Published in final edited form as:

J Chem Thermodyn. 2012 September 1; 52: 95–107. doi:10.1016/j.jct.2012.02.008.

Strategies for assessing proton linkage to bimolecular interactions by global analysis of isothermal titration calorimetry data

Nathan P. Coussens^a, Peter Schuck^b, and Huaying Zhao^{b,*}

^aLaboratory of Cellular and Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, U.S.A

^bDynamics of Macromolecular Assembly, Laboratory of Cellular Imaging and Macromolecular Biophysics, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD 20892, U.S.A

Abstract

Isothermal titration calorimetry (ITC) is a traditional and powerful method for studying the linkage of ligand binding to proton uptake or release. The theoretical framework has been developed for more than two decades and numerous applications have appeared. In the current work, we explored strategic aspects of experimental design. To this end, we simulated families of ITC data sets that embed different strategies with regard to the number of experiments, range of experimental pH, buffer ionization enthalpy, and temperature. We then re-analyzed the families of data sets in the context of global analysis, employing a proton linkage binding model implemented in the global data analysis platform SEDPHAT, and examined the information content of all data sets by a detailed statistical error analysis of the parameter estimates. In particular, we studied the impact of different assumptions about the knowledge of the exact concentrations of the components, which in practice presents an experimental limitation for many systems. For example, the uncertainty in concentration may reflect imperfectly known extinction coefficients and stock concentrations or may account for different extents of partial inactivation when working with proteins at different pH values. Our results show that the global analysis can yield reliable estimates of the thermodynamic parameters for intrinsic binding and protonation, and that in the context of the global analysis the exact molecular component concentrations may not be required. Additionally, a comparison of data from different experimental strategies illustrates the benefit of conducting experiments at a range of temperatures.

Keywords

protein interactions; thermodynamics; proton linkage; isothermal titration calorimetry; global analysis; SEDPHAT

*corresponding author: Huaying Zhao, Ph.D., National Institutes of Health, Bldg. 13, Rm. 3N17, 13 South Drive, Bethesda, MD 20892, phone: 301-594-5450, fax: 301-480-1242, zhaoh3@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Reversible macromolecular interactions of proteins are of great interest in physiology, as they play critical roles, for example, in molecular recognition, signal transduction and most other cellular subsystems. The affinity of association between molecular binding partners is a key parameter for understanding their role in these systems. For the detailed interpretation of the thermodynamic binding parameters from a structural perspective, electrostatic contributions and protonation/deprotonation processes play a crucial role. These physicochemical aspects can be probed by studying the pH dependence of the interaction. The pH is a critical factor in determining the state of the ionisable groups, *via* protonation/deprotonation processes, which may be involved in the macromolecular interactions under study. Theoretical modeling of protein-protein complexes based on 75 structures of protein complexes quantitatively predicted the impact of protonation on the binding energy, suggesting a contribution of at least $5.9 \text{ kJ}\cdot\text{mol}^{-1}$ to the apparent free energy [1] for ~15 % of the structures under investigation. This corresponds to one order of magnitude change in the equilibrium binding constant at $T = 298 \text{ K}$. Therefore, if this factor is not accounted for in the measurement of the equilibrium association constant (K_{obs}), the observed value at a given pH may not reflect well the intrinsic macromolecular binding constant (K_{int}), and may provide inaccurate or insufficient information for understanding the intrinsic physicochemical driving forces for the interactions that control the reaction mechanisms. Similarly, studies of binding induced changes in protonation have been applied in a variety of different contexts, including the determination of affinity constants too large to measure at physiological pH [2] and deconvolution of thermodynamic contributions to protein-protein [3; 4], protein-DNA and protein- small molecule interactions [5; 6]. In the present work, we focus on a protein interaction with proton linkage, although the methodology and conclusions will equally apply to other bimolecular interactions coupled with proton linkage.

Binding involving protonation is a question of linked equilibria. On a molecular level, proton linkage of protein-ligand and protein-protein interactions is resulting from a binding-induced change in the pK_a values of one or more specific ionisable groups, which may not be directly identified from crystal or solution structures. On the other hand, protonation is accompanied by an enthalpy change that is strongly dependent on the chemical environment of the functional group. This characteristic of protonation lends itself particularly well to the application of isothermal titration calorimetry (ITC), from which the heat involved in binding as well as in protonation can be monitored under different buffer conditions. ITC has the potential to provide a complete thermodynamic description of a binding reaction with linked protonation through the determination of the equilibrium dissociation constant, K_d , the enthalpy change, ΔH° , the entropy change, ΔS° and the heat capacity change, ΔC_p . For proton linked binding, the observed ΔH is dependent on the intrinsic enthalpy of binding as well as the protonation and buffer ionization enthalpies, as outlined in a theoretical framework by Baker and Murphy [7] and Doyal *et al.*[2], Bradshaw *et al.*[8]. Consideration of proton linkage will allow determination of the intrinsic macromolecular change of enthalpy, change of entropy, and change of heat capacity upon binding, which have been correlated with structural features, such as buried solvent accessible surface areas [4; 9; 10; 11; 12; 13]. Accordingly, the analysis of proton linkage by ITC is an active field, with more than 100 publications in the recent literature. Examples include the binding between aspartic protease and its inhibitors [14; 15], T cell receptor recognition of major histocompatibility complex (MHC) proteins [3], and ligand binding to trypsin and thrombin [16; 17]. Furthermore, deconvolution of the thermodynamic contributions to a bimolecular interaction, including proton linkage or other equilibria, is critical during the process of rational drug design, since the thermodynamic components can be further attributed to

energetics of hydrogen bonding, electrostatic interactions, hydrophobic interactions, conformational changes, molecular flexibility and solvent effects [18; 19; 20; 21; 22].

Typically, the determination of proton linkage requires measurements of equilibrium binding constants at a number of pH values, at which the protonated and deprotonated species, as well as the free (un-liganded by the macromolecular binding partner) and bound (*i.e.* macromolecular complex) species can be populated. Furthermore, in order to dissect the contribution of protonation to the enthalpy change of the reaction, experiments with different buffer ionization enthalpies are necessary. Finally, in order to assess the heat capacity change, which relates to structural features of the macromolecular complex, binding experiments must be conducted at different temperatures. It was shown previously by Armstrong *et al.* [3] that fitting a single, unambiguous mathematical binding model to all available experimental data at the different conditions is the most powerful analysis approach. Global analysis has also been found to be a very powerful ITC data analysis concept for the study of two- and three-component complexes with multi-site binding and cooperativity [23; 24; 25; 26; 27; 28; 29]. For the latter purpose we have previously introduced the global modeling capabilities of the software SEDPHAT [26], which is a public domain analysis tool for the global and multi-method modelling of biophysical binding data from various techniques besides ITC, including analytical ultracentrifugation, light scattering, surface plasmon resonance biosensing, and different spectroscopy techniques. It offers many binding models, is flexible and user-friendly in that it has a graphical user interface that does not require any scripting, and has found various applications in ITC to study protein interactions [23; 27; 29; 30; 31; 32; 33; 34]. For the present work we have implemented two protonation models for global analysis of ITC data with multiple buffer ionization enthalpies, temperatures, and/or pH values.

For any study of molecular interactions by ITC, it is critical to precisely know the concentrations of the molecules under study. Unfortunately, in practice this is far from trivial especially for proteins, since accurate protein dry weight measurements would require > 10 mg of highly pure and soluble material, which is usually prohibitive. The prediction of protein extinction coefficients at 280 nm (provided there are aromatic amino acids) is imprecise and easily carries at a (5 to 10) % or higher error when predicted from the amino acid composition [35]. More importantly, proteins often contain fractions of misfolded and inactive material, which do not participate in the interactions of interest but contribute to the spectroscopic or other measurements of concentration. Historically, for simple 1:1 binding analyses this problem is often captured in an '*n*-value' that subsumes errors of both binding partner concentrations in the syringe and in the cell, as well as empirically describing the number of sites on the titrant in the cell and/or the inverse of the number of sites on the injectant in the syringe. SEDPHAT deviates from this tradition in that the number of sites always assumes integral values in binding models that are chosen *a priori* (and may be tested against each other in their performance of fitting the data). Errors in the active concentrations are either described with concentration correction factors or with incompetent fractions (*f*) of each component (such that the effective concentration is the apparent value times 1- *f*) [26]. The ambiguity in the interpretation of an '*n*-value' does not completely vanish, however, but translates to the problem that when a single 1:1 binding model is fit to a single data set and neither the cell nor the syringe concentration are known accurately *a priori*, refinement of the concentration parameters will be completely correlated with the macromolecular binding parameters of K_d and ΔH° . Nevertheless, beyond the fact that the explicit 'incompetent fraction' and 'concentration correction' picture is more satisfying because it is physically more accurate, it has the real advantage that, in contrast to the '*n*-value' approach, it is straightforward to apply in the context of global modelling. Here, dependent on the nature of the binding partners, different incompetent fractions may be ascribed to all experiments individually for example, when different protein preparations

are involved in each experiment. Alternatively, concentration errors may be considered as global parameters; for example, when the only uncertainty that needs to be considered might be an error in an extinction coefficient or in the stock concentration of a protein preparation or a ligand solution. The local and global estimates of incompetent fractions are bounded by non-negativity constraints (or concentration error parameters can be bounded to assume values only in a certain range). In the context of a global model, it is not obvious and has not yet been examined to what extent concentration uncertainties translate into uncertainties of macromolecular binding parameters.

