Ensemble Modeling of Cancer Metabolism

by

Tahmineh Khazaei

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science

Chemical Engineering and Applied Chemistry
University of Toronto

© Copyright by Tahmineh Khazaei 2011
Ensemble Modeling of Cancer Metabolism

Tahmineh Khazaei

Master of Applied Science

Chemical Engineering
University of Toronto

2011

Abstract

Metabolism in cancer cells is adapted to meet the proliferative needs of these cells, with notable changes such as enhanced lactate secretion and glucose uptake rates. In this work, we use the Ensemble Modeling (EM) framework to gain insight and predict potential drug targets for tumor cells. A metabolic network consisting of 58 reactions is considered which accounts for glycolysis, the pentose phosphate pathway, lipid metabolism, amino acid metabolism, and includes allosteric regulation. Experimentally measured metabolite concentrations are used for developing the ensemble of models along with information on established drug targets. The resulting models predicted transaldolase (TALA) and succinate-CoA ligase (SUCOAS1m) to display a significant reduction in growth rate when repressed relative to currently known drug targets. Furthermore, the synergetic repression of transaldolase and glycine hydroxymethyltransferase (GHMT2r) showed a three fold decrease in growth rate compared to the repression of single enzyme targets.
Acknowledgments

I wish to express my sincere gratitude to my supervisor, Professor Radhakrishnan Mahadevan, for his excellent guidance and advice throughout the course of this work. His supervision, motivation, enthusiasm and encouragement have been a major contributing factor in the successful completion of this project. Furthermore, I am greatly indebted to him for his valuable advice for my professional career and his tremendous support in reaching my future career goals.

Many thanks are due to my Defence Committee members, Professor Alison McGuigan and Professor William Cluett, for their constructive feedback and comments.

I would like to thank Biozone for providing the facilities to carry out this work. I am thankful to all my colleagues in the Laboratory of Metabolic Systems Engineering for their friendship and support over the course of this project. Special thanks are due to my colleagues, Nicholas Bourdakos, and Nikolaos Anesiadis for their guidance and assistance with the metabolomics experiments. Many thanks are due to Ana Paz and Sahar Javaherian, for their guidance and support with the cell culturing experiments. Finally, special thanks are due to my colleague, Sarat Chandra, for his support, encouragement, and collaboration in the computational aspects of the project.

This project was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant. I would also like to acknowledge Ontario Graduate Scholarship (OGS) for the financial support provided for this project.

Last but not least, I wish to pay very special thanks to my parents who are always there for me with their unconditional love and support. This thesis is dedicated to them.
Table of Contents

List of Tables .......................................................................................................................... vii
List of Figures .......................................................................................................................... viii

Chapter 1 ................................................................................................................................. 1
Introduction ............................................................................................................................... 1

Chapter 2 ................................................................................................................................. 4
Literature Review ...................................................................................................................... 4

2.1 Cancer Metabolism ............................................................................................................ 4
   2.1.1 Cancer Cells Display Aerobic Glycolysis ................................................................. 5
   2.1.2 TCA Cycle and Cell Proliferation ............................................................................ 8
   2.1.3 Metabolic Regulation of Cellular Proliferation ....................................................... 11

2.2 Flux Balance Analysis ...................................................................................................... 13

2.3 Ensemble Modeling ......................................................................................................... 14

Chapter 3 ................................................................................................................................. 16
Knowledge Gaps and Statement of Objectives ....................................................................... 16

Chapter 4 ................................................................................................................................ 18
Methodology ............................................................................................................................ 18

4.1 Experimental Methods ..................................................................................................... 18
   4.1.1 Cell Line ................................................................................................................... 18
   4.1.2 Subculturing ............................................................................................................. 18
   4.1.3 Experimental Protocol ............................................................................................ 19

4.2 Computational Methods .................................................................................................. 21
   4.2.1 Overview of Flux Balance Analysis ....................................................................... 22
Chapter 6 .................................................................................................................. 54

Conclusion and Recommendations........................................................................ 54

References .................................................................................................................. 56

Appendix ..................................................................................................................... 61
List of Tables

Table 1. Decomposition of Enzymatic Reactions into Elementary Reaction Form .......... 28
Table 2. Uptake and Secretion Fluxes ........................................................................... 40
Table 3. Allosteric Regulation Considered in the Model .............................................. 45
Table 4. Perturbations Used for Screening the Ensemble .............................................. 46
# List of Figures

Figure 1. The Difference Metabolic States of Unicellular vs. Multicellular Organisms .............................................................. 5

Figure 2. A Schematic Representation of Oxidative Phosphorylation, Anaerobic Glycolysis, and Aerobic Glycolysis ............................................................. 6

Figure 3. A Schematic of Carbon Flux in Proliferating Cells .................................................. 9

Figure 4. Glutamine Dependent Anaplerosis Replenishes TCA Cycle Intermediates ................................................................. 11

Figure 5. Regulatory Signaling Network in Proliferating Cells .............................................. 13

Figure 6. Experimental Setup .................................................................................................................. 20

Figure 7. The Conceptual Basis for Flux Balance Analysis .................................................. 22

Figure 8. Visualization of the Ensemble Modeling Methodology ........................................... 24

Figure 9. Summary of the Main Steps Involved in EM .................................................. 27

Figure 10. Growth Curve for the colo205 Cell Line ............................................................. 34

Figure 11. Exponential Fit to Growth Phase .................................................................................. 35

Figure 12. Extracellular Amino Acid Concentration Profiles .................................................. 36

Figure 13. Extracellular Carbohydrate Concentration Profiles Measured by HPLC ......... 37

Figure 14a. Experimentally Measured Intracellular Metabolite Profiles ................................ 39

Figure 14b. Experimentally Measure Intracellular Metabolite Profiles ................................ 40

Figure 15. Parameter Optimization for Lactate Secretion Using SensSB Toolbox ........... 41

Figure 16. Reconstructed Model Network .................................................................................. 42

Figure 17. Flux Map at Reference Steady State For colo205 Cells .......................................... 44

Figure 18. Screening of the Ensemble of Models ........................................................................ 47

Figure 19. Perturbation Targets for Reduction of Growth Rate ........................................... 49

Figure 20. Effect of Synergetic Perturbations on Growth Rate ............................................. 49
List of Appendices

Appendix A: List of Enzymes and Abbreviations .......................................................... 61
Appendix B. Extracellular Metabolite Concentrations ................................................. 62
Appendix C. Model Parameters ..................................................................................... 63
Chapter 1
Introduction

Metabolic regulation plays a key role in fulfilling the metabolic demands of proliferative vs. differentiated tissue. While differentiated tissue catabolize nutrients to meet bioenergetic needs (i.e. ATP synthesis), proliferating cells channel much of the glucose derived carbon to anabolic pathways to meet biosynthetic needs (i.e. lipid, protein, and nucleic acids synthesis) (Christofk et al., 2008). The distinguished metabolism of cancer cells compared to differentiated tissue was first observed by Otto Warburg in the 1920s where he noted high rates of glucose consumption and lactate secretion, regardless of oxygen availability. This metabolic behavior in cancer cells is termed aerobic glycolysis or the “Warburg effect.” (Vander Heiden et al., 2009). The importance of aerobic glycolysis for cancer cells has been proven experimentally (Schulz et al., 2006) and this metabolic adaptation has been thought to facilitate the incorporation of nutrients into biomass necessary to produce new cells (Vander Heiden et al., 2009). Indeed, glycolytic intermediates are precursors for biomass production. However, the biosynthetic benefits obtained through high glycolytic fluxes do not explain why such high lactate production rates are observed when more pyruvate could be utilized for ATP production through oxidative phosphorylation (DeBerardinis et al., 2008). Otto Warburg proposed that damage in oxidative metabolism caused the high rates of glycolysis, however later studies have proven otherwise and revealed the mitochondria to be functional (Moreno-Sanchez et al., 2007). It has been suggested that there is a limitation in the maximal velocity of pyruvate oxidation and hence cells must eliminate pyruvate by conversion to lactate (Curi et al. 1988). In terms of control, it has been suggested that the high glycolytic flux allows the biosynthetic pathways, which stem from glycolytic intermediates, to be fine tuned and thereby, high lactate production compensates for the maintenance of biosynthetic fluxes during proliferation (Newsholme et al., 1985).

---

1 Introduction adapted from paper “Ensemble Modeling of Cancer Metabolism” being submitted to BMC Systems Biology
The metabolic switch that occurs in cancer cells suggests that computational modeling approaches could provide insight and further understanding of the complex metabolic interactions. Previous computational work on cancer metabolism has demonstrated that aerobic glycolysis enables the maximal production of the biomass precursor palmitate considering the stoichiometry of only a few central metabolic pathways (Vander Heiden et al., 2009). Another study considering two lumped reactions representing aerobic glycolysis and oxidative phosphorylation constrained by the cell’s glucose uptake capacity and solvent capacity showed that at high glucose uptake rates, aerobic glycolysis provides the cell with the highest rate of ATP production (Vazquez et al., 2010). Recently, Folger et al. developed the first genome scale model of cancer metabolism based on the genome scale human metabolic network (Duarte et al., 2007). This model was used to predict cytostatic drug targets, of which 40% were known targets and 60% new targets. In addition, combinations of synthetic legal drug targets were identified (Folger et al., 2011).

In this study we use an ensemble modeling approach to investigate the metabolism of cancer cells in order to gain further insight into their metabolism and determine potential enzymatic drug targets. Ensemble modeling has been successfully used in the past to improve L-lysine production in E. coli (Contador et al., 2009) and identify regulatory mechanisms in hepatic cells (Dean et al., 2010). Ensemble modeling generates kinetic models while bypassing the need for detailed kinetic parameters (Yan et al., 2011) which are often unknown or difficult to determine experimentally (Lee et al., 2006). EM generates an ensemble of models by sampling for kinetic parameters under thermodynamic and steady state constraints. Each model that is generated has a unique set of kinetic parameters and displays unique dynamic behavior; however, all models are anchored to the same steady state. The models in the ensemble are computationally perturbed and the model-predicted steady state fluxes are compared to experimental perturbation results. Models that capture the experimental results are retained. Continual screening of the models as experimental data becomes available allows for the
convergence to an increasingly realistic and predictive subset of models (Contador et al., 2009). In ensemble modeling, the reactions in the network are formulated in the elementary reaction form, which allows the true mechanism of the enzymatic reactions to be captured. Furthermore, because of the elementary reaction formulation, regulatory information can also be incorporated in the models (Dean et al., 2010).
Chapter 2
Literature Review

2.1 Cancer Metabolism

Unicellular organisms have evolutionary had the need to proliferate and reproduce as quickly as possible in the presence of an adequate nutrient supply. The regulatory control systems present in these organisms have evolved to sense nutrients and channel the supply of carbon, nitrogen, and free energy into pathways that result in the production of biomass. When low nutrient supply is sensed, the metabolism in these unicellular organisms switches to utilize the free energy available for the purposes of surviving the starvation period and therefore cease biomass production. These varying metabolic needs of the cells have evolved the metabolic control systems of these organisms to handle proliferating vs. non-proliferating conditions correspondingly (Vander Heiden et al., 2009).

In contrast to unicellular organisms, multi-cellular organisms are usually exposed to a constant supply of nutrients and very scarcely do they experience nutrient starvation. Their control systems have evolved to prevent aberrant nutrient uptake when nutrient supplies are greater than bioenergetic and biosynthetic needs. In these cells, nutrient uptake is controlled through growth factors, which are proteins that act as stimulants. It is only in the presence of these growth factors that nutrient uptake can take place. Figure 1 depicts the different metabolic states of unicellular and mutlicellular organisms (Vander Heiden et al., 2009).

