1996a; Park et al., 2012). However, *E.coli*potentially does not require entire operon for PHB biosynthesis. *E.coli* produces (S) enantiomer of 3HB from central metabolism and fatty acid degradation and that can be converted to (R) isomer by 3-hydroxybutaryl-CoA epimerase (FadJ;5.1.2.3). Since *E.coli* can produce 3HB, instead of introducing entire PHB producing operon, enzymes augmenting the incomplete PHB producing pathway were introduced into*E.coli*. Here we have used the synthetic genetic circuit (Figure 4.1) constructed using two enzymes, including Propionate-CoA Transferase (Pct_{ap}; EC 2.8.3.8) from *Acetobacterpasteurianus* transfers CoA group from acyl-CoA (mostly Butanoyl-CoA) to acetate to give acetyl-CoA and Polyhydroxyalkanoate polymerase ($Phac_{cv}$) from *Chromobacteriumviolaceium*, thatpolymerizes the monomer into PHB. The Pct_{ap} incorporated to increase the pool of acetyl-CoA by utilizing acetate produced during *E.coli* growth glucose and glycerol supplemented LB media, due tooverflow metabolism.

4.2 Genetic circuit design

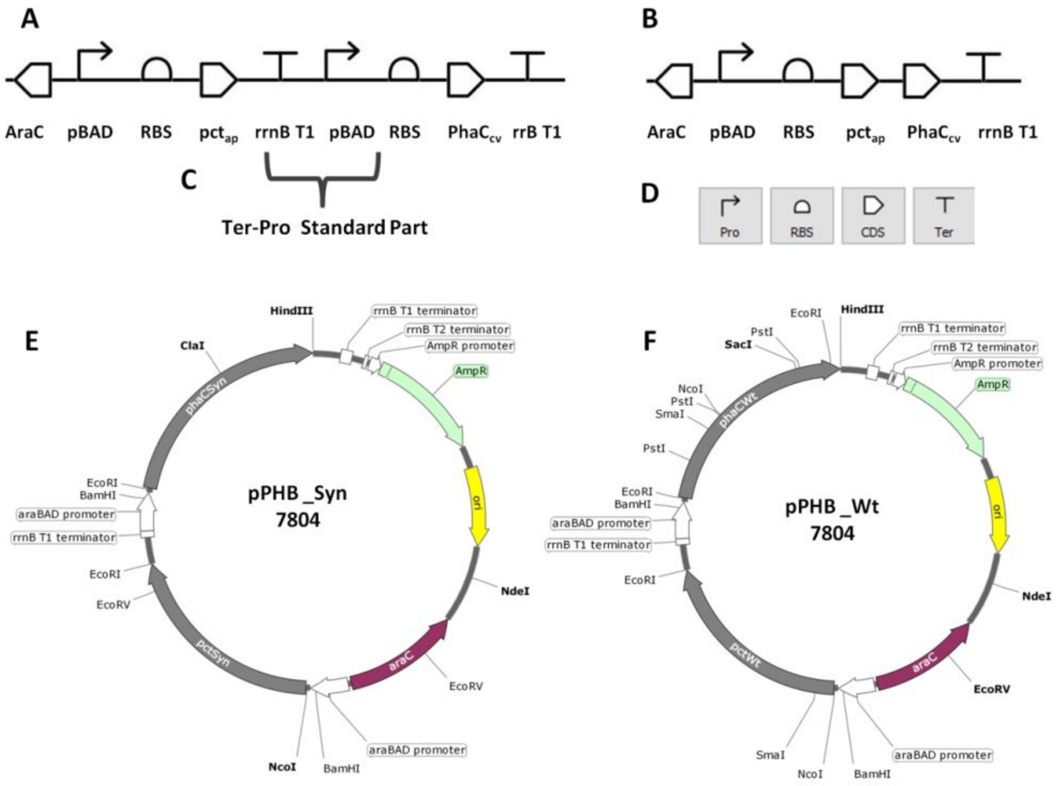

A
AraC
pBAD
RBS
pct$_{ap}$
rrnB T1
pBAD
RBS
PhaC$_{cv}$
rrB T1
C
Ter-Pro Standard Part
B
AraC
pBAD
RBS
pct$_{ap}$
PhaC$_{cv}$
rrnB T1
D
Pro
RBS
CDS
Ter
E
pPHB _Syn
7804
HindIII
ClaI
rrnB T1 terminator
rrnB T2 terminator
AmpR promoter
AmpR
ori
NdeI
araC
EcoRV
araBAD promoter
BamHI
NcoI
pctSyn
EcoRV
EcoRI
rrnB T1 terminator
araBAD promoter
BamHI
EcoRI
phaCSyn
F
pPHB _Wt
7804
EcoRI
HindIII
PstI
SacI
rrnB T1 terminator
rrnB T2 terminator
AmpR promoter
AmpR
NcoI
PstI
SmaI
PstI
phaCWt
EcoRI
BamHI
araBAD promoter
rrnB T1 terminator
EcoRI
ori
NdeI
pctWt
araC
EcoRV
SmaI
NcoI
BamHI
araBAD promoter

Figure 4.1 Genetic circuit design

A) Monocistronic genetic circuit design B) Polycistronic Genetic geneticcircuit design C) Ter-Pr standard part for the design of monocistronic design in one plasmid with multiple genes D) SBOL 2. Standard symbols for standard p parts E) Synthetic plasmid construct (pBAD_Syn) F) Wild typ plasmid construct (pBAD_Wt)

Unlike natural operon structure, synthetic biology approach was used to create the genetic circuit from the standard parts. Both monocistronic design (Figure 4.1 A) (with standard part Ter-Pro Section 2.2.4) and polycistronic design (Figure 4.1 B) (without standard part Ter-Pro (Figure 4.1 C) Section 2.2.4) were investigated using RBS calculator (https://salislab.net/software/) and to mimic the 1:1 translational coupling between two genes monocistronic design was adopted, such that the expression of preceding gene (pct$_{ap}$) should not affect the translation of following genes.

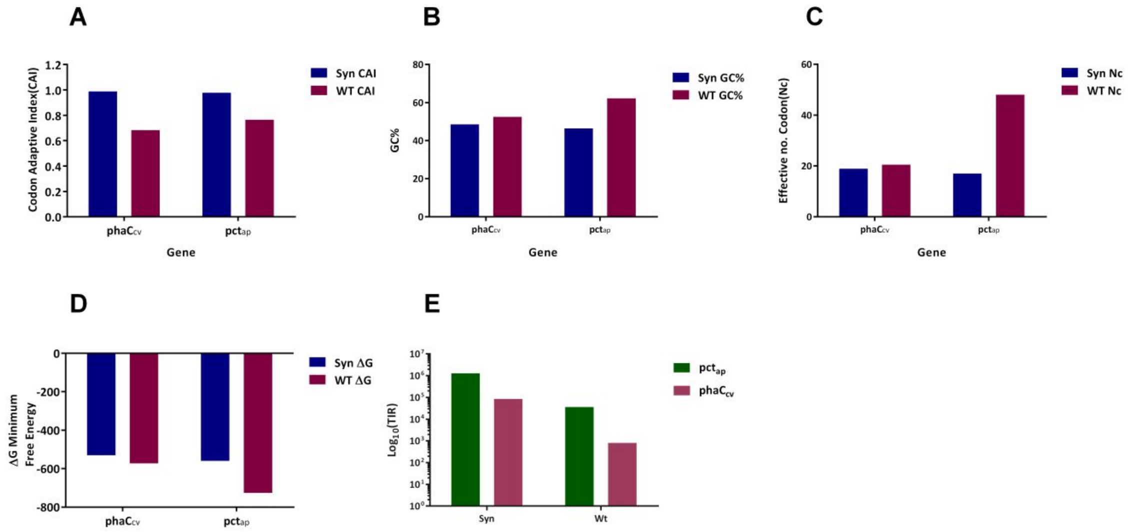


Figure 4.2 Sequence optimization for PHB_Wt and PHB_Syn genetic circuit
(A) Codon adaptive index (CAI) B) GC% (C) Effective no of codons (D) ΔG Minimum free energy
(E) Transcription Initiation Rate (TIR)

\Two gene pathways for biosynthesis of PHB in *E.coli* i.e each gene is controlled by separate pBAD promoter and terminator and placed in tandem. The Gibson assembly method(Gibson et al., 2009) was successful for stitching the wild-type operon and synthetic was chemically synthesized and cloned by Geneart® Gene synthesis services. The gene sequences of synthetic PHB operon were optimized to maximize Codon Adaptive Index (CAImax ~1), optimum GC content between 40% - 50% as compared to the wild-type PHB operon. It was observed that the CAI for synthetic phac$_{cv}$ and pct$_{ap}$ is 0.98 and 0.97 respectively as compared to wild-type *phac*$_{cv}$ and *pct*$_{ap}$i.e 0.68 and 0.76 respectively (Figure 4.2 A). The GC% for synthetic *phac*$_{cv}$and *pct*$_{ap}$ was optimized to 48.47 and 46.45 respectively as compared to wild-type *phac*$_{cv}$ and *pct*$_{ap}$i.e 52.84 and 62.26 respectively (Figure 4.2 B). The effective no. of codons (Nc) for synthetic *phac*$_{cv}$ and *pct*$_{ap}$ was optimized to 20.55 and 48.05 respectively as compared to wild-type phac$_{cv}$ and pct$_{ap}$i.e 18.88 and 17.02 respectively (Figure 4.2 C). The analysis of mRNA sequence for the (Minimum Free Energy) and mRNA structure for wild-type and synthetic mRNA was predicted by using RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) The free energy reduced significantly in both the genes from -725.6 and – 572.9 to -558.6 and -530.4 $phac_{cv}$ and pct_{ap} respectively (Figure 4.2 D). As a result of optimization Translation Initiation Rate (TIR) for synthetic *$phaC_{cv}$* and *pct_{ap}* increased by 105 fold and 35 fold respectively as compared to wild type genes (Figure 4.2 E) while maintaining 1:1 translational coupling due to monocistronic design (*Appendix II*)

4.3 PHB production via metabolic engineering of E. coli

The designed plasmid constructs expressed in *E. coli* host via induction with L-arabinose were tested for protein expression and production of PHB. Expression was confirmed by running cell lysate 2100 Bioanalyzer with protein 80 chip and 10% SDS PAGE. Both the proteins, $PhaC_{cv}$ (mol. Wt.63.27 kDa) and Pct_{ap} (mol Wt.) were over expressed in induced samples (Section 2.3.2.1). PhaC cannot be detected in 55.45 kDa bioanalyzer, because ofinsolublity of its active form and potential carbonosome structure. The $PhaC_{cv}$ expression was confirmed by running whole cell lysate on SDS-PAGE. Therefore the high throughput screening and quantitation for protein expression was done on bioanalyzer only for $Pct_{ap.}$ Glucose (substrate) concentrations were optimized to prevent catabolite repression effect on pBAD operon. With 2 mg/ml glucose, Pct_{ap} expression was 607 ng/ml and 428 ng/ml for pPHB_Wt and pPHB_Syn respectively which was optimum for biosynthesis objective (Figure 4.3).

