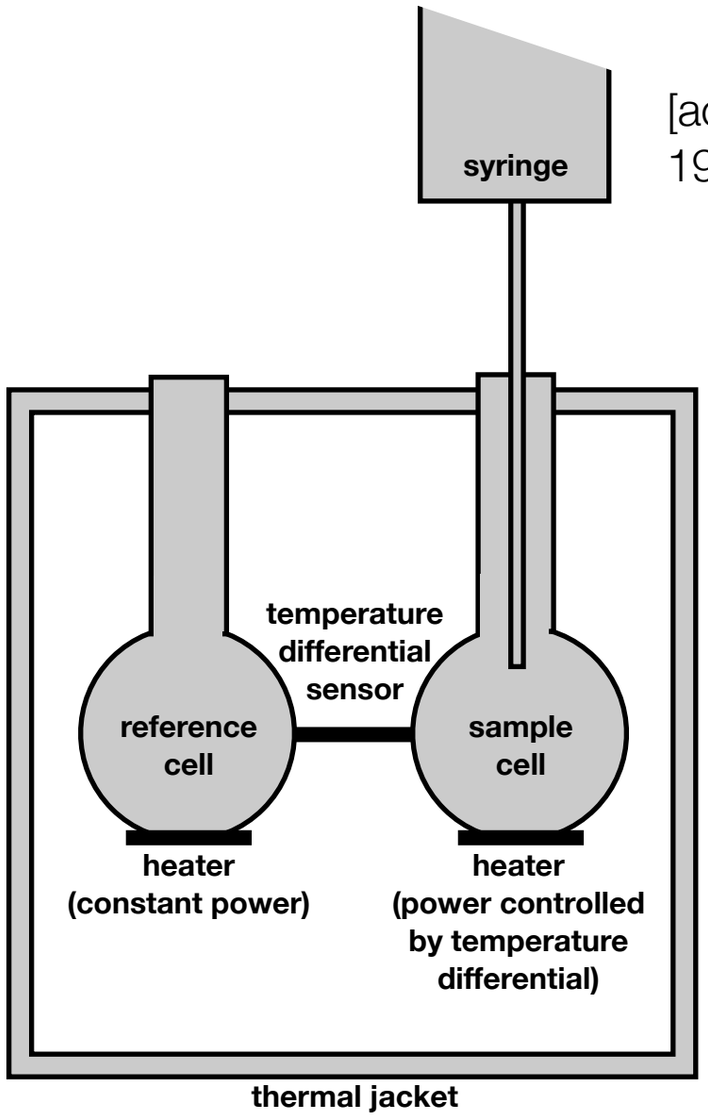
Bayesian analysis of isothermal titration calorimetry data

John D. Chodera and **Vijay S. Pande**
Pande group, Department of Chemistry, Stanford University

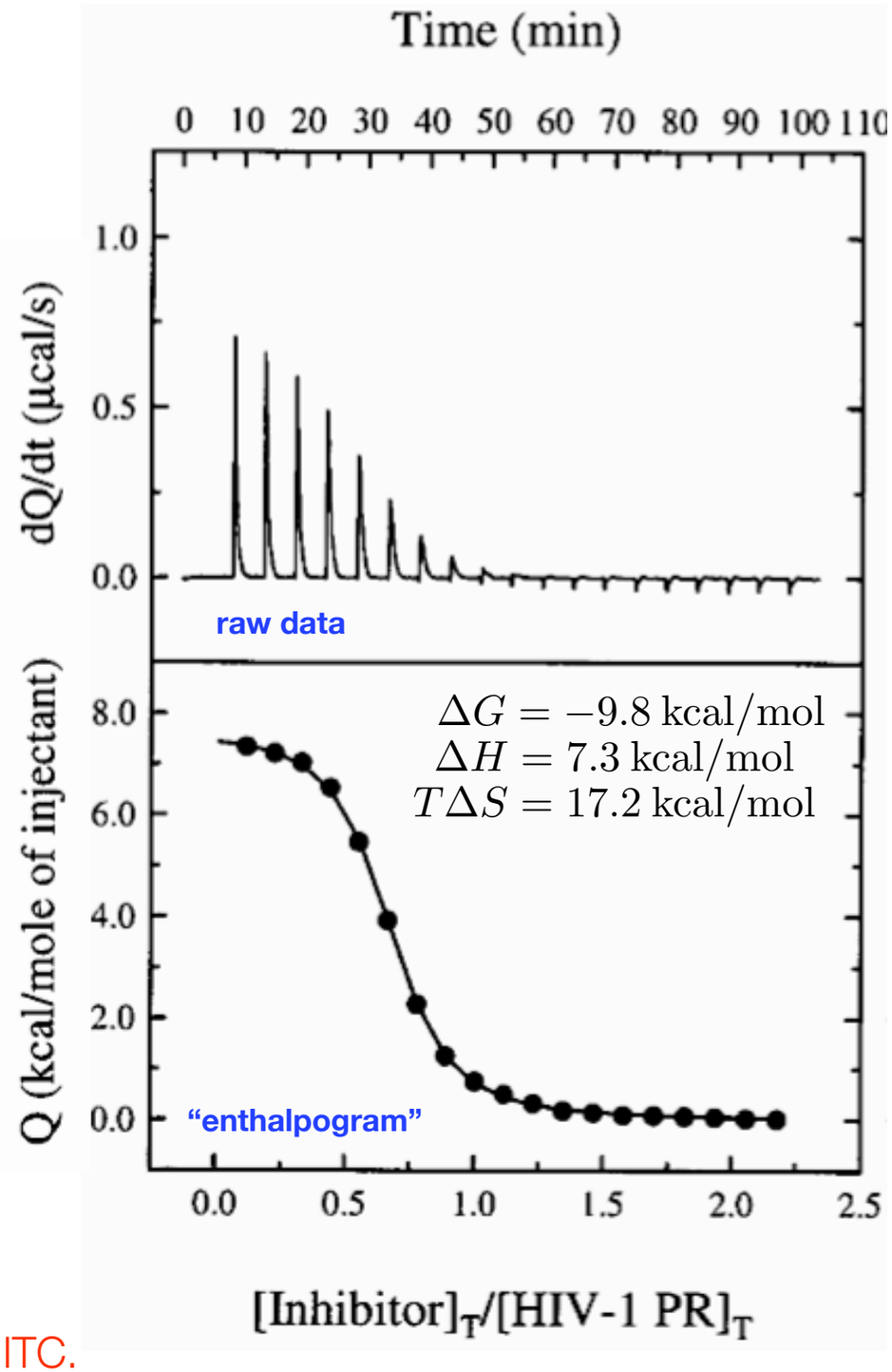
jchodera@stanford.edu

Isothermal titration calorimetry (ITC)



[acetyl pepstatin] = 200 uM
19 injections @ 10 uL/injection

[HIV-1 PR] = 20 uM in 1.4 mL
(0.6 mg/cell)



Only method that simultaneously provides estimates of both ΔG and ΔH.

Some reactions have no measurable change in heat, and are not measurable by ITC.

Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

Curvature changes with binding affinity

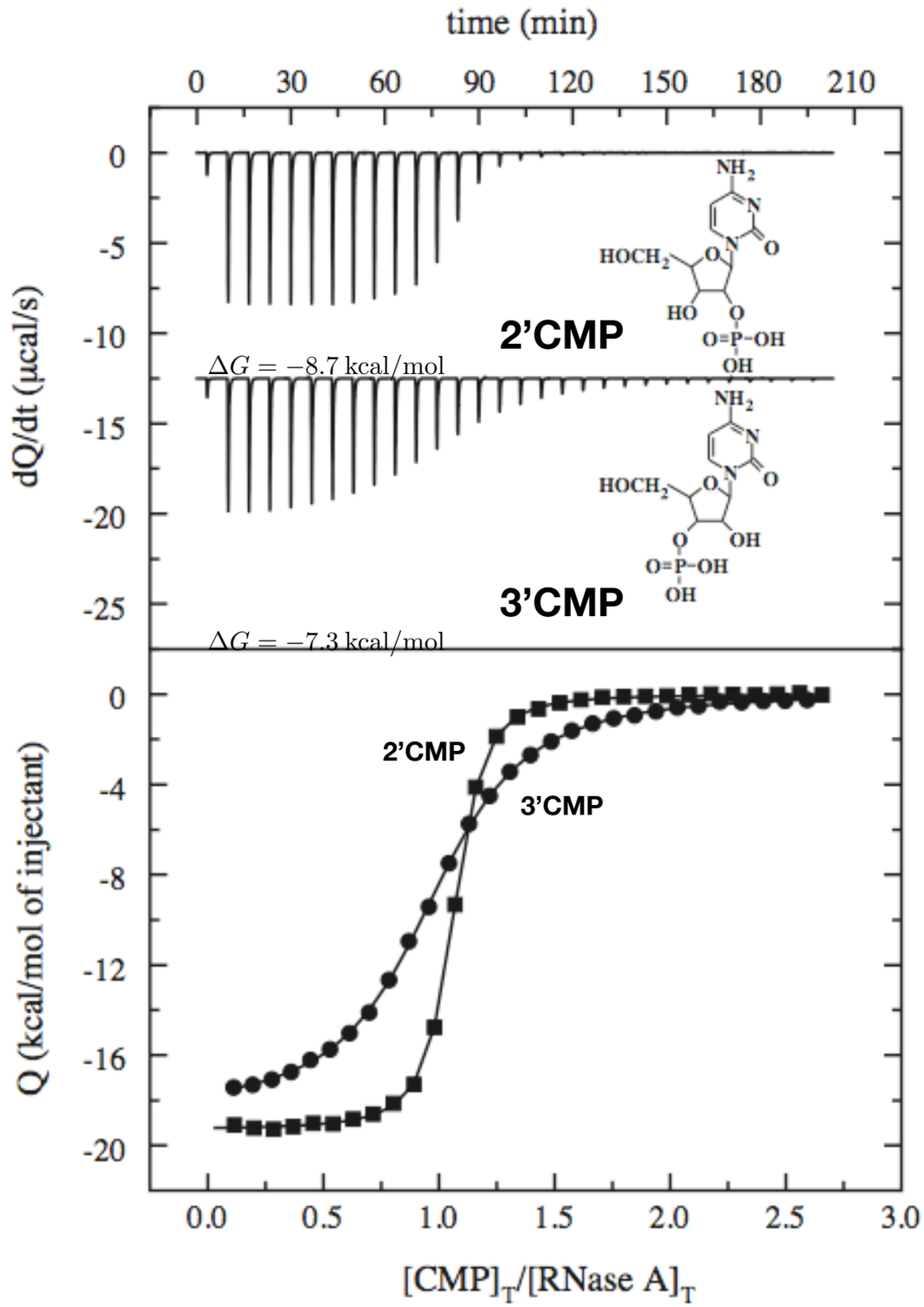
tighter binding gives higher curvature,
limiting effective measurement range:

$$0.1 < [P]_T/K_d < 1000$$

for $[P]_T$ ~ 10 uM, this limits practical K_d range from 10 nM - 100 uM (5 - 11 kcal/mol)

because they are determined by very different features, ΔG, ΔH, and TΔS may have very different (and coupled) uncertainties

What are the real uncertainties in ΔG, ΔH, and TΔS?
How are they related?



Binding model

simple association is described by equilibrium dissociation constant

$$K_d = \exp[\beta \Delta G_b](1\,M) = \frac{[P]_n[L]_n}{[PL]_n}$$ (in units of molarity M)

concentrations after *n* injections $[P]_n, [L]_n, [PL]_n$ are determined by solving nonlinear equations

given additional constraints on total quantity of ligand and protein:

$$L_n = n\Delta V[L]_s = V_n([L]_n + [PL]_n)$$ total quantity of ligand in cell after *n* injections

$$P = V_0[P]_0 = V_n([P]_n + [PL]_n)$$ total quantity of protein in cell (constant)

input → output

$$n, \beta, \Delta G, P, \Delta V, V_0 \longrightarrow [P]_n, [L]_n, [PL]_n$$

These equations can be solved to produce an analytical solution:

$$V_n[PL]_n = \frac{1}{2}\left\{(V_nK_d + L_n + P) - \left[(V_nK_d + L_n + P)^2 - 4L_nP\right]\right\}^{1/2}$$

Binding thermodynamics

binding can either liberate heat (exothermic) or consume heat (endothermic):

$$P + L \overset{\Delta H}{\rightarrow} PL$$

the heat evolved can be written as

$$Q_n = \Delta H \cdot V_n [PL]_n + n\Delta H_0 \qquad \text{heat potential}$$

$$q_n^* = Q_n - Q_{n-1} \qquad \text{heat liberated on injection } n$$

where

ΔH enthalpy of binding

ΔH_0 heat of dilution / mechanical heat

$[PL]_n$ concentration of complex after injection *n*

$V_n = V_0 + n\Delta V$ volume of solution in cell after injection *n*

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

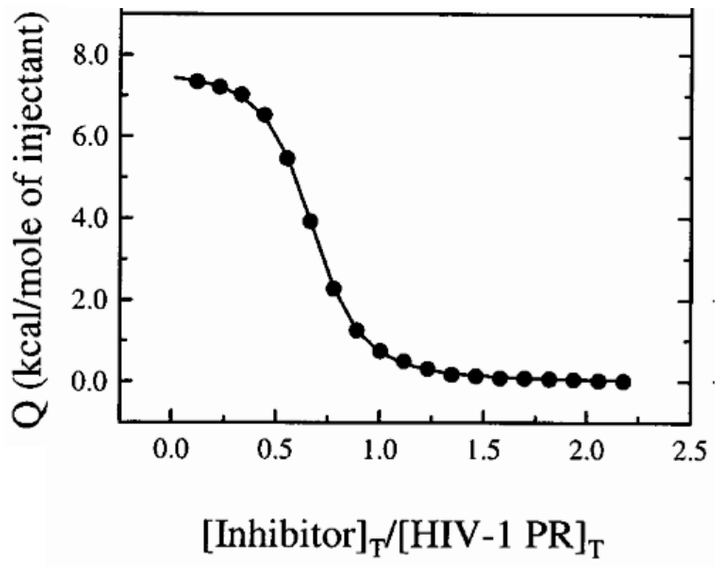
$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

$\mathcal{D} = \{q_1, q_2, \ldots, q_N\}$ measurements of evolved heat



Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

$\theta = \{\Delta G, \Delta H, T\Delta S, \Delta H_0\}$ thermodynamic parameters

$\Delta G = -kT \ln K_a$	free energy of binding
ΔH	enthalpy of binding
$T\Delta S$	entropic contribution to binding
ΔH_0	heat of dilution

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

$\mathcal{D}$ data

θ model parameters

$p(\theta|\mathcal{D})$ posterior

$p(\mathcal{D}|\theta)$ sampling distribution (model)

$p(\theta)$ prior

$$p(\mathcal{D}|\theta) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(q_n - q_n^*)^2}{2\sigma^2}\right]$$ Gaussian error model

q_n measured heat of injection *n*

q_n^* true heat of injection *n*

σ std dev of error in measured heat (nuisance parameter)

$$P + L \overset{\Delta H}{\rightarrow} PL$$

$$q_n^* = Q_n - Q_{n-1}$$

$$Q_n = \Delta H \cdot V_n[PL]_n + n\Delta H_0$$ heat potential

concentrations after *n* injections $[P]_n, [L]_n, [PL]_n$ determined by solving nonlinear equations given $\beta, \Delta G, P, n, \Delta l, V_n$

$L_n = n\Delta l[L]_s = V_n([L]_n + [PL]_n)$ total quantity of ligand in cell after *n* injections

$P = V_0[P]_0 = V_n([P]_n + [PL]_n)$ total quantity of protein in cell (constant)

$K_a = \exp[-\beta\Delta G] = \frac{[PL]_n}{[P]_n[L]_n}$ binding model

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

$p(\Delta G, \Delta H, \Delta H_0, \sigma) \propto \sigma^{-1}$ Jeffreys prior

$\Delta G, \Delta H, \Delta H_0$ can be of any sign and value

$\sigma > 0$ scale parameter; can be of any magnitude

(Later, could build in some *a priori* knowledge of instrument error or calibration runs.)