From a practical perspective, the question of active protein concentrations is particularly relevant for studies of the proton linkage of protein interactions. When experiments have to be conducted over a range of different pH values and in different buffers, not all conditions may be equally suitable for protein stability and solubility. Furthermore, the buffer exchange requires an additional preparative step, which may easily lead to a certain degree of protein degradation, especially when conducted by equilibrium dialysis with its obligatory time requirement. Therefore, it is important to know how tolerant or sensitive the global proton linkage analysis by ITC is towards the presence of unknown concentration errors.

In the present work, we aimed at answering the questions of which experiments would need to be performed, and how well would the active protein concentrations need to be known, in order to either reliably determine the protonation and intrinsic binding parameters in the global analysis, or to at least confidently detect the presence of linked protonation. Towards this end, we focused on a simple bimolecular interaction with linked protonation as a model system, and generated with SEDPHAT multiple ITC data sets designed to correspond to certain practically feasible experimental conditions with regard to pH, buffer ionization enthalpy and temperature. Data sets were grouped into families for global analysis according to different 'experimental' strategies. By re-analyzing these global analysis sets with regard to their statistical parameter uncertainties and parameter cross-correlations, we examined the information content of the simulated experiment series, provided different assumptions on the knowledge of concentration parameters. Our results confirm that proton linkage can be qualitatively diagnosed with ITC experiments at a single pH with different buffer ionization enthalpies, even in the presence of uncertainties of reactant concentrations, but that the quantitative determination of the intrinsic binding and protonation properties requires at least three pH values and multiple buffers. We found that the extension of the experimental data set to include various temperatures can greatly improve the precision of all binding parameters, even though the determination of heat capacity changes from protonation may prove difficult if below $0.4 \text{ kJ}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$. These analyses also illustrate the performance of the global analysis platform and the application of the statistical tools available in SEDPHAT for fitting and examining complex binding models.

2. Theory

2.1 Protonation linked binding for a bimolecular hetero-association

The theory for protonation-linked binding of macromolecules has been outlined previously [2; 3; 7; 8]. For clarity of the current work, in the following we briefly recapitulate the theoretical framework.

As indicated in the reaction scheme figure 1, we consider a protein P and a ligand L, where the ligand could be any molecule binding to the protein, either a small compound or another macromolecule such as a protein, and either P or L may be initially in the cell or in the syringe in the ITC experiment. The binding of P and L is governed by an intrinsic equilibrium association constant K_{int} and an enthalpy change of binding ΔH_{int}^0 . The protonation event can take place on the unliganded (*i.e.* free) protein, with an equilibrium

association constant $K_p^{(f)}$ and the enthalpy change $\Delta H_p^{(f)}$ where the subscript 'p' indicates the protonation reaction, and the superscript '(f)' specifies the free protein. We assume in the following the presence of a single ionisable group (see **Discussion**). Although protonation leaves the activity of protons a_{H^+} unchanged by virtue of the sufficiently high buffer concentration in solution, account must be made of the enthalpy of deprotonation of the buffer substance in solution, ΔH_{buffer} . The protonated protein can also bind the same ligand, but now with a different affinity constant, $K_{int,p}$, and enthalpy change $\Delta H_{int,p}$. Finally, completing the cycle, the proton can also bind to the liganded protein (or protein complex), with the equilibrium constant $K_p^{(c)}$ and the enthalpy change $\Delta H_p^{(c)}$, with the superscript 'c' specifying the protein/ligand complex.

In the presence of proton linkage, the observed binding constant K_{obs} between the protein and ligand follows the mass action law

$$K_{obs} = \frac{([PL] + [P^+L])}{([P] + [P^+]) \cdot [L]} \quad (1)$$

which following figure 1 can be expressed as

$$K_{obs} = K_{int} \frac{1 + K_p^{(c)} \times 10^{-pH}}{1 + K_p^{(f)} \times 10^{-pH}} \quad (2)$$

where we have substituted 10^{-pH} for the proton activity a_{H^+} . Equation (2) indicates that the intrinsic binding constant K_{int} can be resolved by varying pH.

From ITC data, the observed enthalpy, ΔH_{obs}^o is the sum of the binding enthalpy independent of the buffer used in the experiment, ΔH_0^o , and the contribution of the proton ionization of the buffer, described as

$$\Delta H_{obs}^o = \Delta H_0^o + \Delta n_{H^+} \Delta H_{buffer} \quad (3)$$

where Δn_{H^+} is the change in the fractional proton occupancy of the ionisable group in the complex relative to the free state. In principle, equation (3) may be used solely to determine Δn_{H^+} by linear regression of ΔH_{obs}^o as a function of buffer ionization enthalpy [7]. An alternative global analysis based on equation (3) for data at the same pH and temperature but with different buffer ionization enthalpies has been implemented in SEDPHAT, treating K_{obs} , Δn_{H^+} and ΔH_0^o as global unknown parameters. Such a simplified analysis will not reveal intrinsic binding parameters, but may have the potential to prove the presence of protonation linkage. Furthermore, absolute Δn_{H^+} values greater than 1.0 would indicate the presence of more than one ionisable group that are linked to binding (see **Discussion**). Generally, continuing with the assumption of a single ionisable group, Δn_{H^+} will be less than 1.0 [7], since it will be dependent on the difference in the average occupation of the ionisable group in complex, $\bar{H}^{(c)}$, versus free state, $\bar{H}^{(f)}$:

$$\begin{aligned} \bar{H}^{(f)} &= \frac{K_p^{(f)} \times 10^{-pH}}{1 + K_p^{(f)} \times 10^{-pH}} \\ \bar{H}^{(c)} &= \frac{K_p^{(c)} \times 10^{-pH}}{1 + K_p^{(c)} \times 10^{-pH}} \end{aligned} \quad (4)$$

$$\Delta n_{H^+} = \bar{H}^{(c)} - \bar{H}^{(f)} \quad (5)$$

which depends on the shift in the pK_a of the ionisable group upon binding, as well as the solution pH. Therefore, it is advantageous to conduct experiments with different buffer ionization enthalpies also at different pH values, to the extent possible. With this in mind, we can further decompose the component of the enthalpy change that is independent of the buffer ΔH_0^o into the contributions from changes in the fractional occupation of the ionisable groups from $\bar{H}^{(f)}$ to $\bar{H}^{(c)}$ with the associated heats of protonation in the two states $\Delta H_p^{(f)}$ and $\Delta H_p^{(c)}$, respectively, and the contribution intrinsic to the protein-ligand interaction (such as the heats arising from their contact area), and write the total heat of binding as

$$\Delta H_{obs}^o = \Delta H_{int}^o + \bar{H}^{(c)} \Delta H_p^{(c)} - \bar{H}^{(f)} \Delta H_p^{(f)} + (\bar{H}^{(c)} - \bar{H}^{(f)}) \Delta H_{buffer}. \quad (6)$$

In the implementation of these relationships in SEDPHAT, a nomenclature was adopted to express the protonation affinity $K_p^{(f)}$ as pK_p and the shift upon complex formation from $K_p^{(f)}$ to $K_p^{(c)}$ also in logarithmic units as $\Delta pK_a^{(c-f)}$. This is in keeping with the custom to phrase binding constants, such as K_{int} , as a base-10 logarithm of the constant in molar units, *i.e.*, $\log_{10}(K_{int})$, for better scaling. Similarly, instead of treating the enthalpy changes $\Delta H_p^{(f)}$ and $\Delta H_p^{(c)}$ as two independent parameters, the analysis was described as $\Delta H_p^{(f)}$ and $\Delta \Delta H_p^{(c-f)} = \Delta H_p^{(c)} - \Delta H_p^{(f)}$. This simplifies constraining the model to the null hypothesis that there is no proton linkage ($\Delta pK_a^{(c-f)} = 0$ and $\Delta \Delta H_p^{(c-f)} = 0$) with unknown $K_p^{(f)}$ and $\Delta H_p^{(f)}$, in order to contrast the quality of fit with that of the model allowing for proton linkage. It should also be noted that, without restrictions of the generality of the model, the protein component in the proton linkage model is designated as component 'A' and the other species as 'B'.

