J Comput Aided Mol Des (2015) 29:867–883
DOI 10.1007/s10822-015-9867-y

Thermodynamics of protein–ligand interactions as a reference for computational analysis: how to assess accuracy, reliability and relevance of experimental data

Stefan G. Krimmer[1] · Gerhard Klebe[1]

Received: 30 March 2015 / Accepted: 5 September 2015 / Published online: 16 September 2015
© Springer International Publishing Switzerland 2015

Abstract For a conscientious interpretation of thermodynamic parameters (Gibbs free energy, enthalpy and entropy) obtained by isothermal titration calorimetry (ITC), it is necessary to first evaluate the experimental setup and conditions at which the data were measured. The data quality must be assessed and the precision and accuracy of the measured parameters must be estimated. This information provides the basis at which level discussion of the data is appropriate, and allows insight into the significance of comparisons with other data. The aim of this article is to provide the reader with basic understanding of the ITC technique and the experimental practices commonly applied, in order to foster an appreciation for how much measured thermodynamic parameters can deviate from ideal, error-free values. Particular attention is paid to the shape of the recorded isotherm (*c*-value), the influence of the applied buffer used for the reaction (protonation reactions, pH), the chosen experimental settings (temperature), impurities of protein and ligand, sources of systematic errors (solution concentration, solution activity, and device calibration) and to the applied analysis software. Furthermore, we comment on enthalpy–entropy compensation, heat capacities and van't Hoff enthalpies.

Keywords Isothermal titration calorimetry · Data quality and accuracy · Good measuring practice · Data interpretation and correlation · Heat of ionization · van't Hoff evaluation

Electronic supplementary material The online version of this article (doi:10.1007/s10822-015-9867-y) contains supplementary material, which is available to authorized users.

✉ Gerhard Klebe
klebe@staff.uni-marburg.de

Stefan G. Krimmer
krimmer@staff.uni-marburg.de

[1] Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

Introduction: basic thermodynamic relationships

Many computational approaches make use of thermodynamic properties. Most important among all is the binding affinity, usually the target property used for scoring and ranking solutions generated in computational docking simulations and key to all virtual screening applications [1]. However, what kind of a property is "affinity" and how do we obtain experimental information to characterize this property? How good is the quality of the experimental data usually consulted to describe the affinity of a compound, how is its precision and accuracy, particularly if such data are intended for further usage in the development of computational models [2–5]? The aim of this article is to provide foundations necessary to understand which experimental protocols are commonly applied to perform an isothermal titration calorimetry (ITC) measurement and how critically different setups can influence the recorded binding parameters.

The affinity of a ligand binding to its target protein is described by the change in the Gibbs free energy of the system before and after the binding event. Only changes in the Gibbs free energy are detectable, whereas absolute values for individual states cannot be measured. Once equilibrium is attained for the reaction between protein '*P*' and ligand '*L*' forming the protein–ligand complex '*PL*', $P + L \rightleftharpoons PL$, the association constant K_a (L mol^{-1} or M^{-1}) describes the ratio between the concentration of the protein–ligand complex [*PL*] and the product of the free protein [*P*] and free ligand [*L*] concentrations:

$K_a = [PL]/[P][L]$. In contrary, the dissociation equilibrium constant K_d (mol L^{-1} or M) is the inverse of the association constant K_a, i.e. $K_d = [P][L]/[PL]$.

$$\Delta G^\circ = -RT \ln K_a \quad (1)$$

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (2)$$

As described in Eq. 1, at equilibrium the Gibbs free energy of binding ΔG° (kJ mol^{-1}) is logarithmically related to the association constant K_a, weighted by the ideal gas constant R (8.314 J mol^{-1} K^{-1}) and the absolute temperature T (K). It consists of two components (Eq. 2): a change in enthalpy ΔH° (kJ mol^{-1}) and a change in entropy ΔS° (kJ mol^{-1}), the latter weighted by the absolute temperature. The change in enthalpy describes the amount of heat released (exotherm, negative ΔH°) or absorbed (endotherm, positive ΔH°) as bonds and intermolecular contacts (e.g. hydrogen bonds, electrostatic interactions, van der Waals contacts) are established and broken between protein, ligand, water and other buffer components resulting in the formation of a protein–ligand complex. The difference in entropy describes the change in ordering parameters and the distribution of the system over multiple accessible states. A positive ΔS° describes an increase in entropy and thus an increase in disorder and in the number of accessible states. The change in entropy is not only related to conformational changes of the ligand and the protein, but for instance also to the water molecules which play a major role in the binding process. A classic example is the displacement of water molecules from apolar surfaces and the related increase in entropy, which is considered to be the driving force of association in the so-called hydrophobic effect [6]. It should be mentioned that ΔG°, ΔH° and ΔS° are all state functions – their values depend only on the two thermodynamic equilibrium states referred to, and not on the route by which these states are accessed.

The superscript '°' (pronounced "naught") is attached to indicate that the binding free energy value refers to its standard state. However, this sign is frequently omitted. The necessity for referring to a standard state is to achieve comparability between measurements on the same scale. At standard state, the binding free energies are described for the conversion of 1 mole protein and 1 mole ligand to 1 mole of protein–ligand complex, in a hypothetical ideal solution (infinitely diluted), with a unit activity coefficient at a constant pressure of $p^\circ = 10^5$ Pa. The temperature is not part of the standard state and therefore has to be specified. While K_a and ΔH° are determined experimentally in an ITC experiment (see below), ΔG° is calculated according to Eq. 1. This requires the use of the natural logarithm of K_a, which makes it necessary to convert K_a to a unitless value. To achieve this, the standard concentration c° is used, which is by convention 1 M. Depending on the reference concentration scale (e.g. M, mM, μM), the magnitude of the calculated ΔG changes. For example, for a K_a of 10^6 M^{-1} and a reference concentration of 1 M, the result for ΔG is $-13.8 \times RT$, calculated from $-RT \times \ln(10^6\ M^{-1} \times 1\ M)$. On the other hand, for the same K_a, applying a reference concentration of 1 mM, ΔG results in $-6.9 \times RT$, calculated from $-RT \times \ln(10^3\ mM^{-1} \times 1\ mM)$. Consequently, it is necessary to specify the reference concentration applied, which in the case of the standard state is $c^\circ = 1$ M. With the information of ΔG°, the standard entropy change ΔS° is calculated according to Eq. 2.

At this point, the first approximation must already be regarded. In principle, we try to describe the number of particles actively involved in the considered equilibrium by the "concentrations". However, this is only correct if we are dealing with so-called ideal solutions, which correspond to infinite dilutions. Real solutions deviate in their actual concentrations and instead we would have to consider "activities" which usually correspond to a smaller number of particles compared to the theoretically achievable concentrations [7]. In biological systems only a small number of validation studies have been performed to estimate how "ideal" the investigated solutions really are. It has been suggested to perform ITC investigations at different concentrations to estimate the extent to which the measured properties are affected. In one reported study the binding of 2′-cytidine monophosphate to ribonuclease was investigated [8]. Deviations in the binding constants as large as 40 % were reported by increasing the used protein concentration from diluted 0.0145 mM to more concentrated 0.65 mM, which corresponds to a 44-times higher protein concentration. Further down in this article, we will describe another measurement from our own research giving an idea by how much the thermodynamic signature can vary on an absolute scale with concentration. In practical application, an appropriate amount of protein is dissolved in a buffer by the experimentalist and the resulting thermodynamic properties are referenced to this "concentration". As long as data are compared relative to each other across a series using unchanged protein "concentrations" (better: "activities"), data interpretation will unlikely be strongly affected. However, if data are taken from different proteins and measured at largely deviating concentrations, analysis on an absolute scale can easily become quite problematic.

