Calculation of Rate Spectra from Noisy Time Series Data

Vincent A. Voelz[*] and **Vijay S. Pande**

Department of Chemistry, Stanford University, Stanford, California

Abstract

As the resolution of experiments to measure folding kinetics continues to improve, it has become imperative to avoid bias that may come with fitting data to a predetermined mechanistic model. Towards this end, we present a *rate spectrum* approach to analyze timescales present in kinetic data. Computing rate spectra of noisy time series data via numerical discrete inverse Laplace transform is an ill-conditioned inverse problem, so a regularization procedure must be used to perform the calculation. Here, we show the results of different regularization procedures applied to noisy multi-exponential and stretched exponential time series, as well as data from time-resolved folding kinetics experiments. In each case, the rate spectrum method recapitulates the relevant distribution of timescales present in the data, with different priors on the rate amplitudes naturally corresponding to common biases toward simple phenomenological models. These results suggest an attractive alternative to the "Occam's razor" philosophy of simply choosing models with the fewest number of relaxation rates.

Keywords

protein folding kinetics; rate spectra; discrete inverse Laplace transform; regularization; Bayesian inference

Introduction

Complex kinetics often arises in systems with large of numbers of chemical states. Because of this, one might expect protein folding kinetics to be complex, but remarkably, many proteins appear to fold in a two-state manner ($U \rightleftharpoons N$), with barrier-limited kinetics predicted by a simple Arrhenius model. As the time resolution accessible by folding kinetics experiments continues to improve, however, evidence of more complex folding kinetics can now be observed [1, 2].

Traditionally, the interpretation of experimental kinetics have required fitting to some predetermined mechanistic model. A common procedure is to describe the kinetics with the fewest number of number of relaxation rates, thus producing the "simplest" possible network of barrier-limited states. Herein lies a problem, because the simplest model that explains the data may not necessarily be the model with the fewest number of rates. For

[*]Corresponding Author. Current address: 240 Beury Hall, Department of Chemistry, Temple University, Philadelphia, PA 19122, (Phone) 415-318-6417 (Fax) 215-204-1532, voelz@temple.edu.

example, the kinetics may not be in the barrier-limited regime to begin with. Another problem is that kinetics experiments usually measure a single reaction coordinate, and it is very difficult (especially in folding studies) to design experimental probes that can unambiguously report on specific conformational states. In this case, fitting kinetic data to a simple phenomenological model may not capture important details of the underlying dynamics.

With these issues in mind, we present a *rate spectrum* approach to analyze timescales present in kinetic data, without imposing any pre-existing mechanistic model. Instead of fitting to a model with a predetermined number of rates, the rate spectrum approach can be thought of as fitting to a *distribution* of rate amplitudes, in a way that is informed by some simple prior knowledge about the shape of this distribution. In some ways, this approach is in the same spirit of model-free approaches in interpreting NMR experiments [3].

The rate spectrum approach

The rate spectrum approach is akin to calculating a inverse Laplace transform for discrete time series: we assume the signal can be described as a sum of exponential-decay basis functions, and our main task is to estimate the amplitude of each. A somewhat analogous experimental method is modulation fluorometry, in which distributions of relaxation rates can be extracted from fluorescence decays [4]. For a time series (t_i, y_i), $i = 1, \ldots, N$, the rate spectrum approach can be posed as the problem of finding the coefficients β_j, $j = 1, \ldots, K$ corresponding to a set of rates k_j, (of our choosing), to find the best-fit model:

$$\text{(1)}$$

Computing rate spectra is an example of an ill-conditioned inverse problem, so a regularization scheme is needed to make the solution of this inverse problem well-behaved, especially in the presence of noise. Once a rate spectrum (i.e. set of β_j) is obtained, however, one can identify all the timescales present in the data, and use this information as a starting point for inferring mechanistic models.

The regularized rate spectrum can be obtained as the set of β_j that minimize the quantity

$$\text{(2)}$$

where $X_{ij} = e^{-k_j t_i}$. The left-hand term is simply the residual sum of squared errors, while the right-hand term is a penalty for large (positive or negative) values of β_j. The strength of this penalty is controlled by λ, the *regularization parameter*, and $f(\beta)$ defines the type of regularization. We look at three different types of regularization, each of which impose

different penalties on the magnitude of the rate amplitudes: ridge regression ,

lasso regression ($f = \sum_j |\beta_j|$), and elastic net regression (, for

mixing parameter $0 \leq \rho \leq 1$). It can be shown (see Methods) that these different regularization penalties correspond to different prior distributions on the set of rate amplitudes, and (as we show below) are useful in different situations. (For an excellent in-depth discussion of these and other regularization schemes, see *Elements of Statistical Learning* by Hastie, Tibshirani and Friedman [5].)

What do regularized spectra look like? Without regularization, we obtain the least-squares solution, which usually results in a spiky, irregular spectrum, with $\hat{\mathbf{y}}$ overfit to the noise. As λ increases, the sensitivity to noise decreases, and the spectra broaden (Supplementary Figure S1). When the regularization parameter is very large, or the data very noisy, spectra can become broadened considerably, however, in most cases peaks in the spectra can be recovered that correspond to the most important timescales present in the data (Figure 2).

One of the crucial questions in regularization problems is how to choose the regularization parameter λ. Our method treats this as a Bayesian inference problem, in which the posterior likelihood of possible models (i.e. sets of rate amplitudes) can be sampled, along with nuisance parameters σ and τ, which are width parameters for the noise and rate amplitude priors, respectively (see Methods). In this scheme, it can easily be shown that $\lambda = \sigma^2/\tau^2$ (see Supplementary Information). We use a Monte Carlo algorithm to sample over σ, τ (and ρ) to obtain estimates of the posterior $P(\lambda|\mathbf{y})$ (Supplementary Figure S2).

Below, we apply regularization to several examples of synthetic and actual experimental data. Our results show that the rate spectra approach can robustly discern multiple timescales from very noisy data. Moreover, the form of the solutions obtained by these methods sheds light on typical biases in fitting simple models to data.

