Inês Azevedo Isidro
Mestre em Engenharia Biológica

Cell functional enviromics: Applications to *Pichia pastoris*

Dissertação para obtenção do Grau de Doutor em Bioengenharia

Orientador: Rui Manuel Freitas Oliveira, Professor Associado, FCT-UNL

Júri:

Presidente: Doutor Pedro Manuel Corrêa Calvente Barahona
Arguentes: Doutora Susana de Almeida Mendes Vinga Martins
Doutor Armindo José Alves da Silva Salvador
Vogais: Doutor Rui Manuel Freitas Oliveira
Doutora Ana Margarida Palma Teixeira

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2015

Cell functional enviromics: Applications to *Pichia pastoris*

Acknowledgments

I would like to express my gratitude to my supervisor, Prof. Rui Oliveira, without whose vision and support this thesis wouldn't have been possible. My sincere thanks to Dr. Susana Vinga and Dr. Ana Teixeira, my thesis committee, for their valuable feedback and insight.

This PhD thesis was undertaken with financial support from the Fundação para a Ciência e a Tecnologia through scholarship SFRH/BD/70768/2010. I am grateful to the MIT-Portugal for the opportunity of being part of the Bioengineering Systems Doctoral Program, and to my host institutions FCT-UNL and IBET. I am particularly grateful to Dr. José Silva Lopes for the precious guidance through the necessary formalities.

My special thanks go to my colleagues at MIT-Portugal, SBE and IBET-UP for the cherished companionship and memorable moments.

Finally I thank my family, especially my parents for their care and support throughout so many years, and dedicate this thesis to my grandparents.

Abstract

The present PhD thesis develops the cell functional enviromics (CFE) method to investigate the relationship between environment and cellular physiology. CFE may be defined as the envirome-wide cellular function reconstruction through the collection and systems-level analysis of dynamic envirome data. Throughout the thesis, CFE is illustrated by two main applications to cultures of a constitutive *P. pastoris* X-33 strain expressing a scFv antibody fragment.

The first application addresses the challenge of culture media development. A dataset was built from 26 shake flask experiments, with variations in trace elements concentrations and basal medium dilution based on the standard BSM+PTM1. Protein yield showed high sensitivity to culture medium variations, while biomass was essentially determined by BSM dilution. High scFv yield was associated with high overall metabolic fluxes through central carbon pathways concomitantly with a relative shift of carbon flux from biosynthetic towards energy-generating pathways. CFE identified three cellular functions (growth, energy generation and by-product formation) that together described 98.8% of the variance in observed fluxes. Analyses of how medium factors relate to identified cellular functions showed iron and manganese at concentrations close to PTM1 inhibit overall metabolic activity.

The second application addresses bioreactor operation. Pilot 50 L fed-batch cultivations, followed by ^{1}H-NMR exometabolite profiling, allowed the acquisition of data for 21 environmental factors over time. CFE identified five major metabolic pathway groups that are frequently activated by the environment. The resulting functional enviromics map may serve as template for future optimization of media composition and feeding strategies for *Pichia pastoris*.

The present PhD thesis is a step forward towards establishing the foundations of CFE that is still at its infancy. The methods developed herein are a contribution for changing the culture media and process development paradigm towards a holistic and systematic discipline in the future.

Keywords

Pichia pastoris · cell functional enviromics · projection to latent pathways · culture media design

Resumo

A presente tese de doutoramento desenvolve a envirómica funcional celular (EFC) como método para investigar a relação entre a fisiologia da célula e o seu ambiente. A EFC pode ser definida como a reconstrução da função celular através da aquisição e da análise sistemática de dados dinâmicos do ambiente extracelular, considendo o ambiente como um todo (o enviroma). A metodologia é ilustrada através de duas aplicações principais a culturas de uma estirpe de *Pichia pastoris* X-33 com expressão constitutiva de um fragmento de anticorpo do tipo scFv.

A primeira aplicação foi para o desenvolvimento de meios de cultura. Experiências em *shake flasks* foram sujeitas a 26 variações nas concentrações de micro-minerais e na diluição do meio basal, com base no standard BSM+PTM1. O rendimento de produto mostrou uma grande sensibilidade às variações no meio, enquanto que a biomassa variou apenas com a diluição do meio basal. Um rendimento de produto elevado mostrou-se associado a fluxos superiores em todas as vias metabólicas centrais, em simultâneo com um desvio do fluxo de carbono de vias biosintéticas para vias energéticas. Foram identificadas três funções celulares que em conjunto explicam 98.8% da vâriancia observada nos fluxos. A análise da influência de factores do meio mostrou que o ferro e o manganésio em concentrações próximas do standard inibem a actividade metabólica em geral.

A segunda aplicação efectuou-se sobre culturas em bioreactor. Culturas piloto de 50 L foram complementadas com análise metabolómica por ^{1}H-NMR, permitindo a aquisição de dados para 21 factores ambientais ao longo do tempo. Foram identificados cinco grupos de caminhos metabólicos que são frequentemente activados pelo ambiente. O mapa de envirómica funcional resultante pode servir como base para a optimização do meio de cultura e de estratégias de alimentação em culturas de *Pichia pastoris*.

Esta tese é um passo em frente no sentido estabelecer os fundamentos da EFC. Os métodos aqui desenvolvidos constituem uma contribuição para mudar o actual paradigma do desenvolvimento de meio e de processos de cultura celular para o de uma metodologia sistemática e holística.

Palavras-chave

Pichia pastoris · envirómica funcional celular · projecção a vias metabólicas latentes · deselvolvimento de meios de cultura

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Ala	alanine
Arg	arginine
Asp	aspartate
BSM	basal salts media
CFE	cell functional enviromics
DCW	dry cell weight
EFM	elementary flux mode (or elementary mode)
EM	elementary mode (or elementary flux mode)
EMP	Embden-Meyerhof-Parnas pathway
Glu	glutamate
GlyOH	glycerol
GPC	glycerophosphocholine
LV	latent variable
Lys	lysine
MFA	metabolic flux analysis
MFD	metabolic flux distribution
nRMSE	normalized root mean squared error
OD	optical density
PCA	principal component analysis
PLP	projection to latent pathways
PLS	partial least squares regression (or projection to latent structures)
pO_2	dissolved oxygen (as % of saturation)
PPP	pentose phosphate pathway
Pro	proline
PTM1	*Pichia* trace minerals 1
RMSE	root mean squared error
RQ	respiratory quotient
scFv	single-chain variable fragment
TCA	tricarboxylic acids (cycle)
WCW	wet cell weight

Chapter 1
Introduction

1.1 Context and motivation

Competition in the production of public domain biotherapeutics will set forth the need for global process optimization to increase titers, productivity and to control product quality. For that purpose, the integration of classical process optimization and control schemes with emerging systems biology methodologies is a rational step forward to improve the current standard in the industrial biotechnology.

Bioprocess optimization and control currently rely on empirical models that describe the bioreactor at the macroscopic scale. This approach neglects the intricate details of the cell factory and is thus limited in its potential applications. One example is the control of product quality wherein molecular-level properties come into play. For such problems, models that integrate macro-, micro- and molecular-level process parameters are essential for a successful optimization and/or control outcome.

It should be also recognized that in any cell culture process there is a large number of environmental variables that shape cellular physiology. One important implication is that the design space for process development, namely culture medium optimization and process control, is potentially very large. Empirical methods are not well suited to handle high-dimensional design spaces unless a substantial level of reductionism is applied, that however may result in suboptimal performance. Both industry and academia are now acknowledging the need to use state-of-the-art, systems-level genomics, proteomics and metabolomics tools for bioprocess optimization and control [1, 2].

For the purpose of process optimization and control, metabolomics, and specially exometabolomics, are of particular interest. Even though small changes on the regulation of a few transcripts (and on the concentration of the enzymes they code for) usually lead to small variations on the fluxes through metabolic pathways, it has substantial effects on the concentration of target metabolites. As a consequence, changes in the transcriptome and proteome are actually amplified in the

metabolome, rendering its analysis a valuable tool for the detection of metabolic adaptations [3, 4]. Additionally, the metabolites excreted by the cell to the culture medium are more readily available for analysis and quantification. Kell and co-workers [5] focused on what they called the "exometabolome", defending it provides a "metabolic footprint" of cellular metabolism, by reflecting which pathways are active within the cellular compartment at any given moment. Metabolic profiling methods are relatively fast, cheap and high-throughput, and, therefore, are eligible for on-line monitoring. The detection of extracellular metabolites other than the desired end-product shows which non-productive pathways are active and if possible should be down-regulated through appropriate control.

1.2 Cell functional enviromics

Functional enviromics is an emerging systems biology methodology that aims at discovering the function of the entirety of environmental factors – the envirome – on cellular regulation. It is the "environmental analog" of functional genomics. Functional genomics is now a well-established discipline that aims at unravelling gene functions and gene-gene interactions and how these set phenotypic traits. The concept of functional enviromics has been first set forth as a counterpart to functional genomics in tackling mental disorders such as schizophrenia [6]. Only very recently functional enviromics has been addressed in the context of cell physiology [7].

While the genome sets the phenotypic space of a cell, particular trajectories within it are primarily driven by the environment. Such gene-environment interactions are still poorly understood [8]. Cell functional enviromics may contribute to fill that gap. Cell functional enviromics may be defined as the envirome-wide cellular function reconstruction through the collection and systems-level analysis of dynamic envirome data. The key steps for the realization of a functional enviromics study are:

(i) setting the universe of cellular functions and envirome components,
(ii) collecting informative envirome data over time, and
(iii) systems-level analysis of dynamic envirome data to find relationships between environmental variables and cellular functions.

With the advances in systems biology, accurate genome-scale mechanistic models are becoming available for several microorganisms of industrial interest. Such metabolic networks contain the required information to enumerate all possible operational modes of the cells. With adequate systems biology tools, such as functional enviromics, one can investigate how those operational modes are controlled by the environment and/or how they modify the environment. The knowledge of which pathways are dominant as a function of environmental conditions would be a major advantage as it could allow the implementation of on-line pathway-oriented macroscopic control methods that actuate on the cellular system towards the global optimization of product titer and product quality.

1.3 *Pichia pastoris* expression system

The yeast *Pichia pastoris* was the chosen model organism to illustrate the application of cell functional enviromics in this thesis. More specifically, a *P. pastoris* X-33 strain constitutively expressing a scFv-type antibody under the pGAP promoter. A single-chain variable fragment (scFv), which for simplicity is usually described as an antibody fragment, is actually a fusion protein of the variable regions of heavy (V_H) and light (V_L) chains of immunoglobulins, joined together by a flexible linker peptide. Single-chain variable fragments and other recombinant antibody fragments are emerging as a more economical and customizable alternative to monoclonal antibodies (mAbs) for a wide range of diagnostic and therapeutic applications [9].

The yeast *Pichia pastoris* is now established in the academia and in the industry as an expression system of choice. The key aspects that explain the success of *P. pastoris* have been recently reviewed in Ahmad et al. [10]. Among those key aspects is the fast and inexpensive growth that reaches extremely high cell densities on inexpensive, chemically defined media [11, 12]. It can produce foreign proteins at very high levels both intracellularly or secreted to the medium [13] and is able to carry out post-translational modifications. Unlike other yeasts, it shows a strong preference for respiratory growth and a limited tendency for fermentation [14]. Techniques for genetic manipulation are simple and similar to *Saccharomyces cerevisiae,* with commercial expression kits available for both intracellular and secretory expression. Furthermore, the *P. pastoris* genome sequence and annotation is publically available [15] and curated genome-scale metabolic models can be found in the literature [16, 17].

The *Pichia* platform currently has over 70 products on the market or in late stage of development [18]. Applications in the industry range from treatment agents in chemical processes (e.g. phospholipase C, nitrate reductase), to enzymes for the food and feed industries [19] (e.g. trypsin, phytase), and research reagents (e.g. collagen, cystatin C). In addition, several biopharmaceutical products have been approved by regulatory agencies for human use. These include vaccines (Shanvac™ from Shanta/Sanofi), two therapeutic antibody fragments (Nanobodies® ALX00171 and ALX0061 from Ablynx) and several therapeutic proteins (e.g. Kabitor® from Dyax, Insugen® from Biocon, Shanferon™ from Shanta/Sanofi, Ocriplasmin from ThromboGenics, and heparin-binding EGF-like growth factor from Trillium).

Despite the growing importance of the *Pichia pastoris* expression system as an industrial workhorse, the literature is almost absent in systematic studies on how environmental factors affect cellular physiology.

The present PhD thesis focuses on the topic of functional enviromics applied to *Pichia pastoris* cultures. This work is motivated by the perception that the power of the envirome is many times underestimated in cell culture engineering in detriment to genetic engineering. Functional enviromics provides the ideal scientific basis for the in-depth understanding of the role of the envirome on cellular function.

1.4 Thesis objectives

The central goal of this PhD dissertation was the development of analytical methodologies and computational algorithms for pathway-level optimization of cell culture processes based on functional enviromics.

In order to achieve this, the following intermediate objectives were devised:

- Gather experimental data of recombinant *Pichia pastoris* cultures expressing a scFv antibody fragment. Explore the cell's physiological response to variations in bioreactor operating parameters and culture medium composition while optimizing sampling and analytical techniques for envirome characterization.
- Adapt and expand existent functional enviromics tools to process high-throughput envirome data. Implement algorithms in a robust systems biology framework suitable for large-scale process monitoring and control.
- Develop *Pichia pastoris* functional enviromics models using pathway, sequence and annotation data available in public databases, integrated with the algorithms developed in the previous step.
- Apply the developed functional enviromics models and tools to collected experimental data in order to generate knowledge on how envirome factors correlate and influence cell physiology. This knowledge can serve as basis for the design of optimized culture media formulations and operational strategies.

1.5 Thesis outline

This thesis is structured into six chapters. The current section, Chapter 1, is a short introduction to the thesis topic, scientific context and research objectives.

Chapter 2 is an extensive overview of the state-of-the-art based on a review manuscript previously published as a book chapter. It covers current analytical methodologies to measure the envirome, describes how elementary modes arise from metabolic networks, introduces a mathematical basis for cell functional enviromics and characterizes alternative approaches for the selection of elementary modes.

Chapters 3 to 5 represent the core research results divided into self-contained sections, each addressing a specific research question. In Chapter 3, culture media screening data from shake flask cultures is used to investigate sensitivity of central carbon metabolism and protein yield to variations in culture media composition. The core methodology is hybrid metabolic flux analysis, which combines classic metabolic flux analysis (MFA) – to estimate central carbon fluxes from a few

measured fluxes – with partial least squares regression (PLS) – to establish a link between central carbon fluxes and product yield.

Chapter 4 combines the same media screening data with projection to latent pathways (PLP) – the core functional enviromics algorithm – to assess how culture medium components up-regulate or down-regulate key metabolic pathways. The main novelty is the application of PLP in a way that allows inference of cause-effect relationships for culture medium components. The PLP method is evaluated for consistency on elementary mode selection and compared to the standard PLS regression.

In Chapter 5, cell functional enviromics is applied to *Pichia pastoris* data collected over time in a pilot-scale 50 L bioreactor, complemented with ^{1}H-NMR exometabolite profiling. PLP is used to identify key metabolic pathways with high correlation to the envirome and assess how activation of specific pathways varies over time. The resulting functional enviromics map (FEM) identifies links between envirome factors and cellular operation and may serve as basis to optimize media composition and feeding strategies for *Pichia pastoris* cultivations.

Finally, Chapter 6 summarizes the main findings and presents an integrated discussion with focus on impact and perspectives for future work.

References

1. Otero JM, Nielsen J (2010) Industrial Systems Biology. Biotechnol Bioeng 105:439–460
2. Teixeira AP, Carinhas N, Dias JML, et al. (2007) Hybrid semi-parametric mathematical systems: Bridging the gap between systems biology and process engineering. J Biotechnol 132:418–425
3. Raamsdonk LM, Teusink B, Broadhurst D, et al. (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19:45–50
4. Urbanczyk-Wochniak E, Luedemann A, Kopka J, et al. (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Rep 4:989–993
5. Kell DB, Brown M, Davey HM, et al. (2005) Metabolic footprinting and systems biology: The medium is the message. Nat Rev Microbiol 3:557–565
6. Van Os J, Rutten BP, Poulton R (2008) Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. Schizophr Bull 34:1066–82
7. Teixeira AP, Dias JML, Carinhas N, et al. (2011) Cell functional enviromics: Unravelling the function of environmental factors. BMC Syst Biol 5:92
8. Smith EN, Kruglyak L (2008) Gene-environment interaction in yeast gene expression. PLoS Biol 6:e83
9. Holliger P, Hudson PJ (2005) Engineered antibody fragments and the rise of single domains. Nat Biotechnol 23:1126–1136
10. Ahmad M, Hirz M, Pichler H, Schwab H (2014) Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 98:5301–17
11. Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol 13:329–332

12. Heyland J, Fu J, Blank LM, Schmid A (2010) Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnol Bioeng 107:357–68
13. Cregg JM, Vedvick TS, Raschke WC (1993) Recent Advances in the Expression of Foreign Genes in *Pichia pastoris*. Nat Biotechnol 11:905–910
14. Hagman A, Säll T, Piškur J (2014) Analysis of the yeast short-term Crabtree effect and its origin. FEBS J 281:4805–14
15. Mattanovich D, Callewaert N, Rouze P, et al. (2009) Open access to sequence: Browsing the *Pichia pastoris genome*. Microb Cell Fact 8:53
16. Chung BK, Selvarasu S, Andrea C, et al. (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb Cell Fact 9:50
17. Sohn SB, Graf AB, Kim TY, et al. (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol J 5:705–15
18. Research Corporation Technologies (2015) *Pichia* produced products on the market. http://www.pichia.com/science-center/commercialized-products/.
19. Spohner SC, Müller H, Quitmann H, Czermak P (2015) Expression of enzymes for the usage in food and feed industry with Pichia pastoris. J Biotechnol 202:118–34

Chapter 2
State-of-the-art review [1]

Abstract

In this chapter we explore the basic tools for the design of bioprocess monitoring, optimization, and control algorithms that incorporate a priori knowledge of metabolic networks. The main advantage is that this ultimately enables the targeting of intracellular control variables such as metabolic reactions or pathways directly linked with productivity and product quality. We analyze in particular design methods that target elementary modes of metabolic networks. The topics covered include the analysis of the structure of metabolic networks, computation and reduction of elementary modes, measurement methods for the envirome, envirome-guided metabolic reconstruction, and macroscopic dynamic modeling and control. These topics are illustrated with applications to a cultivation process of a recombinant *Pichia pastoris* X-33 strain expressing a single-chain variable fragment (scFv) antibody.

Keywords

bioprocess control · dynamic modeling · elementary modes · envirome measurement · metabolic networks

2.1 Introduction

Historically, process control for cell culture has relied on empirical models with cells treated as ‘‘black boxes’’. Purely descriptive empirical models based on measurements of the concentrations of biomass and normally only a few extracellular compounds, which completely neglect the

[1] Chapter published as: Isidro IA, Ferreira AR, Clemente JJ, Cunha AE, Dias JML and Oliveira R (2013) Design of pathway-level bioprocess monitoring and control strategies supported by metabolic networks. Adv Biochem Eng Biotechnol 132:193–215.

structure of the intracellular compartment, have been widely used for bioprocess optimization and control [1]. With the advances in systems biology, molecular biology data and mechanistic models for microorganisms of industrial interest are becoming available. Systems biology is expected to have a great impact on biotechnological processes including process control, enough to justify the coining of the term ‘‘industrial systems biology’’ [2].

Cell factories consist of complex, intricate networks of a large number of genes, proteins, and metabolites. At a higher hierarchical level, cells are part of larger networks comprising the environment as well as other cells or organisms [3]. As we learn more from genome-scale network reconstruction projects, it becomes apparent that the number of molecular interactions between the extracellular and intracellular environments is very large. Borenstein et al. [4] estimated that 8–11% of the metabolites in the metabolic networks of prokaryotic species originate from the environment. Indeed, cells take a large number of compounds from the environment to carry out their metabolic activity. These include inorganic ions and a large array of low molecular weight organic molecules such as sugars, vitamins, fatty acids and amino acids. As a consequence, cells leave a complex and informative metabolic footprint in the environment, which in yeast cultures may account for more than 100 metabolites [5]. Moreover, experiments with single-gene deletion mutants have shown that the metabolic footprint was sufficiently informative to classify the different mutants [6]. Larger macromolecules present in the environment, such as proteins, carbohydrates, and lipids, also play an important role in signal transduction pathways. Both the low and high molecular weight extracellular molecules form a natural extension of the intracellular biochemical networks of considerable complexity. Understanding the molecular interplay between extra- and intracellular components is essential to engineer the environment of cells more efficiently, namely for optimization of culture medium composition and design of process monitoring and control strategies that target intracellular control variables.

Metabolic networks can be used to interpret metabolic footprinting data and to study how the extracellular environment can be manipulated to control intracellular processes [7]. A few studies have addressed the development of dynamic macroscopic models for process control derived from metabolic networks. Haag et al. [8] showed that, for a class of macroscopic dynamic models, systems with complex intracellular reaction networks can be represented by macroscopic reactions relating extracellular components only with equivalent ‘‘input-output’’ behavior. Following a similar approach, Provost and Bastin [9] have developed macroscopic dynamic models for Chinese hamster ovary (CHO) cultures wherein the reaction mechanism is defined by the elementary modes of the metabolic network. An elementary mode (EM) can be defined as a minimal set of metabolic reactions able to operate at steady-state, with the enzymes weighted by the relative flux they need to carry for the mode to function [10]. As a result, each elementary mode can be viewed as a metabolic subnetwork, which, under the steady-state assumption, can be equivalently represented by a macroscopic reaction involving only extracellular substrates and end-products.

The main difficulty in macroscopic dynamic modeling based on elementary modes lies in the definition of the elementary mode weighting factors. As discussed later, any particular set of metabolic fluxes, or fluxome (i.e., phenotypic state), can be represented as a weighted sum of elementary modes. The magnitude of a weighting factor thus quantifies the contribution of the particular elementary mode to the overall phenotypic state. In Provost and Bastin [9], the elementary mode weighting factors were modeled by Michaelis-Menten kinetic laws as functions of extracellular concentrations. The analogy between Michaelis-Menten kinetics and elementary mode weighting factors is, however, not founded on mechanistic principles. Moreover, this approach gives rise to very complex nonlinear systems, which are difficult to identify. In Teixeira et al. [11] we developed hybrid macroscopic models structured by elementary flux modes for baby hamster kidney (BHK) cells. Instead of Michaelis-Menten kinetic laws, empirical modeling, namely artificial neural networks, was employed to model the elementary mode weighting factors as functions of extracellular physicochemical variables.

Another difficulty in macroscopic dynamic modeling based on elementary modes lies in the typically very high number of elementary modes. Indeed, the number of elementary modes increases exponentially with the size and complexity of the network [12]. However, most of these elementary modes are not active at preset environmental conditions [13]. It is thus not necessary to use the full set of elementary modes for a specific application. Of particular interest is the subset of elementary modes describing a collection of measured phenotypic data. The importance of this lies in the fact that the internal fluxes are not independently distributed but strictly constrained by external fluxes through the pathways at steady-state [14]. Therefore, the challenge is how to select the subset of elementary modes that describe a physiological state of interest. Effective reduction of elementary modes is mandatory to reduce the complexity of the final model and facilitate the design of process control.

In this chapter we explore the basic tools to design bioprocess modeling, monitoring, and control algorithms based on metabolic networks. We start by reviewing basic properties of metabolic networks, metabolic modeling, and elementary modes. The envirome layer of information affects critical bioprocess monitoring and control challenges. The envirome consists of the total quantitative collection of physicochemical properties that define the extracellular environment. These are the properties that can be individually or collectively manipulated in a process and also the ones that are more easily measured in real time. We thus dedicate a section to the measurement of the envirome. In a recent paper we explored the possibility of metabolic reconstruction from dynamic envirome data. We have named this technique ‘‘cell functional enviromics’’ [7]. We show here how this methodology can be used to design bioprocess control algorithms that target intracellular control variables such as fluxes or pathways.

2.2 Genome-scale networks lay the foundation

In January 2012, the genome online database (GOLD) recorded 3 065 completed bacterial genome sequences and 7 755 more sequencing projects underway [15]. Furthermore, the metagenomes (genome of mixed cultures) of 340 sample communities were also recorded in the same database, with 9% of them from engineered mixed microbial systems (wastewater, solid waste, or bioremediation) [16]. Genome-scale networks are constructed on the basis of the complete genome annotation. Identified genes may be associated with metabolic enzymes, membrane transporters, signal transduction, or regulatory control. Combining genome annotation with basic biochemical information currently available in several databases (e.g., KEGG [17] and BioCyc [18] databases), it is possible to reconstruct the majority of the metabolic reactions network and also the associated exometabolome [19]. At least 62 genome-scale metabolic models have been reconstructed for single organisms, representing 37 genera [20, 21], including organisms of industrial relevance such as *Escherichia coli* [22], *Saccharomyces cerevisiae* [23], *Pichia pastoris* [24, 25], and many others. Metabolic networks convey critical information about the interaction between the extra and intracellular phases, which is essential for design of advanced process control strategies that target intracellular control variables.

2.2.1 Structure of metabolic networks

Studies on the architecture of metabolic networks of microorganisms from the different domains of life (Eukarya, Bacteria, and Archaea) have shown that cellular metabolism has a scale-free topology, which means that most metabolites participate in only one or two reactions, while a few, such as adenosine triphosphate (ATP) or pyruvate, are metabolic hubs participating in dozens of metabolic reactions [26]. In the context of bioprocess control it is particularly relevant to analyze how metabolic networks interact with the extracellular environment. Bernhardsson et al. [27] have analyzed the metabolic networks of 134 bacterial species and concluded that common reactions are found at the center of the network and decrease as we move to the periphery of the metabolic network, i.e., closer to the metabolites that cross the cellular membrane. Borenstein et al. [4] have determined the seed set compounds (i.e., exogenously acquired compounds) for each of the 478 prokaryotic species with metabolic networks available in the KEGG database. They found that about 8–11% of the compounds in the whole metabolic network correspond to the seed set and that each organism possesses a characteristic seed set. Moreover, comparing seed sets of different organisms enabled them to trace the evolutionary history of both metabolic networks and growth environments across the tree of life, supporting the ‘‘reverse ecology’’ principle. These structural features are pivotal for the design of process control strategies based on metabolic networks. On the one hand, given the high number and specificity of metabolites that cross the cellular membrane, the measurement of the metabolic footprint, i.e., the complete set of extracellular metabolites, might carry sufficient information to reconstruct a large number of intracellular metabolic processes. On the other hand, the concentrations of many such

extracellular metabolites can be manipulated in order to control intracellular processes linked to product yield and quality.

2.2.2 Material balances

The list of metabolic reactions identified in a genome-scale reconstruction project can be translated into a stoichiometric matrix, $\mathbf{A}$, with dim($\mathbf{A}$) = $m \times q$, where m is the number of intracellular metabolites and q is the number of metabolic reactions. The material balances over the intracellular metabolites take the following general form:

$$\frac{d\mathbf{c}^i}{dt} = \mathbf{A} \times \mathbf{v} - \mu\, \mathbf{c}^i \tag{2.1}$$

where $\mathbf{c}^i$ is the vector of intracellular concentrations with dim($\mathbf{c}^i$) = m, $\mathbf{v}$ is the vector of intracellular fluxes, and μ is the specific growth rate. Under the pseudo-steady-state hypothesis, intracellular metabolites do not accumulate and the dilution term is much smaller than the net turnover of metabolites, thus Eq. (2.1) simplifies to

$$\begin{cases} \mathbf{0} = \mathbf{A} \times \mathbf{v} \\ \mathbf{v}_j \geq 0 \end{cases} \tag{2.2}$$

The inequality constraints in Eq. (2.2) refer to the subset j of irreversible reactions with nonnegative flux values. Equation (2.2) expresses an undetermined system of algebraic equations because $q \gg m$, and thus it has no unique solution. The universe of solutions of Eq. (2.2) forms a polyhedral cone in the fluxome solution space whose edges correspond to independent elementary modes (elementary modes are discussed in more detail in the next section).

Equation (2.2) applies only to balanced intracellular metabolites. For extracellular metabolites the net accumulation is nonzero and the following equation applies:

$$\begin{cases} \mathbf{b} = \mathbf{A}' \times \mathbf{v} \\ \mathbf{v}_j \geq 0 \end{cases} \tag{2.3}$$

with $\mathbf{b}$ the vector of fluxes of extracellular metabolites across the cellular membrane and $\mathbf{A'}$ the stoichiometric matrix of such extracellular metabolites.

2.2.3 Elementary modes

Elementary mode analysis has become a widespread technique for systems-level metabolic pathway analysis [28, 29]. An elementary mode can be defined as a minimal set of enzymes able to operate at steady state, with the enzymes weighted by the relative flux they need for the

mode to function [10]. The universe of elementary modes of a given metabolic network defines the full set of non-decomposable steady-state flux distributions that the network can support. Any particular steady-state flux distribution can be expressed as a nonnegative linear combination of elementary modes.

As such, the phenotype of a cell, as defined by its fluxome, $\mathbf{v}$, can be expressed as a weighted sum of the contribution of each elementary mode:

$$\mathbf{v} = \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \cdots + \lambda_k \mathbf{e}_k = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \tag{2.4}$$

where $\mathbf{e}_i$ is an elementary mode vector with dim($\mathbf{e}_i$) = q, λ_i is the weighting factor of $\mathbf{e}_i$, k is the number of elementary modes, and dim($\mathbf{v}$) = dim($\mathbf{e}_i$) = q is the number of metabolic reactions on the metabolic network. Geometrically the elementary modes correspond to the edges of the polyhedral cone in the fluxome solution space (Figure 2.1).

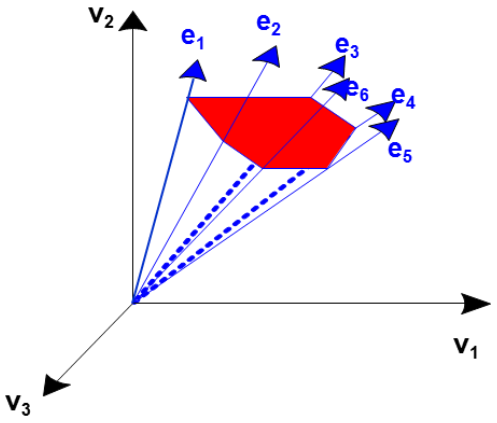


Figure 2.1 – Fluxome solution space of a metabolic network in steady state obeying to the material balances of Eq. (2.2). The solution space has the shape of a polyhedral cone whose edges are the elementary modes.

The elementary mode matrix, $\mathbf{EM}$, is obtained by concatenating all the $\mathbf{e}_i$ vectors into a $q \times k$ matrix:

$$\mathbf{EM} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \ldots \ \mathbf{e}_k] \tag{2.5}$$

Multiplying the $\mathbf{EM}$ matrix by the stoichiometric matrix of the extracellular metabolites, $\mathbf{A'}$, one obtains the elementary mode stoichiometric matrix:

$$\mathbf{A}_{\mathbf{EM}} = \mathbf{A}' \times \mathbf{EM} \tag{2.6}$$

The dimension of $\mathbf{A_{EM}}$ is $m' \times k$, where m' is the number of extracellular metabolites. Each column of $\mathbf{A_{EM}}$ contains the stoichiometry of extracellular metabolites for the particular elementary mode. This matrix holds critical information for process control, since it defines the theoretical metabolic footprint of each elementary biochemical state of the cell. The specific reaction rates of extracellular compounds are given by

$$\mathbf{b} = \mathbf{A_{EM}} \times \lambda \tag{2.7}$$

As shown later, these rates can be used to formulate macroscopic dynamic models of extracellular compounds.

2.2.4 Example: elementary modes of *P. pastoris*

To illustrate the elementary mode concept, we have built a *P. pastoris* metabolic network based on the KEGG database and papers by Chung et al. [24] and Çelik et al. [30]. The genes associated with each reaction are in most cases known and can be found in [24]. The network included the following processes/pathways: uptake reactions (glycerol, sulfate, phosphate, and ammonia), glycolysis/gluconeogenesis, pentose phosphate pathway, tricarboxylic acid cycle (TCA), biosynthesis of amino acids, biosynthesis of macromolecular components of biomass (nucleotides, lipids, carbohydrates, and proteins), biosynthesis of a single-chain variable fragment (scFv), interconversion of folate compounds, oxidative phosphorylation, and energy interconversions. The metabolic network was further simplified by lumping together in single reactions the consecutive reactions in the pathways for synthesis and degradation of biomass and product precursors. The stoichiometry of ATP, nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADPH), flavin adenine dinucleotide ($FADH_2$), and H_2O were also accounted for in the metabolic reactions in order to close the balance of oxygen, hydrogen, and phosphorus. It was assumed a fixed P/O ratio of 2 mol-ATP/mol-NAD(P)H and of 1 mol-ATP/mol-$FADH_2$. The resulting metabolic network for glycerol as carbon source has 104 reactions (thus 104 fluxes), 90 intracellular metabolites, and 16 extracellular metabolites (17% of all metabolites).

The open source bioinformatics software METATOOL 5.0 [31] was used to compute the elementary modes for the *P. pastoris* metabolic network. The total number of elementary modes was 4 119. Figure 2.2 shows a representation of the yields of biomass and product on glycerol and oxygen obtained through stoichiometric analysis of elementary modes.

The number of elementary modes for glycerol feeding was 2 520, 960, 600, and 39 for biomass growth, scFv synthesis, simultaneous biomass growth and scFv synthesis, and catabolism, respectively. From Figure 2.2 it can be seen that the yields on glycerol for biomass growth and scFv synthesis vary from 0.42 to 0.73 g-X/g-S and 0.26 to 0.64 g-X/g-S, respectively, and that the yields on oxygen increase with the yields on glycerol. The elementary modes with lowest yields on glycerol

and oxygen are those that include the metabolic reactions involved in the secretion of organic acids from TCA, namely succinate for biomass growth and citrate for scFv synthesis. On the other hand, the elementary modes with the highest yields on glycerol and oxygen involve the metabolic reaction of the pentose phosphate pathway.

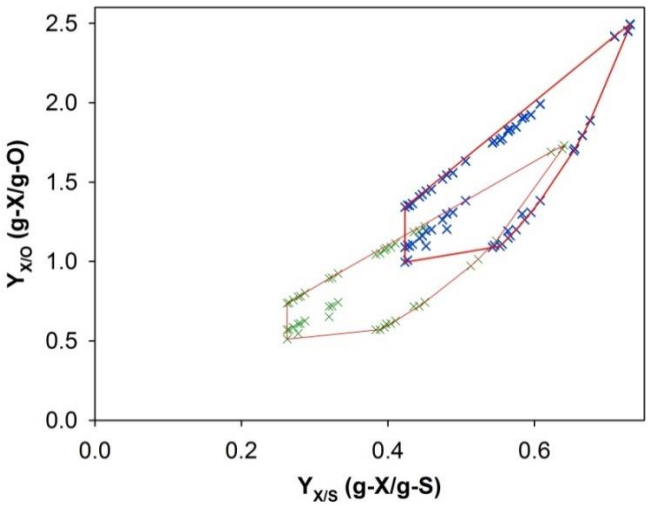


Figure 2.2 – Bounded convex hull in the space of yields of biomass (blue crosses) and product (green crosses) on glycerol and oxygen.

2.3 Measuring the envirome

The whole set of physical and chemical properties that define the environment of cells is known as the ‘‘envirome’’. The envirome is the critical layer of information for bioprocess monitoring and control, since it can be readily measured and/or manipulated in real time. The vast majority of envirome components are also metabolites, of which some are provided by the culture medium and many others are produced inside the cells and then secreted or excreted into the environment. The ‘‘metabolome’’ was defined by Oliver et al. [32] as the qualitative and quantitative collection of all metabolites, that is, all the low molecular weight molecules present in a cell, which are also participants in general metabolic reactions and that are required for the maintenance, growth, and normal function of the cell. The metabolome can be subdivided into the endometabolome (intracellular metabolites) and exometabolome (extracellular metabolites) [33]. Depending on whether the analysis is being targeted for the endo or exometabolome and depending on the analytical detail and quantitative power desired, there are several measurement strategies available (Table 2.1); some such strategies have the potential for real time monitoring and are therefore suitable for process control.

Table 2.1 – Different levels of metabolome analysis

Metabolite target analysis	Identification and quantification focused on one or a few metabolites related to a specific pathway [34]
Metabolite profiling (or metabolic profiling)	Identification and quantification of a selected group of metabolites, e.g., metabolites from a specific metabolic pathway or a specific compound class, such as amino acids, organic acids, or carbohydrates [34]
Metabolomics	Identification and quantification of all metabolites in a biological system. Sample preparation method must retain all metabolites. Analytical technique must be suited to measure metabolites over a broad range of concentrations and needs high discriminatory power
Metabonomics	Analysis of tissues and/or biological fluids to detect changes caused by disease or therapeutic treatments [35]
Metabolic fingerprinting	Fast, high-throughput analysis of intracellular metabolites to provide a characterization of the cells for sample classification. Analytical technique must allow sample discrimination, but it is not required to identify and quantify all the metabolites individually [34]
Metabolic footprinting	Fast, high-throughput analysis of the surrounding medium to characterize the cells based on their exometabolome. As with fingerprinting, it is not necessary to identify and quantify all the metabolites individually to allow sample discrimination [6]

Measuring the endometabolome (metabolic fingerprinting) is not as straightforward as measuring the exometabolome (metabolic footprinting) given the complex sample preparation protocols and the higher number of intracellular metabolites. Endometabolome analysis requires separation of cells from extracellular medium followed by cell breakage. In addition, the rapid turnover inherent to intracellular metabolites, which can be under one second for microbial systems [36], results in the need for a rapid quenching step to halt metabolism. In contrast, the turnover rates of exometabolites are much lower given the low volume ratio between the intracellular and extracellular phases.

Current analytical techniques for exo- or endometabolome analysis include nuclear magnetic resonance (NMR) spectrometry [37, 38] and mass spectrometry (MS) [39]. Either of them can be coupled to separation methods for higher resolution. These hyphenated methods include, for instance, capillary electrophoresis coupled to mass spectrometry (CE-MS) [40], gas chromatography mass spectrometry (GC-MS) [41], and liquid chromatography coupled to nuclear magnetic resonance spectrometry (LC-NMR) [42]. For a detailed review on the application of such methods to metabolomics refer to [5, 43, 44].

A fast and low-cost technique is ^{1}H-NMR. The time of spectral acquisition ranges from 2 to 10 minutes per sample, and automatic samplers can be used. Figure 2.3 shows a ^{1}H-NMR spectrum of the supernatant of *P. pastoris* culture samples. In preliminary offline tests with this technique we detected over 20 metabolites in the extracellular phase. Bundy and co-workers [45] have detected over 80 metabolites in the extracellular medium of *P. pastoris* cultures using ^{1}H-NMR and GC-MS as complementary techniques.

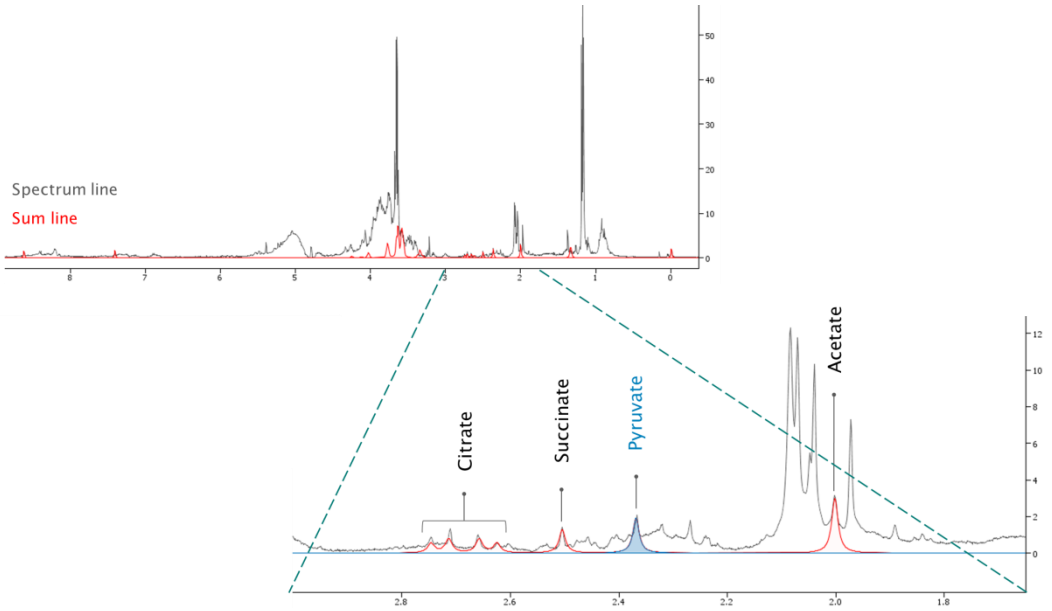


Figure 2.3 – ^{1}H-NMR spectrum for a *P. pastoris* supernatant sample. The black line is the acquired spectrum, whereas the red line is the estimated sum of the individual spectra for identified metabolites. Looking more closely into specific regions of the spectrum it is possible to identify key metabolites.

Knowledge of the metabolome is useful since it is very closely related to cellular phenotype. Because changes upstream accumulate downstream, changes in the transcriptome and proteome are found amplified in the metabolome. As a result, the metabolome allows the detection of changes that have a very small effect on metabolic fluxes [37, 46]. Metabolic fluxes, which can be regarded as the phenotype of a cell, are regulated not only at transcription and translation levels, but also by means of posttranslational events, and as such the metabolome is considered closer to the phenotype than the transcriptome or proteome [34, 47].

Moreover, metabolites are not organism specific, which means the techniques are equally applicable to prokaryotic, fungal, plant and animal cells. Even though a lot of progress has been made towards enabling whole metabolome quantification, these techniques still face challenges related to the inherent characteristics of the metabolome. The size of the metabolome varies greatly, depending on the organism studied. The nature of the metabolites, whether they are polar or nonpolar, volatile or nonvolatile, also influences the analysis, and most methods are biased towards some group of metabolites. In addition, the concentration of different metabolites extends over several orders of magnitude [48], thus adding difficulty to the task of quantifying all metabolites with a single technique. However, quantification of the whole metabolome is not essential for the purposes of process monitoring and control, as a subset of key metabolites is enough to infer cell function.

2.4 Elementary mode reduction

The number of elementary modes increases geometrically with the size of the network. The typically very high number of elementary modes denotes the innate adaptability and robustness of biological networks. As a consequence, the computation of elementary modes suffers from combinatorial explosion, particular for genome-scale networks. The central carbon metabolism of a genome-scale reconstructed *E. coli* metabolic network has approximately 26 million EMs [49]. It is essential to reduce such large numbers of elementary modes according to some criterion in order to decrease the computational power requirements. Indeed, not all calculated elementary modes are thermodynamically feasible or even physiologically reachable [50]. Several methods have been developed to reduce the number of elementary modes, founded on different principles. In what follows we review some of them.

2.4.1 Reduction based on network structural properties

Elementary modes can be reduced on the basis of structural information of the metabolic network without the use of experimental data. Figueiredo et al. [51] presented a method based on the ranking of elementary modes in increasing order of number of reactions. This approach enables identification of the K shortest elementary modes, which are in principle energetically more efficient. Song and Ramkrishna [14] proposed a reduction algorithm based on the effect of elementary modes on the convex hull volume. The principle consists in removing the elementary modes with negligible contribution to the convex hull volume of the original network. This allowed a priori reduction from the initial 369 to a final set of 35 elementary modes for a yeast metabolic network fermenting both glucose and xylose without using experimental data.

2.4.2 Reduction based on thermodynamic properties

Elementary modes can also be discriminated and reduced on the basis of metabolic reaction thermodynamics. The main assumption is that metabolic networks have evolved over time in the sense that cellular regulatory mechanisms were created that favor efficient pathways with low entropy generation. Wlaschin et al. [52] demonstrated, with experimentally determined intracellular fluxes, that elementary mode weighting factors are inversely correlated with the entropy generated by the involved metabolic reactions. Zhao and Kurata [53] proposed a method for correlating enzyme activity and flux distribution which uses Shannon's maximum-entropy principle, a measure of system complexity, as an objective function to estimate the enzyme control flux.

2.4.3 Reduction based on flux data

Several methods have been proposed to eliminate elementary modes on the basis of measured flux data. The equation that applies here is Eq. (2.4); however, the number of elementary mode weighting factors is in general much larger than the number of metabolic fluxes, thus the system is largely undetermined. Palsson and co-workers [54, 55] suggested linear optimization methods to determine how extreme pathways (the systemically independent subset of elementary modes) contribute to a given (measured) steady-state flux distribution. There is a range of possible nonnegative weighting values associated to extreme pathways that produce a given steady-state flux distribution. This range was calculated by maximizing and minimizing the extreme pathway weighting factors, resulting in the so-called α-spectrum. The allowable ranges for the values of λ_i were computed as:

$$\begin{aligned} &\max \lambda_i \\ &\text{subject to Eq. (2.4)}, \quad i = 1, \ldots, k, \quad 0 \le \lambda_i \le 1 \\ &\min \lambda_i \\ &\text{subject to Eq. (2.4)}, \quad i = 1, \ldots, k, \quad 0 \le \lambda_i \le 1 \end{aligned} \tag{2.8}$$

Wang et al. [56] presented a method to calculate the elementary mode coefficients for a large set of elementary modes by devising a quadratic program to explore the possibility and performance of using a subset of the elementary modes to reconstruct flux distributions. Alternatively, a framework based on elementary mode analysis and the convex properties of elementary modes was developed to calculate flux regulation coefficients (FRC) corresponding to an appropriate fractional operation of this mode within the complete set of elementary modes [57].

Schwartz and Kanehisa [58] showed that a combination of structural and kinetic modeling in yeast glycolysis significantly constrains the range of possible behaviors of a metabolic system. All elementary modes are not equal contributors to physiological cellular states, and this approach may open a direction towards a broader identification of physiologically relevant elementary modes among the very large number of stoichiometrically possible modes.

Very recently, Bastin et al. [59] developed a methodology to compute a decomposition of admissible flux vectors in a minimal number of elementary modes without explicitly enumerating all of them. They demonstrated that the vector of admissible weighting vectors ($\boldsymbol{\lambda}$) rewritten as

$$\boldsymbol{\lambda} = \sum_{i=1}^{k} \beta_i \, \mathbf{h}_i, \quad \beta_i \ge 0, \quad \sum_{i=1}^{k} \beta_i = 1 \tag{2.9}$$

is necessarily an admissible $\boldsymbol{\lambda}$ satisfying Eq. (2.7). In this case, the convex polytope, $\mathbf{H} = [h_1 \; h_2 \; \ldots \; h_k]$, contains a number of solutions equal to the number of measurements. Each polytope solution represents a minimal flux distribution given by $\hat{\mathbf{v}}_i = \mathbf{EM} \times \mathbf{h}_i$ and may be viewed

as the simplest pathways that satisfy the pseudo-steady-state assumption and the constraints imposed by the extracellular measurements defined in Eq. (2.7).

2.4.4 Example: reduction of the elementary modes by weighting factor minimization

Here we illustrate the method proposed by Schwartz and Kanehisa [58] for elementary mode reduction. This method identifies a subset of elementary modes by minimizing the sum of weighting factors (λ)

$$\min \sum_{i=1}^{k} \lambda_i \quad \text{subject to Eq. } (2.4) \tag{2.10}$$

This method was applied to the previously described *P. pastoris* metabolic network including 4 119 elementary modes. The results are shown in Table 2.2 for three distinct time points. Only 17 elementary modes were obtained with nonzero weighting factors. Several of these are selected at least twice in the three different phases of the culture. Note that this method basically selects the elementary modes which are closest to the actual biological state by minimizing the sum of weighting factors.

Table 2.2 – Elementary mode reduction results for three distinct culture time points

t = 30.7 h		t = 71 h		t = 119 h	
EM	λ	EM	λ	EM	λ
EM29	0.5324	EM228 [a]	0.9691	EM228 [a]	0.0745
EM153	0.2148	EM193 [a]	0.4303	EM193 [a]	0.0370
EM128	0.1480	EM219	0.1038	EM144	0.0110
EM189 [a]	0.1259	EM185 [a]	0.0773	EM189 [a]	0.0075
EM159	0.0021	EM189 [a]	0.0554	EM18 [a]	0.0067
EM51	0.0002	EM290 [a]	0.0449	EM290 [a]	0.0052
		EM206	0.0028	EM185 [a]	0.0010
		EM18 [a]	0.0002	EM45	0.0007
				EM120	0.0006
				EM177	0.0003

[a] Elementary modes that are selected at least twice

2.5 Pathway-level process control

Upon the identification of the most significant elementary modes, macroscopic dynamic models can be derived with implicit intracellular structure, which can then be used for process monitoring and control. For a stirred tank bioreactor, such material balance equations take the following general form:

$$\frac{d\mathbf{c}}{dt} = \mathbf{b}X - D(\mathbf{c} - \mathbf{c_{in}}) + \mathbf{Q} \tag{2.11}$$

In Eq. (2.11), the state space vector, $\mathbf{c}$, is formed by the concentrations of extracellular compounds, X is the biomass concentration, D is the dilution rate, $\mathbf{c_{in}}$ is the concentration of extracellular compounds in the inlet stream, and $\mathbf{Q}$ is the vector of gas-liquid transfer rates of volatile extracellular compounds. Note that Eq. (2.11) is analogous to the state-space equation proposed by Dochain and Bastin to design adaptive state estimation and control algorithms [1]. The main difference lies in the fact that the extracellular fluxes, $\mathbf{b}$, and the intracellular fluxes, $\mathbf{v}$, are functions of the elementary flux mode weighting factors instead of the traditional reaction kinetics:

$$\begin{bmatrix} \mathbf{v} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{EM} \\ \mathbf{A'} \times \mathbf{EM} \end{bmatrix} \lambda \tag{2.12}$$

An important implication is that any state-space solution of Eqs. (2.11-2.12) obeys the steady-state stoichiometric constraints imposed by the metabolic network.

The main difficulty in deriving these models is the definition of the elementary mode weighting factors as functions of environmental properties. Provost and Bastin [9] employed Michaelis-Menten kinetic laws, resulting in very complex nonlinear systems, which are very difficult to identify. Teixeira et al. [11] have developed hybrid macroscopic models structured by elementary modes using neural networks to model the respective weighting factors. In any case, the effective reduction of the initially very large number of elementary modes is critical to decrease the complexity and to obtain a final parsimonious model.

In a set of recent studies [11, 13] we investigated the systematic reconstruction of metabolic processes based on regression analysis of elementary mode weighting factors against measured environmental effectors. We have called this technique ‘‘cell functional enviromics’’ [7]. The principle is depicted in Figure 2.4. While the genome sets the structure of elementary modes, the envirome sets the relative contribution of each elementary mode to a given flux phenotype observation. While functional genomics studies genome-wide cellular function reconstruction through the collection and analysis of transcriptome or proteome data over time, functional

enviromics studies the reconstruction of cellular function through the collection and analysis of dynamic envirome data. Functional enviromics applies the following main steps:

(i) Compute the elementary mode matrix, $\mathbf{EM}$, from the microorganism metabolic network;

(ii) Acquire informative envirome data over time and organize it in the form of an envirome data matrix $\mathbf{X} = \{c_{i,j}\}$, a $M \times N$ matrix of N envirome factors, and respective measured flux data, $\mathbf{R} = \{\mathbf{b}_i\}$, a $M \times q'$ matrix of measured fluxes;

(iii) Apply systems-level analysis of dynamic envirome data $\mathbf{X}$ and $\mathbf{R}$ to find relationships between environmental variables, $c_{i,j}$, and elementary modes weighting factors, $\lambda_{i,j}$.

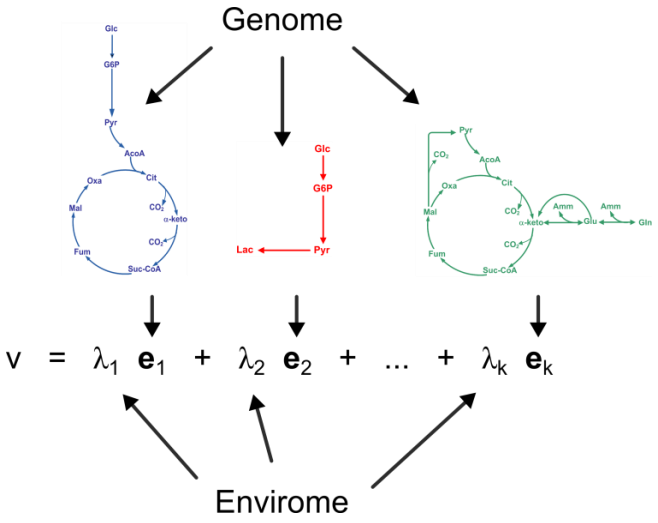


Figure 2.4 – Functional genomics versus functional enviromics. The genome sets the structure of elementary modes. The activation of elementary modes is controlled by the environment.

In what follows, we describe a possible functional enviromics algorithm based on the previous work by Ferreira et al. [13].

2.5.1 Functional enviromics algorithm

Among the whole set of elementary modes, the subset of elementary modes which is tightly linked to the envirome can be effectively determined by regression analysis of flux data, $\mathbf{R} = \{\mathbf{b}_i\}$ against envirome data, $\mathbf{X} = \{c_{i,j}\}$, satisfying the following criteria:

a) Maximize the captured variance of envirome data $\mathbf{X} = \{c_{i,j}\}$ and of flux data $\mathbf{R} = \{\mathbf{b}_i\}$;

b) Maximize correlation between elementary modes weighting factors and envirome variables;

c) Minimize the number of elementary modes required to capture a given variance of $\mathbf{R} = \{\mathbf{b}_i\}$ and $\mathbf{X} = \{c_{i,j}\}$, i.e. minimize redundancy.

These criteria can be fulfilled by maximizing the covariance between envirome data, $\mathbf{X} = \{c_{i,j}\}$, and respective measured flux data, $\mathbf{R} = \{\mathbf{b}_i\}$, according to the formula:

$$\underset{\mathbf{I}}{\text{Maximize}} \quad \text{cov}(\mathbf{X}, \mathbf{R}) \qquad \text{s.t.} \quad \begin{cases} \mathbf{R} = \mathbf{\Lambda} \times \mathbf{EM}^T \\ \mathbf{\Lambda} = \mathbf{X} \times \mathbf{I}^T \end{cases} \tag{2.13}$$

with $\mathbf{EM} = \{\mathbf{e}_i\}$ a $q \times K$ matrix of K elementary cellular functions, $\mathbf{e}_i$, each of length q, $\mathbf{\Lambda} = \{\lambda_i\}$ a $M \times K$ matrix of weight vectors λ_i of elementary modes (dim($\boldsymbol{\lambda}_i$) =M), and $\mathbf{I} = \{I_{i,j}\}$ a $K \times N$ matrix of intensity parameters, which are the degrees in Eq. (2.13). Several methods can be used to solve Eq. (2.13). One efficient method consists in a one by one decomposition of elementary modes according to Eqs. (2.14-2.16):

$$\mathbf{X} = \mathbf{T} \times \mathbf{W}^T + \mathbf{EF}_\mathbf{X} \tag{2.14}$$

$$\mathbf{R} = \mathbf{\Lambda} \times \mathbf{EM}^T + \mathbf{EF}_\mathbf{R} \tag{2.15}$$

$$\mathbf{\Lambda} = \mathbf{T} \times \mathbf{B}^T + \mathbf{EF}_\mathbf{\Lambda} \tag{2.16}$$

with $\mathbf{EF}_i$ the residuals matrices that are minimized, $\mathbf{W}$ a matrix of loading coefficients and $\mathbf{B}$ a matrix of regression coefficients. Finally, the intensity matrix $\mathbf{I}$ is given by:

$$\mathbf{I} = \mathbf{B} \times \mathbf{W}^T \tag{2.17}$$

The result of this procedure is the discrimination of a minimal set of elementary modes that is tightly linked with medium composition. The information can finally be organized in a $N \times K$ data array, called functional enviromics map:

$$\text{Functional enviromics map} = \mathbf{I}^T = \{I_{j,i}\}, \quad j = 1, \ldots, N, \quad i = 1, \ldots, K \tag{2.18}$$

The rows represent envirome factors, columns represent elementary modes and $I_{j,i}$ the relative "intensity" of up- or down-regulation of elementary cellular functions i by medium factor j.

2.5.2 Example: metabolic process control of *P. pastoris* cultures

We study here the optimization and control of a pilot 50 L fermentation of a constitutive *P. pastoris* X-33 strain expressing a single-chain antibody fragment; for details see [60]. The reactor was inoculated at a starting volume of 15 L. Cultivation temperature was controlled at 30 °C, and pH was controlled at 5.0 with addition of ammonium hydroxide 25%, which was also the main nitrogen source for the culture. The airflow rate was kept constant at 1 800 L/h throughout

the fermentation. Overhead pressure was controlled at 200 mbar. Glycerol feeding and dissolved oxygen (DO) control was divided into three phases:

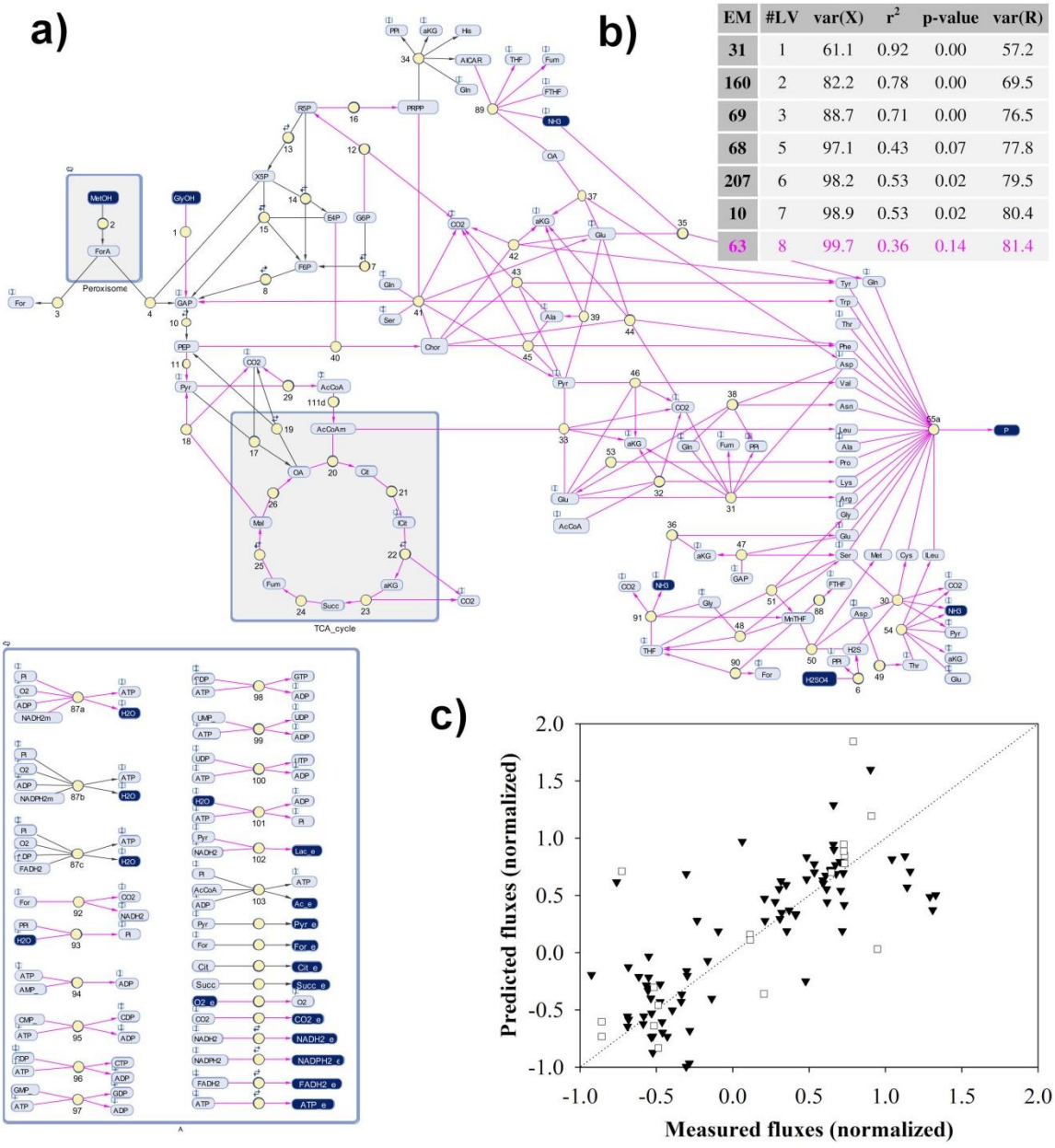
(i) Glycerol batch phase – The reactor was operated initially in batch mode, starting with a glycerol concentration of 40 g/L. DO drops very slowly and remains close to saturation levels;
(ii) Glycerol fed-batch phase – An exponential feeding program is initiated once the concentration of biomass reaches the level of 18 g-DCW/L. It is in this phase that cell density increases significantly and DO decreases more rapidly. Once the DO reaches 50%, it is kept at that level by automatic closed-loop control, manipulating the stirrer speed between 300 and 1 000 rpm;
(iii) Oxygen transfer limitation phase – Once the stirrer speed reaches the maximum level of 1 000 rpm, DO decreases very rapidly and the glycerol feeding program is aborted. From this point on, the DO is kept constant at a low level (e.g., 3–5%) by closed-loop manipulation of the glycerol feeding rate (DO-stat feeding control strategy).