2.2 Temperature dependence of proton linkage and heat capacity changes

When temperature is included as another variable for experimental conditions, we can account for the corresponding changes of binding affinities and enthalpy changes with the expression

$$K(T) = K(T_{ref}) e^{\left[\frac{\Delta H(T_{ref})}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}} \right) + \frac{\Delta C_p}{R} \left(\frac{T_{ref}}{T} - 1 + \ln \frac{T}{T_{ref}} \right) \right]}, \quad (7)$$

based on the modified Gibbs-Helmholtz equation [3], and assuming a temperature-independent heat capacity change

$$\Delta H(T) = \Delta H(T_{ref}) + \Delta C_p (T - T_{ref}). \quad (8)$$

Equations (7) and (8) can be applied to all binding constants and enthalpy changes in equations (1–6), introducing the new parameters $\Delta C_{p,int}$ for the heat capacity change of the intrinsic binding, $\Delta C_{p,p}^{(f)}$ and $\Delta C_{p,p}^{(c)}$ for the heat capacity changes of protonation of the free and liganded protein, respectively. Consistent with the nomenclature of the other binding parameters, the change of the heat capacities upon protein/ligand complex formation, $\Delta \Delta C_{p,p}^{(c-f)}$, is treated as an independent parameter, alongside $\Delta C_{p,p}^{(f)}$.

2.3 ITC linked protonation analysis in SEDPHAT

SEDPHAT allows one to load a large number of experimental ITC titration data into a single global analysis. For intrinsic macromolecular binding and protonation parameters, the observed affinity and enthalpy are calculated at the experimental temperature, buffer pH and ionization enthalpy of the individual experiments with equations (2), (6), (7), and (8), using a reference temperature of 298 K. For convenience, pre-defined ionization enthalpies and heat capacity changes are implemented for several common buffers, using the tables of [36], or they may be user-determined. Dependent on the specified concentrations of binding partners in the cell and the syringe, as well as the injection schedule, changes in reactant concentrations before and after each injection are calculated. From the changes in fractional saturation of complex formation, the observed heats of binding are then calculated for each injection [26]. In the current version, considering the different instrumental designs of the ITC units from the major manufacturers, two different models are implemented for the treatment of the amount of reaction volume being expelled out of the calorimeter cell during the injection: a model for a completely mixed neck [26], and unmixed neck [37] (the default) that can be selected depending on the instrument used. Also in the current version, concentration errors can be directly treated as unknown concentrations, optionally bound by a certain parameter range, with a correction factor refined either as a local parameter specific to each data set, or globally as a joint parameter shared between several data sets.

Alternatively, errors in the active concentration can also be described as non-negative incompetent fractions, which likewise can be either globally shared unknowns, or unknown parameters specific to a particular titration. Although these concepts currently co-exist, they are redundant and either one or the other should be used. In the current work, only the incompetent fractions were applied to describe errors in active concentrations. In addition, baseline offset parameters can optionally be added to all data sets in order to describe imperfect buffer matching and residual contributions of trivial heats of dilution to the measured data. This was applied in the current work throughout.

For a given set of global binding parameters and local experimental parameters, the model is calculated for each isotherm, a local root-mean-square deviation (rmsd) is calculated, and a normalized chi-square is calculated as a measure of the global fit as

$$\chi_r^2 = \frac{1}{N_{tot}} \sum_e \sum_{i=1}^{N_e} \frac{(a_i^{(e)} - f_i^{(e)})^2}{\sigma_e^2} \quad (9)$$

where $a_i^{(e)}$ and $f_i^{(e)}$ are the experimental data point and theoretical prediction of data point i from the experimental set e , which has an error of data acquisition of σ_e and contains N_e data points, with a total of N_{tot} data points globally. Non-linear least squares regression by Marquardt-Levenberg, simplex, and simulated annealing are available for finding the best global fit, allowing for optional user-defined bounds on all non-linear fitting parameters. For the statistical error analysis, a variety of tools are available, including the co-variance matrix, the error projection method with F-statistics [38; 39; 40], plotting of one-dimensional and two-dimensional error contours, and Monte-Carlo analyses. For the present work we mapped error contours and calculated critical confidence levels of equation (9) based on [38]. SEDPHAT can be downloaded from [41], and a detailed description of the practical steps of ITC analysis in SEDPHAT can be found at [42].

3. Results

3.1 Simulated data sets for the global protonation analysis

In order to investigate different experimental approaches for the study of proton linked binding, we utilized the simulation function of SEDPHAT to generate multiple ITC data sets with the hetero-association global proton linkage analysis model. Values for the intrinsic binding parameters K_{int} , $\Delta H_{\text{int}}^{\circ}$, and $\Delta C_{\text{p,int}}$, were taken to mimic those of typical protein interactions, and those describing the protonation $\text{p}K_{\text{a}}^{(\text{f})}$, $\Delta H_{\text{p}}^{(\text{f})}$, $\Delta C_{\text{p,p}}^{(\text{f})}$, $\Delta \text{p}K_{\text{a}}^{(\text{c-f})}$, $\Delta \Delta H_{\text{p}}^{(\text{c-f})}$, and $\Delta \Delta C_{\text{p,p}}^{(\text{c-f})}$ were taken similar to those previously obtained experimentally (table 1) [43; 44]. For simplicity, a single set of concentrations was used, with species A of 10 μM in the cell, and B of 100 μM in the syringe. A 15 % incompetent fraction was assigned to both A and B, in order to make the analysis realistic and avoid non-negativity constraints of the incompetent fraction parameter to impact the statistical analysis. Normally a distributed noise of 418 $\text{J}\cdot\text{mol}^{-1}$ was added to each data set. Values for buffer properties ΔH_{buffer} and $\Delta C_{\text{p,buffer}}$ were taken from the table of Goldberg *et al.* [36].

Simulated experimental conditions were applied in different families of data sets to represent five experimental strategies of proton linkage studies (table 2). Strategy #1 confines the experimental conditions to the variation of buffer ionization enthalpy, at a single pH value equal to the $\text{p}K_{\text{a}}$ of the bound protein complex. Strategy #2 includes three different buffer ionization experiments at two different pH values, adding a lower pH that would better populate the protonated state of the complex. Strategy #3 expands on this further, by adding a set of buffer ionization experiments at a high pH that will nearly fully populate the de-protonated state of the protein complex. Strategy #4 increases the number of pH values in the global set to 6 by adding a very low and two intermediate pH data sets, but has only a single buffer at each of those additional pH values. Furthermore, it explores another experimental dimension by including experiments at three temperatures of 283 K, 298 K, and 310 K. Finally, strategy #5 is similar to strategy #4 in the number and pH of experimental data sets, but limited to a single temperature, and instead varying the buffer ionization enthalpy. For each family, the respective simulated experimental data were assembled into a global analysis in SEDPHAT, for re-analysis and statistical evaluation of information content. For example, the data and fit of the global analysis of strategy #4 are shown in figure 2.

3.2 Diagnosing the presence of proton linkage from observed quantities $\Delta H_{\text{obs}}^{\circ}$ and K_{obs}

The simplest way to determine whether an interaction involves proton linkage is to perform identical ITC titrations at a single pH in multiple buffers with different ionization enthalpies. As shown in equation (3), the observed enthalpy change $\Delta H_{\text{obs}}^{\circ}$ follows a linear relationship with ΔH_{buffer} , with the slope $\Delta n_{\text{H}^{+}}$ reflecting the change upon ligand binding in average proton occupancy of the ionisable group. We employed the data of strategy #1 to represent this approach. Rather than fitting each experiment to obtain $\Delta H_{\text{obs}}^{\circ}$ for the subsequent linear extrapolation, we employed the corresponding global analysis model in SEDPHAT in which $\Delta n_{\text{H}^{+}}$ is considered a global parameter, along with ΔH_0° and a binding affinity constant, K_{obs} . Specifically, $\Delta n_{\text{H}^{+}}$ is a parameter of particular interest in this context because any value with $|\Delta n_{\text{H}^{+}}| > 0$ will qualitatively show the presence of proton linkage. We tested the information content for all parameters by tracing the error surface projection onto the parameter axis, *i.e.*, by calculating the global χ_r^2 in equation (9) for fixed values of the parameter of interest, while all other unknowns are allowed to be optimized to try compensating the constraint. The resulting 1-dimensional maps of the error surface projections show well-defined minima for all global parameters when the incompetent fractions f_A and f_B of the macromolecule and the ligand, respectively, were fixed to the correct values underlying the simulation (figure 3, black line). The $\Delta n_{\text{H}^{+}}$ value was 0.269

(95 % confidence interval 0.239 – 0.298), which is also consistent with the parameters underlying the simulation.

