Data assisted-modeling of the liver organ on a chip dynamics

by

Connor Smith

B.S., North Carolina State University, 2017

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Tim Taylor Department of Chemical Engineering
Carl R. Ice College of Engineering

Kansas State University
Manhattan, Kansas

2022

Approved by:

Davood B. Pourkargar
Assistant Professor
Tim Taylor Dept. of Chemical Engineering

Copyright

Abstract

The concept of mathematical modeling to develop a digital twin of biological processes is in a nascent state. Despite the limited understanding of many biological pathways, a well-trained mathematical model can successfully conduct many experiments to guide development more efficiently. The Organ On a Chip (OOC) concept is a feat of engineering and biology where organs can be simulated in vitro for drug discovery, toxicology, and pharmacokinetic/pharmacodynamic (PK/PD) studies. My Master's Report aims to conduct a literature review on OOCs and the kinetics of the biological pathways in the liver related to Non-alcoholic Fatty Liver Disease (NAFLD). The findings will be applied to develop a data-assisted mathematical modeling framework. A liver OOC for NAFLD is used as the base model where the kinetics of the reactions of metabolites, substrates, and drugs are simulated. A hybrid modeling approach combining first-principle-based models, PK/PD, and machine learning will enable a better understanding of factors impacting NAFLD.

Table of Contents

List of Figures

List of Tables

Introduction

Studying the liver *in vitro* remains a significant challenge in the biomedical industry for understanding liver cytotoxicity, disease progression, drug metabolism kinetics and drug discovery. The heterogeneous composition of the liver plays a crucial part in metabolite regulation and detoxification of endogenous and exogenous substances. To characterize these biological pathways more sophisticated tools are required for successful scale-up to human physiology. However, due to the complex nature of the human liver, mammal liver models are rarely clinically relevant and present a source of error when evaluating the same conditions for a human patient. Microphysiological systems such as perfusion bioreactors, 2D planar primary rat hepatocytes, 3D printed liver tissue, liver organoids, and liver-on-a-chip devices make up the currently used in vitro technologies.

Organ-on-a-chip devices enable a physiological microenvironment via microfluidic technology to allow for multiple cell-cell interactions and intravascular networks that can be monitored by biosensors and other real-time monitoring tools. These miniaturized human organ models combine advanced microfluidics and tissue engineering strategies to overcome the limitations of static cell cultures and dissimilar mammal models to accurately predict human body responses. Typical organ-on-a-chip devices employ oxygen, carbon dioxide, temperature, pH, and substrate supplement control with integrated bio-sensors to recreate and study various environmental conditions. Online biosensors can measure dissolved gas, cellular byproducts (glucose, lactate, pyruvate), electrolytes, protein biomarkers, and cell morphology. Additionally, offline measurements can be performed using analytical assays that are difficult to integrate online such as cytokine and albumin ELISAs, cell counters, fluorescence microscopy, and HPLC. Despite the abundance of analytical resources available for evaluating these

microphysiological environments, organ on a chip (OOC) devices are intensive experiments requiring cell culture optimization, microfluidic device fabrication and analytical assay validation. Integrating advanced mathematical modeling for tissue culture systems is an upcoming field that may provide a higher throughput alternative.

The application of digital twins in tissue culturing is in a nascent state where other areas of the biotech field have successfully implemented the concept such as in pharmacodynamic/pharmacokinetic studies of drugs and large-scale bioprocess modelling. Digital twins vastly expand the experimental design space at only the cost of computational resources. However due to the sophisticated and unknown nature of most biological processes in human organs, developing an accurate digital twin has proven to be quite challenging. Machine learning algorithms can effectively combine the large experimental data sets with comprehensive mathematical models as an alternative to relying solely on adequate mathematical formulas or in vitro tissue cultures that are complex to design, setup, and run. Physics informed machine learning models are a multidisciplinary approach enabling a much more versatile, cost and time-effective strategy for biomedical practices. Tissue culture systems and publicly available patient data provide the necessary means to train and validate ML models for accurate predictions of different experimental conditions. With a sensitive, specific, and accurate model, many environmental conditions can be simulated on an OOC and scaled to the human body to predict potential disease pathways, adverse drug effects or aid in drug discovery.

A human liver-on-a-chip modeling Non-Alcoholic Fatty Liver Disease (NAFLD) developed by Lasli et al. was chosen as the basis for developing a computational model to be further optimized by ML techniques. NAFLD is the most common type of liver disease, affecting 1 in 4 adults [1]. Despite the global prevalence, there is no medical treatment commercially

available. NAFLD begins with steatosis of the liver and progresses slowly to cirrhosis or hepatocellular carcinoma. Free fatty acids released in excess from the adipose tissue can exert hepatoxicity resulting in cell damage. Continuous saturation of excess free fatty acids and other toxic substances disturbs the balance between tissue damage and repair. Ultimately this disturbance shifts in favor of damage. The macroscale heterogeneity of the liver, in combination with the biological diversity of humans and their different lifestyle choices, presents many challenges in deciphering the mechanisms of NAFLD. Given the significant impact that this has on the global population and the abundant data accumulated from patients, there is great potential for developing a predictive model for the onset, progression, and treatment of this disease. Through this approach, a broad set of parameters can be evaluated *in silico*, applied to the *in vitro* organ-on-a-chip device and scaled appropriately to the human body for clinical trial studies. This report prepared a literature review on organ-on-a-chip devices, liver biological pathways, disease modeling, and machine learning. These topics provided the necessary information for developing a data-assisted digital twin of a liver-on-a-chip device. In addition to the review, results from a mathematical model of the nonalcoholic fatty liver disease (NAFLD) biological pathways and drug metabolism kinetics are shown. This model provides the basis for implementing machine learning for further parameter refinement and determining liver kinetic trends.

Review of the Literature

Liver and Nonalcoholic Fatty Liver Disease Biology

The liver is a complex organ that acts as the filter of the entire body and is responsible for regulating most of the chemical levels in the blood. When blood leaves the stomach and intestines, it passes through the liver prior to returning to the other organs. About 13% of the body's blood at any given time is held in liver for it to perform more than 500 vital functions [2]. The human adult liver can weigh from 1300 – 1700g, roughly 2% of the total body weight, which is relatively small compared to rodents whose liver's weight is 5% of the total body weight [3]. This bidirectional biofiltration capability is the only mechanism of the blood that can remove non-water-soluble substances such as nutritional compounds and exogeneous and endogenous materials [3]. The liver is constantly exposed to harsh biological and chemical conditions and requires assistance from the immune system to maintain its performance. Some of the most well-characterized functions of the liver include the production of cholesterol, conversion of excess glucose into glycogen, conversion of ammonia to urea, production of proteins for blood plasma and the metabolism of drugs [2].

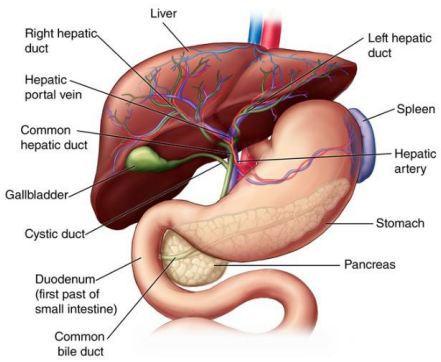


Figure 1: Liver Anatomy [2]

This multifunctioning organ is made up of 2 unique cell types – hepatocytes and biliary cells. Hepatocytes are a highly heterogeneous cell population whose functions depend on their morphology and location in the parenchyma. The biliary cells form the channels for transporting toxic waste out of the parenchyma. While the liver is only made up of two very heterogenous cell types, several communal and high specialized cells reside in the organ, such as endothelial cells, hepatica myofibroblasts, and Kupffer cells. The makeup of the extracellular matrix is unique to the liver allowing it to perform highly diverse metabolic reactions [3]. These different cell types require an organized structure to carry out their essential functions. The portal vein and the hepatic artery supply blood to the liver at a reduced oxygen content and pressure but are enriched in nutrients, toxic materials, hormones, and growth factors. The liver acts like a sponge-like mass connected by tunnels containing an intertwined network of blood vessels. Due to the large surface area and porosity of the liver, it allows for high flow rates and capacities of blood to be processed at once. Inside the liver, the sinusoids form a vascular system that allows for blood perfusion across vessels. The disruption of the sinusoidal membrane prevents efficient communication between the blood and hepatocytes leading to liver dysfunction.

Nonalcoholic Fatty Liver Disease (NAFLD) is a metabolic syndrome of the liver that has been increasing in cases worldwide. The onset of this disease occurs because of excessive accumulation of free fatty acids (FFA) in the hepatic tissue with a wide spectrum of severities. Simple steatosis, nonalcoholic steatohepatitis (NASH), cirrhosis, and hepatocellular carcinoma are the various stages of NAFLD. Despite the rapid increase in the prevalence of NAFLD, pathogenesis is not very well understood. It is known to be associated with type 2 diabetes and coronary artery disease, and there is more recent evidence linking it to multiple cancers [4]. There are no commercially available treatments, and many ongoing clinical trials are attempting

to demonstrate a safe, efficacious and effective drug for NAFLD. The current recommendation from the NIH for treatment is to improve diet and exercise. Proper diet and exercise can reduce the liver's fat, inflammation, and fibrosis.

The onset of hepatic steatosis initiates when the rate of consumption and synthesis of fatty acids (FA) exceeds the rate of catabolism. Four possible mechanisms of fatty acid accumulation within the liver are from increased uptake of FA by hepatocytes, increased *de novo* synthesis of FA and triglycerides (TG), and reduced very low-density lipoprotein (VLDL) synthesis and TGs export, and impaired hepatic mitochondrial β-oxidation [1]. After steatosis has been induced, the liver is much more sensitive to additional stresses resulting in further inflammation and fibrosis. NAFLD and NASH can be differentiated by the presence of increased reactive oxygen species and mitochondrial uncoupling in NASH.

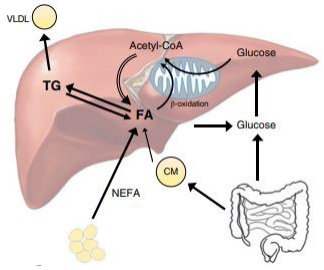


Figure 2: Mechanisms of free fatty acid accumulation [2]

Two leading theories in the mechanism of steatohepatitis are the two-hit and the multi-hit hypotheses. As previously mentioned, the liver can be sensitized by the initial accumulation of triglycerides in the hepatocytes via four different mechanisms. The proposed two-hit mechanism suggests a that second hit from drugs, gut-derived endotoxins, iron overload or oxidative stress would progress the disease. In comparison, the multi-hit hypothesis suggests that continuous lipotoxicity from FFAs by any combination of the degradative pathways drives the injury [4].

These two pathways have been studied in various of *in vitro* and *in silico* models where neither has been proven to be more correct than the other [1, 5, 6, 7, 8, 9, 10]. A clinically relevant NAFLD model should ideally satisfy the biological fidelity and quality by design requirements necessary for Investigation New Drug (IND) and Biologics License Application (BLA) filings. The biological components, chemical reactions, and transport phenomena should be scalable to the human body to be clinically relevant for drug screening. To validate such models, *in vivo* human pathogenic responses are necessary to demonstrate that the drug is non-toxic and potent. Due to the global prevalence of NAFLD, well-established *in vitro* models, and large data generated from clinical trials, a hybrid digital twin model could be successfully implemented. Established biochemical mechanisms and fluid dynamics of an organ system can be optimized further by statistical regression methods.

Liver Organ-On-a-Chip Devices

As previously described, the liver is a highly functionalized organ with many different cell types. Designing in vitro models to accurately describe its many functions is difficult due to its complex biological pathways, reaction kinetics and mass transfer phenomena occurring in the organ simultaneously. While tissue culture systems enable biological functions to be described, the lack of control of the thermodynamics and mass transfer of the small molecules limits these systems from being scaled to the human body. The OOC devices have been shown to be superior organ *in vitro* models in comparison to traditional tissue culture techniques. In pharmaceutical development there is less than a 10% success rate for a drug to be approved by the FDA, with the leading cause of failure to be safety-related [11]. Approximately half of the drugs that resulted in liver injury during clinical trials did not display liver toxicity in *in vivo* mammal models, further

demonstrating the poor translation to the human body from animals [12]. These statistics prove that more effective drug and disease development models are necessary for early-stage research. To bridge the human body, successful scale down of the organs impacted must demonstrate comparable microenvironments which can be monitored in real-time.

