Modeling fructose-load-induced hepatic de-novo lipogenesis by model simplification

Richard J Allen and Cynthia J Musante

Cardiovascular and Metabolic Diseases, Pfizer Inc., USA.

Gene Regulation and Systems Biology
Volume 11: 1–10

DOI: 10.1177/1177625017690133

ABSTRACT: Hepatic de-novo lipogenesis is a metabolic process implemented in the pathogenesis of type 2 diabetes. Clinically, the rate of this process can be ascertained by use of labeled acetate and stimulation by fructose administration. A systems pharmacology model of this process is desirable because it facilitates the description, analysis, and prediction of this experiment. Due to the multiple enzymes involved in de-novo lipogenesis, and the limited data, it is desirable to use single functional expressions to encapsulate the flux between multiple enzymes. To accomplish this we developed a novel simplification technique which uses the available information about the properties of the individual enzymes to bound the parameters of a single governing 'transfer function'. This method should be applicable to any model with linear chains of enzymes that are well stimulated. We validated this approach with computational simulations and analytical justification in a limiting case. Using this technique we generated a simple model of hepatic de-novo lipogenesis in these experimental conditions that matched prior data. This model can be used to assess pharmacological intervention at specific points on this pathway. We have demonstrated this with prospective simulation of acetyl-CoA carboxylase inhibition. This simplification technique suggests how the constituent properties of an enzymatic chain of reactions gives rise to the sensitivity (to substrate) of the pathway as a whole.

KEYWORDS: Linear chain, de-novo lipogenesis, model simplification

RECEIVED: September 30, 2016. **ACCEPTED:** December 05, 2016.

PEER REVIEW: Five peer reviewers contributed to the peer review report. Reviewers' reports totaled 673 words, excluding any confidential comments to the academic editor.

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: RJA, and CJM were employees of Pfizer Inc. during the completion and/or analyses of these studies. Pfizer Inc. has provided all funding for these studies and was responsible for the study design, data analysis, decision to publish, and preparation of the manuscript.

CORRESPONDING AUTHOR: Richard J Allen, Cardiovascular and Metabolic Diseases, Internal Medicine Research Unit, Pfizer Inc., 1 Portland Street, Cambridge, MA 02139, USA. Email: richard.allen@pfizer.com

Introduction

De-novo lipogenesis (DNL) is a complicated hormonally regulated pathway that is active in humans in liver and adipose tissue,[1] and to a lesser extent in skeletal muscle.[2] This pathway is one mechanism for converting carbohydrates into fat, which is eventually exported, or stored, as triglycerides (TGs). Hepatic DNL has been implicated in the pathogenesis of diabetes and hepatic steatosis.[3,4] In type 2 diabetics, where homeostatic mechanisms fail to control the level of glucose and gluconeogeneic precursors,[5] hepatic DNL may act as a way of diverting these substrates away from gluconeogenesis and hepatic glucose output.[6] However, the increased lipid load is an independent risk factor for cardiovascular disease.[7]

A mechanistic, quantitative, understanding of the relationship between substrate load and lipid accumulation via DNL is fundamental to fully appreciating the role of DNL in disease. Fructose is a lipogenic dietary substrate that is up-taken into hepatocytes rapidly on the first pass through the liver.[8] Hence, one technique to stimulate and assay DNL is repeated oral dosing of fructose in the presence of exogenous C14 labeled acetate. Then DNL is measured as the fraction of plasma TGs that have incorporated C14 acetate, using the MIDA method. Hudgins et al.[9] reported that this technique increased de-novo lipids in the plasma by approximately 15%. The purpose of this paper is to develop and use a mathematical 'systems pharmacology' model of this protocol to gain a quantitative understanding of the process. This will allow us, in the future, to predict the effect of putative pharmacological agents in this context.

The development of systems pharmacology models is a time-consuming process. This is because they attempt to faithfully represent key physiological mechanisms mathematically. Hence, they are often large and complex, requiring data from multiple species and experiments. A crucial challenge within the development process is accommodating a balance between mechanistic detail and available experimental knowledge to inform that detail. Therefore, we believe that tools which resolve this tension by systematic simplification of models are particularly valuable. Often the reduction to simplified representations occurs in the development and design phase of the process. In this case it is usually difficult to retain the correspondence to physiology and experimental data.

In the protocol applied by Hudgins et al.,[9] it is plausible to assume that metabolism of fructose to plasma TGs can be well approximated by a linear network (Figure 1). The most direct route from fructose to plasma TGs involves 16 intermediaries: see the supplementary file. However, even in a model where each reaction is described by Hill kinetics this will give rise to at least a 32 parameter model, whereas if we seek to simply retain the boxed species (Figure 1), and construct transfer functions between them, the size of the model will be reduced dramatically. As such, it is desirable to design a simplified representative model of this process incorporating key mechanistic features and known experimental data regarding enzyme properties of the individual steps, while not modeling each enzyme explicitly.

Model reduction will often be an esoteric case-dependent exercise. However, there are some general approaches that have been developed. For example, one way to simplify a model is to separate time-scales, so that interactions faster than the time-scale of interest can be assumed to be in a quasi steady-state.[10–12] Another approach is to reduce parameters based on Hankel singular values, and/or singular value decomposition.[13] This

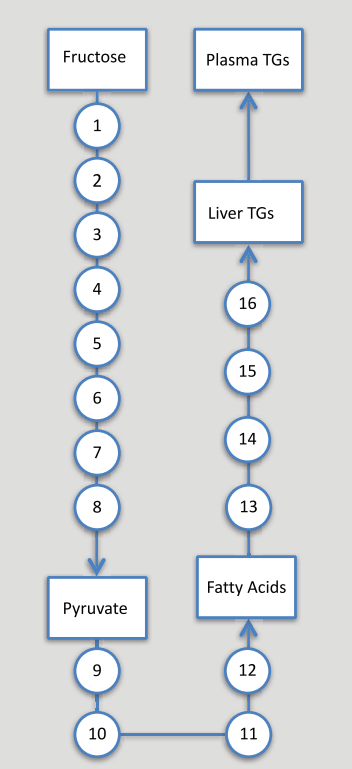

Figure 1. The shortest route from fructose to plasma triglycerides (TGs) includes 16 steps (see Supplementary Table 2). In the experimental conditions of Hudgins et al.,[9] where hepatic fructose exposure is high for a period of approximately 8 hours, we assume that the overwhelming majority of fructose-derived TGs has been synthesized by this pathway.

