



Challenges in quantitative approaches to protein-ligand binding: Studies using isothermal titration calorimetry and alchemical binding free energy calculations

Bas Rustenburg¹, John Chodera¹, and David Minh²

¹Memorial Sloan-Kettering Cancer Center

²Illinois Institute of Technology

Introduction

Among the most fundamental molecular interactions in biology are those of small molecules with their proteins. Endogenous small molecules play the role of messengers in many signaling pathways of the cell, and small molecule drugs interact with proteins in signaling cascades to modulate their function. **Despite having catalogued many of the physical driving forces behind small molecule recognition, there are enormous gaps in our knowledge preventing us from articulating a quantitative, predictive understanding of small molecule affinities and selectivities for biomolecules.** In principle, *alchemical free energy calculations* provide a framework for quantitatively describing all aspects of the thermodynamics of small molecule recognition. However, deficiencies in our quantitative understanding of binding create large challenges in the ability of these calculations to reproduce experimental affinities in many systems, holding back their use in probing function and aiding design.

Issues with experiments: Improper uncertainty estimation

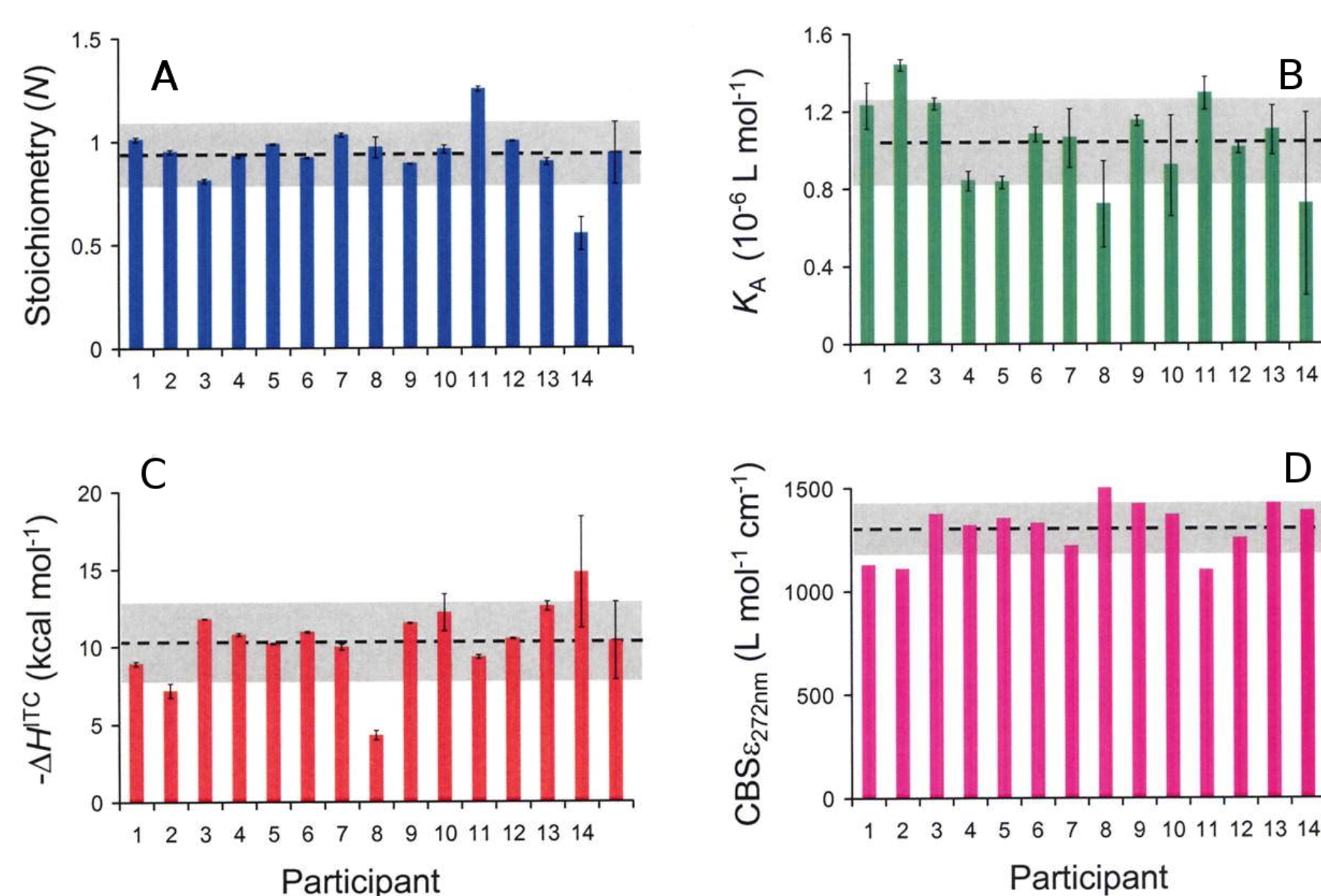


Fig. 1: Binding measurements of CBS to bovine carbonic anhydrase II from the ABRF-MIRG'2 study. **Independent ITC as performed by 14 labs shows individual error estimates are orders of magnitude smaller than actual experimental variation.** A: Predicted stoichiometry of binding. B: Association constant K_A . C: enthalpic contribution to binding. D: Measured extinction coefficient of ligand CBS, as reported by the 14 participants [1].

Reliable baseline estimates

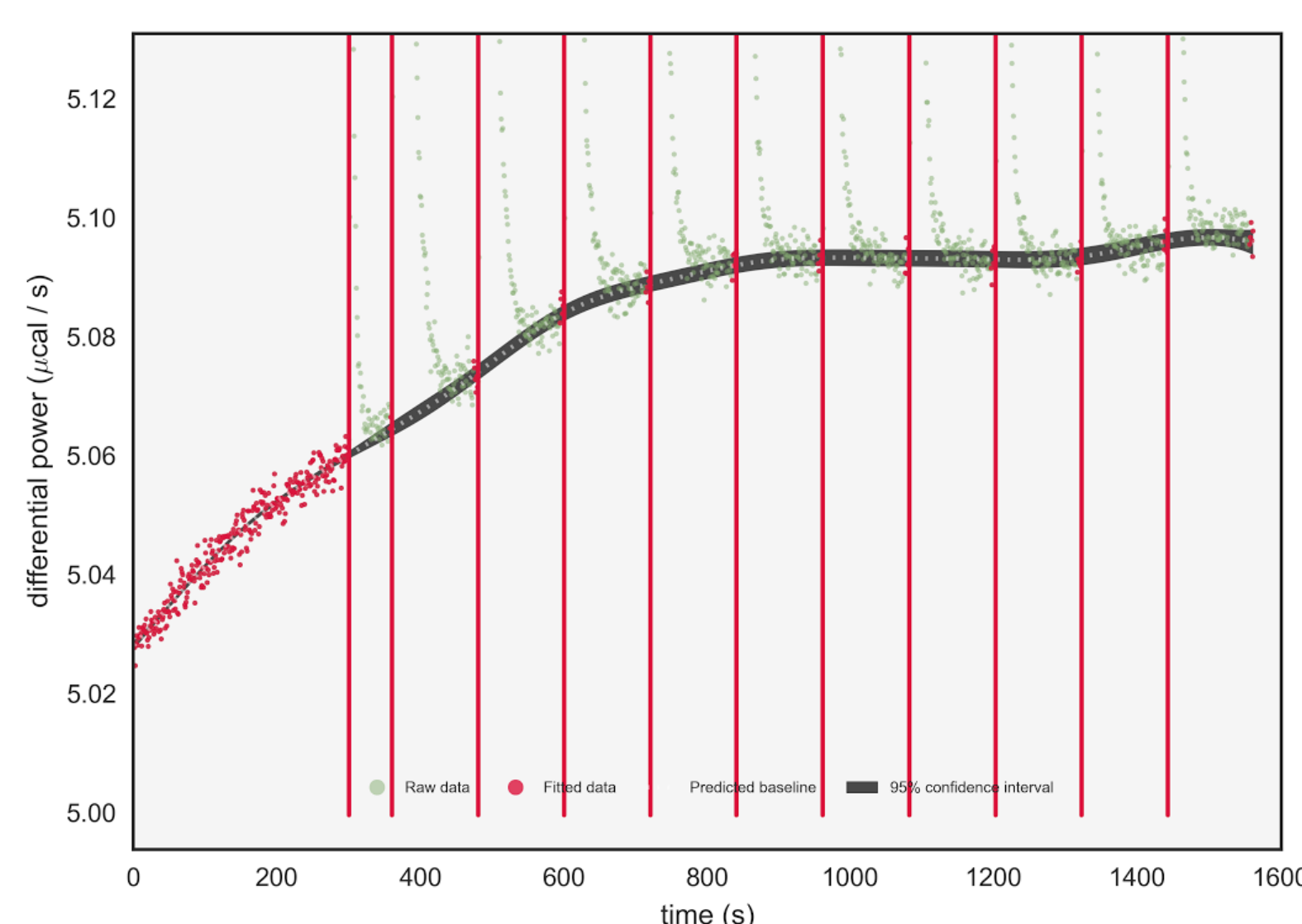


Fig. 2: We apply Gaussian process regression to increase the reliability of our baseline estimates.