Methods

Calculation of rate spectra from a time series (t_i, y_i), $i = 1, \ldots, N$, can be posed as the problem of finding the coefficients β_j, $j = 1, \ldots, K$ corresponding to a set of rates k_j, (of our choosing) such that

$$ \tag{3} $$

The problem can be cast as a linear regression model, if we transform the times t_i into a series of "regressors" $\mathbf{x}_i = (e^{-k_1 t_i}, e^{-k_2 t_i}, \ldots, e^{-k_K t_i})$, such that

$$ \tag{4} $$

where $X_{ij} = e^{-k_j t_i}$.

In simple least-squares regression, the best-fit $\hat{\beta}$ are the values that minimize the residual sum of squared error (RSS):

As long as N K, and **X** has full column rank, **y** = **X**β is an over-constrained system of linear equations, and the β that minimizes the *RSS* is given by $\hat{\beta} = \mathbf{X}^+\mathbf{y}$, where $\mathbf{X}^+$ is the Moore-Penrose pseudo-inverse $\mathbf{X}^+ = (\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\mathbf{X}^\mathrm{T}$.

The regression problem, however, is ill-conditioned because the transformation of the time series into $\mathbf{x}_i$ is highly nonlinear, so solutions of fitted values β_j may be very sensitive to small changes in the inputs y_i. Additionally, for large values of K, over-fitting becomes a problem, with large positive values of β_j balancing out large negative values in such a way as to tightly fit the y_i training data, regardless of noise.

To deal with these problems, a method of *regularization* is required. A typical procedure is a method called Tikhonov regularization, also known as *ridge regression*. In ridge regression, a constrained optimization is used to find the solution for $\hat{\beta}$

$$ \text{(5)} $$

The Lagrangian form of this optimization problem gives a quantity that can be explicitly minimized in order to find $\hat{\beta}$, using the Lagrange multiplier λ.

$$ \text{(6)} $$

Here, λ is a *regularization parameter*, and its role is to penalize large (positive or negative) values of β_j. The optimal β can be obtained from [image], where [image] is the pseudo-inverse of an $(N+K) \times K$ matrix $\mathbf{X}_\lambda$

$$ $$

Two related schemes for regularization are *lasso regression*, in which a penalty is imposed on the sum of the absolute values of β_j,

$$ \text{(7)} $$

and *elastic net regression*, in which a combination of penalties is used, with the mixture controlled by an additional parameter ρ:

$$ (8) $$
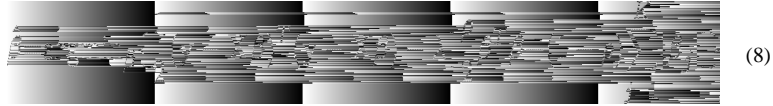
As we will see below, each of these schemes has a particular probabilistic interpretation.

We also note that, unlike ridge regression, solving for β within these other regularization schemes does not come straightforwardly from application of the pseudoinverse. Instead, more general convex optimization schemes must be employed. In the case of lasso regression, for instance, the problem can be reduced to a least-squares problem with additional inequality constraints [6]. In practice, we use a coordinate-descent algorithm to perform minimizations for lasso and elastic net regression.

A Bayesian regularization scheme

An all-important question in regularization problems is: to what extent should our solutions for β be regularized? Ideally, best estimates for λ could be made using the training data itself, with various in-data schemes to avoid overfitting, for example, using generalized cross-validation [5, 7]. Alternatively, a Bayesian approach can be used to obtain the posterior likelihood $P(\lambda|\mathbf{y})$ of a given regularization parameter λ given the training data $\mathbf{y}$. According to Bayes' rule, the posterior probability $P(\lambda|\mathbf{y})$ is proportional to the likelihood of observing the data $\mathbf{y}$ given a model parameterized by λ, multiplied by some prior probability $P(\lambda)$.

A probabilistic interpretation of λ comes from the following model (see Supplementary Information): If we assume that (1) the coefficients β_j are normally distributed with variance τ^2, i.e. $\beta \sim N(0, \tau^2\mathbf{I})$, and (2) that the training data y_i are contaminated with Gaussian noise of variance σ^2, i.e. $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$, it can be shown that the λ parameter in ridge regression corresponds to:

This provides an intuitive interpretation of the meaning of λ: it is the Lagrange multiplier used to simultaneously enforce the variance (τ^2) of the amplitudes β_j, as well as the variance (σ^2) in the resulting fit $\hat{\mathbf{y}}$.

Similarly, lasso regression has the interpretation $\lambda = \sigma^2/\tau^2$, but with the amplitudes β_j drawn from the prior distribution

Elastic net regression has the interpretation that the prior for each β_j is

where ρ is a mixing parameter and $A(\tau, \rho)$ is a normalization constant (see Supplementary Information).

A visual inspection of these prior distributions (Figure 1) shows that the lasso regression should produce more exaggerated peaks in rate spectrum: as compared to ridge regression, it favors β_j either close to zero, or with large amplitude. The prior for elastic net lies somewhere between these two limiting cases.

Although we don't know the values of σ and τ *a priori*, we can make estimates from the training data, by considering the posterior likelihood of σ and τ given the data **y**. From Bayes' rule, this can be computed as:


Here $P(\sigma)$ and $P(\tau)$ are prior distributions for σ and τ, respectively, assuming σ and τ are independent. Conjugate priors are often chosen for the sake of obtaining analytical solutions. We can avoid unnecessary assumptions regarding the prior distributions by sampling numerically from the posterior. We use the non-informative (scale-invariant) Jeffrey's priors $P(\sigma) \sim 1/\sigma$ and $P(\tau) \sim 1/\tau$ to obtain a complete expression for the (unnormalized) posterior likelihood [8]:

We assume that the distribution $P(\beta)$ is very sharply peaked at , so that β is determined completely from λ and the training data **y**


The posterior can then be written

We use this expression to obtain posterior estimates of $P(\lambda|\mathbf{y})$ by simulation-based methods. With enough samples, λ can be treated as a *nuisance parameter* to obtain an expectation estimate for the set of relaxation amplitudes β_j.
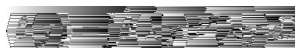
To sample from the posterior distribution, we use a Monte Carlo algorithm using $-\log P(\sigma, \tau|\mathbf{y})$ as an effective energy function, which (up to a constant) is easily shown for ridge regression to be:
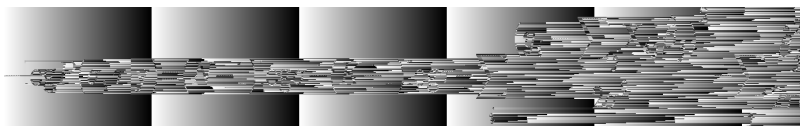
The first two terms in this expression favor small values of σ and τ, while the last two terms favor large values of σ and τ, so the posterior has a maximum at some intermediate value of $\lambda = \sigma^2/\tau^2$.