Cancer cells have shown to overcome the growth factor dependence that controls cellular proliferation and therefore display high rates of cellular proliferation irrespective of the physiological conditions. The mechanism that integrates signal transduction and cell metabolism is largely conserved between cancer cells and normal cells. What is different between these two cells is that in normal cells signaling is initiated by extracellular stimuli, whereas in cancer cells mutations have chronically enhanced these pathways such that they continuously remain in a proliferative state. For example, oncogenic mutations
have been observed to increase the uptake rate of glucose, resulting in nutrient uptake in excess of the bioenergetic demands for cell growth and proliferation (Hsu et al., 2008 and Deberardinis et al., 2008).

![Figure 1](image_url)

**Figure 1.** The different metabolic states of unicellular vs. multicellular organisms. Unicellular organisms adapt their metabolism based on the availability of nutrients in the environment and are stimulated to grow when nutrient abundance is sensed. Multicellular organisms are usually exposed to an abundant supply of nutrients and are simulated to grow only in the presence of growth signals (Vander Heiden et al., 2009).

### 2.1.1 Cancer Cells Display Aerobic Glycolysis

In the 1920s Otto Warburg made the observation that tumor cells uptake glucose at a much higher rate than normal cells. In addition, he observed that most of the glucose derived carbon is used to produce lactate and thus tumor cells exhibit high rates of lactate secretion. This metabolic behavior of cancer cells is termed aerobic glycolysis or the “Warburg effect” (Warburg, 1956).
In the presence of oxygen, differentiated cells (non-proliferating cells) metabolize glucose through glycolysis to pyruvate. Pyruvate subsequently enters the TCA cycle in the mitochondria and is oxidized to produce carbon dioxide. This oxidation in the TCA cycle generates NADH (nicotinamide adenine dinucleotide), which fuels oxidative phosphorylation for the maximal production of ATP with minimal lactate production. Oxygen is essential for the complete oxidization of glucose as it is the final electron acceptor during oxidative phosphorylation. In the absence of oxygen, however, the differentiated cells produce minimal ATP through the process of anaerobic glycolysis. In anaerobic glycolysis pyruvate is channeled away from the TCA cycle and is used for lactate production. Generation of lactate allows for glycolysis to continue (by cycling NADH back to NAD+). However, in cancer cells, most of the glucose is converted to lactate regardless of the presence of oxygen. This property of cancer cells is shared by normal proliferative tissue as well (Vander Heiden et al., 2009). A schematic of the differences between oxidative phosphorylation, anaerobic glycolysis and aerobic glycolysis is presented in Figure 2.

![Figure 2](image_url)  
**Figure 2.** A schematic representation of oxidative phosphorylation, anaerobic glycolysis, and aerobic glycolysis. In the presence of oxygen differentiated tissue undergo oxidative phosphorylation to produce the maximum amount of ATP. When oxygen is limiting, cells redirect the pyruvate generated through glycolysis away from mitochondrial oxidative phosphorylation by generating lactate. Proliferative tissue or tumors have been observed to convert most glucose to lactate regardless of the availability of oxygen (Vander Heiden et al., 2009).
The Warburg effect presents a paradox that has not been completely resolved to this date: why cancer cells use such wasteful form of metabolism while requiring a greater need for ATP. Otto Warburg had proposed that damage in oxidative metabolism results in a high rate of glycolysis and accounts for the high concentrations of lactate produced (Warburg, 1956). However, later studies ruled out this possibility by showing that primary lymphocytes demonstrate the same metabolic behavior, whereby more than 90% of the glucose carbon is converted to lactate, suggesting that aerobic glycolysis is not unique to cancer cells (Brand, 1985; Hedekov, 1968; Roos and Loos, 1973; Wang et al., 1976). In addition, recent studies on cancer cells have proven oxidative metabolism to be functional and no defects have been observed (Moreno-Sanchez et al., 2007).

The high rate of glycolysis observed in cancer cells does provide some advantages for the proliferating cells. Glycolytic intermediates are precursors for biosynthetic pathways. These metabolites can enter the pentose phosphate pathway to produce NADPH and ribose sugars for the generation of nucleotides. Glycerol is generated from glycolytic intermediates for the production of lipids, and nonessential amino acid production pathways also stem from glycolytic intermediates (DeBerardinis et al., 2008). Furthermore, if high enough rates of glycolysis are reached the percentage of cellular ATP produced from glycolysis can exceed that from oxidative phosphorylation (Guppy et al., 1993).

The biosynthetic benefits obtained through high glytolytic fluxes however do not explain why such high rates of lactate production are observed when more of the pyruvate could be utilized for ATP production through oxidative phosphorylation. Curi et al. suggested that there is a limitation in the maximal velocity of pyruvate oxidation, and hence cells must eliminate pyruvate by conversion to lactate to compensate for the high rates of glycolysis. The oxidation of pyruvate involves transport from the cytoplasm to the mitochondria followed by the initial reaction of pyruvate to acetate which is catalyzed by the highly regulated enzyme, pyruvate dehydrogenase (PDH). The activity of PDH is influenced by phosphorylation, the levels of free CoA present, and the NAD+/NADH ratios which could all be limiting relative to the glycolytic flux. The glycolytic flux is
estimated to be greater than an order of magnitude compared to the maximum flux of PDH (Curi et al., 1988). Proliferating cells express the enzyme LDH-A (lactate dehydrogenase A) which rapidly converts pyruvate to lactate. This enzyme is induced by the oncogenes c-Myc and Her2 amongst others (Fantin et al., 2006; Marjanovic et al., 1990; Shim et al., 1997).

A further explanation for the high rates of lactate production was given by Newsholme et al. in terms of metabolic control. In metabolic control, whenever a high-flux metabolic pathway branches into a low flux pathway, the ability to control the low flux pathway is maximized when the flux through the high flux pathway is highest. The high glycolytic flux allows the biosynthetic pathways which stem from glycolytic intermediates to be fine tuned. The high rate of lactate production thus occurs to compensate for the maintenance of biosynthetic fluxes during proliferation (Newsholme et al., 1985).

2.1.2 TCA Cycle and Cell Proliferation

In non-proliferating cells, the TCA cycle mainly serves to derive maximal ATP production, while generating 2 carbon dioxide molecules per each cycle. In non-proliferating cells, however, the carbon that enters the TCA cycle is used for biosynthetic purposes, whereby TCA cycle intermediates enter biosynthetic pathways that consume rather than produce ATP (DeBerardinis et al., 2008).

Reactions which utilize TCA cycle intermediates for biosynthetic production purposes are referred to as cataplerotic reactions. The TCA cycle intermediate citrate is a major cataplerotic substrate. Citrate is transferred out of the mitochondria to the cytoplasm where it is converted to oxaloacetate and the lipogenic precursor acetyl-CoA by the enzyme ATP citrate lyase (ACL) (DeBerardinis et al., 2008). The enzymes ACL and FAS (fatty acid synthase) are shown to be induced in cancer cells and proliferating hematopoietic cells (Buer et al., 2005; Hatzivassiliou et al., 2005; Kuhajda et al., 1994; Pizer et al., 1996). This de novo synthesis of fatty acids is required for creating the lipid membranes during cellular proliferation (Oookhtens et al., 1984). The efflux of citrate results in a “truncated” cycle whereby only a reduced fraction of citrate is oxidized in the
cycle. Studies on hepatoma cells have shown that during their proliferation, higher cholesterol synthesis was associated with a higher citrate efflux and inversely proportional to citrate stimulated oxidation (Parlo and Coleman, 1986). Other intermediates of the TCA cycle such as oxaloacetate and α-ketoglutarate participate in cataplerosis and are precursors for the synthesis of nonessential amino acids used for protein and nucleotide synthesis (DeBerardinis et al., 2008). The cataplerotic pathway of citrate utilization for macromolecular biosynthesis is shown in Figure 3.

![Figure 3](image-url)

**Figure 3.** A schematic of carbon flux in proliferating cells. During proliferation there is a large increase in glycolytic flux and most of the resulting pyruvate is converted to lactate by lactate dehydrogenase (LDH-A). Some of the pyruvate is converted to acetyl-CoA (Ac-CoA) by pyruvate dehydrogenase (PDH) and enters the TCA cycle where it is converted to intermediates such as citrate (cit), which can be used for biosynthesis. Citrate is required for synthesis of fatty acids used to make the lipid membranes (DeBerardinis et al., 2008).
There are notable exceptions to the finding that the TCA cycle is functional in cancer cells. Mutations in the TCA enzyme succinate dehydrogenase (SDH) have shown to cause pheochromocytoma (Astuti et al., 2001; Gimm et al., 2000) and mutations in the enzyme fumarate hydratase (FH) have shown to cause leiomyomata, papillary renal cell cancer, and a dominant syndrome of uterine fibroids (Tomlinson et al., 2002). In all these cases, with the presence of mutations, the cells were not only able to survive but would proliferate at a pathological rate (DeBerardinis et al., 2008).

It is necessary to replenish the supply of intermediates used for cataplerosis by a matching influx through anaplerosis reactions. High anaplerotic flux rates are more indicative of cellular proliferation than high glycolytic fluxes, as the latter can be achieved by hypoxia and other stresses independent of macromolecular synthesis. The major source for anaplerosis is the amino acid glutamine. Glutamine replenishes the TCA intermediate α-ketoglutarate through its conversion to glutamate in the mitochondria by the enzyme PDG (phosphate-dependant glutaminase). Moreover, in the cytosol glutamine donates nitrogen for the production of nucleotides during the formation of glutamate. Glutamate is subsequently metabolized to α-ketoglutarate which can enter the mitochondria (DeBerardinis et al., 2008). Figure 4 shows a schematic for this anaplerotic process. Experimental studies through $^{13}$C-labeled substrates have verified glutamine to be the major anaplerotic precursor in proliferating glioma cells in both rats (Portais et al., 1996) and humans (DeBerardinis et al., 2007). Another suggested anaplerotic mechanism is the reaction catalyzed by pyruvate carboxylase (PC), which directly generates oxaloacetate from pyruvate within the mitochondria. However this hypothesis has not been proven to be a universal component of anaplerosis. Mitogens have shown to increase PC activity in lymphocytes proving that PC does in fact play a role in proliferative metabolism (Curi et al., 1988). However, decreased PC expression in hepatomas (Chang and Morris 1973) and decreased PC/PDH ratios in glioma and neuroblastoma cells compared to normal glia and neuronal tissue (Brand et al., 1992) prove otherwise (DeBerardinis et al., 2008).
Figure 4. Glutamine dependent anaplerosis replenishes TCA cycle intermediates. In the cytosol glutamine is converted to glutamate by donating nitrogen to purines and pyrimidines. Glutamate can then either enter the mitochondria or is converted to α-ketoglutarate, which can subsequently enter the mitochondria. Glutamine can also be converted to glutamate in the mitochondrial matrix by phosphate-dependent glutaminase (PDG). Mitochondrial glutamate is converted to α-ketoglutarate by glutamate dehydrogenase (PDG) and supplies the TCA cycle (DeBerardinis et al., 2008).

2.1.3 Metabolic Regulation of Cellular Proliferation

Normal cells enter proliferative metabolism only when instructed to do so by growth factors and upstream signaling pathways that control gene expression and cell physiology. The main signal transduction pathways that have a role in cellular proliferation include the PI3K/Akt/mTOR pathway along with the transcription factors HIF-1α and Myc (DeBerardinis et al., 2008).

The PI3K/Akt/mTOR pathway is a highly conserved pathway that responds to growth factor binding. When a growth factor binds to its receptor PI3K is activated which results in the phosphorylation of phosphotidylinositol lipids in the plasma membrane that activate the Akt and mTOR effectors. Activation of these effectors result in the increased expression of nutrient transporters in the plasma membrane allowing for higher uptake rates of glucose, amino acids, and other nutrients (DeBerardinis et al., 2008). In addition,
Akt has been experimentally shown to be capable of enhancing lactate production and glycolysis in both nontransformed cells and cancer cells (Elstrom et al., 2004; Plas et al., 2001; Rathmell et al., 2003). Moreover, PI3K and Akt have been shown to stimulate the expression of genes involved in lipid synthesis. In cancer cells, mutations trigger the activation of the PI3K/Akt/mTOR pathway. Mutations can occur through a variety of mechanisms. For example, one previously identified mutation mechanism is the elimination of the activity of PTEN, which is a negative regulator for PI3K (Buzzai et al., 2005).