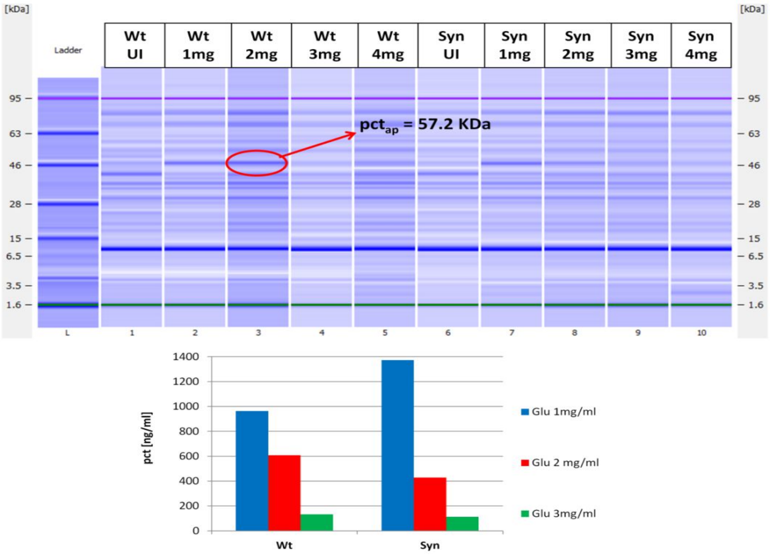


Figure 4.3 Glucose titer for to determine the optimum glucose concentration in LB broth

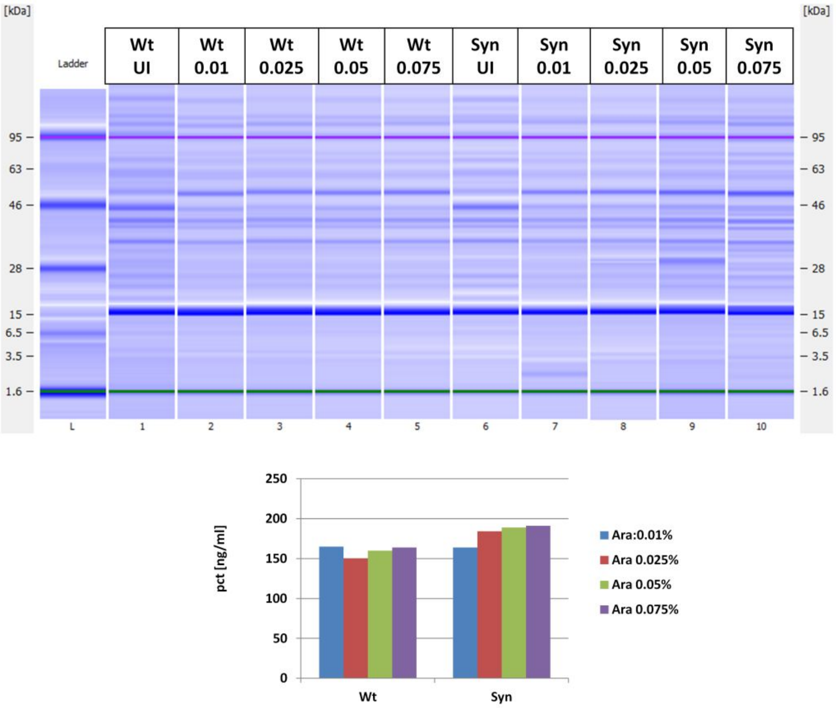


Figure 4.4 Arabinose titration in LB media supplemented with 2 mg/ml glucose

The titration for inducer L- arabinose in the range of 0.01 to 0.075 was done in the presence of 2 mg /ml glucose and no variation was found in the expression level of the protein with average value 153.4 ng/ml (Figure 4.4).

4.4 Substrate utilization and PHB production

PHA biosynthesis on two carbon sources glucose and glycerol were tested. Glycerol and glucose uptake were measured in addition to PHB and biomass production in the tested cells. Strain design and optimization can be done at low substrate concentrations or in buffered media as high concentrations would result in higher substrate level phosphorylation and acid fermentation inhibiting growth. With 2mg/ml glucose as substrate, all carbon was consumed during initial exponential growth (2 to 3 h) (Figure 4.5 B,D) with rate very similar for all *E.coli* strains even while over expressing proteins. Glycerol gets consumed completely in a similar fashion in the early exponential phase (2 to 5 h) (Figure 4.5 A C), but at a slower rate. The substrate utilization rates and growth rates are similar for cells growing on glucose and glycerol at a concentration of 2mg/ml, however final biomass and product yields vary. Growth and PHA production was measured for 24 hrs at regular intervals in LB media and LB media supplemented with glycerol and glucose.

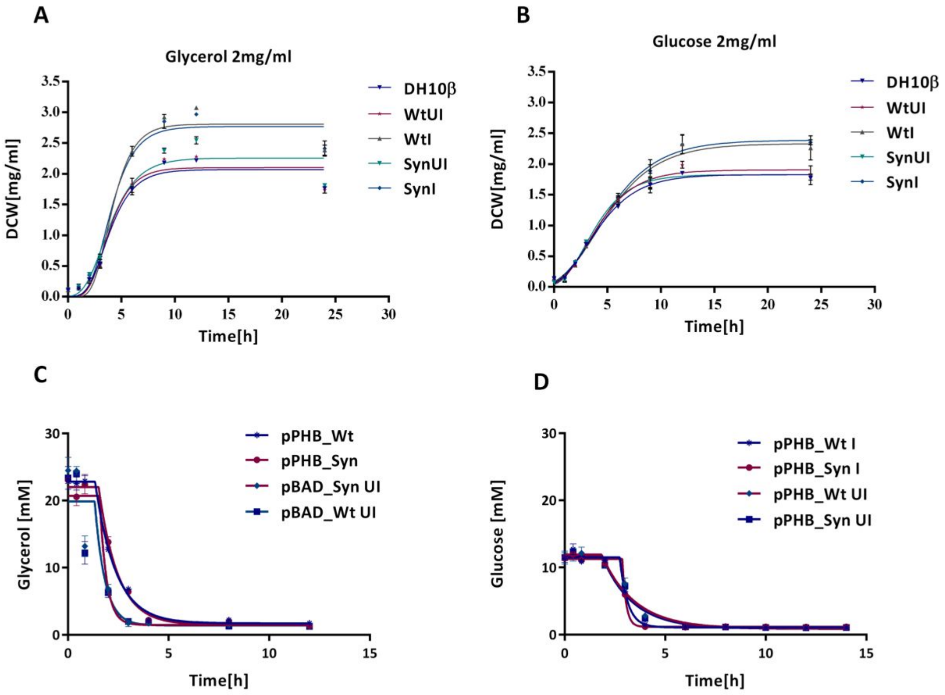


Figure 4.5 Growth and Uptake rate for pPHB_Syn and pPHB_Wt transformed E. coli strains (A) Growth on glycerol (B) Growth on glucose (C) Glycerol uptake rate (D) Glucose uptake rate

Growth was measured using OD at 600nm and converted to gram Dry Cell Weight (gDCW) using a calibration curve. PHB quantitation was done using the developed Nile Red based spectrofluorometric method (Rajankar et al., 2018). With 2mg/ml Glycerol, all carbon was completely consumed by 4 hr. Growth plateaued at 8 hrs, where as maximum PHB was quantitated at 12 hrs yielding 445.12 and 485.08 µg/mgDCW at 24 hrs in pPHB_Syn and pPHB_Wt respectively (Figure 4.6). A minimal amount of PHB was also produced during exponential growth of pPHB_Syn and pPHB_Wt corresponding to 51.35 ± 4.09 µg/mgDCW h^{-1} (Table 4.2)

Table 4.1 Growth rate, Substrate uptake rate and PHB yields for plasmid transformed *E. coli* Strain on Glycerol

	Synthetic		**Wild-type**	
	UI	I	UI	I
Glycerol Uptake rate [mM/gDWh h-^{1}]	-6.75 ± 0.36	-6.98 ± 0.30	-6.91 ± 0.32	-7.05 ± 0.26
Growth Rate [h^{-1}]	0.56	0.66	0.70	0.74
Maximum Biomass Yield [mg]	1.81	2.41	1.46	2.38
PHB YieldMax [μg/mgDCW]	445.12		485.08	
PHBProduction Rate [μg/mgDCWh-^{1}]	51.35 ± 4.09		49.98 ± 3.15	

Table 4.2 Growth rate and Substrate uptake rate for plasmid transformed E. coli Strain on Glucose

	Synthetic		**Wild-type**	
	UI	I	UI	I
Glucose Uptake rate [mM/gDW h^{-1}]	-12.22 ± 0.22	-12.51 ± 0.26	-12.12 ± 0.25	-12.46± 0.28
Growth Rate [h^{-1}]	0.46	0.31	0.42	0.32
Maximum Biomass Yeild [mg]	1.40	1.80	1.23	1.82
PHB Production Max [μg/mgDCW]	-	-	-	-
PHB Production Rate [μg/mgDCW.Hr]	-		-	

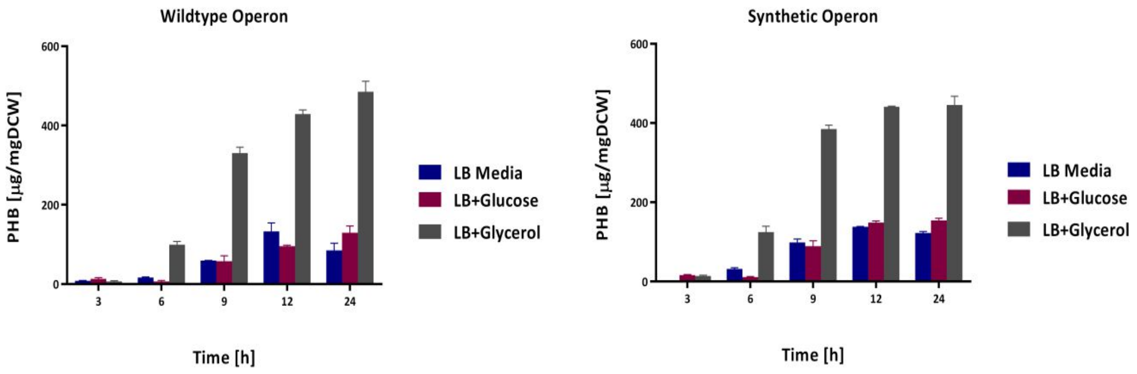


Figure 4.6 PHB yields in *E. coli* strains transformed with plasmid A) pPHB_Wt and pPHB_Syn

4.5 Confirmation of PHA producing cells and Carbonosomes using microscopy

Nile red stained PHB producing cells show intracellular fluorescent red carbonosomes under the fluorescent microscope when excited with 548 nm wavelength (Berlanga et al., 2006; Gorenflo et al., 1999) (Section 2.3.2.3). Nile red is lipophilic dye as discussed in the methods used for the estimation intracellular hydrocarbon and fatty acid(Greenspan et al., 1985; Greenspan and Fowler, 1985; Maes et al., 2017).

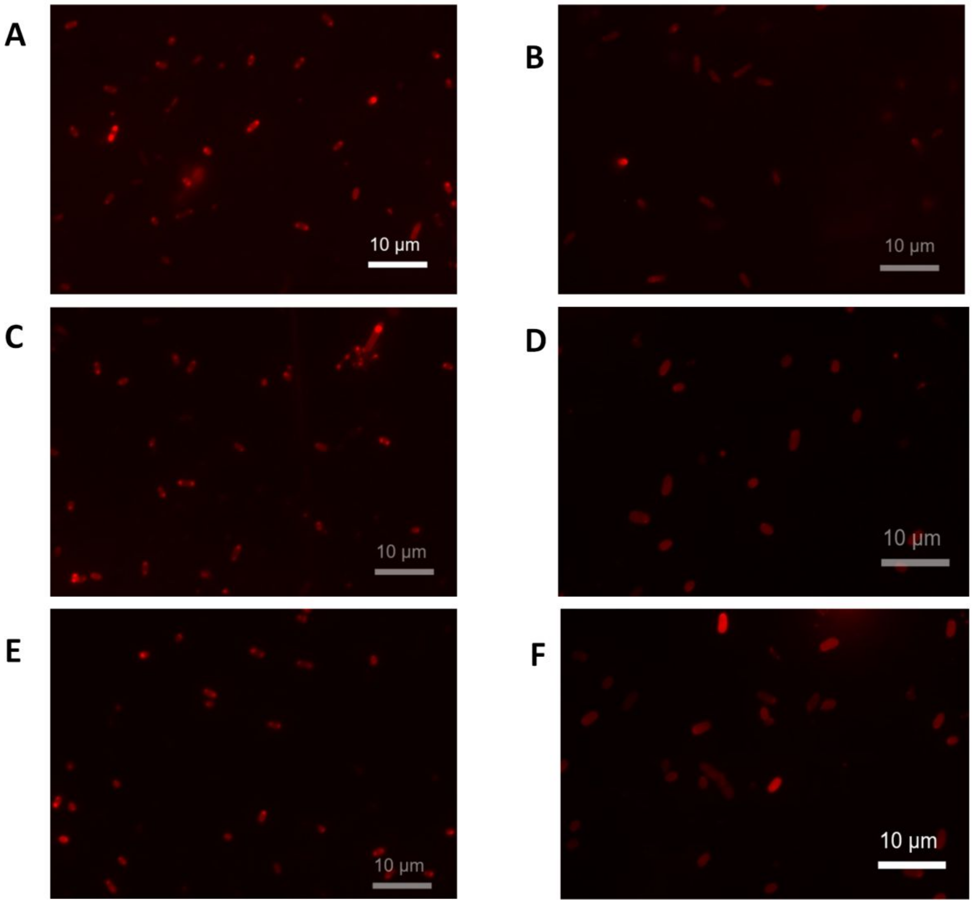


Figure 4.7 Fluorescence microscopy of PHB producing E.coli transformed with pPHB_Syn (A) LB media with glycerol and induced (B) LB media with glycerol un-induced (C) LB media with glucose induced (D) LB media with glucose un-induced (E) Plain LB media induced and (F) Plain LB media un-induced

Carbonosomes were produced by *E.coli* K12 transformed with pPHB_Syn and pPHB_Wt plasmid with a PHA producing genetic circuit inducible by L-arabinose. Glycerol and glucose supplemented LB media and plain LB media showed the presence of carbonosome(Figure 4.7 A,C,E). The un-induced cells did not form acarbonosome(Figure 4.7 B,D,F).

