



Published in final edited form as:

J Proteome Res. 2016 July 01; 15(7): 2115–2122. doi:10.1021/acs.jproteome.5b00990.

Gaussian Process Modeling of Protein Turnover

Mahbubur Rahman^{†,‡}, Stephen F. Previs[§], Takhar Kasumov^{||,⊥}, and Rovshan G. Sadygov^{*,†,‡}

[†]Department of Biochemistry and Molecular Biology, The University of Texas Medical Branch, Galveston, Texas 77555, United States

[‡]Sealy Center for Molecular Medicine, The University of Texas Medical Branch, Galveston, Texas 77555, United States

[§]Merck Research Laboratories 2015 Galloping Hill Road Kenilworth, New Jersey 07033, United States

^{||}Department of Gastroenterology & Hepatology, Cleveland Clinic Cleveland, Ohio 44195, United States

[⊥]Department of Pharmaceutical Sciences School of Pharmacy, Northeast Ohio Medical University Rootstown, Ohio 44225, United States

Abstract

We describe a stochastic model to compute in vivo protein turnover rate constants from stable-isotope labeling and high-throughput liquid chromatography–mass spectrometry experiments. We show that the often-used one- and two-compartment nonstochastic models allow explicit solutions from the corresponding stochastic differential equations. The resulting stochastic process is a Gaussian processes with Ornstein–Uhlenbeck covariance matrix. We applied the stochastic model to a large-scale data set from ¹⁵N labeling and compared its performance metrics with those of the nonstochastic curve fitting. The comparison showed that for more than 99% of proteins, the stochastic model produced better fits to the experimental data (based on residual sum of squares). The model was used for extracting protein-decay rate constants from mouse brain (slow turnover) and liver (fast turnover) samples. We found that the most affected (compared to two-exponent

*Corresponding Author: Tel 409-772-3287; fax 409-772-9670; rovshan.sadygov@utmb.edu.

Notes

The authors declare no competing financial interest.

Supporting Information

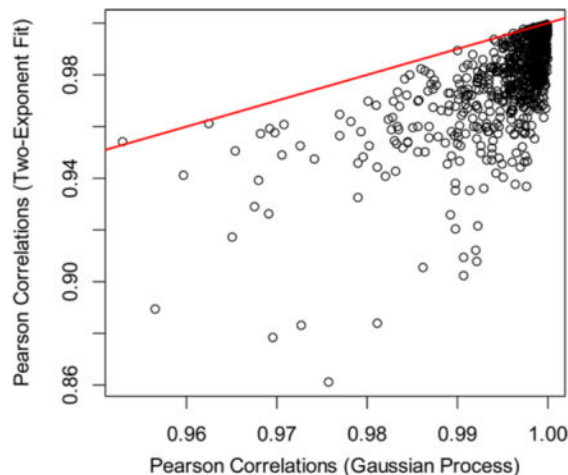
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteo-me.5b00990. Information on the derivation of the GP model from the stochastic differential equation of protein turnover kinetics. Figures showing one- and two-compartment models of protein turnover; a pairwise scatter plot of the RSS for mouse brain proteins; the pairwise scatter plots of the Pearson correlations obtained from GP and two-exponent curve fitting for mouse liver and brain proteins, respectively; 95% confidence interval of the mean of GP model for mitochondrial trifunctional protein, subunit α (Q8BMS1); the scatter plot of the decay-rate constants obtained from the GP and two-exponent curve fitting for mouse brain proteins; the density of the difference between degradation rate constants computed by the GP and two-exponent curve fitting; the boxplots of decay-rate constants computed by two-exponent curve fitting and GP model for mouse brain and liver proteomes; and the density of standard deviations of the model distributions for mouse liver proteins. (PDF)

A table showing synthesis and degradation rate constants as computed by GP and two-exponent curve fit for mouse liver proteins. (XLSX)

A table showing synthesis and degradation rate constants as computed by GP and two-exponent curve fit for mouse brain proteins. (XLSX)

curve fitting) results were those for liver proteins. The ratio of the median of degradation rate constants of liver proteins to those of brain proteins increased 4-fold in stochastic modeling compared to the two-exponent fitting. Stochastic modeling predicted stronger differences of protein turnover processes between mouse liver and brain than previously estimated. The model is independent of the labeling isotope. To show this, we also applied the model to protein turnover studied in induced heart failure in rats, in which metabolic labeling was achieved by administering heavy water. No changes in the model were necessary for adapting to heavy-water labeling. The approach has been implemented in a freely available R code.

Graphical Abstract



Keywords

dynamic proteome; protein degradation rate constant; protein turnover rate constant; Ornstein–Uhlenbeck process; Gaussian process; stochastic differential equation for protein turnover rate constant; mass spectrometry; stable isotope labeling

INTRODUCTION

The continuous degradation and synthesis of proteins plays an important role in maintaining cellular homeostasis.¹ The dynamic equilibrium of cellular protein abundances can change due to, for example, external stimuli, developmental programs, onset of diseases, or aging. The new equilibrium is achieved by balancing between alterations in the rate of synthesis and the rate of degradation. The combined process of synthesis and degradation termed as protein turnover refers to the movement of a protein through its pool and is often expressed as a first-order rate constant.² It is thought that protein degradation is a stochastic process; all proteins (newly synthesized and old) have the same probability of being degraded.²

The labeling of proteins with radioactive amino acids is considered to be a “golden standard” for studying proteome dynamics. In this “pulse-chase” method, proteins of interest are radioactively labeled over a brief period of time (the pulse), and then the decay in radioactivity is measured over time (the chase) and is used to determine the protein half-

life.³ This assay is accurate and minimally perturbative but requires a specific antibody for each protein, which makes it difficult for scaling to multiple proteins. Fluorophore tagging of proteins have also been used for proteome dynamics studies.^{4,5} Although these methods are real-time, it has been argued that the tagging may affect protein turnover properties, both through changing the protein structure and by altering their abundance and stoichiometry relative to interaction partners.⁶

Mass-spectrometry-based proteomics is a widely used, high-throughput technology to identify and quantify proteins.⁷ Most of the quantitative proteomics applications so far have been used for measuring the net change in protein abundance.⁸ The technology has recently been enhanced and applied⁹ in studies aimed at measuring the rate of protein turnover, both in vivo^{10–15} and in vitro.^{16–19} Protein turnover studies using metabolic labeling and liquid chromatography–mass spectrometry (LC–MS) analyses have a complex design that includes a critical bioinformatics component to extract protein-decay rate constants.²

The work discussed herein focuses on the application of a stochastic model to extract protein turnover rate constants from in vivo protein turnover studies. Briefly, the metabolic incorporation of stable isotope and LC–MS measurements of time course peptide and protein labeling provide data on protein kinetics. Traditionally, exponential decay curves have been used to fit the experimental data and infer the protein turnover rate constants.^{11,18,20,21} Single-compartment (single exponential) and multicompartiment (first-order synthesis and multiexponential)^{21,22} models have been employed. Time-dependent ordinary differential equations (ODE) have also been used to model the stable isotope incorporation in metabolic labeling.^{22,23} However, the general applicability of these models to in vivo studies are limited because they may result in negative decay rates (for multicompartiment systems) and are not able to explain fluctuations in the experimental data.²¹