We have investigated how the calculated elementary modes in the example in Section 2.2.4 correlate with the environmental parameters by applying the previously described functional enviromics algorithm. Measured environmental parameters comprised the temperature (T), pH, stirrer speed, pressure, and the concentrations of dissolved oxygen, glycerol, biomass, product (scFv), and inorganic compounds (NH_4^+, K^+, Ca^{2+}, Mg^{2+}, S, and P). From the concentrations of inorganic salts we have calculated the ionic strength and the osmolarity. The measured rates were those of biomass, glycerol, oxygen, carbon dioxide, and product.

The initial number of elementary modes was 4 119, of which a small set of eight elementary modes were discriminated by functional enviromics (Figure 2.5b). The total explained variance of measured fluxes was 81.4%, and the correlation between predicted and measured fluxes was acceptable (Figure 2.5c) given the high level of noise in the measured rates.

Of the eight identified elementary modes, three made by far the largest contribution. Elementary modes 68 and 69 describe the biomass growth, and elementary mode 63 describes the product synthesis (marked in pink in the network of Figure 2.5). We have built a macroscopic model with these three elementary modes. The comparison between the model simulation and experimental measurements is shown in Figure 2.6.

From the initial 4 119 elementary modes, 2 520, 960, 600, and 39 are for biomass growth, scFv synthesis, simultaneous biomass growth and scFv synthesis, and catabolism, respectively. It is interesting to note that the functional enviromics algorithm identified EM 63 for product synthesis, which belongs to the second group of 960 elementary modes. This suggested that the product is cell growth dissociated, which is in agreement with our previous study [60]. Analysis of the weighting factor dynamics (Figure 2.6b) clearly shows that the product synthesis elementary mode



EM	#LV	var(X)	r^2	p-value	var(R)
31	1	61.1	0.92	0.00	57.2
160	2	82.2	0.78	0.00	69.5
69	3	88.7	0.71	0.00	76.5
68	5	97.1	0.43	0.07	77.8
207	6	98.2	0.53	0.02	79.5
10	7	98.9	0.53	0.02	80.4
63	8	99.7	0.36	0.14	81.4

Figure 2.5 – Functional enviromics of a *P. pastoris* X-33 strain expressing a scFv antibody: a) metabolic network, **b)** subset of elementary modes with highest correlation with the environmental variables. Elementary mode 63 (marked in pink in the network and in the table) refers to the scFv biosynthesis. **c)** Predicted versus measured fluxes by the method of Eq. (2.11)**.**

peaks when the weighting factors for cell growth are almost zero. The analysis of three additional fermentations further confirmed that product synthesis is cell growth dissociated and that the final product titer increases with the biomass concentration time integral. The maximum product titer and productivity could be achieved by maximizing the glycerol feeding rate, through an accurate DO-stat glycerol feeding controller, at very low DO set-points in the range of 3-5% [60].

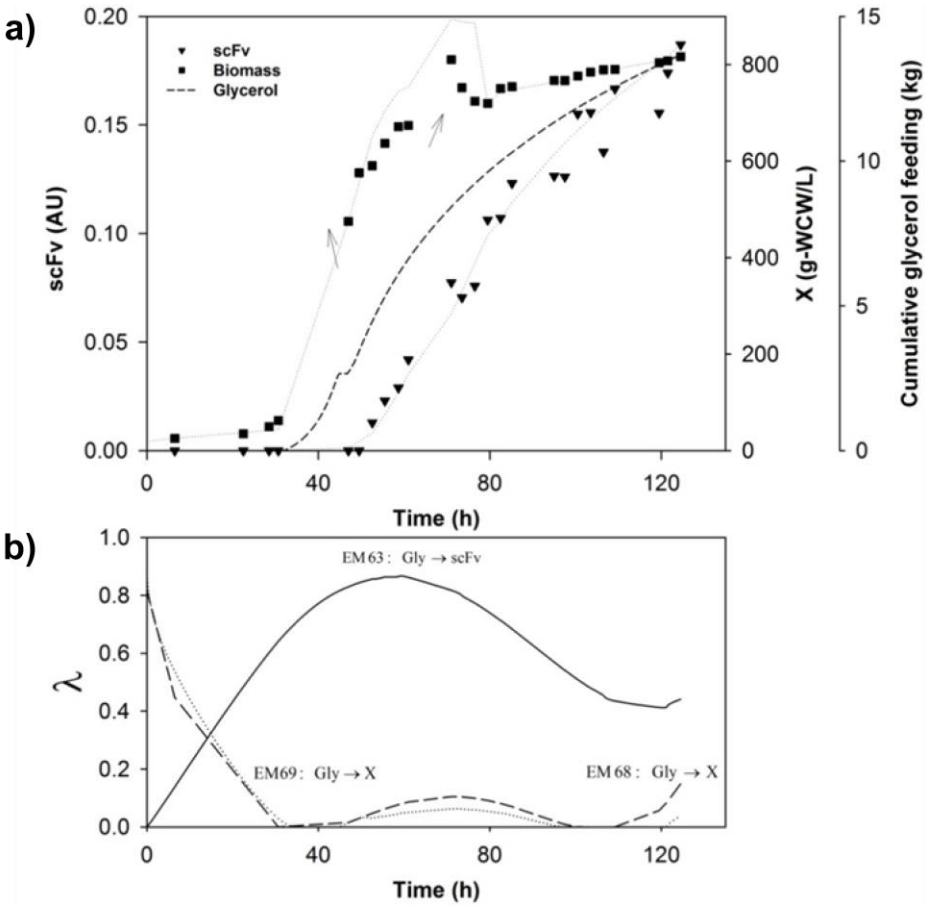


Figure 2.6 – Simulation of biomass and product dynamic profiles on the basis of three elementary modes (EM 63, EM 68, and EM 69) identified by functional enviromics: **a)** biomass and product concentration over time, and **b)** elementary mode weighting factors over time.

2.6 Conclusions

In a typical cell culture process there is a large number of environmental variables that shape cellular physiology. One important implication is that the design space for process development, namely culture medium optimization and process control, is potentially very large. Current process development methodologies in the industry are essentially of empirical nature. Empirical methods are not well suited to handle high-dimensional design spaces unless a substantial level of reductionism is applied, and even then with potential reduction of performance.

With the advances in systems biology, accurate genome-scale metabolic networks are becoming available for several microorganisms used in industry. Such metabolic networks contain the required information to enumerate all the operational modes of cells (i.e., elementary modes). With adequate systems biology tools such as functional enviromics, one can learn how such operational modes are controlled by the environment and/or how they modify the environment.

This paves the way for pathway-level process development strategies, which are much more efficient than traditional empirical methods.

Here we have laid out a process development methodology that can be summarized in the following main steps:

(i) Formulation of an accurate (genome-scale) metabolic network;
(ii) Computation of the elementary modes and pre-reduction of those elementary modes;
(iii) Discrimination of elementary modes with high correlation with environmental variables by functional enviromics;
(iv) Formulation of macroscopic material balances with explicit envirome-correlated elementary modes;
(v) Process optimization oriented to the manipulation of elementary mode weighting factors.

Such design tools can be used to optimize culture media and for advanced process control. A main advantage is the significant reduction of the number of experiments for very large design spaces. This is possible because the structure of the metabolic network constrains the manipulation of the environment. Another big benefit is the possibility to target intracellular control variables such as metabolic reactions or metabolic pathways directly linked with productivity and product quality. All in all, such techniques have the potential to considerably accelerate process development speed, to improve the mechanistic interpretability, and to increase process performance.

References

1. Bastin G, Dochain D (1990) On-line estimation and adaptive control of bioreactors. Elsevier, Amsterdam
2. Otero JM, Nielsen J (2010) Industrial systems biology. Biotechnol Bioeng 105:439–460
3. Talarek N, Cameroni E, Jaquenoud M, et al. (2010) Initiation of the TORC1-regulated G(0) program requires Igo1/2, which license specific mRNAs to evade degradation via the 5^0-3^0 mRNA decay pathway. Mol Cell 38:345–355
4. Borenstein E, Kupiec M, Feldman MW, Ruppin E (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc Natl Acad Sci U S A 105:14482–14487
5. Kell DB, Brown M, Davey HM, et al. (2005) Metabolic footprinting and systems biology: the medium is the message. Nat Rev Microbiol 3:557–565
6. Allen J, Davey HM, Broadhurst D, et al. (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. Nat Biotechnol 21:692–696
7. Teixeira AP, Dias JML, Carinhas N, et al. (2011) Cell functional enviromics: unravelling the function of environmental factors. BMC Syst Biol 5:92
8. Haag JE, Wouwer AV, Bogaerts P (2005) Dynamic modeling of complex biological systems: a link between metabolic and macroscopic description. Math Biosci 193:25–49
9. Provost A, Bastin G (2004) Dynamic metabolic modelling under the balanced growth condition. J Process Contr 14:717–728

10. Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18:326–332
11. Teixeira AP, Alves C, Alves PM, et al. (2007) Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control. BMC Bioinform 8:30
12. Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotech 26:659–667
13. Ferreira A, Dias J, Teixeira A, et al. (2011) Projection to latent pathways (PLP): a constrained projection to latent variables (PLS) method for elementary flux modes discrimination. BMC Syst Biol 5:181
14. Song H-S, Ramkrishna D (2009) Reduction of a set of elementary modes using yield analysis. Biotechnol Bioeng 102:554–568
15. Pagani I, Liolios K, Jansson J, et al. (2012) The genomes online database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 40:D571–D579
16. Tajparast M, Frigon D (2012) Genome-scale metabolic modeling to provide insight into the production of storage compounds during feast-famine cycles of activated sludge. In: Paper presented at the 3rd IWA/WEF wastewater treatment modeling seminar, Mont-Sainte-Anne, Quebec, 26–28 February 2012
17. Kanehisa Laboratories (2010) Kyoto encyclopedia of genes and genomes (KEGG). http://www.genome.jp/kegg/
18. SRI International (2010) BioCyc database collection. http://biocyc.org/
19. Janga SC, Babu MM (2008) Network-based approaches for linking metabolism with environment. Genome Biol 9:239
20. Feist AM, Herrgard MJ, Thiele I, et al. (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7:129–143
21. Orth JD, Palsson BO (2010) Systematizing the generation of missing metabolic knowledge. Biotechnol Bioeng 107:403–412
22. Feist AM, Henry CS, Reed JL, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121
23. Herrgard MJ, Swainston N, Dobson P, et al. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26:1155–1160
24. Chung BKS, Selvarasu S, Andrea C, et al. (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb Cell Fact 9:50
25. Sohn SB, Graf AB, Kim TY, et al. (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol J 5:705–715
26. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:U101–U115
27. Bernhardsson S, Gerlee P, Lizana L (2011) Structural correlations in bacterial metabolic networks. BMC Evol Biol 11:20
28. Schuster S, Dandekar T, Fell DA (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17:53–60
29. Papin JA, Stelling J, Price ND, et al. (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol 22:400–405
30. Celik E, Calik P, Oliver SG (2010) Metabolic Flux Analysis for recombinant protein production by *Pichia pastoris* using dualcarbon sources: effects of methanol feeding rate. Biotechnol Bioeng 105:317–329
31. von Kamp A, Schuster S (2006) Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22:1930–1931

32. Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16:373–378
33. Nielsen J, Oliver S (2005) The next wave in metabolome analysis. Trends Biotechnol 23:544–546
34. Fiehn O (2002) Metabolomics—the link between genotypes and phenotypes. Plant Mol Biol 48:155–171
35. Nicholson JK, Lindon JC, Holmes E (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29:1181–1189
36. de Koning W, van Dam K (1992) A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Anal Biochem 204:118–123
37. Raamsdonk LM, Teusink B, Broadhurst D, et al. (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19:45–50
38. Duarte I, Barros A, Belton PS, et al. (2002) High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer. J Agric Food Chem 50:2475–2481
39. Castrillo JI, Hayes A, Mohammed S, et al. (2003) An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. Phytochemistry 62:929–937
40. Soga T, Ohashi Y, Ueno Y, et al. (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res 2:488–494
41. Fiehn O, Kopka J, Dormann P, et al. (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol 18:1157–1161
42. Wolfender JL, Ndjoko K, Hostettmann K (2003) Liquid chromatography with ultraviolet absorbance-mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on-line structural investigation of plant metabolites. J Chromatogr 1000:437–455
43. Hollywood K, Brison DR, Goodacre R (2006) Metabolomics: current technologies and future trends. Proteomics 6:4716–4723
44. Oldiges M, Lutz S, Pflug S, et al. (2007) Metabolomics: current state and evolving methodologies and tools. Appl Microbiol Biotechnol 76:495–511
45. Tredwell GD, Edwards-Jones B, Leak DJ, Bundy JG (2011) The Development of metabolomic sampling procedures for *Pichia pastoris* and baseline metabolome data. PLoS One 6:e16286
46. Urbanczyk-Wochniak E, Luedemann A, Kopka J, et al. (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Rep 4:989–993
47. ter Kuile BH, Westerhoff HV (2001) Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. FEBS Lett 500:169–171
48. Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. Trends Anal Chem 24:285–294
49. Terzer M, Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24:2229–2235
50. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 97:5528–5533
51. de Figueiredo LF, Podhorski A, Rubio A, et al. (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 25:3158–3165
52. Wlaschin AP, Trinh CT, Carlson R, Srienc F (2006) The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. Metab Eng 8:338–352
53. Zhao Q, Kurata H (2009) Maximum entropy decomposition of flux distribution at steady state to elementary modes. J Biosci Bioeng 107:84–89

54. Wiback SJ, Mahadevan R, Palsson BO (2003) Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. J Theor Biol 224:313–324
55. Wiback SJ, Mahadevan R, Palsson BO (2004) Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the *Escherichia coli* spectrum. Biotechnol Bioeng 86:317–331
56. Wang Q, Yang Y, Ma H, Zhao X (2007) Metabolic network properties help assign weights to elementary modes to understand physiological flux distributions. Bioinformatics 23:1049–1052
57. Nookaew I, Meechai A, Thammarongtham C, et al. (2007) Identification of flux regulation coefficients from elementary flux modes: a systems biology tool for analysis of metabolic networks. Biotechnol Bioeng 97:1535–1549
58. Schwartz J-M, Kanehisa M (2006) Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinform 7:186
59. Jungers RM, Zamorano F, Blondel VD, et al. (2011) Fast computation of minimal elementary decompositions of metabolic flux vectors. Automatica 47:1255–1259
60. Ferreira AR, Ataide F, von Stosch M, et al. (2012) Application of adaptive DO-stat feeding control to *Pichia pastoris* X33 cultures expressing a single chain antibody fragment (scFv). Bioprocess Biosyst Eng 35:1603–1614

Chapter 3
Prediction of heterologous protein expression by hybrid metabolic flux analysis [2]

Abstract

Despite the growing importance of the *Pichia pastoris* expression system as an industrial workhorse, the literature is almost absent in systematic studies on how culture medium composition affects central carbon fluxes and heterologous protein expression. In this study we investigate how 26 variations of the standard BSM+PTM1 medium impact central carbon fluxes and protein expression in a *P. pastoris* X-33 strain expressing a scFv antibody. To achieve this goal, we adopted a hybrid metabolic flux analysis (hMFA) methodology, which is a modification of standard MFA to predict the rate of synthesis of heterologous proteins. Hybrid MFA combines the traditional parametric estimation of central carbon fluxes with non-parametric statistical modeling of product-related quantitative or qualitative measurements as a function of central carbon fluxes. Protein yield was much more sensitive to variations in medium composition (with a variability of 53.6% around the mean) than biomass growth, which was essentially determined by the dilution of basal media. Hybrid MFA was able to describe accurately the protein yield with normalized RMSE of 6.3% over 5 independent experiments. The metabolic state that promoted high protein yields was characterized by high overall metabolic rates through main central carbon pathways concomitantly with a relative shift of carbon flux from biosynthetic towards energy generating pathways.

Keywords

Pichia pastoris X-33 · hybrid metabolic flux analysis · central carbon fluxes · protein expression prediction

---

[2] Based on the paper: Isidro IA, Ferreira AR, Clemente JJ, Cunha AE, and Oliveira R (2015) Hybrid metabolic flux analysis and recombinant protein prediction in *Pichia pastoris* X-33 cultures expressing a single-chain antibody fragment. (manuscript submitted for publication)

3.1 Introduction

The yeast *Pichia pastoris* is now established in academia and industry as an expression system of choice. The key aspects that explain the success of *P. pastoris* have been recently reviewed by Ahmad et al. [1]. Among those key aspects is the fast and inexpensive growth that reaches extremely high cell densities, up to 200 g-DCW L^{-1} on chemically defined media [2, 3]. It can produce foreign proteins at levels as high as 80% total secreted protein or up to 30% total cell protein [4] and is able to carry out post-translational modifications. Unlike other yeasts, it shows a strong preference for respiratory growth and a limited tendency for fermentation [5].

The *Pichia pastoris* MUT+ phenotype uses a very strong and tightly regulated promoter, pAOX1, that is induced by methanol [6]. For constitutive heterologous protein expression the most widely used promoter is pGAP [7], which is the promoter used in this study. These and other promoters used in *Pichia* expression systems have been extensively reviewed elsewhere [8]. In general, techniques for genetic manipulation are simple and similar to *Saccharomyces cerevisiae,* with commercial expression kits available for both intracellular and secretory expression. Furthermore, the *P. pastoris* genome sequence and annotation is publically available [9] and curated genome-scale metabolic models can be found in the literature [10, 11], which will likely trigger rational synthetic biology projects using *P. pastoris* as a platform organism in the near future.

Several studies focused on the metabolism of *P. pastoris* were performed in order to better understand how the central carbon flux distribution and expression of heterologous protein respond to different culture conditions. Such studies have led to important insights into the effect of oxygenation [12], temperature [13], and the use of glucose [3], methanol [14] and mixed glucose:methanol [15] as carbon sources. Notably, a recent study [16] combining Monte Carlo simulation and principal component analysis (PCA) showed that the reactions that are directly linked to ATP production account for most of the variability observed in heterologous protein production rate. In another study [17], the authors were able to estimate recombinant protein production rate based on the distribution of energy resources between growth, ATP production and maintenance using ordinary least squares regression. However, despite the advances in genomics and metabolomics, the literature is almost absent in systematic studies on how culture medium composition affects metabolism and heterologous protein expression.

In this study, we perform metabolic flux analysis of *P. pastoris* cultivated on different culture medium compositions. More specifically, we study how the concentration of trace elements of the PTM1 solution and the BSM dilution (the standard culture media for *P. pastoris* cultures) affects metabolic flux distributions of a *Pichia pastoris* X-33 strain constitutively expressing a heterologous single-chain variable fragment (scFv) under the pGAP promoter. We adopted a simplified network that has been shown to represent central carbon metabolism accurately [18]. Furthermore, in order to model and understand how heterologous protein expression is affected by the culture medium, we used the hybrid MFA methodology [19] due to limitations of traditional MFA methods when

modeling the expression of heterologous proteins. Specifically, the carbon flux used for synthesis of heterologous product is several orders of magnitude lower than the fluxes through main carbon pathways, rendering classic MFA not suitable to estimate heterologous protein synthesis. The hybrid MFA approach can be used instead, combining the parametric estimation of fluxes by MFA with a non-parametric statistical model to establish a link between a product-related quantitative or qualitative measurements and selected fluxes through the metabolic network [19]. Here we use the well-established partial least squares (PLS) regression as the statistical model to predict heterologous protein yield.

3.2 Materials and methods

3.2.1 DOE of medium composition

Eleven (N = 11) medium factors were selected as design parameters, the 10 components of the *Pichia* trace minerals 1 (PTM1) solution and an additional factor for the dilution of the basal salts medium (BSM) solution (see Table 3.1). Statistical design-of-experiments (DoE) was adopted to design medium compositions so as to maximize metabolic information generated while minimizing the number of experiments. Each medium factor has a baseline value taken from the Invitrogen guidelines [20]. The DoE adopted is a two-level D-optimal design with the +1 level coincident with the baseline value and the -1 level defined as 10 times lower than the baseline value. This resulted in $2\times(N+1) = 24$ screening compositions determined by D-optimal design using the MATLAB

Table 3.1 – Design factors for medium screening. List of medium factors with respective baseline values [20], as well as upper and lower levels used for shake flask experiments.

	Medium factors	Units	Baseline	-1 level	+1 level
PTM solution					
Fac1	$CuSO_4 \cdot 5H_2O$	g/L	6.00	0.60	6.00
Fac2	NaI	g/L	0.08	0.008	0.08
Fac3	$MnSO_4 \cdot H_2O$	g/L	3.00	0.30	3.00
Fac4	$Na_2MoO_4 \cdot 2H_2O$	g/L	0.20	0.02	0.20
Fac5	H_3BO_3	g/L	0.02	0.002	0.02
Fac6	$CoCl_2$	g/L	0.50	0.05	0.50
Fac7	$ZnCl_2$	g/L	20.00	2.00	20.00
Fac8	$FeSO_4 \cdot 7H_2O$	g/L	65.00	6.50	65.00
Fac9	Biotin	g/L	0.20	0.02	0.20
Fac10	H_2SO_4	mL/L	5.00	0.50	5.00
BSM solution					
Fac11	BSM dilution factor	v/v	1:1	1:2	1:1

Table 3.2 – Design of experiments for medium screening. D-optimal experimental design for linear function identification with 11 factors and 24 experiments. The actual composition corresponding to +1 and -1 levels is specified in Table 3.1. Experiment 25 is the upper/baseline control and experiment 26 is the lower level control.

Experiment	Medium factor										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	-1	-1	1	-1	1	-1	-1	-1	1
2	1	-1	-1	1	-1	-1	-1	1	-1	-1	1
3	1	-1	1	1	1	-1	1	1	1	-1	-1
4	1	1	-1	-1	1	1	-1	1	1	1	-1
5	1	1	1	1	1	-1	-1	-1	1	1	1
6	-1	1	-1	1	1	1	1	-1	1	-1	-1
7	-1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	-1	1	1	1	-1	-1	-1
9	1	-1	-1	1	-1	-1	1	1	1	1	-1
10	-1	1	1	-1	-1	-1	-1	-1	1	-1	-1
11	-1	1	-1	1	-1	1	-1	1	-1	-1	1
12	-1	-1	-1	-1	1	-1	1	1	1	-1	1
13	-1	-1	1	1	1	1	-1	-1	-1	-1	1
14	1	1	-1	1	-1	1	1	-1	1	-1	1
15	1	-1	-1	1	1	1	-1	-1	-1	1	-1
16	1	1	1	-1	1	-1	-1	1	-1	-1	-1
17	1	-1	1	-1	-1	1	-1	-1	1	-1	1
18	-1	1	1	-1	-1	1	1	1	-1	1	1
19	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1
20	-1	-1	1	1	-1	-1	1	-1	-1	1	-1
21	-1	1	-1	1	1	-1	-1	1	-1	1	1
22	-1	-1	-1	-1	-1	1	-1	1	1	1	-1
23	1	1	-1	-1	-1	-1	1	-1	-1	1	1
24	1	-1	1	-1	1	1	1	1	-1	1	1
25	1	1	1	1	1	1	1	1	1	1	1
26	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

function *rowexch* for linear function identification. Two additional control compositions were added for screening, the lower (-1) level and upper (+1) level. The final medium screening design is shown in Table 3.2.

3.2.2 Shake flask cultures

Medium compositions specified in Table 3.2 were used in 26 screening experiments in 250 mL shake flasks. Basal salts media (BSM) was prepared according to Invitrogen guidelines [20] by mixing 26.7 mL H_3PO_4 85%, 0.93 g $CaSO_4 \cdot 2H_2O$, 18.2 g K_2SO_4, 14.9 g $MgSO_4 \cdot 7H_2O$, 4.13 g KOH, and 40.0 g glycerol in water up to 1 L. After autoclaving at 121 ºC for 30 minutes the pH was adjusted to 5.0 with 25% NH_4OH, which also serves as the nitrogen source. According to the BSM

dilution factor specified in Table 3.2, for each experiment either 40 mL of BSM solution or 20 mL BSM plus 20 mL distilled water were added to the shake flask. The 26 trace elements solutions specified in Table 3.2 were prepared by mixing the components in the specified concentrations and then sterilized by filtration with a 0.22 nm pore size filter. 174 µL of each trace elements solution was added to the corresponding flask.

The inoculum was prepared in a 250 mL shake flask with 40 mL BSM supplemented with 174 µL of PTM1. The cell bank of the *Pichia pastoris* X-33 strain constitutively expressing a single-chain variable fragment (scFv) with a GAP promoter was stored at -80 ºC and thawed at room temperature. After adding 1 mL of cell stock the inoculum was incubated at 30 ºC and 150 rpm on an Innova 4 300 orbital incubator. When the inoculum achieved exponential growth, after approximately 26 hours of incubation, aliquots of 2 mL were used to inoculate the 26 shake flasks.

The 26 shake flasks were incubated for 110 hours at 30 ºC and 150 rpm on the orbital incubator. Samples were taken at intervals of 24 hours. Product concentration was determined by enzyme-linked immunosorbent assays (ELISA) as described in [21]. Biomass concentration was determined by optical density at 600 nm and by centrifugation at 15 000 rpm for 20 min followed by weighting without drying. Wet cell weight (WCW) was converted to dry cell weight (DCW) by multiplying by a factor of 0.28 previously calibrated for similar culture densities [22].

3.2.3 Shake flasks data pre-processing

For each shake flask experiment, specific rates for biomass growth (μ), glycerol consumption (q_S), oxygen consumption (q_O), and CO_2 production (q_{CO}) were estimated using the kinetic model proposed by Jahic and co-workers [23] (Eqs. 3.1-3.6). This model describes carbon flux through anabolic and catabolic pathways, leading to biomass growth and oxygen consumption.

$$q_S = q_{S,max} \frac{S}{S + K_S} \quad \textbf{(3.1)}$$

$$\mu = (q_S - q_m)\, Y_{em} \quad \textbf{(3.2)}$$

$$q_{S,an} = (q_S - q_m)\, Y_{em}\, \frac{C_X}{C_S} \quad \textbf{(3.3)}$$

$$q_{S,en} = q_S - q_{S,an} \quad \textbf{(3.4)}$$

$$q_O = q_{S,an}\, Y_{O/S,an} + q_{S,en}\, Y_{O/S,en} \quad \textbf{(3.5)}$$

$$q_{CO} = q_{S,en}\, \frac{C_S}{C_X} \quad \textbf{(3.6)}$$

Specific protein synthesis rate (q_P) was defined as growth-dissociated using the Luedeking-Piret relation:

$$q_P = \alpha\mu + \beta \ , \quad \alpha = 0 \tag{3.7}$$

The material balance equations for biomass (*X*), glycerol (*S*) and heterologous protein (*P*) concentrations were set as follows:

$$\frac{dX}{dt} = \mu X \tag{3.8}$$

$$\frac{dS}{dt} = -q_S X \tag{3.9}$$

$$\frac{dP}{dt} = q_P X \tag{3.10}$$

The model was numerically integrated and fitted against experimental measurements of *X*, *S* and *P* in the least-squares sense using MATLAB functions *ode45* and *lsqnonlin*. The fitting parameters were $q_{S,max}$, Y_{em}, α and β, while K_S, q_m , $Y_{O/S,an}$, $Y_{O/S,en}$, C_X and C_S were fixed to previously validated values for *P. pastoris* growth on glycerol [23]. Parameter 95% confidence intervals were estimated using a numerical approximation to the Jacobian matrix with MATLAB function *nlparci*, and then used for error propagation to kinetic rates μ, q_O, and q_{CO}.

3.2.4 *Pichia pastoris* metabolic network

The metabolic network model for central carbon metabolism used here was largely based on the *P. pastoris* network described in [18], with adaptations based on other central carbon [12] and genome scale [10] networks. The network consists of 43 metabolic reactions, 34 internal metabolites, and 10 exchange rates (including biomass synthesis). Additional file 3.1 lists metabolic reactions, metabolites and the respective stoichiometric matrix. The network comprehends the main catabolic pathways, namely glycolysis and gluconeogenesis pathways (abbreviated here as EMP for Embden-Meyerhof-Parnas), the tricarboxylic acid cycle (TCA), the pentose phosphate pathway (PPP), anaplerotic, fermentative and phosphorylative oxidation pathways. It supports growth on glucose, glycerol and methanol (or mixed feeds of these substrates), which are the main carbon sources used in *P. pastoris* cultures. The model includes NADH and NADPH coenzyme balancing and a biomass formation reaction from selected internal metabolites based on known macromolecular composition of *P. pastoris* cells [13].

3.2.5 Metabolic flux analysis

Applying a balanced growth condition to the *P. pastoris* network results in a 34 x 43 stoichiometric matrix, *S*, with 34 balanced metabolites and a vector of 43 fluxes *v*, which obey to the following system of algebraic equations:

$$S \cdot v = \varepsilon \tag{3.11}$$

The flux vector *v* is partitioned into two sets: the measured fluxes v_m, which are known, and the remaining calculated fluxes v_c that are unknown. This way, the system above can be equivalently written as:

$$S_m \cdot v_m + S_c \cdot v_c = \varepsilon \tag{3.12}$$

When the steady state condition is obeyed and all metabolites are perfectly balanced then $\varepsilon = 0$. However, when the system is redundant and the measured fluxes are corrupted by error then $\varepsilon \neq 0$ and the consistency index, *h*, can be calculated by Eq. (3.13), with *P* the variance-covariance matrix of the metabolite balancing error. The consistency index *h* expresses the variance-weighted sum of metabolite balancing errors and should fall below a given threshold χ^2 value [24].

$$h = \varepsilon^T \cdot P^{-1} \cdot \varepsilon \tag{3.13}$$

The stoichiometric matrix S_c has one redundant row, corresponding to 34 – 1 = 33 independent conservation equations. The number of unknown fluxes is given by the total number of fluxes minus the measured fluxes, so 43 – 10 = 33 unknowns. With 33 independent equations and 33 unknowns, the system is determined for v_c and all remaining fluxes can be calculated. Given that S_c has one degree of redundancy, at least one of the measured fluxes is balanceable and a consistency index can be calculated [25]. For this system, a calculability and balanceability analysis performed using the null space method [26] showed that all measured rates are balanceable.

The MFA adopted in this work implements a constrained nonlinear optimization using MATLAB function *fmincon* to minimize the consistency index, *h*, under the constraint of nonnegative irreversible fluxes and of measured fluxes v_m, which are allowed to vary within a given (narrow) interval around the measured values. Finally, the errors in estimated fluxes where calculated according to [25] based on the estimated error for measured fluxes.

The subset v_m comprised a total of 10 known fluxes described next. Biomass growth, glycerol uptake, O_2 consumption, and CO_2 production (μ, q_S, q_O and q_{CO} respectively) are known from shake flasks data and their values were allowed to vary within one standard error interval in the sense of *h* minimization. Glucose and methanol are not present in the medium so their uptake rate was

considered known and fixed to zero. Since *P. pastoris* has a low fermentative profile [5] we considered the accumulation of ethanol in the medium to be known and negligible. Citrate, pyruvate and acetate accumulation in the extracellular medium were also considered to be known and negligible. MFA was allowed to adjust the rate of one of these by-products within the interval of 0 ± 1 mmol g-DCW^{-1} h^{-1} until maximum consistency was reached.

3.2.6 Hybrid metabolic flux analysis

Hybrid MFA combines traditional parametric estimation of central carbon fluxes with non-parametric statistical modeling of a product-related quantitative or qualitative measurements as a function of central carbon fluxes [19], as represented in Figure 3.1. With hybrid MFA, different variables can be used to represent productivity, be it the specific rate of product synthesis, heterologous product activity, glycosylation pattern, or a combination of these factors. In this study we used the product yield (mg-heterologous protein/g-DCW), as calculated directly from experimental data, as the response variable. The predictor matrix consisted of central carbon fluxes selected from the MFA-calculated metabolic flux distribution that are representative of distinct metabolic pathways, as well as some ratios of these fluxes. The predictor and response variables were normalized by the total available carbon source at the start of the experiment and the response variable distribution was approximated to a normal distribution using a logarithmic transformation.

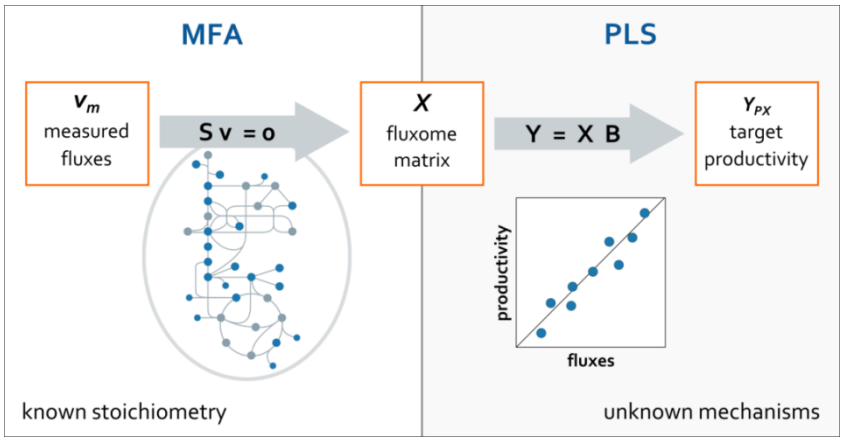


Figure 3.1 – Hybrid metabolic flux analysis framework. Hybrid metabolic flux analysis combines a classic MFA step, where global fluxes are derived from a subset of measured fluxes, with a step of statistical prediction of recombinant product synthesis from the global fluxes.

PLS regression [27] decomposes the predictor matrix *X* into a set of uncorrelated variables as represented by Eq. (3.14). This projection of the fluxes into latent variables (LV) eliminates redundant information and is done by maximizing covariance with the response matrix. Latent variables are represented in the columns of the X-scores matrix *T*, whereas W^T is the matrix of X-weights. Similarly, the response matrix *Y* can also be decomposed into Y-scores and Y-loadings (3.15), although this will not apply here as the response is a single variable. Finally, the inner

regression (3.16) closes the relation between predictor and response. The symbols ε_x, ε_y and ε_B represent the residuals in each relation.

$$T = X \cdot W + \varepsilon_X \quad \textbf{(3.14)}$$

$$Y = U \cdot Q^T + \varepsilon_Y \quad \textbf{(3.15)}$$

$$U = T \cdot B + \varepsilon_B \quad \textbf{(3.16)}$$

After fitting the PLS regression to a test set, the product yield (response vector) can then be estimated from the central carbon fluxes (predictor matrix) for new data sets as shown in Eq. (3.17). The regression coefficients *C* indicate how each flux correlates and influences the product yield.

$$\hat{Y}_{PX} = fluxes \cdot C \quad \textbf{(3.17)}$$

PLS regression was implemented in MATLAB using a variation of the NIPALs algorithm [28]. Confidence intervals for the regression coefficients were calculated by Monte Carlo simulation in which 10 000 bootstrapping samples were used. The predictive power of the hybrid MFA method was evaluated by leaving out 20% of the dataset for validation and fitting the statistical model to the training set.

3.3 Results and discussion

3.3.1 Shake flasks results

Figure 3.2 shows the biomass yield (g-DCW/g-glycerol) and product yield (mg-protein/g-DCW), as calculated directly from experimental measurements, for the 26 shake flask experiments. Biomass yield shows little variability despite the large variations in the medium composition (10-fold variations in trace elements and 2-fold variations in main salts and glycerol concentrations). The observed variability around the mean (as measured by the relative standard deviation) was 4.5% and closely matches the experimental error (estimated to be 4.3%). In contrast, the product yield seems to be rather sensitive to the medium composition, showing a variability of 53.6%, which is significantly higher than the ELISA measurement error (estimated to be 15-20%). This result suggests that, within the concentration ranges tested, the expression of the heterologous protein is much more sensitive to medium composition than biomass growth, which seems to be mainly determined by carbon source availability.

All shake flask experiments with product yield over 0.5 mg/g were cultivated with 1:2 dilution of the main salts solution (BSM), with the exception of experiment 1, with a yield of 0.61 mg/g. This result

seems to support previously published data that showed a positive effect of BSM dilution in the expression of heterologous proteins [29, 30]. It should however be noted that the variability in the product yield within these two groups is significant, especially for the higher yield group. This suggests that the concentrations of trace elements are equally important for the final product yield.

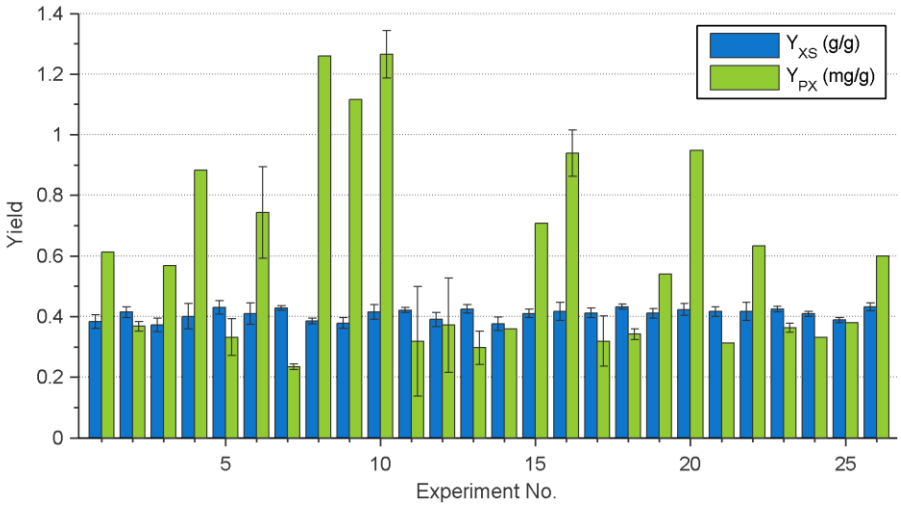


Figure 3.2 – Biomass and product yields for shake flask experiments. Biomass yield Y_{XS} (g-DCW/g-glycerol) shows little variation across experiments, but product yield Y_{PX} (mg-protein/g-DCW) varies up to 3-fold.

The kinetic model previously described in Eqs. (3.1-3.10) was fitted to shake flask measurements of biomass, glycerol and heterologous protein concentration. The model was fitted to each of the 26 experiments, yielding estimates of key specific reaction rates at exponential biomass growth conditions, namely the fluxes of biomass growth, glycerol consumption, oxygen consumption, carbon dioxide production and protein expression (μ, q_S, q_O, q_{CO} and q_P). As illustrative example, Figure 3.3A shows the fitting of the kinetic model to the data of experiment 1. The full set of kinetic parameters for all shake flask experiments is shown in Table 3.3. To note that the model is determined, redundant and consistent with the measured dynamic profiles resulting in narrow confidence intervals of kinetic parameters estimates (5-10% average error). One exception is the estimate of product synthesis rate, which has wider confidence intervals (17% average error), most likely due to the higher measurement error associated with the ELISA assays. As in the case of yields, the determined kinetic rates show large variations among the shake flask experiments denoting high sensitivity to the medium composition. In particular, we can observe that the specific product synthesis rate is the most sensitive parameter to the medium composition, with variations of 55.3% around the mean. The specific substrate consumption rate is also affected by medium composition and has a variation of 18.9% around the mean. As before, the biomass yield Y_{em} has the lowest variability, namely 6.6% around the mean, which is comparable to its 5.6% average error.

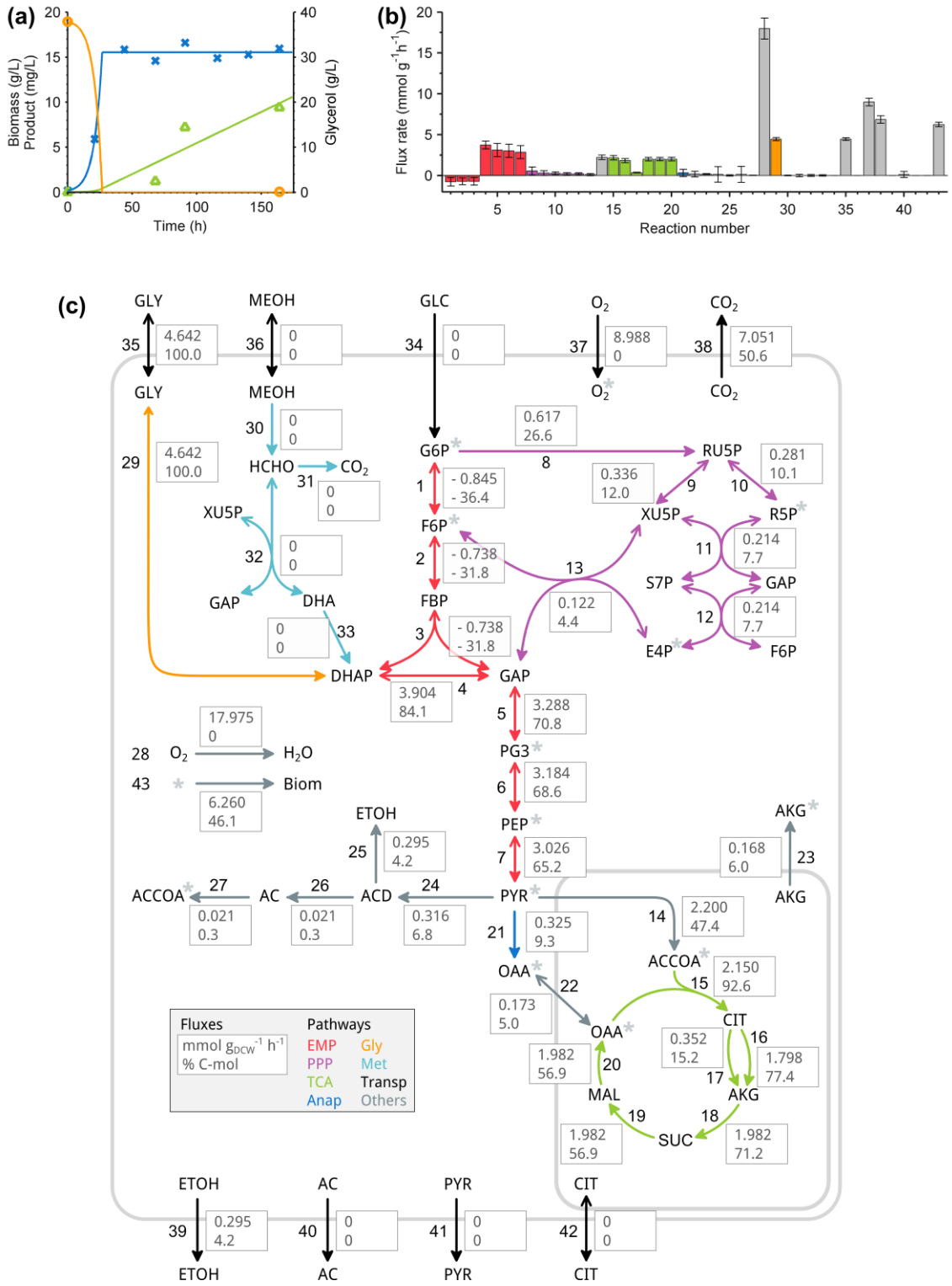


Figure 3.3 – Model fitting and metabolic flux distribution for shake flask experiment 1. (a) Model fit to experimental data: experimental measurements for (o) substrate, (×) biomass and (△) product concentration, with fitted model (lines) **(b)** Metabolic flux distribution as calculated by classic MFA with 95% confidence intervals **(c)** Flux distribution through main carbon pathways. The first line shows metabolic fluxes in mmol g-DCW^{-1} h^{-1}, the second line shows the metabolic fluxes normalized with respect to carbon uptake (% C-mol/C-mol glycerol). For simplicity, decarboxylations and NADH/NADPH formation and consumption are not represented.

Table 3.3 – Model-adjusted fluxes for shake flask experiments. Adjusted model parameters ($q_{S,max}$, Y_{em}, q_P) and calculated seed fluxes (μ, q_S, q_O and q_{CO}) for each experiment with 95% confidence interval.

Exp. No.	$q_{S,max}$ (g g^{-1} h^{-1})	Y_{em} (g g^{-1})	q_P (µg g^{-1} h^{-1})	μ (g g^{-1} h^{-1})	q_O (g g^{-1} h^{-1})	q_{CO} (g g^{-1} h^{-1})
1	0.42 ± 0.04	0.41 ± 0.02	4.4 ± 0.7	0.17 ± 0.02	0.27 ± 0.05	0.32 ± 0.06
2	0.34 ± 0.02	0.43 ± 0.01	2.9 ± 0.3	0.15 ± 0.01	0.21 ± 0.03	0.25 ± 0.03
3	0.35 ± 0.03	0.34 ± 0.01	4.1 ± 0.5	0.12 ± 0.01	0.26 ± 0.04	0.30 ± 0.04
4	0.32 ± 0.06	0.38 ± 0.03	9 ± 2	0.12 ± 0.03	0.22 ± 0.08	0.26 ± 0.10
5	0.39 ± 0.02	0.44 ± 0.01	2.7 ± 0.3	0.17 ± 0.01	0.24 ± 0.03	0.28 ± 0.03
6	0.50 ± 0.04	0.38 ± 0.03	4.9 ± 0.8	0.19 ± 0.02	0.35 ± 0.06	0.41 ± 0.07
7	0.30 ± 0.02	0.43 ± 0.01	2.0 ± 0.3	0.13 ± 0.01	0.19 ± 0.03	0.22 ± 0.03
8	0.33 ± 0.03	0.39 ± 0.02	11 ± 1	0.13 ± 0.01	0.22 ± 0.04	0.26 ± 0.05
9	0.32 ± 0.09	0.37 ± 0.05	11 ± 2	0.12 ± 0.04	0.2 ± 0.1	0.3 ± 0.1
10	0.44 ± 0.04	0.42 ± 0.03	10 ± 1	0.19 ± 0.02	0.28 ± 0.06	0.33 ± 0.07
11	0.32 ± 0.04	0.43 ± 0.02	3.0 ± 0.6	0.14 ± 0.02	0.20 ± 0.05	0.23 ± 0.06
12	0.35 ± 0.03	0.39 ± 0.01	3.0 ± 0.4	0.14 ± 0.01	0.23 ± 0.04	0.28 ± 0.04
13	0.41 ± 0.04	0.44 ± 0.02	2.6 ± 0.6	0.18 ± 0.02	0.25 ± 0.05	0.29 ± 0.06
14	0.48 ± 0.03	0.39 ± 0.02	2.9 ± 0.6	0.19 ± 0.02	0.33 ± 0.05	0.39 ± 0.06
15	0.45 ± 0.03	0.42 ± 0.02	6.7 ± 0.9	0.19 ± 0.02	0.29 ± 0.04	0.34 ± 0.05
16	0.35 ± 0.06	0.42 ± 0.05	8 ± 2	0.15 ± 0.03	0.22 ± 0.08	0.26 ± 0.10
17	0.39 ± 0.04	0.44 ± 0.02	2.7 ± 0.6	0.17 ± 0.02	0.24 ± 0.05	0.28 ± 0.06
18	0.28 ± 0.03	0.44 ± 0.01	2.9 ± 0.4	0.12 ± 0.01	0.17 ± 0.04	0.21 ± 0.04
19	0.48 ± 0.02	0.39 ± 0.01	4.8 ± 0.5	0.19 ± 0.01	0.33 ± 0.02	0.39 ± 0.03
20	0.42 ± 0.02	0.43 ± 0.02	8.6 ± 0.8	0.18 ± 0.01	0.26 ± 0.04	0.31 ± 0.04
21	0.38 ± 0.04	0.42 ± 0.02	4.0 ± 1.0	0.16 ± 0.02	0.24 ± 0.06	0.29 ± 0.07
22	0.35 ± 0.03	0.42 ± 0.02	3.7 ± 0.7	0.15 ± 0.01	0.22 ± 0.04	0.26 ± 0.04
23	0.40 ± 0.05	0.45 ± 0.03	3.4 ± 0.8	0.18 ± 0.03	0.24 ± 0.07	0.28 ± 0.09
24	0.31 ± 0.03	0.42 ± 0.01	2.7 ± 0.6	0.13 ± 0.01	0.20 ± 0.04	0.23 ± 0.05
25	0.22 ± 0.01	0.38 ± 0.01	3.4 ± 0.4	0.084 ± 0.004	0.16 ± 0.01	0.18 ± 0.01
26	0.49 ± 0.09	0.40 ± 0.06	7 ± 2	0.20 ± 0.05	0.3 ± 0.1	0.4 ± 0.2

3.3.2 Central carbon fluxes determined by classical MFA

Metabolic flux analysis (MFA) was employed to estimate the 43 metabolic fluxes for each of the shake flask experiments using the 10 known rates as input to the MFA calculations. The obtained system is redundant with one degree of freedom (see Materials and methods), which enabled the determination of the consistency index, *h*. The overall results of estimated flux data, with 95% confidence intervals, and respective consistency index are provided in Additional file 3.2.

Figure 3.3 shows the metabolic flux distribution (MFD) obtained for experiment 1. The analysis of Figure 3.3C reveals that *Pichia pastoris* growing exponentially on glycerol has typically very high fluxes through central carbon pathways, namely the Embden-Mayerhof-Parnas (EMP, r4 to r7) and TCA cycle (r14 to r20) corresponding to 47.4-84.1% of total carbon intake. Another core feature is

the gluconeogenesis seen as a reverse flux through part of the EMP pathway (r1 to r3), diverting part of the carbon in glycerol towards the pentose phosphate pathway (PPP) at the triose phosphate branch point (r3). The pentose phosphate (r8 to r13) and anaplerotic pathways (r21, r22) which provide precursors for biomass synthesis have lower carbon fluxes, ranging from 5.0 to 26.6% of carbon intake. The high peak for oxidative phosphorylation flux (r28) seen in Figure 3.3B is typical for *P. pastoris* metabolic flux distribution on different substrates [18] due to its preference for aerobic respiration as energy source. All in all, these metabolic flux distribution patterns are in agreement with previously published studies of *P. pastoris* grown on glycerol [17].

Best overall consistency was reached by allowing MFA to adjust the flux for ethanol production around the assumed zero. Only in this case the consistency index criterion was obeyed for all shake flask experiments. These results suggest a low but positive fermentative activity (r25) ranging from 3.9 to 10.4% of carbon intake. Even though *P. pastoris* is known for its preference for oxidative respiration, other studies have found a low but consistent production of fermentative by-products [3], and we have detected the presence of ethanol from NMR analysis (see example in Figure 3.4). These observations are consistent with the hypothesis that ethanol is being produced in small quantities and possibly also other by-products such as acetate and arabinol [31].

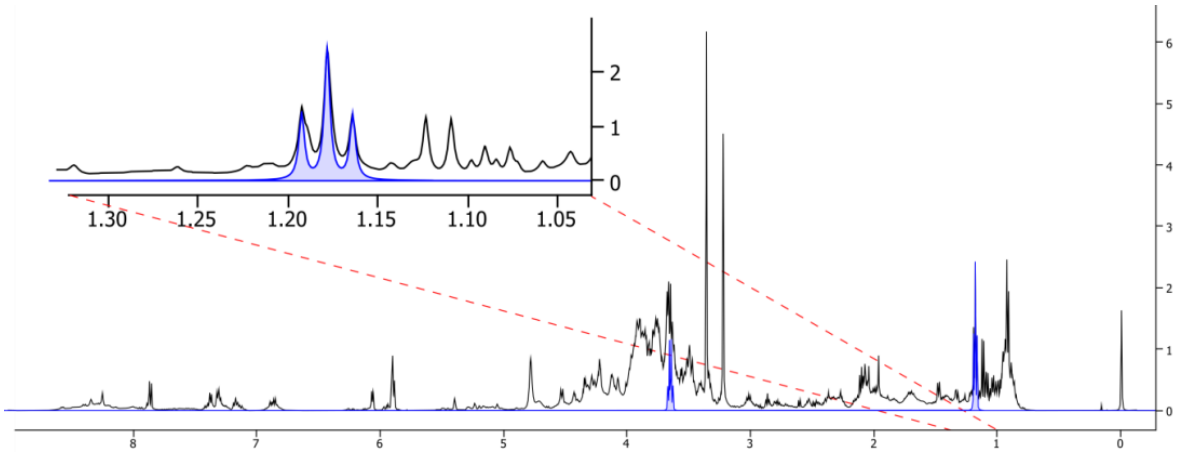


Figure 3.4 – ^{1}H-NMR spectra of culture broth supernatant shows presence of ethanol. Even though *P. pastoris* has a generally low fermentative profile, ethanol has been detected in similar experiments

As expected, the metabolic flux distributions show a considerable variability among the 26 shake flask experiments (Figure 3.5). Most fluxes show 18.6-22.3% variation around the mean. The TCA cycle rate ranged between 1.5 and 2.5 mmol g-DCW^{-1} h^{-1}, showing a good correlation with carbon uptake (R^2 = 0.882). Unlike a previous study with unlimited (batch) growth on glucose [32], our data does not suggest an upper limit around 2.1 ± 0.1 mmol g-DCW^{-1} h^{-1} for the TCA cycle. TCA rate variability has also been observed previously in glucose-limited fed-batch growth [15] and batch growth on glucose for reference and recombinant strains [3].

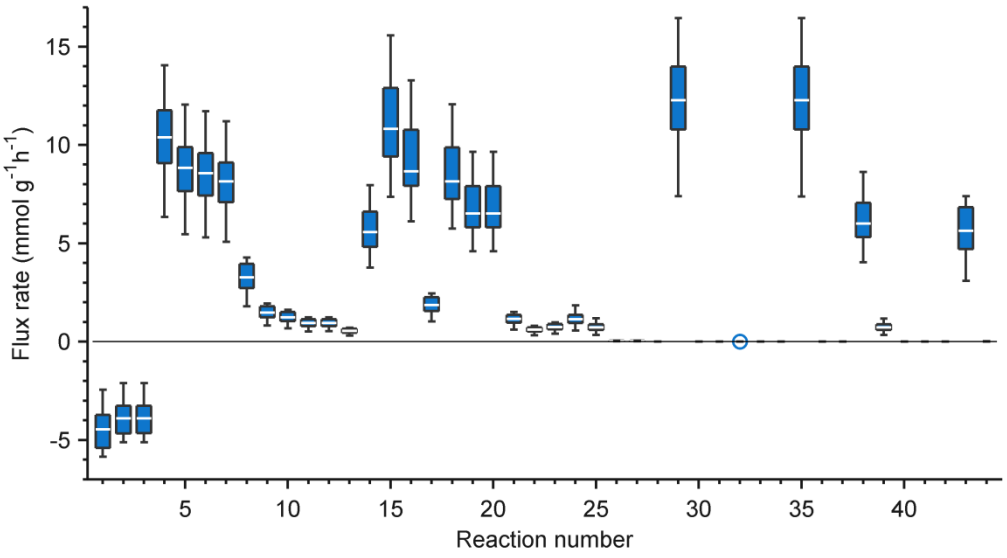


Figure 3.5 – Variability in MFD for different medium compositions. Boxplots for 44 fluxes across 26 shake flask experiments show different variability depending on the metabolic pathway.

The flux through the pentose phosphate pathway, normalized by carbon uptake, showed a strong positive correlation ($R^2 = 0.994$) with biomass yield, as expected since the PPP is the supplier of NADPH and precursors needed for biosynthesis. Comparing the fluxes through the PPP and the TCA cycle, we observe that carbon flux through the pentose phosphate pathway (r29 – r5) increases as flux through TCA cycle (r14) increases, suggesting that higher metabolic rates lead to consistently higher fluxes throughout the metabolic network. However, the normalized carbon flux (% of carbon intake) diverted through the TCA cycle shows an inverse relationship to relative carbon flux through the PPP (Table 3.4). This observation suggests that normalized carbon fluxes (%C) and carbon flux ratios (C-ratios) at main branch points may provide important insight into the metabolic state that cannot be inferred from the absolute fluxes alone. Other studies have demonstrated the possible impact of manipulating flux ratios, particularly at the pyruvate branch point [14, 33].

Overall, these metabolic flux distribution patterns denote a variability between biosynthetic and energy generation metabolic profiles which can only be explained by the different medium compositions.

Table 3.4 – Carbon flux though main pathways normalized by carbon intake. Carbon fluxes through main metabolic pathways, normalized by carbon intake (%C) to facilitate comparison between pathways.

Exp. No.	EMP	PPP	TCA	CO_2	Biomass
1	70.8	29.2	47.4	50.6	45.9
2	68.8	31.2	45.9	48.6	49.2
3	75.9	24.1	53.5	57.2	38.0
4	72.7	27.3	45.7	50.4	43.0
5	69.1	30.9	43.3	47.0	48.7
6	73.2	26.8	48.3	52.4	42.1
7	69.6	30.4	44.1	47.8	47.8
8	72.8	27.2	47.4	51.6	42.9
9	74.1	25.9	43.3	49.5	40.9
10	70.3	29.7	45.2	48.9	46.8
11	70.4	29.6	40.8	46.0	46.6
12	71.3	28.7	50.1	52.7	45.1
13	69.1	30.9	41.5	45.8	48.6
14	72.6	27.4	49.0	52.6	43.1
15	70.0	30.0	46.1	49.3	47.3
16	70.4	29.6	43.6	47.9	46.6
17	68.6	31.4	43.7	47.1	49.5
18	69.0	31.0	41.6	45.8	48.9
19	73.3	26.7	49.3	53.4	43.2
20	69.5	30.5	44.7	48.2	48.1
21	70.9	29.1	41.1	46.4	45.8
22	70.8	29.2	44.2	48.4	46.1
23	67.7	32.3	40.8	44.7	50.8
24	70.7	29.3	42.9	47.6	46.2
25	73.9	26.1	50.8	54.6	41.9
26	71.2	28.8	43.6	48.3	45.3

3.3.3 Heterologous protein yield predicted by hybrid MFA

As previously discussed, classical MFA is limited to central carbon fluxes and it is not suitable to estimate heterologous protein synthesis fluxes. One important limitation is the low amount of carbon diverted for the synthesis of heterologous protein, often much lower than the measurement error of fluxes. For this reason, instead of including a heterologous protein synthesis reaction in the metabolic network, we have adopted the hybrid MFA approach [19], in which we link the central carbon fluxes with the heterologous protein yield by means of a statistical model (Figure 3.1). In this study, we used partial least squares regression (PLS) to model the product to biomass yield Y_{PX} (response vector) as a function of selected fluxes representative of the main metabolic pathways and carbon ratios at the main branch points (predictor matrix).