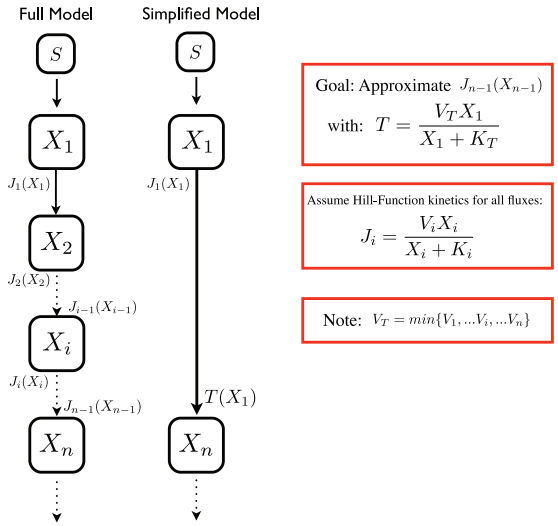

Figure 2. Approximating linear chains with transfer functions. Often it is desirable to reduce the 'full model' motif, to a 'simplified model' motif. In this case, the input species X_1 and the output species X_n are retained in the simplification, and the intermediary fluxes approximated by the total flux, captured by $T(X_1)$, which must approximate $J_{n-1}(X_{n-1})$.

method ranks the parameters based on their effect on the input–output response of the model. The approach we are proposing here is a novel contribution because it uses available experimental information to establish approximate values for lumped parameters that otherwise do not have a direct biochemical and experimental meaning.

An obvious way to simplify a model is to compact linear-chain motifs (synonymously, path graph or linear tree) of a network into the input and output of the motif. If we believe the intermediaries of this graph are not rate-limiting and we do not have time-course data for those nodes, then all we are really interested in, or can constrain, is the 'transfer function' from the first to the last node (Figure 2).

For enzymatic reaction networks a biologically relevant transfer function is a Hill function,[14] which arises naturally in the case of Michaelis–Menten kinetics. This rate law is typically a reasonable assumption for metabolic networks, but for signal transduction networks, where interactions are more complex (for example, when proteins can be both enzymes and substrates) there are superior approximations.[15] Here we assume that Hill functions (with Hill coefficient one) can capture the experimental dynamics and still be predictive, even if the assumptions required for Michaelis–Menten kinetics are not precisely met. Hill functions, similar to metabolic reactions, are saturable, and can give rise to rich dynamics,[16] which we presume includes the requisite model behavior. Here we consider Hill functions with coefficient one, but our analysis shows that similar relationships to the one we develop will hold true at higher values.

Our approach to simplifying linear-chain motifs uses prior knowledge of the system under consideration. For metabolic pathways it is often the case that the maximum rate of a reaction (V_{max}) is not known, but enzyme–substrate affinity is known (related to the Michaelis constant K_m in the case of Michaelis–Menten kinetics). This is because K_m can be measured *in vitro*, whereas V_{max} is a function of in-vivo enzyme expression and regulation, and is less readily known in vivo. Furthermore, in dysregulated disease states V_{max} for given reactions could be, for example, greatly attenuated. In systems pharmacology models these parameters are often using experiments that capture the whole system behavior (i.e. the system the model is describing), rather than informed from direct measurements of the behavior of constitutive component of the model.

A linear chain of n reactions modeled with Michaelis–Menten kinetics will have $2n$ parameters. By introducing a transfer function from the first to last node we can simplify this motif to being determined by either two or four parameters, depending on if we assume conservation of mass (so that the flux leaving the first node is matched by flux into the last). In this case, mass is not conserved in the model because it is 'lost' to the intermediary substrates that we are not considering. However, due to the linear structure of the pathway, we gain a constraint that the maximum rate, V_T, of the transfer function should be less than or equal to V_1 (the maximum flux out of the first node).

While this simplification is often necessary, it is still highly desirable to use all prior knowledge to constrain the transfer function. As mentioned, it is unusual for detailed information

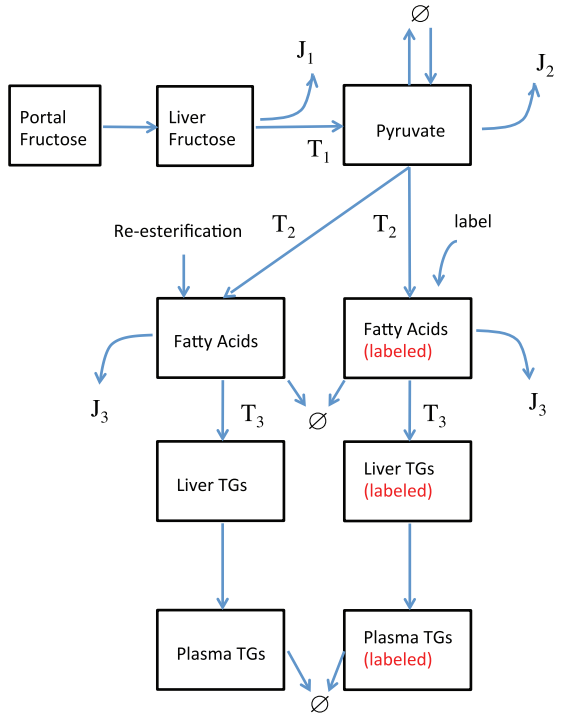

Figure 3. Simplified model of DNL in fructose-stimulated conditions. This model structure was designed based on known biology, experimental protocol,[9] and the simplification of a larger metabolic pathway (Figure 2).

regarding the V_i to be available (however, the maximum rate, V_T, of the transfer function is more readily measured or constrained by experimental data). Although there are graph theoretic approaches to solving models with simpler flux terms,[17] as far as we are aware there are no existing methods for using data about individual reactions for constraining K_T, even though detailed measurements of these values are often available. How to use available information to estimate K_T is the focus of this work.

To motivate our approach for simplification we first introduce a novel model of hepatic DNL which includes the aforementioned 'transfer functions'. Following this we describe our novel approach to parameterizing these functions, which is justified by both computational and analytical rationale.

Modeling hepatic DNL in response to high fructose levels

Model equations

The model detailed in Figure 3 was described mathematically as follows. It was assumed that all reactions could be captured by Hill functions (i.e. can be approximated by Michaelis–Menten-type kinetics). This approximation will not always be valid, and certainly the Michaelis–Menten interpretation of their meaning is not necessarily valid unless specific assumptions are met. We assume that in conditions where the pathway is operating unidirectionally (i.e. highly stimulated at the initial node) that this functional form sufficiently captures the flux from node to node.