For lasso regression, the expression for $-\log P(\sigma, \tau|\mathbf{y})$ is the nearly same, but with the terms involving τ slightly modified:
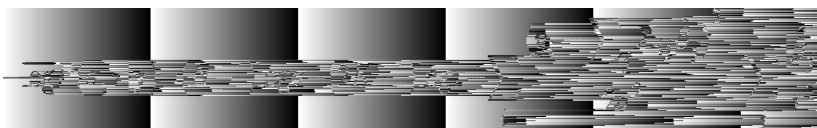
For elastic net regression, $-\log P(\sigma, \tau|\mathbf{y})$ is found to be (see Supplementary Information):

where ⬚⬚⬚⬚⬚⬚. This expression comes from the normalization of $P(\sigma, \tau|\mathbf{y})$.

As discussed in [8], sampling from the posterior in this way has a number of advantages compared to cross-validation. For example, the Bayesian approach is completely determined by the data, avoids the need to define any particular partition of the data, and is robust to sparse data. More importantly, a key advantage of Bayesian regression, as opposed to ridge regression (which requires some in-data estimate of optimal $\lambda = \sigma^2/\tau^2$, typically by cross-validation), is that the posterior provides some interpretable evaluation of how ill-posed the problem is – i.e. the posterior will be broad if the problem is ill-posed. Performing ridge regression for an ill-conditioned problem using a single λ value, even after optimizing to find the best value of λ, could still result in either a very shrunk estimate with a poor fit, or wildly different estimates that all fit the data. While this too is akin to having a wide posterior, it is much harder to interpret.

Also note that the posterior likelihood function depends on the number of rates K used in computing the spectrum. Thus, when K is large (comparable to N) the regularization penalty in the posterior likelihood becomes much larger, and may result in a broadened spectrum. This reflects the amount of statistical evidence available for the calculation; i.e. if one attempts to find $K = 1000$ rate coefficients β_j to explain only $N = 1000$ samples in a time series, the posterior will be more heavily weighted by the prior distribution of β_j. With more evidence (say, using only $K = 100$ rates), the calculation is more heavily weighted towards the data.

Sampling from the posterior

In practice, to sample from the posterior, we use the following procedure, starting with an initial guess of (σ, τ):

1. propose new candidate (σ', τ') as a small perturbation from (σ, τ)

2. calculate ▨, and use it to compute $-\log P(\sigma, \tau|\mathbf{y})$

3. accept or reject the move according the Metropolis criterion (i.e. $(\sigma', \tau') \to (\sigma, \tau)$ with probability $\min(1, P(\sigma', \tau'|\mathbf{y})/P(\sigma, \tau|\mathbf{y}))$)

4. repeat

(For posterior sampling using the elastic net prior, the algorithm is the same, but includes sampling over an additional parameter, ρ.)

For each accepted sample $(\sigma, \tau)_m$ obtained via the Monte Carlo sampling algorithm, we keep track of the predicted $\hat{\beta}_m$, and corresponding weight $w_m = P(\sigma, \tau|\hat{\beta}_m)$, so that we can compute the expectation $\langle\beta\rangle$ using the entire collection of posterior samples:

Uncertainty in $\langle\beta\rangle$ can be computed as

For all the results shown here, twenty thousand steps of Gibbs sampling in (σ, τ) was performed, with moves in σ and τ generated by random perturbations by $\varepsilon \sim N(0, a^2)$ where a typical value of a was 0.005. (ρ was sampled similarly for elastic net regression.)

Results

Rate spectra of noisy multi-exponential time series

To test the rate spectra approach, we considered the following noisy tri-exponential time series:

where $(\tau_1, \tau_2, \tau_3) = (10^{-6}s, 10^{-4}s, 5 \times 10^{-3}s)$, $(a_1, a_2, a_3) = (0.3, 0.3, 0.4)$, and $N(0, s^2)$ denotes normally-distributed noise of variance s^2. One thousand samples y_i were generated at time points t_i exponentially-spaced between 10^{-9} second and 1 second (i.e. log t_i is linearly-spaced). Rate spectra were calculated using 100 rates, exponentially-spaced (i.e. linear on a log-scale) over a range of $1\ s^{-1}$ and $10^9\ s^{-1}$. Note that if we wish to compare these results to a continuous inverse Laplace transform, each coefficient β_j can be thought of as carrying an additional factor $d(\log t_i)$ to account for the change of integration variable.

When we compare the results of ridge, lasso and elastic net regression for the $s = 0.05$ test data (Figure 3), we see that the lasso regression predicts a much more peaked rate spectrum, more closely approximating the "true" rate spectrum we would expect for a pure tri-exponential decay with no noise. This agreement is because the corresponding prior for the β_j more closely matches our preconceived expectations (that the time series contains a small number of high-amplitude timescales.) The elastic net regression produces rate spectra somewhere in between those of the ridge and lasso. Posterior sampling of ρ shows a range of likely mixing parameters, between 0.2 and 0.5. One attractive feature of both the lasso and elastic net is the prediction of rate coefficients that are exactly zero, due to the presence of terms $|\beta_j|$ in the regularization constraint.

Rate spectra of noisy stretched-exponential time series

For non-exponential kinetics, a common procedure is to fit data with a so-called "stretched exponential" function.
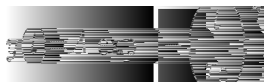
where γ is a *stretching parameter*, $0 < \gamma \leq 1$. The stretched exponential can be thought of as arising from kinetic phenomena on a broad range of timescales, as in the relaxation of glassy materials [9]. An analytical solution for the (continuous) rate spectrum $H_\gamma(k)$ exists [10] (see Supplementary Information), which we compare to calculated rate spectra of stretched exponentials in the presence of noise.