Liver on a chip devices of various cell types, process controls schemes and microfluidic designs have been engineered to study drug screening, toxicity testing, metabolism predictions, liver disease models, and integration in multi-organ devices [12]. Better management of the liver microenvironments and more accurate mimicking of the liver's physiology and pathology can be achievable with the incorporation of microfluidics. The development of these devices is focused on cell type selection, culture geometry, sensor selection, and module design. Cell type is a major consideration for accurately simulating the pathological processes of the functions being investigated. Most of the liver on a chip devices utilize some mixture of HepG2, Kupffer cells, HSCs, HUVECs, or LSECs [10, 13, 14, 15]. The cell culture geometry is of equal importance as the strategy of how the cells are suspended or adhered, which in turn impacts their viability, morphology, and productivity. Liver on a chip devices are superior to other cell culture techniques because of the dynamic flow environment, increasing hepatocytes' albumin synthesis and urea excretion [12]. In the field of liver on a chip fabrication, many things are considered to ensure the reproducibility of the devices such as the length of the culture period, cell-cell communication, substrate feed, contamination control, high throughput screening capabilities and economics. Strategies that are relatively easy and are best for short-term cell culturing include 2D planar and hanging drop cultures. Long-term liver on a chip cell cultures that are used for high throughput studies are microarray systems and 3D bioprinted systems. Other long-

term methods that allow for efficient cell communication are layer deposited, 3D spheroid, and matrix-dependent 3D cultures.

Although the previously mentioned strategies encompass a wide variety of biomedical technologies, there is not one that perfectly suits all OOC applications. Bhise et al. constructed a liver on a chip platform with 3D human HepG2/C3A spheroids encapsulated in a gelatin methacryloyl hydrogel [13]. This proof of concept device correctly predicted drug toxicity from 15mM acetaminophen described in other *in vitro* models. In another study, Banaeiyan et al. designed a device to recreate a large-scale liver-lobule that could be readily customized with different volumes and cell densities to understand the anatomy of the liver and its interactions with other organs [14]. This device contained an integrated network of hexagonal lobule chambers with a central outlet to mimic the nutrient mass transport through the central vein. This 3D layered device allowed for real-time monitoring of cell viability, morphology, and functionality making it a powerful tool for high throughput dose response testing, co-culturing different liver cell types and integrating them into multi-organ platforms.

This Master's Report focuses on the liver on a chip platforms described by Lasli et al. and Suurmond et al. from the Department of Bioengineering, UCLA [8, 9]. These two papers describe a complex liver environment that studies the onset and progression of NAFLD and NASH through online and offline measurements. Hepatocyte steatosis was induced by supplementing the free fatty acids, palmitic and oleic acids, in the media. Experimental parameters were monitored by fluorescent and confocal microscopy, a blood gas analyzer, a lipid accumulation assay, reactive oxygen species production assay, and human albumin ELISA. For both platforms, 3D spheroids were formed in microwells with best seeding density of 200 cells per spheroid. Spheroid size is a critical optimizing parameter to enable high productivity without

forming a necrotic core. Circularity is also an essential characteristic due to the linear gradient of oxygen tension, which is crucial for liver sinusoid homeostasis [8]. Optimal liver cell spheroids have a diameter of 200-300um and circularity close to one.

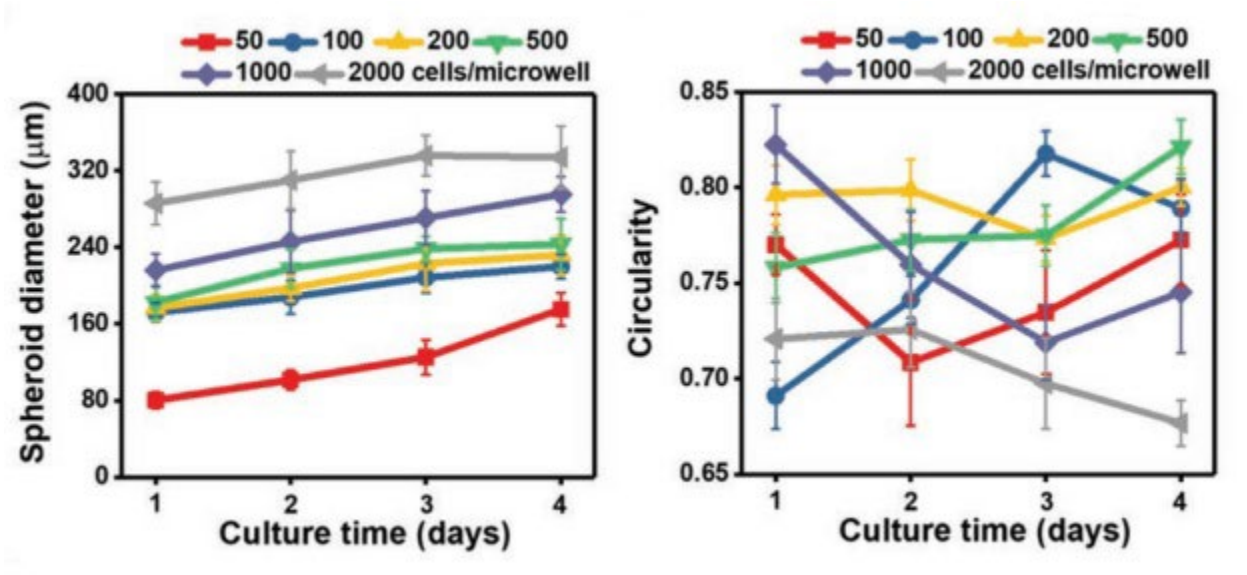


Figure 3: Spheroid Development of cells per microwell [8]

Another important consideration in developing the NAFLD liver on a chip platform from the UCLA group was the cell types, the source, and the ratio of each cell type. As previously mentioned, animal liver models are generally unsuccessful in scaling to the human liver but researchers are unsure if this phenomenon is due to the biological or mass transfer differences. Hepatocellular carcinoma cells (HepG2), umbilical vein endothelial cells (HUVECs) and Kupffer cells (HKCs) were used in different ratios during spheroid formation. The initial study by Lasli et al. demonstrated a ratio of 80% HepG2 and 20% HUVECs to be a suitable composition of the spheroid [8]. Kupffer cells are proposed to communicate with other liver cells via secreted chemicals. Kup5 cells were cultured with the media from healthy and unhealthy liver on a chip conditions to simulate this phenomenon. Then the supernatants from the Kup5 cell cultures were used to re-culture the spheroids with the signaling components to be present or

not. The results indicated that macromolecules in the free fatty acid media activate Kup5 cells to signal the cells downstream to initiate steatosis. Suurmond et al. showed that a ratio of 70% HepG2, 15% HUVECs, and 15% HKCs is more representative of the percentages in vivo [9]. This spheroid makeup allowed for more direct communication between the HKCs, HepG2 and HUVECs cells as compared to the previously described scenarios.

Both systems showed significant differences in cell morphology and lipid accumulation between the control and the steatosis-inducing diet, signifying that the accumulation of FFAs contributes to disease onset. The consistent levels of albumin and urea secretion demonstrated that whether the spheroids included HKCs or were supplemented with the fatty acids, they remained functional. Lasli et al. used the liver on a chip platform to screen two antisteatotic drugs, namely metformin and pioglitazone. Each drug was administered independently to study its impact on lipid accumulation. Both drugs reversed steatosis after 2 days and agreed with previous studies [8].

In this report, the mathematical models of the kinetic pathways of urea, glucose, metformin, and fatty acids demonstrated similar functionality and health of a modeled liver on a chip system. Additionally, the impact of diet and lifestyle are incorporated to demonstrate similar observations as those described by NAFLD researchers.

Mathematical Modeling of Liver and Drug Metabolism Kinetics

In vitro models are labor and resource intensive, restricting a researcher from studying within a small design space. However, a computational model can be used to explore experimental design spaces that are unachievable in a laboratory by executing a small number of experiments for parameter estimation and knowing the mathematical equations that govern the

mechanisms. The concept of applying laboratory insights to a mathematical model has been coined as the digital twin concept or mechanistic modeling [15], Digital twins have made exciting contributions to chemical engineering research and are beginning to gain popularity for biomedical applications. Much of the biological phenomena in nature and disease are still fundamentally challenging to study, but advanced scientific techniques and improved computational power have narrowed many of the gaps that limited this research decades ago. This theoretical tool has been proven to successfully predicting many biological processes *in silico*, such as organ function, disease progression, and metabolic pathways.

One of the earliest applications of digital twins in biomedical sciences was for pharmacokinetic and pharmacodynamic (PK/PD) models. Basic PK/PD models of drug delivery have been used to gain further *in vivo* understanding of complex molecules and disease treatment pathways. While PK modeling quantitatively describes the mechanisms of drug absorption and disposition in the body, PD modeling portrays the time course of drug effects by considering the drug kinetics and rate-limiting step of biological pathways [16]. The pharmacokinetic model is generally compartmentalized into multiple steps to demonstrate the conversion of the initial drug product into intermediates in various organs. The pharmacodynamic model describes the kinetics of intermediate reactions. The combination of these two models enables quantitating the transport of a drug's active ingredient through the different organs and intermediate reactions. These models are crucial for understanding drug dosing, delivery, and organ toxicity.

Lyons et al. showed that a physiologically based PK/PD model was helpful for determining the optimal dose of rifampin in a mouse tuberculosis model [17]. The dose-ranging model allowed for identifying doses that were efficacious and non-toxic while also evaluating anti-microbial resistance and drug-drug interactions. The model parameters were refined using a

Monte Carlo simulation in the software MCSim on data from a previously conducted mouse experiment [17].

A PK model of metformin for treating Type II diabetes provided a baseline for the model presented in this report Chakraborty et al. derived a mass balance-based model for the transient state kinetics of the drug in the stomach, liver, intestine, tissue and kidney [18]. First-order reaction and Michaelis-Menten kinetics were used to describe the consumption and production of metformin intermediates in the different organs. This model adequately describes the metformin active ingredient concentration at different times throughout the body. Michaelis-Menten kinetics provide a mathematical formula for cellular reaction pathways as a function of a substrate concentration. Generally, enzymatic reactions are initiated by a binding interaction between the reaction site and the substrate. Although this equation was originally applied to enzymatic reactions, it has been successfully used to characterize first-order kinetics of compounds in the human body at low concentrations [19]. The generic form of the Michaelis-Menten equation is written as:

$$v_i = \frac{v_{\max}[S]}{K_m + [S]}$$

Figure 4: Michaelis-Menten Equation [19]

With v_i being the reaction rate, v_{max} being the maximum reaction velocity, K_m being the reaction constant, and S being the substrate concentration. This mathematical principle is helpful in simulating various biological reactions in the proposed liver model.

The digital twin concept in biomedicine is achievable due to the improved understanding of biological pathways and the ability to simultaneously solve multiple complex ordinary differential equations at a small computational expense. As for tissue culturing, researchers have

just begun to explore how digital twins can be incorporated into their process workflow. Digital twins act as a process design and refinement tool by profoundly investigating the thermodynamics, mass transfer, and shear effects of cell culturing. For the case of OOC platforms, *in silico* development of multiple organs would allow for the evaluation of organ-organ interactions without requiring sophisticated microfluidic connections. Additionally, the feed media, which is one of the main difficulties of multi-organ platforms, would be overcome since the standalone platform data would be enough to develop comprehensive *in silico* models. Scaling up an organ device can be achieved more efficiently through a computational tool rather than increasing the volume of the reactor, cell mass, and flow rates. Four proposed scaling methods for all *in vitro* and *in silico* models are direct scaling, residence time-based scaling, allometric scaling, and multifunctional scaling. In direct scaling, the ratio of the organ modeled to the human body have a linear correlation. Residence time scaling requires that all components have a constant interaction time with the organ. The allometric scaling law derives a direct correlation between the mass of organisms and its physiological parameters. The multifunctional scaling method relies on the idea that organs carry out many functions simultaneously, so multivariate optimization is required especially for multi organ models.

Glucose metabolism is an essential biological pathway for most cell lifecycles. Berndt et al. described a multiscale metabolic model of glucose metabolism by singular hepatocytes integrated through blood perfusion and tissue transport. Parameters for this model were taken from data in the literature and the v_{max} values of the reaction kinetics were left to be adjustable. v_{max} values were determined by fitting simulated values to steady state solutions [20]. López-Palau et al. established a mathematical glucose metabolism model to understand the pathophysiology of type II diabetes. A critical improvement to the model from those previously

published was the inclusion of blood glucose and insulin dynamics within the pancreas. In previous models, a single function was used to describe the release rate of pancreatic insulin which would be a poor representation of human physiology [10]. With the updates from López-Palau et al., blood glucose dynamics after glucose consumption can be more properly evaluated. Data from these models provided proper parameter values to be incorporated in the liver on a chip mathematical model. Additionally, they are valuable resources for future work for the compartmentalization of a multi-organ on a chip platform.