Analysis of ITC experiments: The Bayesian way

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

$$p(\theta|\mathcal{D}) = (2\pi)^{-N/2}\sigma^{-(N+1)}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(q_n - q_n^*)^2\right]$$ posterior

$\mathcal{D}$ data

θ model parameters

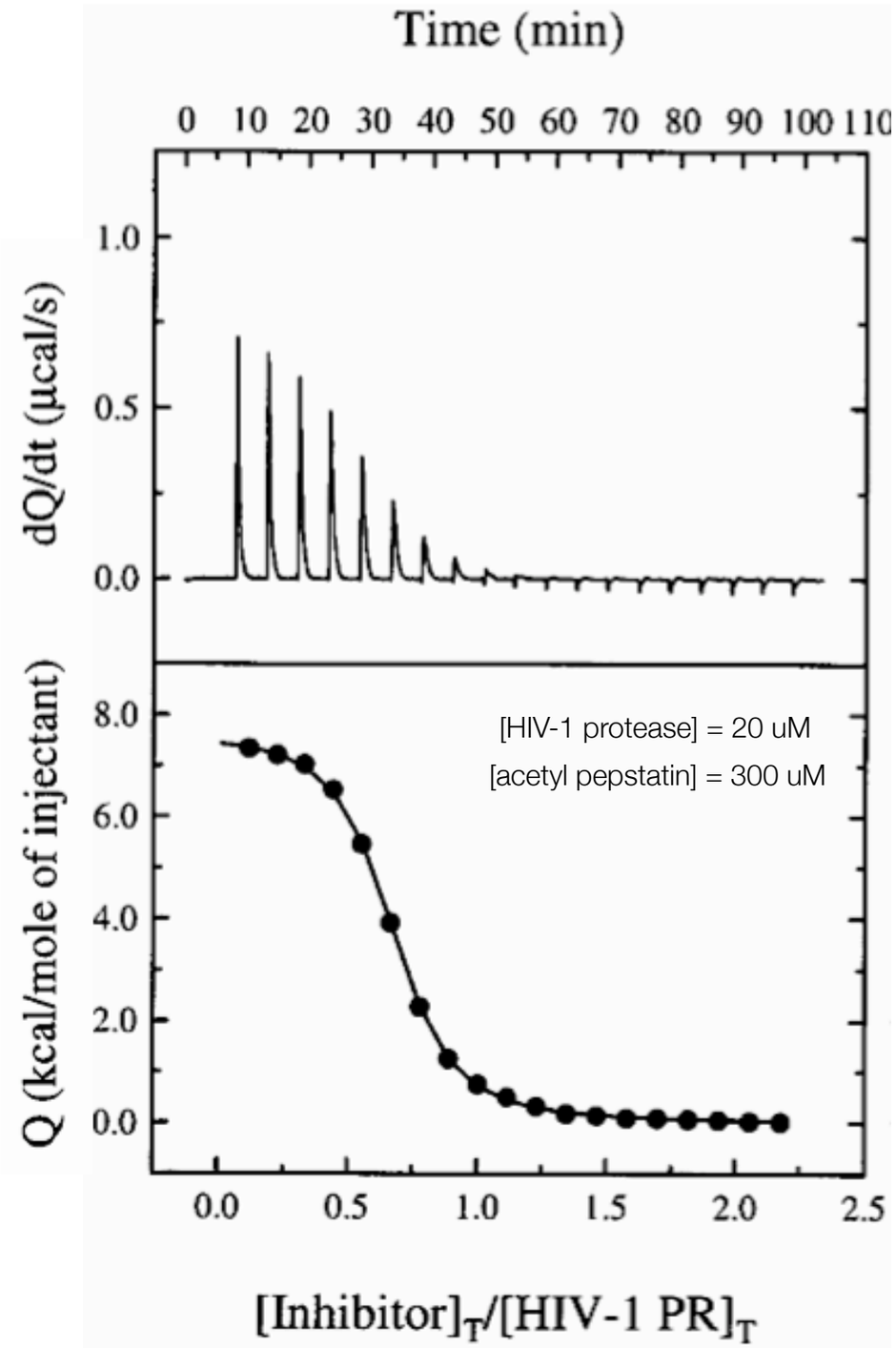
$p(\theta|\mathcal{D})$ posterior

$p(\mathcal{D}|\theta)$ sampling distribution (model)

$p(\theta)$ prior

Illustration of Bayesian method

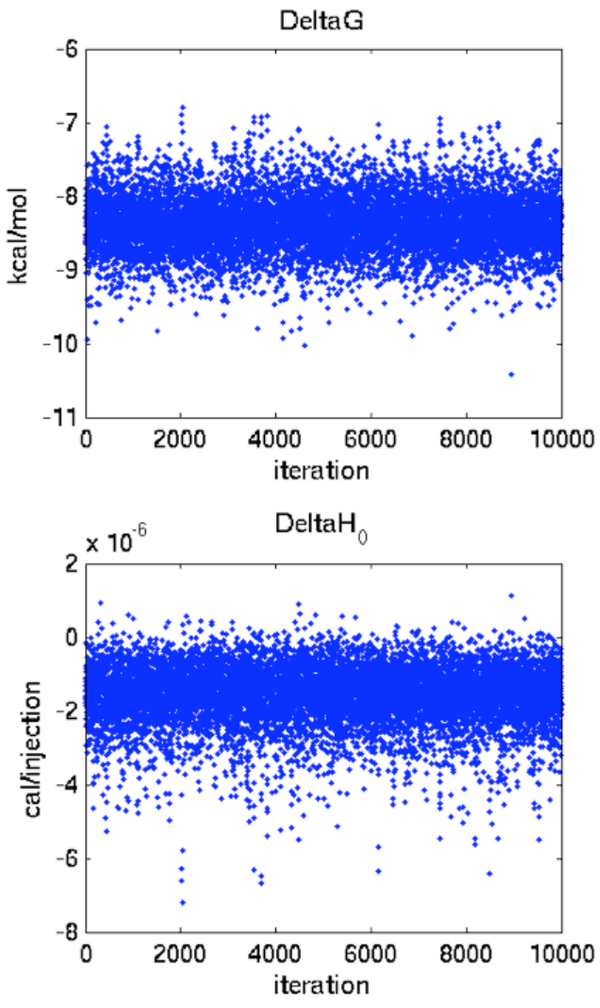
Consider the case of **acetyl pepstatin** binding to **HIV-1 protease**.