We computed rate spectra using ridge, lasso and elastic net regression for various stretching parameters (Figure 4). For these tests, 1000 samples were generated from a stretched exponential time course (with τ_0 set to 1 second), with synthetic noise added ($s = 0.05$), over timescales between 10^{-3} and 10^3 seconds, for various stretching parameters. As in the tri-exponential tests, 100 rates were used to compute spectra, and posterior sampling of $P(\lambda|\mathbf{y})$ was used to determine the extent of regularization. Our results confirm that that regularized rate spectra for ridge and elastic net regression can indeed recapitulate rate spectra very similar to the analytic result $H\gamma(k)$. Lasso regression, however, consistently predicts spectra dominated by a limited number of timescales. As the stretching parameter decreases, lasso regression essentially predicts the minimum number of timescales needed to explain the data: two timescales for $\gamma = 0.9$, three for $\gamma = 0.7$, four for $\gamma = 0.5$, and five for $\gamma = 0.3$.

Like the tri-exponential case, the rate spectrum that best approximates the analytic result comes using the prior distribution of rate amplitudes most closely approximating our expectations–in this case, ridge regression, which generates broad, continuous spectra. The lasso regression results, however, show that a small number of exponential relaxations may fit the data equally well. This highlights why stretched-exponential fits must be performed with caution: such models are predicated upon the existence of a broad, continuous spectrum of relaxation rates; this assumption must be justified for reasons other than simply a good fit to the data (for example, by physical argument).

Application to analysis of folding kinetics

While describing folding kinetics using a stretched exponential is common, it is usually an acknowledgement that the kinetic data are sufficiently complex to warrant more than one timescale, yet lack statistical power to make more definite claims about the nature of such timescales. The stretching parameter γ in this case provides a convenient single-parameter model to describe the complex kinetics.

Liu et al. (2008) published temperature-jump (T-jump) fluorescence traces of WW domain folding kinetics at different temperatures, for several sequences having a range of thermal stabilities. Figure 3 of Li et al. shows examples of contrasting traces: a mutant whose kinetics is well-described by a single-exponential (an "apparent two-state folder" [11]), versus a trace that can be fit more accurately to a single-exponential plus a stretched-exponential component. The purpose of this fitting was (mainly) to identify the faster of two relaxation timescales (the so-called molecular timescale, fit by a stretched-exponential) that the authors consistently observed across many experiments.

We were interested to see if our rate spectra method could identify these fast timescales, and perhaps extract more information about them. Ten thousand steps of posterior sampling (using 100 rates between 1 and 10^8 s^{-1}) were performed for the time traces in Figures 3a and

3b of Li et al., each of which consist of 357 samples between 0 and 100 µ*s*. For clarity, we will refer to these data as traces *A* and *B*.

While it is very difficult to discern deviations from single-exponential kinetics by visual inspection (Figure 5), ridge regression is indeed able to reveal a broadened (negative amplitude) peak near a timescale of ∼ 1 µ*s* for trace *B*, corresponding well to the molecular timescale of ∼ 1.2 ± 0.5 µ*s* originally found by curve-fitting. The width of this peak also is consistent with the stretching parameter originally found by fitting, γ ∼ 0.4.

Neither lasso nor elastic net regression supports more than a single timescale given the available data, although we note that the elastic net method predicts a broadened rate distribution for the low-temperature T-jump. This situation could, of course, change with more data at early timescales, as there are many more points sampled in the trace after 10^{-5} s than before this time. (Obtaining such experimental data, of course, is very challenging, and a motivation for this analysis in the first place.)

Discussion

In this paper, we have presented a new approach for calculating rate spectra from noisy time series data. We have shown that when rate spectra are calculated with a suitable Bayesian regularization scheme, relaxation timescales can be robustly recovered from noisy time traces. As is clear from the examples we present, the statistical power of such spectral analysis depends on the extent of available data, and our prior assumptions about the distribution of rate amplitudes.

What regulation scheme should be used in a given situation? This depends on how much information is known *a priori* about the distribution of rates. For situations in which little is known about the underlying rate processes contributing to the observed data, ridge regression works well. While the rate spectra obtained from ridge regression are more broad, our results show how this regularization procedure recapitulates the relevant timescales as maxima in the rate spectra.

In situations where we are much more confident in expecting a limited number of timescales, lasso regression does an excellent job of reproducing the true theoretical rate spectra (a combination of three delta functions) from noisy tri-exponential test data. In a traditional curve-fitting scenario, fitting the fewest number of exponential relaxations would be deemed the "simplest" model, but in the context of rate spectra calculation, lasso regression achieves this by enforcing our prior assumptions that distribution of timescales should be dominated by only a few relevant timescales. The elastic net method, using a mixture of ridge and lasso priors, produces rate spectra that lie between these two extremes.

Our results calculating rate spectra for noisy stretched-exponential time courses suggest that, although such time courses can be fit with curves having very few free parameters (namely τ_0 and γ), this does not imply that a stretched-exponential is the most parsimonious model that explains the data. Rather, the stretched-exponential is a phenomenological law that serves as a proxy to describe complicated distributions of timescales. The analytical rate distributions corresponding to a perfect stretched-exponential are not extracted very

efficiently with lasso and elastic net regularization methods, because these methods impose strict penalties on broad distributions of relaxation timescales. Despite this, lasso regression for a noisy stretched-exponential curve ($\gamma = 0.3$) shows that a narrow rate spectrum consisting of five main timescales can explain the time course equally well. If our prior expectation is that only a few well-defined relaxation rates exist, lasso regression may be preferred.

This work highlights the importance of model-independent analysis of time series data, and the dangers of conflating two very different questions. One question is: which model best describes my data? This can be answered using any number of hypothesis testing schemes. A very different question, addressed by our work here, is: what timescales are present in my data? The rate spectrum method provides a framework illuminating these timescales from the available data. Unbiased by any preconceived model, rate spectra can thus serve as a very useful starting point for guide more thorough studies of folding kinetics and other dynamic systems.

A growing number of studies have reported that Markov State Models (MSMs) of protein folding, despite having a large number of metastable states and relaxation time scales, predict simple apparent kinetics for experimental observables. These studies have found that projections of the MSM folding dynamics onto experimental observables may be sensitive to only a small number of timescales, masking the underlying complexity [12, 13, 14]. For this and other reasons, we think it is very likely that as the temporal and structural resolution of experiments and simulations continue to increase, using rate spectra to analyze folding kinetics will be increasingly useful.

