ABSTRACT

Title of Thesis:	MULTI-ELEMENT AND INSTATIONARY ISOTOPE-ASSISTED METABOLIC FLUX ANALYSIS FOR IMPROVED EVALUATION OF CELLULAR CARBON FLOW
	Shriramprasad Venkatesan, Master of Science, 2018
Thesis Directed By:	Dr. Ganesh Sriram, Chemical and Biomolecular Engineering Department

Isotope-assisted metabolic flux analysis (isotope MFA) combines experimental and computational techniques to analyze complex metabolic networks and thus to evaluate intracellular carbon flow and partitioning. However, the use of this methodology to increasingly complex networks can compromise flux identifiability and precision. In this thesis, we investigate two unique techniques that hold promise in overcoming this drawback.

The first of these is multi-element isotope MFA. Traditionally, isotope MFA has traditionally utilized only ^{13}C. Multi-element isotope MFA is a novel and untapped field of research, with less than a handful of studies which have utilized tracer elements other than ^{13}C. This thesis explores the possibility of tracking different elements simultaneously fed either as a mixture or as multiple

labels within the same molecule (multi-element tracer). Our analysis shows that multi-element tracking within a network allows for better information yield, which is necessary for resolution of fluxes more accurately for the network model. Also, our study shows that there are advantages to using several multi-element isotope labeling experiments in collaboration with parallel labeling experiments to provide better flux resolution at relatively less expensive costs.

Instationary analysis is another topic of interest for this thesis. In this study, we create an instationary framework in our native flux analysis software adapting prior work done in the field. We also implement the method based on existing guidelines on a Pt diatom to attempt to explain an existing controversy regarding its carbon concentrating mechanism. Our brief work supports earlier work claiming that Pt operates a C4 mechanism.

MULTI-ELEMENT AND INSTATIONARY ISOTOPE-ASSISTED METABOLIC FLUX ANALYSIS FOR IMPROVED EVALUATION OF CELLULAR CARBON FLOW

by

Shriramprasad Venkatesan

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2018

Advisory Committee:
Dr. Ganesh Sriram, Associate Professor and Committee Chair
Dr. Nam Sun Wang, Associate Professor
Dr. Taylor Woehl, Assistant Professor

Acknowledgements

None of this work would have been possible without the help of Dr. Ganesh Sriram who has been a very supportive guide and provided timely feedback helping to identify the problem statement and executing a full – fledged study on the topic. I would also like to thank my lab members for the academic support they provided. I am particularly thankful to Navadeep Boruah who has taught me all the necessary computational techniques when in need and helped me from time to time to expand this work in the right direction and to provide appropriate criticism on my work to help me mold it to the state it is currently. Finally, I would like to thank my family without whom I wouldn't have been able to complete my Masters.

Table of Contents

List of Figures

List of Abbreviations

EMU – Elementary metabolite units

ILE – Isotope Labeling experiment

MID – Mass isotopomer distribution

MFA – Metabolic Flux Analysis

$\widetilde{\widetilde{Cov}}$ – Covariance matrix of the error

$\widetilde{\widetilde{Hes}}$ – Hessian matrix of the error

Chapter 1: Introduction

1.1 Background

A major driving force for chemical engineers is to manufacture valuable chemicals at a large scale and thus, economically profitable manner. Towards achieving this, there have been countless innovations to optimize manufacturing practices, reduce waste and expand the sources of raw materials. The field of synthetic biology arose from this search for alternative sources and procedures of converting raw materials into products. There have been several advances over the past two decades in understanding and modifying the complex mechanisms occurring within the microorganisms to enable them to make a range of desirable products from a reasonably inexpensive raw material feed.

Modeling the biological organism plays a crucial part in helping us interpret the behavior of the system prior to modification. At a fundamental level, a cell is a hotpot of hundreds of intracellular metabolites interacting through thousands of reactions simultaneously. These metabolites which are the building blocks of the components of the cell are a direct indicator of the phenotype of the cell. Hence, gaining knowledge of the reactions at this scale is vital.

Isotope-assisted Metabolic flux analysis (MFA) is a tool which was specifically developed for this purpose. It has been a popular and rapidly evolving technique from its initial application during 1982[1]. As obtaining the kinetics of every single reaction is not practically feasible, this method analyzes non-kinetically the response of a few metabolite products to certain parameters to guesstimate the reaction rates also known as fluxes. By capturing a snapshot of the fluxes at various points, we could pressure-test the organism to various cellular and external stimuli and modify the organism accordingly to maximize the production of desirable chemicals.

The lack of kinetic data implies that the estimation of fluxes depends on the accurate measurement of the product metabolites experimentally and the availability of sufficient data for the estimation of the fluxes to be within the acceptable error limit. But this may not always be possible due to the influence of a variety of external factors and the scope of the model to be predicted. There have been several methods developed to improve the confidence of the flux estimates like modifying the choice of substrate fed to the system[1,2] as well as in adapting novel optimization methods to get a faster convergence rate[3,4]. In this thesis, we show in-silico that simultaneous monitoring of both the carbon as well as the oxygen rearrangements within the cell helps in improving the confidence of flux estimates.

1.2 Thesis Overview

As previously stated, there are two areas where the existing MFA methods could be expanded to increase the use-case scenarios – Flux Identifiability and Dynamic Labeling.

In Chapter 3 under our method of flux identifiability, we evaluate the possibility of using multi-element MFA that improves upon standard ^{13}C-MFA method. To do this, we generate synthetic labeling data for a model network and employ this data for identifying the fluxes under single-element and multi-element MFA. Further to this, we compare the error of estimation for different labeling data-sets.

Chapter 4 deals with our implementation of 'In-stationary Labeling Metabolic Flux Analysis'. Instationary labeling is a well-established and popular method of metabolic flux analysis. However, in this thesis, we have expanded its scope to include analysis of the marine diatom Phaeodactylum Tricornutum (Pt) for the first time. The simulation results that we have obtained help us to understand more about the phenotype of the organism under 'mixotrophic' conditions. Besides, our analysis has thrown open the complexities involved in employing instationary labeling method for Pt analysis. The list of complexities involved are enumerated towards the end of the chapter.

Chapter 2: Isotope-Assisted Metabolic Flux Analysis – An Overview

2.1 Experiment

Metabolic flux analysis involves the calculation of all the metabolic fluxes in a reaction network from experimental data obtained of certain metabolite products produced by the organism. To achieve this, the organic substrates typically fed to the system is labeled at specific atom positions with an isotope. Typically, non-radioactive isotopes such as ^{13}C, ^{15}N and ^{2}H are used as labels to replace the respective atoms within the substrate with ^{13}C being the most widely used. But there have been a few experiments which have even used radioactive isotopes such as ^{14}C under special circumstances.[5] After the cell is allowed time to metabolize the modified substrate into various metabolite products, the amino acids, soluble metabolites and fatty acids are extracted and analyzed through either nuclear magnetic resonance spectroscopy or mass spectroscopy (GC/MS)[1]. Due to the incorporation of the isotope into the metabolites, the measurement of the instruments occurs as a mass isotopomer distribution (MID). Typically, when the substrate is fed unlabeled, the MID reflects the natural abundance of the atom along with a distribution reflecting the effect of the derivatizing agent used prior.

2.2 Isotopomer and Cumomer Representation

A mass isotopomer (also known as isotopomer) is the isotopic variant of the metabolite arising from the incorporation of the isotope into the compound. As

every carbon atom of the metabolite (in the case of a carbon- labeled substrate) has a chance of being labeled, there are a total of 2^n isotopomers possible for a n-chained metabolite. They are typically represented as a binary number of n digits since the inception of the notation through the work of Wiechert et al[6]. For example, a 4-carbon metabolite with its second and fourth carbons from the left labeled is denoted as 0101 with a decimal equivalent of 10. Also, the weight of the isotopomer is defined as the number of labeled positions the metabolite has. So, in the above example, the weight of the isotopomer is 2.[6,7]

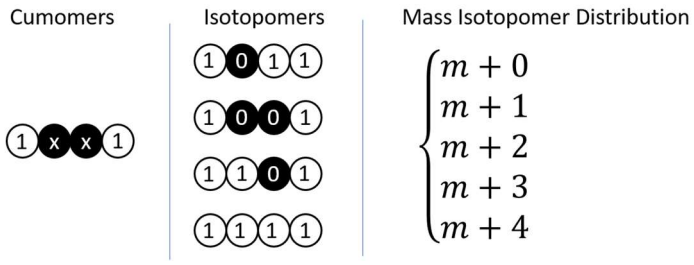

Fig. 2.1: Representation of conversion between isotopomers, cumomers and MIDs. The cumomer 1xx1 can belong to either of 4 different isotopomers which in turn can contribute to a MID based on their weights.

For large complex networks, the equations utilizing the isotopomer notation is a little unwieldy and increases the computational burden. To address this, there have been several alternate notation systems including cumomers[7], elementary metabolite units (EMUs)[8] and bondomers[9]. The program NMR2Flux+ used throughout this thesis uses the cumomer notation during its implementation and therefore, we will be referring to the same henceforth. In the cumomer system, each atom position can exist as either labeled (1) or an 'any' (x) mode. So, for the

same example used as in isotopomers, the isotopomer 0101 can belong to 4 different (xxxx, x1xx, xxx1, x1x1) cumomers.

By being represented this way, the number of cumomers is equivalent to the number of isotopomers (2^n), but the cumomers of equal weight can be grouped together for computation, hence reducing the computational burden. Therefore, this notation system helps us in handling the experimental data more easily when feeding it to the computational algorithm.

The network model which is created to simulate the organism's reaction pathway is a subjective and conscious curation of reactions. It is dependent on the reactions and metabolites which are relevant to the study and on the number of reliable fragment MID data which we could obtain experimentally. Realistically, a network model consisting of at most 40-60 reactions and the corresponding metabolites can be solved in a reliable and consistent manner.

The mass-balance equations of both the overall metabolites as well as that of the individual cumomers are solved simultaneously by the algorithm (Wiechert et al. 2001). The equations to consider are:

$$\frac{d\overline{M}}{dt} = \tilde{\tilde{S}}.\overline{v} \tag{2.1}$$

$$diag(\overline{M}).\frac{d^{n}\overline{x}}{dt} = \tilde{\tilde{A}}(\overline{v}).^{n}\overline{x} + \tilde{\tilde{b}}(\overline{v}, {}^{n-1}\overline{x}, {}^{n-2}\overline{x}, {}^{n-3}\overline{x}, \ldots) \tag{2.2}$$

where $\tilde{\tilde{S}}$ is the stoichiometric matrix of the network, $\overline{v}$ is the vector of network reaction fluxes, $\overline{M}$ is the vector of metabolite pool sizes (the quantity of each metabolite present), ${}^{n}\overline{x}$ is the vector of all cumomers of weight n. $\tilde{\tilde{A}}$ and $\tilde{\tilde{b}}$ are characteristic functions unique to a network and depending on the atom rearrangements of the metabolites occurring within the reactions. For a steady state assumption, the terms on the left side of the equations vanish making them algebraic equations which are easier to solve than the more expensive differential equations otherwise. Additionally, there are a few more equations to convert the cumomers to MIDs, but they are not represented as they are dependent equations and don't add to the complexity of the system.

The program NMR2Flux+ which is utilized throughout this thesis employs a bootstrap Metropolis-Monte Carlo algorithm for repeatedly guessing a set of fluxes which is used to simulate the MIDs of the metabolites which is compared to the experimental data to generate a sum-squared residual (SSR). The program then utilizes a simulated annealing method (Press et al.)[10] for global optimization of the system of equations followed by a Powell algorithm coupled with a Brent's root finding algorithm for further local optimization of the fluxes. The confidence interval while estimating the fluxes is a major topic which will be covered in Chapter 3.

The estimation could also be done in an instationary mode wherein we will need to solve the differential equations solved above. To do so, we will also need to obtain the experimental mass isotopomer distribution data at various time points.