NAFLD results from lipid build-up in the liver, where the fundamental understanding of onset and treatment are relatively unknown. Due to the small size of the sinusoid and the vast number of potential variables, computational models allow for the exploration and interpretation of data sets unavailable in traditional *in vitro* models. Ashworth et al. developed a model focused on the optimization of oxygen input, blood flow and establishing a dimensional framework of the liver for metabolite gradients [6]. In this model, eight compartments were used to describe the periportal blood, pericentral blood, and blood in the rest of the body. Schleicher et al. adapted the Ashworth model to include a concentration-dependent zonation of fatty acid accumulation in the pericentral and periportal blood [22]. Four models were implemented to evaluate the zonation pattern of triglycerides with each one focusing on a specific mechanism The first iteration was a simplified single compartment that was useful for parameter calibration and a sensitivity analysis. The model was further increased in complexity to the final one demonstrating fatty acid and oxygen gradients with linear and nonlinear rate kinetics. Holzhütter et al.'s model broke down the liver into the adipose tissue, extrahepatic organs, plasma, and several liver units. The macroscale heterogeneity of metabolic processes could be mathematically evaluated by further compartmentalizing the liver [7]. Many intrinsic and environmental factors contribute to

steatosis, making this approach highly sophisticated and potentially most physiologically relevant. However complex, this model includes a huge variety of molecular and cellular processes, making data interpretation and confirmation very difficult. One clinically relevant outcome from the macro-scale model demonstrated the potential for steatosis to be reversible by diet and lifestyle change. This finding agreed with the *in vitro* liver on a chip platform developed by Lasli et al. All these models provided the basis for liver biological pathology model development using a single compartment [6, 7, 22].

Qualitative assessment and reproducibility checks are necessary for mathematical model validation. The general practice for assessing a model is to compare the data generated by the simulation against an available data set from *in vitro* or *in vivo* studies [23]. From this, a deviation score can be generated to show model appropriateness. In the next section, applications of advanced statistical techniques using artificial intelligence (AI) in biomedicine will be discussed. These techniques allow for the refinement of tissue culture models by computationally improving the appropriateness of trend predictions.

Machine Learning Applications in Biomedical Sciences

In the previous section, many applications of mathematical models for tissue culture systems were described; however, most of these model designs are mechanistic. Mechanistic modeling is beneficial for extensive studying within a design space where the mechanisms of the process are well known. However, the technique is generally unable to extrapolate and predict results outside the boundaries of the data used for regression. Machine learning (ML), a form of AI, has become a prominent technology in the process control and development space since large datasets of different formats can be used to build predictive models for new research focuses

[24]. Machine learning algorithms allow for sophisticated data mining, optimization, and generalization algorithm-building tasks that are unachievable by traditional computational statistics. With the evolution of computers, these once considered lengthy and time-consuming algorithms can be utilized on any modern device. The computational tools employed are different for mechanistic modeling and machine learning. In utilizing the two methods synergistically, a mathematical model can satisfy engineering principles while applying to complex datasets. This kind of hybrid approach can be referred to as an auditable black box where big data and advanced computing algorithms highly refine a simple mechanistic model. There have been no published hybrid mechanistic-ML models for an *in silico* organ on a chip platform, and the mechanistic model developed for this report will serve as the basis for future implementation.

The standard strategy for ML analysis begins with identifying and formatting a dataset by removing any rows with missing data and scoring class variables with numbers. Secondly, the training task is initiated using a generalized linear model as a good starting point or by more complex methods such as support vector machines (SVM) and Artificial Neural Networks (ANN). Once the algorithms from the training data have been generated, they must be tested against a separate data table and assessed for sensitivity, specificity, and accuracy. Receiver operating characteristic (ROC) curves are useful for explaining the probability that the model could accurately describe a random sample. The final part of an ML model lifecycle is continuous improvement and integration with new datasets to ensure that the model can adapt to predict future datasets.

$$\text{Sensitivity} = \text{true positives} / \text{actual positives}$$
$$\text{Specificity} = \text{true negatives} / \text{actual negatives}$$
$$\text{Accuracy} = (\text{true positives} + \text{true negatives}) / \text{total predictions}$$

Figure 5: ML training functions a) Sensitivity, b) Specificity, c) Accuracy [25]

ML algorithms operate as supervised or unsupervised methods and the strategy taken generally depends on the dataset features and whether an outcome is available for training purposes. Unsupervised ML methods are beneficial in dimensionally reducing and analyzing data with many features, but there are a few instances where linear and logistic regression techniques fail [25]. Critical unobservable patterns in single data analysis are developed by integrating many different data types at once. For instance, blood pressure was predicted from electrocardiographic data as well as a regression analysis of multiple factors, which in turn generated accurate birthweight predictions [25]. By integrating many different data types at once, important patterns that are unobservable in single data analysis can be developed. Although the biomedical data are highly dimensional, they are generally sparse and lack consistency for all instances. Data sparsity arises from measurement technology limitations, physical constraints, and investigative biases [26]. Additionally fundamental differences between various *in vitro* models and human biology can cause misleading algorithm generation for disease diagnosis and treatment if the datasets are not representative. To address this concern, the best practice for developing an adequately optimized and scalable model would be to utilize human patient data sets for *in silico* model development.

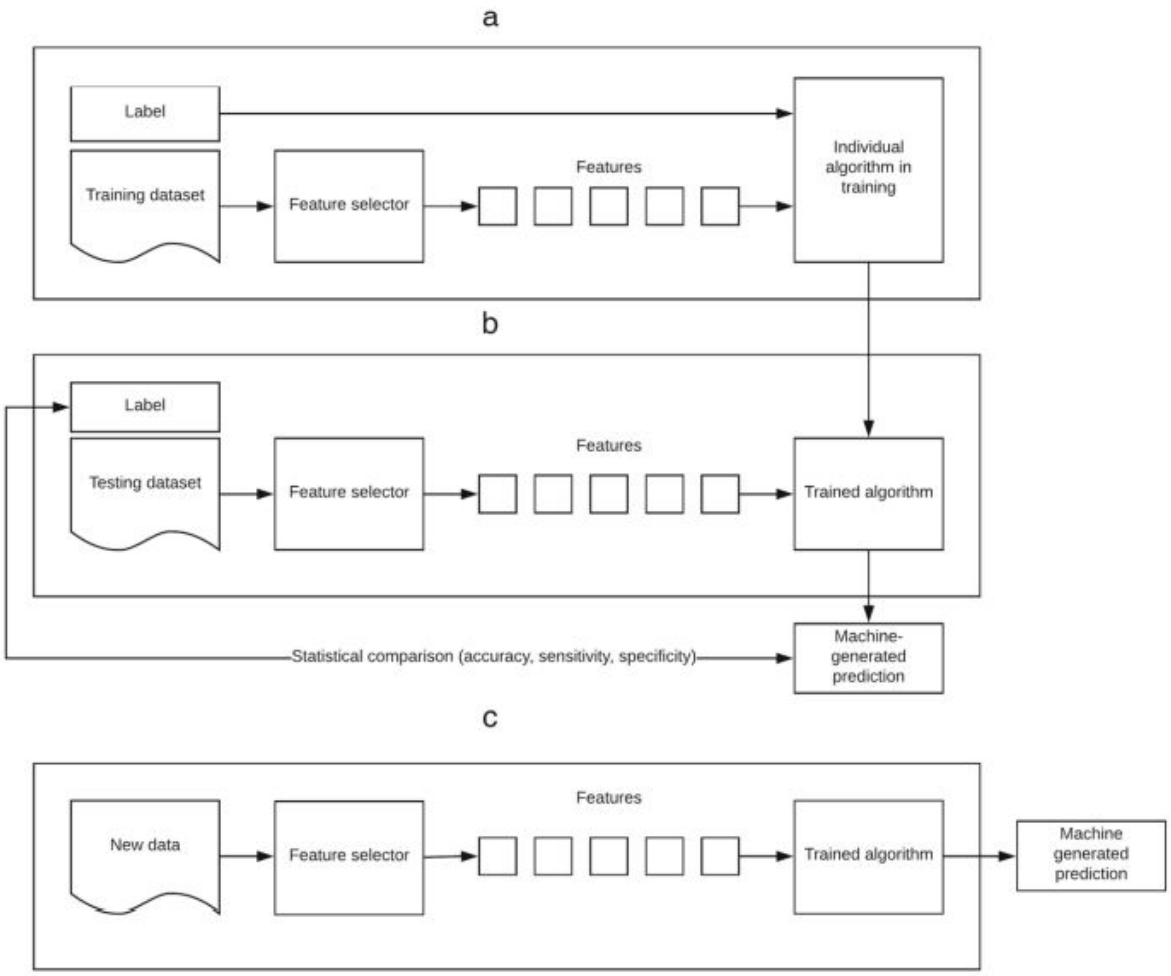


Figure 6: Box Diagram of Supervised ML Model development a) Training b) Validation c) Application [26]

In the case of biological and biomedical sciences where the mechanism of a process is ambiguously defined, machine learning algorithms can generate functions based on new insights. Artificial neural networks have become very popular in medical diagnostics. These computing systems are designed to mimic the biological neural networks of animal brains by learning to perform tasks without a defined task. Neural networks have been shown to be effective for analyzing medical images and contractility profiles [27, 28]. An ML algorithm can build a predictive model based on previous image datasets for disease prognosis and diagnosis by automating the workflow and eliminating human bias. Medical images such as MRIs, cardiac

images, ultrasounds, and functional near-infrared spectroscopies represent a sizeable amount of the big data in medicine [29]. Traditional statistical methods cannot build predictive models to enable doctors to understand patient health trajectories. Time is a essential variable for treatable diseases, so having the capability to identify disease onset and progression could be lifesaving. Machine learning models can increase the success of new drug development by evaluating the vast amount of *in vitro* data, clinical trial data and self-submitted patient data for patterns. The major difficulty arises in the compilation of data because it is common for competitor equipment vendors to use non-standardized data formats. Machine learning algorithms are best trained using data in a format that is the same as the implementation.

Since protected patient data will be the primary source of model training in biomedicine, regulation and oversight are extremely important. The application of machine learning in this field will require legal frameworks and local practices to ensure the safe development, use and monitoring of the systems [30]. Also due to the highly sophisticated algorithms generated by a model, patients may not have ways to verify the claims made so clinician verification is the best practice.

Physics-informed machine learning models of *in vitro* tissue culture systems increase the throughput evaluation of experimental parameters in biomedical sciences and provide more accurate parameterization of models and new functions based on elaborate trends. In traditional physics-based models one major limitation that has continuously improved with the increase in computing power is the accuracy of numerical discretization of partial differential equations. In physics-based models, multiple partial differential equations must be solved numerically, requiring expensive code and sacrificing accuracy. Statistical regression and curve fitting techniques allowed for the inference of relationships within a data set. Machine learning

algorithms integrate big data into these mathematical models, to develop stronger predictions that can be applied outside of the data set. Because of this, operators must limit machine learning algorithms from overstepping into the unnatural through boundary conditions. Through this hybrid approach, physical variables can be identified and fitted to a mathematical system. Physics-informed networks (PINS) incorporate biases by three separate pathways: observational, inductive, and learning biases [24]. Observational biases allow an ML model to learn functions that reflect the physical nature of the database but require a large volume of data for over-parameterized systems. Inductive biases incorporate prior assumptions which implicitly satisfy a set of physical laws; however, they are restrictive in interpreting aberrations in data. Learning biases restrict a system to a broad class of physical constraints expressed in the form of integrals and differential equations. Hybrid approaches to bias implementation have been shown to be successful for low-fidelity models of complex systems.

As previously mentioned, the success of a machine learning model is determined by the relativeness of the data used to train the model. In the USA, the NIH has control of extensive databases of patient demographics, medical history, symptoms, medication use, diet, and exercise habits that is tracked for diseases through clinical trials and by clinicians. One database of particular interest for this research project is the Nonalcoholic Fatty Liver Disease Adult Database provided by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The dataset variables align well with the models described in the methods section.