Notation

We use F, P, $[FA]$, and $[TG]$ to denote fructose, pyruvate, fatty acids, and TGs, respectively. Square brackets denote species concentration. Subscripts refer to different pools for each species (bolus, plasma, or liver). A superscript D denotes the metabolites which have derived from DNL (identified experimentally with C14 acetate) a superscript T denotes total metabolite concentration (unlabeled + labeled):

$$\frac{dF_B}{dt} = D.\delta(\mathbf{T},t) - \beta F_B \tag{1}$$

$$\frac{d[F_P]}{dt} = \frac{(1-\rho_1)\beta F_B}{V_P^F} - \gamma[F_P] + \alpha_1 \tag{2}$$

$$\frac{d[F_L]}{dt} = \frac{\rho_1 \beta F_B}{V_L^F} - J_1 \tag{3}$$

$$\frac{d[P_L]}{dt} = T_1 - J_2 + \alpha_2 \tag{4}$$

$$\frac{d[FA]_L}{dt} = (1-R)T_2 - r_1 J_3 - \sigma_1[FA]_L + \alpha_3 \tag{5}$$

$$\frac{d[TG]_L}{dt} = r_1 T_3 - r_2 J_4 \tag{6}$$

$$\frac{d[TG]_P}{dt} = V_r r_2 J_4 - \sigma_2[TG]_P + \alpha_4 \tag{7}$$

$$\frac{d[FA]_L^D}{dt} = RT_2 - (1-r_1)J_3 - \sigma_1[FA]_L^D \tag{8}$$

$$\frac{d[TG]_L^D}{dt} = (1-r_1)T_1 - (1-r_2)J_4 \tag{9}$$

$$\frac{d[TG]_P^D}{dt} = V_r(1-r_2)J_4 - \sigma_2[TG]_P^D \tag{10}$$

Rationale for equations (1)–(3). These equations predict plasma and liver fructose following oral administration of fructose at discrete times given by the set $\mathbf{T}$. Here D is the dose of fructose administered at each time $T_i \in \mathbf{T}$, and δ is the dirac delta function. We use F_B to denote the amount of fructose bolus that has not yet appeared in plasma in units of mmol and $[F_P]$ is the plasma fructose concentration in units of mmol/L. Here V_p^F is the volume of distribution of fructose in plasma. Plasma fructose is cleared at a rate γ in units of min^{-1}. In this model, the fructose bolus is partitioned into either liver or plasma ρ_1 is the fraction absorbed into the liver, with volume of distribution V_L^F. Here V_r is the ratio of liver and plasma volume of distributions for TGs.

Rationale for equations (4)–(10). The flux terms in these expression are defined as outlined in Figure 2, more specifically:

$$J_1 = \frac{V_1[F_L]}{K_1 + [F_L]} \quad (11)$$

$$T_1 = \frac{V_2[F_L]}{K_2 + [F_L]} \quad (12)$$

$$J_2 = \frac{V_3[P_L]}{K_3 + [P_L]} \quad (13)$$

$$T_2 = \frac{V_4[P_L]}{K_4 + [P_L]} \quad (14)$$

$$J_3 = \frac{V_5[FA]_L^T}{K_5 + [FA]_L^T} \quad (15)$$

$$T_3 = \frac{V_6[FA]_L^T}{K_6 + [FA]_L^T} \quad (16)$$

$$J_4 = \frac{V_7[TG]_L^T}{K_7 + [TG]_L^T} \quad (17)$$

When applied, J_3, T_3, and J_4 are scaled by r_1 and r_2 which are defined as the fraction unlabeled metabolite in each case (see equations (6), (7), (9), and (10)). Due to the constraints $T_i \leqslant J_i$ (see equation (22)), we can rewrite $V_2 = \rho_2 V_1$, $V_4 = \rho_3 V_3$ and $V_6 = \rho_4 J_5$, where $\rho_i \in [0,1]$.

Hudgins et al.[9] measured DNL by using C14 acetate, which identifies the de-novo lipids using the MIDA method. In the model we can consider that $R(t)$ is the percentage of fatty acids from DNL which are 'labeled' at time t, and that $R(t)$ is dependent solely on how much of the acetate pool is labeled. Note that in actuality the MIDA method is a calculation to correct for the combinatorics of combining endogenous and exogenously derived acetyl-CoA to form lipids, however in the model we can identify the metabolites originating from de-novo lipids directly. To allow for the tracer kinetics, we calculate $R(t)$, which is at steady state until the acetate infusion is stopped, as

$$R = \frac{A_L(t)}{A_E(t) + A_L(t)}$$

$$R = \frac{1}{1 + \frac{k_E}{k_I} e^{\lambda(t-T)}}$$

where $A_L(t) = (k_I / \lambda) e^{-\lambda(t-T)}$ and $A_E(t) = k_E / \lambda$ are the labeled and endogenous acetate, respectively, and T is the time that infusion is stopped. Prior to time T we assume that R is at steady state $= \frac{K_I}{K_I + K_E}$. The endogenous rate of appearance of acetate (k_E) and its clearance (λ) are taken from literature values.

Parameter estimates

Parameter ranges, units, and final model fits are reported in the supplementary material. Here we detail our strategy for generating estimates for the parameter ranges, and fitting the model.

