INVESTIGATION OF METABOLIC MODELING STRATEGIES TO PREDICT PHENOTYPES OF MICROBIAL MUTANTS UNDER DATA SCARCITY

VERİ YETERSİZLİĞİ ALTINDA MİKROBİYAL MUTANTLARIN FENOTİPLERİNİ TAHMİN ETMEK İÇİN METABOLİK MODELLEME STRATEJİLERİNİN ARAŞTIRILMASI

ECE ANTURAN

ASSOC. PROF. DR. EDA ÇELİK AKDUR

Supervisor

DR. BIRGITTA EBERT

Co-Supervisor

Submitted to

Graduate School of Science and Engineering of Hacettepe University

As a Partial Fulfillment to the Requirements

for the Award of the Degree of Master of Science

in Chemical Engineering

2020

To my family, friends, Aslı Bakal and Mehmet Can Cömert...

ETHICS

In this thesis study, prepared in accordance with the spelling rules of Institute of Graduate School of Science and Engineering of Hacettepe University,

I declare that

- All the information and documents have been obtained in the base of the academic rules
- All audio-visual and written information and results have been presented according to the rules of scientific ethics
- In case of using other works, related studies have been cited in accordance with the scientific standards
- All cited studies have been fully referenced
- I did not do any distortion in the data set
- And any part of this thesis has not been presented as another thesis study at this or any other university.

10/02/2020

Ece Anturan

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

☐ Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.

☐ Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir.

☐ Tezim ile ilgili gizlilik kararı verilmiştir.

10.02.2020

E. Anturan

ECE ANTURAN

ÖZET

VERİ YETERSİZLİĞİ ALTINDA MİKROBİYAL MUTANTLARIN FENOTİPLERİNİ TAHMİN ETMEK İÇİN METABOLİK MODELLEME STRATEJİLERİNİN ARAŞTIRILMASI

Ece Anturan

Yüksek Lisans, Kimya Mühendisliği Bölümü
Tez Danışmanı: Doç. Dr. Eda ÇELİK AKDUR
Eş Danışman: Dr. Birgitta EBERT
Ocak 2020, 133 sayfa

Mikroorganizmaların fenotip tahminleri, bilişim çalışmaları arasında gelişen bir araştırma alanıdır. Bu tahminler için birçok modelleme stratejisi geliştirilmiştir; ancak bu stratejilerin çoğu, mikroorganizmaların doğal türü için tahminlerde bulunmaktadır. Özellikle asetat ve biyokütle verimleri için yüksek doğrulukta bir mutant fenotipi tahmini yönteminin bulunmaması nedeniyle, bu çalışmada, *Escherichia coli* suşuna ait 21 mutantın fenotip tahminleri için protein ayırım kısıtlamalarının genom ölçekli model (GEM) üzerine eklenmesinin etkisi; literatürde bulunan çeşitli modelleme stratejilerini birleştirerek ve aynı zamanda sadece doğal tür için bilgi sağlama veya mutant için minimum bilgi sağlama koşulu gerçekleştirilerek çalışılmıştır.

Bu çalışmada, Akı Dengesi Analizi (FBA), Metabolit Dengesinin Minimize Edilmesi (MiMBI) ve Metabolik Düzenleme Minimizasyon Yöntemi (MOMA) modelleme stratejilerinin, Kısıtlanmış Dağılım (CA) ve Kinetik ve Omik verileri kullanarak Enzimatik Kısıtlamaları olan GEM (GECKO) yöntemleri ile kombinasyonları sunulmuştur. Kombine stratejiler ile optimal akı dağılımını elde

etmek için doğrusal (LP) ve quadratik programlama (QP) çözüm yöntemleri kullanılmıştır. Genel olarak, bu kombine stratejilerden gelen mutant fenotip tahminleri, özellikle çoğalma hızı tahminleri için, önceki çalışmalarda uygulanan diğer stratejilerden daha doğru bir şekilde gerçekleşmiştir. Pearson Korelasyon Katsayılarına (PCC, ρ) göre, mutantların substrat alımı hakkında bilgi sağlanmadığı koşulda (veri yetersizliği altında), en iyi asetat verimi (ρ = 0.53), biyokütle verimi (ρ = 0.71) ve çoğalma hızı (ρ= -0.53), bu çalışmada geliştirilen "GECKO-MiMBI-Reaksiyon" kombine algoritması kullanılarak elde edilmiştir. Bu değerler literatürde benzer senaryolar için mevcut olan PCC değerlerine göre %13-64 daha yüksektir. Özetle, GEM kullanan kombine yöntemler için yeni geliştirilen algoritmalar ile *Escherichia coli* 'nin mutant fenotipleri için daha doğru tahminler elde edilmiştir.

Anahtar Kelimeler: Akı Denge Analizi, CAFBA, MiMBI, MOMA, Fenotip Tahmini, *Escherichia coli*, Gen Nakavtı, Genom Ölçekli Metabolik Model

ABSTRACT

INVESTIGATION OF METABOLIC MODELING STRATEGIES TO PREDICT PHENOTYPES OF MICROBIAL MUTANTS UNDER DATA SCARCITY

Ece Anturan

Master of Science, Department of Chemical Engineering
Supervisor: Assoc. Prof. Dr. Eda ÇELİK AKDUR
Co-Supervisor: Dr. Birgitta EBERT
January 2020, 133 pages

Phenotype predictions of microorganisms are a developing research area in computational studies. For these predictions, many modeling strategies have been developed; however, most of these strategies are only suitable for wild-type predictions. In this study, because of the lack of methods to predict an accurate mutant phenotype, especially for acetate and biomass yields, the influence of adding protein allocation constraints to Genome-scale Models (GEM) for phenotype predictions of 21 mutant strains of *Escherichia coli* were studied by combining various modelling strategies available in the literature. Conditional information for only the wild-type or minimal information for the mutants were supplied for the predictions.

Here, we provide combinations of Flux Balance Analysis (FBA), Minimization of Metabolites Balance (MiMBl) and Metabolic Adjustment Minimization Method (MOMA) modeling strategies with Constrained Allocation (CA) and GEM with Enzymatic Constraints using Kinetic and Omics data (GECKO) methods. In order to solve for the optimal flux distribution using these combined strategies, linear (LP) and quadratic programming (QP) were used. Generally, with the combined

strategies, the mutant phenotype was predicted with more accuracy compared to using the other strategies in the previous studies, especially for growth rate predictions. According to Pearson Correlation Coefficients (PCC, ρ), when no information about nutrient uptake of mutants were supplied (i.e. under data scarcity), the best acetate yield (ρ=0.53), biomass yield (ρ=0.71) and growth rate (ρ=-0.53) predictions of the experimental data were obtained, using the "GECKO-MiMBI-Reaction" combined algorithm developed in this study. These values were 13 - 64% higher, compared to the PCC values present in the literature for similar scenarios. Overall, with the newly developed algorithms for combined methods using GEM, more accurate predictions for mutant phenotypes of *Escherichia coli* have been obtained.

Keywords: Flux Balance Analysis, CAFBA, MiMBI, MOMA, Phenotype Prediction, *Escherichia coli*, Gene Knockout, Genom Scale Metabolic Model

ACKNOWLEDGMENTS

First of all, I would like to thank both of my thesis advisors, Assoc. Prof. Dr. Eda ÇELİK AKDUR and Dr. Birgitta EBERT, for their support, understanding, knowledge and guidance throughout this research.

Sincere thanks to Prof. Dr. Lars BLANK for accepting me in the EBERT group at Institute of Applied Microbiology (iAMB), RWTH Aachen University, as a researcher and influenced me to develop the thesis during my ERASMUS period, also supporting our team throughout this project. Especially, Tobias ALTER in the EBERT Group was always helpful in my project, teaching me the modelling strategies and combining them with protein limitation methods and even correcting my thesis draft. Also I would like to thank the whole iAMB family for their support and kindness.

I am thankfull to have a side project and financial support provided by TÜBİTAK-1001 Grant # 118Z798. In addition, I am thankfull to Hacettepe University EU Office for the Erasmus exchange program opportunity.

I would like to thank members of thesis examining committee for their time, invaluable suggestions and comments which make the final version of the thesis better.

I am really indebted to my labmates in ÇELİK LAB for sharing their motivation. Being part of this team was pleasure to me.

I am thankful to Mehmet Can CÖMERT for his knowledge, support and understanding especially in hard times of this thesis. I am sure that I can't find any appropriate words to express my gratitude to my family and friends for their never ending support, and caring even as I followed my ambitions and dreams. They are only with me to encourage, cheer up and love. When I have to cope with difficulties, talking to them and even thinking of them are enough for me. I hope I make them proud at the end. In addition, I am thankful to Umut BAYKARA that he supplied support and advices in the final period this master thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

<w>	The distribution's mean value
CA	Constrained Allocation
Enz	Enzyme
FBA	Flux Balance Analysis
GECKO	GEM with Enzymatic Constraints using Kinetic and Omics data
GECKO-MiMBI- Enz	GECKO-MiMBI strategy was applied by excluding reactions except enzyme reactions
GECKO-MiMBI-Rxn	GECKO-MiMBI strategy was applied by excluding enzyme and exchange reactions
GEM	Genome-scale Model
Hete	Heterogenous Case
Homo	Homogenous Case
LP	Linear Programming
ME	Macromolecular Expression
MiMBI	Minimization of Metabolites Balance
MOMA	Metabolic Adjustment Minimization Method
PCC, ρ	Pearson Correlation Coefficient
QP	Quadratic Programming
Rxn	Reaction
w_C	Weight Factor of C-sector
w_E	Weight Factor of E-sector
w_R	Weight Factor of R-sector
Φ_{max}	The proteome fraction that can reach the protein sector components dependent on the specific growth rate

1. INTRODUCTION

Systems biology is a rapidly developing area [1, 2] based on the production and verification of *in silico* models for biological systems by utilizing experimental data. These models can be used to produce new, testable and generally quantitative predictions of cellular behavior [2, 3]. Constraint-based modeling, which allows calculation of metabolic fluxes from intracellular metabolites' mass balances and reaction stoichiometry, has widely been utilized to achieve this aim for the last decade [2, 4, 5].

Genome-wide types of constraint-based models are called genome-scale models (GEMs) and they have been extensively used to predict phenotypes of the knockout mutants, especially production rates of some metabolite or the biomass yield. These models consist of information such as reactions, genes, metabolites, and reaction stoichiometry of a microorganism [2, 6, 7].

One of the latest constraints is proteome allocation. The proteome allocation is based upon the 'supply and demand' analogy. In this analogy, cells handle challenges to growth-limiting perturbations, for instance, nutrient limitation, by raising the number of enzymes allocated to overcoming the limited process. A cell has only a limited amount of protein, and the cell has to distribute this protein amount from different proteome sectors according to different metabolic activities in the cell [8].

Analysis methods, which use the metabolic models, have been improved to be able to make accurate phenotypic predictions. The Flux Balance Analysis (FBA) method calculates the flux of metabolites [9], while Constrained Allocation Flux Balance Analysis (CAFBA) combines conventional FBA with proteome allocation constraints. Its aim is to make more accurate predictions of acetate excretion rate and biomass yield [10]. The GEM with Enzymatic Constraints using Kinetic and Omics data (GECKO) method improves a GEM by explicitly describing enzymes

as part of reactions. By this way, each metabolic flux is constrained to be at or below the maximal catalytic capacity of the corresponding enzyme, which equals the product of a turnover number and the enzyme's abundance or molar concentration [2]. On the other hand, Minimization of Metabolites Balance (MiMBI) minimizes metabolite balance between wild-type and mutant [11], and Metabolic Adjustment Minimization Method (MOMA) does a minimum re-distribution of metabolic fluxes of the knockout relative to the flux distribution of the wild-type [12]. There are several methods that claim to predict mutant strain behaviors, but so far none of them has proven to make sufficiently accurate predictions. Also, in the literature, there are studies predicting the mutant phenotype of *Escherichia coli*; however, in these studies, protein limitation constraints were not applied.

In this study, investigation of the influence of adding protein allocation constraints to GEM on phenotype predictions of *Escherichia coli* mutant strains was studied by providing information for only the wild-type. In addition, several combinations of some of the aforementioned methods were investigated to achieve more accurate predictions, especially for biomass and acetate yields of mutants.

2. LITERATURE SURVEY

2.1. Genome-Scale Metabolic Models

Over the past 10 years, there has been steady progress in improving genome-scale models of metabolism (Metabolic Model, M-Models) for fundamental research and industrial implementations [13-16].

M-Models are stoichiometric representations of the enzymatic and biochemical reactions related to a microorganism's genome-scale metabolic network; however, M-Models do not explicitly define gene expression [16]. In recent years, M-Models have been enlarged to contain the process of gene expression to obtain Macromolecular Expression (ME) models [16, 17], opening up a completely new perspective in the improvement of microbial systems biology [18]. ME-Models supply a chemically and genetically consistent definition of an organism, therefore, they start to bridge the gap between molecular biology and cellular physiology [16].

Genome-scale metabolic models (GEMs) and Genome-scale network reconstructions, helps for (i) explaining bio-molecular functions [19, 20], (ii) predicting phenotypic behavior [20, 21], (iii) finding new biological knowledge [19, 20, 22], and (iv) designing experiments to apply them in engineering applications [20, 23, 24].

Firstly, a component list is created from the genome annotation and information taken from different databases. After that, the prepared component list is manually curated with the help of a fundamental review of previous studies. By this way, mass and charge balanced reactions are obtained. This is combined into pathways. The stoichiometric matrix represents the network. In this matrix, reactions are represented in columns and metabolites are represented in rows. The stoichiometric coefficients are the elements of the matrix. These values are defined for each metabolite in each reaction (reactants are shown as negative values, and products are shown as positive values). The collection of these reactions, represented according to this definition is called the stoichiometric matrix, *S* (Fig. 2.2). At this step condition, special constraints such as measured

uptake and secretion rates, or known regulatory constraints are going to be applied to external and internal reactions. By this way, the resulting set becomes a distinct, condition-specific set. In the case of a distinct set of constraints is used for the same reconstruction, distinct models will be obtained [20].

As the third step, the debugging of the reconstructions is executed by computationally testing the capabilities and properties of the model to guarantee that the derived reconstructions and real organisms have similar capabilities. Gaps in the pathway can be resolved by the help of provided experimental work. In the last step, analysis and simulation are done to reconcile the reconstruction model with experimental data. This can lead to further refinements of the model (Fig. 2.1) [20].

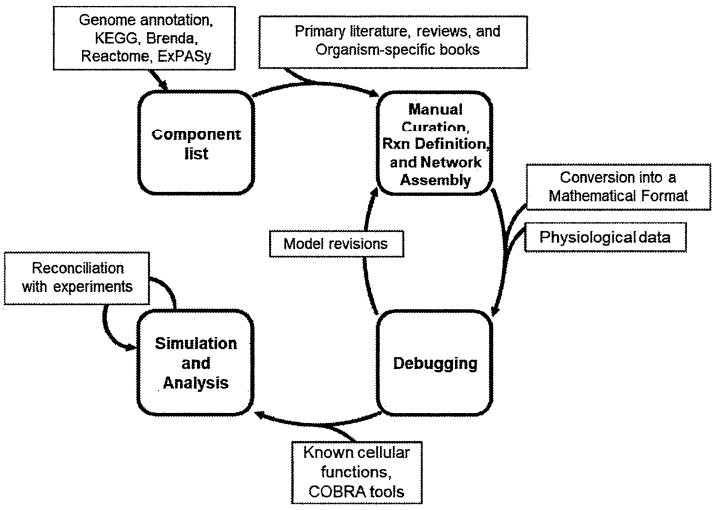

Figure 2.1 The metabolic network reconstruction process summary [20].

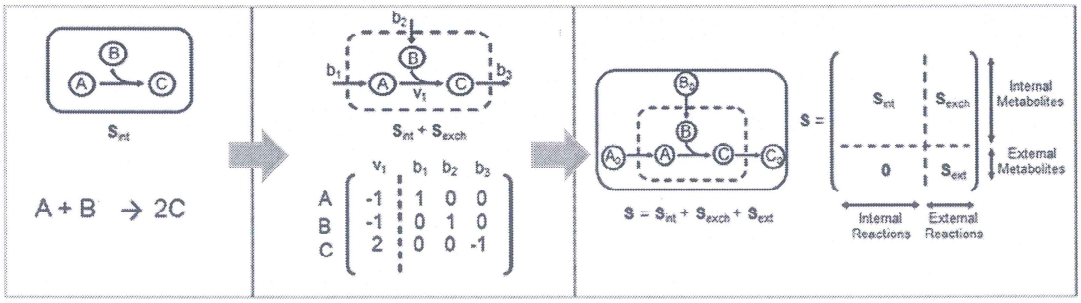

Figure 2.2 Metabolic network reconstruction. Figure modified from [20].

A GEM has been used to apply *in silico* gene knockout studies [25, 26], it was also used to predict both growth behavior of the cell under different environmental conditions [26, 27] and the results of adaptive evolution [26, 28].

Metabolism's genome-scale network reconstructions are freely available from the biochemical, genetic, and genomic (BiGG) knowledge-base [6]. The BiGG database contains information on the biochemical, genomic and proteomic components which includes relationships between each component in a specific microorganism and cell [20].

Genome-scale models are constrained via (i) the stoichiometry of a metabolic network; (ii) pre-set upper and lower boundaries for chosen reactions (usually substrate uptake reactions, or can also describe thermodynamics); (iii) a steady-state assumption is required. The assumption accepts that all metabolite concentrations are constant over time. This status would be achieved in chemostat cultivation experiments. Theoretically, throughout the controlled batch, the logarithmic growth phase and the regulation of growth will not be affected by the change of substrate concentration or product formation [7].

2.1.1. *Escherichia coli* GEMs

The genome-scale reconstruction of *E. coli* defines a milestone in systems biology. It will make a possible quantitative combination of '-omics' datasets. Also with the help of this milestone, studies about the mechanic principles underlying the genotype-phenotype relationship become achievable [29].

Genome-scale reconstructions are very important in quantifying genotype-phenotype relations. Applications of implementations of the genome-scale model (GEM) of *E. coli* are separated into the following five categories [15]:

1. A predicted aspect from a loss-of-function mutation in a basic system is utilized. *In silico* strain design and metabolic engineering studies have been investigated by using *E. coli* metabolic reconstructions to over-produce desired products [15, 23, 24, 30-38].
2. A metabolic network along with biochemical and genetic knowledge is used to drive biological invention, for example predicting genes responsible for orphan reactions [15, 19, 39-42].
3. Computational analyses have investigated a gene [15, 43-47], metabolite [15, 48, 49] and reaction [15, 50-53] essentiality along with considering thermodynamics [15, 44, 52, 54-60] by using the reconstruction in phenotypic studies. These studies try to make better predictions about the physiological status of a cell for a given static environmental condition [15].
4. The *E. coli* reconstructions are used to analyze the actual properties of biological networks. For example finding coupled reaction activities across distinct growth conditions [15, 61].
5. Evolutionary studies investigate the cellular network [15, 62] in terms of adaptive evolution cases [15, 62] or horizontal gene transfer [15, 62, 63]. There are studies with minimal metabolic network evolution by using network reconstruction [15, 64].

iML1515 is the newest genome-scale metabolic model of *E. coli*, which allows an analysis of several data types [65]. These data types include proteomes, transcriptomes, and metabolomes, contain 1,515 open reading frames and 2,719 metabolic reactions containing unique metabolites. The iML1515 information base is connected to 1,515 protein structures for supplying an integrated modeling framework, which combines systems and structural biology [65].

2.2. Genetic Knockout

Studying the cellular system after a genetic knockout is a very effective method for getting new information on network structure. It is also possible to gain

information about the regulations and dynamics of the cell. *Escherichia coli* is easy to handle in the laboratory studies and it has established tools for its strains's genetic manipulation. *Escherichia coli* is an ideal model prokaryotic organism and suited well for gene knockout studies. A number of studies have been performed on *E. coli* knockouts [66]. Libraries of numerous viable *E. coli* single-gene knockouts, such as the Keio collection, are now simplifying these studies [66, 67]. Relevant data for metabolic engineering and biotechnological applications in addition to having significant value in biological sciences are provided by quantitative studies related to gene knockouts [66].

Knockouts created by modifying the central carbon metabolism enzymes in *E. coli* causes significant physiological and metabolic changes, which can be also called as metabolic rewiring [68]. Large differences in substrate uptake rates, biomass yield, acetate yield, the specific growth rate, and changes in biomass composition can be observed. The external metabolic fluxes of biomass generation, glucose uptake, and acetate generation provide a sign for the regulatory adjustments and kinetic limitations occurring in these knockouts [68].

2.3. Metabolic Modelling Strategies

The development of computational methods utilizing metabolic models has been done to predict and interpret a microbe's performance under different ranges of growth conditions [21]. The prediction and interpretation of metabolic flux distributions require computer simulation and mathematical modeling. There exists a prolonged history of quantitative metabolic modeling [21, 69]. A few well-developed mathematical approximations are available to analyze cellular metabolism and its regulation [21, 70-78]. The bigger part of these methods needs comprehensive kinetic and concentration information about enzymes and varied cofactors. Cellular components information is fast developing, however, the implementation of a lot of methods of mathematical modeling is set back by an absence of the needed kinetic information and enzyme concentration data [21, 79-81]. Several methods (Table 2.1) have been developed for the analysis of metabolic networks via constraint-based reconstructions using empirical data to answer biological questions [20].

Table 2.1 Methods for analyzing metabolic networks. The table has been taken from Lewis et al. 2008 [20].

Method	Basis	Example
Alternate Optima	Determining the number of flux status that can be reached by maximization or minimization of an objective function.	[82-86]
Extreme pathways (ExPa)/Elementary Modes (EIMo)	Describing a biochemically possible, unique set of reactions that span the steady-state solution space.	[87-89]
Flux Balance Analysis (FBA)	Determining the maximum or minimum value of a cellular objective function.	[90-93]
Flux Confidence Interval	Determining the reliability intervals for fluxes when fluxomic data is used with the constraint-based model.	[94]
Flux Coupling	Determining the sets of network reactions that are completely, partly or directionally coupled.	[61]
Flux Variability Analysis	Determining the maximum and minimum flux (i.e., determining the limiting box of the solution space) for each reaction under a certain set of constraints.	[84]
Minimization of Metabolites Balance (MiMBl)	Separating the artifacts of stoichiometry representment from the requested objective functions formulation by casting objective functions utilizing metabolite turnovers instead of fluxes.	[11]
Minimization of Metabolic Adjustment (MOMA)	Detection of the possibility for suboptimal growth predictions to be more consistent with experimental data in wild-type and knock-out strains.	[12]
Objective function finder/ ObjFind	Determining the distinct probable cellular objectives.	[95-97]
OptKnock / OptGene	Designing a knock-out strain optimized for the secretion of a byproduct coupled to cellular growth.	[38, 98]
OptReg	Determining the optimum reaction activations/inhibitions and eliminations to enhance biochemical production.	[33]
OptStrain	Detection of reactions (not genome encoded) needs to be added for enabling a strain to produce a foreign compound.	[36]
Regulatory Flux Balance Analysis	Detection of how transcriptional regulatory rules influence the range of probable *in silico* phenotypes.	[22]
Regulatory On/Off Minimization (ROOM)	Determining the most probable flux distribution that requires a minimal change in transcriptional regulation following a gene knockout.	[99]

Robustness analysis	Determining how an objective function changes as a function of another network flux.	[100]
Steady-state Regulatory Flux Balance Analysis	Determining the levels of metabolic and transcriptional regulatory constraints that will influence the metabolic behavior.	[101]
Stable Isotope Tracers	Determining how intracellular flux predictions can be experimentally validated, and which pathways are active under distinct conditions.	[102]
Thermodynamics-based Metabolic Flux Analysis	Determining how to use thermodynamic data to generate thermodynamically probable flux profiles.	[56]
Uniform Random Sampling	Determining the distributions of network states that have not been removed based on physicochemical constraints and/or experimental measurements and determining the fully or partially related reaction sets.	[103-106]

2.3.1. Flux Balance Analysis

The most widely implemented metabolic modeling framework is called Flux Balance Analysis (FBA). This method predicts flux over a metabolic network quantitatively, which will maximize a given criterion thought to resemble the evolved microbial nature [107, 108]. One of the most important elements for FBA method is the objective function. FBA predicts optimal flux distribution depending on this function.

The objective function has been generally described as maximal cellular growth by the studies involving microbial cells. Even though such optimal solutions can be computed, these flux distributions may predict the actual cell behavior incorrectly. If the actual and predicted metabolic fluxes are similar, the data will promote the hypothesis [21].

In the conventional FBA method, metabolic reactions are indicated as a stoichiometric matrix (**S**) in the range of **m** x **n** [9]. Each row of this matrix indicates a unique metabolite (for **m** metabolite system), and each column indicates a reaction (**n** reaction). Stoichiometric coefficients of the metabolites which take part in the reaction are put as inputs in each column. There is a negative and positive coefficient for every consumed and produced metabolite, respectively. For each metabolite that does not take part in a given reaction, zero is used as

stoichiometric coefficient. Throughout all the reactions in a network, the flux is represented by a vector **v** of length n. Metabolite accumulation vector is represented by **x**, with a length of m [9]. The mass flux balance equations are shown as linear vectors of differential equations as follows [109]:

$$\mathbf{S}^* \mathbf{v}(t) = \mathbf{x}(t) \tag{2.1}$$

x(t) in Equation 2.1 can be divided into two sub-vectors:

$$\mathbf{x}(t) = \mathbf{x}_1(t) + \mathbf{x}_2(t) \tag{2.2}$$

where, $\mathbf{x}_1(t)$ & $\mathbf{x}_2(t)$ are accumulation vectors of extracellular and intracellular metabolites, respectively.

The concentration of different metabolite pools is rapidly adjusted to new levels, even after large deviations in the fermentation broth, due to the high turnover rate of most metabolite pools. Thus, the pseudo-steady state approach can be used for intracellular metabolites, $\mathbf{x}_2(t) = 0$ is acceptable [109].

As shown in Equation 2.3, the system is assumed to be in a steady position.

$$dX_i/dt = 0 \tag{2.3}$$

$$\mathbf{S}^*\mathbf{v}(t) = 0 \tag{2.4}$$

There are infinite solutions to Equation 2.4. The number of reactions is higher than the number of compounds (n>m) in any realistic large-scale metabolic model. That is, there are more unknown variables than equations , therefore no unique solution can be found for this equation system. A biological network's flux distribution can be found at any position in a solution space with no constraints. The constraints of mass balance implemented by the stoichiometric matrix **S** and the capacity constraints implemented by the lower and upper limits describe an allowable solution space. Any flux distribution can be computed by FBA within

this space, but the points outside of the space are not accepted by the constraints [9].

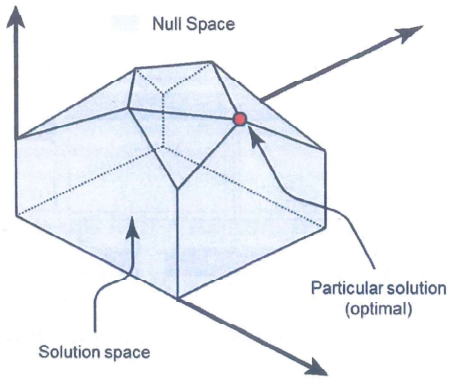

Figure 2.3 A schematic representation of solution space and optimal solution [4, 133].

The FBA can find an optimal flux distribution at the edge of the allowed solution space by optimizing an objective function [9]. FBA method is used for identifying an optimal point inside of the constrained space [9]

$$Z = \mathbf{c}^T\mathbf{v} \tag{2.5}$$

where,
Z: a linear combination of fluxes
c: reaction weight vector
v: flux vector

FBA aims to optimize an objective function as shown in Equation 2.5, which may be any combination of linear functions. Here, '**c**' indicates a weight vector displaying how much each reaction (for example, biomass formation reaction) contributes to the objective function [9].

Optimizing this kind of system is performed by linear programming. Therefore, FBA can be described as a linear program for solving Equation 2.4.

Application of FBA is the predicting of phenotypic impacts deriving from partial or complete metabolic gene knockouts [12, 25, 100, 110]. A gene knockout is applied by equating the corresponding flux value to zero. Then for the new genotype, the flux distribution and maximum growth yield are supplied by linear programming [12].

2.3.2. Metabolic Adjustment Minimization Method

Mutant flux distributions can be computed by method of minimization of metabolic adjustment (MOMA). MOMA supplies a mathematically observable approach for this intermediate suboptimal status, based on the presumption that the mutant flux distribution remains initially as close as possible to the wild-type optimum flux distribution (Fig. 2.4) [12, 111, 112].

In summary, MOMA is defined by the following optimization problem [4]:

$$\min \quad ||v-w||^2 \quad \text{(i) } S.v = 0 \tag{2.6}$$

$$\text{(ii) } v_{j,min} \leq v_j \leq v_{j,max} \tag{2.7}$$

$$\text{(iii) Gene knockout : } v_j = 0 \quad j \in A \tag{2.8}$$

where,

v: flux vector

w: wild-type flux distribution

S: stoichiometric matrix

A: a series of reactions associated with deleted genes

MOMA minimizes distance between a given reference wild-type and a gene deletion mutant flux distribution. Since distance in flux space is a quadratic function, for solving this problem, quadratic programming (QP) is used [12, 113]. MOMA and FBA use the same stoichiometric constraints, but MOMA relaxes the assumption of optimal growth flux used for gene deletions, therefore it can be used to better figure out the flux status of mutants [12]. Furthermore, for the analysis of artificially engineered bacteria, suggesting MOMA as a computational tool can help to figure out how a cell fits the loss of a gene by regulation and evolutionary optimization [12].

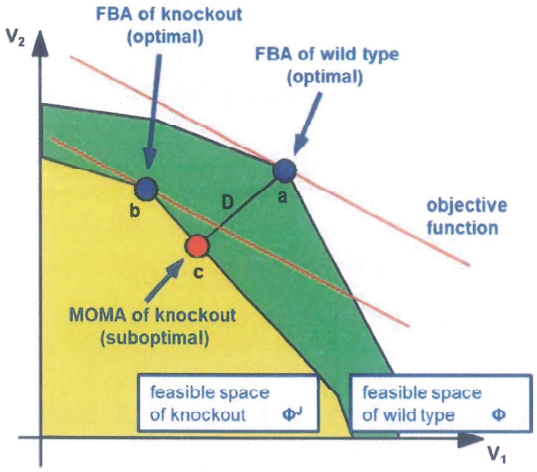

Figure 2.4 Schematic representation of the Metabolic Adjustment Minimization method [4, 12].

2.3.3. Minimization of Metabolites Balance

Minimization of Metabolites Balance (MiMBI) is a method which works like MOMA, but minimizes the differences in the metabolite balances instead of fluxes. A metabolite balance is the sum of absolute fluxes producing or consuming a metabolite. These metabolite balances are set up for each metabolite of the reconstruction. MiMBI takes a reference flux distribution as an input, computes all metabolite balances from it, and then minimizes the distance to this metabolite balance distribution considering the perturbed metabolic network (constraints). MiMBI separates the artifacts of stoichiometry representment from the requested objective functions formulation by casting objective functions utilizing metabolite turnovers instead of fluxes [11, 112] .

MiMBI provides an approximate solution for a suboptimal growth flux condition for mutant by making a metabolite balance between wild-type and mutant (Fig. 2.5).

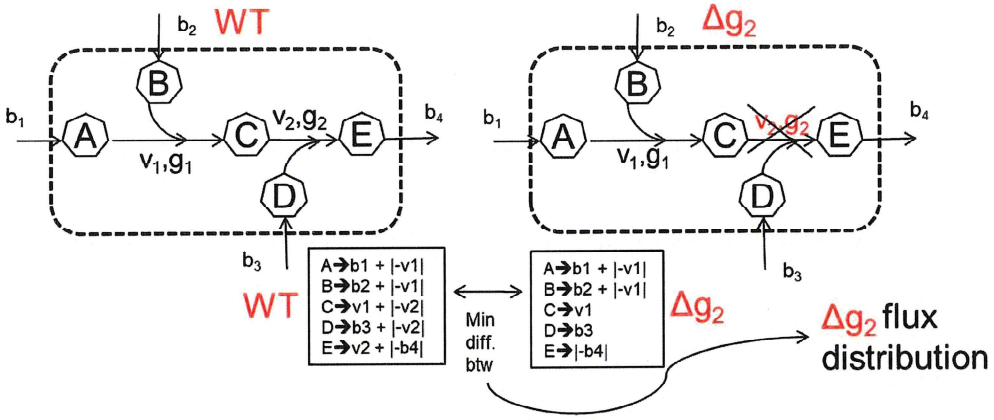

Figure 2.5 Schematic representation of Minimization of Metabolites Balance method

MiMBI explains reaction stoichiometry in the objective function by matching the flux space into the metabolite balance space. By this way, it makes metabolite-usage as a determinant for the prediction of metabolic phenotypes. MiMBI displays robust predictions independent from the stoichiometry representment. In MiMBI, the objective function is formulated as a linear combination of metabolite balances (t_m). The metabolite balance is explained as the summation of all fluxes producing (or consuming) it, multiplied by the corresponding stoichiometric coefficients. The objective function for minimizing metabolic adjustment is formulated again in order to contain metabolite turnovers instead of fluxes [11].

$$t_m = \sum_{i \in N_m} \alpha_{m,i} v_i \quad \forall m \in M, N_m \subset N \tag{2.9}$$

where,

N_m: the subset of N producing or consuming metabolite m

$\alpha_{m,i}$: the stoichiometric coefficient of metabolite m in reaction i (α_m, i is all the time a positive number in this description, regardless of m being a substrate or a product)

MiMBI consists of two optimizations. The first optimization minimizes the summation of metabolite turnovers differences between wild-type and knockout.

Then, the second optimization minimizes summation of flux difference between wild-type and knockout [11].

First optimization [11]:

$$\min \sum_{m \in M} |t_m^{WT} - t_m| \tag{2.10}$$

s.t

$$S.v = 0 \tag{2.11}$$

$$t_m = \sum_{i \in N_m} \alpha_{m,i} v_i \quad \forall i \in N_m, N_m \subset N : S_{m,i} \neq 0 \tag{2.12}$$

$$\alpha_{m,i} \equiv |S_{m,i}| \qquad \forall i \in N, m \in M \tag{2.13}$$

$$v_i^{lb} \leq v_i \leq v_i^{ub} \qquad \forall i \in N \tag{2.14}$$

$$v_i \geq 0 \qquad \forall i \in N \tag{2.15}$$

Second optimization [11]:

$$\min \sum_i \frac{1}{|v_i^{WT}|} |v_i^{WT} - v_i| \quad \forall i \in N : v_i^{WT} \neq 0 \tag{2.16}$$

$$s.t \qquad S.v = 0 \tag{2.17}$$

$$v_i^{lb} \leq v_i \leq v_i^{ub} \qquad \forall i \in N \tag{2.18}$$

$$\sum_{m \in M}^{\cdot} |t_m^{WT} - t_m| = \min \sum_{m \in M} |t_m^{WT} - t_m| \tag{2.19}$$

where,

N: the set of all matrix

v_i: the flux for reactions

M: the set of all intracellular metabolites

S: the stoichiometric reaction i

WT: stands for wild-type (or reference)

v_i^{lb} and v_i^{ub}: the lower and upper bounds for the flux of reaction i

t_m: metabolite balance

$\alpha_{m,i}$: the stoichiometric coefficient of metabolite m in reaction i (αm, i is all the time a positive number in this description, regardless of m being a substrate or a product)

2.3.4. Proteome Allocation

The proteome can be separated into a few coarse-grained sectors. The growth-variant components of proteome fractions comprise approximately 50% of the proteome in terms of mass. The improved predictive capability of the model arise from strict coordination between proteome allocation and metabolism, proposing rules for resource allocation in the proteome economy of the cell. This strategy of global gene regulation shall benefit as a base for future studies on gene expression and establishing synthetic biological circuits [114]. A widely kept perspective, supported by a many of observations, is the idea that cells handle challenges to growth-limiting perturbations, for instance, nutrient limitation, by raising the number of enzymes allocated to overcoming the limited process in analogy with 'supply and demand' (Fig. 2.6) [8, 114-117].

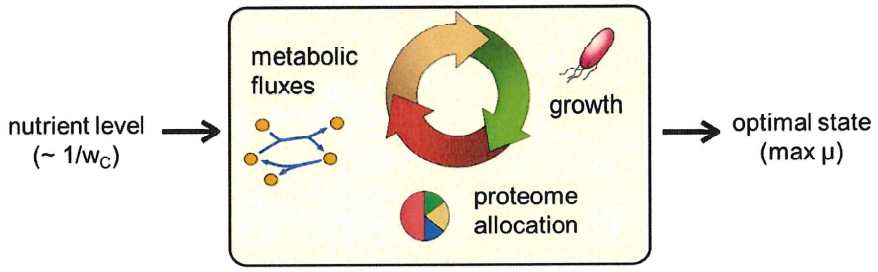

Figure 2.6 Schematic view of the proteome allocation [136].