Though, this method is much more computationally expensive, for some complex networks such as photoautotrophic organisms, this method provides valuable information which may not be available through steady state data. For example, for photosynthesis, the labeled $^{13}CO_2$ which is consumed as a substrate wouldn't yield any information as it labels the products uniformly at all positions whereas instationary analysis would provide us with information regarding how the CO_2 propagates through the network at various time points. This method has been utilized in several studies[11,12], but in Chapter 4, we have made our implementation of the method for the diatom Phaeodactylum Tricornutum to help identify the specific cycles it employs.

Chapter 3: Multi-element Metabolic Flux Analysis

3.1 Introduction

Quantitative flux estimation in a metabolic network is important for us to understand the cellular processes in a target organism to be able to optimize product yields. Metabolic flux analysis has been the gold standard among stable isotope tracing estimation methods. To date, ^{13}C has been the most popular isotopic tracer element used. Since carbon atoms form the backbone of all the organic metabolites, employing ^{13}C in metabolic analysis helps us to track their rearrangements during cellular processes providing us with a comprehensive picture. Besides, the methods and effects involved are exhaustively studied and used during the past decade making it a reliable and inexpensive label to use.

In certain cases, however, ^{13}C labeling may not be effective in determining the reaction fluxes. In networks involving oxygen metabolism like in respiratory cycles or in scenarios like nitrogen fixing environments, tracking the carbon atom rearrangements alone do not provide a complete picture of the undergoing processes. There are also certain chains of reactions wherein there is minimal rearrangement of carbon atoms like in the case of the TCA cycle metabolites. In such cases, the flux identifiability of the network suffers while performing a ^{13}C isotope labeling experiment (^{13}C-ILE) for most of the fed ^{13}C labels. So, it is prudent to consider using alternate elements for labeling along with ^{13}C for those networks. Elements like oxygen and hydrogen need not undergo the same rearrangements

as that of carbon despite them undergoing the same reactions. So, there is a definite possibility that they can provide us with data unable to be inferred from a conventional ^{13}C-ILE. Therefore, in this thesis, we investigate the usage of non-^{13}C tracers individually as well as in tandem with conventional ^{13}C-ILEs to verify the extent of additional information yield we could obtain.

There have been a few studies which have made use of alternative isotopic tracers in specific processes. Lippens et al. (2017) have done a few experiments using nuclear magnetic resonance spectroscopy on a simplified glycolysis and pentose phosphate network with both ^{13}C and ^{15}N as tracers[13]. Also, Zheng et al, 2013 have used a similar method on the marine diatom Phaeodactylum *tricornutum* to identify unusual pathway of urea metabolism using instationary analysis on ^{13}C and ^{15}N tracers[14]. More recently, Nilsson et al. expanded and applied this method for the analysis of human cells to show the feasibility of the experiment[15]. They introduced the concept of 'Two-dimensional metabolite isotope distributions (2D-MID) to enable simultaneous tracking of both carbon and nitrogen through the network.

Apart from using nitrogen, there have been alternate elements utilized as well for monitoring specific compartmentalized reactions. There have been studies which used deuterium labeling to monitor redox reactions involving the cofactor NADPH in the cytosol and mitochondrial sections of mammalian cells[16]. There have also been experiments involving labeled sulphur to examine the metabolism

taken by cysteine, methionine and homolanthionine along with other metabolites involved in isoleucine biosynthesis in a genetic variant of Corynebacterium *glutamicum*.[17]

These studies do inform us that there have been attempts at understanding the mechanisms of cellular processes from a non-carbon point of view. But these studies have been few and far between primarily because carbon is able to estimate fluxes in a network with adequate confidence for most common networks. Also, compared to carbon labeled tracers, the alternate labels need to be tailor made as per the lab's requirements driving their costs up. Despite this, alternate labeling has potential to be utilized under certain circumstances. This study tries to analyze these cases in a more quantitative way and to study the intricacies of doing a multi element analysis in-silico. Preferring to do a multi element analysis is entirely dependent on the scope of the study and other experimental and economic factors. To help with the decision making, we provide a set of factors to consider and illustrate how and when to approach performing a multi element flux analysis. This work is also motivated by the work of Nargund and Sriram, 2013 which identified designer labels specific for plant metabolites through a rigorous analysis of the effect of various parameters on the flux identifiability. We aim to do a similar analysis but for multi - element tracers.

For the scope of this thesis, we focus on tracking the labeling patterns of one other atom in combination with carbon. We chose oxygen for the study

primarily because unlike nitrogen or phosphorous which are situationally present in certain metabolites, oxygen and hydrogen are primary constituents along with carbon in most intracellular metabolites. This allows us to measure the isotope labeling pattern of the same metabolites which we conventionally observe during ^{13}C-MFA. Also, as there are far more hydrogen atoms compared to oxygen atoms, the tracking of hydrogen across every reaction is a lot harder without a much more rigorous network. This increases the complexity and thus, the computation times. Also, experimentally, the sensitivity and accuracy of MID fragment measurements would need to be a lot higher to account for the various proton rearrangements while measurement thus making hydrogen a non-ideal choice. Meanwhile, oxygen participates in fewer reactions and can be expected to provide us with information complementary to that obtained through ^{13}C-isotopomer labeling experiments (^{13}C-ILE). Also, as the weight of cumomer fragments of oxygen is similar in magnitude, the computation times would also be similar in duration to a ^{13}C-ILE.

Flux identifiability is the confidence with which the fluxes in a reaction network are identified by a given isotopic labeling experiment (ILE). It depends on several factors like the structure of the network, the rearrangement patterns of the metabolites within the reactions, the fragmentation and distributions of the metabolite products being analyzed, the labeling pattern of the substrate being fed to the system, the composition of labels and the errors involved in measuring. Each of these factors apart from the structure of the network can be altered during experimentation. The effects of these factors for a ^{13}C-ILE has been discussed by

Nargund and Sriram, 2013 but a few of these factors will be revisited during this study as well[2].

3.2 Methods

3.2.1 Combined mass isotopomer distributions

In order to enable in-silico assessment, we used a modified framework of our computational platform, *NMR2FLUX+* in C. To separately track carbon and oxygen rearrangements, the reactions were mirrored into two separate sets of reactions in the model with each set highlighting the rearrangement patterns of an element. As the two sets are representations of the same reactions, the common reactions are linked to each other to have the same flux value but allowed to have different cumomer fractions arising from the different rearrangements. Additional reactions pertaining to just a single element like the influx of water are not shared between the sets.

The typical experiment of MID data from the GC/MS is a combined labeling distribution from both ^{13}C and ^{18}O. To account for it, we use the concept of multivariate mass isotopomer distributions (MMIDs) introduced by Nilsson et al.,2016. A brief summary of the concept is as highlighted below. For the combined labeling of a metabolite G with e=1, 2…E be the different elements being tracked, then the MMID for the metabolite is represented as:

$$MMID_G = MMID_G^{e=1} \otimes MMID_G^{e=2} \ldots \otimes MMID_G^{e=E} \quad (3.1)$$

For example, for the metabolite Serine ($C_3H_7NO_3$), the combined MID is represented as a convolution of the two individual MIDs,

$$MID_{combined} = \begin{Bmatrix} m+0 \\ m+1 \\ m+2 \\ m+3 \end{Bmatrix} \otimes \begin{Bmatrix} m+0 \\ m+1 \\ m+2 \\ m+3 \\ m+4 \\ m+5 \\ m+6 \end{Bmatrix} \quad (3.2)$$

Convolution in this case provides the contribution of the individual MIDs towards the overall MID values. For example, in the case above, the m+0 component of $MID_{combined}$ occurs only through the contribution of both the m+0 of carbon and oxygen occurring simultaneously. Therefore, $(m+0)_{combined}$ is calculated by the multiplication of the values of $(m+0)_{carbon}$ and $(m+0)_{oxygen}$. Similarly, the $(m+4)_{combined}$ gets its contribution through either of several different combinations of $(m+0)_{carbon}*(m+4)_{oxygen}$, $(m+1)_{carbon}*(m+3)_{oxygen}$, $(m+2)_{carbon}*(m+2)_{oxygen}$ or $(m+3)_{carbon}*(m+1)_{oxygen}$. Therefore, the $(m+4)_{combined}$ is an addition of all of these terms. By doing so for every possible term of the combined MID, we can obtain the convolution of the individual element terms.

So, by convolving the MIDs obtained from each element, we can emulate the MID of a combined label.

3.2.2 Network models

We modeled the glycolysis, pentose phosphate pathway and TCA cycle to evaluate the multi-element labeling on a biologically relevant network. The reaction stoichiometries and the corresponding rearrangements for both carbon and oxygen atoms were obtained through the Metacyc database and manually curating the models for relevancy to the scope of the study. The network contains all major metabolites in the central carbon metabolism pathway along with lumped reactions for biosynthesis of amino acids. The model also accounts for influx and efflux of H_2O and atmospheric CO_2. The labels simulated in this model are based on commercially available tracers for ^{13}C. For ^{18}O, as there are only a couple tracers which are commercially available, we also considered a few custom-made tracer combinations which we feel could be made economically. Further, the ILE has also been corrected for presence of natural labeling of the substrate.

3.2.3 Network decomposition

The network decomposition was done using the NMR2Flux+ software. It uses the cumomer method of decomposition introduced by Wiechert et al. which has been briefly explained in Chapter 2.[6]

3.2.4 Simulation of isotope labeling experiments

For performing the identifiability analysis, we require simulated data of the MIDs for the respective tracers used. For the networks used, we identified a set of metabolites from the network whose isotopic labeling data could be measured by

mass spectroscopy (MS) and did a forward simulation of the ILE which assumes an existing set of fluxes for the network and solves the stoichiometric and cumomer equations to get a simulated set of MIDs.

3.2.5 Flux identifiability evaluation

The mathematical techniques for the evaluation of fluxes have been established previously. The method which we adopt have been used by Nargund and Sriram, 2013 to evaluate tracer combinations for a plant network. A brief outline of the concepts involved is illustrated below[2].

The sensitivity of fluxes with variance in isotopomer measurements can be identified through the covariance matrix of the fluxes. The diagonal elements of the covariance matrix ($\widetilde{\widetilde{Cov}}$) provide the variance of each flux. Therefore, the trace of $\widetilde{\widetilde{Cov}}$ is a good indicator of the identifiability of the fluxes for the provided tracer. We utilize the parameter which uses this concept A_{crit} to evaluate the tracers. Also, the determinant of $\widetilde{\widetilde{Cov}}$ termed as the D-criterion provides us valuable information regarding the variation of each individual flux. Therefore, the premise of an identifiability analysis entails guessing a set of i free fluxes (f_i) to compute the covariance matrix. Meanwhile, the covariance matrix is the inverse of the Hessian matrix of the chi-squared error ($\tilde{\chi}^2$) between experimentally measured and simulated values of the isotopomers.

$$\tilde{\chi}^2_{sim} = \frac{(I_{sim} - I_{meas})^2}{\sigma^2} \tag{3.3}$$

$$\overset{\approx}{Hes}(\tilde{\chi}^2_{sim}) = \frac{\partial^2 \tilde{\chi}^2_{sim}}{\partial f_i \partial f_j} \tag{3.4}$$

$$\overset{\approx}{Cov} = \overset{\approx}{Hes}(\tilde{\chi}^2_{sim})^{-1} \tag{3.5}$$

$$A_{crit} = \frac{tr(\overset{\approx}{Cov})}{n} \tag{3.6}$$

$$D_{crit} = \det(\overset{\approx}{Cov}) \tag{3.7}$$

where I_{sim} is the simulated set of MIDs. During the calculation of the Hessian matrix, the value the experimentally measured MIDs (I_{meas}) disappears. Therefore, we can obtain the value of A_{crit} without knowing the experimental measurements, allowing for an *'a* priori' analysis.

The value of A_{crit} which is defined as the arithmetic mean of the variances is inversely proportional to the sensitivity of the flux variances. Therefore, for simplicity, we evaluate its inverse coined as Information yield (A_{info}). A_{info} is an indicator of the amount of information we could obtain regarding the fluxes from the isotopomer measurements and the tracer label used.