Hypoxia inducible factor 1 (HIF1) is a transcription factor complex that is induced during hypoxia in normal cells and results in the expression of genes encoding glucose transporters as well as LDH-A. HIF1 is activated when the subunit HIF1α binds to it. HIF1α is in turn expressed by the growth factor signaling pathway, particularly the PI3K/Akt/mTOR pathway. During normal oxygen conditions, HIF1α is negatively regulated by the VHL (von Hippel Lindau) tumor suppressor which targets HIF1α for ubiquitination and subsequent degradation. During hypoxia, this degradation is prevented by the action of reactive oxygen species (ROS) generated in the mitochondria (DeBerardinis et al., 2008). Cancer cells have been observed to have mutations in the VHL tumor suppressor which results in the continual expression of HIF1α. Other regulators of HIF1α include the enzymes succinate dehydrogenase (SDH) and fumarate hydratase (FH) of the TCA cycle. The substrates for these two enzymes if accumulated, which would occur when SDH and FH are not functioning properly, impair HIF1α degradation, allowing it to enter the nucleus and engage in transcriptional activity. Mutations in SDH and FH subunits have been observed in some cancer cells (Isaacs et al., 2005; Selak et al., 2005).

Similar to HIF1α and the transcription factor Myc increases the expression of glycolytic enzymes and LDH-A. However, in addition the Myc family of genes also enhances the expression of specialized metabolic activities needed to duplicate the genome (DeBerardinis et al., 2008). Figure 5 provides an illustration of the regulatory signaling effects of the PI3K/Akt/mTOR pathway, HIF1α, and Myc.
Figure 5. Regulatory signaling network in proliferating cells. The PI3K/Akt/mTOR pathway, HIF1α, and Myc participate in the activation of various metabolic pathways to enhance cell proliferation (DeBerardinis et al., 2008).

2.2 Flux Balance Analysis

Flux Balance Analysis (FBA) also known as constraint based modeling (CBM), is a widely used method for studying metabolic networks and allows for the calculation of fluxes for the reactions within a metabolic network. Constraint based modeling analyzes the behavior of a metabolic network without the need for information on kinetic constants or enzyme and metabolite intracellular concentrations but relies solely on simple physical-chemical constraints (Shlomi et al., 2008). Constraint based modeling has been successfully used in many studies to predict the metabolic state of various organisms (Price et al., 2004).
CBM has also been used for studying human metabolism. Duarte et al. developed the first global human metabolic network consisting of 3,311 metabolic reactions in 2007 and since then this model has been used for many studies on human metabolism. The model by Duarte et al. has also been used very recently for the development of the first genome scale network of cancer metabolism (Folger et al., 2011). In this work, the enzyme-coding genes that are highly expressed across 90% of all cancer cell lines were first collected from the NCI-60 collection (Grever et al., 1992) and the reactions associated with these enzymes collected from the human model by Duarte et al. The Model Building Algorithm (MBA) was then used to add the minimal number of reactions required to activate the initial core set of reactions. This model was used to predict cytostatic drug targets, of which 40% were known targets and 60% new targets. In addition combinations of synthetic legal drug targets were also identified (Folger et al., 2011).

2.3 Ensemble Modeling

Detailed kinetic models (Wang et al., 2004; Chassagnole et al., 2002) have been difficult to develop because kinetic parameters are often unknown and obtaining these parameters requires detailed characterization of enzyme kinetics (Yan et al., 2011). Ensemble modeling is a modeling approach that bypasses the need for detailed characterization of kinetic parameters while developing kinetic models. Because this modeling approach is relatively new and was developed in 2008 by Tran et al. there are relatively few works that have used this computational methodology to date. A study by Contador et al. used ensemble modeling for the development of better strains of L-lysine producing Escherichia coli. In the study ensemble modeling allowed for a more predictive model of E. coli metabolism to be developed which was able to identify three new gene targets that could improve L-lysine production when overexpressed (Contador et al., 2009). In another study by Dean et al. ensemble modeling was used to better understand hepatic cell metabolism specifically in regards to fatty acid metabolism. Through ensemble modeling a metabolic inhibitor in the pathway was identified (Dean et al., 2010). To our
knowledge ensemble modeling has not yet been used for the study of cancer metabolism. A detailed description of the ensemble model methodology is presented in chapter 4.
Chapter 3
Knowledge Gaps and Statement of Objectives

The control systems present in mammalian cells can redirect the cellular response based on specific regulatory signals. These changes occur in the form of protein phosphorylation, protein glycosylation, or through alteration of gene expression levels and result in a specific phenotype. The resulting integrated end point phenotype after the many signaling changes have taken place can be represented through the metabolic profile (Boros et al., 2002). Therefore, metabolic profiling data are suitable for studies where the signaling and regulatory mechanisms are not fully understood, as is the case with cancer cells. Why cancer cells use the distinguished form of metabolism they do and the signaling changes responsible for their phenotype is not fully understood to date (DeBerardinis et al., 2008).

This knowledge gap has prompted us to utilize a metabolomics based approach for studying cancer metabolism. Metabolomics studies have been conducted in the past and include qualitative studies comparing the relative concentration of metabolites in cancerous vs. noncancerous cells (Chan et al., 2009; Denkert et al., 2006) as well a recent quantitative study measuring metabolite concentrations in the tumor tissue of colon and stomach cancer patients (Hirayama et al., 2009). However, these purely experimental studies are limited in that they cannot provide information regarding the flow of metabolites in the metabolic network of the cells. Because of the highly complex and integrated network of metabolism, computational means are necessary to analyze the metabolomics data. Previous computational studies on cancer metabolism have been purely computational without incorporation of metabolomics data (Folger et al., 2011).

In this study, we aim to utilize in vitro metabolomics data measured during the growth phase of a cancerous cell line in conjunction with the Ensemble Modeling (EM) computational algorithm to characterize the metabolism in cancer cells. We hypothesize that incorporating experimental metabolomics results into a mathematical model that
captures the metabolism in cancer cells would allow us to identify new enzymatic drug targets. Enzymatic drug targets refer to enzymes whose repression of activity would result in a reduction in growth rate. This work focuses solely on the identification of the enzymes. The mechanisms for which the enzyme activity could be biologically repressed (i.e. regulation at the transcriptional, post translational, or at the enzymatic level) are not considered. This discovery driven study could establish promising new insights into the metabolism of cancer cells and provide grounds for further investigations.

The objectives of this study can be summarized as follows:

1. Collect experimentally measured metabolomics data from cancer cells during the growth phase.

2. Develop a kinetic model of metabolism in cancer cells using the Ensemble Modeling algorithm and the metabolomics data collected.

3. Use the kinetic model developed to predict potential enzymatic drug targets.
4.1 Experimental Methods

The experimental component of the project aims to acquire the necessary data to complement and develop the model. This data includes intracellular and extracellular measurements of metabolite concentrations collected at regular time intervals during the growth phase of the cells. The samples are collected at the Laboratory for Metabolic Systems Engineering (LMSA) at the University of Toronto and partly measured at LMSE using high-performance liquid chromatography (HPLC) and partly sent to CENOMX [CHENOMX Inc., Edmonton, Alberta] for measurement. CHENOMX is a metabolomics company that uses nuclear magnetic resonance (NMR) spectroscopy in combination with advanced algorithms to generate, classify, and interpret metabolomic information. The favorable output of CHENOMX’s metabolomic analysis is that the metabolites are measured and reported in concentrations, as compared to fold-changes and ratios usually reported by other metabolomics companies.

4.1.1 Cell Line

In this study colo205 cell line (colorectal adenocarcinoma cells collected from the colon) are used. The cells were purchased from ATCC and cultured according to ATCC protocols. The mixed growth properties of these cells, which allow them to grow adherently or in suspension, make them suitable for metabolomics studies. For metabolomics purposes, growing the cells in suspension facilitates the data collection process.

4.1.2 Subculturing

Cells are grown in T70 flasks and RPMI-1640 medium. When an approximate 60% coverage of cells is observed within the flask under the microscope, the cells are split at a subcultivation ratio of 1:2. For splitting of the cells, floating cells are initially transferred
to a new flask. Cells that remain attached are removed by rinsing the bottom of the flask with 0.25% trypsin 0.53 mM EDTA solution. Additional trypsin-EDTA solution is added to the flask where then the flask is placed in an incubator (at 37ºC) for approximately 3 minutes in order for the cells to detach. Fresh RPMI-1640 medium is then added to the flask, aspirated, and dispensed into new culture flasks.

### 4.1.3 Experimental Protocol

The equipment and materials necessary for the experiments are as follows:

**Equipment:**

1. Five spinner flasks
2. Magnetic Spinner
3. Fumehood
4. Centrifuge tubes
5. Microcentrifuge tubes
6. Pipette
7. Microscope
8. Hemocytometer

**Material:**

1. SMEM medium supplemented with 0.292g/l L-glutamine, 10% FBS, and 1% Penicillin/Streptomycin
2. RPMI Medium supplemented with 10% FBS, and 1% Penicillin/Streptomycin
3. COLO205 cells
4. 0.25% trypsin 0.53 mM EDTA solution
5. Trypan blue

The experiments were conducted in five spinner flasks. The experimental set up is shown in figure 6. The following experimental procedure was followed:

1. Cells (from the T70flask) were placed in each of the five spinner flasks at a concentration of $2 \times 10^5$ cells/mL and a volume of 150mL of SMEM medium was added.
2. Every twelve hours from each spinner flask three 0.3mL samples and one 1.1mL sample of cell solution was withdrawn and put in a microcentrifuge vial.

3. The three 0.3mL samples were well mixed by pipetting and 90µL of cell solution was removed for cell counts.

4. The 1.1mL sample is centrifuged and the supernatant collected for extracellular metabolite measurement.

5. Two spinner flasks (#1 and #2) were used exclusively for extracellular metabolite measurements and steps 2-4 are repeated for these flasks up to the point of clumping (which was observed at ~108 hours).

6. Three spinner flasks (#3, #4, and #5) were used for intracellular metabolite measurements whereby at the 72\textsuperscript{nd} hour (middle of growth phase) 40mL of cell solution was collected. Extracellular metabolite measurements were also made for these flasks and steps 2-4 were repeated up to the 72\textsuperscript{nd} hour.

![Figure 6. Experimental setup; five flasks placed on a magnetic spinner inside an incubator](image-url)
For *intracellular metabolite measurements*, the samples collected were sent to CHENOMX whereby NMR spectroscopy was used to obtain concentration measurements. For intracellular metabolite sample preparation, the 40mL (~15 million cells) solution that was collected in step 6 of the protocol was centrifuged at 0°C. All but 1mL of the supernatant was then removed. The cells were resuspended in the remaining 1mL of supernatant, transferred to a microvial, and subsequently quenched in liquid nitrogen. The frozen samples were shipped to CHENOMX for analysis.

*Extracellular metabolite measurements* were made partly by CHENOMX and partly at the Laboratory of Metabolic Systems Engineering using HPLC. For CHENOMX sample preparation, the 1.1mL samples collected at hour 24, 72, and 96 in step 2 of the protocol were quenched in liquid nitrogen, and subsequently shipped to CHENOMX. For HPLC measurements, the samples were membrane filtered and ~0.7mL of filtered sample was placed in the HPLC vials for measurement. In this study 30mM solution of H$_2$SO$_4$ was used as the effluent. The temperature of the column was set to 60°C and an effluent flow rate of 0.3ml/min for the optimal separation of the four metabolites of interest: citrate, pyruvate, glucose, and glutamine. Both UV and RI refractive index measurements were made, and concentrations were obtained using the calibration curve initially created with standards.

