Challenges in quantitative approaches to protein-ligand binding: Studies using isothermal titration calorimetry

Bas Rustenburg[1,2], John Chodera[1,2], and David Minh[3]

[1] Computational Biology, Memorial Sloan Kettering Cancer Center
[2] Physiology, Biophysics and Systems Biology, Weill Cornell Medical College
[3] Chemistry Division, Illinois Institute of Technology

Introduction

Among the most fundamental molecular interactions in biology are those of small molecules with their proteins. **However, deficiencies in the estimation of uncertainty create large challenges in the ability of these ITC experiments to be used in a quantitative way, holding back their use in probing function and aiding design.** For instance, most existing analysis procedures fail to incorporate errors in reagent concentrations, which is commonly kept a fixed parameter, whereas previous studies indicate possible errors of 10 % that are not propagated.

A typical ITC experiment

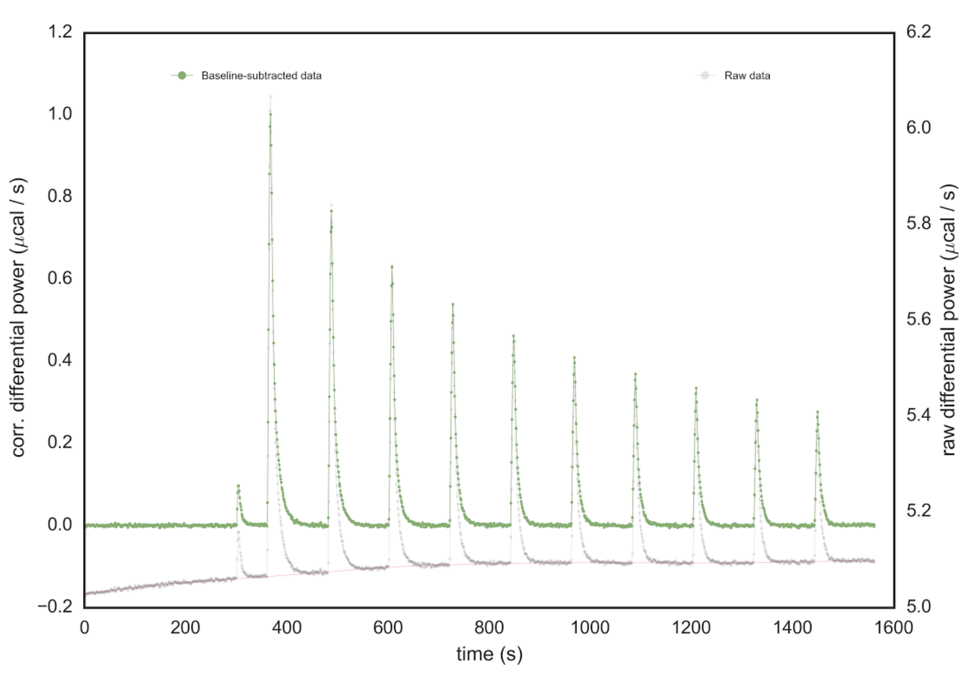

Fig. 1: In an ITC experiment, we inject from a syringe into a sample cell several times, measuring a differential power, and then integrating over that to obtain the heat of the injection, q_n^{obs}.

Bayesian ITC

The posterior distribution is defined as

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

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

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

which we use to propagate instrumental errors.

