Quantifying cooperative multisite binding through Bayesian inference

Aidan B Estelle[1], August George[2], Elisar J Barbar*[1], Daniel M Zuckerman*[2]

[1]Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331, United States

[2]Department of Biomedical Engineering, School of Medicine, Oregon Health and Science University, Portland, Oregon 97239, United states

*corresponding authors:

DMZ: Department of Biomedical Engineering, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland OR 97239. E-mail: Zuckermd@ohsu.edu phone: (503) 418-4159

EJB: Department of Biochemistry and Biophysics, Oregon State University, 2011 ALS, Corvallis, OR 97331. E-mail: barbare@oregonstate.edu Telephone: (541) 737-4143

Abstract

Multistep protein-protein interactions underlie most biological processes, but their characterization through methods such as isothermal titration calorimetry (ITC) is largely confined to simple models that provide little information on the intermediate, individual steps. We examine the hub protein LC8, which binds to disordered regions of 100+ client proteins in a wide range of stoichiometries. Despite evidence that LC8 binds clients cooperatively, prior ITC thermodynamic analyses have relied on models that do not accommodate allostery, and furthermore do not account for critical uncertainties in analyte concentrations. To characterize allostery in a more rigorous fashion, we build on existing Bayesian approaches to ITC to quantify thermodynamic parameters for multi-step binding interactions impacted by significant uncertainty in protein concentration. Notably, we account for a previously unrecognized intrinsic ambiguity in concentrations in standard binding models and clarify how this ambiguity impacts the extent to which binding parameters can be determined in cases of highly uncertain analyte concentrations. Our approach is applicable to a host of multi-step binding interactions, and we use it to investigate two systems. First, we deeply examine 2:2 LC8 binding and find it to be significantly positively cooperative with high confidence for multiple clients. Building on observations in the LC8 system, we develop a system-agnostic 'phase diagram' calculated from synthetic data demonstrating that certain binding parameters intrinsically inflate parameter uncertainty in ITC analysis, independent of experimental uncertainties. Second, we study 2:2 binding between the dynein intermediate chain and binding protein NudE, where in contrast, we find little evidence of allostery.

Introduction

Intracellular processes frequently depend on complex, multistep interactions between proteins or between protein and small-molecule ligands (1– 3). The hub protein LC8 provides an extreme example of binding complexity, accommodating over 100 client proteins via two symmetrical binding grooves (4, 5) – often binding in multivalent fashion with a range of stoichiometries (6–10). LC8 is found throughout the eukaryotic cell, and involved in a host of cell functions, with client proteins including transcription factors (7, 9), tumor suppressors and oncogenes (11, 12), viral proteins (13–15), and cytoskeletal proteins (6, 16).

Structurally, LC8 forms a small 20 kDa homodimer (Fig. 1a), with two identical binding grooves formed at the dimer interface (4, 5). These binding sites induce a beta-strand structure in a well-characterized linear motif anchored by a TQT amino acid sequence within disordered regions of client proteins (6, 9). Despite extensive studies (9, 16, 17), the mechanisms and

thermodynamics of LC8 binding are still not fully understood, due to the difficulty of deconvoluting a multiplicity of microscopic states in its complex binding processes.

While usually fit to a simple model, LC8-client binding is likely impacted by allostery. The first evidence indicating allosteric behavior arose from nuclear magnetic resonance (NMR) titrations of peptides with LC8. A partially bound intermediate was detected half-way through titrations, with an estimated 2.5 to 6-fold higher affinity for the second binding step relative to the first (18). Further evidence of allostery emerged from isothermal titration calorimetry (ITC) studies. Although ITC data are commonly fit to a simple *n*-independent sites binding model (19, 20), this model is inadequate for a number of LC8-client systems that exhibit non-sigmoidal behavior, dipping slightly in heat per injection during early titration points instead of forming a flat plateau (Fig. 1b)(9). This non-canonical behavior raises the possibility that these isotherms may fit well to a two-step model of binding, more representative of the expectation of dimeric LC8-client binding (21).

The use of ITC to interrogate complex systems and multi-step binding is challenging, as ITC data is of relatively low information, and individual isotherms often fit well to varied model parameters (21, 22). Despite this, well designed experiments can utilize ITC to measure cooperativity or allostery (23, 24), entropy-enthalpy compensation (17, 25), changes in protonation state (26, 27), and competition between multiple ligands (28, 29). In general, these studies rely on fitting data globally to a model that includes several isotherms collected at varied conditions to reduce ambiguity of fit parameters (21, 22), or a ‘divide and conquer’ type approach, where subsections of a complex binding network can be isolated and examined (16, 23).

Concentration uncertainty is a critical concern in analysis of ITC data. In principle, accurate determination of protein and ligand concentration is a prerequisite for obtaining reliable thermodynamic quantities by ITC, yet these values are challenging if not impossible to obtain for many systems (20, 30–32). The most common software package for fitting ITC data, built in Origin 7.0 and distributed with calorimeters, attempts to account for this uncertainty in its simplest multi-ligand model through the stoichiometric parameter *n*, which can fit to non-integer values to correct for error in cell concentrations (19, 33). However, this implementation ignores uncertainty in concentration of the titrant in the syringe, and is only applicable to the simple binding model, as complex binding models in Origin have no comparable correction factor. The popular and highly flexible fitting software SEDPHAT significantly improves on Origin’s capabilities, allowing for both explicit or implicit (i.e. an ‘inactive fraction’ correction) uncertainty corrections (22, 34). As the authors note, however, allowing for variation in both analyte

concentrations makes binding constants indeterminable within SEDPHAT due to correlative effects among model parameters.

Bayesian analysis offers a natural framework for incorporating uncertainty in concentration measurements in ITC analysis (20, 35). In a Bayesian framework, thermodynamic parameter determination is guided by a mix of experimental data and 'prior' information, such as uncertainty ranges/models, that weights the overall 'posterior' probability of a given set of thermodynamic parameters. The posterior distribution of estimated binding parameters generated through Bayesian analysis is a complete description of the probability range of each model parameter – and correlations among parameters – based on the input data and priors. With a meaningful prior description of concentration uncertainty, there is reduced risk of underestimating uncertainty in thermodynamic binding parameters.

We build on earlier applications of Bayesian inference to ITC. Nguyen et al. (2018) studied 1:1 binding using a Bayesian statistical framework accounting for concentration uncertainty, and performed sensitivity analysis on concentration priors. For a two-site binding model, Duvvuri et al. (2018) demonstrated that a Bayesian method can accurately and precisely determine two separate affinities when applied as a global model to several isotherms, but the work assumes no uncertainty in measured concentrations (36), raising the possibility that parameter uncertainty is underestimated (20, 22). Cardoso et al. (2020) used a simplified 4-step binding model with a single common binding enthalpy for a set of isotherms to determine 3 of 4 distinct affinities between protein and ligand, with the fourth being uncertain across a range of several orders of magnitude. Although Cardoso et al. (2020) include concentrations as model parameters, they greatly narrow concentration priors using a preliminary 'calibration' assuming independent sites. We note that the *n*-independent sites model likely is not appropriate for complex systems. A sensitivity analysis regarding concentration uncertainty was not performed in either multisite study, and neither work probed the information content of single isotherms for multisite systems.

Here, we report a Bayesian analysis of allosteric effects in two-site systems with a careful accounting of concentration effects critical for reliable analysis. We show that LC8-client interactions unambiguously exhibit positive allostery, driving binding towards a fully bound state. In contrast, symmetric two-site binding between the coiled coil domain of the dynein cargo adaptor NudE and the intermediate chain (IC) of dynein(37) shows no significant evidence for allostery.

We also provide methodological advances. First, we derive simple mathematical relations that govern the influence of concentration uncertainties on different binding

parameters, providing a fundamental basis for the previously noted strong sensitivity of enthalpies – but not free energies – to concentration uncertainty (22). Second, by using synthetic models, we systematically characterize the causes of binding-parameter uncertainties in two ways: we demonstrate that substantial uncertainty can result from the binding parameters themselves, e.g., strong vs. weak binding; and we also determine the effects of different prior functional forms and uncertainty ranges in a multisite context, extending the work of Nguyen et al. (2018). Finally, we outline best practices for determining model parameters and uncertainties in a multisite Bayesian framework.

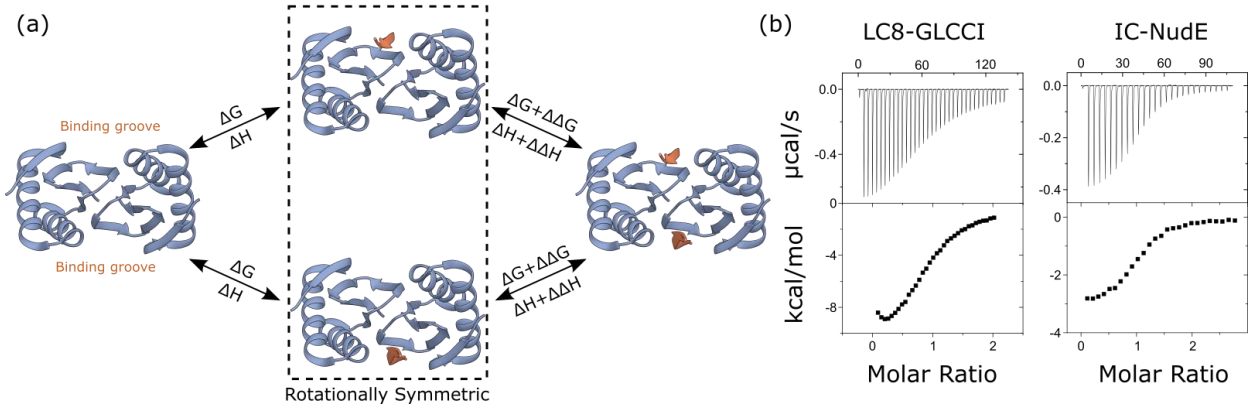

Figure 1: LC8 binds clients through a two-step mechanism. (a) Diagram of LC8-client binding, showing a structure of apo LC8 on the left, and a fully bound structure (PDB 3E2B) on the right. Intermediates are boxed to indicate they are symmetric and indistinguishable species. (b) example isotherms for binding between LC8 and client peptide taken from GLCCI (left) and binding between the intermediate chain (IC) and partner NudE (right).

Results

A mathematical "degeneracy" in thermodynamic parameters impacts analysis at any stoichiometry

We first present a simple mathematical analysis that explains previously reported correlation effects among titrant and titrand concentrations (22), and which significantly impacts the overall analysis of ITC data. Importantly, our analysis applies to monovalent or multivalent binding. Specifically, when the concentrations are uncertain, as is common in analysis of ITC data (20, 22), we show below that only the *ratio* of titrant:titrand concentrations can be estimated, rather than the individual values, and this ambiguity propagates to all thermodynamic parameters. Hence, there is a "degeneracy" in that multiple solutions (sets of concentration values and thermodynamic parameters) will equally describe even idealized ITC data lacking experimental noise (Fig. 2).