Which energetic contributions of the protein–ligand binding reaction are measured by ITC?

Isothermal titration calorimetry (ITC) allows the determination of K_a, ΔG°, ΔH°, ΔS° and n of a binding event in a single experiment at one given measurement temperature.

This is done without any need for labeling by simply measuring heat changes related to a reaction [8–12]. By performing the measurement at varying temperatures, the heat capacity change ΔC_p can also be obtained. This review will focus on the thermodynamics measured by ITC as a source of experimental information about protein–ligand interactions, assuming a single-site 1:1 binding event without major conformational changes of the target protein. It is important to note that ITC records the entire binding event, starting with the separately solvated binding partners (ligand and protein), and detects any alteration giving rise to a heat signal until the formation of the final complex. The process is affected by all changes involving the surrounding buffer, conformational transitions and, importantly enough, modulations of the solvation structure. The picture produced becomes quite complex as many steps on the molecular level can compensate in their thermodynamic signature and thus make it extremely difficult to factorize the ITC results into the discrete contributions of each separate interaction formed between a ligand and its target protein. Correlation of thermodynamic parameters with structural features [13], for example those obtained by X-ray crystallography, must therefore be performed very carefully. This often requires the interpretation within a narrow ligand series involving small variations, for instance the exchange of a moiety, a functional group or even only the addition or removal of a single methyl group [14]. The small variations between two ligands can then be attributed to the observed changes in their thermodynamic profiles. Conversely, unchanged thermodynamic signatures of two closely related ligands do not necessary mean that the binding modes of these ligands are identical, as seen in a series of thrombin inhibitors [15]. The mutual compensation of thermodynamic effects can result in identical thermodynamic signatures with simultaneous changes in the ligands' binding modes. Therefore, a structural inspection is essential. Considering the classification of binding as enthalpy or entropy-driven binding, the selection of enthalpically favored lead structures for subsequent affinity optimization has been suggested as desirable [16–18]. However, an unambiguous classification with respect to such profiles is rather problematic in light of the large impact the rearrangement of the residual water solvation pattern has on the thermodynamic signature, e.g. for the binding of low affinity fragments [14, 19, 20]. Introduction of only small structural modifications can lead to major changes in the fragments' thermodynamic binding signatures. In a recent review [21], it is even concluded that thermodynamically guided compound optimization is not feasible in most cases due to the complexity of the parameters enthalpy and entropy and the difficulties with their assignment to specific interactions.

How does an ITC measurement work and how does ITC raw data look like?

The general principle of an ITC measurement is that two reaction partners, for instance a protein and its ligand, are mixed with each other in a step-wise fashion, and the heat signal associated with the binding event is recorded. Figure 1a displays a schematic representation of an ITC device.

The instrument consists of a sample cell and a reference cell, both in a jacket which is kept below measurement temperature. Both cells are maintained at the constant measurement temperature by applying a thermal heating device using very sensitive and highly regulated electric heating control units. The reference cell contains a solvent of a similar heat capacity to the one used in the sample cell (usually water or buffer). For the measurement, the protein solution is released into the sample cell and the ligand solution is gradually added via a rotating syringe, which also functions as a stirring rod. Typically, about 10–30 injections of the ligand solution are added into the sample cell until all active sites of the protein are saturated. The change in the heat signal in the sample cell resulting from the complex formation is quantified by analyzing the difference in thermal power ($\mu J\ s^{-1}$) necessary to keep the sample cell at the same temperature as the reference cell. For an exothermic reaction in the sample cell, the required thermal power is reduced compared to the reference cell, whereas for endothermic reactions the required thermal power increases. These differences in power over time are recorded and evaluated to quantify the event in the sample cell. Differences in heat as low as 0.1 μJ are detectable with the most sensitive ITC devices. In the ITC thermogram (Fig. 1b), an exothermic binding reaction between protein and ligand is indicated by a series of "downward" peaks, whereas an endothermic reaction produces "upward" peaks. For the first injections of the measurement, the protein in the sample cell has a sufficient amount of unoccupied binding sites so that all injected ligand molecules can find a vacant binding pocket. This results in equally large heat signals. With increasing amount of injected ligand, the concentration of uncomplexed protein molecules becomes progressively smaller, allowing fewer ligand molecules to bind, which results in a gradual decrease of the heat signal. Due to chemical equilibrium conditions, further added free ligand molecules start to displace already bound ligand molecules from the protein. After several further ligand injections, all protein molecules are saturated by ligand molecules and under the regime of equilibrium, an increasing concentration of uncomplexed ligand molecules builds up. At the end of the titration, well beyond the 1:1 binding stoichiometry, only