Reliable baseline estimates

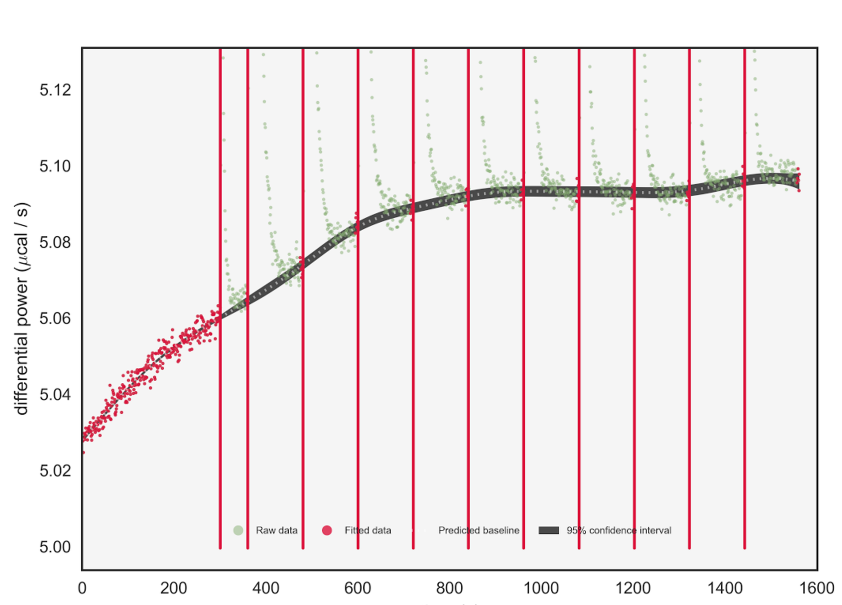

Fig. 2: We apply Gaussian process regression to increase the reliability of our baseline estimates, using scikit learn [1].

Uncertainty estimation

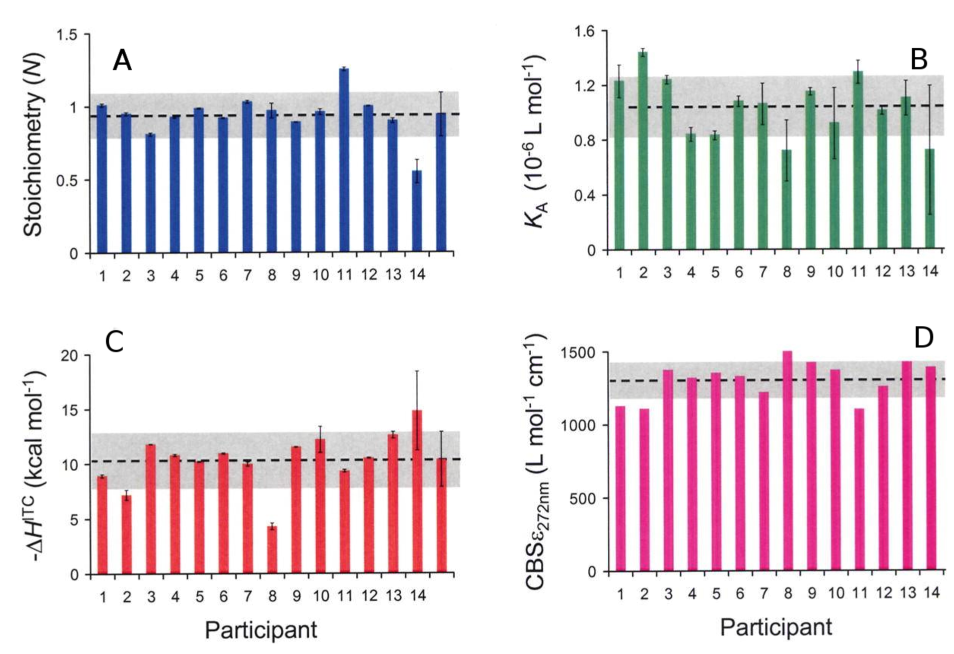

Fig. 3: Binding measurements of CBS to bovine carbonic anhydrase II from the ABRF-MIRG'2 study. A: Stoichiometry. B: Association constant. C: Binding enthalpy. D: Extinction coefficient of CBS, as reported by 14 participants [2].