A typical ITC experiment

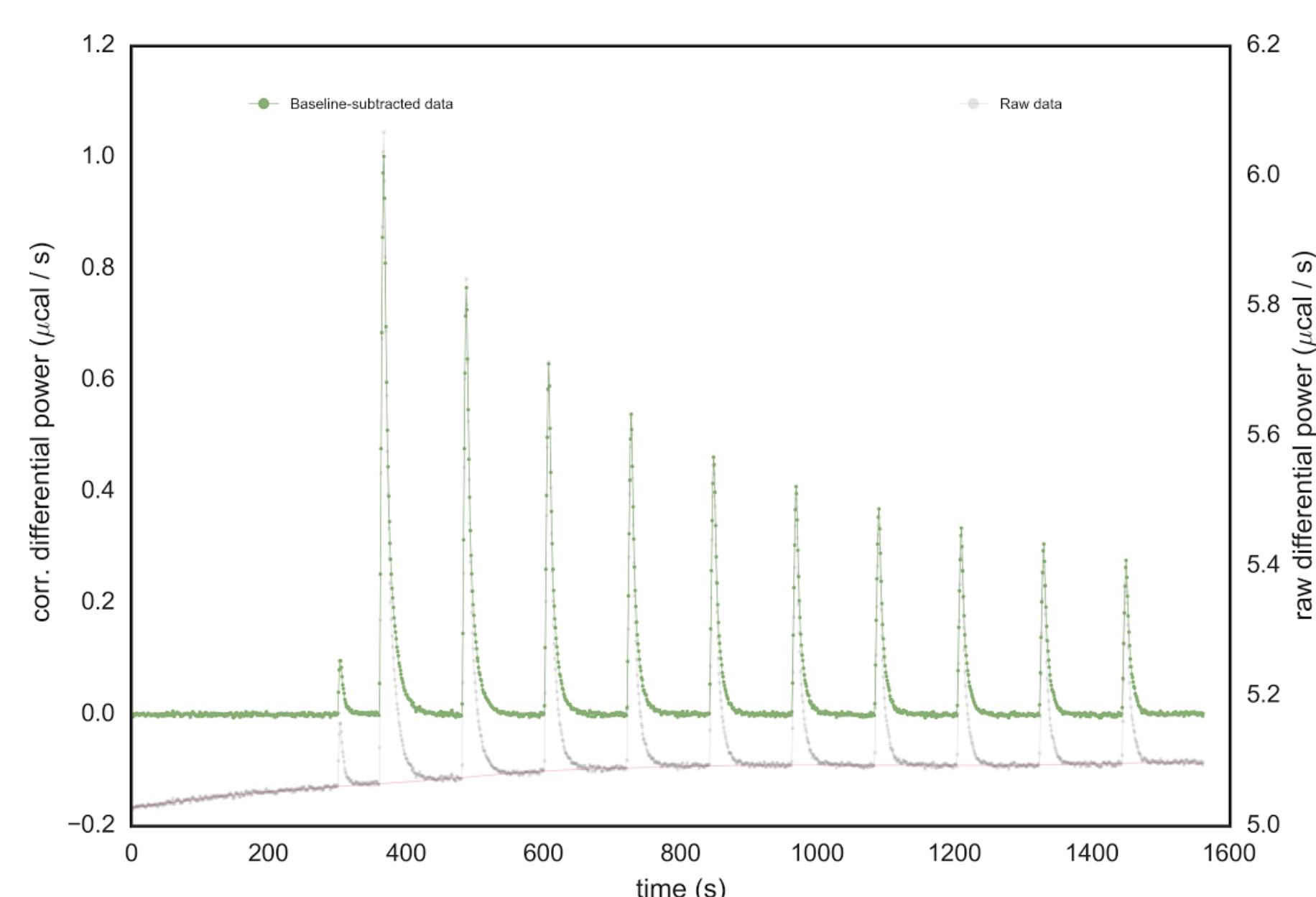


Fig. 3: In an ITC experiment, we inject from a syringe into a sample cell several times, measuring a differential power, and then integrating over that to obtain the heat of the injection, q_n^{obs} .