The 26 points dataset was randomly partitioned in a calibration dataset (21 points) and a validation dataset (5 points). The PLS regression model was fitted to the calibration dataset only and validated

against the validation dataset. Table 3.5 shows the cumulative explained variance in predictor matrix and response vector with up to 6 latent variables. Latent variable 1 accounts for most of the variance in the response variable (80.7%), while latent variables 2, 3 and 4 add 8.2%, 3.4% and 2.4% explained variance, respectively. Using 5 or more latent variables does not add much to the target explained variance (< 0.2%) in either the calibration or validation datasets.

Table 3.5 – PLS regression to predict product yield (*Y*) from selected fluxes (*X*). Cumulative goodness-of-fit measures for the PLS regression model using up to 6 latent variables. V_X and V_Y are the explained variance in the predictor variable (central carbon fluxes) and target variable (product yield), respectively. *nRMSE* is the root mean square error normalized by the range of the target variable.

#LV	V_X (%)	V_Y (%)	R^2	p-value	R^2_{adj}	nRMSE (%)
Calibration dataset (21 points)						
1	58.2	80.7	0.808	< 0.001	0.800	15.0
2	69.8	89.0	0.892	< 0.001	0.883	11.4
3	89.5	92.4	0.924	< 0.001	0.914	9.5
4	99.3	94.8	0.949	< 0.001	0.939	7.8
5	99.7	94.9	0.949	< 0.001	0.936	7.7
6	99.9	95.0	0.950	< 0.001	0.934	7.7
Validation dataset (5 points)						
1	73.1	80.7	0.847	0.03	0.841	15.4
2	81.3	81.9	0.875	0.02	0.864	14.9
3	94.9	94.7	0.954	0.004	0.948	8.1
4	99.1	96.8	0.977	0.001	0.973	6.3
5	99.6	97.0	0.979	0.001	0.973	6.1
6	99.7	96.6	0.979	0.002	0.972	6.4

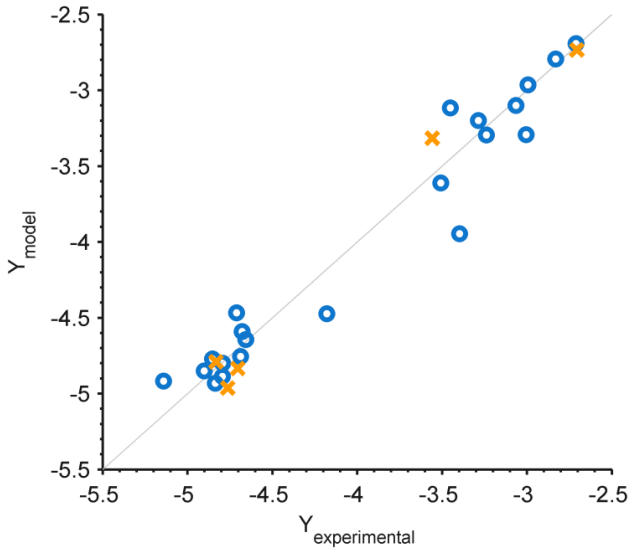


Figure 3.6 – Predictive power of hybrid MFA. Product yield predictions made by hybrid MFA were evaluated by leaving out 25% of the dataset for (×) validation and fitting the statistical model to the (o) training set.

A PLS regression model using only the first latent variable has an adjusted R^2 of 0.841 and a normalized RMSE of 15.4% for validation, similar to those found in [17] who used ordinary least squares regression based on the distribution of energy resources between growth, ATP production and maintenance for heterologous protein estimation. The final PLS model adopted in this work uses 4 latent variables and has a normalized RMSE of 6.3% for the validation dataset. Figure 3.6 shows the fitting of the model to the calibration set and prediction for the validation set.

The PLS model regression coefficients, *C* in Eq. (3.17), denote how each flux/ratio correlates to the target product yield. The analysis of such regression coefficients can unravel the metabolic state that favors the expression of the heterologous protein. Confidence intervals for the regression coefficients of the first latent variable were calculated at 90% confidence level from bootstrapping simulation with 10 000 runs. To evaluate the influence of each pathway flux (or C-ratio) on product yield, regression coefficients and their confidence intervals were classified according to strength of association. The strength of association α was defined as the ratio between the absolute value of a regression coefficient and its confidence interval. For a flux to be statistically meaningful the span of the confidence interval cannot include zero, thus the strength of association must be higher than one ($\alpha > 1$). Figure 3.7 shows the strength of association regions for the main pathways and C-ratios selected for the model.

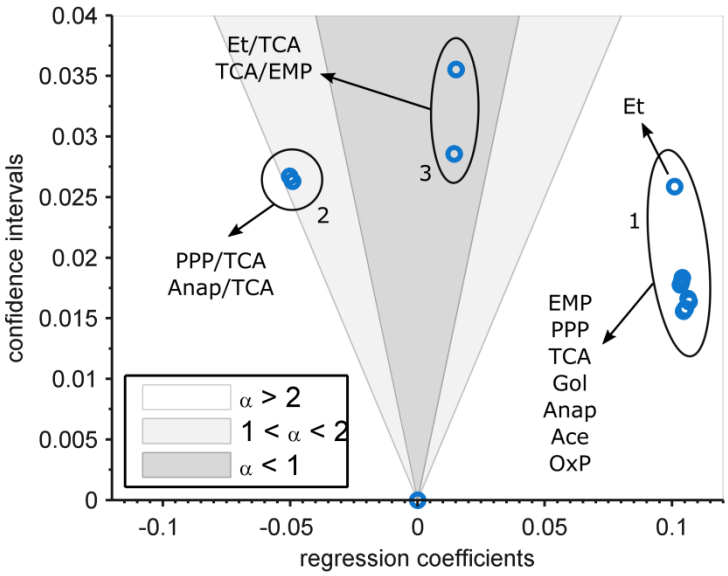


Figure 3.7 – Bootstrapping of predictor weights for the first latent variable. Applying hybrid MFA to 10 000 bootstrapping samples we can estimate the mean regression coefficients and respective confidence intervals for each selected flux/ratio. The α regions plot shows if the relationship between the predictor flux/ratio and the target variables is significant ($\alpha > 1$), highly significant ($\alpha > 2$) or non-conclusive ($\alpha < 1$).

The influence of predictor variables in the first component (Figure 3.7A) can be roughly divided into three groups. One group comprehends predictors with positive impact on product yield and is located on the right-hand side white area of the plot. It includes fluxes through all the main carbon pathways. This suggests that fluxes through main carbon pathways tend to change in the same

direction, and that product yield also follows that tendency. A second group can be seen on the left-hand side of the plot which includes C-ratios PPP/TCA and anaplerotic/TCA. This group has a negative impact on product yield with roughly half the magnitude of that of the first group (main pathways). Physiologically, these ratios represent the use of carbon for synthesis of precursors needed for biomass growth rather than for the energy-generating pathways (TCA cycle). The third and last group is located in the inner α region, which means that the influence of the ethanol/TCA ratio on product synthesis is inconclusive. This is likely the result of the high variability of ethanol rates across experiments.

These results show that the carbon ratios at main branch points are strongly correlated with the heterologous product synthesis. For the strain used in this work, we can thus conclude that high protein yields are associated with two key metabolic features: i) high overall metabolic rates through main carbon pathways and ii) relative shift of carbon flux from biosynthetic towards energy generating pathways. A similar conclusion was reached by Heyland et al. [32], who have studied amino acids supplementation in the culture medium of *Pichia pastoris*. They used a haploid *P. pastoris* strain SMD1168H (Invitrogen, Carlsbad, NM) with glucose feeding, harboring the genetic code for the expression of b-peptidyl aminopeptidase (BapA). They have concluded that higher recombinant protein yields are associated with higher percentage of carbon in the TCA, i.e. higher rates of ATP generation per unit carbon processed, which in turn is associated with lower specific growth rates.

3.4 Conclusions

The main objective of this study was to investigate the sensitivity of *P. pastoris* central carbon metabolism, and of the expression of heterologous protein in particular, to variations in medium composition. With this purpose in mind, we have cultivated a *Pichia pastoris* X-33 strain constitutively expressing a single-chain variable fragment (scFv) antibody in 26 independent shake flask experiments. Each culture was subject to varying concentrations of trace elements, main salts and glycerol, taking as reference composition the standard BSM+PTM1 medium. Thereafter, central carbon fluxes were quantified by classical MFA while heterologous protein yield was modeled by the hybrid MFA method. MFA calculations were shown to obey the consistency index criterion for all experiments. From the analysis of results we can take the following main conclusions:

(i) Final biomass yield shows little variability despite the large variations in the medium composition (10-fold variations in trace elements and 2-fold variations in the main salts and glycerol concentrations). In contrast, the product yield seems to be rather sensitive to the medium composition, showing a variability of 53.6% around the mean, which is significantly higher than the ELISA measurement error (estimated to be 15-20%);

(ii) Central carbon fluxes show a high variability among the different medium compositions. Most fluxes show 18.6-22.3% variation around the mean. Noticeably, there is a significant

variability between biosynthetic and energy generation metabolic profiles, which can only be attributed to the variations in medium compositions;

(iii) The hybrid MFA predicted the antibody fragment yield in 5 independent experiments with a normalized RMSE of 6.3%. Furthermore, analysis of PLS regression coefficients showed that the metabolic state that promotes high protein yields is characterized by high overall metabolic rates through main central carbon pathways concomitantly with a relative shift of carbon flux from biosynthetic towards energy generating pathways.

All in all this study shows that, for the *P. pastoris* strain used and concentration ranges tested, cell growth metabolism is relatively insensitive to the medium composition while the synthesis and expression of the antibody fragment is much more sensitive to the culture medium. This result seems to suggest the customization of culture medium may be an important factor to optimize the expression of specific heterologous proteins in *P. pastoris* cultures.

References

1. Ahmad M, Hirz M, Pichler H, Schwab H (2014) Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol 98:5301–17
2. Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol 13:329–332
3. Heyland J, Fu J, Blank LM, Schmid A (2010) Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. Biotechnol Bioeng 107:357–68
4. Cregg JM, Vedvick TS, Raschke WC (1993) Recent Advances in the Expression of Foreign Genes in *Pichia pastoris*. Nat Biotechnol 11:905–910
5. Hagman A, Säll T, Piškur J (2014) Analysis of the yeast short-term Crabtree effect and its origin. FEBS J 281:4805–14
6. Tschopp JF, Brust PF, Cregg JM, et al. (1987) Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. Nucleic Acids Res 15:3859–3876
7. Waterham HR, Digan ME, Koutz PJ, et al. (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene 186:37–44
8. Vogl T, Glieder A (2013) Regulation of *Pichia pastoris* promoters and its consequences for protein production. N Biotechnol 30:385–404
9. Mattanovich D, Callewaert N, Rouze P, et al. (2009) Open access to sequence: Browsing the *Pichia pastoris genome*. Microb Cell Fact 8:53
10. Chung BK, Selvarasu S, Andrea C, et al. (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb Cell Fact 9:50
11. Sohn SB, Graf AB, Kim TY, et al. (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol J 5:705–15
12. Baumann K, Carnicer M, Dragosits M, et al. (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. BMC Syst Biol 4:141

13. Dragosits M, Stadlmann J, Albiol J, et al. (2009) The effect of temperature on the proteome of recombinant *Pichia pastoris*. J Proteome Res 8:1380–92
14. Unrean P (2014) Pathway analysis of *Pichia pastoris* to elucidate methanol metabolism and its regulation for production of recombinant proteins. Biotechnol Prog 30:28–37
15. Jordà J, Jouhten P, Cámara E, et al. (2012) Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. Microb Cell Fact 11:57
16. González-Martínez JM, Folch-Fortuny A, Llaneras F, et al. (2014) Metabolic flux understanding of *Pichia pastoris* grown on heterogenous culture media. Chemom Intell Lab Syst 134:89–99
17. Tortajada M, Llaneras F, Ramón D, Picó J (2012) Estimation of recombinant protein production in *Pichia pastoris* based on a constraint-based model. J Process Control 22:1139–1151
18. Tortajada M, Llaneras F, Picó J (2010) Validation of a constraint-based model of *Pichia pastoris* metabolism under data scarcity. BMC Syst Biol 4:115
19. Carinhas N, Bernal V, Teixeira A, et al. (2011) Hybrid metabolic flux analysis: combining stoichiometric and statistical constraints to model the formation of complex recombinant products. BMC Syst Biol 5:34
20. Invitrogen Life Technologies (2000) *Pichia* Fermentation Process Guidelines.
21. Ferreira AR, Ataíde F, von Stosch M, et al. (2012) Application of adaptive DO-stat feeding control to *Pichia pastoris* X33 cultures expressing a single chain antibody fragment (scFv). Bioprocess Biosyst Eng 35:1603–14
22. Ferreira AR, Dias JML, von Stosch M, et al. (2013) Fast development of Pichia pastoris GS115 Mut(+) cultures employing batch-to-batch control and hybrid semi-parametric modeling. Bioprocess Biosyst. Eng.
23. Jahic M, Rotticci-Mulder JC, Martinelle M, et al. (2002) Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. Bioprocess Biosyst Eng 24:385–393
24. Wang NS, Stephanopoulos G (1983) Application of macroscopic balances to the identification of gross measurement errors. Biotechnol Bioeng 25:2177–208
25. Van der Heijden RT, Heijnen JJ, Hellinga C, et al. (1994) Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. Biotechnol Bioeng 43:3–10
26. Klamt S, Schuster S, Gilles ED (2002) Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. Biotechnol Bioeng 77:734–51
27. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185:1–17
28. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130
29. d'Anjou MC, Daugulis AJ (2000) Mixed-feed exponential feeding for fed-batch culture of recombinant methylotrophic yeast. Biotechnol Lett 22:341–346
30. Brady CP, Shimp RL, Miles AP, et al. (2001) High-level production and purification of P30P2MSP1(19), an important vaccine antigen for malaria, expressed in the methylotropic yeast *Pichia pastoris*. Protein Expr Purif 23:468–75
31. Carnicer M, Baumann K, Töplitz I, et al. (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. Microb Cell Fact 8:65
32. Heyland J, Fu J, Blank LM, Schmid A (2011) Carbon metabolism limits recombinant protein production in *Pichia pastoris*. Biotechnol Bioeng 108:1942–53
33. Nocon J, Steiger MG, Pfeffer M, et al. (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. Metab Eng 24:129–138

Additional files

Additional file 3.1 – *Pichia pastoris* network model for central carbon metabolism

Includes description of metabolites, reactions and stoichiometric network.

Additional file 3.2 – Metabolic flux distribution

Metabolic flux distribution for all 26 shake flask experiments with 95% confidence intervals and respective consistency index.

Chapter 4
Analysis of culture media screening data by projection to latent pathways [3]

Abstract

Cell culture media formulations contain hundreds of individual components in water solutions which have complex interactions with metabolic pathways. The currently used statistical design methods are empirical and very limited to explore such a large design space. In a previous work we developed a computational method called projection to latent pathways (PLP), which was conceived to maximize covariance between envirome and fluxome data under the constraint of metabolic network elementary flux modes (EFMs). More specifically, PLP identifies a minimal set of EFMs (i.e. pathways) with the highest possible correlation with envirome and fluxome measurements. In this paper we extend the concept for the analysis of culture media screening data to investigate how culture medium components up-regulate or down-regulate key metabolic pathways. A *P. pastoris* X-33 strain was cultivated in 26 shake flask experiments with variations in trace elements concentrations and basal medium dilution, based on the standard BSM+PTM1 medium. PLP identified 3 EFMs (growth, maintenance and by-product formation) describing 98.8% of the variance in observed fluxes. Furthermore, PLP presented an overall predictive power comparable to that of PLS regression. Our results show iron and manganese at concentrations close to the PTM1 standard inhibit overall metabolic activity, while the main salts concentration (BSM) affected mainly energy expenditures for cellular maintenance.

Keywords

Pichia pastoris X-33 · culture media design · elementary flux modes · projection to latent pathways

[3] Based on the paper: Isidro IA, Ferreira AR, Clemente JJ, Cunha AE, and Oliveira R (2016) Analysis of culture media screening data by projection to latent pathways: the case of Pichia pastoris X-33. J Biotechnol 217: 82–89

4.1 Introduction

Cell culture media formulations consist of nutrients, such as peptones, amino acids, meat- and yeast-extracts, minerals and vitamins, inhibitors and solidifying agents. Some medium components may be critical for cell growth or productivity, others may be toxic at certain levels, and many may be involved in complex interactions in the same or competing pathways within the cell. Traditionally, medium components are screened individually (one factor at a time – OFAT) or in small combinations in parallel shake flask experiments, which significantly limits the ability to discover complex interactions between medium components. Statistical design of experiments (DoE) has been applied in numerous studies of media optimization for bacterial, fungal, mammalian and stem cell cultures supported by reactor or shake flask experiments [1–3]. It has been also applied to improve the standard basal salt medium (BSM) in *P. pastoris* cultivations [4].

The most common statistical DoE method is reduced factorial design with two levels of concentrations, which permits a preliminary screening of between five and ten medium factors in a limited number of experiments [1, 5]. With this method, several medium factors are simultaneously compared and their effects are measured and ranked based on analyzed phenotypic response variables, which should be as relevant as possible to the intended purpose of the medium design. For example, if efficient cell propagation is the primary aim, cell counts would be the appropriate response; if a certain metabolic state of the culture is the aim, characteristic metabolites should be analysed and used as response variables [1]. Then, the medium factors are ranked in relation to their influence on the phenotypic target response, and the most effective factors are selected and further tested in additional experiments. In a last stage, when sufficient data is available, a regression model is used to determine optimal levels of the medium factors and a final optimal medium formulation is tested. Regression models maybe linear or quadratic depending on the number of levels of the DoE adopted. It is however generally accepted that medium development by statistical DoE is expensive when applied to many factors with potential interactions. Moreover, these methods are eminently empirical and limited mechanistic knowledge is used or produced in the medium optimization process.

In this study we explore a more rational culture medium design method based on the concept of elementary flux modes (EFM). EFM analysis is currently a popular technique to study the topology of metabolic networks [6–8]. A EFM may be defined as a minimal set of enzymes able to operate at steady state, with the enzymes weighted by the relative flux they need to carry for the mode to function [9]. The universe of EFMs of a given metabolic network defines the full set of non-decomposable steady-state flux distributions that the network can support. Any particular steady-state flux distribution can be expressed as a non-negative linear combination of EFMs:

$$\mathbf{r} = \lambda_1 \mathbf{e_1} + \lambda_2 \mathbf{e_2} + \cdots + \lambda_{nem} \mathbf{e_{nem}} = \sum_{i=1}^{nem} \lambda_i \mathbf{e}_i \qquad \textbf{(4.1)}$$

where $\mathbf{r}$ is the vector of nr metabolic fluxes, $\mathbf{e_i}$ are nem elementary flux modes vectors of length nr, and λ_i are nonnegative EFM weighting factors (scalar variables) which denote the contribution of each EFM to the flux-phenotype of the cells. The nem number of EFMs of a metabolic network is in general very high, denoting the innate adaptability and robustness of cellular systems. According to a study by Zanghellini and co-workers [10], the genome-scale metabolic network of *Pichia pastoris* can have billions of elementary flux modes. In this study we will focus on the central carbon metabolism and thus a core network is sufficient. We used the network described in [11] (Chapter 3), which has a total number of 158 EFMs.

In a previous work we developed a computational method called projection to latent pathways (PLP), which was conceived to minimize the number of active EFMs while maximizing the correlation of envirome and fluxome measurements [12, 13]. When applied to dynamic envirome data in bioreactor experiments, PLP identifies metabolic pathways activation/deactivation dynamic patterns that are the consequence of changing environmental conditions [12]. Here we extend the PLP concept to the analysis of culture media data and compare it to other factorization regression methods, namely PLS. We use as illustrative example the data of a previous *Pichia pastoris* X-33 culture media optimisation study [11] (Chapter 3), where we have investigated the effect of trace elements and main salts concentrations on the metabolism of *P. pastoris* cells.

4.2 Materials and methods

4.2.1 *Pichia pastoris* culture media screening data

A *Pichia pastoris* X-33 strain was grown in 26 shake flask medium screening experiments. These experiments resulted in a data set composed of a 26 × 12 matrix of medium compositions, $\mathbf{X}$, and a 26 × 43 matrix of response metabolic fluxes, $\mathbf{R}$. Briefly, the reference medium composition is the basal salt medium (BSM) with *Pichia* trace minerals 1 (PTM1) [14]. Each experiment explores a different combination of two possible levels (high or low) for 11 medium components, namely the concentrations of the 10 trace components in PTM1 (which can be 1:1 or 1:10 of the reference formulation) and the dilution of BSM (1:1 or 1:2). The medium factors thus include the concentrations of 10 ions that result from the dissociation of trace salts, the concentration of biotin, and the dilution used for the BSM solution. Table 4.1 shows the concentration of these components in the different medium solutions. For each shake flask experiment, the central carbon metabolic flux distribution was determined by metabolic flux analysis (MFA) at conditions of exponential cell growth (data available as Additional file 4.1). Details regarding the experimental method and metabolic flux analysis are provided in a previous paper [11] (Chapter 3).

Table 4.1 – Medium composition in dissociated salts. Concentration (mg/L) for trace salts ions and biotin in final medium solution, and basal salts dilution for each shake flask experiment. Experiment 25 is the BSM+PTM1 reference (upper level) control and experiment 26 is the lower level control.

Exp. ID	Cu^{2+}	Na^{+}	I^{-}	Mn^{2+}	MoO_4^{2-}	B	Co^{2+}	Zn^{2+}	Cl^{-}	Fe^{2+}	Biotin	BSM dil.
1	6.642	0.070	0.295	0.424	0.058	0.015	0.099	41.73	45.37	5.679	0.087	1:1
2	6.642	0.171	0.029	0.424	0.575	0.002	0.099	4.173	4.644	56.79	0.087	1:1
3	6.642	0.171	0.029	4.242	0.575	0.015	0.099	41.73	45.37	56.79	0.870	1:2
4	6.642	0.070	0.295	0.424	0.058	0.015	0.987	4.173	5.713	56.79	0.870	1:2
5	6.642	0.219	0.295	4.242	0.575	0.015	0.099	4.173	4.644	5.679	0.870	1:1
6	0.664	0.219	0.295	0.424	0.575	0.015	0.987	41.73	46.44	5.679	0.870	1:2
7	0.664	0.219	0.295	4.242	0.575	0.015	0.987	41.73	46.44	56.79	0.870	1:1
8	6.642	0.219	0.295	4.242	0.575	0.002	0.987	41.73	46.44	56.79	0.087	1:2
9	6.642	0.171	0.029	0.424	0.575	0.002	0.099	41.73	45.37	56.79	0.870	1:2
10	0.664	0.070	0.295	4.242	0.058	0.002	0.099	4.173	4.644	5.679	0.870	1:2
11	0.664	0.219	0.295	0.424	0.575	0.002	0.987	4.173	5.713	56.79	0.087	1:1
12	0.664	0.022	0.029	0.424	0.058	0.015	0.099	41.73	45.37	56.79	0.870	1:1
13	0.664	0.171	0.029	4.242	0.575	0.015	0.987	4.173	5.713	5.679	0.087	1:1
14	6.642	0.219	0.295	0.424	0.575	0.002	0.987	41.73	46.44	5.679	0.870	1:1
15	6.642	0.171	0.029	0.424	0.575	0.015	0.987	4.173	5.713	5.679	0.087	1:2
16	6.642	0.070	0.295	4.242	0.058	0.015	0.099	4.173	4.644	56.79	0.087	1:2
17	6.642	0.022	0.029	4.242	0.058	0.002	0.987	4.173	5.713	5.679	0.870	1:1
18	0.664	0.070	0.295	4.242	0.058	0.002	0.987	41.73	46.44	56.79	0.087	1:1
19	0.664	0.022	0.029	0.424	0.058	0.015	0.987	41.73	46.44	5.679	0.087	1:2
20	0.664	0.171	0.029	4.242	0.575	0.002	0.099	41.73	45.37	5.679	0.087	1:2
21	0.664	0.219	0.295	0.424	0.575	0.015	0.099	4.173	4.644	56.79	0.087	1:1
22	0.664	0.022	0.029	0.424	0.058	0.002	0.987	4.173	5.713	56.79	0.870	1:2
23	6.642	0.070	0.295	0.424	0.058	0.002	0.099	41.73	45.37	5.679	0.087	1:1
24	6.642	0.022	0.029	4.242	0.058	0.015	0.987	41.73	46.44	56.79	0.087	1:1
25	6.642	0.219	0.295	4.242	0.575	0.015	0.987	41.73	46.44	56.79	0.870	1:1
26	0.664	0.022	0.029	0.424	0.058	0.002	0.099	4.173	4.644	5.679	0.087	1:2

4.2.2 *Pichia pastoris* metabolic elementary flux modes

The universe of possible EFMs was generated for the *P. pastoris* central carbon stoichiometric network described in [11] (Chapter 3). This network describes the main catabolic pathways, namely glycolysis and gluconeogenesis, the tricarboxylic acid (TCA) cycle, the pentose phosphate, anaplerotic, fermentative and phosphorylative oxidation pathways. It supports growth on glucose, glycerol and methanol (or mixed feeds of these substrates), which are the main carbon sources used in *P. pastoris* cultures. The model includes NADH and NADPH coenzymes balancing and a biomass formation reaction from selected internal metabolites based on known macromolecular composition of *P. pastoris* cells [15]. The network consists of 43 metabolic reactions, 34 internal metabolites, and 9 exchange metabolites. The freely available EFMtool [16] was used to calculate 158 possible elementary flux modes for main carbon pathways in *Pichia pastoris*. The universe of

possible EFMs is represented as a $nr \times nem$ elementary mode matrix ($\mathbf{KEM}$), where nr is the number of reactions in the metabolic network and nem is the number of elementary modes. The elementary flux modes matrix is supplied as Additional file 4.2, together with the stoichiometric network it is based on.

4.2.3 Projection to latent pathways

Projection to latent pathways (PLP) is a constrained version of the widely used partial least squares (PLS) regression, also known as projection to latent structures [17]. PLS relates two data matrices $\mathbf{X}$ (predictor) and $\mathbf{Y}$ (response) with a multivariate linear model, with the two matrices decomposed into scores and weights in such a way that maximizes covariance between the scores. As a consequence, the latent variables (LV) extracted from $\mathbf{X}$ will be the ones with higher predictive power for $\mathbf{Y}$. This covariance-guided decomposition gives PLS the ability to analyze noisy, highly collinear data, and with more variables than observations. PLS was implemented in MATLAB based on the NIPALS algorithm [18].

Projection to latent pathways differs from PLS in that it constrains Y-weights to structures – the elementary flux modes – that represent metabolic pathways. Equation (4.2) illustrates the decomposition of $\mathbf{Y}$ in PLS, where $\mathbf{U}$ are the scores and $\mathbf{Q}$ are the weights. In PLP, Eq. (4.4), $\mathbf{Y}$ is a matrix of metabolic rates, $\mathbf{R}$, and the Y-weights matrix is a ranked subset, $\mathbf{REM}$, of the elementary flux modes matrix. This subset represents EFMs that are active under the experimental conditions, ordered by the amount of variance explained in the measured flux data. So the Y-scores $\mathbf{\Lambda}$ represent the strength of activation of each elementary flux mode by environmental factors in $\mathbf{X}$ (4.5), using a PLS-like approach (4.3), where $\mathbf{C}$ is a matrix of regression coefficients.

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^{\mathrm{T}} + \boldsymbol{\varepsilon}_{\mathbf{Y}} \tag{4.2}$$

$$\widehat{\mathbf{U}} = \mathbf{X} \cdot \mathbf{C} \tag{4.3}$$

$$\mathbf{R} = \mathbf{\Lambda} \cdot \mathbf{REM}^{\mathrm{T}} + \boldsymbol{\varepsilon}_{\mathbf{R}} \tag{4.4}$$

$$\widehat{\mathbf{\Lambda}} = \mathbf{X} \cdot \mathbf{C} \tag{4.5}$$

The PLP algorithm was implemented in MATLAB as described in a previous paper [13]. Running PLP requires three pieces of information:

(i) Envirome data: the 26 × 12 data matrix $\mathbf{X}$ of medium factors shown in Table 4.1 and Additional file 4.1.

(ii) Fluxome data: the 26 × 43 data matrix $\mathbf{R}$ of central carbon fluxes provided in Additional file 4.1.

(iii) Universe of EFM candidates: the 158 × 43 $\mathbf{KEM}$ matrix provided as Additional file 4.2

The envirome and fluxome data was randomly partitioned into a 21 data points training set and a 5 data points validation set. $\mathbf{X}$ is normalized by unit variance and mean-centering, $\mathbf{R}$ is normalized only by unit variance in PLP and by unit variance and mean-centering in PLS. The reason why $\mathbf{R}$ is not mean centered in PLP is to maintain consistency with the definition of EFM in Eq. (4.1).

A bootstrapping analysis was performed to evaluate the consistency of EFM selection in PLP. For each of the 1000 runs, a sample of size 26 was selected from the original 26 observations using random sampling with replacement and a PLP model with 5 components was fitted to that sample. For each selected EFM the bootstrapping analysis allowed the determination of the average regression coefficients ($\bar{\mathbf{C}}$) and the respective 90% confidence intervals ($\mathbf{CI}$).

4.3 Results and discussion

4.3.1 *Pichia pastoris* elementary flux modes

The central carbon network of *Pichia pastoris* yielded 158 EFMs, representing all possible metabolic operational modes of *Pichia pastoris* cells for substrates uptake, energy generation, cell growth and by-products synthesis (full description provided in Additional file 4.2). Among these, 119 EFMs result in biomass growth. Biomass can be synthesized from all possible combinations of the three carbon sources included in the network: glucose, glycerol and methanol. Only 23 EFMs rely on a single carbon source, among these 8 use glucose, 8 use glycerol and 7 use methanol. The network also supports the synthesis of glycerol from other sources and there are 68 elementary flux modes that lead to glycerol being produced rather than consumed. Given that glycerol was the only carbon source fed to the cells, only the 8 glycerol-using elementary flux modes are good candidates for representing the metabolism of cultured *P. pastoris* cells. These comprise EFMs 9, 10, 13 and 18 for biomass growth and EFMs 147, 149, 150 and 153 for catabolic activity with or without by-product formation (ethanol, citrate or pyruvate). Theoretical biomass yield varies between 0.52 and 0.72 g-biomass/g-glycerol, given that some elementary modes for growth also produce fermentative by-products. The experimental biomass yield estimated as the ratio of final biomass to initial glycerol ranges from 0.37 to 0.43 g/g, which is significantly lower than the theoretical maximum. This essentially means that a significant percentage of glycerol is spent in catabolic EFMs with the resulting energy used for maintenance processes.

4.3.2 Projection to latent pathways

Projection to latent pathways was used to discriminate a minimal set of EFMs that best explain the observed metabolic flux distribution (response variable) while maximizing correlation with culture medium composition (predictor variable). PLP regression was performed over a training set of 21 points, i.e. 21 different medium compositions, and was thereafter used to predict fluxes of 5

independent medium compositions (validation set). The results for models with 1 to 5 EFMs are shown in Table 4.2. The first three EFMs individually explain 65.6, 27.5, and 5.8% of variance in the observed fluxes, whereas the fourth EFM explains only 0.4%. Based on the inflection of the RMSE curve for the training and validation data sets (see Table 4.2) the optimal PLP model has 3 EFMs. The remaining EFMs explain less than 1% of the target variance, which can be reasonably assumed to fall within the experimental error.

Table 4.2 – PLP regression results. Cumulative goodness-of-fit measures for the PLP regression model using up to 5 elementary flux modes (latent variables). V_X and V_R are the explained variance in the predictor (medium factors) and target variables (fluxes). $RMSE_{val}$ is the root mean square error for the validation set.

#EFM	V_X (%)	V_R (%)	R^2	p-value	R^2_{adj}	RMSE	$RMSE_{val}$
1	70.1	65.6	0.589	< 0.001	0.569	2.66	2.83
2	75.0	93.0	0.878	< 0.001	0.868	1.20	1.12
3	77.5	98.8	0.977	< 0.001	0.974	0.488	0.452
4	81.2	99.2	0.984	< 0.001	0.981	0.400	0.408
5	86.0	99.3	0.985	< 0.001	0.982	0.387	0.423

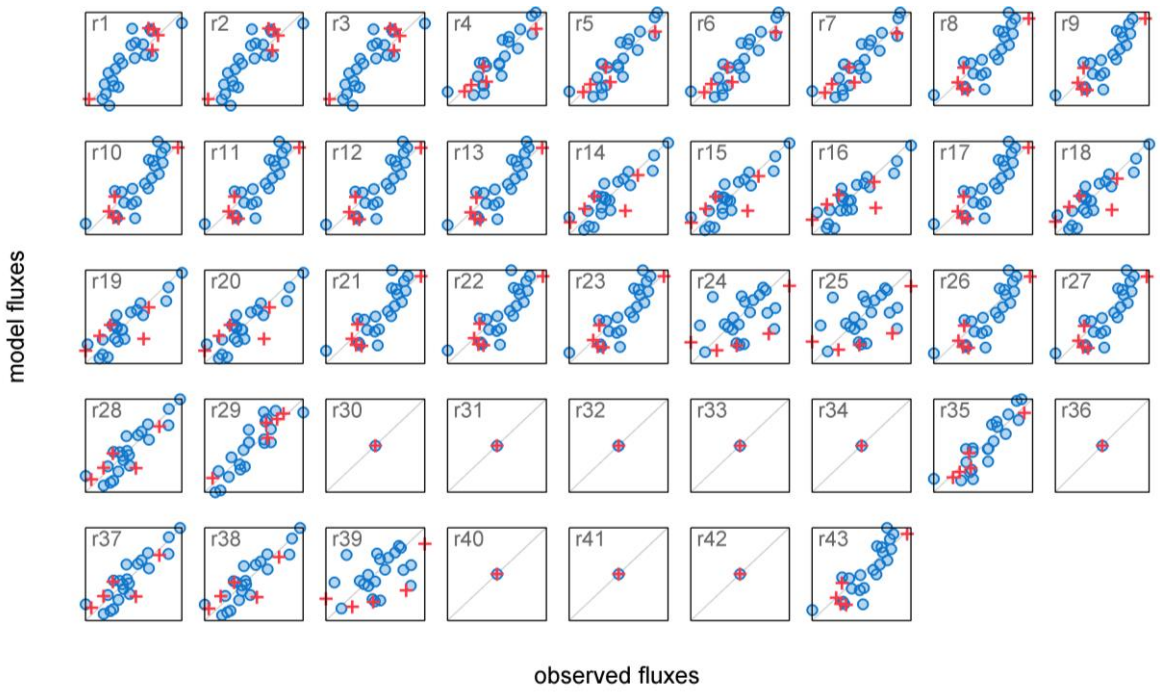


Figure 4.1 – PLP regression results. Predictive power of final PLP model with three elementary flux modes. Correlation between observed and predicted central carbon fluxes for (o) training and (+) validation data sets.

From the universe of 158 possible EFMs, PLP has discriminated 3 that account for 98.8% of the total variance in observed metabolic fluxes and 77.5% of the variance in the envirome factors. Figure 4.1 shows the generally good correlation between the observed fluxes and PLP predictions both for the training and for the validation sets. The possible exception is for reactions related to

ethanol production (r24-25 and r39) where there is a higher dispersion of residuals which are however near Gaussian (results not shown).

The three selected elementary modes, ranked in order of descending importance for explained variance were EFM 10, 147 and 149 (described in Table 4.3). EFM 10, which individually explains 65.6% of flux variation, represents biomass growth from glycerol as a sole carbon source. This result supports, as expected, that most of the carbon intake is directed towards biomass synthesis. A breakdown of explained variance by reaction (Figure 4.2) shows that EFM 10 explains variance in all key reactions needed for biomass growth (gluconeogenesis, pentose phosphate and anaplerotic pathways), and less in the key catabolic reactions for energy generation (glycolysis, TCA cycle and phosphorylative oxidation). The second elementary mode (EFM 147), which adds 27.5% explained variance, represents aerobic respiration from glycerol as carbon source and explains variance in the energy-generating reactions. Finally, EFM 149 as a third component adds 5.8% explained variance by accounting for some of the variability observed in fluxes related to ethanol synthesis.

Table 4.3 – Description of frequently selected EFMs. Description of frequently selected elementary flux modes in terms of their stoichiometry for exchange reactions. EFMs are ordered by selection frequency in bootstrapping (only EFMs with Freq. > 10% are shown). The full stoichiometry of each EFM is described in Additional file 4.2.

Freq. (%)	EFM	Stoichiometry of macroscopic reactions *
100	10	0.406 Gly +0.405 O_2 → Biom + 0.217 CO_2
100	147	0.143 Gly + 0.500 O_2 → 0.429 CO_2
100	149	Gly + 0.500 O_2 → Et + CO_2
38.2	72	0.083 Glu + 0.500 O_2 → 0.500 CO_2
22.8	2	0.386 Gly + 0.058 Meth + 0.500 O_2 → Biom + 0.217 CO_2
18.4	111	0.171 Gly + 0.117 Glu + 0.288 O_2 → Biom + 0.217 CO_2
12.0	46	0.405 Gly + 0.503 Glu + 0.058 Meth + 0.500 O_2 → Biom + 0.215 Et + 0.432 CO_2
12.0	38	Meth + 0.500 O_2 → 0.405 Gly + 0.067 Cit
10.7	131	0.067 Glu + -0.100 Meth + 0.550 O_2 → 0.500 CO_2

*Gly – glycerol, Glu – glucose, Meth – methanol, Cit – citrate, Et – ethanol, Biom – biomass. Units are mol for all metabolites and C-mol for biomass.

Figure 4.3 shows the biplot of the response fluxes for the first two model components. Because PLP latent structures are EFMs determined a priori from the network topology, the loadings and scores have metabolic meaning. In Figure 4.3 the loadings represent the contribution of each reaction in the metabolic network to the first two selected elementary modes (EFM 10 for growth and EFM 147 for energy production). These reactions can be roughly categorized into (i) reactions that contribute to both elementary flux modes, (ii) reactions that contribute only for one of the two, and (iii) reactions that do not participate in either. The first group includes glycerol metabolism and most reactions in glycolysis and TCA cycle, or in other words, reactions that are necessary both for growth and for energy production. The second group includes gluconeogenesis, the pentose phosphate and anaplerotic pathways which are used to generate precursors and reduced

coenzymes for biomass synthesis. The last group comprises reactions for methanol utilization and exchange reactions for other carbon sources and by-product excretion. The scores in Figure 4.3 represent the EFM weights, Λ, of EFM 10 and EFM 147 in each shake flask experiment. The results show a positive correlation (R^2 = 0.813) between the two in the direction of the common reaction loadings, suggesting higher growth rates and higher catabolic rates are generally observed together.

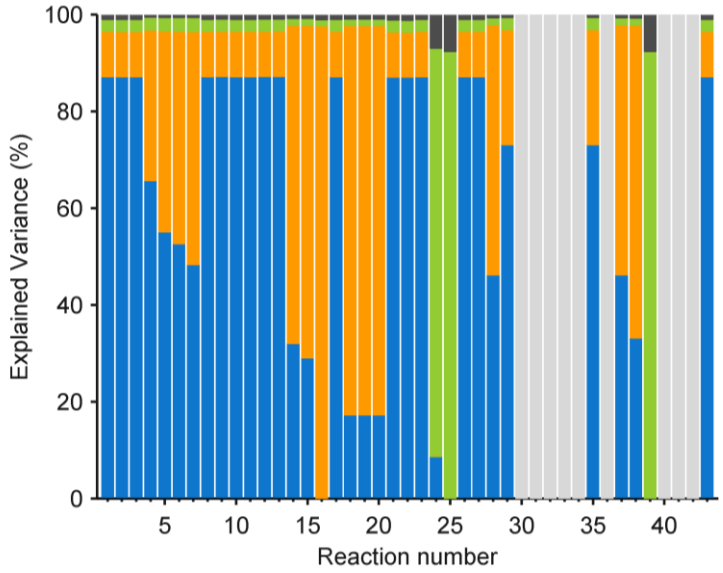


Figure 4.2 – Explained variance by reaction for the PLP model. Explained variance for each reaction in the central carbon network. Breakdown by component: ■ EFM 10, ■ EMF 147, ■ EFM 149, ■ unexplained variance, ■ non defined (flux is 0).

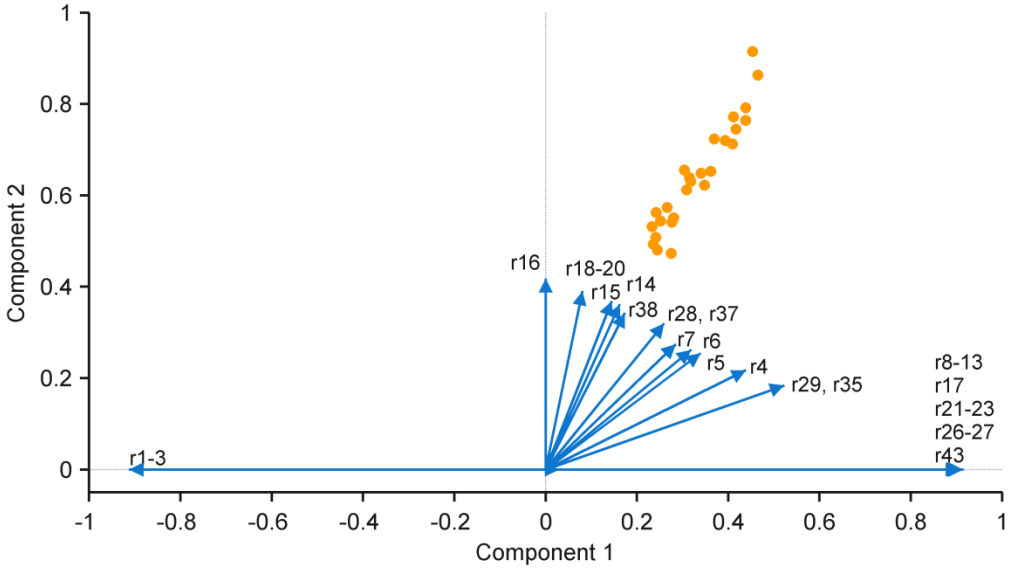


Figure 4.3 – R-biplot for the PLP model. Decomposition of target fluxes for the first two components: (▸) loadings represent the contribution of each reaction to the first and second selected EFMs, (o) scores are the EFM weights, Λ, for each experiment.

4.3.3 Consistency of EFM selection

PLS-like regressions are generally well-suited for modeling high dimensional data sets with few observations. Nevertheless, for a reduced number of observations data partitioning into training and validation sets may have a significant impact in the final regression model. A bootstrapping analysis was therefore performed to access the consistency of EFM selection. Table 4.3 lists all elementary modes with a frequency of selection over 10%. EFMs 10, 147 and 149 are always selected. They all use glycerol as single substrate for growth, maintenance and ethanol fermentation, respectively. Other selected elementary modes had frequencies under 40%. These use glucose or mixed feeds of glucose, glycerol and/or methanol as carbon source. Since these carbon sources were not fed to the cells, we know the corresponding EFMs are not feasible. The results show less frequently selected EFMs are very sensitive to the data partitioning and to experimental noise and thus less reliable to interpret and they should be excluded from the final model.

4.3.4 Effect of medium factors on metabolism

PLP regression coefficients set a linear model for EFM weighting factors, $\mathbf{\Lambda}$, as function of medium factors (Eqs. 4.4-4.5), denoting either a causal or effector relationship between environmental state and active pathways. These regression coefficients, together with respective confidence intervals, can be used to identify meaningful relationships, although this interpretation should be done with care as they cannot discriminate between a cause and an effect. In this case however, the measured fluxes reflect the metabolic state at exponential growth (before nutrient depletion), while culture media components are design parameters set a priori. This essentially means that medium components cannot be a consequence of cellular metabolism but rather the cause of the observed differences in metabolic flux distribution.

The bootstrapping analysis provided estimates of the mean value of regression coefficients and respective confidence intervals. The strength of association, α, was defined as the ratio between the absolute value of a regression coefficient and its confidence interval. For a flux to be statistically meaningful the span of the confidence interval cannot include zero, thus the strength of association must be higher than one ($\alpha > 1$) to be meaningful, or ideally $\alpha > 2$ for a strong association. Figure 4.4 shows these relationships and their strength of association areas for each of the three selected elementary modes. Moreover, the overall importance of each medium component was quantified in terms of variable importance on projection (Figure 4.5). VIP is calculated as the sum of regression coefficients (for each environmental factor) weighted by the explained variance of each EFM and may be interpreted as the importance of each medium component to the full model (with the 3 EFMs). Figure 4.5 shows that iron, manganese and BSM dilution have much higher VIP in comparison to the other culture medium factors. The analysis of these results leads to the following main conclusions:

- Concentrations of biotin, copper, sodium, iodine, molybdate and boron have no consistent impact in the metabolism of *P. pastoris* cells in the concentration ranges studied;
- Zinc and chlorine have a less significant correlation with cell growth rate and no correlation at all with maintenance and therefore it is difficult to extract a meaningful interpretation;
- Iron presents a strong negative correlation to all selected EFMs suggesting that iron inhibits overall metabolic activity and thus its concentration should be reduced;
- Manganese presents a moderate negative correlation to all three EFMs suggesting that manganese should also be reduced;
- BSM dilution has no effect on specific cellular growth rate but correlates negatively with cellular maintenance, suggesting that decreasing overall basal medium concentrations increases energy expenditures for cellular maintenance;
- Finally, it is very significant that iron and manganese effect on metabolism supersede that of main salts concentration (BSM), showing that trace elements composition is highly important for *Pichia pastoris* cultures.

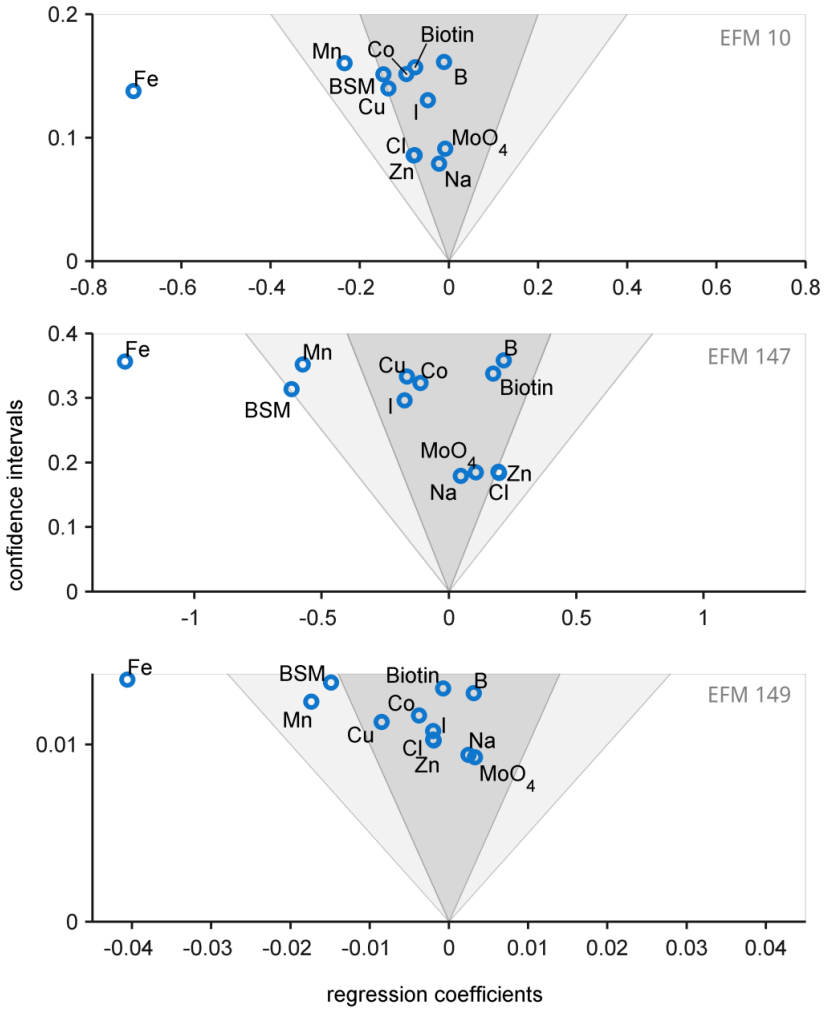


Figure 4.4 – Bootstrapping of medium factors weight on selected elementary flux modes. The α regions plot shows if the relationship between the predictor and the target is ▢ non-conclusive ($\alpha < 1$), ▢ significant ($\alpha > 1$), or ▢ highly significant ($\alpha > 2$).

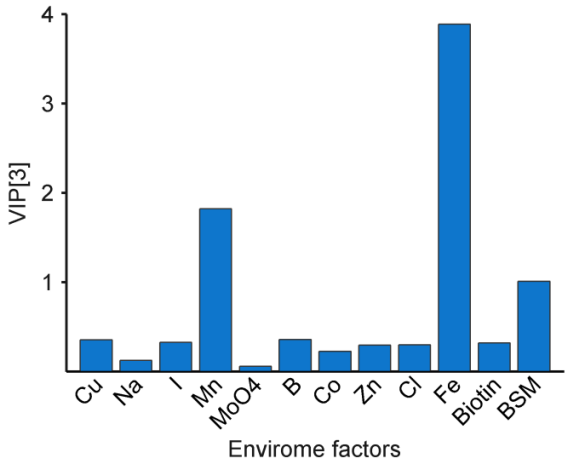


Figure 4.5 – Variable importance on projection for the PLP model with 3 EFMs. Weighted sum of regression coefficients for each environmental factor, normalized by the explained variance of each component

4.3.5 Comparison to PLS

Projection to latent pathways works in a similar way to PLS regression. Both methods attempt to describe a set of response variables from a set of predictor variables by projecting them into latent input and output variables, which are regressed against each other, in a way that maximizes co-variance. The fundamental difference is that for PLP the latent variables for the response variables (fluxes) are constrained to fixed structures, the EFMs [13]. In order to evaluate how this extra constraint affects the predictive power of PLP, a PLS regression was applied to the same data with the same partitioning between training and validation data points.

PLS regression models with 1 to 5 components were fit to the data with results shown in Table 4.4. The explained variance of target fluxes is 98.7% for the first component alone, whereas the second component adds 0.4% explained variance. Using the same criteria as for PLP, based on RMSE curves and 1% explained variance cut-off, the optimal number of components for the PLS model is one. The final PLS model explains 98.7% of the variance in target fluxes based on 72.7% of variance in predictor medium factors. Given that PLP is able to predict 98.8% of target variance based on 77.5% of predictor variance, this result suggests that the predictive capability of the final PLS and PLP models is not affected by the extra constraints. A key difference however is that the PLP needs three components (i.e. EFMs) to explain a similar variance to PLS with just one component, which may be seen as disadvantage of PLP in terms of model complexity. This result also suggests that the first PLS latent variable lumps the contributions of the 3 EFMs identified by PLP. This is coherent with the fact that each EFM explains variance in a distinct group of metabolic reactions, as discussed before.

Two other details are worth analysis in the comparison between PLP and PLS. The first refers to the prediction of ethanol-producing reactions (r24-25 and r39). As seen in Figure 4.6, ethanol related fluxes predictions by PLS are clearly biased in comparison to PLP (Figure 4.1). In fact PLS prediction bias in ethanol fluxes is eliminated only with 4 latent variables (results not shown) while PLP needs only 3 EFMs to achieve the same. The second observation refers to the handling of outliers in the data. Taking as example reactions r31 and r32, which are part of the methanol utilization pathway, we know that their fluxes should be zero as methanol was not fed to the cells. However, due to MFA error propagation, some observations attained nearly zero (e.g. 10^{-6}) flux values, which are clearly data outliers. While PLS normalizes these outliers and attempts to predict them (Figure 4.6), PLP takes advantage of the metabolic-meaningful latent variables and is able to discard those observations because they are inconsistent with the rest of the network (Figure 4.1).

Table 4.4 – PLS regression results. Cumulative goodness-of-fit measures for the PLS regression model using up to 5 latent variables. V_X and V_Y are the explained variance in the predictor (medium factors) and target variables (fluxes). $RMSE_{val}$ is the root mean square error for the validation set.

#LV	V_X (%)	V_Y (%)	R^2	p-value	R^2_{adj}	RMSE	$RMSE_{val}$
1	72.7	98.7	0.975	< 0.001	0.974	0.508	0.359
2	77.9	99.1	0.982	< 0.001	0.980	0.430	0.398
3	81.4	99.3	0.986	< 0.001	0.984	0.379	0.406
4	85.1	99.3	0.987	< 0.001	0.984	0.366	0.431
5	88.3	99.4	0.987	< 0.001	0.984	0.362	0.485

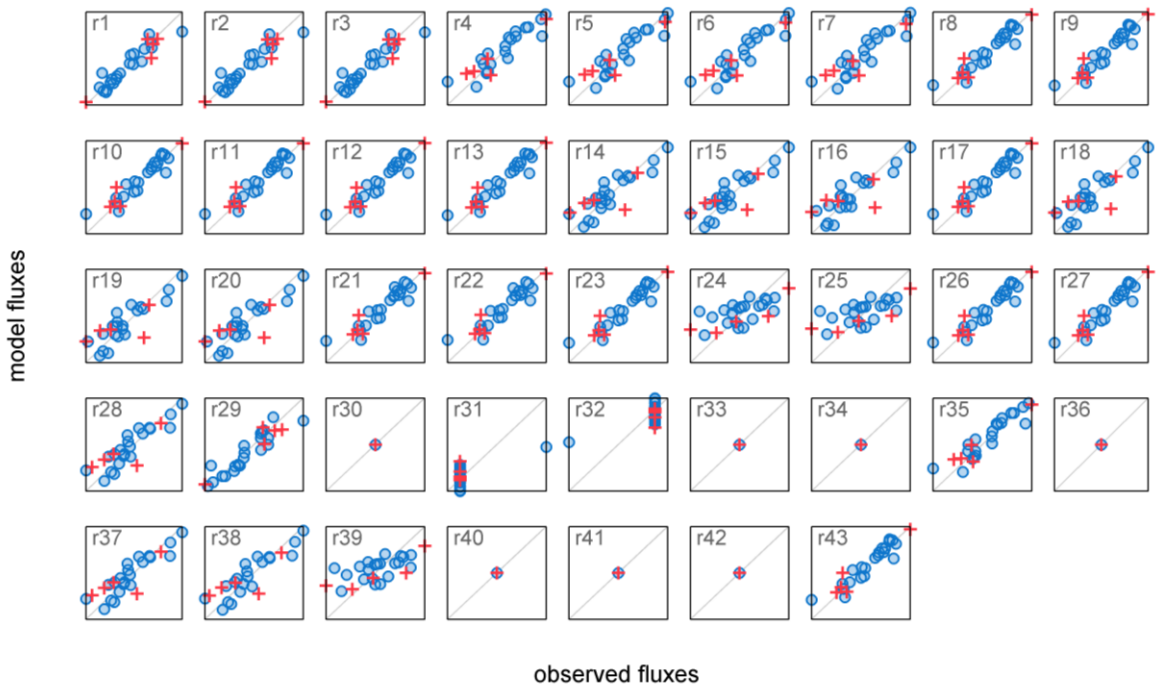


Figure 4.6 – PLS regression results. Predictive power of final PLS model with one latent variable. Correlation between observed and predicted central carbon fluxes for (o) training and (+) validation data sets.

4.4 Conclusions

Current methods of culture media optimization are based on empirical models that relate culture media composition parameters with target phenotypic response variables. In this paper, we investigated the use of PLP as a method to extract meaningful metabolic knowledge from culture media screening data. The key advantage of PLP is the ability to deconvolute contributions of different metabolic pathways to the observed metabolic fluxes and to study the impact of culture media composition at the level of regulation of metabolic pathways. When applied to a *P. pastoris* X-33 strain, PLP consistently identified 3 EFMs that explain most of the variance in observed fluxes: growth (65.6%), maintenance (27.5%) and by-product formation (5.8%). Furthermore, the results show that iron and manganese at concentrations in the range 5.7-56.8 and 0.4-4.2 mg/L, respectively, inhibit overall metabolic activity. The 1:2 dilution of BSM has no clear effect on growth but increases energy expenditures for cellular maintenance. Variations of biotin and remaining trace salts have negligible effect on *P. pastoris* metabolism. A remarkable result of this study is that trace elements components showed a higher influence on metabolism than the dilution of BSM. In the direct comparison with PLS, we have concluded that PLS generates more parsimonious regression models (with less latent variables), while PLP generates robust metabolic meaningful models less prone to data inconsistencies such as data outliers.

References

1. Mandenius C-F, Brundin A (2008) Bioprocess optimization using design-of-experiments methodology. Biotechnol Prog 24:1191–203
2. Deshpande RR, Wittmann C, Heinzle E (2004) Microplates with integrated oxygen sensing for medium optimization in animal cell culture. Cytotechnology 46:1–8
3. Chang KH, Zandstra PW (2004) Quantitative screening of embryonic stem cell differentiation: Endoderm formation as a model. Biotechnol Bioeng 88:287–298
4. Ghosalkar A, Sahai V, Srivastava A (2008) Optimization of chemically defined medium for recombinant *Pichia pastoris* for biomass production. Bioresour Technol 99:7906–10
5. Montgomery DC (2005) Design and analysis of experiments, 6th ed. John Wiley & Sons Inc
6. De Figueiredo LF, Schuster S, Kaleta C, Fell DA (2008) Can sugars be produced from fatty acids? A test case for pathway analysis tools. Bioinformatics 24:2615–21
7. Papin JA, Stelling J, Price ND, et al. (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol 22:400–405
8. Trinh CT, Wlaschin A, Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81:813–826
9. Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18:326–332
10. Zanghellini J (2012) Refined genome scale model of *Pichia pastoris*. Abstr. book, Pichia 2012 Conf. Feb 29 – Mar 31, Alpbach, Austria

11. Isidro IA, Ferreira AR, Clemente JJ, et al. (2015) Hybrid metabolic flux analysis and recombinant protein prediction in *Pichia pastoris* X-33 cultures expressing a single-chain antibody fragment. (manuscript submitted for publication)
12. Teixeira AP, Dias JML, Carinhas N, et al. (2011) Cell functional enviromics: Unravelling the function of environmental factors. BMC Syst Biol 5:92
13. Ferreira A, Dias J, Teixeira A, et al. (2011) Projection to latent pathways (PLP): a constrained projection to latent variables (PLS) method for elementary flux modes discrimination. BMC Syst Biol 5:181
14. Invitrogen Life Technologies (2000) *Pichia* Fermentation Process Guidelines.
15. Dragosits MM (2009) The Interrelation of Environmental Factors and Heterologous Protein Production in *Pichia pastoris* - A Systems Biology Approach. Universität für Bodenkultur Wien
16. Terzer M, Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24:2229–35
17. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185:1–17
18. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130

Additional files

Additional file 4.1 – Culture media screening data

Envirome and fluxome for all 26 shake flask experiments

Additional file 4.2 – *Pichia pastoris* elementary modes network

Description of the 158 elementary flux modes calculated for *Pichia pastoris* main carbon pathways. Includes original stoichiometric network with the description of reactions and metabolites.

Chapter 5
Cell functional enviromics of *Pichia pastoris* cells

Abstract

Cell functional enviromics is a novel technique for cellular function reconstruction through the collection and analysis of dynamic envirome data rather than the typical transcriptomic data used in functional genomics. The key principle is the collection of detailed quantitative data of how cells change their environment and then inference of the biological mechanisms that produced such footprints using appropriate systems analysis tools. In this paper we apply cell functional enviromics to *Pichia pastoris* cells constitutively expressing a single-chain variable fragment (scFv) antibody. Two 50 liter cultivations were performed in fed-batch mode with and without amino acids supplementation. Exometabolite profiling was performed by ^{1}H-NMR, which allowed, together with other standard analytic techniques, the acquisition of data for 21 environmental factors over time. Acquired data was subject to systems analysis using the projection to latent pathways method to identify key metabolic pathways regulated by key environmental factors. We have identified 5 major metabolic pathway groups that are frequently activated by the environment. The resulting *Pichia pastoris* functional enviromics map may serve as template for future optimization of media composition and feeding strategies for *Pichia pastoris*.