For *cell counts* 0.3mL samples were collected in triplets every 12 hours from each spinner flask and were well mixed by pipetting. 90µL of solution was then removed, mixed with 10µL of trypan blue and cell counts were made using a hemocytometer. When clumping was observed under the microscope cell counting was terminated as the cell count would no longer be accurate.

### 4.2 Computational Methods

In this study two distinct modeling approaches were combined to develop a representative model of the system under study: Flux Balance Analysis and Ensemble Modeling.
4.2.1 Overview of Flux Balance Analysis

Flux Balance Analysis determines the rate of flow of metabolites, in other words the flux, through a metabolic network subject to imposed mass balance and capacity constraints. The reactions in the system are mathematically represented in the form of a numerical matrix. The constraints of the system include the matrix of stoichiometries which impose mass balance constraints on the system and the upper and lower bounds for each reaction, which define the maximum and minimum allowable fluxes of the reactions. Within the space of the constraints, a range of possible flux distributions is obtained. The system can be optimized for a specific phenotype in terms of a biological objective, such as biomass production. Optimizing the objective function then selects for a single set of flux values in the solution space. The conceptual basis of constraint based modeling is shown in Figure 7. As can be seen in Figure 7 when no constraints are imposed, the flux distribution of a biological network can lie at any point in the solution space. When mass balance and capacity constraints are introduced a solution space is defined, and when the solution space is optimized for a specific objective function, the single point on the edge of the solution space is identified as the single optimal flux distribution (Orth et al., 2010).

![Figure 7. The conceptual basis for flux balance analysis. Mass balance and enzyme capacity constraints introduce a solution space. Optimization of an objective function finds the an optimum solution within the solution space (Orth et al., 2010).](image-url)
4.2.2 Overview of Ensemble Modeling

The goal of EM is to generate a set of kinetic models whereby each model has different values for its kinetic parameters but all models retain the same model structure and are anchored to the same steady state value. If each enzymatic reaction in the model is represented as a nonlinear ordinary differential equation with the following mathematical representation:

\[
\frac{dx_i}{dt} = \sum v_p(x, k) - \sum v_c(x, k)
\]

(1)

where \(x_i\) is the species/metabolites in the model, and \(v_p\) and \(v_c\) are the the enzyme kinetic functions for the production and consumption of species \(x_i\) respectively, then the ensemble modeling problem is stated mathematically as sampling and obtaining all the different sets of kinetic parameters such that,

\[
v(x_{ss}, k) = v_{net}^{ref}
\]

(2)

where \(v_{net}^{ref}\) is the known steady state flux, and \(x_{ss}\) is the the steady state concentration of the metabolites in the model (Dean et al., 2010). In this study \(v_{net}^{ref}\) is obtained from the steady state fluxes calculated through FBA, and \(x_{ss}\) is obtained from the experimentally measured metabolite concentrations. For constrained sampling of kinetic parameter values, EM first breaks down each reaction into a series of elementary reactions. Elementary reactions are the most fundamental description of enzyme kinetics which follows mass-action kinetics. This conversion is necessary, as it is difficult and time consuming to meet the constraints of equation (2) when dealing with nonlinear functions. The log-linear properties of mass action kinetics make parameter sampling more efficient (Dean et al., 2010). Furthermore, the reactions in the network are also constrained thermodynamically to account for thermodynamic feasibility of the network. Once the initial ensemble of models is constructed, the possible phenotype of the system due to flux perturbations within the system can be examined (Tran et al., 2008). In this case, the fluxes perturbed are the previously studied and known cancer drug targets. The set of ordinary differential equations are then resolved with the same set of kinetic parameters.
but with repression of the enzymatic drug targets. The models that predict the experimental data (reduced biomass production/growth of cancer cells) are then retained. With more data, the ensemble converges to a set of models that more accurately describe the system and become more predictive. The advantage of using the ensemble modeling approach is that due to the elementary reaction form that each reaction is represented as, details about the true mechanism of enzymatic reactions such as enzyme regulation, thermodynamics, and steady-state metabolite levels can be incorporated into the model (Tran et al., 2008). An illustration of the EM approach is shown in Figure 8.

Figure 8. Visualization of the Ensemble Modeling methodology. An ensemble of models is initially generated, whereby each model displays different kinetic parameters but all models are anchored to the same reference flux value. A perturbation is introduced to the system computationally as well as examined experimentally. The models that display the same steady state behavior as the experimental results are retained (Tran et al., 2008).

As can be seen in Figure 8, in the original ensemble all models reach the same steady state value but have different kinetic parameters (left panel). When a perturbation (overexpression or underexpression) is introduced to the system all models reach a
different steady state value (top panel). The same perturbation is performed experimentally (bottom panel) in order to screen the models. The models that exhibit the same outcome as the experiments after perturbation are retained (right panel) (Contador et al., 2009).

4.2.3 Uptake and Secretion Fluxes

Uptake and secretion flux values are necessary as input data for FBA. The SensSB (Sensitivity Analysis for Systems Biology) toolbox, a MATLAB based software toolbox, is used to determine the uptake and secretion fluxes of the metabolites into and out of the system based on the extracellular metabolites measurements. Calculation of fluxes is based on the following mathematical principle:

\[
\frac{dy}{dt} = \rho y \\
\frac{dx_i}{dt} = v_i y
\]

where \( y \) is the concentration of cells, \( \rho \) is the growth rate of cells, \( x_i \) is the concentration of species/metabolite \( i \), and \( v_i \) is the uptake or secretion rate of metabolite \( i \) into the cells.

In this study, uptake fluxes were calculated for the metabolites glucose, and glutamine, which are taken up by the cell. Secretion fluxes were calculated for the metabolites: pyruvate, citrate, lactate, glutamate, and alanine, which experimentally proved to be secreted by the colo205 cells. Within SensSB, SSglobal optimization and fmincon local optimization settings were selected for carrying out parameter estimation.

4.2.4 Flux Balance Analysis Methodology

The internal steady state fluxes of the system are calculated using Flux Balance Analysis (FBA). In FBA the metabolic reactions are represented in an \( m \times n \) stoichiometric matrix of \( m \) metabolites and \( n \) reactions (Schilling et al., 1999). The flux through the reactions in the network are represented by the \( n \times 1 \) vector, \( v \). The internal fluxes are calculated by solving the system of mass balance equations at steady state:

\[
Sv = 0
\]
The solution space is constrained by upper and lower flux bounds, $v_{ub}$ and $v_{lb}$. Moreover, through optimization of an objective function, $c$, FBA identifies a single optimal flux distribution amongst the many flux distributions within the solution space. These constraints are mathematically represented as follows (Orth et al., 2010):

\[
\begin{align*}
    \max_{} & \quad c^T v \\
    \text{s.t.} & \quad S v = 0 \\
    & \quad v_{lb} \leq v \leq v_{ub}
\end{align*}
\]

### 4.2.5 Ensemble Modeling Methodology

An illustration of the EM algorithm is presented in Figure 9 (Contador et al., 2009). The EM methodology involves three main steps: model structure development, building the ensemble of models, and perturbing and screening the ensemble of models. The inputs to the system are shown by the dashed lines and include: reference steady state fluxes, standard Gibbs free energies, and enzyme tuning data. The steady state fluxes are obtained through FBA. The standard Gibbs free energy values for the reactions involved in the network are collected from literature (Jankowski et al., 2008). Reversibilities and enzyme fractions are sampled and used for the calculation of kinetic parameters. Each model in the ensemble of models is then represented by a set of ordinary differential equations based on the kinetic parameters calculated. Finally, the models are screened based on available enzyme tuning data. Detailed descriptions of each step are presented in the sections that follow.

#### 4.2.5.1 Establishing Model Structure Using Elementary Reactions

Every enzymatic reaction in the network is first broken down to its corresponding set of elementary reactions. Elementary reactions are the most fundamental form of reaction and represent events at the molecular level. For simple enzymatic reaction, where metabolite $X_i$ is converted by enzyme $E_i$ to metabolite $X_{i+1}$,

\[
X_i \xrightleftharpoons{E_i} X_{i+1},
\]

the reaction is broken down to six elementary reactions illustrated by:
Each elementary reaction, $v_{i,K}$, follows mass action kinetics,

$$v_{i,1} = k_{i,1}[X_i][E_i].$$

where $k_{i,1}$ is the rate constant for the first elementary reaction, $[X_i]$ is the concentration of metabolite $i$, and $[E_i]$ is the concentration of enzyme $i$. For more complex reactions which involve more substrates and products in the enzymatic reactions, corresponding elementary steps are added. Table 1 includes the list of elementary reaction mechanisms for different reaction types (Dean et al., 2010).

---

**Figure 9.** Summary of the main steps involved in EM. The models in the network are initially broken down into the elementary reaction form. Through sampling of reversibilities and enzyme fractions kinetic parameters are calculated. Each of the models in the ensemble of models is computationally perturbed by repressing the enzymes that have previously been identified as drug targets. The models that show a decrease in growth rate after perturbation are retained and used to predict new enzymatic drug targets (adapted from Contador et al., 2009).
### Table 1. Decomposition of enzymatic reactions into elementary form (Dean et al., 2010)

<table>
<thead>
<tr>
<th>Reaction Type</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Substrate, 1 product</td>
<td>$x_i + e_i \xrightarrow{\eta_{i,1}} x_{i+1} + e_i$</td>
</tr>
<tr>
<td>1 Substrate, 2 products</td>
<td>$x_i + e_i \xrightarrow{\eta_{i,1}} x_{i+1} + e_i$, $x_i + e_{i+1} \xrightarrow{\eta_{i,2}} x_{i+2} + e_i$</td>
</tr>
<tr>
<td>2 Substrates, 1 product</td>
<td>$x_i + e_i \xrightarrow{\eta_{i,1}} x_{i+1} + e_i$, $x_i + e_{i+1} \xrightarrow{\eta_{i,2}} x_{i+2} + e_i$</td>
</tr>
<tr>
<td>2 Substrates, 2 products</td>
<td>$x_i + e_i \xrightarrow{\eta_{i,1}} x_{i+1} + e_i$, $x_i + e_{i+1} \xrightarrow{\eta_{i,2}} x_{i+2} + e_i$</td>
</tr>
</tbody>
</table>

#### Competitive inhibition

- $x_i + e_i \xrightarrow{\eta_{i,1}} x_{i+1} + e_i$
- $x_i + e_{i+1} \xrightarrow{\eta_{i,2}} x_{i+2} + e_i$

#### Noncompetitive inhibition

- $x_i + e_i \xrightarrow{\eta_{i,1}} x_{i+1} + e_i$
- $x_i + e_{i+1} \xrightarrow{\eta_{i,2}} x_{i+2} + e_i$

In order to allow for easier and more accurate numerical simulations, equation (11) is normalized by scaling the concentration of metabolites with the steady state metabolite concentration, $X_i^{ss,ref}$, and by scaling the concentration of the free enzyme and enzyme complexes with the total enzyme concentration at steady state, $E_{i,total}^{ref}$ (Carr et al., 2001). Equation (11) then becomes

$$v_{i,1} = (k_{i,1}E_{i,total}^{ref}X_i^{ss,ref}) \cdot \frac{[X_i]}{X_i^{ss,ref}} \cdot \frac{[E_i]}{E_{i,total}^{ref}} = \bar{R}_{i,1}^{ref} \cdot \bar{X}_i \cdot \bar{E}_{i,1}.$$  \hspace{1cm} (12)

In log-linear form equation (12) has the following form

$$\ln v_{i,1} = \ln \bar{R}_{i,1}^{ref} + \ln \bar{X}_i + \ln \bar{E}_{i,1}.$$ \hspace{1cm} (13)

At reference steady state, $\bar{X}_i^{ss,ref}$ equals 1 and equation (13) becomes (Tran et al., 2008)

$$\ln v_{i,1}^{ref} = \ln \bar{R}_{i,1}^{ref} + \ln \bar{E}_{i,1}^{ref}.$$ \hspace{1cm} (14)
4.2.5.2 Sampling reversibilities of elementary steps and enzyme fractions