3.2.6 Flux estimation

Once a suitable tracer has been identified, in order to test the simulated flux values, we input the simulated set of isotopomers generated previously. The

software NMR2Flux+ uses a bootstrap simulated annealing global optimization algorithm which assumes an initial guess value of fluxes and iteratively minimizes the $\tilde{\chi}^2_{sim}$ values.

3.3 Results

3.3.1 Simulated ILE in a simple network

We first illustrate our approach using a hypothetical network Consider the hypothetical network in Fig. 3.1. The network consists of six (A, B, E, F, G, H) metabolites and seven (v_0, v_1, …v_6) shared reactions including one reversibility (r_5) with an influx of water and efflux of a product metabolite C existing only in the oxygen network (see Fig. 3.1). The degrees of freedom for both networks is 2. So, the entire network is determinable if we know the value of two independent free fluxes. In this case, all the stoichiometric reactions can be rewritten with respect to the fluxes v_1 and v_2 representative of each of the 2 out of 3 pathways available to produce B from A.

To benchmark the flux identifiability, we first test out several ^{13}C - labels and ^{18}O - labels on the network individually. The metabolite fragments being observed determines the elements of the covariance matrix and therefore, we keep the common metabolite fragments for both ^{13}C and ^{18}O ILEs made on this network consistent. Fig.3.2a shows the information yield for the various tracers tried. There was no information gained from A being labeled individually at the positions 1,2,3 and 4. This is as expected as the cumomer fractions obtained from these labels

are not affected by the flux values in the network and the same MID will be achieved irrelevant of which pathway the process goes through. The maximum yield was obtained for ^{13}C-ILEs is from 1,2 ^{13}C-A substrate being fed.

We observe that due to the different rearrangements occurring in oxygen atoms, the information yield from each ^{18}O-label is varied compared to their ^{13}C counterparts. Overall, the information gained from ^{18}O-ILEs is significantly higher compared to their carbon counterparts. This can be attributed to the presence of an additional metabolite product C being measured. As metabolite C contains only oxygen atoms in its' structure, it would be 'invisible' when performing a ^{13}C-ILE. Here, the maximum yield observed was from 2,3,4-^{18}O A being fed to the system.

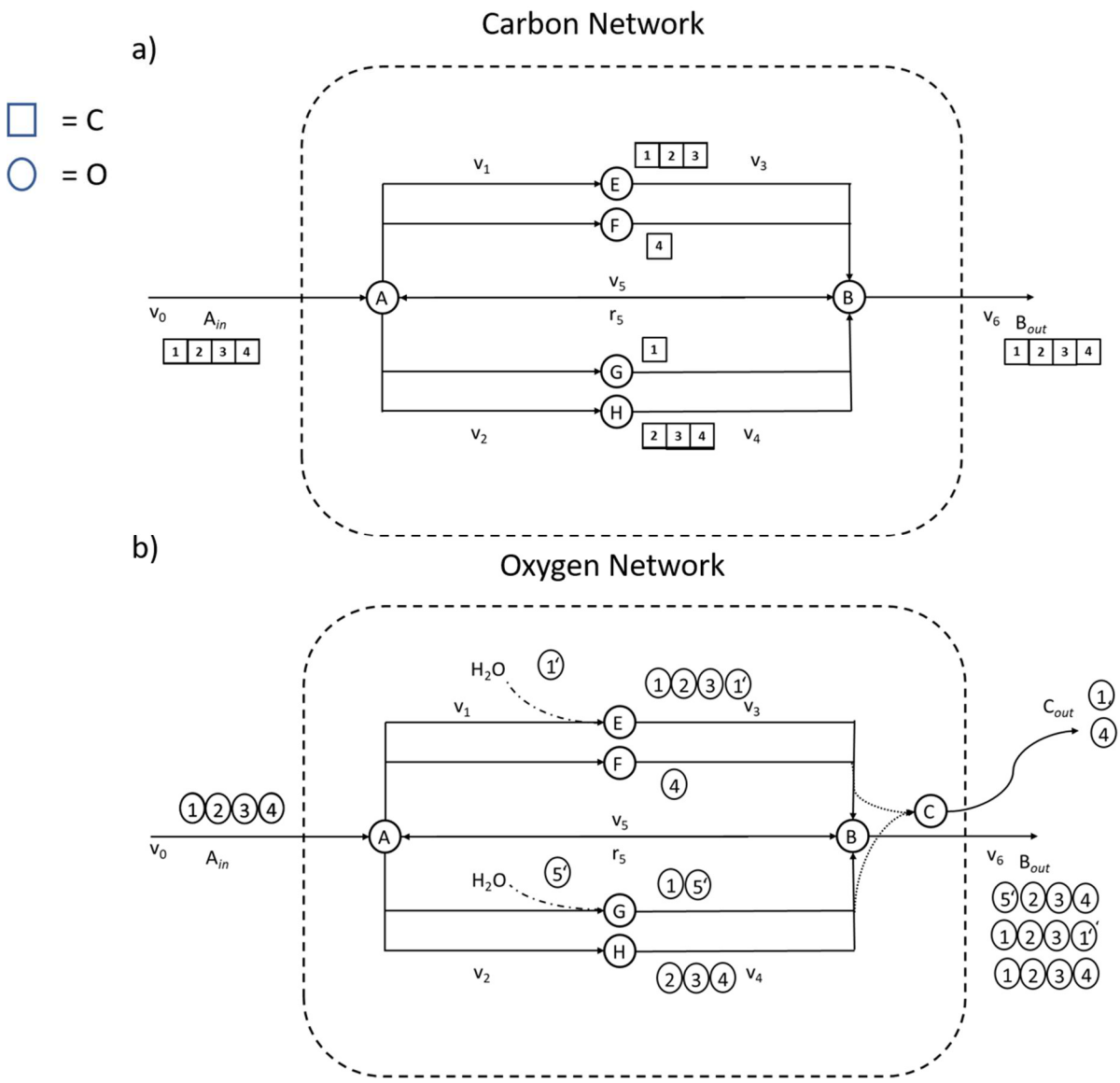

Fig. 3.1: Metabolic reaction map for a hypothetical network. Part a) represents the carbon atom rearrangements whereas b) provides the oxygen atom rearrangements. The metabolite C and H_2O contain only oxygen atoms. Only the rearrangements of the metabolites after the first pass is shown for clarity. Due to the presence of the reversible reaction r_5, there are additional configurations possible in subsequent passes. Note that due to the differing rearrangements, there are 3 configurations of B in the oxygen network whereas only 1 is possible in the carbon network even in the 1st pass.

Now that we have obtained a list of tracers from both elements which can solve their respective element networks, it makes sense that a tracer which has been labeled both with ^{13}C and ^{18}O would perform better. To test it out we created a combined network having both the reactions of both elements. As there is no interaction between the two networks, testing out a single-element tracer on this combined network would not provide enough information to solve both the networks simultaneously and thus, have not been simulated for the combined network. Also, the isotopomer distribution of the measured metabolites A and B obtained would be a combined MID with contributions from both elements. For simplicity in representation, we took the six best tracers from each single-element ILE performed (see Fig 3.2a) and performed the analysis with their combinations to form a multi-element tracer.

From the map in Fig. 3.2b, we can infer that the determination of the optimal multi-element tracer is non-trivial and is not a simple case of choosing a combination of the best tracer for each element. In the case above, we can see that the combination of the best tracers from each element (1,2-^{13}C and 2,3,4-^{18}O) provide only 43% of the maximum yield we could obtain for this network. This is because there is a dilution effect like that experienced in a ^{13}C - combined labeling experiment because of the combined MIDs taking the contributions from both oxygen and carbon. Also, another contributing factor is due to the A-criterion being a trace of the variances of the individual fluxes, if both the single-element tracers are considerably better at estimation of one of the free fluxes than the other, they

tend to mask each other leading to the other free flux not being estimated very well reducing the overall A-criterion value. This could be verified by performing a backward simulation to determine the free flux variances. Fig. 3.3a. displays a free flux scatter plot of two different multi-element labels, namely A labeled at1,3 ^{13}C-1,2 ^{18}O (label 1) and at 1,2-^{13}C 4-^{18}O (label 2). We generated simulated MIDs for both combination of labels assuming v_2 = 0.4 and v_4 = 0.4. Then we did a backwards estimation of the fluxes by perturbing the simulated MIDs 50 times. We can see that while label 1 is able to estimate v_2 accurately, the confidence with which it is able to predict v_4 is poor. Meanwhile, label 2 has a very good confidence estimate of both v_2 and v_4, thus correlating well with its expected behavior according to its A-criterion value.

To verify if indeed, a multi-element tracer was better than their single-element counterparts, we did a backwards simulation on both a combined network and each individual network. For this example, we took a 1,2-^{13}C 4-^{18}O label as it displays the minimum masking effect among the list of labels checked. Fig. 3.3b. shows that individually taken, v_4 is evaluated better by an oxygen ILE whereas v_2 is predicted well by a ^{13}C-ILE, though outpacing the other just by a mild margin. In the case of the multi-element ILE, it simulates v_4 better than both the individual labels, whereas it is very similar in estimation of v_2 to the single element tracers. This shows that a multi-element ILE is better in estimation owing to its better accuracy overall.

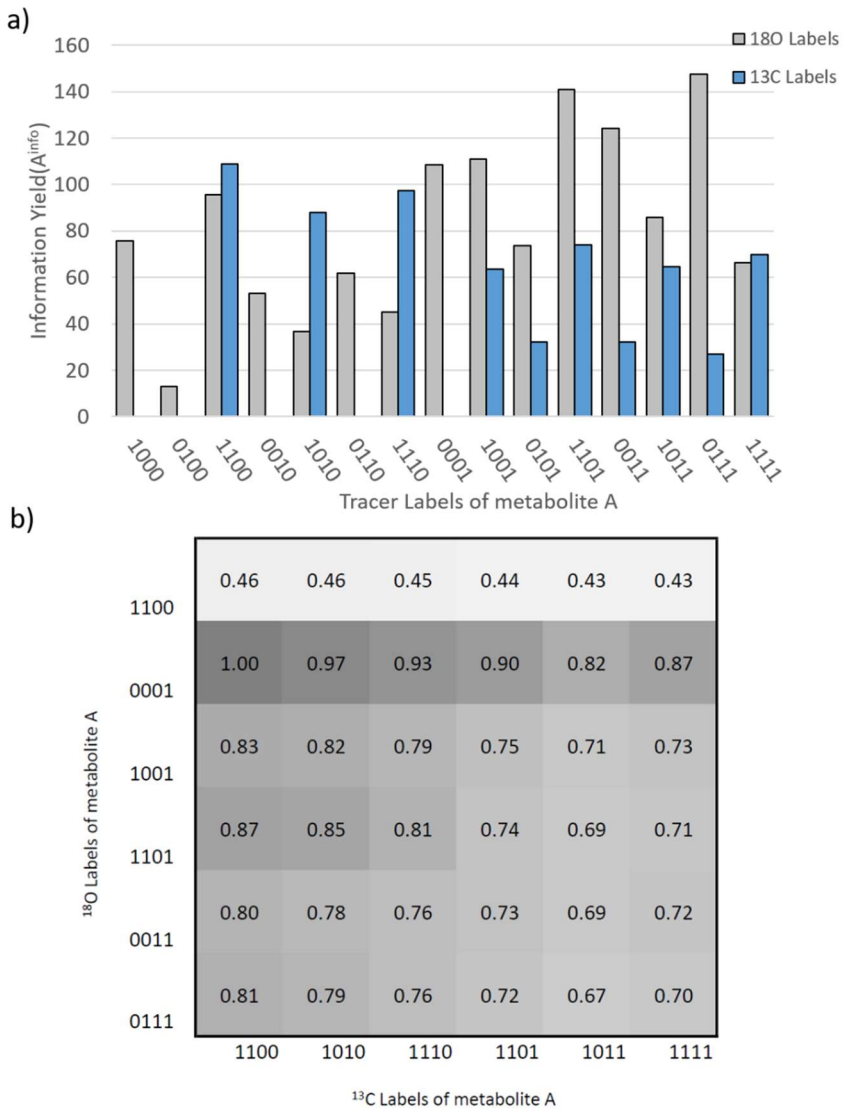

Fig. 3.2: Flux identifiability analysis of a hypothetical network. Part a) shows a comparative plot of the information yield obtained for the network through single-element tracking of the metabolites. Each of the labels are at 50% purity. Part b) provides a heat map of the information yield obtained from a multi-element ILE with each tile representing a 50% purity of the six best tracers for each element. The values within each tile depicts the relative information yield which is the ratio of the A_{info} of each combination to that of the maximum yield obtained from these multi-element ILEs. In this case, the denominator is that of 1,2-^{13}C 4-^{18}O A which is the label with highest information yield possible in the network.