Stochastic models were recently applied to high-throughput time-course data in proteome studies.^{24,25} SORAD²⁴ used a Gaussian process (GP) to describe time-course data of phosphor proteins from enzyme-linked immunosorbent assay experiments. Similar models have been used for chromatographic alignment²⁵ and beta-binomial process modeling of high-throughput sequencing.²⁶ The stochastic processes add the flexibility of accounting for fluctuations in the data. Here, we describe a nonparametric, data-driven stochastic approach, GP, for modeling time-course data from metabolic labeling experiments. We show that the GP with the Ornstein–Uhlenbeck (OU) kernel is a natural solution to the protein turnover kinetics in one- or two-compartment models. This approach produces fits that are highly improved compared to the traditional curve-fitting models; it overcomes problems with negative decay rates and random fluctuations in decay rates. Last, the model is isotope-independent (i.e., ¹⁵N or ²H labeling). The R code implementing the model is available at <https://ispace.utmb.edu/users/rgsadygo/Proteomics/ProteinKinetics>.

METHODS

Protein turnover studies using metabolic labeling, mass spectrometry and bioinformatics have a complex design and comprise multiple stages.²⁷ There are some variations dependent

on what type of labeling (e.g., ^{15}N , ^2H , or ^{13}C) is used and how a tracer is administered. For, e.g., studies that rely on heavy-water labeling, the animals are provided heavy water using a “primed-infusion” logic. For example, animals are given an initial loading bolus to instantaneously raise the enrichment of body water to a desired level. They are simultaneously given labeled water in the drinking bottles to maintain the enrichment. At predetermined time points, the animals are sacrificed, and the organs or tissues are collected. LC–MS/MS is used to identify proteins and measure the level of incorporation of heavy isotopes.^{28–30} The time-course data of the heavy isotope incorporations is then modeled to extract the protein turnover rate constant (Figure 1). The last stage is the focus of this study. In the following text, we will first review the current models and then describe our stochastic model to extract protein turnover rate constants from heavy-isotope labeling data.

Ordinary differential equations have previously been used to model the incorporation of the heavy-isotope-labeled species in metabolic labeling and LC–MS.^{21,23} As a prototypical study, Guan and colleagues²¹ have used one- and two-compartment models to determine protein turnover rate constants from experiments using an ^{15}N -labeled algae diet (produced from ^{15}N -enriched NaNO_3) and mass spectrometry. In brief, the one-compartment model assumes that proteins are synthesized at a constant rate, k_{syn} , and degraded proportional to their abundance levels with the degradation rate constant of k_{deg} . With the additional assumption of the steady-state for a protein concentration, it was shown that the fraction of the heavy-isotope-labeled proteins, $\beta(t)$ (characterized by the relative isotope fraction, RIF), is described by a first-order deterministic equation:

$$\dot{\beta}(t) = k_{\text{deg}} - k_{\text{deg}}\beta(t) \quad (1)$$

where t is the time, and the initial value condition is $\beta(0) = 0$ (the baseline, before the tracer administration). The solution of this equation is a growth curve with a single exponent:

$$\beta(t) = 1 - e^{-tk_{\text{deg}}}$$

The details of all formula derivations are provided in the Supporting Information. We note that under the steady-state assumption (which is used in this work), the protein degradation rate constant is equal to the fractional protein synthesis rate (ratio of the synthesis rate over the protein concentration). Thus, throughout the paper, we refer to the protein turnover rate constant and degradation rate constant interchangeably.

Several other publications^{11,12,18,20,31} modeling protein turnover from stable isotope labeling experiments have used equations similar to eq 1. However, when applied to the data sets from mouse protein turnover studies that used ^{15}N labeling, the fit to the data was often complicated (in particular, for fast protein turnover).²¹ Better results were achieved by using a two-compartment model. In this model, the first compartment considers the kinetics of amino acids, and the second compartment accounts for the amino acids incorporation into proteins. Protein synthesis is assumed to be a first-order process. It is modeled to depend on

the concentrations of heavy-isotope-labeled amino acids (precursor pool). The model leads to a system of equations for the fraction of heavy-isotope-labeled amino acids, $\alpha(t)$ and $\beta(t)$:

$$\begin{cases} \dot{\alpha}(t) = k_{\text{syn}} - k_{\text{syn}}\alpha(t) \\ \dot{\beta}(t) = k_{\text{deg}}\alpha(t) - k_{\text{deg}}\beta(t) \end{cases} \quad (2)$$

where k_{syn} is the rate constant of protein synthesis from the precursor amino acids. Eq 2 was derived under the assumption of the steady-state for the total (labeled and unlabeled) protein level. The amino acids are synthesized with the constant rate, k_0 , and degraded proportional to their abundances with the rate constant, k_{syn} . Protein degradation is characterized with the rate constant, k_{deg} . The equation for $\alpha(t)$ is solved first, with the initial value for stable-isotope-labeled amino acids, $\alpha(0)$, set equal to zero:

$$\alpha(t) = 1 - e^{-tk_{\text{syn}}}$$

which leads to the following equation for the $\beta(t)$:

$$\dot{\beta}(t) = k_{\text{deg}}(1 - e^{-tk_{\text{syn}}}) - k_{\text{deg}}\beta(t) \quad (3)$$

The solution of eq 3 ($\beta(0) = 0$) is the two-exponent growth curve:

$$\beta(t) = 1 - (k_{\text{syn}}e^{-tk_{\text{deg}}} - k_{\text{deg}}e^{-k_{\text{syn}}t}) / (k_{\text{syn}} - k_{\text{deg}})$$

The solutions for $\alpha(t)$ and $\beta(t)$ were obtained with the assumption that the heavy-isotope labeling is asymptotically complete. This assumption, while somewhat reasonable for incorporation of ^{15}N -labeled amino acids into proteins, will not be generalizable for incorporation of deuteriums during the administration of heavy water. This is because, in contrast to the fully ^{15}N -labeled diet, the metabolic labeling with heavy water involves administration of only ~5% $^2\text{H}_2\text{O}$ in the drinking water. Metabolic labeling with 100% ^{15}N in the diet may lead to the complete labeling of all nitrogen atoms in a protein, but using 5% heavy water will only result in partial labeling of the proteins. The incomplete labeling was also observed⁶ in labeling by SILAC.³² It was observed that unless the cell medium is regularly renewed by labeled amino acids, complete labeling of proteins will not be achieved. In animal experiments, it is not feasible to renew the cellular environment.