4.6 Confirmation of PHB by NMR spectroscopy

The PHA was extracted for NMR analysis using Trypsin-Sodium hypochlorite method Section (2.3.2.2). ^{1}H and ^{13}C peak characterstic of PHB polymer was observed in the polymer extracted from PHA producing cells. NMR spectrum at 200 Mz was noisy because of contaminations from cells (Figure 4.8 A and B).

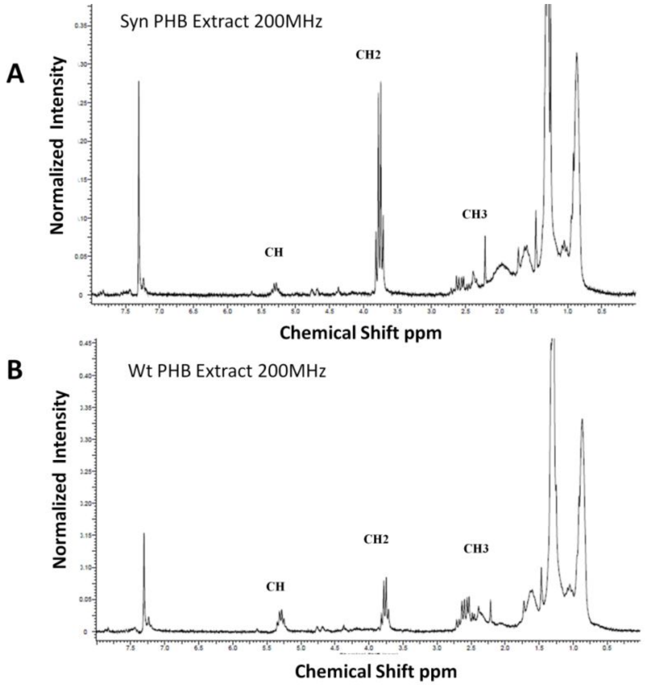


Figure 4.8 1H NMR spectrum for polymer extracted from PHB Producing *E.coli* A) pPHB_Syn and B) pPHB_Wt growing in LB media supplemented with glycerol and induced with L-arabinose

To alleviate this diffusion filter NMR spectroscopy was performed (Figure 4.9 A,B,C). The high gradient strength is sufficient to filter out signals from all low molecular weight species in the solution, however the signals from PHB is observed in the spectrum due to the slow diffusion of the polymer. The diffusion coefficient determined for PHB is ~ 2.35 × 10-11 m2/s which is typical for polymeric species corresponding to 1450 3HB units which suggests the Number of average molecular weight (Mn) for PHB is 150,000 KDa . The 2D correlation spectroscopy (COSY) experiment was recorded with 2K × 128 data points and 16 scans (Figure 4.9D). The connectivity pattern between the different types of protons clearly shows that the polymeric species is PHB and there are no co-polymers. In PHB_GISyn cells PHB was not produced in initial 24 hrs of growth rather the PHB was produced only after addition of adding glycerol continuously only after 36 hrs.

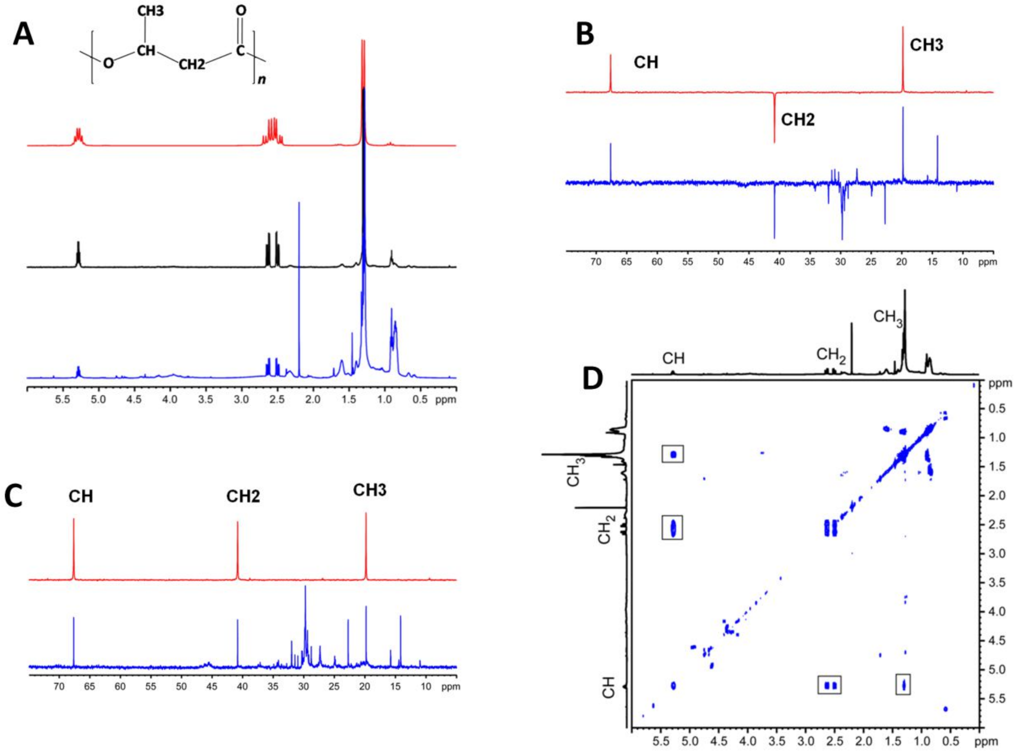


Figure 4.9 NMR Spectroscopy for structural confirmation

(A) 1H NMR spectrum of extract from recombinant E.coli grown on glycerol (blue, 500MHz), diffusion filtered spectrum of the extract (Black, 500MHz) and spectrum of PHB standard (Red, 200MHz) in CDCl3 at 25C. The diffusion filtered spectrum selects signals from macromolecules and eliminates signals from low molecular weight species clearly shows the presence of PHB (B) DEPT spectra (C) 13C spectra of cell extract and PHB standard recorded on a 500 MHz spectrometer. Spectra of cell extract and PHB standard are show in blue and red respectively. The different types of carbons in PHB are indicated (D) COSY spectrum peaks in 2D marked with square

NMR analysis suggested that *E. coli* transformed with pBAD_Syn and pBAD_Wt produce PHB only while growing on glycerol. Cells growing on glucose and LB broth did not produce any PHB. This suggests that formation of carbonosome does not ensure polymerization and potentially monomer levels are very important as well. Integration of heterologous DNA coding for the PHB producing circuit into the genome of *E. coli* would allow for the stable expression without the need for antibiotic selection. Site specific integration of DNA was achieved using through homologous recombination techniques.

4.7 Genome Integration of the PHB circuit into *E. coli*

A preliminary screen of PHA production was conducted on LB plates. Colonies with potential genome integration of the PHB circuit were screened using Agilent DNA 12000 Kit chip based assay after colony PCR (Figure 4.10). Those found positive were tested for expression of Pct_{ap}gene using Agilent Protein 230 Kit chip based assay (Figure 4.10). The three clones found positive were for the PHB_Syn. The genome integration with PHB_Wt was not successful because of large size of the genome integration and heterologous wild type genes could not survive wild type nucleases and recombinases.

The complete lysate was run on SDS-PAGE to detect PhaCcv in insoluble fraction and unlike plasmid does not show the expression of both the protein due to the single copy of the gene (Figure 4.12)

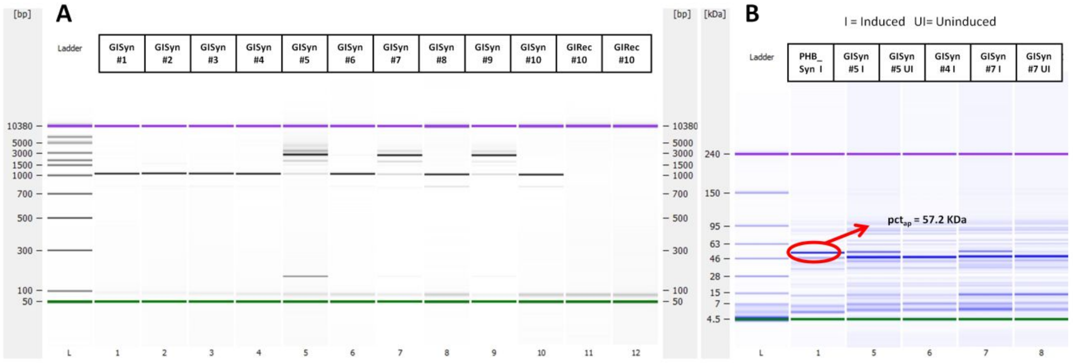


Figure 4.10 Genome Integration of PHB_Syn operon in E.coli genome by Tn7 Transposon

4.7.1 Confirmation of carbonosome formation in the genome integrated PHB producing *E. coli* strain

Fluorescent microscopy of the genome integrated strain PHB_GISyngrown on glycerol supplemented LB media and LB media after induction produced the carbonosome (Figure 4.11A, E). Growth on glucose however did not produce any carbonosome even after induction (Figure 4.11C, D).