Keywords

Pichia pastoris X-33 · ^{1}H-NMR metabolomics · cell functional enviromics · projection to latent pathways

5.1 Introduction

Microorganisms constantly monitor their surroundings for the availability of nutrients and other chemicals, using both external and internal sensors to respond dynamically to environmental changes. Integration of the external environment with metabolism occurs through the intake of

compounds from the environment and results, for example, in a transcriptional response or an allosteric interaction with an enzyme. Functional genomics is an expanding discipline that aims at unravelling gene functions and gene-gene interactions and how these set phenotypic traits. However, the effect of genetic variants on phenotypic traits is highly dependent on environmental and physiological conditions, but such gene-environment interactions are still poorly understood [1]. Functional genomics is a landmark because it is holistic in the sense that all genes are analyzed simultaneously, which is a critical factor for success given the massive interactions of genes in a genetic network. Equivalent holistic methods for the analysis of the interaction between environment and biological function are needed. Functional enviromics is the environmental analog to functional genomics. It may be defined as the discipline that studies the function of environmental variables using holistic principles. The concept of functional enviromics has been first set forth as a counterpart to functional genomics in tackling mental disorders such as schizophrenia [2]. Only very recently functional enviromics has been addressed in the context of cell physiology [3]. A basic premise in functional enviromics is that while a cell's genome uniquely sets the phenotypic space of a cell, particular trajectories within it are primarily driven by the environment (Figure 5.1).

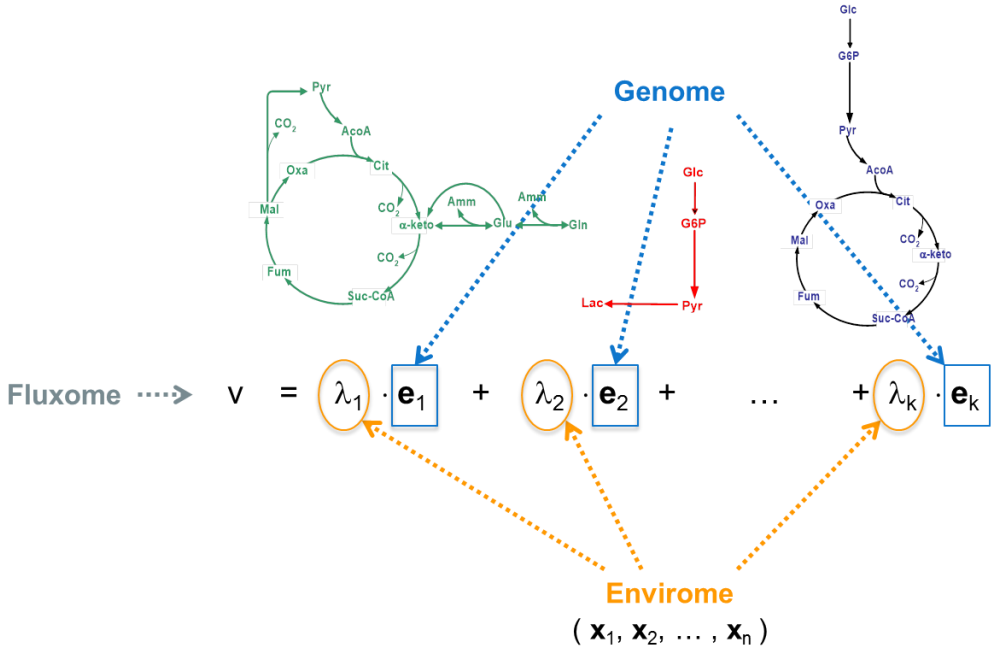


Figure 5.1 – Functional genomics versus functional enviromics. The genome sets all operation modes of a cell. The operation modes are represented here by the $\mathbf{e}_i$ vectors which are elementary flux modes. Genes code for all the enzymes and regulatory mechanisms for a given mode to operate. The activation of operational modes is however controlled by the environment by setting the relative weight, λ_i, with which elementary modes contribute to flux phenotype.

The key steps for functional enviromics analysis are: i) setting the universe of cellular functions, $\mathbf{e}_i$, and envirome components, $\mathbf{x}_i$, ii) collecting informative envirome data over time, and iii) systems level analysis of dynamic envirome data to find relationships between environmental variables and cellular functions. In step i) the structure of the biological system may be set through genome-scale

network reconstruction. The identity of exogenous factors is primarily “engraved” in the DNA sequence of cells. The annotated full genomic sequence is becoming available for a growing number of organisms with industrial interest, paving the way for genome-wide reconstruction of the cellular environment. Examples are *Escherichia coli*, *Sacharomyces cerevisiae* and *Pichia pastoris*, but also animal cells such as Chinese hamster ovary (CHO) cells [4, 5]. Genes may be associated with metabolic enzymes, membrane transporters, signal transduction or regulatory control. Combined with basic biochemical information currently available in several databases (e.g. KEGG [6] and BioCyc [7] databases), it is possible to reconstruct the majority of the metabolic reaction network and also many of the regulatory mechanisms. From this knowledge it is possible to reconstruct the latent exometabolome, and also exoproteome, with a critical function in regulatory processes. Borenstein et al. [8] proposed a graph-theoretical approach to define these exogenously acquired compounds – they called it the seed set of an organism - and have identified their repertoire across the tree of life. This is one of the most comprehensive studies so far that links organisms' metabolic circuitry with their environment.

Quantitative analysis of cellular environment (step ii) is nontrivial and generally requires the combination of different measurement techniques. Exometabolomics and exoproteomics can be used to quantify metabolites and proteins [9], but also the inorganic footprint should be measured using for instance ICP-MS techniques.

For step iii) we have previously proposed the projection to latent pathways (PLP) algorithm [3, 10]. PLP was designed to maximize the covariance between environmental state (envirome data sets) and observed phenotypic trait (rate of change of envirome variables) under the constraint of known genes translated into a plausible set of cellular functions. More specifically, PLP finds a minimal set of elementary modes [11] based on two criteria: a) variance of measured phenotypic trait explained by each elementary mode and b) degree of correlation of each elementary mode with the environmental state. By maximizing these two criteria, the algorithm delivers a ranking of elementary modes in decreasing order of explained variance in the measured flux data and a functional enviromics map (FEM) representing the strength of up- or down-correlation of EFMs and environmental factors [3]. Mathematical details of the method can be found in [10].

PLP has been used before in the context of *Pichia pastoris* cultures for the analysis of culture media composition impact on *Pichia* physiology [12] (Chapter 4). In this paper we augment significantly this analysis by measuring many more environmental factors using ^{1}H-NMR in 50 L reactor experiments. The so collected envirome data enabled the creation of a functional enviromics map that may serve in future studies to design specialty culture media and advanced feeding strategies for *Pichia pastoris* cultures.

5.2 Materials and methods

5.2.1 Strain and cultivation conditions

Two pilot-scale cultivations of a *Pichia pastoris* X-33 strain constitutively expressing a single-chain variable fragment (scFv) with a GAP promoter were performed with and without amino acid supplements in the medium. Basal media (BM) composition was as follows (per liter): 26.7 mL H_3PO_4 85%, 0.65 g $CaSO_4 \cdot 2H_2O$, 13.29 g K_2SO_4, 18.33 g $MgSO_4 \cdot 7H_2O$, 4.13 g KOH, and 40.0 g glycerol. For the experiment with amino acid supplementation (BM+AA) the basal media included additionally 2.36 g alanine, 0.82 g arginine, 3.69 g aspartate, 3.01 g glutamate, 2.02 g glutamine, 26.70 g glycine, 0.55 g lysine, and 0.69 g proline per liter. The BM solution was sterilized by filtration with a 0.2 nm pore size, after which the pH was adjusted to 5.0 with 25% NH_4OH. 4.35 mL of trace minerals (TM) solution was added per liter of BM with the following composition: 12.0 g $CuSO_4 \cdot 5H_2O$, 0.16 g NaI, 6.00 g $MnSO_4 \cdot H_2O$, 0.40 g $Na_2MoO_4 \cdot 2H_2O$, 0.001 g H_3BO_3, 0.25 g $CoCl_2$, 40.00 g $ZnCl_2$ 3.25 g $FeSO_4 \cdot 7H_2O$, 0.40 g biotin, and 10.0 mL H_2SO_4 (per liter of TM). The TM solution was also sterilized by filtration with a 0.2 nm pore size.

Cultivations were carried out according to Invitrogen guidelines [13]. The pre-inoculum was incubated in a 250 mL shake flask with 40 mL of culture medium and inoculated with 1 mL cell bank, previously stored at -80 ºC and thawed at room temperature. Incubation proceeded at 30.0 ºC and 150 rpm for 55.5 h (BM+AA) or 63.0 h (BM), until the 600 nm optical density (OD_{600}) reached 16.0 (BM+AA) or 40.4 (BM). The inoculum was prepared by inoculating a 2 L culture flask containing 750 mL of culture medium with 25 mL of the pre-inoculum, which was further incubated at 30.0 ºC and 150 rpm for 66.5 h (BM+AA) or 73.0 h (BM), until OD_{600} was 3.73 (BM+AA) or 3.98 (BM). The fermentations were carried out in a 50 L BioEngineering LP351 bioreactor, where 15 L of culture medium with 0.2 g/L of SAG 471 antifoam were inoculated with the totality of the inoculum. Temperature, pressure and airflow were kept constant at 30 ºC, 100 mbarg and 30 lpm, respectively, and pH was controlled at 5.0 by the addition of a 25% NH_4OH solution. Agitation was initially set to 300 rpm. The bioreaction carried out for 172 h, subdivided into three operation phases with different control strategies:

(i) Batch phase (approximate duration 30 h): pO_2 (dissolved oxygen as % of saturation) gradually decreased until it reached 50%, where it was kept constant by increasing the agitation rate up to the maximum of 1000 rpm;
(ii) Exponential feeding (approximate duration 10 h): when the agitation rate reached 440 rpm, glycerol feeding began according to the formula $Q(g/h) = 65.0\ e^{0.16\ tfeed}$;
(iii) Controlled feeding (approximate duration 130 h): after pO_2 decreased below 5.0%, which is usually quickly followed by a pO_2 spike, a model reference adaptive controller (MRAC) [14] was used to adjust glycerol addition rate in order to keep pO_2 stable at 5.0%.

Three to four samples were taken every day at intervals of approximately 3 hours. Heterologous protein concentration was determined by enzyme-linked immunosorbent assays (ELISA) as described in [15]. Biomass wet cell weight (WCW) concentration was determined by optical density at 600 nm and by centrifugation at 15000 rpm for 20 min followed by weighting without drying. Dry cell weight (DCW) was estimated from wet cell weight by multiplying by a factor of 0.258 previously calibrated for similar culture densities [16]. Concentration in the supernatant of glycerol, amino acids and other exometabolites was determined by ^{1}H-NMR.

5.2.2 Core metabolic network

The core metabolic network model for *Pichia pastoris* used in this work was based on genome-scale networks [17, 18], which were simplified by selecting core metabolic pathways, and by aggregating sequential reactions and restricting to growth on glycerol. The core metabolic network consisted of 98 reactions and 92 metabolites, of which 8 are exchanged with the extracellular medium. It includes glycolysis/gluconeogenesis and pentose phosphate pathways, tricarboxylic acid (TCA) cycle, synthesis of amino acids, interconversion of folate compounds, synthesis of biomass from macromolecular components (nucleotides, lipids, carbohydrates, and proteins), synthesis of a single-chain variable fragment (scFv) and exchange reactions (glycerol, O_2, CO_2, H_2O, H_2SO_4, NH_4, scFv and biomass). The core metabolic network was then extended to include uptake reactions for the 7 amino acids in the medium supplement and detected by ^{1}H-NMR analysis. It was also extended to include synthesis and secretion of 3 other metabolites detected by ^{1}H-NMR (as explained later in the Results section). The final network thus consists of 116 reactions and 99 metabolites of which 18 are exchange reactions/metabolites and is supplied in Additional file 5.1.

5.2.3 Determination and reduction of elementary flux modes

Based on the core metabolic model described above, elementary flux modes (EFM) were calculated using the freely available EFMtool [19]. The obtained EFMs are represented as a large $nr \times nem$ matrix ($\mathbf{KEM}$), with nr the number of reactions in the metabolic network and nem the number of elementary modes. The first reduction step consisted in the elimination of all intracellular reactions resulting in the EFM footprint, represented by the $nexch \times nem$ matrix, $\mathbf{KEM_{footprint}}$, where rows correspond to exchange metabolites and columns to elementary flux modes. The $\mathbf{KEM_{footprint}}$ matrix was determined by Eq. (2.6),

$$\mathbf{KEM_{footprint}} = \mathbf{S_{exch}} \cdot \mathbf{KEM} \qquad \textbf{(5.1)}$$

with $\mathbf{S_{exch}}$ the stoichiometric matrix of exchange metabolites.

The second step of reduction was performed by a tailored clustering method, where the $\mathbf{KEM}_{\text{footprint}}$ is further reduced into groups with the same metabolite exchange patterns. The grouping criterion was the input/output pattern of each elementary mode. This means that all EFMs belonging to one group must have the same exchange metabolites and the same exchange directionality (in or out of the cell). The centroids were then calculated and taken as representative of each group. We call this method 'EFM pattern-based clustering'.

5.2.4 Projection to latent pathways

Projection to latent pathways (PLP) was used to identify EFM activation by environmental factors [3, 10, 12]. PLP is a constrained version of the widely used partial least squares (PLS) regression, where a predictor envirome matrix ($\mathbf{X}$) is related to response fluxome matrix ($\mathbf{R}$) through a multilinear regression under the constraint of an existing set of EFMs. The key difference is that PLP constrains predictor weights to structures that represent metabolic pathways, i.e. the elementary flux modes (EFM). The PLP regression model is described in Eqs. (5.2-5.3). Fluxes in $\mathbf{R}$ are the product of the subset of ranked elementary modes, $\mathbf{REM}$, by how active those elementary modes were in each observation (represented by the EFM-weights, $\mathbf{\Lambda}$). The EFM-weights are estimated from the envirome variables, $\mathbf{X}$, with a multilinear regression model represented in Eq. (5.3) where $\mathbf{C}$ is the regression coefficients matrix.

$$\mathbf{R} = \mathbf{\Lambda} \cdot \mathbf{REM}^{\mathrm{T}} \tag{5.2}$$

$$\widehat{\mathbf{\Lambda}} = \mathbf{X} \cdot \mathbf{C} \tag{5.3}$$

PLP (as PLS) decomposes predictor and response matrices into scores and weights in such a way that covariance between the scores is maximized. This covariance-guided decomposition enables the analysis of noisy, highly collinear data, with more variables than observations.

To mitigate the effects of different orders of magnitude in the variables, $\mathbf{X}$ was normalized by unit variance and mean-centering, while $\mathbf{R}$ was normalized by dividing by the maximum absolute value (so that for each reaction the maximum absolute value is 1). For consistency, the same normalization applied to $\mathbf{R}$ columns is also applied to $\mathbf{KEM}$ rows. The reason why $\mathbf{R}$ is not mean-centered in PLP is to maintain the direction of the reactions in the EFM definition, which is necessary for physical interpretation. Before normalization, experimental error in measured concentrations ($\mathbf{X}$) was filtered with smoothing splines and fluxes ($\mathbf{R}$) were estimated from the derivative of those splines.

The optimal number of components for the PLP regression model was determined by leaving out 6 data points for validation and fitting the model to the remaining 32 data points. Because fermentation data are generally monotonous time series and because differences between batches

are expected, low-discrepancy sequences (more specifically, Halton sequences with a random offset) were used to obtain a homogeneous distribution of the validation points.

A bootstrapping analysis was used to assess the consistency of EFM identification and respective coefficients in the bootstrapping models. For that purpose, 500 bootstrapping runs were performed using random sampling with replacement to select a sample of size N from the original N observations. The bootstrapping analysis allowed the calculation of the frequency of selection of each EFM and the average regression coefficients ($\bar{\mathbf{C}}$).

5.3 Results and discussion

5.3.1 Reactor experiments

The two pilot-scale fermentations were followed for approximately 175 h, with 19 samples taken over time in each. Maximum biomass concentration for experiments with and without amino acid supplementation was 182.25 and 193.33 g-DCW/L, respectively. The maximum heterologous protein titer was 256.49 and 241.84 mg/L, corresponding to a maximum protein yield of 1.48 and 1.28 mg-protein/g-DCW, respectively. Figure 5.2 shows the concentration profiles for biomass, carbon source (glycerol) and heterologous protein (scFv) throughout the cultures. The concentration profiles and final concentration values were similar for both experiments.

Several exometabolites were quantified by ^{1}H-NMR analysis of the supernatant. This high-throughput technique was able to detect and quantify glycerol, 7 of the 8 supplemented amino acids, 5 additional amino acids and 8 other metabolites or by-products.

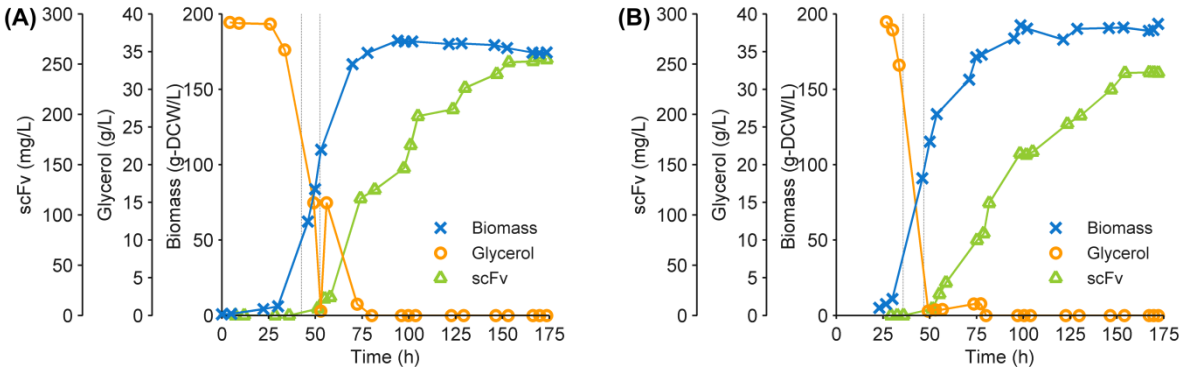


Figure 5.2 – Culture profiles over time. Biomass, glycerol and heterologous protein (scFv) concentrations over 175 h of culture with **(A)** and without **(B)** amino acid supplementation. Gray lines indicate the three fed-batch phases.

Figure 5.3A-B shows the consumption of amino acid supplements. Five of the supplemented amino acids were clearly identified and quantified (alanine, asparagine, glutamate, arginine and proline),

lysine concentration was below the detection limit and was considered zero. For glycine some measurements were affected by partial signal overlaps, and glutamine could not be measured at all due to complete signal overlaps.

Five other amino acids (isoleucine, phenylalanine, threonine, tyrosine and valine) were detected in the supernatant showing an accumulative profile. Other exometabolites detected were, in decreasing order of concentration: allantoin, glycerophosphocholine (GPC), uridine, inosine, isovalerate, acetate, isobutyrate, and 2-oxoisovalerate. Allantoin and GPC were the most abundant by-products, with all others having between 3 to 20% (w/w) of allantoin concentration. Figure 5.3C-D shows the profiles for the main by-products (allantoin and GPC) compared to the typical fermentative by-product acetate. It should be noted that ^{1}H-NMR allowed the detection of unexpected by-products that can be present in significant concentrations in the environment of *P. pastoris* cells. Detailed sampling data, including on-line fermentation measurements, at-line analytics and ^{1}H-NMR analytics, is supplied as Additional file 5.2.

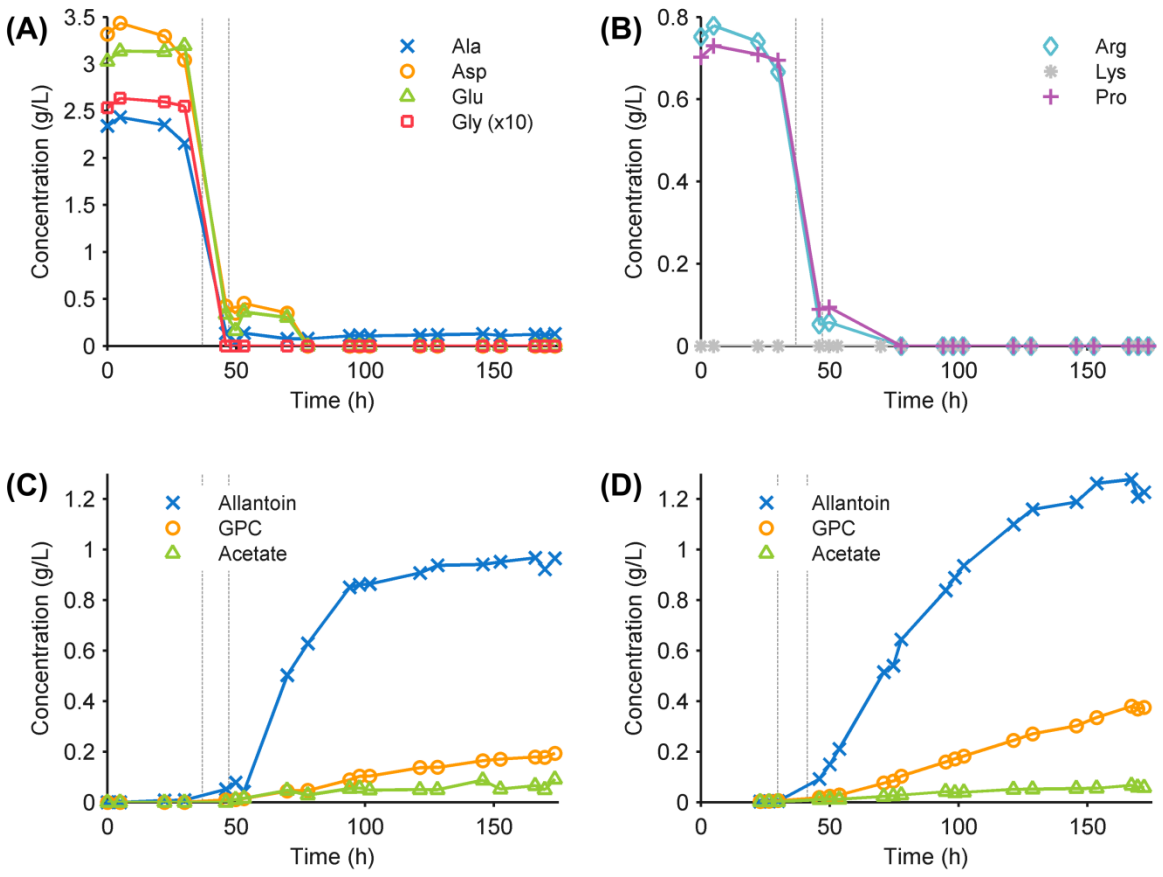


Figure 5.3 – Profiles for exometabolites over time as determined by ^{1}H-NMR. (A-B) Concentrations of supplemented amino acids: Ala – alanine, Asp – aspartate, Glu – glutamate, Gly – glycine (concentration was 10x the scale shown), Arg – arginine, Lys – lysine, and Pro – proline. **(C-D)** Concentrations of allantoin and GPC, the by-products with higher concentration, compared with acetate as a common by-product, for cultures with **(C)** and without **(D)** amino acid supplementation. Gray lines indicate the three fed-batch phases.

5.3.2 Elementary flux modes

The results of ^{1}H-NMR were used to update the core metabolic network described previously. This was done by (i) adding uptake reactions for 7 amino acids supplements (alanine, arginine aspartate, glutamate, glycine, lysine and proline), and (ii) including production and excretion of the two most abundant by-products (allantoin and GPC), as well as of the common fermentative by-product acetate. The core network extended with ^{1}H-NMR identified metabolites consists of 99 metabolites and 116 reactions, from which 18 are exchange. Elementary flux modes calculation based on this network lead to 83 116 EFMs, of which 66 080 are for biosynthesis, 24 720 for heterologous scFv production and 7 700 for simultaneous synthesis of biomass and scFv.

The EFM footprint was calculated as shown in Eq. (2.6), following the elimination of unused fluxes for H_2O, H_2SO_4, and NH_3. The resulting EFM footprint contains the same elementary modes, but is limited to 15 exchange reactions (see comparison in Table 5.1).

Table 5.1 – EFM footprinting and clustering. Descriptive properties (including number of reactions, metabolites and EFMs producing biomass and recombinant product) for three EFM representations of the same metabolic network: original EFM, EFM footprint and clustered EFM footprint.

	EFM	EFM footprint *	Clustered EFM footprint *
Number of reactions			
Total	116	15	15
Intracellular	98	0	0
Exchange	18	15	15
Number of metabolites			
Total	99	15	15
Intracellular	81	0	0
Exchange	18	15	15
Number of EFMs			
Total	83 116	83 116	967
Biomass	66 080	66 080	707
Product	24 720	24 720	474
Both	7 700	7 700	218

* for the footprint the number of reactions is always coincident with the number of metabolites, since each row on the $\mathbf{KEM}_{\mathbf{footprint}}$ matrix is an exchange metabolite (Eq. 5.1)

Even with a metabolic model limited to one carbon source and a few selected by-products, the number of EFMs can easily become computationally expensive due to combinatorial explosion. To circumvent this limitation, the universe of possible elementary modes was reduced by EFM pattern-based clustering (if the inputs and outputs were the same, the elementary modes were put in the same cluster). The number of clusters obtained was 967, corresponding to only 1.16% of the original number of EFMs (Table 5.1), yet the captured variance was 99.5%. The number of EFMs producing biomass, scFv or both was reduced respectively to 1.07, 1.92, and 2.83% of the original

number. The average carbon accuracy in the clustered footprint was 99.1% which shows cluster centroids can be used to represent EFMs clusters without significant loss of accuracy. The reduced EFM footprint is available as Additional file 5.3.

Principal component analysis (PCA) was used to visualize the efficiency of the multi-dimensional clustering of 83 116 elementary flux modes across 15 reactions. For that end, PCA was applied to the EFM footprint (pre-clustering data) with the 15 reactions as variables. This transforms the data into principal components in such a way that each component captures progressively less variance. The first two components cumulatively captured 40.3% of the data variance. The same transformation was then applied to the clustered data so they could be visualized together in a biplot (Figure 5.4A). The biplot shows a good overall consistency between the distribution of the original EFM footprint and the clustered footprint for the first two principal components. Moreover, this consistency is evident even though the number of EFMs in each cluster varied from 1 to 400 (Figure 5.4B). Overall, these results support the reduced universe of 967 EFMs as a good representation of the original 83 116. The final clustered EFM footprint was used as input to the PLP regression that is discussed in the next section.

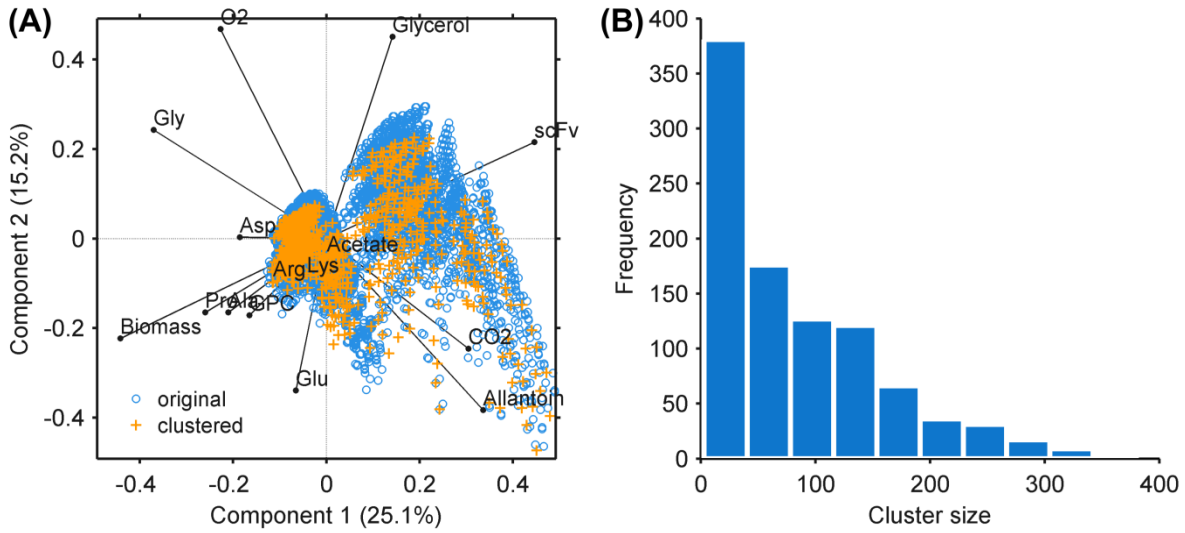


Figure 5.4 – Clustering of the EFM footprint. (A) Clustering PCA biplot for first two principal components: points show EFM-scores before (o) and after (+) clustering, while arrows show the loadings or weights of the original variables on the principal components. **(B)** Histogram showing distribution of cluster size (number of EFMs in the cluster).

5.3.3 Projection to latent pathways

Projection to latent pathways identifies a minimal set of elementary modes that best explains the observed fluxes (response variables) while maximizing the correlation between observed fluxes and envirome factors (predictor variables). PLP regression thus requires three pieces of information (envirome, fluxome and EFM candidates):

(i) Envirome ($\mathbf{X}$): consists of a 38 × 21 matrix with 38 observations (19 for each experiment) and 21 measured variables ('envirome' in Additional file 5.2). The measured variables selected as predictors included on-line bioreaction measurements (dissolved oxygen and exhaust gas composition) and concentrations for biomass, scFv, glycerol, supplemented amino acids and other exometabolites;

(ii) Fluxome ($\mathbf{R}$): 38 × 13 matrix with measured fluxes for biosynthesis, scFv production, glycerol consumption, consumption of supplemented amino acids and accumulation of representative by-products, namely allantoin, GPC and acetate ('fluxome' in Additional file 5.2);

(iii) Universe of possible elementary modes ($\mathbf{KEM}$): the 15 × 967 clustered EFM footprint ($\mathbf{KEM}_{\mathbf{footprint}}$) was used as the matrix of elementary modes candidates (Additional file 5.3). To note that the columns for O_2 intake and CO_2 consumption were not considered for PLP.

The PLP model was fitted to the training data set and used to predict fluxes for the validation data set. The validation results (Figure 5.5A) show the optimal number of components for the final model is 5, with 75.8% explained flux variance for the training and 75.2% for the validation set. Adding more components leads to model over-fitting as shown by the decreasing explained variance for validation.

Figure 5.5B shows the fit of model-predicted and observed fluxes (with all variables normalized) for the final 5-component model. The prediction shows a high correlation, with some outliers. The visible horizontal tendency is because acetate fails to be predicted, most likely due to very low measured concentrations (2.09-71.24 mg/L), significantly lower than other by-products and lower than envirome factor concentrations which are mostly over 1 g/L. Outliers close to the vertical axis correspond to four points with lower product synthesis rate and the points for the peak amino acid supplement consumption, in the transition between the batch and fed-batch phases.

The final 5-component model includes, in descending order of explained variance, EFMs 704, 836, 158, 600 and 167. A bootstrapping analysis was performed to check the consistency of EFM selection in PLP models with 5 components. Figure 5.5C shows the frequency of selection of each EFM across multiple bootstrap runs and highlights the elementary modes selected for the previous 5-component model. The elementary mode more often selected is EFM 167, with a 51.0% relative frequency. Four other elementary modes are selected between 20 to 50% of the runs, including EFM 836 (42.0%) and EFM 600 (31.6%). The total number of different elementary modes selected by bootstrapping was 195, and the vast majority has a selection frequency under 20%. The low consistency in the selection of elementary modes suggests a high degree of redundancy between different elementary mode combinations.

(A)

#LV	V_X (%)	V_Y (%)	R^2	p-value	R^2_{adj}
Training					
1	27.0	14.8	0.128	< 0.001	0.104
2	68.8	39.1	0.361	< 0.001	0.325
3	87.5	56.8	0.538	< 0.001	0.498
4	95.7	66.5	0.647	< 0.001	0.604
5	96.9	75.8	0.743	< 0.001	0.703
6	98.0	79.0	0.780	< 0.001	0.738
7	99.3	81.9	0.808	< 0.001	0.764
Validation					
1	29.4	16.6	0.148	0.001	0.124
2	73.0	43.7	0.421	< 0.001	0.388
3	94.7	48.8	0.486	< 0.001	0.440
4	97.1	58.8	0.620	< 0.001	0.574
5	98.3	75.2	0.754	< 0.001	0.716
6	98.6	71.0	0.722	< 0.001	0.668
7	99.6	71.0	0.721	< 0.001	0.656

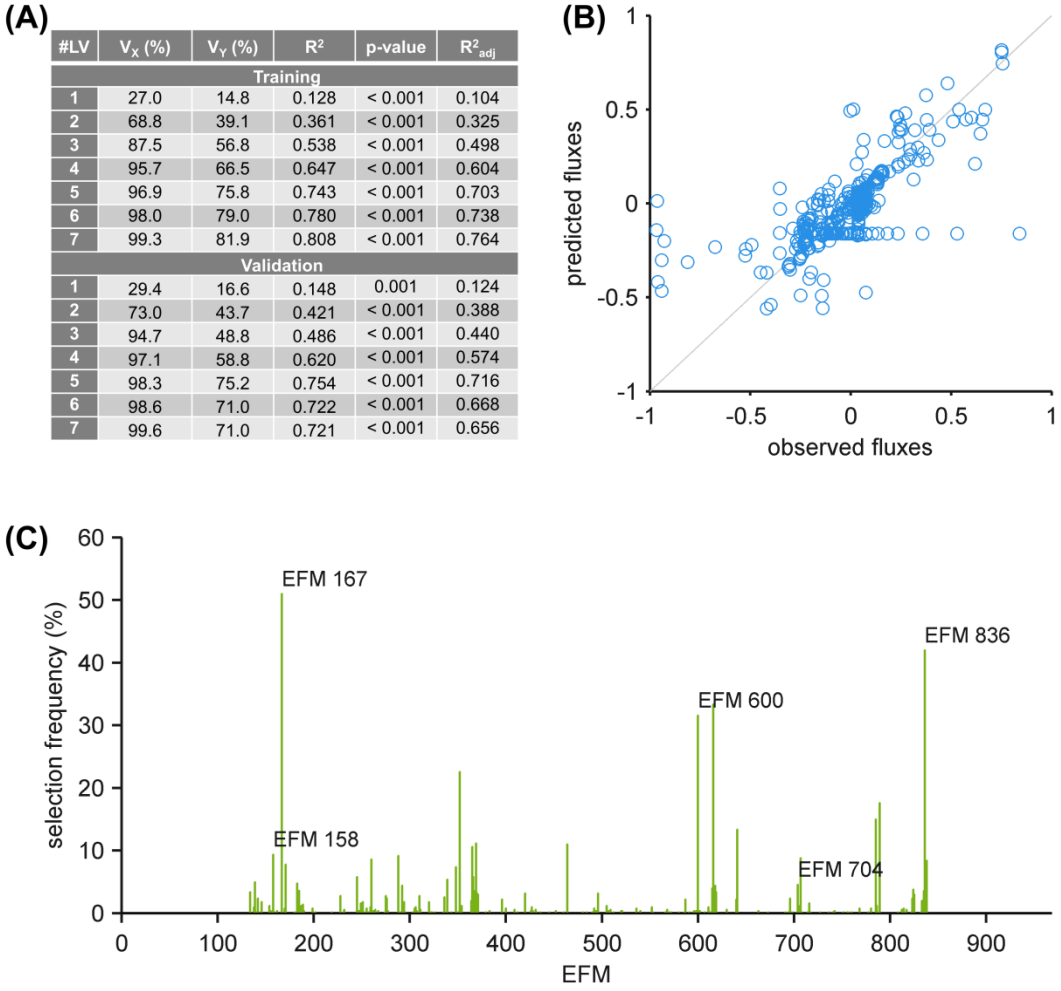


Figure 5.5 – PLP regression and bootstrapping. (A) Cumulative goodness-of-fit measures for the PLP regression model using up to 7 latent variables: V_X and V_Y are the explained variance in the predictor (envirome factors) and target variables (fluxes). R^2_{adj} is the adjusted R^2 considering the number of predictors. **(B)** Correlation between model-predicted and observed fluxes (normalized) for the final 5-component model with EFMs 704, 836, 158, 600 and 167. **(C)** Frequency of selection for all possible EFMs in bootstrapping with 5 latent variables, with EFMs from previous model highlighted in the plot.

5.3.4 Analysis of selected EFMs

The elementary modes selected in the 5-component model are shown in Table 5.2. All selected elementary modes lead to biomass production and glycerol consumption, with different by-products and additional substrate combinations. The first two selected elementary modes (EFMs 704 and 836) lead to simultaneous production of scFv and allantoin, while the third (EFM 158) is the only selected elementary mode with GPC production. The fourth and fifth elementary modes represent biomass synthesis without by-product formation, with more or less amino acids as substrates. The major difference between the first and second elementary modes (producing allantoin and scFv), is that EFM 704 uses amino acids as substrates while EFM 836 does not. Similarly, for biosynthesis without by-product formation, EFM 600 uses only glycine, while EFM 167 uses five different amino acids.

Table 5.2 – Description of the 5 selected EFMs. Exchange stoichiometry of EFMs selected for the final 5-component model in decreasing order of explained variance (stoichiometry for all EFMs is supplied as Additional file 5.3). Units are C-mol for biomass and scFv, and mol for all other metabolites.

EFM	Stoichiometry of exchange reactions
704	0.358958 Glycerol + 0.0167094 O_2 + 0.0116741 Ala + 0.00323236 Pro → 0.140939 CO_2 + 1 Biomass + 0.000290963 scFV + 0.00058698 Allantoin
836	0.377423 Glycerol + 0.0167411 O_2 → 0.144292 CO_2 + 1 Biomass + 0.00106197 scFV + 0.000608137 Allantoin
158	0.293845 Glycerol + 0.0158289 O_2 + 0.00850275 Ala + 0.0059306 Arg + 0.0207707 Asp + 0.0215831 Glu → 0.144481 CO_2 + 1 Biomass + 0.00058288 GPC
600	0.398093 Glycerol + 0.0158289 O_2 + 0.024955 Gly → 0.259707 CO_2 + 1 Biomass
167	0.327046 Glycerol + 0.0158289 O_2 + 0.0108294 Ala + 0.0059306 Arg + 0.021651 Asp + 0.0158231 Gly + 0.00310997 Pro → 0.198527 CO_2 + 1 Biomass

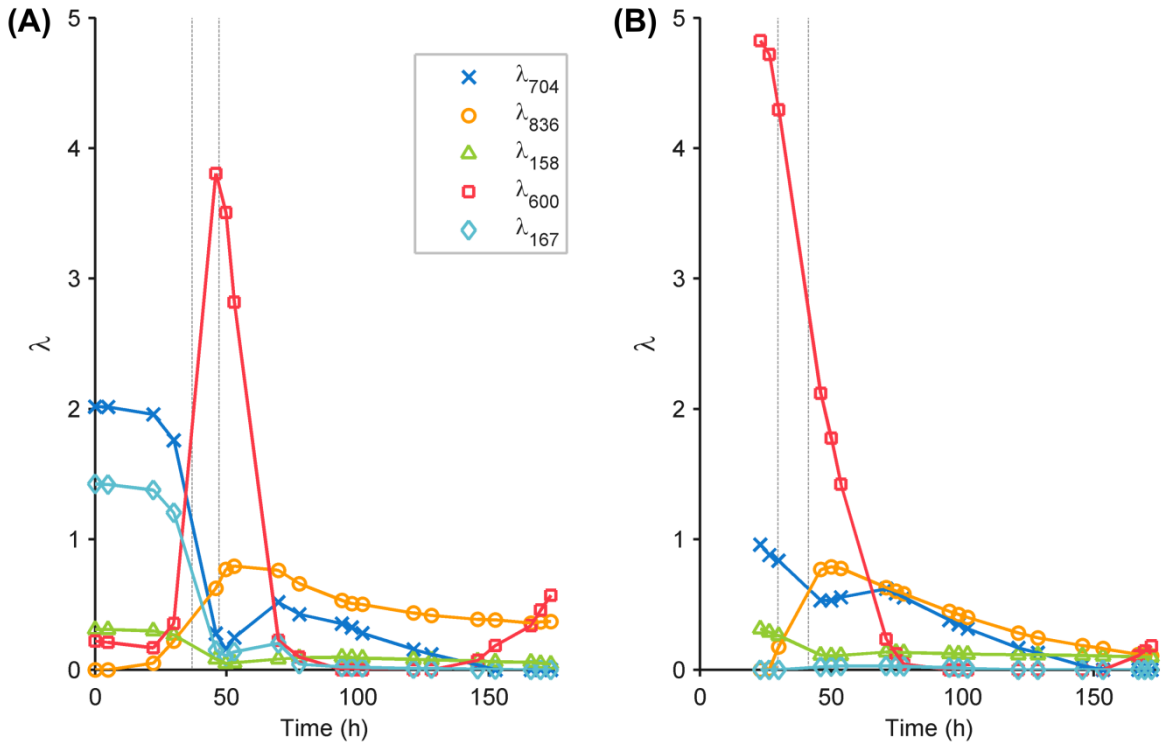


Figure 5.6 – EFM activation profiles over time. EFM-weights, λ, over time for cultures with **(A)** and without **(B)** amino acid supplementation. Gray lines indicate the three fed-batch phases.

An interesting feature of PLP is the ability to plot EFM-weights, $\boldsymbol{\Lambda}$ in Eq. (5.2), over time, thus giving a visual representation of how active each elementary mode (or metabolic function) is in different stages of the fermentation. Figure 5.6 shows different EFM-weights profiles for the experiments with and without amino acid supplementation. For the experiment with amino acid supplements (Figure 5.6A), EFMs 704 and 167 (both consuming multiple amino acids) are the more active at the beginning of the fermentation. As the culture reaches exponential growth, EFM 600 (biosynthesis

from glycerol and glycine as only amino acid) becomes more active, reaching a peak half-way through the exponential phase as amino acids are depleted. In comparison, for the experiment without amino acid supplements (Figure 5.6B), EFM 600 is the most active through both lag and exponential phases. EFM 836 which leads to scFv synthesis without intake of amino acids has a similar profile in both experiments: it reaches its peak half-way through the exponential phase (~50 h) and becomes the predominant elementary mode when growth slows down into stationary phase (~75 h).

5.3.5 Assessment of envirome-EFM relationship

A metric commonly used with PLS regression to estimate the importance of each predictor variable in the projection model is the variable importance in projection (VIP) score. VIP was calculated for the PLP model as the sum of regression coefficients for each environmental factor, weighted by the explained variance of each component. Figure 5.7 shows overall VIP score for each predictor/envirome variable as well as its breakdown across the 5 model components. The envirome factors with higher overall impact in the final model are glycerol and biomass concentrations, and exhaust gas composition (%O_2 and %CO_2). Another observation is that variable importance varies a lot across components, i.e., different envirome factors are more correlated with certain elementary modes than others. This is to be expected from a metabolic point of view and is clear, for example, for EFM 167 (biosynthesis with amino acids uptake) which depends mostly on amino acid availability.

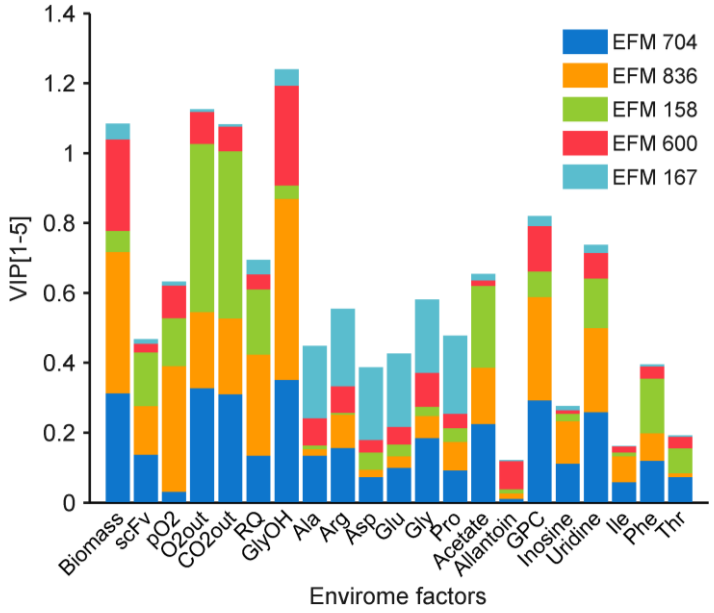


Figure 5.7 – Variable importance in projection (VIP). The VIP score shows the contribution of individual envirome factors to the final 5-component model. The total contributions are broken down by model components, each corresponding to a selected EFM.

Another way to study the role of envirome factors is to look at the average regression coefficients for all EFMs selected during bootstrapping. These can be represented in the form of a functional enviromics map or FEM (Figure 5.8), which shows how each envirome factor up- or down-correlates with selected elementary modes. In the FEM, the heat map represents the regression coefficients, normalized by the norm of each elementary mode. To improve readability, only EFMs that were selected in more than one bootstrap run were included in the figure (121 out of 195). The hierarchical clustering of both envirome factors and elementary flux modes undercovers patterns in the data and allows the subdivision of the elementary modes into groups. Each of the five identified groups has a corresponding EFM in the 5-component model (shown in the bar on the right-hand side of Figure 5.8). The emergence of these groups supports the previous observation that the high degree of redundancy in the universe of possible elementary modes leads to alternative models with different sequences of selected EFMs but similar overall behavior.

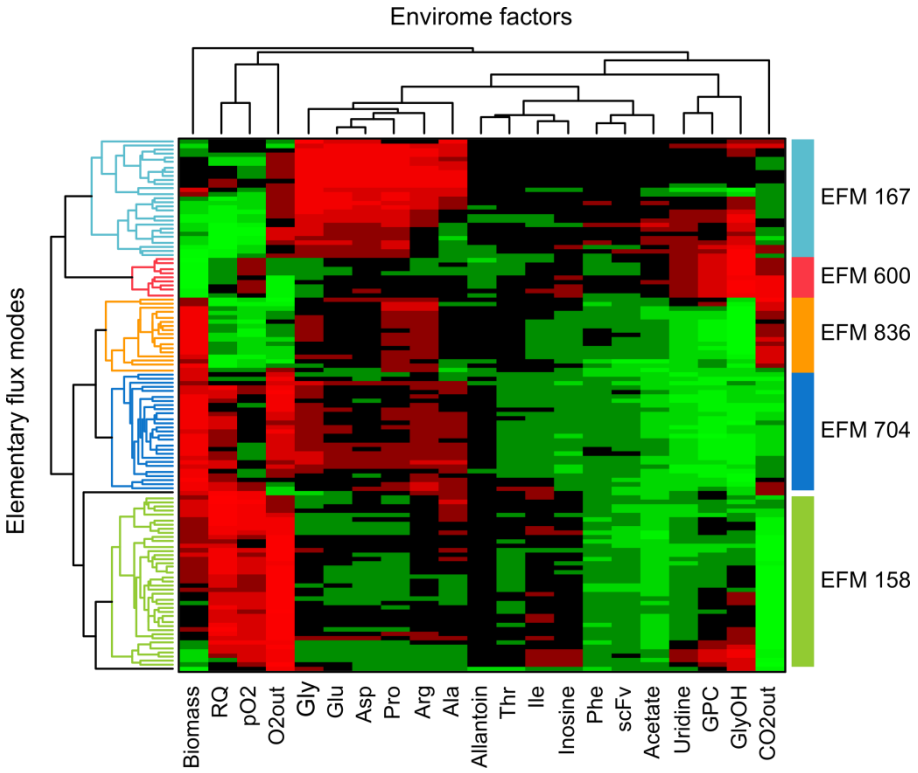


Figure 5.8 – Functional enviromics map (FEM). The heat map shows how each envirome factor up (red) or down (green) correlates with selected elementary modes. The bar to the right shows different groups of EFMs, identified by the corresponding EFM on the 5-component model.

5.4 Conclusions

In this paper we have outlined how cell functional enviromics can be applied to analyze envirome patterns in *Pichia pastoris* cultures. Two 50 liter cultivations were performed in fed-batch mode with

and without amino acids supplementation. The experiments were followed over time by ^{1}H-NMR exometabolite profiling. We have observed that final biomass concentration and scFv titer were not significantly affected by amino acids supplementation, however, some differences in metabolites uptake and excretion are observed in the two experiments. We have detected substantial accumulation of allantoin and GPC over time, as well as the accumulation of multiple amino acids in cultivations with or without amino acids supplementation.

The measured metabolite profiles were analyzed by projection to latent pathways (PLP) using a core metabolic network with 99 metabolites, 116 metabolic reactions and 83 116 elementary flux modes. We have developed a reduction method called EFM pattern-based clustering, based on the conservation of the exchange reactions pattern, which resulted in 967 clusters that explain 99.5% of the variance on the original EFM matrix. Among these EFM candidates, PLP identified 5 groups of EFMs with high correlation with envirome measurements. All identified elementary modes produce biomass from glycerol, with varying by-products and additional substrate combinations. The first two selected elementary modes lead to simultaneous production of scFv and allantoin, which seem to be correlated in *Pichia pastoris* cultures. Noticeable differences between the selected EFMs are the amino acids uptake patterns akin to the amino acids supplementation in the experiments.

The analysis of the envirome-EFM relationship has shown that the envirome factors with higher overall impact in the final model are glycerol and biomass concentrations, and exhaust gas composition (%O_2 and %CO_2). Another observation is that different envirome factors are more correlated with certain elementary modes than others. This resulted in the functional enviromics map that depicts the relationship between individual EFMs and envirome components. This map may serve as template to optimize media composition and feeding strategies for *Pichia pastoris* in the future.

References

1. Smith EN, Kruglyak L (2008) Gene-environment interaction in yeast gene expression. PLoS Biol 6:e83
2. Van Os J, Rutten BP, Poulton R (2008) Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. Schizophr Bull 34:1066–82
3. Teixeira AP, Dias JML, Carinhas N, et al. (2011) Cell functional enviromics: Unravelling the function of environmental factors. BMC Syst Biol 5:92
4. Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. (2008) GenBank. Nucleic Acids Res 36:D25–30
5. Pagani I, Liolios K, Jansson J, et al. (2012) The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 40:D571–9
6. Kanehisa Laboratories (2010) Kyoto encyclopedia of genes and genomes (KEGG). http://www.genome.jp/kegg/.
7. SRI International (2010) BioCyc database collection. http://biocyc.org/.
8. Borenstein E, Kupiec M, Feldman MW, Ruppin E (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc Natl Acad Sci U S A 105:14482–14487

9. Oldiges M, Lutz S, Pflug S, et al. (2007) Metabolomics: current state and evolving methodologies and tools. Appl Microbiol Biotechnol 76:495–511
10. Ferreira A, Dias J, Teixeira A, et al. (2011) Projection to latent pathways (PLP): a constrained projection to latent variables (PLS) method for elementary flux modes discrimination. BMC Syst Biol 5:181
11. Schuster S, Dandekar T, Fell DA (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17:53–60
12. Isidro IA, Ferreira AR, Clemente JJ, et al. (2015) Analysis of culture media screening data by projection to latent pathways: the case of *Pichia pastoris* X-33. J Biotechnol (in press)
13. Invitrogen Life Technologies (2000) *Pichia* Fermentation Process Guidelines.
14. Oliveira R, Clemente JJ, Cunha AE, Carrondo MJT (2005) Adaptive dissolved oxygen control through the glycerol feeding in a recombinant *Pichia pastoris* cultivation in conditions of oxygen transfer limitation. J Biotechnol 116:35–50
15. Ferreira AR, Ataíde F, von Stosch M, et al. (2012) Application of adaptive DO-stat feeding control to *Pichia pastoris* X33 cultures expressing a single chain antibody fragment (scFv). Bioprocess Biosyst Eng 35:1603–14
16. Jahic M, Rotticci-Mulder JC, Martinelle M, et al. (2002) Modeling of growth and energy metabolism of *Pichia pastoris* producing a fusion protein. Bioprocess Biosyst Eng 24:385–393
17. Chung BK, Selvarasu S, Andrea C, et al. (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb Cell Fact 9:50
18. Sohn SB, Graf AB, Kim TY, et al. (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol J 5:705–15
19. Terzer M, Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24:2229–35

Additional files

Additional file 5.1 – Core metabolic network model for *Pichia pastoris*

Reaction list for the core metabolic network with 116 reactions and 99 metabolites.

Additional file 5.2 – Sampling data

Full sampling data (on-line bioreaction measurements, biomass, scFV and ^{1}H-NMR analysis), envirome and fluxome matrixes

Additional file 5.3 – Elementary flux modes clustered footprint

Description of the 967 elementary flux modes calculated for *Pichia pastoris* core metabolic network.

Chapter 6
Conclusions

6.1 General discussion

In a typical cell culture process there is a large number of environmental variables that shape cellular physiology. One important implication is that the design space for process development, namely culture medium optimization and process control, is potentially very large. While current process development methodologies in the biotechnology industry are essentially of empirical nature, empirical methods are not well suited to handle high-dimensional design spaces unless a substantial level of reductionism is applied. That however may result in suboptimal performance.

With the advances in systems biology, accurate genome-scale metabolic networks are becoming available for several microorganisms used in the industry. Such metabolic networks contain the required information to enumerate all possible operational modes of the cells (i.e., elementary flux modes). With adequate systems biology tools, such as functional enviromics, one can investigate how those operational modes are controlled by the environment and/or how they modify the environment.

In this PhD thesis the biological system of choice was a *Pichia* pastoris strain expressing a scFv antibody fragment. Despite the growing importance of the *Pichia pastoris* expression system as industrial workhorse, the literature is almost absent in systems-level studies on how culture medium composition affects central carbon fluxes and heterologous protein expression. As a basis to address this gap, a culture media screening dataset was built from 26 independent shake flask experiments. Each culture was subject to varying concentrations of trace elements, main salts and glycerol, taking as reference composition the standard BSM+PTM1 medium.

In Chapter 3, hybrid metabolic flux analysis was used to investigate the sensitivity of *P. pastoris* central carbon metabolism and heterologous protein expression to variations in medium composition. For the concentration ranges tested, cell growth seems relatively insensitive to the medium composition, depending essentially on available carbon source, while the expression of heterologous protein is much more sensitive to culture medium variations. This suggests

customization of culture medium might be an important factor to optimize the expression of specific heterologous proteins in *P. pastoris* cultures. The metabolic state that promotes high protein yields seems to be characterized by high overall metabolic rates through main central carbon pathways concomitantly with a relative shift of carbon flux from biosynthetic towards energy generating pathways.

In Chapter 4, we investigated the use of projection to latent pathways (PLP) as a method to extract meaningful metabolic knowledge from culture media screening data. Results show PLP is able to deconvolute contributions of multiple metabolic pathways to the observed metabolic fluxes. It is also able to generate knowledge on the impact of culture media composition at the level of metabolic pathway regulation. Additionally, PLP seems to be more robust to inconsistencies in the data (e.g. outliers or high variability) than the standard, purely statistical PLS regression. A remarkable result is that some trace elements showed a higher influence on metabolism than the dilution of BSM. Our results show iron and manganese at concentrations close to the PTM1 standard inhibit overall metabolic activity and should therefore be decreased for optimized culture media formulations. BSM concentration showed no effect on specific growth rate but correlated negatively with cellular maintenance, suggesting that decreasing overall basal medium concentrations shifts energy expenditures towards cellular maintenance. This is consistent with Chapter 3 findings that (i) a shift of carbon flux from biosynthetic towards energy generating pathways promotes high protein yields and (ii) higher protein yields were observed in experiments with BSM dilution.

While Chapters 3 and 4 were able to focus on how the environment, more specifically culture medium composition, affects cell physiology, it is important to note that the relationship between cellular operation and the environment is not unidirectional. Not only do environmental variables shape cell physiology, the cellular operation itself changes the environment, leaving a characteristic footprint. This problem can be addressed from a systems perspective using cell functional enviromics.

Cell functional enviromics may be defined as the envirome-wide cellular function reconstruction through the collection and systems-level analysis of dynamic envirome data. The key steps for the realization of a functional enviromics study are: (i) setting the universe of cellular functions and envirome components, (ii) collecting informative envirome data over time, and (iii) systems-level analysis of dynamic envirome data to find relationships between environmental variables and cellular functions.

Chapter 5 finally aggregates all previous developments in the core concept of the thesis, cell functional enviromics. To illustrate the concept, it was applied to bioreactor cultures of a recombinant *P. pastoris* strain constitutively expressing a scFv antibody. The concentrations of key metabolites present in the culture medium were quantified by ^{1}H-NMR exometabolomics, adding essential detail to envirome data collected over time. EFM pattern-based clustering was introduced as a reduction method based on the conservation of the exchange reactions pattern. This clustering method was able to achieve a sharp reduction in the total number of possible elementary flux

modes, without significantly compromising how accurately they describe cell physiology. Projection to latent pathways identified major metabolic pathway groups with high correlation with envirome measurements. Detailed analysis of the PLP model showed different envirome factors are more correlated with certain elementary modes than others, reflecting how cellular pathways are activated or repressed by the presence of specific cues in the environment (and vice-versa). The resulting functional enviromics map depicts the relationship between identified cellular operation modes and envirome factors. This map may serve as template to optimize media composition and feeding strategies for *Pichia pastoris*.

Finally, the present PhD thesis is a step forward towards establishing the foundations of the cell functional enviromics methodology. Cell functional enviromics is motivated by the very complex relationship between cells and their environment, characterized by high dimensionality and network-type interdependencies with other cellular components (such as metabolites, proteins and genes). Current techniques to study the effect of environment on cellular physiology, be it at the level of culture media design or process control, are still based on empiricism and reductionism, thus not adequate to the challenge of the quest. The developments in this PhD thesis may provide a relevant contribution for changing the culture media optimization and process control paradigm towards a holistic and systematic discipline in the future.

6.2 Future work

This PhD thesis addresses a novel systems biology methodology, cell functional enviromics, which may be considered still at its infancy. There are thus many challenges to be addressed before this new approach eventually becomes wide spreading in the academia and industry. This thesis focused in particular on the computational challenges, through several practical applications of the projection to latent pathways method, which is the key computational tool for the reconstruction of cells from the side of the environment.

It is important to mention that PLP presupposes a linear model between the environmental parameters and the cellular response in terms of fluxes. Based on well-established models for enzyme kinetics (e.g. Michaelis-Menten) it can be expected that biological systems will tend towards nonlinearity more often than not. While in studies with a limited parameter space it is reasonable to assume that a linear model can be a good approximation to the biological system, it would be interesting to address this issue in future studies. Even if the linear approximation limits the generalization of conclusions outside the range of parameters studied, it identifies critical environmental parameters (e.g. the concentration of iron and manganese in culture media). Such parameters can be the target of subsequent studies with a broader parameter scope. Other approach that would be worth investigating is the use of different data transformations on environment factors and metabolic fluxes. In the same way the Lineweaver-Burk plot linearizes a Michaelis-Menten-type kinetics, such transformations could potentially make the relationship

between fluxes and environmental factors linear (or nearly linear) in a much broader parameter space. This could be easily integrated in PLP with the use of approximative kinetics (e.g. lin-log transformation) [1].

The use of elementary flux modes for large metabolic networks can easily become computationally expensive due to combinatorial explosion, which is especially critical for genome-scale networks. The consequence is that complex metabolic networks are not useful for fast process optimization or online process control. It would be interesting to assess how the simplification of metabolic networks affects the ability of PLP to accurately model cells in culture.

Overall, this PhD thesis shows the potential of cell functional enviromics, as a systems biology approach, to elucidate the ways in which cell physiology is affected by the surrounding environment and how cellular activity itself modifies that environment. This paves the way for rational culture media design and pathway-level process development strategies to replace traditional empirical methods.