A python package is available for computing rate spectra using the methods described in this paper. Code and examples are freely downloadable at https://simtk.org/home/ratespec.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Reiner A, Henklein P, Kiefhaber T. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. PNAS. 2010; 107(11)

2. Neuweiler H, Banachewicz W, Fersht AR. Kinetics of chain motions within a protein-folding intermediate. PNAS. 2010; 107(51):22106–22110. [PubMed: 21135210]

3. Peng JW, Wagner G. Mapping of spectral density functions using heteronuclear nmr relaxation measurements. Journal of Magnetic Resonance (1969). 1992; 98(2):308–332.

4. Gratton E, Hall RD, Jameson DM. Multifrequency phase and modulation fluorometry. Annual Review of Biophysics and Bioengineering. 1984; 13:105–124.

5. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. 5th edition. New York: Springer-Verlag; 2011.

6. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B. 1996; 58(1):267–288.

7. Golub G, Heath M, Wahba G. Generalized cross-validation as a method for estimating a good ridge parameter. Technometrics. 1979; 21(2):215–223.

8. Habeck M, Rieping W, Nilges M. Weighting of experimental evidence in macromolecular structure determination. PNAS. 2006; 103(6):1756–1761. [PubMed: 16446450]

9. Palmer RG, Stein DL, Abrahams E, Anderson PW. Models of hierarchically constrained dynamics for glassy relaxation. Physical Review Letters. 1984; 53(10):958–961.

10. Berberan-Santos MN, Bodunov EN, Valeur B. Mathematical functions for the analysis of luminescence decays with underlying distributions 1. kohlrausch decay function (stretched exponential). Chemical Physics. 2005; 315:171–182.

11. Nguyen H, Jäger M, Moretto A, Gruebele M, Kelly JW. Tuning the free-energy landscape of a ww domain by temperature, mutation, and truncation. PNAS. 2003; 100(7):3948–3953. [PubMed: 12651955]

12. Bowman GR, Voelz VA, Pande VS. Atomistic folding simulations of the five helix bundle protein λ_{6-85}. JACS. 2011; 133(4):664–667.

13. Noé F, Doose S, Daidone I, Löllmann M, Chodera JD, Sauer M, Smith JC. Dynamical fingerprints: Understanding biomolecular processes in microscopic detail by combination of spectroscopy, simulation and theory. PNAS. 2011 submitted.

14. Voelz VA, Jäger M, Yao S, Chen Y, Zhu L, Waldauer SA, Bowman GR, Friedrichs M, Bakajin O, Lapidus LJ, Weiss S, Pande VS. Surprising complexity to protein folding revealed by theory and experiment. 2011 submitted.

15. Pollard H. The representation of e^{-x^λ} as a laplace integral. Bulletin of the American Mathematical Society. 1946; 52:908–910.