When both incompetent fractions were allowed to be optimized in the global fit, only a lower limit for Δn_{H^+} of 0.18 could be determined. Thus, on a 95 % confidence level, the uptake of protons upon binding is statistically supported by the data. The inability to determine the parameter values is due to the high correlation among them. In this case, the correlation is introduced by the incompetent fraction parameters, which can compensate for changes in the binding parameters. Only when an incompetent fraction parameter reaches zero and its non-negativity constraint becomes effective is the correlation broken, and altered binding parameters will have an impact on the quality of fit. This is when the error contour map can have a discontinuous derivative and enter a steeper branch, which in this case provides the one-sided error limit (figure 3, blue line). We observed an error contour map with a very similar shape when the protein in the cell is allowed to have incompetent fractions that may vary from experiment to experiment in the different buffers, while the concentration of the ligand in the syringe is described by one single global incompetent fraction parameter (figure 3, green line), but only a much more shallow map that does not reach the 68% confidence level is obtained when all incompetent fraction parameters freely float as local parameters in each individual experiment. Thus, our ability to reliably detect protonation is intimately linked with our confidence in the active concentrations of protein and ligand in the cell, and in the bounds that we set for these parameters. Analogous observations with regard to the influence of constraints on the confidence interval of proton linkage and intrinsic binding parameters were made with all other models, as described in the following.

3.3 Determination of intrinsic binding affinities and the proton affinity for the free and complex states

Due to the limited information of the data at a single pH, strategy #1 is primarily used only to verify that a binding event is linked to protonation and to determine whether protons are absorbed or released upon binding. Next, we explored the global analysis of the 6 experiments of strategy #2 covering 2 pH conditions. For this, we switched the SEDPHAT model from that of equation (3), with its three parameters for the observed binding to a larger model embedding equations (2) and (6–8), with its nine mechanistic parameters. When fixing f_A and f_B at the correct values and adjusting the 6 global parameters $\log_{10}(K_{\text{int}})$, $\Delta H_{\text{int}}^{\circ}$, $pK_a^{(f)}$, $\Delta H_p^{(f)}$, $\Delta pK_a^{(c-f)}$, and $\Delta \Delta H_p^{(c-f)}$, we found all affinity parameters to be well determined by the data, but not the enthalpy parameters $\Delta H_{\text{int}}^{\circ}$, $\Delta H_p^{(f)}$, and $\Delta \Delta H_p^{(c-f)}$ (see table 1). Due to the absence of temperature variation in this global data set, heat capacity parameters are meaningless and were ignored. Among the affinity parameters, the most important one is $\Delta pK_a^{(c-f)}$, since any significant non-zero value demonstrates the presence of proton linkage. Its error projection map is shown as a blue line in figure 4a. The failure to determine the enthalpy parameters with strategy #2 (figure 4e, magenta line) can be understood conceptually from the inspection of equation (6), where three enthalpy parameters all contribute to $\Delta H_{\text{obs}}^{\circ}$, such that a minimum of three pH conditions are necessary in order to provide three different sets of coefficients to have a well-determined linear system for the unknown enthalpy parameters.

When examining the condition under which the concentrations for both A and B are uncertain, we allowed f_A and f_B to be refined as global parameters, and found that $\log_{10}(K_{\text{int}})$, $pK_a^{(f)}$ and $\Delta pK_a^{(c-f)}$ were still well determined by the data although with significantly wider minimum and correspondingly larger error intervals than with fixed f_A and f_B (table 1, and figure 4a, black line), dependent on the bounds of f_A and f_B . With the non-negativity constraints for incompetent fractions, the 95% confidence interval for f_A was 0 – 0.50, and for f_B 0 – 0.48. Next, we further relaxed the assumptions on the knowledge of

reactant concentrations, by considering f_A an adjustable local parameter, potentially different from experiment to experiment, while f_B is still considered a global unknown common to all experiments. The resulting parameter uncertainties become only slightly larger, as shown, for example in figure 4a, red line, and table 1. In this case, there is a strong correlation between the $\Delta pK_a^{(c-f)}$ parameter and f_B , as illustrated in the diagonal features of the two-dimensional error projection map figure 5a. Finally, in the case of both f_A and f_B being locally adjustable unknowns, the model was found too flexible to determine any of the parameters (figure 4a, magenta line). Overall, strategy #2 clearly illustrated that even in the presence of substantial uncertainty in the species concentrations, with experiments at two pH values and buffers with three different ionization enthalpies at each pH, the intrinsic binding constant and proton binding constants can be resolved reliably.

Next, we considered strategy #3 that with its three pH values at 6.5, 7.5, and 8.5 has the potential to resolve the three enthalpy parameters. For the simplest scenario assuming we knew the correct incompetent fractions f_A and f_B of 15 % each underlying the simulation, all 6 adjustable parameters for affinity and enthalpy ($\log_{10}(K_{int})$, ΔH_{int}° , $pK_a^{(f)}$, $\Delta H_p^{(f)}$, $\Delta pK_a^{(c-f)}$, $\Delta \Delta H_p^{(c-f)}$) were determined within 95 % confidence intervals as presented in table 1. For the $\Delta pK_a^{(c-f)}$ parameter the error contour map is shown in figure 4b, blue line. Treating f_A and f_B as globally adjustable parameters also produced good results, slightly better than strategy #2 with the same types of concentration uncertainties. However, for ΔH_{int}° the error intervals are much wider than those from the fixed incompetent fractions ((-84.1, -52.3) compared to (-64.4, -61.5) kJ·mol⁻¹). The dominant impact of the incompetent fractions on ΔH_{int}° is due to their correlation with this parameter, which can be seen from the covariance matrix (Supplement table 4). The correlation coefficients between ΔH_{int}° and f_A or f_B are both -1.0 indicating a complete correlation unless a non-negative constraint for f_A or f_B takes effect. Further, we applied more freedom to the incompetent fractions in the global analysis, while optimizing f_A as a local parameter for each titration and f_B as a global parameter. This produces similarly well-determined values for all parameters, but with slightly larger error intervals, as indicated for $\Delta pK_a^{(c-f)}$ in figure 4b, red line. Also a cross-correlation exists between $\Delta pK_a^{(c-f)}$ and f_B , as can be discerned in figure 5b. Thus, the results of strategy #3 indicate that it is possible to obtain accurate values for all affinity and enthalpy parameters, with the exception of ΔH_{int}° , even when the values are uncertain as long as at least one of the incompetent fractions is adjusted as a global parameter common to all experiments. In the case that both f_A and f_B are treated as locally adjustable parameter, no determination of binding parameters is possible (figure 4b, magenta line).

Strategy #5 has more pH values and data sets than strategy #3. However, based on the error projection analysis for the thermodynamic parameters (figure 4b and 4d, table 1), these two strategies behave very similarly regarding the determination of the parameter uncertainty.

3.3 Determination of the heat capacity change for intrinsic and proton binding

As shown in table 2, strategy #4 was introduced with temperature as an additional variable. With 12 ITC data sets, we used global analysis with the proton linkage model to determine values for 9 thermodynamic parameters: $\log_{10}(K_{int})$, ΔH_{int}° , $\Delta C_{p,int}$, pK_a^f , ΔH_p^f , $\Delta C_{p,p}^{(f)}$, $\Delta pK_a^{(c-f)}$, $\Delta \Delta H_p^{(c-f)}$, and $\Delta \Delta C_{p,p}^{(c-f)}$ (table 1). When f_A and f_B were fixed, the error projection maps showed that all 9 parameters yielded well-defined global minima at the 95 % confidence level. Similar results were also obtained either when f_A and f_B were both optimized globally (figure 6) or when optimizing f_A locally and f_B globally. When optimized as global parameters, the values of f_A and f_B were primarily determined between 0.1 and 0.2 (table 1), which is substantially better than any of the previous analyses.