To account for the incomplete labeling, we introduced a new parameter, the heavy-isotope-enrichment fraction at the asymptote, λ . With this parameter, the solution for $\beta(t)$ becomes

$$\beta(t) = \lambda - \lambda(k_{\text{syn}}e^{-tk_{\text{deg}}} - k_{\text{deg}}e^{-k_{\text{syn}}t}) / (k_{\text{syn}} - k_{\text{deg}}) \quad (4)$$

A more descriptive text on the one- and two-compartment models and their solutions is provided in the Supporting Information. Even though the two-compartment model produced significantly improved fits compared to the experimental data, it was observed that the model was not able to account for the stochastic fluctuations in the experimental data.²¹ It has subsequently lead to high synthesis rate constants in the two-compartment model and negative decay rate constants for the expanded, three-compartment model. Here, we will show that stochastic differential equations based on either eq 1 or eq 2 allow the exact solution in the form of a GP with OU kernel. The stochastic equation uses a model fluctuation parameter that allows one to account for and estimate the effects of stochastic fluctuations in experimental data.

The stochastic analog of eq 3 for $\beta(t)$ (with the inclusion of the incomplete isotope labeling at the asymptote) is

$$dX(t) = k_{\text{deg}}(\lambda(1 - e^{-tk_{\text{syn}}}) - X(t)) dt + \Gamma(k_{\text{deg}}, k_{\text{syn}}, t, X, \lambda, \sigma_{\gamma}) dB(t) \quad X(0) = 0 \quad (5)$$

where $X(t)$ is a stochastic variable denoting the fraction of the heavy-isotope-enriched protein ($\beta(t)$ in the deterministic model), $\Gamma(k_{\text{deg}}, k_{\text{syn}}, t, X, \lambda, \text{ and } \sigma_{\gamma})$ is the volatility, which, in general, is dependent on the model parameters k_{syn} , k_{deg} , λ , and $X(t)$, and the variance of the model fluctuations, σ_{γ} ; $B(t)$ is the Brownian motion (white Gaussian noise). If we assume that, in addition to the model fluctuations, there are also measurement errors that are described by a white Gaussian noise, ε (with standard deviation σ_{ε}), the final model is

$$Y(t) = X(t) + \varepsilon \quad \varepsilon \sim N(0, \sigma_{\varepsilon}^2)$$

The model depends on the stochastic variable, $X(t)$, which is the solution to eq 5. We assume a homoscedastic variance, which allows for an explicit solution to eq 5. Thus, if the volatility, Γ , depends only on the model fluctuations, σ_{γ} , then eq 5 becomes

$$dX(t) = k_{\text{deg}}(\lambda(1 - e^{-tk_{\text{syn}}}) - X(t)) dt + \sigma_{\gamma} dB(t) \quad (6)$$

Eq 6 can be solved (using the assumption of Itô's process^{33,34} for $X(t)$ and the transformation $g(X, t) = X(t) \cdot \exp(k_{\text{deg}} t)$. The details of the derivation are provided in the Supporting Information). The solution of eq 6 is

$$X(t) = \lambda - \lambda(k_{\text{syn}} e^{-tk_{\text{deg}}} - k_{\text{deg}} e^{-k_{\text{syn}} t}) / (k_{\text{syn}} - k_{\text{deg}}) + \sigma_{\gamma} \int_0^t e^{-k_{\text{deg}}(t-s)} dB(s) \quad (7)$$

Note that the solution, (7), is the OU process^{34,35} with a mean value, $\beta(t)$, in eq 4 and covariance matrix, $K(t,s)$ (also called Gram matrix), expressed as

$$K(s,t) = \sigma_\gamma^2 e^{(-|s-t|k_{\text{deg}})} / k_{\text{deg}}$$

where t and s are time points. The result allows the direct application of the OU process to the proteome turnover dynamics. Because the solution is exact, it is not only methodologically preferable but also offers practical advantages to the Gaussian kernel model that was previously used for the first-order equation;²⁴ this stems from the fact that for the exact solution, there is one less parameter. The scaling factor in the Gaussian kernel is an independent parameter that needs to be estimated from the data. In the exact solution, the scaling parameter in OU kernel is the degradation rate constant, k_{deg} . In addition, it is natural that for the processes that behave exponentially, eq 7, the covariance terms between the time points will also be exponential (depending on the absolute value of the time difference) rather than Gaussian. Thus, our final model for the heavy-isotope-labeled protein proportions becomes

$$\vec{Y} \sim \text{MVN}(\vec{\mu}, \Sigma) \quad \vec{Y}, \vec{\mu} \in \mathbb{R}^n \quad (8)$$

where MVN stands for multivariate normal distribution, n is the number of data points (time points at which heavy isotope levels have been measured), μ is equal to β , which is determined in eq 4, and the covariance matrix, Σ is an n by n matrix defined as

$$\Sigma(s,t) = K(s,t) + \sigma_\varepsilon^2 \delta(s,t) \quad (9)$$

where δ is the Kronecker's δ . The model (eq 8) with the covariance matrix (eq 9) is a GP with OU kernel, and numerical methods exist^{36,37} for its solution to obtain k_{deg} and k_{syn} . The approaches to extract the parameters from the data use the posterior distributions or likelihood function and find parameter values that maximize them. In total, there are five hyperparameters in our model, $\theta = (k_{\text{deg}}, k_{\text{syn}}, \sigma_\gamma, \sigma_\varepsilon, \lambda)$. The log-likelihood function is

$$\log(\mathcal{L}(\theta)) = -0.5 \log(\det(\Sigma)) - 0.5 (\vec{y} - \vec{\mu})^T \Sigma^{-1} (\vec{y} - \vec{\mu}) - 0.5n \log(2\pi) \quad (10)$$

The posterior distribution, $p(\theta|y)$, of the parameters is proportional to the product of the likelihood function and the prior distributions, $\pi(\theta)$. The log of the posterior distribution of the hyperparameters is

$$\log(p(\theta|y)) \propto \log(\mathcal{L}(\theta)) + \log(\pi(\theta))$$

The hyperparameters can be estimated from the data either by maximizing the posterior distribution with respect to the hyperparameters³⁶ or by sampling from the posterior distribution with Markov chain Monte Carlo methods.²⁵ We used a quasi-Newton approach, a Broyden–Fletcher–Goldfarb–Shanno algorithm³⁸ available in R,³⁹ to maximize the posterior distribution function and obtain the hyperparameters. For the prior distributions (except for λ), we chose exponential distributions with the prior decay rates estimated from the nonstochastic model. A uniform prior was chosen for the asymptotic enrichment level, λ . Note that the derivative of the log-likelihood function, eq 10, with respect to the every hyperparameter is in a closed form and is used in the optimizations. The corresponding formulas are shown in the Supporting Information.

The approach is implemented in an R code³⁹ and is available in the Supporting Information.