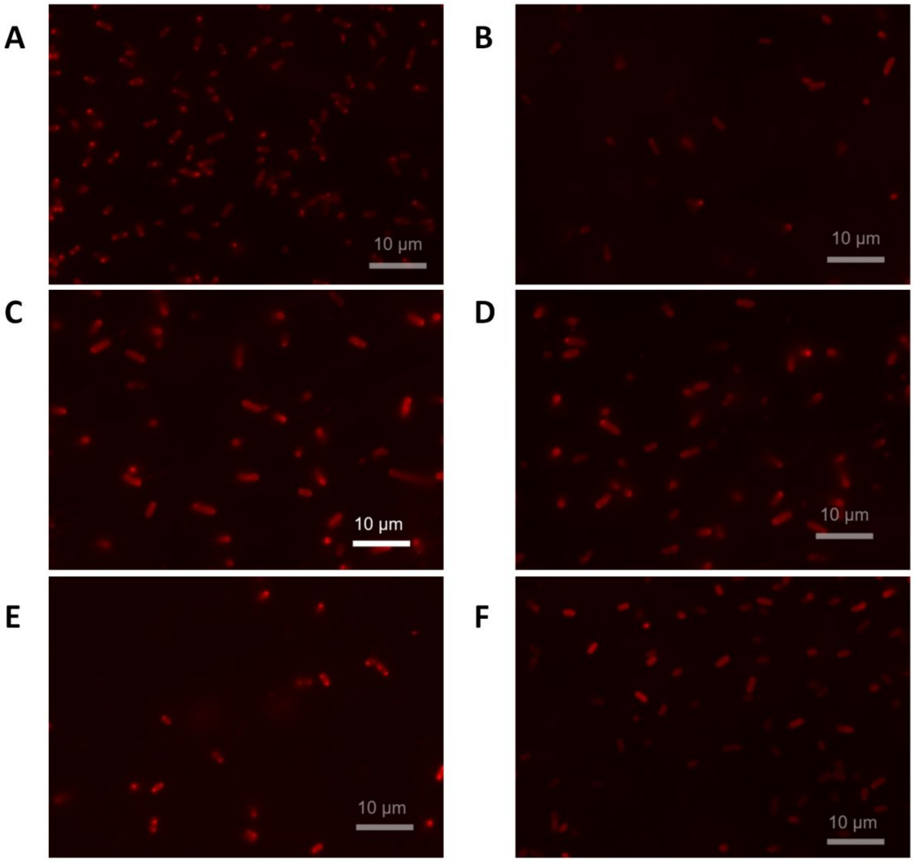
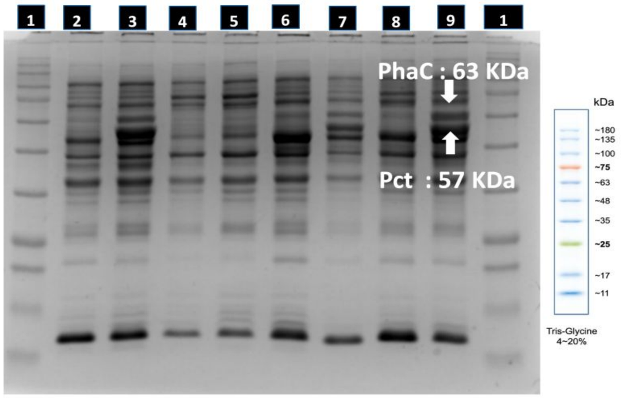

A
B
C
D
E
F
10 µm
10 µm
10 µm
10 µm
10 µm
10 µm

Figure 4.11 Fluorescence microscopy for PHB producing genome integrated (PHB_GISyn) A) LB media with glycerol induced B) LB media with glycerol uninduced C) LB media with glucose induced D) LB media with glucose uninduced E Plain LB media induced and F Plain LB.


1
2
3
4
5
6
7
8
9
1
PhaC : 63 KDa
Pct : 57 KDa
kDa
~180
~135
~100
~75
~63
~48
~35
~25
~17
~11
Tris-Glycine
4~20%

Lane	Strain	Substrate	Induction
1	Ladder	-	-
2	DH10β+pPHB_Syn	LB+Glycerol	+
3	DH10β+pPHB_Syn	LB+Glycerol	-
4	GI_Syn#7	LB+Glucose	+
5	GI_Syn#7	LB+Glucose	-
6	GI_Syn#7	LB Broth	+
7	GI_Syn#7	LB Broth	-
8	GI_Syn#7	LB+Glycerol	+
9	GI_Syn#7	LB+Glycerol	-
10	Ladder	-	-

Figure 4.12 Genome Integrated strain SDS-PAGE Expression of Pctap and PhaC in induced cel

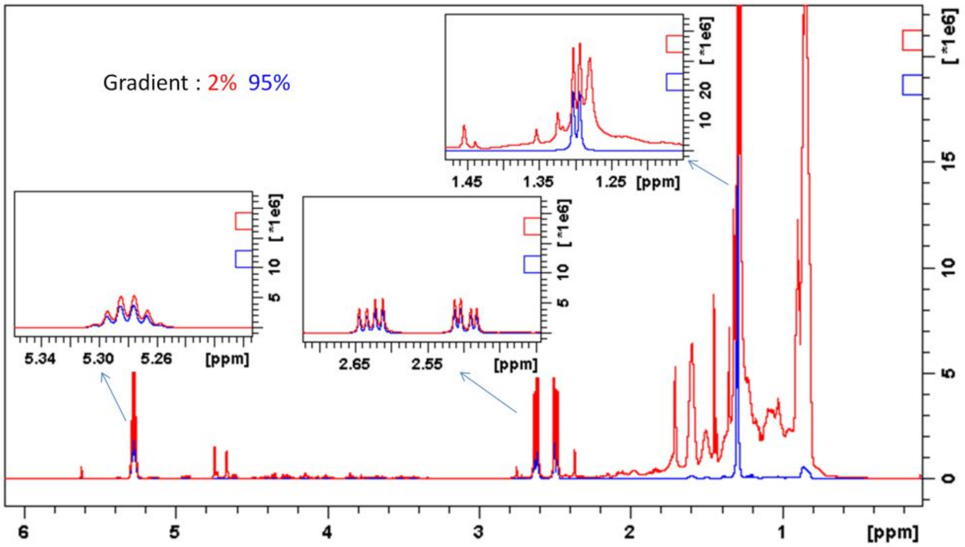


Figure 4.13 ^{1}H NMR spectrum for polymer extracted from *E.coli* transformed

4.8 Confirmation of PHB produced in the genome integrated *E. coli* strain

NMR Spectrometry further confirmed the identity of the PHA as PHB when glycerol was the carbon source. No polymer was detected during growth on Glucose supplemented LB media and LB media. One of the potential causes includes catabolite repression of pBAD promoter in presence of glucose. Another reason could potentially be the differential impact of an indigenous AraC in *E.coli K12* in PHB_GISyn.

4.9 Growth and PHB production in the genome integrated PHB producing *E. coli* strain

Growth experiments on glucose and glycerol indicated growth rates and yields similar to plasmid strains. The PHB was produced to a high concentration of 795 µg/ml when glycerol was used as carbon source at 24 hours of growth. Growth rates were similar for PHB_GISyn grown on high and low substrate concentrations. The final biomass yields were higher at 20mg/ml substrate concentrations. No PHB was formed while growing on glucose. Although, the glycerol uptake rates were similar, the maximum yield of PHB was dependent on concentration of substrate (higher on 20 mg/ml glycerol).

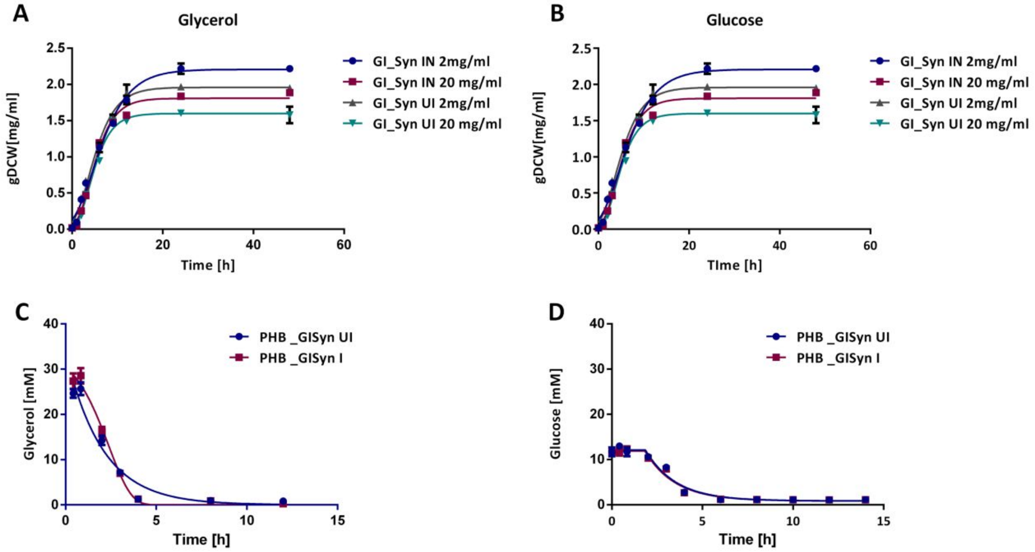


Figure 4.14 Growth and substrate utilization for PHB_GISyn on LB media supplemented with glucose and glycerol

Table 4.3 Growth rate, Substrate uptake rate and PHB yields for PHB_GISyn

Glycerol	**PHB_GISyn**	
	UI	**I**
Uptake rate [mmols/h]	7.71	8.39
Growth Rate [h^{-1}]	0.42	0.42
Maximum Biomass Yeild [mgDCW]	3.13	3.28
PHB Production [µg/mgDCW]	-	395 .41

Glucose	**PHB_GISyn**	
	UI	**I**
Uptake rate [mmols/h]	2.75	3.83
Growth Rate [h^{-1}]	0.34	0.34
Maximum Biomass Yeild [mgDCW]	2.80	2.66
PHB Production [µg/mgDCW]	-	-

4.10 *Insilico* analysis of PHB producing strain

A four reaction module for augmenting the genome scale *E. coli* model iJO1366 (Orth et al., 2014) with PHB production capabilities through action of Acyl-CoA transferase and Polyhydroxylakanoate Synthase reactions was developed (Table 4.4). The degree of polymerization calculated from the NMR data was 1700. This was incorporated into the stoichiometric reaction vector specific for our system. The set of constraints obtained from the experimental data and robustness analysis. For simulating LB media condition reduced costs on catabolizable 11 amino acids exchange reactions were used as a fixed uptake. Reduced cost is the sensitivity of the objective function in response to change in fluxes of a particular reaction and its effect on the objective these included. The +ve value of reduced cost suggests the secretion of these amino acid at the cost of growth which is the objective (Table 4.5). Thus by fixing uptake value to the –ve of the reduced cost we are reliving the *in silico* model from extra cost of making these amino acids. Other constraints were taken from experimental results (i) Substrate (Glucose and Glycerol) uptake rates (ii) Growth yields (iii) PHB secretion associated with molar growth yields of each strain (Table 4.6).

Table 4.4 Reactions added in iJO1366 GSM for simulating for PHB production for experimental growth and substrate uptake rate

Reaction name	Formula	Reaction Stoichiometric vector
Reactions predicted to be catalyzed by Pct$_{ap}$		
Pct_but	'accoa[c]'.'but[c]'.'ac[c]'.'btcoa[c]	[1,1,-1,-1]
Pct_acac	btcoa[c]'.'acac[c]'.'but[c]'.'aacoa[c]	[1,1,-1,-1]
Pct_ppa	accoa[c]','ppa[c]'.'ac[c]'.'ppaco[c]	[1,1,-1,-1]
Reaction annotated to be catalyzed by Pct$_{ap}$		
Act_but	ac[c]','butcoa[c]'accoa[c]','but[c]	[1,1,-1,-1]
Reaction catalyzed by PhaC$_{cv}$		
PhaC	3hbcoa[c]','PHB[c]','coa[c]	[-1700,1,,1700]
Ex_PHB[c]	'PHB[c] -> ' 'PHB[e] <=> '	-
Ex_PHB[e]	'PHB[e] <=> '	-

The incorporation of the reactions from genetic circuit in GSM has estimated the difference in the growth yield for the glucose and glycerol with the accuracy of 80 % and above. The flux balance analysis model clearly predicts better yield of PHB with better accuracy on glycerol as the C-source than glucose.