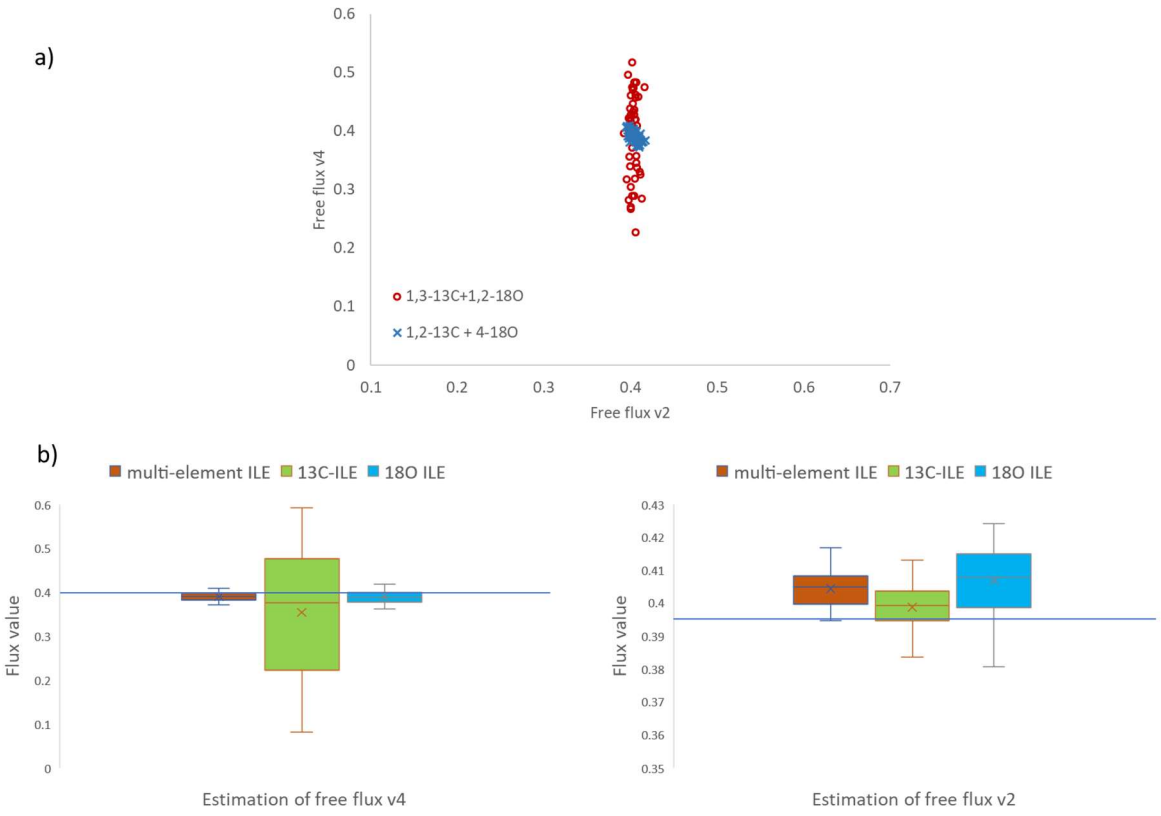

Fig. 3.3: a) Comparison of Flux values of two different combination of labels. The free fluxes v_2 and v_4 are estimated by the global optimization solver 50 times for every label (True value: v_2=0.4, v_4=0.4). This figure shows how for one set of combined labels, there is complimentary information leading to better convergence whereas in the other, the inability of estimation of one flux carries over to the combined label leading to ineffective convergence. b) Comparison of a multi-element label and their corresponding single-element labels. The label considered in this case is 1,2-^{13}C A and 4-^{18}O A which gives the best information yield in the combined network.

Finally, from the list of obtained A_{info} values for our multi-element ILEs, it is revealed that while the exact combination that contributes to an optimal tracer is

hard to determine, there is an inclination to favor the combinations where both the single element tracers were sufficiently good at determining their individual networks. So, for searching for combination labels for more complex models, we could search from a much tighter solution space by performing separate single – element ILEs and choosing a selection of the better tracers from each.

3.3.2 Implementation on a simulated central carbon metabolism network

Thus far, we have established the unique advantages that a multi-element ILE provides and demonstrated the feasibility of doing a multi-element isotope labeling experiment. Therefore, we wanted to extend our analysis to a more practically useful network. The central carbon metabolic pathway is a network which is present in a variety of organisms and provides most of the primary metabolic products necessary for the survival of the organism like amino acids, fatty acids and nucleotides. Therefore, it is an important network to understand for any metabolic study. Until now, there have been several in-depth ^{13}C-MFA studies made on the network, both in-vitro as well as in-silico. Our study aims to expand our understanding through a multi-element labeling experiment on a more realistic network which we believe could provide some unique insights as to the type of tracers that could aid in better flux estimation. Also, obtaining similar results to the hypothetical network is an indication that this method would work better even for more complex scenarios and not just a serendipitous result.

For this study, we manually curated a list of reactions and metabolites which are important and that a multi-element ILE would be insightful on. It consists of the basic glycolysis, pentose phosphate pathway and the (tricarboxylic acid, Krebs) TCA cycle reactions. As the Krebs cycle is an important part of aerobic respiration in cells, doing an oxygen ILE could reveal a lot more information than a carbon ILE. The pentose phosphate pathway and the latter half of the TCA cycle were assumed to be reversible. The amino acids Alanine, Serine, Glutamine, Malate, Aspartate and Glycine form the list of metabolite products formed from this network assuming glucose and unlabeled CO_2 being the only sources of carbon entering the system. There is also an unlimited supply of unlabeled H_2O assumed for this network which is typically the case in most lab-grown culture studies. There is a total of 31 reactions and 15 non-product metabolites involved in this network. The total degrees of freedom in this network equals 2 which are the reactions of glucose-6-phosphate entering the pentose phosphate pathway (*g6pdh*) and the reaction involving pyruvate entering the TCA cycle through oxaloacetate(*anaf*). For any forward simulations to simulate MIDs of the metabolite products, we assume all reversible reactions to be operating at 50% and the free fluxes to be taking 50% of the input flux to the reaction. As the flux values do affect the isotopomer distributions, these values are kept constant across all the ILEs performed.

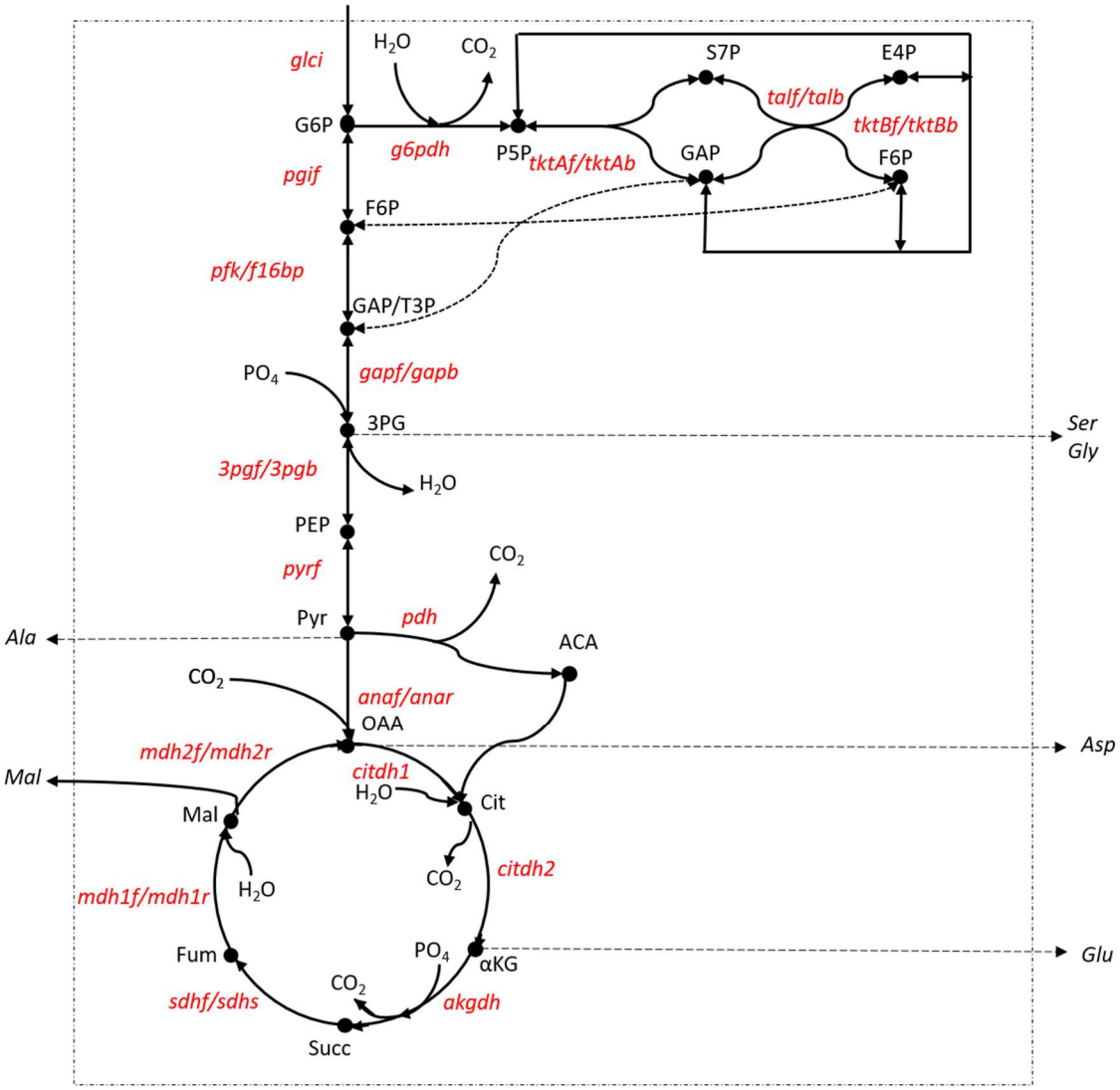

Fig. 3.4: Reaction network diagram for a glycolysis, pentose phosphate and TCA cycle model. All the metabolites coming out of the system (dashed lines and Mal) are measured during the ILE. H_2O, PO_4 and CO_2 which initially enters the system is assumed to be unlabeled. Both the pentose phosphate as well as the latter half of the TCA cycle is assumed to be reversible. The names of the fluxes are as shown(red). The free fluxes in this network are the reactions *g6pdh* and *anaf*.

Like the previous analysis on the hypothetical network, we initially verify the viability of the single-element tracers on the network. As we are dealing with a realistic network, it is more useful to test out commercially available tracers compared to the more convoluted ones which need to be tailor-made. Therefore, we tested the tracers which were labeled at either one or two positions on the substrate backbone for simplicity.

There were 28 different labels from each element analyzed for the single element network model. The results for the single element ILEs performed are as shown in Fig. 3.5a. and Fig. 3.5b. Compared to the hypothetical network, the rearrangements are not drastically different from oxygen to carbon in this network. Yet, as observed in Fig 3.5, the information yield is still higher due to the contributions of the unlabeled H_2O, PO_4 and CO_2 towards providing more valuable data.

As before, we selected the six best tracers from each single-element experiment to implement in a multi-element analysis. Fig. 3.5c illustrates the 36 different combinations of labels possible and the relative information yields obtained. For this network and the metabolites being analyzed, we found that the most information could be obtained for a 1,6-^{13}C 3,4-^{18}O glucose. The non-linear correlation between the choices of tracers used and the respective information yield is similar as in the case of the hypothetical network. This demonstrates the

need for us to perform an in-silico analysis and tailor our multi-element tracers wisely before experimentation for each individual network.