The first sampling step involves sampling of the reversibilities for each of the enzymatic reactions. The sampling involves a Monte Carlo algorithm, and the reversibility values range from 0, a completely irreversible reaction, to 1, a completely reversible reaction at equilibrium. The reversibility is defined as the ratio of the smaller value for the forward and reverse reaction rates over the larger value,

\[ R_{i,j} = \frac{\min(v_{i,2j-1}, v_{i,2j})}{\max(v_{i,2j-1}, v_{i,2j})} \]  

(15)

where, \( v_{i,2j-1} \) and \( v_{i,2j-1} \) are the forward and backward rates of step \( j \) in reaction \( i \). From the reversibilities, the forward and backward elementary reaction rates can be calculated using the additional constraint,

\[ v_{i,2j-1}^{ref} - v_{i,2j-1}^{ref} = V_{i,net}^{ref} \]  

(16)

where \( V_{i,net}^{ref} \) is the net flux of reaction \( i \) at reference steady state (Tran et al., 2008). The reversibilities are constrained using Gibbs free energy to ensure that the steady state is thermodynamically feasible,

\[ \sum_{j=1}^{n_i} \ln R_{i,j} = \text{sign}(V_{i,net}) \frac{\Delta G_i}{RT} \]  

(17)

where \( n_i \) represents the number of elementary steps present in the enzymatic reaction \( i \) and \( \text{sign}(V_{i,net}) \) is the direction of the reaction: +1 for forward reactions and -1 for reverse reactions. Equation (17) ensures that that for each reaction, the net flux is positive if \( \Delta G < 0 \) and is negative if \( \Delta G > 0 \). Since the values for the Gibbs free energy depend on the metabolite concentrations, an exact value cannot be calculated and a range is established (Dean et al., 2010),

\[ \left( \frac{\Delta G}{RT} \right)_{\text{lower bound}} \leq \text{sign}(V_{i,net}^{ref}) \cdot \sum_j \ln R_{i,j}^{ref} \leq \left( \frac{\Delta G}{RT} \right)_{\text{upper bound}}. \]  

(18)
The different reversibilities are a representation of different kinetic states. For example, if within the enzymatic reaction \( i \), \( R_{i,j} \) for step \( j \) is close to zero while that of the next step is near 1, step \( j \) is determined to be the rate limiting step (Tran et al., 2008).

In addition to reversibilities, the enzyme fractions are also sampled. At steady state, the total enzyme concentration for each reaction is the sum of the free enzymes and bound enzymes. The distribution of the total enzyme amount over the free enzymes and bound enzymes affects the kinetics of the system and for this reason the enzyme fractions are considered. The enzyme fractions are sampled using a Monte Carlo algorithm with the constraint that the total enzyme amount is conserved. In other words, the sum of the enzyme fractions of the elementary reactions for each enzymatic reaction must equal one (Contador et al., 2008),

\[
\sum_{j=1}^{n_i} \bar{e}_{i,j}^{\text{ref}} = 1. \quad (19)
\]

### 4.2.5.3 Calculating the Kinetic Parameters

Once the reversibilities and enzyme fractions are sampled and their values generated, the kinetic parameters, \( \bar{R}_{i,k}^{\text{ref}} \), can be computed. The rates of elementary reactions are computed from the reversibilities as follows:

\[
\bar{v}_{i,2j-1}^{\text{ref}} = \frac{\bar{v}_{i,\text{net}}^{\text{ref}}}{1 - R_{i,j}}. \quad (20)
\]

\[
\bar{v}_{i,2j}^{\text{ref}} = \frac{\bar{v}_{i,\text{net}}^{\text{ref}} \cdot R_{i,j}}{1 - R_{i,j}}. \quad (21)
\]

Finally \( \bar{R}_{i,k}^{\text{ref}} \) is calculated from equation (14) based on the elementary reaction rates and enzymes fractions (Tran et al., 2008).
4.2.5.4 ODE Models with Kinetic Parameters

The process of calculating kinetic variables based on sampled reversibilities and enzyme fractions can be repeated thousands of times to obtain thousands of models. All models reach the same steady state however each model has a unique set of kinetic parameters,

\[
Model_k = f(R_k^{ref}, e_k^{ref}).
\]  

(22)

Once the kinetic parameters are calculated, the net steady state fluxes are calculated. To calculate the net steady state flux, the metabolic network is first described as a system of ordinary differential equations, with the metabolite concentrations, \( \bar{X}_i \), and the enzyme fractions, \( \bar{e}_{i,j} \), as the ODE variables:

\[
\frac{d\bar{X}_i}{dt} = \frac{1}{x_i^{s,ref}} \left( \sum v_{generation} - \sum v_{consumption} \right)
\]

(23)

\[
\frac{d\bar{e}_i}{dt} = \frac{1}{E_{i,Total}^{ref}} \left( \sum v_{generation} - \sum v_{consumption} \right)
\]

(24)

The ODEs are solved using the ode15s solver in MATLAB with an integration time of 1000 (time units) and a step size of 10 (time units). The concentrations obtained are input into equation (12) to calculate the steady state fluxes (Contador et al., 2008).

4.2.5.5 In Silico Enzyme Tuning

After generating the ensemble of models, the models are perturbed to study the behavior of each model under the perturbations. Perturbations involve overexpressing or repressing specific enzymes in the model by a factor of \( n \). In this study the enzymes of interest are drug targets and are therefore repressed with an \( n \) value of 0.1 to study their respective effects on the growth rate of the cancer cells. The models that after perturbation show decreased growth rate behavior are retained and as such the ensemble of models are screened again. For the perturbation case equation (13) is rewritten as follows:

\[
\ln v_{i,1} = \ln \bar{R}_{i,1}^{ref} + \ln \bar{X}_i + \ln \bar{e}_{i,1} + \ln E_{i,r}
\]

(25)
where \( E_{r} \) represents the fold change in the total enzyme concentration over the reference steady state value (\( E_{r} = 0.1 \) for this study). Furthermore, for the perturbation studies, if the metabolic network contains moiety conservation relationships such as cofactors, the initial conditions are set such that the sum of the cofactors and their intermediates are conserved before and after the perturbation (Contador et al., 2009).

### 4.2.6 Major Assumptions and Limitations of the Models

In this study, the development and screening of the models using the ensemble modeling approach is based on steady state conditions. Initially, the intracellular steady state flux distribution is calculated using FBA. FBA takes in as input, the steady state uptake and secretion fluxes of metabolites that are transported across the membrane of the cell. These fluxes are calculated from the experimentally measured extracellular metabolite concentration profiles. In this study, it is assumed that there is a constant steady state flux for the uptake and secretion of metabolites during the growth phase. The basis of this assumption lies in the fact that the metabolic changes which take place inside of the cell occur very rapidly and almost instantaneously, even though the large-scale changes measured in the medium take place in the time scale of hours to days. This assumption allows us to capture steady state flux values from dynamic concentration profiles. Since the uptake and secretion rates are assumed to be at steady state, the intracellular flux distribution calculated using FBA is also assumed to be at steady state. Indeed, the FBA does have inherent steady state assumptions built into the methodology (Equation (7)).

There are, however, drawbacks to the FBA approach. FBA does not take into account regulation. Furthermore, for this specific study the enzyme capacity constraints are not known for the reactions in the system. These drawbacks would cause FBA to possibly over predict the metabolic capacity of the system under study.

The purpose of ensemble modeling is to develop kinetic/dynamic models based on the steady state reaction fluxes. This dynamic allows us to introduce perturbations to the system and obtain a new steady state. In this study, the screening of the models is based on the final steady state behavior. For example, when the enzyme lactate dehydrogenase (LDH) is mathematically repressed, and the system of ordinary differential equations is
solved, the behavior of the system only after it has reached steady state is of interest. The models that show a reduced growth rate at steady state, compared to the case where no enzyme is repressed, are kept.

There are certain limitations to the ensemble modeling methodology. One of the limitations is that EM is a computationally intensive algorithm. Depending on the size of the metabolic network considered, the computational time could take between hours to days. In this study, considering a network of 58 reactions and screening using four perturbations, it took approximately 3 days to run the code. Therefore, if larger genome-scale systems were to be studied, the computational time could introduce challenges that need to be overcome. Furthermore, the EM approach is limited by one's access or ability to calculate the reference steady state fluxes, as these fluxes constitute the primary input into the algorithm (Tran et al., 2008). In this study, FBA is used to calculate the steady state fluxes from experimentally measured uptake and secretion rates. If the internal fluxes were directly measured a more accurate flux distribution could be obtained, resulting in a more representative and realistic ensemble of models. However, experimentally measuring the internal fluxes using $^{13}$C labeling for all the pathways in the metabolic network is both difficult and costly. As the metabolic network is expanded to include more reactions this limitation is further pronounced.
5.1 Experimental Results

5.1.1 Growth Curve

The growth rate of the colo205 cells is an essential input for the Flux Balance Analysis computations, whereby the values for the internal fluxes are computed such that the computationally simulated growth rate matches what is experimentally observed. The growth curve for the colo205 cells is presented in Figure 10.

As seen in Figure 10, the experiments gave quite reproducible results. The error bars represent the standard deviation between the five spinner flasks. Samples were taken at 12 hour intervals until clumping was observed under the microscope whereby sampling was stopped due to inaccuracy of the counts (hour 120 onward). The shape of the growth curve obtained agrees with the sigmoidal behavior for growth of mammalian cells in suspension noted in literature (Jakoby 1979). The initial section of the growth curve, where the slope
is close to zero, represents the “lag phase” in which the cells undergo little or no division. The exponential part of the curve, from hour ~24 to 97, is termed the “growth phase” and the cells are undergoing division at the maximum rate. The last section of the curve is the “stationary phase” where growth decreases due to nutrient depletion or accumulation of deltarious waste products (Jakoby 1979). The growth rate for the colo205 cells was obtained by fitting equation (3) to the cell count data during the “growth phase” (Figure 11). To obtain growth rate in units of gDW/hr the cell number was first converted to mass. The mass of a human colon adenocarcinoma cell during the growth phase has been reported in literature to vary between ~0.5-1.8 ng (Park et al., 2010). An average value of 1.15ng/cell was used for the calculations.

![Figure 11. Exponential fit to growth phase](image.png)

From the fit of the curve with $R^2$ value of 0.9917, parameter $\rho$ was determined to be 0.0224 gDWhr$^{-1}$.

### 5.1.2 Internal and External Metabolite Concentrations

Extracellular and intracellular metabolite concentrations were collected as explained in chapter 4. These samples were sent to CHENOMX for measurement using NMR mass
spectroscopy. CHENOMX reported the concentrations for 35 intracellular and extracellular metabolites. The measured extracellular concentration profiles for the amino acids are presented in Figure 12.

![Graph showing extracellular amino acid concentration profiles](image)

**Figure 12.** Extracellular amino acid concentration profiles. The error bars represent the standard deviation between the samples.

The five spinner flasks gave quite reproducible results as can be seen by the standard deviation error bars in Figure 12. Of the 14 amino acids measured most of them are
shown to be taken up by the colo205 cells except for the nonessential amino acids, alanine and glutamate. For the purposes of this study, only the amino acids alanine, glutamate, and glutamine were used for uptake and secretion rate calculations, as not all of the metabolic pathways have been incorporated in the model. However, for future studies where a more expanded metabolic network is considered, this additional data could be incorporated. Concentration measurements for the carbohydrates citrate, pyruvate, lactate, and glucose were made using HPLC, and are presented in Figure 13.

![Figure 13](image)

**Figure 13.** Extracellular carbohydrate concentration profiles measured by HPLC. The error bars represent the standard deviation between the samples.
The carbohydrate concentration profiles also demonstrated reproducibility, with the error bars representing standard deviation between the samples. The data for the four carbohydrates presented in the Figure 19 have been incorporated into the model. The full list of extracellular metabolites measured by CHENOMX and their concentration profiles is presented in appendix B.