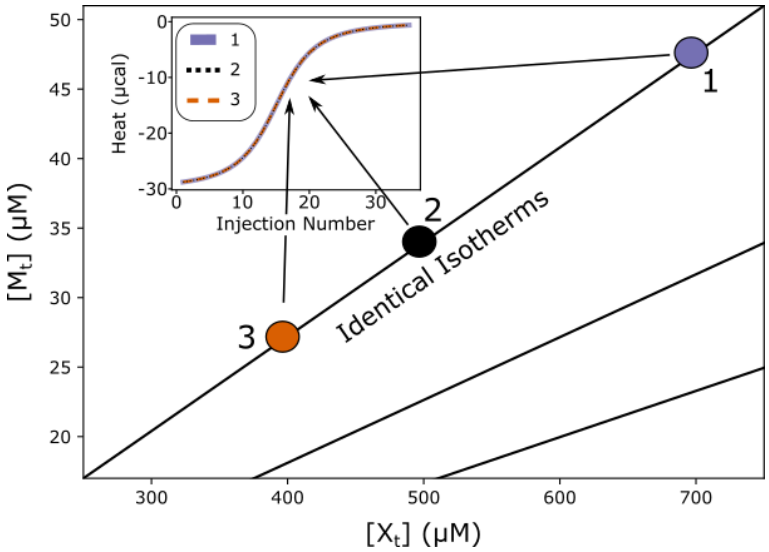

Figure 2: Exact degeneracy in binding isotherms. Based on the scaling relations of Eq (2), for any set of ligand and macromolecule concentrations (X_t, M_t), there are infinitely many alternative concentrations (e.g., filled circles) on a diagonal line in the ([X_t], [M_t]) plane which yield exactly equivalent isotherms (inset) for a fixed set of thermodynamic parameters. For any given point in parameter space, equivalent degenerate lines can be drawn in a radial manner (e.g. the two additional solid black lines), passing through the point and the origin. The plotted synthetic isotherms are for 1:1 binding, but analogous degeneracy also holds for multivalent binding - see text.

We first describe the degeneracy for standard 1:1 binding between a macromolecule M and ligand X, following the scheme

$$M + X \rightleftharpoons MX \tag{1}$$

The heat, Q, of a 1:1 binding system at any titration point can be described using the standard quadratic binding equation used in the independent sites model (19, 38):

$$\frac{Q}{V_0} = \frac{[M_t]\Delta H}{2}\left\{1 + \frac{[X_t]}{[M_t]} + \frac{K_d}{[M_t]} - \sqrt{\left(1 + \frac{[X_t]}{[M_t]} + \frac{K_d}{[M_t]}\right)^2 - \frac{4[X_t]}{[M_t]}}\right\} \tag{2}$$

where [M_t] and [X_t] are the concentrations of macromolecule and ligand (i.e., cell component and syringe component) respectively, while K_d and ΔH are the binding affinity and enthalpy.

The degeneracy is demonstrated by introducing a linear scaling of all parameters by an arbitrary number denoted α. Specifically, we apply the following transformations:

$$[M_t] \longrightarrow \alpha[M_t]$$
$$[X_t] \longrightarrow \alpha[X_t]$$
$$K_d \longrightarrow \alpha K_d$$
$$\Delta H \longrightarrow \frac{\Delta H}{\alpha} \tag{3}$$

Applying this set of transformations, we can rewrite the binding equation:

$$\frac{Q}{V_0} = \frac{\alpha[M_t]\frac{\Delta H}{\alpha}}{2}\left\{1 + \frac{\alpha[X_t]}{\alpha[M_t]} + \frac{\alpha K_d}{\alpha[M_t]} - \sqrt{\left(1 + \frac{\alpha[X_t]}{\alpha[M_t]} + \frac{\alpha K_d}{\alpha[M_t]}\right)^2 - \frac{4\alpha[X_t]}{\alpha[M_t]}}\right\} \tag{4}$$

Regardless of the value of the factor α, all introduced factors cancel leaving Q unchanged.

Nearly identical considerations apply in the two-step binding model of primary interest here. As detailed in the SI, the value of Q is unchanged when both concentrations and both K_d values are multiplied by α and both ΔH values are divided by α. The underlying model is more complex as it requires solving a system of nonlinear equations (see Methods for details), but the result is that α is propagated through the nonlinear equation solutions, and once again cancels in the calculation of Q, leaving the heat value unchanged.

For reference, the corresponding concentration degeneracy scaling relations for 2:2 binding derived in Methods are as follows:

$$\begin{gathered}
[M_t] \longrightarrow \alpha[M_t] \\
[X_t] \longrightarrow \alpha[X_t] \\
K_{d1} \longrightarrow \alpha K_{d1} \\
\Delta G_1 \longrightarrow \Delta G_1 + RT\log\alpha \\
K_{d2} \longrightarrow \alpha K_{d2} \\
\Delta G_2 \longrightarrow \Delta G_2 + RT\log\alpha \\
\Delta H_1 \longrightarrow \frac{\Delta H_1}{\alpha} \\
\Delta H_2 \longrightarrow \frac{\Delta H_2}{\alpha}
\end{gathered} \tag{5}$$

To facilitate analysis and discussion of allostery below, from this point on we parameterize our model using ΔG, ΔΔG, ΔH and ΔΔH. The ΔΔG and ΔΔH value correspond to the differences between the first and second binding steps. Thus $K_{d1} = e^{\Delta G/RT}$, $K_{d2} = e^{(\Delta G+\Delta\Delta G)/RT}$, and $\Delta H_2 - \Delta H_1 = \Delta\Delta H$. The energy-like formulation allows for easy assessment of allostery, as ΔΔG is the free energy of allostery (which will be zero in the absence of allostery and positive or negative for negative or positive cooperativity, respectively), and ΔΔH is the change in enthalpy between binding steps with analogous characterization.

The degeneracy and associated scaling relationships in Eq (3) provide important insight into assessment of thermodynamic parameters inferred from ITC data. We see directly that binding enthalpy changes proportionately to concentrations of titrant and titrand. That is, a

given percent error in an assumed concentration of either ligand (characterized by alpha) translates to the same scale of error in ΔH. On the other hand, the binding free energy ΔG, is less sensitive to concentration errors, due to scaling with ln(α), rather than directly multiplied by α.

The scaling relationships of Eq. (3) also presage a significant issue in Bayesian inference, namely, sensitivity to the choice of priors. Within the set of degenerate solutions (diagonal lines of concentration pairs in Fig. 2), the Bayesian 'likelihood' probability – which describes how well a parameter set fits the data in the absence of prior information – will be constant, as solutions are mathematically identical. Thus, within any degenerate set, the assumed prior distributions for concentrations, will determine the overall posterior distributions (see Methods). Because the posterior distributions ultimately determine the uncertainty ranges, this is a key point.

Below, we continue to examine the ramifications of the concentration degeneracy, demonstrating concretely that enthalpy is more impacted by uncertainty in concentrations than free energy. We also examine the influence of priors on parameter distributions, and discuss parameter distributions determined from isotherms in cases of high concentration uncertainty.

Validation of Bayesian inference pipeline with synthetic data

To test our Bayesian pipeline (Methods), we generated 'synthetic' simulated isotherms using hand-chosen sets of thermodynamic parameters ΔG, ΔΔG, ΔH, ΔΔH (see Fig. 1) inserted in Methods Eq (15) with added Gaussian noise. Following an exploration using synthetic data of how allostery impacts binding isotherms (e.g. Fig. 3a), we selected synthetic model parameters to mimic the isotherm shape seen in LC8-peptide binding examples. Specifically, slight positive allostery (ΔΔG = -1, ΔΔH = -1.5 kcal/mol) was best-suited to imitating real LC8-peptide isotherms, along with ΔG = -7 and ΔH = -10 kcal/mol. Synthetic noise is taken from a Gaussian distribution with a zero mean and standard deviation σ = 0.2 μcal. As shown in Fig. 3, we used our pipeline to sample posterior distributions for these isotherms. For concentrations, we chose uniform prior distributions of ±10% of the true value (which simply limits sampled concentration values to these ranges). The choice of 10% approximates what we view to be an attainable level of uncertainty for experimental protein concentrations.