MCMC sampling

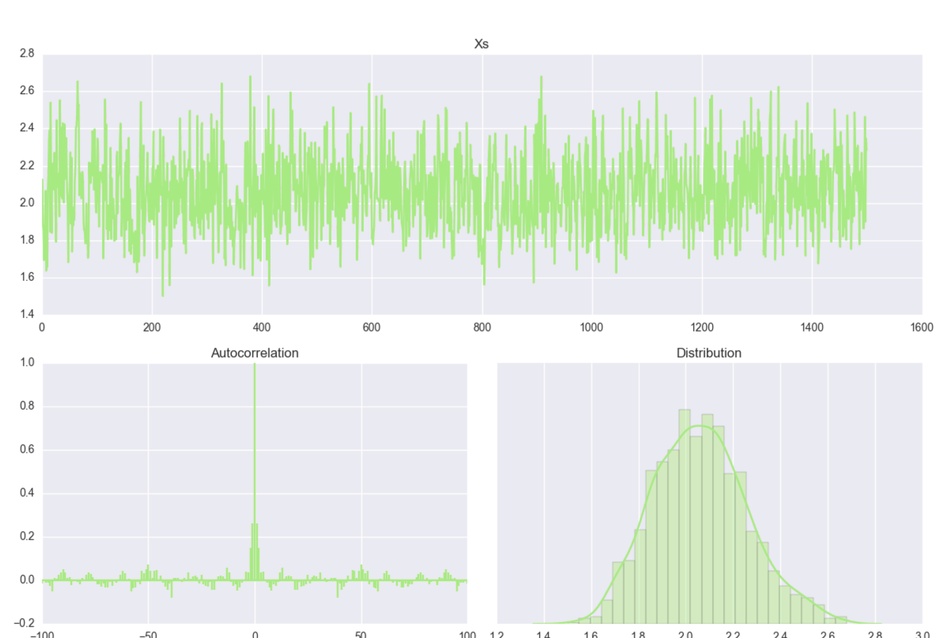

Fig. 4: An example distribution sampled for the syringe concentration using pymc [3].

Conclusions

- Not propagating errors in concentrations leads to large underestimation of uncertainty.
- Using Bayesian inference, we can incorporate prior information into our modeling
- MCMC will then give us posterior distributions with more accurate uncertainty estimates
- This allows us to use ITC experiments as a means of validating free energy calculations

Observation model

We model the integrated heats as being samples from a normal distribution $\mathcal{N}$,

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

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

Posterior predictions

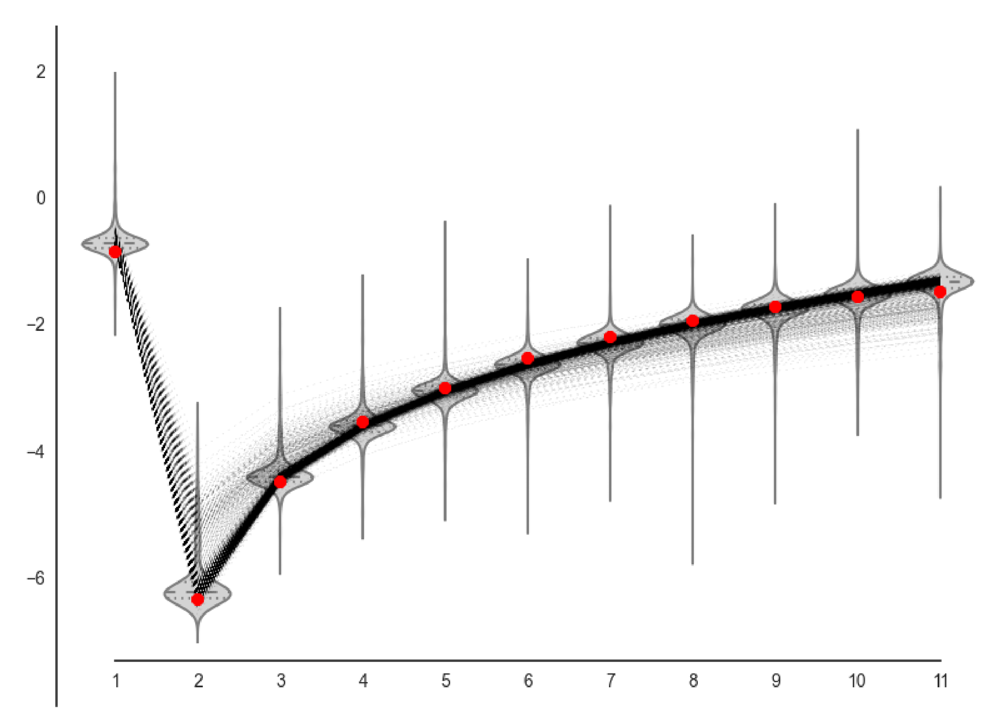

Fig. 5: Our observations (red dots) and our sampled posterior heats (violins) provide us with new estimates plus credible intervals. Model traces are shown as dotted lines.

References

[1] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J Mach Learn Res* 12 (2011), pp. 2825–2830.

[2] D. G. Myszka et al. "The ABRF-MIRG'02 study: assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction." eng. In: *J Biomol Tech* 14.4 (Dec. 2003), pp. 247–269.

[3] A. Patil, D. Huard, and C. J. Fonnesbeck. "PyMC: Bayesian Stochastic Modelling in Python." eng. In: *J Stat Softw* 35.4 (July 2010), pp. 1–81.

LaTeX TikZposter