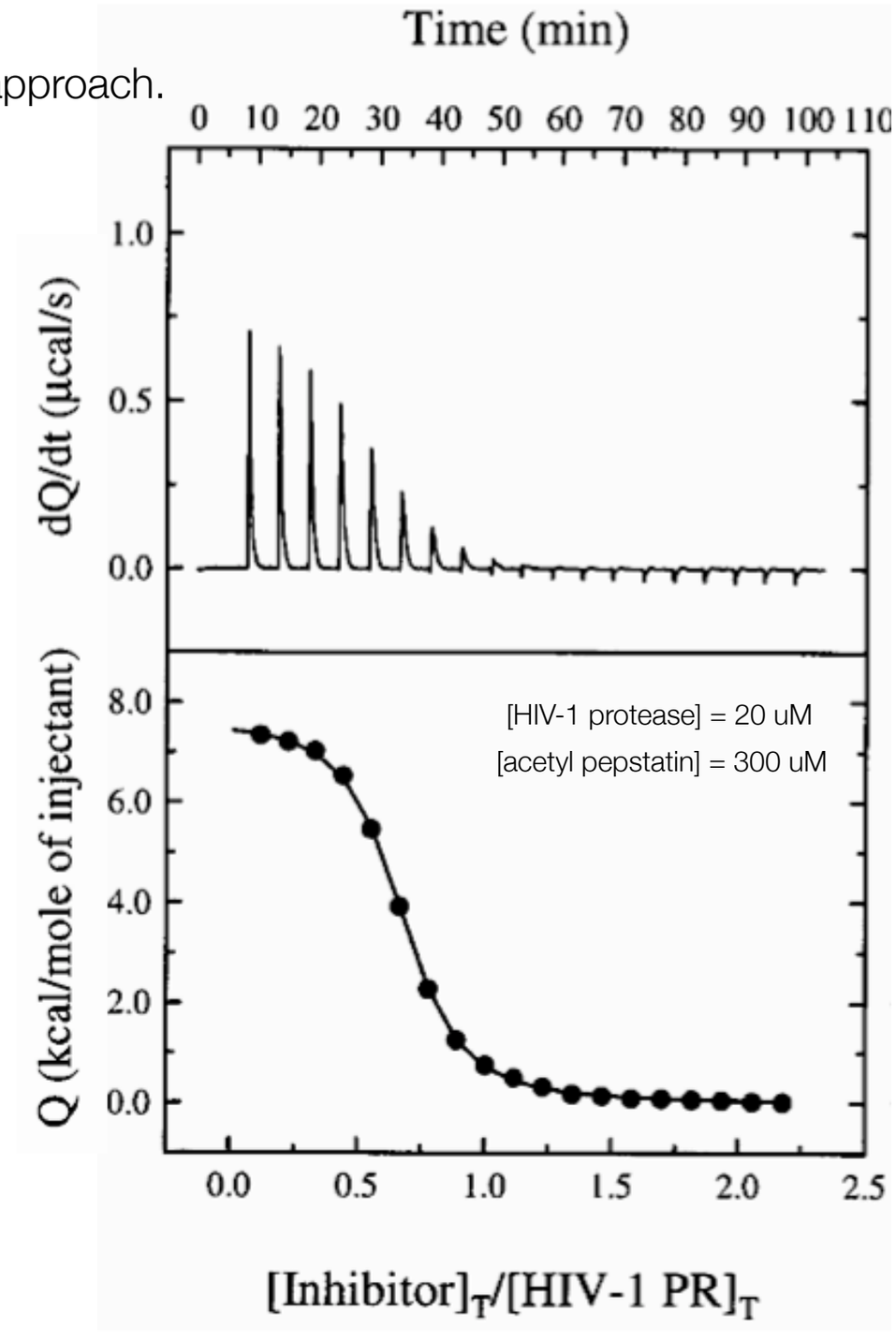


Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

Illustration of Bayesian method

We sample from the posterior density using a Metropolis Monte Carlo approach.