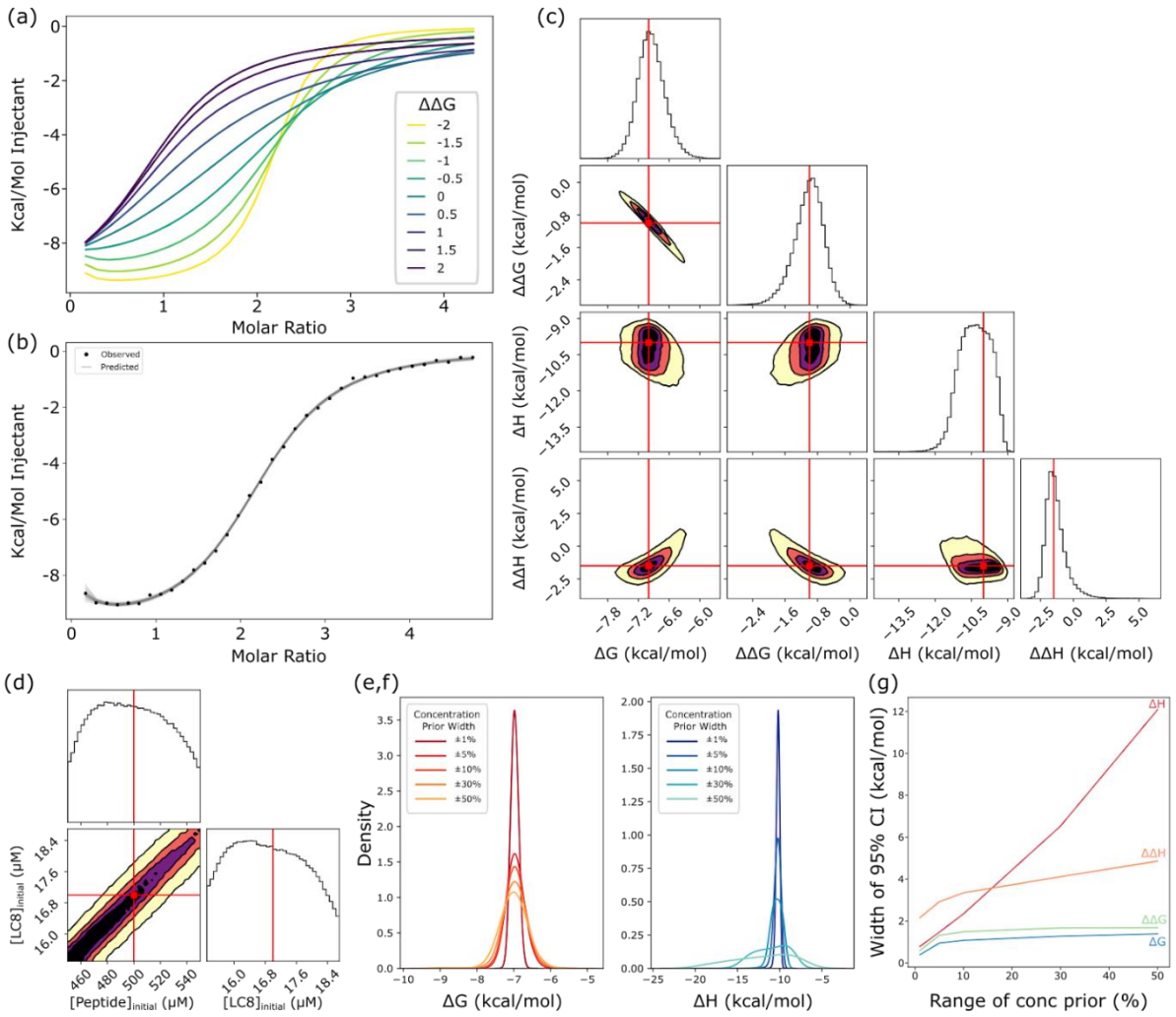

Figure 3: Analysis of two-step model using synthetic isotherms. (a) A set of synthetic isotherms for two-step binding with varied ΔΔG parameters demonstrating how allostery changes isotherm shape. Thermodynamic parameters are ΔG = -7, ΔH=-10, and ΔΔH=0. Concentrations are set at 17 and 500 μM for cell and syringe respectively, and injection volumes are 6 μL. (b) A synthetic isotherm with added gaussian noise (points) with 50 fitted isotherms (lines) generated through the Bayesian pipeline, i.e., sampled from the posterior. (c) One and two-dimensional marginal distributions for thermodynamic parameters, with contours in the two-dimensional plots set at 95 (yellow), 75 (orange), 50(purple) and 25%(black) confidence. Red lines and dots indicate true values for the synthetic isotherm. Marginal distributions, along with MCMC chains for all eight model parameters, including nuisance parameters can be found in Supplemental Figure S1. (d) Marginal distributions for concentration parameters, exhibiting characteristic diagonal shape (Fig. 2) with contours as in (c). (e,f) One-dimensional distributions for ΔG (e) and ΔH (f) plotted for models with prior ranges for concentrations of 1, 5, 10, 30 and 50% of the stated concentration. (g) Width of the 95% Bayesian credibility region, akin to a confidence interval, for thermodynamic parameters as a function of the width of the concentration prior used in modeling, plotted from models with prior ranges for concentrations of +/-1, 5, 10, 30 and 50% of the stated concentration.

Under these representative conditions, we validated our pipeline by calculating posterior distributions that recapitulated known model parameters, as well as closely capturing the isotherm shape (Fig. 3b,c). The finite widths of the distributions are due to synthetic experimental noise. The posterior distribution for ΔG covers a range of ~1 kcal/mol distributed

around the true value of -7 kcal/mol. Examination of the distribution lets us define a 'credibility region,' that contains 95% of the distribution probability (i.e., from the 2.5 to 97.5%ile of the distribution), which is directly analogous to a frequentist confidence interval. For ΔG, the 95% credibility region is -7.5 to -6.4 kcal/mol. Similarly, the 95% credibility region for ΔΔG covers a range of ~1.5 kcal/mol, evenly distributed around -1 kcal/mol. ΔH and ΔΔH both have slightly wider credibility regions, with widths of 2.3 and 3.3 kcal/mol respectively, but both are distributed around the true values of -10 and -1.5 kcal/mol respectively.

One benefit of Bayesian inference is the ability to examine multi-dimensional likelihood distributions to obtain correlations between model parameters. For example, in our two-dimensional distributions for the thermodynamic parameters, the ΔG and ΔΔG values are strongly negatively correlated (Fig. 3c), indicating a compensatory effect in the model, where increases in ΔG can be compensated by decreases in ΔΔG to arrive at similar solutions. Resultantly, the distribution for both ΔG and ΔΔG are broader than the 'total' free energy (i.e., $2\Delta G + \Delta\Delta G$), evidence that we can know the overall energy of binding more precisely than we can know the energy of each step (Supp. Fig. S2). Additionally, the mathematical degeneracy for concentrations described above can clearly be seen in these two-dimensional correlations: the two-dimensional marginal distribution for each concentration is a noise-broadened straight line covering the entire prior range (Fig. 3d). The scaling relationship of the model parameters outlined previously means that each point along this diagonal corresponds to a degenerate solution, i.e., each point has equivalent likelihood based on the data.

Impact of concentration degeneracy on two-site thermodynamic parameters assessed via synthetic data.

Bayesian inference enables determination of distributions for thermodynamic parameters even in cases of a concentration degeneracy. The net result, as will be seen, is a broadening of (posterior) parameter distributions based on multiple equally likely solutions, constrained by the priors used. Despite intrinsic limitations surrounding concentrations, the *ratio* of concentrations can be quantified with relatively high precision even when individual concentrations are highly uncertain.

To quantify the impacts of the concentration degeneracy within a Bayesian inference pipeline, we examined a series of uniform prior distributions for concentrations, ranging from ±1% to ±50% for both concentrations. These priors were applied to a synthetic isotherm mimicking experimental parameters, as described in the pipeline validation above. The choice of concentration prior – which embodies assumed experimental uncertainty – significantly impacts

the predicted uncertainty of thermodynamic parameters. The distributions for ΔG and ΔH, not surprisingly, both widen as the prior range is increased (Fig. 3e,f). As anticipated by the degeneracy scaling relations of Eq (3), the width of the distributions for ΔH and ΔΔH increases roughly linearly with the concentration prior range, while the distributions for ΔG and ΔΔG increase initially at low concentration ranges then level off. This can be explained by the logarithmic relationship between the K_d (which is what scales with the degeneracy) and free energy. Functionally, high uncertainty in concentrations therefore only slightly increases uncertainty in binding free energy, while having a more significant impact on binding enthalpy.

The concentration degeneracy of the model limits the degree to which erroneously determined individual concentrations can be corrected. As discussed above, the fact that the Bayesian likelihood is equal at any point along the degeneracy lines (Fig. 2) means that the data have little impact on the posterior distributions for *individual* concentrations, which instead takes the shape of the prior used. This can be seen in the model validation example (Fig. 3d), where the posterior distribution is approximately uniform, echoing the uniform prior.

The ratio of concentrations ('macromolecule' to 'ligand'), on the other hand, is a more meaningful parameter, and the quality of the ratio can improve a single uncertain concentration. For example, when we sample the posterior for the same isotherm, but use a normal (i.e. Gaussian) distribution for one concentration prior and a uniform distribution for the other, both posteriors take the shape of a normal distribution (Supp. Fig. S3). This is a direct result of the degeneracy identified above. Supp. Table S1 shows concentration ratio credibility regions for the experimental systems.

Because of the nearly determinative relationship between the prior and posterior concentration distributions, we elected to use uniform priors for concentrations throughout this work to avoid undue influence on our results from model priors. We believe varying the widths of uniform priors is the best way to probe concentration uncertainty effects.

For completeness, we also assayed 1:1 binding with synthetic data. Overall, the impact of the concentration degeneracy on model parameters is similar (Supp. Fig. S4): binding enthalpy posterior distributions are significantly wider than free energy distributions. In response to changes in concentration prior ranges, the posterior for ΔG is more impacted than in the two-step model, but the distribution remains much narrower than that of the enthalpy in every case.

Application to 2:2 LC8:IDP Systems

We applied the Bayesian analysis pipeline to a set of 7 experimental isotherms of binding between LC8 and client peptides, all of which bind in a 2:2 ratio. Note that the two LC8's form a strong homodimer (Kd ~ 60 nM (39)) and this initial homodimer formation is excluded from our analysis. The systems were selected from a prior study (9) for tight binding and their deviation from the standard sigmoidal isotherm shape. As noted above, the user-supplied uncertainties for concentrations significantly impact uncertainty in other parameters. Following analysis with priors of ±10% and ±20% of the measured LC8 concentration as determined by absorbance at 280 nm, we have elected to focus on results at ±10%, as moving to ±20% does not significantly impact the posterior distributions (Table 1, Supp. Table S2), and we believe ±10% to be an achievable uncertainty in protein concentration for most cases. The high degree of purity (>95%) and high number of chromophores (1 Trp, 5 Tyr) allow relatively high confidence in the LC8 concentration measurement. Comparatively, because of the difficulty in accurately measuring concentration for peptides with few or no chromophores (31, 40), we used a prior of increased width for the peptide concentration, up to a limit of ±50% of the initially measured value estimated by absorbance at 280 nm. As discussed above, the posterior distributions processed through the Bayesian pipeline are limited by the most restrictive prior used, so this approach ensures that posterior distributions are limited to the range around the measured concentration of LC8, allowing us to effectively infer the ratio of LC8:peptide (Supp. Table S1).

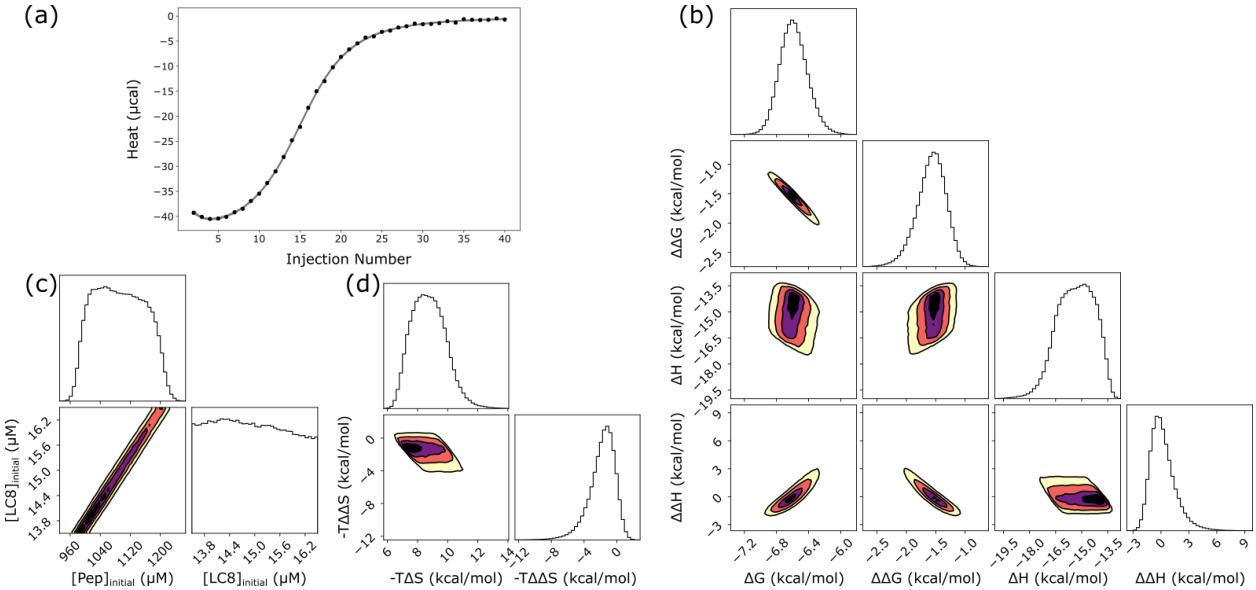

Figure 4: LC8 binding to a peptide from the protein SPAG5. (a) Experimental titration isotherm of SPAG5 into LC8 (points) with 50 example traces (lines) drawn from the posterior distribution of thermodynamic parameters and concentrations. (b) One and two-dimensional marginal distributions for thermodynamic parameters, with contours in the two-dimensional plots set at 95 (yellow), 75 (orange), 50(purple) and 25%(black) confidence. (c) Marginal distributions for concentrations of LC8 and peptide, showing a line of degenerate solutions, which may be compared to Fig. 2. (d) Marginal distributions for entropy (-TΔS) and entropy of allostery (-TΔΔS).