RESULTS

Performance Metrics Using ¹⁵N Labeling Data

The protein turnover experiments using ¹⁵N labeling were used to test and compare the performance of the stochastic model with the results from exponential curve fitting. The ¹⁵N labeling data is freely available online.²¹ The data consists of two parts, a brain and a liver proteome. For every protein, RIF values of all peptides (for the protein) have been averaged to obtain protein RIF. We used these values to test our model. On average, protein turnover rates in brain are longer than those in the liver. For the nonstochastic model, the fitting of liver data has been more difficult and required three exponential curves, some of which produced negative rate constants.²¹ As a goodness of fit metric, we chose the residual sum of squares (RSS) for each model:

$$\text{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where y_i are the observed values and \hat{y}_i are the corresponding theoretical predictions. In Figure 2, we show the comparison of the RSS values between the stochastic modeling (x axis) and the two-exponent curve fitting (y -axis) for the mouse liver proteome (the red line is the line of unity). For the data points above the line of unity, the RSSs of the two-exponent curves are larger than the corresponding RSSs from GP. The fits obtained by the GPs resulted in smaller RSS values for 794 out of 797 proteins. Unlike the exponential curve fitting, GP allows for correlations between the measured RIFs in the time-course data. It is a desirable feature because the data points that are nearby in the time domain are expected to be correlated, and ideally, a model will be able to account for the correlation. In the case of the GP, the time correlations are provided in the structure of the Gram matrix. The diagonal elements of the matrix are the sum of the white-noise variance and amplitude of the OU kernel. The nondiagonal elements of the matrix are the covariance between the different time points. The corresponding figure for brain proteome is shown in Figure S1. Another

goodness-of-fit measure in time course models is the Pearson correlation between the predicted values and experimental observations along the time axis. In Figure S2 (mouse liver) and Figure S3 (mouse brain), we show the scatter plot of the Pearson correlation coefficients for each protein obtained using predictions from two-exponent fitting and GP modeling. Similar to RSS, GP showed substantially improved correlation coefficients compared to the two-exponent fitting.

To explicitly compare the fits from the two models, we have chosen a liver mitochondrial protein, trifunctional protein subunit α (Q8BMS1). The experimental data (empty circles) and fits (green line, GP; blue line, exponential curves) are shown in Figure 3. The GP results in a better fit (an approximately 70 times smaller RSS value compared to the two-compartment exponential curve fit). The k_{syn} rate constant computed in the exponential curve fit is very large, in the order of 10^{12} (day^{-1}), and the GP produced 1.4 day^{-1} . The degradation rate constant, k_{deg} , in GP was 0.15 day^{-1} , which was twice higher than the corresponding value from the two-exponent curve fit. The 95% confidence interval of the GP predicted fit is shown in Figure S4.

Comparison with the ODE Results Using ^{15}N Labeling Data

The scatter plot of the computed decay rate constants using a GP model and two-exponent curve fitting for mouse liver proteins is shown in Figure 4. The relevant data for mouse brain are shown in Figure S5. In general, the GP model produced decay rate constants that are larger than those from the two-exponent curve fitting. The correlation between the rate constants was 0.81. The density plot of the differences between k_{deg} values produced by the GP and two-exponent curve fit are shown in Figure S6. The distribution peaks at about 0.18 day^{-1} and extends up to 1 day^{-1} . For liver proteins, the median degradation rate constants were 0.091 and 0.228 day^{-1} for the two-exponent curve fit and GP, respectively. Thus, GP modeling, on average, predicts more than 2-fold faster degradation rate constants for liver proteins. For the mouse brain proteins, the effects were not so drastic. The median rate constants for brain proteins were 0.040 and 0.054 day^{-1} in the two-exponent curve fit and GP modeling, respectively. It has been noted previously that the exponent curve-fitting models were less accurate for modeling faster protein turnover (liver proteins).²¹ The ratio of the median turnover rate constants of liver proteins to brain proteins increased from 2.25 in two-exponent fitting to 4.23 in GP modeling. Thus, the GP modeling predicts a larger difference between the protein turnovers in liver and brain. The ratio of the medians of turnover rate constants predicted by the GP modeling compares well to earlier studies that estimated the rate of incorporation of [^{14}C] tyrosine in mouse liver (2.8% per hour) and brain (0.6% per hour).⁴⁰ Note that in these studies, the calculation of the rate constants are based on the tracer incorporation rate and the steady-state assumption, i.e., the fractional synthesis rate of proteins are equal to the fractional degradation rate. Under the steady-state assumption, the degradation rate constant is linearly dependent on the synthesis rate, as can be seen from the one-compartment model (solution of eq 1b in the Supporting Information). Figure S7 summarizes decay rate constants of brain and liver proteomes computed using two-exponent curve fitting and GP modeling.

Stronger differences were observed in cases where one estimates the synthesis rate constants, k_{syn} (Figure 5). Unlike the two-exponent curve fit, there are no k_{syn} values by the GP that are larger than 10 day^{-1} . For the two-exponent curve fitting k_{syn} , values as large as 10^{14} day^{-1} have been produced. The median values of k_{syn} for the two-exponent curve and GP were 9.12×10 and 1 day^{-1} , respectively. We note that in the labeling experiments where the first isotope distribution pattern is measured after about 9 h of metabolic labeling, the accurate determination of the rate constants of this much higher magnitude are unrealistic. The GP, however, yielded a maximum synthesis rate constant less than 10 day^{-1} . In addition, as it is shown in Figures 2 and S1–S3, the goodness of fit is substantially improved for the GP. The synthesis and degradation rate constants as computed by GP, and two-exponent curve fittings are provided in Tables S1 and S2 for liver and brain proteins, respectively.

The stochastic model produces a volatility standard deviation, σ_{γ} , for each protein rate constant estimate. The higher volatility standard deviation means a stronger random fluctuation effects. In Figure S8, we show the distribution of model standard deviations for the liver proteins. The mode of the standard deviation is nonzero at 0.018, and the highest values are at 0.08.

As was mentioned above, one way to improve the model fitting to the data has been to add more compartments. However, we believe that one should consider the development of a rationale solution and interpretation, e.g., make sure the modeling is done properly (as Guan et al.²¹ have shown) but that it is supported by some strong physiological underpinnings. It would appear that the addition of more compartments solved one problem (i.e., gave a better fit of the data) but failed to address what these compartments would represent from a biochemical perspective.

Our approach simply considers another way to align the biology and the modeling. We suspect that the approach described herein provides a better fit than previous ODE methods, in part because of instability in the ^{15}N -precursor precursor labeling (note that a major assumption is a constant precursor labeling). For example, if the amino-acid enrichment is constant, one would expect that the use of an ODE method should produce a good fit of the apparent protein-decay rate constant. Presumably, the use of ^{15}N -labeling (via the diet) results in pulsatile (or oscillatory) changes in the ^{15}N -labeling of proteogenic amino acids. One would expect a maximum ^{15}N -enrichment immediately following the consumption of the food and a lower degree of labeling at times when food is not consumed (given that during those periods, ^{14}N would be appearing via endogenous protein degradation). In addition, one needs to consider that there are gradients between plasma amino acid labeling and the labeling of free amino acids in different tissues. Effectively, the “free amino acid” compartment is further compartmentalized, and the “mixing” between different tissues and variable plasma labeling is not necessarily equal across sites. We suspect that this is what led Guan and colleagues²¹ to use models of increasing complexity when they fit the tissue-specific protein labeling. The observations would be expected to play a greater role in cases where protein turnover is faster (e.g., liver versus brain) and therein requires more complex models (e.g., three- versus two-compartment) to get a better fit.