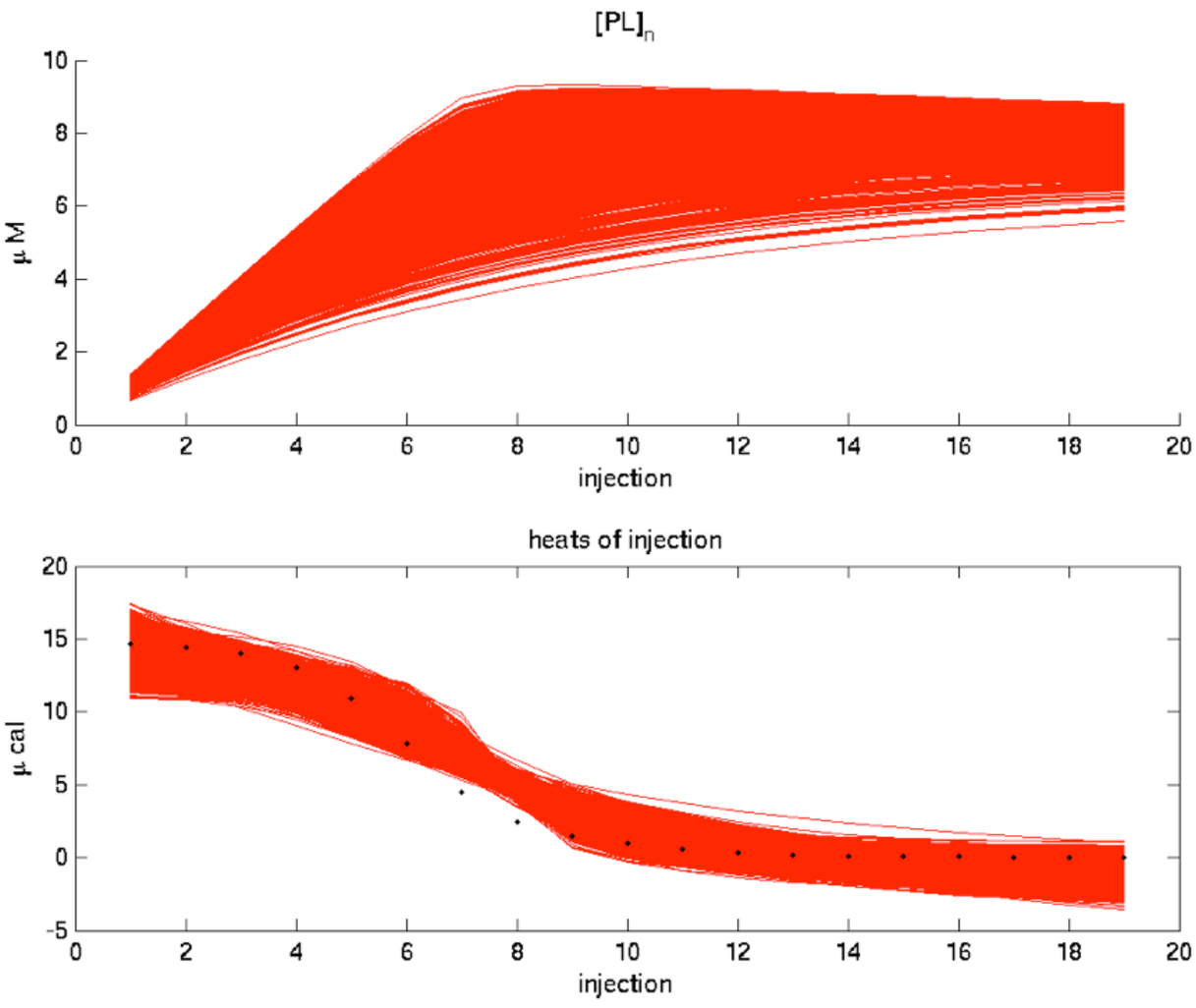


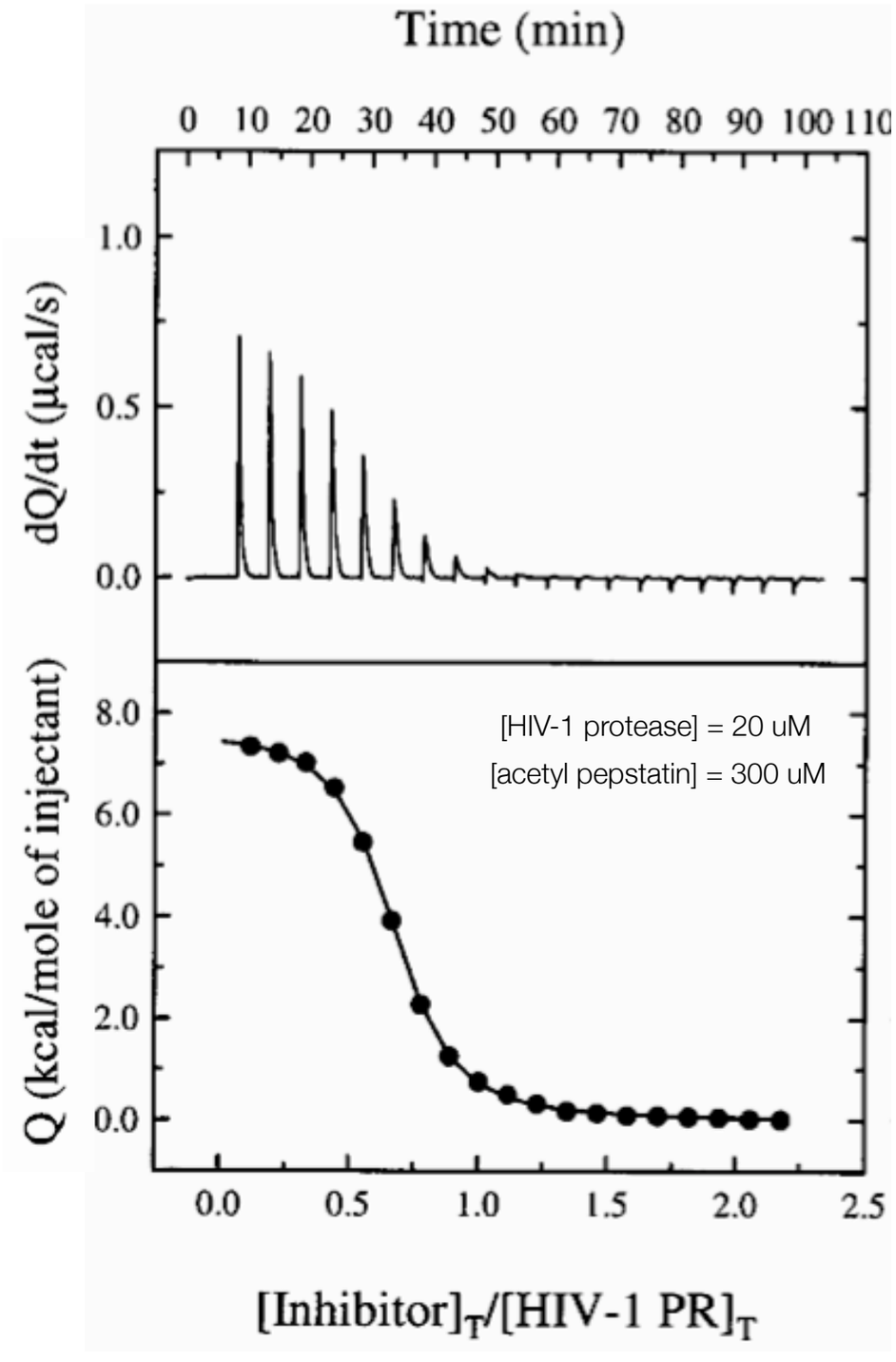


Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

Illustration of Bayesian method

The fit of each model to the data can be examined.