We further investigated whether the robust determination of thermodynamic parameters by strategy #4 was due to the increased number and pH range of datasets or to the introduction of temperature, whereby $\Delta C_{p,\text{buffer}}$ is able to contribute to the analysis for experiments carried out at multiple temperatures. For this purpose, we devised strategy #5 to be comparable with strategy #4, but without variations in temperature. In the configuration with fixed f_A and f_B , strategy #5 was slightly better in determining $\Delta pK_a^{(c-f)}$ (figure 4). However, when f_A and f_B were optimized globally, strategy #4 allowed for better determination of the thermodynamic parameters than any other strategies. This includes the enthalpy parameters, as shown for $\Delta H_{\text{int}}^\circ$ in figure 4e. Therefore, including data with multiple temperatures could improve the determination of thermodynamic parameters when the exact values of molecular concentrations are not ensured. This advantage of strategy #4 is especially valuable for the studies using materials with limited stability as is often the case with proteins.

Table 3 shows the covariance matrix for strategy #4 with both incompetent fractions optimized as global parameters in the analysis. All the thermodynamic parameters show low correlation with each other, except $\Delta H_{\text{int}}^\circ$ and $\Delta C_{p,\text{int}}$, which is expected since the $\Delta C_{p,\text{int}}$ is the temperature derivative of $\Delta H_{\text{int}}^\circ$. Moreover, the incompetent fractions f_A and f_B have low correlation coefficients with most of the other parameters, suggesting that they have minimal impact on the uncertainty of these parameters. For example, when examining the cross-correlation between $\Delta pK_a^{(c-f)}$ and f_B , while fitting f_A as a local parameter and f_B as global parameter (figure 5), strategy #4 is the only one not showing a diagonal feature, demonstrating independence of the uncertainties for these two parameters under study. Therefore, strategy #4 seems to have sufficient information to determine most of the parameters, including the incompetent fractions.

4. Discussion

4.1 Comparison of the experimental strategies to study proton linkage to binding

The question of which pH values would be optimal for determining the relevant thermodynamic parameters of protonation linked binding has been previously examined theoretically using simulated data [7], with the focus on the pH dependence of the observed enthalpy and heat capacity change. In the present study, we extended the theoretical exploration of proton linkage binding to the context of global analysis, with the specific viewpoint of comparing the information content provided by different possible experimental strategies, and with regard to the impact of experimental imperfections in our knowledge of reactant concentrations.

For studying the thermodynamics of a system with linked equilibria such as protonation, many parameters are involved in the equations as shown in Theory. It is essential to understand the practical resolution limits of the parameters, and how the information in a global context arises from the individual experiments. We found that strategy #1 of conducting experiments at a single pH value in four different buffers with different ΔH_{buffer} allows the qualitative identification of proton linkage to binding; strategy #2 with experiments at two pH values and in 3 buffers at each pH is sufficient to measure K_{int} , pK_a^f and $\Delta pK_a^{(c-f)}$; an additional pH is required to determine the intrinsic binding and protonation enthalpies, $\Delta H_{\text{int}}^\circ$, $\Delta H_p^{(f)}$ and $\Delta \Delta H_p^{(c-f)}$ as shown by strategies #3 and #5; finally, the inclusion of data at multiple temperatures is required to possibly determine the heat capacity change values, $\Delta C_{p,\text{int}}$, $\Delta C_{p,p}^{(f)}$ and $\Delta \Delta C_{p,p}^{(c-f)}$, as illustrated in strategy #4. These observations are consistent with the predictions of the minimal experimental conditions necessary based on the underlying equations by Baker & Murphy [7], and demonstrate that

most of these parameters can be obtained reliably in the global analysis (table 2) despite experimental noise and concentration uncertainties (see below).

4.2 Resolution of the heat capacity change of protonation

Heat capacity changes of protonation are the most difficult parameters to determine precisely. Error projection maps of $\Delta C_{p,p}^{(f)}$ and $\Delta\Delta C_{p,p}^{(c-f)}$ yielded a global minimum at expected values and are well-formed to allow the determination of 95% confidence intervals, as shown in table 1 and figure 6. However, the resulting parameter uncertainties of $(-0.84, 0.79) \text{ kJ}\cdot\text{K}^{-1} \text{ mol}^{-1}$ and $(-0.18, 0.23) \text{ kJ}\cdot\text{K}^{-1} \cdot \text{mol}^{-1}$ for $\Delta C_{p,p}^{(f)}$ and $\Delta\Delta C_{p,p}^{(c-f)}$ respectively, are very large (fixed f_A and f_B). These error intervals indicate that no reliable determination of these two parameters can be made with useful precision. Although the analytical inspection of the equations governing the temperature dependence of the binding parameters (equations (7) and (8)) would suggest even two experiments at different temperatures to be sufficient to reveal heat capacity changes, the limited precision of experimental data and error amplification obscures this information: In order to resolve $\Delta C_{p,p}^{(f)}$ reliably, the determination of $\Delta H_p^{(f)}$ needs to have an error less than $\Delta C_{p,p}^{(f)} \times \Delta T$. In our simulated system, for a temperature difference such as 15 K (298K - 283K), $\Delta\Delta H_p^{(f)}$ is only $0.63 \text{ kJ}\cdot\text{mol}^{-1}$, whereas the 95 % confidence interval for $\Delta H_p^{(f)}$ is $(-39.1, -22.7) \text{ kJ}\cdot\text{mol}^{-1}$ even when f_A and f_B are fixed. Therefore, the precision in the $\Delta H_p^{(f)}$ measurement does not allow for the determination of $\Delta C_{p,p}^{(f)}$. In principle, of course, given a sufficiently large value for $\Delta C_{p,p}^{(f)}$ and low experimental noise, it could be determined using strategy #4. For example, in the study of a T-cell receptor/peptide-MHC complex formation by Armstrong and Baker [3], the $\Delta C_{p,p}^{(f)}$ was large enough to be determined from their experimental data by global analysis, with a value of $-(3.3 \pm 0.84) \text{ kJ}\cdot\text{K}^{-1} \cdot \text{mol}^{-1}$. On the other hand, for typical protein interactions involving the protonation of a histidine side chain, $\Delta C_{p,p}^{(f)}$ values range from -0.029 to $-0.63 \text{ kJ}\cdot\text{K}^{-1} \text{ mol}^{-1}$ [43; 44], and accordingly we have simulated our data with a value for $\Delta C_{p,p}^{(f)}$ of $-0.042 \text{ kJ}\cdot\text{K}^{-1} \cdot \text{mol}^{-1}$. In any case, assuming an error of $4.18 \text{ kJ}\cdot\text{mol}^{-1}$ in the value of $\Delta H_p^{(f)}$ resulted from ITC analysis, it is theoretically possible to determine the value of a $\Delta C_{p,p}^{(f)}$ with an absolute value higher than $0.42 \text{ kJ}\cdot\text{K}^{-1} \cdot \text{mol}^{-1}$, by using data under two temperatures with a 10 K difference. It is obvious that the determination of $\Delta\Delta C_{p,p}^{(c-f)}$ will pose even more difficulties, since it is composed of differences of protonation heat capacity changes between the free species and the complex (or equivalently, involves the analysis of the temperature-dependence of $\Delta\Delta H_p^{(c-f)}$, which always has an even larger error than $\Delta H_p^{(f)}$).

4.3 Global ITC analysis using the proton linkage model

Because of the complexity of the model and the unfavourable error amplifications for the determination of protonation enthalpies and heat capacity changes, the data analysis method and the assessment of uncertainties is critical. All of the thermodynamic parameters can be directly calculated from K_{obs} and $\Delta H_{\text{obs}}^{\circ}$ values derived from the independent analysis of individual data sets using the traditional single-site analysis methods. However, in such a treatment, the errors in K_{obs} and $\Delta H_{\text{obs}}^{\circ}$ from each data set add to the uncertainty of the final parameters of interest. This is especially critical for linear regression analysis, when for practical reasons the experimental parameters such as temperature, pH, and buffer ionization enthalpy can only be varied in a limited range. Global analysis increases the rigidity of the model by adding the relationships between K_{obs} and $\Delta H_{\text{obs}}^{\circ}$ as constraints into global non-linear regression optimization. For example, it embeds the linear relationship of equation (3) for identifying the presence of proton involvement in strategy #1.

Generally, global analysis of ITC data is becoming increasingly recognized as a powerful tool to optimally leverage the information from multiple experiments under different

configurations, for example, in the context of multi-site binding [23; 24; 26; 27; 28; 29]. For proton linkage studies, Armstrong and Baker [3] first demonstrated that experimental ITC data on proton-linked binding can be described well with direct global modelling. In the present work, we describe for the first time an analysis software with pre-implemented proton-linked binding models that provides an efficient and convenient approach for globally analyzing multiple ITC data sets with various local parameters.