Bayesian analysis of the seven systems reveals significant heterogeneity in the precision with which binding parameters can be determined (Table 1). As will be described in detail below, this is only partially reflective of apparent data quality (e.g., noise level). Instead, certain binding parameters, particularly binding enthalpies, are intrinsically more difficult to characterize. Variations in precision do not stem from inadequate sampling in the Bayesian pipeline: triplicate runs are performed to confirm sampling quality (see Methods) (example in Supp Fig. S5).

In particularly tractable cases, such as for SPAG5 binding in Figure 4, the analysis provides marginal distributions of similar precision to those seen with synthetic data. For binding between a peptide from the protein SPAG5 and LC8, Bayesian analysis yields a 95% credibility region of -6.9 to -6.2 for ΔG (Table 1), equivalent to a range for K_{d1} of 8.7 μM to 27 μM. The 95% credibility region for ΔΔG, the allosteric difference between the first and second binding event, is -2.1 to -1.1, roughly equivalent to a 6 to 30-fold increase in affinity for the second binding step relative to the first. The change in binding enthalpy between first and second events, ΔΔH, is distributed around zero (Fig. 4c), with significant uncertainty, meaning we are unable to discern conclusively if there is any allosteric change in enthalpy between binding steps. From ΔG and ΔH values for both binding steps, we can additionally calculate -TΔS and -TΔΔS, for the entropy of binding and the change in entropy across binding steps respectively. Although the marginal distributions for these terms are quite broad (Fig. 4d), the -TΔΔS mostly sits at negative values, indicating that binding allostery is likely to be entropically driven. See Table 1 for the full set of credibility regions.

	ΔG		$\Delta\Delta G$		ΔH		$\Delta\Delta H$		-T ΔS		-T $\Delta\Delta S$	
Peptide	min	max	min	max	min	max	min	max	min	max	min	max
SPAG5	-6.9	-6.2	-2.1	-1.1	-18	-14	-1.8	3.6	6.9	11	-5.7	0.6
BSN (I)	-5.6	-4.9	-2.1	-0.9	-37	-14	0.1	49	8.8	32	-51	-2.6
BSN (II)	-7.1	-6.5	-1.2	-0.4	-5.8	-4.8	-9.1	-7.0	-1.9	-1.0	6.1	8.3
SLC9A2	-6.8	-5.5	-2.6	-0.5	-24	- 10	-5.0	23	3.7	18.3	-26	4.4
Ebola VP35	-7.4	-6.8	-1.7	-0.9	-14	-11	-0.7	1.5	4.0	6.7	-3.0	-0.2
GLCCI	-6.1	-5.1	-2.6	-1.0	-27	-9.7	-0.5	36	3.7	22	-39	-0.5
BIM	-9.5	-7.1	-2.0	0.8	-12	-10	2.9	-0.5	1.4	4.4.	-0.7	2.2

Table 1: Ranges for thermodynamic parameters for LC8-client binding. Values delineate 95% Bayesian credibility regions from sampled posterior distributions, which are akin to 95% confidence intervals.

Some general conclusions about allostery are apparent from the full set of data (Table 1). In all cases except one (binding to BIM), the distribution for ΔΔG is negative, indicating that all isotherms exhibit some positive cooperativity. Even for BIM, which has the widest ΔΔG distribution, the range predominantly covers negative values. All isotherms exhibit precisely determined free energies: 95% credibility regions cover a range of 2 kcal/mol or less for all

cases except BIM. A common feature among some isotherms, seen clearly in the 'fair' and 'poor' examples in Figure 5, is an apparent loss of precision in our ability to determine model enthalpies, as both show wide distributions for ΔH and ΔΔH. For these isotherms (e.g., SLC9A2, GLCCI, and BIM), the two-dimensional marginal distribution for ΔH and ΔΔH shows a clear correlative effect (Supp. Fig. S6), and the one-dimensional distribution for the 'total' enthalpy (i.e. $2\Delta H + \Delta\Delta H$) is significantly narrower than the individual parameter distributions (Supp. Fig. S2). In sum, the wide enthalpy distributions represent an inability to determine 'microscopic' enthalpies for individual binding events, even when the overall enthalpy is determinable.

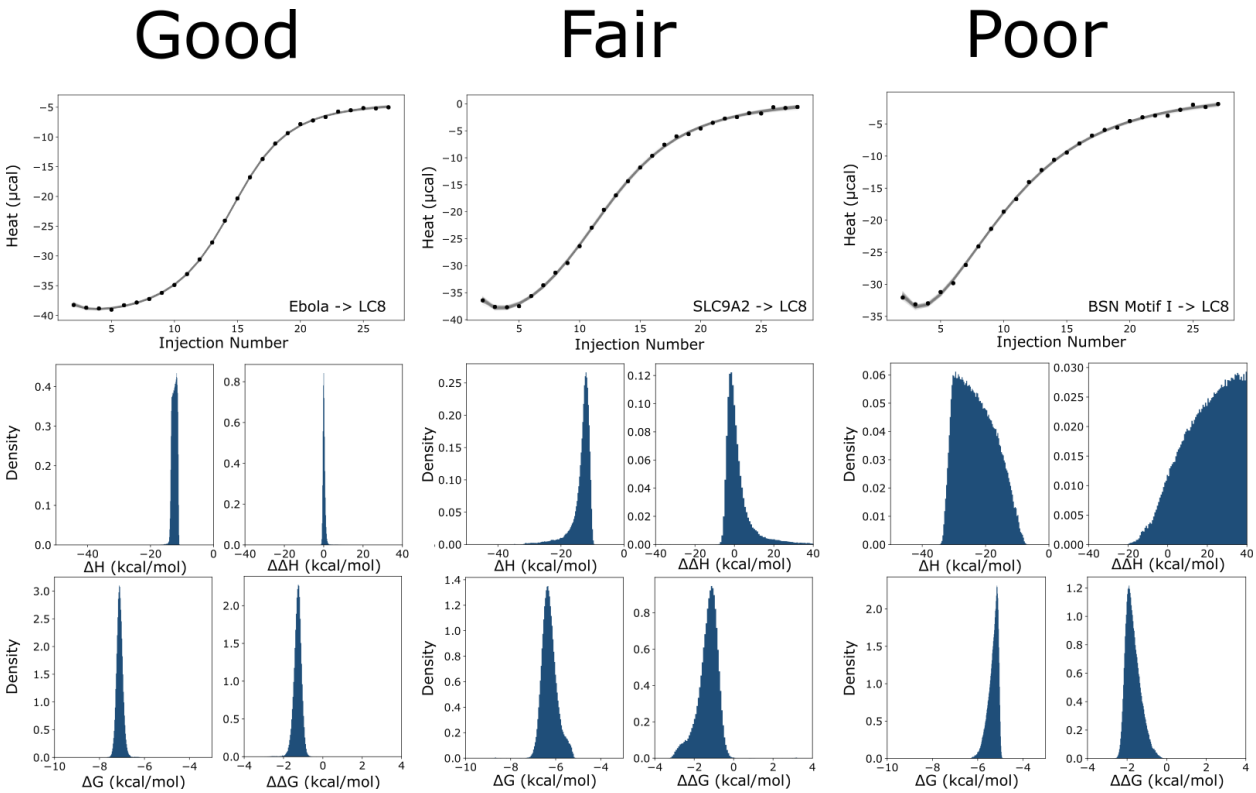

Figure 5: Example distributions for thermodynamic parameters from 3 LC8-peptide isotherms. Binding between LC8 and peptides from Ebola VP35 (left), SLC9A2 (middle) and motif 1 from BSN (right). Isotherms are shown at the top, and distributions for thermodynamic parameters are shown below. Horizontal axes represent the full width of the prior for each parameter to allow for direct comparison between each isotherm.

Parameter inference from multiple isotherms

The use of additional experimental information is expected to increase the precision of parameter determination, and Bayesian inference is readily adapted to employ multiple isotherms, whether at matching or different experimental conditions (36). For some systems where two isotherms were available, we therefore used a 'global' model that included both isotherms. Despite the higher dimensionality resulting from additional nuisance parameters (see Methods), we found it relatively easy to sample the parameter space for a two-isotherm model

(Supp. Fig. S7). In some cases, such as for GLCCI, the addition of a second isotherm usefully narrowed posterior distributions, while in others (e.g. BSN motif I) it proved less impactful, largely just taking the same shape as the distribution for individual isotherms. We note that the isotherms examined were designed as technical replicates, not as optimized isotherms at different conditions for a global model. Results on multiple isotherms with varied experimental setups, e.g., different concentrations, would likely be more consistently valuable. Nevertheless, the global models demonstrate our ability to apply the pipeline to multiple isotherms simultaneously, a key step toward improved precision going forward.

IC-NudE binding

To confirm the utility of the Bayesian pipeline for a range of systems, we tested it on binding of the intermediate chain of dynein (IC) to the non-dynein protein NudE. Binding between IC and NudE can be described by the same model as binding between LC8 and clients – NudE forms a dimeric coiled-coil structure which then accommodates two strands of monomeric disordered IC for a 2:2 complex stoichiometry (Fig. 6a)(37). Prior characterization of NudE-IC binding used a simple independent sites binding model without taking in consideration any binding allostery, and thus provides a good system for re-analysis as well as for comparison to LC8-client binding (37, 41).

For IC-NudE binding, a two-step model recapitulates the parameters determined in fits to independent sites modeling with little evidence of allostery. For high confidence in model parameters, we applied a global model, identical to the one used in LC8-client binding, to two titrations of IC into NudE. Bayesian sampling returns distributions that are narrowly dispersed for all thermodynamic parameters, both for individual-isotherm models (Supp. Fig. S8), and for the global, 2-isotherm model (Fig. 6). As expected from the lack of prior evidence of allostery, neither ΔΔG nor ΔΔH are significantly shifted from a distribution around zero, suggesting little, if any, allostery in binding. Simple models on these data indicate that binding has an enthalpy of -3.1 kcal/mol, and an affinity of 2.3 uM (i.e. a ΔG of of -7.6 kcal/mol) implying a TΔS value of 4.5 kcal/mol, meaning binding is entropically favored (42). Our two-step model predicts a ΔH distribution centered near -3 kcal/mol, and a ΔG distribution centered near -7.5 kcal/mol, aligning well with the published values. This binding interaction works well as a counterexample to LC8-client binding: distributions for allosteric terms are centered around zero, and determined distributions match closely to reported values modeled from an independent sites model.