Kinetics of Heart Mitochondrial Carnitine Palmitoyltransferase 1

In a recent study,⁴¹ we used heavy-water labeling to determine the changes in half-lives of mitochondrial proteins in cardiomyocytes due to the induced heart failure in rats. One of the proteins that we studied was carnitine palmitoyltransferase (CPT) 1. The experimental data (crosses, normal heart; triangles, induced heart failure) and corresponding GP fits (black line, normal heart; blue line, induced heart failure) are shown in Figure 6 (the data is for the CPT 1 from subsarcolemmal mitochondria (SSM) of cardiomyocytes). The computed turnover rate constants for the normal and induced heart failure samples were practically the same and equal to 0.09 day^{-1} . The result is in agreement with the finding that turnover rates for CPT 1 in SSM were not affected by the induced heart failure. The k_{syn} rate constant was substantially larger in heart-failure-induced SSM CPT 1 than in the protein of the normal heart: 2.2 versus 0.09 day^{-1} .

CONCLUSIONS AND FUTURE WORK

We have applied GP modeling to extract protein turnover rates from the time-course data of isotope incorporation in stable-isotope labeling and LC-MS experiments. It was shown that an exact solution to the stochastic differential equation yields a GP solution with OU kernel. The approach is preferable to the more widely used GP with the Gaussian kernel because the OU solution is exact and the number of parameters is less. We then used this approach to analyze a large-scale data set obtained from mice fed a diet containing ^{15}N -labeled algae. The results have demonstrated that GP modeling substantially improves the fit to the experimental data, particularly for protein turnover in liver, which in a previous model produced negative rate constants. We also applied the model to the carnitine palmitoyltransferase of SSM of cardiomyocytes that were studied to infer the effects of the induced heart failure. This experiment had used heavy-water labeling.

In future studies, we plan to extend the model to include relative isotope fractions for each peptide of a protein. A similar approach has employed a mixed-effects modeling for growth curves resulting from the Gompertz model.⁴² Using each point of peptide data separately rather than averaging the peptide data will allow us to estimate the variability due to the heavy-isotope labeling of peptides. In addition, we will extend the model to describe the turnover of proteins that are not in a steady-state condition.⁴³

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under award number R01GM112044.

ABBREVIATIONS

GP	Gaussian process
SSM	subsarcolemmal mitochondria

LC-MS	liquid-chromatography–mass spectrometry
ODE	ordinary differential equation
OU	Ornstein–Uhlenbeck

References

- Hinkson IV, Elias JE. The dynamic state of protein turnover: It's about time. *Trends Cell Biol.* 2011; 21:293–303. [PubMed: 21474317]
- Claydon AJ, Beynon R. Proteome dynamics: revisiting turnover with a global perspective. *Mol Cell Proteomics.* 2012; 11:1551–1565. [PubMed: 23125033]
- Schimke RT, Doyle D. Control of enzyme levels in animal tissues. *Annu Rev Biochem.* 1970; 39:929–976. [PubMed: 4394639]
- Yen HC, Xu Q, Chou DM, Zhao Z, Elledge SJ. Global protein stability profiling in mammalian cells. *Science.* 2008; 322:918–923. [PubMed: 18988847]
- Eden E, Geva-Zatorsky N, Issaeva I, Cohen A, Dekel E, Danon T, Cohen L, Mayo A, Alon U. Proteome half-life dynamics in living human cells. *Science.* 2011; 331:764–768. [PubMed: 21233346]
- Boisvert FM, Ahmad Y, Gierlinski M, Charriere F, Lamont D, Scott M, Barton G, Lamond AI. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol Cell Proteomics.* 2012; 11:M111011429.
- Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR III. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev.* 2013; 113:2343–2394. [PubMed: 23438204]
- Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem.* 2012; 404:939–965. [PubMed: 22772140]
- Li Q. Advances in protein turnover analysis at the global level and biological insights. *Mass Spectrom Rev.* 2010; 29:717–736. [PubMed: 19757418]
- McClatchy DB, Dong MQ, Wu CC, Venable JD, Yates JR III. ¹⁵N metabolic labeling of mammalian tissue with slow protein turnover. *J Proteome Res.* 2007; 6:2005–2010. [PubMed: 17375949]
- Kim TY, Wang D, Kim AK, Lau E, Lin AJ, Liem DA, Zhang J, Zong NC, Lam MP, Ping P. Metabolic labeling reveals proteome dynamics of mouse mitochondria. *Mol Cell Proteomics.* 2012; 11:1586–1594. [PubMed: 22915825]
- Price JC, Guan S, Burlingame A, Prusiner SB, Ghaemmaghami S. Analysis of proteome dynamics in the mouse brain. *Proc Natl Acad Sci US A.* 2010; 107:14508–14513.
- Claydon AJ, Thom MD, Hurst JL, Beynon RJ. Protein turnover: measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. *Proteomics.* 2012; 12:1194–1206. [PubMed: 22577021]
- Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR III. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem.* 2004; 76:4951–4959. [PubMed: 15373428]
- Savas JN, Toyama BH, Xu T, Yates JR 3rd, Hetzer MW. Extremely long-lived nuclear pore proteins in the rat brain. *Science.* 2012; 335:942. [PubMed: 22300851]
- Schwanhauser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature.* 2011; 473:337–342. [PubMed: 21593866]
- Boisvert FM, Ahmad Y, Gierlinski M, Charriere F, Lamont D, Scott M, Barton G, Lamond AI. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol Cell Proteomics.* 2012; 11:M111.
- Hsieh EJ, Shulman NJ, Dai DF, Vincow ES, Karunadharma PP, Pallanck L, Rabinovitch PS, MacCoss MJ. Topograph, a software platform for precursor enrichment corrected global protein turnover measurements. *Mol Cell Proteomics.* 2012; 11:1468–1474. [PubMed: 22865922]