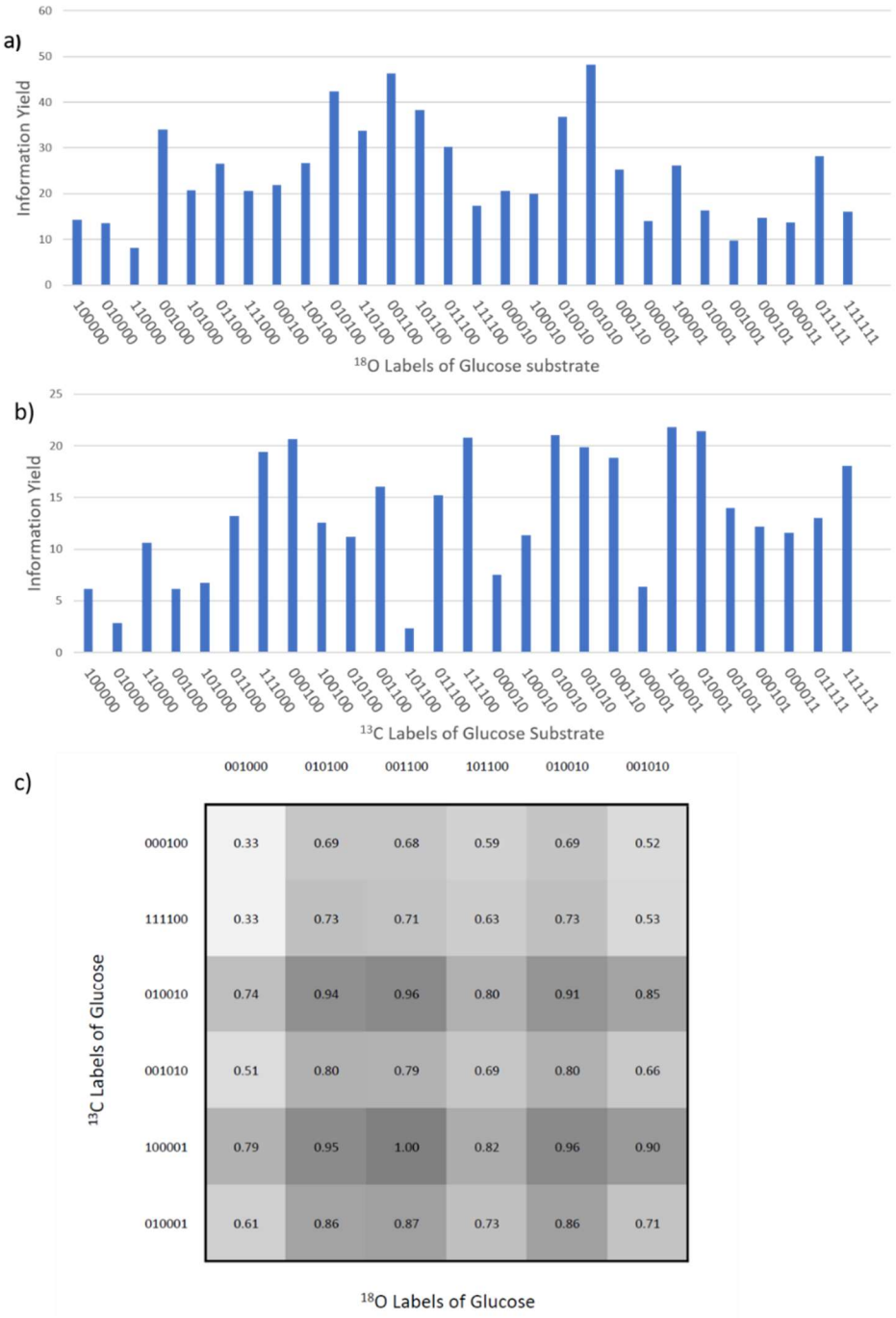

Fig 3.5: Information yield data obtained for combination of tracers. Part a) represents the various ^{18}O tracers whereas part b) displays for the various ^{13}C labels. Part c) provides the relative information yield values which is the ratio of A_{info} values for each label vs. the maximum possible yield. In this case, the denominator is 1,6-^{13}C 3,4-^{18}O glucose which provides maximum information yield for the network.

3.3.3 Comparison to parallel labeling experiments

A parallel labeling experiment is generally done with two or more replicates of the same culture simultaneously in separate flasks, but which are fed different labels. The data which is obtained from each is combined for flux estimation. It is a well-studied and utilized method for providing complimentary information to increase the resolution of fluxes significantly (Crown SB et al.,2012)[18].

From a modeling and resolution point of view, there is just a minor difference between multi-element and parallel labeling experiments. Whereas in a multi-element experiment, there are several separate networks which are solved simultaneously to provide a single set of data, in parallel labeling, a single network is solved in tandem to provide different sets of data. Owing to the very similar method of computation and no interactions occurring between the carbon and oxygen networks, in-silico analysis of a multi-element experiment is carried out very similar to that of a parallel labeling experiment. Therefore, based on simulated data, there is very minor differences between a multi-element ILE and a parallel ILE. In effect, this means that an ILE with two single-element labels provided in combination or as labeled within the same molecule is identical to an ILE with the single-element labels provided in parallel. This also implies that when impractical or uneconomical, it is better to perform a parallel labeling experiment with different

single-element labels (e.g. ^{13}C and ^{18}O), as it provides the same information as a multi-element ILE. Also, the conclusions that we gained from the simulations we ran for a multi-element ILE still holds true for a parallel ILE with more than 1 elements being tracked. However, there may be differences experimentally between both the ILEs which could lead to different flux estimation. For example, as a multi-element ILE has only a single set of MID data which is a convolution of the single-element MIDs, the standard deviations of measurements may vary when compared to a parallel labeling experiment with two measurements. But, verifying that is beyond the scope of this study.

Given that, it is also possible to combine both their benefits by conducting separate multi-element ILEs in parallel. To verify this, we ran a simulation using the two best tracers we obtained from the previous analysis on the central carbon metabolism network (3,4-^{18}O 1,6-^{13}C glucose and 2,5-^{18}O 1,6-^{13}C glucose). We initially performed the multi-element ILE on each of these combinations separately and then simulated a parallel experiment with both labels.

Fig. 3.6 illustrates our simulations of both a parallel multi-element ILE as well as an individual multi-element ILE. It displays the residual of the intracellular fluxes in the network as estimated by the program as opposed to the true value. As observed, the fluxes are estimated closest to the true value in the case of the parallel labeling experiment. Both the individual multi element labels behaved

similarly with mildly better resolution achieved in the case of 3,4-^{18}O 1,6-^{13}C glucose which corroborates with the prediction from Fig. 3.5c.

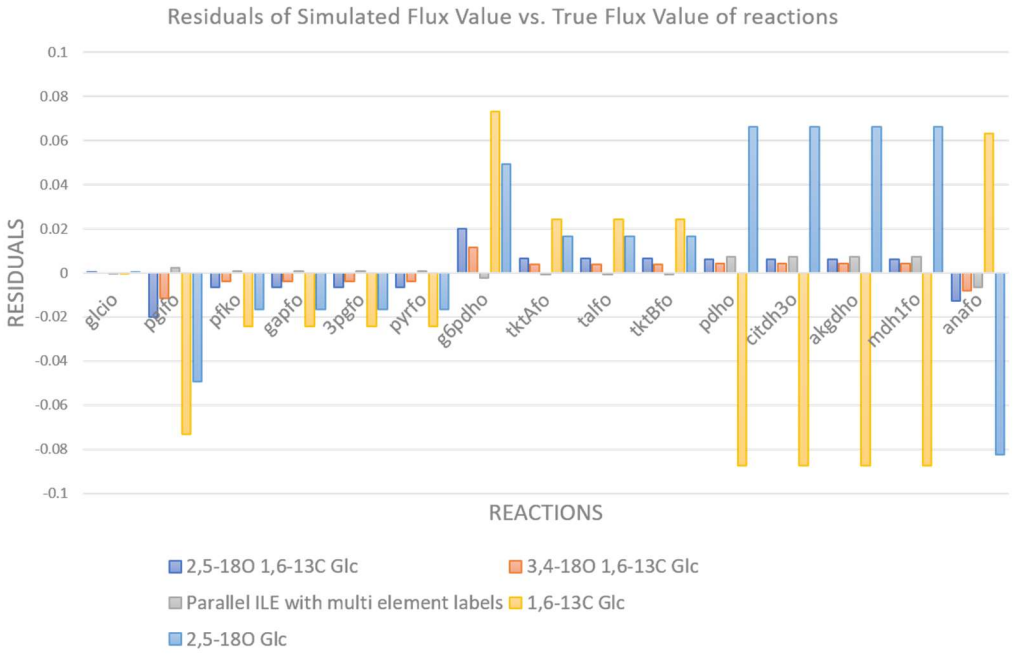

Fig 3.6: Residual chart comparing single element, multi element and parallel ILEs for the realistic network using the two most optimal multi-element tracers. A value of 0 implies that the simulated flux values match exactly with the true values of the intracellular fluxes. The signs indicate whether the simulated flux values are higher or lower than the true values and do not provide any useful conclusions for this study.

Further, the single element tracer estimates relatively poorly with the 13C tracer resolving much worse than the others. As far as the reactions are considered, both the single element tracers primarily fail at estimating the fluxes in the TCA cycle. This may be due to the primarily unaltered rearrangements between the metabolites in the TCA cycle. Succinate being a symmetric molecule, along with the additions of H_2O and CO_2 provide the only sources of differing

rearrangements. The multi element ILEs mitigate this problem by combining both the individual element labeling information to constrain their bounds of evaluation better.

3.4 Discussion

3.4.1 Practical Impact

Metabolic Flux Analysis has grown a lot in terms of scope and technological capability over the past decade. With improvements made in the estimation algorithms used, it has pushed the boundaries of the complexities of cellular networks that it can measure. Now that it has neared a plateau, it is important to look for alternative perspectives to improve flux estimation as well. This study on multi-element flux analysis is an attempt to push that boundary further using alternate element tracking which have been strangely absent in modern metabolic flux analysis studies despite being a very lucrative option. Our rigorous computational study shows that a multi-element flux analysis indeed holds promise as far as better flux resolution is concerned.

Though this study hasn't incorporated any experimental results to verify the practical reliability of a multi-element study, our results both on a hypothetical network as well as a realistic network show that an alternate element can be used as an alternative to ^{13}C in some cases. There are several instances where this is very useful as we have demonstrated using the hypothetical network as a plausible

scenario. Even in the real world, there are several complex networks where this holds true as well. In photo-autotrophic organisms which take in HCO_3 to produce their food source, a normal ^{13}C labeling experiment may not provide a complete picture. But when used in tandem with other elements like oxygen or deuterium may reveal far more information. We haven't tested our method on such a network, but it is a future area of study which could be significant.

Further, we showed that the design of tracers for a multi element (single label, two elements labeled) or combined labeling (more than two labels, single-element labeled) experiment is not a trivial task of choosing the best tracers from each single element labeling experiment. Rather, there is a careful consideration to be made to check for the complementarity of the labels towards one another. If both estimate a certain free flux parameter poorly, the combination of both cannot be expected to resolve the flux any better than their components. This is not always apparent from the information yield of the single-element labeling experiments and thus, the design of a multi element tracer needs to be chosen with greater care.

Our results also demonstrate the plausibility of a multi-element parallel and simultaneous labeling experiment. The results show that both these methods have a much better estimation rate of fluxes than a single element method. This is especially significant in the case of a parallel experiment because this implies that in cases where a multi-element tracer is prohibitively expensive or need to be tailor

made, single element experiments can be carried out in parallel to fulfill our experimental needs.