Proteomic resource allocation models can quantitatively take into account all of the observed behaviors, and accurately predict responses to new perturbations [118]. External conditions and intracellular demands can coordinate protein production in a living cell. This regulation guarantees that required proteins are generated for growth-limiting processes [119].

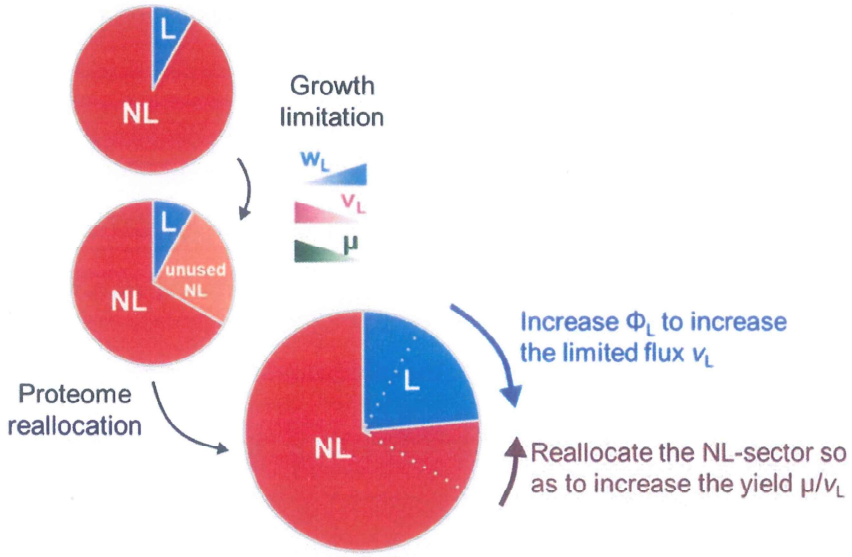

Figure 2.7 Effect of growth limitation on proteome composition. Figure modified from [136].

Protein production is important for the cell physiology and is closely related with cell division rate. Initiatives to model cellular physiology are restricted by the lack of empirical data describing the molecular processes constraining protein expression [119, 120].

The ME-models incorporate the whole translation as well as transcription process, and by constraining the model to accurately allocate proteome to large sectors, the growth rate and metabolism of the wild-type *E. coli* can be predicted better [121]. Due to the total proteome is constrained in size, resource allocation towards one sector may decrease protein availability for other sectors. For this reason, the sector constraints display the costs of cellular determinations to over-allocate proteomic resources for aims other than maximum growth on a minimum medium [121]. One of the causes for use of ME-models in prior studies has been to gain insight into macromolecular expression processes of *E. coli* under various conditions [121, 122].

Even though their big predictive power has already been proven, for studying whole biological processes related to metabolism, using resource allocation models is not possible. For example, such models every time suppose optimality

of growth. For starvation, this may not be a correct assumption when the objective varies more against maximization survival rather than growth. In addition, starvation starts distinct kinetics of uptaking of nutrients, where substrate affinity has a role. Therefore, it is not anymore sufficient to take into account just upper limits on uptake rates based upon turnover rate and enzyme amount, like in models of resource allocation [123].

2.3.5. Overflow Metabolism

Fast-growing microorganisms attempt to refrain using high-yield respiratory pathways to produce ATP, even if oxygen is present. These microorganisms partly switch to fermentation instead of protein-costly respiration under restrictions in the protein availability [10, 124-129]. The choice for low-yield pathways becomes apparent by the presence of fermentation products such as acetate for *E. coli* or ethanol for *S. cerevisiae* [10, 118, 126, 127, 129, 130], which is called overflow metabolism. This behavior is predicted by standard FBA schemes at a qualitative level only when added capacity constraints on respiratory pathways [10, 131] or density constraints for soluble [10, 132, 133] or membrane-bound [10, 134] enzymes are incorporated [10].

2.3.6. Constrained Allocation Flux Balance Analysis

Constrained Allocation Flux Balance Analysis, CA-FBA, is a computational genome-scale framework that combines proteome allocation constraints into traditional FBA [10, 135]. CA-FBA supplies an elementary, numerical approximation to balancing the exchange between growth and proteome allocation constraints. The central focus area of CA-FBA is to investigate cellular strategies for energy production. By effectively modeling the trade-off between growth and its biosynthetic costs, CA-FBA naturally produces cellular states with suboptimal growth yields, where carbon overflow is obtained with quantitative accuracy [10].

Constraint-based modeling's active area is a combination of omics data to improve or make context-specific models. A general formalism for integrating omics data from any empirical condition into constraint-based metabolism and

ME-models is indicated by the sector constraints. The constraints can exist as fine-grained (individual proteins) or coarse-grained (protein groups with a functional relationship). An approximation for narrowing the gap between governing principles of proteome allocation defined by systems-level models and the complexity captured by omics data is supplied by this changeable formalism [121, 136].

CA-FBA applies proteome allocation constraints on the fluxes. These constraints define competition in the distribution of cellular resources. CA-FBA based on the growth of *E. coli* in carbon-bound media constitutes a 4-sector segmentation of the proteome.

These sectors are:

- ribosome-bound proteins (R-sector);
- biosynthetic enzymes (E-sector);
- proteins dedicated to carbon uptake and transport (C-sector);
- expression levels are core cleaning proteins (Q-sector) that are independent of specific growth rate (Fig. 2.8) [10].

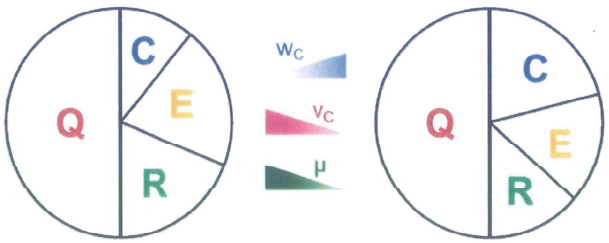

Figure 2.8 Schematic representation of proteome sectors in a cell [136]

$$\phi_C + \phi_R + \phi_E + \phi_Q = 1 \qquad (2.20)$$

Corresponding proteome fractions (indicated by ϕX for the X-sector) should add up to 1 as shown in Equation 2.20. Proteome fractions of the sectors are calculated as shown in Equations 2.21, 2.22 and 2.23 [10].

$$\phi_R = \phi_{R,0} + w_R\lambda \quad (2.21)$$

where,

λ: specific growth rate (h^{-1})

w_R: weight factor, specific growth rate proteome fraction separated into ribosomal proteins per unit

$\phi_{R,0}$: a constant bound to the strain indicating the estimated ribosomal proteome fraction when specific growth rate equals to zero

$$\phi_C = \phi_{C,0} + w_C v_C \quad (2.22)$$

where,

v_C: carbon uptake flux (mmol/g_{DW}h)

w_C: weight factor, proteome fraction separated into units per carbon flow to the C-sector

$\phi_{C,0}$: offset independent of specific growth rate

$$\phi_E = \phi_{E,0} + \Sigma_i w_i |v_i| \quad (2.23)$$

$$\phi_{E,0} = \Sigma_i \phi_{i,0} \quad (2.24)$$

where,

v_i: uptake flux of the i reaction (mmol/g_{DW}h)

w_i: weight factor, the proteome fraction to be deposited in enzyme E_i in the unit flow of i reaction

$\Phi_{i,0}$: offset

In CA-FBA, the E-sector represents the proteome sector for biosynthetic enzymes, and also the proteome fraction of this sector is represented via the weight w_i to be invested in enzyme E_i per unit flux of reaction i. CA-FBA strategy contains two different cases for the simulation of the strategy. In the homogenous case, w_i values for all E-sector related reactions were set to the same value w_E. In the heterogenous case, 1000 sets of randomly chosen w_i were generated, to

represent one's knowledge deficiency of their specific value. More obviously, w_i is gained from a determined probability distribution without any dependence between reactions. The distribution's mean value <w> is remarkably close to w_E, which is common weight for E-sector related reactions in the homogenous case, for this, it sets the average specific growth rate (μ) achieved for $w_C = 0$ to λ_{max}. 1000 CA-FBA simulations with 1000 sets of randomly chosen w_i were done for w_C value and average flux distribution (average of 1000 simulations for each flux) was the output of this CA-FBA heterogenous case simulation [10].

$$w_R \lambda + w_C v_C + \Sigma_i w_i |v_i| = \Phi_{max} \quad (2.25)$$

$$\Phi_{max} = 1 - \phi_Q - \phi_{R,0} - \phi_{C,0} - \phi_{E,0} \quad (2.26)$$

Here, Φ_{max} refers to the proteome fraction that can reach the protein sector components depending on the specific growth rate and is estimated to be approximately 50% for *E. coli* [10].

In summary, CA-FBA is defined by the following optimization problem. When a metabolic network encoded by a stoichiometric matrix (S = (S)) is given, CA-FBA, in the case of carbon limitation, is indicated as in Equations 2.27, 2.28 and 2.29 [10].

$$\max_v \lambda \text{ subject to } \quad \text{(i)} \ \Sigma_i S_{\mu i} v_i = 0 \ \forall \mu \quad (2.27)$$

$$\text{(ii)} \ l_i \leq v_i \leq u_i \ \forall i \quad (2.28)$$

$$\text{(iii)} \ w_R \lambda + w_C v_C + \Sigma_i w_i |v_i| = \Phi_{max} \quad (2.29)$$

Where λ; specific growth rate, v = (v_i); the flux vector and (l_i, u_i) indicate the lower and upper limits for each vi flux, respectively, the elements of the stoichiometric matrix of the metabolic network (With μ; indexing metabolites and i; indexing induction reactions) [10].

2.3.7. GEM with Enzymatic Constraints using Kinetic and Omics data

The GEMs are widely used to predict the production rate of a metabolite and specific growth rate by assuming that the carbon uptake rate applies some

constraints to production. However, rates are also constrained by enzyme abundances. Since enzymes are not explicitly considered in classical GEMs, there is attention in improving novel modeling notions that are going to make possible the connection of enzyme levels and flux rates. Proteomics data have until now commonly been associated with GEMs indirectly by being connected with protein levels to corresponding fluxes [2, 137, 138]. GECKO is a method which improves a GEM to take into explaining enzymes as a component of reactions, by this way providing that every metabolic reaction operates within its enzyme's catalytic capacity, equivalent to the product of the enzyme's abundance and turnover number (Fig. 2.9) [2].

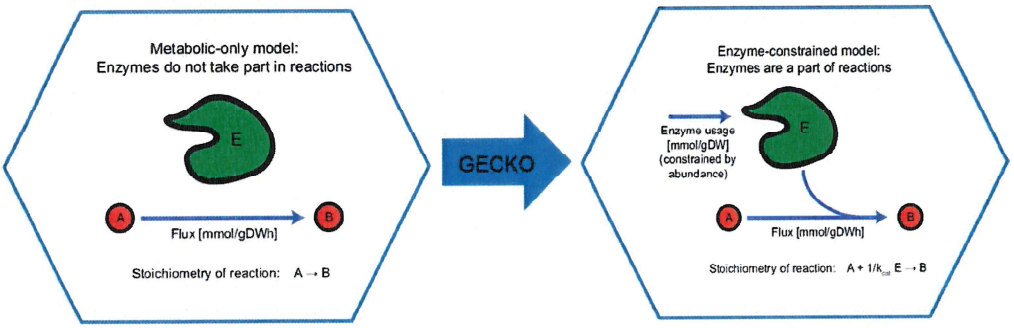

Figure 2.9 The difference between GECKO and traditional GEMs. Figure modified from [2].

GECKO expands the stoichiometric matrix by adding new columns representing the enzymes and new columns representing the use of each enzyme in GECKO [2]. Kinetics info, in terms of kcat values, is added as stoichiometric coefficients to turn into the metabolic flux in mmol/g_{DW}h and into the enzyme use needed in mmol/gDW. The upper left submatrix equals the original stoichiometric matrix and the lower left submatrix represents the kinetic data (Fig. 2.10). The protein level is added as an upper limit for every enzyme used; in this way, the constraint is applied to every flux. A cell has limited total protein. Therefore, total enzyme usage concentrations given as an upper limit are adjusted according to proteomic data. Without proteomic data, a maximal total enzyme constraint is applied. Turnover numbers are taken from databases such as BRENDA [2].

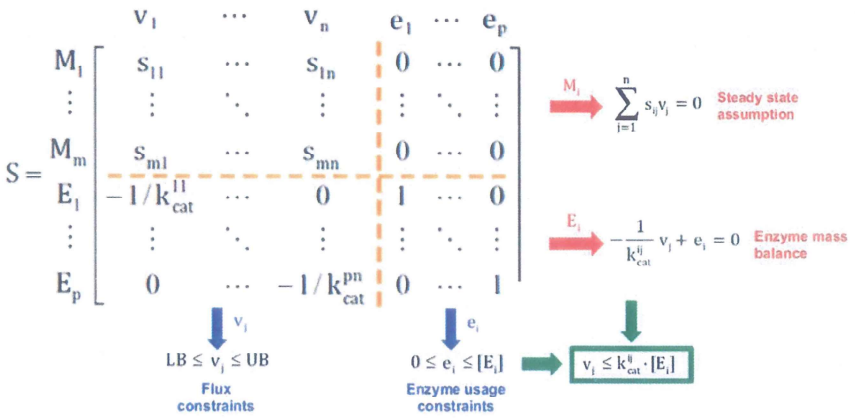

Figure 2.10 Explanation of the stoichiometric matrix used in GECKO method. M: metabolites; E: enzymes; v: metabolic fluxes; e: enzyme usage [2].

3. MATERIALS AND METHODS

In this study, several metabolic modeling strategies were combined with each other and these combined strategies were used to predict mutant phenotypes. The methods for the combination of these modeling strategies and the model used for the prediction of phenotypes are explained in the next subsections.

3.1. iML1515 model of *Escherichia coli* K-12 MG1655

iML1515 model of *Escherichia coli* K-12 MG1655 was obtained as a MATLAB file (http://bigg.ucsd.edu/models/iML1515) from BIGG database [139], which has 2712 reactions and 1877 metabolites. The model contains not only irreversible reactions but also reversible reactions. An objective function of the iML1515 model from BIGG database is the maximization of the biomass formation reaction. In the iML1515 model, 'Cob(I)alamin exchange' was an irreversible reaction. Its reaction's lower bound was equated to 0. However, with the new objective function reaction, 'Cob(I)alamin exchange' reaction created no specific growth rate. Therefore, this reaction's lower bound was adjusted to -1000.

All combined modeling strategies perfomed in this study were done by converting the iML1515 model into the model using only irreversible reactions. For the application of GECKO method, enzymatic constraints were added to iML1515 as one more row to the stoichiometric matrix in a GECKO fashion and the new models were saved as 'iML1515_GECKO_BRENDA' (for enzyme molar concentration) and 'iML1515_GECKO_BRENDA_2' (for enzyme mass concentration) by Tobias Alter via guidelines of the previous study [2]. Then, 21 genes were selected for mutant phenotype prediction studies; gene names, subsystems and related reactions are shown in Table 3.1 and Figure A.1 according to experimental data [68]. Mutant phenotype predictions for single-gene knockout was obtained for each gene in this study.

Table 3.1 Gene list used in the model to predict mutant phenotypes

Gene	Subsystem in the model	Reactions in the model
glk	Cofactor and Prosthetic Group Biosynthesis	HEX1
zwf	Cobifactor and Prosthetic Group Biosynthesis	G6PDH2r
gnd	Cofactor and Prosthetic Group Biosynthesis	GND
rpe	Methylglyoxal Metabolism	RPE
edd	Nucleotide Salvage Pathway	EDD
eda	Glycolysis/Gluconeogenesis	EDA, OAADC
pgi	Extracellular exchange	PGI
tpiA	Transport, Outer Membrane Porin	TPI
ptsG	Alternate Carbon Metabolism	ACGAptspp, GLCptspp
crr	Cofactor and Prosthetic Group Biosynthesis	ACGAptspp, GLCptspp SUCptspp, TREptspp ACMUMptspp, MALTptspp
pgm	Alternate Carbon Metabolism	PGMT
rpiB	Folate Metabolism	RPI, ALLPI
tktA	Pentose Phosphate Pathway	TKT1, TKT2
tktB	Glycolysis/Gluconeogenesis	TKT1, TKT2
talA	Glycolysis/Gluconeogenesis	TALA
talB	Glycolysis/Gluconeogenesis	TALA
pfkA	Tyrosine, Tryptophan, and Phenylalanine Metabolism	PFK, PFK_2, PFK_3
pfkB	Cysteine Metabolism	PFK
fbp	Alternate Carbon Metabolism	FBP
fbaB	Extracellular exchange	FBA
sgcE	-	RPE

3.2. Constrained Allocation Flux Balance Analysis

CA-FBA (protein allocation constraints added version of FBA method) simulations can be done in a homogenous and heterogenous fashion (mentioned in Section 2.3.6). The examination of CA-FBA (mentioned in Section 4.1) was started with a test-case to understand whether the code taken from the previous study [10] runs properly and gives similar results. Therefore, a specialized version of the iJR904 model (outdated GEM model of *E. coli* K-12 MG1655) supplied in the supplementary materials of the previous study [10] were taken and used.

3.2.1. CA-FBA Homogenous Case

For CA-FBA homogenous case simulations, the model for wild-type phenotype prediction was modified for constrained allocation by equating glucose uptake rate, w_C, w_E, w_R, and Φ_{max} to 8.5 mmol/g_{DW}.h, $2x10^{-3}$ g_{DW}h/mmol, $1.3x10^{-3}$ g_{DW}h/mmol, 0.169 h, and 0.484, respectively. Only glucose uptake backward reaction rate's upper bound and lower bound in the model for single-gene knockout mutant were equated to 1000 mmol/g_{DW}.h and 0 mmol/g_{DW}.h, respectively as distinct from the model for wild-type.

3.2.2. CA-FBA Heterogenous Case

For CA-FBA heterogenous case simulations, the model for wild-type and single-gene knockout mutant phenotype prediction was modified for constrained allocation with the same w_E, w_R, Φ_{max} and glucose uptake rate as in the homogenous case. In the heterogenous case, <w> was calculated as $1.5x10^{-3}$ g_{DW}h/mmol. For the w_i value, 1000 models were generated. Each model was generated with a random set of weights which independently drawn from the same probability density $p(w) \sim 1/w$, $w_{min} \leq w \leq w_{max}$, which corresponds to a uniform density for the logarithm of w. p(w) is completely decided by its average <w> and width $\delta \equiv \log10(w_{max}/w_{min})$. In accordance with this purpose, 'setWeights_rand' function from the previous study [10] was used to set w_i values randomly.

3.3. Flux Balance Analysis via GEM with Enzymatic Constraints using Kinetic and Omics Data

For a simulation of GECKO-FBA (enzymatic constraints added version of FBA method) with molar/mass enzyme usage concentration, the model for wild-type and single-gene knockout mutant of *E. coli* glucose uptake rate and total enzyme amount were constrained to 8.5 mmol/g_{DW}.h and 29.65 μmol/g_{DW} (enzyme usage molar concentration) or 2163.5 mg/g_{DW} (enzyme usage mass concentration), respectively.

3.4. Minimization of Metabolites Balance

3.4.1. CA-MiMBI

For a simulation of CA-MiMBI (protein allocation constraints added version of MiMBI method), all exchange reactions in the model and the last metabolite line in the model's S matrix, which represents constrained allocation equation, were excluded. Reference flux distribution was calculated by CA-FBA homogenous case simulation for the wild-type of *E. coli* and 'solMiMBI.x' named solution vector was saved with the 'iML1515_MiMBI_fluxdist_WT' name.

3.4.2. GECKO-MiMBI with Enzyme Usage Molar/Mass Concentration

GECKO-MiMBI (enzymatic constraints added version of MiMBI method) simulation with enzyme usage molar concentration/mass concentration has been constructed in 2 different ways: GECKO-MiMBI-Reaction and GECKO-MiMBI-Enzyme. For a simulation of GECKO-MiMBI-Reaction, all exchange reactions in the model, the last metabolite line in the model's S matrix, which represents enzymes and total enzyme usage amount, were excluded.

For a simulation of GECKO-MiMBI-Enzyme, all reactions in the model, and all metabolite lines in the model's S matrix which represents metabolites and total enzyme usage amount were excluded. Reference flux distribution was calculated by GECKO-FBA simulation.

3.5. Metabolic Adjustment Minimization Method

The source code for the solution function of MOMA was taken from the COBRA Toolbox library. However, metabolite accumulation vector which means right hand side of 'S.v=0' equation (it is represented as 'b' vector in the MOMA function code) is formed from zeros vector in the code. This vector does not contain any total proteome or enzyme usage amount. Therefore, in the code, the b vector was equated to the model b vector instead of zeros vector, and this new code was saved and used in this study.

3.5.1. CA-MOMA

For a simulation of CA-MOMA (protein allocation constraints added version of MOMA method), the model for wild-type of *E. coli* and single-gene knockout mutant was modified for constrained allocation as in CA-FBA homogenous case. MOMA solution for the closest mutant phenotype prediction to wild-type prediction was calculated by using these modified models for wild-type and single-gene knockout mutant.

3.5.2. GECKO-MOMA with Molar/Mass Enzyme Usage Concentration

For a simulation of GECKO-MOMA (enzymatic constraints added version of MOMA method) with molar/mass enzyme usage concentration, the model for wild-type *E. coli* and single-gene knockout mutant was modified for constrained allocation as in GECKO-FBA homogenous case.

3.6. Single Gene Deletion to Predict Mutant Phenotype with Unconstrained/Constrained Model

Single gene deletions in each modeling strategy were done by setting upper and lower bounds of the corresponding reactions to zero. Related reactions can be detected according to the gene-reaction rules in a model. All mutant phenotype predictions were made with the unconstrained model (unconstrained substrate uptake rate for each mutant) and constrained model (constrained model substrate uptake rate for each mutant), separately. All predictions with unconstrained model have occurred as mentioned in the previous sections (by constraining glucose uptake rate for only wild-type). For predictions with

constrained model, glucose and oxygen uptake rates were taken from a previous study, specific for each single-gene knockout mutant [68] and these values were used in determining the upper and lower bounds. Other details were applied same as the procedure given for unconstrained model.

3.7. MATLAB and COBRA Toolbox

All modeling strategies (Figure 3.1) were applied via COBRA Toolbox v2.0 [140] in MATLAB R2018b (MathWorks Inc., Natick, MA, USA) by using the adjusted version of the *E. coli* iML1515 genome-scale model.

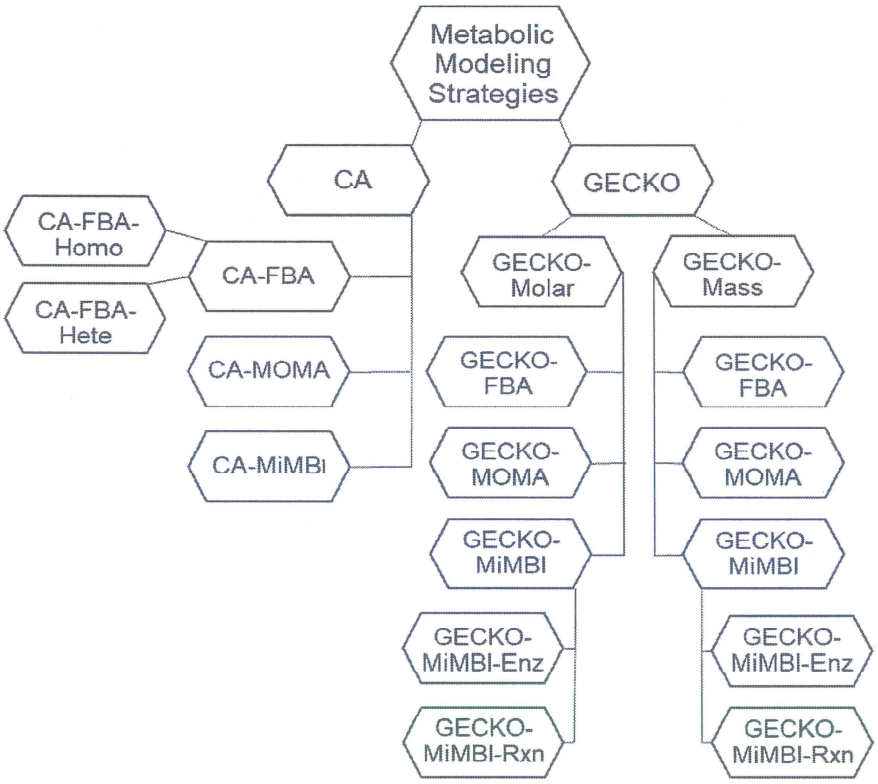

Figure 3.1 Applied metabolic modeling strategies in this study

Gurobi 7.5.1 under an academic license was used for linear programming (LP) and quadratic programming (QP) problem solutions. The general application algorithm for these modeling strategies are shown in Figure 3.2. The predicted results were then plotted against the experimental data using Origin Pro 2017.

3.8. Pearson Correlation Coefficient

Pearson Correlation Coefficient (PCC, ρ) was used to estimate accuracy of the predictions in this thesis and compare to the experimental results from the previous study (Table A.1) [68]. The formula of PCC is given in Equation 3.1. x_i and y_i represent i^{th} observed result and i^{th} predicted result, respectively. ρ_{xy} represents pearson correlation coefficient between x and y. n represents the number of results.

$$\rho_{xy} = \frac{n\sum_i x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2} - \sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (3.1)$$

PCC value varies between -1 and +1 to express the power of relationship. A '+' sign indicated a positive relationship and a '–' sign indicates a negative relationship between the compared values.

Figure 3.2 General application algorithm for metabolic modeling strategies in this study

4. RESULTS AND DISCUSSION

In this study, investigation of the influence of adding protein allocation constraints to GEM on phenotype predictions of *Escherichia coli* mutant strains was studied by providing information on the condition for only the wild-type strain. In addition, several combinations of computational analysis methods were investigated to achieve more accurate predictions than previous methods, especially for specific growth rate, biomass and acetate yields of mutants.

For all the methods tested in this study, experimental data (Table A.1) [68] is plotted against predicted results, where each data point indicates a different mutant and a linear relationship is expected between the predicted and experimental values. Then, pearson correlation coefficient (PCC) is used to determine how well the predictions match with the experimental data. The PCC depends on both variables being normally distributed. This means that the variables should have a linear relationship. In addition, this coefficient is very sensitive to outliers. The outliers can cause a lower coefficient than expected. Therefore, correlation coefficients for non-parametrics were not used. All PCC values for the combined strategies studied in this thesis are given in Table B.1.

4.1. Validation of CA-FBA

Firstly, the CA-FBA modeling strategy was examined and validated. The aim of this part was to test the source code from a previous study [10] for CA-FBA to understand whether the code runs properly and gives similar results; where, a specialized version of the iJR904 model of *E. coli* K-12 MG1655 [10] was taken and used. The source for the application of CA-FBA and the specialized version of iJR904 model were taken from supplementary materials of the work by Mori et al., 2016 [10]. In the homogenous case with nutrient limitation, w_C was varied in the range from zero to one. This range was divided into 100 equal values. For each w_C, flux distributions were calculated by CA-FBA. In the heterogenous cases, w_C was also set in the range from zero to one (at w_C=0, maximum specific growth rate can be found) by dividing into 100 values. For each w_C value, 1000 CA-FBA solutions were found with random w_i's, and an average value for each flux was calculated as the output. The same values for the protein allocation

constraints were applied in both homogenous and heterogenous cases, as in the previous study [10]. Similar results were calculated with the method of previous study [10] as shown in Figure 4.1. The results proved that the code and the model worked properly.

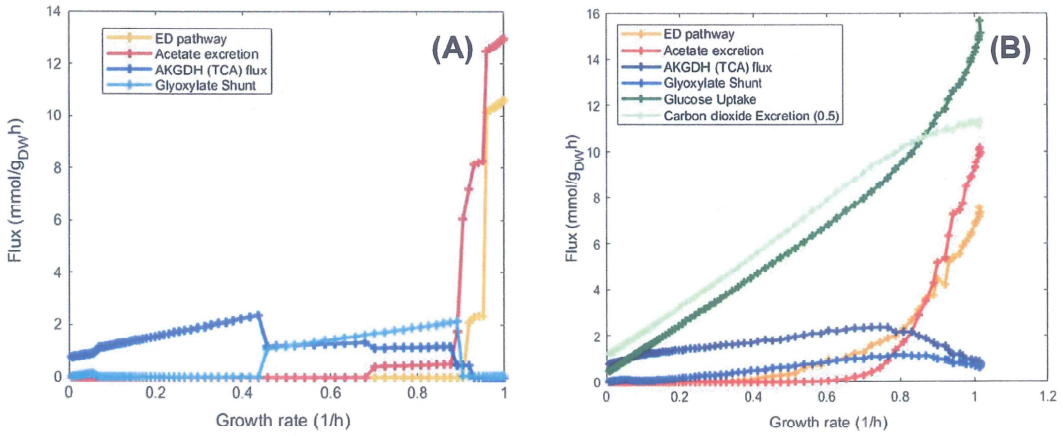

Figure 4.1 Examination of CA-FBA with the same model and conditions as in the previous study [10]. (A) The homogenous case of CA-FBA when the carbon source is limited. Results are CA-FBA fluxes as a function of specific growth rate (μ), obtained by varying the degree of carbon (glucose) limitation (C-lim), by adjusting w_C value from zero to one $g_{DW}h/mmol$. w_E and w_R values were constant, $8.3x10^{-4}$ $g_{DW}h/mmol$ and 0.169 h, respectively. (B) The heterogenous case of CA-FBA when the carbon source is limited. Results are average CA-FBA fluxes as a function of μ, obtained by varying the degree of carbon (glucose) limitation (C-lim) by adjusting w_C value 0-1 $g_{DW}h/mmol$. <w> and w_R values were constant, $8.8x10^{-4}g_{DW}h/mmol$ and 0.169 h, respectively. In both cases, $\Phi_{max} = 0.484$.

4.2. Studying CA-FBA with iML1515 Model of *Escherichia coli* K-12 MG1655

The latest model of *E. coli* K-12 MG1655 is the iML1515 model [65]. Therefore, as the next part of this study, CA-FBA modeling strategy was studied with iML1515 model. Between the iJR904 and iML1515 model of *E. coli* K-12 MG1655, there are approximately 1000 extra reactions. Because of this difference, new w_E (for the homogenous case) and <w> (for the heterogenous

case) values were needed and both were set at $7.8x10^{-4}$ $g_{DW}h/mmol$. These new values were set based on specific growth rate, so that it reached its maximum value (~1 h^{-1}) when w_C was zero. After the integration, in the homogenous case using the iML1515 model, acetate overflow metabolism started at approximately a specific growth rate of μ=0.7 h^{-1} and acetate excretion rate reached 14 $mmol/g_{DW}h$. These results (Figure 4.2.A) were similar to the previous study [10]. In the heterogenous case using the iML1515 model, similar results (Figure 4.2.B) were also obtained until μ reached 0.8 h^{-1}. The differences from the previous study [10] after μ=0.8 h^{-1} might be caused by the setting of w_i values randomly. The exactly same w_i values as w_i values setted randomly in the previous study [10] might not be setted so this situation might created this differences.

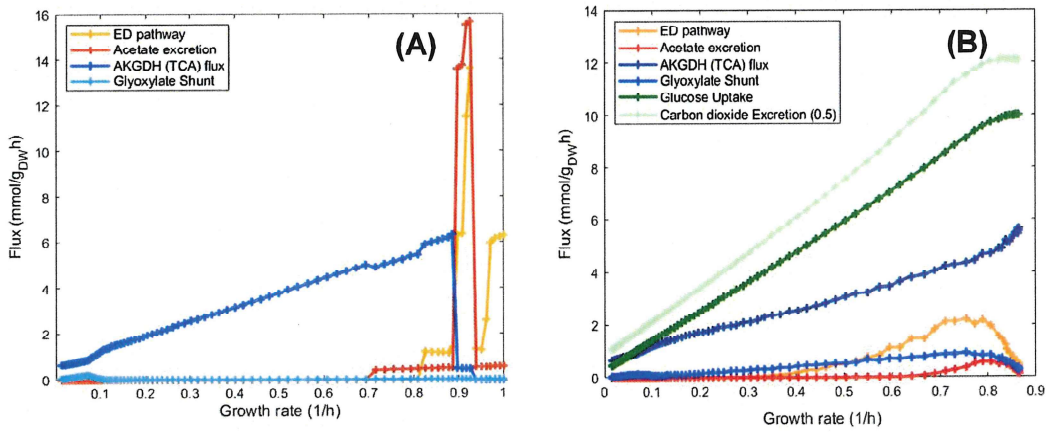

Figure 4.2 Studying CA-FBA with the iML1515 model.(A) The homogenous case of CA-FBA, when the carbon source is limited. Results are CA-FBA fluxes as a function of μ, obtained by varying the degree of carbon (glucose) limitation (C-lim) by adjusting w_C value 0-1 $g_{DW}h/mmol$. (B) The heterogenous case of CA-FBA when carbon source is limited. Results are average CAFBA fluxes as a function of μ, obtained by varying the degree of carbon (glucose) limitation (C-lim) by adjusting w_C value 0-1 $g_{DW}h/mmol$. w_E or <w>, w_R and Φ_{max} values were constant, $7.8x10^{-4}$ $g_{DW}h/mmol$ and 0.169 h and 0.484, respectively.

4.3. Prediction of Mutant Phenotype Using Metabolic Modeling Strategies with Constrained Allocation Method and the Unconstrained iML1515 Model

Phenotypes of mutants of *E. coli* K-12 MG1655 were predicted via metabolic modeling strategies, which were combined with constrained allocation (CA) method by using an unconstrained iML1515 model. For the mutants in the unconstrained model, data were not defined for nutrient uptake. For knockout predictions, as a starting point, 21 genes (without ybhE) were selected which are the same genes from a previous study [68]. ybhE gene was not in the iML1515 model, thus, *ΔybhE* phenotype predictions were not included. Also, experimental results for glucose uptake, growth and acetate secretion rate of these genes were taken from the previous study [68]. For *ΔtktA* and *ΔtktB* knockout phenotype, biomass and acetate yields were calculated as zero, even though these predictions were different from zero in the simulation results of the previous study [68]. The reason for this can be the use of iML1515 model. Because, in the previous study by Long et al., 2016 [68], iAF1260 genome-scale model of *E. coli* was used and they have simulated growth and acetate secretion rate of *ΔtktA* and *ΔtktB* knockout as different from zero. Therefore, in the iML1515 model (latest model of *E. coli*), added new reactions distinct from the iAF1260 model may prevent the accurate predictions for these knockouts.

According to results of Long et al. 2016 [68], wild-type of *E. coli* K-12 MG1655 for 8.5 mmol/g_{DW}h as glucose uptake rate, μ was reported to be 0.63 h^{-1}. In the prediction strategies with CA method, w_C, w_E and <w> values were set to $2x10^{-3}$ g_{DW}h/mmol, $1.3x10^{-3}g_{DW}$h/mmol and $1.5x10^{-3}$ g_{DW}h/mmol, respectively, to reach the same glucose uptake rate and specific growth rate as the experimental values [68].

In CA simulations for the wild-type flux distribution, both upper and lower bounds of glucose uptake rates were set to 8.5 mmol/g_{DW}h as distinct from GECKO simulations in order to get more accurate wild-type flux distribution. In GECKO simulations for wild-type flux distribution, only the upper bound for glucose uptake rate was equated to 8.5 mmol/g_{DW}h.

4.3.1. CA-FBA Predictions

Mutant phenotype predictions by CA-FBA were computed by modifying the model according to constrained allocation method as homogenous and heterogenous cases, separately. In both cases, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted, and biomass and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are shown in Figure 4.3, Figure 4.4, Table 4.1 and Table 4.2. All results are given in detail in Table C.1 and Table C.2.

Pearson correlation coefficient (PCC, ρ) values of biomass and acetate yield predictions of CA-FBA method homogeneous case were 0.11 and 0.12, respectively. PCC values of biomass and acetate yield predictions from Long et al., 2016 [68] were 0.17 and -0.12, respectively. Therefore, biomass and acetate yield predictions by homogenous case of CA-FBA method were calculated to be close to the FBA predictions in the previous study [68] according to PCC values (Figure 4.3.A, Figure 4.3.B). However, growth rate predictions (ρ=-0.28) were not well as the FBA predictions in the previous study [68] (ρ=0.43) according to PCC values (Figure 4.3.C). Also, acetate yield predictions of *Δedd* and *Δeda* knockouts were found close to the experimental data in the homogenous case (Figure 4.3.A).

In the heterogenous case, PCC values of biomass yield and growth rate predictions were 0.25 and -0.27, respectively. PCC values of biomass yield and growth rate predictions from Long et al., 2016 [68] were 0.17 and 0.43, respectively. Therefore, biomass yield predictions by heterogenous case of CA-FBA method were calculated better than the FBA predictions in the previous study [68] according to PCC values (Figure 4.4.A). However, growth rate predictions were not well as the FBA predictions in the previous study [68] (Figure 4.4.B).