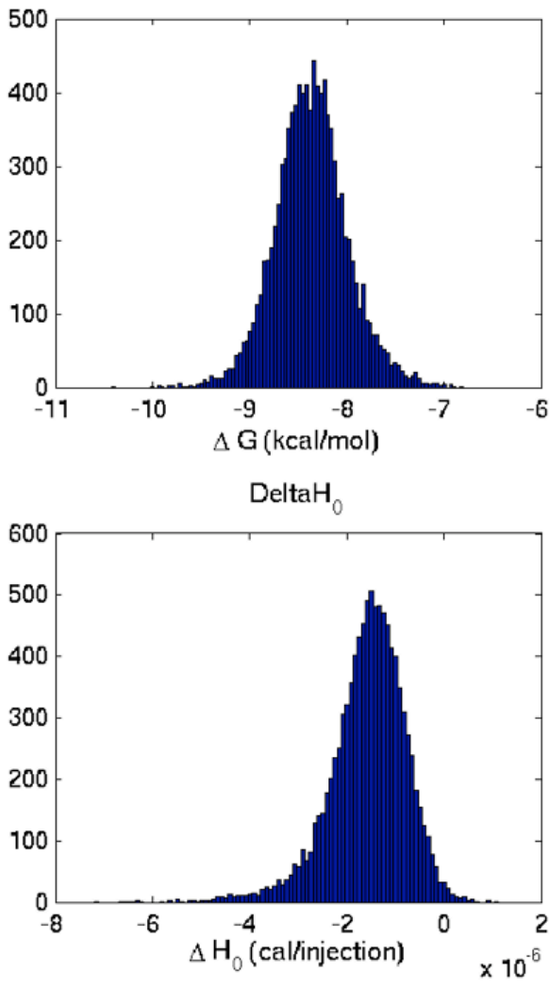


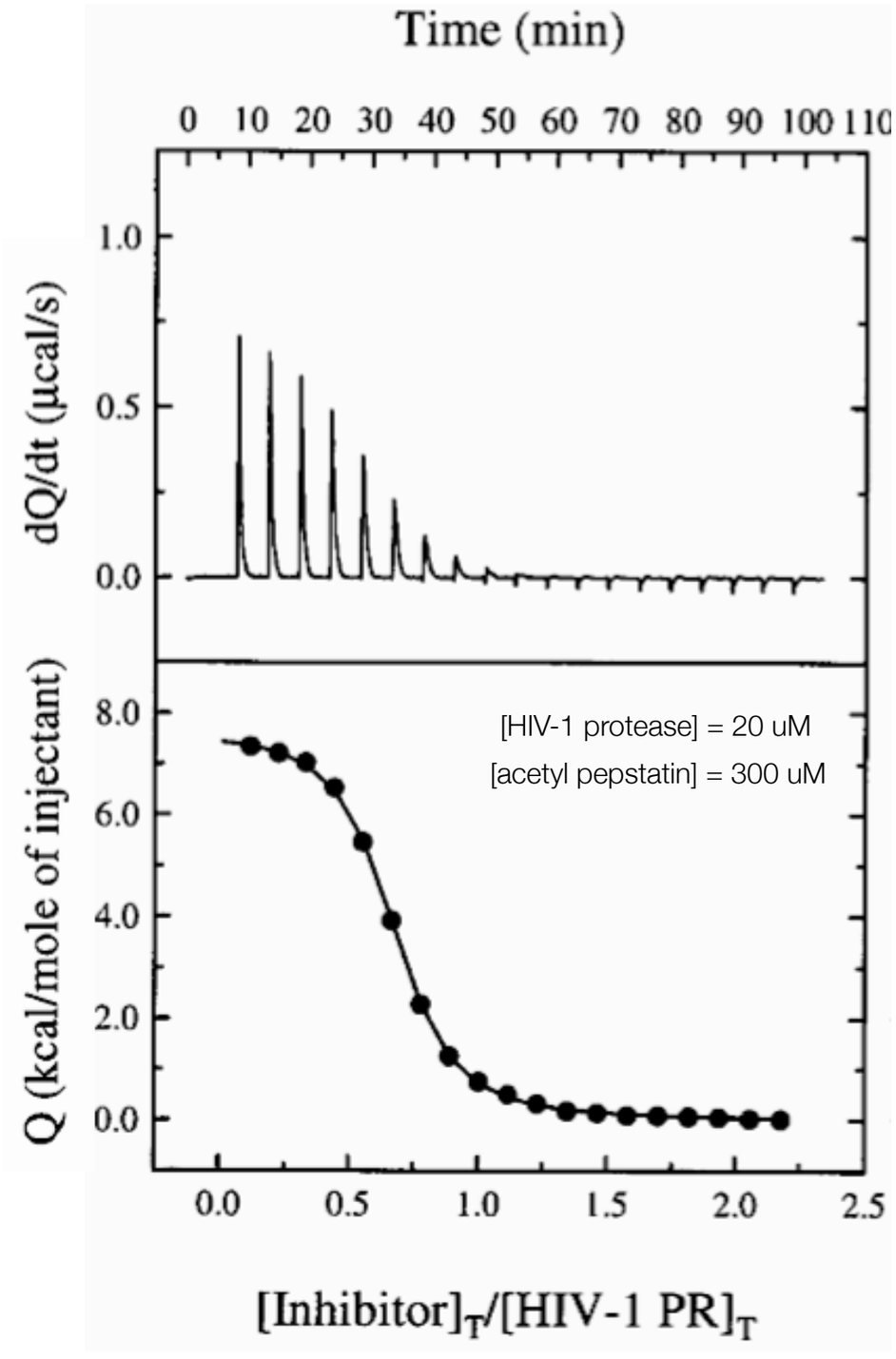


Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

Illustration of Bayesian method

We can compute **marginal posterior probability distributions** for each thermodynamic parameter to assess the uncertainty.

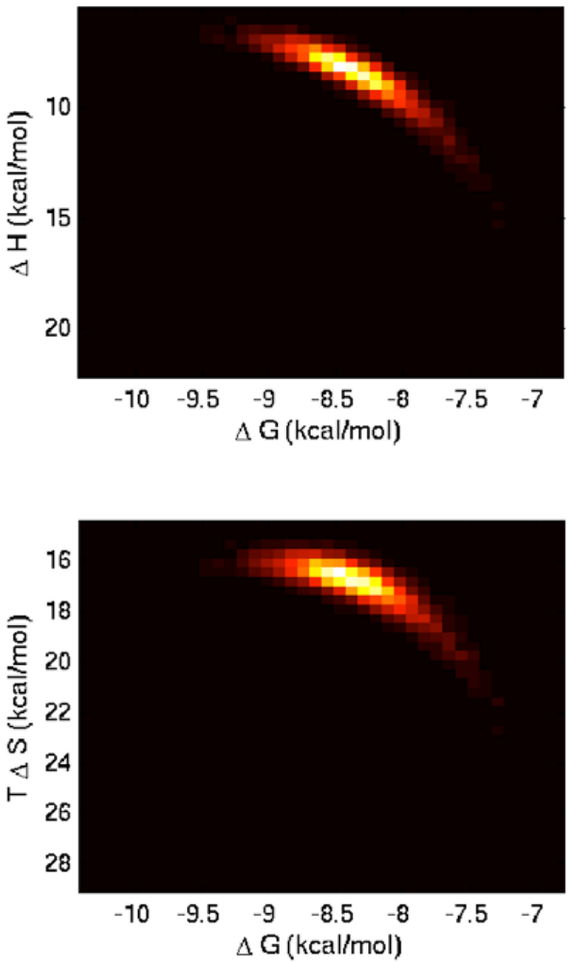


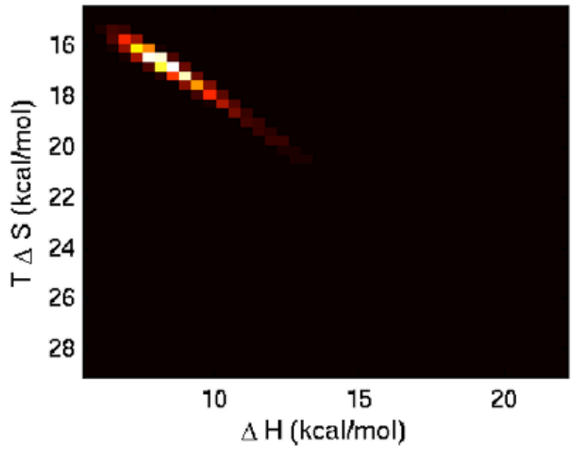


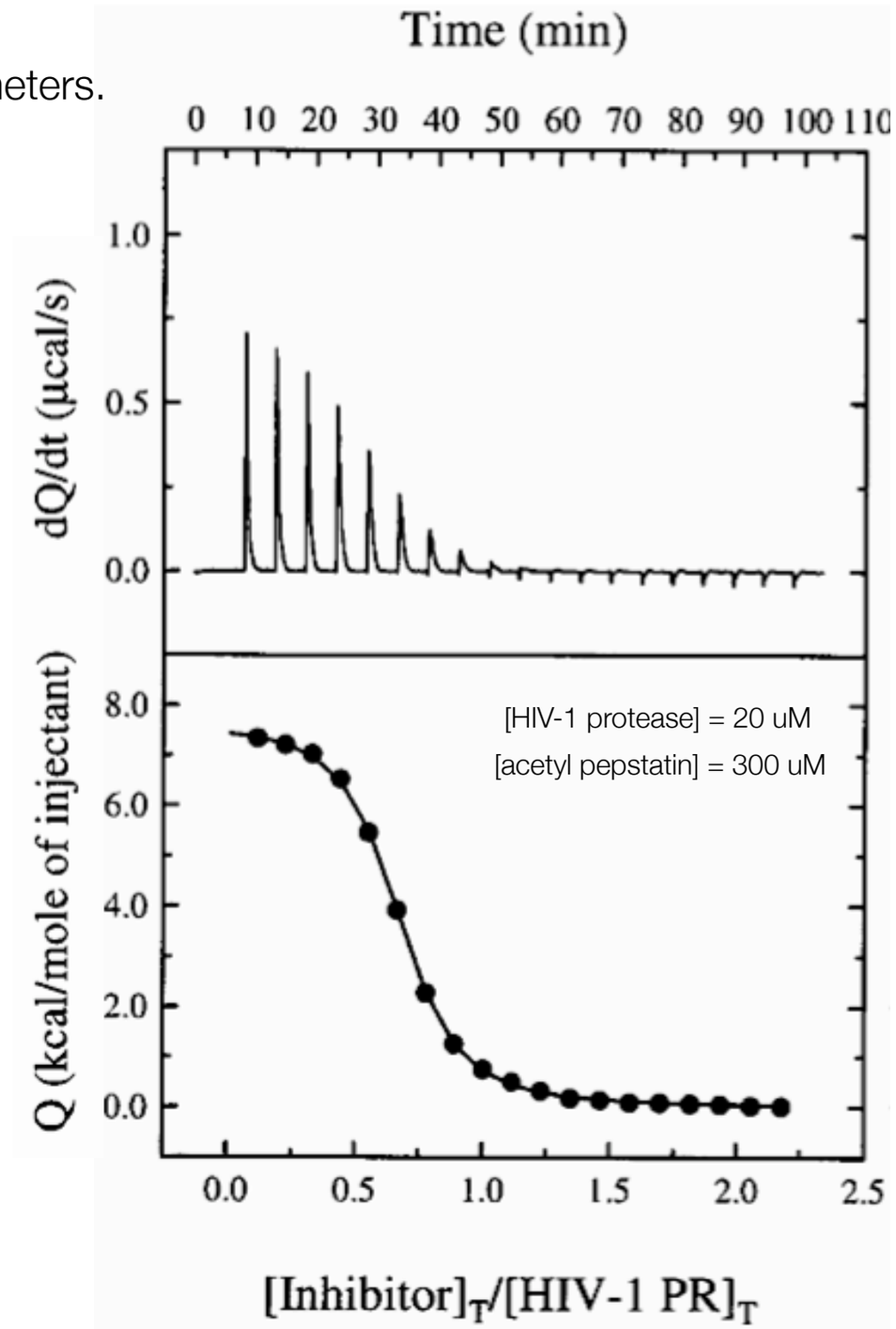
Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

Illustration of Bayesian method

Or we can obtain **joint distribution functions** for any subset of parameters.