Notable trends include that of the high glutamine and glucose uptake rates, further proving them as the main source of nutrient uptake in cancer cells. Glucose is the major lipogenic substrate for cancer cells as highlighted in green in Figure 3 (DeBerardinis et al., 2008) and therefore high uptake rates are essential. Glutamine is an anaplerotic source during cell proliferation, replenishing the TCA cycle carbon that is used for biosynthesis, and therefore proposed to be an essential nutrient for cancer metabolism (DeBerardinis et al., 2008). The high rates of glutamine observed in the experiments, further verify this finding. In addition, the high secretion rate of lactate observed in cancer cells, the prominent indicator of the Warburg effect, is also observed in the experimental results. An interesting experimental finding is the accumulation of pyruvate in the medium. The high rates of glycolysis in cancer cells result in high pyruvate production. However, paradoxically, most of the pyruvate does not enter the TCA cycle for ATP production, but is converted to lactate by the highly expressed enzyme LDH-A which is induced by oncogenes during proliferation (Vander Heiden et al., 2009). The experimentally observed secretion of pyruvate out of the cells could further prove the hypothesis made by Curi et al. that there is a limitation to the maximum rate of pyruvate oxidation (Curi et al., 1988). However, in terms of LDH-A activity the question of whether there is a limitation also in the maximum rate of lactate production from pyruvate could be addressed.

Intracellular metabolite concentrations were also measured in this study and are inputs for ensemble modeling. The experiments gave quite reproducible results with low standard deviation. The full list of intracellular metabolites measured and their corresponding concentrations are presented in Figure 14. The metabolite concentrations used in this study for developing the model include pyruvate, glutamine, malate, lactate, glucose, citrate, fumarate, succinate, alanine, and glutamate. Samples for intracellular
measurements were collected during hour 72 to capture the steady state behaviour of the colo205 cells during the growth phase.

**Figure 14a.** Experimentally measured intracellular metabolite profiles
Figure 14b. Experimentally measured intracellular metabolite profiles

5.2 Computational Results

5.2.1 Uptake and Secretion Rate

As described in chapter 4, the SensSB toolbox was used for calculation of uptake and secretion rates from concentration profiles of selected metabolites. The calculated fluxes are presented in Table 2.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Uptake Rate (mmol/gDWhr)</th>
<th>Secretion Rate (mmol/gDWhr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>0.672</td>
<td>-</td>
</tr>
<tr>
<td>Glutamine</td>
<td>0.346</td>
<td>-</td>
</tr>
<tr>
<td>Pyruvate</td>
<td>-</td>
<td>0.027</td>
</tr>
<tr>
<td>Citrate</td>
<td>-</td>
<td>0.012</td>
</tr>
<tr>
<td>Lactate</td>
<td>-</td>
<td>0.820</td>
</tr>
<tr>
<td>Glutamate</td>
<td>-</td>
<td>0.032</td>
</tr>
<tr>
<td>Alanine</td>
<td>-</td>
<td>0.106</td>
</tr>
</tbody>
</table>

An example of the SensSB fitting is shown in Figure 15 for lactate. The red line shows the optimal fit for the lactate data. The blue line is the simulation for growth rate that was simultaneously solved according to equations (3) and (4).
5.2.2 Model Reconstruction

A reconstructed model of human central metabolism is generated in this study to further understand the metabolic behavior of cancer cells. The model is a reconstruction of the previously developed 3,742 reaction genome-scale human metabolic network (Duarte et al., 2007). In this study, the model includes 58 reactions and 57 metabolites and only the major metabolic pathways related to the area under study are incorporated. These pathways include glycolysis, TCA cycle, the pentose phosphate pathway, pyruvate metabolism, amino acid metabolism for selective amino acids, lipid biosynthesis, and nucleotide biosynthesis. The metabolic network is compartmentalized to account for the mitochondria, cytoplasm, and the extracellular environment. Metabolites are exchanged between compartments through exchange reactions. However, in this model cofactors are assumed to be freely transported between compartments and no exchange reactions are considered for them. Furthermore, balancing of the cofactors (ATP/ADP, NADH/NAD, NADPH/NADP, GTP/GDP, and FADH2/FAD) is taken into account in this network. A depiction of the metabolic network reconstructed for this study is shown in Figure 16.

A biomass equation is included in the model which is based on a previous study that computationally verified a biomass equation for cancer cells (Shlomi et al., 2011). The
biomass equation accounts for lipid, protein and nucleotide synthesis. In this study, the rate of biomass production is assumed to be equal to the growth rate obtained experimentally.

Figure 16. Reconstructed metabolic network used for the study of cancer metabolism. Metabolic reactions are shown in black arrows, transport reactions between compartments are shown in red arrows, and selected enzymes are reported in blue. Appendix A provides the list of enzymes and their abbreviations.

5.2.3  **Flux Balance Analysis**

Flux balance analysis was conducted in the MATLAB programming environment. The internal fluxes were calculated such that the known biomass, uptake, and secretion fluxes that were measured experimentally met the computationally simulated values. This was
done by minimizing the error between the computationally simulated and experimentally measured fluxes for the metabolites whose flux was experimentally measured (Table 2). The objective function for FBA then takes following form:

$$f = \sum_{i=1}^{8} abs \left( \frac{v_{i,s} - v_{i,m}}{v_{i,m}} \right)$$  \hspace{1cm} (26)$$

where $v_{i,s}$ is the simulated flux value and $v_{i,m}$ is the experimentally measured flux value for reaction $i$. The eight reactions considered include glucose uptake, glutamine uptake, pyruvate secretion, citrate secretion, lactate secretion, glutamate secretion, alanine secretion, and biomass production. These are reactions for which the rates were experimentally measured as described in sections 5.2.1. and 5.1.1. The internal fluxes were then calculated using FBA by minimizing the objective function subject to mass balance constraints. The upper and lower bounds for fluxes were also constrained to ensure that the direction of the fluxes were thermodynamically feasible. The steady state fluxes of the reference state are reported in the flux map presented in Figure 17.

As demonstrated in Figure 17 the values calculated capture the reported behavior of metabolism in cancer cells. The highest numerical flux values in the metabolic network are observed in glycolysis. In addition, a large portion of the glycolytic flux is drawn for lipid metabolism. This agrees with the explanation that the high glycolytic rate observed in cancer cells is due to the necessary biosynthetic requirement for proliferation, whereby glycolytic intermediates are used as biosynthetic precursors (DeBerardinis et al., 2008).

The pyruvate produced through glycolysis can enter many pathways: lactate production, amino acid synthesis, TCA cycle, and secretion out of the cell. The numerical results demonstrate this split quantitatively and suggest that 66.1% of the pyruvate is used for lactate production, 9.4% used for synthesis of the amino acid alanine, 19.3% enters the TCA cycle while 5.2% is secreted out of the cell. It is clear that most of the pyruvate is converted to lactate as also described by the Warburg effect (Warburg 1956). Figure 17 shows that there is in fact flux flow through the TCA cycle agreeing that the mitochondrion is functional. A notably high flux of glutamine synthesis is observed in the
cytoplasm. This glutamine enters the mitochondria and feeds the TCA cycle through conversion to glutamate.

Figure 17. The flux map at reference steady state for the colo205 cells. Black arrows show metabolic reactions and red arrows show transport reactions between compartments. Flux values are reported in units of mmol/gDW hr.

5.2.4 Construction of the Initial Ensemble
An ensemble of 1000 models was constructed as described in the Methods section. In addition to the steady state input obtained through FBA, other inputs include Gibbs free
energies for the net reactions which were collected from literature (Jankowski et al., 2008) as well as experimentally measured intracellular metabolite concentrations at reference state (Figure 14). Allosteric regulation is also included in the model, a list of which is presented in Table 3. When all the allosteric regulations listed in Table 3 were initially included, the models did not describe the cancer metabolic phenotype and did not converge to the steady state reference flux values. Each of the regulatory components were then examined individually and we found that the models would not converge when the activation of pyruvate kinase by fructose-1,6-bisphosphate was included. Indeed studies have shown that in cancer cells the growth factor signalling pathway activates tyrosine kinases which bind to pyruvate kinase and result in the release of the otherwise tightly bound fructose-1,6-bisphosphate. The low activity form of pyruvate kinase has been shown to be essential for aerobic glycolysis (Christofk et al., 2008). The ensemble of models was then generated with the allosteric regulators listed in Table 3 excluding the regulation of pyruvate kinase by fructose-1,6-bisphosphate.

Table 3. Allosteric regulation considered in the model

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Regulatory Metabolite</th>
<th>Activator (+)/Inhibitor(-)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hexokinase</td>
<td>Glucose-6-phosphate</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td>Phosphofructokinase</td>
<td>ATP</td>
<td>-</td>
<td>Elliot, W. H. 2009</td>
</tr>
<tr>
<td></td>
<td>Citrate</td>
<td>-</td>
<td>Elliot, W. H. 2009</td>
</tr>
<tr>
<td></td>
<td>Phosphoenolpyruvate</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td>Pyruvate kinase</td>
<td>Fructose-1,6-bisphosphate</td>
<td>+</td>
<td>Elliot, W. H. 2009</td>
</tr>
<tr>
<td></td>
<td>ATP</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>Alanine</td>
<td>-</td>
<td>Elliot, W. H. 2009</td>
</tr>
<tr>
<td></td>
<td>Acetyl-CoA</td>
<td>-</td>
<td>Elliot, W. H. 2009</td>
</tr>
<tr>
<td>Pyruvate dehydrogenase</td>
<td>NADH</td>
<td>-</td>
<td>Elliot, W. H. 2009</td>
</tr>
<tr>
<td></td>
<td>ATP</td>
<td>-</td>
<td>Elliot, W. H. 2009</td>
</tr>
<tr>
<td>Citrate Synthase</td>
<td>Citrate</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>Succinyl-CoA</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>NADH</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>ATP</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>ADP</td>
<td>+</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td>Isocitrate dehydrogenase</td>
<td>ATP</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>NADH</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>ADP</td>
<td>+</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td>α-ketoglutarate dehydrogenase</td>
<td>NADH</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
<tr>
<td></td>
<td>Succinyl-CoA</td>
<td>-</td>
<td>Michal, G., 1999</td>
</tr>
</tbody>
</table>
5.2.5 Perturbation and Screening of the Ensemble

The ensemble of models is screened based on experimental drug target data (Table 4) collected from DrugBank, which is a database that includes drug data and drug target information (Wishart et al., 2008). The enzymes presented in Table 4 have been proposed to reduce the proliferation rate of cancer cells when the activity of the enzymes is repressed. The information for drug targets available in DrugBank is based on multiple studies on the drug targets. It is therefore important to note that the studies were not conducted specifically on a colon cancer cell line. For example, one of the studies that identified lactate dehydrogenase (LDH) as a potential drug target, was carried out using an A549 (adenocarcinomic human alveolar basal epithelial) cell line. A shRNA construct for the LdhA gene was used to repress LDH activity through gene silencing (Xie et al., 2009). As such, a major assumption in our study is that the models obtained after perturbation screenings are generalized models of cancer metabolism and are not specific to the colo205 cells.