16. Berberan-Santos MN. Analytical inversion of the laplace transform without contour integration: application to luminescence decay laws and other relaxation functions. Journal of Mathematical Chemistry. 2005; 38(2)
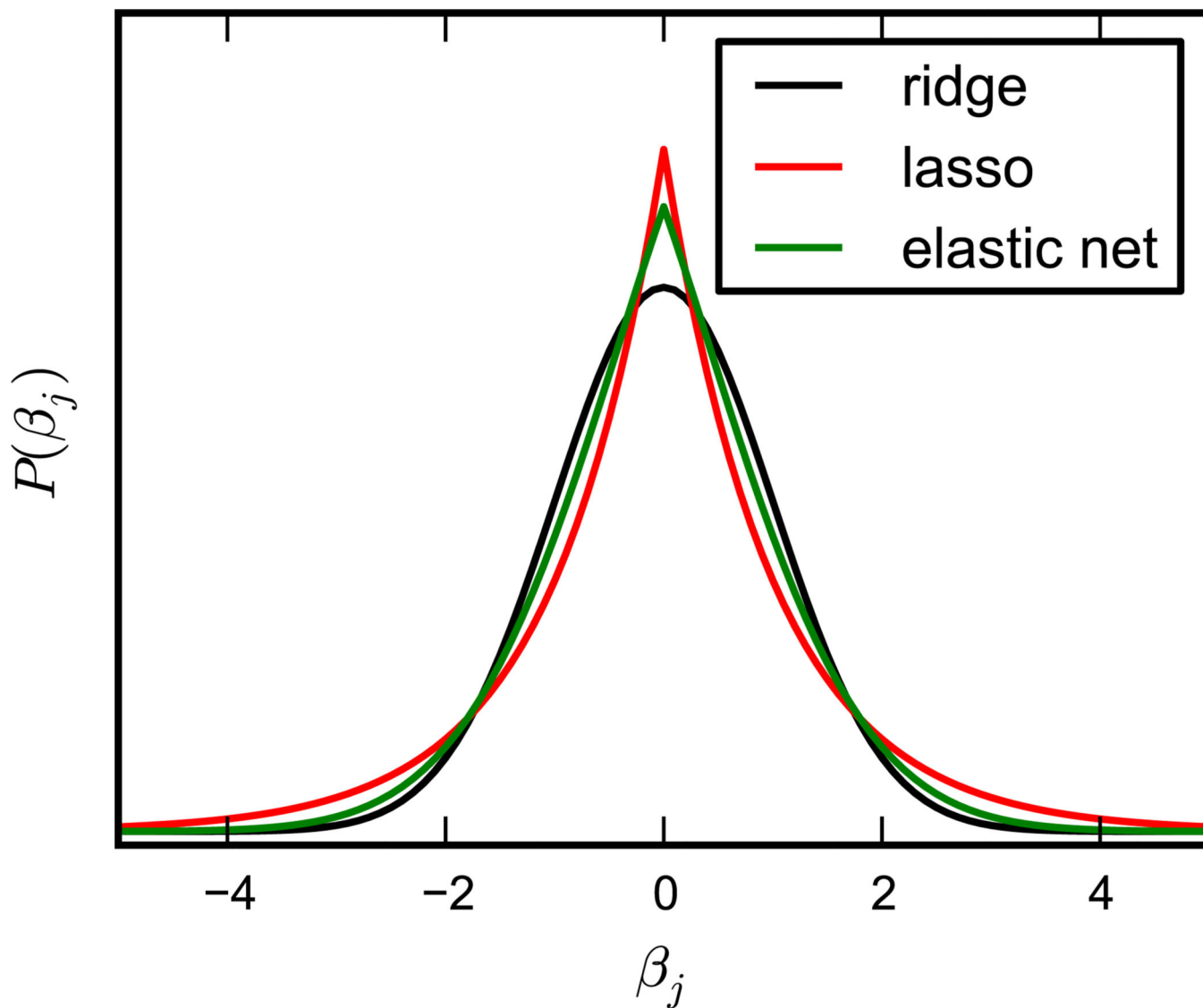
Figure 1.
Prior distributions of rate amplitudes β_j according to ridge regression (black), lasso regression (red), and elastic net regression (green; shown for $\rho = 0.5$).
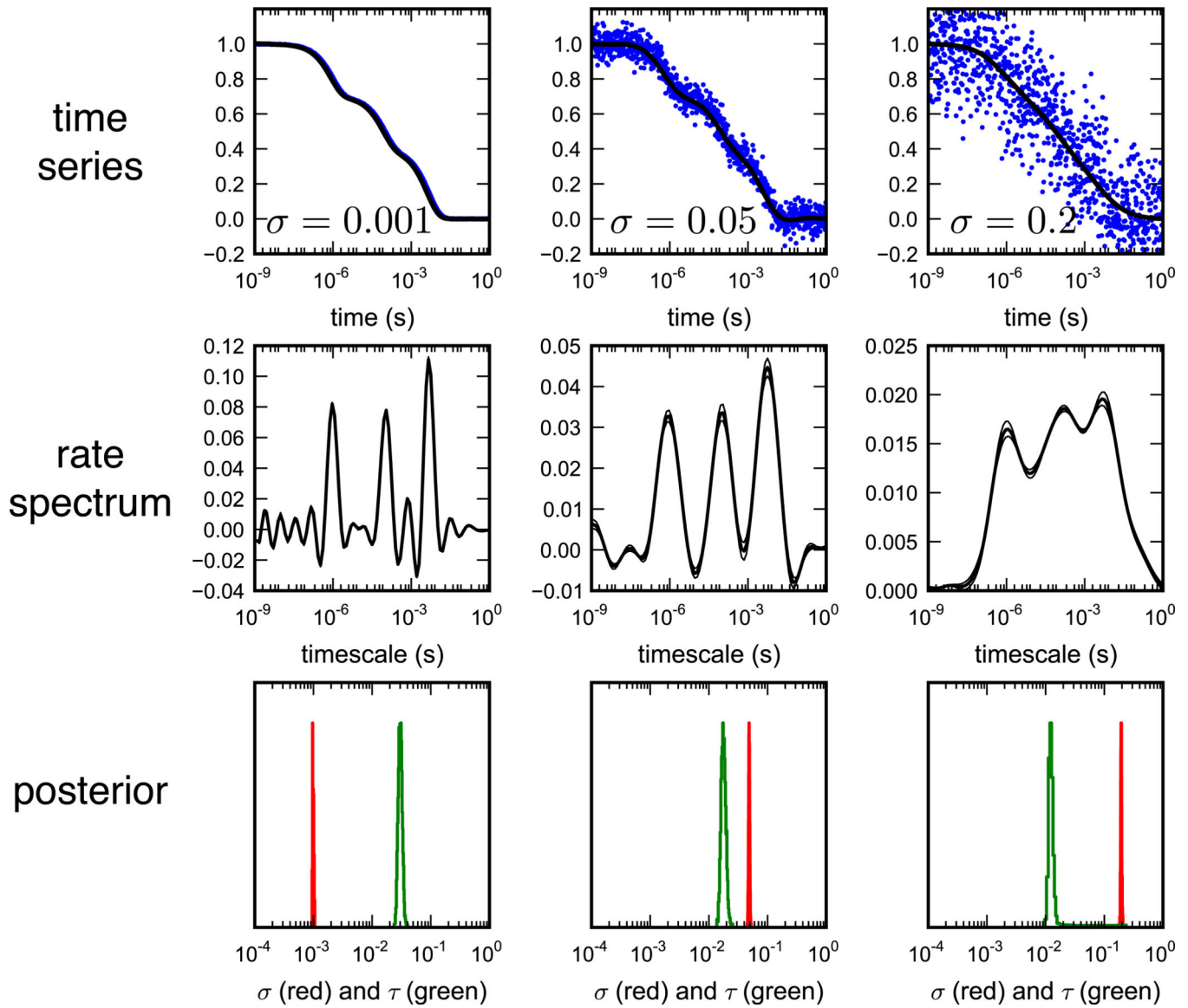
Figure 2.