19. Nelson CJ, Li L, Jacoby RP, Millar AH. Degradation rate of mitochondrial proteins in *Arabidopsis thaliana* cells. *J Proteome Res.* 2013; 12:3449–3459. [PubMed: 23750852]
20. Zhang Y, Reckow S, Webhofer C, Boehme M, Gormanns P, Egge-Jacobsen WM, Turck CW. Proteome scale turnover analysis in live animals using stable isotope metabolic labeling. *Anal Chem.* 2011; 83:1665–1672. [PubMed: 21314131]
21. Guan S, Price JC, Ghaemmaghami S, Prusiner SB, Burlingame AL. Compartment modeling for mammalian protein turnover studies by stable isotope metabolic labeling. *Anal Chem.* 2012; 84:4014–4021. [PubMed: 22444387]
22. Ramakrishnan R, Ramakrishnan JD. Using mass measurements in tracer studies—a systematic approach to efficient modeling. *Metab, Clin Exp.* 2008; 57:1078–1087. [PubMed: 18640385]
23. Brunengraber DZ, McCabe BJ, Kasumov T, Alexander JC, Chandramouli V, Previs SF. Influence of diet on the modeling of adipose tissue triglycerides during growth. *Am J Physiol Endocrinol Metab.* 2003; 285:E917–925. [PubMed: 12799315]
24. Aijo T, Granberg K, Lahdesmaki H. Sorad: a systems biology approach to predict and modulate dynamic signaling pathway response from phosphoproteome time-course measurements. *Bioinformatics.* 2013; 29:1283–1291. [PubMed: 23505293]
25. Tsai TH, Tadesse MG, Di Poto C, Pannell LK, Mechref Y, Wang Y, Ransom HW. Multi-profile Bayesian alignment model for LC-MS data analysis with integration of internal standards. *Bioinformatics.* 2013; 29:2774–2780. [PubMed: 24013927]
26. Topa H, Jonas A, Kofler R, Kosiol C, Honkela A. Gaussian process test for high-throughput sequencing time series: application to experimental evolution. *Bioinformatics.* 2015; 31:1762–1770. [PubMed: 25614471]
27. Shekar KC, Li L, Dabkowski ER, Xu W, Ribeiro RF Jr, Hecker PA, Recchia FA, Sadygov RG, Willard B, Kasumov T, Stanley WC. Cardiac mitochondrial proteome dynamics with heavy water reveals stable rate of mitochondrial protein synthesis in heart failure despite decline in mitochondrial oxidative capacity. *J Mol Cell Cardiol.* 2014; 75:88–97. [PubMed: 24995939]
28. Rachdaoui N, Austin L, Kramer E, Previs MJ, Anderson VE, Kasumov T, Previs SF. Measuring proteome dynamics in vivo: as easy as adding water? *Mol Cell Proteomics.* 2009; 8:2653–2663. [PubMed: 19724074]
29. Kasumov T, Ilchenko S, Li L, Rachdaoui N, Sadygov RG, Willard B, McCullough AJ, Previs S. Measuring protein synthesis using metabolic (²H) labeling, high-resolution mass spectrometry, and an algorithm. *Anal Biochem.* 2011; 412:47–55. [PubMed: 21256107]
30. Busch R, Kim YK, Neese RA, Schade-Serin V, Collins M, Awada M, Gardner JL, Beysen C, Marino ME, Misell LM, Hellerstein MK. Measurement of protein turnover rates by heavy water labeling of nonessential amino acids. *Biochim Biophys Acta, Gen Subj.* 2006; 1760:730–744.
31. Cambridge SB, Gnad F, Nguyen C, Bermejo JL, Kruger M, Mann M. Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J Proteome Res.* 2011; 10:5275–5284. [PubMed: 22050367]
32. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics.* 2002; 1:376–386. [PubMed: 12118079]
33. Øksendal, BK. *Stochastic Differential Equations: An Introduction with Applications.* 6. Springer; New York: 2003.
34. Iacus, SM. *Springer Series in Statistics.* Springer; New York: 2008. *Simulation and inference for stochastic differential equations with r examples.*
35. Huynh-Thu VA, Sanguinetti G. Combining tree-based and dynamical systems for the inference of gene regulatory networks. *Bioinformatics.* 2015; 31:1614–1622. [PubMed: 25573916]
36. Zurauskiene J, Kirk P, Thorne T, Pinney J, Stumpf M. Derivative processes for modelling metabolic fluxes. *Bioinformatics.* 2014; 30:1892–1898. [PubMed: 24578401]
37. Neal, RM. *Bayesian Learning for Neural Networks.* Springer; New York: 1996.
38. Byrd RH, Lu PH, Nocedal J, Zhu CY. A Limited Memory Algorithm for Bound Constrained Optimization. *Siam J Sci Comput.* 1995; 16:1190–1208.
39. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing; Vienna, Austria: 2010.

40. Lajtha A, Litzkovits L, Toth J. Comparison of turnover rates of proteins of the brain, liver and kidney in mouse in vivo following long term labeling. *Biochim Biophys Acta, Nucleic Acids Protein Synth.* 1976; 425:511–520.
41. Kasumov T, Dabkowski ER, Shekar KC, Li L, Ribeiro RF Jr, Walsh K, Previs SF, Sadygov RG, Willard B, Stanley WC. Assessment of cardiac proteome dynamics with heavy water: slower protein synthesis rates in interfibrillar than subsarcolemmal mitochondria. *Am J Physiol Heart Circ Physiol.* 2013; 304:H1201–H1214. [PubMed: 23457012]
42. Donnet S, Foulley JL, Samson A. Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics.* 2010; 66:733–741. [PubMed: 19912169]
43. Tuvdendorj D, Chinkes DL, Herndon DN, Zhang XJ, Wolfe RR. A novel stable isotope tracer method to measure muscle protein fractional breakdown rate during a physiological non-steady-state condition. *Am J Physiol Endocrinol Metab.* 2013; 304:E623–E630. [PubMed: 23321475]
44. Guan S, Price JC, Prusiner SB, Ghaemmaghami S, Burlingame AL. A data processing pipeline for mammalian proteome dynamics studies using stable isotope metabolic labeling. *Mol Cell Proteomics.* 2011; 10:M111010728.

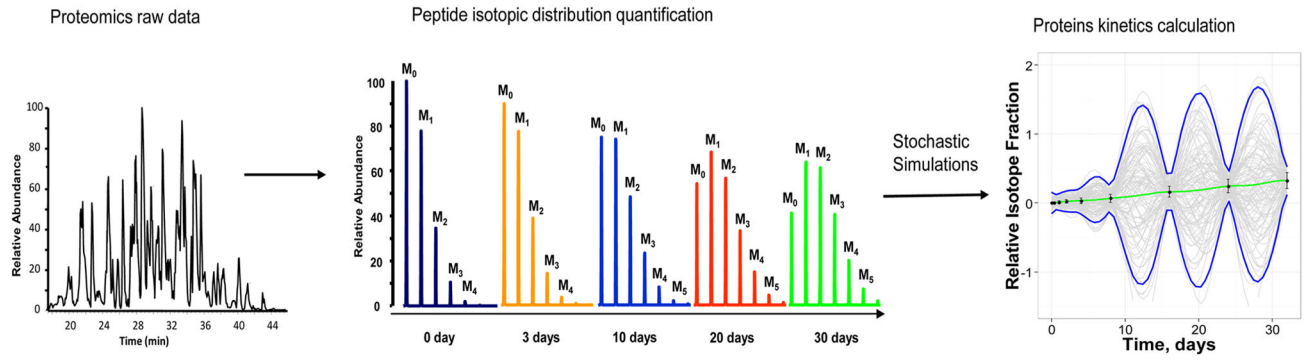


Figure 1. Modeling of protein turnover rate constants from stable-isotope-labeling experiments. The time-course of heavy-isotope labeling, extracted from the full scan spectra of tryptic peptides, is modeled as a GP with OU kernel.