Another factor of ITC data analysis that global analysis permits to accommodate naturally is the uncertainty in molecular concentrations for both binding partners. ITC titrations are highly sensitive to the precise knowledge of protein/ligand molar ratios. This constitutes a principal problem when working with interacting components where our knowledge of concentrations is not completely accurate (which is usually the case when working with biological macromolecules). In the present work we studied the impact of incompetent fractions on the resulting uncertainty of the thermodynamic parameters.

4.4 Effect of uncertainty in the active concentrations of molecular binding partners

To this end, incompetent fraction (f) has been introduced in the global model of the present work. Our simulated data were assigned 15% incompetent fractions of the nominal loading concentrations, with non-negativity constraints for f_A and f_B ($0 < f < 1$). (Alternatively, errors in active concentrations can be equivalently accounted for by correction factors, and the non-negativity constraint of f_A and f_B would correspond to an equivalent one-sided constraint in the concentration correction factor to be not less than 0.85.) For strategies #2–#5, when f_A and f_B were allowed to be optimized during the non-linear regression, either as global or local parameters, the resulting error intervals for the parameters were larger than those with f_A and f_B fixed at the correct values. In spite of the larger errors, most of the thermodynamic parameters still had well-formed minima that allowed determination of a 95 % confidence interval while globally adjusting f_A and f_B . For strategy #1, with globally optimized f_A and f_B , the lower limit of Δn_{H^+} can be obtained, showing a positive value as a clear indication for proton uptake coupled to binding.¹ Remarkably, with the more information rich data sets of strategy #4, the incompetent fractions can even be determined with some error, but with well-formed minima in the error projection map (table 1 and figure 6). Furthermore, one of the reactants may be allowed to assume different incompetent fractions in the individual titrations (f as local parameter), which can be particularly important in the planning of experiments with proteins that may have different stability at different pH values.

The possibility to consider imperfections in concentrations by the incompetent fraction provides great flexibility for parameter determination in global modelling. Nevertheless, the utilization of this parameter needs practical consideration in data analysis. Generally, for a bimolecular interaction, the incorporation of incompetent fractions for both components should be avoided due to their high correlation with binding parameters [26]. As discussed by Houtman *et al.* in [26], the choice of either fixing or optimizing incompetent fraction for a specific molecule depends on the experimental conditions, the stability of the molecules and the their preparation. An additional opportunity for SEDPHAT analysis implemented more recently, and not examined in the present paper, is the introduction of parameter constraints. While incompetent fractions can assume values between 0 and 1, different

¹Considering this problem in the framework of the traditional single data set analysis accounting for concentration errors by ‘n-values’, one would be required to assign *ad hoc* correct concentrations to one of the binding partners, and, in the case of 1:1 reactions, absorb n-values deviating from 1.0 into concentration errors of the other binding partner. This would be equivalent in the SEDPHAT global model with globally fixing one incompetent fraction to zero, and allowing the other to be refined as a local parameter. While this can be employed in the SEDPHAT analysis, it may not always correspond to our knowledge of precise active reactant concentrations. In practice, we may rather assign a possible range, which can be entered as constraints to the concentration parameters in the global SEDPHAT analysis.

constraints may be used, especially when they are phrased in terms of concentration errors. Such constraints limit the additional flexibility introduced by the additional parameters to narrow parameter regions, and may prevent them from increasing the confidence intervals of the parameters of interest. In practice, we recommend to fully take the advantage of any prior knowledge for the macromolecular concentrations and their plausible range and to be conservative when locally optimizing the incompetent fractions.

4.5 Estimating parameter uncertainties in the global analysis

A crucial difference between protonation analysis by global modelling and the hierarchical approach based on a first stage of individual titration analyses appears in the opportunities of error analysis. In contrast to the individual analysis approach, the global modelling approach allows one to assume any fixed value of a thermodynamic parameter and then calculate how consistent this value is with the raw experimental data (by allowing all other parameters to be adjusted and compensate). This allows one to construct a map of global best-fit χ_r^2 values as a function of the value of the fixed parameter. These error projection maps are rigorous tools to analyze the information content of the raw data, and can be quantitatively evaluated with F-statistics [38; 40; 45].

A second, alternative approach to obtain parameter uncertainties is the Monte-Carlo method. However, in our experience it performs best for simple error surfaces. For complicated error surfaces, the difficulty of the Monte-Carlo approach lies in the reliable optimization of the global fit at each iteration. In contrast to the error projection maps, incomplete optimization is difficult to catch in the Monte-Carlo method and will cause incorrect parameter distributions. This would lead to too small confidence limits, consistent with our findings in the present study (not shown). For example, when using strategy #4 with fixed f_A and f_B , the 95 % confidence interval for $\Delta H_p^{(f)}$ is $(-32.2, -26.8)$ and $(-39.1, -22.7)$ kJ·mol⁻¹ from Monte Carlo and error projection analysis respectively.

A third approach is the calculation of a covariance matrix. Although it is well-known to provide unreliable parameter error estimates for non-linear regression, it can be used to calculate the cross-correlation matrix, which is highly useful to describe the rigidity of the model. The correlation coefficient matrix provides direct information of the local dependence of a pair of best-fit parameter values on each other (for example, at the overall best-fit). As shown in the Supplement tables, in the configuration of floating f_A and f_B as global parameters, strategy #1 showed complete correlation for all the parameters, suggesting that the data do not have sufficient information to determine them. Strategy #3 and #5 showed very similar correlation pattern, illustrating that more pH conditions do not improve the information content of the data. This information is consistent with the error intervals of the parameters resulting from the error projection maps using these two strategies. Interestingly, compared to other strategies, strategy #4 showed a lower correlation for some of the parameters, such as ΔH_{int}^o and $\Delta pK_a^{(c-f)}$, as well as f_A or f_B with $\Delta pK_a^{(c-f)}$. The latter pairs yielded a correlation coefficient of 0.23, much lower than those in strategy #2 (0.93), #3 (0.82) and #5 (0.81). This strongly supports the conclusion that strategy #4 provides sufficient information for resolving the thermodynamic parameters even in the presence of high uncertainty in molecular concentrations. This possibly is due to the inclusion of another dimension of information, temperature in the current case, which adds constraints to the error surfaces and provides better determination of the parameters. Armstrong *et al.* [3] shows high correlation between ΔH_{int}^o , $\Delta H_p^{(f)}$ and $\Delta \Delta H_p^{(c-f)}$, which was not observed in our analysis. This might be related to more data sets with different buffer enthalpies for the third pH in our analysis and lower noise in our simulated data.

4.6 Conclusions

The present work has illustrated the advantages of the global analysis for the study of proton-linked binding. We have outlined experimental strategies for the reliable determination of intrinsic binding parameters and proton binding parameters, and demonstrated that concentration uncertainties can be accounted for naturally in the context of the global modelling. In order to apply the result of the present study to the planning of an experiment series for a specific system, simulations for the designed experimental conditions can help to evaluate whether the parameters of interest can be resolved reliably. The required simulation tools are part of SEDPHAT functions.

The protonation model currently implemented is restricted to a single linked protonation site. Future extension to multiple protonation sites would be useful, especially when the pH range to be studied does not allow restriction of the analysis to a single protonation event [8]. Another possibility for extension of the analysis presented here is to take advantage of the global multi-method modelling capabilities of SEDPHAT to incorporate data from other techniques that determine K_{obs} with higher confidence, which might allow reducing parameter uncertainties in the enthalpy values, in turn.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Lawrence E. Samelson for his encouragement and critical reading of the manuscript. This work was supported by the Intramural Research Programs of NCI and NIBIB, National Institutes of Health.