Noise broadens regularized rate spectra. (top) Regularized rate spectra (using ridge regression) for the tri-exponential data with increasing amounts of noise $N(0, s^2)$, for $s = 0.001$, 0.05 and 0.2. (middle) Rate spectra shown are expectation values from posterior sampling, shown with a 95% confidence interval (in all cases very small). Despite this spectral broadening, regularized rate spectra show three peaks corresponding to each timescale, even for very noisy data. (bottom) The posterior distributions sampled for σ and τ give additional information about the nature of the data being fit. While the posterior distribution of τ remains robust for different values of synthetic noise (reflecting the regularization penalty), the posterior distributions of σ are very narrow and predict extremely well the variance of the artificial noise used to product the time series.
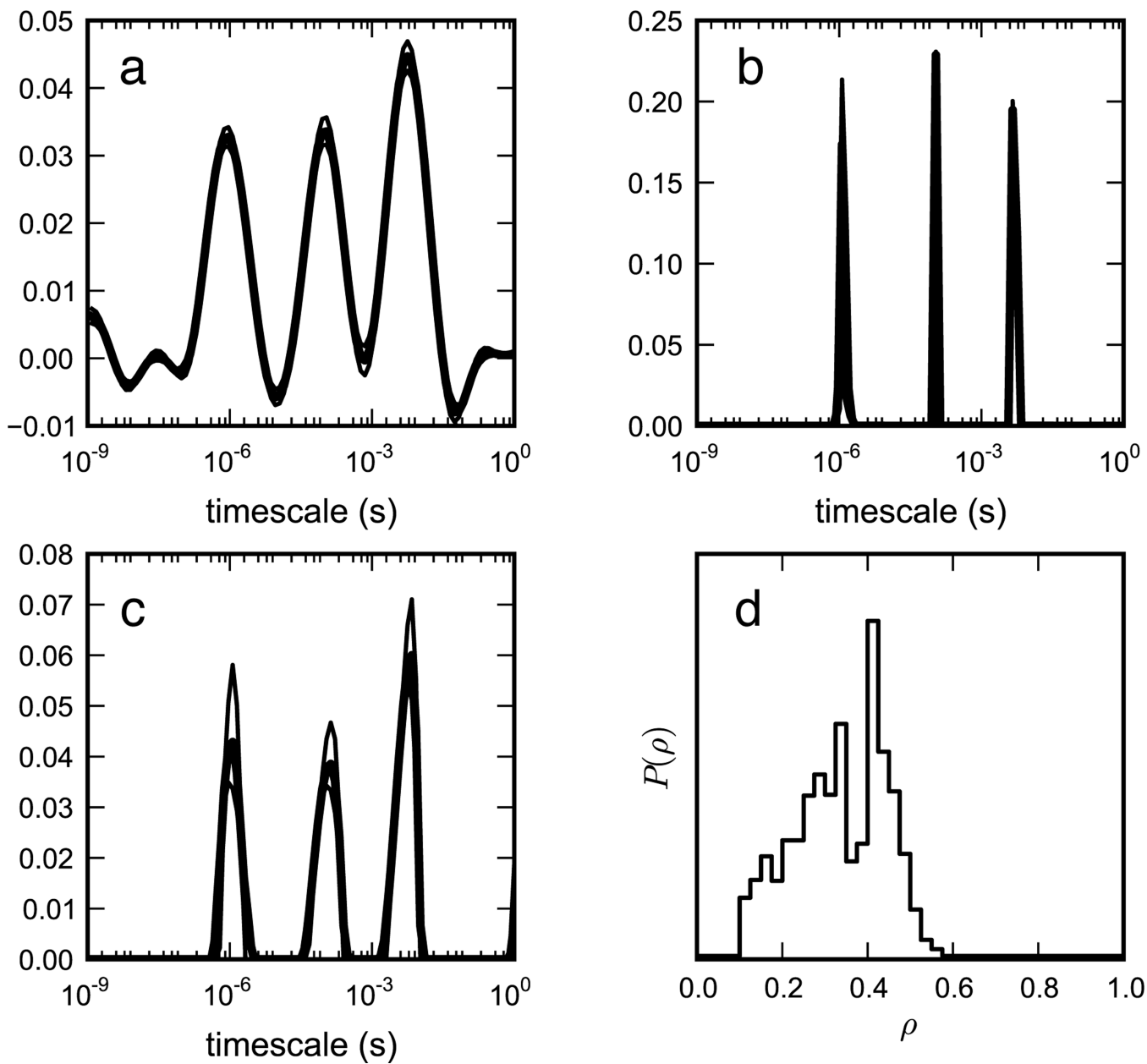
Figure 3.
Rate spectra for the tri-exponential data with artificial noise $N(0, s^2)$, $s = 0.05$, using (a) ridge regression, (b) lasso regression and (c) elastic net regression. Rate spectra shown are expectation values from posterior sampling, shown with a 95% confidence interval. (d) The sampled posterior distribution of mixing parameter ρ in the elastic net regression.
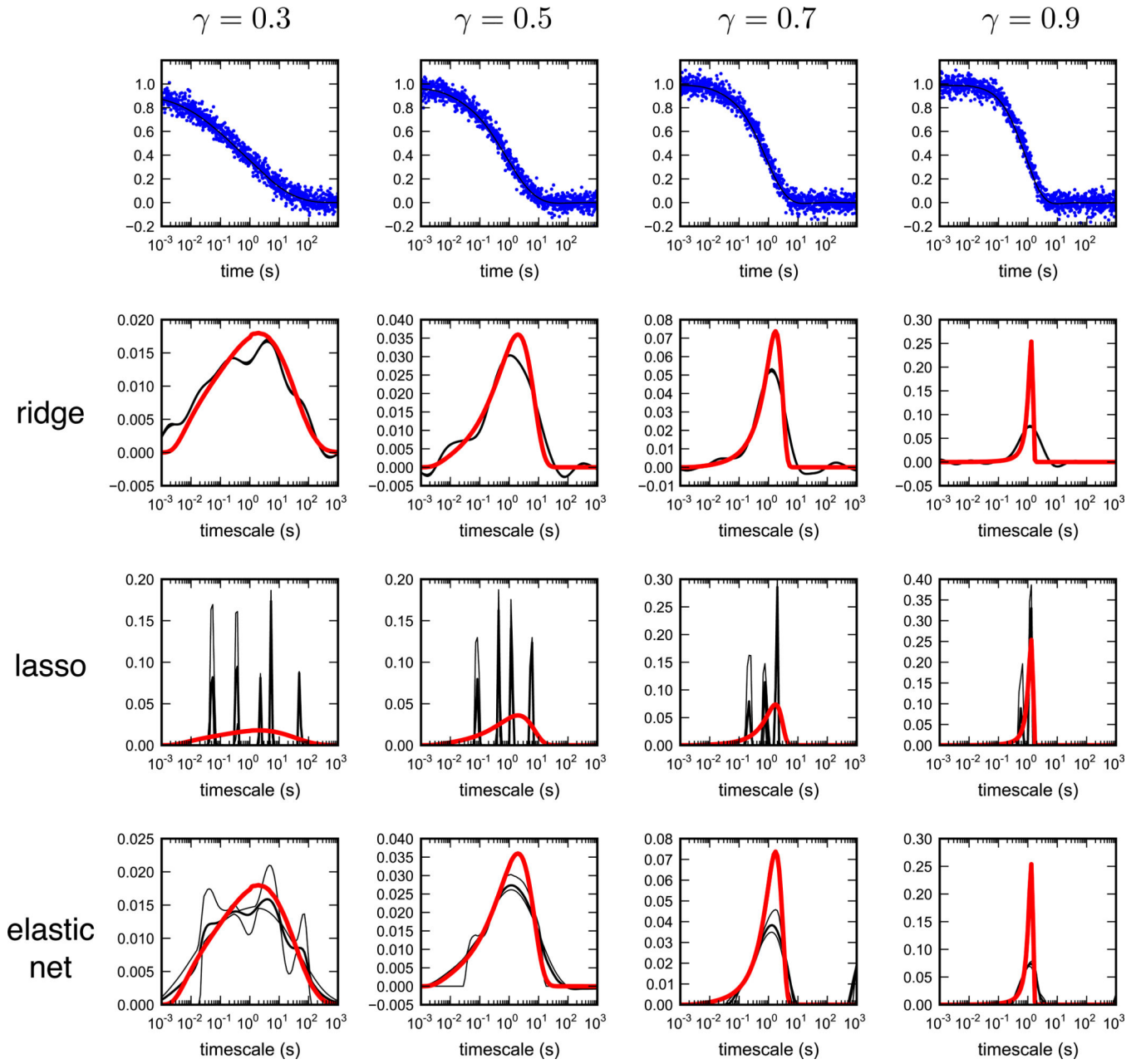
Figure 4.