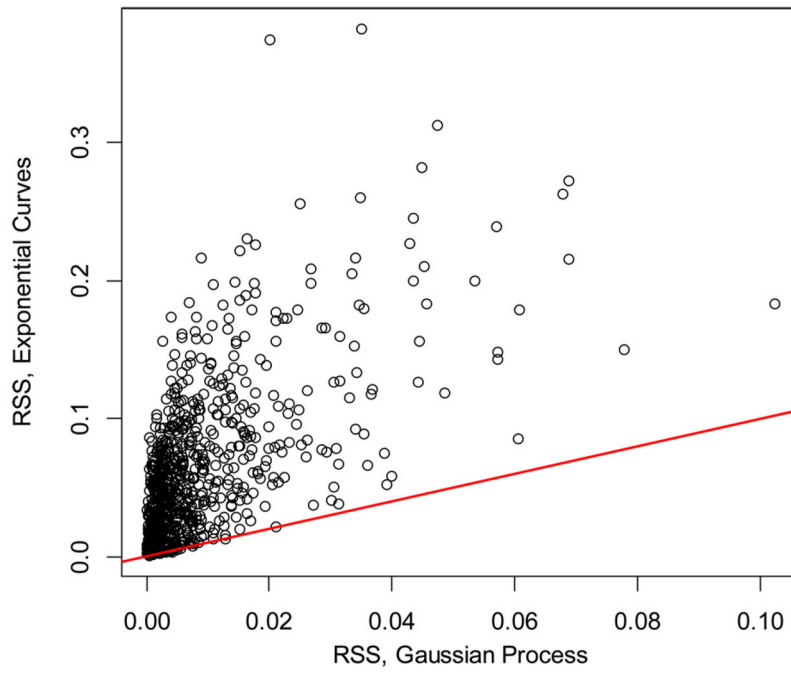


Figure 2. Comparison of the residual sum of squares obtained by using exponential curve fitting⁴⁴ (y -axis) and GP (x -axis). The red line is the line of unity. The shown data is for the liver proteome.

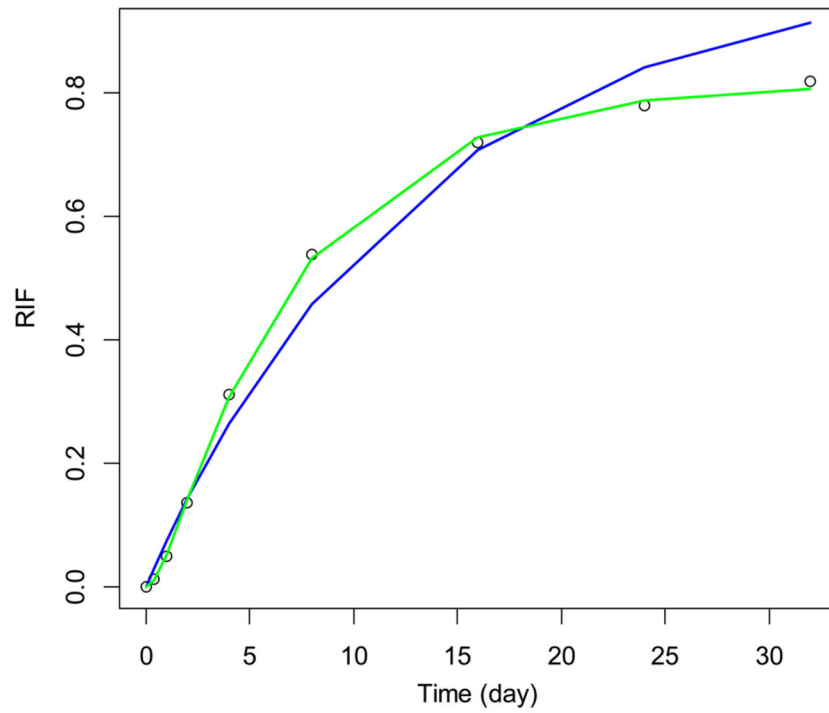


Figure 3. Experimental (empty circles) data and model fits (green line, GP; blue line, exponential curve fitting) for mitochondrial trifunctional protein, subunit α (Q8BMS1). The GP produced a fit with RSS about 70 times smaller than that from the two-exponent fit.

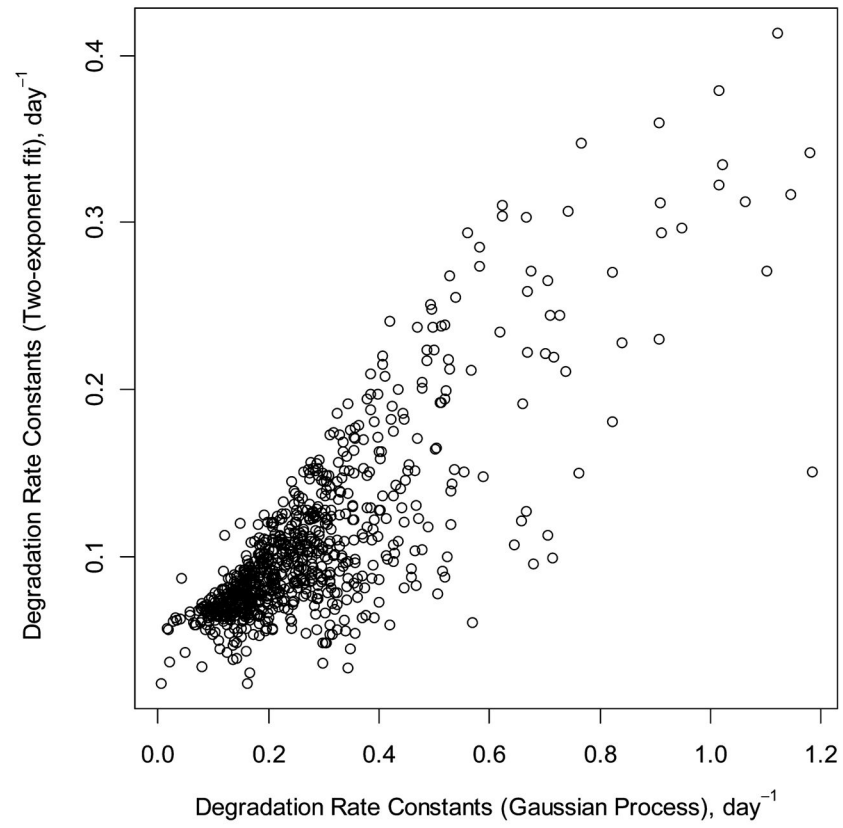


Figure 4. Scatter plot of the decay rates constants as computed by the GP (x -axis) and two-exponent curve fitting (y -axis). In general, the rate constants computed by the GP model were larger (mouse liver proteins).