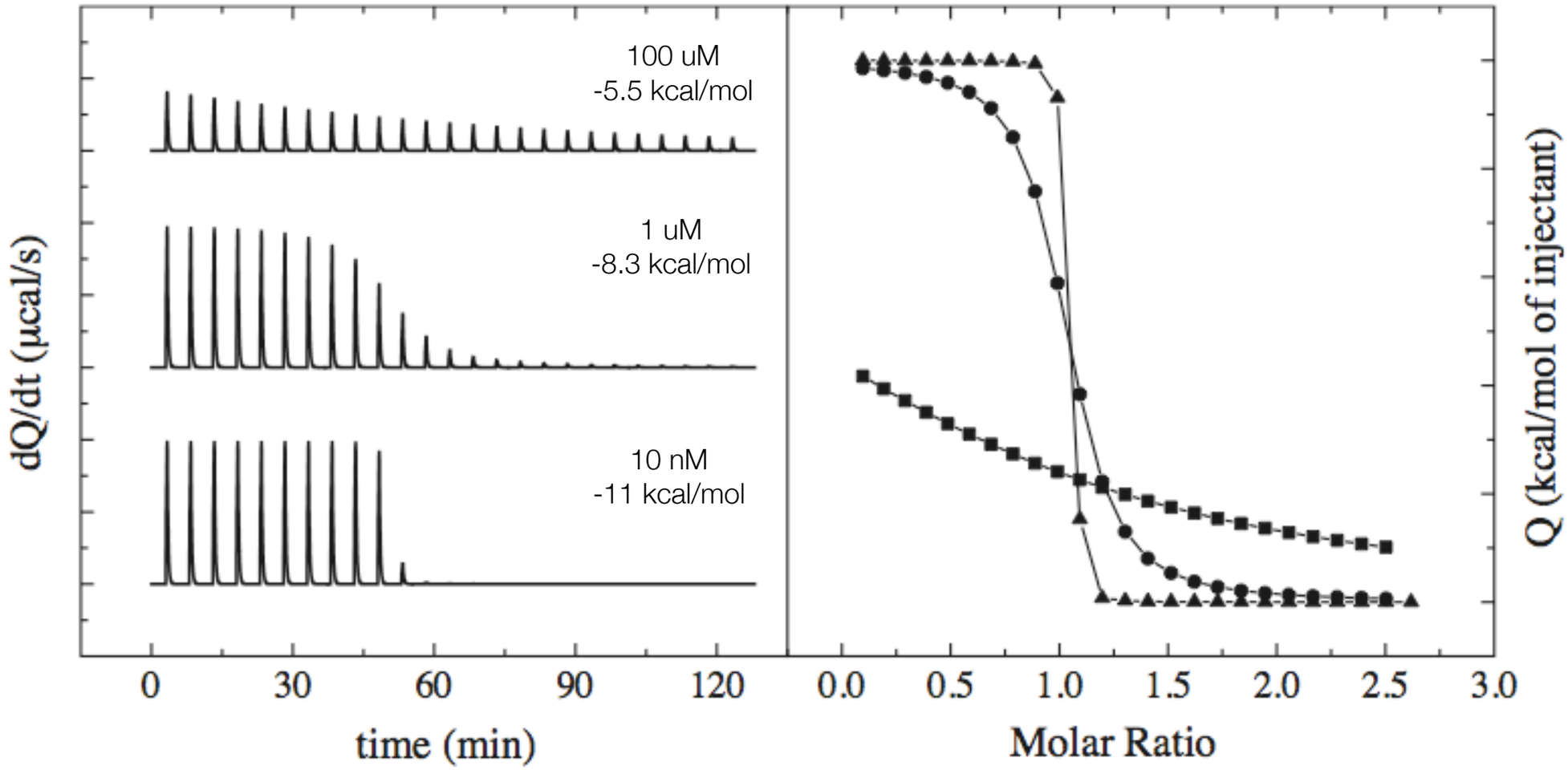




Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

High-affinity ligands can be troublesome

Ligands with very high binding affinity cause increase in thermogram curvature.

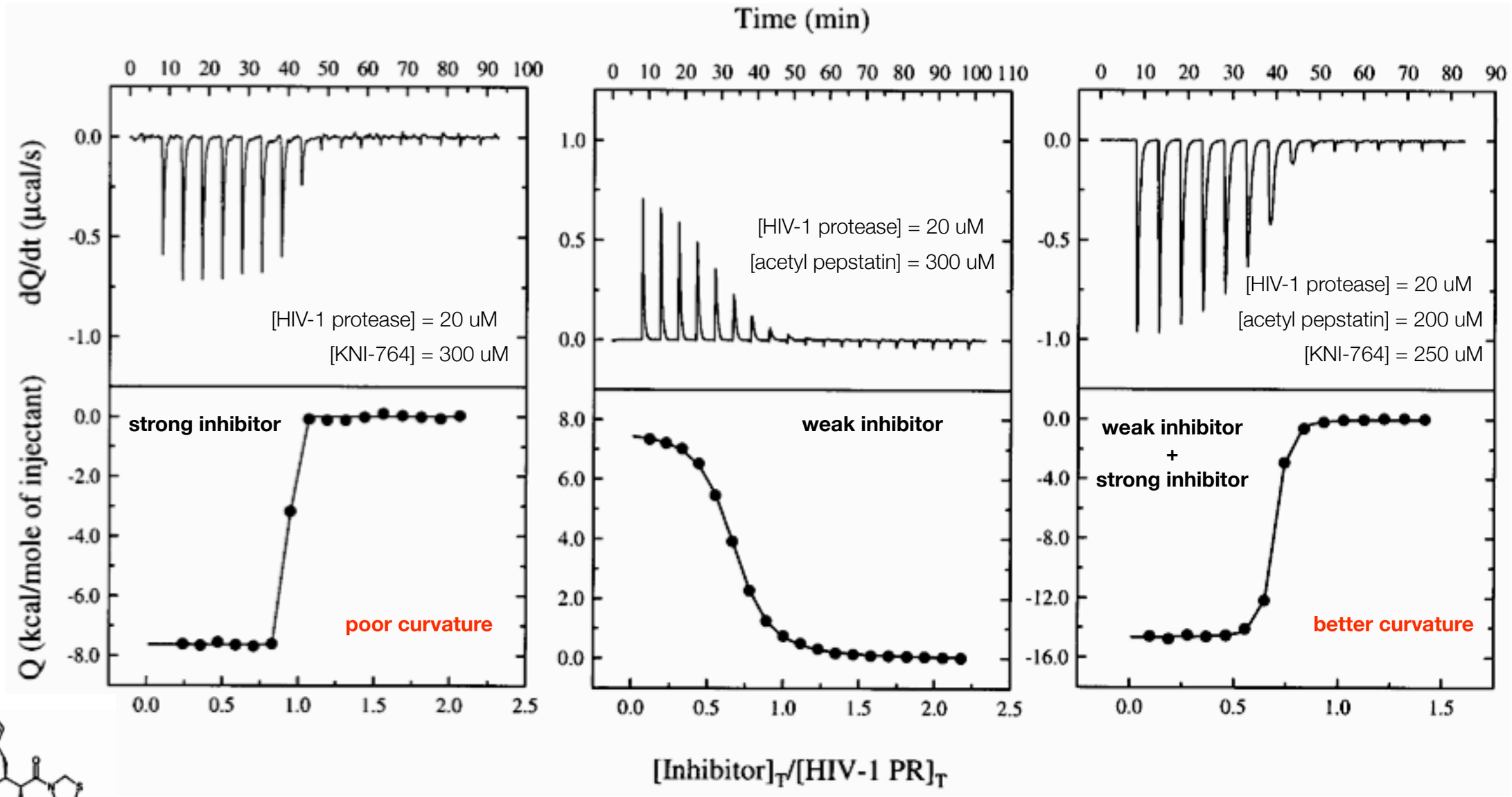


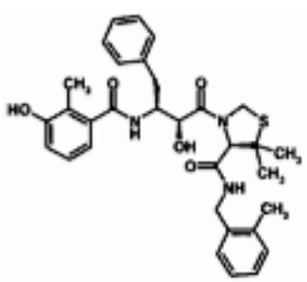
Lack of information near inflection point means that, even though enthalpy (ΔH) can still be accurately extracted, binding affinity (ΔG) becomes extremely uncertain.