Especially, for *Δpgi* and *Δcrr* knockouts, highly accurate biomass yield predictions were obtained in the heterogenous case (Figure 4.4) and for the '*ΔptsG*' knockout, a highly accurate acetate yield prediction was attained in the

homogenous case (Figure 4.3.A) compared with the experimental data. Acetate yield predictions of knockouts (except *ΔtktA* and *ΔtktB* knockouts) in heterogenous cases were calculated as zero (data not plotted).

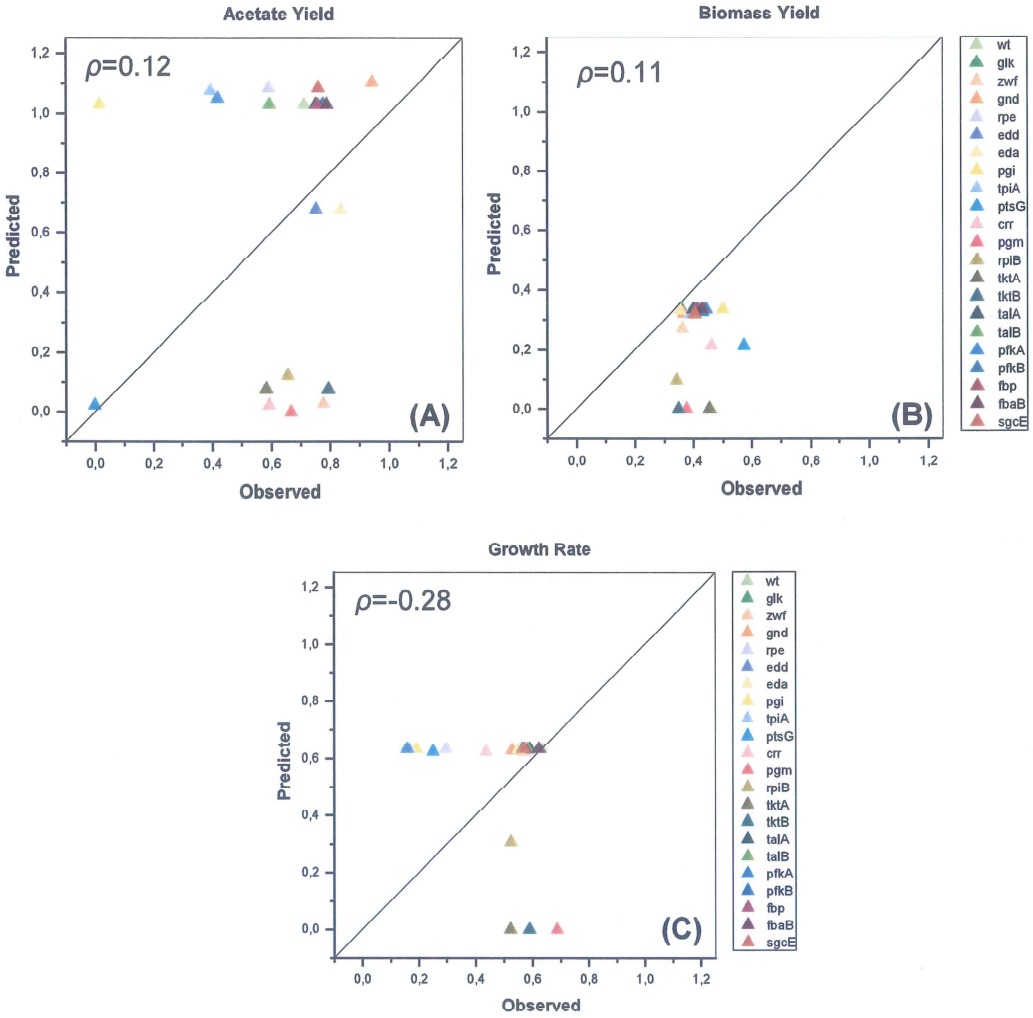

Figure 4.3 Mutant phenotype prediction of unconstrained iML1515 model via the homogenous case of the CA-FBA method. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data, while w_C, w_E, w_R and Φ_{max} values are 2×10^{-3} $g_{DW}h/mmol$, 1.3×10^{-3} $g_{DW}h/mmol$, 0.169 h and 0.484 respectively.

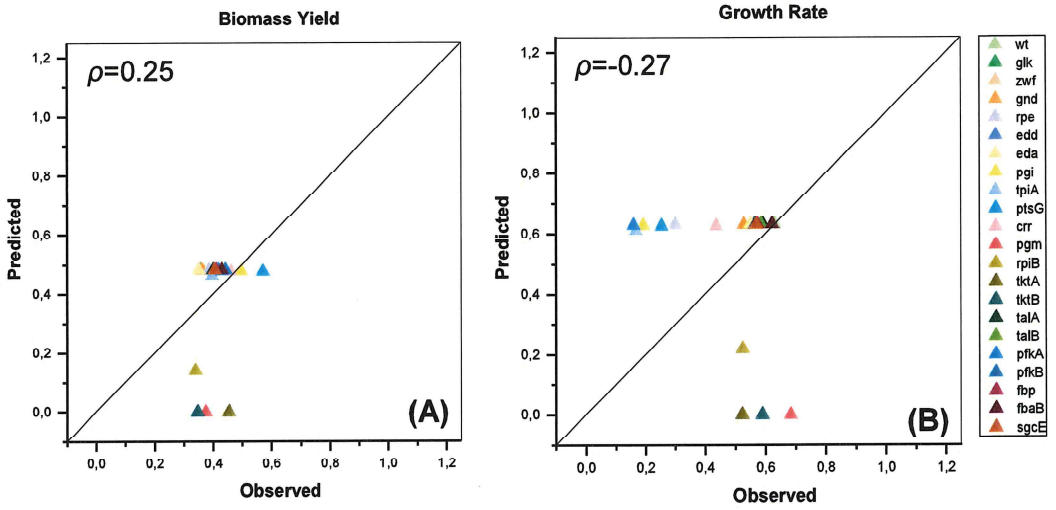

Figure 4.4 Mutant phenotype prediction of unconstrained iML1515 model via the heterogenous case of the CA-FBA method. The comparison of simulation results of (A) biomass yield and (B) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data while w_C, <w>, w_R and Φ_{max} values are $2x10^{-3}$ $g_{DW}h/mmol$, $1.5x10^{-3}$ $g_{DW}h/mmol$, 0.169 h and 0.484 respectively.

4.3.2. CA-MOMA Predictions

Mutant phenotype predictions via CA-MOMA strategy were applied by modifying the model according to constrained allocation method. In CA-MOMA, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are given in Figure 4.5, Table 4.1 and Table 4.2. All results are given in detail in Table C.3.

PCC values of biomass yield predictions of CA-MOMA method and the MOMA predictions in the previous study [68] were 0.30 and 0.07, respectively. Therefore, biomass yield predictions by CA-MOMA method were calculated better than the MOMA predictions in the previous study [68] according to PCC values (Figure 4.5.B). PCC values of acetate yield and growth rate predictions were -0.02 and -0.21, respectively. PCC values of acetate yield and growth rate predictions from Long et al., 2016 [68] were 0.24 and 0.51, respectively. Unfortunately, acetate yield and growth rate predictions were not well as the MOMA predictions in the

previous study [68] according to PCC values (Figure 4.5.A, Figure 4.5.C). There is not much difference in the phenotypes between the various mutants, so the slight growth defects observed for the mutant strains may not have been predicted in detail.

Highly accurate acetate yield prediction (Figure 4.5.A) for *ΔtpiA* knockout was calculated and also for *ΔpfkA* knockout, acetate yield prediction was close to the experimental data. Highly accurate biomass yield predictions of most knockouts were calculated in CA-MOMA, which was a better strategy than CA-FBA and CA-MiMBI for biomass prediction.

4.3.3. CA-MiMBI Predictions

Mutant phenotype predictions via CA-MiMBI strategy were applied by calculating reference flux distribution via CA-FBA homogenous case simulation for the wild-type of *E. coli*. In CA-MiMBI, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are given in Figure 4.6 and Table 4.1 and Table 4.2 and also all results are given in Table C.4.

PCC values of biomass yield, acetate yield and growth rate predictions of CA-MiMBI method were 0.34, -0.10 and -0.19, respectively. In the combinations of CA methods with modelling strategies with unconstrained model, the best biomass yield predictions according to PCC value was obtained with CA-MiMBI.

Highly accurate acetate yield prediction (Figure 4.6.A) for *ΔtktA* knockout was calculated, however, biomass yield of this mutant equals to zero. Thus, the overall phenotype is unconclusive. Also, for *ΔpfkA*, *ΔtpiA* and *Δcrr* knockouts, acetate yield predictions were close to the experimental data. Highly accurate biomass yield predictions of most knockouts were calculated with CA-MiMBI.

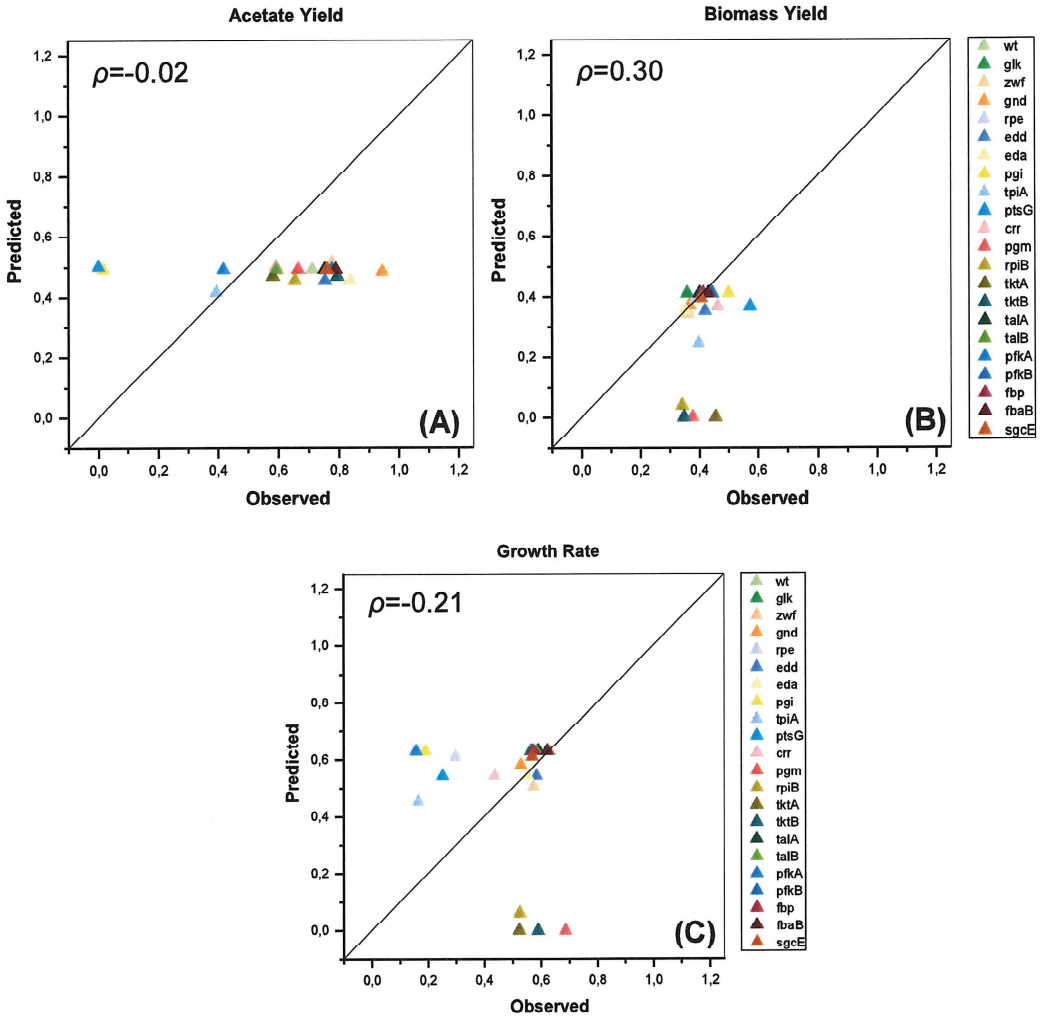

Figure 4.5 Mutant phenotype prediction of unconstrained iML1515 model via CA-MOMA method. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. FBA simulations of reference wild-type and single-gene deletion knockouts were done by CA-FBA, while w_C, w_E, w_R and Φ_{max} values are 2×10^{-3} $g_{DW}h/mmol$, 1.3×10^{-3} $g_{DW}h/mmol$, 0.169 h and 0.484 respectively. Glucose uptake rate for reference wild-type CA-FBA simulation is fixed to 8.5 mmol/ $g_{DW}h$.

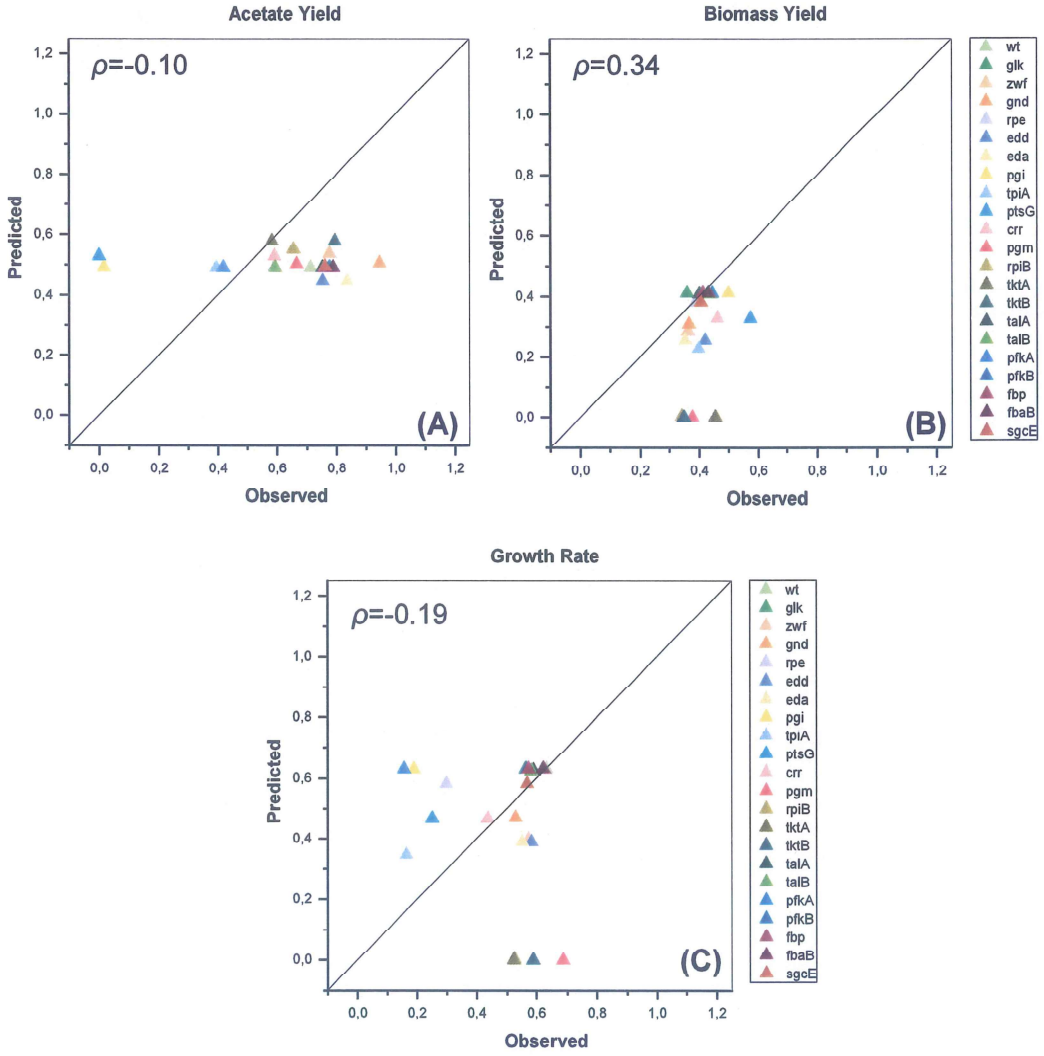

Figure 4.6 Mutant phenotype prediction of unconstrained iML1515 model via CA-MiMBI method. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by CA-FBA while w_C, w_E, w_R, Φ_{max} values and glucose uptake rate are $2x10^{-3}$ $g_{DW}h/mmol$, $1.3x10^{-3}$ $g_{DW}h/mmol$, 0.169 h, 0.484 and 8.5 mmol/ $g_{DW}h$ respectively.

Table 4.1 Biomass yield results of mutants phenotype predictions of unconstrained iML1515 model by modeling methods with the combination of constrained allocation, compared with the experimental data (Exp.) from [68].

Biomass Yield					
Gene	**CA-FBA-Homo**	**CA-FBA-Hete**	**CA-MOMA**	**CA-MiMBI**	**Exp.**
wt	0.335	0.483	0.412	0.412	0.415
glk	0.335	0.483	0.412	0.412	0.359
zwf	0.271	0.481	0.343	0.285	0.363
gnd	0.321	0.481	0.371	0.308	0.366
rpe	0.32	0.482	0.395	0.381	0.388
edd	0.331	0.483	0.353	0.255	0.419
eda	0.331	0.483	0.353	0.255	0.353
pgi	0.335	0.479	0.412	0.412	0.499
tpiA	0.32	0.464	0.246	0.227	0.398
ptsG	0.215	0.478	0.368	0.329	0.573
crr	0.215	0.478	0.368	0.329	0.462
pgm	0	0	0	0	0.377
rpiB	0.096	0.141	0.039	0.003	0.342
tktA	0	0	0	0	0.455
tktB	0	0	0	0	0.349
talA	0.335	0.483	0.412	0.409	0.401
talB	0.335	0.483	0.412	0.409	0.441
pfkA	0.328	0.48	0.412	0.412	0.431
pfkB	0.335	0.483	0.412	0.412	0.444
fbp	0.335	0.483	0.412	0.412	0.412
fbaB	0.335	0.483	0.412	0.412	0.431
sgcE	0.32	0.482	0.395	0.381	0.406

Table 4.2 Acetate yield results of mutants phenotype prediction of unconstrained iML1515 model by modeling methods with the combination of constrained allocation, compared with the experimental data (Exp.) from [68].

Acetate Yield					
Gene	CA-FBA-Homo	CA-FBA-Hete	CA-MOMA	CA-MiMBI	Exp.
wt	1.030	0	0.492	0.492	0.714
glk	1.030	0	0.492	0.492	0.779
zwf	0.029	0	0.514	0.538	0.778
gnd	1.105	0	0.484	0.507	0.946
rpe	1.087	0	0.491	0.495	0.594
edd	0.679	0	0.455	0.448	0.755
eda	0.679	0	0.455	0.448	0.839
pgi	1.030	0	0.492	0.492	0.017
tpiA	1.077	0	0.416	0.492	0.397
ptsG	0.023	0	0.501	0.529	0.000
crr	0.023	0	0.501	0.529	0.594
pgm	0	0	0.492	0.503	0.667
rpiB	0.123	0	0.454	0.551	0.657
tktA	0.077	2	0.469	0.577	0.584
tktB	0.077	2	0.469	0.577	0.796
talA	1.030	0	0.492	0.492	0.754
talB	1.030	0	0.492	0.492	0.596
pfkA	1.051	0	0.492	0.492	0.420
pfkB	1.030	0	0.492	0.492	0.777
fbp	1.030	0	0.492	0.492	0.760
fbaB	1.030	0	0.492	0.492	0.790
sgcE	1.087	0	0.491	0.495	0.762

In summary, mutant phenotype prediction results by applying combination of modelling strategies with CA method were given when nutrient uptakes were not constrained for mutants. Within these four combined methods, CA-FBA

homogenous case for both acetate yield and growth rate, and CA-MiMBI for biomass yield predictions were the best methods according to their PCC values. In the previous studies [68, 141, 142], for different engineered *E. coli* strains and several mutants of *E. coli*, PCC values for FBA were varied in 0.10-0.30 range and for MOMA in the 0.20-0.50 range. Therefore, these modelling strategies in this section shows similar accuracy with the previous studies according to PCC values. No overflow metabolism were seen in CA-FBA heterogenous case. This situation might be due to the absence of protein limitation, which would have enforced ATP production via fermentation pathways. In addition, the situation might be that this combined strategy could not reach the acetate yield of the experimental data for wild-type, therefore the insufficiency of prediction might have continued for mutants.

4.4. Prediction of Mutant Phenotype Using GECKO Model and the Unconstrained iML1515 Model

Phenotypes of mutants of *E. coli* K-12 MG1655 were predicted via metabolic modeling strategies, which were combined with a model modified from the iML1515 model according to the GECKO method. Predictions were done with the unconstrained model. For mutants of unconstrained model, data were not defined for nutrient uptake. For knockout predictions, as a starting point, 21 genes (without ybhE) were selected which are the same genes from the previous study [68]. ybhE gene was not in the iML1515 model and also in the model modified by GECKO. Hence, ybhE knockout phenotype predictions were not done. Also, experimental results for glucose uptake, growth and acetate secretion rate of these genes were taken from the previous study [68].

According to the results of Long et al. 2016 [91], for 8.5 mmol/g_{DW}h as glucose uptake rate of wild-type of *E. coli* K-12 MG1655, μ is 0.63 h^{-1}. The prediction strategies with the GECKO method, total enzyme usage concentrations were adjusted to reach the same glucose uptake and μ as these experimental values.

4.4.1. Total Enzyme Usage Concentration Determination

Because the total protein in the cell is limited, total enzyme usage concentration can be provided as input to the model as an upper bound for each enzyme usage in GECKO method [2]. Therefore, total enzyme usage was calculated as 29.65 μmol/g_{DW} (Figure 4.7, Table C.5) and applied to the model in order to reach experimental phenotypes of wild-type of *E. coli* K-12 MG1655, mentioned in Section 4.4.

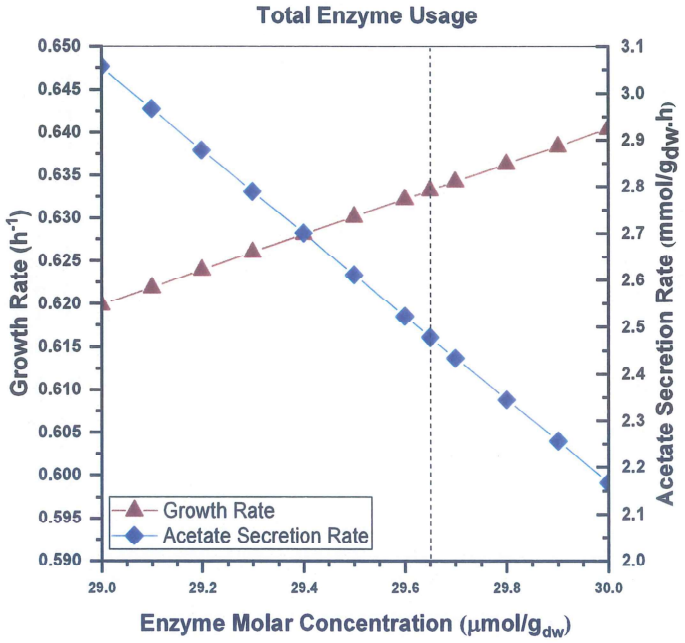

Figure 4.7 Sensitivity analysis for enzyme molar concentration as total enzyme usage amount.

The total enzyme usage molar concentration is related to the number of enzymes. However, the intracellular space is limited and enzymes occupied volume varies depending on the amino acid chain length. Therefore, not only total enzyme usage molar concentration, but also total enzyme usage mass concentration was calculated, which better reflects the total occupied space of all enzymes. In this case, by the same way mentioned in Section 4.4, the total enzyme usage mass concentration was determined as 318 mg/g_{DW} which were taken from the previous study [143]. However, this value could not supply growth rate as 0.63 h^{-1}. That growth rate was supplied by 2163.5 mg/g_{DW} (Figure 4.8, Table C.6). Therefore,

the value from the previous study [143] was multiplied with a correction factor, which was calculated as 6.8 to reach 2163.5 mg/g_{DW}. The value of the total enzyme usage mass concentration was seen as impossible in terms of biological meaning. However, this value was calculated to reach the experimental value. This situation might be because of underestimated k_{cat} values of the enzymes. From now on, it should be thought as a scalar number which was calculated to provide the aim instead of its biological meaning.

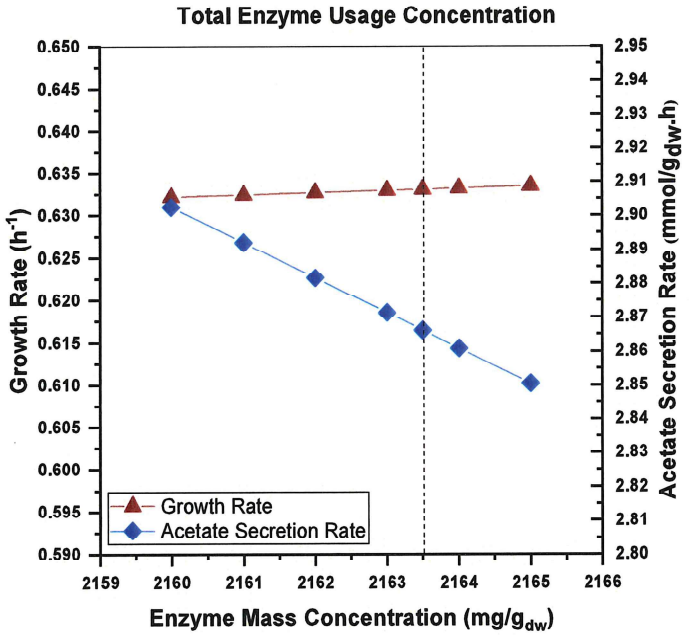

Figure 4.8 Sensitivity analysis for enzyme mass concentration as total enzyme usage amount.

4.4.2. GECKO-FBA Predictions

Mutant phenotype predictions via GECKO-FBA strategy were applied as GECKO-FBA method by using enzyme molar concentration and enzyme mass concentration, separately. In both cases, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are given in Figure 4.9, Figure 4.10 and Table 4.3 and Table 4.4 and also all results are given in Table C.7 and Table C.8.

PCC values of biomass yield predictions of GECKO-FBA method by using enzyme molar concentration, enzyme mass concentration and the FBA predictions in the previous study [68] were 0.28, 0.28 and 0.17, respectively. Therefore, biomass yield predictions of GECKO-FBA method by using both type of concentration were calculated better than the FBA predictions in the previous study [68] according to PCC values (Figure 4.9.B, Figure 4.10.B). PCC values of acetate yield and growth rate predictions of GECKO-FBA method by using enzyme molar concentration were 0.05 and -0.27, respectively (Figure 4.9.A, Figure 4.9.C). PCC values of acetate yield and growth rate predictions of GECKO-FBA method by using enzyme mass concentration were 0.10 and -0.27, respectively (Figure 4.10.A, Figure 4.10.C). PCC values of acetate yield and growth rate predictions from Long et al., 2016 [68] were -0.12 and 0.43, respectively. Unfortunately, acetate yield and growth rate predictions of GECKO-FBA method by using both type of concentration were not well as the FBA predictions in the previous study [68] according to PCC values.

Biomass yield predictions for most knockouts except *ΔtktA*, *ΔtktB*, *ΔrpiB*, and *Δpgm* knockouts were close to experimental data in both enzyme concentration type (Figure 4.9.B, Figure 4.10.B). Acetate yield predictions for *ΔtpiA* and *ΔpfkA* knockouts were close to experimental data in molar enzyme concentration type (Figure 4.9.A) and for *ΔtpiA*, *ΔpfkA* and *ΔtalB* knockouts were close to experimental data in enzyme mass concentration type (Figure 4.10.A).

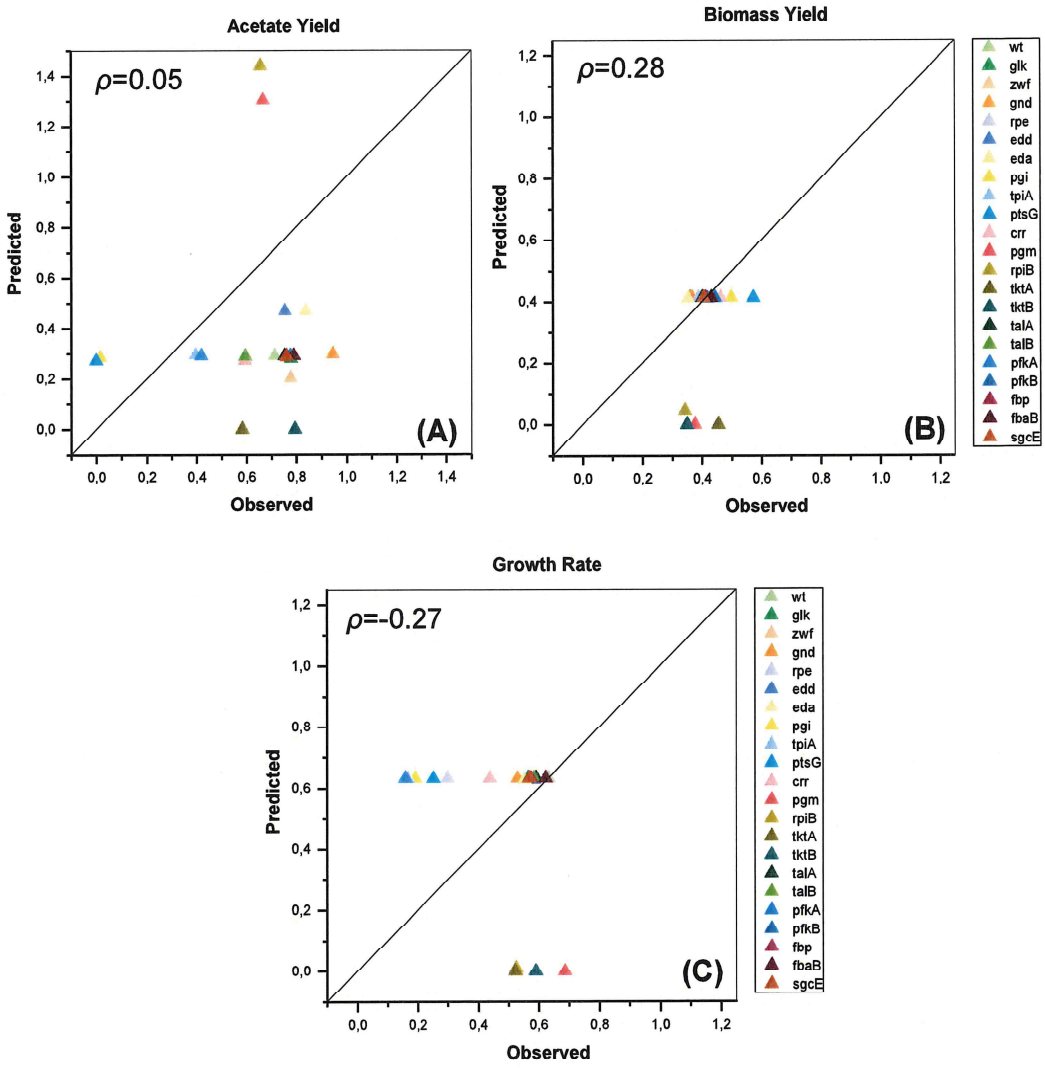

Figure 4.9 Mutant phenotype prediction of unconstrained iML1515 model via the GECKO-FBA method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate of *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data, while total usage amount equals to 29.65 μmol/g_{DW} and upper bound of glucose uptake rate equals to 8.5 mmol/g_{DW}h, are shown.

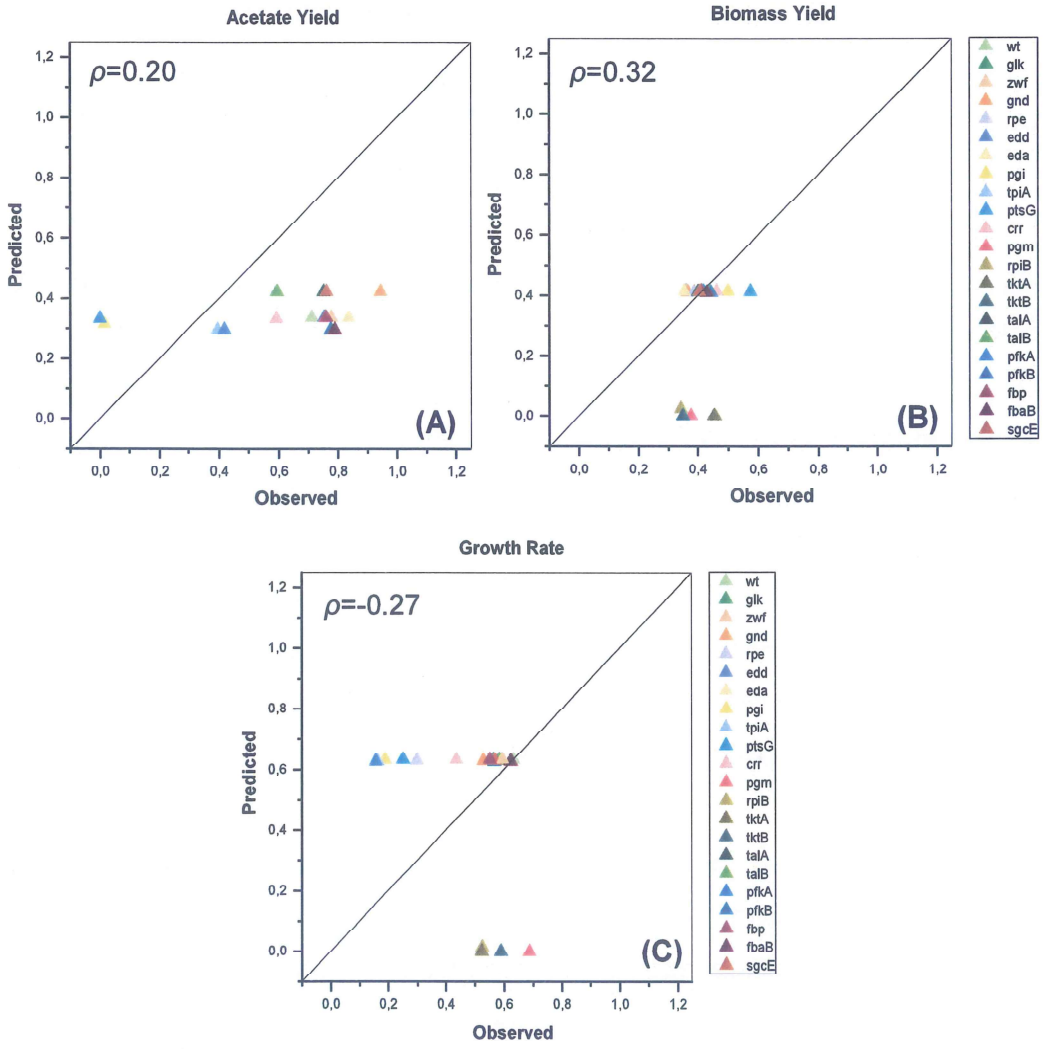

Figure 4.10 Mutant phenotype prediction of unconstrained iML1515 model via GECKO-FBA method by using enzyme mass concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data, while total usage amount equals to 2163.5 mg/g_{DW} and upper bound of glucose uptake rate equals to 8.5 mmol/g_{DW}h, are shown.

4.4.3. GECKO-MOMA Predictions

Mutant phenotype predictions by GECKO-MOMA strategy were applied as GECKO-MOMA method by using enzyme molar concentration and enzyme mass concentration, separately. In both cases, μ, glucose uptake rate, acetate

secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results from GECKO-MOMA method by using enzyme molar concentration are given in Figure 4.11, Table 4.3 and Table 4.4 and also all results are given in Table C.9. In the GECKO-MOMA method by using enzyme mass concentration, the prediction could be done only for *Δpgi* knockout. The result of this phenotype prediction is given in Table C.10.

PCC values of biomass yield predictions of GECKO-MOMA method by using enzyme molar concentration and the MOMA predictions in the previous study [68] were 0.22 and 0.07, respectively. Therefore, biomass yield predictions of GECKO-MOMA method by using enzyme molar concentration were calculated better than the experimental data according to PCC values (Figure 4.11.B). PCC values of acetate yield and growth rate predictions of GECKO-MOMA method by using enzyme molar concentration were -0.16 and -0.38, respectively (Figure 4.11.A, Figure 4.11.C). PCC values of acetate yield and growth rate predictions from Long et al., 2016 [68] were 0.24 and 0.51, respectively. Unfortunately, acetate yield and growth rate predictions of GECKO-MOMA method by using using enzyme molar concentration were not well as the MOMA predictions in the previous study [68] according to PCC values.