The models are computationally perturbed by repressing each of the enzymes listed in Table 4 by a factor of 0.1. This is achieved by reducing the concentration of the enzymes of interest by a factor of 0.1 as described in section 4.2.5.5 (refer to Equation (25)). As demonstrated by Figure 18 the ensemble of 1000 models converges to a set of 4 models capable of describing the perturbation phenotypes. As targeting the enzymes listed in Table 4 would reduce the tumor size, it is assumed that targeting the enzymes in this study reduces biomass production. The results of the screening are presented in Figure 18. The kinetic parameters for the four models are presented in appendix C.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Enzyme</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hk1</td>
<td>Hexokinase (HEX1)</td>
<td>Wishart et al., 2008</td>
</tr>
<tr>
<td>LdhA</td>
<td>Lactate dehydrogenase (LDH)</td>
<td>Wishart et al., 2008</td>
</tr>
<tr>
<td>Shmt1</td>
<td>Glycine hydroxymethyltransferase (GHMT2r)</td>
<td>Wishart et al., 2008</td>
</tr>
<tr>
<td>Nme1</td>
<td>Nucleoside-diphosphate kinase (NDPK1)</td>
<td>Wishart et al., 2008</td>
</tr>
</tbody>
</table>
5.2.6 Predicting Cancer Drug Targets

The models obtained through screening of the ensemble were used to identify potential drug targets. This was accomplished by perturbing every enzyme in the network and identifying enzymes that showed a greater reduction in the biomass production rate compared to the previously identified drug targets used for screening the models (Figure 19). Two enzymes were identified to have a greater decrease in biomass production than glycine hydroxymethyltransferase (GHMT2r). The enzyme target with the highest reduction in biomass production when repressed was the transaldolase (TALA) enzyme from the pentose phosphate pathway. The gene encoding for transaldolase, *Taldo1*, is not a known, approved, or experimental anticancer drug target. In this study, we identify transaldolase to be a potential drug target. Transaldolase is a key enzyme in the nonoxidative pentose phosphate pathway that provides ribose-5-phosphate for nucleic
acid synthesis and NADPH for lipid biosynthesis (Ma et al., 2009). In addition, the product of the transaldolase reaction, erythrose-4-phosphate, is used for the synthesis of the three amino acids tyrosine, phenylalanine, and tryptophan (Samland and Sprenger, 2009). This enzyme has been found to be overexpressed in all cancer cells (Lee et al., 2007) and has been suggested as a cancer biomarker specifically for colon cancer (Peng et al., 2009). These results highlight the potential of transaldolase as a target. Further experimental studies are required to evaluate the impact of repressing the enzyme in normal proliferating cells in vivo.

The enzyme that showed the second highest decrease in biomass production was succinate-CoA ligase (SU.COAS1m). Like transaldolase, succinate-CoA ligase has not been reported as a known, approved, or experimental drug target (Wishart et al., 2008), however, the gene associated with the enzyme has been shown to be overexpressed in cancer cells (Lee et al., 2007).

The effect of simultaneously repressing the enzymes was also studied by repressing two enzymes at a time. The effect of TALA-SU.COAS1m repression and TALA-GHMT2r repression was considered (Figure 20). It was found that the TALA-SU.COAS1m combination did not result in a much greater decrease in growth rate compared to TALA alone. However the combination of TALA and GHMT2r showed a three fold decrease in growth rate relative to the solo case. These results suggest the synergetic effectiveness of the potential candidates. Further experimental studies are required to evaluate their potential in vivo.
5.3 Discussion

In this work intracellular and extracellular metabolite concentrations were measured along with the growth rate of colo205 cells to develop a kinetic model of cancer metabolism using the EM computational methodology. An ensemble of
1000 models was generated. Each of the models in the ensemble consisted of different kinetic parameters, but all models were anchored to the same steady state flux. The steady state reference fluxes were calculated by FBA using experimentally measured uptake and secretion rates. In order to screen for models which more accurately represented the system under study, perturbation data were used. Four enzymes (LDH, GHMT2r, HEX1, and NDPK1) have previously been noted in literature to show a reduction in cancer cell proliferation rate when their activity was repressed (Wishart et al., 2008). In this work, we computationally reduced the activity for each of these enzymes to analyze the effect of this reduction on growth rate. The models that showed a reduction in growth rate were kept. Four models out of the 1000 initial models were able to accurately display a reduction in growth rate when the aforementioned enzymes were repressed.

Ensemble modeling generates kinetic models and determines the kinetic parameter for each reaction in the network. These models can be used to study the dynamic behavior as well as the steady state behavior of metabolism. Previous studies on computational modeling of cancer metabolism have only used FBA for metabolic studies and are therefore limited to steady state conditions. This work is the first study on cancer metabolism using the ensemble modeling approach. Besides being able to address dynamic behavior, the advantage of ensemble modeling is that any type of experimental information can be incorporated in the model to obtain a more predictive model. For example, in this work we’ve incorporated steady state intracellular metabolite concentrations during growth as well as regulatory information for cancer metabolism. Quantitative or qualitative information on experimentally measured enzyme concentrations, metabolite concentrations, or kinetic parameter values before and after perturbation could also be used to screen the models. For this study, we have used the relative growth rate changes that occur after repressing enzyme activity to screen the models. Moreover, as the ensemble modeling technique captures dynamic behavior, if experimental information
regarding the dynamics of cell metabolism is available, it could additionally be used to screen the ensemble.

The four models that remained after screening and determined to be most representative of the actual metabolic behavior in cancer cells were used for predictive studies. Our models predicted that repressing the activity of the transaldolase enzyme would result in the greatest reduction in growth rate. Transaldolase is a key enzyme in the pentose phosphate pathway and takes part in nucleotide, lipid, and amino acid synthesis. As biomass production is also a function of nucleotide, amino acid, and lipid production rates, it seems reasonable that transaldolase has a substantial role in biosynthesis. Transaldolase has not previously been reported as an experimental or approved drug target (Wishart et al., 2008). Previous computational studies on cancer metabolism using FBA alone have also not identified transaldolase as a potential drug target (Folger et al., 2011). Interestingly, the gene associated with the transaldolase enzyme is overexpressed in all cancer cell lines (Lee et al., 2007) and has been established as a cancer biomarker for colon cancer patients (Peng et al., 2009).

The ensemble modeling approach in this study has predicted information about effective enzymatic targets that have not previously been established. Because computational methods are capable of capturing the complex nature of the metabolic interactions, they can provide insight into information that might not be obvious without the aid of computational tools. Conclusions made from computational studies can establish new hypothesis and provide direction for further experimental studies. From the result of our computational studies, we suggest that targeting transaldolase would be an effective means for reducing growth rate in cancer cells. Our models further predicted that the synergetic repression of transaldolase with glycine hydroxymethyltransferase would result in a three fold greater reduction in growth rate compared to the solo repression of
transaldolase. This prediction suggests the potential efficacy of designing drugs that would target more than one enzyme simultaneously.

The next direct experimental step to complement the study would be to examine the transaldolase enzyme in terms of its composition and properties in cancerous versus noncancerous cells. Any differences in the enzyme composition within these two cells would make this enzyme an ideal candidate as a drug target. For example, there are many interests in the enzyme pyruvate kinase, as its composition in cancerous cells varies from that of normal proliferating cells. Such enzymes are sought after as they can be targeted without affecting the metabolism of the normal proliferating cells present in the body.

To enhance the predictive capability of the models developed in our study, the most significant improvement would be to measure the steady state fluxes using $^{13}$C labeling. As the ensemble of models that is initially generated is anchored to the steady state fluxes, it is critical that these fluxes be as close to their real values as possible. The assumptions inherent in FBA are responsible for the greatest fraction of uncertainty in the overall uncertainty present in the methodology used in this study. Flux balance analysis is based on mass balance constraints and does not account for regulatory constraints. In addition, enzyme capacity constraints are usually unknown. Moreover, there is a possibility that due to the redundancies in the metabolic pathways, multiple optimal steady states could exist. However, when experiments are conducted, the steady state fluxes measured would capture the metabolic behavior of the cells the most accurately.

In this study, the models generated were used to predict potential enzymatic drug targets. There are many other studies that can be conducted using these models. For example, they can be used for understanding the mode of action of a drug under study. If the mode of action of a drug is unknown, or the exact target of the drug is unknown, ensemble modeling could predict which perturbations would result in the
experimentally observed phenotype. Furthermore, if the target of the drug is known the efficacy of the drug can be determined by comparing the steady state fluxes observed after perturbing the enzyme target computationally to the steady state fluxes observed experimentally. This could then give insight into possible side effects by showing what other pathways are affected.
Chapter 6
Conclusion and Recommendations

Metabolic profiling provides information that complements signal transduction and genetic studies, and provides information on whether the adjustments on carbon flow that occur are actually taking place (Boros et al., 2002). In this study, we used experimentally measured metabolomic data obtained from a colon cancer cell line to construct a kinetic model of cancer metabolism using an ensemble modeling methodology. The ensemble of 1000 models was screened by repressing known cancer drug targets. After four screening steps the ensemble converged to a set of models that become more predictive. We used the models to identify enzyme targets that would reduce the growth rate of cancer cells. Two enzymatic targets, transaldolase and succinate-CoA ligase, computationally showed a greater potential decrease in growth rate compared to currently available enzymatic drug targets. Furthermore, we studied the effect of simultaneous targeting of the enzymes identified and found a 3 fold increase in effectiveness when transaldolase and glycine hydroxymethyltransferase were synergistically repressed.

The ensemble modeling methodology has demonstrated to be suitable in this study for studying metabolic perturbations such as repression targets for drug discovery. Further experimental work is necessary to complement the computational predictions made in this study. In vitro gene silencing studies could initially be conducted to study the accuracy of the model predictions. If successful, further studies are required to identify a means of repressing the proposed targets in vivo. As the proposed enzymatic targets are enzymes that are conserved in cancer cells and normal cells, the structure of the enzymes should be studied carefully to identify potential discrepancies between the two enzymes for designing targets specific for cancer cells.

In terms of computational work, further expansion of the metabolic and regulatory network considered would allow for a more precise representation of the metabolic system of interest. This would allow for more predictions that would become increasingly realistic. However, this expansion would highlight a potential limitation of EM
methodology due to the large parameter spaces and relatively long computation times (hours to days) that is involved. This limitation could be overcome by improvements in the EM algorithm and enhanced computing capabilities (Dean et al., 2010).
References


Appendix A: Enzyme Abbreviations

A list of enzymes include in the model along with the abbreviations are presented below.

<table>
<thead>
<tr>
<th>Enzyme Name</th>
<th>Enzyme Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hexokinase</td>
<td>HEX1</td>
</tr>
<tr>
<td>glucose-6-phosphate isomerase</td>
<td>PGI</td>
</tr>
<tr>
<td>phosphofructokinase</td>
<td>PFK</td>
</tr>
<tr>
<td>D-fructose-bisphosphate aldolase</td>
<td>FBA</td>
</tr>
<tr>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
<td>GAPD</td>
</tr>
<tr>
<td>phosphoglycerate kinase</td>
<td>PGK</td>
</tr>
<tr>
<td>phosphoglycerate mutase</td>
<td>PGM</td>
</tr>
<tr>
<td>enolase</td>
<td>ENO</td>
</tr>
<tr>
<td>pyruvate kinase</td>
<td>PYK</td>
</tr>
<tr>
<td>lactate dehydrogenase</td>
<td>LDH_L</td>
</tr>
<tr>
<td>triose-phosphate isomerase</td>
<td>TPI</td>
</tr>
<tr>
<td>L-alanine transaminase</td>
<td>ALATA_L</td>
</tr>
<tr>
<td>aspartate transaminase</td>
<td>ASPTA</td>
</tr>
<tr>
<td>glutamine synthetase</td>
<td>GLNS</td>
</tr>
<tr>
<td>ATP-Citrate lyase</td>
<td>ACITL</td>
</tr>
<tr>
<td>phosphoglycerate dehydrogenase</td>
<td>PGCD</td>
</tr>
<tr>
<td>phosphoserine transaminase</td>
<td>PSERT</td>
</tr>
<tr>
<td>phosphoserine phosphatase</td>
<td>PSP_L</td>
</tr>
<tr>
<td>glycine hydroxymethyltransferase</td>
<td>GHMT2r</td>
</tr>
<tr>
<td>ribose-5-phosphate isomerase</td>
<td>RPI</td>
</tr>
<tr>
<td>ribulose-5-phosphate-3-epimerase</td>
<td>RPE</td>
</tr>
<tr>
<td>transketolase 1</td>
<td>TKT1</td>
</tr>
<tr>
<td>transketolase 2</td>
<td>TKT2</td>
</tr>
<tr>
<td>transaldolase</td>
<td>TALA</td>
</tr>
<tr>
<td>malate dehydrogenase</td>
<td>MDH</td>
</tr>
<tr>
<td>pyruvate dehydrogenase</td>
<td>PDHm1</td>
</tr>
<tr>
<td>citrate synthase</td>
<td>CSm</td>
</tr>
<tr>
<td>aconitase hydratase</td>
<td>ACONTAm</td>
</tr>
<tr>
<td>isocitrate dehydrogenase</td>
<td>ICDHxm</td>
</tr>
<tr>
<td>2-oxoglutarate dehydrogenase</td>
<td>AKGDm</td>
</tr>
<tr>
<td>succinate-CoA ligase</td>
<td>SUCOAS1m</td>
</tr>
<tr>
<td>succinate dehydrogenase</td>
<td>SUCD1m</td>
</tr>
<tr>
<td>fumarase</td>
<td>FUMm</td>
</tr>
<tr>
<td>malate dehydrogenase, mitochondrial</td>
<td>MDHm</td>
</tr>
<tr>
<td>glutamate dehydrogenase</td>
<td>GLUDxm</td>
</tr>
<tr>
<td>glutaminase</td>
<td>GLUNm</td>
</tr>
<tr>
<td>pyruvate carboxylase</td>
<td>PCm</td>
</tr>
</tbody>
</table>
Appendix B: Extracellular Metabolite Concentrations

The extracellular metabolite concentrations measured by CHENOMX are presented in the table below. Concentrations are reported in units of μM and the time reported is in units of hours.