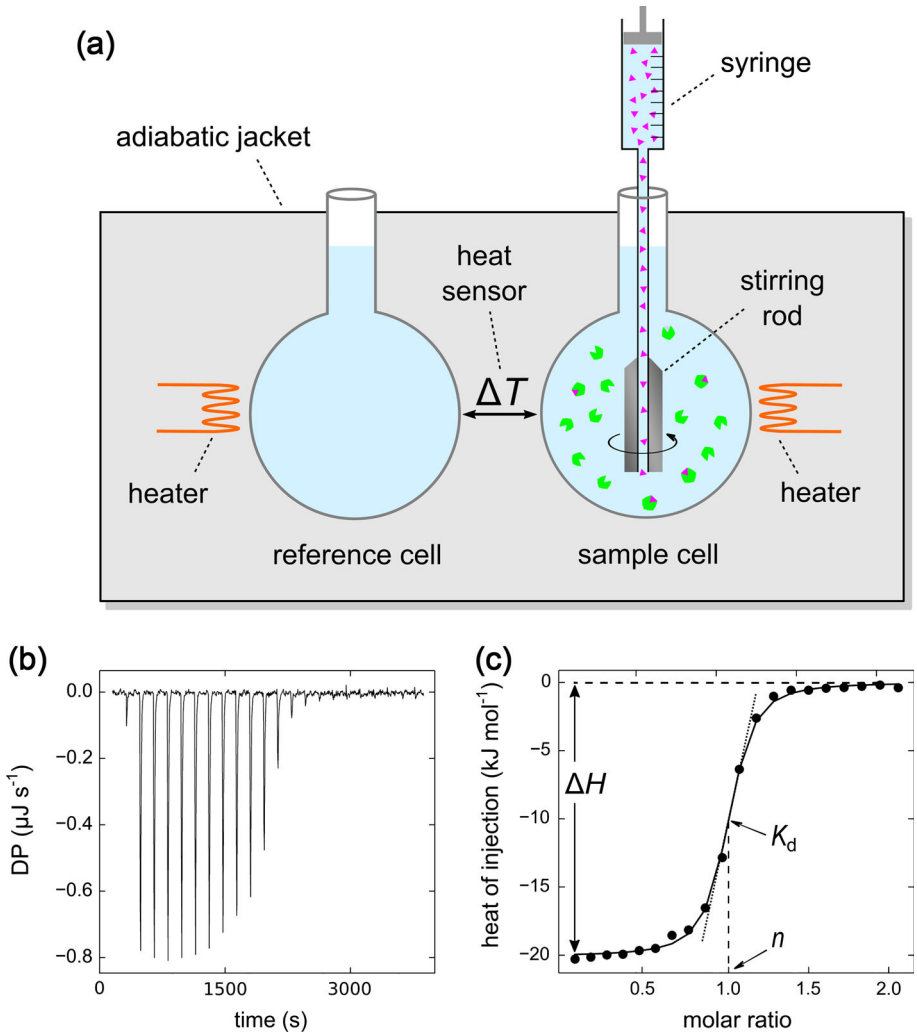

Fig. 1 **a** Schematic depiction of an ITC device. A solution containing dissolved ligand molecules (*magenta*) is stepwise injected into the sample cell containing a solution with dissolved protein (*green*). The heat released from the binding reaction in the sample cell between protein and ligand is recorded with respect to a reference cell. **b** Raw thermogram of an ITC measurement, the differential power (DP) in μJ s^{-1} of the electric device keeping both cells at constant temperature is plotted against time. **c** Integrated raw data and isotherm. The molar change in enthalpy observed for the injections is plotted against the molar ratio of the binding reaction. A 1:1 binding model is fitted to the data, from which $\Delta H°$, K_a and the stoichiometry n of the reaction are extracted

very small peaks of equal size remain which represent the heat of mixing of the solutions in the cell and in the syringe. Integration over these peaks can be used to define the zero baseline and to correct for the heat of dilution. For data analysis, all measured peaks until those purely resulting from dilution must be integrated. Integration in this manner gives the total amount of heat originating from each injection, which is then related to the amount of injected ligand. To achieve this, the measured heats are plotted against the molar ratio between ligand and protein concentration in the sample cell (Fig. 1c). An appropriate binding model is fitted to the data points, in the simplest case a single-site 1:1 binding model. More complicated cases such as a two-site or triple-site binding or a competitive binding require different models [22, 23]. The selection of the binding model must be performed carefully and ideally under the control of independent experimental results obtained by other techniques, e.g. knowledge of the binding mode from a crystal structure. After curve fitting, the thermodynamic parameters are then extracted from the model curve.

How to get which data from the ITC isotherm?

In Fig. 1c, a typical ITC isotherm is displayed. An appropriate model, in this case clearly a one-site binding model, was fitted to the data points of the integrated heat peaks via a nonlinear least squares fitting process. From the curve fitted to the data points, we obtain the change in enthalpy $\Delta H°$, the equilibrium constant K_a, and, by use of the latter value and application of Eq. 1, the Gibbs free energy $\Delta G°$ of the studied reaction [12, 24]. The change in enthalpy is related to the observed heat signal, while the K_a value is obtained from the slope at the inflection point. The location of the inflection point on the molar ratio axis describes the binding stoichiometry n, which is also referred to as the "site parameter". Importantly, the entropic term $-T\Delta S°$ of

binding is not available from an independent experiment but must be calculated as the numerical difference between $\Delta G°$ and $\Delta H°$, using Eq. 2. Accordingly, any error affecting the experimental determination of $\Delta G°$ or $\Delta H°$ will directly influence the calculated magnitude of the entropy.

Which requirements must a curve fulfill to enable the extraction of reliable thermodynamic parameters?

Optimally, a binding isotherm should show a sigmoidal curvature with plateaus at the beginning and end of the titration. Experimental uncertainties can be further reduced during the integration step by ensuring an adequate signal-to-noise level, as well as by observing significant differences between peaks resulting from the binding reactions compared to peaks from buffer mismatch reactions. A buffer mismatch reaction between syringe and sample cell buffer can result in huge mixing heat signals in addition to those originating from ligand binding. In order to avoid buffer mismatch, dialysis of protein and ligand solutions against the same buffer can be performed. A buffer mismatch is often the result of an inappropriate adjustment of the dimethyl sulfoxide (DMSO) concentration in cell and syringe, or due to a mismatch of the buffers' pH value. DMSO is a dipolar, low reactive solvent frequently added to increase ligand solubility. Furthermore, it is used for the preparation of ligand stock solutions (pure solutions of DMSO containing a high amount of ligand, typically between 10 and 100 mM). Such solutions are used for storage and efficient use of sometimes precious compound material. Prior to measurement, the stock solution is diluted with buffer to obtain the desired concentration, usually resulting in high concentration accuracy. However, it is recommended to keep the concentration of DMSO during the ITC measurement as low as possible. A maximum concentration of 5 % (v/v) should not be exceeded, which already corresponds to the high molar concentration of 0.7 mol L^{-1}. It must be kept in mind that even concentrations of 0.5–1 % (v/v) DMSO were reported to significantly influence protein–ligand binding parameters [25]. Furthermore, in one of our thermolysin crystal structures (PDB code 4D91), DMSO was found in complex with the active site of the protein. Therefore, at least for the metalloprotease thermolysin, DMSO actively competes for the binding site with any other ligand present and hence influences the measured binding parameters of the ligand. A further source of dilution peaks is the dissociation of ligand aggregates within the sample cell, which can occur upon injection into the larger volume of the sample cell.

Owing to these experimental deficiencies, an extraction of the heat signals originating solely from the protein–ligand complex formation is necessary. For the correction of the buffer mismatch peaks, it is considered as best practice to subtract the average of all constant dilution peaks recorded, which appear after the system has reached sufficient saturation, from all measured peaks [26]. Another possibility for correction is to perform control titrations: for the first control, the ligand solution is titrated into the sample cell containing pure buffer. For the second control, pure buffer is titrated into the sample cell containing the protein. Both control titrations are then subtracted from the actual titration curve of ligand into protein. Either way, the corrected integrals of the peaks are then fitted to an appropriate model curve. For the extraction of reliable thermodynamic parameters from the binding isotherm, the shape of the curve resulting from the fit is critical and can be described by evaluating the so-called c-value [8]:

$$c = nK_aM_{tot} \quad (3)$$

The parameter n describes the stoichiometry of the reaction (the molar ratio between syringe reactant and cell reactant inside the sample cell at which the inflection point of the titration curve occurs). K_a refers to the association constant of the ligand and M_{tot} (mol L^{-1}) to the total concentration of the macromolecule in the sample cell. In Fig. 2a, ITC isotherms with c-value between 0.01 and 1000 are shown.