References

1. Mason AC, Jensen JH. *Proteins*. 2008; 71:81–91. [PubMed: 17932920]
2. Doyle ML, Louie G, Dal Monte PR, Sokoloski TD. *Methods in enzymology*. 1995; 259:183–194. [PubMed: 8538453]
3. Armstrong KM, Baker BM. *Biophysical Journal*. 2007; 93:597–609. [PubMed: 17449678]
4. Horn JR, Ramaswamy S, Murphy KP. *Journal of Molecular Biology*. 2003; 331:497–508. [PubMed: 12888355]
5. Fisher HF, Singh N. *Methods Enzymol*. 1995; 259:194–221. [PubMed: 8538455]
6. Garcia-Fuentes L, Reche P, Lopez-Mayorga O, Santi DV, Gonzalez-Pacanowska D, Baron C. *Eur J Biochem*. 1995; 232:641–645. [PubMed: 7556218]
7. Baker BM, Murphy KP. *Biophysical journal*. 1996; 71:2049–2055. [PubMed: 8889179]
8. Bradshaw JM, Waksman G. *Biochemistry*. 1998; 37:15400–15407. [PubMed: 9799501]
9. Connelly PR, Varadarajan R, Sturtevant JM, Richards FM. *Biochemistry*. 1990; 29:6108–6114. [PubMed: 2383573]
10. Murphy KP, Freire E. *Advances in Protein Chemistry*. 1992; 43:313–361. [PubMed: 1442323]
11. Spolar RS, Livingstone JR, Record MT. *Biochemistry*. 1992; 31:3947–3955. [PubMed: 1567847]
12. Spolar RS, Record MT. *Science (New York, N.Y.)*. 1994; 263:777–784.
13. Xie D, Freire E. *Proteins*. 1994; 19:291–301. [PubMed: 7984625]
14. Xie D, Gulnik S, Collins L, Gustchina E, Suvorov L, Erickson JW. *Biochemistry*. 1997; 36:16166–16172. [PubMed: 9405050]
15. Gomez J, Freire E. *J Mol Biol*. 1995; 252:337–350. [PubMed: 7563055]
16. Czodrowski P, Sottriffer Ca, Klebe G. *Journal of molecular biology*. 2007; 367:1347–1356. [PubMed: 17316681]

17. Dullweber F, Stubbs MT, Musil D, Stürzebecher J, Klebe G. *Journal of molecular biology*. 2001; 313:593–614. [PubMed: 11676542]
18. Brik A, Wong CH. *Organic & biomolecular chemistry*. 2003; 1:5–14. [PubMed: 12929379]
19. Lin JH, Lu AYH. *Pharmacological reviews*. 1997; 49:403–449. [PubMed: 9443165]
20. Ting A, McGuire R, Johnson A, Green S. *Journal of chemical information and computer sciences*. 2000; 40:347–353. [PubMed: 10761138]
21. Holdgate GA, Ward WHJ. 2005:10.
22. Meek TD, Dayton BD, Metcalf BW, Dreyer GB, Strickler JE, Gorniak JG, Rosenberg M, Moore ML, Magaard VW, Debouck C. *Proceedings of the National Academy of Sciences of the United States of America*. 1989; 86:1841–1845. [PubMed: 2648384]
23. Barda-Saad M, Shirasu N, Pauker MH, Hassan N, Perl O, Balbo A, Yamaguchi H, Houtman JCD, Appella E, Schuck P, Samelson LE. *The EMBO journal*. 2010; 29:2315–2328. [PubMed: 20562827]
24. Brown A. *International Journal of Molecular Sciences*. 2009; 10:3457–3477. [PubMed: 20111687]
25. Henzl MT, Larson JD, Agah S. *Analytical Biochemistry*. 2003; 319:216–233. [PubMed: 12871715]
26. Houtman JCD, Brown PH, Bowden B, Yamaguchi H, Appella E, Samelson LE, Schuck P. *Protein Science*. 2007; 16:30–42. [PubMed: 17192587]
27. Houtman JCD, Yamaguchi H, Barda-Saad M, Braiman A, Bowden B, Appella E, Schuck P, Samelson LE. *Nature structural & molecular biology*. 2006; 13:798–805.
28. Muscroft-Taylor AC, Soares da Costa TP, Gerrard Ja. *Biochimie*. 2010; 92:254–262. [PubMed: 20025926]
29. Pellizzaro ML, McGhee AM, Renton LC, Nix MG, Fisher J, Turnbull WB, Wilson AJ. *Chemistry (Weinheim an der Bergstrasse, Germany)*. 2011; 17:14508–14517.
30. Johnston JW, Coussens NP, Allen S, Houtman JCD, Turner KH, Zaleski A, Ramaswamy S, Gibson BW, Apicella Ma. *The Journal of biological chemistry*. 2008; 283:855–865. [PubMed: 17947229]
31. Moncrieffe MC, Grossmann JG, Gay NJ. *The Journal of biological chemistry*. 2008; 283:33447–33454. [PubMed: 18829464]
32. Zhi L, Mans J, Paskow MJ, Brown PH, Schuck P, Jonji S, Natarajan K, Margulies DH. *Biochemistry*. 2010; 49:2443–2453. [PubMed: 20166740]
33. Duff Michael JR, Grubbs J, Howell EE. *J Vis Exp*. 2011:e2796.
34. Tang LY, Yamashita M, Coussens NP, Tang Y, Wang X, Li C, Deng CX, Cheng SY, Zhang YE. *The EMBO journal*. 2011; 30:4777–4789. [PubMed: 22045334]
35. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. *Protein science: a publication of the Protein Society*. 1995; 4:2411–2423. [PubMed: 8563639]
36. Goldberg RN, Kishore N, Lennen RM. *Journal of Physical and Chemical Reference Data*. 2002; 31:231–370.
37. ITC Data Analysis in Origin® Tutorial Guide. MicroCal Inc;
38. Johnson ML. *Analytical biochemistry*. 1992; 206:215–225. [PubMed: 1443589]
39. Johnson, ML.; Straume, M. *Modern Analytical Ultracentrifugation*. 1994. p. 37-65.
40. Bevington, PR.; Robinson, DK. *Data reduction and error analysis for the physical sciences*. McGraw-Hill; 1969.
41. <https://sedfitsedphat.nibib.nih.gov/software/default.aspx>.
42. http://www.analyticalultracentrifugation.com/sedphat/isothermal_titration_calorimetry.htm.
43. Baker BM, Murphy KP. *Journal of molecular biology*. 1997; 268:557–569. [PubMed: 9159490]
44. Bastiansen O. *Berichte der Bunsengesellschaft für physikalische Chemie*. 1977; 81:540.
45. Heerklotz H, Tsalaloukas AD, Keller S. *Nature Protocols*. 2009; 4:686–697.

Highlights

- We demonstrate use of global analysis of ITC data for proton-linked binding study.
- Various experimental strategies are evaluated for their information content.
- Data at multiple temperatures might improve the precision of binding parameters.
- The methods for detailed error analysis of parameter uncertainties are discussed.
- By global modeling uncertainty in molecular concentrations might be accounted for.

FIGURE 1.
Scheme of binding with a single proton-linked event.

FIGURE 2.

Global fit of 12 ITC data sets by SEDPHAT. Shown is a global fit to the 12 data sets of strategy #4 using the Hetero-Association Global Proton Linkage Analysis model of SEDPHAT. Strategy #4 spans 6 pH values (6.0, 6.5, 7.0, 7.5, 8.0, 8.5) and 3 temperatures (283 K, 298 K, 310 K) with multiple buffers of different ionization enthalpies. As shown from top left to bottom right: pH 6.0 Cacodylate 298 K, pH 6.5 Bis-Tris 283 K, pH 6.5 Bis-Tris 298 K, pH 6.5 Bis-Tris 310 K, pH 7.0 Imidazole 298 K, pH 7.5 PBS 283 K, pH 7.5 PBS 298 K, pH 7.5 PBS 310 K, pH 8.0 Tris 298 K, pH 8.5 HEPES 283 K, pH 8.5 HEPES 298 K, and pH 8.5 HEPES 310 K. Solid lines show the fit to individual data points represented by red dots. In this SEDPHAT screenshot, for each titration the ordinate is normalized heat in kcal mol^{-1} and the abscissa is the molar ratio of B and A in the cell. Residuals of the data points for each titration are plotted below the binding isotherms.

FIGURE 3.

Error projection maps for Δn_{H+} in the global model of equation (3) applied to strategy #1, with fixed incompetent fractions (black), global incompetent fractions (blue), local incompetent A and global incompetent B (green), and local incompetent A and B (magenta). The dotted lines represent the χ_r^2 values corresponding to 95% confidence levels by F-statistics.

FIGURE 4.

(a–d) Error projection maps of ΔpK_a for strategy #2 (a), #3 (b), #4(c) and #5 (d) with fixing or optimizing for incompetent fractions: fixed f_A and f_B (blue), globally optimizing f_A and f_B (black), locally fitting f_A and globally fitting f_B (red), and locally optimizing f_A and f_B (magenta). (e) Overlay of error projection maps of ΔH°_{int} in analyses when f_A and f_B were optimized as global parameters, strategy #2 (magenta), strategy #3 (green), strategy #4 (red), strategy #5 (black), guideline for normalized global $\chi^2_{r,P=0.95}$ at 95 % confidence level (gray). For all error projection maps, in order to simplify the comparison of the different error projection maps $\chi^2_r(x)$ in their correspondence to different confidence levels $\chi^2_{r,P=0.95}$ and different best-fit value $\chi^2_{r,0}$, their presentation here was normalized according to $\chi^2_{r,norm} = (\chi^2_r(x) - \chi^2_{r,0}) / (\chi^2_{r,P=0.95} - \chi^2_{r,0})$. In this presentation, for all curves the best-fit value is at the ordinate value of 0, and the limits of the 95 % confidence interval are reached at the ordinate value of 1.0.