Posterior predictions of injected heats

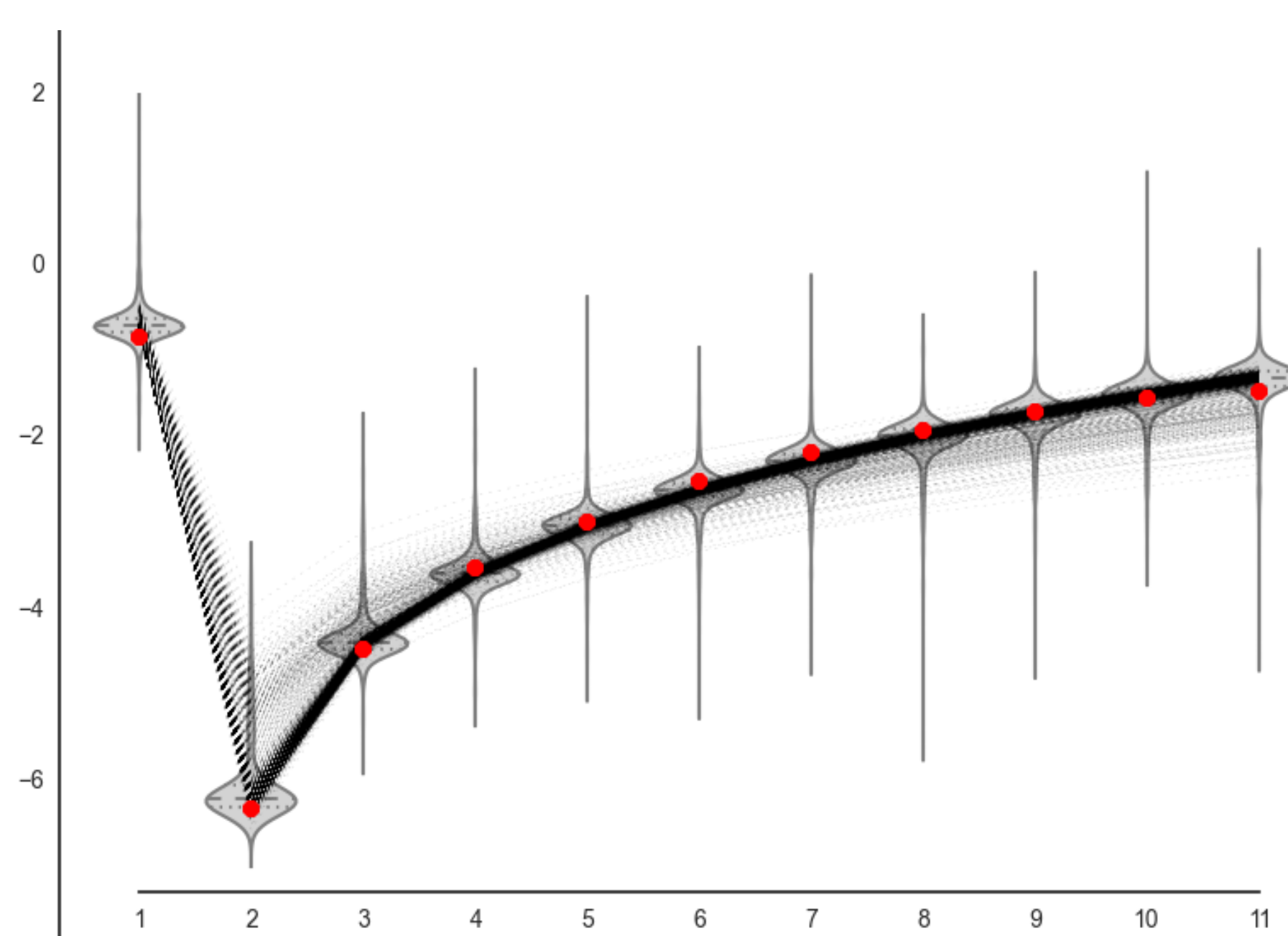


Fig. 4: Our observations (red dots) and our sampled posterior heats (violin plots) provide us with new estimates plus credible intervals.

From likelihood to posterior density

Using Bayes rule, we can obtain a posterior distribution $\mathcal{P}(\theta|\mathcal{D})$

$$\mathcal{P}(\theta|\mathcal{D}) \propto \mathcal{P}(\mathcal{D}|\theta)\mathcal{P}(\theta) \quad (1)$$

Here, $\mathcal{P}(\theta)$ is a prior density of our parameters:

$$\theta = \{\Delta G_{\text{bind}}, \Delta H_{\text{bind}}, \Delta H_0, [\text{X}_{\text{syr}}], [\text{M}_{\text{cell}}], \sigma\} \quad (2)$$

for which the distributions can be defined using prior knowledge about errors made by instrumentation.