Table 4.5 Constraints to simulate LB media conditions to derived from reduced cost analysis

Amino Acids	Exchanges	-ve Reduced Cost
L-Alanine	EX_ala-L(e)	-0.046857
L-Aspartate	EX_asp-L(e)	-0.052064
L-Arginine	EX_arg-L(e)	-0.112458
L-Aspargine	EX_asn-L(e)	-0.056229
L-Cysteine	EX_cys-L(e)	-0.089550
L-Glutamine	EX_gln-L(e)	-0.076013
L-Glutamate	EX_glu-L(e)	-0.074972
Glycine	EX_gly(e)	-0.025511
L-Lysine	EX_lys-L(e)	-0.122871
L-Proline	EX_pro-L(e)	-0.094757
L-Serine	EX_ser-L(e)	-0.043734
L-Threonine	EX_thr-L(e)	-0.076013
L- Tryptophan	EX_trp-L(e)	-0.211381

Table 4.6 Constraints used to simulate Induced and Un-induced condition

Constraints for Induced condition		
Reaction	**Lower bound (lb) mmole.gDCW^{-1} h^{-1}**	**Upper Bound (ub)mmole.gDCW^{-1} h^{-1}**
Glc/Gly Exchange	-1	-1
PhaC	1	1000
PHB Exchange	Fixed Experimental value	
3-HB Exchange	0	1000
Constraints for Uninduced condition		
Glc/Gly Exchange	-1	-1
PhaC	0	0
PHB Exchange	0	1000

Table 4.7 FBA prediction after applying experimental constraints

Glycerol				
Strain	**GR/SUR**	**PHB Production**	**Robustness 3-HB**	**Model Prediction**
SynUI	0.065	-	0.0834	0.0647
SynI	0.068	3.4×10^4	-	0.0637
WtUI	0.067	-	0.0755	0.0691
WtI	0.070	3.3×10^4	-	0.0650
Glucose				
SynUI	0.114	-	0.0782	0.1137
SynI	0.089	0	0.1229	0.1077
WtUI	0.104	-	0.0950	0.1044
WtI	0.084	0	0.1313	0.1077

Flux Variability Analysis was performed to understand the redundancy & flexibility in the network comparison of FVA with the two C-sources (Glucose and Glycerol) 41 out of 1608 internal metabolic reactions change their flux direction. Constraints-based methods like Flux variability analysis (FVA) and correlated sets of analysis can identify the redundancy of the network and identify establish flux redistribution or metabolic rewiring towards fatty acid metabolism critical for monomer 3HB-CoA synthesis.

4.11 Conclusions

This chapter discussed the need to evaluate pathways in the context of the host organism in order to relieve the metabolic burden associated with import of biosynthetic potential. Understanding genetic circuits and standardizing their function is paramount for strain design. In many previous studies the entire operon containing genes for PHB production have been cloned and expressed into the host organism (S. Y. Lee et al., 1994; Mahishi et al., 2003). Herein, we have shown how minimal genes from a pathway not native to E. coli can be introduced to complement host metabolism and allow for product synthesis. The problems of plasmid not being a good choice for the metabolic engineering due to maintenance costs and antibiotic selection markers were also addressed by integrating the PHB production circuit into the *E. coli*. Methods for screening clones and product evaluation are critical and this chapter used one such scalable method for PHB quantitation from cells developed in this thesis. It is necessary to understand the carbon and energy source utilization and limitations for a system to augment the genetic circuit design, but the extensive media optimization and bioprocess optimization by integrating design concepts, mechanistic modeling and high throughput screening is the essential aspect of strain development.

Chapter 5

Conclusion and Future scope

> *Metabolic Engineers have all the tools—*
>
> *Biology, computing, and engineering rules,*
>
> *Knowledge, experience, perspective on detail.*
>
> *Let's help Metabolic Genomics to set sail.*
>
> *Opportunity's here . . . but now it's time for a beer!*
>
> *For the times, they are a changing'.*
>
> *-James Bailey*

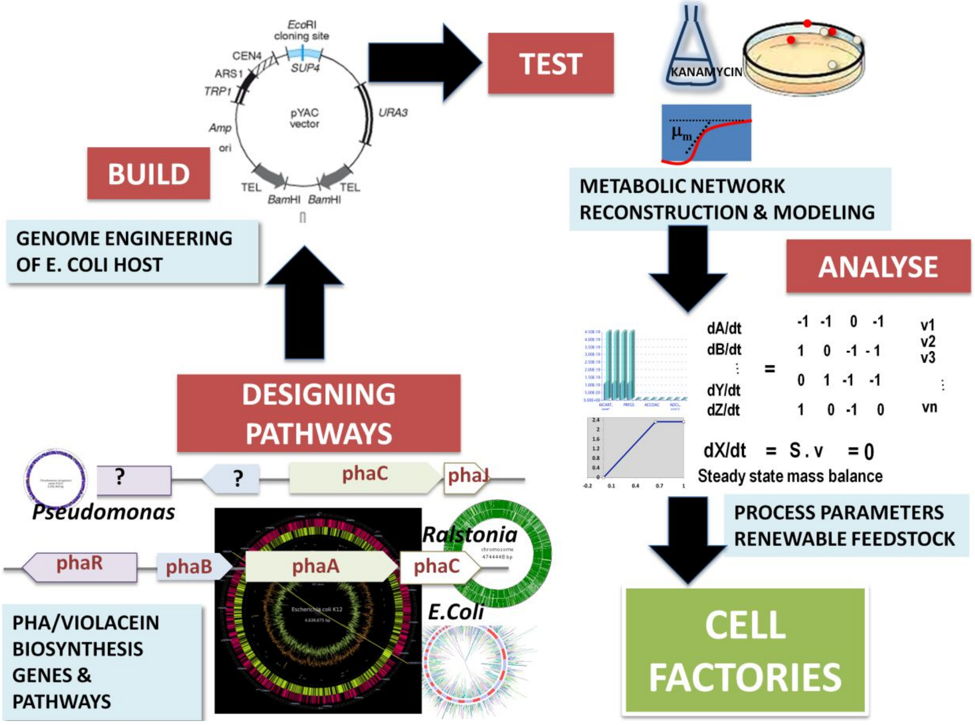

EcoRI
cloning site
CEN4
SUP4
ARS1
TRP1
pYAC
vector
URA3
Amp
ori
TEL
TEL
BamHI
BamHI
TEST
KANAMYCIN
BUILD
GENOME ENGINEERING OF E. COLI HOST
METABOLIC NETWORK RECONSTRUCTION & MODELING
ANALYSE
dA/dt
dB/dt
dY/dt
dZ/dt
=
-1 -1 0 -1
1 0 -1 -1
0 1 -1 -1
1 0 -1 0
v1
v2
v3
vn
dX/dt = S . v = 0
Steady state mass balance
DESIGNING PATHWAYS
?
?
phaC
phaJ
Pseudomonas
Ralstonia
phaR
phaB
phaA
phaC
E.Coli
PROCESS PARAMETERS
RENEWABLE FEEDSTOCK
PHA/VIOLACEIN BIOSYNTHESIS GENES & PATHWAYS
CELL FACTORIES

Figure5.1 Synthetic biology paradigm culminating into cell factories; End of the loop

Microbes can be used as factories- cell factories for bio-based production of chemicals. However, naturally occurring microorganisms in most cases are not capable of producing desired value added products at high efficiencies. Metabolic engineering and synthetic biology and systems biology together can enhance the ability of bio-based product formation at both the cellular and the bioprocess levels. The advent of technologies to understand the molecular mechanisms and regulatory processes underlying the bacterial synthesis of drugs and biopolymers over the past decade has resulted in powerful tools to engineer host organisms that are capable of not only efficient native molecule production but also the production of modified molecules exhibiting unique catalytic or material properties for specific applications. These designer organisms can efficiently convert different carbon sources into a diverse range of polymers with varying chemical and material properties, making the process sustainable with the promise of attainable and viable economic cost. In my thesis I have looked at targeting cellular levels of value added products ranging two ends of the molecular spectrum (a low molecular weight drug and a bioplastic polymer). The work described in this thesis uses a synergistic approach combining the three fields of metabolic engineering, systems biology and synthetic biology using computational and experimental tools for the first time to explore the potential of making a drug molecule violacein and a biopolyester, Poly hydroxy butyrate successfully.

5.1 Recapitulation

Some of the main outcomes described in this thesis are recapitulated here.

- Genes, operons and pathways were designed for drug molecule violacein using first principles for expression in *E. coli*
- Genes were designed as a circuit for biopolymer molecule PHB for expression in *E. coli*
- A novel scalable method was developed for absolute quantitation of PHB
- Constraints-based analysis identified tryptophan and NADPH as limiting for violacein biosynthesis
- Gene sequence optimization of vioABCDE operon increases TIR by a maximum of 50 fold

- Altered growth rates and growth yields in synthetic operon constructs as compared to recombinant or wild type constructs
- Integrative strategies for strain design and development improved yields
- Integration of PHB circuit genes into *E. coli* genome eliminates antibiotic selection and produces PHB

5.2 Future of Metabolic Engineering

Synthetic biology lends itself elegantly to metabolic engineering and microbial cell factories. The design principles and engineering tools to build and optimize cellular processes are state of the art today and encompass the field of synthetic biology. This process allows access to 3--tiers of optimization: the central dogma, intrinsic regulatory interactions and extrinsic environmental interactions leading to their seamless integration. Since synthetic biology in its broadest sense aims to harness the emergent properties of the central dogma for biotechnology and human use, recent paradigm shift towards a systemic approach is central. It can rely heavily on the databases of biochemical, molecular and genomic information based on large--scale experimentation and computational analysis of biochemical networks. The reconstruction of in silico genome scale stoichiometric models of metabolic networks that compute the functional state of the cell accurately facilitate a priori determination of effects of gene deletions and insertions, redox and growth parameters ideally lends itself to this effort. However, the complexity of metabolic pathways, along with regulation often results in unexpected outcomes for metabolic engineering and the resulting strains may require extensive fine--tuning to be economically viable.

Appendix I

Table I List of Strains

Name	Genotypes	Comment
E.coli DH10β	**Genotype:**F^{-} endA1 deoR^{+} recA1 galE15 galK16 nupG rpsL Δ(lac)X74 φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) Str^{R} λ^{-}	Invitrogen
Chromobacterium violaceum	ATCC 12472	Acquired from ATCC
E.coli K12 MG1655 (ATCC47076)	F- lambda- ilvG- rfb-50 rph-1	Acquired from ATCC
E.coli K12 MG1655 ΔtrpR	F- lambda- ilvG- rfb-50 rph-1Δ*trpR*	Ghenerated in Lab from *E.coli K12*
E.coli K12 MG1655 ΔpheA	F- lambda- ilvG- rfb-50 rph-1Δ*trpR*	Ghenerated in Lab from *E.coli K12*
E.coli K12 MG1655 Δpgi	F- lambda- ilvG- rfb-50 rph-1Δ*trpR*	Ghenerated in Lab from *E.coli K12*
PHB_GISyn	F- lambda- ilvG- rfb-50 rph-1Δ*trpR*+ *Genome integrate PHB producing operon PHB-Syn*	Ghenerated in Lab from *E.coli K12*

Table: II List of primers,

Name	**Sequences and Genotypes**	**Comment**
Ter-Pro_FW	GGTTTGAAACTCTAAGGTACGAATTCTTTGGCGGATGAGA Tm: 56.4 ° C	Terminator-Promoter Forward Primer
Ter-Pro_RV	GAGTTGACGAACTGCTGCATGGTGAATTCCTCCTGTTAGC Tm: 58.4 ° C	Terminator-Promoter Reverse Primer
PhaC_FW	GCTAACAGGAGGAATTCACCATGCAGCAGTTCGTCAACTC Tm: 58.4 ° C	phaC $_{cv}$ Forward Primer
PhaC_RV	CTCATCCGCCAAAACAGCCATCATTGCAGGCTGGCGGCG Tm: 63.0 ° C	phaC $_{cv}$ Reverse Primer
Ter_pro	rrnB terminator-araBAD promoter sequence; double stranded Linear DNA	This Study Chemical Synthesis by GeneArt® Gene synthesis Services
SynPHB	Synthetic Genetic circuit containing PHB Biosynthesis operon	This Study Chemical Synthesis by GeneArt® Gene synthesis Services
WtPHB	Genetic Circuit with Wildtype PHB biosynthesis operon	This Study Chemical Synthesis by GeneArt® Gene synthesis Services
VA_FW	AATTCGAGCTCGGTACAACTGCTCCAGCCTTTGCCGC Tm 65.3	Wt VioA Gene Forward Prrimer
VA_RV	CGAACGGTCATGGCTGACAAGC Tm 65.8	Wt VioA Gene Reverse Primer
VB_FW	GCTTGTCAGCCATGACCGTTCG Tm 65.8	Wt Vio BC Forward Primer