For equations (1)–(3) parameters β, γ, and α_1 were independently fit using data from an oral fructose test.[18] Fractional absorption of fructose, ρ, was estimated as 0.72, based on a reported 72% of oral fructose being up taken into the liver of fasted rats, and in humans at least half of intravenously delivered fructose being metabolized hepatically.[19] Here α_1 was estimated assuming a low basal concentration of plasma fructose of 51 μmol/L.[18]

Feasible ranges for the zeroth-order source terms in the model (α_i) were made as follows. Here α_1 was assumed to be known with high confidence (see above) and α_2, a source term for non-fructose derived pyruvate, was estimated by noting that rate of hepatic glucose output is approximately 0.8 mmol/min,[20] a significant portion of this will be from glycogen. Assuming that α_2 would be within a reasonable range of this value, we constrained α_2 to be within $[10^{-6}, 5 \times 0.8 \times V_L^F]$ mmol/L/min. The uptake rate of fatty acids, α_3, was estimated to be a maximum of 0.6 mmol/min,[21] similarly a wide including this value was confined: $[10^{-6}, 2 \times 0.6 / V_L^F]$ mmol/L/min. The basal source term for plasma TGs, α_4 was estimated from the estimate for clearance σ_2 (estimated from the fructose DNL experiment[9]), and reported values for plasma TG concentrations (1.2 mmol/L).[22] Total whole body fatty acid oxidation was estimated to be between $[0, 995]$ mg/kg/day as reported by Aarsland et al.,[23] depending on substrate availability. Given the unique nature of the protocol being simulated here, and the large variability in the reported values, this range was used to calculate the first-order rate constant (σ_2), by dividing by reported typical hepatic fatty acid levels.[24]

Apart from the Michaelis constants in the transfer functions (K_2, K_4, K_6), for which we describe a novel approach in Section 3, the remaining parameters of the DNL model were estimated using a bounded simulated annealing routine (simulannealingbnd[25]) which is an algorithm for bounded global optimization. For each parameter a range was estimated to pass to the routine.

If a range was not estimable from experimental data, a reasonably wide-range around the reported value was passed to the optimization routine: this range spanned one, two, or three orders of magnitude, the latter being assigned to the V_i parameters. This is representative of the large degree of uncertainty in their value.

Computational approach to estimate parameters of the transfer functions

In this section we describe a novel computational approach to estimate a value of K_2, K_4, and K_6 in the transfer functions (equations (12), (14), and (16)). These parameters have no direct biological interpretation because the transfer functions (12), (16), and (16) are lumped representations of multiple metabolic steps. Nevertheless, if there are data regarding the individual reactions then it is desirable to apply that data to inform the possible values of these parameters.

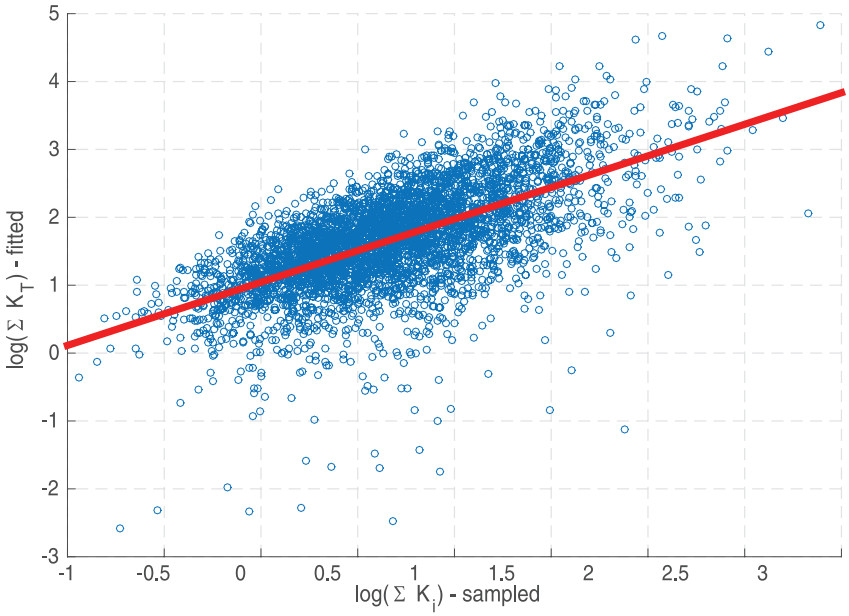

Figure 4. Predicting transfer function kinetic parameters from individual flux parameters. For a high stimulus level and given chain length ($n = 8$) $\log(\Sigma K_i)$ is linearly correlated ($r^2 = 0.52$) with $\log(K_T)$.

Below we show that with limited assumptions about the nature of the pathway and experimental conditions, a rule can be constructed to estimate the expected value, $E(K_T)$, of the transfer-function parameter K_T , based solely on the values of the K_i (the Michaelis–Menten constant of each enzymatic/metabolic step, see Figure 1). This rule is analogous to the well-known result that $V_T = \min(V_i)$, i.e. the maximum rate of the pathway is controlled by the rate-limiting step.

It is imperative to point out that the rule developed in this section only provides an estimate, or bounds (depending on prior knowledge), for $E(K_T)$ that is required to be refined based on fitting of experimental data (specifically data where input and output to the system are measured). Nevertheless, this rule provides a rationale for incorporating prior knowledge about the motif into the lumped parameter estimation. For the model introduced in Section 2 this method was used to estimate a plausible range for the fitted K_T. This range was then used in the fitting procedure as a bound on K_T. To the best of the authors' knowledge there was no other way to construct a relatively narrow, but physiologically relevant, range for K_T without this analysis.

Our strategy for approaching this problem was to simulate the full model in Figure 2, and attempt to fit the flux into the final node as function of the concentration of the first node (Figure 2, inset). In this manner we could attempt to establish a relationship between the set $\{K_i\}$ and K_T. Computational simulation (see Appendix A.2) illustrated a reproducible relationship between ΣK_i and K_T . Specifically this relationship is that $\log(E(K_T))$ is a linear function of $\log(\Sigma K_i)$, where $E(K_T)$ is the expected value of the transfer function (assuming that the V_i are unknown). Figure 4 is an example of this relationship for a chain of length 8. Note that this relationship provides an order-of-magnitude estimate for K_T which, although broad, is an improvement on the a priori range for human enzyme K_m (which can be estimated as ≈10 order of magnitudes, as assessed by the BRENDA tool[26]). This estimate provides a rationale for establishing a suitable, and computationally feasible, range to pass to a constrained optimization routine (or to evaluate model performance using unconstrained optimization). While this estimate may not lead to improved model fits, it will lead to a model where the parameter choices have improved experimental justification.

To investigate how robust this relationship is, we repeated this procedure for different lengths of chains ($n = 3$ to 24) and stimulus strengths (over five orders of magnitude). We noticed that the correlation between $\log(E(K_T))$ and $\log(\Sigma K_i)$ was obscured at lower stimulus levels (Figure 5A), but was largely independent of the length of the chain (Figure 5B and C). The y-intercept of the regression line is determined by the stimulus level (Figure 5C).

In the appendix we provide an analytical analysis of this problem, which provides a direct interpretation of why this relationship holds in a limiting case.