Obviously, curves with high c-values show a clear sigmoidal curvature, whereas curves described by low c-values appear flat. In practical experience, ligands with an affinity of about 10^4 M^{-1}–10^8 M^{-1}, corresponding to K_d values between 100 μM and 10 nM, yield curves with c-values between 10 and 500, which is frequently considered as optimal by experimenters [8, 10, 28–31]. For such compounds, the experimental setup is often designed according to the so-called prevailing "standard-protocol" [32], consisting of about 25 injections and a molar ratio between ligand and protein of two at the end of the titration (Fig. 1b). For compounds in the "optimal" affinity window, this protocol usually results in sufficient heat signals for each injection and in sufficient protein saturation at the end of the titration, leading to well-analyzable titration curves (Fig. 1c). However, rather than the usually applied 25 injections, it was found that titration curves with the highest precision are achievable by designing the measurement with only 10 injections of equal volume, also resulting in a reduced runtime of the measurement [33]. A first, small injection, as visible in Fig. 1b, is usually performed and later discarded for data evaluation due to inaccuracy in the heat signal frequently observed for the first injection. However, it was shown that the inaccuracy in the heat signal is the result of an injection volume error originating from the syringe plunger's drive mechanism. The inaccuracy is observed directly after the drive direction

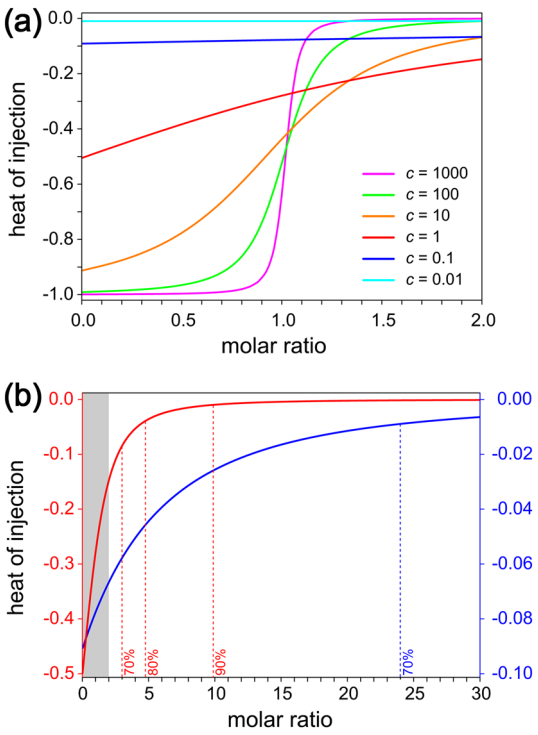

Fig. 2 **a** ITC isotherms of exothermic 1:1 binding reactions showing curvatures with *c*-values between 0.01 and 1000. The titration curves are shown up to a molar excess of two of the ligand over the protein. The arbitrarily chosen heat of injection of −1 corresponds to the exothermic heat signal for complete binding of the injected ligand. The isotherms were simulated with a modified version of a tool for modeling ITC curves of a perfusion calorimeter [27], not considering volume change or overflow of the sample cell. **b** Curves with *c*-values of 1 (*red*) and 0.1 (*blue*), displayed with different scales for the heat of injection, are titrated up to a molar ratio of 30 between ligand and protein. The part of the curves in the *grayed area* describes the curves resulting from a titration up to a molar ratio of two between ligand and protein. The *dashed lines* indicate the degree of protein saturation for a given molar ratio. Protein saturation was calculated with the fractional occupancy calculator from [19]

of the plunger changed, as is the case between the filling (up) and the ejection (down) movement of the plunger [34]. Thus, even if the first injection is deleted from the data evaluation, it is inaccurate to assume that the whole volume was actually injected into the sample cell. A simple solution to this problem is to perform a short 'down' movement of the plunger after the syringe filling but prior to the actual measurement. Thereby, volume errors can be significantly reduced [34].

For ligands with affinities lower than 10^4 M^{-1} (100 µM), titration curves exhibiting *c*-values below 10 are usually observed. As a matter of fact, such curves do not show a clear sigmoidal shape but rather a more simple one (Fig. 2a) without a clearly defined inflection point or a baseline at the beginning of the titration. In theory the *c*-value can be adjusted for every K_a by simply adjusting the concentration of the macromolecule participating in the reaction, according to Eq. 3. However, for the analysis of a ligand with the low affinity of 1 mM, this would require a protein concentration in the sample cell of about 10 mM in order to achieve a *c*-value of 10 [29]. In practice, this strategy is usually hampered owing to too low protein solubility and limited availability of protein material. In addition, at such high concentrations, deviation from an ideal solution will likely occur, resulting in reduced protein activity [7]. Therefore, a modified experimental setup must be applied—the so-called "low *c*-value titration" [30, 35]. In such a scenario, the low *c*-value curves are actually used for parameter analysis. The critical step of such a titration is to achieve sufficient reaction between protein and ligand [30]. A protein saturation of at least 70 % at the end of the titration has been suggested to be the lower limit [19, 30]. In order to achieve sufficient saturation, decreasing ligand affinity must be compensated with increasing ligand excess to favor the formation of the protein–ligand complex. For a low affinity ligand giving rise to a curve with a *c*-value of 0.01, this corresponds to a 24-fold molar ligand excess over the macromolecule in the sample cell (Fig. 2b). Consequently, ligand solubility is the main issue at this point for achieving the required ligand concentration in the syringe solution [19, 29, 30]. As mentioned before, curves exhibiting low *c*-values below 10 do not show a clear sigmoidal shape but a more simple one (Fig. 2a, b). Because a clear inflection point is missing, it is impossible to determine the value of *n* experimentally [29]. Nevertheless, in order to still get access to the thermodynamic parameters, the stoichiometry *n* of the binding reaction must be fixed according to an independently determined value. For an accurate determination of *n*, the concentration of protein and ligand as well as the binding ratio between protein and ligand must be exactly known. It has been shown that the error in ΔH° is strongly dependent on the error in *n* [33]. On the other hand, the determination of the affinity constant K_a turns out to be almost independent from the stoichiometry *n* [35]. Therefore, in a low *c*-value titration, even if the determination of accurate values for *n* and ΔH° fails, the affinity constant K_a can still be measured. Furthermore, because for low affinity ligands only a fraction of the injected ligand actually binds to the protein and thus produces a measurable heat signal, the observed signals are usually very low. Accordingly, low *c*-value titrations should be performed with a small amount of injections (e.g. only 4–5), but with a large injection volume [36]. Additionally, it is of advantage to vary the injection volume during the titration. As more protein becomes saturated and as a smaller fraction of the injected ligand binds, the gradual decrease in the heat signal can be compensated by increasing the injection volume.

Conversely to low *c*-value curves, curves with *c*-values above 1000 also create some problems in the analysis. For

curves with c-values >500, the uncertainty for the K_a determination increases [31]. Such curves no longer show sigmoidal curvature (Fig. 2a), but instead a more rectangular shape, which makes the determination of the slope at the inflection point unreliable. According to Eq. 3, in order to obtain an optimal c-value for high affinity ligands ($K_a > 10^8$ M^{-1}), measurement with very low protein concentration is required. This, however, can lead to injection peaks below the sensitivity range of the ITC instrument. In contrast to the assignment of K_a, the enthalpy of binding $\Delta H°$ is easily determinable for curves with high c-values. For $\Delta H°$, the molar heat signal of a complete binding reaction of all injected ligand molecules has to be determined. This can be reliably extracted from the step-like titration curves which show clearly defined plateaus at the beginning and end of the titration.