VB_RV	GCTTTCCACAAGCCAAATCCAGC	Tm 64.6	Wt Vio BC Reverse Primer
VCD_FW	GCTGGATTTGGCTTGTGGAAAGC	Tm 64.6	Wt Vio DE Genes Forward Primer
VCD_RV	GCCAAGCTTGCATGCCAACATCCCGCCAATTTCCTGGC	Tm: 64.2	Wt Vio DE Genes Reverse Primer
VA_FWSyn	GACCATGATTACGCCAATGAAACATAGCAGCGATATT	Tm: 64.2	Syn VioA Gene Forward Prrimer
VA_RVSyn	AACGGTATTGCTTGCCATATC	Tm: 59.1	Syn VioA Gene Reverse Primer
VB_FWSyn	TATGGCAAGCAATACCGTT	Tm: 60.4	Syn VioB Gene Forward Prrimer
VB_RVSyn	CGTTTCATTTTATGCTTCACG	Tm: 61.2	Syn VioB Gene Reverse Primer
VC_FWSyn	CGTGAAGCATAAAATGAAACG	Tm:58.2	Syn VioC Gene Forward Prrimer
VC_RVSyn	ACCAATAACCAGGATTTTCAT	Tm: 57.9	Syn VioC Gene Reverse Primer
VDE_FWSyn	ATGAAAATCCTGGTTATTGGT	Tm: 61.7	Syn VioDE Gene Forward Prrimer
VCD_RV	GCCAAGCTTGCATGCCAACATCCCGCCAATTTCCTGGC	Tm: 62.2	Syn VioDE Gene Reverse Primer
PLAGI2_Fw:	**AGTAAGCCACGTTTTAATTAATCAGATCCC**CATACTCCCGCCATTCAGAG	Tm:	Forward primer for PHB Genetic circuit Genome Integration
PLAGI2_Fw:	**CGAGCGCGCCGTGGCGCGCCTCCTAGGTGC**GAGCGTTCACCGACAAACAAC		Reverse primer for PHB Genetic circuit Genome Integration

Table: III List of plasmids

Name	**Sequences and Genotypes**	**Comment**
pBAD_pct	Propionyl-CoA Transferase pct_{ap} from *Acetobacter pasteurianus 386B* cloned in pBADHis_C_A238	This Study Chemical Synthesis by GeneArt® Gene synthesis Services
pPHB_WT	Wild type Propionyl-CoA Transferase pct_{ap} from *Acetobacter pasteurianus* and wild type $phaC_{cv}$ from *Chromobacterium violaceum* expressed under pBAD promoter	Generated in Lab
pPHB_Syn	Synthetic Propionyl-CoA Transferase pct_{ap} from *Acetobacter pasteurianus* and wild type $phaC_{cv}$ from *Chromobacterium violaceum* expressed under pBAD promoter	This Study Chemical Synthesis by GeneArt® Gene synthesis Services
pSYN	Synthetic Violacein operon $_{cv}$ from *Chromobacterium violaceum* expressed under cloned in pUC19	Generated in Lab
pCVW	Wild type Violacein operon $_{cv}$ from *Chromobacterium violaceum* expressed under cloned in pUC19	Generated in Lab
pGRG36	Tn7 transposon carrying plasmid for genome integration of PHB Operon	Gift from Nancy Craigs lab

Bibliography

Acevedo, A., Aroca, G., Conejeros, R., 2014. Genome-scale NAD(H/(+)) availability patterns as a differentiating feature between Saccharomyces cerevisiae and Scheffersomyces stipitis in relation to fermentative metabolism. PLoS One 9, e87494. doi:10.1371/journal.pone.0087494

Alkhalaf, L.M., Ryan, K.S., 2015. Biosynthetic manipulation of tryptophan in bacteria: Pathways and mechanisms. Chem. Biol. 22, 317–328. doi:10.1016/j.chembiol.2015.02.005

Alper, H., Jin, Y.-S., Moxley, J.F., Stephanopoulos, G., 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. Metab. Eng. 7, 155–164. doi:10.1016/j.ymben.2004.12.003

Andrighetti-Fröhner, C.R., Antonio, R. V., Creczynski-Pasa, T.B., Barardi, C.R.M., Simões, C.M.O., 2003. Cytotoxicity and Potential Antiviral Evaluation of Violacein Produced by Chromobacterium violaceum. Mem. Inst. Oswaldo Cruz. doi:10.1590/S0074-02762003000600023

Appleton, E., Tao, J., Haddock, T., Densmore, D., 2014. Interactive assembly algorithms for molecular cloning. Nat. Methods 11, 657–662. doi:10.1038/nmeth.2939

Au, L.-C., Yang, F.-Y., Yang, W.-J., Lo, S.-H., Kao, C.-F., 1998. Gene Synthesis by a LCR-Based Approach: High-Level Production of Leptin-L54 Using Synthetic Gene in Escherichia coli 8 units Pfu DNA ligase (Statagene La Jolla, CA) and. Biochem. Biophys. Res. Commun. 248, 200–203.

August, P.R., Grossman, T.H., Minor, C., Draper, M.P., MacNeil, I.A., Pemberton, J.M., Call, K.M., Holt, D., Osburne, M.S., 2000. Sequence analysis and functional characterization of the violacein biosynthetic pathway from Chromobacterium violaceum. J. Mol. Microbiol. Biotechnol. 2, 513–9. doi:11075927

Avery, O.T., 1944. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES: INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. J. Exp. Med. 79, 137–158.

doi:10.1084/jem.79.2.137

Balibar, C.J., Walsh, C.T., 2006. In Vitro Biosynthesis of Violacein from L -Tryptophan by the Enzymes VioA - E from Chromobacterium V iolaceum † 15444–15457.

Banerjee, D., Parmar, D., Bhattacharya, N., Ghanate, A.D., Panchagnula, V., Raghunathan, A., 2017. A scalable metabolite supplementation strategy against antibiotic resistant pathogen Chromobacterium violaceum induced by NAD+/NADH+ imbalance. BMC Syst. Biol. 11, 51. doi:10.1186/s12918-017-0427-z

Berlanga, M., Montero, M.T., Fernández-Borrell, J., Guerrero, R., 2006. Rapid spectrofluorometric screening of poly-hydroxyalkanoate-producing bacteria from microbial mats. Int. Microbiol. 9, 95–102.

Bonnet, J., Yin, P., Ortiz, M.E., Subsoontorn, P., Endy, D., 2013. Amplifying genetic logic gates. Science (80-.). 340, 599–603. doi:10.1126/science.1232758

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657. doi:10.1002/bit.10803

Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94. doi:10.1006/jmbi.1997.0951

Carr, P.A., Park, J.S., Lee, Y.-J., Yu, T., Zhang, S., Jacobson, J.M., 2004. Protein-mediated error correction for de novo DNA synthesis. Nucleic Acids Res. 32, e162. doi:10.1093/nar/gnh160

Cello, J., Paul, A. V., Wimmer, E., 2002. Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. Science (80-.). doi:10.1126/science.1072266

Charusanti, P., Conrad, T.M., Knight, E.M., Venkataraman, K., Fong, N.L., Xie, B., Gao, Y., Palsson, B., 2010. Genetic basis of growth adaptation of Escherichia coli after deletion of pgi, a major metabolic gene. PLoS Genet. 6. doi:10.1371/journal.pgen.1001186

Choi, J.H., Lee, S.J., Lee, S.J., Lee, S.Y., 2003. Enhanced production of insulin-like growth factor I fusion protein in Escherichia coli by coexpression of the downregulated genes identified by transcriptome profiling. Appl. Environ. Microbiol. doi:10.1128/AEM.69.8.4737-4742.2003

Choi, S.Y., Park, S.J., Kim, W.J., Yang, J.E., Lee, H., Shin, J., Lee, S.Y., 2016. One-step fermentative production of poly(lactate-co-glycolate) from carbohydrates in Escherichia coli. Nat. Biotechnol. 34, 1–8. doi:10.1038/nbt.3485

Choi, S.Y., Yoon, K.H., Lee, J. Il, Mitchell, R.J., 2015. Violacein: Properties and production of a versatile bacterial pigment. Biomed Res. Int. doi:10.1155/2015/465056

Chubukov, V., Mukhopadhyay, A., Petzold, C.J., Keasling, J.D., Martín, H.G., 2016. Synthetic and systems biology for microbial production of commodity chemicals. npj Syst. Biol. Appl. 2, 16009. doi:10.1038/npjsba.2016.9

Chung, B.K., Lee, D., 2012. Computational codon optimization of synthetic gene for protein expression. BMC Syst. Biol. 6, 1. doi:10.1186/1752-0509-6-134

Dunn, W.B., Bailey, N.J.C., Johnson, H.E., 2005. Measuring the metabolome: current analytical technologies. Analyst 130, 606. doi:10.1039/b418288j

Durán, N., Justo, G.Z., Durán, M., Brocchi, M., Cordi, L., Tasic, L., Castro, G.R., Nakazato, G., 2016. Advances in Chromobacterium violaceum and properties of violacein-Its main secondary metabolite: A review. Biotechnol. Adv. doi:10.1016/j.biotechadv.2016.06.003

Durán, N., Justo, G.Z., Ferreira, C. V, Melo, P.S., Cordi, L., Martins, D., 2007. Violacein: properties and biological activities. Biotechnol. Appl. Biochem. 48, 127–133. doi:10.1042/BA20070115

Durán, N., Menck, C.F., 2001. Chromobacterium violaceum: a review of pharmacological and industrial perspectives. Crit. Rev. Microbiol. 27, 201–222. doi:10.1080/20014091096747

Edwards, J.S., Palsson, B.O., 1999. Systems properties of the Haemophilus influenzae Rd metabolic genotype. J. Biol. Chem. 274, 17410–17416. doi:10.1074/jbc.274.25.17410

Edwards, J.S., Ramakrishna, R., Palsson, B.O., 2002. Characterizing the metabolic phenotype: A phenotype phase plane analysis. Biotechnol. Bioeng. 77, 27–36. doi:10.1002/bit.10047

Espah Borujeni, A., Channarasappa, A.S., Salis, H.M., 2014. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. 42, 2646–2659. doi:10.1093/nar/gkt1139

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3. doi:10.1038/msb4100155

Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., Palsson, B.O., 2009. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7.

Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Herrgard, M.J., Palsson, B.Ø., 2010. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metab. Eng. 12, 173–186. doi:10.1016/j.ymben.2009.10.003

Ferreira, C.V., Bos, C.L., Versteeg, H.H., Justo, G.Z., Durán, N., Peppelenbosch, M.P., 2004. Molecular mechanism of violacein-mediated human leukemia cell death. Blood 104, 1459–1464. doi:10.1182/blood-2004-02-0594

Gibson, D.G., 2009. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res. doi:10.1093/nar/gkp687

Gibson, D.G., Benders, G. a, Andrews-Pfannkoch, C., Denisova, E. a, Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M. a, Merryman, C., Young, L., Noskov, V.N., Glass, J.I., Venter, J.C., Hutchison, C. a, Smith, H.O., 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319, 1215–1220. doi:10.1126/science.1151721

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E.A., Young, L., Qi, Z.-Q., Segall-Shapiro, T.H., Calvey, C.H., Parmar, P.P., Hutchison, C.A., Smith, H.O., Venter, J.C., 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science (80-.). 329, 52–56. doi:10.1126/science.1190719

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C. a, Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. doi:10.1038/nmeth.1318

Gonzalez, R., Tao, H., Purvis, J.E., York, S.W., Shanmugam, K.T., Ingram, L.O., 2003. Gene Array-Based Identification of Changes That Contribute to Ethanol Tolerance in Ethanologenic Escherichia coli: Comparison of KO11 (Parent) to LY01 (Resistant Mutant). Biotechnol. Prog. 19, 612–623. doi:10.1021/bp025658q

Gorenflo, V., Steinbüchel, A., Marose, S., Rieseberg, M., Scheper, T., 1999. Quantification of bacterial polyhydroxyalkanoic acids by Nile red staining. Appl. Microbiol. Biotechnol. 51, 765–772. doi:10.1007/s002530051460

Grantham, R., Gautier, C., Gouy, M., Mercier, R., Pavé, A., 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. 8, 197. doi:10.1093/nar/8.1.197-c

Greenspan, P., Fowler, S.D., 1985. Spectrofluorometric studies of the lipid probe, nile red. J. Lipid Res. 26, 781–789.

Greenspan, P., Mayer, E.P., Fowler, S.D., 1985. Nile Red " A Selective Fluorescent Stain for Intracellular Lipid Droplets. J. Cell Biol. 100, 965–973.