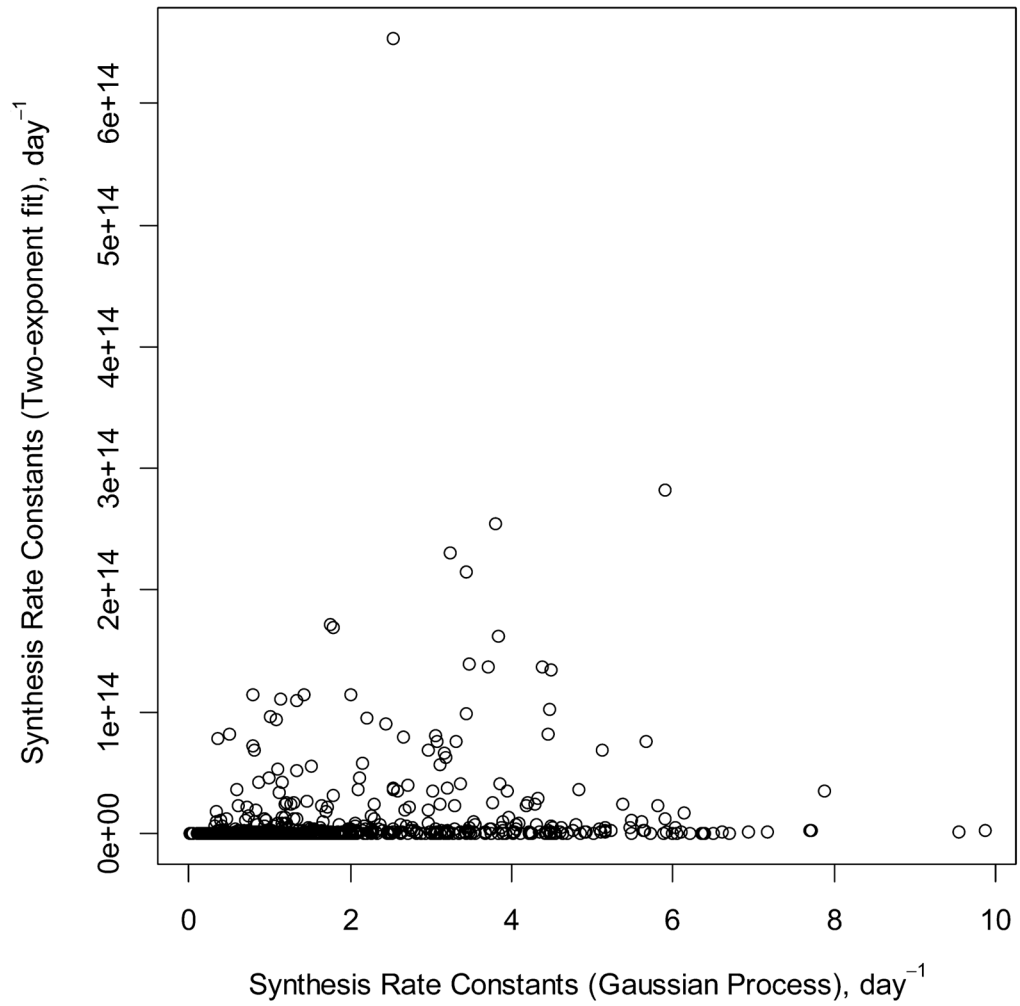


Figure 5. Synthesis rate constants as computed using two-exponent curves (y -axis) and GP (x -axis) for mouse liver data proteome.²¹

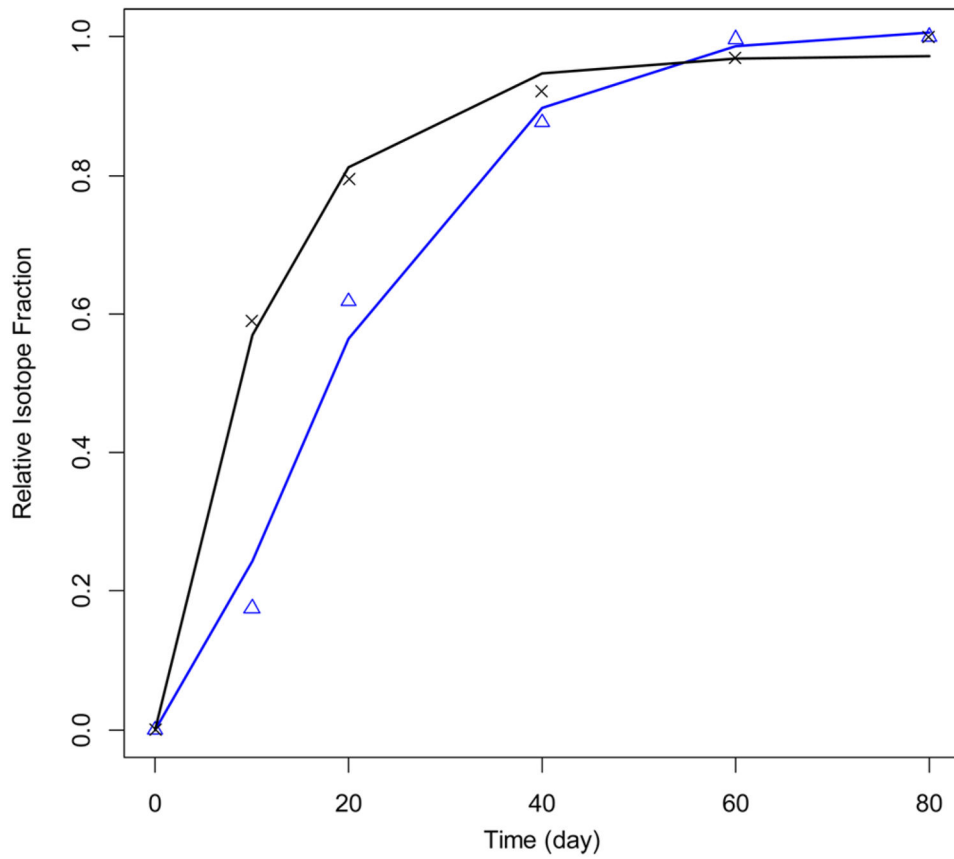


Figure 6. GP modeling of RIF for CPT protein from normal (experimental data is shown with blue triangles and the fit in blue line) and induced heart failure (the experimental data is shown with black crosses and the fit in black line) interfebrullar mitochondria.⁴¹