3.4.2 Cost of Tracers

The study we have done shows that there is an untapped potential in utilizing alternate elements apart from ^{13}C to use in metabolic flux analysis. But there is also the aspect of economics to be considered when bring it to practicality. Especially, considering we suggest using an alternate atom labeling, many of which not available for sale commercially, they would need to be custom made driving up the costs. Nevertheless, considering the benefit of performing a multi-element ILE, there is an acceptable limit of expense versus utility that is to be evaluated before performing a multi-element ILE.

To understand the cost of doing a multi-element ILE, we obtained the prices of several different labels available commercially. All prices were obtained from Omicron Biochemicals Inc. We did the standard forward simulations on both the individual labels as well as in a parallel labeling experiment with each element tracer in an experiment to obtain the information yields for each. Further, we plotted the prices of each label against their corresponding A_{info} values.

As we do not have sufficient information on the commercial labels available for the alternate element labeled tracers, it was not possible for us to plot all the possible points of cost versus utility on the Pareto front graph. Despite that, the

available points gave us some good points to consider when selecting our designer tracers.

The Pareto efficiency is a metric to evaluate a set of actions which is dependent on more than one parameter. A 100% Pareto efficiency would indicate that beyond that point, the optimization of one parameter would lead to compromises in the other whereas a low efficiency implies that we could improve one parameter with no impact on the other. For our study, we illustrated a Pareto front of the parallel labels which could lead to maximum information yield at the lowest cost possible as given in Fig. 3.7

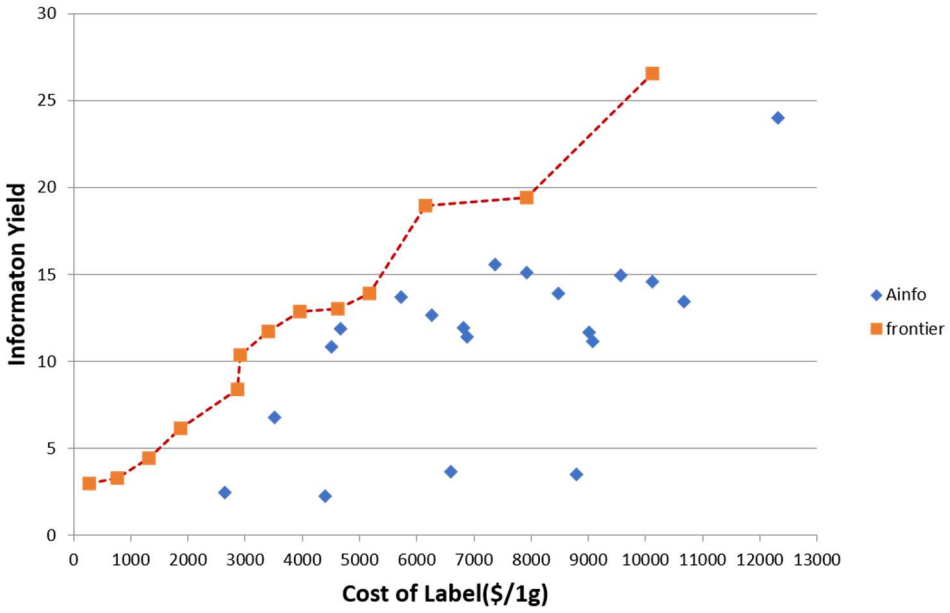

Fig. 3.7: Pareto frontier plot of parallel labels with multiple elements. The factors they are evaluated are the combined cost of both tracers used and the information yields obtained for the realistic network we used. The 100% Pareto efficiency labels are shown by the red dotted line.

From the figure, we can observe that as the information yield increases, the respective cost for purchasing the labels also increase. Starting from a minimum of $220 for a 1-^{13}C glucose label, the minimum cost for better resolution of fluxes increases nearly linearly. Any label which is below the Pareto frontier is not optimal for the money invested on it. So, this gives us a clear idea on what to expect of an optimal tracer as far as resolution of fluxes is concerned.

3.5 Conclusions and Future Work

We have shown through this study that a multi element study could be made to mitigate the problem of poor resolution of fluxes in ^{13}C ILEs. We demonstrated how to design a multi element tracer and evaluated its performance compared to single element tracers. We have also compared how a multi element labeling experiment performed in parallel could help combine the benefits of both multiple element tracing and parallel experimentation.

We also provided a brief economic analysis based on the available information regarding commercially sold labels and provided a way to decide when a multi element tracer could be economical for the study and, to evaluate the labels based on both their costs as well as their ability to ability to resolve fluxes effectively. This is to be used as a guideline for any experiment involving more than one element or when accurate values of intracellular fluxes are required for the analysis.

From our analysis, we can provide a list of guidelines to choose a suitable tracer in-silico to perform a multi-element analysis.

1. Choose a list of available commercial substrate labels available.
2. Determine the information yields obtained through single-element tracer experiments for the elements to be tracked as per the equations outlined in Section 3.2.
3. Choose a suitable threshold limit for the information yield desired. Choosing this threshold is subject to the sensitivity of the flux analysis as well as the budget of the experiment and varies on a case-by-case basis.
4. Select the single-element tracers which are beyond the threshold limit and perform an in-silico multi-element analysis with each possible combination of the tracers to determine the combined information yields.
5. Perform a Pareto analysis similar to Fig. 3.7 to obtain the cost versus information yields obtained for the chosen network and substrate.
6. Depending on the economic limitations associated with the study, choose an appropriate tracer from the list of combinations which lie on the Pareto frontier.

Provided that, there are also a few ways by which this area of research could be moved forward. As this is an in-silico analysis, there have been no specific scenarios considered during this study and have been largely based on a prokaryotic cell network. For future work, this method could be utilized in cases like anaerobic glycolysis occurring in cancer cells wherein anaerobic production of ATP occurs even in the presence of oxygen.[2] This has been a well-researched topic, but due to constraints in flux estimation through MFA, fitting a model to a more realistic tumor cell network has been difficult. By utilizing multi element MFA methods, the flux estimation could be improved significantly along with tracking of ATP molecules to quantify the alternate pathway fluxes and understanding the behavior better.

Also, we have shown only the use of ^{13}C along with ^{18}O in this study for ease of explanation of the method, but this could be used along with other elements as well depending on the network requirements and measurement data accessible. With the current methodology of tracking the elements and the optimization algorithms used, the computational burden is increased significantly. This is especially the case for elements like hydrogen which are much more abundant compared to carbon in carbohydrate molecules. Therefore, the number of cumomers or EMUs required to be calculated are several orders of magnitude higher for hydrogen tracking. This is an area which could be addressed by careful design of deuterium tracers. As previously cited, there has been a previous study using deuterium tracers to track NADPH which participates in very few reactions[16].

Therefore, the goal of a future study should be to identify similar metabolites which have the capability to provide complimentary information without sacrificing computational cost for hydrogen.

Chapter 4: Instationary ^{13}C Metabolic Flux Analysis

4.1 Introduction

Recently, instationary ^{13}C metabolic flux analysis method has gained interest since it addresses some of the critical problems of the stationary method during the implementation stages. Especially, in the case of autotrophs which make use of single atom metabolites like CO_2, a stationary metabolic flux analysis wouldn't yield much information due to uniform accumulation of the labeled carbon in all the metabolites. Instationary analysis makes use of time-variant datapoints rather than steady state datapoint and hence, it can track the rates of accumulation of label in metabolites over time. This allows observation of the contributing factors to the production of certain metabolites and the preference of certain networks towards specific cycles. This will enable us to gain deeper understanding of the behavior of the cell and new perspective into ways of engineering the cell lines.

One of the initial works on instationary metabolic flux analysis has been the work by Wiechert et al.[7] which introduced the method of approaching an instationary method and to provide respective guidelines to follow both experimentally as well as computationally. There have been several studies later using instationary state analysis method which highlight several important deductions which were not

possible previously. Jazmin et al. applied an isotopically instationary flux analysis on a photoautotrophic organism (Arabidopsis *thaliana*) to map their network by feeding labeled $^{13}CO_2$. Due to the substrate being a single carbon molecule, performing a stationary MFA would lead to all the metabolites being uniformly labeled leading to no information gain.[19] Also, there have been studies on performing instationary MFA on a high yield strain of E.*coli* to analyze fermentation process to produce 1,3-propanediol.[20] Even in our group, there have been studies which utilize an instationary MFA to analyze the operation of Phaeodactylum *tricornutum* (Pt) by feeding labeled $H^{13}CO_3$. The method used measurements at various time points and fit to a stationary network model to approximate the MIDs and to draw inferences from it.[21]

One problem which is faced during this method is that the computational load is much higher than that of the normal stationary method.[7] This occurs because the normal algebraic equations,

$$0 = \tilde{\tilde{A}}(\overline{v}).{}^{n}\overline{x} + \tilde{\tilde{b}}(\overline{v}, {}^{n-1}\overline{x}, {}^{n-2}\overline{x}, \ldots) \tag{4.1}$$

used in stationary method, is replaced by non-linear differential equations.

$$diag(\overline{M}(t)).\frac{d^{r}\overline{x}}{dt} = \tilde{\tilde{A}}(\overline{v}).{}^{r}\overline{x} + \tilde{\tilde{b}}(\overline{v}, {}^{r-1}\overline{x}, {}^{r-2}\overline{x}, \ldots) \tag{4.2}$$

Here, $\tilde{\tilde{A}}$ is a matrix of the fluxes between cumomers of the same order ($^{r}\overline{x}$) whereas $\tilde{\tilde{b}}$ is a matrix of fluxes with cumomers of lower order than the one being evaluated currently on the LHS. For reactants, the flux values are given a negative

sign and a positive sign given for products by convention. $\overline{M}(t)$ is the vector of metabolite pool sizes of each cumomer metabolite.

Due to the significant increase to the computational load to solve the differential equations, true instationary flux analysis is limited to very small networks. To surpass the problem, studies on instationary MFAs typically make a few assumptions. Studies have shown that organisms could be brought to a pseudo steady state wherein the system is at metabolic steady state, but isotopically non-stationary.[22] Metabolic steady state condition implies that the organism is allowed enough time after the substrate is fed, so that the reaction fluxes(*v*) reach a steady value and there is no accumulation of intermediate metabolites within the cell. This also means that the metabolite pool sizes(M) which are the concentrations of metabolites present in the cell are constant with time. So, the resultant equations boil down to:

$$diag(\overline{M}).\frac{d\overline{x}}{dt} = \tilde{\tilde{A}}(\overline{v}).^{n}\overline{x} + \tilde{\tilde{b}}(\overline{v}, {}^{n-1}\overline{x}, {}^{n-2}\overline{x}, \ldots) \tag{4.3}$$

Analyzing at isotopically instationary conditions implies that the isotope concentrations within the cell is still time variant. This means that the analysis starts from time t=0 at which point, the isotopic label is fed instead of the unlabeled substrate. This isotopic instationary operating condition is much simpler to implement under this assumption, yet more computationally demanding than a fully steady state scenario.

The diatom, Pt, is a mixotrophic organism which uses photosynthesis for its survival. Photosynthetic carbon assimilation has been done primarily through the Calvin-Benson-Bassham cycle which uses an enzyme Ribulose 1,5-bisphosphate carboxylase (RuBisCO) to catalyze CO_2 carboxylation. The rate of fixation depends on the concentration of CO_2 and relative amounts of O_2 present around the enzyme. But, due to the low amounts of solubility of CO_2 in water (1.45g/L), the dissolved CO_2 present in the vicinity of RuBisCO is much lesser than the Michaelis constant for the enzyme. This leads to the cycle performing below its optimal performance limit.[21]

To overcome this, some plants use a carbon concentrating mechanism (CCM) which converts entering CO_2 into an intermediate, pumps it into a separate organelle and converts the intermediate back into CO_2 which is trapped in the vicinity of RuBisCO by a CO_2 impenetrable membrane. One of the CCMs is a biophysical pumping mechanism which converts CO_2 and water into a HCO_3 anion which is a transported across the membrane by a transporter enzyme and converted back to CO_2 through a carbonic anhydrase. This mechanism is called a C3 mechanism due to the inorganic CO_2 being fixed first into a 3-carbon molecule called 3-phosphoglycerate.[23]

The other mechanism converts a 3-carbon molecule such as PEP or pyruvate into a 4-carbon molecule such as oxaloacetate which is transported across the membrane.[24] There, it is decarboxylated to produce CO_2 and the

remaining 3-carbon molecule is transported back across the membrane completing the cycle.