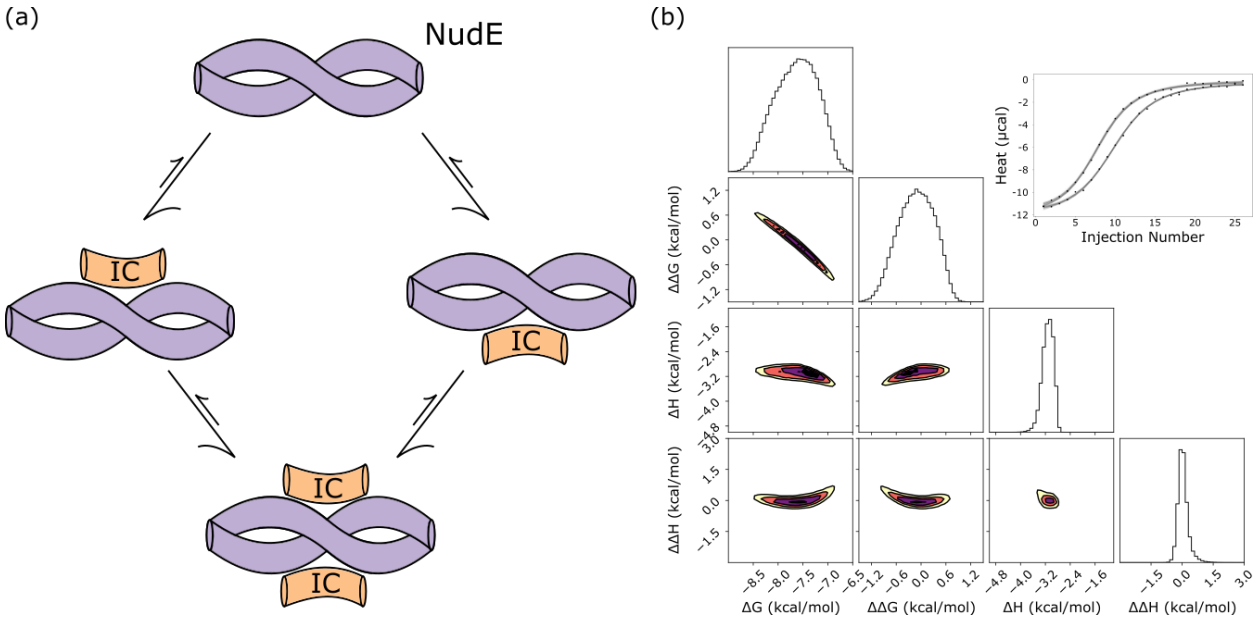

Figure 6: binding between the intermediate chain (IC) and NudE. (a) A model of IC-NudE binding, which forms a 2:2 complex, similar to what is seen in LC8. A cartoon diagram of NudE is shown in purple and IC in orange. (b) Sampled distributions modeled from two isotherms for binding between IC and NudE from yeast. Marginal distributions for thermodynamic parameters are shown on the left, and the top right corner contains the experimental isotherms (points) with model values (lines) drawn from the posterior.

'Phase diagram' analysis reveals weak binding affinities underlie loss of precision in binding enthalpies

We exploit synthetic isotherms to systematically survey binding parameters and determine the extent to which the physical parameters themselves intrinsically lead to lower precision in parameter inference. That is, for a fixed level of experimental noise, we quantify the widths of posterior marginal distributions and array the information in an interpretable 'phase diagram.' This effort was motivated by initial anecdotal observations that weaker binding was correlated with increased uncertainty, i.e., broader posterior marginals, in binding parameters, especially ΔH and ΔΔH. We created a series of synthetic isotherms on a grid of ΔG and ΔΔG values and determined posterior distributions for each isotherm. Two-dimensional plots of the width of these distributions (Fig. 7) as a function of ΔG and ΔΔG capture trends in our ability to determine model parameters.

Generally, we lose precision in binding enthalpy in situations of weaker binding. Interestingly, the relationship appears to differ somewhat between ΔH (Fig. 7a) and ΔΔH (Fig. 7b). For ΔH, the primary dependence appears to be on the value of ΔG, with precision decreasing when ΔΔG is 0 or negative (top left corner of 7a). Conversely, the precision for ΔΔH appears dependent on both ΔG and ΔΔG, with the worst precision found in the top right quarter of the plot, where binding is weak and allostery is positive. This is consistent with our

experimental results, wherein tighter-binding isotherms, higher magnitude ΔG value) such as Ebola VP35 and SPAG5 (Fig. 4, Fig. 5) perform better in terms of enthalpy determination than slightly weaker-binding isotherms such as SLC9A2 (Fig. 5), and much better than weakly-binding isotherms such as for BSN motif I (Fig. 5).

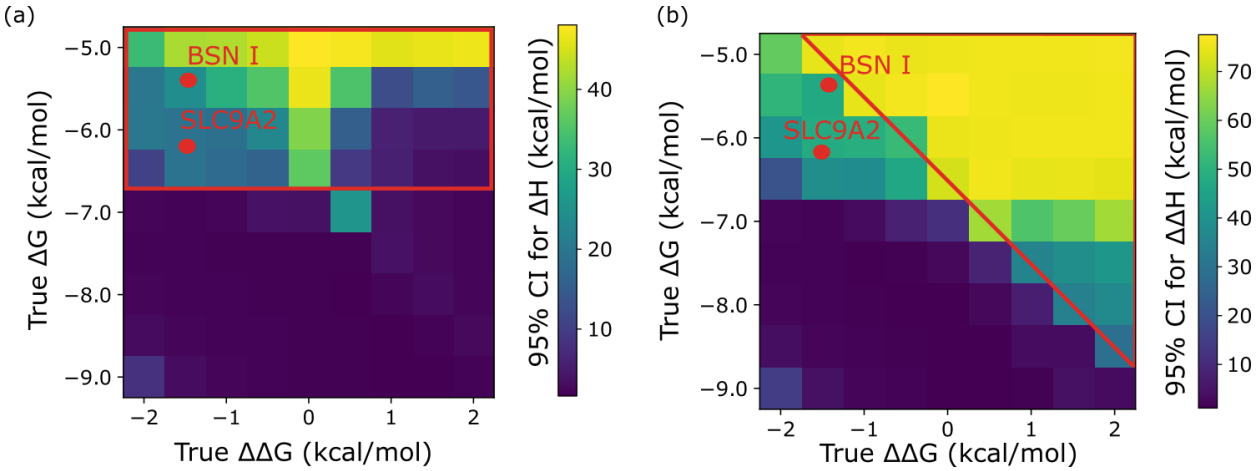

Figure 7: Phase diagram of width of posterior distributions as a function of model parameters. Two-dimensional plots along axes of ΔG and ΔΔG, wherein synthetic data with those parameters were generated, then sampled for posterior distributions at each point. Boxes are colored by the width of the 95% credibility region for ΔH (a) and ΔΔH (b), with lighter colors correspond to wider credibility regions (color bars). Red polygons demonstrate where each K_d (K_{d1} for left, K_{d2} for right) is greater than 17 μM, the cell concentration set for these synthetic isotherms. Red dots indicate mean values for experimental isotherms for binding for BSN motif I and SLC9A2

This overall loss in precision in cases of weaker binding is consistent with the concept of c in ITC experimental design (38, 43), with additional considerations. The parameter $c = n[\text{cell}]/K_d$, where n is binding stoichiometry, is a guide for setting experimental concentrations: ideally, c should be between 10 and 1000. For example, when n=1, the concentration in the cell should be 10-1000 times the K_d. In a multisite system, the exact relationship between each K_d and enthalpy determination is complicated by the existence of two binding constants, both of which may be outside of the relevant range and therefore limit precision. Conversely, model precision appears highest when both K_{d1} and K_{d2} are within the 10 to 1000 range for their respective c values. Concretely, the cell concentration for these simulated experiments is 17 μM, just below the K_d at a ΔG value of -6.5 kcal/mol. This neatly explains the steep drop-off in precision we see above -6.5 kcal/mol ΔG values for precision in ΔH, boxed in red in Figure 6. Similarly, when ΔG2 (i.e. ΔG + ΔΔG) is above -6.5, the precision drops off steeply for ΔΔH, seen along a diagonal boxed in red in the Figure.

For weak-binding isotherms such as BSN motif I, the lack of precision in enthalpy determination can be explained by a value for K_{d1} that is above the experimental concentration. For BSN motif I (location marked in Fig. 7, along with SLC9A2), ΔG is near -5 kcal/mol,

compensated for by a negative ΔΔG value such that overall binding still appears relatively strong. Functionally, this means that the first-glance evaluation of binding affinity excluding allosteric considerations can be somewhat misleading, hiding the fact that ΔG (and therefore K_{d1}) is relatively weak. Even hidden in this fashion, the weakness of K_{d1} nonetheless hurts our ability to accurately determine binding enthalpies and explains the variation we see in precision in enthalpy parameters throughout our tested data.

Discussion

This work examines the application of two-step binding models to isothermal titration calorimetry binding data, focusing on the hub protein LC8 and on accounting for critical uncertainties. LC8 binds over 100 client proteins in eukaryotic cells, and is involved in regulating a host of cell functions, motivating the detailed mechanistic study of LC8 binding. Using a Bayesian framework, we sought to determine precisely how much information can be extracted from a single isothermal titration (ITC) calorimetry isotherm, and examine how uncertainty in analyte concentration impacts model parameters, an investigation greatly aided by the use of simulated 'synthetic' isotherms with known parameters. Building on prior work (20, 35, 36), we have advanced Bayesian analysis of binding, and applied it to rigorous biophysical characterization of binding between LC8 and client peptides, as well as binding between the intermediate chain of dynein and the coiled-coil domain of NudE. We also used synthetic data to unambiguously separate effects of experimental error from intrinsic limitations, and we systematically surveyed the latter to generate a 'phase diagram' of intrinsic (in)tractability.

Allostery in LC8 binding

Our data show that LC8 can bind client proteins with significant positively cooperative allostery. Of the 7 peptides examined here, Bayesian analysis for all except one (BIM) yields a highly certain negative ΔΔG value, in agreement with early NMR studies that suggested positive allostery in LC8 binding (18). Further investigation is required to determine whether such behavior is universal for LC8 client peptides. For the present study, we selected test isotherms with preference for two criteria we anticipated would leverage Bayesian modeling: (1) tight-binding to LC8 and (2) an isotherm shape that breaks from a strict sigmoid. Because this was neither a comprehensive nor random selection of systems, more work will be needed to determine conclusively whether LC8 binding is uniformly positively cooperative and whether the degree of allostery is sequence dependent.