The displacement titration is an alternative strategy that has been developed in order to yield reliable microcalorimetry data from ligands across a wider range of affinities. This strategy is available for both low [19, 37] and high affinity binders [38]. For weak binding ligands, the protein is first saturated with the low-affinity ligand of interest, which is subsequently displaced by a previously characterized high-affinity reference ligand. Therefore, the reference ligand must bind competitively to the same protein site as the low-affinity ligand. As a result, a thermogram in the optimal c-value range is obtained. The amount by which this new competitive binding signal differs from the signal of the reference ligand alone depends on the amount of heat required to displace the low affinity ligand, which in turn relates to the latter ligand's binding signature. As a disadvantage, any uncertainties and experimental errors in the determination of the thermodynamic parameters of the reference ligand will also affect the parameters of the low-affinity ligand. The displacement strategy for high-affinity ligands follows the same concept as the displacement titration of weak binders, however with the important difference that a weak to medium potent ligand is used to preincubate the protein. This ligand must be previously characterized thermodynamically and serves as a reference ligand [38]. This strategy also allows the titration curve to shift into a c-value range that results in proper sigmoidal isotherms. From this, the stoichiometry and the K_a value are extracted. Unlike the characterization of weak-binding ligands, the $\Delta H°$ value is taken from a separate titration curve of the strong binding ligand showing a rectangular shape. As mentioned earlier, the rectangular shape is no obstacle for the $\Delta H°$ determination, and using this curve avoids error propagation that can occur due to uncertainties in the characterization of the reference ligand. Therefore, displacement titrations applied to ligands with a high affinity yield much more accurate data than displacement titrations of weak binders.

The interdependence of enthalpy and entropy

As mentioned, the ITC measurement determines only $\Delta H°$ and K_a experimentally, while changes in entropy are calculated from the numerical difference between $\Delta G°$ and $\Delta H°$ according to Eq. 2 [1]. Thus, considering that the values determined for K_a and, in consequence, $\Delta G°$ are less error-prone than those for $\Delta H°$ [29], the error of $-T\Delta S°$ will always depend on the error of $\Delta H°$. Calculating the $-T\Delta S°$ value from the numerical difference will propagate any error affecting the $\Delta H°$ measurement. In consequence, inaccurate measurement of $\Delta H°$ can lead to "artificial" enthalpy–entropy compensation (EEC) [39, 40]. In contrast to artificial EEC, "intrinsic" EEC describes the phenomenon by which enthalpy and entropy are really compensating each other and ultimately hardly influence the overall Gibbs free energy of binding, for example in a drug optimization scenario [41, 42]. It is intuitive to understand that during ligand optimization, EEC occurs at least to some extent: stronger fixation, which leads to a higher enthalpic contribution, leads to less flexibility and therefore lowers entropy and vice versa. However, due to inaccuracies in $\Delta H°$ and thus in $-T\Delta S°$ determination, the extent of EEC can be overestimated. An EEC purely imposed owing to experimental inaccuracies is particularly dangerous if a global analysis [43] of available thermodynamic data (for instance derived from the *BindingDB* [44], the *SCORPIO* [45] or from the *PDBcal* [46] online databases) is conducted, and thermodynamic data across different proteins and ligand series are compared on an absolute scale [14]. It must be considered that such thermodynamic data result from measurements conducted under deviating experimental conditions. The experiments are possibly performed at different temperatures, and buffers of deviating ionization enthalpies are used without applying the required correction. Protein concentrations are selected in very different ranges making direct comparison on absolute scale problematic. Furthermore, the experiments are obviously performed by different persons in different laboratories using different devices, leading to systematic deviations and uncertainties in the data of an unknown magnitude. A remarkable test case on ITC data accuracy has been studied across several laboratories. In the ABRF-MIRG'02 study [47], identical samples were thermodynamically characterized by 14 independent laboratories. Surprisingly enough, a plot of the determined $-T\Delta S°$ versus $\Delta H°$ values of the identical reaction performed 14 times suggests a nice EEC [39]. This study demonstrates how careful one must be when making a comparison of global data, and how easily such comparisons can be misleading. Before discussing the corrections necessary to reveal accurate comparative information, we want to argue

that data evaluated and compared relative to each other across a congeneric series of ligands can yield reliable information. In the case when congeneric ligand series are measured under the same conditions with concentrations falling in a narrow window and corrected for putative differences in the heats of ionization (see below), ΔH° can be determined with very high precision but also accuracy and the influence of an error-prone EEC can be minimized. In one of our studies of a congeneric ligand series binding to the well-established model system thermolysin [48], measurements were performed by the same operator over a short period of time using the same ITC device. Experimental conditions such as pH and temperature were kept constant. Moreover, it was also possible to keep the concentration of the protein and ligand solutions constant. The applied ligands were checked for high purity and the protein solutions were all prepared from the same batch. DMSO-free protein and ligand solutions were freshly prepared for each measurement. As a result, the extended ligand series showed ITC isotherms with c-values in the narrow range between 11 and 158 and an average stoichiometry n of 0.753 ± 0.04. The important characteristic of the observed stoichiometry is that it remains constant throughout all measurements. Deviations from the theoretical value of 1.000 are due to partial protein inactivity, which, however, has no effect on the accuracy of the measurement parameters and could easily be corrected by adjusting the protein concentration to the measured activity level of the respective batch. Again, this has no advantage for the measurement itself, but would mask the otherwise obvious partial protein activity. In our opinion, the ITC isotherms of such studies must be documented in the supporting information of a publication as they can be a proof of the accuracy of the measured data [48]. In our study, we examined how the rearrangement of surface water molecules during the protein–ligand binding reaction affects the thermodynamic signature of the complex formation by analyzing a series of closely related ligands. Because we were dealing with relatively small changes in their thermodynamic profiles, high precision—particularly with respect to a relative comparability—was very important.

Heat effects from proton transfer reactions between protein, ligand and buffer

The observed heat signal resulting from an ITC measurement (ΔH°_{obs}) is the sum of the heat signals produced by the actual binding event (the intrinsic change in binding enthalpy ΔH°_{bind}) plus any additional effects contributed by the entire system [49]. Most important are heat changes resulting from a proton transfer (protonation or deprotonation) between the formed protein–ligand complex and the surrounding buffer ($n_{H+}\Delta H^\circ_{ion}$):

$$\Delta H^\circ_{obs} = \Delta H^\circ_{bind} + n_{H+}\Delta H^\circ_{ion} \quad (4)$$

The explanation for such an occurrence, also known as 'proton linkage', can be found in a shift of the pK_a value of ionizable functional groups of the protein residues and/or the ligand during complex formation, as these groups are brought into a novel environment with different dielectric properties [50]. Depending on the buffer compounds (different buffers show different ionization enthalpies ΔH°_{ion} upon proton exchange) and the involved functional groups of protein and/or ligand, the heat of ionization ($n_{H+}\Delta H^\circ_{ion}$) can have a significant impact on the observed heat of binding (ΔH°_{obs}). By running the binding reaction in buffers exhibiting deviating heats of ionization, the resulting enthalpies will be different between buffers, but the association constant K_a and thus affinity data (ΔG°) are usually not significantly affected [51, 52]. The thermodynamic binding profiles of a ligand measured in different buffers showing proton linkage thus show similar affinities, but their enthalpic and entropic terms vary depending on the applied buffer, and resemble enthalpy−entropy compensation. This exemplifies how arbitrary an absolute scale comparison of such data would be—the uncorrected enthalpies are rather meaningless on such a scale due to the superimposed buffer effects. At best, it is still possible to uncover trends, but it is difficult to detect more subtle correlations. Therefore, if proton linkage occurs, enthalpies must be corrected for the heat of ionization before they are ready for comparative analysis on an absolute scale, or even a relative scale in some cases (see below). By measuring the heat signal of ligand binding in buffers of different ionization enthalpies at the same pH, e.g. in Tris, ACES, HEPES and PIPES buffer as performed in the experiment described in Fig. 3a–e, the enthalpic contribution from each buffer's heat of ionization can be determined. The number of protons exchanged during the reaction (n_{H+}) can also be identified (Fig. 3d), and the observed enthalpies can be corrected for their buffer contributions (Fig. 3e). To achieve this correction, the experimentally obtained enthalpies are plotted against the ionization enthalpies of the buffers considering the values referenced in literature [53]. A linear regression is performed, and its intercept with the y-axis reveals the buffer-corrected enthalpy.