There have been ongoing discussions among the academia relating to the mechanism that Pt uses. There are two primary schools of thought- namely that Pt uses C4 mechanism and the other that it uses a C3 mechanism. There have been strong arguments made on either side. Putative enzymes which support C3 mechanism were found through enzyme annotation[25] which were also strengthened by analysis which showed increased HCO_3 uptake efficiency with decreased atmospheric CO_2 levels.[26]

On the other hand, stable ^{13}C isotope labeling showed that the C4 metabolites which are aspartate, malate and succinate were enriched rapidly whereas the C3 metabolites were enriched at a much slower pace indicating a C4 mechanism operating. Also, the isoforms of pyruvate carboxylase enzyme were found to be stimulated by light which is consistent with what is expected of a C4 mechanism.[27,14]

With conflicting evidence on both sides, there is a need to study the organism in depth to identify the underlying mechanism that Pt undergoes. Prior research in our lab have performed rapid instationary studies using $NaH^{13}CO_3$ to identify enrichment of various metabolites indicative of the C4 or C3 pathways. But

there has not been a flux analysis fit using the data to verify these results which is the motivation for this minor study.

4.2 Methods

4.2.1 Instationary Flux Analysis

To perform MFA, we use the program NMR2Flux+. To enable instationary metabolic flux analysis, the software had to be altered to accommodate solving the differential equations illustrated prior. To iteratively solve the differential equations, we adapted the 4th order explicit Runge-Kutta method to solve the entire set of equations at each time step.[28] The equations are as follows:

$$f(x,t) = diag(M)^{-1}.(\tilde{\tilde{A}}(\overline{v}).{}^{r}\overline{x}(t) + \tilde{\tilde{b}}(\overline{v}, {}^{r-1}\overline{x}(t), {}^{r-2}\overline{x}(t),...)) \tag{4.4}$$

$$k_1 = h * f(x_n, t_n) \tag{4.5}$$

$$k_2 = h * f(x_n + \frac{k_1}{2}, t_n + \frac{h}{2}) \tag{4.6}$$

$$k_3 = h * f(x_n + \frac{k_2}{2}, t_n + \frac{h}{2}) \tag{4.7}$$

$$k_4 = h * f(x_n + k_3, t_n + h) \tag{4.8}$$

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{4.9}$$

$$t_{n+1} = t_n + h \tag{4.10}$$

Here, n ranges from 0 to the maximum time until which we wish to evaluate the MIDs for ’h’ is the time-step of every iteration. A more sensitive experiment would require a smaller time-step, but it simultaneously increases the number of

iterations, thus raising the computational time k_1, k_2, k_3, k_4 are parameters called Runge-Kutta increments which are evaluated at every time-step. As the evaluation at each time-step(t_n) is done from cumomer (r) ranging from $t_n = 1$ to the maximum order determined by the network, at each order increase the matrix $\tilde{\tilde{b}}$ is a constant which is evaluated at the end of the previous iteration of cumomer order. So, the evaluation at each time-step is done incrementally for each order of cumomers. The computational time that this method takes increases by an order of O($\frac{t_{end}}{h}$) where t_{end} is the time until which the instationary method runs and h is the time step.

Due to the nature of the equations, all the cumomers when operated in instationary mode are exponential functions of the time period t with a growth rate of $diag\left(M\right)^{-1} A$. The enrichment of ^{13}C for all intermediate metabolites rises until the growth decays to 0 when the system reaches a steady state value.

4.2.2 Evaluation of MIDs

For all calculations of the instationary method, we run the system in forward simulations and attempt to manually match the experimental data to the simulated set of MIDs

4.2.3 Network Model

For creating the network model, we manually created a set of metabolites that characterize the Pt network. As the instationary method, takes a much longer time to compute, we created a network with as few metabolites to fully identify the processes.

4.2.4 Experimental MID Data

The experimental MID data required to fit the experimental data was obtained from a prior study performed by our lab by Andrew Quinn and Dr. Ganesh Sriram. The instationary data was used previously as evidence to support C4 mechanism occurring in Pt cells. But due to limitations in the program then, was not used to estimate the network fluxes.

The experiment was performed by providing $H^{13}CO_3$ as a pulse label to the four Pt culture replicates. At time t=0, $H^{13}CO_3$ was provided in batch to the replicates. Then the cell was allowed to metabolize. At each time point of 3 hours, 24 hours, 48 hours and 72 hours, one of the replicates was derivatized and analyzed using the GC/MS to obtain the amino acid MIDs. Then, the data was computationally corrected for natural labeling of carbon.

4.3 Results and Discussion

We created a Pt network consisting of the Calvin cycle and a section of the reversible TCA cycle to understand the network. $H^{13}CO_3$ was fed as the only

source of carbon to the network whereas Alanine, Malate, Glutamine and Aspartate. Further, only the latter half of the TCA cycle and the Calvin cycle metabolites are assumed to be reversible. (see Fig 4.1)

The experiment was performed as a pulse experiment where the labeled bicarbonate was fed in batch to the cell. As there is no method currently available to emulate pulse-labeling experiments in-silico, we initially assumed the cell to grown on labeled bicarbonate until the 3-hour point and later to grow on unlabeled bicarbonate until steady state at 72 hours.

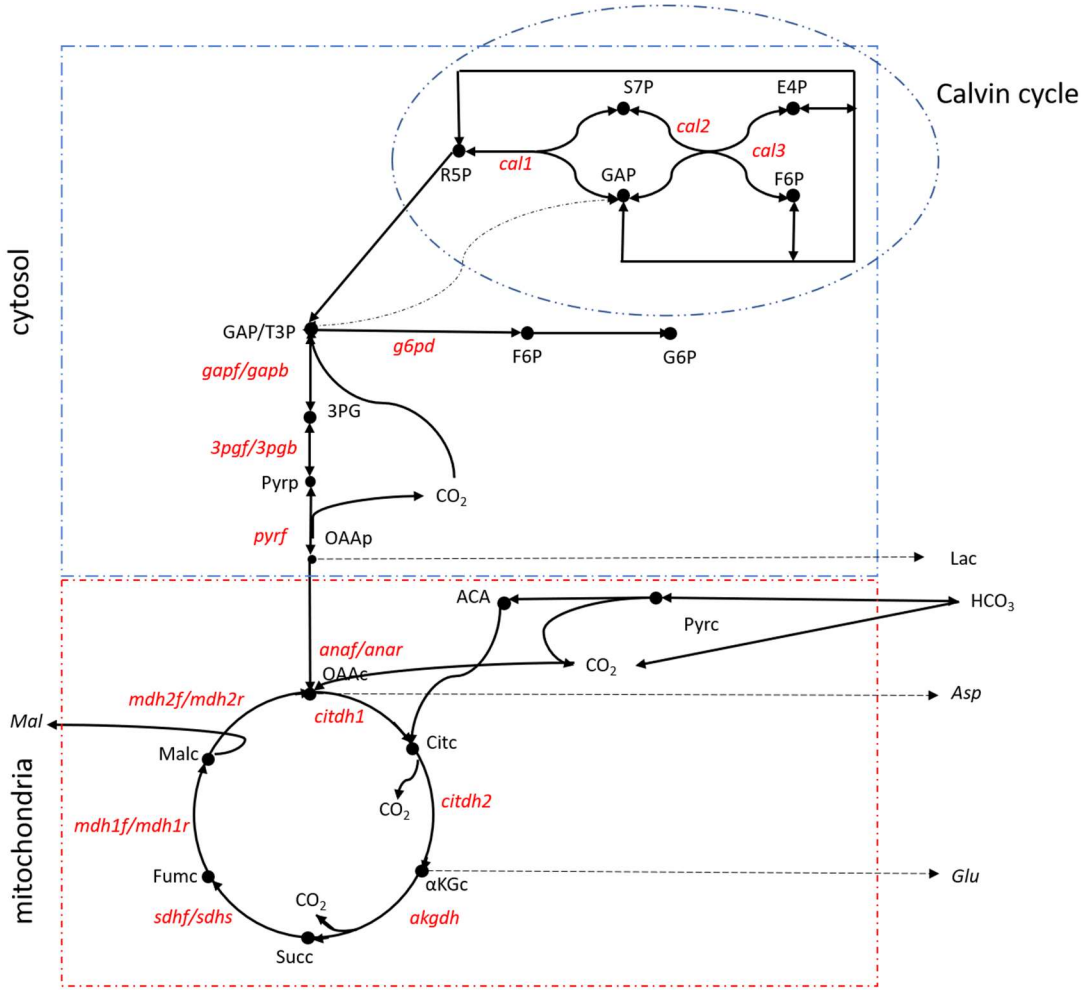

Fig. 4.1: Network model of Phaeodactylum *tricornutum.* The fluxes are as shown in red, and the metabolites are written in black. The dashed lines indicate the fluxes which are experimentally measured.

Initially, we modeled the data and checked the limits of each flux to obtain the solution space that we could look for the fluxes. The free fluxes for the network are the transport reaction of oxaloacetate from the mitochondria to the cytosol (oaat), the conversion of oxaloacetate to citrate (citdh1) and the reaction of 3-phosphoglycerate entering the Calvin cycle(cal0). We found that for the modeled network, it was only possible for oxaloacetate to transport 30% of the total

bicarbonate entering the system onto the cytosol. Any higher would lead to the flux values being infeasible for the network. Therefore, we need to assume a set of fluxes which are feasible. Accordingly, the free flux values were assumed to be oaat = 0.3, citdh1= 0.25 and cal0 = 0.25. In order to fit the experimental data, we would need to assume a set of metabolite pool sizes as well which hugely influences the MID variation with time. By trial and error for every metabolite measurement identified, we varied the metabolite pool sizes of the precursors to match the measurements as close as possible. This calibration gives us a set of relative pool size measurements which can be assumed as constant due to the pseudo steady approximation we made for analyzing in instationary mode.

Once we got the set of metabolite pool sizes which closest resemble the experimental MID data, we used the calibrated network to simulate the MIDs with unlabeled HCO_3 set as the feed from time t = 3 hours until 72 hours. For this second part of the simulations, the labeling at 3 hours was taken as the initial enrichment. Though, there weren't as many time steps of simulations done, it gives us a very good idea of the path that the ^{13}C takes during its metabolism. Fig. 4.2 illustrates the simulated enrichment of the measured metabolites for the pulse labeling experiment.

From the simulated enrichment data, we can conclude that while malate, alanine and glutamine weren't sufficiently labeled at the 3-hour mark, aspartate

demonstrates rapid enrichment which then stabilized at nearly the same enrichment until the eventual steady state at 72 hours.