<table>
<thead>
<tr>
<th>Time</th>
<th>Acetate</th>
<th>Acetone</th>
<th>Alanine</th>
<th>Arginine</th>
<th>Betaine</th>
<th>Choline</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>39.1</td>
<td>7.9</td>
<td>248.85</td>
<td>645.4</td>
<td>19.05</td>
<td>8.2</td>
</tr>
<tr>
<td>72</td>
<td>65.74</td>
<td>12.38</td>
<td>730.56</td>
<td>506.14</td>
<td>22.58</td>
<td>6</td>
</tr>
<tr>
<td>96</td>
<td>83.65</td>
<td>6.4</td>
<td>729</td>
<td>437.7</td>
<td>24.65</td>
<td>2.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Creatine</th>
<th>Creatinine</th>
<th>Ethanol</th>
<th>Formate</th>
<th>Fructose</th>
<th>Fumarate</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>33.55</td>
<td>34.95</td>
<td>658.95</td>
<td>40.45</td>
<td>546.45</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>35.84</td>
<td>41.18</td>
<td>731.16</td>
<td>95.24</td>
<td>629.96</td>
<td>3.82</td>
</tr>
<tr>
<td>96</td>
<td>36.35</td>
<td>38.75</td>
<td>757.2</td>
<td>247.95</td>
<td>590.2</td>
<td>6.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Glutamine</th>
<th>Glycerol</th>
<th>Histidine</th>
<th>Isoleucine</th>
<th>Isopropanol</th>
<th>Lactate</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>1693.6</td>
<td>104.35</td>
<td>221.65</td>
<td>501.1</td>
<td>14.45</td>
<td>2192.9</td>
</tr>
<tr>
<td>72</td>
<td>384.16</td>
<td>59.24</td>
<td>153.12</td>
<td>367.26</td>
<td>12.68</td>
<td>4371.12</td>
</tr>
<tr>
<td>96</td>
<td>0</td>
<td>54.2</td>
<td>153.2</td>
<td>267.9</td>
<td>13.05</td>
<td>6723.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Leucine</th>
<th>Lysine</th>
<th>Methionine</th>
<th>Phosphocholine</th>
<th>Phenylalanine</th>
<th>Pyroglutamate</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>391.9</td>
<td>430.7</td>
<td>122.75</td>
<td>23.85</td>
<td>242.35</td>
<td>235.55</td>
</tr>
<tr>
<td>72</td>
<td>292.12</td>
<td>377.06</td>
<td>93.8</td>
<td>10.2</td>
<td>207.02</td>
<td>491.1</td>
</tr>
<tr>
<td>96</td>
<td>188.6</td>
<td>344.95</td>
<td>71.25</td>
<td>5.45</td>
<td>185.15</td>
<td>530.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Succinate</th>
<th>Threonine</th>
<th>Tryptophan</th>
<th>Tyrosine</th>
<th>Valine</th>
<th>myo-Inositol</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>91.05</td>
<td>501.1</td>
<td>43.5</td>
<td>284.4</td>
<td>535.8</td>
<td>196.65</td>
</tr>
<tr>
<td>72</td>
<td>105.66</td>
<td>346.2</td>
<td>34.38</td>
<td>243.94</td>
<td>415.34</td>
<td>190.22</td>
</tr>
<tr>
<td>96</td>
<td>111.55</td>
<td>398.15</td>
<td>33.7</td>
<td>233.2</td>
<td>343.5</td>
<td>183.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Glucose</th>
<th>Leucine</th>
<th>Pyruvate</th>
<th>Citrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>6633.05</td>
<td>391.9</td>
<td>104.9</td>
<td>47.25</td>
</tr>
<tr>
<td>72</td>
<td>5048.1</td>
<td>292.12</td>
<td>330.22</td>
<td>78.18</td>
</tr>
<tr>
<td>96</td>
<td>2880.8</td>
<td>188.6</td>
<td>428.55</td>
<td>105.1</td>
</tr>
</tbody>
</table>
Appendix C: Model Parameters

In the ensemble modeling approach the reactions of the system are broken down into elementary reactions, whereby each of the reactions takes the following form:

\[ v_{i,n} = \bar{K}_{i,n}^{ref} \bar{X}_i \bar{e}_{i,n} \]  

(27)

where \( v_{i,n} \) is the reaction rate, \( \bar{K}_{i,n}^{ref} \) is the normalized kinetic parameter, \( \bar{X}_i \) is the normalized concentration, and \( \bar{e}_{i,n} \) is the normalized enzyme fraction for the elementary step \( n \) in the overall reaction \( i \). The overall lumped kinetic parameter for reaction \( i \), can be calculated from the lumped kinetic parameters of the elementary reactions as follows (Tran et al., 2008):

\[ \bar{K}_i^{ref} = \frac{\prod_{n} \bar{K}_{i,n}^{ref}}{\prod_{n} \bar{K}_{i,n}^{ref}} \]  

(28)

The overall lumped kinetic parameters for the 4 models are presented in the table below:
<table>
<thead>
<tr>
<th>Reaction</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEX1</td>
<td>64.62</td>
<td>8.46</td>
<td>1.78</td>
<td>1.37</td>
</tr>
<tr>
<td>PGI</td>
<td>0.026</td>
<td>0.00027</td>
<td>0.00037</td>
<td>0.17</td>
</tr>
<tr>
<td>PFK</td>
<td>5.17</td>
<td>0.015</td>
<td>0.12</td>
<td>7.79</td>
</tr>
<tr>
<td>FBA</td>
<td>372.31</td>
<td>1755.04</td>
<td>52008.76</td>
<td>86.26</td>
</tr>
<tr>
<td>GAPD</td>
<td>43.87</td>
<td>651.35</td>
<td>1318.54</td>
<td>47.75</td>
</tr>
<tr>
<td>PFK</td>
<td>262.59</td>
<td>1927.42</td>
<td>8563.52</td>
<td>0.22</td>
</tr>
<tr>
<td>PGM</td>
<td>0.26</td>
<td>0.24</td>
<td>0.011</td>
<td>110.31</td>
</tr>
<tr>
<td>ENO</td>
<td>0.00071</td>
<td>0.00023</td>
<td>0.000003</td>
<td>0.0021</td>
</tr>
<tr>
<td>PYK</td>
<td>246.90</td>
<td>1446.42</td>
<td>3.78</td>
<td>15.76</td>
</tr>
<tr>
<td>LDH_L</td>
<td>0.14</td>
<td>13.71</td>
<td>0.47</td>
<td>0.13</td>
</tr>
<tr>
<td>TPI</td>
<td>239.11</td>
<td>638.85</td>
<td>4492.39</td>
<td>5.40</td>
</tr>
<tr>
<td>ALATA_L</td>
<td>1.20</td>
<td>0.46</td>
<td>0.00025</td>
<td>0.32</td>
</tr>
<tr>
<td>ASPTA</td>
<td>0.10</td>
<td>0.12</td>
<td>0.04</td>
<td>0.15</td>
</tr>
<tr>
<td>GLNS</td>
<td>1.33</td>
<td>4.98</td>
<td>12.48</td>
<td>1.90</td>
</tr>
<tr>
<td>ACITL</td>
<td>0.00</td>
<td>0.05</td>
<td>3.46</td>
<td>0.02</td>
</tr>
<tr>
<td>PGCD</td>
<td>5.68</td>
<td>263.55</td>
<td>86.29</td>
<td>0.10</td>
</tr>
<tr>
<td>PSERT</td>
<td>943.86</td>
<td>2334.09</td>
<td>2817.30</td>
<td>20.00</td>
</tr>
<tr>
<td>PSP_L</td>
<td>0.00023</td>
<td>0.000062</td>
<td>0.000038</td>
<td>0.0009</td>
</tr>
<tr>
<td>GHMT2r</td>
<td>0.004</td>
<td>0.00052</td>
<td>0.0019</td>
<td>0.0038</td>
</tr>
<tr>
<td>RPI</td>
<td>16.04</td>
<td>0.03</td>
<td>105.87</td>
<td>0.97</td>
</tr>
<tr>
<td>RPE</td>
<td>0.93</td>
<td>16.20</td>
<td>0.02</td>
<td>1.53</td>
</tr>
<tr>
<td>TKT1</td>
<td>1212.44</td>
<td>488.65</td>
<td>82821.49</td>
<td>2.39</td>
</tr>
<tr>
<td>TKT2</td>
<td>0.00001</td>
<td>0.0013</td>
<td>0.00003</td>
<td>0.22</td>
</tr>
<tr>
<td>TALA</td>
<td>0.000028</td>
<td>0.0019</td>
<td>3.77</td>
<td>0.0069</td>
</tr>
<tr>
<td>MDH</td>
<td>29.74</td>
<td>45.57</td>
<td>20.75</td>
<td>6.02</td>
</tr>
<tr>
<td>PDHm</td>
<td>0.15</td>
<td>16.23</td>
<td>20.44</td>
<td>0.033</td>
</tr>
<tr>
<td>CSm</td>
<td>790.31</td>
<td>5541.58</td>
<td>248493.24</td>
<td>208.99</td>
</tr>
<tr>
<td>ACONTAm</td>
<td>1484.98</td>
<td>3102.02</td>
<td>1.15</td>
<td>250.75</td>
</tr>
<tr>
<td>ICDHxm</td>
<td>0.09</td>
<td>0.03</td>
<td>0.29</td>
<td>1.84</td>
</tr>
<tr>
<td>AKGDm</td>
<td>0.15</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>SUCOAS1m</td>
<td>50092.87</td>
<td>2513.30</td>
<td>12205401.11</td>
<td>387.62</td>
</tr>
<tr>
<td>SUCD1m</td>
<td>0.0041</td>
<td>1.37</td>
<td>0.00028</td>
<td>3.89</td>
</tr>
<tr>
<td>FUMm</td>
<td>0.00049</td>
<td>0.00061</td>
<td>1.80E-05</td>
<td>0.0034</td>
</tr>
<tr>
<td>MDHm</td>
<td>0.04</td>
<td>0.04</td>
<td>0.053</td>
<td>0.28</td>
</tr>
<tr>
<td>GLUDxm</td>
<td>0.11</td>
<td>0.01</td>
<td>0.069</td>
<td>0.16</td>
</tr>
<tr>
<td>GLUNm</td>
<td>0.33</td>
<td>0.55</td>
<td>2.42</td>
<td>3.20</td>
</tr>
<tr>
<td>PCm</td>
<td>0.35</td>
<td>1.91</td>
<td>1.44</td>
<td>0.48</td>
</tr>
</tbody>
</table>