In GECKO-MOMA method by using enzyme mass concentration, highly accurate biomass yield predictions of knockouts except *ΔtktA*, *ΔtktB*, *Δpgm*, *Δedd*, *Δcrr*, *Δpgi* and *ΔptsG* knockouts were calculated (Figure 4.11.B). Biomass yields of *Δedd*, *Δcrr*, *Δpgi* and *ΔptsG* knockouts were close to experimental data (Figure 4.11.B).

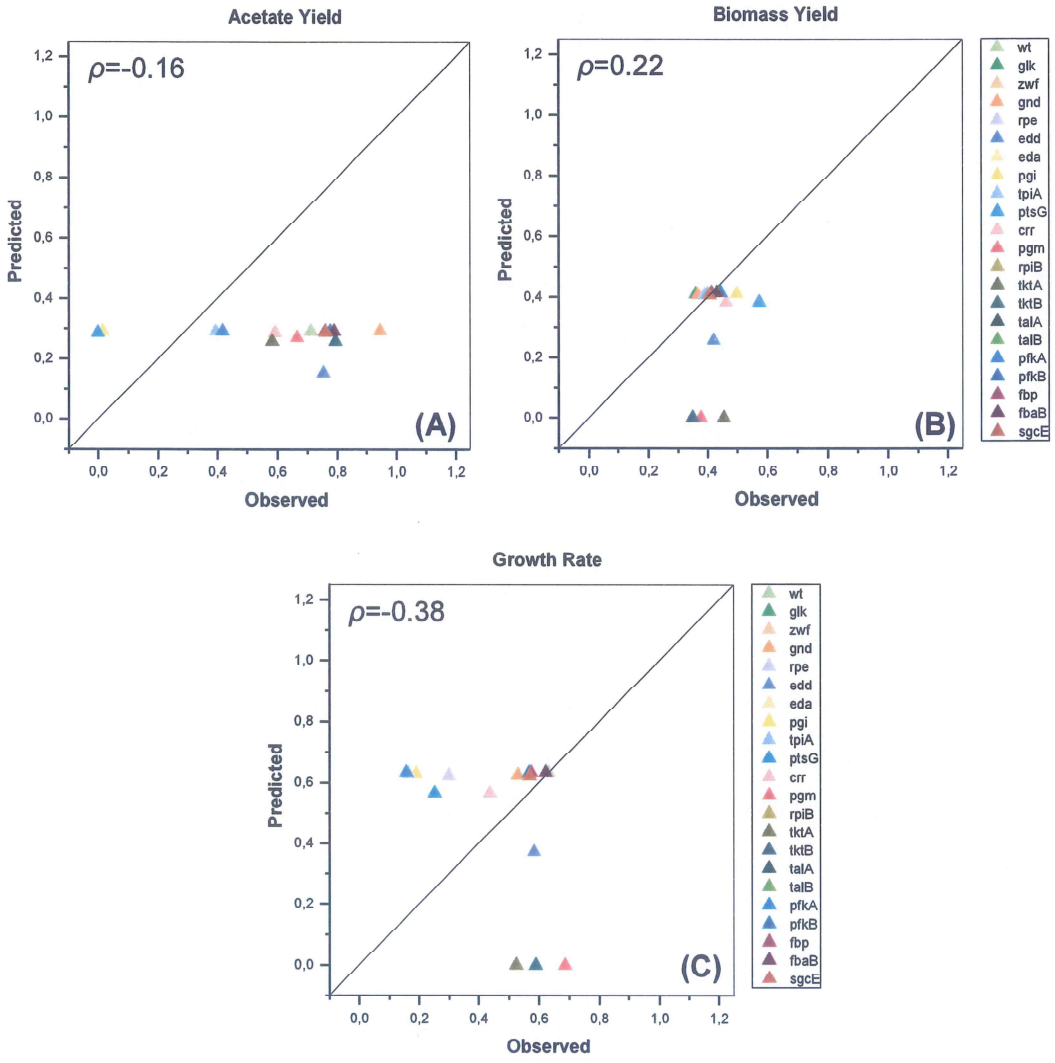

Figure 4.11 Mutant phenotype prediction of unconstrained iML1515 model via GECKO-MOMA method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown FBA simulations of reference wild-type and single-gene deletion knockouts were done by GECKO-FBA, while glucose uptake rate for reference wild-type CA-FBA simulation is fixed to 8.5 mmol/g_{DW}h. The total usage amount equals to 29.65 μmol/g_{DW}.

4.4.4. GECKO-MiMBI Predictions

Mutant phenotype predictions by the GECKO-MiMBI strategy were done by excluding enzyme and exchange reactions (GECKO-MiMBI-Reaction) and excluding reactions except the enzyme reactions (GECKO-MiMBI-Enzyme), separately. In both methods, predictions were done by considering upper bounds for the total enzyme molar concentration and enzyme mass concentration, separately. In both cases, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results from GECKO-MiMBI-Reaction method (with enzyme molar concentration) are given in Figure 4.12, results from GECKO-MiMBI-Reaction method (with enzyme mass concentration) are given in Figure 4.14. Growth rate, acetate and biomass yield results from the GECKO-MiMBI method-Enzyme (with enzyme molar concentration) are given in Figure 4.13, results from GECKO-MiMBI-Enzyme method (with enzyme mass concentration) are given in Figure 4.15. The summary of GECKO-MiMBI is given in Table 4.3 and Table 4.4 and also all results are given in Table C.11, Table C.12, Table C.13, and Table C.14.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Reaction method by using enzyme molar concentration were 0.34, -0.18 and -0.33, respectively (Figure 4.12.A, Figure 4.12.B, Figure 4.12.C). With the GECKO-MiMBI-Reaction method by using enzyme molar concentration, highly accurate biomass yield predictions of knockouts except *ΔtktA*, *ΔtktB*, *Δpgm*, *Δedd*, *ΔrpiB*, *Δpgi*, *Δcrr*, *ΔptsG* and *Δzwf* knockouts were calculated (Figure 4.12.B). Biomass yields for *Δpgi*, *Δcrr*, *ΔptsG* and *Δzwf* knockouts were close to experimental data (Figure 4.12.B). Acetate yields of *ΔtpiA* and *ΔpfkA* knockouts were predicted better than other knockouts and also they were close to experimental data (Figure 4.12.A).

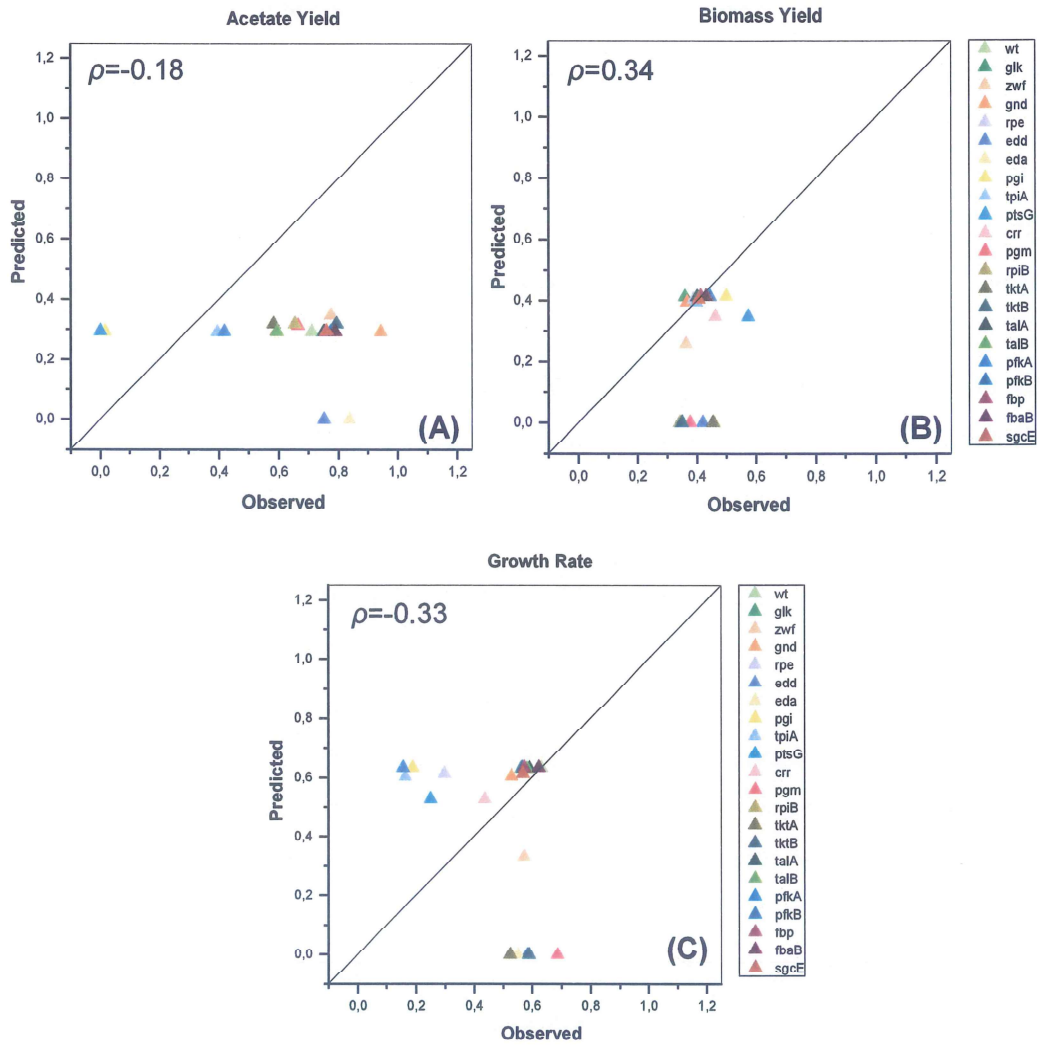

Figure 4.12 Mutant phenotype prediction of unconstrained iML1515 model via GECKO-MiMBI-Reaction method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose uptake rate is fixed to 8.5 mmol/g_{DW}h. The total usage amount equals to 29.65 μmol/g_{DW}.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Enzyme method by using enzyme molar concentration were 0.34, -0.21 and -0.30, respectively (Figure 4.13.A, Figure 4.13.B, Figure 4.13.C). With the GECKO-MiMBI-Enzyme method by using enzyme molar concentration, highly accurate biomass yield predictions of knockouts except *ΔtktA*, *ΔtktB*, *Δpgm*, *ΔrpiB*, *Δedd*, *Δpgi*, *Δcrr*, and *ΔptsG* knockouts were calculated (Figure 4.13.B). Biomass yields of *Δpgi*, *Δcrr*, *ΔptsG* and *Δedd* knockouts were close to experimental data (Figure 4.13.B). Acetate yield of *ΔpfkA* knockout was predicted better than other knockouts and also it was close to experimental data (Figure 4.13.A).

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Reaction method by using enzyme mass concentration were 0.53, 0.71 and -0.53, respectively (Figure 4.14.A, Figure 4.14.B, Figure 4.14.C). For three type predictions, GECKO-MiMBI-Reaction method by using enzyme mass concentration was the best method in the combinations of GECKO methods with modelling strategies using unconstrained model.

In GECKO-MiMBI-Reaction method by using enzyme mass concentration, biomass yield predictions of knockouts except *ΔtktA*, *ΔtktB*, *Δpgm*, *ΔrpiB*, *ΔtalA*, *ΔtalB* and *ΔsgcE* knockouts were close to experimental data (Figure 4.14.B). Acetate yields of most knockouts were close to experimental data (Figure 4.14.A).

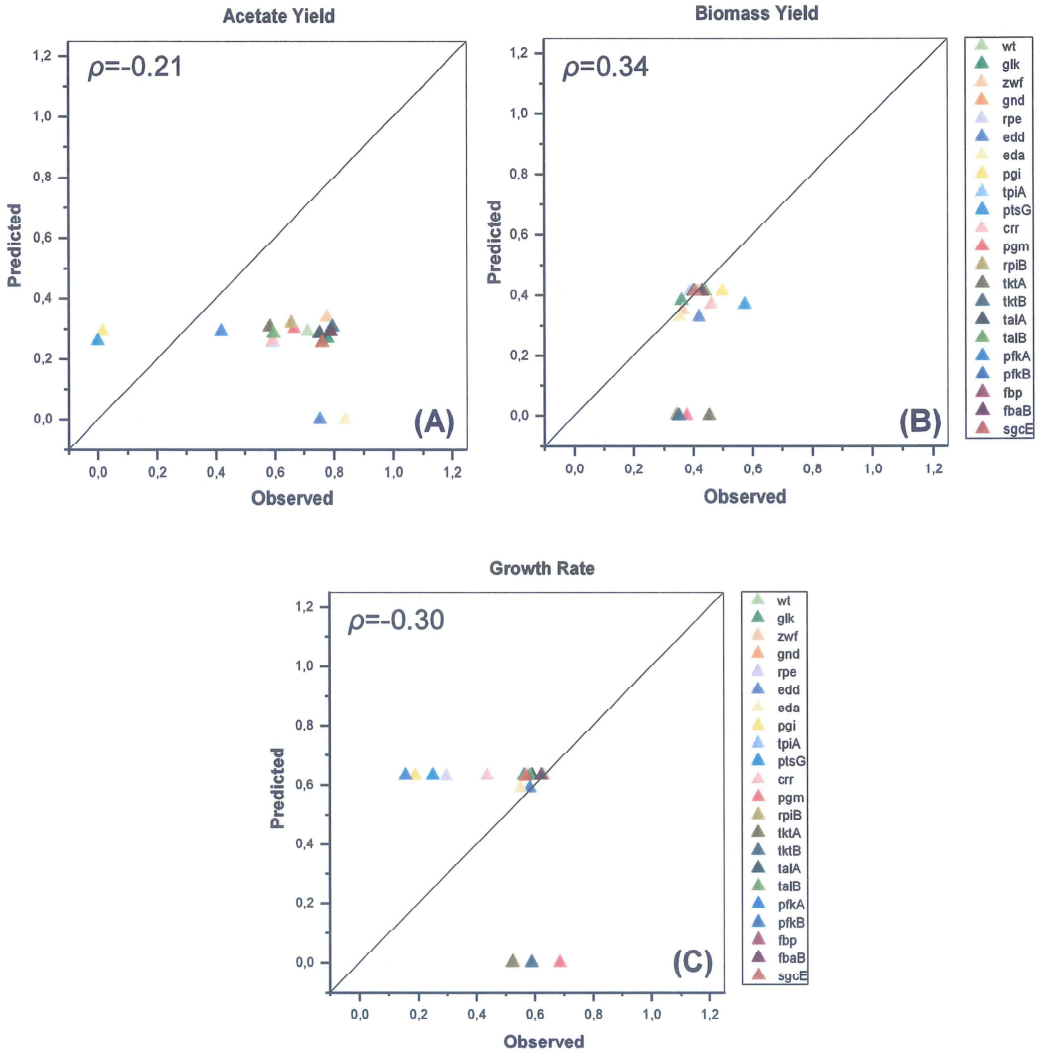

Figure 4.13 Mutant phenotype prediction of unconstrained iML1515 model via GECKO-MiMBI-Enzyme method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose uptake rate is fixed to 8.5 mmol/g_{DW}h. The total usage amount equals to 29.65 µmol/g_{DW}.

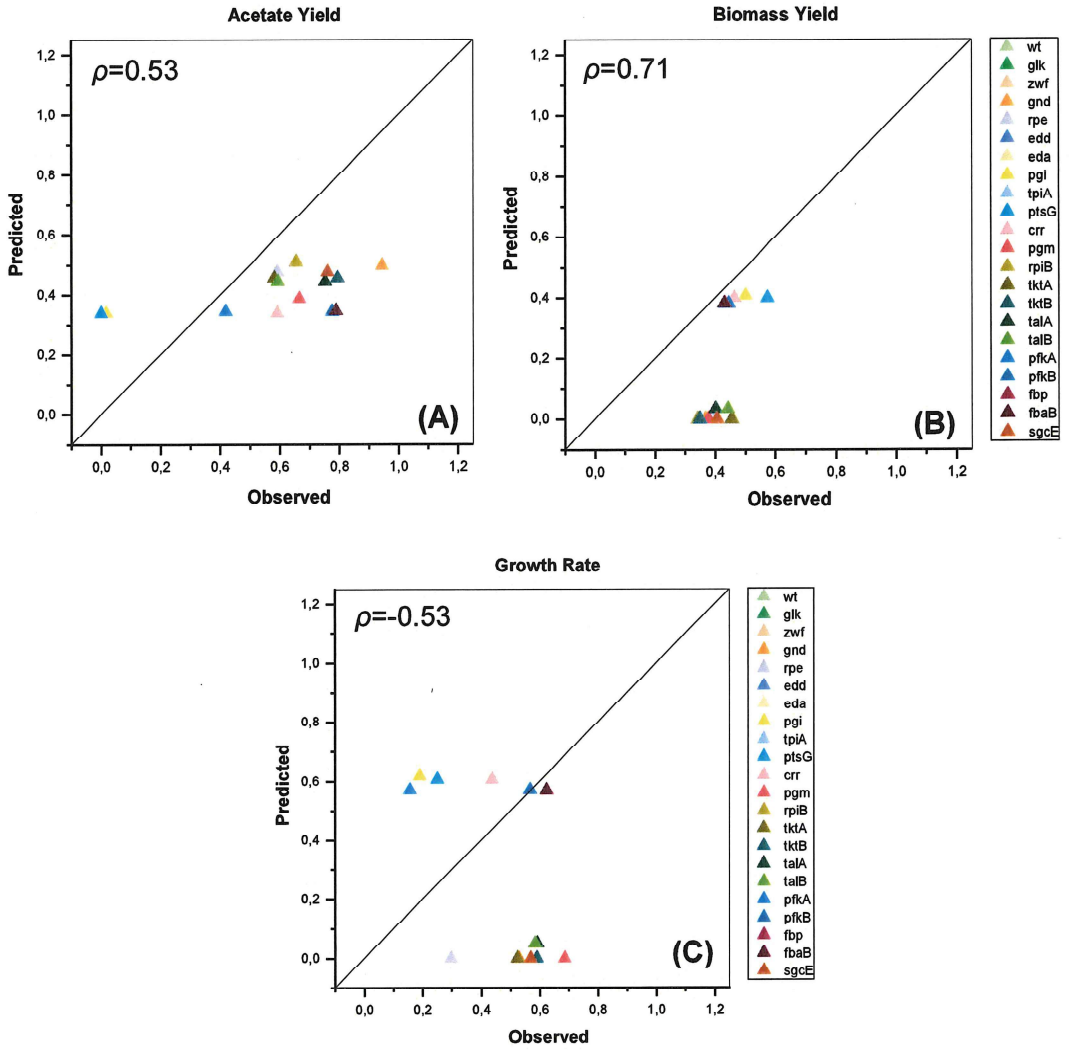

Figure 4.14 Mutant phenotype prediction of unconstrained iML1515 model via GECKO-MiMBI-Reaction method by using enzyme mass concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose uptake rate is fixed to 8.5 mmol/g_{DW}h. The total usage amount equals to 2163.5 mg/g_{DW}.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Enzyme method by using enzyme mass concentration were 0.29, 0.20 and -0.25, respectively (Figure 4.15.A, Figure 4.15.B, Figure 4.15.C).

In the GECKO-MiMBI-Enzyme method by using enzyme mass concentration, biomass yield predictions of knockouts except *ΔtktA*, *ΔtktB*, *Δpgm* and *ΔrpiB* knockouts were close to experimental data (Figure 4.15.B). In addition, highly accurate biomass yields of *Δfbp* and *Δedd* knockouts were predicted (Figure 4.15.B). Acetate yields of *ΔpfkA*, *ΔtktB*, *Δgnd* and *ΔtalA* knockouts were close to experimental data (Figure 4.15.A). Since biomass yield of *ΔtktB* was zero, although acetate yield predictions were good, overall phenotype was falsely predicted.

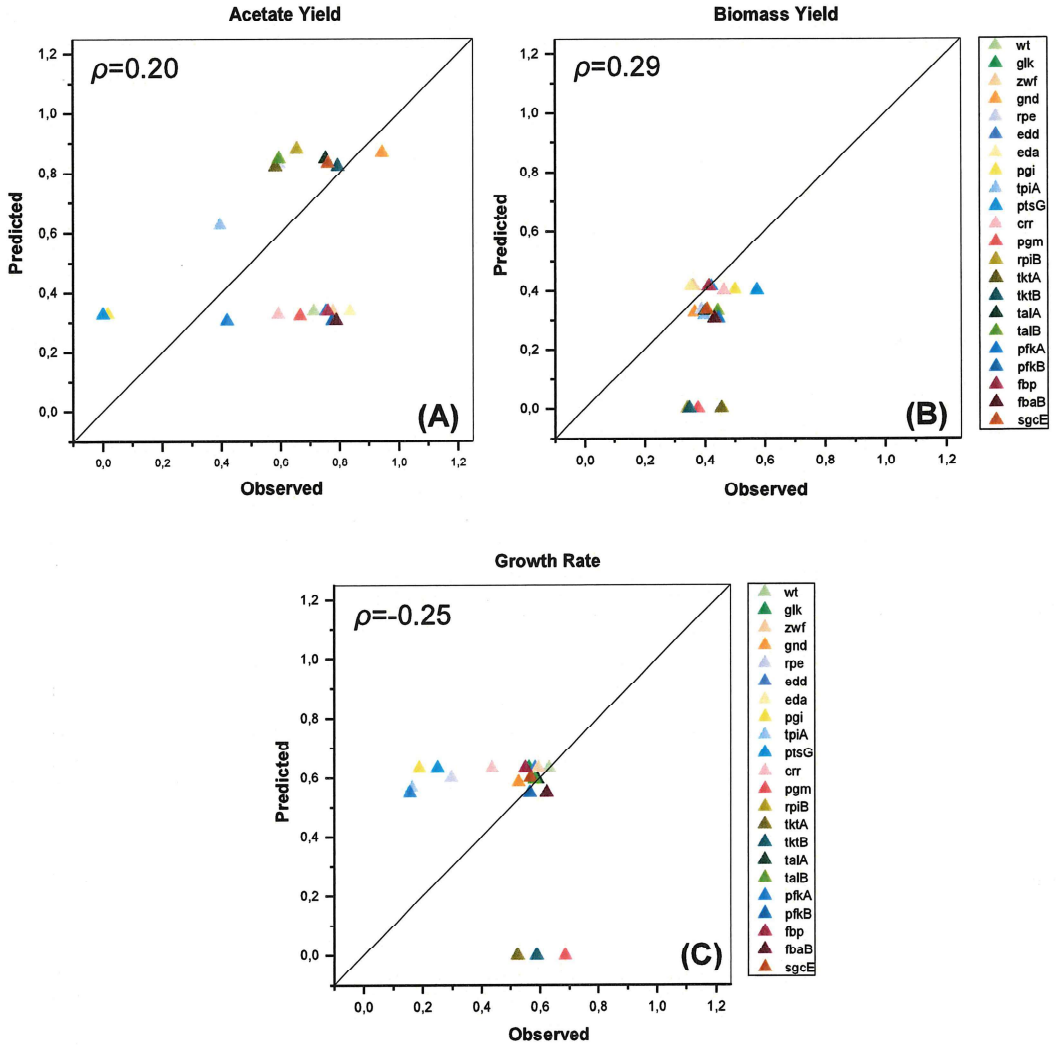

Figure 4.15 Mutant phenotype prediction of unconstrained iML1515 model via GECKO-MiMBI-Enzyme method by using enzyme mass concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate of *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose uptake rate is fixed to 8.5 mmol/g_{DW}h. The total usage amount equals 2163.5 mg/g_{DW}.

Table 4.3 Biomass yield results of mutants phenotype predictions of unconstrained iML1515 model by modeling methods with the combination of GECKO, compared with the experimental data (Exp.) from [68].

Biomass Yield								
	Enzyme Usage Molar Concentration				Enzyme Usage Mass Concentration			
Gene	GECKO -FBA	GECKO -MOMA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	GECKO -FBA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	Exp.
wt	0.414	0.414	0.414	0.414	0.414	--	0.414	0.415
glk	0.414	0.41	0.413	0.382	0.414	--	0.414	0.359
zwf	0.414	--	0.258	0.352	0.414	--	0.414	0.363
gnd	0.414	0.408	0.395	--	0.412	0	0.326	0.366
rpe	0.414	0.407	0.405	0.413	0.412	0	0.334	0.388
edd	0.411	0.256	0	0.328	0.414	--	0.414	0.419
eda	0.411	--	0	0.328	0.414	--	0.414	0.353
pgi	0.414	0.411	0.414	0.414	0.414	0.407	0.402	0.499
tpiA	0.414	0.411	0.395	--	0.41	--	0.315	0.398
ptsG	0.414	0.383	0.348	0.37	0.414	0.4	0.401	0.573
crr	0.414	0.383	0.348	0.37	0.414	0.4	0.401	0.462
pgm	0	0	0	0	0	0	0	0.377
rpiB	0.045	--	0	0.003	0.024	0	0	0.342
tktA	0	0	0	0	0	0	0	0.455
tktB	0	0	0	0	0	0	0	0.349
talA	0.414	--	0.412	0.413	0.412	0.033	0.33	0.401
talB	0.414	--	0.412	0.413	0.412	0.033	0.33	0.441
pfkA	0.414	0.414	0.414	0.414	0.41	0.382	0.305	0.431
pfkB	0.414	0.414	0.414	--	0.41	0.382	0.305	0.444
fbp	0.414	0.414	0.414	--	0.414	--	0.414	0.412
fbaB	0.414	0.414	0.414	0.414	0.41	0.383	0.305	0.431
sgcE	0.414	0.407	0.405	0.413	0.412	0	0.334	0.406

Table 4.4 Acetate yield results of mutants phenotype predictions of unconstrained iML1515 model by modeling methods with the combination of GECKO, compared with the experimental data (Exp.) from [68].

Acetate Yield								
	Enzyme Usage Molar Concentration				Enzyme Usage Mass Concentration			
Gene	GECKO -FBA	GECKO -MOMA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	GECKO -FBA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	Exp.
wt	0.292	0.291	0.292	0.292	0.337	--	0.337	0.714
glk	0.28	0.29	0.292	0.269	0.337	--	0.337	0.779
zwf	0.204	--	0.348	0.339	0.337	--	0.337	0.778
gnd	0.298	0.293	0.292	--	0.423	0.498	0.866	0.946
rpe	0.286	0.288	0.295	0.255	0.424	0.477	0.832	0.594
edd	0.47	0.152	0	0	0.337	--	0.337	0.755
eda	0.47	--	0	0	0.337	--	0.337	0.839
pgi	0.285	0.29	0.292	0.292	0.315	0.339	0.328	0.017
tpiA	0.296	0.291	0.291	--	0.297	--	0.626	0.397
ptsG	0.272	0.287	0.294	0.261	0.333	0.34	0.326	0.000
crr	0.272	0.287	0.294	0.261	0.333	0.34	0.326	0.594
pgm	1.304	0.271	0.312	0.301	2	0.388	0.322	0.667
rpiB	1.44	--	0.318	0.319	1.705	0.51	0.878	0.657
tktA	0	0.258	0.318	0.306	2	0.456	0.823	0.584
tktB	0	0.258	0.318	0.306	2	0.456	0.823	0.796
talA	0.291	--	0.292	0.285	0.423	0.445	0.848	0.754
talB	0.291	--	0.292	0.285	0.423	0.445	0.848	0.596
pfkA	0.292	0.292	0.292	0.292	0.297	0.345	0.305	0.420
pfkB	0.292	0.292	0.292	--	0.297	0.345	0.305	0.777
fbp	0.292	0.291	0.292	--	0.337	--	0.337	0.760
fbaB	0.292	0.292	0.292	0.292	0.297	0.347	0.305	0.790
sgcE	0.286	0.288	0.295	0.255	0.424	0.477	0.832	0.762

In summary, mutant phenotype prediction results were obtained using a combination of modelling strategies with GECKO method. The results from two different types of total enzyme concentrations were given when nutrient uptakes were not constrained for mutants. Within these seven combined methods, the best method was GECKO-MiMBI-Reaction-Mass for acetate yield, biomass yield and growth rate predictions according to their PCC values. In previous studies [68, 141, 142], for different engineered *E. coli* strains and several mutants of *E. coli*, PCC values for FBA were varied in 0.10-0.30 range and for MOMA 0.20-0.50. Therefore, these modelling strategies in this section except GECKO-MiMBI-Reaction-Mass show similar accuracy with the previous studies according to PCC values. It can be clearly seen from PCC values that GECKO-MiMBI-Reaction-Mass strategy was better than these previous studies. GECKO-FBA predictions with both type of enzyme concentration showed difference for only acetate yield predictions. Based on the PCC values, it can be said that enzyme mass concentration was better than molar concentration for acetate yield predictions when GECKO-FBA were applied.

However, the results did not match completely with the experimental data. As a further study (in Section 4.5 and 4.6), mutant phenotypes were predicted by supplying more information about the mutants. In this way, it should be deduced if a certain minimum of mutant data need to be supplied to facilitate accurate predictions.

4.5. Prediction of Mutant Phenotypes Using CAFBA and the constrained iML1515 Model

Phenotypes of mutants of *E. coli* K-12 MG1655 were predicted via metabolic modeling strategies, which were combined with a model modified from the iML1515 model according to the GECKO method. Predictions were done with the constrained model. In the constrained model, the experimental data was used to constrain substrate uptake rates for each mutant. For knockout predictions, as a starting point 21 genes (without ybhE) are selected which are same genes from the previous study [68]. Also, experimental results for growth and acetate secretion rate (in order to compare with predictions) were taken.

According to the results of Long et al. 2016 [68], for 8.5 mmol/g_{DW}h as glucose uptake rate of wild-type of *E. coli* K-12 MG1655, μ is 0.63 h^{-1}. In the prediction strategies with constrained allocation method, w_C, w_E and <w> values were set to $2x10^{-3}$ g_{DW}h/mmol, $1.3x10^{-3}$ g_{DW}h/mmol and $1.5x10^{-3}$ g_{DW}h/mmol respectively to reach the same glucose uptake and μ as these experimental values. As these values were used in the predictions with the unconstrained model, same values were used in the predictions with constrained model.

4.5.1. CA-FBA Predictions

Mutant phenotype predictions via CA-FBA strategy were applied by modifying model according to constrained allocation method as homogenous and heterogenous cases, separately. In both cases, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are given in Figure 4.16, Figure 4.17 and Table 4.5 and Table 4.6 and also all results are given in Table D.1 and Table D.2.

PCC values of biomas yield, acetate yield and growth rate predictions of CA-FBA method homogeneous case were 0.30, 0.54 and 0.44 respectively. In the heterogenous case, PCC values of biomas yield, acetate yield and growth rate predictions were 0.26, 0.45 and 0.63 respectively. PCC values of biomass and acetate yield predictions from Long et al., 2016 [68] were 0.17, -0.12 and 0.43 respectively. Therefore, prediction results by homogenous case of CA-FBA method were calculated better than the FBA predictions in the previous study [68] according to PCC values for these three type of predictions, especially acetate yield predictions (Figure 4.16.A, Figure 4.16.B, Figure 4.16.C). Prediction results by heterogenous case of CA-FBA method were calculated better than the FBA predictions in the previous study [68] according to PCC values for these three type of predictions, especially growth rate predictions (Figure 4.17.A, Figure 4.17.B, Figure 4.17.C).

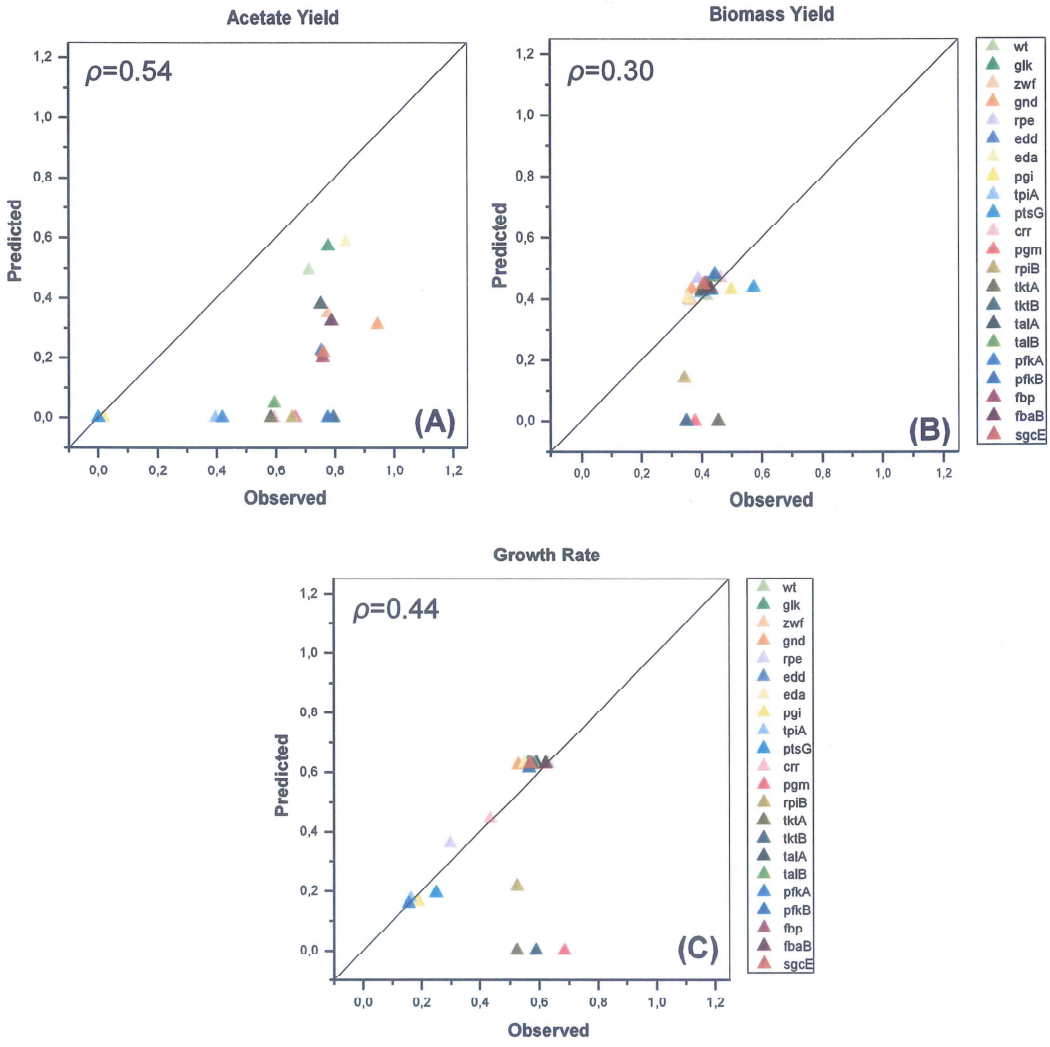

Figure 4.16 Mutant phenotype prediction of constrained iML1515 model via the homogenous case of the CA-FBA method. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data, while w_C, w_E, w_R and Φ_{max} values are $2x10^{-3}$ $g_{DW}h/mmol$, $1.3x10^{-3}$ $g_{DW}h/mmol$, 0.169 h and 0.484 respectively, are shown. Glucose and oxygen uptake rate of wild-type and single-gene deletion knockouts are fixed to the same value as experimental data during simulations are performed.