Interestingly, ligands with more entropy-driven binding profiles are better measureable if they have an ionization reaction superimposed onto the actual binding event. Without the ionization reaction, the enthalpic signal of the binding reaction can be below the detection limit of the ITC device. This was the case in a ligand binding reaction to thrombin, which showed a buffer corrected enthalpy of

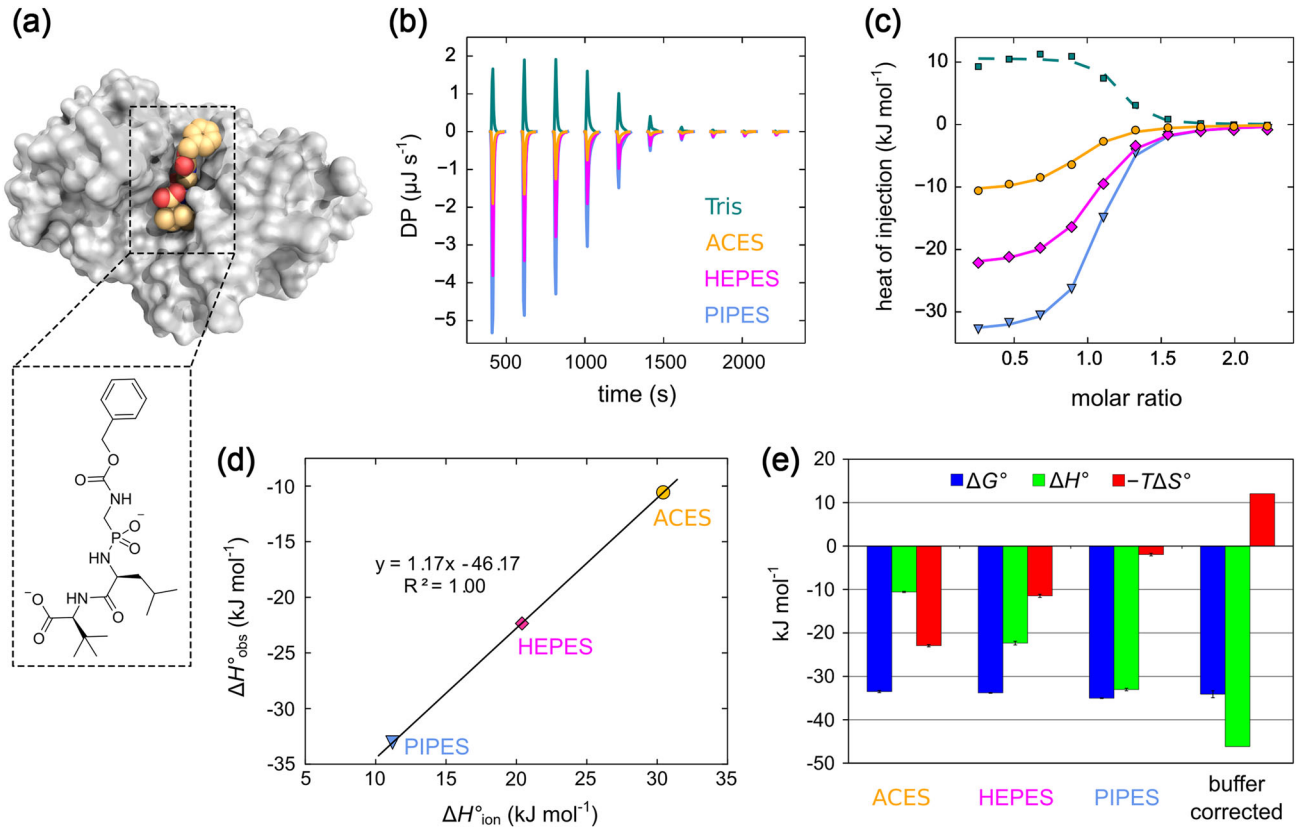

Fig. 3 Determination of the heat of ionization for the binding reaction between thermolysin and a phosphonamidate ligand. **a** The crystal structure of thermolysin (Connolly surface in *white*) in complex with the analyzed ligand clearly reveals a 1:1 binding mode (PDB code 5DPE). **b** Overlay of the ITC raw thermograms of the binding reaction measured in Tris, ACES, HEPES and PIPES buffer. Only extracted heat peaks (without baselines) are displayed, as performed by the peak shape analysis algorithm of *NITPIC* [54]. Except for the buffer substance, an identical experimental setup was applied for all titrations in order to guarantee comparability of the resulting heat signals. Whereas the binding reaction in Tris buffer results in an overall endothermic reaction (*upward peaks*), the complex formation in the other buffers is exothermic (*downward peaks*), its signal increasing from ACES to HEPES to PIPES. **c** Integrated data of the heat signals observed for the measurements in the four different buffers. The legend for the data is shown in *panel b*. The 1:1 binding model curve does not fit perfectly to the integrated data points of the titration in Tris buffer (*dotted lines*), suggesting a more complex scenario, likely due to an active displacement of Tris from the active site of the protein during ligand binding. Consequently, the titration in Tris buffer was not considered for the calculation of the heat of ionization. In contrast, the 1:1 binding model perfectly fits to the data points of ACES, HEPES and PIPES, confirming the chosen model in these cases. **d** Calculation of the heat of ionization. The experimentally observed enthalpies $\Delta H°_{obs}$ are plotted against the heat of ionization $\Delta H°_{ion}$ of the respective buffers. The slope of the *straight line* describes the proton uptake during the formation of the protein–ligand complex (on average 1.17 moles), whereas its interception with the y-axis describes the buffer corrected enthalpy of the binding reaction ($\Delta H°_{corrected} = -46.2$ kJ mol^{-1}). **e** Thermodynamic profiles of the complex formation in ACES, HEPES and PIPES buffer as well as the buffer corrected thermodynamic profile. For the buffer corrected profile, the change in the Gibbs free energy $\Delta G°$ is calculated as the average of $\Delta G°$ observed in the three buffers, $\Delta H°$ is derived as described in *panel d*, and the entropic term is calculated from the numerical difference between $\Delta G°$ and $\Delta H°$. More experimental details are given in the supplementary material

-1.4 kJ mol^{-1}—a value impossible to detect, if not the nicely measurable buffer uncorrected heat signals of -29.0 (Tris buffer), -17.4 (TRICINE buffer) and -14.3 kJ mol^{-1} (HEPES buffer) [55] would have occurred.