Simplification parameters for the hepatic DNL model

An unknown in applying this technique to the DNL model is the stimulus strength relative to the $E(V_i)$. Given the high levels of fructose loading in the clinical experiment,[9] and based on our analytical analysis, we assumed a range of [0,1] for the y-intercept (equation (21)). This corresponds to a range of stimulus of $S \approx E(V_i)$ (the stimulus is likely to be saturating some reactions on the pathway) to $S \approx 10E(V_i)$ (the stimulus is very likely to be saturating reactions on the pathway).

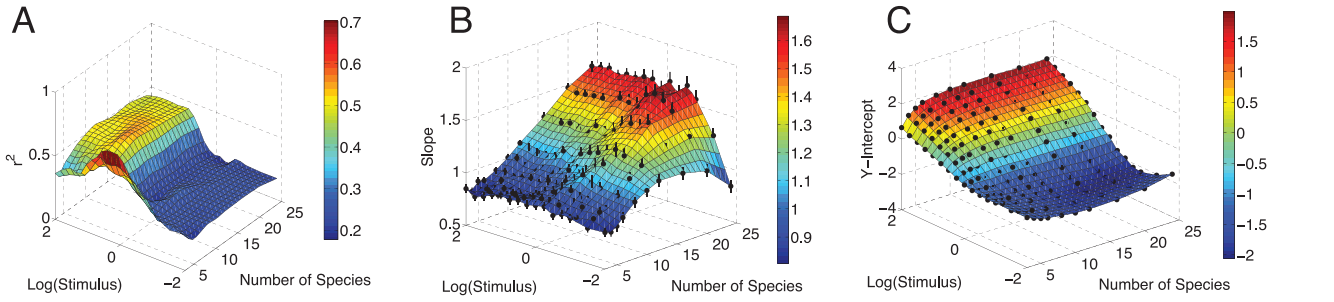

Figure 5. The effect of chain length and stimulus level on the relationship between K_T and ΣK_i, see Figure 4. A relationship and correlation between transfer function kinetic parameters from individual flux parameters is maintained over a wide range of chain-lengths and stimulus level. For weak stimuli the correlation is lost.

Application to and simulation of a model of hepatic DNL

To demonstrate this technique we applied it to the model of hepatic DNL presented in Figure 5. Note that this model represents a significant simplification of the 16 enzyme path (Figure 1). The individual reactions and estimates for their K_m are given in Supplementary Table 2. In this case, we have assumed that we can daisy chain this approximation together by applying this simplification three times. This allowed us to take account of the major sources and sinks in this pathway, and retain a suitable level of detail for future mechanistic questions. Applying the method above to this data gave a range of values for the K_T for each of the transfer functions in this model (Supplementary Table 3). The range for each K_i is generated by both experimental uncertainty and uncertainty estimating the stimulus level relative to $\{V_i\}$ in the pathway. We could have considered the additional uncertainty due to $\mathrm{Var}(K_T)$, given that equation (21), only provides an estimate for the expected value of K_T, but this was not necessary in this example. In this case we took the maximum value of the *y*-intercept in the relationship as being in the range [0,1], which in our prior simulations (Figure 5) corresponded to a range where the stimulus level ≈[0,1] times the average V_i in the pathway. This range represents mildly saturating to highly saturating conditions, which seemed reasonable based on observed experimental data. Since we only established a search range for the K_i parameters, and since there were still unconstrained parameters in the model (for example, the V_{max} parameters), it was necessary to fit the model against further experimental, dynamic, data. To do this we applied this model to previously published human data, where subjects were repeatedly fed fructose over an 8 hour time period and the appearance of de-novo lipids in blood was measured.[9] The model could be fit to the available data but not uniquely (Figure 6). This is typical for a model of this size applied to limited human data. However, this did test that estimating the K_i parameters of the transfer functions did not lead to constraints that were discordant with available dynamic data.

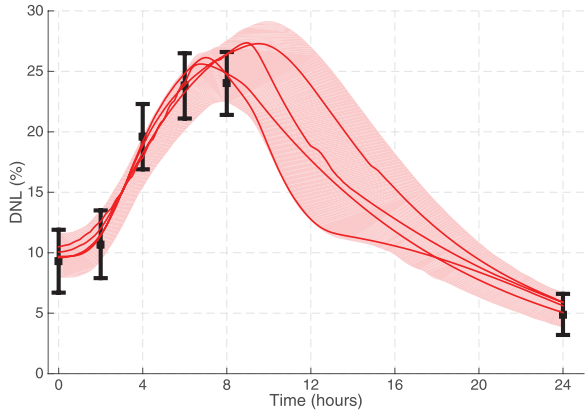

Figure 6. Summary of model performance compared with clinical data from Hudgins et al.9 (black points). Multiple fits (≈1800) were found that could describe the data. The upper-quartile of fits were found to cluster into four distinct groups (cluster centroid, solid red lines). The shaded region summarizes the upper and lower bounds of all fits with outliers removed (defined as the top and bottom percentiles).

Given the remaining uncertainty we captured multiple parameter sets (≈3000), so that we can propagate the uncertainty forward for future simulations (Supplementary Table 1).

Because novel inhibitors of acetyl-CoA carboxylase (ACC) have been developed,[27,28] we wanted to investigate the effect of inhibiting ACC in this model. This was simulated by re-running the model with a reduction of parameter V_4 by 90% and 50%. Recalling the simplification paradigm, this parameter is the $V_{\max}$ of the rate limiting step of the reactions between pyruvate and fatty acids (Figure 3). In this case, the rate-limiting step of this part of the pathway, fat synthesis, is ACC.[29] Hence, this modification represents inhibition of ACC.

Figure 7 shows the effect of ACC inhibition in the model. Inhibition of ACC is predicted to have a potent effect on fractional DNL as assayed by this clinical methodology. The model also predicts that in a clinical trial the variability in response will be higher for doses that induce 50% inhibition than placebo or higher doses (as shown by wider prediction intervals for the 50% case). This prediction illustrates

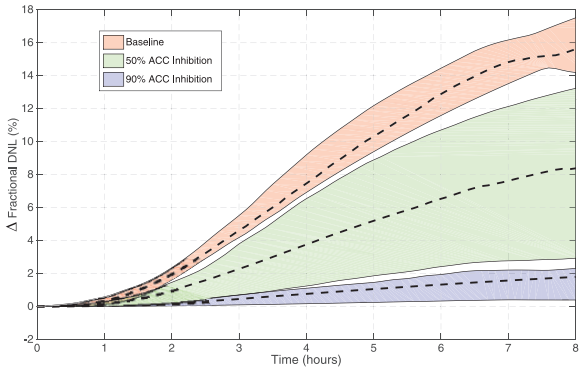

Figure 7. Evaluation of the model response to 50% and 90% inhibition of acetyl-CoA carboxylase (ACC) in comparison with the baseline response (shaded area is the 10–90% prediction interval). The model predicts that the fractional DNL response to fructose stimulation can be effectively modulated by inhibition of ACC. Variability in predicted response at 50% inhibition is significantly larger than the baseline and 90% case.

the utility in generating multiple parameter sets to explore variability in response, and also has implications for how such a study should be conducted. However, this effect could be masked by other sources of variability.