Additional Applications for OOC Modeling

The liver is a highly functional organ performing many complex tasks that are not fully understood. Building *in silico* models to characterize these biological pathways will be a highly

efficient strategy for gaining new insights into human nature. The earliest explorations of human anatomy required the dissection of human cadavers, and only limited knowledge can be gained from a nonfunctioning organ. Computer and coding improvements have enabled the success of digital twins in many disciplines. Biomedical sciences, compared to chemical sciences, are much less well known and much more complicated given the highly interactive nature of heterogeneous organs in the body. Biological pathways such as the citric acid cycle, cellular respiration, mitosis, and DNA replication have been thoroughly studied but still have high degrees of freedom when modeling the process at a large scale. Developing mathematical models for *in vitro* systems can reduce the uncertainty of defining the biological mechanisms and early-stage drug development at a small scale. Combining the kinetics of biological processes and mimicking the fluid dynamics of the human body, *in silico* organ on chip devices are scalable and high throughput systems. Mathematical modeling can address many specific tasks such as cell kinetics, bioreactor fluid modeling, pharmacology, and image-based modeling [15].

Many other in vitro systems should be explored in addition to the proposed NAFLD liver on a chip model. Drug and toxicity screening testing generally performed on rodents does not provide the most reproducible results when a drug is in clinical trials. Liver diseases such as Type I and II diabetes with a large patient population, would greatly benefit from representative ML-based mathematical models [6]. More than 240 million people are infected by HBV, and liver on a chip models are being successfully implemented for preclinical platform research [31].

The ability for digital twin implementations does not end with the liver, as many research groups are exploring many organ systems and could benefit from *in silico* replication. For example, Lee et al. are exploring a cardiotoxicity platform that will help with the early detection of breast cancer and the prediction of chemotherapy-induced cardiotoxicity [32].

Cardiomyocytes' structure and beating alignment for drug screening have been recapitulated on a heart on a chip device which has been demonstrated to be valid in vitro tool for high throughput drug screening [33]. Combined with the many published PK/PD models and the clinical trial data, a variety of drugs can be screened mathematically with a precise heart model. Multi-organ on the chip platforms are highly sophisticated networks of different cell types, microfluidic devices, and sensor specifications which would be much more economical for experimentation as a digital twin. A liver and intestine co-culture platform was evaluated for systemic repeated dose substance testing. If a base model for the affected organs and biological pathways could be developed, then any combination of experimental conditions could be evaluated at a much higher throughput than all other in vitro models.

The power of biomedical digital twins is nascent, but with the wide availability of patient data and clinical trial data, there is a great potential for physics-based ML modeling. The main challenges of this strategy are database sparsity, ethical restrictions, and inter/intra biological diversity.

Methodology

The mathematical model used for our disease simulations consists of four modules: (1) the free fatty acid module describes the metabolism of fatty acids to phospholipids and triglycerides; (2) the urea pathway module describes the biochemical synthesis of urea in hepatocytes; (3) the glucose module describes the metabolic pathway of glucose to glycogen within the liver; (4) the PD/PK module describes the reaction kinetics of a drugs metabolism and interactions with the body for controlling free fatty acid accumulation. In this model, the patient's health can be manipulated through the feed conditions, the "activity coefficient", and drug concentration, all of which directly influences the metabolic efficiency of the components in the system. NAFLD onset and progression are linked to lifestyle choices (diet and exercise), which is why it is crucial to demonstrate changes in this model based on a human's health. To mathematically weigh patients' health, an activity coefficient of 0 corresponds to someone in poor health, whereas an activity coefficient of 1 corresponds to a very healthy person. Currently, the activity coefficient is a representation of exercise and diet, but in future iterations, it may be worthwhile to split these into separate parameters.

Lasli et al. and Suurmond et al. have published a liver on a chip monitoring several physical, chemical, and biological substances that can be mathematically modeled by PDEs. In the liver on a chip system, lipid accumulation is induced by supplementing the free fatty acids, palmitic acid, and oleic acid. The activation of Kupffer Cells for instigating the defense and repair process are accompanied by elevated levels of ROS (reactive oxygen species) and proinflammatory cytokines (TNF-α and IL-6). To monitor the functionality of the spheroids (HepG2 + HUVECs) with and without Human Kupffer Cells (HKCS) albumin and urea concentrations were monitored. Metformin and Pioglitazone have demonstrated clinical success

in treating excessing lipid accumulation and have been evaluated for safety and efficacy. Glucose metabolism measurements were not provided in the proposed model but are an effective indicator for cellular health and NAFLD progression. In MATLAB the differential equations describing the biological reactions of the first order were simulated using the code as shown in Appendix 1. 22 differential equations were solved simultaneously using the ode15s solver to enable a numerically stable solution.

Module 1: Free fatty acid accumulation in the liver has been proposed as one of the main reasons for steatosis onset. Figure 7 describes the biological pathway and the free fatty acid intermediates that are useful for understanding the systems kinetics.

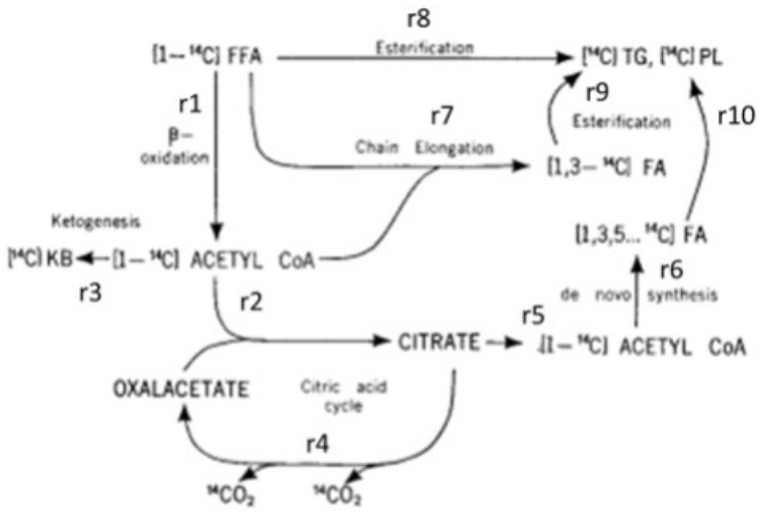


Figure 7: Biological Pathway of Free Fatty Acid Production [34]

Table 1: Reaction Components of the Free Fatty Acid Module

Reaction	Mechanism	Substrate
r1	B-oxidation	FFA
r2	CA Cycle	Oxalacetate, Acetyl CoA
r3	Ketogenesis	Acetyl CoA
r4	CA Cycle	Citrate
r5	CA Cycle	Citrate
r6	de novo synthesis	Acetyl CoA
r7	Chain Elongation	Acetyl CoA, FFA
r8	Esterification	FFA
r9	Esterification	FA_EL
r10	TG Production	FA_DNS

The reaction parameter results of the free fatty acid model is based on 20 milligrams of a harvested rat liver at 117e6 cells/g [34, 35]. The original rate data for the FFA metabolic process was presented in disintegrations per minute (DPM). This radiochemical concentration allows for the quick measurement of carbon-based molecules. However, simple linear regression was implemented to convert to millimolar units. Figure 8 and Table 2 show the 3 data points that we compared from 2 sources Ontko and Ashworth et al. to determine the correlative function for converting DPM to mM/min.

Table 2: DPM to mM/min data points

Constant	Value [DPM]	Source	Value [mM/min]	Source2	Step
Vm1	5.00E-05	ONTKO 1972 [34]	0.20	Ashworth et al. 2016 [6]	B-oxidation
Vm10	6.00E-04	ONTKO 1972 [34]	0.51	Ashworth et al. 2016 [6]	Triglyceride synthesis
Vm9	3.30E-03	ONTKO 1972 [34]	2.10	Ashworth et al. 2016 [6]	Esterification

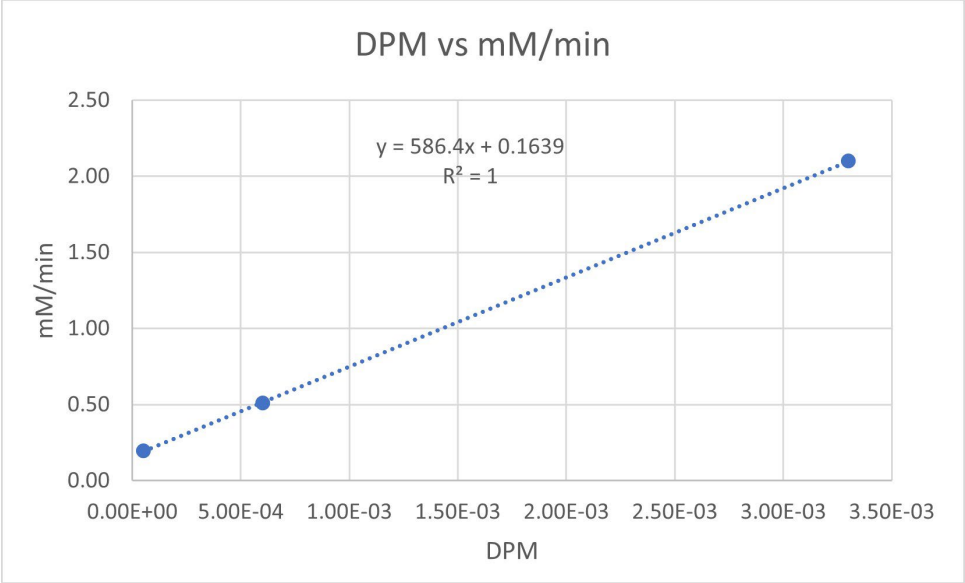


Figure 8: Linear plot for converting DPM to mM/min in the FFA Model

Module 2: The urea biological pathway is an essential indicator of cell health and productivity. Urea is the main product of ammonia metabolism in mammals and is associated with loss of liver function. A cell's inability to detoxify ammonia into urea results in hyperammonemia crises and can ultimately lead to apoptosis. This detoxification is highly dependent on the enzymatic reactions of the urea cycle intermediate molecules; ornithine, citrulline, arginosuccinate, arginine, and urea. Figure 9 shows the simplified reaction mechanism which was selected for the liver on a chip MATLAB model [36]. The kinetic parameter data was based on an entire mouse liver weighing approximately 2g at 135e6 cells/g.

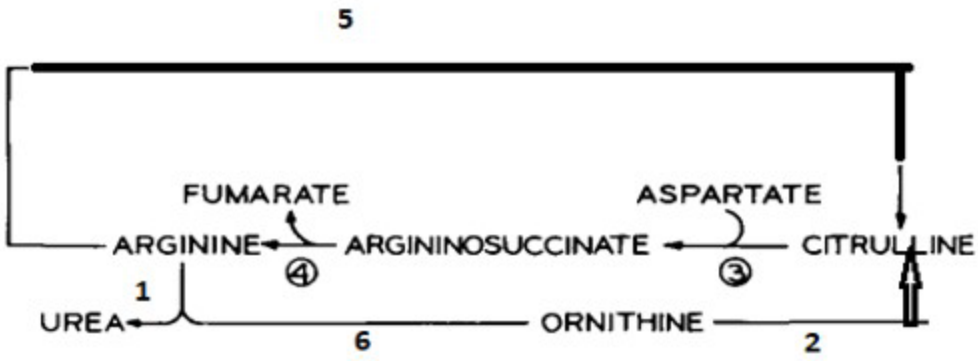


Figure 9: Urea Cycle Reaction Pathway [36]

Table 3: Urea Cycle Components

Reaction	Substrate
r1	Arginine
r2	Ornithine
r3	Citrulline
r4	Arginosuccinate
r5	Arginine
r6	Arginine

Module 3: A mathematical model of glucose in various organs was developed using type II diabetes data [21]. The similarities between NALFD and type 2 diabetes allowed this model adaption to be helpful in incorporating glucose in the NAFLD liver OOC model. However, glucose metabolite intermediates were not incorporated into the reaction scheme.

Module 4: PK/PD of modeling of drug kinetics allows for a mathematical understanding of the toxicity and potency effects on human organs. The PK model of metformin was adapted from Chakraborty et al., and a first-order reaction kinetics were assumed for the PD model. The pathway simulates the reactions of the solid dose drug to its active form in the liver. In addition, a PK/PD model was developed for evaluating the concentration of Pioglitazone intermediate

metabolites and active ingredients in the liver after dosing. These mathematical schemes allowed for a more accurate representation of the medicinal component's concentration in the liver after digestion. However, the metformin model was used in the overall kinetic model code shown in Appendix 1 due to the reduced degrees of freedom in parameterization.