Our findings are a step toward understanding the underlying biological function of LC8 allostery. While LC8-client complexes are varied, the putative functional unit of many LC8-client interactions is a 2:2 bound structure, where LC8 promotes dimerization in client proteins (44–46). Further, In proteins where LC8-binding plays a structural role, such as at the nuclear pore in yeast (10, 47), fully bound states driven by cooperativity are more likely to be highly rigidified. In both the case of 2:2 binding, and structural complexes, then, the functional state is promoted by cooperativity. We have proposed that positive allostery could drive the formation of homodimeric complexes (18), with the same client bound to each LC8 motif, and that allostery could effectively encourage homologous complexes while discouraging heterologous ones. In the future, ITC or nuclear magnetic resonance titrations of a second peptide into a bound LC8-peptide complex could be used to examine whether heterologous binding is indeed disfavored. The picture of LC8-client binding is additionally complicated by the discovery over the last decade of multivalent LC8-binding proteins such as the nucleoporin NUP159 or the transcription factor ASCIZ. Complexes between LC8 and multivalent clients are often highly heterogeneous in both stoichiometry and conformation (7, 48, 49), and particularly in the case of ASCIZ, the fully bound state is highly disfavored by some form of negative cooperativity, ensuring ASCIZ is sensitive to LC8 even at high LC8 concentrations. The role that allostery of LC8 binding plays in these interactions is likely very complicated, and its relationship to effects seen in multivalent binding that rely on the length and structure of linkers between motifs (49, 50), remains to be seen.

The mechanism of allostery appears to be entropically driven. While entropy is often the term with the widest distribution (table 1), owing to its dependence on both the free energy and the enthalpy, there is a clear trend in our results towards positive $T\Delta\Delta S$ values, which equates to the second binding step being more entropically favorable than the first. Relatedly, NMR dynamics measurements indicate LC8's flexible core is rigidified on binding to clients (17, 51). Since LC8-binding allostery necessarily requires some change in the structural ensemble of LC8, it is possible that the first binding step can be thought of as 'paying up-front' for the entropic cost of both binding steps–i.e., rigidifying the whole LC8 core. This mechanism would also allow for variation in allostery on a per-peptide basis, as the degree of rigidification in the core seen by NMR is dependent on client sequence (17). Future molecular dynamics simulations can examine the differences in rigidity of the LC8 core in different bound states and across binding to different peptides.

Bayesian inference in binding analysis

"How much information is contained in an ITC isotherm?" is a fundamental biophysics question that Bayesian inference is uniquely suited to answer. Building on prior work (20, 35, 36), we have improved the ability of the Bayesian approach to account for the uncertainty that intrinsically occurs in *both* titrant and titrand concentrations. Our approach was motivated in large part by the apparently novel recognition of a mathematical "degeneracy" in ITC analysis, i.e., the existence of multiple solutions even in the absence of experimental noise, which prevents inference of a fully unique set of thermodynamic parameters. This degeneracy holds for simple 1:1 binding and apparently for arbitrary stoichiometry, as described in the Results.

While fitting ITC data to multi-step binding and other complex models is challenging, Bayesian inference allows for quantified "posterior" probability distributions for model parameters, reducing the risk of overfitting a complex model to insufficient data. These posterior distributions – or more accurately, the joint distribution over all binding parameters – fundamentally answer the question of the information contained in an ITC isotherm (20, 36). Bayesian inference is particularly powerful both for handling the degenerate nature of binding models that account for concentrations, and for experiments like ITC, which is very 'low-information' by nature (21, 22).

The Bayesian approach offers several advantages over frequentist fitting methods (20, 22) that are particularly useful in the case of ITC-measured binding: (1) Bayesian inference is not hampered by correlative or degenerate model solutions, allowing for inclusion of concentration parameters, (2) inferred distributions offer insight into correlative relationships in model parameters, and (3) the Bayesian model is highly flexible, and allows for incorporation of additional experimental data, whether additional isotherms or through the implementation of a variety of prior distributions.

Our investigation of how concentrations impact model parameters has shown that, as expected from the correlative relationships of the model parameters, that uncertainty in concentration significantly increases uncertainty in binding enthalpy, and has a reduced impact on free energy. This agrees well with Nguyen et al. (2018) who reported results on 1:1 binding, suggesting the concentration-enthalpy relationship is likely to be generic to all binding models. We have shown that while the individual concentrations may be indeterminable from the model alone, the ratio of concentrations can be readily determined (Supp Table S2).

Our primary goal has been to quantify uncertainty as completely as possible in determination of thermodynamic binding parameters. From a single isotherm, we sample

marginal posterior distributions with widths on the scale of 1-2 kcal/mol for a two-step model of binding with four thermodynamic parameters, consistent with prior Bayesian analysis (35, 36). Although this is much higher uncertainty than the fractions of a kcal/mol usually reported in the analysis of ITC data (23, 24), the difference can be explained to a great extent by the complexity of the two-step model, which intrinsically includes other correlative effects, e.g., between ΔG and ΔΔG, which are not accounted for in frequentist fitting methods. Additional uncertainty, beyond what can be attributed to the two-step binding model, arises from our 'skeptical' consideration of analyte concentrations, modeled by realistically wide concentration priors (± 10% for LC8, up to ±50% for peptides) that contribute to uncertainty in determined parameters. While 'microscopic' free energy and enthalpy parameters for individual binding steps cannot always be determined with good precision, the total values accounting for both steps show improved precision (Supp. Table S2, Supp. Fig. S2).

We believe that, in cases where higher precision is required for binding parameters, uncertainty can be decreased through the use of careful concentration determination through multiple methods, and the use of global models derived from multiple isotherms at varied concentrations.

Synthetic datasets guide experimentation

Our investigation has benefited significantly from the use of synthetic isotherms. Built from known thermodynamic parameters, and modeled using our Bayesian pipeline, the value of synthetic isotherms as an aid in experimental design is well-established (20, 22). They are particularly valuable in cases of complex binding, where it is not necessarily clear how determinable model parameters are, such as in our case. Synthetic isotherms have allowed us to test and troubleshoot our pipeline (Fig. 3), probe the information content of isotherms under variable conditions of concentration and priors (Fig. 3, Supp. Fig. S4), and examine how thermodynamic parameters themselves impact our ability to determine information from isotherms, resulting in the 'phase diagram' of relative tractability (Fig. 7). In the context of multi-isotherm modeling, utilizing synthetic data to design new experiments, such as is done with frequentist fitting in the program SEDPHAT (22, 34), is likely to be of significant value.

Practical limitations of Bayesian sampling and global modeling

Bayesian statistical analysis is much more computationally expensive than frequentist fitting methods. It usually relies on Markov chain Monte Carlo (MCMC) sampling, which requires simulating a sufficient number of steps to adequately explore the parameter space, potentially

including a need to locate and sample multiple probability peaks (akin to energy basins in conformation space). For our ITC model, simple MCMC sampling methods proved unable to adequately sample the model space, even following sampling times of several days and over 4 million samples. While the ensemble sampler(52) used by us and others applying Bayesian models to ITC (35, 36) has been robust for our purposes, sampling continues to be an important consideration, especially when considering future study of more complex models. For all work presented here, wall-clock sampling times were on the scale of hours, and hence readily feasible. More complex models could require significantly more sampling, although there is no simple scaling law that applies because of the uncertain nature of the parameter-space 'landscape'. Global modeling of multiple isotherms may also require additional sampling: as additional isotherms are added to a global model, each one brings with it a new set of nuisance parameters (4 per isotherm in our work - see Methods). In our hands, global models of two isotherms could be well-sampled within half a day. While global models of technical replicates may improve signal to noise ratios, ideally, global experiments should be designed with the intent of covering several experimental conditions (35, 36), and all experiments must be high quality to ensure they contribute to global fits.

As an aid to investigators employing Bayesian inference in future studies, we have developed a set of guiding best practices (see Supplemental Information).

Concluding remarks and future steps

Bayesian inference has allowed us to characterize the binding and allostery with high confidence for two different protein-protein interactions of 2:2 stoichiometry, despite significant uncertainties in analyte concentrations and inherent limitations of isothermal titration calorimetry. Our analysis was enabled by improvements to prior work (20, 35, 36) in treating concentration uncertainties, and further demonstrates the value of Bayesian inference to ITC analysis. We used synthetic data to systematically characterize the uncertainty landscape for 2:2 binding based on intrinsic binding properties, an approach that readily can be extended to other models.

We examined two multi-step binding systems, the hub protein LC8 and the dynein intermediate chain (IC), For LC8, every client peptide studied showed significant evidence of allostery, corroborating hypotheses from a decade ago (18), and thus serving as an important step toward quantitative characterization of more complex LC8-client complexes. In contrast, the dynein IC/NudE complex showed minimal evidence of allostery.

While our focus here has been on two-step symmetric-site binding systems, Bayesian methods can be applied to other complex models investigated by ITC. Measurement of complex multivalent systems, enthalpy-entropy compensation, and ternary complexes or competition binding are all likely to benefit from analysis under a Bayesian framework. Although there is a limit on how much information can be gained from individual isotherms, investigation utilizing synthetic data can guide design, to help determine experimental conditions that maximize gain from additional ITC experiments within a given system.

Acknowledgements

This work was supported by the U.S. National Institutes of Health grant R01-GM141733 and by U.S. National Science Foundation Grant MCB 2119837. We additionally acknowledge the support in the form of computational resources of the Oregon State University NMR Facility funded in part by the National Institutes of Health, HEI Grant 1S10OD018518, and by the M. J. Murdock Charitable Trust grant #2014162. We appreciate helpful discussions with John Chodera, David Minh, and Trung Hai Nguyen.

Author Contributions

Estelle, A.B.: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft

George, A.: Methodology, Software, Validation, Formal Analysis

Barbar. E.J.: Conceptualization, Supervision, Funding Acquisition, Project Administration, Writing – Reviewing and Editing

Zuckerman, D.M.: Conceptualization, Supervision, Funding Acquisition, Project Administration, Writing – Reviewing and Editing

Competing Interests

The authors declare no competing interests.

References

1. B. Alberts, *et al.*, *Molecular Biology of the Cell*, 4th Ed. (Garland Science, 2002).

2. R. Schneider, M. Blackledge, M. R. Jensen, Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Current Opinion in Structural Biology* **54**, 10–18 (2019).

3. J. Weng, W. Wang, Dynamic multivalent interactions of intrinsically disordered proteins. *Current Opinion in Structural Biology* **62**, 9–13 (2020).

4. E. Barbar, Dynein Light Chain LC8 Is a Dimerization Hub Essential in Diverse Protein Networks. *Biochemistry* **47**, 503–508 (2008).

5. P. Rapali, *et al.*, DYNLL/LC8: a light chain subunit of the dynein motor complex and beyond. *The FEBS Journal* **278**, 2980–2996 (2011).

6. S. Clark, A. Nyarko, F. Löhr, P. A. Karplus, E. Barbar, The Anchored Flexibility Model in LC8 Motif Recognition: Insights from the Chica Complex. *Biochemistry* **55**, 199–209 (2016).

7. S. Clark, *et al.*, Multivalency regulates activity in an intrinsically disordered transcription factor. *eLife* **7**, e36258 (2018).