Recapitulation of stretched-exponential rate spectra from noisy time series. (top) Noisy data sets ($s = 0.05$, blue) with best-fit time traces $\hat{y}$ (shown are the results from ridge regression; best-fits from lasso and elastic net are nearly indistinguighable) for various stretching parameters $\gamma = 0.3$, 0.5, 0.7 and 0.9. (bottom) Regularized rate spectra (shown here as the expectation over all posterior samples of λ) for ridge, lasso and elastic net regression. The ridge and elastic net rate spectra correspond well to the analytical results (red line), while lasso regression consistently describes the data with the smallest possible number of timescales. Note that the elastic net posterior takes longer to converge, as can be seen in by the larger error estimates from posterior sampling.
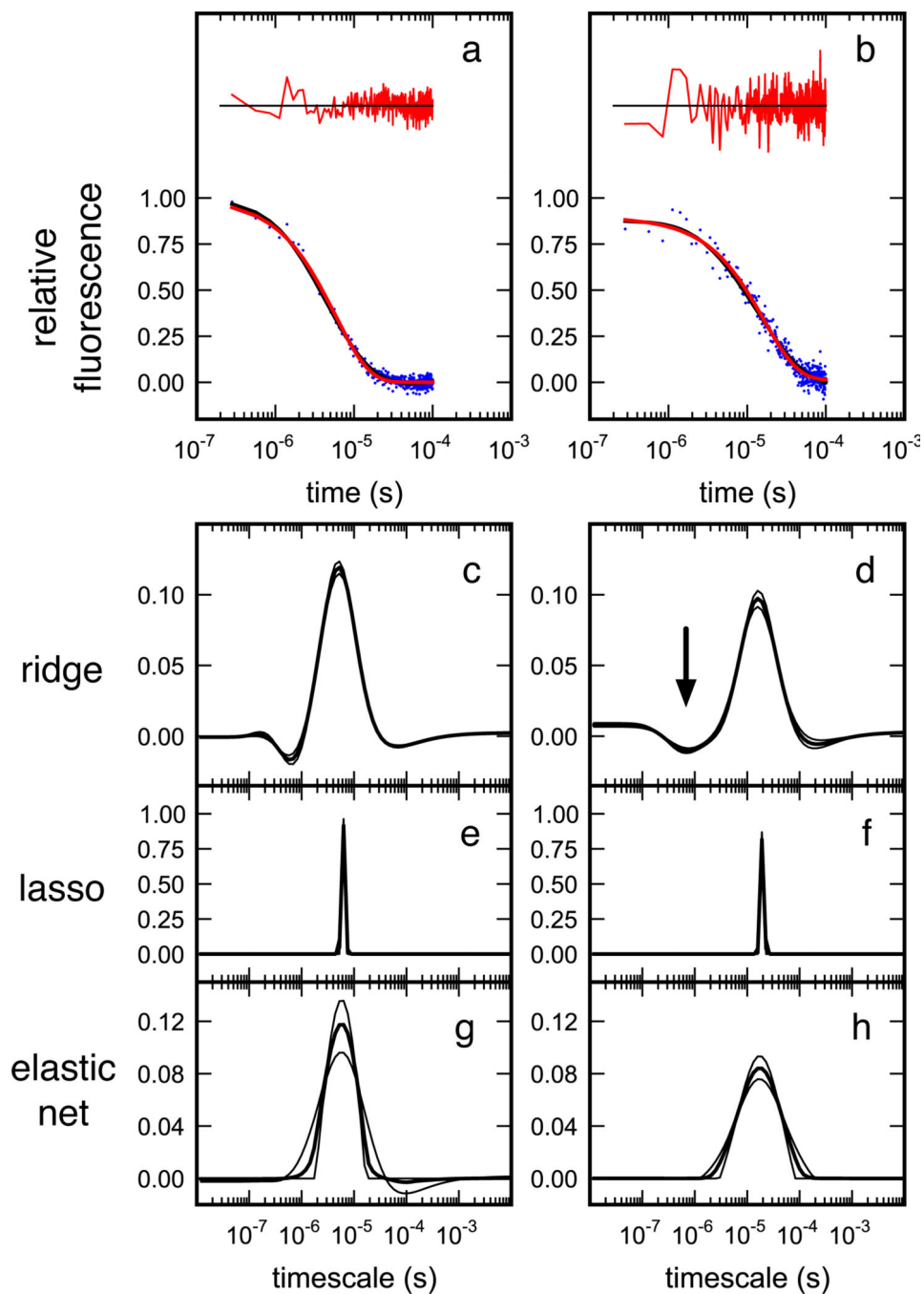
Figure 5.
Relative fluorescence time traces reprinted from Liu et al. (2008): Relaxation kinetics of a
WW domain (variant 20) after T-jump to (a) the melting temperature, and (b) 19° C below
the melting temperature. Single-exponential fits to the time traces and their residuals are
shown in red; ridge regression fits are shown in black. Rate spectra for these two traces are
shown for (c,d) ridge regression, (e,f) lasso regression, and (g,h) elastic net regression.

Ridge regression identifies a broadened (negative amplitude) peak near the 1 μs timescale (marked with arrow).