References

1. Heijnen JJ (2005) Approximative kinetic formats used in metabolic network modeling. Biotechnol Bioeng 91:534–45

Appendix

Additional file 3.1 – *Pichia pastoris* network model for central carbon metabolism

Description of stoichiometric network used in Chapters 3 and 4

File: addfile3.1_network.xlsx
Sheet: network
Page: 1/2

Stoichiometric Network

	Reaction_number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Reversibility	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**1**
1	G6P[c]	-1	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0
2	F6P[c]	1	-1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
3	FBP[c]	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	DHAP[c]	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	GAP[c]	0	0	1	1	-1	0	0	0	0	0	1	-1	1	0	0	0	0	0	0	0
6	PG3[c]	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	PEP[c]	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
8	PYR[c]	0	0	0	0	0	0	1	0	0	0	0	0	0	-1	0	0	0	0	0	0
9	RU5P[c]	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0
10	XU5P[c]	0	0	0	0	0	0	0	0	1	0	-1	0	-1	0	0	0	0	0	0	0
11	R5P[c]	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0
12	S7P[c]	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0
13	E4P[c]	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
14	ACCOA[m]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0
15	OAA[m]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	1
16	CIT[m]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	0	0	0
17	AKG[m]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	-1	0	0
18	SUC[m]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0
19	MAL[m]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1
20	OAA[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	AKG[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	ACD[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	ETOH[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	AC[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	ACCOA[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	GLY[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	MEOH[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	HCHO[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	DHA[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	NADH	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1
31	NADPH[c]	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
32	NADPH[m]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
33	CO2[i]	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	1	0	0
34	O2[i]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	GLC[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	GLY[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	MEOH[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	O2[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	CO2[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	ETOH[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	AC[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	PYR[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	CIT[e]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	BIOMASS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

File: addfile3.1_network.xlsx
Sheet: network
Page: 2/2

21	22	23	24	25	26	27	28	29	30	31	32	33	GLC[e]	GLY[e]	MEOH[e]	O2[e]	CO2[e]	ETOH[e]	AC[e]	PYR[e]	CIT[e]	BIOMASS
0	**1**	**1**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	-0.0364
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0364
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0165
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0252
-1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	-0.0294
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0107
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0146
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0080
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0008
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0
0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0242
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0267
0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0
0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0033
0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	0	0	0	0	0	0	0	0.0000
0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	-1	0	0	-1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0.0627
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.1995
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.0561
-1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	-1	0	0	0	0	0.0177
0	0	0	0	0	0	0	-0.5	0	-0.5	0	0	0	0	0	0	1	0	0	0	0	0	0.0000
0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

File: addfile3.1_network.xlsx
Sheet: reactions
Page: 1/1

Reactions in this network

RxID	Description	Rev	nC
Embden Meyerhoff Parnas (Glycolysis and Gluconeogenesis)			
r1	G6P[c] ↔ F6P[c]	1	6
r2	F6P[c] + ATP ↔ FBP[c] + ADP	1	6
r3	FBP[c] ↔ DHAP[c] + GAP[c]	1	6
r4	DHAP[c] ↔ GAP[c]	1	3
r5	GAP[c] + NAD + ADP ↔ PG3[c] + NADH + ATP	1	3
r6	PG3[c] ↔ PEP[c] + H2O	1	3
r7	PEP[c] + ADP ↔ PYR[c] + ATP	1	3
Pentose phosphate pathway			
r8	G6P[c] + 2 NADP[c] → RU5P[c] + CO2[i] + 2 NADPH[c]	0	6
r9	RU5P[c] ↔ XU5P[c]	1	5
r10	RU5P[c] ↔ R5P[c]	1	5
r11	R5P[c] + XU5P[c] ↔ S7P[c] + GAP[c]	1	5
r12	S7P[c] + GAP[c] ↔ E4P[c] + F6P[c]	1	5
r13	E4P[c] + XU5P[c] ↔ F6P[c] + GAP[c]	1	5
TCA cycle			
r15	ACCOA[m] + OAA[m] + H2O → CIT[m] + COA	0	6
r16	CIT[m] + NAD[m] → AKG[m] + CO2[i] + NADH[m]	0	6
r17	CIT[m] + NADP[m] → AKG[m] + CO2[i] + NADPH[m]	0	6
r18	AKG[m] + NAD[m] → SUC[m] + CO2[i] + NADH[m] + ATP	0	5
r19	SUC[m] + FAD[m] + H2O → MAL[m] + FADH2[m]	1	4
r20	MAL[m] + NAD[m] → OAA[m]+ NADH[m]	1	4
Anaplerotic pathways			
r21	PYR[c] + CO2[i] + ATP → OAA[c] + ADP	0	4
Cytosol-mitochondria transport			
r14	PYR[c] + COA + NAD[m] → ACCOA[m] + CO2[i] + NADH[m]	0	3
r22	OAA[c] ↔ OAA[m]	1	4
r23	AKG[m] ↔ AKG[c]	1	5
Fermentative pathways and pyruvate metabolism			
r24	PYR[c] → ACD[c] + CO2[i]	0	3
r25	ACD[c] + NADH ↔ ETOH[c] + NAD	1	2
r26	ACD[c] + NADP[c] + H2O → AC[c] + NADPH[c]	0	2
r27	AC[c] + COA[c] + ATP → ACCOA[c] + AMP	0	2
Oxidative phosphorylation			
r28	NADH + 1/2 O2[i] → NAD + H2O	0	0
Glycerol metabolism			
r29	GLY[c] + NAD ↔ DHAP[c] + NADH	1	3
Methanol metabolism			
r30	MEOH[c] + 1/2 O2 → HCHO[c] + H2O	0	1
r31	HCHO[c] + 2 NAD → 2 NADH + CO2[i]	0	1
r32	HCHO[c] + XU5P[c] ↔ DHA[c] + GAP[c]	1	6
r33	DHA[c] + ATP → DHAP[c] + ADP	0	3
Exchange reactions			
r34	GLC[e] + ATP → G6P[c] + ADP	0	6
r35	GLY[e] ↔ GLY[c]	1	3
r36	MEOH[e] ↔ MEOH[c]	1	1
r37	O2[e] ↔ O2[i]	0	0
r38	CO2[i] ↔ CO2[e]	0	1
r39	ETOH[c] ↔ ETOH[e]	0	2
r40	AC[c] ↔ AC[e]	0	2
r41	PYR[c] ↔ PYR [e]	0	3
r42	CIT[m] ↔ CIT[e]	0	6
Biomass formation			
r43	0,0033 ACCOA[c] + 0,008 ACCOA[m] + 0,0267 AKG[c] + 0,0363 F6P[c] + 0,0165 PG3[c] + 0,0363 G6P[c] + 0,00000003 GLY[c] + 0,00002 O2[i] + 0,0242 OAA[c] + 0,0008 OAA[m] + 0,0252 PEP[c] + 0,0294 PYR[c] + 0,0107 R5P[c] + 0,0146 E4P[c] + 0,199 NADPH[c] + 0,0561 NADPH[m] + 0,0626 NAD → BIOM + 0,0127 CO2[i] + 0,0626 NADH + 0,0033 COA[c] +	0	1.018

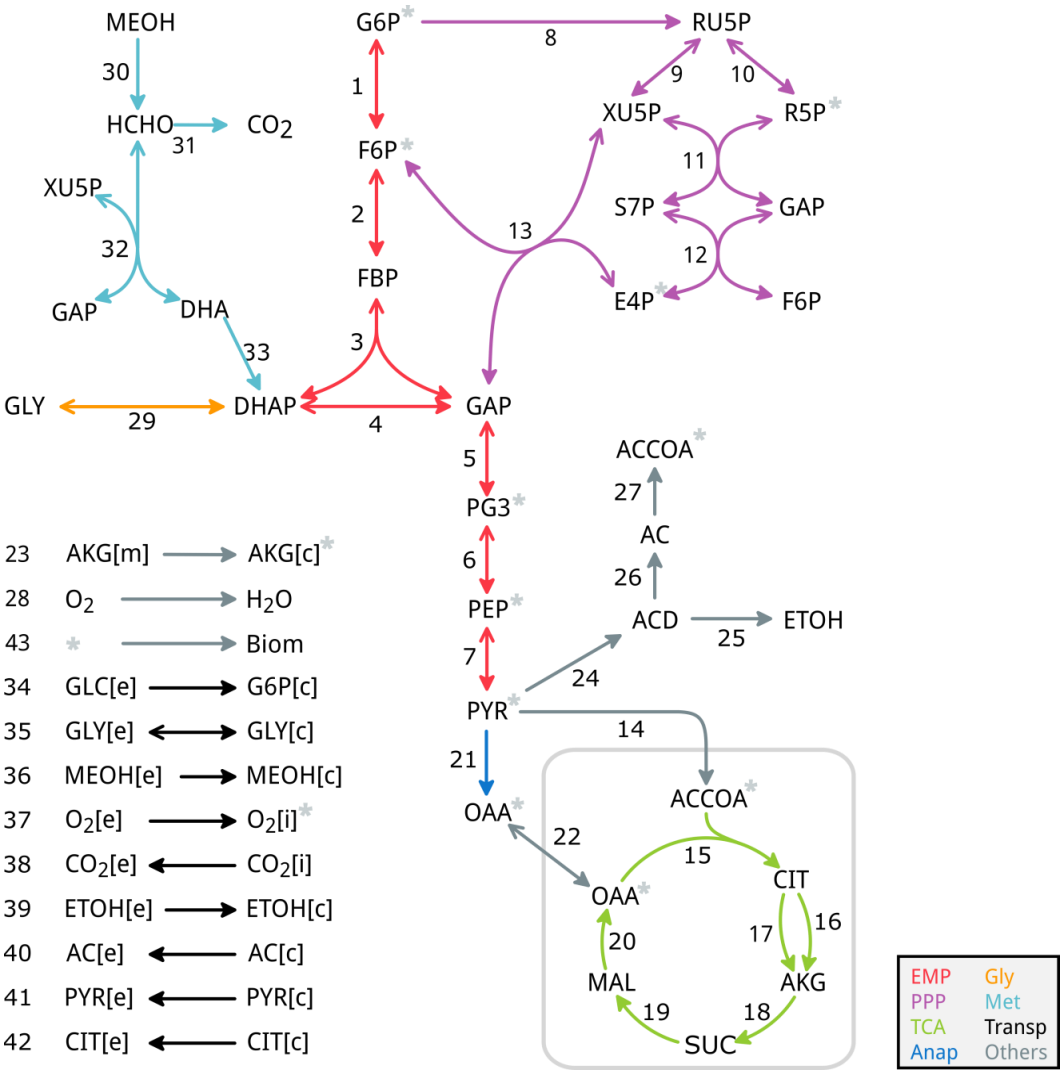
File: addfile3.1_network.xlsx
Sheet: metabolites
Page: 1/1

Metabolites in this network

Abbreviation	Compartment	Name
PG3[c]	cytosolic	3-phosphoglycerate
ACD[c]	cytosolic	acetaldehyde
AC[c]	cytosolic	acetate
ACCOA[m]	mitochondrial	acetyl-coenzyme-A
ACCOA[c]	cytosolic	acetyl-coenzyme-A
AKG[m]	mitochondrial	alpha-ketoglutarate
AKG[c]	cytosolic	alpha-ketoglutarate
CO2[i]	intracellular	carbon dioxyde
CIT[m]	mitochondrial	citrate
DHA[c]	cytosolic	dihydroxyacetone
DHAP[c]	cytosolic	dihydroxyacetone phosphate
E4P[c]	cytosolic	erytrose-4-phosphate
ETOH [c]	cytosolic	ethanol
HCHO[c]	cytosolic	formaldehyde
FBP[c]	cytosolic	fructose-1,6-biphosphate
F6P[c]	cytosolic	fructose-6-phosphate
G6P[c]	cytosolic	glucose-6-phosphate
GAP[c]	cytosolic	glyceraldehyde-3-phosphate
GLY[c]	cytosolic	glycerol
MAL[m]	mitochondrial	malate
MEOH[c]	cytosolic	methanol
NADPH[c]	cytosolic	nicotinamide adenine dinucleotide phosphate
NADPH[m]	mitochondrial	nicotinamide adenine dinucleotide phosphate
NADH	cytosolic	nicotinamide adenine dinucleotide, reduced
OAA[c]	cytosolic	oxaloacetate
OAA[m]	mitochondrial	oxaloacetate
O2[i]	intracellular	oxygen
PEP[c]	cytosolic	phosphoenolpyruvate
PYR[c]	cytosolic	pyruvate
R5P[c]	cytosolic	ribose-5-phosphate
RU5P[c]	cytosolic	ribulose-5-phosphate
S7P[c]	cytosolic	septulose-7-phosphate
SUC[m]	mitochondrial	succinate
XU5P[c]	cytosolic	xylulose-5-phosphate
AC[e]	extracellular	acetate
CIT[e]	extracellular	CIT
CO2[e]	extracellular	carbon dioxyde
ETOH[e]	extracellular	ethanol
GLC[e]	extracellular	glucose
GLY[e]	extracellular	glycerol
MEOH[e]	extracellular	methanol
O2[e]	extracellular	oxygen
PYR[e]	extracellular	pyruvate
BIOMASS		biomass

File: addfile3.1_network.xlsx
Sheet: diagram
Page: 1/1

Network diagram



Additional file 3.2 – Metabolic flux distribution

Metabolic flux distribution for all 26 shake flask experiments with 95% confidence intervals and respective consistency index

Metabolic Flux Distribution

MFA-ajusted fluxes

Units: mmol/(gDCWh) except product in mg/(gDCWh)

RxID		r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20	r21	r22
Rev	ExpID	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	1	1	0	1
nC		6	6	6	3	3	3	3	6	5	5	5	5	5	3	6	6	6	5	4	4	4	4
	1	-0.845	-0.738	-0.738	3.904	3.288	3.184	3.026	0.617	0.336	0.281	0.214	0.214	0.122	2.2	2.15	1.798	0.352	1.982	1.982	1.982	0.325	0.173
	2	-0.711	-0.621	-0.621	3.028	2.509	2.422	2.288	0.519	0.282	0.236	0.18	0.18	0.102	1.673	1.631	1.335	0.296	1.49	1.49	1.49	0.273	0.145
	3	-0.583	-0.509	-0.509	3.364	2.939	2.867	2.757	0.425	0.232	0.194	0.148	0.148	0.084	2.073	2.039	1.796	0.243	1.923	1.923	1.923	0.224	0.119
	4	-0.599	-0.523	-0.523	2.99	2.553	2.479	2.366	0.437	0.238	0.199	0.152	0.152	0.086	1.605	1.57	1.32	0.25	1.451	1.451	1.451	0.23	0.122
	5	-0.84	-0.733	-0.733	3.614	3.002	2.899	2.741	0.613	0.333	0.279	0.212	0.212	0.121	1.881	1.831	1.481	0.35	1.664	1.664	1.665	0.323	0.171
	6	-0.915	-0.799	-0.799	4.683	4.015	3.903	3.731	0.668	0.364	0.304	0.232	0.232	0.132	2.65	2.595	2.214	0.382	2.414	2.414	2.414	0.352	0.187
	7	-0.629	-0.549	-0.549	2.769	2.311	2.233	2.115	0.459	0.25	0.209	0.159	0.159	0.091	1.463	1.425	1.163	0.262	1.301	1.301	1.301	0.242	0.128
	8	-0.609	-0.532	-0.532	3.052	2.608	2.533	2.419	0.444	0.242	0.202	0.154	0.154	0.088	1.699	1.662	1.409	0.254	1.542	1.542	1.542	0.234	0.124
	9	-0.599	-0.523	-0.523	3.175	2.738	2.664	2.552	0.437	0.238	0.199	0.151	0.151	0.086	1.6	1.565	1.315	0.25	1.446	1.446	1.446	0.23	0.122
	10	-0.906	-0.791	-0.791	4.091	3.43	3.319	3.149	0.661	0.36	0.301	0.229	0.229	0.131	2.208	2.154	1.776	0.378	1.974	1.974	1.974	0.348	0.185
	11	-0.664	-0.58	-0.58	3.015	2.531	2.449	2.324	0.485	0.264	0.221	0.168	0.168	0.096	1.468	1.429	1.152	0.277	1.297	1.297	1.297	0.255	0.136
	12	-0.677	-0.591	-0.591	3.194	2.7	2.617	2.49	0.494	0.269	0.225	0.171	0.171	0.098	1.894	1.854	1.572	0.282	1.72	1.72	1.72	0.26	0.138
	13	-0.885	-0.773	-0.773	3.818	3.173	3.064	2.898	0.646	0.351	0.294	0.224	0.224	0.128	1.904	1.852	1.483	0.369	1.676	1.676	1.676	0.34	0.181
	14	-0.915	-0.799	-0.799	4.557	3.89	3.777	3.605	0.668	0.364	0.304	0.232	0.232	0.132	2.624	2.569	2.188	0.382	2.388	2.388	2.388	0.352	0.187
	15	-0.913	-0.797	-0.797	4.072	3.407	3.295	3.123	0.666	0.362	0.303	0.231	0.231	0.132	2.243	2.188	1.808	0.381	2.007	2.007	2.007	0.351	0.186
	16	-0.709	-0.619	-0.619	3.224	2.707	2.62	2.487	0.517	0.282	0.236	0.18	0.18	0.102	1.675	1.633	1.337	0.296	1.492	1.492	1.492	0.273	0.145
	17	-0.843	-0.736	-0.736	3.561	2.946	2.842	2.684	0.615	0.335	0.28	0.213	0.213	0.122	1.879	1.829	1.478	0.352	1.662	1.662	1.662	0.324	0.172
	18	-0.609	-0.532	-0.532	2.615	2.171	2.096	1.982	0.445	0.242	0.203	0.154	0.154	0.088	1.308	1.272	1.018	0.254	1.151	1.151	1.151	0.234	0.124
	19	-0.905	-0.777	-0.774	4.529	3.886	3.774	3.604	0.662	0.364	0.3	0.23	0.234	0.136	2.616	2.562	2.189	0.378	2.39	2.39	2.392	0.325	0.168
	20	-0.881	-0.769	-0.769	3.852	3.21	3.101	2.936	0.643	0.35	0.293	0.223	0.223	0.127	2.068	2.016	1.648	0.367	1.841	1.841	1.841	0.339	0.18
	21	-0.773	-0.675	-0.675	3.58	3.016	2.921	2.776	0.564	0.307	0.257	0.196	0.196	0.111	1.747	1.701	1.379	0.322	1.548	1.548	1.548	0.297	0.158
	22	-0.716	-0.625	-0.625	3.299	2.777	2.689	2.554	0.523	0.284	0.238	0.181	0.181	0.103	1.734	1.691	1.393	0.299	1.549	1.549	1.549	0.275	0.146
	23	-0.902	-0.788	-0.788	3.69	3.033	2.922	2.752	0.658	0.358	0.3	0.228	0.228	0.13	1.827	1.774	1.398	0.376	1.595	1.595	1.595	0.347	0.184
	24	-0.623	-0.544	-0.544	2.859	2.405	2.328	2.211	0.455	0.248	0.207	0.158	0.158	0.09	1.46	1.423	1.164	0.26	1.3	1.3	1.3	0.24	0.127
	25	-0.408	-0.352	-0.351	2.11	1.818	1.767	1.691	0.298	0.163	0.135	0.103	0.105	0.06	1.251	1.227	1.058	0.17	1.148	1.148	1.148	0.151	0.079
	26	-0.977	-0.853	-0.853	4.586	3.874	3.754	3.571	0.713	0.388	0.325	0.247	0.247	0.141	2.369	2.311	1.904	0.407	2.118	2.118	2.118	0.376	0.2

File: addfile3.2_mfd.xlsx
Sheet: MFD
Page: 2/2

	r23	r24	r25	r26	r27	r28	r29	r30	r31	r32	r33	r34	GLC	GLY	MEOH	O2	CO2	ETOH	AC	PYR	CIT	BIOMASS		h	
	1	0	1	0	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0			
	5	3	2	2	2	0	3	1	1	1	6	3	6	3	1	0	1	2	2	3	6	1.018			
1	0.168	0.316	0.295	0.021	0.021	17.98	4.642	0	0	0	0	0	0	4.642	0	8.988	7.051	0.295	0	0	0	6.287	0.004	1E-08	
2	0.141	0.187	0.169	0.018	0.018	13.8	3.648	0	0	0	0	0	0	3.648	0	6.899	5.319	0.169	0	0	0	5.288	0.003	4E-08	
3	0.116	0.332	0.318	0.014	0.014	16.4	3.873	0	0	0	0	0	0	3.873	0	8.202	6.646	0.318	0	0	0	4.337	0.004	1E-07	
4	0.119	0.4	0.385	0.015	0.015	13.24	3.513	0	0	0	0	0	0	3.513	0	6.619	5.311	0.385	0	0	0	4.457	0.009	4E-07	
5	0.167	0.354	0.333	0.021	0.021	15.76	4.347	0	0	0	0	0	0	4.347	0	7.882	6.13	0.333	0	0	0	6.246	0.003	4E-08	
6	0.182	0.529	0.506	0.023	0.023	21.52	5.482	0	0	0	0	0	0	5.482	0	10.76	8.624	0.506	0	0	0	6.81	0.005	7E-08	
7	0.125	0.273	0.258	0.016	0.016	12.19	3.318	0	0	0	0	0	0	3.318	0	6.096	4.762	0.258	0	0	0	4.678	0.002	5E-08	
8	0.121	0.353	0.338	0.015	0.015	13.87	3.584	0	0	0	0	0	0	3.584	0	6.935	5.546	0.338	0	0	0	4.529	0.011	9E-08	
9	0.119	0.591	0.576	0.015	0.015	13.39	3.697	0	0	0	0	0	0	3.697	0	6.696	5.487	0.576	0	0	0	4.453	0.011	3E-07	
10	0.18	0.395	0.372	0.022	0.022	18.27	4.882	0	0	0	0	0	0	4.882	0	9.135	7.162	0.372	0	0	0	6.74	0.01	3E-07	
11	0.132	0.455	0.439	0.016	0.016	12.51	3.595	0	0	0	0	0	0	3.595	0	6.255	4.966	0.439	0	0	0	4.943	0.003	2E-08	
12	0.134	0.187	0.17	0.017	0.017	15.26	3.785	0	0	0	0	0	0	3.785	0	7.628	5.979	0.17	0	0	0	5.034	0.003	5E-08	
13	0.176	0.46	0.438	0.022	0.022	16.15	4.591	0	0	0	0	0	0	4.591	0	8.077	6.314	0.438	0	0	0	6.583	0.003	2E-07	
14	0.182	0.429	0.407	0.023	0.023	21.24	5.356	0	0	0	0	0	0	5.356	0	10.62	8.446	0.407	0	0	0	6.809	0.003	1E-07	
15	0.181	0.33	0.308	0.023	0.023	18.47	4.869	0	0	0	0	0	0	4.869	0	9.233	7.204	0.308	0	0	0	6.789	0.007	7E-08	
16	0.141	0.384	0.366	0.018	0.018	14	3.844	0	0	0	0	0	0	3.844	0	7.003	5.522	0.366	0	0	0	5.277	0.008	6E-09	
17	0.167	0.296	0.275	0.021	0.021	15.71	4.298	0	0	0	0	0	0	4.298	0	7.853	6.069	0.275	0	0	0	6.274	0.003	2E-08	
18	0.121	0.306	0.291	0.015	0.015	11.09	3.147	0	0	0	0	0	0	3.147	0	5.545	4.327	0.291	0	0	0	4.533	0.003	4E-08	
19	0.179	0.465	0.443	0.022	0.022	21.15	5.303	0	0	0	0	0	0	5.301	0	10.57	8.498	0.443	0	0	0	6.748	0.005	0.029	
20	0.175	0.336	0.315	0.022	0.022	17.17	4.622	0	0	0	0	0	0	4.622	0	8.584	6.681	0.315	0	0	0	6.555	0.009	8E-08	
21	0.153	0.563	0.544	0.019	0.019	14.86	4.255	0	0	0	0	0	0	4.255	0	7.428	5.927	0.544	0	0	0	5.748	0.004	3E-08	
22	0.142	0.388	0.371	0.018	0.018	14.44	3.925	0	0	0	0	0	0	3.925	0	7.219	5.704	0.371	0	0	0	5.328	0.004	1E-07	
23	0.179	0.381	0.358	0.022	0.022	15.58	4.478	0	0	0	0	0	0	4.478	0	7.792	6.007	0.358	0	0	0	6.711	0.003	1E-07	
24	0.124	0.375	0.359	0.015	0.015	12.26	3.404	0	0	0	0	0	0	3.404	0	6.132	4.856	0.359	0	0	0	4.637	0.003	1E-07	
25	0.081	0.2	0.19	0.01	0.01	10.03	2.461	0	0	0	0	0	0	2.461	0	5.015	4.028	0.19	0	0	0	3.039	0.003	0.002	
26	0.194	0.612	0.588	0.024	0.024	19.81	5.439	0	0	0	0	0	0	5.439	0	9.904	7.876	0.588	0	0	0	7.266	0.007	3E-07	

Metabolic Flux Distribution

95% CI for MFA-ajusted fluxes

Units: mmol/(gDCWh) except product in mg/(gDCWh)

RxID		r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20	r21	r22
Rev	ExpID	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	1	1	0	1
nC		6	6	6	3	3	3	3	6	5	5	5	5	5	3	6	6	6	5	4	4	4	4
	1	0.191	1.064	1.23	1.372	2.649	2.604	2.56	0.066	0.179	0.065	0.123	0.31	0.419	0.45	0.449	0.422	0.136	0.57	0.584	0.707	1.281	0.991
	2	0.162	0.961	1.114	1.266	2.46	2.427	2.395	0.036	0.164	0.045	0.099	0.275	0.377	0.291	0.268	0.282	0.112	0.497	0.503	0.613	0.977	0.746
	3	0.172	1.007	1.166	1.315	2.548	2.51	2.473	0.043	0.17	0.054	0.109	0.29	0.395	0.361	0.348	0.342	0.122	0.528	0.538	0.653	1.107	0.851
	4	0.241	1.266	1.456	1.581	3.023	2.955	2.888	0.102	0.208	0.096	0.166	0.375	0.5	0.69	0.71	0.637	0.178	0.709	0.735	0.884	1.796	1.403
	5	0.164	0.967	1.121	1.271	2.47	2.437	2.404	0.039	0.165	0.047	0.101	0.277	0.379	0.302	0.282	0.292	0.114	0.501	0.508	0.618	0.997	0.762
	6	0.201	1.088	1.256	1.392	2.684	2.635	2.588	0.079	0.183	0.071	0.131	0.318	0.429	0.493	0.497	0.46	0.143	0.589	0.605	0.731	1.366	1.061
	7	0.164	0.963	1.116	1.265	2.457	2.423	2.39	0.039	0.164	0.047	0.101	0.277	0.378	0.308	0.289	0.297	0.114	0.5	0.507	0.617	1.004	0.768
	8	0.184	1.047	1.21	1.357	2.627	2.585	2.544	0.058	0.177	0.06	0.118	0.303	0.411	0.409	0.402	0.386	0.13	0.555	0.566	0.687	1.204	0.929
	9	0.305	1.53	1.754	1.857	3.515	3.415	3.317	0.144	0.246	0.133	0.218	0.461	0.605	0.969	1.008	0.889	0.231	0.888	0.93	1.113	2.422	1.902
	10	0.21	1.108	1.278	1.412	2.72	2.67	2.619	0.089	0.188	0.075	0.136	0.325	0.437	0.52	0.526	0.485	0.149	0.603	0.621	0.75	1.422	1.106
	11	0.194	1.066	1.232	1.37	2.645	2.599	2.554	0.07	0.179	0.067	0.125	0.311	0.42	0.464	0.464	0.434	0.138	0.573	0.588	0.711	1.305	1.011
	12	0.171	1	1.158	1.307	2.534	2.498	2.462	0.044	0.169	0.052	0.108	0.288	0.392	0.347	0.333	0.331	0.12	0.523	0.532	0.646	1.082	0.831
	13	0.195	1.065	1.231	1.369	2.644	2.598	2.553	0.073	0.18	0.067	0.125	0.311	0.419	0.462	0.463	0.433	0.138	0.573	0.587	0.711	1.303	1.009
	14	0.187	1.05	1.214	1.357	2.623	2.58	2.538	0.062	0.177	0.062	0.12	0.305	0.413	0.429	0.425	0.403	0.132	0.56	0.572	0.693	1.238	0.957
	15	0.183	1.026	1.187	1.333	2.582	2.542	2.503	0.062	0.175	0.058	0.115	0.297	0.404	0.393	0.385	0.371	0.128	0.543	0.553	0.672	1.168	0.9
	16	0.251	1.253	1.44	1.56	2.986	2.918	2.85	0.12	0.21	0.098	0.167	0.373	0.495	0.692	0.713	0.64	0.18	0.704	0.731	0.879	1.796	1.404
	17	0.195	1.07	1.236	1.374	2.651	2.605	2.56	0.072	0.18	0.068	0.126	0.312	0.421	0.468	0.47	0.438	0.139	0.576	0.591	0.715	1.315	1.019
	18	0.172	0.993	1.15	1.297	2.515	2.479	2.442	0.048	0.168	0.053	0.107	0.286	0.39	0.35	0.337	0.333	0.12	0.52	0.529	0.643	1.083	0.832
	19	0.163	0.973	1.128	1.284	2.498	2.466	2.434	0.032	0.166	0.044	0.099	0.278	0.381	0.276	0.247	0.27	0.112	0.5	0.506	0.616	0.963	0.733
	20	0.176	1.009	1.168	1.319	2.56	2.524	2.488	0.052	0.172	0.053	0.109	0.291	0.396	0.346	0.331	0.331	0.122	0.527	0.535	0.651	1.086	0.833
	21	0.203	1.101	1.271	1.408	2.715	2.666	2.617	0.077	0.185	0.072	0.132	0.322	0.434	0.501	0.505	0.467	0.145	0.596	0.613	0.74	1.386	1.076
	22	0.178	1.017	1.178	1.328	2.575	2.538	2.5	0.054	0.173	0.054	0.111	0.293	0.4	0.361	0.348	0.344	0.124	0.533	0.542	0.659	1.113	0.855
	23	0.23	1.186	1.365	1.491	2.862	2.801	2.741	0.102	0.199	0.087	0.152	0.351	0.468	0.616	0.63	0.57	0.165	0.657	0.68	0.819	1.627	1.269
	24	0.183	1.042	1.206	1.353	2.619	2.578	2.538	0.055	0.176	0.059	0.116	0.302	0.41	0.402	0.394	0.379	0.129	0.551	0.563	0.682	1.191	0.918
	25	0.154	0.939	1.09	1.247	2.43	2.401	2.372	0.015	0.16	0.038	0.092	0.267	0.368	0.221	0.178	0.223	0.105	0.478	0.481	0.588	0.871	0.658
	26	0.343	1.593	1.824	1.922	3.642	3.535	3.429	0.184	0.265	0.144	0.233	0.484	0.633	1.031	1.073	0.947	0.249	0.931	0.976	1.168	2.566	2.016

File: addfile3.2_mfd.xlsx
Sheet: MFD CI
Page: 2/2

													GLC	GLY	MEOH	O2	CO2	ETOH	AC	PYR	CIT	BIOMASS	
	r23	r24	r25	r26	r27	r28	r29	r30	r31	r32	r33	r34	r35	r36	r37	r38	r39	r40	r41	r42	r43	rProd	
		1	0	1	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	
		5	3	2	2	2	0	3	1	1	6	3	6	3	1	0	1	2	2	3	6	1.018	
1	0.137	1.726	1.954	0.111	0.056	2.899	0.433	0.159	0.444	0.279	0.218		0	0.388	0	1.234	0.852	0.669	0	0	0	0.61	7E-04
2	0.099	1.684	1.928	0.102	0.051	1.584	0.27	0.136	0.357	0.204	0.139		0	0.205	0	0.657	0.448	0.382	0	0	0	0.333	3E-04
3	0.115	1.711	1.95	0.106	0.053	2.181	0.35	0.146	0.394	0.236	0.174		0	0.3	0	0.924	0.633	0.508	0	0	0	0.393	5E-04
4	0.198	1.802	1.993	0.129	0.064	4.705	0.68	0.202	0.598	0.401	0.336		0	0.623	0	1.961	1.384	1.024	0	0	0	0.941	0.002
5	0.101	1.686	1.928	0.103	0.051	1.69	0.281	0.137	0.363	0.209	0.145		0	0.219	0	0.706	0.482	0.388	0	0	0	0.362	3E-04
6	0.147	1.718	1.937	0.113	0.056	3.241	0.469	0.165	0.468	0.299	0.239		0	0.424	0	1.377	0.955	0.74	0	0	0	0.734	8E-04
7	0.102	1.672	1.912	0.102	0.051	1.751	0.289	0.137	0.364	0.211	0.148		0	0.23	0	0.734	0.501	0.404	0	0	0	0.363	3E-04
8	0.127	1.739	1.976	0.11	0.055	2.566	0.391	0.154	0.423	0.26	0.198		0	0.343	0	1.092	0.751	0.596	0	0	0	0.54	0.001
9	0.272	1.873	1.995	0.151	0.076	6.741	0.965	0.257	0.788	0.548	0.472		0	0.857	0	2.701	1.965	1.325	0	0	0	1.307	0.002
10	0.154	1.723	1.938	0.115	0.057	3.446	0.486	0.169	0.485	0.313	0.252		0	0.441	0	1.461	1.017	0.781	0	0	0	0.833	0.001
11	0.14	1.712	1.935	0.111	0.055	3.011	0.444	0.16	0.45	0.285	0.225		0	0.4	0	1.281	0.886	0.692	0	0	0	0.653	6E-04
12	0.112	1.709	1.949	0.106	0.053	2.071	0.332	0.144	0.388	0.23	0.167		0	0.278	0	0.876	0.6	0.481	0	0	0	0.412	4E-04
13	0.14	1.712	1.935	0.111	0.055	3.003	0.439	0.16	0.45	0.284	0.224		0	0.394	0	1.278	0.884	0.69	0	0	0	0.676	6E-04
14	0.131	1.722	1.953	0.11	0.055	2.732	0.412	0.156	0.432	0.268	0.207		0	0.366	0	1.163	0.801	0.632	0	0	0	0.572	6E-04
15	0.123	1.716	1.951	0.108	0.054	2.446	0.365	0.15	0.412	0.251	0.189		0	0.315	0	1.041	0.716	0.566	0	0	0	0.583	9E-04
16	0.199	1.767	1.95	0.127	0.063	4.733	0.656	0.201	0.596	0.401	0.337		0	0.6	0	1.969	1.395	1.021	0	0	0	1.114	0.002
17	0.141	1.713	1.936	0.111	0.056	3.05	0.447	0.161	0.453	0.287	0.227		0	0.403	0	1.297	0.898	0.7	0	0	0	0.673	6E-04
18	0.112	1.692	1.93	0.105	0.052	2.102	0.331	0.143	0.387	0.23	0.168		0	0.278	0	0.89	0.61	0.488	0	0	0	0.448	4E-04
19	0.096	1.72	1.971	0.104	0.052	1.414	0.253	0.137	0.356	0.2	0.131		0	0.178	0	0.576	0.392	0.318	0	0	0	0.301	5E-04
20	0.112	1.729	1.973	0.107	0.053	2.05	0.318	0.145	0.39	0.23	0.166		0	0.26	0	0.867	0.594	0.474	0	0	0	0.491	8E-04
21	0.149	1.736	1.957	0.114	0.057	3.294	0.481	0.167	0.475	0.304	0.243		0	0.436	0	1.399	0.97	0.752	0	0	0	0.719	1E-03
22	0.115	1.731	1.974	0.107	0.054	2.176	0.335	0.147	0.397	0.237	0.174		0	0.281	0	0.923	0.633	0.504	0	0	0	0.506	7E-04
23	0.179	1.746	1.945	0.121	0.06	4.164	0.587	0.186	0.546	0.361	0.299		0	0.538	0	1.749	1.228	0.921	0	0	0	0.95	8E-04
24	0.125	1.737	1.976	0.11	0.055	2.509	0.386	0.153	0.419	0.256	0.194		0	0.337	0	1.068	0.734	0.583	0	0	0	0.515	6E-04
25	0.084	1.693	1.945	0.101	0.05	0.822	0.197	0.13	0.33	0.177	0.104		0	0.085	0	0.274	0.187	0.152	0	0	0	0.14	4E-04
26	0.289	1.897	2.002	0.157	0.078	7.193	0.979	0.269	0.833	0.581	0.502		0	0.87	0	2.855	2.099	1.376	0	0	0	1.681	0.002

Additional file 4.1 – Culture media screening data

Envirome and fluxome for all 26 shake flask experiments

Culture media dissociated salts

Units: g/L (in the final medium solution)

ExpID	PTM											BSM (and PTM)						
	Cu	Na	I	Mn	MoO4	B	Co	Zn	Cl	Fe	Biotin	Gly	Ca	K	Mg	PO4	SO4	NH4
1	0.00664	7E-05	0.00029	0.00042	5.8E-05	1.5212E-05	9.9E-05	0.04173	0.04537	0.00568	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.3836	2.61744
2	0.00664	0.00017	2.9E-05	0.00042	0.00058	1.5212E-06	9.9E-05	0.00417	0.00464	0.05679	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.4715	2.61744
3	0.00664	0.00017	2.9E-05	0.00424	0.00058	1.5212E-05	9.9E-05	0.04173	0.04537	0.05679	0.00087	18.9474	0.10255	5.0274	0.69601	17.8146	7.86812	1.23984
4	0.00664	7E-05	0.00029	0.00042	5.8E-05	1.5212E-05	0.00099	0.00417	0.00571	0.05679	0.00087	18.9474	0.10255	5.0274	0.69601	17.8146	7.89672	1.23984
5	0.00664	0.00022	0.00029	0.00424	0.00058	1.5212E-05	9.9E-05	0.00417	0.00464	0.00568	0.00087	40	0.21649	10.6134	1.46936	37.6086	16.4255	2.61744
6	0.00066	0.00022	0.00029	0.00042	0.00058	1.5212E-05	0.00099	0.04173	0.04644	0.00568	0.00087	18.9474	0.10255	5.0274	0.69601	17.8146	7.76449	1.23984
7	0.00066	0.00022	0.00029	0.00424	0.00058	1.5212E-05	0.00099	0.04173	0.04644	0.05679	0.00087	40	0.21649	10.6134	1.46936	37.6086	16.5044	2.61744
8	0.00664	0.00022	0.00029	0.00424	0.00058	1.5212E-06	0.00099	0.04173	0.04644	0.05679	8.7E-05	18.9474	0.10255	5.0274	0.69601	17.8146	7.86812	1.23984
9	0.00664	0.00017	2.9E-05	0.00042	0.00058	1.5212E-06	9.9E-05	0.04173	0.04537	0.05679	0.00087	18.9474	0.10255	5.0274	0.69601	17.8146	7.89672	1.23984
10	0.00066	7E-05	0.00029	0.00424	5.8E-05	1.5212E-06	9.9E-05	0.00417	0.00464	0.00568	0.00087	18.9474	0.10255	5.0274	0.69601	17.8146	7.77116	1.23984
11	0.00066	0.00022	0.00029	0.00042	0.00058	1.5212E-06	0.00099	0.00417	0.00571	0.05679	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.4625	2.61744
12	0.00066	0.00002	2.9E-05	0.00042	5.8E-05	1.5212E-05	9.9E-05	0.04173	0.04537	0.05679	0.00087	40	0.21649	10.6134	1.46936	37.6086	16.4625	2.61744
13	0.00066	0.00017	2.9E-05	0.00424	0.00058	1.5212E-05	0.00099	0.00417	0.00571	0.00568	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.3812	2.61744
14	0.00664	0.00022	0.00029	0.00042	0.00058	1.5212E-06	0.00099	0.04173	0.04644	0.00568	0.00087	40	0.21649	10.6134	1.46936	37.6086	16.3836	2.61744
15	0.00664	0.00017	2.9E-05	0.00042	0.00058	1.5212E-05	0.00099	0.00417	0.00571	0.00568	8.7E-05	18.9474	0.10255	5.0274	0.69601	17.8146	7.8088	1.23984
16	0.00664	7E-05	0.00029	0.00424	5.8E-05	1.5212E-05	9.9E-05	0.00417	0.00464	0.05679	8.7E-05	18.9474	0.10255	5.0274	0.69601	17.8146	7.86812	1.23984
17	0.00664	2.2E-05	2.9E-05	0.00424	5.8E-05	1.5212E-06	0.00099	0.00417	0.00571	0.00568	0.00087	40	0.21649	10.6134	1.46936	37.6086	16.3903	2.61744
18	0.00066	7E-05	0.00029	0.00424	5.8E-05	1.5212E-06	0.00099	0.04173	0.04644	0.05679	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.5044	2.61744
19	0.00066	2.2E-05	2.9E-05	0.00042	5.8E-05	1.5212E-05	0.00099	0.04173	0.04644	0.00568	8.7E-05	18.9474	0.10255	5.0274	0.69601	17.8146	7.76449	1.23984
20	0.00066	0.00017	2.9E-05	0.00424	0.00058	1.5212E-06	9.9E-05	0.04173	0.04537	0.00568	8.7E-05	18.9474	0.10255	5.0274	0.69601	17.8146	7.80644	1.23984
21	0.00066	0.00022	0.00029	0.00042	0.00058	1.5212E-05	9.9E-05	0.00417	0.00464	0.05679	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.4977	2.61744
22	0.00066	2.2E-05	2.9E-05	0.00042	5.8E-05	1.5212E-06	0.00099	0.00417	0.00571	0.05679	0.00087	18.9474	0.10255	5.0274	0.69601	17.8146	7.88768	1.23984
23	0.00664	7E-05	0.00029	0.00042	5.8E-05	1.5212E-06	9.9E-05	0.04173	0.04537	0.00568	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.4189	2.61744
24	0.00664	2.2E-05	2.9E-05	0.00424	5.8E-05	1.5212E-05	0.00099	0.04173	0.04644	0.05679	8.7E-05	40	0.21649	10.6134	1.46936	37.6086	16.5135	2.61744
25	0.00664	0.00022	0.00029	0.00424	0.00058	1.5212E-05	0.00099	0.04173	0.04644	0.05679	0.00087	40	0.21649	10.6134	1.46936	37.6086	16.5135	2.61744
26	0.00066	2.2E-05	2.9E-05	0.00042	5.8E-05	1.5212E-06	9.9E-05	0.00417	0.00464	0.00568	8.7E-05	18.9474	0.10255	5.0274	0.69601	17.8146	7.76449	1.23984

Metabolic Flux Distribution

Units: mmol/(gDCW h) except product in mg/(gDCW h)

RxID	ExpID	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20	r21	r22
Rev		1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	1	1	0	1
	1	-0.845	-0.738	-0.738	3.904	3.288	3.184	3.026	0.617	0.336	0.281	0.214	0.214	0.122	2.2	2.15	1.798	0.352	1.982	1.982	1.982	0.325	0.173
	2	-0.711	-0.621	-0.621	3.028	2.509	2.422	2.288	0.519	0.282	0.236	0.18	0.18	0.102	1.673	1.631	1.335	0.296	1.49	1.49	1.49	0.273	0.145
	3	-0.583	-0.509	-0.509	3.364	2.939	2.867	2.757	0.425	0.232	0.194	0.148	0.148	0.084	2.073	2.039	1.796	0.243	1.923	1.923	1.923	0.224	0.119
	4	-0.599	-0.523	-0.523	2.99	2.553	2.479	2.366	0.437	0.238	0.199	0.152	0.152	0.086	1.605	1.57	1.32	0.25	1.451	1.451	1.451	0.23	0.122
	5	-0.84	-0.733	-0.733	3.614	3.002	2.899	2.741	0.613	0.333	0.279	0.212	0.212	0.121	1.881	1.831	1.481	0.35	1.664	1.664	1.665	0.323	0.171
	6	-0.915	-0.799	-0.799	4.683	4.015	3.903	3.731	0.668	0.364	0.304	0.232	0.232	0.132	2.65	2.595	2.214	0.382	2.414	2.414	2.414	0.352	0.187
	7	-0.629	-0.549	-0.549	2.769	2.311	2.233	2.115	0.459	0.25	0.209	0.159	0.159	0.091	1.463	1.425	1.163	0.262	1.301	1.301	1.301	0.242	0.128
	8	-0.609	-0.532	-0.532	3.052	2.608	2.533	2.419	0.444	0.242	0.202	0.154	0.154	0.088	1.699	1.662	1.409	0.254	1.542	1.542	1.542	0.234	0.124
	9	-0.599	-0.523	-0.523	3.175	2.738	2.664	2.552	0.437	0.238	0.199	0.151	0.151	0.086	1.6	1.565	1.315	0.25	1.446	1.446	1.446	0.23	0.122
	10	-0.906	-0.791	-0.791	4.091	3.43	3.319	3.149	0.661	0.36	0.301	0.229	0.229	0.131	2.208	2.154	1.776	0.378	1.974	1.974	1.974	0.348	0.185
	11	-0.664	-0.58	-0.58	3.015	2.531	2.449	2.324	0.485	0.264	0.221	0.168	0.168	0.096	1.468	1.429	1.152	0.277	1.297	1.297	1.297	0.255	0.136
	12	-0.677	-0.591	-0.591	3.194	2.7	2.617	2.49	0.494	0.269	0.225	0.171	0.171	0.098	1.894	1.854	1.572	0.282	1.72	1.72	1.72	0.26	0.138
	13	-0.885	-0.773	-0.773	3.818	3.173	3.064	2.898	0.646	0.351	0.294	0.224	0.224	0.128	1.904	1.852	1.483	0.369	1.676	1.676	1.676	0.34	0.181
	14	-0.915	-0.799	-0.799	4.557	3.89	3.777	3.605	0.668	0.364	0.304	0.232	0.232	0.132	2.624	2.569	2.188	0.382	2.388	2.388	2.388	0.352	0.187
	15	-0.913	-0.797	-0.797	4.072	3.407	3.295	3.123	0.666	0.362	0.303	0.231	0.231	0.132	2.243	2.188	1.808	0.381	2.007	2.007	2.007	0.351	0.186
	16	-0.709	-0.619	-0.619	3.224	2.707	2.62	2.487	0.517	0.282	0.236	0.18	0.18	0.102	1.675	1.633	1.337	0.296	1.492	1.492	1.492	0.273	0.145
	17	-0.843	-0.736	-0.736	3.561	2.946	2.842	2.684	0.615	0.335	0.28	0.213	0.213	0.122	1.879	1.829	1.478	0.352	1.662	1.662	1.662	0.324	0.172
	18	-0.609	-0.532	-0.532	2.615	2.171	2.096	1.982	0.445	0.242	0.203	0.154	0.154	0.088	1.308	1.272	1.018	0.254	1.151	1.151	1.151	0.234	0.124
	19	-0.905	-0.777	-0.774	4.529	3.886	3.774	3.604	0.662	0.364	0.3	0.23	0.234	0.136	2.616	2.562	2.189	0.378	2.39	2.39	2.392	0.325	0.168
	20	-0.881	-0.769	-0.769	3.852	3.21	3.101	2.936	0.643	0.35	0.293	0.223	0.223	0.127	2.068	2.016	1.648	0.367	1.841	1.841	1.841	0.339	0.18
	21	-0.773	-0.675	-0.675	3.58	3.016	2.921	2.776	0.564	0.307	0.257	0.196	0.196	0.111	1.747	1.701	1.379	0.322	1.548	1.548	1.548	0.297	0.158
	22	-0.716	-0.625	-0.625	3.299	2.777	2.689	2.554	0.523	0.284	0.238	0.181	0.181	0.103	1.734	1.691	1.393	0.299	1.549	1.549	1.549	0.275	0.146
	23	-0.902	-0.788	-0.788	3.69	3.033	2.922	2.752	0.658	0.358	0.3	0.228	0.228	0.13	1.827	1.774	1.398	0.376	1.595	1.595	1.595	0.347	0.184
	24	-0.623	-0.544	-0.544	2.859	2.405	2.328	2.211	0.455	0.248	0.207	0.158	0.158	0.09	1.46	1.423	1.164	0.26	1.3	1.3	1.3	0.24	0.127
	25	-0.408	-0.352	-0.351	2.11	1.818	1.767	1.691	0.298	0.163	0.135	0.103	0.105	0.06	1.251	1.227	1.058	0.17	1.148	1.148	1.148	0.151	0.079
	26	-0.977	-0.853	-0.853	4.586	3.874	3.754	3.571	0.713	0.388	0.325	0.247	0.247	0.141	2.369	2.311	1.904	0.407	2.118	2.118	2.118	0.376	0.2

File: addfile4.1_plp data.xlsx
Sheet: fluxome
Page: 2/2

												GLC	GLY	MEOH	O2	CO2	ETOH	AC	PYR	CIT	BIOMASS		
	r23	r24	r25	r26	r27	r28	r29	r30	r31	r32	r33	r34	r35	r36	r37	r38	r39	r40	r41	r42	r43	rProd	h
	1	0	1	0	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	
1	0.168	0.316	0.295	0.021	0.021	17.98	4.642	0	0	0	0	0	4.642	0	8.988	7.051	0.295	0	0	0	0	6.287	0.004	1E-08
2	0.141	0.187	0.169	0.018	0.018	13.8	3.648	0	0	0	0	0	3.648	0	6.899	5.319	0.169	0	0	0	0	5.288	0.003	4E-08
3	0.116	0.332	0.318	0.014	0.014	16.4	3.873	0	0	0	0	0	3.873	0	8.202	6.646	0.318	0	0	0	0	4.337	0.004	1E-07
4	0.119	0.4	0.385	0.015	0.015	13.24	3.513	0	0	0	0	0	3.513	0	6.619	5.311	0.385	0	0	0	0	4.457	0.009	4E-07
5	0.167	0.354	0.333	0.021	0.021	15.76	4.347	0	0	0	0	0	4.347	0	7.882	6.13	0.333	0	0	0	0	6.246	0.003	4E-08
6	0.182	0.529	0.506	0.023	0.023	21.52	5.482	0	0	0	0	0	5.482	0	10.76	8.624	0.506	0	0	0	0	6.81	0.005	7E-08
7	0.125	0.273	0.258	0.016	0.016	12.19	3.318	0	0	0	0	0	3.318	0	6.096	4.762	0.258	0	0	0	0	4.678	0.002	5E-08
8	0.121	0.353	0.338	0.015	0.015	13.87	3.584	0	0	0	0	0	3.584	0	6.935	5.546	0.338	0	0	0	0	4.529	0.011	9E-08
9	0.119	0.591	0.576	0.015	0.015	13.39	3.697	0	0	0	0	0	3.697	0	6.696	5.487	0.576	0	0	0	0	4.453	0.011	3E-07
10	0.18	0.395	0.372	0.022	0.022	18.27	4.882	0	0	0	0	0	4.882	0	9.135	7.162	0.372	0	0	0	0	6.74	0.01	3E-07
11	0.132	0.455	0.439	0.016	0.016	12.51	3.595	0	0	0	0	0	3.595	0	6.255	4.966	0.439	0	0	0	0	4.943	0.003	2E-08
12	0.134	0.187	0.17	0.017	0.017	15.26	3.785	0	0	0	0	0	3.785	0	7.628	5.979	0.17	0	0	0	0	5.034	0.003	5E-08
13	0.176	0.46	0.438	0.022	0.022	16.15	4.591	0	0	0	0	0	4.591	0	8.077	6.314	0.438	0	0	0	0	6.583	0.003	2E-07
14	0.182	0.429	0.407	0.023	0.023	21.24	5.356	0	0	0	0	0	5.356	0	10.62	8.446	0.407	0	0	0	0	6.809	0.003	1E-07
15	0.181	0.33	0.308	0.023	0.023	18.47	4.869	0	0	0	0	0	4.869	0	9.233	7.204	0.308	0	0	0	0	6.789	0.007	7E-08
16	0.141	0.384	0.366	0.018	0.018	14	3.844	0	0	0	0	0	3.844	0	7.003	5.522	0.366	0	0	0	0	5.277	0.008	6E-09
17	0.167	0.296	0.275	0.021	0.021	15.71	4.298	0	0	0	0	0	4.298	0	7.853	6.069	0.275	0	0	0	0	6.274	0.003	2E-08
18	0.121	0.306	0.291	0.015	0.015	11.09	3.147	0	0	0	0	0	3.147	0	5.545	4.327	0.291	0	0	0	0	4.533	0.003	4E-08
19	0.179	0.465	0.443	0.022	0.022	21.15	5.303	0	0	0	0	0	5.301	0	10.57	8.498	0.443	0	0	0	0	6.748	0.005	0.029
20	0.175	0.336	0.315	0.022	0.022	17.17	4.622	0	0	0	0	0	4.622	0	8.584	6.681	0.315	0	0	0	0	6.555	0.009	8E-08
21	0.153	0.563	0.544	0.019	0.019	14.86	4.255	0	0	0	0	0	4.255	0	7.428	5.927	0.544	0	0	0	0	5.748	0.004	3E-08
22	0.142	0.388	0.371	0.018	0.018	14.44	3.925	0	0	0	0	0	3.925	0	7.219	5.704	0.371	0	0	0	0	5.328	0.004	1E-07
23	0.179	0.381	0.358	0.022	0.022	15.58	4.478	0	0	0	0	0	4.478	0	7.792	6.007	0.358	0	0	0	0	6.711	0.003	1E-07
24	0.124	0.375	0.359	0.015	0.015	12.26	3.404	0	0	0	0	0	3.404	0	6.132	4.856	0.359	0	0	0	0	4.637	0.003	1E-07
25	0.081	0.2	0.19	0.01	0.01	10.03	2.461	0	0	0	0	0	2.461	0	5.015	4.028	0.19	0	0	0	0	3.039	0.003	0.002
26	0.194	0.612	0.588	0.024	0.024	19.81	5.439	0	0	0	0	0	5.439	0	9.904	7.876	0.588	0	0	0	0	7.266	0.007	3E-07

Additional file 4.2 – *Pichia pastoris* elementary modes network

Description of the 158 elementary flux modes calculated for *Pichia pastoris* central carbon network

File: addfile4.2_ems158.xlsx
Sheet: EMs
Page: 1/6

Elementary Modes

EM	Reaction	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	r1	0	-0.134	0	-0.134	0	0	0	0	-0.075	-0.134	-0.042	0	-0.049	-0.031	0	0	0	-0.031	0.154	-0.021	0	0	-0.11	-0.055	0.448	0
	r2	-0.022	-0.156	-0.036	-0.185	-0.051	-0.044	-0.049	-0.022	-0.096	-0.117	-0.087	-0.047	-0.082	-0.075	-0.257	-0.656	-0.699	-0.072	-0.298	-0.716	-0.681	-0.257	-0.763	-0.715	-0.003	-0.699
	r3	-0.022	-0.156	-0.036	-0.185	-0.051	-0.044	-0.049	-0.022	-0.096	-0.117	-0.087	-0.047	-0.082	-0.075	-0.257	-0.656	-0.699	-0.072	-0.298	-0.716	-0.681	-0.257	-0.763	-0.715	-0.003	-0.699
	r4	0.154	0.288	0.394	0.288	0.154	0.408	0.403	0.328	0.395	0.288	0.411	0.409	0.406	0.414	0.154	0.095	0.301	0.414	0	0.284	0.136	0.154	0.237	0.119	0	0.301
	r5	0.19	0.19	0.35	0.19	0.19	0.352	0.354	0.306	0.299	0.19	0.324	0.352	0.317	0.331	0.19	0.117	0.259	0.33	0.19	0.226	0.117	0.19	0.156	0.078	0.386	0.259
	r6	0.174	0.174	0.334	0.174	0.174	0.337	0.338	0.29	0.283	0.174	0.309	0.337	0.302	0.316	0.174	0.107	0.248	0.316	0.174	0.217	0.112	0.174	0.143	0.071	0.37	0.248
	r7	0.148	0.148	0.309	0.148	0.148	0.314	0.313	0.265	0.259	0.148	0.287	0.314	0.279	0.295	0.148	0.091	0.231	0.294	0.148	0.202	0.105	0.148	0.122	0.061	0.345	0.231
	r8	0.098	0.098	0.018	0.098	0.098	0	0.01	0.04	0.039	0.098	0.009	0	0.016	0	0.098	0.06	0	0	0.098	0	0	0.098	0.081	0.04	0	0
	r9	0.073	0.073	0	0.087	0.087	-0.011	0	0.015	0.014	0.053	0	-0.01	0	-0.009	0.19	0.366	0.325	-0.01	0.288	0.326	0.33	0.19	0.377	0.355	0.19	0.325
	r10	0.025	0.025	0.018	0.011	0.011	0.011	0.01	0.025	0.025	0.045	0.009	0.01	0.016	0.009	-0.092	-0.306	-0.325	0.01	-0.19	-0.326	-0.33	-0.092	-0.297	-0.315	-0.19	-0.325
	r11	0.015	0.015	0.007	0	0	0.001	0	0.015	0.014	0.034	0	0	0.007	0	-0.103	-0.312	-0.332	0.001	-0.2	-0.333	-0.333	-0.103	-0.305	-0.319	-0.2	-0.332
	r12	0.015	0.015	0.007	0	0	0.001	0	0.015	0.014	0.034	0	0	0.007	0	-0.103	-0.312	-0.332	0.001	-0.2	-0.333	-0.333	-0.103	-0.305	-0.319	-0.2	-0.332
	r13	0	0	-0.007	-0.015	-0.015	-0.012	-0.014	0	0	0.019	-0.013	-0.013	-0.007	-0.013	-0.117	-0.321	-0.342	-0.011	-0.215	-0.341	-0.337	-0.117	-0.317	-0.325	-0.215	-0.342
	r14	0.064	0.064	0.064	0.064	0.064	0.058	0.062	0.064	0.063	0.064	0.057	0.058	0.058	0.055	0.064	0.039	0.043	0.055	0.064	0.038	0.019	0.064	0.053	0.026	0.064	0.043
	r15	0.056	0.056	0.056	0.056	0.056	0.051	0.054	0.056	0.055	0.056	0.05	0.051	0.051	0.048	0.056	0.034	0.038	0.048	0.056	0.033	0.017	0.056	0.046	0.023	0.056	0.038
	r16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	r17	0.056	0.056	0.056	0.056	0.056	0.051	0.054	0.056	0.055	0.056	0.05	0.051	0.051	0.048	0.056	0.034	0.038	0.048	0.056	0.033	0.017	0.056	0.046	0.023	0.056	0.038
	r18	0.029	0.029	0.029	0.029	0.029	0.027	0.028	0.029	0.029	0.029	0.026	0.027	0.027	0.025	0.029	0.018	0.02	0.025	0.029	0.017	0.009	0.029	0.024	0.012	0.029	0.02
	r19	0.029	0.029	0.029	0.029	0.029	0.027	0.028	0.029	0.029	0.029	0.026	0.027	0.027	0.025	0.029	0.018	0.02	0.025	0.029	0.017	0.009	0.029	0.024	0.012	0.029	0.02
	r20	0.029	0.029	0.029	0.029	0.029	0.027	0.028	0.029	0.029	0.029	0.026	0.027	0.027	0.025	0.029	0.018	0.02	0.025	0.029	0.017	0.009	0.029	0.024	0.012	0.029	0.02
	r21	0.052	0.052	0.052	0.052	0.052	0.047	0.05	0.052	0.05	0.052	0.046	0.047	0.047	0.044	0.052	0.032	0.035	0.044	0.052	0.03	0.016	0.052	0.042	0.021	0.052	0.035
	r22	0.027	0.027	0.027	0.027	0.027	0.025	0.027	0.027	0.027	0.027	0.024	0.025	0.025	0.023	0.027	0.017	0.018	0.023	0.027	0.016	0.008	0.027	0.023	0.011	0.027	0.018
	r23	0.027	0.027	0.027	0.027	0.027	0.024	0.026	0.027	0.026	0.027	0.024	0.024	0.024	0.023	0.027	0.016	0.018	0.023	0.027	0.016	0.008	0.027	0.022	0.011	0.027	0.018
	r24	0.003	0.003	0.164	0.003	0.003	0.182	0.173	0.12	0.117	0.003	0.158	0.182	0.148	0.171	0.003	0.002	0.134	0.17	0.003	0.117	0.061	0.003	0.003	0.001	0.199	0.134
	r25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	r26	0.003	0.003	0.164	0.003	0.003	0.182	0.173	0.12	0.117	0.003	0.158	0.182	0.148	0.171	0.003	0.002	0.134	0.17	0.003	0.117	0.061	0.003	0.003	0.001	0.199	0.134
	r27	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.003	0.003	0.002	0.001	0.003	0.003	0.001	0.003	0.002
	r28	0.523	0.791	0.996	0.777	0.508	1	1	0.871	1	0.811	1	1	1	1	0.405	0	0.404	1	0	0.352	0	0.405	0.333	0	0.404
	r29	0.117	0.386	0.43	0.372	0.103	0.452	0.438	0.35	0.491	0.406	0.485	0.452	0.488	0.486	0	-0.249	0	0.486	-0.405	0	-0.183	-3E-08	-2E-08	-0.167	-0.601	-2E-08
	r30	0.058	0.058	0	0.102	0.102	0	0.014	0	0	0	0.013	0.004	0	0.003	0.41	1	1	0	0.703	1	1	0.41	1	1	0.605	1
	r31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	r32	0.058	0.058	0	0.102	0.102	0	0.014	0	0	0	0.013	0.004	0	0.003	0.41	1	1	0	0.703	1	1	0.41	1	1	0.605	1
	r33	0.058	0.058	0	0.102	0.102	0	0.014	0	0	0	0.013	0.004	0	0.003	0.41	1	1	0	0.703	1	1	0.41	1	1	0.605	1
	r34	0.134	0	0.054	0	0.134	0.033	0.046	0.076	0	0	0	0.033	0	0	0.134	0.083	0.024	0	0.288	0	0.011	0.134	0	0	0.484	0.024
	r35	0.117	0.386	0.43	0.372	0.103	0.452	0.438	0.35	0.491	0.406	0.485	0.452	0.488	0.486	3E-08	-0.249	2E-08	0.486	-0.405	2E-08	-0.183	0	0	-0.167	-0.601	0
	r36	0.058	0.058	0	0.102	0.102	0	0.014	0	0	0	0.013	0.004	0	0.003	0.41	1	1	0	0.703	1	1	0.41	1	1	0.605	1
	r37	0.29	0.425	0.498	0.439	0.305	0.5	0.507	0.436	0.5	0.405	0.507	0.502	0.5	0.502	0.408	0.5	0.702	0.5	0.351	0.676	0.5	0.408	0.666	0.5	0.302	0.702
	r38	0.217	0.217	0.297	0.217	0.217	0.287	0.295	0.275	0.269	0.217	0.27	0.287	0.269	0.269	0.217	0.133	0.212	0.269	0.217	0.184	0.096	0.217	0.178	0.089	0.315	0.212
	r39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	r40	0	0	0.16	0	0	0.179	0.169	0.116	0.114	0	0.155	0.179	0.145	0.168	0	0	0.132	0.168	0	0.115	0.06	0	0	0	0.196	0.132
	r41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	r42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	r43	1	1	1	1	1	0.911	0.969	1	0.977	1	0.888	0.912	0.905	0.855	1	0.615	0.672	0.854	1	0.586	0.304	1	0.822	0.411	1	0.672
GLC																											
GLY																											
MEOH																											
O2																											
CO2																											
ETOH																											
AC																											
PYR																											
CIT																											
BIOMA																											