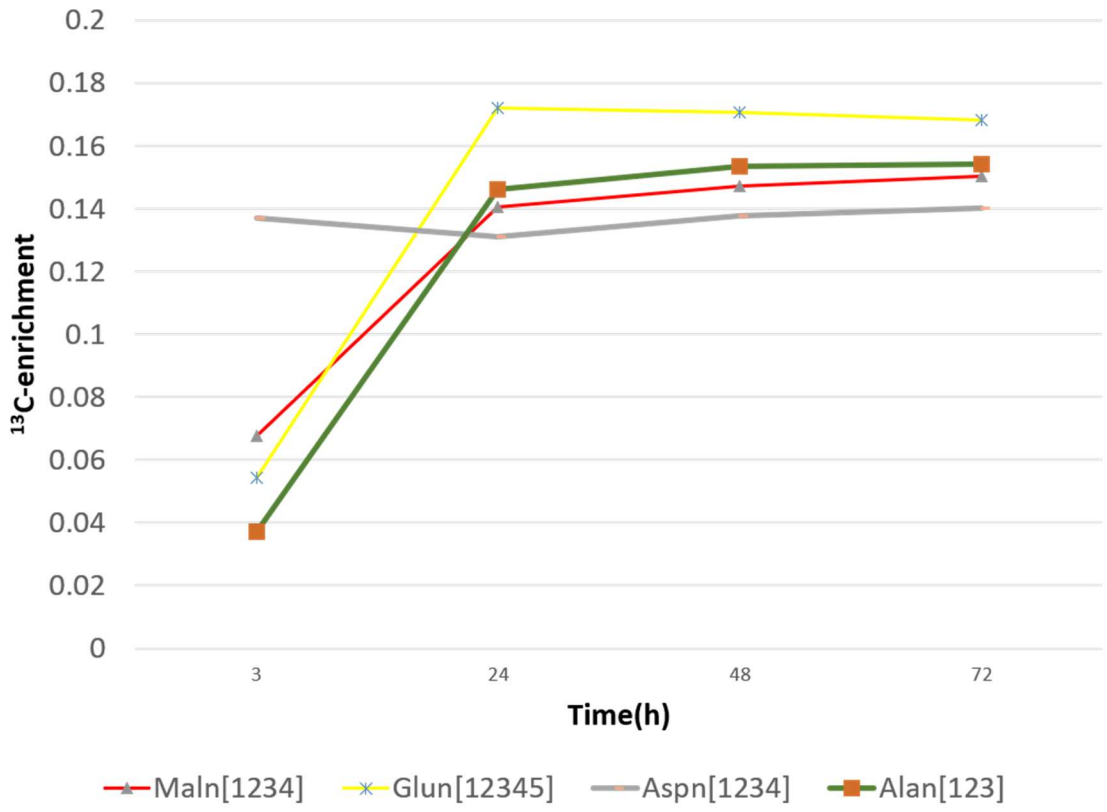

Fig. 4.2: Plot of ^{13}C enrichment as a function of time. There are four primary metabolites which fit well with the model. Malate, Glutamine, Aspartate and Alanine.

This indicates that the bicarbonate labels the oxaloacetate first which gets transported across the membrane. This is indicative of a C4 mechanism happening within the organism. If the organism were to have operated with a C3 mechanism, Alanine would have been initially enriched followed by the other metabolites due to the formation of 3PG and pyruvate through the biophysical method. In fact, as

observed, alanine is the last to get enriched during the pulse experiment providing additional evidence supporting the same hypothesis.

4.4 Conclusions

During this study, we adapt and implement the instationary metabolic flux analysis method. Previous experimental data from the lab, despite being performed under instationary conditions were not able to fit on an instationary network model due to lack of a computational framework. This thesis study thus created the necessary computational framework from existing theoretical methods, thus enabling any future experiments to be performed more conveniently.

We also took a simple network and utilized existing data to fit a model network. The enrichment data simulated for a wild-type Pt cell line also indicates an underlying C4 mechanism at play. This supports prior studies performed in the lab. Though, this work and evidence may not be sufficient to conclusively prove the working of the cell. However, this lays the groundwork for further experiments involving genetic variants of the Pt cell line especially targeting the silencing of vital enzymes like phosphoenolpyruvate carboxylase and pyruvate carboxylase would demonstrate more conclusive proof regarding the mechanism that Pt employs.

Chapter 5: Conclusions and Future Work

During this thesis, we introduced a method – multi-element flux analysis – which is an enhanced tracking mechanism while performing isotope-assisted metabolic flux analysis. In this method, we proposed the idea of tracking elements apart from carbon to improve the visibility of the network during stable isotope techniques. This allows for higher resolution of fluxes than was possible from the ^{13}C flux analysis techniques normally used.

To mitigate the high costs of custom-made multi-element tracers, we also observed the differences between a parallel labeling experiment and a combined labeling experiment for this scenario. The results showed that a parallel labeling experiment with more than one element being tracked improves the flux

estimation significantly. This result gives us greater economic freedom while designing our ideal multi-element labeling experiment.

In the latter part of this thesis, we looked at the instationary metabolic flux analysis method. This has been a rapidly growing field of research in isotope assisted flux analysis techniques. But this thesis focuses on our implementation of the method on our native flux analysis software NMR2Flux+. Also, using this technique we analyzed a diatom network model which has long been a point of controversy with regards to its carbon concentrating mechanism. Our results though brief seem to support the claim that the organism undergoes a C4 mechanism to improve its efficiency in producing energy from CO_2 through photosynthesis.

Both these methods are unique in their own way. But they have a common goal of improving flux resolution under circumstances which do not allow steady state ^{13}C MFA. The multi-element flux analysis method is still a nascent field and hasn't been explored much. We expect future work to utilize this method on a realistic method using experimental data for non – conventional networks. There is a special use case in application to the urea cycle where a nitrogen label would provide much deeper insight than has been possible using ^{13}C-MFA. More diverse applications for this method have been discussed in Chapter 3.

Our instationary method is more of a personal goal. There has been instationary methods implemented by other groups previously. But the platform we created helps us to implement the method in networks of interest to us. Especially in co-culture models of Pt and a Bacillus *thuringiensis* network is of interest to analyze the symbiosis between both while grown together, we hope that the instationary method helps us in identifying the inter-cellular metabolites aiding us to understand microbial consortiums in greater detail.

References

1. Walker T. E.; Han C. H.; Kollman V. H.; London R. E.; Matwiyoff N. A. 13C Nuclear Magnetic Resonance Studies of the Biosynthesis by Microbacterium *ammoniaphilum* of L-Glutamate Selectively Enriched with Carbon-13. *The Journal of Biological Chemistry* **1982**, 257 (3), 1189–1195.

2. Nargund, S.; Sriram, G. Designer Labels for Plant Metabolism: Statistical Design of Isotope Labeling Experiments for Improved Quantification of Flux in Complex Plant Metabolic Networks. *Molecular Biosystems* **2013**, 9 (1), 99–112.

3. Antoniewicz M. R. 13C Metabolic Flux Analysis: Optimal Design of Isotopic Labeling Experiments. *Current Opinion in Biotechnology* **2013**, 24 (6), 1116–1121

4. Rantanen, A. Algorithms for 13c Metabolic Flux Analysis. dissertation, University of Helsinki, **2006**, 106.

5. Dieuaide-Noubhani M; Raffard G; Canioni P; Pradet A; Raymond P. Quantification of Compartmented Metabolic Fluxes in Maize Root Tips Using Isotope Distribution from 13C- or 14C-Labeled Glucose. *The Journal of Biological Chemistry* **1995**, 270 (22), 13147–13159.

6. Wiechert, W. 13c Metabolic Flux Analysis. *Metabolic Engineering* **2001**, 3 (3), 195–206.

7. Wiechert, W.; Siefke, C.; de Graaf, A. A.; Marx, A. Bidirectional Reaction Steps in Metabolic Networks: II. Flux Estimation and Statistical Analysis. *Biotechnology and Bioengineering* **1997**, 55 (1), 118–135.

8. Antoniewicz, M. R.; Kelleher, J. K.; Stephanopoulos, G. Elementary Metabolite Units (EMU): A Novel Framework for Modeling Isotopic Distributions. *Metabolic Engineering* **2007**, 9 (1), 68–86.

9. van Winden W. A.; Heijnen J. J.; Verheijen P. J. Cumulative Bondomers: A New Concept in Flux Analysis from 2D [13C,1H] Cosy NMR Data. *Biotechnology and Bioengineering* **2002**, *80* (7), 731–745.

10. Press, W. H. Numerical Recipes in C : The Art of Scientific Computing, 2nd ed.; Cambridge University Press: Cambridge, England, **1994**.

11. Nöh Katharina; Wiechert, W. Experimental Design Principles for Isotopically Instationary 13C Labeling Experiments. *Biotechnology and Bioengineering* **2006**, *94* (2), 234–251.

12. Nöh Katharina; Wahl, A.; Wiechert, W. Computational Tools for Isotopically Instationary 13C Labeling Experiments Under Metabolic Steady State Conditions. *Metabolic Engineering* **2006**, 8 (6), 554–577.

13. Millard, P.; Cahoreau, E.; Heuillet, M.; Portais, J.-C.; Lippens, G. ^{15}N-NMR-Based Approach for Amino Acids-Based 13C-Metabolic Flux Analysis of Metabolism. *Analytical Chemistry* **2017**, *89* (3), 2101–2106.

14. Zheng Y; Quinn AH; Sriram G. Experimental Evidence and Isotopomer Analysis of Mixotrophic Glucose Metabolism in the Marine Diatom Phaeodactylum *tricornutum*. *Microbial Cell Factories* **2013**, 12, 109–109.

15. Nilsson, R.; Jain, M. Simultaneous tracing of carbon and nitrogen isotopes in human cells. *Molecular Biosystems* **2016**, 12, 1929–1937.

16. Lewis C. A.; Parker S. J.; Fiske B. P.; McCloskey D.; Gui D. Y.; Green C. R.; Vokes N. I.; Feist A. M.; Vander Heiden M. G.; Metallo C. M. Tracing Compartmentalized Nadph Metabolism in the Cytosol and Mitochondria of Mammalian Cells. *Molecular Cell* **2014**, *55* (2), 253–263.

17. de Graaf, A. A.; Eggeling, L.; & Sahm, H. Metabolic engineering for L-lysine production by Corynebacterium *glutamicum*. *Advanced Biochemical Engineering Biotechnology* **2001,** 73, 9–29.

18. Crown S.B; Antoniewicz M.R. Parallel Labeling Experiments and Metabolic Flux Analysis: Past, Present and Future Methodologies. *Metabolic Engineering* **2013**, 16 (1), 21–32.

19. Jazmin, L. J. & Young, J. D. Isotopically nonstationary 13C metabolic flux analysis methods. *Molecular Biology* **2013***,* 985, 367–390.

20. Antoniewicz, M. R.; Kraynie, D. F.; Laffend, L. A.; González-Lergier Joanna; Kelleher, J. K.; Stephanopoulos, G. Metabolic Flux Analysis in a Nonstationary System: Fed-Batch Fermentation of a High Yielding Strain of Escherichia *coli* Producing 1,3-Propanediol. *Metabolic Engineering* **2007**, *9* (3), 277–292.

21. Zheng Y; Quinn A. H; Sriram G. Experimental Evidence and Isotopomer Analysis of Mixotrophic Glucose Metabolism in the Marine Diatom Phaeodactylum *tricornutum*. *Microbial Cell Factories* **2013**, *12*, 109–109.

22. Mahadevan R; Edwards J. S; Doyle F. J. 3rd. Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia *coli*. *Biophysical Journal* **2002**, *83* (3), 1331–1340.

23. Cerón-García M. C; Fernández-Sevilla J. M; Sánchez-Mirón A; García-Camacho F.; Contreras-Gómez A.; Molina-Grima, E. Mixotrophic Growth of Phaeodactylum *tricornutum* on Fructose and Glycerol in Fed-Batch and Semi-Continuous Modes. *Bioresource Technology* **2013**, *147*, 569–576.

24. Fabris, M.; Matthijs, M.; Rombauts, S.; Vyverman, W.; Goossens, A.; Baart, G. J. E. The Metabolic Blueprint of Phaeodactylum *tricornutum* Reveals a Eukaryotic Entner-Doudoroff Glycolytic Pathway. *The Plant Journal* **2012**, *70*(6), 1004–1014.

25. Tachibana M; Kikutani S; Endo Y; Matsuda Y; Allen A.E; Bowler C. Localization of Putative Carbonic Anhydrases in Two Marine Diatoms, Phaeodactylum *tricornutum* and Thalassiosira *pseudonana*. *Photosynthesis Research* **2011**, *109* (1-3), 205–221.

26. Hopkinson B.M. A Chloroplast Pump Model for the CO_2 Concentrating Mechanism in the Diatom Phaeodactylum *tricornutum. Photosynthesis Research* **2014**, *121* (2-3), 223–233.

27. Haimovich-Dayan M.; Garfinkel N.; Ewe D.; Marcus Y.; Gruber A.; Wagner H.; Kroth P. G.; Kaplan A. The role of C_4 metabolism in the marine diatom Phaeodactylum *tricornutum*. *New Phytologist* **2013,** 197, 177–185.

28. Butcher, J. C. Numerical Methods for Ordinary Differential Equations, 3rd ed.; Wiley: Chichester, UK, **2016**.