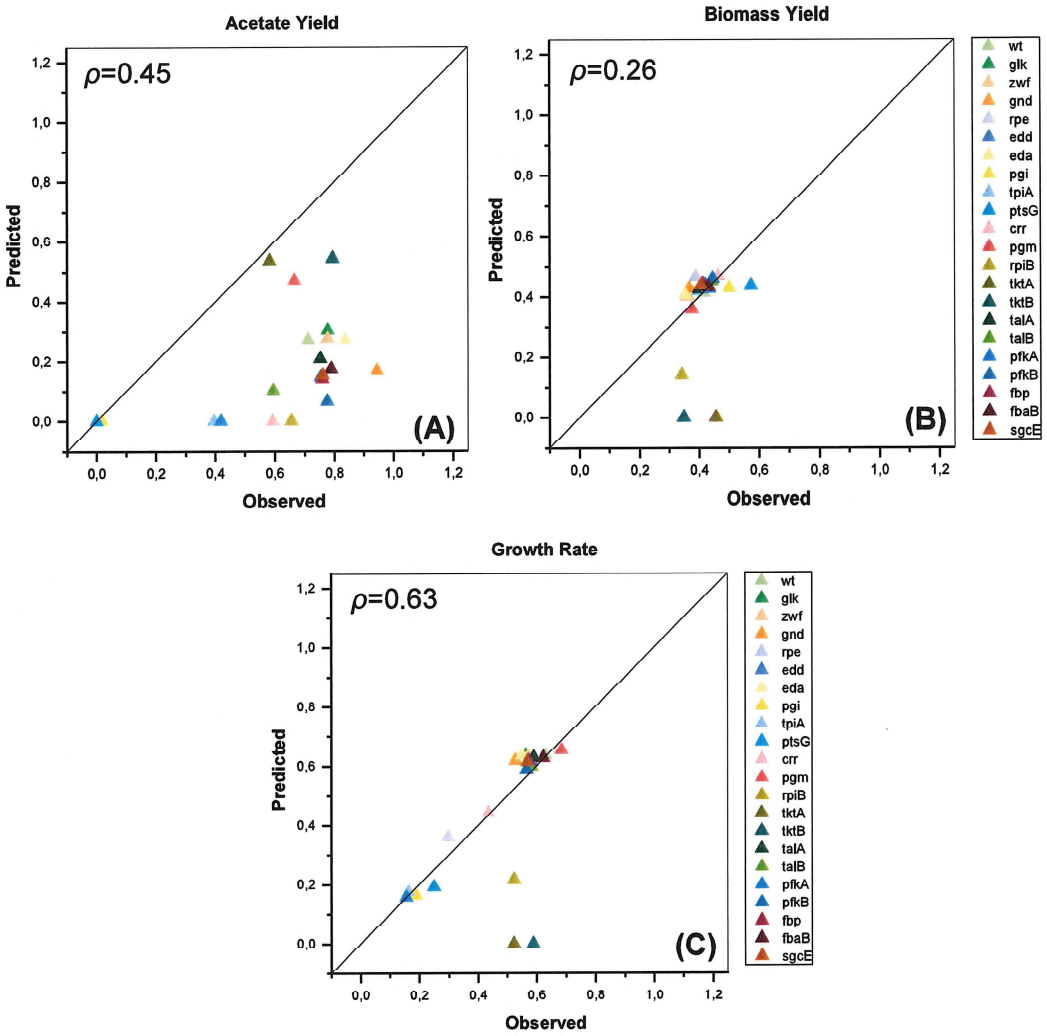

Figure 4.17 Mutant phenotype prediction of constrained iML1515 model via the heterogenous case of the CA-FBA method. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data, while w_C, <w>, w_R and Φ_{max} values are 0 $g_{DW}h/mmol$, $1.4x10^{-3}$ $g_{DW}h/mmol$, 0.169 h and 0.484 respectively. Glucose and oxygen uptake rate of wild-type and single-gene deletion knockouts are fixed to the same value as experimental data during simulations are performed.

Biomass yield predictions of most knockouts were calculated with a high accuracy according to experimental data except *ΔtktA*, *ΔtktB* and *ΔrpiB* in both cases (Figure 4.16.B, Figure 4.17.B). Highly accurate biomass yield of *Δpgm* knockout

was predicted in CA-FBA heterogenous case, even though it predicted as zero in the homogenous case. Acetate yield predictions of knockouts were calculated as similar in both cases (Figure 4.16.A, Figure 4.17.A).

4.5.2. CA-MOMA Predictions

Mutant phenotype predictions via CA-MOMA strategy were applied by modifying model according to constrained allocation method. In this strategy, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are given in Figure 4.18, Table 4.5 and Table 4.6 and also all results are given in Table D.3.

PCC values of biomass yield predictions of CA-MOMA method and experimental data were -0.32 and 0.07, respectively. Therefore, biomass yield predictions by CA-MOMA method were more accurate than the MOMA predictions in the previous study [68] according to PCC values (Figure 4.18.B). However, except *Δglk*, *Δzwf*, *Δeda*, *ΔtalA* knockouts and wild-type, the biomass yield predictions of most knockouts, CA-MOMA (using constrained model) (Figure 4.18.B) did not predict with high accuracy. In spite of, its PCC value is not lower than many of them. This can be due to principle of Pearson correlation phenomena mentioned in second paragraph of Section 4. PCC values of acetate yield and growth rate predictions were 0.76 and 0.54, respectively. PCC values of of acetate yield and growth rate predictions from Long et al., 2016 [68] were 0.24 and 0.51, respectively (Figure 4.18.A, Figure 4.18.C). According to the PCC values, while the growth rate predictions were close to the MOMA predictions in the previous study [68], acetate yield estimates were significantly better than the previous study [68] (Figure 4.18.A, Figure 4.18.C).

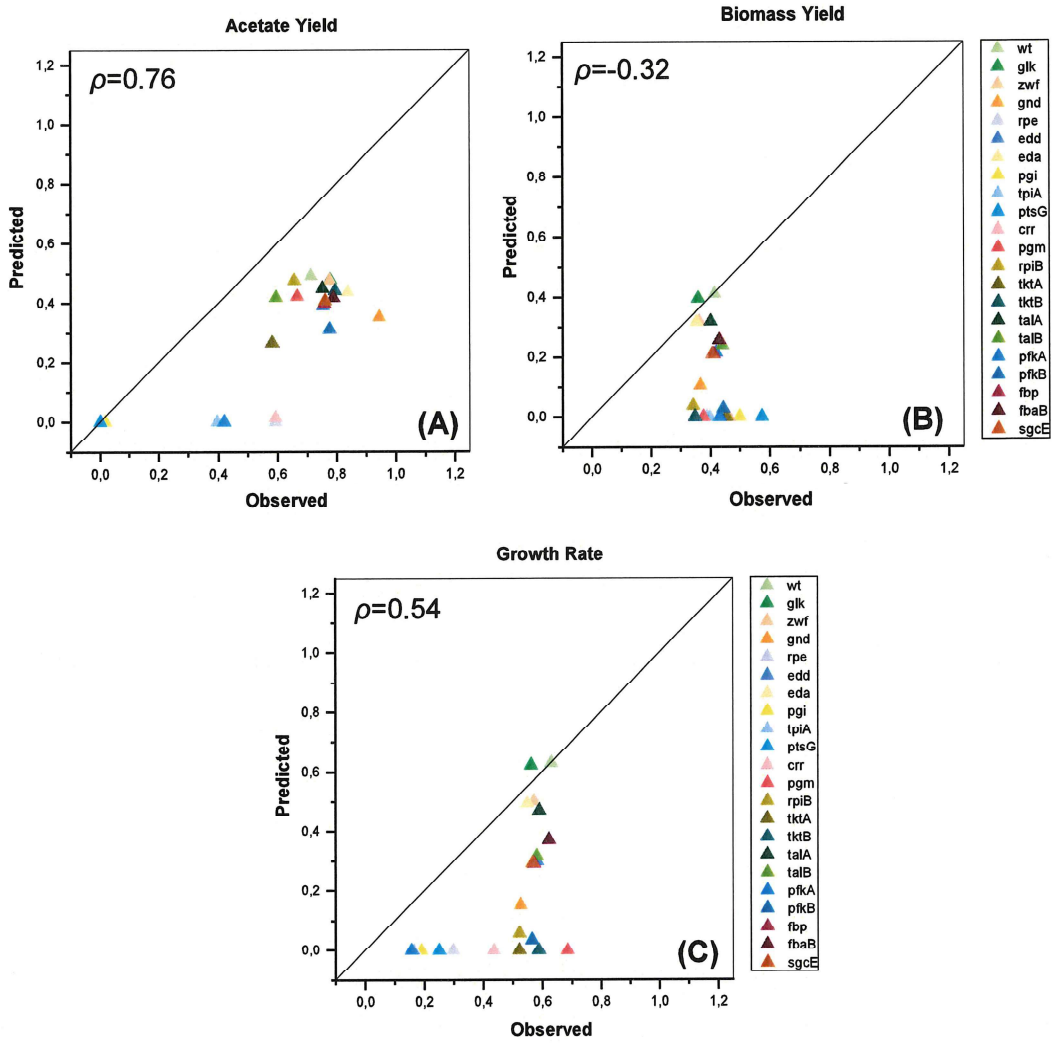

Figure 4.18 Mutant phenotype prediction of constrained iML1515 model via CA-MOMA method. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. FBA simulations of reference wild-type and single-gene deletion knockouts were done by CA-FBA while w_C, w_E, w_R and Φmax values are $2x10^{-3}$ gDWh/mmol, $1.3x10^{-3}$ gDWh/mmol, 0.169 h and 0.484 respectively. Glucose and oxygen uptake rate of wild-type and single-gene deletion knockouts are fixed to the same value as experimental data during simulations are performed.

4.5.3. CA-MiMBI Predictions

Mutant phenotype predictions via CA-MiMBI strategy were applied by modifying the model according to constrained allocation method. In this strategy, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are given in Figure 4.19, Table 4.5 and Table 4.6 and also all results are given in Table D.4.

PCC values of biomass yield, acetate yield and growth rate predictions of CA-MiMBI method were 0.60, -0.17 and 0.61, respectively (Figure 4.19.A, Figure 4.19.B, Figure 4.19.C) .

Accurate acetate yield predictions for *Δpgi*, *ΔptsG*, *ΔtalB*, *ΔtpiA* and *ΔpfkA* knockouts in CA-MiMBI with the constrained model (Figure 4.19.A) were calculated. The prediction of biomass yield predictions of approximately half of the knockouts were accurate according to experimental data in CA-MiMBI with constrained model (Figure 4.19.B) strategies.

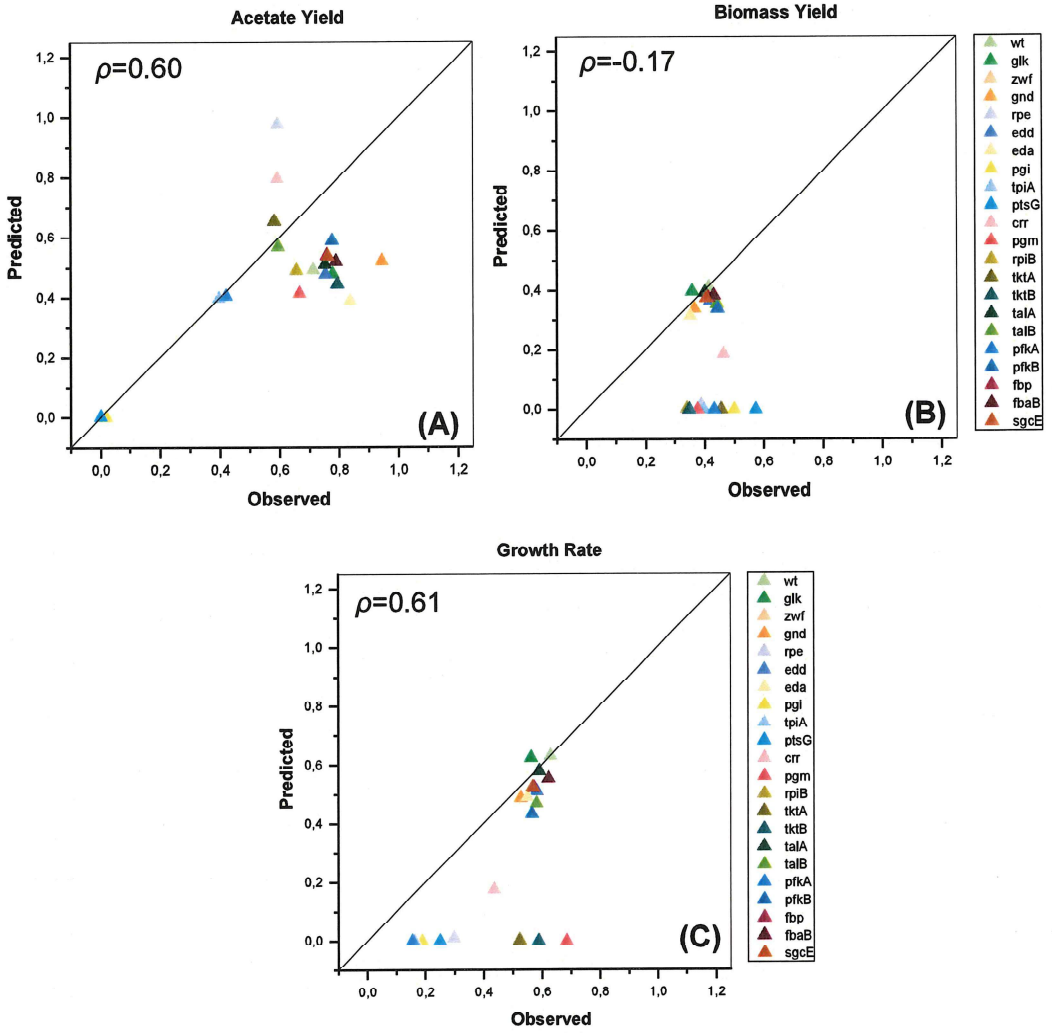

Figure 4.19 Mutant phenotype prediction of constrained iML1515 model via CA-MiMBI method. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by CA-FBA while w_C, w_E, w_R, Φ_{max} values, glucose and oxygen uptake rate are 2×10^{-3} $g_{DW}h/mmol$, 1.3×10^{-3} $g_{DW}h/mmol$, 0.169 h, 0.484, 8.5 $mmol/g_{DW}h$ and 38.74 $mmol/g_{DW}h$ respectively. Glucose and oxygen uptake rate of single-gene deletion knockouts are fixed to the same value as experimental data during simulations are performed.

Table 4.5 Biomass yield results of mutants phenotype predictions of constrained iML1515 model by modeling methods with the combination of constrained allocation, compared with the experimental data (Exp.) from [68].

Biomass Yield					
Gene	**CA-FBA-Homo**	**CA-FBA-Hete**	**CA-MOMA**	**CA-MiMBI**	**Exp.**
wt	0.412	0.415	0.412	0.412	0.415
glk	0.401	0.405	0.396	0.397	0.359
zwf	0.396	0.401	0.318	--	0.363
gnd	0.433	0.429	0.105	0.338	0.366
rpe	0.469	0.467	0	0.014	0.388
edd	0.45	0.44	0.215	0.366	0.419
eda	0.403	0.407	0.317	0.315	0.353
pgi	0.431	0.429	0	0	0.499
tpiA	0.421	0.419	0	0	0.398
ptsG	0.439	0.438	0	0	0.573
crr	0.47	0.469	0	0.186	0.462
pgm	0	0.359	0	0	0.377
rpiB	0.141	0.141	0.037	0.003	0.342
tktA	0	0	0	0	0.455
tktB	0	0	0	0	0.349
talA	0.427	0.426	0.319	0.393	0.401
talB	0.473	0.453	0.239	0.355	0.441
pfkA	0.43	0.429	0	0	0.431
pfkB	0.481	0.46	0.028	0.339	0.444
fbp	0.451	0.443	0.209	0.376	0.412
fbaB	0.435	0.434	0.256	0.383	0.431
sgcE	0.448	0.438	0.209	0.374	0.406

Table 4.6 Acetate yield results of mutants phenotype prediction of constrained iML1515 model by modeling methods with the combination of constrained allocation, compared with the experimental data (Exp.) from [68].

Acetate Yield					
Gene	**CA-FBA-Homo**	**CA-FBA-Hete**	**CA-MOMA**	**CA-MiMBI**	**Exp.**
wt	0.492	0.272	0.492	0.492	0.714
glk	0.571	0.304	0.478	0.479	0.779
zwf	0.349	0.276	0.476	--	0.778
gnd	0.312	0.168	0.353	0.521	0.946
rpe	0	0	0	0.975	0.594
edd	0.224	0.147	0.391	0.477	0.755
eda	0.588	0.272	0.437	0.389	0.839
pgi	0	0	0	0	0.017
tpiA	0	0	0	0.398	0.397
ptsG	0	0	0	0.002	0.000
crr	0	0	0.014	0.796	0.594
pgm	0	0.471	0.422	0.413	0.667
rpiB	0	0	0.476	0.491	0.657
tktA	0	0.536	0.265	0.653	0.584
tktB	0	0.544	0.441	0.445	0.796
talA	0.38	0.208	0.449	0.51	0.754
talB	0.048	0.102	0.419	0.569	0.596
pfkA	0	0	0	0.405	0.420
pfkB	0	0.065	0.312	0.588	0.777
fbp	0.201	0.141	0.397	0.54	0.760
fbaB	0.323	0.175	0.418	0.52	0.790
sgcE	0.492	0.272	0.492	0.492	0.762

In summary, mutant phenotype prediction results by applying combination of modelling strategies with CA method with two different types of total enzyme concentrations were given when nutrient uptakes were constrained for mutants. Within these four combined methods, CA-FBA heterogenous case for growth rate and CA-MOMA for both acetate yield and biomass yield predictions were the best methods according to their PCC values. In the previous studies [68, 141, 142], for different engineered *E. coli* strains and several mutants of *E. coli*, PCC values for FBA were varied in 0.10-0.30 range and for MOMA, in the 0.20-0.50 range. Therefore, these modelling strategies in this section show slightly more accurate predictions than in the previous studies according to PCC values. It can be clearly seen from PCC values that CA-MOMA for acetate yield and CA-FBA heterogenous case for growth rate were better than these previous studies.

4.6. Prediction of Mutant Phenotypes Using GECKO and the Constrained iML1515 Model

The phenotypes of mutants of *E. coli* K-12 MG1655 were predicted via metabolic modeling strategies which combined with a model modified from the iML1515 model according to GECKO method. Predictions were done with constrained model. For each mutant of constrained model, special data, which were taken from the experimental data, were defined about uptaking of nutrients. For knockout predictions, as a starting point 21 genes (without ybhE) are selected which are same genes from the previous study [68]. 'ybhE' gene was not in the iML1515 model and also in the model modified by GECKO. Also, experimental results for glucose uptake, growth and acetate secretion rate were taken from the previous study [68].

According to their results of wild-type of *E. coli* K-12 MG1655 for 8.5 mmol/g_{DW}h as glucose uptake rate, μ is 0.63 h^{-1}. In the prediction strategies with the GECKO method, total enzyme usage amount was adjusted to reach the same glucose uptake and μ as these experimental values.

4.6.1. Total Enzyme Usage Concentration Determination

As mentioned in Section 4.4.1, the total enzyme usage molar concentration and total enzyme usage mass concentration were calculated to reach experimental data of wild-type of *E. coli* K-12 MG1655. However, these values did not reach to required values, which were decided according to experimental data when the constrained model was used. For this reason, new values were calculated as 30.9 $\mu mol/g_{DW}$ for the total enzyme usage molar concentration and 2225 mg/g_{DW} for the total enzyme usage mass concentration. These values were used in strategies with the constrained model modified by the GECKO method.

4.6.2. GECKO-FBA Predictions

Mutant phenotype predictions via GECKO-FBA strategy were applied as GECKO-FBA method by using enzyme molar concentration and enzyme mass concentration, separately. In both cases, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results are given in Figure 4.20, Figure 4.21 and Table 4.7 and Table 4.8 and also all results are given in Table D.5 and Table D.6.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-FBA method by using enzyme molar concentration were 0.32, 0.02 and 0.42, respectively (Figure 4.20.A, Figure 4.20.B, Figure 4.20.C). PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-FBA method by using enzyme mass concentration were 0.32, 0.20 and 0.42, respectively (Figure 4.21.A, Figure 4.21.B, Figure 4.21.C). PCC values of biomass yield, acetate yield and growth rate predictions from Long et al., 2016 [68] were 0.17, -0.12 and 0.43, respectively. Due to these PCC values, biomass yield predictions of GECKO-FBA method by using both type of concentrations were better than the FBA predictions in the previous study [68]. Growth rate predictions of GECKO-FBA method by using both type of concentration were similar with the FBA predictions in the previous study [68] according to PCC values. Acetate yield predictions of GECKO-FBA method by using enzyme mass concentration was

better than the FBA predictions in the previous study [68] while GECKO-FBA method by using enzyme molar concentration was worse. Therefore, mass concentration supplies more accurate results than molar concentration for GECKO-FBA strategy when the constrained model was used.

Accurate biomass yield predictions for most knockouts except *ΔtktA*, *ΔtktB*, *ΔrpiB*, *Δpgm*, *Δpgi* and *ΔptsG* knockouts were calculated in both enzyme concentration and *Δpgi* and *ΔptsG* knockouts were close to experimental data in both enzyme concentration type (Figure 4.20.B, Figure 4.21.B). However, there were no acetate secretion from knockouts (Figure 4.20.A) except *ΔrpiB* knockout in GECKO-FBA method by using enzyme molar concentration. Therefore, GECKO-FBA predictions applying enzyme molar concentration are less accurate than other strategies.

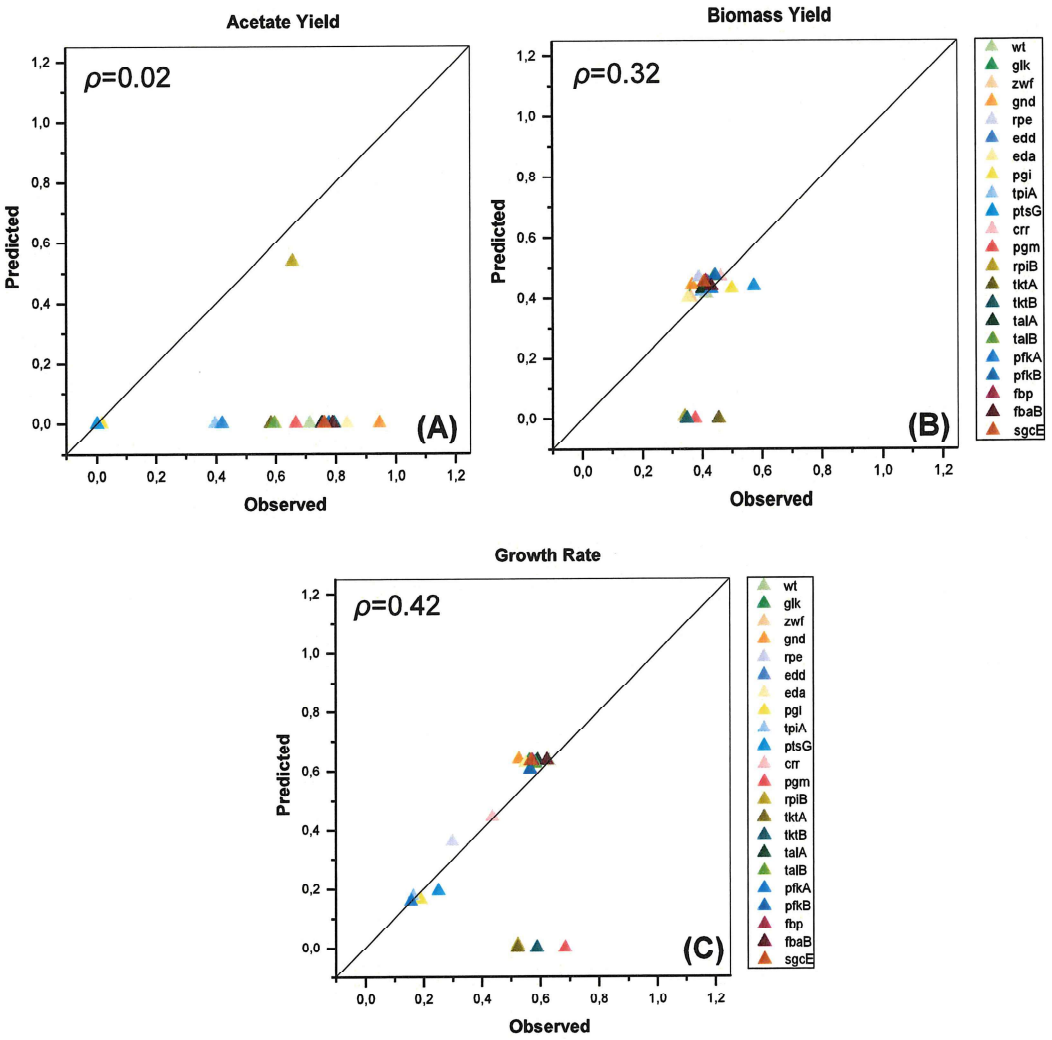

Figure 4.20 Mutant phenotype prediction of constrained iML1515 model via GECKO-FBA method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data, while total usage amount equals to 30.9 μmol/g_{DW} and glucose and oxygen uptake rate of wild-type and single-gene deletion knockouts are fixed to the same value as experimental data are shown.

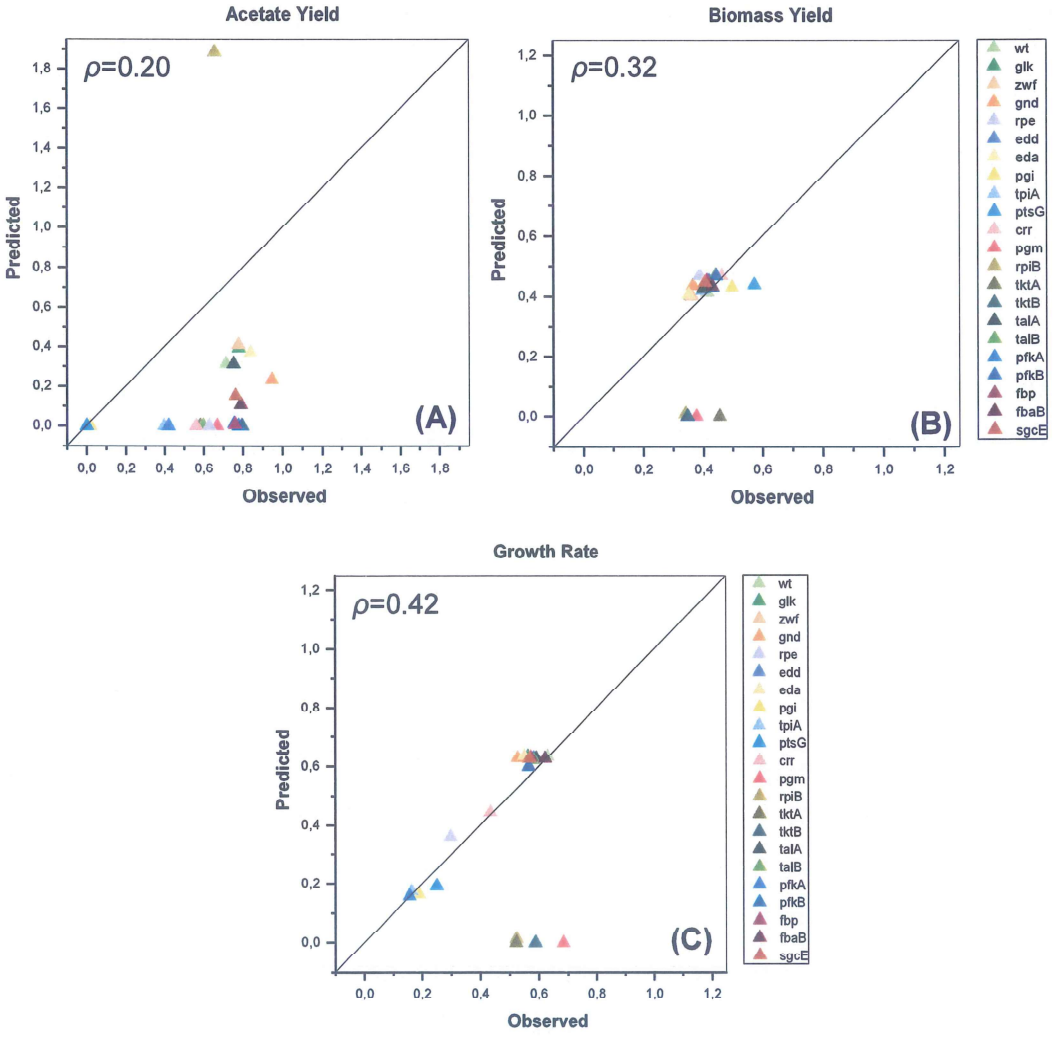

Figure 4.21 Mutant phenotype prediction of constrained iML1515 model via GECKO-FBA method by using enzyme mass concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data, while total usage amount equals to 2225 mg/g_{DW} and glucose and oxygen uptake rate of wild-type and single-gene deletion knockouts are fixed to the same value as experimental data are shown.

4.6.3. GECKO-MOMA Predictions

Mutant phenotype predictions by GECKO-MOMA strategy were applied as GECKO-MOMA method by using enzyme molar concentration and enzyme mass concentration, separately. In both cases, μ, glucose uptake rate, acetate

secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results from the GECKO-MOMA method by using enzyme molar concentration are given in Figure 4.22, Figure 4.23, Table 4.7 and Table 4.8 and also all results are given in Table D.7 and Table D.8.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MOMA method by using enzyme molar concentration were 0.17, 0 and 0.61, respectively (Figure 4.22.A, Figure 4.22.B, Figure 4.22.C). PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MOMA method by using enzyme mass concentration were 0.19, 0.49 and 0.50, respectively (Figure 4.23.A, Figure 4.23.B, Figure 4.23.C). PCC values of biomass yield, acetate yield and growth rate predictions from Long et al., 2016 [68] were 0.07, 0.24 and 0.51, respectively. According to these values, biomass yield predictions of GECKO-MOMA method by using both types of concentration were more accurate than the MOMA predictions in the previous study [68]. Growth rate predictions of GECKO-MOMA method by using enzyme molar concentration was better than the MOMA predictions in the previous study [68], while GECKO-MOMA method by using enzyme mass concentration was similar. Also, acetate yield predictions of GECKO-MOMA method by using enzyme mass concentration were better than the MOMA predictions in the previous study [68], in terms of PCC value.

In the GECKO-MOMA method by using enzyme molar concentration and by using enzyme mass concentration, biomass yield predictions of knockouts except *ΔtktA* and *ΔtktB* knockouts were close to experimental data (Figure 4.22.B, Figure 4.23.B). It can be said that these strategies were appropriate to predict biomass yields of knockouts. However, there were no acetate secretion from knockouts (Figure 4.22.A) in GECKO-MOMA method by using enzyme molar concentration. Therefore, the acetate yield predictions by GECKO-MOMA applying enzyme molar concentrations are less accurate than other strategies.

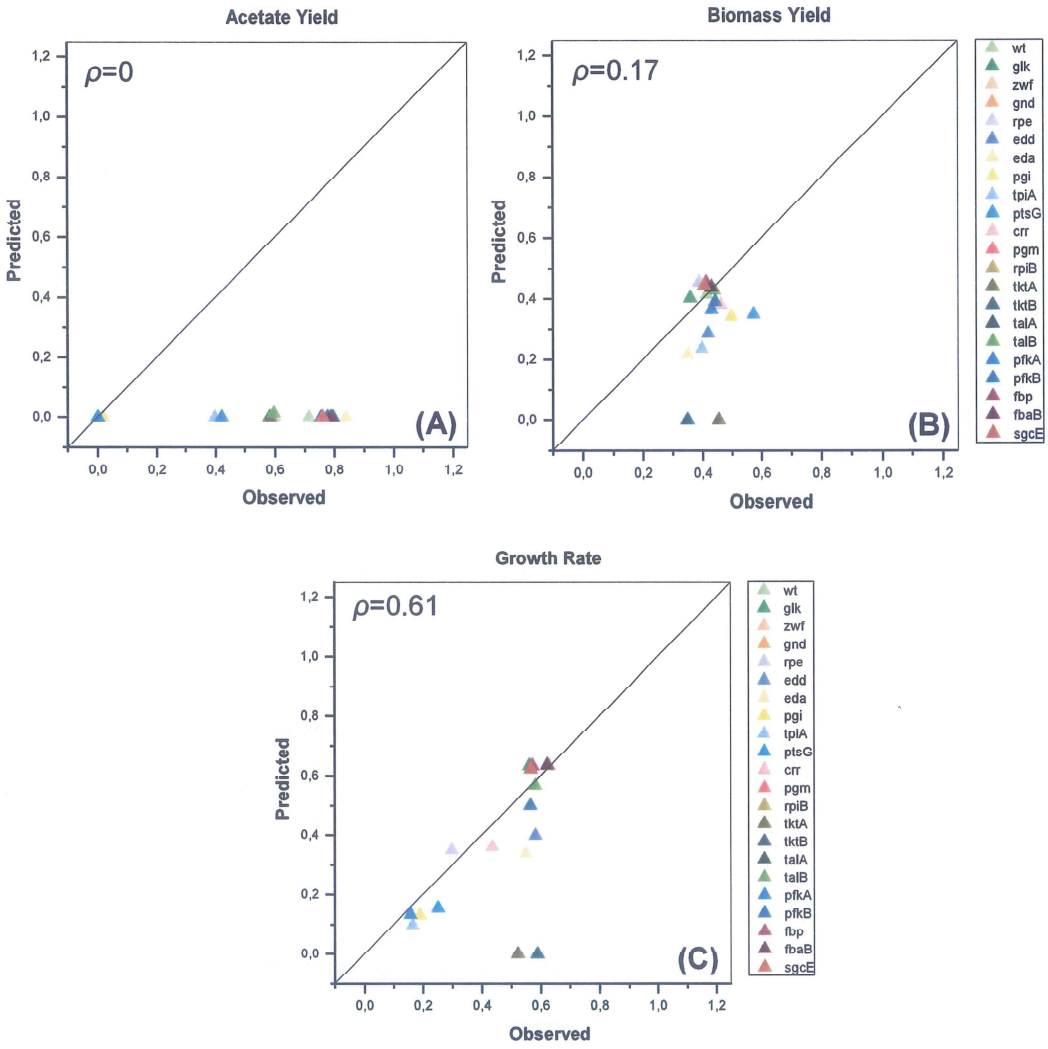

Figure 4.22 Mutant phenotype prediction of constrained iML1515 model via GECKO-MOMA method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. FBA simulations of reference wild-type and single-gene deletion knockouts were done by GECKO-FBA while total usage amount equals to 30.9 μmol/g_{DW} and glucose and oxygen uptake rate of wild-type and single-gene deletion knockouts are fixed to the same value as experimental data.

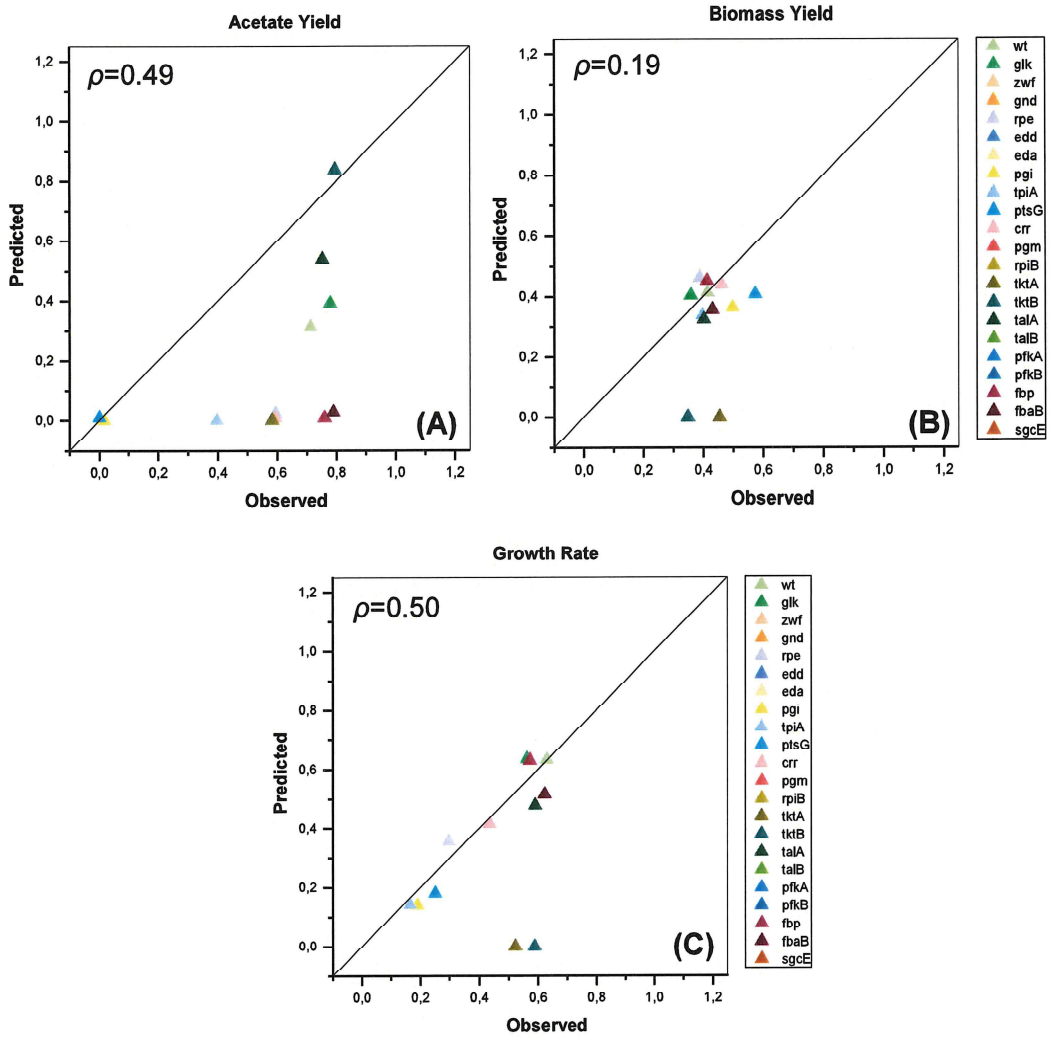

Figure 4.23 Mutant phenotype prediction of constrained iML1515 model via GECKO-MOMA method by using enzyme mass concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. FBA simulations of reference wild-type and single-gene deletion knockouts were done by GECKO-FBA while total usage amount equals to 2225 mg/g_{DW} and glucose and oxygen uptake rate of wild-type and single-gene deletion knockouts are fixed to the same value as experimental data.