File: addfile4.2_ems158.xlsx
Sheet: EMs
Page: 2/6

	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
r1	-0.021	0.046	0	0.046	-0.11	0	-0.134	0	-0.01	0.067	0	0	-0.031	0	0	0	0.0000	0.059	0.154	0.369	0.261	0.051	0.154	0	0.354	0.051	0.154
r2	-0.716	-0.026	-0.072	-0.026	-0.763	-0.067	-0.207	-0.667	-0.691	-0.6	-0.667	-0.667	-0.094	-0.333	-0.667	-0.571	-0.6667	0.037	0.132	0.347	0.24	4E-04	0.103	-0.667	0.303	4E-04	0.081
r3	-0.716	-0.026	-0.072	-0.026	-0.763	-0.067	-0.207	-0.667	-0.691	-0.6	-0.667	-0.667	-0.094	-0.333	-0.667	-0.571	-1	0.037	0.132	0.347	0.24	4E-04	0.103	-0.667	0.303	4E-04	0.081
r4	0.284	0.108	0.154	0.108	0.237	0.413	0.288	0.048	0.138	0	0.333	0.133	0.418	0.167	0.333	0	0	0.095	0	0	-0.107	0.102	0	0.333	0	0.102	0
r5	0.226	0.19	0.19	0.19	0.156	0.356	0.19	0.048	0.11	0.067	0.333	0.133	0.334	0.167	0.333	0	0	0.19	0.19	0.405	0.19	0.19	0.19	0.333	0.39	0.19	0.19
r6	0.217	0.174	0.174	0.174	0.143	0.341	0.174	0.048	0.105	0.067	0.333	0.133	0.319	0.167	0.333	0	0.1667	0.174	0.174	0.389	0.174	0.174	0.174	0.333	0.374	0.174	0.174
r7	0.202	0.148	0.148	0.148	0.122	0.317	0.148	0.048	0.098	0.067	0.333	0.133	0.297	0.167	0.333	0	0.1667	0.148	0.148	0.363	0.148	0.148	0.148	0.333	0.349	0.148	0.148
r8	0	0.098	0.098	0.098	0.081	0	0.098	0	0	0	0	0	0	0	0	0	0.0000	0.098	0.098	0.098	0.098	0.098	0.098	0	0.098	0.098	0.098
r9	0.326	0.098	0.098	0.098	0.377	0	0.098	0.333	0.33	0.333	0.333	0.333	0	0.167	0.333	0.286	0	0.073	0.073	0.073	0.073	0.087	0.087	0.333	0.087	0.087	0.098
r10	-0.326	0	0	0	-0.297	0	0	-0.333	-0.33	-0.333	-0.333	-0.333	0	-0.167	-0.333	-0.286	0	0.025	0.025	0.025	0.025	0.011	0.011	-0.333	0.011	0.011	0
r11	-0.333	-0.011	-0.011	-0.011	-0.305	-0.01	-0.011	-0.333	-0.333	-0.333	-0.333	-0.333	-0.009	-0.167	-0.333	-0.286	-0.3333	0.015	0.015	0.015	0.015	0	0	-0.333	0	0	-0.011
r12	-0.333	-0.011	-0.011	-0.011	-0.305	-0.01	-0.011	-0.333	-0.333	-0.333	-0.333	-0.333	-0.009	-0.167	-0.333	-0.286	0	0.015	0.015	0.015	0.015	0	0	-0.333	0	0	-0.011
r13	-0.341	-0.025	-0.025	-0.025	-0.317	-0.023	-0.025	-0.333	-0.337	-0.333	-0.333	-0.333	-0.022	-0.167	-0.333	-0.286	-0.3333	0	0	0	0	-0.015	-0.015	-0.333	-0.015	-0.015	-0.025
r14	0.038	0.064	0.064	0.064	0.053	0.059	0.064	0.048	0.018	0.067	0.167	0.067	0.055	0.167	0	0	0.0000	0.064	0.064	0.064	0.064	0.064	0.064	0	0.064	0.064	0.064
r15	0.033	0.056	0.056	0.056	0.046	0.052	0.056	0.048	0.016	0.067	0.167	0.067	0.048	0.167	0	0	0.0000	0.056	0.056	0.056	0.056	0.056	0.056	0	0.056	0.056	0.056
r16	0	0	0	0	0	0	0	0.048	0	0.067	0	0	0	0.167	0	0	0	0	0	0	0	0	0	0	0	0	0
r17	0.033	0.056	0.056	0.056	0.046	0.052	0.056	0	0.016	0	0	0	0.048	0	0	0	0	0.056	0.056	0.056	0.056	0.056	0.056	0	0.056	0.056	0.056
r18	0.017	0.029	0.029	0.029	0.024	0.027	0.029	0.048	0.008	0.067	0	0	0.025	0.167	0	0	0	0.029	0.029	0.029	0.029	0.029	0.029	0	0.029	0.029	0.029
r19	0.017	0.029	0.029	0.029	0.024	0.027	0.029	0.048	0.008	0.067	0	0	0.025	0.167	0	0	0	0.029	0.029	0.029	0.029	0.029	0.029	0	0.029	0.029	0.029
r20	0.017	0.029	0.029	0.029	0.024	0.027	0.029	0.048	0.008	0.067	0	0	0.025	0.167	0	0	0.0000	0.029	0.029	0.029	0.029	0.029	0.029	0	0.029	0.029	0.029
r21	0.03	0.052	0.052	0.052	0.042	0.048	0.052	0	0.015	0	0.167	0.067	0.045	0	0	0	0.0000	0.052	0.052	0.052	0.052	0.052	0.052	0	0.052	0.052	0.052
r22	0.016	0.027	0.027	0.027	0.023	0.025	0.027	0	0.008	0	0.167	0.067	0.024	0	0	0	0	0.027	0.027	0.027	0.027	0.027	0.027	0	0.027	0.027	0.027
r23	0.016	0.027	0.027	0.027	0.022	0.025	0.027	0	0.008	0	0	0	0.023	0	0	0	0	0.027	0.027	0.027	0.027	0.027	0.027	0	0.027	0.027	0.027
r24	0.117	0.003	0.003	0.003	0.003	0.184	0.003	0	0.057	0	0	0	0.172	0	0.333	0	0	0.003	0.003	0.218	0.003	0.003	0.003	0	0.204	0.003	0.003
r25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.333	0	0.0000	0	0	0.215	0	0	0	0	0.2	0	0
r26	0.117	0.003	0.003	0.003	0.003	0.184	0.003	0	0.057	0	0	0	0.172	0	0	0	0.0000	0.003	0.003	0.003	0.003	0.003	0.003	0	0.003	0.003	0.003
r27	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0	9E-04	0	0	0	0.003	0	0	0	0	0.003	0.003	0.003	0.003	0.003	0.003	0	0.003	0.003	0.003
r28	0.352	0.405	0.497	0.405	0.333	1	0.766	0	0	0	0.5	0	1	1	0	0	0	0.405	0.215	0	0	0.405	0.2	0.333	0	0.405	0.19
r29	-2E-08	0	0.092	-3E-08	0	0.446	0.361	-0.286	-0.171	-0.4	0	-0.2	0.481	0	0	-0.286	0	-3E-08	-0.19	-0.405	-0.405	0	-0.205	0	-0.405	-3E-08	-0.216
r30	1	0.134	0.134	0.134	1	0.033	0.134	1	1	1	1	1	0.031	0.5	1	1	1.0000	0.058	0.058	0.058	0.058	0.102	0.102	1	0.102	0.102	0.134
r31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.143	0.0000	0	0	0	0	0	0	0	0	0	0
r32	1	0.134	0.134	0.134	1	0.033	0.134	1	1	1	1	1	0.031	0.5	1	0.857	1.0000	0.058	0.058	0.058	0.058	0.102	0.102	1	0.102	0.102	0.134
r33	1	0.134	0.134	0.134	1	0.033	0.134	1	1	1	1	1	0.031	0.5	1	0.857	1.0000	0.058	0.058	0.058	0.058	0.102	0.102	1	0.102	0.102	0.134
r34	0	0.18	0.134	0.18	0	0.033	0	0	0	0.067	0	0	0	0	0	0	0.0000	0.193	0.288	0.503	0.396	0.186	0.288	0	0.488	0.186	0.288
r35	0	3E-08	0.092	0	2E-08	0.446	0.361	-0.286	-0.171	-0.4	0	-0.2	0.481	0	0	-0.286	0	0	-0.19	-0.405	-0.405	3E-08	-0.205	0	-0.405	0	-0.216
r36	1	0.134	0.134	0.134	1	0.033	0.134	1	1	1	1	1	0.031	0.5	1	1	1	0.058	0.058	0.058	0.058	0.102	0.102	1	0.102	0.102	0.134
r37	0.676	0.27	0.316	0.27	0.666	0.517	0.45	0.5	0.5	0.5	0.75	0.5	0.516	0.75	0.5	0.5	1	0.232	0.137	0.029	0.029	0.254	0.151	0.667	0.051	0.254	0.162
r38	0.184	0.217	0.217	0.217	0.178	0.29	0.217	0.143	0.09	0.2	0	0	0.272	0.5	0.333	0.143	0	0.217	0.217	0.432	0.217	0.217	0.217	0	0.417	0.217	0.217
r39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.333	0	0	0	0	0.215	0	0	0	0	0.2	0	0
r40	0.115	0	0	0	0	0.181	0	0	0.056	0	0	0	0.169	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.333	0	0	0
r42	0	0	0	0	0	0	0	0	0	0	0.167	0.067	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r43	0.586	1	1	1	0.822	0.921	1	0	0.285	0	0	0	0.863	0	0	0	0	1	1	1	1	1	1	0	1	1	1

File: addfile4.2_ems158.xlsx
Sheet: EMs
Page: 3/6

	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
r1	0.249	0.343	0.059	0	0.068	0.154	0.388	0.271	0.068	0.175	0.328	0.543	0.436	0.448	0.242	0.242	0.543	0.5	0.083	0.333	0.638	0.254	1	0.2	0.75	0.583	1
r2	0.176	0.271	0.037	0.017	0.085	0.171	0.405	0.288	0.085	0.153	0.306	0.521	0.414	0.376	0.17	0.17	0.47	0.5	0.083	0.333	0.565	0.203	1	0.2	0.75	0.583	1
r3	0.176	0.271	0.037	0.017	0.085	0.171	0.405	0.288	0.085	0.153	0.306	0.521	0.414	0.376	0.17	0.17	0.47	0.5	0.083	0.333	0.565	0.203	1	0.2	0.75	0.583	1
r4	-0.095	0	0.095	0.154	0.085	0	0	-0.117	0.085	0.153	0	0	-0.107	0	0.206	0.206	-0.095	0.5	0.083	0.333	0	-0.1	0	0	0	-0.417	0
r5	0.19	0.38	0.19	0.19	0.19	0.19	0.424	0.19	0.19	0.306	0.306	0.521	0.306	0.386	0.386	0.386	0.386	1	0.167	0.667	0.576	0.19	1	0.2	0.75	0.167	1
r6	0.174	0.363	0.174	0.174	0.174	0.174	0.408	0.174	0.174	0.29	0.29	0.505	0.29	0.37	0.37	0.37	0.37	1	0.167	0.667	0.559	0.174	1	0.2	0.75	0.167	1
r7	0.148	0.338	0.148	0.148	0.148	0.148	0.383	0.148	0.148	0.265	0.265	0.48	0.265	0.345	0.345	0.345	0.345	1	0.167	0.667	0.534	0.148	1	0.2	0.75	0.167	1
r8	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.04	0.04	0.04	0.04	0	0	0	0	0	0	0	0	0.098	0	0	0	0	0
r9	0.098	0.098	0.073	0.053	0.053	0.053	0.053	0.053	0.053	0.015	0.015	0.015	0.015	0	0	0	0	0	0	0	0	0.087	0	0	0	0	0
r10	0	0	0.025	0.045	0.045	0.045	0.045	0.045	0.045	0.025	0.025	0.025	0.025	0	0	0	0	0	0	0	0	0.011	0	0	0	0	0
r11	-0.011	-0.011	0.015	0.034	0.034	0.034	0.034	0.034	0.034	0.015	0.015	0.015	0.015	-0.011	-0.011	-0.011	-0.011	0	0	0	-0.011	0	0	0	0	0	0
r12	-0.011	-0.011	0.015	0.034	0.034	0.034	0.034	0.034	0.034	0.015	0.015	0.015	0.015	-0.011	-0.011	-0.011	-0.011	0	0	0	-0.011	0	0	0	0	0	0
r13	-0.025	-0.025	0	0.019	0.019	0.019	0.019	0.019	0.019	0	0	0	0	-0.025	-0.025	-0.025	-0.025	0	0	0	-0.025	-0.015	0	0	0	0	0
r14	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0	0.167	0.333	0.064	0.064	0	0.2	0.125	0.167	0.5
r15	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0	0.167	0.333	0.056	0.056	0	0.2	0.125	0.167	0.5
r16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.167	0	0	0	0	0.2	0.125	0.167	0
r17	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0	0	0	0.056	0.056	0	0	0	0	0
r18	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0	0.167	0	0.029	0.029	0	0.2	0.125	0.167	0
r19	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0	0.167	0	0.029	0.029	0	0.2	0.125	0.167	0
r20	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0	0.167	0	0.029	0.029	0	0.2	0.125	0.167	0
r21	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0	0	0.333	0.052	0.052	0	0	0	0	0.5
r22	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0	0	0.333	0.027	0.027	0	0	0	0	0.5
r23	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0	0	0	0.027	0.027	0	0	0	0	0
r24	0.003	0.193	0.003	0.003	0.003	0.003	0.238	0.003	0.003	0.12	0.12	0.334	0.12	0.199	0.199	0.199	0.199	1	0	0	0.389	0.003	0	0	0.625	0	0
r25	0	0.19	0	0	0	0	0.234	0	0	0	0	0.215	0	0	0	0	0	1	0	0	0.19	0	0	0	0.625	0	0
r26	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.12	0.12	0.12	0.12	0.199	0.199	0.199	0.199	0	0	0	0.199	0.003	0	0	0	0	0
r27	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0	0	0	0.003	0.003	0	0	0	0	0
r28	0	0	0.405	0.542	0.405	0.234	0	0	0.405	0.521	0.215	0	0	0.19	0.601	0.601	0	0	1	1	0	0	0	1	0	0	0.5
r29	-0.405	-0.405	0	0.137	0	-0.171	-0.405	-0.405	-3E-08	0	-0.306	-0.521	-0.521	-0.412	-3E-08	0	-0.601	0	0	0	-0.601	-0.405	-1	-0.2	-0.75	-1	-1
r30	0.134	0.134	0.058	0	0	0	0	0	0	0	0	0	0	0.036	0.036	0.036	0.036	0	0	0	0.036	0.102	0	0	0	0	0
r31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r32	0.134	0.134	0.058	0	0	0	0	0	0	0	0	0	0	0.036	0.036	0.036	0.036	0	0	0	0.036	0.102	0	0	0	0	0
r33	0.134	0.134	0.058	0	0	0	0	0	0	0	0	0	0	0.036	0.036	0.036	0.036	0	0	0	0.036	0.102	0	0	0	0	0
r34	0.383	0.478	0.193	0.134	0.203	0.288	0.522	0.405	0.203	0.251	0.404	0.619	0.512	0.484	0.279	0.279	0.579	0.5	0.083	0.333	0.674	0.388	1	0.2	0.75	0.583	1
r35	-0.405	-0.405	3E-08	0.137	3E-08	-0.171	-0.405	-0.405	0	3E-08	-0.306	-0.521	-0.521	-0.412	0	3E-08	-0.601	0	0	0	-0.601	-0.405	-1	-0.2	-0.75	-1	-1
r36	0.134	0.134	0.058	0	0	0	0	0	0	0	0	0	0	0.036	0.036	0.036	0.036	0	0	0	0.036	0.102	0	0	0	0	0
r37	0.067	0.067	0.232	0.271	0.203	0.117	2E-05	2E-05	0.203	0.261	0.107	2E-05	2E-05	0.113	0.319	0.319	0.018	0	0.5	0.5	0.018	0.051	0	0.5	0	0	0.25
r38	0.217	0.406	0.217	0.217	0.217	0.217	0.451	0.217	0.217	0.275	0.275	0.49	0.275	0.315	0.315	0.315	0.315	1	0.5	0	0.505	0.217	0	0.6	1	0.5	0
r39	0	0.19	0	0	0	0	0.234	0	0	0	0	0.215	0	0	0	0	0	1	0	0	0.19	0	0	0	0.625	0	0
r40	0	0	0	0	0	0	0	0	0	0.116	0.116	0.116	0.116	0.196	0.196	0.196	0.196	0	0	0	0.196	0	0	0	0	0	0
r41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
r42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.333	0	0	0	0	0	0	0.5
r43	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0	0

File: addfile4.2_ems158.xlsx
Sheet: EMs
Page: 4/6

	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107
r1	1	0.833	0.667	0.5	0.394	0.5	0.175	0.215	0.215	0.498	0.602	0.248	0.248	0.65	0.549	0.416	0.226	0.226	0.616	0.516	0.248	0.448	0.648	0.548	0.448	0.248	0.022
r2	1	0.833	0.667	0.5	0.358	0.5	0.153	0.179	0.179	0.461	0.565	0.2	0.2	0.601	0.5	0.365	0.175	0.175	0.565	0.465	0.197	0.4	0.597	0.497	0.397	0.197	0
r3	1	0.833	0.667	0.5	0.358	0.5	0.153	0.179	0.179	0.461	0.565	0.2	0.2	0.601	0.5	0.365	0.175	0.175	0.565	0.465	0.197	0.4	0.597	0.497	0.397	0.197	0
r4	0	-0.167	0	-0.5	0	0.5	0.153	0.179	0.179	-0.104	0	0.2	0.2	0	-0.101	0	0.19	0.19	0	-0.1	0.201	0	0	-0.1	0	0.201	0.132
r5	1	0.667	0.667	0	0.35	1	0.306	0.35	0.35	0.35	0.558	0.386	0.386	0.588	0.386	0.365	0.365	0.365	0.565	0.365	0.386	0.386	0.587	0.386	0.386	0.386	0.19
r6	1	0.667	0.667	0	0.334	1	0.29	0.334	0.334	0.334	0.541	0.37	0.37	0.571	0.37	0.349	0.349	0.349	0.549	0.349	0.37	0.37	0.57	0.37	0.37	0.37	0.174
r7	1	0.667	0.667	0	0.309	1	0.265	0.309	0.309	0.309	0.516	0.345	0.345	0.546	0.345	0.323	0.323	0.323	0.523	0.323	0.345	0.345	0.545	0.345	0.345	0.345	0.148
r8	0	0	0	0	0.018	0	0.04	0.018	0.018	0.018	0.018	0	0	0	0	0.011	0.011	0.011	0.011	0.011	0	0	0	0	0	0	0.098
r9	0	0	0	0	0	0	0.015	0	0	0	0	-0.012	-0.012	-0.012	-0.012	0	0	0	0	0	-0.011	-0.012	-0.011	-0.011	-0.011	-0.011	0.073
r10	0	0	0	0	0.018	0	0.025	0.018	0.018	0.018	0.018	0.012	0.012	0.012	0.012	0.011	0.011	0.011	0.011	0.011	0.011	0.012	0.011	0.011	0.011	0.011	0.025
r11	0	0	0	0	0.007	0	0.015	0.007	0.007	0.007	0.007	0.001	0.001	0.001	0.001	0	0	0	0	0	0	0.001	0	0	0	0	0.015
r12	0	0	0	0	0.007	0	0.015	0.007	0.007	0.007	0.007	0.001	0.001	0.001	0.001	0	0	0	0	0	0	0.001	0	0	0	0	0.015
r13	0	0	0	0	-0.007	0	0	-0.007	-0.007	-0.007	-0.007	-0.013	-0.013	-0.013	-0.013	-0.015	-0.015	-0.015	-0.015	-0.015	-0.015	-0.013	-0.015	-0.015	-0.015	-0.015	0
r14	0.333	0.333	0	0	0.064	0	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064
r15	0.333	0.333	0	0	0.056	0	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056
r16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r17	0	0	0	0	0.056	0	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056
r18	0	0	0	0	0.029	0	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
r19	0	0	0	0	0.029	0	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
r20	0	0	0	0	0.029	0	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
r21	0.333	0.333	0	0	0.052	0	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
r22	0.333	0.333	0	0	0.027	0	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027
r23	0	0	0	0	0.027	0	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027
r24	0.333	0	0.667	0	0.164	0	0.12	0.164	0.164	0.164	0.371	0.199	0.199	0.401	0.199	0.178	0.178	0.178	0.378	0.178	0.199	0.199	0.4	0.199	0.199	0.199	0.003
r25	0.333	0	0.667	0	0	0	0	0	0	0	0.208	0	0	0.202	0	0	0	0	0.2	0	0	0	0.2	0	0	0	0
r26	0	0	0	0	0.164	0	0.12	0.164	0.164	0.164	0.164	0.199	0.199	0.199	0.199	0.178	0.178	0.178	0.178	0.178	0.199	0.199	0.199	0.199	0.199	0.199	0.003
r27	0	0	0	0	0.003	0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
r28	0	0	0	0	0.208	1	0.521	0.565	0.565	0	0	0.601	0.601	0	0	0.2	0.58	0.58	0	0	0.601	0.202	0	0	0.2	0.601	0.479
r29	-1	-1	-0.667	-1	-0.358	0	-3E-08	-3E-08	0	-0.565	-0.565	-3E-08	0	-0.601	-0.601	-0.38	-3E-08	0	-0.58	-0.58	-3E-08	-0.4	-0.601	-0.601	-0.401	0	0.074
r30	0	0	0.333	0.5	0	0	0	0	0	0	0	0	0	0	0	0.015	0.015	0.015	0.015	0.015	0.004	0	0.004	0.004	0.004	0.004	0.058
r31	0	0	0.333	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.015	0.015	0.015	0.015	0.015	0.004	0	0.004	0.004	0.004	0.004	0.058
r33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.015	0.015	0.015	0.015	0.015	0.004	0	0.004	0.004	0.004	0.004	0.058
r34	1	0.833	0.667	0.5	0.448	0.5	0.251	0.27	0.27	0.552	0.666	0.285	0.285	0.686	0.585	0.463	0.273	0.273	0.663	0.563	0.284	0.484	0.685	0.584	0.484	0.284	0.156
r35	-1	-1	-0.667	-1	-0.358	0	0	0	3E-08	-0.565	-0.565	0	3E-08	-0.601	-0.601	-0.38	0	3E-08	-0.58	-0.58	0	-0.4	-0.601	-0.601	-0.401	3E-08	0.074
r36	0	0	0.333	0.5	0	0	0	0	0	0	0	0	0	0	0	0.015	0.015	0.015	0.015	0.015	0.004	0	0.004	0.004	0.004	0.004	0.058
r37	0	0	0.167	0.25	0.104	0.5	0.261	0.283	0.283	2E-05	2E-05	0.301	0.301	2E-05	2E-05	0.107	0.297	0.297	0.007	0.007	0.303	0.101	0.002	0.002	0.102	0.303	0.269
r38	0.333	0	1	0.5	0.297	0	0.275	0.297	0.297	0.297	0.505	0.315	0.315	0.517	0.315	0.304	0.304	0.304	0.505	0.304	0.315	0.315	0.515	0.315	0.315	0.315	0.217
r39	0.333	0	0.667	0	0	0	0	0	0	0	0.208	0	0	0.202	0	0	0	0	0.2	0	0	0	0.2	0	0	0	0
r40	0	0	0	0	0.16	0	0.116	0.16	0.16	0.16	0.16	0.196	0.196	0.196	0.196	0.175	0.175	0.175	0.175	0.175	0.196	0.196	0.196	0.196	0.196	0.196	0
r41	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r42	0.333	0.333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r43	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

File: addfile4.2_ems158.xlsx
Sheet: EMs
Page: 5/6

	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134
r1	0.051	0	0.667	-0.017	0.051	0.022	0.208	0	0.072	0.051	0.208	0.072	0.072	0.154	0.448	0.253	0.449	0.213	0.449	0.072	0.072	0.072	0.667	0.067	0.072	0.467	0.222
r2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r4	0.103	0.154	0	0.171	0.103	0.306	0.24	0.23	0.121	0.103	0.24	0.121	0.081	0	0	0	0	-0.06	-7E-04	0.134	0.134	0.134	0	0.1	0.134	-0.3	0
r5	0.19	0.19	0.667	0.19	0.19	0.306	0.386	0.241	0.217	0.19	0.386	0.217	0.19	0.19	0.386	0.289	0.388	0.19	0.386	0.243	0.243	0.243	0.667	0.167	0.243	0.167	0.222
r6	0.174	0.174	0.667	0.174	0.174	0.29	0.37	0.225	0.2	0.174	0.37	0.2	0.174	0.174	0.37	0.273	0.371	0.174	0.37	0.226	0.226	0.226	0.667	0.167	0.226	0.167	0.222
r7	0.148	0.148	0.667	0.148	0.148	0.265	0.345	0.2	0.175	0.148	0.345	0.175	0.148	0.148	0.345	0.248	0.346	0.148	0.345	0.201	0.201	0.201	0.667	0.167	0.201	0.167	0.222
r8	0.098	0.098	0	0.098	0.098	0.04	0	0.073	0.085	0.098	0	0.085	0.098	0.098	0	0.098	0	0.098	0	0.098	0.098	0.098	0	0	0.098	0	0
r9	0.087	0.062	0.333	0.053	0.088	0.015	0.068	0.036	0.085	0.088	0.068	0.085	0.098	0.139	0.188	0.188	0.188	0.169	0.188	0.098	0.098	0.098	0.333	0.033	0.098	0.233	0.111
r10	0.011	0.036	-0.333	0.045	0.011	0.025	-0.068	0.036	0	0.011	-0.068	0	0	-0.041	-0.188	-0.09	-0.188	-0.071	-0.188	0	0	0	-0.333	-0.033	0	-0.233	-0.111
r11	0	0.025	-0.333	0.034	-2E-04	0.015	-0.079	0.025	-0.011	-2E-04	-0.079	-0.011	-0.011	-0.051	-0.199	-0.101	-0.199	-0.081	-0.199	-0.011	-0.011	-0.011	-0.333	-0.033	-0.011	-0.233	-0.111
r12	0	0.025	-0.333	0.034	-2E-04	0.015	-0.079	0.025	-0.011	-2E-04	-0.079	-0.011	-0.011	-0.051	-0.199	-0.101	-0.199	-0.081	-0.199	-0.011	-0.011	-0.011	-0.333	-0.033	-0.011	-0.233	-0.111
r13	-0.015	0.011	-0.333	0.019	-0.015	0	-0.093	0.011	-0.025	-0.015	-0.093	-0.025	-0.025	-0.066	-0.213	-0.116	-0.214	-0.096	-0.214	-0.025	-0.025	-0.025	-0.333	-0.033	-0.025	-0.233	-0.111
r14	0.064	0.064	0.333	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.064	0.117	0.117	0.09	0.167	0.167	0.09	0.167	0.222
r15	0.056	0.056	0.333	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.109	0.109	0.082	0.167	0.167	0.082	0.167	0.222
r16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.053	0.053	0	0.167	0.167	0	0.167	0.222
r17	0.056	0.056	0	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0	0	0.056	0	0
r18	0.029	0.029	0	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.082	0.082	0.029	0.167	0.167	0.029	0.167	0.222
r19	0.029	0.029	0	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.082	0.082	0.029	0.167	0.167	0.029	0.167	0.222
r20	0.029	0.029	0	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.082	0.082	0.029	0.167	0.167	0.029	0.167	0.222
r21	0.052	0.052	0.333	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.078	0	0	0.078	0	0
r22	0.027	0.027	0.333	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.054	0	0	0.054	0	0
r23	0.027	0.027	0	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0	0	0.027	0	0
r24	0.003	0.003	0	0.003	0.003	0.12	0.199	0.054	0.03	0.003	0.199	0.03	0.003	0.003	0.199	0.103	0.201	0.003	0.199	0.003	0.003	0.003	0.5	0	0.003	0	0
r25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.099	0.001	0	0	0	0	0	0.5	0	0	0	0
r26	0.003	0.003	0	0.003	0.003	0.12	0.199	0.054	0.03	0.003	0.199	0.03	0.003	0.003	0.199	0.003	0.199	0.003	0.199	0.003	0.003	0.003	0	0	0.003	0	0
r27	0.003	0.003	0	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0	0	0.003	0	0
r28	0.406	0.533	0	0.576	0.405	0.828	0.601	0.687	0.431	0.405	0.601	0.431	0.353	0.149	0.002	0	0	0	0	0.72	0.72	0.484	0	1	0.484	0	1
r29	8E-04	0.128	-1	0.171	0	0.306	0	0.23	0	-3E-08	-3E-08	-3E-08	-0.053	-0.256	-0.6	-0.405	-0.601	-0.405	-0.601	0	-3E-08	-3E-08	-1	0	0	-1	-0.333
r30	0.102	0.026	1	0	0.103	0	0.24	0	0.121	0.103	0.24	0.121	0.134	0.256	0.6	0.405	0.601	0.346	0.601	0.134	0.134	0.134	1	0.1	0.134	0.7	0.333
r31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r32	0.102	0.026	1	0	0.103	0	0.24	0	0.121	0.103	0.24	0.121	0.134	0.256	0.6	0.405	0.601	0.346	0.601	0.134	0.134	0.134	1	0.1	0.134	0.7	0.333
r33	0.102	0.026	1	0	0.103	0	0.24	0	0.121	0.103	0.24	0.121	0.134	0.256	0.6	0.405	0.601	0.346	0.601	0.134	0.134	0.134	1	0.1	0.134	0.7	0.333
r34	0.185	0.134	0.667	0.117	0.186	0.098	0.245	0.109	0.194	0.186	0.245	0.194	0.207	0.288	0.484	0.387	0.486	0.348	0.485	0.207	0.207	0.207	0.667	0.067	0.207	0.467	0.222
r35	8E-04	0.128	-1	0.171	3E-08	0.306	3E-08	0.23	3E-08	0	0	0	-0.053	-0.256	-0.6	-0.405	-0.601	-0.405	-0.601	3E-08	0	0	-1	0	3E-08	-1	-0.333
r36	0.102	0.026	1	0	0.103	0	0.24	0	0.121	0.103	0.24	0.121	0.134	0.256	0.6	0.405	0.601	0.346	0.601	0.134	0.134	0.134	1	0.1	0.134	0.7	0.333
r37	0.254	0.279	0.5	0.288	0.254	0.414	0.421	0.343	0.276	0.254	0.421	0.276	0.243	0.203	0.301	0.203	0.301	0.173	0.3	0.427	0.427	0.309	0.5	0.55	0.309	0.35	0.667
r38	0.217	0.217	0	0.217	0.217	0.275	0.315	0.242	0.23	0.217	0.315	0.23	0.217	0.217	0.315	0.316	0.316	0.217	0.315	0.375	0.375	0.217	1	0.5	0.217	0.5	0.667
r39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.099	0.001	0	0	0	0	0	0.5	0	0	0	0
r40	0	0	0	0	0	0.116	0.196	0.051	0.026	0	0.196	0.026	0	0	0.196	0	0.196	0	0.196	0	0	0	0	0	0	0	0
r41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r42	0	0	0.333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.026	0	0	0.026	0	0
r43	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0

File: addfile4.2_ems158.xlsx
Sheet: EMs
Page: 6/6

	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158
r1	0.072	0.072	0.072	0.267	0.451	0.072	0.4	0.667	0.4	0.444	0.667	0.364	0	0.451	0	0	0	0.571	0	0.072	0.048	0.036	0.051	0.051
r2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r4	0.134	0.134	0.134	0.4	0	0.134	0.6	0	0.6	0	0.167	-0.364	0.143	0	1	0.4	0	0	0.5	0.376	0.399	0.358	0.365	0.397
r5	0.243	0.243	0.243	0.667	0.488	0.243	1	0.667	1	0.444	0.833	0	0.143	0.39	1	0.4	0	0.571	0.5	0.386	0.386	0.35	0.365	0.386
r6	0.226	0.226	0.226	0.667	0.471	0.226	1	0.667	1	0.444	0.833	0	0.143	0.373	1	0.4	0	0.571	0.5	0.37	0.369	0.334	0.349	0.37
r7	0.201	0.201	0.201	0.667	0.446	0.201	1	0.667	1	0.444	0.833	0	0.143	0.348	1	0.4	0	0.571	0.5	0.345	0.344	0.309	0.323	0.345
r8	0.098	0.098	0.098	0	0.098	0.098	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.018	0.011	0
r9	0.098	0.098	0.098	0.133	0.288	0.098	0.2	0.333	0.2	0.222	0.333	0.182	0	0.19	0	0	0	0.286	0	0	-0.012	0	0	-0.011
r10	0	0	0	-0.133	-0.19	0	-0.2	-0.333	-0.2	-0.222	-0.333	-0.182	0	-0.19	0	0	0	-0.286	0	0	0.012	0.018	0.011	0.011
r11	-0.011	-0.011	-0.011	-0.133	-0.2	-0.011	-0.2	-0.333	-0.2	-0.222	-0.333	-0.182	0	-0.2	0	0	0	-0.286	0	-0.011	0.001	0.007	0	0
r12	-0.011	-0.011	-0.011	-0.133	-0.2	-0.011	-0.2	-0.333	-0.2	-0.222	-0.333	-0.182	0	-0.2	0	0	0	-0.286	0	-0.011	0.001	0.007	0	0
r13	-0.025	-0.025	-0.025	-0.133	-0.215	-0.025	-0.2	-0.333	-0.2	-0.222	-0.333	-0.182	0	-0.215	0	0	0	-0.286	0	-0.025	-0.013	-0.007	-0.015	-0.015
r14	0.064	0.064	0.064	0.333	0.064	0.064	0	0.067	0	0	0	0	0.143	0.064	0	0.2	0	0	0	0.064	0.064	0.064	0.064	0.064
r15	0.056	0.056	0.056	0.333	0.056	0.056	0	0.067	0	0	0	0	0.143	0.056	0	0.2	0	0	0	0.056	0.056	0.056	0.056	0.056
r16	0	0	0	0	0	0	0	0.067	0	0	0	0	0.143	0	0	0	0	0	0	0	0	0	0	0
r17	0.056	0.056	0.056	0	0.056	0.056	0	0	0	0	0	0	0	0.056	0	0	0	0	0	0.056	0.056	0.056	0.056	0.056
r18	0.029	0.029	0.029	0	0.029	0.029	0	0.067	0	0	0	0	0.143	0.029	0	0	0	0	0	0.029	0.029	0.029	0.029	0.029
r19	0.029	0.029	0.029	0	0.029	0.029	0	0.067	0	0	0	0	0.143	0.029	0	0	0	0	0	0.029	0.029	0.029	0.029	0.029
r20	0.029	0.029	0.029	0	0.029	0.029	0	0.067	0	0	0	0	0.143	0.029	0	0	0	0	0	0.029	0.029	0.029	0.029	0.029
r21	0.052	0.052	0.052	0.333	0.052	0.052	0	0	0	0	0	0	0	0.052	0	0.2	0	0	0	0.052	0.052	0.052	0.052	0.052
r22	0.027	0.027	0.027	0.333	0.027	0.027	0	0	0	0	0	0	0	0.027	0	0.2	0	0	0	0.027	0.027	0.027	0.027	0.027
r23	0.027	0.027	0.027	0	0.027	0.027	0	0	0	0	0	0	0	0.027	0	0	0	0	0	0.027	0.027	0.027	0.027	0.027
r24	0.056	0.003	0.003	0	0.003	0.056	1	0	0	0.444	0	0	0	0.199	1	0	0	0	0	0.199	0.199	0.164	0.178	0.199
r25	0.053	0	0	0	0	0.053	1	0	0	0.444	0	0	0	0	1	0	0	0	0	0	0	0	0	0
r26	0.003	0.003	0.003	0	0.003	0.003	0	0	0	0	0	0	0	0.199	0	0	0	0	0	0.199	0.199	0.164	0.178	0.199
r27	0.003	0.003	0.003	0	0.003	0.003	0	0	0	0	0	0	0	0.003	0	0	0	0	0	0.003	0.003	0.003	0.003	0.003
r28	0.405	0.458	0.458	1	0	0.405	0	0	1	0	0	0	1	0	1	0	1	0	1	0.941	1	0.923	0.93	0.994
r29	0	0	-3E-08	0	-0.703	-3E-08	0	-1	0	-0.667	-0.833	-0.909	0.143	-0.605	1	0.4	0	-0.857	0.5	0.34	0.399	0.358	0.35	0.393
r30	0.134	0.134	0.134	0.4	0.703	0.134	0.6	1	0.6	1	1	1	0	0.605	0	0	0.5	1	0	0.036	0	0	0.015	0.004
r31	0	0	0	0	0	0	0	0	0	0	0	0.455	0	0	0	0	0.5	0.143	0	0	0	0	0	0
r32	0.134	0.134	0.134	0.4	0.703	0.134	0.6	1	0.6	0.667	1	0.545	0	0.605	0	0	0	0.857	0	0.036	0	0	0.015	0.004
r33	0.134	0.134	0.134	0.4	0.703	0.134	0.6	1	0.6	0.667	1	0.545	0	0.605	0	0	0	0.857	0	0.036	0	0	0.015	0.004
r34	0.207	0.207	0.207	0.267	0.586	0.207	0.4	0.667	0.4	0.444	0.667	0.364	0	0.488	0	0	0	0.571	0	0.109	0.085	0.091	0.098	0.087
r35	3E-08	3E-08	0	0	-0.703	0	0	-1	0	-0.667	-0.833	-0.909	0.143	-0.605	1	0.4	0	-0.857	0.5	0.34	0.399	0.358	0.35	0.393
r36	0.134	0.134	0.134	0.4	0.703	0.134	0.6	1	0.6	1	1	1	0	0.605	0	0	0.5	1	0	0.036	0	0	0.015	0.004
r37	0.27	0.296	0.296	0.7	0.351	0.27	0.3	0.5	0.8	0.5	0.5	0.5	0.5	0.302	0.5	0.5	0.75	0.5	0.5	0.488	0.5	0.462	0.473	0.499
r38	0.269	0.217	0.217	0	0.217	0.269	1	0.2	0	0.778	0	0.455	0.429	0.315	1	0	0.5	0.143	0	0.315	0.315	0.297	0.304	0.315
r39	0.053	0	0	0	0	0.053	1	0	0	0.444	0	0	0	0	1	0	0	0	0	0	0	0	0	0
r40	0	0	0	0	0	0	0	0	0	0	0	0	0	0.196	0	0	0	0	0	0.196	0.196	0.16	0.175	0.196
r41	0	0.053	0.053	0	0.298	0	0	0.6	1	0	0.833	0	0	0.003	0	0	0	0.571	0.5	0	0	0	0	0
r42	0	0	0	0.333	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0	0	0	0	0	0
r43	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0.999	1	1	1

Additional file 5.1 – *Pichia pastoris* core metabolic network

Reaction list for the core metabolic network with 116 reactions and 99 metabolites

File: addfile5.1_network.m

```
% ------------------------------------------------------------------------
% Stoichiometric network for P. pastoris
% Extended central model
% ------------------------------------------------------------------------
%
% Description: Extended central model with amino acid support and incorporation
of NMR results
% Units: mol (except Biomass and Product in C-mol)
%
% Includes:
% - glycolitic/EMP, PPP, TCA, anaplerotic, folate
% - amino acid synthesis
% - biomass synthesis equation from bio precursors (protein, phospholipids,
%   carbohydrates, RNA and DNA)
% - protein synthesis from amino acids
% - recombinant product synthesis (scFv) based on amino acid composition
% - added frem NMR data: acetate output reaction, allantoin synthesis and
%   output, GPC synthesis and output
%
% Currently off:
% - methanol and glucose uptake and metabolism
% - amino acid export
%
% Authors:
%      Filipe Ataide, Nuno Carinhas, Joao Dias & Rui Oliveira
%
% References:
%      Chung BKS et al. 2010. Genome-scale metabolic reconstruction and in silico
analysis of methylotrophic yeast Pichia pastoris for strain improvement.
Microbial Cell Factories 9
%      Çelik E, Calik P, Oliver SG. 2010. Metabolic Flux Analysis for Recombinant
Protein Production by Pichia pastoris Using Dual Carbon Sources: Effects of
Methanol Feeding Rate. Biotechnology and Bioengineering 105(2): 317-329
%      Kyoto Encyclopedia of Genes and Genomes (KEGG) developed by Kanehisa
Laboratories, 2010. http://www.genome.jp/kegg/
%      Carnicer M et al. 2009. Macromolecular and elemental composition analysis
and extracellular metabolite balances of Pichia pastoris growing at different
oxygen levels. Microb. Cell Fact. 8, 65.
%      Heyland J et al. 2011. Carbon metabolism limits recombinant protein
production in Pichia pastoris. Biotechnol. Bioeng. 108, 1942:53.
%      Sherman F. 2002. Getting Started with Yeast. Methods Enzymol. 350.
%      Voet D, Voet JG. 1995, Biochemistry (2nd edition)
%      Caspeta L et al. 2012. Genome-scale metabolic reconstructions of Pichia
stipitis and Pichia pastoris and in silico evaluation of their potentials'
% ------------------------------------------------------------------------


% --- REACTION FORMULAS (NETWORK DEFINITION) ---

    % % % Assimilation
    '1 GlyOH --> 1 G3P'
    '1 G3P <--> 1 DHAP'
    '1 DHAP <--> 1 GAP'
    %     '1 MetOH + 0.5 O2 --> 1 ForA + 1 H2O'
    %     '1 ForA + 1 H2O --> 1 For' % Should be off for glycerol as only feed
    %     '1 X5P + 1 ForA --> 2 GAP' % Should be off for glycerol as only feed
    %     '1 H3PO4 + 2 H2O -->'
    '1 H2SO4 --> 3 H2O + 1 H2S'

    % Detour from glycolytic pathway
    '1 GAP <--> 1 13BPG'
    '1 13BPG <--> 1 3PG'
    '1 13BPG --> 1 23BPG'
    '1 23BPG --> 1 3PG'
    '1 3PG <--> 1 2PG'
```

```
    '1 2PG <--> 1 PEP + 1 H2O'

    % % % Glycolysis/Gluconeogenesis
    '1 H2O + 2 GAP --> 1 F6P'
    '1 GAP --> 1 PEP + 1 H2O' % irreversible on glycerol and/or methanol
    '1 F6P --> 1 G6P'  % irreversible on glycerol and/or methanol
    %     '1 F6P <--> 2 GAP + 1 DHAP' % can be left out on glycerol and/or
methanol
    '1 PEP --> 1 Pyr'

    % % % PP Pathway
    '1 G6P + 1 H2O --> 1 R5P + 1 CO2'
    '1 R5P --> 1 X5P' % irreversible on glycerol and/or methanol
    '1 R5P + 1 X5P --> 1 E4P + 1 F6P' % irreversible on glycerol and opposite
direction for methanol
    '1 E4P + 1 X5P --> 1 GAP + 1 F6P' % irreversible on glycerol and opposite
direction for methanol
    '1 R5P --> 1 PRPP'
    '1 Pyr + 1 H2O + 1 CO2 --> 1 OA'
    % '1 Mal --> 1 Pyr + 1 CO2' % can be left out on glycerol and/or methanol
    %  '1 OA <--> 1 PEP + 1 CO2' % can be left out on glycerol and/or methanol

    % % % Citric acid pathway
    '1 OA + 1 AcCoA + 1 H2O --> 1 Cit + 1 CoA'
    '1 Cit --> 1 ICit'
    '1 ICit --> 1 aKG + 1 CO2' % irreversible on glycerol and/or methanol
    '1 aKG --> 1 CO2 + 1 Succ'
    '1 Succ --> 1 Fum'
    '1 Fum + 1 H2O --> 1 Mal' % irreversible on glycerol and/or methanol
    '1 Mal --> 1 OA'

    % % % Glyoxylate Pathway
    '1 ICit --> 1 Glx + 1 Succ'
    '1 AcCoA + 1 Glx + 1 H2O --> 1 Mal + 1 CoA'

    % % % C3/C4 Metabolism
    '1 Pyr + 1 CoA --> 1 AcCoA + 1 CO2'
    '1 AcCoA + 1 H2O <--> 1 Ac + 1 CoA'

    % % % Aminoacids synthesis (Bioquimica Stryer, p.683)
    % Glutamate
    '1 aKG + 1 NH3 --> 1 Glu + 1 H2O'
    % Glutamine
    '1 aKG + 1 Glu --> 2 Gln'
    % Proline
    '1 Glu --> 1 Pro + 1 H2O'
    % Ornithine
    '2 Glu --> 1 Orn + 1 aKG'
    % Arginine
    '1 Glu + 1 Gln + 1 Asp + 1 CO2 + 2 H2O --> 1 Arg + 1 aKG + 1 Fum'

    % Aspartate
    '1 OA + 1 Glu --> 1 Asp + 1 aKG' %rev
    % Asparagine
    '1 Asp + 1 NH3 --> 1 Asn' %rev --  -->
    % Methionine
    '1 Asp + 1 H2S + 1 MnTHF --> 1 Met + 1 THF + 1 H2O'
    % Threonine
    '1 Asp + 1 H2O --> 1 Thr'
    % Isoleucine
    '1 Glu + 1 Pyr + 1 Thr --> 1 aKG + 1 Ile + 1 H2O + 1 CO2 + 1 NH3'
    % Lysine
    '2 Glu + 1 AcCoA --> 1 Lys + 1 aKG + 1 CO2 + 1 CoA'

    % Alanine
    '1 Pyr + 1 Glu --> 1 Ala + 1 aKG' %rev
    % Valine
    '1 Glu + 2 Pyr --> 1 Val + 1 H2O + 1 CO2 + 1 aKG'
```

```
    % Leucine
    '2 Pyr + 1 AcCoA + 1 Glu --> 1 Leu + 2 CO2 + 1 aKG + 1 CoA'

    % Chorismate (percursor of aminoacids)
    '1 E4P + 2 PEP --> 1 Chor'
    % Histidine
    '1 PRPP + 1 Gln --> 1 His + 1 AICAR + 1 Glu'
    % Phenylalanine
    '1 Chor + 1 Glu --> 1 Phe + 1 aKG + 1 H2O + 1 CO2'
    '1 Chor + 1 Ala --> 1 Phe + 1 Pyr + 1 H2O + 1 CO2'
    % Tyrosine
    '1 Chor + 1 Glu --> 1 Tyr + 1 aKG + 1 CO2'
    '1 Chor + 1 Ala --> 1 Tyr + 1 Pyr + 1 CO2'
    % Tryptophan
    '1 Chor + 1 Gln + 1 PRPP + 1 Ser --> 1 Trp + 1 GAP + 1 Glu + 1 Pyr + 1 H2O +
1 CO2'% Serine
    '1 3PG + 1 Glu + 1 H2O --> 1 Ser + 1 aKG'
    % Cysteine
    '1 Asp + 1 Ser + 1 H2O + 1 H2S --> 1 Cys + 1 Pyr + 1 NH3 + 1 CO2' % review
this
    % Glycine
    '1 Ser + 1 THF --> 1 Gly + 1 MnTHF + 1 H2O'

    % % % Protein synthesis
    '0.1068 Ala + 0.0674 Arg + 0.0439 Asn + 0.0439 Asp + 0.0928 Gln + 0.0928 Glu
+ 0.0712 Gly + 0.0179 His + 0.0412 Ile + 0.0699 Leu + 0.0633 Lys + 0.0077 Met +
0.0153 Orn + 0.0303 Phe + 0.0367 Pro + 0.0643 Ser + 0.0577 Thr + 0.0140 Trp +
0.0213 Tyr + 0.0558 Val --> 4.9193 Prot' % Cmol

    % % % Lipid biosynthesis
    % The elemental composition of lipids was calculated as an average of
    % the major compounds identified (Carnicer et al. 2009)
    '26 AcCoA + 1 GAP + 1 H2O --> 26 CoA + 1 TAG'
    '8 AcCoA + 1 DAG --> 1 TAG + 8 CoA'
    '18 AcCoA + 10 O2 --> 18 CoA + 1 Zym + 7 H2O + 9 CO2'
    '1 Met + 1 Zym + 2 O2 --> 1 Asp + 1 Erg + 2 H2O + 1 H2S'
    '1 Ser + 18 AcCoA + 1 GAP + 1 CTP --> 18 CoA + 1 CMP + 1 PDS'
    '1 PDS --> 1 PDE + 1 CO2'
    '1 PDE + 3 Met + 6 H2O --> 3 Asp + 3 H2S + 1 PDC'
    '0.0662 TAG + 0.0676 Erg + 0.0127 PDC + 0.0070 PDE --> 6.37 Lip'% Cmol

    % % % Carbohydrate biosynthesis
    '1 G6P + 1 UTP + 1 H2O --> 1 Glc + 1 UDP'
    '1 Glc --> 6 Sug' % Cmol

    % % % RNA biosynthesis
    '1 Asp + 2 Gln + 1 Gly + 1 PRPP + 1 FTHF + 1 CO2 + 3 H2O --> 2 Glu + 1 THF +
1 Fum + 1 AICAR'
    '1 AICAR + 1 FTHF --> 1 THF + 1 IMP + 1 H2O'
    '1 IMP + 1 Asp + 1 GTP --> 1 AMP + 1 Fum + 1 GDP' %(Voet and Voet 1995,
p.801)
    '1 IMP + 1 H2O --> 1 XMP' %(Voet and Voet 1995, p.801)
    '1 XMP + 1 Gln + 1 H2O --> 1 GMP + 1 Glu' %(Voet and Voet 1995, p.801)
    '1 PRPP + 1 Asp + 1 Gln + 1 H2O --> 1 UMP + 1 Glu'
    '1 NH3 + 1 UMP --> 1 CMP + H2O'
    '0.2329 AMP + 0.2280 CMP + 0.2329 GMP + 0.3059 UMP --> 9.463 RNA'

    % % % DNA biosynthesis
    '1 CMP --> 1 dCMP + H2O'
    '1 GMP --> 1 dGMP + H2O'
    '1 AMP --> 1 dAMP + H2O'
    '1 UMP + 1 MnTHF --> 1 dTMP + 1 THF + 1 H2O'
    '0.3 dAMP + 0.2 dCMP + 0.2 dGMP + 0.3 dTMP --> 9.8 DNA'

    % % % Biomass synthesis (Source: Carnicer et al. 2009)
    '1.386 Prot + 0.398 Lip + 1.229 Sug + 0.185 RNA + 0.004 DNA --> 3.202 X' % C-
mol
```

```
    % % % Product synthesis
    '0.0658 Ala + 0.0494 Arg + 0.0123 Asn + 0.0329 Asp + 0.0206 Cys + 0.0370 Glu
+ 0.0535 Gln + 0.1358 Gly + 0.0000 His + 0.0329 Ile + 0.0782 Leu + 0.0288 Lys +
0.0082 Met + 0.0453 Phe + 0.0494 Pro + 0.1564 Ser + 0.0782 Thr + 0.0165 Trp +
0.0494 Tyr + 0.0494 Val + 180 H2O --> 4.650 Prod' % C-mol

    % % % Conversion of carbon units
    '1 THF + 1 For --> 1 FTHF'
    '1 MnTHF + 2 H2O --> 1 THF + 1 For'
    '1 THF + 1 Gly --> 1 MnTHF + 1 CO2 + 1 NH3'
    '1 For --> 1 CO2'

    % % % Energy related
    '1 CMP --> 1 CDP'
    '1 CDP --> 1 CTP'
    '1 GMP --> 1 GDP'
    '1 GDP --> 1 GTP'
    '1 UDP --> 1 UTP'

    % % % Added from NMR identified compounds
    % Glycerophosphocholine (GPC) synthesis
    '1 PDC + 1 H2O --> 1 DAG + 1 Choline'
    '1 Choline + 1 G3P --> 1 GPC'
    % Allantoin synthesis (Voet and Voet 1995, p.817)
    '1 IMP + 1 H2O --> 1 Inosine'
    '1 Inosine --> 1 Hypoxanthine + 1 R5P'
    '1 Hypoxanthine + 1 O2 + 1 H2O --> 1 Xanthine + 1 H2O2'
    '1 Xanthine + 1 O2 + 1 H2O --> 1 Urate + 1 H2O2'
    '1 Urate + 2 H2O + 1 O2 --> 1 Allantoin + 1 CO2 + 1 H2O2'
    '2 H2O2 --> 2 H2O + 1 O2'

    % % % Exchange reactions
    '--> 1 GlyOH' % units in mol
    '--> 1 H2SO4'
    '--> 1 NH3'
    '--> 1 O2'
    '<--> 1 H2O'
    '1 CO2 -->'
    '1 X -->' % Cmol
    '1 Prod -->' % Cmol
    % Derived from NMR results
    '--> 1 Ala'
    '--> 1 Arg'
    '--> 1 Asp'
    '--> 1 Glu'
    '--> 1 Gly'
    '--> 1 Lys'
    '--> 1 Pro'
    '1 GPC -->'
    '1 Ac -->'
    '1 Allantoin -->'
```

Additional file 5.2 – Sampling data

Full sampling data (on-line bioreaction measurements, biomass, scFV and ^{1}H-NMR analysis), envirome and fluxome matrices

File: addfile5.2_samplings.xlsx
Sheet: samples
Page: 1/2

Sample measurements and analytics

Source		At-line			On-line									NMR in medium (consumed)					
FieldID	t	X_WCW	X	Prod	pH	pO2	Temp	O2out	CO2out	V	NH3OH	GlyTotal	GlyCons	GlyOH	Ala	Arg	Asp	Glu	Gly
FullName	Time	Biomass V	Biomass	Product	pH	dissolved (	Temperatu	O2 exhau	CO2 exhe	Volume	Total amm	Total glyc	Consumed	Glycerol	Alanine	Arginine	Aspartate	Glutamate	Glycine
Units	CultureID h	gWCW/L	gDCW/L	mg/L		%	°C	%	%	L	g	g	g	g/L	g/L	g/L	g/L	g/L	g/L
	F1868/14 0	3.07	0.77	0	4.99	98.42	30.06	21.09	0.01	15.00	84	600	17	38.87	2.34	0.75	3.32	3.03	25.34
	5	4.23	1.06	0	4.99	97.39	30.03	21.10	0.01	14.96	86	600	20	38.76	2.44	0.78	3.44	3.14	26.34
	22	16.93	4.23	0	5.03	86.71	30.03	21.00	0.10	14.91	85	600	24	38.66	2.35	0.74	3.30	3.13	25.97
	30	24.05	6.01	0	5.14	52.48	30.10	20.69	0.33	14.81	85	600	78	35.23	2.16	0.67	3.04	3.20	25.51
	46	249.10	62.28	6.71	5.10	19.11	30.06	13.20	5.12	15.80	85	1911	1675	14.94	0.14	0.05	0.43	0.34	
	50	334.98	83.75	16.77	4.90	6.64	30.06	11.74	5.47	16.58	130	2703	2693	0.59	0.04	0.06	0.35	0.16	
	53	439.48	109.87	18.21	4.91	4.85	30.03	12.11	5.31	17.98	206	4115	3846	14.96	0.14	0.09	0.45	0.36	
	70	666.45	166.61	116.4	4.93	4.34	29.99	14.47	4.15	23.23	433	9518	9483	1.49	0.08	0.20	0.35	0.30	
	78	696.78	174.20	124.99	4.93	5.11	29.91	15.32	3.77	24.83	499	11288	11288	0	0.08				
	94	729.00	182.25	146.39	4.93	-0.66	29.99	16.39	3.20	27.20	586	13856	13856	0	0.11				
	98	726.73	181.68	169.50	4.95	7.87	29.95	16.44	3.22	27.57	599	14306	14306	0	0.12				
	102	726.90	181.73	198.22	4.93	6.39	29.95	16.37	3.20	27.94	612	14764	14764	0	0.11				
	121	720.10	180.03	205.16	4.95	7.56	29.99	16.43	3.22	29.73	657	16837	16837	0	0.12				
	128	721.90	180.48	226.38	4.95	4.96	30.03	16.39	3.29	30.24	669	17527	17527	0	0.12				
	146	717.18	179.30	240.16	4.94	5.83	29.95	16.31	3.35	31.85	698	19392	19392	0	0.13				
	153	709.48	177.37	251.87	4.94	4.60	30.06	16.11	3.49	32.41	710	20103	20103	0	0.11				
	166	697.50	174.38	252.82	4.95	7.66	29.95	16.10	3.51	33.67	733	21625	21625	0	0.12				
	170	695.55	173.89	256.49	4.94	4.70	29.99	15.78	3.71	33.93	740	22034	22034	0	0.11				
	174	697.68	174.42	254.67	4.95	5.16	30.06	15.52	3.89	34.27	748	22488	22488	0	0.13				
	F1817/14 23	20.00	5.00	0	4.96	81.92	30.03	20.87	0.13	14.87	88	600	5	38.96					
	27	30.00	7.50	0	4.94	68.97	30.06	20.69	0.24	14.84	92	600	38	37.87					
	30	43.50	10.88	0	4.94	49.73	29.99	20.29	0.46	14.86	107	613	119	33.21					
	46	364.00	91.00	6.5	4.92	3.13	30.03	14.19	4.17	17.34	265	2916	2906	0.58	0.019				
	50	461.00	115.25	21.2	4.92	4.96	29.99	14.05	4.22	18.69	342	4244	4230	0.74	0.025				
	54	534.00	133.50	32.55	4.92	4.82	30.06	14.33	4.17	19.84	399	5436	5420	0.79	0.028				
	71	625.50	156.38	74.75	4.93	5.13	30.10	16.38	3.08	23.42	549	9173	9138	1.51	0.06				
	75	685.00	171.25	81.53	4.92	5.13	29.91	16.52	3.01	23.95	571	9771	9734	1.54	0.061				
	78	692.08	173.02	111.63	4.93	5.04	30.06	16.62	2.94	24.35	589	10245	10245	0	0.077				
	95	735.00	183.75	160.98	4.94	4.87	29.95	17.24	2.59	26.51	672	12560	12560	0	0.121				
	99	769.89	192.47	159.75	4.94	5.22	30.03	17.35	2.52	26.85	687	12974	12974	0	0.1				
	102	760.87	190.22	162.67	4.94	4.87	29.95	17.44	2.46	27.17	700	13364	13364	0	0.102				
	121	731.79	182.95	190.43	4.92	5.18	30.06	17.94	2.16	28.86	760	15217	15217	0	0.134				
	129	760.36	190.09	198.73	4.94	5.00	30.03	18.11	2.06	29.36	778	15810	15810	0	0.13				
	146	762.50	190.63	224.83	4.95	4.69	30.03	18.25	1.97	30.43	813	17034	17034	0	0.133				
	154	763.89	190.97	241.24	4.94	5.18	30.10	18.27	1.95	30.84	825	17543	17543	0	0.148				
	167	755.00	188.75	241.84	4.94	5.09	29.99	18.24	2.03	31.53	844	18360	18360	0	0.195				
	170	757.50	189.38	241.58	4.94	4.91	30.06	18.22	2.05	31.60	847	18506	18506	0	0.142				
	172	773.33	193.33	241.32	4.94	5.85	29.69	18.20	2.06	31.69	850	18657	18657	0	0.14				