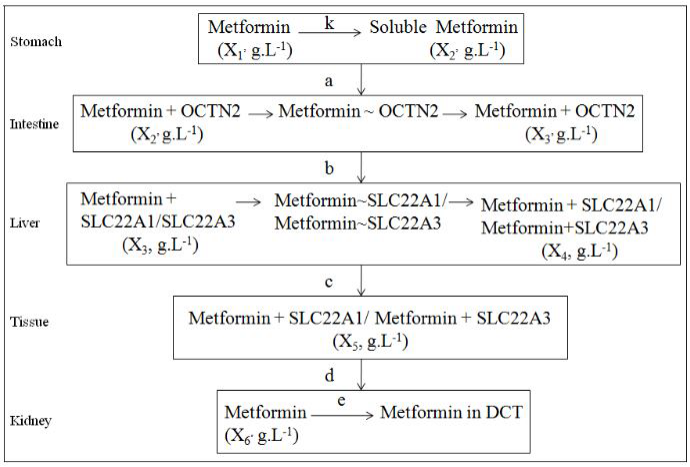


Figure 10: Metformin PK/PD Model [18]

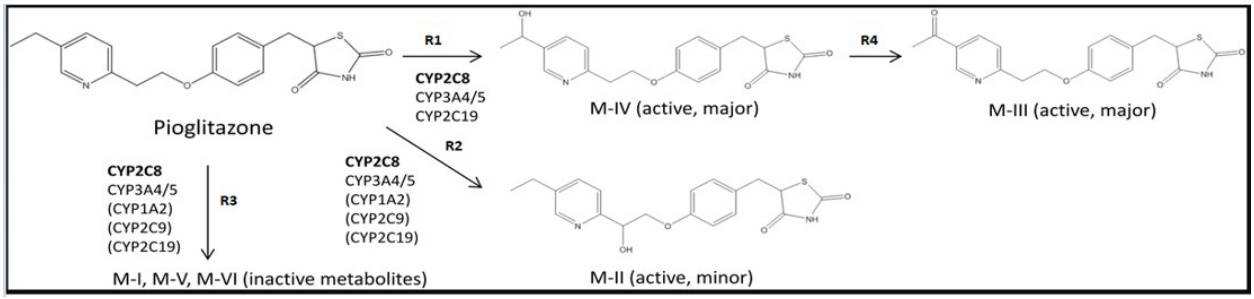


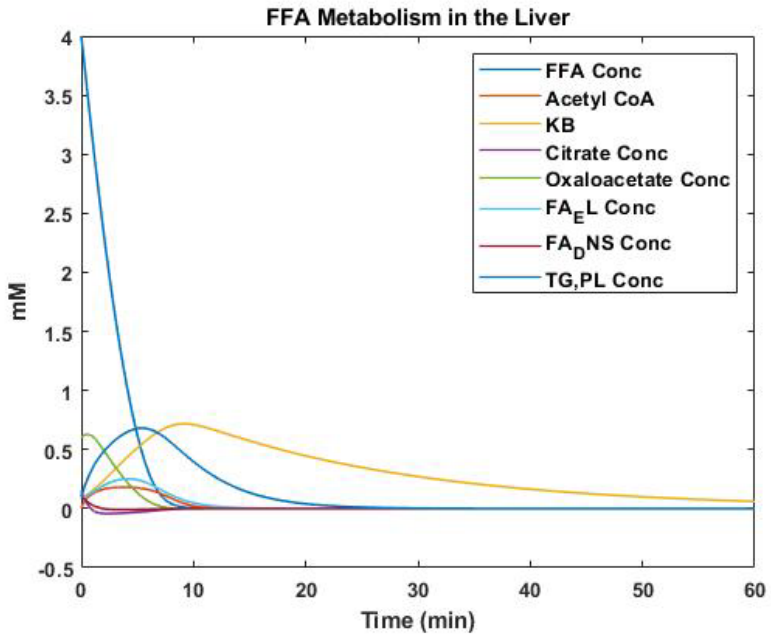
Figure 11: Pioglitazone PK/PD Model [37]

Results

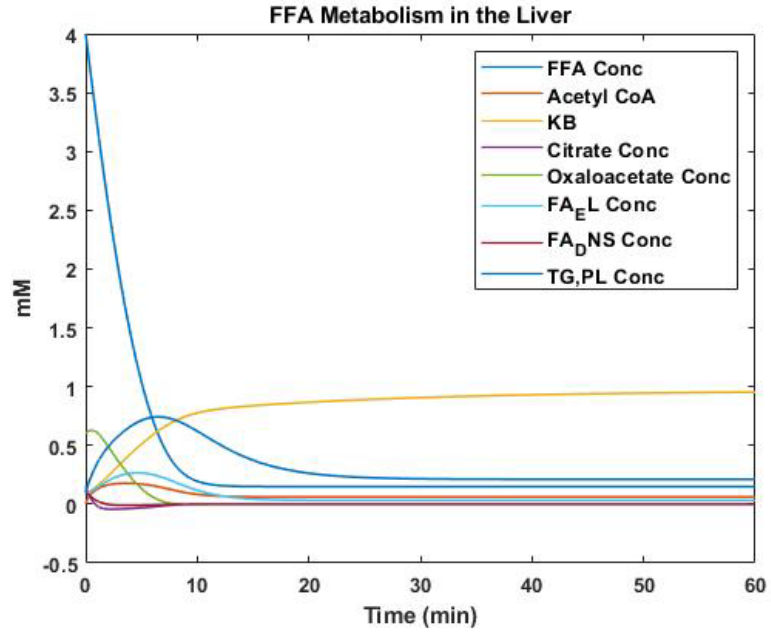
To develop a physics-informed ML model, a baseline mechanistic model is necessary. This project aimed to build the biological kinetic model in MATLAB, which would be incorporated into the CFD liver on a chip model developed by a collaborator, which could then be further optimized through ML algorithms. NAFLD affects a large population of adults and has yet to be dynamically understood so that a commercial drug could be made available for treatment. For a successful in silico model of the organ on a chip in vitro platform, the thermodynamics and mass transfer processes should be simulated precisely. However, due to the limitations of numerical methods for solving partial differential equations, precision can only be gained at the cost of computing time.

Additionally, many biological pathways are so complex involving multiple organs, cell types, enzymes, and other reactions, that accurate determination of kinetic parameters is rather challenging. The design below looks to reduce the limiting factors for enabling a high precision ML model of liver on the chip dynamics by modeling small molecule kinetics with first-order reaction kinetics. The parameters of this model include the kinetic parameters, the initial conditions for each molecule monitored, the feed rate of free fatty acids, the activity level and the drug dosage.

Excessive free fatty acid metabolism into triglycerides is known to be a contributing factor to hepatoxicity. Therefore, this model can be used to predict the concentrations of fatty acid intermediates and triglycerides in batch or continuous feed operations of free fatty acids.



A)

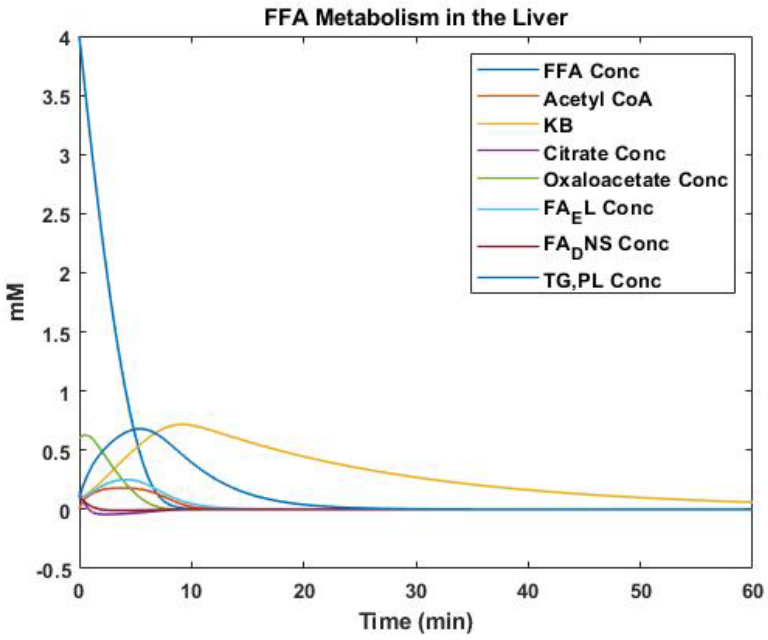


B)

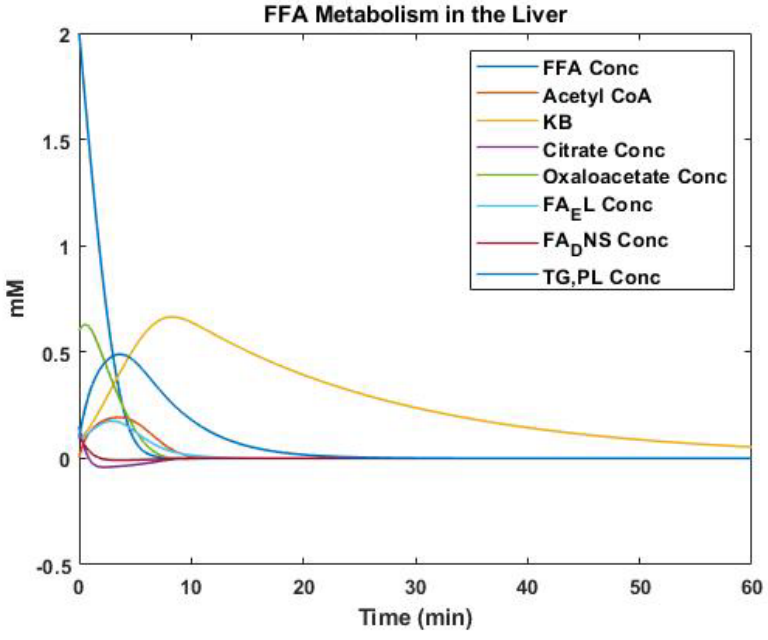
Figure 12: A) Batch supplement of free fatty acids results in the depletion of all components to 0. B) A continuous supplement of free fatty acids achieves steady state based upon the reaction kinetics of each individual component.

By manipulating the feed of free fatty acids in the model, the steady state curves allow for determining the first-order time dynamics of the fatty acid intermediates. The model's

response to the feed change signifies that triglyceride accumulation highly depends on the amount of fatty acids consumed. In the 0 to 10 min time range, there is a nearly linear response to free fatty acid depletion, and by halving the initial batch supplement of the time to reach below 0.5 mM FFAs is halved as shown below.



A)



B)

Figure 13: A) 4mM FFA starting concentration B) 2mM FFA starting condition

Increased exercise and diet change are currently the only two effective ways for a patient to prevent the onset of NAFLD or slow the disease's progression. It is essential to include a variable labeled the activity level for which the lifestyle of the human can be incorporated into the model. This activity coefficient is intended to manipulate the reaction kinetics of free fatty acids and the urea cycle by contributing to their depletion through a first-order reaction.

Equation 1

$$f(activity_{FFAs}) = k_{act} * A * C_{FFAs}$$

Equation 2

$$f(activity_{UREA}) = k_{actU} * A * C_{UREA}$$

Where k_{act} and k_{actU} are the activity reaction constants for the FFA and urea cycles. A is the activity coefficient ranging from 0.0 (unhealthy) to 1.0 (healthy). This range enables for the system's health to impact the metabolism of FFAs. The variables, C_{FFAs} and C_{UREA} are the concentrations of FFAs and urea respectively. These functions enable weights to be placed on the model for evaluating how the disease can impact patients of diverse lifestyles. By increasing the activity coefficient, the impact of depletion of FFAs in the system is observed, signifying more healthy individuals can potentially reverse the impacts of the disease.