4.6.4. GECKO-MiMBI Predictions

Mutant phenotype predictions by GECKO-MiMBI strategy were applied by excluding enzyme and exchange reactions (GECKO-MiMBI-Reaction) and excluding reactions except the enzyme reactions (GECKO-MiMBI-Enzyme), separately. In both methods, predictions were done via enzyme molar concentration and enzyme mass concentration, separately. In both cases, μ, glucose uptake rate, acetate secretion rate, CO_2 exchange rate, and O_2 uptake rate were predicted and biomass yields and acetate yields were calculated for the 21 selected genes. Growth rate, acetate and biomass yield results from the GECKO-MiMBI-Reactionmethod (with enzyme molar concentration) are given in Figure 4.24, results from GECKO-MiMBI-Reaction method (with enzyme mass concentration) are given in Figure 4.26. Growth rate, acetate and biomass yield results from the GECKO-MiMBI-Enzyme method (with enzyme molar concentration) are given in Figure 4.25, results from GECKO-MiMBI-Enzyme method (with enzyme mass concentration) are given in Figure 4.27. The summary of GECKO-MiMBI is given in Table 4.7 and Table 4.8 and also all results are given in Table D.9, Table D.10, Table D.11, and Table D.12.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Reaction method by using enzyme molar concentration were -0.14, 0.32 and 0.52, respectively (Figure 4.24.A, Figure 4.24.B, Figure 4.24.C). With the GECKO-MiMBI-Reaction method by using enzyme molar concentration, accurate biomass yield predictions of approximately half of knockouts were calculated (Figure 4.24.B).

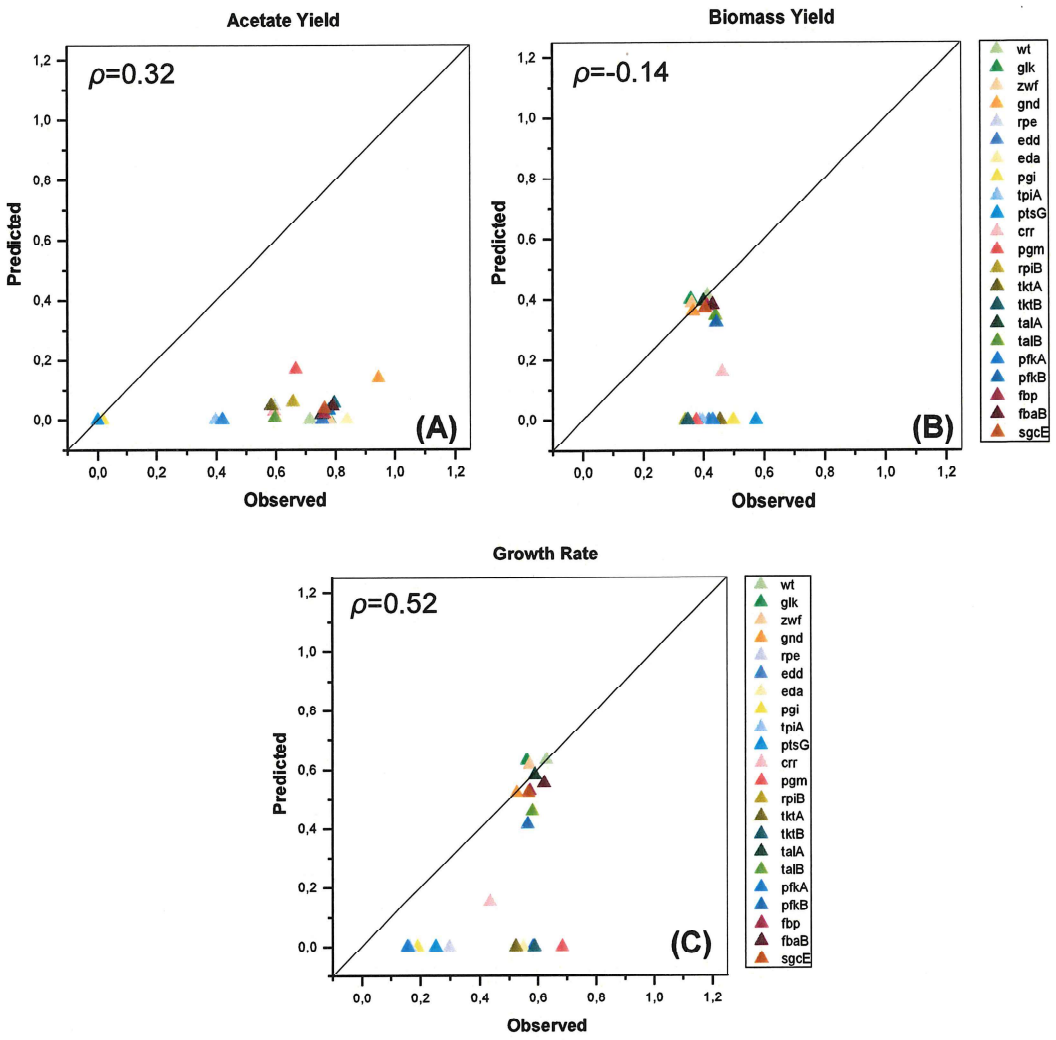

Figure 4.24 Mutant phenotype prediction of constrained iML1515 model via GECKO-MiMBI-Reaction method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate of *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose and oxygen uptake rates are 8.5 mmol/g_{DW}h and 38.74 mmol/g_{DW}h respectively. Glucose and oxygen uptake rate of single-gene deletion knockouts were fixed to the same value as experimental data during the simulations. The total usage amount equals 30.9 μmol/g_{DW}.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Enzyme method by using enzyme molar concentration were -0.22, 0.07 and 0.61, respectively (Figure 4.25.A, Figure 4.25.B, Figure 4.25.C). This strategy has the best growth rate predictions within GECKO method with using constrained model according to PCC values.

In the GECKO-MiMBI-Enzyme method by using enzyme molar concentration, highly accurate biomass yield predictions of less than half of knockouts were calculated (Figure 4.25.B). This strategy was better than the GECKO-MiMBI-Reaction method by using enzyme molar concentration in terms of biomass yield. This strategy is only good for the acetate yield in *Δpgm* knockouts (Figure 4.25.A).

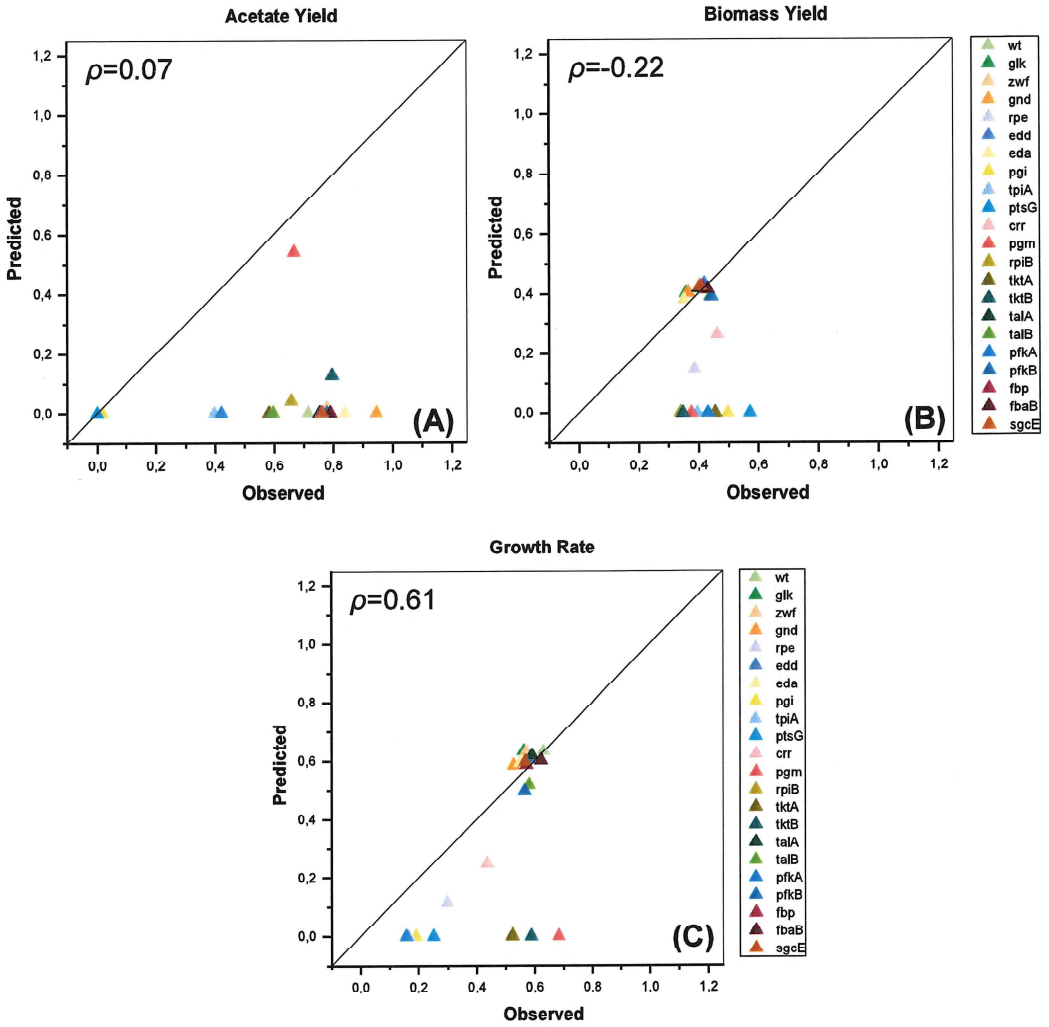

Figure 4.25 Mutant phenotype prediction of constrained iML1515 model via GECKO-MiMBI-Enzyme method by using enzyme molar concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose and oxygen uptake rates are 8.5 mmol/g_{DW}h and 38.74 mmol/g_{DW}h respectively. Glucose and oxygen uptake rate of single-gene deletion knockouts are fixed to the same value as experimental data during simulations are performed. The total usage amount equals 30.9 μmol/g_{DW}.

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Reaction method by using enzyme mass concentration were -0.20, 0.50 and 0.42, respectively (Figure 4.26.A, Figure 4.26.B, Figure 4.26.C). This strategy has the best acetate yield predictions within GECKO method with using constrained model according to PCC values.

In the GECKO-MiMBI-Reaction method by using enzyme mass concentration, accurate biomass yield predictions of less than half of knockouts were calculated (Figure 4.26.B). Predictions of acetate yields of most knockouts were improved compared to other GECKO methods and also they were close to experimental data (Figure 4.26.A).

PCC values of biomass yield, acetate yield and growth rate predictions of GECKO-MiMBI-Enzyme method by using enzyme mass concentration were -0.25, 0.44 and 0.58, respectively (Figure 4.27.A, Figure 4.27.B, Figure 4.27.C). This strategy has the best biomass yield predictions within GECKO method with using constrained model according to PCC values.

In the GECKO-MiMBI-Enzyme method by using enzyme mass concentration, biomass yield predictions of approximately half of knockouts were close to experimental data (Figure 4.27.B). In addition, highly accurate biomass yields of *Δfbp*, *Δedd* knockouts and wild-type were predicted (Figure 4.27.B). Acetate yields of *ΔsgcE*, *Δgnd*, *ΔtalA* and *ΔtalB* knockouts were close to experimental data (Figure 4.27.A).

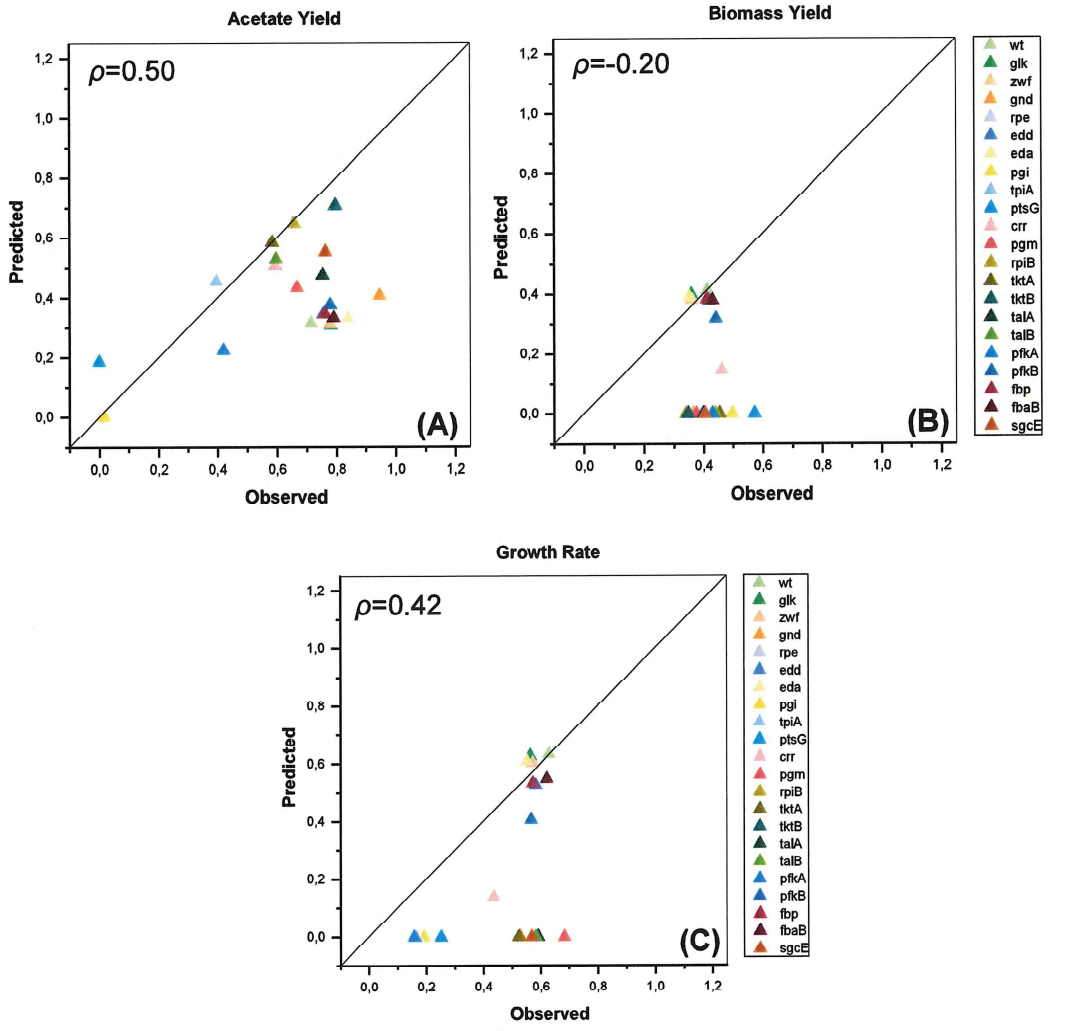

Figure 4.26 Mutant phenotype prediction of constrained iML1515 model via GECKO-MiMBI-Reaction method by using enzyme mass concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose and oxygen uptake rates are 8.5 mmol/$g_{DW}h$ and 38.74 mmol/$g_{D\ Wh}$ respectively. Glucose and oxygen uptake rate of single-gene deletion knockouts are fixed to the same value as experimental data during simulations are performed. The total usage amount equals 2225 mg/g_{DW}.

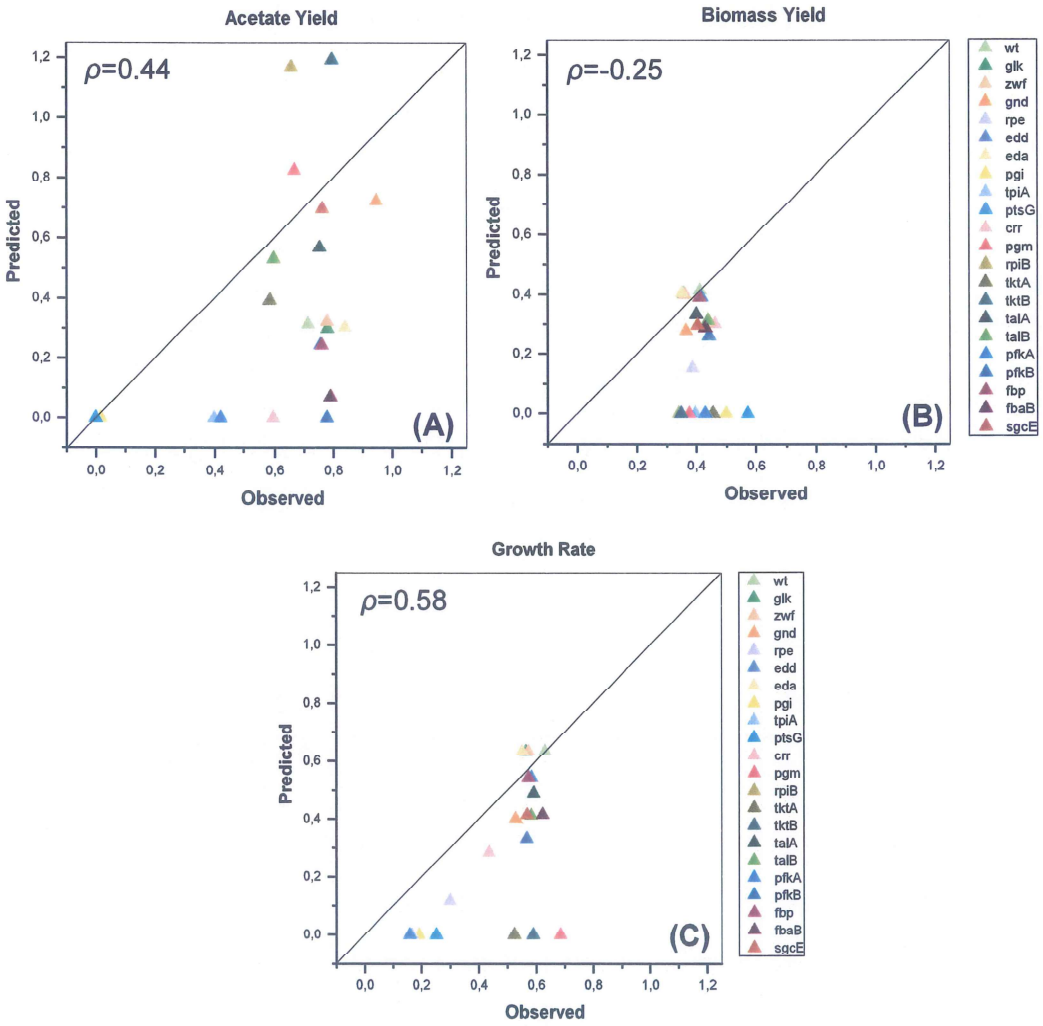

Figure 4.27 Mutant phenotype prediction of constrained iML1515 model via GECKO-MiMBI-Enzyme method by using enzyme mass concentration. The comparison of simulation results of (A) acetate yield, (B) biomass yield and (C) growth rate for *E. coli* K-12 MG1655 strain single gene deletion knockouts to experimental data are shown. Reference wild-type flux distribution was obtained by GECKO-FBA, while glucose and oxygen uptake rates are 8.5 mmol/g_{DW}h and 38.74 mmol/g_{DW}h respectively. Glucose and oxygen uptake rate of single-gene deletion knockouts are fixed to the same value as experimental data during simulations are performed. The total usage amount equals 2225 mg/g_{DW}.

Table 4.7 Biomass yield results of mutants phenotype predictions of constrained iML1515 model by modeling methods with the combination of GECKO, compared with the experimental data (Exp.) from [68].

Biomass Yield									
	Enzyme Usage Molar Concentration				Enzyme Usage Mass Concentration				
Gene	GECKO -FBA	GECKO -MOMA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	GECKO - FBA	GECKO -MOMA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	Exp.
wt	0.414	0.414	0.414	0.414	0.414	0.414	0.414	0.414	0.415
glk	0.403	0.403	0.401	0.403	0.404	0.404	0.398	0.402	0.359
zwf	0.401	--	0.39	0.4	0.402	--	0.381	0.4	0.363
gnd	0.443	--	0.362	0.405	0.436	--	0	0.278	0.366
rpe	0.469	0.455	0	0.148	0.469	0.462	--	0.153	0.388
edd	0.45	0.286	0	0.434	0.452	--	0.378	0.39	0.419
eda	0.401	0.217	0	0.38	0.406	--	0.389	0.405	0.353
pgi	0.431	0.341	0	0	0.431	0.364	0	0	0.499
tpiA	0.421	0.236	0	0	0.421	0.339	0	0	0.398
ptsG	0.439	0.349	0	0	0.439	0.408	0	0	0.573
crr	0.47	0.382	0.161	0.264	0.47	0.439	0.146	0.301	0.462
pgm	0	--	0	0	0	--	0	0	0.377
rpiB	0.007	--	0	0.003	0.009	--	0	0	0.342
tktA	0	0	0	0	0	0	0	0	0.455
tktB	0	0	0	0	0	0	0	0	0.349
talA	0.43	--	0.395	0.419	0.426	0.325	0	0.332	0.401
talB	0.472	0.429	0.347	0.392	0.47	--	0	0.312	0.441
pfkA	0.43	0.365	0	0	0.43	--	0	0	0.431
pfkB	0.474	0.391	0.325	0.389	0.469	--	0.318	0.26	0.444
fbp	0.455	0.455	0.38	0.422	0.452	0.452	0.382	0.39	0.412
fbaB	0.439	0.439	0.383	0.415	0.433	0.357	0.379	0.287	0.431
sgcE	0.452	0.444	0.372	0.426	0.447	--	0	0.296	0.406

Table 4.8 Acetate yield results of mutants phenotype predictions of constrained iML1515 model by modeling methods with the combination of GECKO, compared with the experimental data (Exp.) from [68].

Acetate Yield									
	Enzyme Usage Molar Concentration				Enzyme Usage Mass Concentration				
Gene	GECKO -FBA	GECKO -MOMA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	GECKO -FBA	GECKO -MOMA	GECKO -MiMBI -Rxn	GECKO -MiMBI -Enz	Exp.
wt	0	0	0	0	0.313	0.313	0.313	0.313	0.714
glk	0	0	0.002	0	0.392	0.392	0.307	0.298	0.779
zwf	0	--	0	0.018	0.408	--	0.311	0.321	0.778
gnd	0	--	0.139	0	0.232	--	0.406	0.722	0.946
rpe	0	0.004	0.046	0	0	0.022	--	0	0.594
edd	0	0	0	0	0.011	--	0.344	0.244	0.755
eda	0	0	0	0	0.37	--	0.33	0.302	0.839
pgi	0	0	0	0	0	0	0	0	0.017
tpiA	0	0	0	0	0	0	0.456	0	0.397
ptsG	0	0	0	0	0	0.009	0.185	0	0.000
crr	0	0	0.027	0	0	0.008	0.507	0	0.594
pgm	0	--	0.169	0.54	0	--	0.434	0.824	0.667
rpiB	0.538	--	0.059	0.041	1.884	--	0.645	1.168	0.657
tktA	0	0	0.047	0	0	0	0.584	0.393	0.584
tktB	0	0	0.055	0.126	0	0.838	0.706	1.192	0.796
talA	0	--	0.016	0	0.312	0.539	0.475	0.567	0.754
talB	0	0.014	0.007	0	0	--	0.529	0.53	0.596
pfkA	0	0	0	0	0	--	0.224	0	0.420
pfkB	0	0	0.032	0	0	--	0.375	0	0.777
fbp	0	0	0.02	0	0.007	0.007	0.344	0.243	0.760
fbaB	0	0	0.045	0	0.106	0.027	0.331	0.069	0.790
sgcE	0	0	0.036	0	0.152	--	0.551	0.697	0.762

In summary, mutant phenotype prediction results were obtained using a combination of modelling strategies with GECKO method. The results from two different types of total enzyme concentrations were given when nutrient uptakes were constrained for mutants. Within these eight combined methods, the best methods were GECKO-MiMBI-Reaction-Mass and GECKO-MOMA-Mass for acetate yield, GECKO-FBA with both type of enzyme concentration for biomass yield and GECKO-MiMBI-Enzyme-Molar and GECKO-MOMA-Molar for growth rate predictions according to their PCC values. In the previous studies [68, 141, 142], for different engineered *E. coli* strains and several mutants of *E. coli*, PCC values for FBA were varied in 0.10-0.30 range and for MOMA in the 0.20-0.50 range. Therefore, these modelling strategies in this section show slightly improved predictions than the previous studies according to PCC values. GECKO-FBA predictions with both type of enzyme concentration showed difference for only acetate yield predictions. According to the PCC values, enzyme mass concentration was better than molar concentration for acetate yield predictions when GECKO-FBA were applied. No overflow metabolism were seen in GECKO-FBA and GECKO-MOMA with total enzyme molar concentration. This situation might be because no protein limitations occurred, which would have enforced ATP production via fermentation pathways. In addition, the combined strategy could not reach to acetate yields of the experimental data for wild-type, therefore this insufficiency of prediction might have continued for the mutants.

5. CONCLUSION

Many modeling strategies have been developed for phenotype predictions, yet, there is still a need for modeling strategies for mutant phenotype predictions. Because of the lack of the currently available methods to predict accurate mutant phenotypes when only little or no information about the mutants is available, this study aimed to investigate the influence of adding protein allocation constraints to an *Escherichia coli* GEM on phenotype predictions for 21 mutant strains. Predictions were done by applying previously established, constraint-based optimization methods and supplying condition information for only the wild-type or minimal information for mutants.

In this study, FBA, MiMBI and MOMA were applied by combining with CA and GECKO methods to gain more accurate predictions. Thus, acetate yields, biomass yields and growth rates of *E. coli* mutants were predicted and compared with experimental data from Long et al., 2016 [68]. In addition, CA-FBA method from Mori et al., 2016 [10] was examined with the latest *E. coli* model (iML1515 model) for adaptation the new model to the strategy.

According to PCC values when no information about nutrient uptake of mutants were supplied, for acetate and biomass yields and growth rate predictions, the best strategy was GECKO-MiMBI-Reaction by limited enzyme with mass concentration. In addition, it was obviously better than strategies applied in the experimental study [68], potentially because it depends on metabolite balances between the wild-type and mutants unlike other strategies. By this way, it accounts for not only flux but also stoichiometry, so, by addressing the relationship between the mutant and wild-type in more detail, it supplies the closest optimal flux distribution for mutants. Therefore, it can be said that total enzyme limitation and the MiMBI strategy can supply more accurate predictions for overflow metabolism of *E. coli* and advantages for *in-silico* studies in this field. In addition, this strategy supports the aim of this thesis that mutant phenotypes can be predicted under data scarcity.

In general, in this thesis, according to the predictions applying substrate uptake rate constraints, biomass yield of mutants were higher and acetate yields were lower. Under aerobic conditions, *E. coli* produces acetate for maximization of substrate uptake and growth rate under constrained oxidative phosphorylation capacity. For this reason, when substrate uptake rate constraints for the mutants are significantly lower than the maximum substrate uptake rate used in the unconstrained model, higher oxidative phosphorylation flux can cause lower acetate yields and higher biomass yield for *E. coli* [68, 118]. Therefore, the higher biomass yields and the lower acetate yields of mutants were potentially due to this phenomena.

Some acetate yield predictions (for example; CA-FBA heterogenous case with unconstrained model, GECKO-FBA with constrained model and GECKO-MOMA with constrained model) showed no overflow metabolism. This situation might be due to the absence of protein limitation, which would have enforced ATP production via fermentation pathways. In addition, the situation might be that this combined strategy could not reach the acetate yield of the experimental data for wild-type, therefore the insufficiency of prediction might have continued for mutants.

For combined MOMA strategies with protein limitation, supplying information about nutrient uptake of mutants gave more accurate predictions, especially for growth rate. According to the general concept, MOMA does not need any information about nutrient uptake of mutants, because MOMA predicts mutant phenotypes by taking wild-type phenotype as reference. However, in this study, combined MOMA strategies gave more accurate mutant phenotype predictions when the information about nutrient uptake of mutants were given. Therefore, the results of this thesis study, where combined MOMA strategies were used, does not support this general concept.

ΔtktA and *ΔtktB* knockouts' phenotype predictions could not be made by using iML1515 model in contrast to the outdated model, which was used by Long et al., 2016 [68]. The reason of this can be the related reactions with tktA and tktB genes

producing a precursor for biomass synthesis, which is essential in the iML1515 model, but was not included in the previous model. In addition, the iML1515 model accounts for the function of 1515 metabolic genes and this value corresponds to approximately 30% of the *E. coli* genome [44, 141]. The representation of the *E. coli* metabolism can be improved for the *E. coli* GEMs in the future studies, which in turn would improve these predictions.

In conclusion, modeling strategies combined with protein limitations are more accurate than other established methods for predicting mutant phenotypes. In addition, the predictions in this study were compared with unevolved strain data. Growth phenotypes of genetically modified organisms can evolve over time. Therefore, the complete growth capabilities of the organism could not be performed with more accuracy by the initial phenotype of a newly modified organism. Adaptive evolution could supply large raises in growth rate for gene-deletion strains [144]. Therefore, as future study, growth capabilities of mutants can be increased by adaptive laboratory evolution (ALE) experiments and predictions can be compared with these experimental results to gain more reliable comparison and explanation for accuracy of modelling strategies.

REFERENCES

[1] P. Bork, Is there biological research beyond Systems Biology? A comparative analysis of terms, Molecular Systems Biology, 1 (**2005**) 2005.0012.

[2] B.J. Sanchez, C. Zhang, A. Nilsson, P.J. Lahtvee, E.J. Kerkhoven, J. Nielsen, Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints, Molecular Systems Biology, 13 (**2017**) 935.

[3] S.A. Becker, A.M. Feist, M.L. Mo, G. Hannum, B.O. Palsson, M.J. Herrgard, Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox, Nature Protocols, 2 (**2007**) 727-738.

[4] B.O. Palsson, Systems Biology: Constraint-based Reconstruction and Analysis, University Printing House, Cambridge, Chapter 1, **2015**.

[5] N.E. Lewis, H. Nagarajan, B.O. Palsson, Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods, Nature Reviews Microbiology, 10 (**2012**) 291.

[6] E.J. O'Brien, J.M. Monk, B.O. Palsson, Using Genome-scale Models to Predict Biological Capabilities, Cell, 161 (**2015**) 971-987.

[7] E.J. Kerkhoven, P.J. Lahtvee, J. Nielsen, Applications of computational modeling in metabolic engineering of yeast, Fems Yeast Research, 15 (**2015**) 1-13.

[8] J.H.S. Hofmeyr, A. Cornish-Bowden, Regulating the cellular economy of supply and demand, Febs Letters, 476 (**2000**) 47-51.

[9] J.D. Orth, I. Thiele, B.O. Palsson, What is flux balance analysis?, Nature Biotechnology, 28 (**2010**) 245-248.

[10] M. Mori, T. Hwa, O.C. Martin, A. De Martino, E. Marinari, Constrained Allocation Flux Balance Analysis, Plos Computational Biology, 12 (**2016**) e1004913.

[11] A.R. Brochado, S. Andrejev, C.D. Maranas, K.R. Patil, Impact of Stoichiometry Representation on Simulation of Genotype-Phenotype Relationships in Metabolic Networks, Plos Computational Biology, 8 (**2012**) e1002758.

[12] D. Segre, D. Vitkup, G.M. Church, Analysis of optimality in natural and perturbed metabolic networks, Proceedings of the National Academy of Sciences of the United States of America, 99 (**2002**) 15112-15117.

[13] R. Mahadevan, B.Ø. Palsson, D.R. Lovley, In situ to in silico and back: elucidating the physiology and ecology of *Geobacter spp.* using genome-scale modelling, Nature Reviews Microbiology, 9 (**2010**) 39.

[14] M.A. Oberhardt, B.O. Palsson, J.A. Papin, Applications of genome-scale metabolic reconstructions, Molecular Systems Biology, 5 (**2009**) 320.

[15] A.M. Feist, B.Ø. Palsson, The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*, Nature Biotechnology, 26 (**2008**) 659.

[16] J.A. Lerman, D.R. Hyduke, H. Latif, V.A. Portnoy, N.E. Lewis, J.D. Orth, A.C. Schrimpe-Rutledge, R.D. Smith, J.N. Adkins, K. Zengler, B.O. Palsson, In silico method for modelling metabolism and gene product expression at genome scale, Nature Communications, 3 (**2012**) 929.

[17] I. Thiele, R.M.T. Fleming, R. Que, A. Bordbar, D. Diep, B.O. Palsson, Multiscale Modeling of Metabolism and Macromolecular Synthesis in *E. coli* and Its Application to the Evolution of Codon Usage, Plos One, 7 (**2012**) e45635.

[18] E.J. O'Brien, J.A. Lerman, R.L. Chang, D.R. Hyduke, B.O. Palsson, Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction, Molecular Systems Biology, 9 (**2013**) 693.

[19] J.L. Reed, T.R. Patel, K.H. Chen, A.R. Joyce, M.K. Applebee, C.D. Herring, O.T. Bui, E.M. Knight, S.S. Fong, B.O. Palsson, Systems approach to refining genome annotation, Proceedings of the National Academy of Sciences of the United States of America, 103 (**2006**) 17480-17484.

[20] N.E. Lewis, Thiele, Ines Jamshidi, Neema Palsson, Bernhard O., Metabolic Systems Biology, **2009**, pp. 5535-5552.

[21] J.S. Edwards, R.U. Ibarra, B.O. Palsson, In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data, Nature Biotechnology, 19 (**2001**) 125-130.

[22] M.W. Covert, E.M. Knight, J.L. Reed, M.J. Herrgard, B.O. Palsson, Integrating high-throughput and computational data elucidates bacterial networks, Nature, 429 (**2004**) 92-96.

[23] S.J. Lee, D.Y. Lee, T.Y. Kim, B.H. Kim, J.W. Lee, S.Y. Lee, Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation, Applied and Environmental Microbiology, 71 (**2005**) 7880-7887.

[24] H. Alper, Y.S. Jin, J.F. Moxley, G. Stephanopoulos, Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*, Metabolic Engineering, 7 (**2005**) 155-164.

[25] J.S. Edwards, B.O. Palsson, The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities, Proceedings of the

National Academy of Sciences of the United States of America, 97 (**2000**) 5528-5533.

[26] J.L. Reed, T.D. Vo, C.H. Schilling, B.O. Palsson, An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR), Genome Biology, 4 (**2003**) R54.

[27] J.S. Edwards, R.U. Ibarra, B.O. Palsson, In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data, Nature Biotechnology, 19 (**2001**) 125.

[28] R.U. Ibarra, J.S. Edwards, B.O. Palsson, *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth, Nature, 420 (**2002**) 186-189.

[29] I. Thiele, N. Jamshidi, R.M.T. Fleming, B.O. Palsson, Genome-Scale Reconstruction of *Escherichia coli*'s Transcriptional and Translational Machinery: A Knowledge Base, Its Mathematical Formulation, and Its Functional Characterization, Plos Computational Biology, 5 (**2009**) e1000312.

[30] J.H. Park, K.H. Lee, T.Y. Kim, S.Y. Lee, Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation, Proceedings of the National Academy of Sciences of the United States of America, 104 (**2007**) 7797-7802.

[31] K.H. Lee, J.H. Park, T.Y. Kim, H.U. Kim, S.Y. Lee, Systems metabolic engineering of *Escherichia coli* for L-threonine production, Molecular Systems Biology, 3 (**2007**) 149.

[32] Q.Z. Wang, X. Chen, Y.D. Yang, X.M. Zhao, Genome-scale in silico aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production, Applied Microbiology and Biotechnology, 73 (**2006**) 887-894.

[33] P. Pharkya, C.D. Maranas, An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems, Metabolic Engineering, 8 (**2006**) 1-13.

[34] S.S. Fong, A.P. Burgard, C.D. Herring, E.M. Knight, F.R. Blattner, C.D. Maranas, B.O. Palsson, In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid, Biotechnology and Bioengineering, 91 (**2005**) 643-648.

[35] H. Alper, K. Miyaoku, G. Stephanopoulos, Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets, Nature Biotechnology, 23 (**2005**) 612-616.

[36] P. Pharkya, A.P. Burgard, C.D. Maranas, OptStrain: A computational framework for redesign of microbial production systems, Genome Research, 14 (**2004**) 2367-2376.

[37] P. Pharkya, A.P. Burgard, C.D. Maranas, Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock, Biotechnology and Bioengineering, 84 (**2003**) 887-899.

[38] A.P. Burgard, P. Pharkya, C.D. Maranas, OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization, Biotechnology and Bioengineering, 84 (**2003**) 647-657.

[39] V.S. Kumar, M.S. Dasika, C.D. Maranas, Optimization based automated curation of metabolic reconstructions, Bmc Bioinformatics, 8 (**2007**) 212.

[40] T. Fuhrer, L. Chen, U. Sauer, D. Vitkup, Computational prediction and experimental verification of the gene encoding the NAD(+)/NADP(+)-dependent succinate semialdehyde dehydrogenase in *Escherichia coli*, Journal of Bacteriology, 189 (**2007**) 8073-8078.