It should be noted that the buffer correction is performed under the often unfounded assumption that interactions between protein and ligand do not change with the various buffers and additives [7], even though salts can significantly influence the activity of the protein [7, 56], according to the Hofmeister series [57]. Furthermore, it must also be considered that, for instance in the case of the aspartic proteases [58], the pH used for the measurement can have a significant influence on the actual protonation state of residues and functional groups (e.g. on the catalytic dyad). The protonation state can influence the molar quantity of protons transferred, which in turn affects the heat of ionization. Thus, the enthalpy from the ligand binding process can vary, and the Gibbs free energy can also be altered. Interestingly, a method has been described where the pH dependence of binding affinity is exploited to

provide access to affinity data for binding that is too tight to be measured directly at the pH of interest. In this method, affinities are measured at pH values showing less tight binding, and are subsequently extrapolated to obtain the affinity at the pH of interest [50, 59].

In special cases, buffer-uncorrected enthalpies can be used for a relative comparison, particularly across a narrow compound series and if all studied ligands induce the same change in their protonation states. This may occur if the site where the ligands are structurally modulated is remote from the site where the protonation transfer occurs. All binding events will be influenced by the superimposed protonation change in similar fashion, but in a relative comparison across the series this contribution cancels out. For example, in a study of a congeneric series of phosphonamidate thermolysin ligands like the one shown in Fig. 3a [48, 60], we observed a buffer dependency of the enthalpic term. It was possible to identify Glu143 as the site which entraps one proton upon ligand binding. However, because the parent scaffold of the congeneric ligand series remains unchanged next to Glu143 and varies only at a site remote to it, all ligands are equally affected by the heat of ionization. In this example, only relative changes of the thermodynamic signatures of the ligands were of interest, and not their absolute values and thus the heat of ionization contributions fall out of the correlation. However, it must be underscored that such data cannot be used in a global correlation of thermodynamic properties on an absolute scale. In the described thermolysin example, it was sufficient to perform the ITC measurements in only one buffer. However, it would be rather meaningless and arbitrary to compare these results with data measured in other buffers or with ligands showing a deviating basic scaffold next to Glu143.

The presence of ionization effects upon complex formation is not always so obvious. In a study on thrombin inhibitors [55], mutual compensation of protonation effects between ligand and protein occurred upon ligand binding. The imidazole moiety of thrombin's His57 released 0.6 mole of protons, whereas a primary amino function of the inhibitor picked up an equal amount of protons, resulting in a negligible detectable net proton exchange. However, a ligand where the ionizable amino function was replaced by a non-ionizable amide function revealed the proton exchange upon complex formation—a release of 0.6 mole of protons only attributable to His57. It was gathered that the same proton release occurs during binding of the ligand with the amino function, but in this case it is masked by the superimposed proton uptake of the latter group, and therefore the expected buffer dependence is not apparent. In such a case, buffer ionization corrections will be difficult to make and accordingly cannot be successfully performed without further studies. One strategy to at least reduce the contribution of a superimposed proton linkage is to perform the measurement in a buffer with low heat of ionization (e.g. acetate buffer, $\Delta H^{\circ}_{\mathrm{ion}} = 0.41$ kJ mol^{-1}). Thus, the buffer contribution will be negligible. However, the contribution added by the group of the protein or ligand which displays the partner in the proton exchange reaction will still show a heat effect.

Are any further effects expected to modulate the heat contribution? Ions are often involved in ligand binding, and in some cases can be detected in the formed crystal structure [61]. The entrapment or the release of such ions most likely has a heat contribution, representing a possible artifact superimposed to the binding process which must be corrected. Further influences can originate from the salt as a component of the buffer. In recent studies on a host–guest system comprising a hydrophobic binding site [62, 63], the thermodynamics of binding is strongly influenced by different salts. The measurements in buffers containing NaF or NaCl ('kosmotropic' salts) result only in a slight increase in affinity with minor changes in enthalpy and entropy. However, ITC measurements in buffers containing $NaClO_4$, NaSCN, $NaClO_3$ or NaI ('chaotropic' salts) result in significantly decreasing K_a values, involving a major decrease in enthalpy and increase in entropy. It was shown that the chaotropic anions competitively bind to the hydrophobic pocket of the host and thereby modulate the thermodynamics of binding [63]. In our own studies, we analyzed a thermolysin-ligand binding reaction in buffers containing 200 and 1000 mM NaSCN (Fig. 4a, b). As a result, between 200 and 1000 mM NaSCN, the enthalpic term increases, whereas the entropic term decreases. Nonetheless, the Gibbs free energy is not significantly affected (Fig. 4a).

Therefore, especially if chaotropic salts are used as a buffer additive (e.g. for increasing the solubility of an otherwise not sufficiently soluble protein, so-called 'salting-in' effect of chaotropes) and the active site of a protein contains a hydrophobic concave surface as a binding site for the chaotropic anions [64], the binding profile can be significantly influenced by the added salt. Again, in the case of congeneric series of ligands where all studied ligands show the same effects, the contribution will cancel out in a relative comparison.

Temperature-dependency of ΔH°, change in heat capacity ΔC_p and van't Hoff analysis of ΔH°

It is well recognized that chemical processes are dependent on temperature. In consequence, chemical equilibria and the corresponding association or dissociation constants are temperature-dependent. As the Gibbs free energy is related to the latter constants (Eq. 1), also this property will in

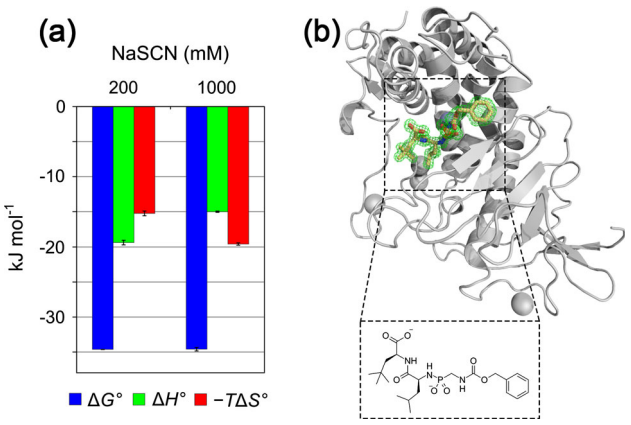

Fig. 4 **a** Effect of salt concentration on the thermodynamics of binding of the same thermolysin-ligand binding reaction performed in buffers containing the chaotropic salt NaSCN at concentrations of 200 and 1000 mM. Standard deviations are given for the measurements performed in duplicate. Experimental details are given in the supplementary material. **b** Crystal structure of the protein in complex with the analyzed ligand (PDB code 5DPF). Thermolysin is displayed as *white ribbon*, the ligand binding to the active site is displayed as stick representation (*orange*) and the $F_o - F_c$ omit electron density of the ligand is shown at a contour level of 3σ as green mesh. The crystal structure clearly reveals a 1:1 binding mode

general be dependent on temperature. ΔG° factorizes into enthalpy and entropy, whereby entropy is weighted with the absolute temperature (Eq. 2). Likewise, ΔH° and ΔS° change with temperature. The partial derivative of the enthalpy with respect to temperature while holding the pressure constant reveals the above-mentioned change in heat capacity ΔC_p (kJ mol^{-1} K^{-1}) of a reaction:

$$\Delta C_p = \left(\frac{\partial \Delta H^\circ}{\partial T}\right)_p \tag{5}$$