Since the expression of ACC can be tightly controlled,[30] it is interesting to interpret these results as the effect of modulating ACC expression to 50% or 10% of baseline (mathematically, these are the same simulations as Figure 7). In this interpretation the results supports the hypothesis that regulation of ACC by SREBP1c[30] is a crucial component of the lipogenic regulatory program controlled by SREBPC1 and regulation of ACC expression directly controls the contribution of DNL to the liver and plasma TG pools.

Discussion

Here we have developed a novel systems pharmacology model of hepatic DNL. In particular, this model is designed to support, and analyze, clinical experiments where hepatic DNL is measured. As such, this model could be a powerful tool to assess novel pharmacological agents that may affect DNL. For example it could be used to infer pharmacologically meaningful parameters such as IC50, or $I_{\max}$ from clinical data, as indicated in Figure 7.

As a focused simulation of a specific experimental protocol we have high confidence in the model when applied in this context, but in other contexts the model would require further validation and refinement before application. However, we believe that within this context the model is a good representation of the underlying biology because of the novel simplification approach we developed. This approach links underlying biochemical constants with a parameter of a lumped representation that hitherto had no biologically interpretable meaning.

The model we have presented here is an acute view of fructose-induced DNL and tracks a major metabolic route of fructose in these conditions. However, fructose metabolism also has a major signaling role which can alter downstream metabolic pathways. These signaling effects include activation of SREBP1c which has a profound effect on the entire lipogenic pathway.[30] Our assumption is that the acute nature of the study allows us to ignore these effects during the course of the simulated trial. However, future work could include analysis and modification of the model to describe patients from a different population, such as those with non-alcoholic fatty liver disease. This population consumes large amounts of dietary fructose and have a higher fasting DNL level. Such an analysis like could facilitate insight to the precise biochemical parameters that are inducing increased DNL in these patients, and how they might respond in this clinical protocol.

Systems pharmacology models are a useful tool to understand biological systems, disease pathophysiology, and the potential effects of pharmacological intervention. As more biological detail is added to a model it becomes less constrained and the balance between the number of parameters and available data becomes less favorable. Therefore it becomes necessary to simplify the model. However, there is a balance between redressing this while retaining enough mechanistic detail to satisfactorily address the biological questions and hypotheses of interest. In this case it is important that all available data is used to establish parameter values.

The model reduction technique applied here simplifies a model while utilizing available data about individual enzyme kinetics. This technique supplements existing approaches of constraining, or reducing, a model based on dynamic data. By bounding the search for lumped parameter values into a biologically and mathematically justified space, physiologically reasonable parameter sets can be identified. Our analysis and simulations show that the Michaelis constant of the transfer function is related to the sum of the individual Michaelis constants and the stimulus level relative to the distribution of V_i in the pathway. We have demonstrated the validity of this approach by analysis of a limiting discontinuous case, and simulation of many parameterizations of an example motif in a continuous case. This motif was based upon a chain of enzymes that can be described by Michaelis–Menten kinetics. Typically, and in the DNL example we give, the regulation of metabolic enzymes and their kinetics are far more complicated than the relatively simple Michaelis–Menten representation (for example, product inhibition, multiple-substrate reactions, co-operative binding). Given our analytical results we hypothesize that these results should be valid as long as the pathway is saturable, and unidirectional (or is so in the conditions of interest). Application of this approach to enzymatic chains with mixed kinetics is a future avenue of research.

We have focused on this method in order to develop rules for fitting parameters. However, there is a biological interpretation of our result which generates an interesting hypothesis for future testing. We can consider the transfer function as a summary of the kinetics of the pathway. As such, K_T is a measure of the sensitivity of the pathway to stimulus. If K_T is small, the

pathway is very sensitive and fully activated for small stimuli. Conversely, if K_T is large the pathway is not sensitive. Equation (21) is then interpreted as follows: the sensitivity of a pathway decreases (larger K_T) in both long reaction chains or in high stimulus conditions, which seems intuitively obvious. The inverse of this is that, according to this analysis, increased enzyme expression sensitizes a pathway. This analysis also indicates that therapeutic intervention designed to improve sensitivity to substrate should be focused on the 'sensitivity limiting step' (i.e. the reaction with the highest K_m). In the future we hope to test these concepts experimentally.

Acknowledgements

We would like to acknowledge Gianluca Nucci (Pfizer Inc.) for his support and motivation of this work. We would also like to thank Theodore Rieger (Pfizer Inc.), and William C. Thompson (formerly Pfizer Inc., currently SAS Institute Inc.) for valuable input and feedback on this work.

Author Contributions

RJA performed the analysis, wrote the manuscript, and approved the final manuscript. CJM reviewed and approved the final manuscript.