[41] P. Kharchenko, L.F. Chen, Y. Freund, D. Vitkup, G.M. Church, Identifying metabolic enzymes with multiple types of association evidence, Bmc Bioinformatics, 7 (**2006**) 177.

[42] L.F. Chen, D. Vitkup, Predicting genes for orphan metabolic activities using phylogenetic profiles, Genome Biology, 7 (**2006**) R17.

[43] A.E. Motter, N. Gulbahce, E. Almaas, A.L. Barabasi, Predicting synthetic rescues in metabolic networks, Molecular Systems Biology, 4 (**2008**) 168.

[44] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, B.O. Palsson, A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information, Molecular Systems Biology, 3 (**2007**) 121.

[45] A.R. Joyce, J.L. Reed, A. White, R. Edwards, A. Osterman, T. Baba, H. Mori, S.A. Lesely, B.O. Palsson, S. Agarwalla, Experimental and computational assessment of conditionally essential genes in *Escherichia coli*, Journal of Bacteriology, 188 (**2006**) 8259-8271.

[46] S. Gerdes, R. Edwards, M. Kubal, M. Fonstein, R. Stevens, A. Osterman, Essential genes on metabolic maps, Current Opinion in Biotechnology, 17 (**2006**) 448-456.

[47] C.M. Ghim, K.I. Goh, B. Kahng, Lethality and synthetic lethality in the genome-wide metabolic network of *Escherichia coli*, Journal of Theoretical Biology, 237 (**2005**) 401-411.

[48] P.J. Kim, D.Y. Lee, T.Y. Kim, K.H. Lee, H. Jeong, S.Y. Lee, S. Park, Metabolite essentiality elucidates robustness of *Escherichia coli* metabolism, Proceedings of the National Academy of Sciences of the United States of America, 104 (**2007**) 13638-13642.

[49] M. Imielinski, C. Belta, A. Halasz, H. Rubin, Investigating metabolite essentiality through genome-scale analysis of *Escherichia coli* production capabilities, Bioinformatics, 21 (**2005**) 2008-2016.

[50] R. Guimera, M. Sales-Pardo, L.A.N. Amaral, A network-based method for target selection in metabolic networks, Bioinformatics, 23 (**2007**) 1616-1622.

[51] A. Samal, S. Singh, V. Giri, S. Krishna, N. Raghuram, S. Jain, Low degree metabolites explain essential reactions and enhance modularity in biological networks, Bmc Bioinformatics, 7 (**2006**) 118.

[52] C.S. Henry, M.D. Jankowski, L.J. Broadbelt, V. Hatzimanikatis, Genome-scale thermodynamic analysis of *Escherichia coli* metabolism, Biophysical Journal, 90 (**2006**) 1453-1461.

[53] A.P. Burgard, S. Vaidyaraman, C.D. Maranas, Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments, Biotechnology Progress, 17 (**2001**) 791-797.

[54] P.B. Warren, J.L. Jones, Duality, Thermodynamics, and the linear programming problem in constraint-based models of metabolism, Physical Review Letters, 99 (**2007**) 108101.

[55] A. Hoppe, S. Hoffmann, H.G. Holzhutter, Including metabolite concentrations into flux balance analysis: Thermodynamic realizability as a constraint on flux distributions in metabolic networks, Bmc Systems Biology, 1 (**2007**) 23.

[56] C.S. Henry, L.J. Broadbelt, V. Hatzimanikatis, Thermodynamics-based metabolic flux analysis, Biophysical Journal, 92 (**2007**) 1792-1805.

[57] M. Ederer, E.D. Gilles, Thermodynamically feasible kinetic models of reaction networks, Biophysical Journal, 92 (**2007**) 1846-1857.

[58] A. Kummel, S. Panke, M. Heinemann, Systematic assignment of thermodynamic constraints in metabolic network models, Bmc Bioinformatics, 7 (**2006**) 512.

[59] A. Kummel, S. Panke, M. Heinemann, Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data, Molecular Systems Biology, 2 (**2006**) 2006.0034.

[60] D.A. Beard, S.C. Liang, H. Qian, Energy balance for analysis of complex metabolic networks, Biophysical Journal, 83 (**2002**) 79-86.

[61] A.P. Burgard, E.V. Nikolaev, C.H. Schilling, C.D. Maranas, Flux coupling analysis of genome-scale metabolic network reconstructions, Genome Research, 14 (**2004**) 301-312.

[62] C. Pal, B. Papp, M.J. Lercher, Adaptive evolution of bacterial metabolic networks by horizontal gene transfer, Nature Genetics, 37 (**2005**) 1372-1375.

[63] C. Pal, B. Papp, M.J. Lercher, Horizontal gene transfer depends on gene content of the host, Bioinformatics, 21 (**2005**) 222-223.

[64] C. Pal, B. Papp, M.J. Lercher, P. Csermely, S.G. Oliver, L.D. Hurst, Chance and necessity in the evolution of minimal metabolic networks, Nature, 440 (**2006**) 667-670.

[65] J.M. Monk, C.J. Lloyd, E. Brunk, N. Mih, A. Sastry, Z. King, R. Takeuchi, W. Nomura, Z. Zhang, H. Mori, A.M. Feist, B.O. Palsson, iML1515, a knowledgebase that computes *Escherichia coli* traits, Nature Biotechnology, 35 (**2017**) 904-908.

[66] C.P. Long, M.R. Antoniewicz, Metabolic flux analysis of *Escherichia coli* knockouts: lessons from the Keio collection and future outlook, Current Opinion in Biotechnology, 28 (**2014**) 127-133.

[67] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, Molecular Systems Biology, 2 (**2006**) 2006.0008.

[68] C.P. Long, J.E. Gonzalez, N.R. Sandoval, M.R. Antoniewicz, Characterization of physiological responses to 22 gene knockouts in *Escherichia coli* central carbon metabolism, Metabolic Engineering, 37 (**2016**) 102-113.

[69] J.E. Bailey, Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities, Biotechnology Progress, 14 (**1998**) 8-20.

[70] B. Novak, A. Toth, A. Csikasz-Nagy, B. Gyorffy, J.J. Tyson, K. Nasmyth, Finishing the cell cycle, J Theor Biol, 199 (**1999**) 223-233.

[71] H.H. McAdams, A. Arkin, It's a noisy business! Genetic regulation at the nanomolar scale, Trends in genetics : TIG, 15 (**1999**) 65-69.

[72] N. Barkai, S. Leibler, Robustness in simple biochemical networks, Nature, 387 (**1997**) 913-917.

[73] D. Fell, Understanding the control of metabolism, Portland Press, London, **1996**.

[74] B.O. Palsson, I.D. Lee, Model complexity has a significant effect on the numerical value and interpretation of metabolic sensitivity coefficients, Journal of Theoretical Biology, 161 (**1993**) 299-315.

[75] J.C. Liao, Modelling and analysis of metabolic pathways, Current Opinion in Biotechnology, 4 (**1993**) 211-216.

[76] B.O. Palsson, A. Joshi, S.S. Ozturk, Reducing complexity in metabolic networks: making metabolic meshes manageable, Federation proceedings, 46 (**1987**) 2485-2489.

[77] D.S. Kompala, D. Ramkrishna, N.B. Jansen, G.T. Tsao, Investigation of bacterial-growth on mixed substrates - experimental evaluation of cybernetic models, Biotechnology and Bioengineering, 28 (**1986**) 1044-1055.

[78] M.M. Domach, S.K. Leung, R.E. Cahn, G.G. Cocks, M.L. Shuler, Computer-model for glucose-limited growth of a single cell of *Escherichia coli* b/r-a, Biotechnology and Bioengineering, 26 (**1984**) 203-216.

[79] N. Jamashidi, Edwards, J., Fahland, T., Church, G. & Palsson, B., A computer model of human red blood cell metabolism, Bioinformatics, (**2001**).

[80] P.J. Mulquiney, W.A. Bubb, P.W. Kuchel, Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: in vivo kinetic characterization of 2,3-bisphosphoglycerate synthase/phosphatase using 13C and 31P NMR, The Biochemical journal, 342 Pt 3 (**1999**) 567-580.

[81] I.D. Lee, B.O. Palsson, A Macintosh software package for simulation of human red blood cell metabolism, Computer methods and programs in biomedicine, 38 (**1992**) 195-226.

[82] T.D. Vo, H.J. Greenberg, B.O. Palsson, Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data, Journal of Biological Chemistry, 279 (**2004**) 39532-39540.

[83] J.L. Reed, B.O. Palsson, Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: Assessment of correlated reaction subsets that comprise network states, Genome Research, 14 (**2004**) 1797-1805.

[84] R. Mahadevan, C.H. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models, Metabolic Engineering, 5 (**2003**) 264-276.

[85] S. Lee, C. Phalakornkule, M.M. Domach, I.E. Grossmann, Recursive MILP model for finding all the alternate optima in LP models for metabolic networks, Computers & Chemical Engineering, 24 (**2000**) 711-716.

[86] A. Varma, B.O. Palsson, Metabolic capabilities of *Escherichia coli* .1. synthesis of biosynthetic precursors and cofactors, Journal of Theoretical Biology, 165 (**1993**) 477-502.

[87] J.A. Papin, J. Stelling, N.D. Price, S. Klamt, S. Schuster, B.O. Palsson, Comparison of network-based pathway analysis methods, Trends in Biotechnology, 22 (**2004**) 400-405.

[88] S. Schuster, C. Hilgetag, J.H. Woods, D.A. Fell, Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism, Journal of Mathematical Biology, 45 (**2002**) 153-181.

[89] C.H. Schilling, D. Letscher, B.O. Palsson, Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from? A pathway-oriented perspective, Journal of Theoretical Biology, 203 (**2000**) 229-248.

[90] A. Varma, B.O. Palsson, Metabolic capabilities of *Escherichia coli* .2. optimal-growth patterns, Journal of Theoretical Biology, 165 (**1993**) 503-522.

[91] J.M. Savinell, B.O. Palsson, Network analysis of intermediary metabolism using linear optimization .1. development of mathematical formalism, Journal of Theoretical Biology, 154 (**1992**) 421-454.

[92] D.A. Fell, J.R. Small, Fat synthesis in adipose-tissue - an examination of stoichiometric constraints, Biochemical Journal, 238 (**1986**) 781-786.

[93] E.T. Papoutsakis, Equations and calculations for fermentations of butyric-acid bacteria, Biotechnology and Bioengineering, 26 (**1984**) 174-187.

[94] M.R. Antoniewicz, J.K. Kelleher, G. Stephanopoulos, Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements, Metabolic Engineering, 8 (**2006**) 324-337.

[95] R. Schuetz, L. Kuepfer, U. Sauer, Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*, Molecular Systems Biology, 3 (**2007**) 119.

[96] A.L. Knorr, R. Jain, R. Srivastava, Bayesian-based selection of metabolic objective functions, Bioinformatics, 23 (**2007**) 351-357.

[97] A.P. Burgard, C.D. Maranas, Optimization-based framework for inferring and testing hypothesized metabolic objective functions, Biotechnology and Bioengineering, 82 (**2003**) 670-677.

[98] K.R. Patil, I. Rocha, J. Forster, J. Nielsen, Evolutionary programming as a platform for in silico metabolic engineering, Bmc Bioinformatics, 6 (**2005**) 308.

[99] T. Shlomi, O. Berkman, E. Ruppin, Regulatory on/off minimization of metabolic flux changes after genetic perturbations, Proceedings of the National Academy of Sciences of the United States of America, 102 (**2005**) 7695-7700.

[100] J.S. Edwards, B.O. Palsson, Robustness analysis of the *Escherichia coli* metabolic network, Biotechnology Progress, 16 (**2000**) 927-939.

[101] T. Shlomi, Y. Eisenberg, R. Sharan, E. Ruppin, A genome-scale computational study of the interplay between transcriptional regulation and metabolism, Molecular Systems Biology, 3 (**2007**) 101.

[102] U. Sauer, Metabolic networks in motion: C-13-based flux analysis, Molecular Systems Biology, 2 (**2006**) 62.

[103] I. Thiele, N.D. Price, T.D. Vo, B.O. Palsson, Candidate metabolic network states in human mitochondria, Journal of Biological Chemistry, 280 (**2005**) 11683-11695.

[104] S.J. Wiback, I. Famili, H.J. Greenberg, B.O. Palsson, Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space, Journal of Theoretical Biology, 228 (**2004**) 437-447.

[105] N.D. Price, J. Schellenberger, B.O. Palsson, Uniform sampling of steady-state flux spaces: Means to design experiments and to interpret enzymopathies, Biophysical Journal, 87 (**2004**) 2172-2186.

[106] E. Almaas, B. Kovacs, T. Vicsek, Z.N. Oltvai, A.L. Barabasi, Global organization of metabolic fluxes in the bacterium *Escherichia coli*, Nature, 427 (**2004**) 839-843.

[107] W.R. Harcombe, N.F. Delaney, N. Leiby, N. Klitgord, C.J. Marx, The Ability of Flux Balance Analysis to Predict Evolution of Central Metabolism Scales with the Initial Distance to the Optimum, Plos Computational Biology, 9 (**2013**) e1003091.

[108] A. Varma, B.O. Palsson, Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110, Applied and Environmental Microbiology, 60 (**1994**) 3724-3731.

[109] P. Calik, T.H. Ozdamar, Bioreaction network flux analysis for industrial microorganisms: A review, Reviews in Chemical Engineering, 18 (**2002**) 553-596.

[110] J.S. Edwards, B.O. Palsson, Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions, BMC Bioinformatics, 1 (**2000**) 1.

[111] M.K. Lee, M.S. Mohamad, Y.W. Choon, K.M. Daud, N.A. Nasarudin, M.A. Ismail, Z. Ibrahim, S. Napis, R.O. Sinnott, A Hybrid of Particle Swarm Optimization and Minimization of Metabolic Adjustment for Ethanol Production of *Escherichia coli*, in: F. FdezRiverola, M. Rocha, M.S. Mohamad, N. Zaki, J.A. CastellanosGarzon (Eds.), Practical Applications of Computational Biology and Bioinformatics, 13th International Conference, Springer International Publishing Ag, Cham, (**2020**), 36-44.

[112] R.S. Costa, S. Vinga, Assessing *Escherichia coli* metabolism models and simulation approaches in phenotype predictions: Validation against experimental data, Biotechnology Progress, 34 (**2018**) 1344-1354.

[113] H.P. Bonarius, V. Hatzimanikatis, K.P. Meesters, C.D. de Gooijer, G. Schmid, J. Tramper, Metabolic flux analysis of hybridoma cells in different culture media using mass balances, Biotechnology and Bioengineering, 50 (**1996**) 299-318.

[114] S. Hui, J.M. Silverman, S.S. Chen, D.W. Erickson, M. Basan, J.L. Wang, T. Hwa, J.R. Williamson, Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria, Molecular Systems Biology, 11 (**2015**) 784.

[115] M. Faizi, R. Steuer, Optimal proteome allocation strategies for phototrophic growth in a light-limited chemostat, Microbial Cell Factories, 18 (**2019**) 165.

[116] M. Faizi, T. Zavrel, C. Loureiro, J. Cerveny, R. Steuer, A model of optimal protein allocation during phototrophic growth, Biosystems, 166 (**2018**) 26-36.

[117] L. Yang, J.T. Yurkovich, Z.A. King, B.O. Palsson, Modeling the multi-scale mechanisms of macromolecular resource allocation, Current Opinion in Microbiology, 45 (**2018**) 8-15.

[118] M. Basan, S. Hui, H. Okano, Z.G. Zhang, Y. Shen, J.R. Williamson, T. Hwa, Overflow metabolism in *Escherichia coli* results from efficient proteome allocation, Nature, 528 (**2015**) 99-+.

[119] M. Kafri, E. Metzl-Raz, G. Jona, N. Barkai, The Cost of Protein Production, Cell Reports, 14 (**2016**) 22-31.

[120] F. Bertaux, S. Marguerat, V. Shahrezaei, Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits, Royal Society Open Science, 5 (**2018**) 172234 .

[121] L. Yang, J.T. Yurkovich, C.J. Lloyd, A. Ebrahim, M.A. Saunders, B.O. Palsson, Principles of proteome allocation are revealed using proteomic data and genome-scale models, Scientific Reports, 6 (**2016**) 36734.

[122] E.J. O'Brien, B.O. Palsson, Computing the functional proteome: recent progress and future prospects for genome-scale models, Current Opinion in Biotechnology, 34 (**2015**) 125-134.

[123] A.M. Reimers, H. Lindhorst, S. Waldherr, A Protocol for Generating and Exchanging (Genome-Scale) Metabolic Resource Allocation Models, Metabolites, 7 (**2017**) 47.

[124] S. Dashko, N. Zhou, C. Compagno, J. Piskur, Why, when, and how did yeast evolve alcoholic fermentation?, FEMS Yeast Res, 14 (**2014**) 826-832.

[125] R. Diaz-Ruiz, M. Rigoulet, A. Devin, The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression, Biochimica Et Biophysica Acta-Bioenergetics, 1807 (**2011**) 568-576.

[126] A.J. Wolfe, The acetate switch, Microbiology and molecular biology reviews : MMBR, 69 (**2005**) 12-50.

[127] C. Wills, Regulation of sugar and ethanol-metabolism in *Saccharomyces cerevisiae*, Critical Reviews in Biochemistry and Molecular Biology, 25 (**1990**) 245-280.

[128] D.P. Clark, The fermentation pathways of *Escherichia coli*, FEMS microbiology reviews, 5 (**1989**) 223-234.

[129] G. Gottschalk, Bacterial Metabolism, Springer Science & Business Media, (**1986**), 208–282.

[130] Y. Chen, J. Nielsen, Energy metabolism controls phenotypes by protein efficiency and allocation, Proceedings of the National Academy of Sciences of the United States of America, 116 (**2019**) 17592-17597.

[131] R.A. Majewski, M.M. Domach, Simple constrained-optimization view of acetate overflow in *E. coli*, Biotechnol Bioeng, 35 (**1990**) 732-738.

[132] A. Vazquez, Q.K. Beg, M.A. Demenezes, J. Ernst, Z. Bar-Joseph, A.L. Barabasi, L.G. Boros, Z.N. Oltvai, Impact of the solvent capacity constraint on *E. coli* metabolism, BMC Syst Biol, 2 (**2008**) 7.

[133] Q.K. Beg, A. Vazquez, J. Ernst, M.A. de Menezes, Z. Bar-Joseph, A.L. Barabasi, Z.N. Oltvai, Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity, Proc Natl Acad Sci U S A, 104 (**2007**) 12663-12668.

[134] K. Zhuang, G.N. Vemuri, R. Mahadevan, Economics of membrane occupancy and respiro-fermentation, Molecular Systems Biology, 7(**2011**) 500.

[135] H. Zeng, A.D. Yang, Flux Balance Analysis Incorporating a Coarse-grained Proteome Constraint for Predicting Overflow Metabolism in *Escherichia coli*, in: A.A. Kiss, E. Zondervan, R. Lakerveld, L. Ozkan (Eds.) 29th European Symposium on Computer Aided Process Engineering, Pt A (**2019**) 865-870.

[136] C. Cheng, E.J. O'Brien, D. McCloskey, J. Utrilla, C. Olson, R.A. LaCroix, T.E. Sandberg, A.M. Feist, B.O. Palsson, Z.A. King, Laboratory evolution reveals a two-dimensional rate-yield tradeoff in microbial metabolism, Plos Computational Biology, 15 (**2019**) e1007066.

[137] B.J. Sanchez, J. Nielsen, Genome scale models of yeast: towards standardized evaluation and consistent omic integration, Integrative Biology, 7 (**2015**) 846-858.

[138] I. Massaiu, L. Pasotti, N. Sonnenschein, E. Rama, M. Cavaletti, P. Magni, C. Calvio, M.J. Herrgard, Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-glutamic acid production strains, Microbial Cell Factories, 18 (**2019**) 3.

[139] Z.A. King, J. Lu, A. Drager, P. Miller, S. Federowicz, J.A. Lerman, A. Ebrahim, B.O. Palsson, N.E. Lewis, BiGG Models: A platform for integrating, standardizing and sharing genome-scale models, Nucleic Acids Research, 44 (**2016**) D515-D522.

[140] J. Schellenberger, R. Que, R.M.T. Fleming, I. Thiele, J.D. Orth, A.M. Feist, D.C. Zielinski, A. Bordbar, N.E. Lewis, S. Rahmanian, J. Kang, D.R. Hyduke, B.Ø. Palsson, Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0, Nature Protocols, 6 (**2011**) 1290-1307.

[141] S.W. Zhang, W.L. Gou, Y. Li, Prediction of metabolic fluxes from gene expression data with Huber penalty convex optimization function, Molecular Biosystems, 13 (**2017**) 901-909.

[142] A. Khodayari, C.D. Maranas, A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains, Nature Communications, 7 (**2016**) 13806.

[143] A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, M. Heinemann, The quantitative and condition-dependent *Escherichia coli* proteome, Nature Biotechnology, 34 (**2015**) 104-110.

[144] S.S. Fong, B.O. Palsson, Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes, Nature Genetics, 36 (**2004**) 1056-1058.

APPENDICES

Appendix A: Experimental Data from the Literature

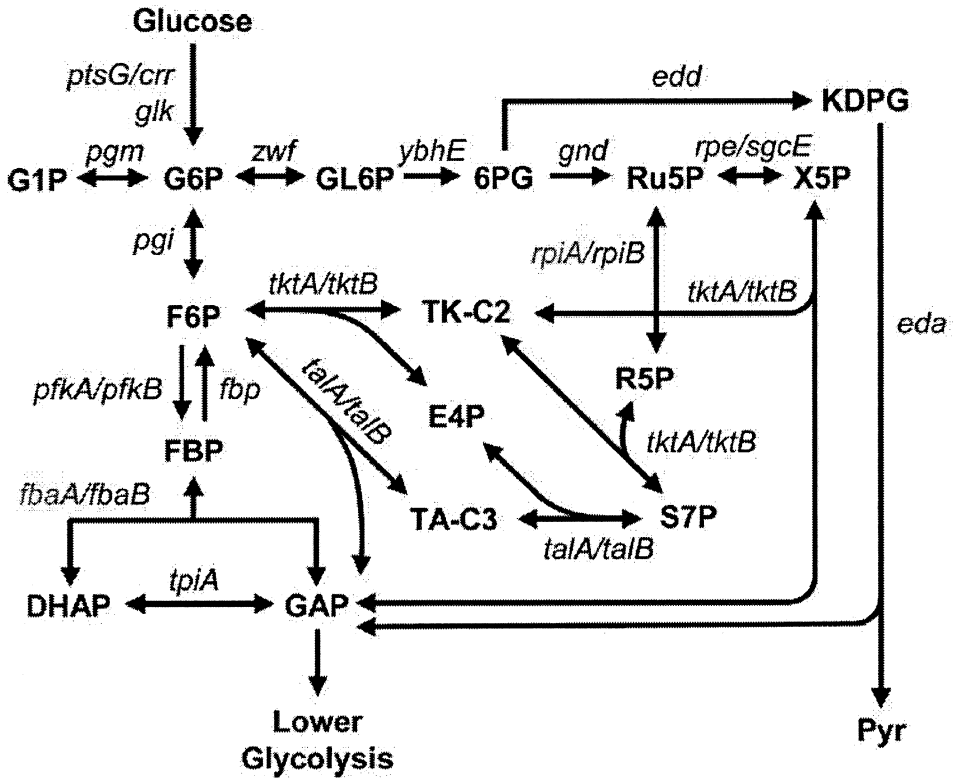

Figure A.1 Upper central carbon metabolism with the all genes which were deleted to create the single gene deletion mutants included in the study (rpiA and fbaA showed red color were not used in this study) [68].

Table A.1 Experimental data used in the study [68]

Gene	Growth Rate	Glucose Uptake Rate	Acetate Secretion Rate	CO2 Rate	Acetate Yield	Oxygen Uptake Rate	Biomass Yield
	mmol/ gdw.h	mmol/gdw.h	mmol/gdw.h	mmol/ gdw.h		mmol/gdw.h	
wt	0.63258	8.478	6.051	38.75	0.71365	38.74	0.4145
glk	0.5643	8.734	6.805	38.77	0.77918	38.77	0.35894
zwf	0.5736	8.783	6.837	39	0.77841	39	0.36281
gnd	0.5284	8.018	7.582	32.93	0.94555	32.93	0.3661
rpe	0.29967	4.286	2.545	20.62	0.59375	20.61	0.38841
edd	0.5846	7.745	5.848	34.75	0.75505	34.75	0.41936
eda	0.55123	8.683	7.282	37.51	0.83865	37.51	0.35269
pgi	0.19093	2.128	0.035	12.69	0.01656	12.69	0.49852
tpiA	0.16647	2.325	0.922	12.1	0.39661	12.1	0.3977
ptsG	0.25233	2.447	0	14.67	0	14.67	0.57289
crr	0.43723	5.255	3.119	25.27	0.59358	25.27	0.46226
pgm	0.68553	10.113	6.747	47.16	0.6671	47.16	0.37659
rpiB	0.5243	8.522	5.596	39.92	0.65671	39.92	0.3418
tktA	0.52373	6.397	3.733	30.89	0.58358	30.89	0.45486
tktB	0.59	9.394	7.476	41.39	0.79579	41.39	0.34892
talA	0.59153	8.189	6.171	36.77	0.75357	36.77	0.4013
talB	0.58323	7.353	4.382	35.33	0.59588	35.33	0.44066
pfkA	0.15847	2.042	0.857	10.53	0.41986	10.53	0.4312
pfkB	0.56732	7.104	5.522	31.56	0.77743	31.55	0.44369
fbp	0.5735	7.736	5.883	34.63	0.76042	34.63	0.41186
fbaB	0.62327	8.04	6.351	35.51	0.78996	35.51	0.43069
sgcE	0.5686	7.79	5.936	34.84	0.76203	34.84	0.40552

Appendix B: Pearson Correlation Coefficients

Table B.1 Pearson Correlation Coefficients

METHOD	BY	AY	μ
CA-FBA-HETE-Unconstrained	0.25	0.07	-0.27
CA-FBA-HOMO-Unconstrained	0.11	0.12	-0.28
CA-MiMBI-Unconstrained	0.34	-0.10	-0.19
CA-MOMA-Unconstrained	0.30	-0.02	-0.21
GECKO-FBA-Molar-Unconstrained	0.28	0.05	-0.27
GECKO-MiMBI-ENZ-Molar-Unconstrained	0.34	-0.21	-0.30
GECKO-MiMBI-RXN-Molar-Unconstrained	0.34	-0.18	-0.33
GECKO-MOMA-Molar-Unconstrained	0.22	-0.16	-0.38
GECKO-FBA-Mass-Unconstrained	0.28	0.10	-0.27
GECKO-MiMBI-ENZ-Mass-Unconstrained	0.29	0.20	-0.25
GECKO-MiMBI-RXN-Mass-Unconstrained	0.71	0.53	-0.53
CA-FBA-HETE-Constrained	0.26	0.45	0.63
CA-FBA-HOMO-Constrained	0.30	0.54	0.44
CA-MiMBI-Constrained	-0.17	0.60	0.61
CA-MOMA-Constrained	-0.32	0.76	0.54
GECKO-FBA-Molar-Constrained	0.32	0.02	0.42
GECKO-MiMBI-ENZ-Molar-Constrained	-0.22	0.07	0.61
GECKO-MiMBI-RXN-Molar-Constrained	-0.14	0.32	0.52
GECKO-MOMA-Molar-Constrained	0.17	0.00	0.61
GECKO-FBA-Mass-Constrained	0.32	0.20	0.42
GECKO-MiMBI-ENZ-Mass-Constrained	-0.25	0.44	0.58
GECKO-MiMBI-RXN-Mass-Constrained	-0.20	0.50	0.42
GECKO-MOMA-Mass-Constrained	0.19	0.49	0.50
FBA-Unconstrained	0.25	-	-0.27
MOMA-Unconstrained	-0.08	0.01	0.11
FBA-Constrained	0.20	-	0.47
MOMA-Constrained	-0.38	0.31	0.30

Appendix C: Mutant Phenotype Predictions with Unconstrained Model

Table C.1 CA-FBA homogenous case results of mutants phenotype predictions of unconstrained iML1515 model

Gene	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/gdw.h	mmol/ gdw.h	mmol/gdw.h			mmol/ gdw.h
wt	0.63443	10.517	10.832	15.469	0.335	1.030	14.186
glk	0.63443	10.517	10.832	15.469	0.335	1.030	14.186
zwf	0.62571	12.839	0.370	14.628	0.271	0.029	19.372
gnd	0.62985	10.902	12.044	15.543	0.321	1.105	14.269
rpe	0.63422	11.026	11.982	16.232	0.320	1.087	14.949
edd	0.63218	10.616	7.205	14.325	0.331	0.679	14.561
eda	0.63218	10.616	7.205	14.325	0.331	0.679	264.559
pgi	0.63443	10.517	10.832	15.469	0.335	1.030	14.186
tpiA	0.6343	11.003	11.850	16.354	0.320	1.077	15.071
ptsG	0.62498	16.157	0.369	12.345	0.215	0.023	20.794
crr	0.62498	16.157	0.369	12.345	0.215	0.023	20.794
pgm	0.000	48.357	0.000	13.720	0.000	0.000	59.791
rpiB	0.30794	17.754	2.185	17.736	0.096	0.123	261.637
tktA	0.000	59.775	4.573	4.573	0.000	0.077	62.062
tktB	0.000	59.775	4.573	4.573	0.000	0.077	62.062
talA	0.63443	10.517	10.832	15.469	0.335	1.030	264.183
talB	0.63443	10.517	10.832	15.469	0.335	1.030	264.183
pfkA	0.63439	10.753	11.304	15.940	0.328	1.051	264.654
pfkB	0.63443	10.517	10.832	15.469	0.335	1.030	14.186
fbp	0.63443	10.517	10.832	15.469	0.335	1.030	14.186
fbaB	0.63443	10.517	10.832	15.469	0.335	1.030	264.183
sgcE	0.63422	11.026	11.982	16.232	0.320	1.087	14.949

Table C.2 CA-FBA heterogenous case results of mutants phenotype predictions of unconstrained iML1515 model

Gene	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.631	7.253	0.000	17.698	0.483	0.000	16.422
glk	0.631	7.253	0.000	17.698	0.483	0.000	266.419
zwf	0.629	7.259	0.000	17.804	0.481	0.000	16.531
gnd	0.629	7.259	0.000	17.804	0.481	0.000	16.531
rpe	0.630	7.256	0.000	17.752	0.482	0.000	16.478
edd	0.631	7.253	0.000	17.698	0.483	0.000	16.422
eda	0.631	7.253	0.000	17.698	0.483	0.000	266.419
pgi	0.627	7.267	0.000	17.951	0.479	0.000	16.683
tpiA	0.611	7.317	0.000	18.883	0.464	0.000	267.644
ptsG	0.626	7.270	0.000	18.018	0.478	0.000	16.753
crr	0.626	7.270	0.000	18.018	0.478	0.000	16.753
pgm	0.000	9.302	0.000	55.814	0.000	0.000	55.814
rpiB	0.218	8.593	0.000	15.527	0.141	0.000	257.224
tktA	0.000	9.302	18.605	18.605	0.000	2.000	18.605
tktB	0.000	9.302	18.605	18.605	0.000	2.000	18.605
talA	0.631	7.253	0.000	17.698	0.483	0.000	266.419
talB	0.631	7.253	0.000	17.698	0.483	0.000	266.419
pfkA	0.628	7.264	0.000	17.894	0.480	0.000	266.622
pfkB	0.631	7.253	0.000	17.698	0.483	0.000	266.419
fbp	0.631	7.253	0.000	17.698	0.483	0.000	266.419
fbaB	0.631	7.253	0.000	17.698	0.483	0.000	266.419
sgcE	0.630	7.256	0.000	17.752	0.482	0.000	16.478

Table C.3 CA-MOMA results of mutants phenotype predictions of unconstrained iML1515 model

Gene	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.630	8.500	4.181	16.850	0.412	0.492	15.576
glk	0.630	8.500	4.180	16.851	0.412	0.492	15.577
zwf	0.507	8.205	4.214	16.663	0.343	0.514	266.018
gnd	0.581	8.715	4.220	16.757	0.371	0.484	265.674
rpe	0.610	8.586	4.214	16.850	0.395	0.491	15.597
edd	0.545	8.574	3.899	16.624	0.353	0.455	15.852
eda	0.545	8.574	3.899	16.624	0.353	0.455	265.839
pgi	0.630	8.500	4.180	16.851	0.412	0.492	265.574
tpiA	0.454	10.261	4.269	17.211	0.246	0.416	16.114
ptsG	0.545	8.216	4.116	16.740	0.368	0.501	265.807
crr	0.545	8.216	4.116	16.740	0.368	0.501	15.820
pgm	0.000	8.549	4.207	17.062	0.000	0.492	15.838
rpiB	0.062	8.781	3.990	16.896	0.039	0.454	265.818
tktA	0.000	8.863	4.158	16.868	0.000	0.469	15.873
tktB	0.000	8.863	4.158	16.868	0.000	0.469	15.873
talA	0.630	8.500	4.180	16.851	0.412	0.492	265.574
talB	0.630	8.500	4.180	16.851	0.412	0.492	265.574
pfkA	0.630	8.500	4.180	16.851	0.412	0.492	265.574
pfkB	0.630	8.500	4.180	16.851	0.412	0.492	265.574
fbp	0.630	8.500	4.180	16.851	0.412	0.492	265.574
fbaB	0.630	8.500	4.181	16.850	0.412	0.492	15.576
sgcE	0.610	8.586	4.214	16.850	0.395	0.491	15.597

Table C.4 CA-MiMBI results of mutants phenotype predictions of unconstrained iML1515 model

Gene	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/gdw.h	mmol/gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.630	8.500	4.180	16.851	0.412	0.492	15.577
glk	0.630	8.500	4.180	16.851	0.412	0.492	15.577
zwf	0.399	7.773	4.180	16.449	0.285	0.538	15.578
gnd	0.471	8.500	4.309	16.127	0.308	0.507	15.577
rpe	0.583	8.500	4.209	16.559	0.381	0.495	15.577
edd	0.390	8.500	3.808	15.871	0.255	0.448	15.577
eda	0.390	8.500	3.808	15.871	0.255	0.448	15.577
pgi	0.630	8.500	4.180	16.851	0.412	0.492	15.577
tpiA	0.348	8.500	4.180	16.355	0.227	0.492	15.578
ptsG	0.468	7.899	4.180	16.672	0.329	0.529	15.577
crr	0.468	7.899	4.180	16.672	0.329	0.529	15.577
pgm	0.000	8.316	4.180	16.795	0.000	0.503	15.577
rpiB	0.005	7.586	4.180	15.783	0.003	0.551	15.577
tktA	0.000	7.241	4.180	16.092	0.000	0.577	15.578
tktB	0.000	7.241	4.180	16.092	0.000	0.577	15.578
talA	0.625	8.492	4.180	16.851	0.409	0.492	15.577
talB	0.625	8.492	4.180	16.851	0.409	0.492	15.577
pfkA	0.630	8.500	4.180	16.851	0.412	0.492	15.577
pfkB	0.630	8.500	4.180	16.851	0.412	0.492	15.577
fbp	0.630	8.500	4.180	16.851	0.412	0.492	15.577
fbaB	0.630	8.500	4.180	16.851	0.412	0.492	15.577
sgcE	0.583	8.500	4.209	16.559	0.381	0.495	15.577

Table C.5 Determination results of total enzyme usage molar concentration

Total Enzyme Usage	Growth Rate	Glucose Uptake Rate-ub=8.5	Acetate Secretion Rate	CO2 Rate
mumol/gdw.h	mmol/gdw.h	mmol/gdw.h	mmol/gdw.h	mmol/gdw.h
29.00	0.620	8.5	3.057	19.511
29.10	0.622	8.5	2.968	19.605
29.20	0.624	8.5	2.879	19.699
29.30	0.626	8.5	2.790	19.792
29.40	0.628	8.5	2.701	19.886
29.50	0.630	8.5	2.612	19.979
29.60	0.632	8.5	2.523	20.073
29.6	0.633	8.5	2.479	20.120
29.70	0.634	8.5	2.434	20.167
29.80	0.636	8.5	2.345	20.260
29.90	0.638	8.5	2.257	20.354
30.00	0.640	8.5	2.168	20.447

Table C.6 Determination results of total enzyme usage mass concentration

Total Enzyme Usage	Growth Rate	Glucose Uptake Rate-ub=8.5	Acetate Secretion Rate	CO2 Rate
mumol/gdw.h	mmol/gdw.h	mmol/gdw.h	mmol/gdw.h	mmol/gdw.h
2160.0	0.632	8.500	2.903	19.314
2161.0	0.632	8.500	2.892	19.324
2162.0	0.633	8.500	2.882	19.334
2163.0	0.633	8.500	2.871	19.344
2163.5	0.633	8.500	2.866	19.349
2164.0	0.633	8.500	2.861	19.354
2165.0	0.634	8.500	2.851	19.364