8. S. A. Clark, N. Jespersen, C. Woodward, E. Barbar, Multivalent IDP assemblies: Unique properties of LC8-associated, IDP duplex scaffolds. *FEBS Letters* **589**, 2543–2551 (2015).

9. N. Jespersen, *et al.*, Systematic identification of recognition motifs for the hub protein LC8. *Life Sci Alliance* **2**, e201900366 (2019).

10. A. Nyarko, Y. Song, J. Nováček, L. Žídek, E. Barbar, Multiple Recognition Motifs in Nucleoporin Nup159 Provide a Stable and Rigid Nup159-Dyn2 Assembly. *Journal of Biological Chemistry* **288**, 2614–2622 (2013).

11. G. Erdős, *et al.*, Novel linear motif filtering protocol reveals the role of the LC8 dynein light chain in the Hippo pathway. *PLOS Computational Biology* **13**, e1005885 (2017).

12. K. L. West, *et al.*, LC8/DYNLL1 is a 53BP1 effector and regulates checkpoint activation. *Nucleic Acids Research* **47**, 6236–6249 (2019).

13. N. E. Jespersen, *et al.*, The LC8-RavP ensemble Structure Evinces A Role for LC8 in Regulating Lyssavirus Polymerase Functionality. *Journal of Molecular Biology* **431**, 4959–4977 (2019).

14. P. Luthra, D. S. Jordan, D. W. Leung, G. K. Amarasinghe, C. F. Basler, Ebola Virus VP35 Interaction with Dynein LC8 Regulates Viral RNA Synthesis. *Journal of Virology* **89**, 5148–5153 (2015).

15. J. Rodriguez Galvan, *et al.*, Human Parainfluenza Virus 3 Phosphoprotein Is a Tetramer and Shares Structural and Interaction Features with Ebola Phosphoprotein VP35. *Biomolecules* **11**, 1603 (2021).

16. J. Hall, P. A. Karplus, E. Barbar, Multivalency in the Assembly of Intrinsically Disordered Dynein Intermediate Chain. *Journal of Biological Chemistry* **284**, 33115–33121 (2009).

17. A. Nyarko, *et al.*, Conformational dynamics promote binding diversity of dynein light chain LC8. *Biophysical Chemistry* **159**, 41–47 (2011).

18. G. Benison, P. A. Karplus, E. Barbar, The Interplay of Ligand Binding and Quaternary Structure in the Diverse Interactions of Dynein Light Chain LC8. *Journal of Molecular Biology* **384**, 954–966 (2008).

19. MicroCal, Data Analysis in Origin (1998).

20. T. H. Nguyen, *et al.*, Bayesian analysis of isothermal titration calorimetry for binding thermodynamics. *PLOS ONE* **13**, e0203224 (2018).

21. L. Freiburger, K. Auclair, A. Mittermaier, Global ITC fitting methods in studies of protein allostery. *Methods* **76**, 149–161 (2015).

22. H. Zhao, G. Piszczek, P. Schuck, SEDPHAT – a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods* **76**, 137–148 (2015).

23. C. Feng, A. Roy, C. B. Post, Entropic allostery dominates the phosphorylation-dependent regulation of Syk tyrosine kinase release from immunoreceptor tyrosine-based activation motifs. *Protein Science* **27**, 1780–1796 (2018).

24. A. L. Lee, P. J. Sapienza, Thermodynamic and NMR Assessment of Ligand Cooperativity and Intersubunit Communication in Symmetric Dimers: Application to Thymidylate Synthase. *Frontiers in Molecular Biosciences* **5** (2018).

25. J. D. Chodera, D. L. Mobley, Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annu Rev Biophys* **42**, 121–142 (2013).

26. P. Czodrowski, C. A. Sotriffer, G. Klebe, Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pKa Calculations and ITC Experiments. *Journal of Molecular Biology* **367**, 1347–1356 (2007).

27. H. Steuber, P. Czodrowski, C. A. Sotriffer, G. Klebe, Tracing Changes in Protonation: A Prerequisite to Factorize Thermodynamic Data of Inhibitor Binding to Aldose Reductase. *Journal of Molecular Biology* **373**, 1305–1320 (2007).

28. S. Leavitt, E. Freire, Direct measurement of protein binding energetics by isothermal titration calorimetry. *Current Opinion in Structural Biology* **11**, 560–566 (2001).

29. A. Velazquez-Campoy, Y. Kiso, E. Freire, The Binding Energetics of First- and Second-Generation HIV-1 Protease Inhibitors: Implications for Drug Design. *Archives of Biochemistry and Biophysics* **390**, 169–175 (2001).

30. S. E. Boyce, J. Tellinghuisen, J. D. Chodera, "Avoiding accuracy-limiting pitfalls in the study of protein-ligand interactions with isothermal titration calorimetry" (Biochemistry, 2015) https:/doi.org/10.1101/023796 (June 6, 2022).

31. S. Contreras-Martos, *et al.*, Quantification of Intrinsically Disordered Proteins: A Problem Not Fully Appreciated. *Frontiers in Molecular Biosciences* **5** (2018).

32. D. G. Myszka, *et al.*, The ABRF-MIRG'02 Study: Assembly State, Thermodynamic, and Kinetic Analysis of an Enzyme/Inhibitor Interaction. *J Biomol Tech* **14**, 247–269 (2003).

33. MicroCal, VP-ITC Users Manual (2001).

34. J. C. D. Houtman, *et al.*, Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: Application to adaptor protein complexes in cell signaling. *Protein Science* **16**, 30–42 (2007).

35. M. V. C. Cardoso, *et al.*, CALX-CBD1 Ca2+-Binding Cooperativity Studied by NMR Spectroscopy and ITC with Bayesian Statistics. *Biophysical Journal* **119**, 337–348 (2020).

36. H. Duvvuri, L. C. Wheeler, M. J. Harms, pytc: Open-Source Python Software for Global Analyses of Isothermal Titration Calorimetry Data. *Biochemistry* **57**, 2578–2583 (2018).

37. A. Nyarko, Y. Song, E. Barbar, Intrinsic Disorder in Dynein Intermediate Chain Modulates Its Interactions with NudE and Dynactin. *Journal of Biological Chemistry* **287**, 24884–24893 (2012).

38. T. Wiseman, S. Williston, J. F. Brandts, L.-N. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical Biochemistry* **179**, 131–137 (1989).

39. G. Benison, M. Chiodo, P. A. Karplus, E. Barbar, Structural, Thermodynamic, and Kinetic Effects of a Phosphomimetic Mutation in Dynein Light Chain LC8. *Biochemistry* **48**, 11381–11389 (2009).

40. N. J. Anthis, G. M. Clore, Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci* **22**, 851–858 (2013).

41. J. Jie, F. Löhr, E. Barbar, Interactions of Yeast Dynein with Dynein Light Chain and Dynactin: GENERAL IMPLICATIONS FOR INTRINSICALLY DISORDERED DUPLEX SCAFFOLDS IN MULTIPROTEIN ASSEMBLIES. *Journal of Biological Chemistry* **290**, 23863–23874 (2015).

42. J. Jie, *Protien Disorder in the Evolution of Dynein Regulation.* (Oregon State University, 2016).

43. W. B. Turnbull, A. H. Daranas, On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **125**, 14859–14866 (2003).

44. J. R. Becker, *et al.*, The ASCIZ-DYNLL1 axis promotes 53BP1-dependent non-homologous end joining and PARP inhibitor sensitivity. *Nat Commun* **9**, 5406 (2018).

45. J.-B. Manneville, M. Jehanno, S. Etienne-Manneville, Dlg1 binds GKAP to control dynein association with microtubules, centrosome positioning, and cell polarity. *Journal of Cell Biology* **191**, 585–598 (2010).

46. L. Wang, M. Hare, T. S. Hays, E. Barbar, Dynein Light Chain LC8 Promotes Assembly of the Coiled-Coil Domain of Swallow Protein. *Biochemistry* **43**, 4611–4620 (2004).

47. M. Gaik, *et al.*, Structural basis for assembly and function of the Nup82 complex in the nuclear pore scaffold. *Journal of Cell Biology* **208**, 283–297 (2015).

48. B. Mostofian, *et al.*, Continuum dynamics and statistical correction of compositional heterogeneity in multivalent IDP oligomers resolved by single-particle EM. *Journal of Molecular Biology* **434**, 167520 (2022).

49. P. N. Reardon, *et al.*, The dynein light chain 8 (LC8) binds predominantly "in-register" to a multivalent intrinsically disordered partner. *Journal of Biological Chemistry* **295**, 4912–4922 (2020).

50. C. S. Sørensen, A. Jendroszek, M. Kjaergaard, Linker Dependence of Avidity in Multivalent Interactions Between Disordered Proteins. *Journal of Molecular Biology* **431**, 4784–4795 (2019).

51. J. Hall, A. Hall, N. Pursifull, E. Barbar, Differences in Dynamic Structure of LC8 Monomer, Dimer, and Dimer−Peptide Complexes. *Biochemistry* **47**, 11940–11952 (2008).

52. D. Foreman-Mackey, D. W. Hogg, D. Lang, J. Goodman, emcee : The MCMC Hammer. *Publications of the Astronomical Society of the Pacific* **125**, 306–312 (2013).

53. E. Freire, A. Schön, A. Velazquez-Campoy, "Chapter 5 Isothermal Titration Calorimetry: General Formalism Using Binding Polynomials" in *Methods in Enzymology*, Biothermodynamics, Part A., (Academic Press, 2009), pp. 127–155.

54. A. Gelman, *et al.*, Bayesian Data Analysis Third edition. 677.

55. D. W. Hogg, D. Foreman-Mackey, Data Analysis Recipes: Using Markov Chain Monte Carlo. *ApJS* **236**, 11 (2018).

56. T. Bayes, An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London* **53**, 370–418 (1763).

57. W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).

58. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 1087–1092 (1953).

59. S. Sharma, Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. *Annual Review of Astronomy and Astrophysics* **55**, 213–259 (2017).

60. J. Goodman, J. Weare, Ensemble samplers with affine invariance. *CAMCoS* **5**, 65–80 (2010).

METHODS

Binding Models

1:1 binding

For a 1:1 binding interactions between some macromolecule M, and ligand X:

$$M + X \rightleftharpoons MX \tag{1}$$

The energy of binding is described by the following quadratic equation (38)

$$\frac{Q}{V_0} = \frac{[M_t]\Delta H}{2}\left\{1 + \frac{[X_t]}{[M_t]} + \frac{K_d}{[M_t]} - \sqrt{\left(1 + \frac{[X_t]}{[M_t]} + \frac{K_d}{[M_t]}\right)^2 - \frac{4[X_t]}{[M_t]}}\right\} \tag{2}$$

$[M_t]$ and $[X_t]$ are the total concentrations of macromolecule and ligand after each injection, ΔH is the binding enthalpy, K_d is the binding affinity, V_0 is the volume of the cell, and Q is the heat of the system. This is directly equivalent to Origin's independent-sites model (19) when n=1. The observed measurement is dQ_i (i.e. the heat of injection i), which is calculated from Q using the following equation:

$$dQ_i = Q_i + \frac{V_i}{V_0}\left(\frac{Q_i + Q_{i-1}}{2}\right) - Q_{i-1} + \Delta H_0 \tag{3}$$

where V_i is the injection volume for injection i, to account for the change in volume associated with the injection. ΔH_0 is a correction term to account for heat of dilution and other effects that can shift, assumed constant over all injections in a given isotherm.