Griffith, F., 1928. The Significance of Pneumococcal Types. J. Hyg. (Lond). 27, 113–159. doi:10.1017/S0022172400031879

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C., Jahn, D., 2005. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 33. doi:10.1093/nar/gki376

H C Crick, W.F., 1953. Molecular Structl,lre of Deoxypentose Nucleic Acids. Astbury, W. “1.”., Syrnp. Soc. Exp. BioI. I, Nucleic Acid 171, 192.

Han, M.J., Jeong, K.J., Yoo, J.S., Lee, S.Y., 2003. Engineering Escherichia coli for Increased Productivity of Serine-Rich Proteins Based on Proteome Profiling. Appl. Environ. Microbiol. doi:10.1128/AEM.69.10.5772-5781.2003

Haselkorn, R., Artur, L., Bataus, M., Batista, S., Teno, C., 2003. The complete genome sequence of Chromobacterium violaceum reveals remarkable and exploitable bacterial adaptability 100.

He, L., Xiu, Y., Jones, J.A., Baidoo, E.E.K., Keasling, J.D., Tang, Y.J., Koffas, M.A.G., 2017. Deciphering flux adjustments of engineered E. coli cells during fermentation with changing growth conditions. Metab. Eng. 39, 247–256. doi:10.1016/j.ymben.2016.12.008

Heinemann, M., Panke, S., 2006. Synthetic biology - Putting engineering into biology. Bioinformatics 22, 2790–2799. doi:10.1093/bioinformatics/btl469

Holley, R.W., Everett, G.A., Madison, J.T., Zamir, A., 1965. Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid* Downloaded from. J. Bux,oorcx~ Chem. 240.

Hoshino, T., 2011. Violacein and related tryptophan metabolites produced by Chromobacterium violaceum: Biosynthetic mechanism and pathway for construction of violacein core. Appl. Microbiol. Biotechnol. 91, 1463–1475. doi:10.1007/s00253-011-3468-z

Hosokawa, K., Soliev, A.B., Kajihara, A., Enomoto, K., 2016. Effects of a microbial pigment violacein on the activities of protein kinases. Cogent Biol. doi:10.1080/23312025.2016.1259863

Howard-Jones, A.R., Walsh, C.T., 2006. Staurosporine and rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. J. Am. Chem. Soc. 128, 12289–12298. doi:10.1021/ja063898m

Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., Pelletier, J.F., Qi, Z.-Q., Richter, R.A., Strychalski, E.A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K.S., Smith, H.O., Glass, J.I., Merryman, C., Gibson, D.G., Venter, J.C., 2016. Design and synthesis of a minimal bacterial genome. Science (80-.). 351, aad6253-aad6253. doi:10.1126/science.aad6253

Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420, 20–23.

doi:10.1038/nature01195.1.

IGSG, 2013. International Gene Synthesis Consortium [WWW Document]. URL https://genesynthesisconsortium.org/ (accessed 8.6.18).

Ikemura, T., 1981. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the E. coli translational system. J. Mol. Biol. 151, 389–409. doi:10.1016/0022-2836(81)90003-6

Immanuel, S.R.C., Banerjee, D., Rajankar, M.P., Raghunathan, A., 2018. Integrated constraints based analysis of an engineered violacein pathway in Escherichia coli. Biosystems 171, 10–19. doi:10.1016/j.biosystems.2018.06.002

Ingram, L.O., Jarboe, L.R., Zhang, X., Wang, X., Moore, J.C., Shanmugam, K.T., 2010. Metabolic engineering for production of biorenewable fuels and chemicals: Contributions of synthetic biology. J. Biomed. Biotechnol. doi:10.1155/2010/761042

Jiang, P.X., Wang, H.S., Zhang, C., Lou, K., Xing, X.H., 2010. Reconstruction of the violacein biosynthetic pathway from Duganella sp. B2 in different heterologous hosts. Appl. Microbiol. Biotechnol. 86, 1077–1088. doi:10.1007/s00253-009-2375-z

Jung, S., Mcdonald, K., 2011. Visual gene developer : a fully programmable bioinformatics software for synthetic gene optimization. BMC Bioinformatics 12, 340. doi:10.1186/1471-2105-12-340

Jung, Y.K., Kim, T.Y., Park, S.J., Lee, S.Y., 2010. Metabolic engineering of Escherichia coli for the production of polylactic acid and its copolymers. Biotechnol. Bioeng. 105, 161–171. doi:10.1002/bit.22548

Kelly Jr, T.J., Smith, H.O., 1970. A restriction enzyme from Hemophilus influenzae: II. Base sequence of the recognition site. J. Mol. Biol. 51, 393–409. doi:http://dx.doi.org/10.1016/0022-2836(70)90150-6

Kim, H.U., Kim, T.Y., Lee, S.Y., 2008. Metabolic flux analysis and metabolic engineering of microorganisms. Mol. BioSyst. 4, 113–120. doi:10.1039/B712395G

Klamt, S., Mahadevan, R., 2015. On the feasibility of growth-coupled product synthesis in microbial strains. Metab. Eng. 30, 166–178. doi:10.1016/j.ymben.2015.05.006

Kodumal, S.J., Patel, K.G., Reid, R., Menzella, H.G., Welch, M., Santi, D. V, 2004. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. Proc Natl Acad Sci U S A 101, 15573–15578. doi:10.1073/pnas.0406911101

Lee, I.Y., Kim, M.K., Park, Y.H., Lee-, S.Y., 1996a. Communication to the Editor Regulatory Effects of Cellular Nicotinamide Recombinant Escherichia coli 52, 707–712.

Lee, I.Y., Kim, M.K., Park, Y.H., Lee, S.Y., 1996b. Regulatory effects of cellular nicotinamide nucleotides and enzyme activities on poly(3-hydroxybutyrate) synthesis in recombinant Escherichia coli. Biotechnol. Bioeng. doi:10.1002/(SICI)1097-0290(19961220)52:6<707::AID-BIT8>3.0.CO;2-S

Lee, J., Sung, B., Kim, M., Blattner, F.R., Yoon, B., Kim, J., Kim, S., 2009. Metabolic engineering of a reduced-genome strain of Escherichia coli for L-threonine production. Microb. Cell Fact. 8, 2. doi:10.1186/1475-2859-8-2

Lee, J.W., Na, D., Park, J.M., Lee, J., Choi, S., Lee, S.Y., 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat. Chem. Biol. 8, 536–46. doi:10.1038/nchembio.970

Lee, S.C., Lee, S.Y., Chang, H.N., Chang, Y.K., Woo, S.I., 1994. Production of Poly(3-Hydroxybutyric Acid) by Fed-Batch Culture of AIcaIigenes eutrophus with Glucose Concentration Control Beom 43, 892–898.

Lee, S.J., Lee, D.-Y., Kim, T.Y., Kim, B.H., Lee, J., Lee, S.Y., 2005. Metabolic Engineering of Escherichia coli for Enhanced Production of Succinic Acid, Based on Genome Comparison and In Silico Gene Knockout Simulation. Appl. Environ. Microbiol. 71, 7880–7887. doi:10.1128/AEM.71.12.7880-7887.2005

Lee, S.Y., Lee, K.M., Chan, H.N., Steinbüchel, A., 1994. Comparison of recombinant Escherichia coli strains for synthesis and accumulation of poly-(3-hydroxybutyric acid) and morphological changes. Biotechnol. Bioeng. 44, 1337–1347. doi:10.1002/bit.260441110

Lehman, I.R., Bessman, M.J., Ernest, S., Simms, E.S., Kornberg, A., 1958. ARTICLE : Enzymatic Synthesis of Deoxyribonucleic SUBSTRATES AND PARTIAL

PURIFICATION OF AN ENZYME FROM ESCHERICHIA COLI of Deoxyribonucleic. J. Biol. Chem. 233, 163–170.

Lim, H.N., Lee, Y., Hussein, R., 2011. Fundamental relationship between operon organization and gene expression. Proc. Natl. Acad. Sci. U. S. A. 108, 10626–10631. doi:10.1073/pnas.1105692108

Maes, T., Jessop, R., Wellner, N., Haupt, K., Mayes, A.G., 2017. A rapid-screening approach to detect and quantify microplastics based on fluorescent tagging with Nile Red. Sci. Rep. 7, 1–10. doi:10.1038/srep44501

Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5, 264–276. doi:10.1016/j.ymben.2003.09.002

Mahishi, L.H., Rawal, S.K., 2002. Effect of amino acid supplementation on the synthesis of poly (3-hydroxybutyrate) by recombinant pha + Sa Escherichia coli 805–806.

Mahishi, L.H., Tripathi, G., Rawal, S.K., 2003. Poly (3-hydroxybutyrate) (PHB) synthesis by recombinant Escherichia coli harbouring Streptomyces aureofaciens PHB biosynthesis genes : Effect of various carbon and nitrogen sources 158, 19–27.

Mandal, M., Breaker, R.R., 2004. Gene regulation by riboswitches. Nat. Rev. Mol. Cell Biol. doi:10.1038/nrm1403

Maxam, A.M., Gilbert, W., 1980. [57] Sequencing end-labeled DNA with base-specific chemical cleavages, in: Methods in Enzymology. pp. 499–560. doi:10.1016/S0076-6879(80)65059-9

Meselson, M., Yuan, R., 1968. DNA Restriction Enzyme from E Coli. Nature 217, 1110-. doi:DOI 10.1038/2171110a0

Murphy, K.C., 1998. Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. J. Bacteriol. doi:0021-9193/98/$04.00+0

Muyrers, J.P.P., Zhang, Y., Testa, G., Stewart, A.F., 1999. Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. 27.

Nakamura, Y., Gojobori, T., Ikemura, T., 1999. Codon usage tabulated from the international DNA sequence databases; its status 1999. Nucleic Acids Res. 27, 292.

doi:10.1093/nar/27.1.292

Nielsen, J., Keasling, J.D., 2016. Engineering Cellular Metabolism. Cell 164, 1185–1197. doi:10.1016/j.cell.2016.02.004

Nielsen, J., Keasling, J.D., 2011. Synergies between synthetic biology and metabolic engineering. Nat. Biotechnol. 29, 693–695. doi:10.1038/nbt.1937

Novogrodsky, A., Tal, M., Traub, A., Hurwitz, J., 1966. The enzymatic phosphorylation of ribonucleic acid and deoxyribonucleic acid. II. Further properties of the 5'-hydroxyl polynucleotide kinase. J. Biol. Chem. 241, 2933–2943.

Ohnishi, J., Mitsuhashi, S., Hayashi, M., Ando, S., Yokoi, H., Ochiai, K., Ikeda, M., 2002. A novel methodology employing Corynebacterium glutamicum genome information to generate a new L-lysine-producing mutant. Appl. Microbiol. Biotechnol. 58, 217–23.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.O., 2014. A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. Mol. Syst. Biol. 7, 535–535. doi:10.1038/msb.2011.65

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat Biotechnol 28, 245–248. doi:10.1038/nbt.1614.What

Palego, L., Betti, L., Rossi, A., Giannaccini, G., 2016. Tryptophan biochemistry: Structural, nutritional, metabolic, and medical aspects in humans. J. Amino Acids. doi:10.1155/2016/8952520

Park, S.J., Lee, S.Y., Kim, T.W., Jung, Y.K., Yang, T.H., 2012. Biosynthesis of lactate-containing polyesters by metabolically engineered bacteria. Biotechnol. J. doi:10.1002/biot.201100070

Patterson, S.D., Aebersold, R.H., 2003. Proteomics: The first decade and beyond. Nat. Genet. doi:10.1038/ng1106

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: A computational framework for redesign of microbial production systems. Genome Res. 14, 2367–2376. doi:10.1101/gr.2872004

Pharkya, P., Burgard, A.P., Maranas, C.D., 2003. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. Biotechnol. Bioeng. 84, 887–899. doi:10.1002/bit.10857

Queiroz, K.C.S., Milani, R., Ruela-de-Sousa, R.R., Fuhler, G.M., Justo, G.Z., Zambuzzi, W.F., Duran, N., Diks, S.H., Spek, C.A., Ferreira, C. V., Peppelenbosch, M.P., 2012. Violacein Induces Death of Resistant Leukaemia Cells via Kinome Reprogramming, Endoplasmic Reticulum Stress and Golgi Apparatus Collapse. PLoS One 7. doi:10.1371/journal.pone.0045362

Raab, D., Graf, M., Notka, F., Wagner, R., Scho, T., 2010. The GeneOptimizer Algorithm : using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization 215–225. doi:10.1007/s11693-010-9062-3

Radwanski, E.R., Last, R.L., 1995. Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. Plant Cell 7, 921–34. doi:10.1105/tpc.7.7.921

Rajankar, M.P., Ravindranathan, S., Rajamohanan, P.R., Raghunathan, A., 2018. Absolute quantitation of poly(R)-3-hydroxybutyric acid using spectrofluorometry in recombinant Escherichia coli. Biol. Methods Protoc. 3. doi:10.1093/biomethods/bpy007

Rehm, B.H.A., 2010. Bacterial polymers: Biosynthesis, modifications and applications. Nat. Rev. Microbiol. doi:10.1038/nrmicro2354

Rehm, B.H.A., Antonio, R. V, Spiekermann, P., Amara, A.A., 2002. Molecular characterization of the poly (3-hydroxybutyrate) (PHB) synthase from Ralstonia eutropha : in vitro evolution , site-speci ¢ c mutagenesis and development of a PHB synthase protein model 1594, 178–190.