Table C.7 GECKO-FBA with enzyme usage molar concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.479	20.120	0.414	0.292	18.839
glk	0.633	8.500	2.382	20.316	0.414	0.280	19.035
zwf	0.633	8.500	1.731	21.638	0.414	0.204	20.358
gnd	0.633	8.500	2.531	20.022	0.414	0.298	18.742
rpe	0.633	8.500	2.433	20.216	0.414	0.286	18.935
edd	0.629	8.500	3.998	17.260	0.411	0.470	15.988
eda	0.629	8.500	3.998	17.260	0.411	0.470	15.988
pgi	0.633	8.500	2.421	20.240	0.414	0.285	18.959
tpiA	0.633	8.500	2.512	20.055	0.414	0.296	18.774
ptsG	0.633	8.500	2.316	20.458	0.414	0.272	19.177
crr	0.633	8.500	2.316	20.458	0.414	0.272	19.177
pgm	0.000	8.500	11.084	28.832	0.000	1.304	28.832
rpiB	0.010	1.261	1.817	2.248	0.045	1.440	1.860
tktA	0.000	1.608	0.000	3.858	0.000	0.000	2.412
tktB	0.000	1.608	0.000	3.858	0.000	0.000	2.412
talA	0.633	8.500	2.471	20.136	0.414	0.291	18.855
talB	0.633	8.500	2.471	20.136	0.414	0.291	18.855
pfkA	0.633	8.500	2.479	20.120	0.414	0.292	18.839
pfkB	0.633	8.500	2.479	20.120	0.414	0.292	18.839
fbp	0.633	8.500	2.479	20.120	0.414	0.292	18.839
fbaB	0.633	8.500	2.479	20.120	0.414	0.292	18.839
sgcE	0.633	8.500	2.433	20.216	0.414	0.286	18.935

Table C.8 GECKO-FBA with enzyme usage mass concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.866	19.349	0.414	0.337	18.068
glk	0.633	8.500	2.866	19.349	0.414	0.337	18.068
zwf	0.633	8.500	2.866	19.349	0.414	0.337	18.068
gnd	0.630	8.500	3.598	18.021	0.412	0.423	16.747
rpe	0.631	8.500	3.604	17.971	0.412	0.424	16.695
edd	0.633	8.500	2.866	19.349	0.414	0.337	18.068
eda	0.633	8.500	2.866	19.349	0.414	0.337	18.068
pgi	0.633	8.500	2.679	19.724	0.414	0.315	18.443
tpiA	0.627	8.500	2.520	20.273	0.410	0.297	19.003
ptsG	0.633	8.500	2.830	19.423	0.414	0.333	18.142
crr	0.633	8.500	2.830	19.423	0.414	0.333	18.142
pgm	0.000	8.500	17.000	17.000	0.000	2.000	17.000
rpiB	0.014	3.385	5.770	6.381	0.024	1.705	5.831
tktA	0.000	8.500	17.000	17.000	0.000	2.000	17.000
tktB	0.000	8.500	17.000	17.000	0.000	2.000	17.000
talA	0.630	8.500	3.596	17.999	0.412	0.423	16.723
talB	0.630	8.500	3.596	17.999	0.412	0.423	16.723
pfkA	0.627	8.500	2.520	20.279	0.410	0.297	19.010
pfkB	0.627	8.500	2.521	20.279	0.410	0.297	19.010
fbp	0.633	8.500	2.866	19.349	0.414	0.337	18.068
fbaB	0.627	8.500	2.520	20.279	0.410	0.297	19.010
sgcE	0.631	8.500	3.604	17.971	0.412	0.424	16.695

Table C.9 GECKO-MOMA with enzyme usage molar concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.477	20.120	0.414	0.291	18.840
glk	0.627	8.484	2.459	20.153	0.410	0.290	18.908
zwf	-	-	-	-	--	--	-
gnd	0.626	8.521	2.501	20.100	0.408	0.293	18.861
rpe	0.623	8.520	2.456	20.172	0.407	0.288	18.882
edd	0.373	8.086	1.231	19.630	0.256	0.152	19.814
eda	--	--	--	--	--	--	--
pgi	0.629	8.500	2.466	20.150	0.411	0.290	18.868
tpiA	0.630	8.509	2.480	20.113	0.411	0.291	18.838
ptsG	0.566	8.215	2.356	20.041	0.383	0.287	19.120
crr	0.566	8.215	2.356	20.041	0.383	0.287	19.120
pgm	0.000	8.744	2.366	20.069	0.000	0.271	19.277
rpiB	--	--	--	--	--	--	--
tktA	0.000	8.788	2.269	20.016	0.000	0.258	19.445
tktB	0.000	8.788	2.269	20.016	0.000	0.258	19.445
talA	--	--	--	--	--	--	--
talB	--	--	--	--	--	--	--
pfkA	0.633	8.500	2.479	20.120	0.414	0.292	18.839
pfkB	0.633	8.500	2.478	20.122	0.414	0.292	18.841
fbp	0.633	8.500	2.477	20.121	0.414	0.291	18.841
fbaB	0.633	8.500	2.478	20.121	0.414	0.292	18.840
sgcE	0.623	8.520	2.456	20.172	0.407	0.288	18.882

Table C.10 GECKO-MOMA with enzyme usage mass concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	-	-	-	-	--	--	-
glk	-	-	-	-	--	--	-
zwf	--	--	--	--	--	--	--
gnd	-	-	-	-	--	--	-
rpe	--	--	--	--	--	--	--
edd	--	--	--	--	--	--	--
eda	--	--	--	--	--	--	--
pgi	0.625	8.468	2.794	19.515	0.410	0.330	18.250
tpiA	--	--	--	--	--	--	--
ptsG	--	--	--	--	--	--	--
crr	--	--	--	--	--	--	--
pgm	--	--	--	--	--	--	--
rpiB	--	--	--	--	--	--	--
tktA	-	-	-	--	--	--	-
tktB	--	--	--	--	--	--	--
talA	--	--	--	--	--	--	--
talB	--	--	--	--	--	--	--
pfkA	--	--	--	--	--	--	--
pfkB	--	--	--	--	--	--	--
fbp	--	--	--	--	--	--	--
fbaB	--	--	--	--	--	--	--
sgcE	-	--	-	-	--	--	-

Table C.11 GECKO-MiMBI-Rxn with enzyme usage molar concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.479	20.120	0.414	0.292	18.839
glk	0.631	8.498	2.479	20.119	0.413	0.292	18.839
zwf	0.332	7.132	2.479	20.120	0.258	0.348	18.839
gnd	0.604	8.500	2.479	20.055	0.395	0.292	18.840
rpe	0.613	8.417	2.479	20.120	0.405	0.295	18.839
edd	0.000	7.624	0.000	17.894	0.000	0.000	18.839
eda	0.000	7.657	0.000	17.558	0.000	0.000	18.839
pgi	0.633	8.500	2.479	20.120	0.414	0.292	18.839
tpiA	0.604	8.500	2.472	20.073	0.395	0.291	18.839
ptsG	0.528	8.430	2.479	19.931	0.348	0.294	18.839
crr	0.528	8.429	2.479	19.932	0.348	0.294	18.839
pgm	0.000	7.943	2.479	19.425	0.000	0.312	18.839
rpiB	0.000	7.804	2.479	19.080	0.000	0.318	18.839
tktA	0.000	7.805	2.479	19.168	0.000	0.318	18.839
tktB	0.000	7.805	2.479	19.168	0.000	0.318	18.839
talA	0.630	8.487	2.479	20.120	0.412	0.292	18.839
talB	0.630	8.487	2.479	20.120	0.412	0.292	18.839
pfkA	0.633	8.500	2.479	20.120	0.414	0.292	18.839
pfkB	0.633	8.500	2.479	20.120	0.414	0.292	18.839
fbp	0.633	8.500	2.479	20.120	0.414	0.292	18.839
fbaB	0.633	8.500	2.479	20.120	0.414	0.292	18.839
sgcE	0.613	8.417	2.479	20.120	0.405	0.295	18.839

Table C.12 GECKO-MiMBI-Enz with enzyme usage molar concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.479	20.120	0.414	0.292	18.839
glk	0.633	9.199	2.474	20.115	0.382	0.269	19.188
zwf	0.633	10.000	3.387	21.028	0.352	0.339	20.270
gnd	--	--	--	--	--	--	--
rpe	0.630	8.481	2.162	20.193	0.413	0.255	18.929
edd	0.590	10.000	0.000	17.903	0.328	0.000	18.161
eda	0.590	10.000	0.000	17.902	0.328	0.000	18.160
pgi	0.633	8.499	2.478	20.119	0.414	0.292	18.839
tpiA	--	--	--	--	--	--	--
ptsG	0.633	9.511	2.479	20.120	0.370	0.261	19.344
crr	0.633	9.512	2.479	20.120	0.370	0.261	19.345
pgm	0.000	8.230	2.479	19.055	0.000	0.301	19.067
rpiB	0.004	7.770	2.479	18.361	0.003	0.319	18.874
tktA	0.000	8.100	2.479	18.558	0.000	0.306	18.979
tktB	0.000	8.100	2.479	18.558	0.000	0.306	18.979
talA	0.633	8.507	2.428	20.131	0.413	0.285	18.854
talB	0.633	8.507	2.428	20.131	0.413	0.285	18.854
pfkA	0.633	8.500	2.479	20.120	0.414	0.292	18.839
pfkB	--	--	--	--	--	--	--
fbp	--	--	--	--	--	--	--
fbaB	0.633	8.500	2.479	20.120	0.414	0.292	18.839
sgcE	0.630	8.481	2.162	20.193	0.413	0.255	18.929

Table C.13 GECKO-MiMBI-Rxn with enzyme usage mass concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	--	--	--	--	--	--	--
glk	--	--	--	--	--	--	--
zwf	--	--	--	--	--	--	--
gnd	0.000	8.500	4.229	17.131	0.000	0.498	18.068
rpe	0.000	8.500	4.051	16.785	0.000	0.477	18.068
edd	--	--	--	--	--	--	--
eda	--	--	--	--	--	--	--
pgi	0.620	8.455	2.866	19.347	0.407	0.339	18.068
tpiA	--	--	--	--	--	--	--
ptsG	0.608	8.438	2.866	19.333	0.400	0.340	18.068
crr	0.608	8.438	2.866	19.333	0.400	0.340	18.068
pgm	0.000	8.199	3.177	18.794	0.000	0.388	18.068
rpiB	0.000	8.500	4.334	16.818	0.000	0.510	18.068
tktA	0.000	8.500	3.875	16.722	0.000	0.456	18.068
tktB	0.000	8.500	3.875	16.722	0.000	0.456	18.068
talA	0.051	8.500	3.782	16.933	0.033	0.445	18.068
talB	0.051	8.500	3.782	16.933	0.033	0.445	18.068
pfkA	0.571	8.316	2.866	19.324	0.382	0.345	18.068
pfkB	0.571	8.316	2.866	19.324	0.382	0.345	18.068
fbp	--	--	--	--	--	--	--
fbaB	0.570	8.266	2.866	19.349	0.383	0.347	18.068
sgcE	0.000	8.500	4.051	16.785	0.000	0.477	18.068

Table C.14 GECKO-MiMBI-Enz with enzyme usage mass concentration results of mutants phenotype predictions of unconstrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.866	19.349	0.414	0.337	18.068
glk	0.633	8.500	2.866	19.349	0.414	0.337	18.068
zwf	0.633	8.500	2.866	19.349	0.414	0.337	18.068
gnd	0.587	10.000	8.663	19.349	0.326	0.866	18.068
rpe	0.602	10.000	8.321	19.349	0.334	0.832	18.068
edd	0.633	8.500	2.866	19.349	0.414	0.337	18.068
eda	0.633	8.500	2.866	19.349	0.414	0.337	18.068
pgi	0.633	8.751	2.866	19.349	0.402	0.328	18.194
tpiA	0.568	10.000	6.257	22.800	0.315	0.626	21.542
ptsG	0.633	8.782	2.866	19.349	0.401	0.326	18.209
crr	0.633	8.782	2.866	19.349	0.401	0.326	18.209
pgm	0.000	8.377	2.697	19.349	0.000	0.322	18.279
rpiB	0.000	10.000	8.779	19.349	0.000	0.878	18.333
tktA	0.000	10.000	8.229	19.349	0.000	0.823	18.333
tktB	0.000	10.000	8.229	19.349	0.000	0.823	18.333
talA	0.595	10.000	8.478	19.349	0.330	0.848	18.068
talB	0.595	10.000	8.478	19.349	0.330	0.848	18.068
pfkA	0.550	10.000	3.055	19.349	0.305	0.305	18.068
pfkB	0.550	10.000	3.055	19.349	0.305	0.305	18.068
fbp	0.633	8.500	2.866	19.349	0.414	0.337	18.068
fbaB	0.550	10.000	3.055	19.349	0.305	0.305	18.068
sgcE	0.602	10.000	8.321	19.349	0.334	0.832	18.068

Appendix D: Mutant Phenotype Predictions with Constrained Model

Table D.1 CA-FBA homogenous case results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.630	8.500	4.180	16.851	0.412	0.492	38.740
glk	0.631	8.734	4.985	16.622	0.401	0.571	38.770
zwf	0.625	8.783	3.069	17.318	0.396	0.349	39.000
gnd	0.625	8.018	2.500	17.520	0.433	0.312	32.930
rpe	0.362	4.286	0.000	10.909	0.469	0.000	20.610
edd	0.628	7.745	1.732	17.306	0.450	0.224	34.750
eda	0.630	8.683	5.102	16.102	0.403	0.588	37.510
pgi	0.165	2.128	0.000	6.018	0.431	0.000	12.690
tpiA	0.176	2.325	0.000	6.742	0.421	0.000	12.100
ptsG	0.193	2.447	0.000	6.767	0.439	0.000	14.670
crr	0.445	5.255	0.000	13.317	0.470	0.000	25.270
pgm	0.000	10.113	0.000	39.980	0.000	0.000	47.160
rpiB	0.217	8.522	0.000	15.409	0.141	0.000	39.920
tktA	0.000	6.397	0.000	30.890	0.000	0.000	30.890
tktB	0.000	9.394	0.000	41.145	0.000	0.000	41.390
talA	0.629	8.189	3.112	17.154	0.427	0.380	36.770
talB	0.626	7.353	0.355	17.764	0.473	0.048	35.330
pfkA	0.158	2.042	0.000	5.787	0.430	0.000	10.530
pfkB	0.615	7.104	0.000	17.441	0.481	0.000	31.550
fbp	0.628	7.736	1.554	17.597	0.451	0.201	34.630
fbaB	0.629	8.040	2.598	17.301	0.435	0.323	35.510
sgcE	0.628	7.790	1.697	17.647	0.448	0.218	34.840

Table D.2 CA-FBA heterogenous case results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.634	8.500	2.314	19.344	0.415	0.272	38.740
glk	0.636	8.734	2.652	19.520	0.405	0.304	38.770
zwf	0.634	8.783	2.426	20.123	0.401	0.276	39.000
gnd	0.619	8.018	1.350	19.320	0.429	0.168	32.930
rpe	0.361	4.286	0.000	10.915	0.467	0.000	20.610
edd	0.613	7.745	1.140	18.272	0.440	0.147	34.750
eda	0.636	8.683	2.365	18.931	0.407	0.272	37.510
pgi	0.164	2.128	0.000	6.019	0.429	0.000	12.690
tpiA	0.176	2.325	0.000	6.747	0.419	0.000	12.100
ptsG	0.193	2.447	0.000	6.768	0.438	0.000	14.670
crr	0.443	5.255	0.002	13.328	0.469	0.000	25.270
pgm	0.653	10.113	4.764	19.412	0.359	0.471	47.160
rpiB	0.217	8.522	0.000	15.423	0.141	0.000	39.920
tktA	0.000	6.397	3.429	20.513	0.000	0.536	30.890
tktB	0.000	9.394	5.106	25.879	0.000	0.544	41.390
talA	0.628	8.189	1.707	19.216	0.426	0.208	36.770
talB	0.599	7.353	0.751	17.846	0.453	0.102	35.330
pfkA	0.158	2.042	0.000	5.787	0.429	0.000	10.530
pfkB	0.589	7.104	0.465	17.456	0.460	0.065	31.550
fbp	0.617	7.736	1.090	18.558	0.443	0.141	34.630
fbaB	0.627	8.040	1.408	19.084	0.434	0.175	35.510
sgcE	0.614	7.790	1.194	18.800	0.438	0.153	34.840

Table D.3 CA-MOMA results of mutants phenotype predictions of constrained iML1515 model

Gene	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.630	8.500	4.180	16.851	0.412	0.492	38.740
glk	0.622	8.734	4.176	16.862	0.396	0.478	38.770
zwf	0.503	8.783	4.177	16.593	0.318	0.476	39.000
gnd	0.152	8.018	2.827	15.069	0.105	0.353	32.930
rpe	0.000	4.286	0.000	6.838	0.000	0.000	20.610
edd	0.300	7.745	3.025	15.543	0.215	0.391	34.750
eda	0.495	8.683	3.797	16.403	0.317	0.437	37.510
pgi	0.000	2.128	0.000	4.256	0.000	0.000	12.690
tpiA	0.000	2.325	0.000	4.336	0.000	0.000	12.100
ptsG	0.000	2.447	0.000	4.894	0.000	0.000	14.670
crr	0.000	5.255	0.074	11.004	0.000	0.014	25.270
pgm	0.000	10.113	4.270	19.054	0.000	0.422	47.160
rpiB	0.057	8.522	4.055	17.040	0.037	0.476	39.920
tktA	0.000	6.397	1.698	13.085	0.000	0.265	30.890
tktB	0.000	9.394	4.143	17.245	0.000	0.441	41.390
talA	0.470	8.189	3.673	16.274	0.319	0.449	36.770
talB	0.317	7.353	3.080	15.393	0.239	0.419	35.330
pfkA	0.000	2.042	0.000	4.084	0.000	0.000	10.530
pfkB	0.035	7.104	2.217	14.589	0.028	0.312	31.550
fbp	0.291	7.736	3.073	15.575	0.209	0.397	34.630
fbaB	0.371	8.040	3.361	15.954	0.256	0.418	35.510
sgcE	0.294	7.790	3.175	15.517	0.209	0.408	34.840

Table D.4 CA-MiMBI results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.630	8.500	4.180	16.851	0.412	0.492	38.740
glk	0.624	8.734	4.180	16.851	0.397	0.479	38.770
zwf	--	--	--	--	--	--	--
gnd	0.488	8.018	4.180	16.393	0.338	0.521	32.930
rpe	0.011	4.286	4.180	15.743	0.014	0.975	20.610
edd	0.510	7.745	3.698	16.690	0.366	0.477	34.750
eda	0.492	8.683	3.378	16.786	0.315	0.389	37.510
pgi	0.000	2.128	0.001	12.614	0.000	0.000	12.690
tpiA	0.000	2.325	0.925	12.100	0.000	0.398	12.100
ptsG	0.000	2.447	0.006	14.670	0.000	0.002	14.670
crr	0.176	5.255	4.180	15.971	0.186	0.796	25.270
pgm	0.000	10.113	4.180	18.346	0.000	0.413	47.160
rpiB	0.005	8.522	4.180	15.903	0.003	0.491	39.920
tktA	0.000	6.397	4.180	16.548	0.000	0.653	30.890
tktB	0.000	9.394	4.180	16.631	0.000	0.445	41.390
talA	0.579	8.189	4.180	16.818	0.393	0.510	36.770
talB	0.469	7.353	4.180	16.528	0.355	0.569	35.330
pfkA	0.000	2.042	0.826	10.589	0.000	0.405	10.530
pfkB	0.433	7.104	4.180	16.476	0.339	0.588	31.550
fbp	0.523	7.736	4.180	16.636	0.376	0.540	34.630
fbaB	0.554	8.040	4.180	16.806	0.383	0.520	35.510
sgcE	0.525	7.790	4.180	16.667	0.374	0.537	34.840

Table D.5 GECKO-FBA with enzyme usage molar concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	0.000	25.091	0.414	0.000	38.740
glk	0.634	8.734	0.000	26.455	0.403	0.000	38.770
zwf	0.634	8.783	0.000	26.754	0.401	0.000	39.000
gnd	0.639	8.018	0.000	21.955	0.443	0.000	32.930
rpe	0.362	4.286	0.000	10.909	0.469	0.000	20.610
edd	0.628	7.745	0.000	20.764	0.450	0.000	34.750
eda	0.627	8.683	0.000	26.442	0.401	0.000	37.510
pgi	0.165	2.128	0.000	6.018	0.431	0.000	12.690
tpiA	0.176	2.325	0.000	6.742	0.421	0.000	12.100
ptsG	0.193	2.447	0.000	6.767	0.439	0.000	14.670
crr	0.445	5.255	0.000	13.317	0.470	0.000	25.270
pgm	0.000	10.113	0.000	49.091	0.000	0.000	47.160
rpiB	0.010	8.522	4.583	40.303	0.007	0.538	39.920
tktA	0.000	6.397	0.000	30.890	0.000	0.000	30.890
tktB	0.000	9.394	0.000	39.540	0.000	0.000	41.390
talA	0.634	8.189	0.000	23.163	0.430	0.000	36.770
talB	0.624	7.353	0.000	18.554	0.472	0.000	35.330
pfkA	0.158	2.042	0.000	5.787	0.430	0.000	10.530
pfkB	0.606	7.104	0.000	17.816	0.474	0.000	31.550
fbp	0.633	7.736	0.000	20.487	0.455	0.000	34.630
fbaB	0.636	8.040	0.000	22.224	0.439	0.000	35.510
sgcE	0.634	7.790	0.000	20.794	0.452	0.000	34.840

Table D.6 GECKO-FBA with enzyme usage mass concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.665	19.767	0.414	0.313	38.740
glk	0.635	8.734	3.426	19.570	0.404	0.392	38.770
zwf	0.635	8.783	3.587	19.533	0.402	0.408	39.000
gnd	0.629	8.018	1.862	18.618	0.436	0.232	32.930
rpe	0.362	4.286	0.000	10.909	0.469	0.000	20.610
edd	0.630	7.745	0.088	20.508	0.452	0.011	34.750
eda	0.635	8.683	3.211	19.667	0.406	0.370	37.510
pgi	0.165	2.128	0.000	6.018	0.431	0.000	12.690
tpiA	0.176	2.325	0.000	6.742	0.421	0.000	12.100
ptsG	0.193	2.447	0.000	6.767	0.439	0.000	14.670
crr	0.445	5.255	0.000	13.317	0.470	0.000	25.270
pgm	0.000	10.113	0.000	49.091	0.000	0.000	47.160
rpiB	0.014	8.522	16.052	16.659	0.009	1.884	39.920
tktA	0.000	6.397	0.000	29.820	0.000	0.000	30.890
tktB	0.000	9.394	0.000	39.251	0.000	0.000	41.390
talA	0.628	8.189	2.557	18.310	0.426	0.312	36.770
talB	0.623	7.353	0.000	18.632	0.470	0.000	35.330
pfkA	0.158	2.042	0.000	5.787	0.430	0.000	10.530
pfkB	0.600	7.104	0.000	18.067	0.469	0.000	31.550
fbp	0.630	7.736	0.055	20.519	0.452	0.007	34.630
fbaB	0.627	8.040	0.856	20.868	0.433	0.106	35.510
sgcE	0.627	7.790	1.187	18.700	0.447	0.152	34.840

Table D.7 GECKO-MOMA with enzyme usage molar concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	0.000	25.091	0.414	0.000	38.740
glk	0.633	8.734	0.000	26.455	0.403	0.000	38.770
zwf	-	-	-	-	--	--	-
gnd	-	-	-	-	--	--	-
rpe	0.351	4.286	0.018	10.874	0.455	0.004	20.610
edd	0.399	7.745	0.000	19.940	0.286	0.000	34.750
eda	0.339	8.683	0.000	25.435	0.217	0.000	37.510
pgi	0.131	2.128	0.000	6.020	0.341	0.000	12.690
tpiA	0.099	2.325	0.000	6.239	0.236	0.000	12.100
ptsG	0.154	2.447	0.000	6.661	0.349	0.000	14.670
crr	0.361	5.255	0.000	13.073	0.382	0.000	25.270
pgm	--	--	--	--	--	--	--
rpiB	--	--	--	--	--	--	--
tktA	0.000	6.397	0.000	15.399	0.000	0.000	30.890
tktB	0.000	9.394	0.000	30.135	0.000	0.000	41.390
talA	--	--	--	--	--	--	--
talB	0.568	7.353	0.100	18.301	0.429	0.014	35.330
pfkA	0.134	2.042	0.000	5.774	0.365	0.000	10.530
pfkB	0.501	7.104	0.000	17.570	0.391	0.000	31.550
fbp	0.633	7.736	0.000	20.488	0.455	0.000	34.630
fbaB	0.635	8.040	0.000	22.224	0.439	0.000	35.510
sgcE	0.623	7.790	0.000	20.814	0.444	0.000	34.840

Table D.8 GECKO-MOMA with enzyme usage mass concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.665	19.767	0.414	0.313	38.740
glk	0.635	8.734	3.424	19.574	0.404	0.392	38.770
zwf	-	-	-	-	--	--	-
gnd	-	-	-	-	--	--	-
rpe	0.357	4.286	0.095	10.867	0.462	0.022	20.610
edd	--	--	--	--	--	--	--
eda	--	--	--	--	--	--	--
pgi	0.140	2.128	0.000	6.100	0.364	0.000	12.690
tpiA	0.142	2.325	0.000	6.390	0.339	0.000	12.100
ptsG	0.180	2.447	0.022	6.718	0.408	0.009	14.670
crr	0.415	5.255	0.039	13.205	0.439	0.008	25.270
pgm	--	--	--	--	--	--	--
rpiB	--	--	--	--	--	--	--
tktA	0.000	6.397	0.000	15.310	0.000	0.000	30.890
tktB	0.000	9.394	7.871	16.730	0.000	0.838	41.390
talA	0.478	8.189	4.410	19.135	0.325	0.539	36.770
talB	--	--	--	--	--	--	--
pfkA	--	--	--	--	--	--	--
pfkB	--	--	--	--	--	--	--
fbp	0.630	7.736	0.055	20.519	0.452	0.007	34.630
fbaB	0.516	8.040	0.217	19.792	0.357	0.027	35.510
sgcE	-	-	-	-	--	--	-

Table D.9 GECKO-MiMBI-Rxn with enzyme usage molar concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	0.000	25.091	0.414	0.000	38.740
glk	0.631	8.734	0.014	24.986	0.401	0.002	38.770
zwf	0.616	8.783	0.000	25.091	0.390	0.000	39.000
gnd	0.522	8.018	1.111	21.978	0.362	0.139	32.930
rpe	0.000	4.286	0.196	16.853	0.000	0.046	20.610
edd	0.000	7.745	0.000	20.408	0.000	0.000	34.750
eda	0.000	8.683	0.000	20.781	0.000	0.000	37.510
pgi	0.000	2.128	0.000	12.561	0.000	0.000	12.690
tpiA	0.000	2.325	0.000	12.351	0.000	0.000	12.100
ptsG	0.000	2.447	0.000	13.111	0.000	0.000	14.670
crr	0.152	5.255	0.144	19.022	0.161	0.027	25.270
pgm	0.000	10.113	1.709	27.011	0.000	0.169	47.160
rpiB	0.000	8.522	0.499	24.843	0.000	0.059	39.920
tktA	0.000	6.397	0.303	22.449	0.000	0.047	30.890
tktB	0.000	9.394	0.517	25.091	0.000	0.055	41.390
talA	0.582	8.189	0.132	24.125	0.395	0.016	36.770
talB	0.460	7.353	0.051	23.058	0.347	0.007	35.330
pfkA	0.000	2.042	0.000	11.126	0.000	0.000	10.530
pfkB	0.415	7.104	0.224	21.669	0.325	0.032	31.550
fbp	0.529	7.736	0.153	23.018	0.380	0.020	34.630
fbaB	0.555	8.040	0.361	--	0.383	0.045	35.510
sgcE	0.522	7.790	0.283	23.174	0.372	0.036	34.840

Table D.10 GECKO-MiMBI-Enz with enzyme usage molar concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	0.000	25.091	0.414	0.000	38.740
glk	0.633	8.734	0.000	24.982	0.403	0.000	38.770
zwf	0.633	8.783	0.155	25.247	0.400	0.018	39.000
gnd	0.585	8.018	0.000	21.541	0.405	0.000	32.930
rpe	0.114	4.286	0.000	12.078	0.148	0.000	20.610
edd	0.605	7.745	0.000	21.027	0.434	0.000	34.750
eda	0.594	8.683	0.000	23.470	0.380	0.000	37.510
pgi	0.000	2.128	0.000	6.127	0.000	0.000	12.690
tpiA	0.000	2.325	0.000	6.773	0.000	0.000	12.100
ptsG	0.000	2.447	0.000	7.505	0.000	0.000	14.670
crr	0.249	5.255	0.000	14.694	0.264	0.000	25.270
pgm	0.000	10.113	5.459	31.905	0.000	0.540	47.160
rpiB	0.004	8.522	0.347	24.394	0.003	0.041	39.920
tktA	0.000	6.397	0.000	15.192	0.000	0.000	30.890
tktB	0.000	9.394	1.188	25.825	0.000	0.126	41.390
talA	0.618	8.189	0.000	23.074	0.419	0.000	36.770
talB	0.518	7.353	0.000	21.518	0.392	0.000	35.330
pfkA	0.000	2.042	0.000	5.421	0.000	0.000	10.530
pfkB	0.498	7.104	0.000	19.906	0.389	0.000	31.550
fbp	0.587	7.736	0.000	21.584	0.422	0.000	34.630
fbaB	0.601	8.040	0.000	21.730	0.415	0.000	35.510
sgcE	0.597	7.790	0.000	21.626	0.426	0.000	34.840

Table D.11 GECKO-MiMBI-Rxn with enzyme usage mass concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.665	19.767	0.414	0.313	38.740
glk	0.626	8.734	2.679	19.646	0.398	0.307	38.770
zwf	0.603	8.783	2.733	19.767	0.381	0.311	39.000
gnd	0.000	8.018	3.257	15.374	0.000	0.406	32.930
rpe	-	-	-	-	--	--	-
edd	0.527	7.745	2.665	18.430	0.378	0.344	34.750
eda	0.609	8.683	2.868	19.361	0.389	0.330	37.510
pgi	0.000	2.128	0.000	10.109	0.000	0.000	12.690
tpiA	0.000	2.325	1.060	9.965	0.000	0.456	12.100
ptsG	0.000	2.447	0.453	10.468	0.000	0.185	14.670
crr	0.138	5.255	2.665	14.872	0.146	0.507	25.270
pgm	0.000	10.113	4.384	22.752	0.000	0.434	47.160
rpiB	0.000	8.522	5.498	18.412	0.000	0.645	39.920
tktA	0.000	6.397	3.735	14.512	0.000	0.584	30.890
tktB	0.000	9.394	6.636	19.551	0.000	0.706	41.390
talA	0.000	8.189	3.888	16.957	0.000	0.475	36.770
talB	0.000	7.353	3.888	16.961	0.000	0.529	35.330
pfkA	0.000	2.042	0.457	9.326	0.000	0.224	10.530
pfkB	0.407	7.104	2.665	17.728	0.318	0.375	31.550
fbp	0.532	7.736	2.665	18.330	0.382	0.344	34.630
fbaB	0.548	8.040	2.665	19.462	0.379	0.331	35.510
sgcE	0.000	7.790	4.295	15.520	0.000	0.551	34.840

Table D.12 GECKO-MiMBI-Enz with enzyme usage mass concentration results of mutants phenotype predictions of constrained iML1515 model

Genes	Growth Rate	Glucose Uptake Rate	Acetate Sec. Rate	CO2 Rate	Biomass Yield	Acetate Yield	O_2 Uptake Rate
	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h	mmol/ gdw.h			mmol/ gdw.h
wt	0.633	8.500	2.665	19.767	0.414	0.313	38.740
glk	0.632	8.734	2.602	19.690	0.402	0.298	38.770
zwf	0.633	8.783	2.823	19.925	0.400	0.321	39.000
gnd	0.401	8.018	5.789	15.399	0.278	0.722	32.930
rpe	0.118	4.286	0.000	10.732	0.153	0.000	20.610
edd	0.544	7.745	1.892	18.286	0.390	0.244	34.750
eda	0.633	8.683	2.625	19.728	0.405	0.302	37.510
pgi	0.000	2.128	0.000	5.764	0.000	0.000	12.690
tpiA	0.000	2.325	0.000	6.775	0.000	0.000	12.100
ptsG	0.000	2.447	0.000	7.507	0.000	0.000	14.670
crr	0.285	5.255	0.000	13.859	0.301	0.000	25.270
pgm	0.000	10.113	8.337	27.266	0.000	0.824	47.160
rpiB	0.000	8.522	9.951	20.299	0.000	1.168	39.920
tktA	0.000	6.397	2.511	10.778	0.000	0.393	30.890
tktB	0.000	9.394	11.200	21.537	0.000	1.192	41.390
talA	0.490	8.189	4.644	17.686	0.332	0.567	36.770
talB	0.413	7.353	3.899	16.187	0.312	0.530	35.330
pfkA	0.000	2.042	0.000	5.955	0.000	0.000	10.530
pfkB	0.332	7.104	0.000	14.647	0.260	0.000	31.550
fbp	0.543	7.736	1.883	18.268	0.390	0.243	34.630
fbaB	0.416	8.040	0.558	16.420	0.287	0.069	35.510
sgcE	0.415	7.790	5.426	15.699	0.296	0.697	34.840

CURRICULUM VITAE

Name and Surname	:	Ece Anturan
Place of Birth	:	Bursa/TURKEY
Date of Birth	:	28.06.1992
Marital Status	:	Single
Address	:	Kavaklıdere Mahallesi Tunalı Hilmi Caddesi 36/6 Küçükesat Çankaya/ANKARA
Cell Phone	:	+90 538 295 15 92
E-mail Address	:	ece.anturan@gmail.com

Foreign Language : English (Upper-intermediate)

German (Intermediate)

Educational Status

B.Sc.	:	Hacettepe University, Faculty of Engineering, Chemical Engineering (2010-2016) (CGPA:3.05)
B.Sc.	:	Hacettepe University, Faculty of Engineering, Food Engineering (2012-2015) (CGPA:2.46) (Minor)
B.Sc.	:	RWTH Aachen University, Faculty of Mechanical Engineering, Mechanical Engineering (2013-2014) (Erasmus Program)
M.Sc.	:	Hacettepe University, Faculty of Engineering, Chemical Engineering (2017-2020) (cgpa:3.87)
M.Sc.	:	RWTH Aachen University, Institute of Applied Microbiology (2017-2018) (Erasmus Program)

Work Experiences

2015-2015 (35 days)	Internship (Summer Internship) OYAK-RENAULT Otomobil Fabrikaları A.Ş. BURSA TURKEY - Paint shop Laboratory
2017-2017 (7 months)	Master of Science Degree Studies

	Dept. of Chemical Eng., Hacettepe Univ., Ankara-TURKEY
	Industrial Biotechnology and Protein Engineering Laboratory
2017-2018 (13 months)	Master of Science Degree Thesis Studies
	Institute of Applied Microbiology, RWTH Aachen University, Aachen-GERMANY

Projects and Budgets

- Design of PET Production (Spring Semester of 2014-2015 Academic Year)
- Investigation of tetracycline loading and releasing to/from magnetic chitosan and chitosan halloysite nanocomposites (Spring Semester of 2015-2016 Academic Year)
- Investigation of metabolic modelling strategies to predict phenotypes of microbial mutants under data scarcity (Spring Semester of 2017-2018 Academic Year - Spring Semester of 2019-2020 Academic Year)
- Budget from TÜBİTAK-1001 Grant # 118Z798

Computer Skills

- MS Office Programs (Advanced)
- MATLAB (Advanced)

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
TO THE DEPARTMENT OF CHEMICAL ENGINEERING

Date: 27/01/2020

Thesis Title / Topic: Investigation Of Metabolic Modeling Strategies To Predict Phenotypes Of Microbial Mutants Under Data Scarcity

According to the originality report obtained by my thesis advisor by using the *Turnitin* plagiarism detection software and by applying the filtering options stated below on 27/01/2020 for the total of 121 pages including the a) Title Page, b) Introduction, c) Main Chapters, d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 9 %.

Filtering options applied:

1. Bibliography/Works Cited excluded
2. Quotes excluded
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Sciene and Engineering Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

27/01/2020

Name Surname:	Ece ANTURAN
Student No:	N16221788
Department:	CHEMICAL ENGINEERING
Program:	CHEMICAL ENGINEERING – M.Sc.
Status:	■ Masters ☐ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED.

Assoc. Prof. Dr. Eda ÇELİK AKDUR