REFERENCES

1. Schutz Y. Concept of fat balance in human obesity revisited with particular reference to de novo lipogenesis *Int J Obesity* 2004;28:S3–S11.
2. Katsuhiko F, Haowei S, Li Y, et al. Muscle lipogenesis balances insulin sensitivity and strength through calcium signaling *J Clin Investig* 2013;123:1229–1240.
3. Fatima A, Lisa S, Shahida H, Hubert K, Nousheen Z. De novo lipogenesis in health and disease *Metabolism* 2014;63:895–902.
4. Ferré P, Foufelle F. Hepatic steatosis: a role for de novo lipogenesis and the transcription factor SREBP-1c. *Diabetes Obesity Metabolism* 2010;12(Suppl. 2): 83–92.
5. Leahy JL. Pathogenesis of type 2 diabetes mellitus. *Arch Med Res* 2005;36:197–209.
6. Hellerstein MK. De novo lipogenesis in humans: metabolic and regulatory aspects *Eur J Clin Nutrition* 1999;1:S53–65.
7. Jackson KG, Poppitt SD, Minihane AM. Postprandial lipemia and cardiovascular disease risk: Interrelationships between dietary, physiological and genetic determinants *Atherosclerosis* 2012;220:22–33.
8. Stanhope KL, Griffen SC, Bair BR, Swarbrick MM, Keim NL, Havel PJ. Twenty-four-hour endocrine and metabolic profiles following consumption of high-fructose corn syrup-, sucrose-, fructose-, and glucose-sweetened beverages with meals. *Amer J Clin Nutrition* 2008;87:1194–1203.
9. Hudgins LC, Parker TS, Levine DM, Hellerstein MK. A dual sugar challenge test for lipogenic sensitivity to dietary fructose *J Clin Endocrinol Metabolism* 2011;96:861–868.
10. Krüger R, Heinrich R. Model reduction and analysis of robustness for the Wnt/. BETA.-catenin signal transduction pathway *Genome Informatics* 2004;1:138–148.
11. Segel LA, Slemrod M. The quasi-steady-state assumption: a case study in perturbation *SIAM Rev* 2006;31:446–477.
12. Alon U. Simplicity in biology *Nature* 2007;446:497–497.
13. Sun C, Hahn J. Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis *Chem Eng Sci* 2006;61: 5393–5403.
14. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nature Cell Biol* 2006;8:1195–1203.
15. Ciliberto A, Capuani F, Tyson JJ. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. *PLoS Computat Biol* 2007;3:e45.
16. Allen RJ, Elston TC. From physics to pharmacology? *Rep Prog Phys* 2010;74:016601.
17. Karmakar S, Mandal B. Graph theoretical analysis on the kinetic rate equations of linear chain and cyclic reaction networks. *J Phys Chem A* 2014;118:7672–7682.
18. Chong MF-F, Fielding BA, Frayn KN. Mechanisms for the acute effect of fructose on postprandial lipemia. *Amer J Clin Nutrition* 2007;85:1511–1520.
19. Mayes PA. Intermediary metabolism of fructose. *Amer J Clin Nutrition* 1993;58:754S–765S.
20. Glauber H, Wallace P, Brechtel G. Effects of fasting on plasma glucose and prolonged tracer measurement of hepatic glucose output in NIDDM *Diabetes* 1987; 10:1187–1194.
21. Hallgreen CE. *The interplay between glucose and fat metabolism: A biosimulation approach.* PhD thesis, Technical University of Denmark, Copenhagen, 2009.
22. Laws A, Reaven GM. Evidence for an independent relationship between insulin resistance and fasting plasma HDL-cholesterol, triglyceride and insulin concentrations. *Acta Medica Scandinavica* 1992;231:25–30.
23. Aarsland A, Chinkes D, Wolfe RR. Hepatic and whole-body fat synthesis in humans during carbohydrate overfeeding. *Amer J Clin Nutrition* 1997;65:1774–1782.
24. Cairns S R, Kark A E, Peters T J. Raised hepatic free fatty acids in a patient with acute fatty liver after gastric surgery for morbid obesity. *J Clin Pathol* 1986;39:647–649.
25. MATLAB. MATLAB and Global Optimization Toolbox Release 2014b, 2014.
26. Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* 2002;30:47–49.
27. Griffith DA, Kung DW, Esler WP, et al. Decreasing the rate of metabolic ketone reduction in the discovery of a clinical acetyl-CoA carboxylase inhibitor for the treatment of diabetes. *J Med Chem* 2014;57:10512–10526.
28. Harriman G, Greenwood J, Bhat S, et al. Acetyl-CoA carboxylase inhibition by ND-630 reduces hepatic steatosis, improves insulin sensitivity, and modulates dyslipidemia in rats. *Proc Natl Acad Sci* 2016;113:E1796–E1805.
29. Donaldson WE. Regulation of fatty acid synthesis. *Federation Proc* 1979;38:2617–2621.
30. Softic S, Cohen DE, Kahn CR. Role of dietary fructose and hepatic de novo lipogenesis in fatty liver disease *Digest Dis Sci* 2016;61:1282–1293.

Appendix

Analytical interpretation

To understand our computational results from Section 3 we analyzed a simpler system, where the Hill functions of the original reaction chain are replaced with step functions (which is a limiting case of the example above). With this approximation we can provide analytical intuition for why the sum of the K_i can be translated to estimate K_T.

Consider a limiting case where the ith flux is a step function:

$$J_i = \begin{cases} V_i & \text{if } X_i \geqslant K_i \\ 0 & \text{otherwise} \end{cases}$$

Also write $V_0 = S$ for notational convenience in this analysis. Then

$$\dot{X}_i = J_{i-1} - J_i$$

Consider X_1. While $X_1 < K_1$, i.e. $t < K_1 / V_0$,

$$\dot{X}_1 = V_0$$

for $t \geqslant K_1 / V_0$

$$X_1 = K_1 + (t - K_1 / V_0)(V_0 - V_1)$$

If we assume $V_{i-1} \geqslant V_i$ for all i (see below for how to remove this assumption), then for $t > \sum_{i=1}^{i} K_i / V_{i-1}$

$$X_i = K_i + (t - \sum_{i=1}^{i} K_i / V_{i-1})(V_{i-1} - V_i)$$

Consider that $J_{n-1} = 0$ unless $X_{n-1} > K_{n-1}$. From the above, this occurs at $\hat{t} = \sum_{i=1}^{n-1} K_i / V_{i-1}$. Therefore, the transfer function from X_1 to X_n should be zero until time $\hat{t}$. From this we infer $K_T = X_1(\hat{t})$:

$$K_T = K_1 + (V_0 - V_1)\left(\sum_{i=2}^{n-1} K_i / V_{i-1}\right)$$

or

$$K_T = (V_0 - V_1)\left(\sum_{i=1}^{n-1} K_i / V_{i-1}\right) + V_1 K_1 / V_0 \qquad (18)$$

Note that, in general, the V_i will not be known a priori. But if we assume that the V_i are identical, then

$$K_T = \left(\frac{V_0}{V} - 1\right)\left(\sum_{i=1}^{n-1} K_i\right) + \frac{V K_1}{V_0} \qquad (19)$$

This expression provides the intuition and an analytical justification for the simulation results.