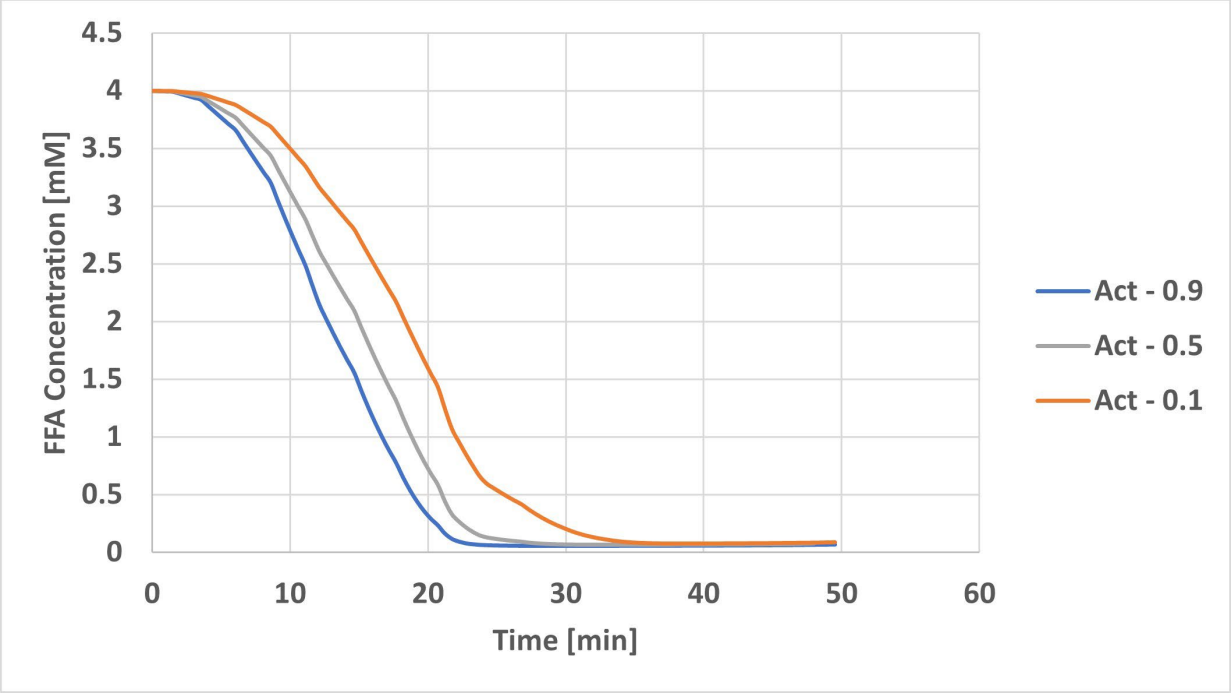


Figure 14: Impact of the activity coefficient on FFA Metabolism

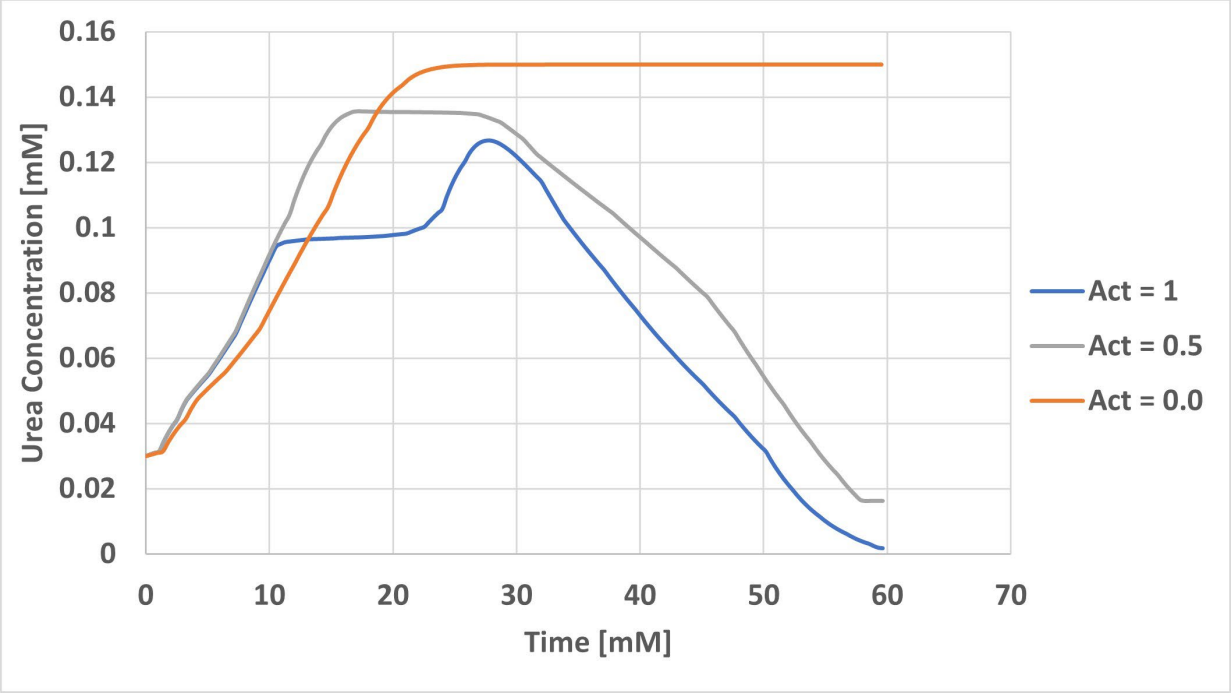


Figure 15: Impact of the activity coefficient on Urea Cycle

Pharmacokinetics of a drug allows for the relationship between the drug dosage and the active ingredients to be determined empirically. Chakraborty et al. established a simplified PK model of Metformin, which was directly integrated into the liver kinetic model to develop a system that responds to potential NAFLD treatments. Three drug concentrations were evaluated at an activity coefficient of 0 to demonstrate the effectiveness of Metformin in reducing FFA accumulation in the liver. As expected, the concentration increase resulted in the rate of FFA depletion. Although a concentration of 0.5 g/L of Metformin was demonstrated to be the most effective dose, further model development will be required to evaluate potential toxicity effects. To develop a toxicity mathematical model, *in vivo* data would be necessary to determine representative reaction rate constants.

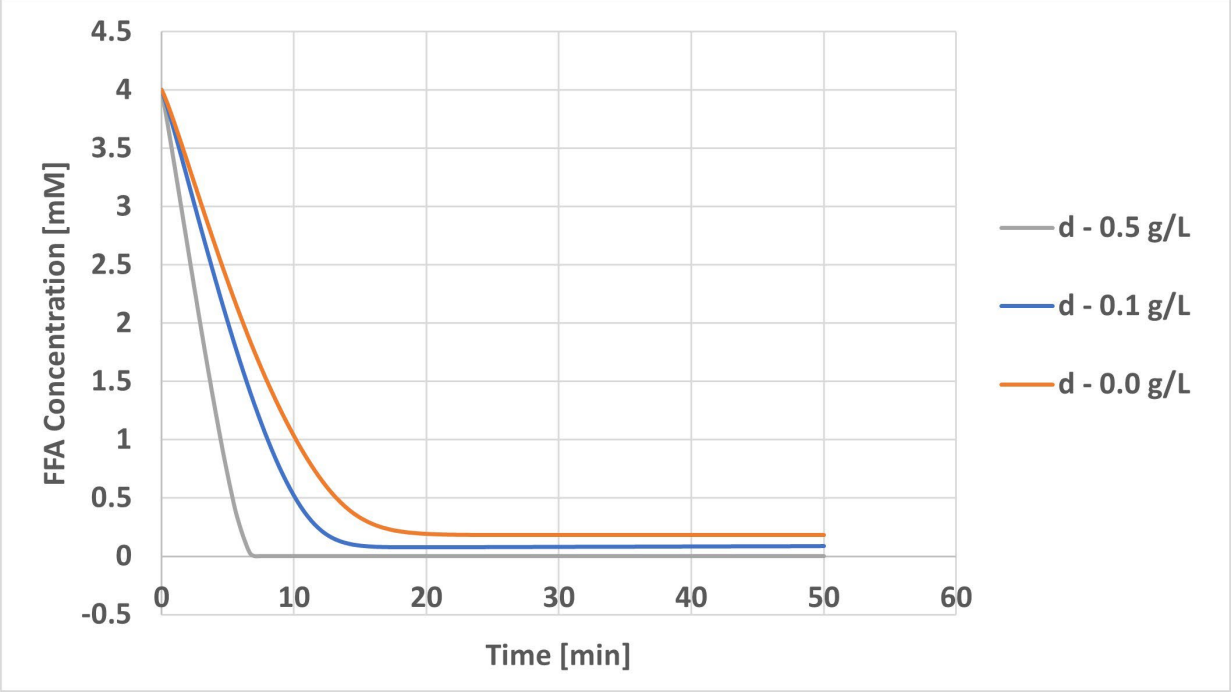


Figure 16: FFA Metabolism at Metformin Concentrations of 0 - 0.5 g/L.

Glucose metabolism is a fundamental part of the cell life cycle and a good indicator of cell growth. However, toxic effects from over-saturation and poor regulation by the body can lead to diabetes. An attempt to depict the effects of excess glucose accumulation is demonstrated by having a continuous feed of glucose in the system. In the model, the activity coefficient depicts the health of the platform and can lead to increase depletion, a lag, or steady state depending on the activity assignment. Within the studied feeds of 0 to 5 mM/min glucose, it is evident that the activity coefficient has a more significant impact on the accumulation of glucose.

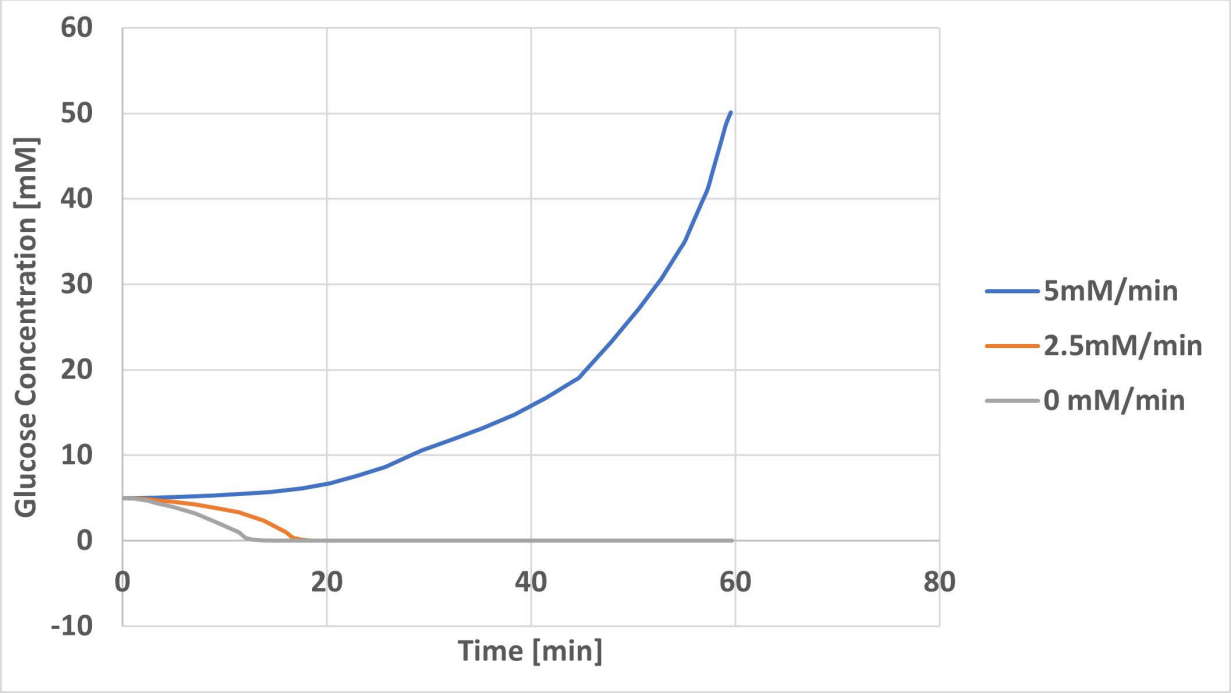


Figure 17: The glucose concentration in the liver at different feed rates of 0, 2.5 & 5 mM/min with the activity coefficient held constant at 0.5.

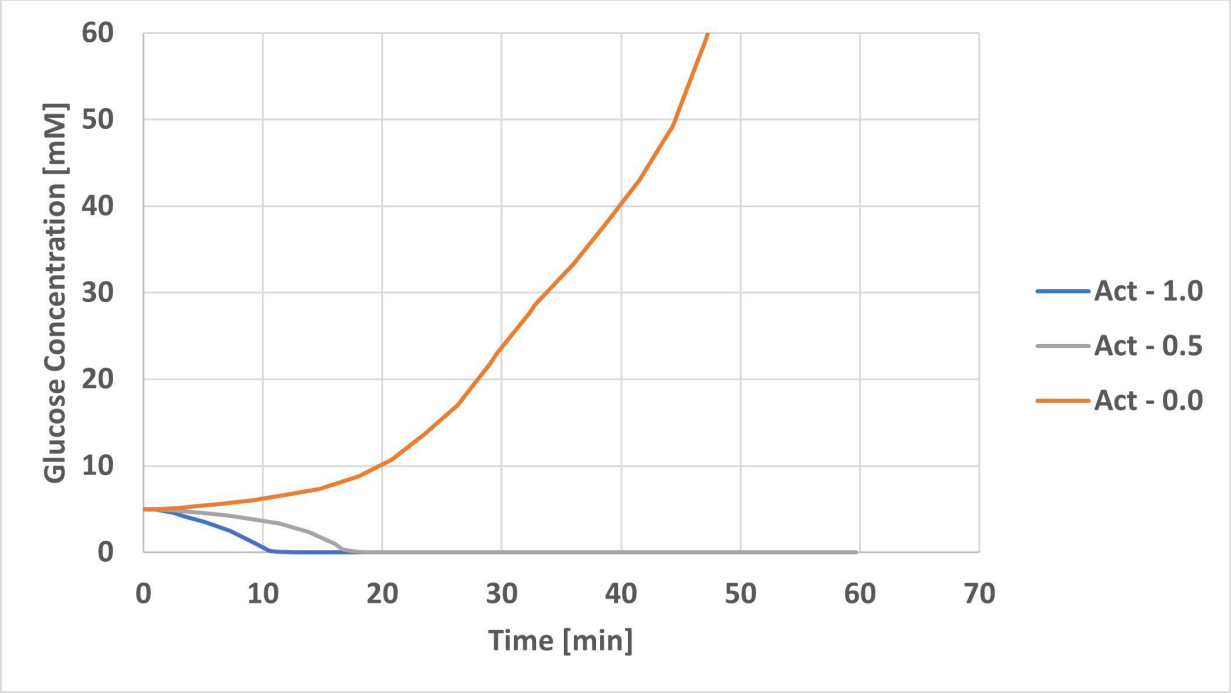


Figure 18: The concentration of glucose at different activity coefficients of 1, 0, & 0.5 with the glucose feed constant at 2.5 mM/min.