FIGURE 5.

Two-dimensional error surface projections maps for examining the cross-correlation between two global parameters, ΔpK_a and f_B of strategies #2 (a), #3 (b) and #4(c) when f_A was floated as a local parameter for each experiment. In these error surface projection maps, values of two parameters were fixed while all other parameter values were allowed to compensate for the constraint and to obtain the best possible fit for the given parameter pair. The colours indicate the normalized global chi square values for combinations of f_B and ΔpK_a in the optimization. Colour bars on the right show the scale of the colour.

FIGURE 6.

Error projection maps for all 9 thermodynamic parameters determined by strategy #4 when both f_A and f_B were floated globally in the analysis. Nine panels show error projection maps of $\log_{10}(K_{int})$, ΔH_{int}^f , $\Delta C_{p, int}$, pK_a^f , ΔH_p^f , $\Delta C_{p, p}^f$, $\Delta pK_a^{(c-f)}$, $\Delta \Delta H_p^{(c-f)}$, and $\Delta \Delta C_{p, p}^{(c-f)}$.

Solid gray lines mark the normalized critical $\chi^2_{r, P=0.95}$. These results indicate that it is possible to quantitatively characterize the thermodynamics of proton linkage, despite considerable uncertainty in the molecular concentrations.

TABLE 1

Values of the thermodynamic parameters for the simulated ITC data, and 95 % confidence intervals in the different analysis strategies.

parameter	simulated value	strategy #2			strategy #3			strategy #4		
		fixed f_A, f_B	global f_A, f_B	local f_A, f_B	fixed f_A, f_B	global f_A, f_B	local f_A, f_B	fixed f_A, f_B	global f_A, f_B	local f_A, f_B
$\log_{10} (K_{int} \times M)$	7.5	(7.41, 7.56)	(7.40, 7.60)	(7.39, 7.61)	(7.44, 7.55)	(7.40, 7.60)	(7.39, 7.63)	(7.46, 7.56)	(7.44, 7.59)	(7.41, 7.61)
$pK_a^{(f)}$	7	(6.84, 7.22)	(6.82, 7.25)	(6.78, 7.30)	(6.83, 7.20)	(6.80, 7.20)	(6.75, 7.27)	(6.74, 7.22)	(6.72, 7.24)	(6.65, 7.25)
$\Delta pK_a^{(c-f)}$	0.5	(0.46, 0.58)	(0.40, 0.82)	(0.38, 0.86)	(0.45, 0.57)	(0.41, 0.63)	(0.37, 0.69)	(0.44, 0.59)	(0.43, 0.61)	(0.41, 0.65)
$\Delta H_{int}^{(f)} / (kJ \cdot mol^{-1})$	-62.8	ND	ND	ND	(-64.4, -61.5)	(-84.1, -52.3)	(-89.1, -51.5)	(-64.0, -61.5)	(-69.0, -56.5)	(-71.1, -54.8)
$\Delta H_p^{(f)} / (kJ \cdot mol^{-1})$	-29.3	ND	ND	ND	(-38.3, -20.7)	(-39.0, -20.1)	(-40.1, -17.7)	(-39.1, -22.7)	(-39.6, -22.5)	(-42.0, -17.4)
$\Delta \Delta H_p^{(c-f)} / (kJ \cdot mol^{-1})$	-4.18	ND	ND	ND	(-6.11, -1.46)	(-6.65, -1.26)	(-7.41, -0.54)	(-6.11, -2.13)	(-6.32, -1.92)	(-6.95, -1.34)
$\Delta C_{p,int} / (kJ \cdot K^{-1} \cdot mol^{-1})$	-2.09	ND	ND	ND	ND	ND	ND	(-2.26, -1.97)	(-2.34, -1.84)	(-2.55, -1.72)
$\Delta C_{p,p}^{(f)} / (kJ \cdot K^{-1} \cdot mol^{-1})$	-0.042	ND	ND	ND	ND	ND	ND	(-0.84, 0.79)	(-0.92, 0.88)	(-1.09, 1.17)
$\Delta \Delta C_{p,p}^{(c-f)} / (kJ \cdot K^{-1} \cdot mol^{-1})$	-0.021	ND	ND	ND	ND	ND	ND	(-0.18, 0.23)	(-0.18, 0.24)	(-0.23, 0.31)
f_A	0.15	ND	(NF*, 0.48)	ND	ND	(NF*, 0.37)	ND	ND	(0.063, 0.23)	ND
f_B	0.15	ND	(NF*, 0.48)	(NF*, 0.51)	ND	(NF*, 0.37)	(NF*, 0.40)	ND	(0.063, 0.23)	(0.03, 0.25)

* ND: Not Determined.

** NF: Not Found.

TABLE 2
Strategies for the choice of experimental conditions in the global proton linkage analysis.

	pH 6.0	pH 6.5	pH 7.0	pH 7.5	pH 8.0	pH 8.5
Strategy1 (1 pH, 4 buffers)				PBS, 298K		
				HEPES, 298K		
				Imidazole,298K		
				Tris, 298K		
Strategy2 (2 pHs, 6 buffers)		Cacodylate,298K		PBS, 298K		
		MES, 298K		HEPES, 298K		
		Bis-tris, 298K		Tris, 298K		
Strategy3 (3 pHs, 6 buffers)		Cacodylate,298K		PBS, 298K		CHES, 298K
		MES, 298K		HEPES, 298K		HEPES,298K
		Bis-tris, 298K		Tris, 298K		Tris, 298K
Strategy4 (6 pHs, 6 buffers, 3 Ts)		Bis-tris, 283K		PBS, 283K		HEPES,283K
		Bis-tris, 298K	Imidazole,298K	PBS, 298K	Tris,298K	HEPES,298K
		Bis-tris, 310K		PBS, 310K		HEPES,310K
Strategy5 (6 pHs, 8 buffers)		Cacodylate,298K		PBS, 298K		CHES, 298K
		MES, 298K	Imidazole,298K	HEPES, 298K	Tris, 298K	HEPES,298K
		Bis-tris, 298K		Tris, 298K		Tris, 298K

TABLE 3
Covariance matrix of the parameters in the global analysis for the data in Strategy 4 with f_A and f_B floated as global parameters.

	$\log_{10}(K_{int})$	ΔH_{int}^0 /(kJ·mol ⁻¹)	$\Delta C_{p,int}^0$ /(kJ·K ⁻¹ ·mol ⁻¹)	$pK_a^{(f)}$	$\Delta pK_a^{(e,f)}$	ΔH_p^0 /(kJ·mol ⁻¹)	$\Delta \Delta H_p^{(e,f)}$ /(kJ·mol ⁻¹)	$\Delta C_{p,p}^{(f)}$ /(kJ·K ⁻¹ ·mol ⁻¹)	$\Delta \Delta C_{p,p}^{(e,f)}$ /(kJ·K ⁻¹ ·mol ⁻¹)	f_A	f_B
$\log K_{int}$	1.00										
ΔH_{int}^0 /(kJ·mol ⁻¹)	-0.70	1.00									
$\Delta C_{p,int}^0$ /(kJ·K ⁻¹ ·mol ⁻¹)	-0.67	0.91	1.00								
$pK_a^{(f)}$	-0.03	-0.19	-0.18	1.00							
$\Delta pK_a^{(e,f)}$	-0.01	-0.27	-0.24	-0.57	1.00						
ΔH_p^0 /(kJ·mol ⁻¹)	-0.15	-0.24	-0.09	0.59	-0.06	1.00					
$\Delta \Delta H_p^{(e,f)}$ /(kJ·mol ⁻¹)	-0.17	0.12	0.16	-0.59	0.43	-0.42	1.00				
$\Delta C_{p,p}^{(f)}$ /(kJ·K ⁻¹ ·mol ⁻¹)	-0.02	-0.22	-0.25	0.53	-0.17	0.57	-0.29	1.00			
$\Delta \Delta C_{p,p}^{(e,f)}$ /(kJ·K ⁻¹ ·mol ⁻¹)	0.07	-0.01	-0.21	-0.33	0.28	-0.41	0.35	-0.56	1.00		
f_A	0.78	-0.97	-0.92	0.19	0.23	0.13	-0.22	0.17	0.04	1.00	
f_B	0.78	-0.97	-0.93	0.19	0.23	0.13	-0.22	0.17	0.04	1.00	1.00