Modeling the likelihood of the integrated heats

According to the central limit theorem (CLT), we can model the integrated heats as being samples from a normal distribution \mathcal{N} ,

$$q_n^{\text{obs}} \sim \mathcal{N}(q_n^{\text{true}}, \sigma^2) \quad (3)$$

with the true heats q_n^{true} as a mean, with a variance of σ^2 .

Markov chain Monte Carlo gives us probabilities for all our parameters

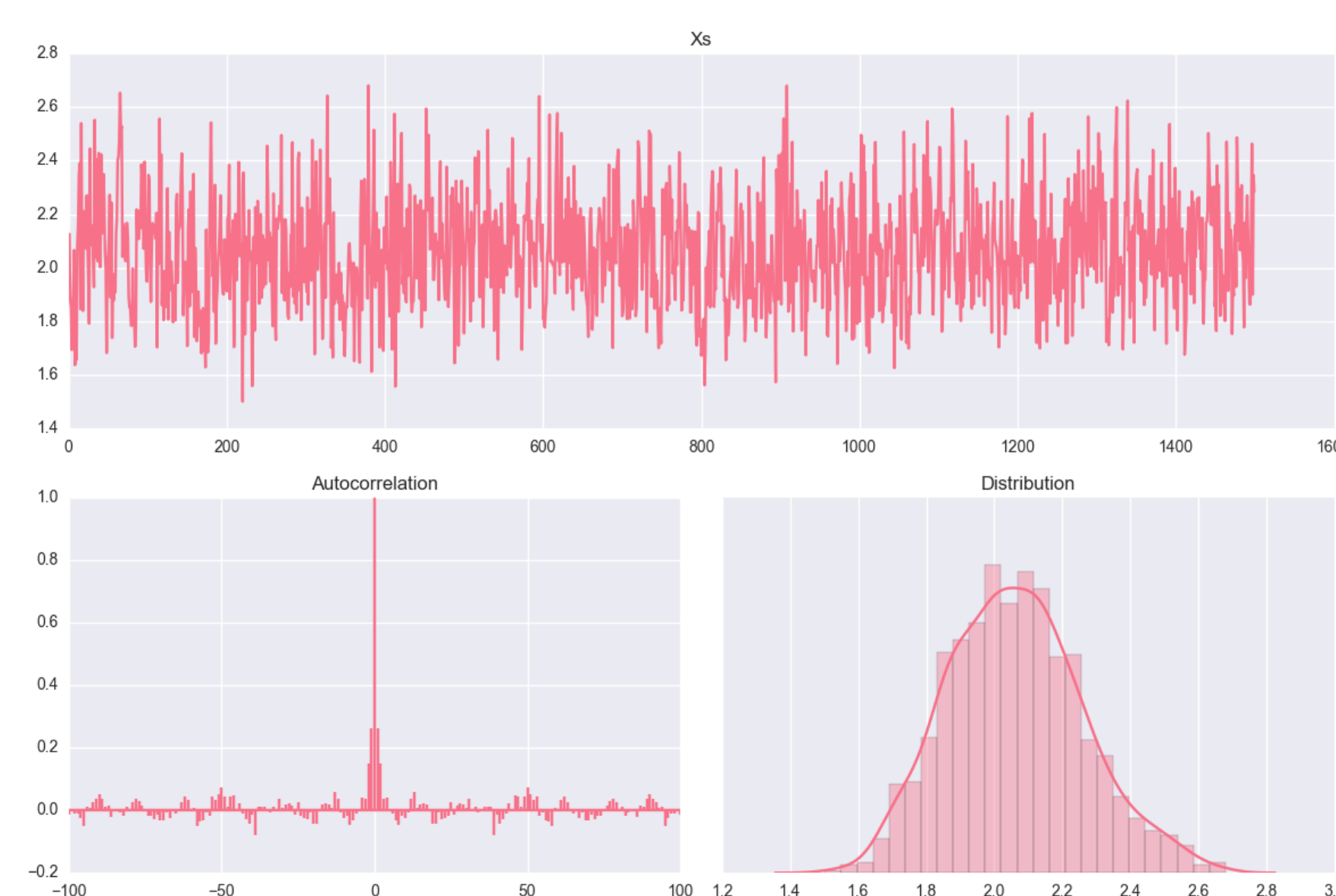


Fig. 5: An example distribution sampled for the syringe concentration.

Sample Block 4