Discussion

Four modules were integrated into a single model to demonstrate the synergy between free fatty acid metabolism, glucose metabolism, the urea cycle, and the pharmacodynamics and kinetics of Metformin. In these modules, first order and Michaelis Menten kinetic equations were used to develop a simplified biological model that could be scaled to a liver on a chip device.In MATLAB, 21 small molecules were modeled to develop a parameterized system that could depict symptoms of NAFLD, the health of liver cells, and the effects of lifestyle choices. One of the main challenges in developing such a model was the limited resources for determining appropriate reaction parameters for the selected reaction pathways. Data was pulled from charts and published kinetic parameters from various *in vitro* models, including rats, mice, and tissue culture systems. Unknown parameters were estimated based upon similar mechanisms to achieve steady state kinetics. To ensure that the initial conditions were representative of a human liver, concentrations of various components in the blood and liver were used. Despite the data sourcing challenges, the model required minimal optimization after module integration. The four modules were selected for their importance in understanding the liver's biology based on data availability. These components, glucose, urea, fatty acids, triglycerides, and metformin are included in the NIH's database which will be used for ML implementation.

The original liver on a chip platform that this digital twin is based on operated more closely to a batch mode than a continuous feed. This model was designed so small molecules such as free fatty acids or glucose could be fed in batches or continuously. In batch mode, steady state equilibrium was achieved much quicker, and most intermediates were depleted to zero. Continuous feed dynamics have a more complex equilibrium and depend on whether the reaction pathway was favored more forwards or backwards. Manipulating the feed process allows for the

evaluation of special cases which are not possible *in vitro*. Especially in the case of high dose studies for understanding drug safety and efficacy, controlling the feed of fatty acids and glucose limits the system from achieving homeostasis rapidly. Performing such experiments *in vitro* and *in vivo* is not possible due to biological and ethical limitations.

The ability to manipulate the model's activity level as a depiction of health is a task which has not been widely used in other liver mathematical models and was an effective tool for controlling the productivity and functionality of the system based on a health score. This strategy should be more broadly applied in disease-based mathematical models due to the major impact that a patient's lifestyle has on their overall health. In the case of NAFLD where there are no commercially available drugs, sufficient exercise and proper diet are two of the most common approaches in disease management and treatment. In the FFA, urea, and glucose modules, increasing the activity coefficient improved the productivity of small molecule clearance. The effects were most prominent in the FFA and glucose modules due to drastic increases in metabolic rates. At high activity coefficients, the impacts of FFA and glucose diets (feeds) are practically insignificant. This aligns with the diets of high-caliber athletes who intake 2 – 3 times daily calories as the recommended 2000 calories/day but maintain healthy organs. For the urea cycle, the effects were minimal when transitioning from highly active to a moderate condition, but steady state excess accumulation was shown for a person of poor health ($A = 0$). Urea is a natural byproduct of even healthy cells; however, increased urea accumulation signifies cells transitioning into the death phase.

In addition to incorporating patient health, drug treatment was modeled for effects in the NAFLD system. Metformin was the model drug used in Lasli et al.'s liver on a chip platform, and a PK model was proposed by Chakraborty et al. [18, 23]. A simple PD model design utilized

first-order reaction kinetics of metformin on FFAs to demonstrate similar findings of the *in vitro* model. By performing a few calibration experiments to determine the dynamics and kinetics of a drug on an organ, sufficient reaction constants of other drugs can be determined and integrated with this model. In such a kinetic model, different drug concentrations at different patient health levels and diets can be evaluated for potency and toxicity effects. The interactive nature of diet, exercise, and drug dose produced useful trends and allowed for a minimum dose determination with respect to the system's health. To make this system clinically relevant, a scoring system should be generated so that the selected activity level is representative of the model or patient. Although this model does not evaluate potential organ toxicity, combining this model with a few toxicology studies allows for a baseline understanding of active ingredient concentration and residence time. This digital twin approach can evaluate experimental parameters at a much higher rate *in silico* than performing a complete factorial parameter evaluation in vitro.

Conclusion

Although the liver is a complex system of many cell types, this model was purposely developed to be simpler as opposed to modeling each reaction pathway in full. In the literature review, an overview of liver and NAFLD biology was described as well as the many *in vitro* systems being used in practice. Additionally, many biological mathematical model strategies were proposed for evaluating different aspects of the liver. ML applications in biomedical sciences were described, but none of these have developed a physics-based ML model for predicting *in vitro* systems using human patient data. This strategy is not restricted to livers since many ongoing studies investigate other organ platforms such as the heart, intestines, and kidneys.

By successfully integrating free fatty acids, urea, glucose, a PK/PD model, and an activity coefficient, realistic trends were produced for many different NAFLD scenarios in the liver. This mechanistic model will serve as a central aspect of a physics-informed ML model. Further optimization will allow for more representative reaction constants determination and NAFLD predictive algorithm development.

References

1. Takahashi, Y., Soejima, Y., & Fukusato, T. (2012). Animal models of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World journal of gastroenterology, 18*(19), 2300–2308. [PubMed]

2. Liver: Anatomy and Functions [Internet]. Baltimore, MA. Johns Hopkins Medicine. Available from: https://www.hopkinsmedicine.org/health/conditions-and-diseases/liver-anatomy-and-functions

3. Nagy, P., Thorgeirsson, S.S. and Grisham, J.W. (2020). Organizational Principles of the Liver. *In The Liver, 6*, 3. [Wiley]

4. Rotman, Y. and Kapuria, D. (2020). Non-alcoholic Fatty Liver Disease. *In The Liver*, 6, 670. [Wiley]

5. Nikmaneshi MR, Firoozabadi B, Munn LL. (2020). A mechanobiological mathematical model of liver metabolism. *Biotechnology and Bioengineering, 117*, 2861–2874. [PubMed]

6. Ashworth WB, Davies NA, Bogle ID. (2016). A Computational Model of Hepatic Energy Metabolism: Understanding Zonated Damage and Steatosis in NAFLD. *PLoS Comput Biol. 12*(9):e1005105. [PubMed]

7. Holzhütter HG, Berndt N. (2021). Computational Hypothesis: How Intra-Hepatic Functional Heterogeneity May Influence the Cascading Progression of Free Fatty Acid-Induced Non-Alcoholic Fatty Liver Disease (NAFLD). *Cells. 10*(3):578. [PubMed]

8. Lasli S, Kim HJ, Lee K, Suurmond CE, Goudie M, Bandaru P, Sun W, Zhang S, Zhang N, Ahadian S, Dokmeci MR, Lee J, Khademhosseini A. (2019). A Human Liver-on-a-Chip Platform for Modeling Nonalcoholic Fatty Liver Disease. *Adv Biosyst. 3*(8), e1900104. [Wiley]

9. Suurmond, C. E., Lasli, S., van den Dolder, F. W., Ung, A., Kim, H. J., Bandaru, P., Lee, K., Cho, H. J., Ahadian, S., Ashammakhi, N., Dokmeci, M. R., Lee, J., & Khademhosseini, A. (2019). In Vitro Human Liver Model of Nonalcoholic Steatohepatitis by Coculturing Hepatocytes, Endothelial Cells, and Kupffer Cells. *Advanced healthcare materials, 8*(24), e1901379. [PubMed]

10. Ramos, M. J., Bandiera, L., Menolascina, F., & Fallowfield, J. A. (2021). In vitro models for non-alcoholic fatty liver disease: Emerging platforms and their applications. *iScience, 25*(1), 103549. [PubMed]

11. Zhang, Y. S., Aleman, J., Shin, S. R., Kilic, T., Kim, D., Mousavi Shaegh, S. A., Massa, S., Riahi, R., Chae, S., Hu, N., Avci, H., Zhang, W., Silvestri, A., Sanati Nezhad, A., Manbohi, A., De Ferrari, F., Polini, A., Calzone, G., Shaikh, N., Alerasool, P., … Khademhosseini, A. (2017). Multisensor-integrated organs-on-chips platform for

automated and continual in situ monitoring of organoid behaviors. *Proceedings of the National Academy of Sciences of the United States of America, 114*(12), E2293–E2302. [PubMed]

12. Deng, J., Wei, W., Chen, Z., Lin, B., Zhao, W., Luo, Y., & Zhang, X. (2019). Engineered Liver-on-a-Chip Platform to Mimic Liver Functions and Its Biomedical Applications: A Review. *Micromachines, 10*(10), 676. [PubMed]

13. Bhise, N. S., Manoharan, V., Massa, S., Tamayol, A., Ghaderi, M., Miscuglio, M., Lang, Q., Shrike Zhang, Y., Shin, S. R., Calzone, G., Annabi, N., Shupe, T. D., Bishop, C. E., Atala, A., Dokmeci, M. R., & Khademhosseini, A. (2016). A liver-on-a-chip platform with bioprinted hepatic spheroids. *Biofabrication, 8*(1), 014101. [PubMed]

14. Banaeiyan, A. A., Theobald, J., Paukštyte, J., Wölfl, S., Adiels, C. B., & Goksör, M. (2017). Design and fabrication of a scalable liver-lobule-on-a-chip microphysiological platform. *Biofabrication, 9*(1), 015014. [PubMed]

15. Möller, J.; Pörtner, R. (2021). Digital Twins for Tissue Culture Techniques—Concepts, Expectations, and State of the Art. *Processes, 9,* 447. [MDPI]

16. Zou, H., Banerjee, P., Leung, S., & Yan, X. (2020). Application of Pharmacokinetic-Pharmacodynamic Modeling in Drug Delivery: Development and Challenges. *Frontiers in pharmacology, 11*, 997. [PubMed]

17. Lyons, M. A., & Lenaerts, A. J. (2015). Computational pharmacokinetics/pharmacodynamics of rifampin in a mouse tuberculosis infection model. *Journal of pharmacokinetics and pharmacodynamics, 42*(4), 375–389. [PubMed]

18. Chakraborty S, Arora K, Sharma P K, Nath A, Bhattacharjee C. (2014). Pharmacokinetics Study of Metformin – Mathematical Modelling and Simulation. *APCBEE Procedia, 9*:151-158. [Elsevier]

19. Roskoski R. (2007). Michaelis-Menten Kinetics. *xPharm: The Comprehensive Pharmacology Reference*, 1-10. [Elsevier]

20. Berndt N, Horger MS, Bulik S, Holzhütter HG. (2018). A multiscale modelling approach to assess the impact of metabolic zonation and microperfusion on the hepatic carbohydrate metabolism. *PLoS Comput Biol. 14*(2):e1006005. [PubMed]

21. López-Palau, N. E., & Olais-Govea, J. M. (2020). Mathematical model of blood glucose dynamics by emulating the pathophysiology of glucose metabolism in type 2 diabetes mellitus. *Scientific reports, 10*(1), 12697. [PubMed]

22. Schleicher, J., Dahmen, U., Guthke, R., & Schuster, S. (2017). Zonation of hepatic fat accumulation: insights from mathematical modelling of nutrient gradients and fatty acid uptake. *Journal of the Royal Society, Interface, 14*(133), 20170443. [PubMed]

23. Scheff, J.D., Foteinou, P.T., Calvano, S.E., Lowry, S.F. and Androulakis, I.P. (2010). Dynamic Models of Disease Progression: Toward a Multiscale Model of Systemic Inflammation in Humans. *In Process Systems Engineering, 7,* 321. [Wiley]