Velazquez-Campoy A and Friere E. Biophys. Chem. 115:115, 2005.

Competition with a weaker ligand helps resolution

Competing off a weaker ligand of known affinity can give much more precise binding affinity (ΔG).





This works especially well if the weaker ligand has opposite enthalpy (ΔH) of binding.

Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

Binding model for competitive binding

Suppose we have M ligands that all bind competitively

$$P + L^{(1)} \overset{K^{(1)}}{\rightleftharpoons} PL^{(1)}$$
$$P + L^{(2)} \overset{K^{(2)}}{\rightleftharpoons} PL^{(2)}$$

...

$$P + L^{(M)} \overset{K^{(M)}}{\rightleftharpoons} PL^{(M)}$$

We can solve the system of nonlinear equations numerically for concentrations $[P]_n, [L^{(m)}]_n, [PL^{(m)}]_n$

total ligand $$L_n^{(m)} = n\Delta V[L^{(m)}]_s = V_n([L^{(m)}]_n + [PL^{(m)}]_n) \qquad m = 1, 2, \ldots, M$$

total protein $$P = V_0[P]_0 = V_n([P]_n + \sum_{m=1}^{M} [PL^{(m)}]_n)$$

binding model $$K_d^{(m)} = \exp[\beta\Delta G_b^{(m)}] = \frac{[P]_n[L^{(m)}]_n}{[PL^{(m)}]_n} \qquad m = 1, 2, \ldots, M$$

What are the advantages of the Bayesian method?

Simple to understand!

Easy to incorporate new binding models.

Don’t necessarily need a baseline “blank” experiment if parameters can be estimated accurately enough from a single run.

Can accurately represent the true posterior joint distribution of all thermodynamic parameters, regardless of the data.

Confidence intervals and marginal distributions? No problem!

Information content of given datasets, expected information content of new experiments

Make joint inferences from data from multiple experiments

Experimental design:

> “Will experiment X give me enough information to make it worthwhile?”
> “What is the best experimental design to reduce the uncertainty in Z?”
> “Do I have to run a baseline for sample X?”

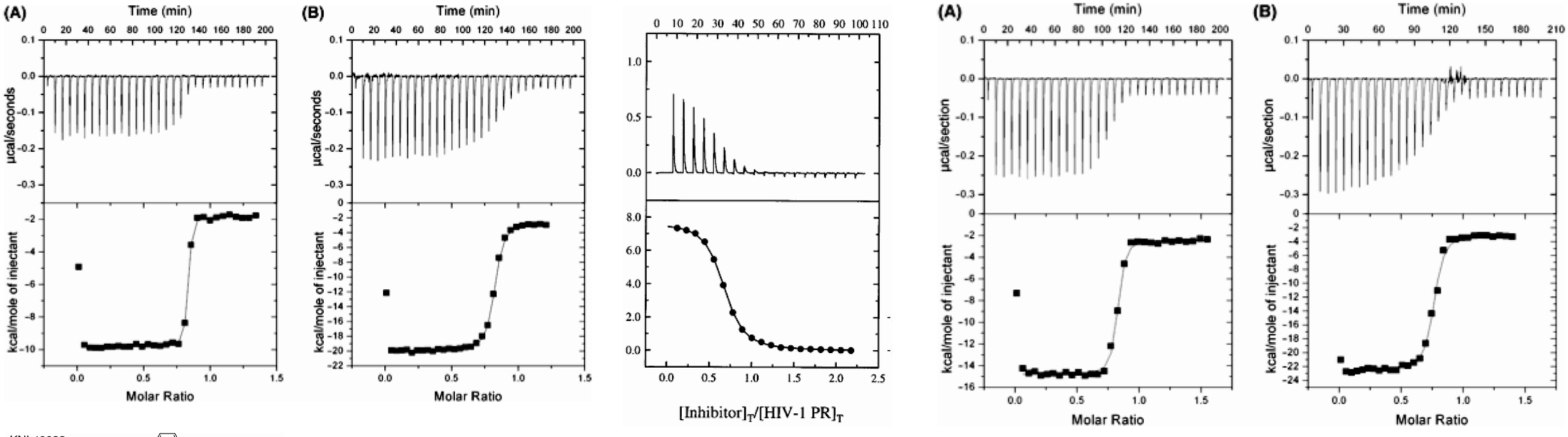
Making joint inferences from multiple experiments

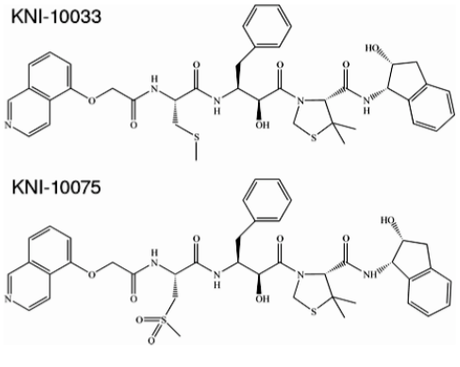
$$P + L_0 \rightleftharpoons PL_0$$

$$P + L_1 \rightleftharpoons PL_1 \qquad P + L_2 \rightleftharpoons PL_2$$

$$PL_0 + L_1 \rightleftharpoons L_0 + P + L_1 \rightleftharpoons PL_1 \qquad PL_0 + L_2 \rightleftharpoons L_0 + P + L_2 \rightleftharpoons PL_2$$

KNI-10033 **acetyl pepstatin** **KNI-10075**





We can use all data simultaneously to determine $\Delta\Delta G = \Delta G_2 - \Delta G_1$

Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

Lafont V, Armstrong AA, Ohtaka H, Kiso Y, Azmel LM, and Freire E. Chem Biol. Drug Des. 69:413, 2007.

Neglected sources of error: The known unknowns

pipetting errors in preparation of samples

errors in the known concentration or activity of samples

variability in the syringe injection volume

separation of heat of dilution into mechanical (stirrer/injector) heat + dilution

need better model for heat of dilution that depends on injection volume

Binding of charged ligands not so straightforward?

Consider the binding of a charged ligand to a charged protein

$$P^+B^- + A^+L^- \overset{K}{\rightleftharpoons} P^+L^- + A^+B^-$$

- Counterions A and B are present because macroscopic systems must be electrically neutral.
- Counterions are dissolved in solution.

We fit this with a model like

$$K' = \frac{[P^+L^-]}{[P^+B^-][A^+L^-]}$$

But a more correct model would be

$$K = \frac{[P^+L^-][A^+B^-]}{[P^+B^-][A^+L^-]} = K' \cdot [A^+B^-]$$

So measured free energies are salt concentration dependent from a purely Le Chatalier's principle effect!

$$\Delta G = \Delta G' + kT \ln \frac{[A^+B^-]}{1M}$$