What about the assumption $V_{i-1} \geqslant V_i$? If we remove this restriction we have to consider carefully the flux out of the nodes where $V_i > V_{i-1}$. In this case, clearly the steady-state value of X_i, $\hat{X}_i$, is K_i, because if $X_i < K_i$, then $\dot{X}_i > 0$, and if $X_i \geqslant K_i$, then $\dot{X}_i \leqslant 0$.

Due to the discontinuity at $X_i = K_i$, at some time points $X_i = K_i - \varepsilon$ and others $X_i = K_i + \varepsilon$; the former giving $J_i = 0$, the latter $J_i = V_i$. Therefore, the average contribution of flux from X_i to X_{i+1} should be scaled by a dimensionless factor α_i representing the proportion of time where $X_i \geqslant K_i$. Consider the time interval $[t, t+T]$ where T is suitably large, then

$$\begin{aligned} &\int_t^{t+T} \dot{X}_i \, dt \approx 0 \\ &\Rightarrow \int_t^{t+T} \alpha_{i-1} V_{i-1} - \alpha_i V_i \, dt \approx 0 \\ &\Rightarrow \alpha_i \approx \alpha_{i-1} \frac{V_{i-1}}{V_i} \end{aligned}$$

Then the V_i can be rewritten as $V_i' = \alpha_i V_i$, which do meet the assumption required above, where

$$\alpha_i = \begin{cases} 1 & \text{if } \alpha_{i-1} V_{i-1} \geqslant V_i \\ \alpha_{i-1} V_{i-1} / V_i & \text{otherwise} \end{cases}$$

Assume that for small chains, the sample mean and variance are approximated by $E(V') \approx E(V)$, and $\mathrm{Var}(V') \approx \mathrm{Var}(V)$, then taking expected values of both sides of equation 18, and assuming independence between V_i and K_i gives

$$E(K_T) = \hat{S}\left(\sum_{i=1}^{n-1} K_i\right) + \frac{E(V) K_1}{V_0} \qquad (20)$$

where $\hat{S} = E(V_0 / V) - E(V) E(1/V)$. Taylor expand $E(1/V)$ around $V = E(V)$:

$$\begin{aligned} E\left(\frac{1}{V}\right) &\approx E\left(\frac{1}{E(V)} - \frac{(V - E(V))}{E(V)^2} + \frac{(V - E(V))^2}{E(V)^3}\right) \\ &= \frac{1}{E(V)} + \frac{\mathrm{Var}(V)}{E(V)^3} \end{aligned}$$

Applying this result to the expression above gives $\hat{S} \approx E(V_0 / V) - (1 + c_v^2)$, where c_v is the coefficient of variation of the V_i. For large stimulus $\hat{S}$ can be thought as the stimulus level normalized to the pathway capacity. The negative term can be interpreted as a measure of information loss down the pathway: as this term approaches $E(V_0 / V)$ the dependence of $E(K_T)$ on $\sum_{i=1}^{n-1} K_i$ is lost. While there is no data, to the best of the authors' knowledge, on typical c_v values, it is reasonable to assume that it is (relative to the first term) small. The consequences of it being large, is that under high flux conditions

some intermediary pools will become deposits for almost all the flux (those where the maximum in rate is very much faster than maximum out rate). We can try to bound S based on prior biological knowledge or experimental data (for example, it could be inferred by perturbing the K_i pharmacologically and use the shift of the K_T to infer $\hat{S}$ in those experimental conditions). Consider the last term to be negligible, then

$$\log(E(K_T)) = \log(\hat{S}) + \log\left(\sum_{i=1}^{n-1} K_i\right) \tag{21}$$

Even if $\hat{S}$ is not estimable, in the context of large stimulus and relatively small c_v (these conditions guarantee flux into the final node), we still gain the important heuristic:

$$E(K_T) > \sum_{i=1}^{n-1} K_i$$

Simulating linear chains motifs

Our method for simulating linear chains is the following. First, a new linear-chain model is automatically generated with n ordinary differential equations (ODEs) representing n species. The flux from the ith species, X_i to the next in the chain X_{i+1} is modeled as a Hill function:

$$J_i = \frac{V_i X_i}{X_i + K_i}$$

Then the complete system of equations is constructed as

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \begin{cases} S - J_i & \text{if } i = 1 \\ J_{i-1} & \text{if } i = n \\ J_{i-1} - J_i & \text{otherwise} \end{cases}$$

where S is an input signal to the system. Parameters (V_i and k_i) for this model were chosen from broad log-normal distributions ($V \sim log_{10}\mathcal{N}(\mu_V, \sigma_V^2)$, $K \sim log_{10}\mathcal{N}(\mu_K, \sigma_K^2)$, with $[\mu_V, \sigma_V] = [0.1, 0.5]$ and $[\mu_K, \sigma_K] = [-1, 1]$, the latter being estimated from BRENDA[26] and the former chosen for computational tractability). The strength of the input signal (S) and the length of the chain (n) were varied in predefined values $n \in \{3,4,5,6,8,10,12,16,20,24\}$. For each S and n, 4000 parameter sets were simulated using the MATLAB ODE15s routine[25] and the results were recorded. We then attempted to simplify the linear chain by only considering the first and last node, that is we tried to replicate the model results by fitting a two-node model to the data. More precisely, we identified K_T by minimizing the function

$$f(T(X_1; V_T, K_T); J_{i-1}) = \sum_{t \in t_{int}} \left| T - J_{i-1} \right|$$

where t_{int} is the set of simulation time points returned by the ODE15s routine. These time-points are adaptive to the integration size. Here T is the transfer function, defined by

$$T(X_1; V_T, K_T) = \frac{V_T X_1}{X_1 + K_T}$$

Since we know that the maximum flux from X_1 to X_n is defined by the rate-limiting step, we fix

$$V_T = \min\{V_i\} \tag{22}$$

Then K_T is found by numerically solving

$$K_T = \arg\min_{K_T} f(T(X_1; V_T, K_T); J_{i-1})$$

This problem was solved using MATLAB's nonlincon function with K_T bounded, the upper-bound (which was necessary for speed) was found empirically.

The resulting collection of fits at the kth stimulus level (S^k) and jth chain length (n^j) were analyzed by fitting the equation

$$\log(K_T^{kj}) = m^{kj} \log\left(\sum_i K_i^{kj}\right) + c^{kj}$$

The coefficient of determination (r^2) was recorded for each k and j. The rationale for this analysis being the analytical results above.