24. Karniadakis, G.E., Kevrekidis, I.G., Lu, L. Perdikaris, P. Wang, S. Yang, L. (2021) Physics-informed machine learning. *Nat Rev Phys, 3,* 422–440. [Nature]

25. Narang R, Deva J, Dwivedi SN. (2019). Machine Learning and Medical Research Data Analysis. *J-PCS, 5*(1), 12-13. [J-PCS]

26. Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. *An international journal on information fusion, 50*, 71–91. [PubMed]

27. Varoquaux, G., & Cheplygina, V. (2022). Machine learning for medical imaging: methodological failures and recommendations for the future. *NPJ digital medicine, 5*(1), 48. [PubMed]

28. Teles, D., Kim, Y., Ronaldson-Bouchard, K., & Vunjak-Novakovic, G. (2021). Machine Learning Techniques to Classify Healthy and Diseased Cardiomyocytes by Contractility Profile. *ACS biomaterials science & engineering, 7*(7), 3043–3052. [PubMed]

29. Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?. *Heart (British Cardiac Society), 104*(14), 1156–1164. [PubMed]

30. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *The New England journal of medicine, 380*(14), 1347–1358. [PubMed]

31. Ortega-Prieto, A. M., Skelton, J. K., Wai, S. N., Large, E., Lussignol, M., Vizcay-Barrena, G., Hughes, D., Fleck, R. A., Thursz, M., Catanese, M. T., & Dorner, M. (2018). 3D microfluidic liver cultures as a physiological preclinical tool for hepatitis B virus infection. *Nature communications, 9*(1), 682. [PubMed]

32. Lee, J., Mehrotra, S., Zare-Eelanjegh, E., Rodrigues, R. O., Akbarinejad, A., Ge, D., Amato, L., Kiaee, K., Fang, Y., Rosenkranz, A., Keung, W., Mandal, B. B., Li, R. A., Zhang, T., Lee, H., Dokmeci, M. R., Zhang, Y. S., Khademhosseini, A., & Shin, S. R. (2021). A Heart-Breast Cancer-on-a-Chip Platform for Disease Modeling and Monitoring of Cardiotoxicity Induced by Cancer Chemotherapy. *Small (Weinheim an der Bergstrasse, Germany), 17*(15), e2004258. [PubMed]

33. Ren, L., Zhou, X., Nasiri, R., Fang, J., Jiang, X., Wang, C., Qu, M., Ling, H., Chen, Y., Xue, Y., Hartel, M. C., Tebon, P., Zhang, S., Kim, H. J., Yuan, X., Shamloo, A., Dokmeci, M. R., Li, S., Khademhosseini, A., Ahadian, S., … Sun, W. (2020). Combined Effects of Electric Stimulation and Microgrooves in Cardiac Tissue-on-a-Chip for Drug Screening. *Small methods, 4*(10), 2000438. [PubMed]

34. Ontko JA. (1972). Metabolism of free fatty acids in isolated liver cells. Factors affecting the partition between esterification and oxidation. *J Biol Chem. 247*(6):1788-800. [PubMed]

35. Sohlenius-Sternbeck, A.K., (2006). Determination of the hepatocellularity number for human, dog, rabbit, rat and mouse livers from protein concentration measurements. *Toxicology in Vitro. 20*, 1582-1586. [PubMed]

36. Cohen P. P. (1981). The ornithine-urea cycle: biosynthesis and regulation of carbamyl phosphate synthetase I and ornithine transcarbamylase. *Current topics in cellular regulation, 18*, 1–19. [PubMed]

37. Kawaguchi-Suzuki, M., & Frye, R. F. (2013). Current clinical evidence on pioglitazone pharmacogenomics. *Frontiers in pharmacology, 4*, 147. [PubMed]

Appendix A - MATLAB Code

```
function ddt = TotalKin(t, y)
%model includes drug and activity coefficient
%drug is assumed to follow first order reaction kinetics
%rate constants have been changed to mM/min from DPM using linear reg

C1=y(1); %FFA
C2=y(2); %Acetyl CoA
C3=y(3); %Ketone Bodies
C4=y(4); %Citrate
C5=y(5); %Oxaloacetate
C6=y(6); %FA Elongation
C7=y(7); %FA de novo synthesis
C8=y(8); %Triglyceride, Phospholipids
C9 = y(9); %activity coefficient from 0- 1.0; 0 = unhealthy 1.0 = most healthy
C10 = y(10); %Glucose in the liver
C11 = y(11); %Glucose consumed
C12=y(12); %Metformin in the Stomach
C13=y(13); %Soluble Metformin in the Stomach
C14=y(14); %Metformin+OCTN2 (Intestine)
C15=y(15); %Metformin+SLC22A1/SLC22A3 (Liver)
C16=y(16); %Metformin (Tissue)
C17=y(17); %Metformin in DCT (Kidney)
C18=y(18); %Ornithine
C19=y(19); %Citrulline
C20=y(20); %Arginosuccinate
C21=y(21); %Arginine
C22=y(22); %Urea
C23=y(23); %FFA Feed Rate

%Constants
Vm1 =.19;%Max rate for reaction1
Vm2 = .2; %Max rate for reaction2
Vm3 = .36; %Max rate for reaction3
Vm4 = .28; %Max rate for reaction4
Vm5 = .2; %Max rate for reaction5
Vm6 = .22; %Max rate for reaction6
Vm7 = .40; %Max rate for reaction7
Vm8 = .19; %Max rate for reaction8
Vm9 = .94; %Max rate for reaction9
Vm10 = .52; %Max rate for reaction10
Vm14 = 1; %Max rate for reaction14
Vm15 = .75; %Max rate for reaction15
Vm19 =3.24; %Max rate for reaction19
Vm20 = .5; %Max rate for reaction20
Vm21 = .5; %Max rate for reaction21
Vm22 = .603; %Max rate for reaction22
Vm23 = .5; %Max rate for reaction23
Vm24 = .169; %Max rate for reaction24
Km1 =.3; %MM Constant for reaction1
Km2_2 = .2; %MM Constant for reaction2_2
Km2_5 = .2; %MM Constant for reaction2_5
Km3 =.4; %MM Constant for reaction3
```

```
Km4 = .2; %MM Constant for reaction4
Km5 = .2; %MM Constant for reaction5
Km6 =0.6; %MM Constant for reaction6
Km7_1 =.8; %MM Constant for reaction7_1
Km7_2 = .2; %MM Constant for reaction_2
Km8 = .6; %MM Constant for reaction8
Km9 = 2; %MM Constant for reaction9
Km10 = .3; %MM Constant for reaction10
Km11 = 1e-2; %MM Constant for reaction11
Km14 = 4; %MM Constant for reaction14
Km15 = 4.5; %MM Constant for reaction15
Km19 =.4; %MM Constant for reaction19
Km20 = .04; %MM Constant for reaction20
Km21 =.1; %MM Constant for reaction21
Km22 = .1; %MM Constant for reaction22
Km23 = .1; %MM Constant for reaction23
Km24 =0.1; %MM Constant for reaction24
Vm25 = 8.5; %MM Constant for reaction25
kact = 2.5e-1; %activity reaction constant for FFA pathway
kactU = 7.5e-2; %activity reaction constant for urea pathway
k9 = 4.5e-3; %drug metabolism rate constant
k10 = .75; %drug FA reaction rate constant
k = 1; %Dissolution rate of Metformin in stomach
a = .1; %linear blood flow rate from stomach to intestine
b = .25; %linear blood flow rate from intestine to liver
c = .5; %linear blood flow rate from liver to tissue
d = 1; %linear blood flow rate from tissue to kidney

%Explicit Equations
V1= Vm1*C1/(Km1+C1); %reaction1
V2= Vm2*C2*C5/(Km2_2*Km2_5 + Km2_2*C5+Km2_5*C2 + C2*C5); %reaction2
V3= Vm3*C2/(Km3+C2); %reaction3
V4= Vm4*C4/(Km4+C4); %reaction4
V5= Vm5*C5/(Km5+C5); %reaction5
V6= Vm6*C4/(Km6+C4); %reaction6
V7= Vm7*C2*C1/(Km7_1*Km7_2+Km7_1*C2+Km7_2*C1+C2*C1); %reaction7
V8= Vm8*C1/(Km8+C1); %reaction8
V9= Vm9*C6/(Km9+C6); %reaction9
V10= Vm10*C7/(Km10+C7); %reaction10
V15= Vm14*C13/(Km14+C13); %MM eq for reaction15
V16= Vm15*C14/(Km15+C14); %MM eq for reaction16
V19= Vm19*C21/(Km19+C21); %reaction19
V20= Vm20*C18/(Km20 + C18); %reaction20
V21= Vm21*C19/(Km21+C19); %reaction21
V22= Vm22*C20/(Km22+C20); %reaction22
V23= Vm23*C21/(Km23+C21); %reaction23
V24= Vm24*C21/(Km24+C21); %reaction24
V25= Vm25*C10/(Km11+C10); %reaction25

%Differential Equations
ddt(1)= C23 -V1 - V7 - V8 - kact*C9*C1 - k10*C17;
ddt(2)= V1 + V5 - V2 -V6 - V7 - V3;
ddt(3)= V3 - .05*C3;
ddt(4)= V2 - V4 - V5;
ddt(5)= V4 - V2;
```

```
ddt(6) = V7 - V9;
ddt(7) = V6 - V10;
ddt(8) = V8 + V9 + V10 - .25*C8;
ddt(9) = 0;
ddt(10) = C11 - C9*V25;
ddt(11) = 0;
ddt(12)= -k*C12;
ddt(13)= k*C12-a*C13;
ddt(14)= V15 - b*C14;
ddt(15)= V16 - c*C15;
ddt(16)= c*C15 - d*C16;
ddt(17) = d*C16;
ddt(18)= V24-V20;
ddt(19)= V20 + V23 - V21;
ddt(20)= V21 - V22;
ddt(21)= V22 - V23 - V24 - V19;
ddt(22)= V19-kactU*C9*C22;
ddt(23)= 0;

ddt = ddt';
end

%Total Kinetic Model Driver
clc; clear;

a = .5; %activity coefficient 0-no physical activity 1-high physical activity
m = 0.1; %Metformin dose
f = 0; %FFA Feed rate
g = 2.5; %glucose feed rate
tspan = [0 60]; %Range of independent variable

y0=[4 .003 0.1 0.15 0.6 0.1 0.1 0.1 a 5 g m 0 0 0 0 0 0.04 0.04 0.01 0.03 0.03 f];

%IVs[mM] FFA (1), Acetyl CoA (2), Ketone Bodies (3), Citrate (4), Oxaloacetate (5),
FA_El (6), FA_DNS (7), TG&PL (8), AC, Act Coef (9),
%Glucose(liver) (10), Glucose(consumed) (11), Metformin(Stomach) (12), Soluble
Metformin(Stomach) (13), Metformin+OCTN2 (Intestine) (14),
%Metformin+SLC22A1/SLC22A3 (Liver) (15), Metformin (Tissue) (16), Metformin in DCT
(Kidney) (17), Ornithine (18), Citrulline (19),
%Arginosuccinate (20), Arginine (21), Urea (22), FFA Feed (23)

[t, y]=ode15s(@TotalKin,tspan, y0);

plot(t,y(:,1), t,y(:,2), t,y(:,3), t,y(:,4), t,y(:,5), t,y(:,6), t,y(:,7), t,y(:,8))
legend( 'FFA Conc', 'Acetyl CoA', 'KB', 'Citrate Conc','Oxaloacetate Conc', 'FA_EL
Conc', 'FA_DNS Conc', 'TG,PL Conc')
xlabel('Time (min)')
ylabel('mM')
title('FFA Metabolism in the Liver
```

Glossary of Terms

3D: Three Dimension

AI: Artificial Intelligence

ANN: Artificial Neural Network

BLA: Biologics Licensing Agreement

DT: Digital Twin

FDA: Food and Drug Administration

FA: Fatty Acid

FFA: Free Fatty Acid

HepG: Hepatocellular Carcinoma Cells

HKC: Human Kupffer Cells

HUVEC: Umbilical Vein Endothelial Cells

in silico: performed via computer simulation

in vitro: performed in a test tube or outside a living organism

in vivo: performed on whole living organisms

ML: Machine Learning

NAFLD: Nonalcoholic Fatty Liver Disease

NASH: Nonalcoholic Steatohepatitis

NIH: National Institute of Health

NN: Neural networks

ODE: Ordinary Differential Equation

OOC: Organ on a Chip

PDE: Partial Differential Equation

PD/PK: Pharmacodynamic/Pharmacokinetic

PINS: Physics Informed Network

SVM: Support Vector Machine

TG: Triglycerides

VLDL: Very Low Density lipoproteins