Rehm, B.H.A., Steinbu, A., 1999. Biochemical and genetic analysis of PHA synthases and other proteins required for PHA synthesis 25, 3–19.

Reznik, E., Mehta, P., Segrè, D., 2013. Flux Imbalance Analysis and the Sensitivity of Cellular Growth to Changes in Metabolite Pools. PLoS Comput. Biol. 9, e1003195. doi:10.1371/journal.pcbi.1003195

Richardson, S.M., Leslie, ‖, Mitchell, A., Stracquadanio, G., Yang, K., Dymond, J.S., Dicarlo, J.E., Lee, D., Cheng, §, Huang, L.V., Chandrasegaran, S., Cai, Y., Boeke, J.D., Bader, J.S., 2017. Design of a synthetic yeast genome Downloaded from. Science (80-.). 355, 1040–1044.

Richmond, K.E., 2004. Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. Nucleic Acids Res. 32, 5011–5018. doi:10.1093/nar/gkh793

Rodrigues, A.L., Trachtmann, N., Becker, J., Lohanatha, A.F., Blotenberg, J., Bolten, C.J., Korneli, C., Souza, A.O. De, Porto, L.M., Sprenger, G.A., Wittmann, C., 2013. Systems metabolic engineering of Escherichia coli for production of the antitumor drugs violacein and deoxyviolacein. Metab. Eng. 20, 29–41. doi:10.1016/j.ymben.2013.08.004

Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. 27, 946–950. doi:10.1038/nbt.1568

Sanger, F., Donelson, J.E., Coulson, A.R., Kosselt, H., Fischert, D., 1973. Use of DNA Polymerase I Primed by a Synthetic Oligonucleotide to Determine a Nucleotide Sequence in Phage fl DNA (octadeoxyribonucleotide/pulse-labeling/homochromatography/intercistronic region/evolution) 70, 1209–1213.

Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.Ø., 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics 11, 213. doi:10.1186/1471-2105-11-213

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.Ø., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6, 1290–1307. doi:10.1038/nprot.2011.308

Schubert, P., Steinbüchel, a, Schlegel, H.G., 1988. Cloning of the *Alcaligenes eutrophus* genes for synthesis of poly-β-hydroxybutyric acid (PHB) and synthesis of PHB in *Escherichia coli*. J. Bacteriol. 170, 5837–5847. doi:10.1128/JB.170.12.5837-5847.1988

Schuster, S., Pfeiffer, T., Fell, D.A., 2008. Is maximization of molar yield in metabolic networks favoured by evolution? J. Theor. Biol. 252, 497–504. doi:10.1016/j.jtbi.2007.12.008

Sebastian Palluk, D.H.A., 2018. full-text. Nat Biotechnol 36, 645–50. doi:10.1038/nbt.4173 NATURE

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. 99, 15112–15117. doi:10.1073/pnas.232349399

Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., Court, D.L., 2009. Recombineering: A homologous recombination-based method of genetic engineering. Nat. Protoc. doi:10.1038/nprot.2008.227

Smith, H.O., Hutchison 3rd, C.A., Pfannkoch, C., Venter, J.C., 2003. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci U S A 100, 15440–15445. doi:10.1073/pnas.2237126100

Smith, H.O., Hutchison, C.A., Pfannkoch, C., Venter, J.C., 2003. Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. Proc. Natl. Acad. Sci. doi:10.1073/pnas.2237126100

Smith, H.O., Welcox, K.W., 1970. A Restriction enzyme from Hemophilus influenzae. J. Mol. Biol. 51, 379–391. doi:10.1016/0022-2836(70)90149-X

Spaans, S.K., Weusthuis, R.A., van der Oost, J., Kengen, S.W.M., 2015. NADPH-generating systems in bacteria and archaea. Front. Microbiol. 6, 1–27. doi:10.3389/fmicb.2015.00742

Spiekermann, P., Rehm, B.H.A., Kalscheuer, R., Baumeister, D., Steinbüchel, A., 1999. A sensitive , viable-colony staining method using Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds. Arch. Microbiol. 171, 73–80.

Stafford, D.E., Stephanopoulos, G., 2001. Metabolic engineering as an integrating platform for strain development. Curr. Opin. Microbiol. 4, 336–40.

Steinbüchel, A., Debzi, E.M., Marchessault, R.H., Timm, A., 1993. Synthesis and production of poly(3-hydroxyvaleric acid) homopolyester by Chromobacterium violaceum. Appl. Microbiol. Biotechnol. 39, 443–449. doi:10.1007/BF00205030

Stemmer, W.P.C., Crameri, A., Ha, K.D., Brennan, T.M., Heyneker, H.L., 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene. doi:10.1016/0378-1119(95)00511-4

Stemmer, W.P.C., Crameri, A., Ha, K.D., Brennan, T.M., Heyneker, H.L., 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene 164, 49–53. doi:10.1016/0378-1119(95)00511-4

Stephanopoulos, G., 2012. Synthetic Biology and Metabolic Engineering. ACS Synth. Biol. 1, 514–525. doi:10.1021/sb300094q

Sun, H., Zhao, D., Xiong, B., Zhang, C., Bi, C., 2016. Engineering Corynebacterium glutamicum for violacein hyper production. Microb. Cell Fact. 15, 148. doi:10.1186/s12934-016-0545-0

Tatum, E.L., Lederberg, J., 1947. Gene Recombination in the Bacterium Escherichia coli. J. Bacteriol. 53, 673–684. doi:10.1038/158558a0

The Nobel Prize in Physiology or Medicine 1968 [WWW Document], n.d. URL https://www.nobelprize.org/nobel_prizes/medicine/laureates/1968/ (accessed 7.22.18).

Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121. doi:10.1038/nprot.2009.203

Thomason, L.C., Costantino, N., Shaw, D. V., Court, D.L., 2007. Multicopy plasmid modification with phage λ Red recombineering. Plasmid. doi:10.1016/j.plasmid.2007.03.001

Tian, J., Gang, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., Church, G., 2004. Accurate multiplex gene synthesis from programmable DNA microchips. Nature. doi:10.1038/nature03151

Tian, J., Ma, K., Saaem, I., 2009. Advancing high-throughput gene synthesis technology. Mol. Biosyst. 5, 714–722. doi:10.1039/b822268c

Tian, T., Salis, H.M., 2015. A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. Nucleic Acids Res. 43, 7137–7151. doi:10.1093/nar/gkv635

Tsuruta, H., Paddon, C.J., Eng, D., Lenihan, J.R., Horning, T., Anthony, L.C., Regentin, R., Keasling, J.D., Renninger, N.S., Newman, J.D., 2009. High-Level Production of Amorpha-4,11-Diene, a Precursor of the Antimalarial Agent Artemisinin, in Escherichia coli. PLoS One 4, e4489. doi:10.1371/journal.pone.0004489

Varma, A., Boesch, B.W., Palsson, B.O., 1993. Biochemical production capabilities of escherichia coli. Biotechnol. Bioeng. doi:10.1002/bit.260420109

Varma, A., Palsson, B.O., 1993. Metabolic Capabilities of Escherichia coli: I. Synthesis of Biosynthetic Precursors and Cofactors. J. Theor. Biol. 165, 477–502. doi:10.1006/jtbi.1993.1202

Villalobos, A., Ness, J.E., Gustafsson, C., Minshull, J., Govindarajan, S., 2006. Gene Designer: A synthetic biology tool for constructuring artificial DNA segments. BMC Bioinformatics. doi:10.1186/1471-2105-7-285

Von Kamp, A., Klamt, S., 2017. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. Nat. Commun. 8, 1–10. doi:10.1038/ncomms15956

Weiss, B., Richardson, C.C., 1967. Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. Proc. Natl. Acad. Sci. 57, 1021–1028. doi:10.1073/pnas.57.4.1021

Wright, F., 1990. The “effective number of codons” used in a gene. Gene 87, 23–29. doi:10.1016/0378-1119(90)90491-9

Wu, G., Yan, Q., Jones, J.A., Tang, Y.J., Fong, S.S., Koffas, M.A.G., 2016. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. Trends Biotechnol. 34, 652–664. doi:10.1016/j.tibtech.2016.02.010

Yadav, V.G., De Mey, M., Giaw Lim, C., Kumaran Ajikumar, P., Stephanopoulos, G., 2012. The future of metabolic engineering and synthetic biology: Towards a systematic practice. Metab. Eng. 14, 233–241. doi:10.1016/j.ymben.2012.02.001

Yang, T.H., Kim, T.W., Kang, H.O., Lee, S.H., Lee, E.J., Lim, S.C., Oh, S.O., Song, A.J., Park, S.J., Lee, S.Y., 2010. Biosynthesis of polylactic acid and its copolymers using evolved propionate CoA transferase and PHA synthase. Biotechnol. Bioeng. 105, 150–160. doi:10.1002/bit.22547

Yu, D., Ellis, H.M., Lee, E.-C., Jenkins, N.A., Copeland, N.G., Court, D.L., 2000. An efficient recombination system for chromosome engineering in Escherichia coli. Proc. Natl. Acad. Sci. doi:10.1073/pnas.100127597

Zhang, Y., Buchholz, F., Muyrers, J.P.P., Francis Stewart, A., 1998. A new logic for DNA engineering using recombination in Escherichia coli. Nat. Genet. doi:10.1038/2417

Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C.-H., Fu, J., Chen, S., Liu, Y., 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proc. Natl. Acad. Sci. U. S. A. 113, E6117–E6125. doi:10.1073/pnas.1606724113

List of publication

Immanuel, S. R. C., Banerjee, D., **Rajankar, M. P**. & Raghunathan, A. Integrated constraints based analysis of an engineered violacein pathway in Escherichia coli. *Biosystems* **171,** 10–19 (2018).

Mayooreshwar P. Rajankar , Sapna Ravindranathan , P.R Rajamohanan and Anu Raghunathan Absolute Quantitation Of Poly(R)-3-Hydroxybutyric Acid Using Spectrofluorometry In Recombinant E*scherichia Coli*. Oxford Academic Biology Methods and Protocols. Biol. Methods Protoc. 3. doi:10.1093/biomethods/bpy007

Mayooreshwar P. Rajankar, Sapna Ravindranathan, P.R Rajamohanan and Anu Raghunathan Synthetic Biology for Augmenting Pathway for the Biosynthesis of Poly(R)-3-Hydroxybutyric Acid in *E.coli* (Manuscript under prepration).

Agrawal, R., Pandey, A., **Rajankar, M.P**., Dixit, N.M. and Saini, D.K. The Two Component Signaling Networks of *Mycobacterium tuberculosis* Displays Extensive Crosstalk *In Vitro*. (2015) *Biochem. J.* 469 (1), 121 – 134.