Two-step binding

Two-step binding is modeled in a standard fashion, such as in the binding polynomial model (53) as:

$$M + X \rightleftharpoons MX + X \rightleftharpoons MX_2 \tag{1}$$

Under this scheme, each binding affinity is as follows:

$$2K_{d1} = \frac{[X][M]}{[MX]}$$

$$\frac{1}{2}K_{d2} = \frac{[X][MX]}{[MX_2]} \tag{5,6}$$

where K_{d1} and K_{d2} are the affinities for the first and second binding step. Factors of 2 and ½ account for the existence of indistinguishable rotationally symmetric intermediates in our model. The total concentrations of X and M can be written as:

$$[M_t] = [M] + [MX] + [MX_2]$$
$$[X_t] = [X] + [MX] + 2[MX_2] \quad (7,8)$$

Through rearrangement and substitution of equations 5 and 6, the total concentration equations can be rewritten only in terms of [M] and [X], the concentrations of free macromolecule and ligand:

$$[M_t] = [M] + 2\frac{[X][M]}{K_{d1}} + \frac{[X]^2[M]}{K_{d1}K_{d2}}$$
$$[X_t] = [X] + 2\frac{[X][M]}{K_{d1}} + 2\frac{[X]^2[M]}{K_{d1}K_{d2}} \quad (9,10)$$

This system of equations is solved numerically for each given injection point to determine the unbound concentrations [M] and [X]. With both free concentrations determined, the system heat can be calculated:

$$\frac{Q}{V_0} = \Delta H_1[MX] + (\Delta H_1 + \Delta H_2)[MX_2] \quad (11)$$

where ΔH_1 and ΔH_2 are the enthalpies of binding step one and two respectively. The concentrations of each bound state can be calculated from [X] and [M] and equations 5 and 6. As in the 1:1 binding model, equation 3 is used to calculate the observed heat of injection, dQ, for each injection.

Degeneracy in two-step binding

When protein concentrations are included as model parameters, degenerate solutions are introduced. As outlined in the manuscript, the degeneracy is exposed from the following transformation:

$$[M_t] \longrightarrow \alpha[M_t]$$
$$[X_t] \longrightarrow \alpha[X_t]$$
$$K_{d1} \longrightarrow \alpha K_{d1}$$
$$\Delta G_1 \longrightarrow \Delta G_1 + RT\log\alpha$$
$$K_{d2} \longrightarrow \alpha K_{d2}$$
$$\Delta G_2 \longrightarrow \Delta G_2 + RT\log\alpha$$

$$\Delta H_1 \longrightarrow \frac{\Delta H_1}{\alpha}$$
$$\Delta H_2 \longrightarrow \frac{\Delta H_2}{\alpha}$$

(12)

Here, α can be any positive number. Following this transformation, the equations used to calculate [X] and [M] (eq. 5 and 6 in the methods) are transformed:

$$\alpha[M_t] = [M] + 2\frac{[X][M]}{\alpha K_{d1}} + \frac{[X]^2[M]}{\alpha^2 K_{d1}K_{d2}}$$
$$\alpha[X_t] = [X] + 2\frac{[X][M]}{\alpha K_{d1}} + 2\frac{[X]^2[M]}{\alpha^2 K_{d1}K_{d2}}$$

(13,14)

In these transformed concentration-sum equations, the new solutions for both [X] and [M] are exactly the previous solutions multiplied by α, as can be verified by substitution. Finally, applying the transformed values into the equation for Q yields

$$\frac{Q}{V_0} = 2\frac{\Delta H_1}{\alpha}\frac{\alpha[X]\alpha[M]}{\alpha K_{d1}} + \frac{(\Delta H_1 + \Delta H_2)}{\alpha}\frac{\alpha[M](\alpha[X])^2}{\alpha^2 K_{d1}K_{d2}}$$

(15)

As in the 1:1 binding model, cancellation of α shows there is no change in the value of Q for any α value. This demonstrates the degeneracy for 2:2 binding, which we can expect to generalize to higher stoichiometries.

Bayesian inference

Bayesian inference is a method to calculate a "posterior" *distribution* of model parameter values based on prior assumptions (encoded as prior distributions for parameters presumed to hold in the absence of data) and the data. In general, as more data is analyzed, the influence of the prior will decrease (54, 55) . The posterior distribution of parameters provides rich information such as the parameter means and confidence intervals (technically "credibility regions"), in addition to correlation information regarding whether and how parameters vary together.

Bayesian inference is based on Bayes' rule (54, 56) which enables us to infer a distribution of parameters θ (e.g., binding free energy and enthalpy, etc.) consistent with a given set of data D (e.g., ITC isotherms):

$$P(\theta|D) = P(D|\theta)\,P(\theta) \,/\, P(D)$$

(16)

where $P(\theta|D)$ is the (posterior) probability distribution of the model parameters, θ, given the data, D; $P(D|\theta)$ (the likelihood) is the probability distribution of the data given the model parameters and is given below; $P(\theta)$ (the prior) is the probability of the model parameters, specified below; and $P(D)$ (the evidence) is the probability of the data. For a given set of data, the unknown denominator $P(D)$ is constant, independent of parameters, so it does not affect the inference of posteriors. Typically, it is not possible to analytically solve Bayes' rule, so numerical methods such as Markov chain Monte Carlo are used to determine the target (posterior) distribution (57–59). Details of our implementation are given below.

Bayesian model

Following prior work (31, 36), we assume the data has Gaussian noise with a mean of zero and an unknown standard deviation. The ITC model parameters θ include concentration terms ($X_{initial}$,$M_{initial}$) and thermodynamic terms (ΔG, ΔΔG, ΔH, ΔΔH), as well as the nuisance parameters (ΔH_0 and σ) for heat of dilution and Gaussian noise. We use uniform prior distributions for the model parameters specified below and the unknown noise standard deviation unless otherwise stated. For global models (e.g. Supp. Fig. S7, Fig. 6), while it may be possible to assume a global noise or concentration model, we instead elected to apply global models with an additional set of concentration and nuisance parameters for each additional isotherm (bringing the total parameter count up to 12 for two-isotherm models). Uniform prior ranges for thermodynamic parameters were identical for all models, listed below in the methods table. For nuisance parameters ΔH_0 and σ, uniform priors of -10 to 10 μcal and 0.001 to 1 μcal respectively were used in all models.

The likelihood for a set of data $D = \{x_1, x_2, \dots\}$, denoted ($p(D|\theta)$), is the product of the probabilities at all data points x_i based on a normal distribution of standard deviation σ centered around $\mu_i(\theta)$, the calculated value of point *i* for the binding model and parameters θ. It therefore takes the following form:

$$p(D|\theta) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \mu_i)^2}{2\sigma^2}\right\} \quad (17)$$

and we note that σ is assumed unknown and sampled as part of the Bayesian inference process. When the priors are uniform, as we most often assume, the posterior is simply proportional to the likelihood given here.

Sampling

We use the affine-invariant Markov chain Monte Carlo sampling method (60) to perform Bayesian inference, as also used by Duvvuri et al. (2018) and Cardoso et al. (2020). The affine-invariant sampler is an ensemble-based method in which multiple walkers move through the sample space in a correlated fashion. We empirically found this method to sample significantly better than the standard Metropolis-Hastings (57, 58) sampler for our model. In our hands, the Metropolis-Hastings method was unable to converge on the target distribution after 4,000,000 sampling steps, whereas the affine-invariant sampler was able to converge after 100,000 sampling steps.

Implementation

We used the EMCEE package (52) in Python to perform the affine sampling, using a 20%:80% mix of the “differential evolution” and “stretch” move sets with 25-50 walkers. For each experiment, 3 replicas are run for 50,000-200,000 sampling steps/replica until convergence. Each replica converged, as determined by the autocorrelation time, where sampled steps must be greater than 50x the autocorrelation. Convergence was additionally assessed through examination of posterior distributions from model replicas, which were nearly identical in all cases (Supp. Fig. S5). This implementation runs at ~9 samples for each walker per second on 4 cores of a node on the Oregon State College of Science computing cluster.

The code, data, and an example notebook are available at:
https://github.com/ZuckermanLab/Bayesian_ITC

Isotherm	# samples	# walkers	dG prior (kcal/mol)	dH prior (kcal/mol)	ddG prior (kcal/mol)	ddH Prior (kcal/mol)	X initial Prior (µcal)	M initial Prior (µcal)
Synthetic isotherm models	50,000	25-50	-3 to -10	-50 to 0	-4 to 4	-40 to 40	Varied - ±10% of stated unless otherwise noted	Varied - ±10% of stated unless otherwise noted
Single isotherm experimental models	100,000	50	-3 to -10	-50 to 0	-4 to 4	-40 to 40	± 10% of stated	Up to ±50% of stated
Two isotherm	200,000	50	-3 to -10	-50 to 0	-4 to 4	-40 to 40	± 10% of	Up to ±50% of

experimental models							stated	stated

Methods table: Model priors and sampling lengths for all isotherms.

Experimental ITC

All isothermal titration calorimetry experiments used here have been previously reported in other publications (9, 42). Briefly, LC8, IC and NudE were all expressed in BL21 or Rosetta cell lines, and purified to 95% purity using a combination of 6xHis TALON affinity purification and size exclusion chromatography. LC8-binding peptides were purchased from Genscript. Proteins were dialyzed prior to calorimetry into a buffer of 50 mM NaPO4, 50 mM NaCl, 5 mM β-mercaptoethanol and 1 mM NaN3, at pH 7.5. In the case of LC8-peptide binding, Peptides were dissolved into buffer following dialysis to ensure minimal buffer mismatch between peptide and protein. All ITC experiments were performed at 25 C, with an initial injection of 2 µL, which was discarded to account for the first injection anomaly. Peak integration was performed in Origin 7.0.

Synthetic ITC isotherms

Synthetic isotherms for 1:1 and two-step binding were generated following equation 2 for 1:1 binding and equations 8,9 and 10 for two-step binding. Parameters were chosen to mimic typical experimental conditions employed in our group. For 1:1 binding (Supp. Fig. S4), we used ΔG and ΔH values of -12 and -8 respectively, and concentrations of 34 µM in the cell and 500 µM in the syringe. For two-step binding, varied thermodynamic parameters were used (e.g. Fig. 3, Fig. 7), but concentrations were fixed at 17 µM in the cell and 500 µM in the syringe. For all synthetic isotherms under both models, we simulated one injection of 2µL followed by 34 injections of 6 µL, with a ΔH_0 of 0 µcal, and added synthetic noise from a Gaussian distribution with standard deviation 0.2 µcal. To accurately replicate experimental conditions, we eliminated the first injection when applying models to this data.