File: addfile5.2_samplings.xlsx
Sheet: samples
Page: 2/2

NMR others non-a.a.											NMR a.a. Others				
Lys	Pro	Ac	For	Allantoin	GPC	Inosine	Isobut	Isoval	aKIV	Uridine	Ile	Phe	Thr	Tyr	Val
Lysine	Proline	Acetate	Formate	Allantoin	Glyceroph	Inosine	Isobutyrati	Isovalerati	2-oxoisov.	Uridine	Isoleucine	Phenylala	Threonine	Tyrosine	Valine
g/L	g/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L
	0.70	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.73	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.71	0	0	7.29	0	4.72	0	0	0	1.17	8.82	0	5.23	0	2.87	0
	0.70	0	0	8.10	0.00	11.25	1.33	2.53	4.051	11.62	3.08	6.71	28.44	5.42	9.26
	0.09	0	0	50.62	9.54	17.92	3.11	26.15	28.821	17.37	8.53	13.69	37.95	15.32	12.58
	0.09	10.52	0	77.40	10.75	17.24	6.75	32.56	28.733	24.86	11.32	14.76	40.33	15.04	14.29
	0.16	15.80	0	41.64	14.34	17.01	3.37	26.35	28.493	22.41	8.65	14.33	45.75	15.06	12.98
	0.24	47.29	0	502.05	44.31	50.31	19.15	85.43	18.022	80.53	49.21	63.55	76.44	61.58	66.19
		28.41	0	627.86	46.51	53.81	17.41	98.20	20.293	87.77	53.23	65.63	80.30	63.30	69.58
		53.77	0	849.42	88.39	84.10	24.64	123.14	22.712	111.86	79.54	103.27	111.85	90.02	107.39
		55.96	0	860.30	102.65	85.66	26.63	126.45	23.498	115.99	85.89	105.50	119.34	96.06	115.94
		47.62	0	862.84	103.52	87.02	24.47	124.54	23.512	112.31	81.97	93.82	113.14	89.29	107.11
		49.66	0	906.41	136.64	107.46	26.98	122.16	23.701	132.67	90.17	96.38	121.00	95.47	116.37
		49.35	0	936.86	137.83	118.30	28.63	121.82	22.637	131.27	93.82	118.12	129.44	95.54	125.88
		87.02	0	940.74	164.17	128.78	35.89	120.87	22.487	147.87	99.56	112.03	129.97	92.57	130.21
		51.39	0	951.01	170.13	129.28	31.63	116.96	28.733	156.61	89.63	88.19	113.91	83.35	108.69
		65.46	0	966.66	178.80	127.78	35.63	122.88	27.453	178.83	96.96	111.73	126.30	84.77	121.70
		49.42	0	921.29	177.60	124.19	33.05	119.07	30.728	172.86	91.06	99.50	119.99	81.51	111.21
		91.36	0	964.61	193.46	135.32	41.12	130.43	25.344	191.37	96.06	86.27	123.48	80.53	112.60
		2.09	0	2.56	3.10	6.08				6.33		3.43	0		0
		2.41	0	3.22	4.18	3.35				4.01	1.03	2.98	0		0.92
		2.42	0	4.37	6.15	6.86				11.96	1.84	4.68	15.82		1.74
		9.91	0	91.73	18.56	27.21				20.82	11.41	10.60	31.72		9.90
		11.75	0	149.21	24.38	35.51				28.55	17.43	16.18	34.71		15.51
		12.06	0	210.91	28.64	39.21				30.64	23.13	20.22	39.49		18.82
		22.56	0	514.07	75.70	81.61				78.95	58.67	38.51	73.81		47.03
		25.09	0	540.14	83.42	77.61				80.28	59.48	37.25	71.42		48.75
		28.50	0	643.42	103.16	94.31				91.11	77.28	53.28	92.34		62.45
		41.67	0	837.75	158.98	132.51				126.65	140.17	117.51	141.57		99.43
		37.60	0	887.96	171.89	136.77				129.07	106.35	60.49	117.93		89.93
		39.30	0	935.29	182.58	138.65				135.84	109.16	64.77	125.37		93.61
		51.06	0	1098.40	245.12	175.80				168.90	152.28	109.11	159.68		127.48
		51.53	0	1158.74	270.70	180.98				180.87	145.11	89.78	158.71		128.47
		54.03	0	1187.52	301.96	194.58				194.86	148.72	87.12	158.85		135.42
		55.00	0	1261.09	335.17	196.75				211.13	169.29	113.29	173.11		148.72
		66.57	0	1277.25	379.45	213.19				241.91	230.63	190.85	226.98		176.26
		59.54	0	1208.61	369.10	193.04				237.04	160.55	107.45	163.36		145.79
		58.11	0	1225.89	374.75	198.94				240.82	158.14	108.14	161.83		144.13

File: addfile5.2_samplings.xlsx
Sheet: envirome
Page: 1/2

Envirome

Source		At-line		On-line				NMR in medium (consumed)									NMR others non-a.a.				
FieldID	t	X	Prod	pO2	O2out	CO2out	RQ	GlyOH	Ala	Arg	Asp	Glu	Gly	Pro	Ac	Allantoin	GPC	Inosine	Uridine
FullName	Time	Biomass	Product	pO2	O2 exhau	CO2 exhau	Resp quo	Glycerol	Alanine	Arginine	Aspartate	Glutamate	Glycine	Proline	Acetate	Allantoin	Glycerop	Inosine	Uridine
Units	CulturelD h	gDCW/L	mg/L	%	%	%		g/L	g/L	g/L	g/L	g/L	g/L	g/L	mg/L	mg/L	mg/L	mg/L	mg/L
	F1868/14 0	0.71	0.00	98.30	20.99	0.01	1.000	39.25	2.51	0.782	3.45	3.16	26.35	0.730	0.00	0.68	0.91	1.50	3.49
	F1868/14 5	1.02	0.00	97.34	21.00	0.01	1.000	39.11	2.49	0.779	3.44	3.16	26.32	0.729	0.00	1.60	1.15	1.95	4.27
	F1868/14 22	4.10	0.67	86.64	20.90	0.10	0.950	37.41	2.28	0.747	3.33	3.16	26.42	0.720	0.69	6.44	2.74	4.76	8.43
	F1868/14 30	9.25	1.26	53.19	20.59	0.33	0.806	34.50	1.77	0.649	2.97	2.83	24.65	0.673	3.20	9.97	4.15	7.08	11.47
	F1868/14 46	59.22	8.61	19.82	13.10	5.12	0.648	19.47	0.19	0.072	0.51	0.39	1.06	0.109	11.54	40.98	9.97	15.50	21.77
	F1868/14 50	82.12	13.94	4.99	11.64	5.47	0.584	15.07	0.11	0.055	0.37	0.23	-0.30	0.103	14.22	56.94	12.30	18.50	25.44
	F1868/14 53	101.77	20.55	4.75	12.01	5.31	0.591	11.63	0.08	0.081	0.40	0.28	-0.33	0.145	16.66	75.76	14.66	21.42	29.03
	F1868/14 70	164.88	91.45	5.10	14.37	4.15	0.626	1.68	0.06	0.187	0.33	0.29	0.00	0.221	30.45	438.59	34.90	42.84	56.12
	F1868/14 78	174.68	125.21	5.19	15.22	3.77	0.652	0.57	0.07	0.017	0.02	0.02	0.00	0.018	36.45	644.92	49.38	55.34	72.45
	F1868/14 94	181.41	166.63	5.34	16.29	3.20	0.679	0.06	0.10	0.000	0.00	0.00	0.00	0.000	46.44	841.36	85.42	80.29	104.34
	F1868/14 98	181.82	174.70	5.58	16.34	3.22	0.691	0.04	0.11	0.001	0.00	0.00	0.00	0.001	48.22	858.54	94.16	85.55	110.69
	F1868/14 102	182.01	182.83	5.56	16.27	3.20	0.677	0.02	0.11	0.001	0.00	0.00	0.00	0.001	49.84	871.18	103.01	90.75	116.73
	F1868/14 121	181.01	213.56	5.32	16.33	3.22	0.689	0.00	0.12	0.000	0.00	0.00	0.00	0.000	55.83	914.71	137.72	111.22	137.66
	F1868/14 128	180.44	223.46	5.25	16.29	3.29	0.699	0.00	0.12	0.000	0.00	0.00	0.00	0.000	57.83	928.89	146.55	116.44	142.86
	F1868/14 146	178.60	243.23	5.58	16.21	3.35	0.698	0.00	0.12	0.000	0.00	0.00	0.00	0.000	63.17	944.81	164.77	125.42	156.06
	F1868/14 153	177.53	248.22	5.33	16.01	3.49	0.698	0.00	0.12	0.000	0.00	0.00	0.00	0.000	64.86	949.10	170.62	127.42	161.85
	F1868/14 166	175.04	253.96	6.09	16.00	3.51	0.702	0.00	0.12	0.000	0.00	0.00	0.00	0.000	68.61	951.95	181.63	130.15	175.13
	F1868/14 170	174.37	255.02	5.28	15.68	3.71	0.696	0.00	0.12	0.000	0.00	0.00	0.00	0.000	69.84	951.68	184.74	130.84	179.12
	F1868/14 174	173.67	255.99	5.50	15.42	3.89	0.697	0.00	0.12	0.000	0.00	0.00	0.00	0.000	71.24	953.52	188.12	131.58	183.46
	F1817/14 23	4.64	0.00	82.55	20.87	0.13	0.947	39.60	0.000	0	0	0	0	0	2.08	2.08	3.49	4.31	5.85
	F1817/14 27	7.84	0.00	68.41	20.69	0.24	0.774	36.79	0.001	0	0	0	0	0	2.60	3.65	4.52	5.59	7.16
	F1817/14 30	13.17	0.00	50.17	20.29	0.46	0.648	30.95	0.003	0	0	0	0	0	3.23	6.45	5.83	7.29	8.77
	F1817/14 46	86.54	14.53	4.40	14.19	4.17	0.612	1.43	0.018	0	0	0	0	0	8.20	78.51	17.81	23.86	21.36
	F1817/14 50	109.61	21.73	4.97	14.05	4.22	0.606	0.75	0.024	0	0	0	0	0	10.05	128.51	23.02	30.75	26.28
	F1817/14 54	126.68	29.67	4.81	14.33	4.17	0.625	0.47	0.029	0	0	0	0	0	11.97	188.07	28.87	37.96	31.60
	F1817/14 71	164.93	78.49	4.99	16.38	3.08	0.667	0.01	0.060	0	0	0	0	0	22.83	518.02	71.48	77.47	66.30
	F1817/14 75	169.87	90.76	4.98	16.52	3.01	0.672	0.00	0.067	0	0	0	0	0	25.34	581.07	83.70	86.22	75.33
	F1817/14 78	173.55	100.88	4.99	16.62	2.94	0.672	0.00	0.073	0	0	0	0	0	27.36	629.19	94.20	93.27	82.81
	F1817/14 95	187.30	151.50	4.99	17.24	2.59	0.689	0.00	0.103	0	0	0	0	0	37.93	856.64	159.52	130.89	124.98
	F1817/14 99	188.55	158.97	5.01	17.35	2.52	0.690	0.00	0.107	0	0	0	0	0	39.75	895.85	172.75	137.57	132.74
	F1817/14 102	189.19	165.49	4.95	17.44	2.46	0.691	0.00	0.111	0	0	0	0	0	41.47	933.54	185.72	143.95	140.13
	F1817/14 121	187.81	192.75	5.06	17.94	2.16	0.706	0.00	0.128	0	0	0	0	0	49.27	1102.89	249.96	173.76	174.07
	F1817/14 129	188.41	202.52	5.00	18.11	2.06	0.713	0.00	0.132	0	0	0	0	0	51.57	1148.31	271.08	182.04	184.54
	F1817/14 146	190.15	225.76	4.98	18.25	1.97	0.717	0.00	0.139	0	0	0	0	0	56.04	1216.56	314.76	194.76	206.38
	F1817/14 154	190.36	234.13	5.00	18.27	1.95	0.714	0.00	0.141	0	0	0	0	0	57.90	1236.04	334.29	198.17	216.56
	F1817/14 167	190.64	241.84	5.03	18.24	2.03	0.733	0.00	0.142	0	0	0	0	0	60.80	1243.04	368.74	201.37	235.47
	F1817/14 170	190.90	242.73	5.07	18.22	2.05	0.736	0.00	0.142	0	0	0	0	0	61.27	1241.29	375.26	201.61	239.13
	F1817/14 172	191.22	243.59	4.77	18.20	2.06	0.735	0.00	0.142	0	0	0	0	0	61.71	1239.25	381.88	201.82	242.85

File: addfile5.2_samplings.xlsx
Sheet: envirome
Page: 2/2

NMR a.a. Others			
Ile	Phe	Thr	Val
Isoleucine	Phenylala	Threonine	Valine
mg/L	mg/L	mg/L	mg/L
0.30	0.00	0.11	0.00
0.46	0.00	0.24	0.00
1.67	2.74	2.67	1.77
2.83	4.88	7.01	3.84
8.28	14.96	30.95	12.73
10.77	19.33	39.01	17.12
13.48	23.78	45.88	21.84
39.78	54.86	77.80	57.40
55.36	69.94	90.27	74.70
78.53	94.35	110.59	103.88
81.85	97.98	113.84	108.47
84.62	100.87	116.63	112.17
92.24	107.61	124.58	121.53
93.70	108.51	125.82	123.18
95.35	105.58	125.38	121.96
95.16	103.20	124.31	119.40
94.80	98.82	123.01	115.48
94.71	97.09	122.73	114.29
94.65	95.03	122.45	112.99
0.84	3.01	0.40	0.75
1.25	3.71	4.33	1.13
1.88	4.57	9.06	1.70
10.57	11.82	30.75	9.34
15.22	14.79	36.61	13.30
20.63	18.01	42.36	17.81
57.59	38.48	73.82	46.90
67.00	43.94	81.38	54.02
74.73	48.61	87.67	59.84
113.82	73.81	121.55	90.41
119.39	77.28	127.06	95.61
124.36	80.38	132.19	100.61
145.32	94.04	154.44	123.94
149.93	96.06	159.42	130.15
157.03	99.49	165.15	140.22
159.49	102.43	166.11	143.22
160.88	107.17	164.82	145.66
160.87	107.98	164.35	145.90
160.85	108.79	163.86	146.13

File: addfile5.2_samplings.xlsx
Sheet: fluxome
Page: 1/1

Fluxes in mg/(gDCW h)

Source		At-line			NMR in medium (consumed)								NMR others non-a.a.	
FieldID	t	X	Prod	GlyOH	Ala	Arg	Asp	Glu	Gly	Lys	Pro	Ac	Allantoin	GPC
FullName	Time	Biomass	Product	Glycerol	Alanine	Arginine	Aspartate	Glutamate	Glycine	Lysine	Proline	Acetate	Allantoin	Glyceroph
Units	h			mg/(gDCW.h)								mg/(gDCW.h)		
F1868/14	0	71.42	0.0000	-51.29	-7.02	-1.386	-5.87	0.00	-13.85	0	-0.436	0.000	0.195	0.058
F1868/14	5	72.27	0.0000	-39.88	-4.08	-0.577	-2.97	0.00	-15.92	0	-0.689	0.000	0.224	0.054
F1868/14	22	94.01	0.0096	-65.64	-8.82	-1.323	-4.18	0.00	9.74	0	0.086	0.010	0.077	0.035
F1868/14	30	113.77	0.0134	-60.73	-11.80	-2.903	-11.55	-4.83	-94.19	0	-2.381	0.041	0.073	0.024
F1868/14	46	109.72	0.0207	-355.51	-0.45	-0.235	-1.19	-1.32	-12.24	0	-0.204	0.014	0.071	0.012
F1868/14	50	93.62	0.0241	-253.18	-0.13	0.060	0.03	0.00	-1.20	0	0.126	0.012	0.072	0.011
F1868/14	53	78.05	0.0278	-241.93	-0.04	0.127	0.23	0.28	0.25	0	0.183	0.011	0.087	0.011
F1868/14	70	20.48	0.0351	-64.66	0.01	-0.076	-0.18	-0.16	0.03	0	-0.103	0.007	0.213	0.012
F1868/14	78	11.88	0.0252	-46.26	0.01	-0.101	-0.17	-0.14	-0.01	0	-0.116	0.006	0.144	0.013
F1868/14	94	4.86	0.0154	-25.17	0.01	0.017	0.03	0.02	0.00	0	0.019	0.004	0.049	0.015
F1868/14	98	4.01	0.0150	-23.06	0.01	0.003	0.01	0.00	0.00	0	0.004	0.003	0.037	0.014
F1868/14	102	3.51	0.0140	-22.44	0.01	-0.001	0.00	0.00	0.00	0	-0.001	0.003	0.032	0.014
F1868/14	121	2.31	0.0114	-18.55	0.00	0.000	0.00	0.00	0.00	0	0.000	0.002	0.026	0.010
F1868/14	128	2.33	0.0112	-18.90	0.00	0.000	0.00	0.00	0.00	0	0.000	0.003	0.024	0.009
F1868/14	146	1.87	0.0084	-17.44	0.00	0.000	0.00	0.00	0.00	0	0.000	0.002	0.018	0.007
F1868/14	153	1.75	0.0072	-19.02	0.00	0.000	0.00	0.00	0.00	0	0.000	0.002	0.018	0.007
F1868/14	166	1.59	0.0056	-18.09	0.00	0.000	0.00	0.00	0.00	0	0.000	0.003	0.014	0.007
F1868/14	170	1.52	0.0053	-19.18	0.00	0.000	0.00	0.00	0.00	0	0.000	0.003	0.015	0.008
F1868/14	174	1.48	0.0051	-19.59	0.00	0.000	0.00	0.00	0.00	0	0.000	0.003	0.017	0.008
F1817/14	23	150.47	0.0000	-114.42	0.000	0	0	0	0	0	0	0.028	0.072	0.055
F1817/14	27	150.39	0.0000	-140.33	0.004	0	0	0	0	0	0	0.021	0.076	0.043
F1817/14	30	151.09	0.0001	-215.49	0.049	0	0	0	0	0	0	0.016	0.082	0.034
F1817/14	46	85.15	0.0208	-114.37	0.018	0	0	0	0	0	0	0.006	0.133	0.016
F1817/14	50	62.34	0.0211	-169.24	0.016	0	0	0	0	0	0	0.006	0.149	0.016
F1817/14	54	45.44	0.0214	-121.45	0.015	0	0	0	0	0	0	0.006	0.159	0.017
F1817/14	71	14.80	0.0228	-42.75	0.014	0	0	0	0	0	0	0.005	0.127	0.022
F1817/14	75	13.34	0.0229	-39.14	0.014	0	0	0	0	0	0	0.005	0.117	0.023
F1817/14	78	12.22	0.0225	-36.39	0.014	0	0	0	0	0	0	0.005	0.110	0.024
F1817/14	95	6.26	0.0154	-24.15	0.009	0	0	0	0	0	0	0.004	0.079	0.024
F1817/14	99	5.08	0.0136	-22.72	0.008	0	0	0	0	0	0	0.003	0.076	0.023
F1817/14	102	4.00	0.0123	-21.23	0.008	0	0	0	0	0	0	0.003	0.073	0.023
F1817/14	121	2.52	0.0094	-15.60	0.005	0	0	0	0	0	0	0.002	0.052	0.019
F1817/14	129	2.92	0.0097	-13.87	0.004	0	0	0	0	0	0	0.002	0.042	0.018
F1817/14	146	2.14	0.0087	-12.22	0.003	0	0	0	0	0	0	0.002	0.028	0.016
F1817/14	154	1.67	0.0066	-10.34	0.002	0	0	0	0	0	0	0.002	0.020	0.016
F1817/14	167	1.86	0.0037	-9.92	0.001	0	0	0	0	0	0	0.001	0.006	0.016
F1817/14	170	1.99	0.0035	-9.79	0.001	0	0	0	0	0	0	0.001	0.005	0.016
F1817/14	172	2.04	0.0035	-9.85	0.001	0	0	0	0	0	0	0.001	0.004	0.017

Additional file 5.3 – Elementary flux modes clustered footprint

Description of the 967 elementary flux modes calculated for *Pichia pastoris* core metabolic network.

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 1/14

EFMs clustered footprint

all metabolites in mol, except biomass and product in C-mol

Metabolite	EM	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
	1	-0.2928	-0.0158	0.15253	1	0	-0.0113	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
	2	-0.3475	-0.0514	0.20851	1	0	-0.0113	0	-0.0528	0	0	0	0	0	0	0.02368
	3	-0.3395	-0.0465	0.20044	1	0	-0.0113	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
	4	-0.3402	-0.0158	0.17622	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00129	0	0
	5	-0.352	-0.0514	0.21174	1	0	-0.0113	0	-0.0528	0	0	0	0	0.00129	0	0.02368
	6	-0.3639	-0.0158	0.24727	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00129	0	0
	7	-0.324	-0.0544	0.24786	0	1	-0.0248	0	-0.0867	0	0	0	0	0	0	0.03628
	8	-0.3337	-0.0158	0.17299	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
	9	-0.2928	-0.0158	0.15253	1	0	-0.0113	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
	10	-0.4222	-0.0158	0.30557	1	0	-0.0113	0	0	0	0	0	0	0.00129	0	0
	11	-0.4125	-0.0158	0.29266	1	0	-0.0113	0	0	0	0	0	-0.0032	0.00129	0	0
	12	-0.344	-0.0465	0.20367	1	0	-0.0113	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
	13	-0.3542	-0.0158	0.23435	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
	14	-0.3076	-0.0465	0.23231	1	0	-0.0113	0	-0.0496	0	-0.0637	0	-0.0032	0	0	0.02045
	15	-0.314	-0.0514	0.242	1	0	-0.0113	0	-0.0528	0	-0.067	0	0	0	0	0.02368
	16	-0.3475	-0.0514	0.20851	1	0	-0.0113	0	-0.0528	0	0	0	0	0	0	0.02368
	17	-0.3115	-0.0465	0.23618	1	0	-0.0113	0	-0.0496	0	-0.065	0	-0.0032	0.00129	0	0.02045
	18	-0.3179	-0.0514	0.24587	1	0	-0.0113	0	-0.0528	0	-0.0683	0	0	0.00129	0	0.02368
	19	-0.3402	-0.0158	0.17622	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00129	0	0
	20	-0.2928	-0.0158	0.15253	1	0	-0.0113	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
	21	-0.3395	-0.0465	0.20044	1	0	-0.0113	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
	22	-0.352	-0.0514	0.21174	1	0	-0.0113	0	-0.0528	0	0	0	0	0.00129	0	0.02368
	23	-0.3639	-0.0158	0.24727	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00129	0	0
	24	-0.324	-0.0544	0.24786	0	1	-0.0248	0	-0.0867	0	0	0	0	0	0	0.03628
	25	-0.3337	-0.0158	0.17299	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
	26	-0.2928	-0.0158	0.15253	1	0	-0.0113	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
	27	-0.344	-0.0465	0.20367	1	0	-0.0113	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
	28	-0.4222	-0.0158	0.30557	1	0	-0.0113	0	0	0	0	0	0	0.00129	0	0
	29	-0.314	-0.0514	0.242	1	0	-0.0113	0	-0.0528	0	-0.067	0	0	0	0	0.02368
	30	-0.4125	-0.0158	0.29266	1	0	-0.0113	0	0	0	0	0	-0.0032	0.00129	0	0
	31	-0.3542	-0.0158	0.23435	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
	32	-0.3115	-0.0465	0.23618	1	0	-0.0113	0	-0.0496	0	-0.065	0	-0.0032	0.00129	0	0.02045
	33	-0.288	-0.0158	0.14959	1	0	-0.0113	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
	34	-0.3179	-0.0514	0.24587	1	0	-0.0113	0	-0.0528	0	-0.0683	0	0	0.00129	0	0.02368
	35	-0.3076	-0.0465	0.23231	1	0	-0.0113	0	-0.0496	0	-0.0637	0	-0.0032	0	0	0.02045
	36	-0.2898	-0.0167	0.15075	1	0	-0.0113	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
	37	-0.2886	-0.0158	0.15017	1	0	-0.0113	0	-0.0292	-0.0237	0	0	0	0	0	0
	38	-0.289	-0.0158	0.14979	1	0	-0.0113	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
	39	-0.336	-0.0158	0.17386	1	0	-0.0113	0	-0.0292	0	0	0	0	0	0	0
	40	-0.3364	-0.0158	0.17347	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00019	0	0
	41	-0.3354	-0.0158	0.17328	1	0	-0.0113	0	-0.0292	0	-0.0006	0	0	0	0	0
	42	-0.36	-0.0158	0.24452	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00019	0	0
	43	-0.3591	-0.0158	0.24433	1	0	-0.0113	0	-0.0292	0	-0.0006	0	0	0	0	0
	44	-0.336	-0.0167	0.17386	1	0	-0.0113	0	-0.0297	0	0	0	0	0	0	0.00058
	45	-0.4174	-0.0158	0.30263	1	0	-0.0113	0	0	0	-0.0006	0	0	0	0	0
	46	-0.3591	-0.0167	0.24317	1	0	-0.0113	0	-0.0297	0	0	0	0	0	0	0.00058
	47	-0.3597	-0.0158	0.24491	1	0	-0.0113	0	-0.0292	0	0	0	0	0	0	0
	48	-0.4185	-0.0167	0.30263	1	0	-0.0113	0	0	0	0	0	0	0	0	0.00058
	49	-0.418	-0.0158	0.30321	1	0	-0.0113	0	0	0	0	0	0	0	0	0
	50	-0.4183	-0.0158	0.30283	1	0	-0.0113	0	0	0	0	0	0	0.00019	0	0
	51	-0.2886	-0.0158	0.15017	1	0	-0.0113	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
	52	-0.289	-0.0158	0.14979	1	0	-0.0113	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
	53	-0.288	-0.0158	0.14959	1	0	-0.0113	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
	54	-0.3299	-0.0158	0.17024	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
	55	-0.3289	-0.0158	0.17005	1	0	-0.0113	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
	56	-0.2898	-0.0167	0.15075	1	0	-0.0113	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
	57	-0.3279	-0.0427	0.2393	0.08981	1	-0.0258	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
	58	-0.3295	-0.0167	0.17063	1	0	-0.0113	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
	59	-0.3295	-0.0158	0.17063	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0	0	0
	60	-0.35	-0.0158	0.23199	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0	0	0
	61	-0.3504	-0.0158	0.2316	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
	62	-0.3494	-0.0158	0.23141	1	0	-0.0113	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
	63	-0.4087	-0.0158	0.28991	1	0	-0.0113	0	0	0	0	0	-0.0032	0.00019	0	0
	64	-0.4077	-0.0158	0.28972	1	0	-0.0113	0	0	0	-0.0006	0	-0.0032	0	0	0
	65	-0.3494	-0.0167	0.23025	1	0	-0.0113	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
	66	-0.2196	0	0.17088	0	1	-0.0248	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
	67	-0.4089	-0.0167	0.28972	1	0	-0.0113	0	0	0	0	0	-0.0032	0	0	0.00058

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 2/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
68	-0.4083	-0.0158	0.2903	1	0	-0.0113	0	0	0	0	0	-0.0032	0	0	0
69	-0.3019	-0.0412	0.22576	0	1	-0.0248	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
70	-0.247	0	0.19832	0	1	-0.0248	0	-0.0504	-0.0363	0	0	0	0	0	0
71	-0.3209	-0.0158	0.17141	1	0.12012	-0.0142	0	-0.0352	-0.028	0	0	0	0.00129	0	0
72	-0.3769	-0.0158	0.19945	1	0.12012	-0.0142	0	-0.0352	0	0	0	0	0.00129	0	0
73	-0.2921	0	0.20716	0	1	-0.0248	0	-0.0504	0	-0.0274	0	0	0	0	0
74	-0.3019	-0.0412	0.22576	0	1	-0.0248	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
75	-0.3284	0	0.316	0	1	-0.0248	0	-0.0504	0	-0.0274	0	0	0	0	0
76	-0.3196	-0.0412	0.2346	0	1	-0.0248	0	-0.0779	0	0	0	0	0	0	0.02744
77	-0.3196	0	0.2346	0	1	-0.0248	0	-0.0504	0	0	0	0	0	0	0
78	-0.3284	-0.0412	0.26112	0	1	-0.0248	0	-0.0779	0	0	0	0	0	0	0.02744
79	-0.3558	0	0.34344	0	1	-0.0248	0	-0.0504	0	0	0	0	0	0	0
80	-0.405	-0.0158	0.28358	1	0.12012	-0.0142	0	-0.0352	0	0	0	0	0.00129	0	0
81	-0.37	0	0.18417	0	1	-0.0248	0	0	0	0	0	0	0	0	0
82	-0.4122	-0.0158	0.16424	1	0.12012	-0.0142	0	0	0	0	0	0	0.00129	0	0
83	-0.3426	0	0.15673	0	1	-0.0248	0	0	0	-0.0274	0	0	0	0	0
84	-0.4754	-0.0158	0.354	1	0.12012	-0.0142	0	0	0	0	0	0	0.00129	0	0
85	-0.4293	0	0.41686	0	1	-0.0248	0	0	0	-0.0274	0	0	0	0	0
86	-0.3974	-0.0412	0.15673	0	1	-0.0248	0	0	0	0	0	0	0	0	0.02744
87	-0.2196	0	0.17088	0	1	-0.0248	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
88	-0.4842	-0.0412	0.41686	0	1	-0.0248	0	0	0	0	0	0	0	0	0.02744
89	-0.4567	0	0.4443	0	1	-0.0248	0	0	0	0	0	0	0	0	0
90	-0.3001	-0.0412	0.22219	0	1	-0.0248	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
91	-0.247	0	0.19832	0	1	-0.0248	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
92	-0.3209	-0.0158	0.17141	1	0.12012	-0.0142	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
93	-0.2983	0	0.22398	0	1	-0.0248	0	-0.0504	0	0	0	-0.0106	0	0	0
94	-0.2965	0	0.27351	0	1	-0.0248	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
95	-0.3679	-0.0158	0.19495	1	0.12012	-0.0142	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
96	-0.2709	0	0.19654	0	1	-0.0248	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
97	-0.3233	-0.042	0.23661	0.07515	1	-0.0256	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
98	-0.3915	-0.0158	0.26556	1	0.12012	-0.0142	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
99	-0.2965	-0.0385	0.22219	0	1	-0.0248	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
100	-0.3213	0	0.14611	0	1	-0.0248	0	0	0	-0.0274	0	-0.0106	0	0	0
101	-0.2983	-0.0385	0.22398	0	1	-0.0248	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
102	-0.324	0	0.30095	0	1	-0.0248	0	-0.0504	0	0	0	-0.0106	0	0	0
103	-0.3762	-0.0412	0.14611	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0.02744
104	-0.3488	0	0.17355	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0
105	-0.4031	-0.0158	0.15974	1	0.12012	-0.0142	0	0	0	0	0	-0.0045	0.00129	0	0
106	-0.4248	0	0.40181	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0
107	-0.4619	-0.0158	0.33598	1	0.12012	-0.0142	0	0	0	0	0	-0.0045	0.00129	0	0
108	-0.3974	0	0.37437	0	1	-0.0248	0	0	0	-0.0274	0	-0.0106	0	0	0
109	-0.289	-0.0158	0.14979	1	0	-0.0113	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
110	-0.288	-0.0158	0.14959	1	0	-0.0113	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
111	-0.4523	-0.0412	0.37437	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0.02744
112	-0.3354	-0.0158	0.17328	1	0	-0.0113	0	-0.0292	0	-0.0006	0	0	0	0	0
113	-0.2898	-0.0167	0.15075	1	0	-0.0113	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
114	-0.2886	-0.0158	0.15017	1	0	-0.0113	0	-0.0292	-0.0237	0	0	0	0	0	0
115	-0.336	-0.0167	0.17386	1	0	-0.0113	0	-0.0297	0	0	0	0	0	0	0.00058
116	-0.336	-0.0158	0.17386	1	0	-0.0113	0	-0.0292	0	0	0	0	0	0	0
117	-0.3364	-0.0158	0.17347	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00019	0	0
118	-0.3597	-0.0158	0.24491	1	0	-0.0113	0	-0.0292	0	0	0	0	0	0	0
119	-0.36	-0.0158	0.24452	1	0	-0.0113	0	-0.0292	0	0	0	0	0.00019	0	0
120	-0.3591	-0.0158	0.24433	1	0	-0.0113	0	-0.0292	0	-0.0006	0	0	0	0	0
121	-0.4183	-0.0158	0.30283	1	0	-0.0113	0	0	0	0	0	0	0.00019	0	0
122	-0.4174	-0.0158	0.30263	1	0	-0.0113	0	0	0	-0.0006	0	0	0	0	0
123	-0.3591	-0.0167	0.24317	1	0	-0.0113	0	-0.0297	0	0	0	0	0	0	0.00058
124	-0.288	-0.0158	0.14959	1	0	-0.0113	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
125	-0.4185	-0.0167	0.30263	1	0	-0.0113	0	0	0	0	0	0	0	0	0.00058
126	-0.418	-0.0158	0.30321	1	0	-0.0113	0	0	0	0	0	0	0	0	0
127	-0.2898	-0.0167	0.15075	1	0	-0.0113	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
128	-0.2886	-0.0158	0.15017	1	0	-0.0113	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
129	-0.289	-0.0158	0.14979	1	0	-0.0113	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
130	-0.3295	-0.0158	0.17063	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0	0	0
131	-0.3299	-0.0158	0.17024	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
132	-0.3289	-0.0158	0.17005	1	0	-0.0113	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
133	-0.3494	-0.0158	0.23141	1	0	-0.0113	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
134	-0.3279	-0.0427	0.2393	0.08981	1	-0.0258	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
135	-0.3295	-0.0167	0.17063	1	0	-0.0113	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
136	-0.3494	-0.0167	0.23025	1	0	-0.0113	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
137	-0.35	-0.0158	0.23199	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0	0	0
138	-0.3504	-0.0158	0.2316	1	0	-0.0113	0	-0.0292	0	0	0	-0.0032	0.00019	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 3/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
139	-0.4083	-0.0158	0.2903	1	0	-0.0113	0	0	0	0	0	-0.0032	0	0	0
140	-0.4087	-0.0158	0.28991	1	0	-0.0113	0	0	0	0	0	-0.0032	0.00019	0	0
141	-0.4077	-0.0158	0.28972	1	0	-0.0113	0	0	0	-0.0006	0	-0.0032	0	0	0
142	-0.3209	-0.0158	0.17141	1	0.12012	-0.0142	0	-0.0352	-0.028	0	0	0	0.00129	0	0
143	-0.2196	0	0.17088	0	1	-0.0248	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
144	-0.4089	-0.0167	0.28972	1	0	-0.0113	0	0	0	0	0	-0.0032	0	0	0.00058
145	-0.3019	-0.0412	0.22576	0	1	-0.0248	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
146	-0.3019	-0.0412	0.22576	0	1	-0.0248	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
147	-0.247	0	0.19832	0	1	-0.0248	0	-0.0504	-0.0363	0	0	0	0	0	0
148	-0.3196	0	0.2346	0	1	-0.0248	0	-0.0504	0	0	0	0	0	0	0
149	-0.3769	-0.0158	0.19945	1	0.12012	-0.0142	0	-0.0352	0	0	0	0	0.00129	0	0
150	-0.2921	0	0.20716	0	1	-0.0248	0	-0.0504	0	-0.0274	0	0	0	0	0
151	-0.405	-0.0158	0.28358	1	0.12012	-0.0142	0	-0.0352	0	0	0	0	0.00129	0	0
152	-0.3284	0	0.316	0	1	-0.0248	0	-0.0504	0	-0.0274	0	0	0	0	0
153	-0.3196	-0.0412	0.2346	0	1	-0.0248	0	-0.0779	0	0	0	0	0	0	0.02744
154	-0.3426	0	0.15673	0	1	-0.0248	0	0	0	-0.0274	0	0	0	0	0
155	-0.3284	-0.0412	0.26112	0	1	-0.0248	0	-0.0779	0	0	0	0	0	0	0.02744
156	-0.3558	0	0.34344	0	1	-0.0248	0	-0.0504	0	0	0	0	0	0	0
157	-0.3974	-0.0412	0.15673	0	1	-0.0248	0	0	0	0	0	0	0	0	0.02744
158	-0.37	0	0.18417	0	1	-0.0248	0	0	0	0	0	0	0	0	0
159	-0.4122	-0.0158	0.16424	1	0.12012	-0.0142	0	0	0	0	0	0	0.00129	0	0
160	-0.4567	0	0.4443	0	1	-0.0248	0	0	0	0	0	0	0	0	0
161	-0.4754	-0.0158	0.354	1	0.12012	-0.0142	0	0	0	0	0	0	0.00129	0	0
162	-0.4293	0	0.41686	0	1	-0.0248	0	0	0	-0.0274	0	0	0	0	0
163	-0.2196	0	0.17088	0	1	-0.0248	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
164	-0.4842	-0.0412	0.41686	0	1	-0.0248	0	0	0	0	0	0	0	0	0.02744
165	-0.3001	-0.0412	0.22219	0	1	-0.0248	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
166	-0.247	0	0.19832	0	1	-0.0248	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
167	-0.3209	-0.0158	0.17141	1	0.12012	-0.0142	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
168	-0.2983	0	0.22398	0	1	-0.0248	0	-0.0504	0	0	0	-0.0106	0	0	0
169	-0.3679	-0.0158	0.19495	1	0.12012	-0.0142	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
170	-0.2709	0	0.19654	0	1	-0.0248	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
171	-0.3915	-0.0158	0.26556	1	0.12012	-0.0142	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
172	-0.2965	-0.0385	0.22219	0	1	-0.0248	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
173	-0.2965	0	0.27351	0	1	-0.0248	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
174	-0.2983	-0.0385	0.22398	0	1	-0.0248	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
175	-0.324	0	0.30095	0	1	-0.0248	0	-0.0504	0	0	0	-0.0106	0	0	0
176	-0.3233	-0.042	0.23661	0.07515	1	-0.0256	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
177	-0.3488	0	0.17355	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0
178	-0.4031	-0.0158	0.15974	1	0.12012	-0.0142	0	0	0	0	0	-0.0045	0.00129	0	0
179	-0.3213	0	0.14611	0	1	-0.0248	0	0	0	-0.0274	0	-0.0106	0	0	0
180	-0.4619	-0.0158	0.33598	1	0.12012	-0.0142	0	0	0	0	0	-0.0045	0.00129	0	0
181	-0.3974	0	0.37437	0	1	-0.0248	0	0	0	-0.0274	0	-0.0106	0	0	0
182	-0.3762	-0.0412	0.14611	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0.02744
183	-0.3449	-0.0514	0.20851	1	0	-0.0139	0	-0.0528	0	0	0	0	0	0	0.02368
184	-0.4523	-0.0412	0.37437	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0.02744
185	-0.4248	0	0.40181	0	1	-0.0248	0	0	0	0	0	-0.0106	0	0	0
186	-0.3375	-0.0158	0.17622	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00129	0	0
187	-0.2902	-0.0158	0.15253	1	0	-0.0139	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
188	-0.3368	-0.0465	0.20044	1	0	-0.0139	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
189	-0.3494	-0.0514	0.21174	1	0	-0.0139	0	-0.0528	0	0	0	0	0.00129	0	0.02368
190	-0.3612	-0.0158	0.24727	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00129	0	0
191	-0.3142	-0.0544	0.24786	0	1	-0.0345	0	-0.0867	0	0	0	0	0	0	0.03628
192	-0.3311	-0.0158	0.17299	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
193	-0.2902	-0.0158	0.15253	1	0	-0.0139	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
194	-0.4195	-0.0158	0.30557	1	0	-0.0139	0	0	0	0	0	0	0.00129	0	0
195	-0.4098	-0.0158	0.29266	1	0	-0.0139	0	0	0	0	0	-0.0032	0.00129	0	0
196	-0.3413	-0.0465	0.20367	1	0	-0.0139	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
197	-0.3515	-0.0158	0.23435	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
198	-0.3166	-0.0514	0.24454	1	0	-0.0139	0	-0.0528	0	-0.0656	0	0	0.00129	0	0.02368
199	-0.3063	-0.0465	0.23097	1	0	-0.0139	0	-0.0496	0	-0.0611	0	-0.0032	0	0	0.02045
200	-0.3127	-0.0514	0.24066	1	0	-0.0139	0	-0.0528	0	-0.0643	0	0	0	0	0.02368
201	-0.3368	-0.0465	0.20044	1	0	-0.0139	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
202	-0.3449	-0.0514	0.20851	1	0	-0.0139	0	-0.0528	0	0	0	0	0	0	0.02368
203	-0.3101	-0.0465	0.23485	1	0	-0.0139	0	-0.0496	0	-0.0624	0	-0.0032	0.00129	0	0.02045
204	-0.3142	-0.0544	0.24786	0	1	-0.0345	0	-0.0867	0	0	0	0	0	0	0.03628
205	-0.3375	-0.0158	0.17622	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00129	0	0
206	-0.2902	-0.0158	0.15253	1	0	-0.0139	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
207	-0.4195	-0.0158	0.30557	1	0	-0.0139	0	0	0	0	0	0	0.00129	0	0
208	-0.3494	-0.0514	0.21174	1	0	-0.0139	0	-0.0528	0	0	0	0	0.00129	0	0.02368
209	-0.3311	-0.0158	0.17299	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00129	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 4/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
210	-0.3515	-0.0158	0.23435	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
211	-0.3612	-0.0158	0.24727	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00129	0	0
212	-0.4098	-0.0158	0.29266	1	0	-0.0139	0	0	0	0	0	-0.0032	0.00129	0	0
213	-0.3127	-0.0514	0.24066	1	0	-0.0139	0	-0.0528	0	-0.0643	0	0	0	0	0.02368
214	-0.3413	-0.0465	0.20367	1	0	-0.0139	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
215	-0.2902	-0.0158	0.15253	1	0	-0.0139	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
216	-0.3101	-0.0465	0.23485	1	0	-0.0139	0	-0.0496	0	-0.0624	0	-0.0032	0.00129	0	0.02045
217	-0.3166	-0.0514	0.24454	1	0	-0.0139	0	-0.0528	0	-0.0656	0	0	0.00129	0	0.02368
218	-0.3063	-0.0465	0.23097	1	0	-0.0139	0	-0.0496	0	-0.0611	0	-0.0032	0	0	0.02045
219	-0.2859	-0.0158	0.15017	1	0	-0.0139	0	-0.0292	-0.0237	0	0	0	0	0	0
220	-0.2863	-0.0158	0.14979	1	0	-0.0139	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
221	-0.2854	-0.0158	0.14959	1	0	-0.0139	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
222	-0.3337	-0.0158	0.17347	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00019	0	0
223	-0.3327	-0.0158	0.17328	1	0	-0.0139	0	-0.0292	0	-0.0006	0	0	0	0	0
224	-0.2871	-0.0167	0.15075	1	0	-0.0139	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
225	-0.3564	-0.0158	0.24433	1	0	-0.0139	0	-0.0292	0	-0.0006	0	0	0	0	0
226	-0.3333	-0.0167	0.17386	1	0	-0.0139	0	-0.0297	0	0	0	0	0	0	0.00058
227	-0.3333	-0.0158	0.17386	1	0	-0.0139	0	-0.0292	0	0	0	0	0	0	0
228	-0.3564	-0.0167	0.24317	1	0	-0.0139	0	-0.0297	0	0	0	0	0	0	0.00058
229	-0.357	-0.0158	0.24491	1	0	-0.0139	0	-0.0292	0	0	0	0	0	0	0
230	-0.3574	-0.0158	0.24452	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00019	0	0
231	-0.4153	-0.0158	0.30321	1	0	-0.0139	0	0	0	0	0	0	0	0	0
232	-0.4157	-0.0158	0.30283	1	0	-0.0139	0	0	0	0	0	0	0.00019	0	0
233	-0.4147	-0.0158	0.30263	1	0	-0.0139	0	0	0	-0.0006	0	0	0	0	0
234	-0.2863	-0.0158	0.14979	1	0	-0.0139	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
235	-0.2854	-0.0158	0.14959	1	0	-0.0139	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
236	-0.4159	-0.0167	0.30263	1	0	-0.0139	0	0	0	0	0	0	0	0	0.00058
237	-0.3263	-0.0158	0.17005	1	0	-0.0139	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
238	-0.2871	-0.0167	0.15075	1	0	-0.0139	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
239	-0.2859	-0.0158	0.15017	1	0	-0.0139	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
240	-0.3268	-0.0167	0.17063	1	0	-0.0139	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
241	-0.3268	-0.0158	0.17063	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0	0	0
242	-0.3272	-0.0158	0.17024	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
243	-0.3477	-0.0158	0.2316	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
244	-0.3467	-0.0158	0.23141	1	0	-0.0139	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
245	-0.3179	-0.0427	0.2393	0.08981	1	-0.0358	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
246	-0.405	-0.0158	0.28972	1	0	-0.0139	0	0	0	-0.0006	0	-0.0032	0	0	0
247	-0.3467	-0.0167	0.23025	1	0	-0.0139	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
248	-0.3473	-0.0158	0.23199	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0	0	0
249	-0.4062	-0.0167	0.28972	1	0	-0.0139	0	0	0	0	0	-0.0032	0	0	0.00058
250	-0.4056	-0.0158	0.2903	1	0	-0.0139	0	0	0	0	0	-0.0032	0	0	0
251	-0.406	-0.0158	0.28991	1	0	-0.0139	0	0	0	0	0	-0.0032	0.00019	0	0
252	-0.2373	0	0.19832	0	1	-0.0345	0	-0.0504	-0.0363	0	0	0	0	0	0
253	-0.317	-0.0158	0.17141	1	0.12012	-0.0181	0	-0.0352	-0.028	0	0	0	0.00129	0	0
254	-0.2098	0	0.17088	0	1	-0.0345	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
255	-0.2824	0	0.20716	0	1	-0.0345	0	-0.0504	0	-0.0274	0	0	0	0	0
256	-0.2922	-0.0412	0.22576	0	1	-0.0345	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
257	-0.2922	-0.0412	0.22576	0	1	-0.0345	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
258	-0.3098	-0.0412	0.2346	0	1	-0.0345	0	-0.0779	0	0	0	0	0	0	0.02744
259	-0.3098	0	0.2346	0	1	-0.0345	0	-0.0504	0	0	0	0	0	0	0
260	-0.3731	-0.0158	0.19945	1	0.12012	-0.0181	0	-0.0352	0	0	0	0	0.00129	0	0
261	-0.3461	0	0.34344	0	1	-0.0345	0	-0.0504	0	0	0	0	0	0	0
262	-0.4011	-0.0158	0.28358	1	0.12012	-0.0181	0	-0.0352	0	0	0	0	0.00129	0	0
263	-0.3187	0	0.316	0	1	-0.0345	0	-0.0504	0	-0.0274	0	0	0	0	0
264	-0.4083	-0.0158	0.16424	1	0.12012	-0.0181	0	0	0	0	0	0	0.00129	0	0
265	-0.3328	0	0.15673	0	1	-0.0345	0	0	0	-0.0274	0	0	0	0	0
266	-0.3187	-0.0412	0.26112	0	1	-0.0345	0	-0.0779	0	0	0	0	0	0	0.02744
267	-0.4195	0	0.41686	0	1	-0.0345	0	0	0	-0.0274	0	0	0	0	0
268	-0.3877	-0.0412	0.15673	0	1	-0.0345	0	0	0	0	0	0	0	0	0.02744
269	-0.3603	0	0.18417	0	1	-0.0345	0	0	0	0	0	0	0	0	0
270	-0.4744	-0.0412	0.41686	0	1	-0.0345	0	0	0	0	0	0	0	0	0.02744
271	-0.447	0	0.4443	0	1	-0.0345	0	0	0	0	0	0	0	0	0
272	-0.4716	-0.0158	0.354	1	0.12012	-0.0181	0	0	0	0	0	0	0.00129	0	0
273	-0.2373	0	0.19832	0	1	-0.0345	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
274	-0.317	-0.0158	0.17141	1	0.12012	-0.0181	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
275	-0.2098	0	0.17088	0	1	-0.0345	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
276	-0.3641	-0.0158	0.19495	1	0.12012	-0.0181	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
277	-0.2611	0	0.19654	0	1	-0.0345	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
278	-0.2904	-0.0412	0.22219	0	1	-0.0345	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
279	-0.2868	-0.0385	0.22219	0	1	-0.0345	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
280	-0.2868	0	0.27351	0	1	-0.0345	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 5/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
281	-0.2886	0	0.22398	0	1	-0.0345	0	-0.0504	0	0	0	-0.0106	0	0	0
282	-0.3142	0	0.30095	0	1	-0.0345	0	-0.0504	0	0	0	-0.0106	0	0	0
283	-0.3133	-0.042	0.23661	0.07515	1	-0.0356	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
284	-0.3876	-0.0158	0.26556	1	0.12012	-0.0181	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
285	-0.3993	-0.0158	0.15974	1	0.12012	-0.0181	0	0	0	0	0	-0.0045	0.00129	0	0
286	-0.3116	0	0.14611	0	1	-0.0345	0	0	0	-0.0274	0	-0.0106	0	0	0
287	-0.2886	-0.0385	0.22398	0	1	-0.0345	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
288	-0.3877	0	0.37437	0	1	-0.0345	0	0	0	-0.0274	0	-0.0106	0	0	0
289	-0.3665	-0.0412	0.14611	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0.02744
290	-0.339	0	0.17355	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0
291	-0.4425	-0.0412	0.37437	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0.02744
292	-0.4151	0	0.40181	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0
293	-0.4581	-0.0158	0.33598	1	0.12012	-0.0181	0	0	0	0	0	-0.0045	0.00129	0	0
294	-0.2859	-0.0158	0.15017	1	0	-0.0139	0	-0.0292	-0.0237	0	0	0	0	0	0
295	-0.2863	-0.0158	0.14979	1	0	-0.0139	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
296	-0.2854	-0.0158	0.14959	1	0	-0.0139	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
297	-0.3337	-0.0158	0.17347	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00019	0	0
298	-0.3327	-0.0158	0.17328	1	0	-0.0139	0	-0.0292	0	-0.0006	0	0	0	0	0
299	-0.2871	-0.0167	0.15075	1	0	-0.0139	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
300	-0.3564	-0.0158	0.24433	1	0	-0.0139	0	-0.0292	0	-0.0006	0	0	0	0	0
301	-0.3333	-0.0167	0.17386	1	0	-0.0139	0	-0.0297	0	0	0	0	0	0	0.00058
302	-0.3333	-0.0158	0.17386	1	0	-0.0139	0	-0.0292	0	0	0	0	0	0	0
303	-0.3564	-0.0167	0.24317	1	0	-0.0139	0	-0.0297	0	0	0	0	0	0	0.00058
304	-0.357	-0.0158	0.24491	1	0	-0.0139	0	-0.0292	0	0	0	0	0	0	0
305	-0.3574	-0.0158	0.24452	1	0	-0.0139	0	-0.0292	0	0	0	0	0.00019	0	0
306	-0.4153	-0.0158	0.30321	1	0	-0.0139	0	0	0	0	0	0	0	0	0
307	-0.4157	-0.0158	0.30283	1	0	-0.0139	0	0	0	0	0	0	0.00019	0	0
308	-0.4147	-0.0158	0.30263	1	0	-0.0139	0	0	0	-0.0006	0	0	0	0	0
309	-0.2863	-0.0158	0.14979	1	0	-0.0139	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
310	-0.2854	-0.0158	0.14959	1	0	-0.0139	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
311	-0.4159	-0.0167	0.30263	1	0	-0.0139	0	0	0	0	0	0	0	0	0.00058
312	-0.3263	-0.0158	0.17005	1	0	-0.0139	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
313	-0.2871	-0.0167	0.15075	1	0	-0.0139	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
314	-0.2859	-0.0158	0.15017	1	0	-0.0139	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
315	-0.3268	-0.0167	0.17063	1	0	-0.0139	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
316	-0.3268	-0.0158	0.17063	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0	0	0
317	-0.3272	-0.0158	0.17024	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
318	-0.3477	-0.0158	0.2316	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
319	-0.3467	-0.0158	0.23141	1	0	-0.0139	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
320	-0.3179	-0.0427	0.2393	0.08981	1	-0.0358	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
321	-0.405	-0.0158	0.28972	1	0	-0.0139	0	0	0	-0.0006	0	-0.0032	0	0	0
322	-0.3467	-0.0167	0.23025	1	0	-0.0139	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
323	-0.3473	-0.0158	0.23199	1	0	-0.0139	0	-0.0292	0	0	0	-0.0032	0	0	0
324	-0.4062	-0.0167	0.28972	1	0	-0.0139	0	0	0	0	0	-0.0032	0	0	0.00058
325	-0.4056	-0.0158	0.2903	1	0	-0.0139	0	0	0	0	0	-0.0032	0	0	0
326	-0.406	-0.0158	0.28991	1	0	-0.0139	0	0	0	0	0	-0.0032	0.00019	0	0
327	-0.2373	0	0.19832	0	1	-0.0345	0	-0.0504	-0.0363	0	0	0	0	0	0
328	-0.317	-0.0158	0.17141	1	0.12012	-0.0181	0	-0.0352	-0.028	0	0	0	0.00129	0	0
329	-0.2098	0	0.17088	0	1	-0.0345	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
330	-0.2824	0	0.20716	0	1	-0.0345	0	-0.0504	0	-0.0274	0	0	0	0	0
331	-0.2922	-0.0412	0.22576	0	1	-0.0345	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
332	-0.2922	-0.0412	0.22576	0	1	-0.0345	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
333	-0.3098	-0.0412	0.2346	0	1	-0.0345	0	-0.0779	0	0	0	0	0	0	0.02744
334	-0.3098	0	0.2346	0	1	-0.0345	0	-0.0504	0	0	0	0	0	0	0
335	-0.3731	-0.0158	0.19945	1	0.12012	-0.0181	0	-0.0352	0	0	0	0	0.00129	0	0
336	-0.3461	0	0.34344	0	1	-0.0345	0	-0.0504	0	0	0	0	0	0	0
337	-0.4011	-0.0158	0.28358	1	0.12012	-0.0181	0	-0.0352	0	0	0	0	0.00129	0	0
338	-0.3187	0	0.316	0	1	-0.0345	0	-0.0504	0	-0.0274	0	0	0	0	0
339	-0.4083	-0.0158	0.16424	1	0.12012	-0.0181	0	0	0	0	0	0	0.00129	0	0
340	-0.3328	0	0.15673	0	1	-0.0345	0	0	0	-0.0274	0	0	0	0	0
341	-0.3187	-0.0412	0.26112	0	1	-0.0345	0	-0.0779	0	0	0	0	0	0	0.02744
342	-0.4195	0	0.41686	0	1	-0.0345	0	0	0	-0.0274	0	0	0	0	0
343	-0.3877	-0.0412	0.15673	0	1	-0.0345	0	0	0	0	0	0	0	0	0.02744
344	-0.3603	0	0.18417	0	1	-0.0345	0	0	0	0	0	0	0	0	0
345	-0.4744	-0.0412	0.41686	0	1	-0.0345	0	0	0	0	0	0	0	0	0.02744
346	-0.447	0	0.4443	0	1	-0.0345	0	0	0	0	0	0	0	0	0
347	-0.4716	-0.0158	0.354	1	0.12012	-0.0181	0	0	0	0	0	0	0.00129	0	0
348	-0.2373	0	0.19832	0	1	-0.0345	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
349	-0.317	-0.0158	0.17141	1	0.12012	-0.0181	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
350	-0.2098	0	0.17088	0	1	-0.0345	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
351	-0.3641	-0.0158	0.19495	1	0.12012	-0.0181	0	-0.0352	0	0	0	-0.0045	0.00129	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 6/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
352	-0.2611	0	0.19654	0	1	-0.0345	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
353	-0.2904	-0.0412	0.22219	0	1	-0.0345	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
354	-0.2868	-0.0385	0.22219	0	1	-0.0345	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
355	-0.2868	0	0.27351	0	1	-0.0345	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
356	-0.2886	0	0.22398	0	1	-0.0345	0	-0.0504	0	0	0	-0.0106	0	0	0
357	-0.3142	0	0.30095	0	1	-0.0345	0	-0.0504	0	0	0	-0.0106	0	0	0
358	-0.3133	-0.042	0.23661	0.07515	1	-0.0356	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
359	-0.3876	-0.0158	0.26556	1	0.12012	-0.0181	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
360	-0.3993	-0.0158	0.15974	1	0.12012	-0.0181	0	0	0	0	0	-0.0045	0.00129	0	0
361	-0.3116	0	0.14611	0	1	-0.0345	0	0	0	-0.0274	0	-0.0106	0	0	0
362	-0.2886	-0.0385	0.22398	0	1	-0.0345	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
363	-0.3877	0	0.37437	0	1	-0.0345	0	0	0	-0.0274	0	-0.0106	0	0	0
364	-0.3665	-0.0412	0.14611	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0.02744
365	-0.339	0	0.17355	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0
366	-0.4425	-0.0412	0.37437	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0.02744
367	-0.4151	0	0.40181	0	1	-0.0345	0	0	0	0	0	-0.0106	0	0	0
368	-0.4581	-0.0158	0.33598	1	0.12012	-0.0181	0	0	0	0	0	-0.0045	0.00129	0	0
369	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
370	-0.3507	-0.0465	0.20044	1	0	0	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
371	-0.3588	-0.0514	0.20851	1	0	0	0	-0.0528	0	0	0	0	0	0	0.02368
372	-0.3633	-0.0514	0.21174	1	0	0	0	-0.0528	0	0	0	0	0.00129	0	0.02368
373	-0.3752	-0.0158	0.24727	1	0	0	0	-0.0292	0	0	0	0	0.00129	0	0
374	-0.3488	-0.0544	0.24786	0	1	0	0	-0.0867	0	0	0	0	0	0	0.03628
375	-0.3655	-0.0158	0.23435	1	0	0	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
376	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
377	-0.4335	-0.0158	0.30557	1	0	0	0	0	0	0	0	0	0.00129	0	0
378	-0.3197	-0.0514	0.24763	1	0	0	0	-0.0528	0	-0.0782	0	0	0	0	0.02368
379	-0.4238	-0.0158	0.29266	1	0	0	0	0	0	0	0	-0.0032	0.00129	0	0
380	-0.3552	-0.0465	0.20367	1	0	0	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
381	-0.3171	-0.0465	0.24182	1	0	0	0	-0.0496	0	-0.0763	0	-0.0032	0.00129	0	0.02045
382	-0.3236	-0.0514	0.25151	1	0	0	0	-0.0528	0	-0.0795	0	0	0.00129	0	0.02368
383	-0.3132	-0.0465	0.23794	1	0	0	0	-0.0496	0	-0.075	0	-0.0032	0	0	0.02045
384	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
385	-0.3507	-0.0465	0.20044	1	0	0	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
386	-0.3588	-0.0514	0.20851	1	0	0	0	-0.0528	0	0	0	0	0	0	0.02368
387	-0.3633	-0.0514	0.21174	1	0	0	0	-0.0528	0	0	0	0	0.00129	0	0.02368
388	-0.4335	-0.0158	0.30557	1	0	0	0	0	0	0	0	0	0.00129	0	0
389	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
390	-0.3752	-0.0158	0.24727	1	0	0	0	-0.0292	0	0	0	0	0.00129	0	0
391	-0.3488	-0.0544	0.24786	0	1	0	0	-0.0867	0	0	0	0	0	0	0.03628
392	-0.4238	-0.0158	0.29266	1	0	0	0	0	0	0	0	-0.0032	0.00129	0	0
393	-0.3197	-0.0514	0.24763	1	0	0	0	-0.0528	0	-0.0782	0	0	0	0	0.02368
394	-0.3132	-0.0465	0.23794	1	0	0	0	-0.0496	0	-0.075	0	-0.0032	0	0	0.02045
395	-0.3552	-0.0465	0.20367	1	0	0	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
396	-0.3655	-0.0158	0.23435	1	0	0	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
397	-0.2993	-0.0158	0.14959	1	0	0	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
398	-0.3171	-0.0465	0.24182	1	0	0	0	-0.0496	0	-0.0763	0	-0.0032	0.00129	0	0.02045
399	-0.3236	-0.0514	0.25151	1	0	0	0	-0.0528	0	-0.0795	0	0	0.00129	0	0.02368
400	-0.301	-0.0167	0.15075	1	0	0	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
401	-0.2999	-0.0158	0.15017	1	0	0	0	-0.0292	-0.0237	0	0	0	0	0	0
402	-0.3003	-0.0158	0.14979	1	0	0	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
403	-0.3709	-0.0158	0.24491	1	0	0	0	-0.0292	0	0	0	0	0	0	0
404	-0.3713	-0.0158	0.24452	1	0	0	0	-0.0292	0	0	0	0	0.00019	0	0
405	-0.3703	-0.0158	0.24433	1	0	0	0	-0.0292	0	-0.0006	0	0	0	0	0
406	-0.4296	-0.0158	0.30283	1	0	0	0	0	0	0	0	0	0.00019	0	0
407	-0.4287	-0.0158	0.30263	1	0	0	0	0	0	-0.0006	0	0	0	0	0
408	-0.3703	-0.0167	0.24317	1	0	0	0	-0.0297	0	0	0	0	0	0	0.00058
409	-0.2993	-0.0158	0.14959	1	0	0	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
410	-0.4298	-0.0167	0.30263	1	0	0	0	0	0	0	0	0	0	0	0.00058
411	-0.4292	-0.0158	0.30321	1	0	0	0	0	0	0	0	0	0	0	0
412	-0.301	-0.0167	0.15075	1	0	0	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
413	-0.2999	-0.0158	0.15017	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
414	-0.3003	-0.0158	0.14979	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
415	-0.3616	-0.0158	0.2316	1	0	0	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
416	-0.3607	-0.0158	0.23141	1	0	0	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
417	-0.3537	-0.0427	0.2393	0.08981	1	0	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
418	-0.419	-0.0158	0.28972	1	0	0	0	0	0	-0.0006	0	-0.0032	0	0	0
419	-0.3607	-0.0167	0.23025	1	0	0	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
420	-0.3612	-0.0158	0.23199	1	0	0	0	-0.0292	0	0	0	-0.0032	0	0	0
421	-0.4201	-0.0167	0.28972	1	0	0	0	0	0	0	0	-0.0032	0	0	0.00058
422	-0.4195	-0.0158	0.2903	1	0	0	0	0	0	0	0	-0.0032	0	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 7/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
423	-0.4199	-0.0158	0.28991	1	0	0	0	0	0	0	0	-0.0032	0.00019	0	0
424	-0.2718	0	0.19832	0	1	0	0	-0.0504	-0.0363	0	0	0	0	0	0
425	-0.3351	-0.0158	0.17141	1	0.12012	0	0	-0.0352	-0.028	0	0	0	0.00129	0	0
426	-0.2443	0	0.17088	0	1	0	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
427	-0.3169	0	0.20716	0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
428	-0.3267	-0.0412	0.22576	0	1	0	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
429	-0.3267	-0.0412	0.22576	0	1	0	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
430	-0.3443	-0.0412	0.2346	0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
431	-0.3443	0	0.2346	0	1	0	0	-0.0504	0	0	0	0	0	0	0
432	-0.3912	-0.0158	0.19945	1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
433	-0.3806	0	0.34344	0	1	0	0	-0.0504	0	0	0	0	0	0	0
434	-0.4192	-0.0158	0.28358	1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
435	-0.3532	0	0.316	0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
436	-0.4264	-0.0158	0.16424	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
437	-0.3673	0	0.15673	0	1	0	0	0	0	-0.0274	0	0	0	0	0
438	-0.3532	-0.0412	0.26112	0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
439	-0.454	0	0.41686	0	1	0	0	0	0	-0.0274	0	0	0	0	0
440	-0.4222	-0.0412	0.15673	0	1	0	0	0	0	0	0	0	0	0	0.02744
441	-0.3948	0	0.18417	0	1	0	0	0	0	0	0	0	0	0	0
442	-0.5089	-0.0412	0.41686	0	1	0	0	0	0	0	0	0	0	0	0.02744
443	-0.4815	0	0.4443	0	1	0	0	0	0	0	0	0	0	0	0
444	-0.4897	-0.0158	0.354	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
445	-0.2718	0	0.19832	0	1	0	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
446	-0.3351	-0.0158	0.17141	1	0.12012	0	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
447	-0.2443	0	0.17088	0	1	0	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
448	-0.3822	-0.0158	0.19495	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
449	-0.3231	0	0.22398	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
450	-0.2957	0	0.19654	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
451	-0.3249	-0.0412	0.22219	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
452	-0.4057	-0.0158	0.26556	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
453	-0.3213	-0.0385	0.22219	0	1	0	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
454	-0.3488	0	0.30095	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
455	-0.3213	0	0.27351	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
456	-0.3231	-0.0385	0.22398	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
457	-0.3489	-0.042	0.23661	0.07515	1	0	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
458	-0.3735	0	0.17355	0	1	0	0	0	0	0	0	-0.0106	0	0	0
459	-0.4174	-0.0158	0.15974	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0
460	-0.3461	0	0.14611	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
461	-0.4761	-0.0158	0.33598	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0
462	-0.4222	0	0.37437	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
463	-0.401	-0.0412	0.14611	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
464	-0.2993	-0.0158	0.14959	1	0	0	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
465	-0.4771	-0.0412	0.37437	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
466	-0.4496	0	0.40181	0	1	0	0	0	0	0	0	-0.0106	0	0	0
467	-0.301	-0.0167	0.15075	1	0	0	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
468	-0.2999	-0.0158	0.15017	1	0	0	0	-0.0292	-0.0237	0	0	0	0	0	0
469	-0.3003	-0.0158	0.14979	1	0	0	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
470	-0.3709	-0.0158	0.24491	1	0	0	0	-0.0292	0	0	0	0	0	0	0
471	-0.3713	-0.0158	0.24452	1	0	0	0	-0.0292	0	0	0	0	0.00019	0	0
472	-0.3703	-0.0158	0.24433	1	0	0	0	-0.0292	0	-0.0006	0	0	0	0	0
473	-0.4296	-0.0158	0.30283	1	0	0	0	0	0	0	0	0	0.00019	0	0
474	-0.4287	-0.0158	0.30263	1	0	0	0	0	0	-0.0006	0	0	0	0	0
475	-0.3703	-0.0167	0.24317	1	0	0	0	-0.0297	0	0	0	0	0	0	0.00058
476	-0.2993	-0.0158	0.14959	1	0	0	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
477	-0.4298	-0.0167	0.30263	1	0	0	0	0	0	0	0	0	0	0	0.00058
478	-0.4292	-0.0158	0.30321	1	0	0	0	0	0	0	0	0	0	0	0
479	-0.301	-0.0167	0.15075	1	0	0	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
480	-0.2999	-0.0158	0.15017	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
481	-0.3003	-0.0158	0.14979	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
482	-0.3616	-0.0158	0.2316	1	0	0	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
483	-0.3607	-0.0158	0.23141	1	0	0	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
484	-0.3537	-0.0427	0.2393	0.08981	1	0	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
485	-0.419	-0.0158	0.28972	1	0	0	0	0	0	-0.0006	0	-0.0032	0	0	0
486	-0.3607	-0.0167	0.23025	1	0	0	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
487	-0.3612	-0.0158	0.23199	1	0	0	0	-0.0292	0	0	0	-0.0032	0	0	0
488	-0.4201	-0.0167	0.28972	1	0	0	0	0	0	0	0	-0.0032	0	0	0.00058
489	-0.4195	-0.0158	0.2903	1	0	0	0	0	0	0	0	-0.0032	0	0	0
490	-0.4199	-0.0158	0.28991	1	0	0	0	0	0	0	0	-0.0032	0.00019	0	0
491	-0.2873	-0.0158	0.14139	1	0	-0.0113	0	-0.0292	-0.0181	0	-0.0056	0	0.00129	0	0
492	-0.32	-0.0382	0.17538	1	0	-0.0113	0	-0.044	0	0	-0.0056	-0.0032	0	0	0.01488
493	-0.328	-0.043	0.18345	1	0	-0.0113	0	-0.0473	0	0	-0.0056	0	0	0	0.01811

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 8/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
494	-0.3416	-0.0158	0.21385	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
495	-0.3023	-0.0451	0.21999	0	1	-0.0248	0	-0.0805	0	0	-0.0062	0	0	0	0.03009
496	-0.3235	-0.0158	0.15951	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
497	-0.2873	-0.0158	0.14139	1	0	-0.0113	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0.00129	0	0
498	-0.3999	-0.0158	0.27215	1	0	-0.0113	0	0	0	0	-0.0056	0	0.00129	0	0
499	-0.3326	-0.043	0.18668	1	0	-0.0113	0	-0.0473	0	0	-0.0056	0	0.00129	0	0.01811
500	-0.3245	-0.0382	0.17861	1	0	-0.0113	0	-0.044	0	0	-0.0056	-0.0032	0.00129	0	0.01488
501	-0.3319	-0.0158	0.20093	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
502	-0.317	-0.0158	0.15628	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
503	-0.2443	0	0.17088	0	1	0	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
504	-0.3902	-0.0158	0.25924	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
505	-0.3462	-0.0158	0.12713	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
506	-0.2718	0	0.19832	0	1	0	0	-0.0504	-0.0363	0	0	0	0	0	0
507	-0.3001	-0.043	0.21136	1	0	-0.0113	0	-0.0473	0	-0.0558	-0.0056	0	0	0	0.01811
508	-0.3351	-0.0158	0.17141	1	0.12012	0	0	-0.0352	-0.028	0	0	0	0.00129	0	0
509	-0.304	-0.043	0.21524	1	0	-0.0113	0	-0.0473	0	-0.0571	-0.0056	0	0.00129	0	0.01811
510	-0.3267	-0.0412	0.22576	0	1	0	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
511	-0.2937	-0.0382	0.20167	1	0	-0.0113	0	-0.044	0	-0.0526	-0.0056	-0.0032	0	0	0.01488
512	-0.32	-0.0382	0.17538	1	0	-0.0113	0	-0.044	0	0	-0.0056	-0.0032	0	0	0.01488
513	-0.328	-0.043	0.18345	1	0	-0.0113	0	-0.0473	0	0	-0.0056	0	0	0	0.01811
514	-0.3169	0	0.20716	0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
515	-0.2975	-0.0382	0.20555	1	0	-0.0113	0	-0.044	0	-0.0539	-0.0056	-0.0032	0.00129	0	0.01488
516	-0.2873	-0.0158	0.14139	1	0	-0.0113	0	-0.0292	-0.0181	0	-0.0056	0	0.00129	0	0
517	-0.3912	-0.0158	0.19945	1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
518	-0.3267	-0.0412	0.22576	0	1	0	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
519	-0.3235	-0.0158	0.15951	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
520	-0.3443	-0.0412	0.2346	0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
521	-0.3443	0	0.2346	0	1	0	0	-0.0504	0	0	0	0	0	0	0
522	-0.3416	-0.0158	0.21385	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
523	-0.3532	0	0.316	0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
524	-0.3023	-0.0451	0.21999	0	1	-0.0248	0	-0.0805	0	0	-0.0062	0	0	0	0.03009
525	-0.3999	-0.0158	0.27215	1	0	-0.0113	0	0	0	0	-0.0056	0	0.00129	0	0
526	-0.3806	0	0.34344	0	1	0	0	-0.0504	0	0	0	0	0	0	0
527	-0.4192	-0.0158	0.28358	1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
528	-0.3326	-0.043	0.18668	1	0	-0.0113	0	-0.0473	0	0	-0.0056	0	0.00129	0	0.01811
529	-0.317	-0.0158	0.15628	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
530	-0.2873	-0.0158	0.14139	1	0	-0.0113	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0.00129	0	0
531	-0.3532	-0.0412	0.26112	0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
532	-0.4264	-0.0158	0.16424	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
533	-0.3319	-0.0158	0.20093	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
534	-0.3673	0	0.15673	0	1	0	0	0	0	-0.0274	0	0	0	0	0
535	-0.4222	-0.0412	0.15673	0	1	0	0	0	0	0	0	0	0	0	0.02744
536	-0.3948	0	0.18417	0	1	0	0	0	0	0	0	0	0	0	0
537	-0.3245	-0.0382	0.17861	1	0	-0.0113	0	-0.044	0	0	-0.0056	-0.0032	0.00129	0	0.01488
538	-0.454	0	0.41686	0	1	0	0	0	0	-0.0274	0	0	0	0	0
539	-0.3902	-0.0158	0.25924	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
540	-0.3462	-0.0158	0.12713	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
541	-0.2937	-0.0382	0.20167	1	0	-0.0113	0	-0.044	0	-0.0526	-0.0056	-0.0032	0	0	0.01488
542	-0.4897	-0.0158	0.354	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
543	-0.3001	-0.043	0.21136	1	0	-0.0113	0	-0.0473	0	-0.0558	-0.0056	0	0	0	0.01811
544	-0.5089	-0.0412	0.41686	0	1	0	0	0	0	0	0	0	0	0	0.02744
545	-0.304	-0.043	0.21524	1	0	-0.0113	0	-0.0473	0	-0.0571	-0.0056	0	0.00129	0	0.01811
546	-0.4815	0	0.4443	0	1	0	0	0	0	0	0	0	0	0	0
547	-0.3351	-0.0158	0.17141	1	0.12012	0	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
548	-0.2443	0	0.17088	0	1	0	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
549	-0.2975	-0.0382	0.20555	1	0	-0.0113	0	-0.044	0	-0.0539	-0.0056	-0.0032	0.00129	0	0.01488
550	-0.2957	0	0.19654	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
551	-0.3249	-0.0412	0.22219	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
552	-0.2718	0	0.19832	0	1	0	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
553	-0.3213	0	0.27351	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
554	-0.3231	0	0.22398	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
555	-0.3822	-0.0158	0.19495	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
556	-0.3489	-0.042	0.23661	0.07515	1	0	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
557	-0.4057	-0.0158	0.26556	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
558	-0.3213	-0.0385	0.22219	0	1	0	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
559	-0.3461	0	0.14611	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
560	-0.3231	-0.0385	0.22398	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
561	-0.3488	0	0.30095	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
562	-0.401	-0.0412	0.14611	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
563	-0.3735	0	0.17355	0	1	0	0	0	0	0	0	-0.0106	0	0	0
564	-0.4174	-0.0158	0.15974	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 9/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
565	-0.4496	0	0.40181	0	1	0	0	0	0	0	0	-0.0106	0	0	0
566	-0.4761	-0.0158	0.33598	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0
567	-0.4222	0	0.37437	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
568	-0.2834	-0.0158	0.13865	1	0	-0.0113	0	-0.0292	-0.0181	0	-0.0056	0	0.00019	0	0
569	-0.2825	-0.0158	0.13845	1	0	-0.0113	0	-0.0292	-0.0181	-0.0006	-0.0056	0	0	0	0
570	-0.4771	-0.0412	0.37437	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
571	-0.3187	-0.0158	0.15657	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	0	0	0	0
572	-0.2842	-0.0167	0.13961	1	0	-0.0113	0	-0.0297	-0.0175	0	-0.0056	0	0	0	0.00058
573	-0.283	-0.0158	0.13903	1	0	-0.0113	0	-0.0292	-0.0181	0	-0.0056	0	0	0	0
574	-0.3193	-0.0167	0.15715	1	0	-0.0113	0	-0.0297	0	0	-0.0056	0	0	0	0.00058
575	-0.3193	-0.0158	0.15715	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0	0	0
576	-0.3196	-0.0158	0.15676	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00019	0	0
577	-0.3374	-0.0158	0.21149	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0	0	0
578	-0.3378	-0.0158	0.2111	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00019	0	0
579	-0.3368	-0.0158	0.21091	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	0	0	0	0
580	-0.3961	-0.0158	0.26941	1	0	-0.0113	0	0	0	0	-0.0056	0	0.00019	0	0
581	-0.3951	-0.0158	0.26921	1	0	-0.0113	0	0	0	-0.0006	-0.0056	0	0	0	0
582	-0.3368	-0.0167	0.20975	1	0	-0.0113	0	-0.0297	0	0	-0.0056	0	0	0	0.00058
583	-0.2825	-0.0158	0.13845	1	0	-0.0113	0	-0.0292	-0.0149	-0.0006	-0.0056	-0.0032	0	0	0
584	-0.3963	-0.0167	0.26921	1	0	-0.0113	0	0	0	0	-0.0056	0	0	0	0.00058
585	-0.3957	-0.0158	0.26979	1	0	-0.0113	0	0	0	0	-0.0056	0	0	0	0
586	-0.2842	-0.0167	0.13961	1	0	-0.0113	0	-0.0297	-0.0143	0	-0.0056	-0.0032	0	0	0.00058
587	-0.283	-0.0158	0.13903	1	0	-0.0113	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0	0	0
588	-0.2834	-0.0158	0.13865	1	0	-0.0113	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0.00019	0	0
589	-0.3128	-0.0158	0.15392	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0	0	0
590	-0.3132	-0.0158	0.15353	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00019	0	0
591	-0.3122	-0.0158	0.15334	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	-0.0032	0	0	0
592	-0.3271	-0.0158	0.19799	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	-0.0032	0	0	0
593	-0.4542	-0.0505	0.29124	0.55772	1	-0.0311	0	-0.0945	0	0	-0.0093	-0.0124	0	0	0.02776
594	-0.3128	-0.0167	0.15392	1	0	-0.0113	0	-0.0297	0	0	-0.0056	-0.0032	0	0	0.00058
595	-0.3271	-0.0167	0.19683	1	0	-0.0113	0	-0.0297	0	0	-0.0056	-0.0032	0	0	0.00058
596	-0.3277	-0.0158	0.19857	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0	0	0
597	-0.3281	-0.0158	0.19819	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00019	0	0
598	-0.342	-0.0158	0.12477	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0
599	-0.3423	-0.0158	0.12438	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00019	0	0
600	-0.3414	-0.0158	0.12419	1	0	-0.0113	0	0	0	-0.0006	-0.0056	-0.0032	0	0	0
601	-0.3864	-0.0158	0.25649	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00019	0	0
602	-0.3854	-0.0158	0.2563	1	0	-0.0113	0	0	0	-0.0006	-0.0056	-0.0032	0	0	0
603	-0.3425	-0.0167	0.12419	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0.00058
604	-0.2134	0	0.15849	0	1	-0.0248	0	-0.0504	-0.0301	-0.0274	-0.0062	0	0	0	0
605	-0.3866	-0.0167	0.2563	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0.00058
606	-0.386	-0.0158	0.25688	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0
607	-0.2957	-0.0412	0.21338	0	1	-0.0248	0	-0.0779	-0.0026	0	-0.0062	0	0	0	0.02744
608	-0.2408	0	0.18594	0	1	-0.0248	0	-0.0504	-0.0301	0	-0.0062	0	0	0	0
609	-0.3145	-0.0158	0.15879	1	0.12012	-0.0142	0	-0.0352	-0.0217	0	-0.0063	0	0.00129	0	0
610	-0.358	-0.0158	0.18051	1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	0	0.00129	0	0
611	-0.2735	0	0.18858	0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	0	0	0	0
612	-0.2957	-0.0412	0.21338	0	1	-0.0248	0	-0.0779	0	0	-0.0062	-0.0026	0	0	0.02744
613	-0.3036	0	0.27884	0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	0	0	0	0
614	-0.301	-0.0412	0.21602	0	1	-0.0248	0	-0.0779	0	0	-0.0062	0	0	0	0.02744
615	-0.301	0	0.21602	0	1	-0.0248	0	-0.0504	0	0	-0.0062	0	0	0	0
616	-0.3036	-0.0412	0.22396	0	1	-0.0248	0	-0.0779	0	0	-0.0062	0	0	0	0.02744
617	-0.3311	0	0.30628	0	1	-0.0248	0	-0.0504	0	0	-0.0062	0	0	0	0
618	-0.3797	-0.0158	0.2457	1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	0	0.00129	0	0
619	-0.3514	0	0.16559	0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0
620	-0.3932	-0.0158	0.1453	1	0.12012	-0.0142	0	0	0	0	-0.0063	0	0.00129	0	0
621	-0.324	0	0.13815	0	1	-0.0248	0	0	0	-0.0274	-0.0062	0	0	0	0
622	-0.4501	-0.0158	0.31612	1	0.12012	-0.0142	0	0	0	0	-0.0063	0	0.00129	0	0
623	-0.4045	0	0.3797	0	1	-0.0248	0	0	0	-0.0274	-0.0062	0	0	0	0
624	-0.3789	-0.0412	0.13815	0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0.02744
625	-0.2134	0	0.15849	0	1	-0.0248	0	-0.0504	-0.0195	-0.0274	-0.0062	-0.0106	0	0	0
626	-0.4594	-0.0412	0.3797	0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0.02744
627	-0.4319	0	0.40714	0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0
628	-0.2877	-0.0412	0.19742	0	1	-0.0248	0	-0.0699	0	0	-0.0062	-0.0106	0	0	0.02744
629	-0.2408	0	0.18594	0	1	-0.0248	0	-0.0504	-0.0195	0	-0.0062	-0.0106	0	0	0
630	-0.3145	-0.0158	0.15879	1	0.12012	-0.0142	0	-0.0352	-0.0172	0	-0.0063	-0.0045	0.00129	0	0
631	-0.2797	0	0.2054	0	1	-0.0248	0	-0.0504	0	0	-0.0062	-0.0106	0	0	0
632	-0.349	-0.0158	0.17601	1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	-0.0045	0.00129	0	0
633	-0.2523	0	0.17796	0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	-0.0106	0	0	0
634	-0.3662	-0.0158	0.22767	1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	-0.0045	0.00129	0	0
635	-0.2718	-0.0292	0.19742	0	1	-0.0248	0	-0.0699	0	-0.008	-0.0062	-0.0106	0	0	0.01946

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 10/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
636	-0.2718	0	0.23634	0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	-0.0106	0	0	0
637	-0.2797	-0.0292	0.2054	0	1	-0.0248	0	-0.0699	0	0	-0.0062	-0.0106	0	0	0.01946
638	-0.2992	0	0.26378	0	1	-0.0248	0	-0.0504	0	0	-0.0062	-0.0106	0	0	0
639	-0.4181	-0.0459	0.27181	0.43884	1	-0.0297	0	-0.0892	0	0	-0.0086	-0.012	0.00057	0	0.02599
640	-0.3507	-0.0465	0.20044	1	0	0	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
641	-0.3027	0	0.12753	0	1	-0.0248	0	0	0	-0.0274	-0.0062	-0.0106	0	0	0
642	-0.3588	-0.0514	0.20851	1	0	0	0	-0.0528	0	0	0	0	0	0	0.02368
643	-0.3576	-0.0412	0.12753	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0.02744
644	-0.3302	0	0.15497	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0
645	-0.3842	-0.0158	0.1408	1	0.12012	-0.0142	0	0	0	0	-0.0063	-0.0045	0.00129	0	0
646	-0.3726	0	0.3372	0	1	-0.0248	0	0	0	-0.0274	-0.0062	-0.0106	0	0	0
647	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
648	-0.3752	-0.0158	0.24727	1	0	0	0	-0.0292	0	0	0	0	0.00129	0	0
649	-0.3488	-0.0544	0.24786	0	1	0	0	-0.0867	0	0	0	0	0	0	0.03628
650	-0.4366	-0.0158	0.2981	1	0.12012	-0.0142	0	0	0	0	-0.0063	-0.0045	0.00129	0	0
651	-0.4275	-0.0412	0.3372	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0.02744
652	-0.3633	-0.0514	0.21174	1	0	0	0	-0.0528	0	0	0	0	0.00129	0	0.02368
653	-0.4001	0	0.36465	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0
654	-0.2834	-0.0158	0.13865	1	0	-0.0113	0	-0.0292	-0.0181	0	-0.0056	0	0.00019	0	0
655	-0.2825	-0.0158	0.13845	1	0	-0.0113	0	-0.0292	-0.0181	-0.0006	-0.0056	0	0	0	0
656	-0.4335	-0.0158	0.30557	1	0	0	0	0	0	0	0	0	0.00129	0	0
657	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
658	-0.283	-0.0158	0.13903	1	0	-0.0113	0	-0.0292	-0.0181	0	-0.0056	0	0	0	0
659	-0.3552	-0.0465	0.20367	1	0	0	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
660	-0.3655	-0.0158	0.23435	1	0	0	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
661	-0.2842	-0.0167	0.13961	1	0	-0.0113	0	-0.0297	-0.0175	0	-0.0056	0	0	0	0.00058
662	-0.3197	-0.0514	0.24763	1	0	0	0	-0.0528	0	-0.0782	0	0	0	0	0.02368
663	-0.3187	-0.0158	0.15657	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	0	0	0	0
664	-0.4238	-0.0158	0.29266	1	0	0	0	0	0	0	0	-0.0032	0.00129	0	0
665	-0.3193	-0.0158	0.15715	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0	0	0
666	-0.3132	-0.0465	0.23794	1	0	0	0	-0.0496	0	-0.075	0	-0.0032	0	0	0.02045
667	-0.3196	-0.0158	0.15676	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00019	0	0
668	-0.3236	-0.0514	0.25151	1	0	0	0	-0.0528	0	-0.0795	0	0	0.00129	0	0.02368
669	-0.3368	-0.0158	0.21091	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	0	0	0	0
670	-0.3193	-0.0167	0.15715	1	0	-0.0113	0	-0.0297	0	0	-0.0056	0	0	0	0.00058
671	-0.3588	-0.0514	0.20851	1	0	0	0	-0.0528	0	0	0	0	0	0	0.02368
672	-0.3378	-0.0158	0.2111	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0.00019	0	0
673	-0.3171	-0.0465	0.24182	1	0	0	0	-0.0496	0	-0.0763	0	-0.0032	0.00129	0	0.02045
674	-0.3368	-0.0167	0.20975	1	0	-0.0113	0	-0.0297	0	0	-0.0056	0	0	0	0.00058
675	-0.3507	-0.0465	0.20044	1	0	0	0	-0.0496	0	0	0	-0.0032	0	0	0.02045
676	-0.3374	-0.0158	0.21149	1	0	-0.0113	0	-0.0292	0	0	-0.0056	0	0	0	0
677	-0.3961	-0.0158	0.26941	1	0	-0.0113	0	0	0	0	-0.0056	0	0.00019	0	0
678	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0237	0	0	0	0.00129	0	0
679	-0.3951	-0.0158	0.26921	1	0	-0.0113	0	0	0	-0.0006	-0.0056	0	0	0	0
680	-0.3752	-0.0158	0.24727	1	0	0	0	-0.0292	0	0	0	0	0.00129	0	0
681	-0.3957	-0.0158	0.26979	1	0	-0.0113	0	0	0	0	-0.0056	0	0	0	0
682	-0.3488	-0.0544	0.24786	0	1	0	0	-0.0867	0	0	0	0	0	0	0.03628
683	-0.4335	-0.0158	0.30557	1	0	0	0	0	0	0	0	0	0.00129	0	0
684	-0.3041	-0.0158	0.15253	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00129	0	0
685	-0.3655	-0.0158	0.23435	1	0	0	0	-0.0292	0	0	0	-0.0032	0.00129	0	0
686	-0.3633	-0.0514	0.21174	1	0	0	0	-0.0528	0	0	0	0	0.00129	0	0.02368
687	-0.3963	-0.0167	0.26921	1	0	-0.0113	0	0	0	0	-0.0056	0	0	0	0.00058
688	-0.3197	-0.0514	0.24763	1	0	0	0	-0.0528	0	-0.0782	0	0	0	0	0.02368
689	-0.4238	-0.0158	0.29266	1	0	0	0	0	0	0	0	-0.0032	0.00129	0	0
690	-0.3552	-0.0465	0.20367	1	0	0	0	-0.0496	0	0	0	-0.0032	0.00129	0	0.02045
691	-0.3171	-0.0465	0.24182	1	0	0	0	-0.0496	0	-0.0763	0	-0.0032	0.00129	0	0.02045
692	-0.3236	-0.0514	0.25151	1	0	0	0	-0.0528	0	-0.0795	0	0	0.00129	0	0.02368
693	-0.3132	-0.0465	0.23794	1	0	0	0	-0.0496	0	-0.075	0	-0.0032	0	0	0.02045
694	-0.283	-0.0158	0.13903	1	0	-0.0113	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0	0	0
695	-0.2834	-0.0158	0.13865	1	0	-0.0113	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0.00019	0	0
696	-0.2825	-0.0158	0.13845	1	0	-0.0113	0	-0.0292	-0.0149	-0.0006	-0.0056	-0.0032	0	0	0
697	-0.3132	-0.0158	0.15353	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00019	0	0
698	-0.3122	-0.0158	0.15334	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	-0.0032	0	0	0
699	-0.2842	-0.0167	0.13961	1	0	-0.0113	0	-0.0297	-0.0143	0	-0.0056	-0.0032	0	0	0.00058
700	-0.4542	-0.0505	0.29124	0.55772	1	-0.0311	0	-0.0945	0	0	-0.0093	-0.0124	0	0	0.02776
701	-0.3128	-0.0167	0.15392	1	0	-0.0113	0	-0.0297	0	0	-0.0056	-0.0032	0	0	0.00058
702	-0.3128	-0.0158	0.15392	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0	0	0
703	-0.3277	-0.0158	0.19857	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0	0	0
704	-0.3281	-0.0158	0.19819	1	0	-0.0113	0	-0.0292	0	0	-0.0056	-0.0032	0.00019	0	0
705	-0.3271	-0.0158	0.19799	1	0	-0.0113	0	-0.0292	0	-0.0006	-0.0056	-0.0032	0	0	0
706	-0.3423	-0.0158	0.12438	1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00019	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 11/14

	GlyOH	O2	CO2	X		Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
707	-0.3414	-0.0158	0.12419		1	0	-0.0113	0	0	0	-0.0006	-0.0056	-0.0032	0	0	0
708	-0.3271	-0.0167	0.19683		1	0	-0.0113	0	-0.0297	0	0	-0.0056	-0.0032	0	0	0.00058
709	-0.3854	-0.0158	0.2563		1	0	-0.0113	0	0	0	-0.0006	-0.0056	-0.0032	0	0	0
710	-0.3425	-0.0167	0.12419		1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0.00058
711	-0.342	-0.0158	0.12477		1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0
712	-0.3866	-0.0167	0.2563		1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0.00058
713	-0.386	-0.0158	0.25688		1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0	0	0
714	-0.3864	-0.0158	0.25649		1	0	-0.0113	0	0	0	0	-0.0056	-0.0032	0.00019	0	0
715	-0.2999	-0.0158	0.15017		1	0	0	0	-0.0292	-0.0237	0	0	0	0	0	0
716	-0.3003	-0.0158	0.14979		1	0	0	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
717	-0.2993	-0.0158	0.14959		1	0	0	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
718	-0.3713	-0.0158	0.24452		1	0	0	0	-0.0292	0	0	0	0	0.00019	0	0
719	-0.3703	-0.0158	0.24433		1	0	0	0	-0.0292	0	-0.0006	0	0	0	0	0
720	-0.301	-0.0167	0.15075		1	0	0	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
721	-0.4287	-0.0158	0.30263		1	0	0	0	0	0	-0.0006	0	0	0	0	0
722	-0.3703	-0.0167	0.24317		1	0	0	0	-0.0297	0	0	0	0	0	0	0.00058
723	-0.3709	-0.0158	0.24491		1	0	0	0	-0.0292	0	0	0	0	0	0	0
724	-0.2134	0	0.15849		0	1	-0.0248	0	-0.0504	-0.0301	-0.0274	-0.0062	0	0	0	0
725	-0.4292	-0.0158	0.30321		1	0	0	0	0	0	0	0	0	0	0	0
726	-0.4296	-0.0158	0.30283		1	0	0	0	0	0	0	0	0	0.00019	0	0
727	-0.2408	0	0.18594		0	1	-0.0248	0	-0.0504	-0.0301	0	-0.0062	0	0	0	0
728	-0.3145	-0.0158	0.15879		1	0.12012	-0.0142	0	-0.0352	-0.0217	0	-0.0063	0	0.00129	0	0
729	-0.4298	-0.0167	0.30263		1	0	0	0	0	0	0	0	0	0	0	0.00058
730	-0.2735	0	0.18858		0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	0	0	0	0
731	-0.2957	-0.0412	0.21338		0	1	-0.0248	0	-0.0779	0	0	-0.0062	-0.0026	0	0	0.02744
732	-0.2957	-0.0412	0.21338		0	1	-0.0248	0	-0.0779	-0.0026	0	-0.0062	0	0	0	0.02744
733	-0.301	0	0.21602		0	1	-0.0248	0	-0.0504	0	0	-0.0062	0	0	0	0
734	-0.358	-0.0158	0.18051		1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	0	0.00129	0	0
735	-0.2993	-0.0158	0.14959		1	0	0	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
736	-0.2999	-0.0158	0.15017		1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
737	-0.301	-0.0412	0.21602		0	1	-0.0248	0	-0.0779	0	0	-0.0062	0	0	0	0.02744
738	-0.3003	-0.0158	0.14979		1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
739	-0.3797	-0.0158	0.2457		1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	0	0.00129	0	0
740	-0.3036	0	0.27884		0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	0	0	0	0
741	-0.301	-0.0167	0.15075		1	0	0	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
742	-0.3607	-0.0158	0.23141		1	0	0	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
743	-0.3311	0	0.30628		0	1	-0.0248	0	-0.0504	0	0	-0.0062	0	0	0	0
744	-0.3537	-0.0427	0.2393	0.08981		1	0	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
745	-0.3616	-0.0158	0.2316		1	0	0	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
746	-0.3036	-0.0412	0.22396		0	1	-0.0248	0	-0.0779	0	0	-0.0062	0	0	0	0.02744
747	-0.419	-0.0158	0.28972		1	0	0	0	0	0	-0.0006	0	-0.0032	0	0	0
748	-0.3607	-0.0167	0.23025		1	0	0	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
749	-0.3612	-0.0158	0.23199		1	0	0	0	-0.0292	0	0	0	-0.0032	0	0	0
750	-0.3932	-0.0158	0.1453		1	0.12012	-0.0142	0	0	0	0	-0.0063	0	0.00129	0	0
751	-0.4199	-0.0158	0.28991		1	0	0	0	0	0	0	0	-0.0032	0.00019	0	0
752	-0.324	0	0.13815		0	1	-0.0248	0	0	0	-0.0274	-0.0062	0	0	0	0
753	-0.4201	-0.0167	0.28972		1	0	0	0	0	0	0	0	-0.0032	0	0	0.00058
754	-0.3514	0	0.16559		0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0
755	-0.4195	-0.0158	0.2903		1	0	0	0	0	0	0	0	-0.0032	0	0	0
756	-0.4501	-0.0158	0.31612		1	0.12012	-0.0142	0	0	0	0	-0.0063	0	0.00129	0	0
757	-0.4045	0	0.3797		0	1	-0.0248	0	0	0	-0.0274	-0.0062	0	0	0	0
758	-0.3789	-0.0412	0.13815		0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0.02744
759	-0.2134	0	0.15849		0	1	-0.0248	0	-0.0504	-0.0195	-0.0274	-0.0062	-0.0106	0	0	0
760	-0.4594	-0.0412	0.3797		0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0.02744
761	-0.4319	0	0.40714		0	1	-0.0248	0	0	0	0	-0.0062	0	0	0	0
762	-0.2443	0	0.17088		0	1	0	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
763	-0.2408	0	0.18594		0	1	-0.0248	0	-0.0504	-0.0195	0	-0.0062	-0.0106	0	0	0
764	-0.3145	-0.0158	0.15879		1	0.12012	-0.0142	0	-0.0352	-0.0172	0	-0.0063	-0.0045	0.00129	0	0
765	-0.3267	-0.0412	0.22576		0	1	0	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
766	-0.2718	0	0.19832		0	1	0	0	-0.0504	-0.0363	0	0	0	0	0	0
767	-0.3351	-0.0158	0.17141		1	0.12012	0	0	-0.0352	-0.028	0	0	0	0.00129	0	0
768	-0.3267	-0.0412	0.22576		0	1	0	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
769	-0.2523	0	0.17796		0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	-0.0106	0	0	0
770	-0.2877	-0.0412	0.19742		0	1	-0.0248	0	-0.0699	0	0	-0.0062	-0.0106	0	0	0.02744
771	-0.2797	0	0.2054		0	1	-0.0248	0	-0.0504	0	0	-0.0062	-0.0106	0	0	0
772	-0.3169	0	0.20716		0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
773	-0.349	-0.0158	0.17601		1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	-0.0045	0.00129	0	0
774	-0.3443	-0.0412	0.2346		0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
775	-0.3443	0	0.2346		0	1	0	0	-0.0504	0	0	0	0	0	0	0
776	-0.3912	-0.0158	0.19945		1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
777	-0.2718	-0.0292	0.19742		0	1	-0.0248	0	-0.0699	0	-0.008	-0.0062	-0.0106	0	0	0.01946

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 12/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
778	-0.2718	0	0.23634	0	1	-0.0248	0	-0.0504	0	-0.0274	-0.0062	-0.0106	0	0	0
779	-0.4181	-0.0459	0.27181	0.43884	1	-0.0297	0	-0.0892	0	0	-0.0086	-0.012	0.00057	0	0.02599
780	-0.3662	-0.0158	0.22767	1	0.12012	-0.0142	0	-0.0352	0	0	-0.0063	-0.0045	0.00129	0	0
781	-0.3532	0	0.316	0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
782	-0.3806	0	0.34344	0	1	0	0	-0.0504	0	0	0	0	0	0	0
783	-0.2992	0	0.26378	0	1	-0.0248	0	-0.0504	0	0	-0.0062	-0.0106	0	0	0
784	-0.4192	-0.0158	0.28358	1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
785	-0.3027	0	0.12753	0	1	-0.0248	0	0	0	-0.0274	-0.0062	-0.0106	0	0	0
786	-0.3532	-0.0412	0.26112	0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
787	-0.2797	-0.0292	0.2054	0	1	-0.0248	0	-0.0699	0	0	-0.0062	-0.0106	0	0	0.01946
788	-0.4264	-0.0158	0.16424	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
789	-0.3842	-0.0158	0.1408	1	0.12012	-0.0142	0	0	0	0	-0.0063	-0.0045	0.00129	0	0
790	-0.3673	0	0.15673	0	1	0	0	0	0	-0.0274	0	0	0	0	0
791	-0.3948	0	0.18417	0	1	0	0	0	0	0	0	0	0	0	0
792	-0.3576	-0.0412	0.12753	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0.02744
793	-0.3302	0	0.15497	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0
794	-0.454	0	0.41686	0	1	0	0	0	0	-0.0274	0	0	0	0	0
795	-0.3726	0	0.3372	0	1	-0.0248	0	0	0	-0.0274	-0.0062	-0.0106	0	0	0
796	-0.4222	-0.0412	0.15673	0	1	0	0	0	0	0	0	0	0	0	0.02744
797	-0.4897	-0.0158	0.354	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
798	-0.4001	0	0.36465	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0
799	-0.4366	-0.0158	0.2981	1	0.12012	-0.0142	0	0	0	0	-0.0063	-0.0045	0.00129	0	0
800	-0.5089	-0.0412	0.41686	0	1	0	0	0	0	0	0	0	0	0	0.02744
801	-0.4815	0	0.4443	0	1	0	0	0	0	0	0	0	0	0	0
802	-0.4275	-0.0412	0.3372	0	1	-0.0248	0	0	0	0	-0.0062	-0.0106	0	0	0.02744
803	-0.2718	0	0.19832	0	1	0	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
804	-0.3351	-0.0158	0.17141	1	0.12012	0	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
805	-0.2443	0	0.17088	0	1	0	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
806	-0.3822	-0.0158	0.19495	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
807	-0.2957	0	0.19654	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
808	-0.3249	-0.0412	0.22219	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
809	-0.3213	-0.0385	0.22219	0	1	0	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
810	-0.3213	0	0.27351	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
811	-0.3231	0	0.22398	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
812	-0.3488	0	0.30095	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
813	-0.3489	-0.042	0.23661	0.07515	1	0	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
814	-0.4057	-0.0158	0.26556	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
815	-0.4174	-0.0158	0.15974	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0
816	-0.3461	0	0.14611	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
817	-0.3231	-0.0385	0.22398	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
818	-0.4222	0	0.37437	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
819	-0.401	-0.0412	0.14611	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
820	-0.3735	0	0.17355	0	1	0	0	0	0	0	0	-0.0106	0	0	0
821	-0.4771	-0.0412	0.37437	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
822	-0.4496	0	0.40181	0	1	0	0	0	0	0	0	-0.0106	0	0	0
823	-0.4761	-0.0158	0.33598	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0
824	-0.2999	-0.0158	0.15017	1	0	0	0	-0.0292	-0.0237	0	0	0	0	0	0
825	-0.3003	-0.0158	0.14979	1	0	0	0	-0.0292	-0.0237	0	0	0	0.00019	0	0
826	-0.2993	-0.0158	0.14959	1	0	0	0	-0.0292	-0.0237	-0.0006	0	0	0	0	0
827	-0.3713	-0.0158	0.24452	1	0	0	0	-0.0292	0	0	0	0	0.00019	0	0
828	-0.3703	-0.0158	0.24433	1	0	0	0	-0.0292	0	-0.0006	0	0	0	0	0
829	-0.301	-0.0167	0.15075	1	0	0	0	-0.0297	-0.0231	0	0	0	0	0	0.00058
830	-0.4287	-0.0158	0.30263	1	0	0	0	0	0	-0.0006	0	0	0	0	0
831	-0.3703	-0.0167	0.24317	1	0	0	0	-0.0297	0	0	0	0	0	0	0.00058
832	-0.3709	-0.0158	0.24491	1	0	0	0	-0.0292	0	0	0	0	0	0	0
833	-0.4298	-0.0167	0.30263	1	0	0	0	0	0	0	0	0	0	0	0.00058
834	-0.4292	-0.0158	0.30321	1	0	0	0	0	0	0	0	0	0	0	0
835	-0.4296	-0.0158	0.30283	1	0	0	0	0	0	0	0	0	0.00019	0	0
836	-0.2999	-0.0158	0.15017	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0	0	0
837	-0.3003	-0.0158	0.14979	1	0	0	0	-0.0292	-0.0205	0	0	-0.0032	0.00019	0	0
838	-0.2993	-0.0158	0.14959	1	0	0	0	-0.0292	-0.0205	-0.0006	0	-0.0032	0	0	0
839	-0.3607	-0.0158	0.23141	1	0	0	0	-0.0292	0	-0.0006	0	-0.0032	0	0	0
840	-0.3537	-0.0427	0.2393	0.08981	1	0	0	-0.0805	0	0	0	-0.0109	0	0	0.02749
841	-0.301	-0.0167	0.15075	1	0	0	0	-0.0297	-0.0199	0	0	-0.0032	0	0	0.00058
842	-0.3607	-0.0167	0.23025	1	0	0	0	-0.0297	0	0	0	-0.0032	0	0	0.00058
843	-0.3612	-0.0158	0.23199	1	0	0	0	-0.0292	0	0	0	-0.0032	0	0	0
844	-0.3616	-0.0158	0.2316	1	0	0	0	-0.0292	0	0	0	-0.0032	0.00019	0	0
845	-0.4195	-0.0158	0.2903	1	0	0	0	0	0	0	0	-0.0032	0	0	0
846	-0.4199	-0.0158	0.28991	1	0	0	0	0	0	0	0	-0.0032	0.00019	0	0
847	-0.419	-0.0158	0.28972	1	0	0	0	0	0	-0.0006	0	-0.0032	0	0	0
848	-0.3351	-0.0158	0.17141	1	0.12012	0	0	-0.0352	-0.028	0	0	0	0.00129	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 13/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
849	-0.2443	0	0.17088	0	1	0	0	-0.0504	-0.0363	-0.0274	0	0	0	0	0
850	-0.4201	-0.0167	0.28972	1	0	0	0	0	0	0	0	-0.0032	0	0	0.00058
851	-0.3267	-0.0412	0.22576	0	1	0	0	-0.0779	0	0	0	-0.0088	0	0	0.02744
852	-0.3267	-0.0412	0.22576	0	1	0	0	-0.0779	-0.0088	0	0	0	0	0	0.02744
853	-0.2718	0	0.19832	0	1	0	0	-0.0504	-0.0363	0	0	0	0	0	0
854	-0.3443	0	0.2346	0	1	0	0	-0.0504	0	0	0	0	0	0	0
855	-0.3912	-0.0158	0.19945	1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
856	-0.3169	0	0.20716	0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
857	-0.4192	-0.0158	0.28358	1	0.12012	0	0	-0.0352	0	0	0	0	0.00129	0	0
858	-0.3532	0	0.316	0	1	0	0	-0.0504	0	-0.0274	0	0	0	0	0
859	-0.3443	-0.0412	0.2346	0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
860	-0.3673	0	0.15673	0	1	0	0	0	0	-0.0274	0	0	0	0	0
861	-0.3532	-0.0412	0.26112	0	1	0	0	-0.0779	0	0	0	0	0	0	0.02744
862	-0.3806	0	0.34344	0	1	0	0	-0.0504	0	0	0	0	0	0	0
863	-0.4222	-0.0412	0.15673	0	1	0	0	0	0	0	0	0	0	0	0.02744
864	-0.3948	0	0.18417	0	1	0	0	0	0	0	0	0	0	0	0
865	-0.4264	-0.0158	0.16424	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
866	-0.4815	0	0.4443	0	1	0	0	0	0	0	0	0	0	0	0
867	-0.4897	-0.0158	0.354	1	0.12012	0	0	0	0	0	0	0	0.00129	0	0
868	-0.454	0	0.41686	0	1	0	0	0	0	-0.0274	0	0	0	0	0
869	-0.3351	-0.0158	0.17141	1	0.12012	0	0	-0.0352	-0.0235	0	0	-0.0045	0.00129	0	0
870	-0.2443	0	0.17088	0	1	0	0	-0.0504	-0.0257	-0.0274	0	-0.0106	0	0	0
871	-0.5089	-0.0412	0.41686	0	1	0	0	0	0	0	0	0	0	0	0.02744
872	-0.2957	0	0.19654	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
873	-0.3249	-0.0412	0.22219	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02744
874	-0.2718	0	0.19832	0	1	0	0	-0.0504	-0.0257	0	0	-0.0106	0	0	0
875	-0.3213	0	0.27351	0	1	0	0	-0.0504	0	-0.0274	0	-0.0106	0	0	0
876	-0.3231	0	0.22398	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
877	-0.3822	-0.0158	0.19495	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
878	-0.3489	-0.042	0.23661	0.07515	1	0	0	-0.0798	0	0	0	-0.0109	9.7E-05	0	0.02719
879	-0.4057	-0.0158	0.26556	1	0.12012	0	0	-0.0352	0	0	0	-0.0045	0.00129	0	0
880	-0.3213	-0.0385	0.22219	0	1	0	0	-0.0761	0	-0.0018	0	-0.0106	0	0	0.02566
881	-0.3461	0	0.14611	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
882	-0.3231	-0.0385	0.22398	0	1	0	0	-0.0761	0	0	0	-0.0106	0	0	0.02566
883	-0.3488	0	0.30095	0	1	0	0	-0.0504	0	0	0	-0.0106	0	0	0
884	-0.401	-0.0412	0.14611	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
885	-0.3735	0	0.17355	0	1	0	0	0	0	0	0	-0.0106	0	0	0
886	-0.4174	-0.0158	0.15974	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0
887	-0.4496	0	0.40181	0	1	0	0	0	0	0	0	-0.0106	0	0	0
888	-0.4761	-0.0158	0.33598	1	0.12012	0	0	0	0	0	0	-0.0045	0.00129	0	0
889	-0.4222	0	0.37437	0	1	0	0	0	0	-0.0274	0	-0.0106	0	0	0
890	-0.3173	-0.0382	0.17538	1	0	-0.0139	0	-0.044	0	0	-0.0056	-0.0032	0	0	0.01488
891	-0.3254	-0.043	0.18345	1	0	-0.0139	0	-0.0473	0	0	-0.0056	0	0	0	0.01811
892	-0.4771	-0.0412	0.37437	0	1	0	0	0	0	0	0	-0.0106	0	0	0.02744
893	-0.2926	-0.0451	0.21999	0	1	-0.0345	0	-0.0805	0	0	-0.0062	0	0	0	0.03009
894	-0.3208	-0.0158	0.15951	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
895	-0.2846	-0.0158	0.14139	1	0	-0.0139	0	-0.0292	-0.0181	0	-0.0056	0	0.00129	0	0
896	-0.35	-0.0158	0.13036	1	0	-0.0139	0	0	0	0	-0.0056	0	0.00129	0	0
897	-0.3299	-0.043	0.18668	1	0	-0.0139	0	-0.0473	0	0	-0.0056	0	0.00129	0	0.01811
898	-0.3389	-0.0158	0.21385	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
899	-0.3144	-0.0158	0.15628	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
900	-0.2846	-0.0158	0.14139	1	0	-0.0139	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0.00129	0	0
901	-0.3972	-0.0158	0.27215	1	0	-0.0139	0	0	0	0	-0.0056	0	0.00129	0	0
902	-0.3435	-0.0158	0.12713	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
903	-0.3218	-0.0382	0.17861	1	0	-0.0139	0	-0.044	0	0	-0.0056	-0.0032	0.00129	0	0.01488
904	-0.3293	-0.0158	0.20093	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
905	-0.2923	-0.0382	0.20034	1	0	-0.0139	0	-0.044	0	-0.0499	-0.0056	-0.0032	0	0	0.01488
906	-0.2988	-0.043	0.21003	1	0	-0.0139	0	-0.0473	0	-0.0532	-0.0056	0	0	0	0.01811
907	-0.3876	-0.0158	0.25924	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
908	-0.3254	-0.043	0.18345	1	0	-0.0139	0	-0.0473	0	0	-0.0056	0	0	0	0.01811
909	-0.2962	-0.0382	0.20422	1	0	-0.0139	0	-0.044	0	-0.0512	-0.0056	-0.0032	0.00129	0	0.01488
910	-0.3027	-0.043	0.2139	1	0	-0.0139	0	-0.0473	0	-0.0544	-0.0056	0	0.00129	0	0.01811
911	-0.3208	-0.0158	0.15951	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
912	-0.2846	-0.0158	0.14139	1	0	-0.0139	0	-0.0292	-0.0181	0	-0.0056	0	0.00129	0	0
913	-0.3173	-0.0382	0.17538	1	0	-0.0139	0	-0.044	0	0	-0.0056	-0.0032	0	0	0.01488
914	-0.3299	-0.043	0.18668	1	0	-0.0139	0	-0.0473	0	0	-0.0056	0	0.00129	0	0.01811
915	-0.3389	-0.0158	0.21385	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0.00129	0	0
916	-0.2926	-0.0451	0.21999	0	1	-0.0345	0	-0.0805	0	0	-0.0062	0	0	0	0.03009
917	-0.2846	-0.0158	0.14139	1	0	-0.0139	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0.00129	0	0
918	-0.3972	-0.0158	0.27215	1	0	-0.0139	0	0	0	0	-0.0056	0	0.00129	0	0
919	-0.35	-0.0158	0.13036	1	0	-0.0139	0	0	0	0	-0.0056	0	0.00129	0	0

File: addfile5.3_ems967.xlsx
Sheet: efms
Page: 14/14

	GlyOH	O2	CO2	X	Prod	Ala	Arg	Asp	Glu	Gly	Lys	Pro	GPC	Ac	Allantoin
920	-0.3218	-0.0382	0.17861	1	0	-0.0139	0	-0.044	0	0	-0.0056	-0.0032	0.00129	0	0.01488
921	-0.3293	-0.0158	0.20093	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
922	-0.3144	-0.0158	0.15628	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0.00129	0	0
923	-0.2988	-0.043	0.21003	1	0	-0.0139	0	-0.0473	0	-0.0532	-0.0056	0	0	0	0.01811
924	-0.3876	-0.0158	0.25924	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
925	-0.3435	-0.0158	0.12713	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0.00129	0	0
926	-0.2962	-0.0382	0.20422	1	0	-0.0139	0	-0.044	0	-0.0512	-0.0056	-0.0032	0.00129	0	0.01488
927	-0.3027	-0.043	0.2139	1	0	-0.0139	0	-0.0473	0	-0.0544	-0.0056	0	0.00129	0	0.01811
928	-0.2923	-0.0382	0.20034	1	0	-0.0139	0	-0.044	0	-0.0499	-0.0056	-0.0032	0	0	0.01488
929	-0.2808	-0.0158	0.13865	1	0	-0.0139	0	-0.0292	-0.0181	0	-0.0056	0	0.00019	0	0
930	-0.2798	-0.0158	0.13845	1	0	-0.0139	0	-0.0292	-0.0181	-0.0006	-0.0056	0	0	0	0
931	-0.316	-0.0158	0.15657	1	0	-0.0139	0	-0.0292	0	-0.0006	-0.0056	0	0	0	0
932	-0.2815	-0.0167	0.13961	1	0	-0.0139	0	-0.0297	-0.0175	0	-0.0056	0	0	0	0.00058
933	-0.2804	-0.0158	0.13903	1	0	-0.0139	0	-0.0292	-0.0181	0	-0.0056	0	0	0	0
934	-0.3166	-0.0167	0.15715	1	0	-0.0139	0	-0.0297	0	0	-0.0056	0	0	0	0.00058
935	-0.3166	-0.0158	0.15715	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0	0	0
936	-0.317	-0.0158	0.15676	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0.00019	0	0
937	-0.3347	-0.0158	0.21149	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0	0	0
938	-0.3351	-0.0158	0.2111	1	0	-0.0139	0	-0.0292	0	0	-0.0056	0	0.00019	0	0
939	-0.3341	-0.0158	0.21091	1	0	-0.0139	0	-0.0292	0	-0.0006	-0.0056	0	0	0	0
940	-0.3463	-0.0167	0.12742	1	0	-0.0139	0	0	0	0	-0.0056	0	0	0	0.00058
941	-0.3457	-0.0158	0.12799	1	0	-0.0139	0	0	0	0	-0.0056	0	0	0	0
942	-0.3341	-0.0167	0.20975	1	0	-0.0139	0	-0.0297	0	0	-0.0056	0	0	0	0.00058
943	-0.393	-0.0158	0.26979	1	0	-0.0139	0	0	0	0	-0.0056	0	0	0	0
944	-0.3936	-0.0167	0.26921	1	0	-0.0139	0	0	0	0	-0.0056	0	0	0	0.00058
945	-0.3934	-0.0158	0.26941	1	0	-0.0139	0	0	0	0	-0.0056	0	0.00019	0	0
946	-0.2808	-0.0158	0.13865	1	0	-0.0139	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0.00019	0	0
947	-0.3924	-0.0158	0.26921	1	0	-0.0139	0	0	0	-0.0006	-0.0056	0	0	0	0
948	-0.2804	-0.0158	0.13903	1	0	-0.0139	0	-0.0292	-0.0149	0	-0.0056	-0.0032	0	0	0
949	-0.2798	-0.0158	0.13845	1	0	-0.0139	0	-0.0292	-0.0149	-0.0006	-0.0056	-0.0032	0	0	0
950	-0.3105	-0.0158	0.15353	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0.00019	0	0
951	-0.3096	-0.0158	0.15334	1	0	-0.0139	0	-0.0292	0	-0.0006	-0.0056	-0.0032	0	0	0
952	-0.2815	-0.0167	0.13961	1	0	-0.0139	0	-0.0297	-0.0143	0	-0.0056	-0.0032	0	0	0.00058
953	-0.443	-0.0505	0.29124	0.55772	1	-0.0423	0	-0.0945	0	0	-0.0093	-0.0124	0	0	0.02776
954	-0.3101	-0.0167	0.15392	1	0	-0.0139	0	-0.0297	0	0	-0.0056	-0.0032	0	0	0.00058
955	-0.3101	-0.0158	0.15392	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0	0	0
956	-0.325	-0.0158	0.19857	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0	0	0
957	-0.3254	-0.0158	0.19819	1	0	-0.0139	0	-0.0292	0	0	-0.0056	-0.0032	0.00019	0	0
958	-0.3244	-0.0158	0.19799	1	0	-0.0139	0	-0.0292	0	-0.0006	-0.0056	-0.0032	0	0	0
959	-0.3397	-0.0158	0.12438	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0.00019	0	0
960	-0.3387	-0.0158	0.12419	1	0	-0.0139	0	0	0	-0.0006	-0.0056	-0.0032	0	0	0
961	-0.3244	-0.0167	0.19683	1	0	-0.0139	0	-0.0297	0	0	-0.0056	-0.0032	0	0	0.00058
962	-0.3827	-0.0158	0.2563	1	0	-0.0139	0	0	0	-0.0006	-0.0056	-0.0032	0	0	0
963	-0.3399	-0.0167	0.12419	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0	0	0.00058
964	-0.3839	-0.0167	0.2563	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0	0	0.00058
965	-0.3393	-0.0158	0.12477	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0	0	0
966	-0.2036	0	0.15849	0	1	-0.0345	0	-0.0504	-0.0301	-0.0274	-0.0062	0	0	0	0
967	-0.3833	-0.0158	0.25688	1	0	-0.0139	0	0	0	0	-0.0056	-0.0032	0	0	0