The change in heat capacity ΔC_p describes the amount of heat which is necessary for a temperature change of the system of 1 K. In other words, it describes how well the system can absorb or release heat, attributable to the available degrees of freedom [14]. Empirically, a correlation of increasing ΔC_p with an increasing burial of apolar and polar surfaces between macromolecules has been found, which is associated to the displacement of water molecules upon complex formation [65]. According to Eq. 5, for the analysis of the change in heat capacity ΔC_p of a protein–ligand complex formation, the change in ΔH° at different temperatures needs to be determined. Interestingly, in biological systems, ΔC_p of a protein–ligand complex formation almost exclusively exhibits negative values, and usually adopts values differing from zero. Accordingly, the complex exhibits a lower heat capacity compared to the sum of heat capacities of protein and ligand in their uncomplexed state. With respect to enthalpy and entropy, this general behavior results in the finding that protein–ligand complex formation becomes more exothermic (enthalpic) with increasing temperature and simultaneously entropically less favorable [49]. This observation can be exploited in ITC measurements. The property measured in an ITC experiment is a heat signal resulting from the enthalpic component of binding. Thus, a predominantly entropically driven process hardly produces any measurable effect. If such a situation is experienced, the titration should be repeated at a temperature 5 or 10 K higher or lower. Then, usually a detectable signal can be recorded. On the other hand, this observation clearly demonstrates that the thermodynamic properties are not temperature independent, even in the small windows accessible with biological systems. It also implies that values of ΔG°, ΔH° and $-T\Delta S^\circ$ measured at different temperatures can hardly be compared directly. Furthermore, it indicates that some care is needed to define a process as 'enthalpy or entropy-driven', as it matters at which temperature the process has been recorded [66]. This means, in a discussion of thermodynamic properties, we should only compare series of complexes measured at the same temperature relative to each other and regard them in the comparative analysis as 'enthalpically or entropically more favored' in their formation.

Popular evaluations of thermodynamic properties make use of the so-called van't Hoff evaluation [14, 67, 68]. For this, usually the biological system under consideration is studied at for example three different temperatures by evaluating well-recordable signals such as a change in a spectroscopic property or shifts in resonance signals. The recorded signals are then used to quantify the concentrations (or better: activities) of the unbound and bound species involved in the equilibrium. However, therefore, the binding event has to follow a two-state transition between the free and bound state and the change in the recorded spectrometric signals, subsequently used to assign the binding constant, has to consider all of the free and bound molecules involved in the complex formation reaction [1, 14, 49]. Considering the recent results found by simulations to describe binding kinetics, at least questions this assumption quite strongly, as usually multistep mechanisms have to be discussed [69]. At this point the burden is on the experimentalist to correctly assign the concentrations at equilibrium, however, usually it is by no means trivial to ensure this assumption. The measurements of the binding constants are in the following performed at different temperatures and for the evaluation the integrated form of the van't Hoff equation is used [14]:

$$\ln\left(\frac{K_2}{K_1}\right) = \frac{1}{R}\int_{T_1}^{T_2} \frac{\Delta H^\circ(T)\mathrm{d}T}{T^2} \tag{6}$$

The binding constants K_1 and K_2 for a reaction describe the measurement at the two different temperatures T_1 and

T_2. Frequently, for the evaluation a "simpler" form, the so-called linear form of the van't Hoff equation is used (Eq. 7), which, however, only arises if ΔH° is assumed to be temperature independent, as only then it can be taken out of the integral:

$$\ln\left(\frac{K_2}{K_1}\right) = \frac{\Delta H^\circ}{R}\int_{T_1}^{T_2}\frac{dT}{T^2} = -\frac{\Delta H^\circ}{R(T_2 - T_1)} \quad (7)$$

Applying this latter form, the binding constants are plotted against the reciprocal of the temperature and evaluated by a linear fit, where the slope of the straight line describes the van't Hoff enthalpy. However, as described above, this is usually a non-valid assumption, as experience shows that ΔC_p deviates from zero and thus ΔH° is actually temperature dependent. To circumvent this, integration of the differential form of the equation requires some kind of approximation to describe the temperature dependency of $\Delta H^\circ(T)$, for example as a Taylor expansion, to achieve a non-linear fit.

The advantage of ITC experiments is that they are performed at one temperature and reflect the entire binding process. They including all heat signals produced, even if binding passes through multiple states. From this, ΔG° and ΔH° become available. At first glance, heat capacity changes appear as an ideal property to relate structural properties and molecular degrees of freedom with thermodynamic entities. However, measurements of ΔC_p require ITC experiments to be performed across a temperature range. As a matter of fact, the complexity of multicomponent systems like protein–ligand complexes, including the surrounding aqueous buffer environment, is so large that the changes of heat capacity are very difficult to interpret on molecular level [14]. It should not be forgotten that the ubiquitously present water in biological systems is a substance with one of the largest heat capacities known, and most likely the changes with temperature while studying biological processes involve major changes in the surrounding water environment superimposed or inherently correlated with the changes of the biological system.

The importance of high ligand purity and accurately known ligand and protein concentrations

The importance of ligand purity and the determination of the exact ligand concentration is well appreciated [14, 30, 47, 56]. Inaccurate ligand concentration can be the result of solution preparation directly based on a ligand sample's weight if the sample contains unexpected impurities. Water is a common impurity for hygroscopic powders in particular; impurities may also originate from the synthesis. Even without impurities, accurate weighting in can be a serious problem, especially for electrostatically charged ligand powders. This problem can be addressed by an antistatic device, which, however, is not available in many laboratories. Another concern is that ligands in solution (for instance in a ligand stock solution) can suffer from chemical instabilities like partial hydrolysis over time during storage. Inaccurate or inadequate methods to determine the ligand solution concentration, for example via HPLC, might also impose a problem. Incorrect ligand and protein concentrations both have only minor consequences for the accurate determination of ΔG° [29, 30, 56]. However, for titrations in the *c*-value range of 10–500, incorrect ligand concentrations have a huge impact on ΔH° in particular [29, 30, 56], because the measured heat signal is attributed to a false amount of injected ligand. Errors resulting from ill-defined ligand concentrations must be classified as systematic, mostly unrecognized errors [47, 56, 70]. On the other hand, for low *c*-value titrations, it is the inaccuracy in the actually active protein concentration that lead to inaccurate ΔH° determinations [29, 30]. In addition to their effects on concentration, ligand impurities also lead to unpredictable heat reactions. The first indication about ligand purity is the stoichiometry *n* of the binding reaction available from proper sigmoidal titration curves (described by the 'incomplete fraction' parameter in the program *SEDPHAT*), especially in studies of ligand series binding to the same protein. Assuming that the protein shows unchanged activity in each measurement (which can be achieved by using protein material from the same batch), the stoichiometry should remain unchanged throughout the measurement of the whole series. If this is not the case, ligand impurity may provide an explanation. It must be noted that the experimentally determined stoichiometry will hardly match exactly 1.00, even in a simple one-site binding reaction, due to partial degradation or denaturation of the protein. If the protein activity is controlled by an independent experiment and *n* is found to be significantly lower than the expected value, the most likely reason is a higher than expected ligand concentration. If the stoichiometry cannot be determined experimentally, as is the case for low *c*-value titrations with fixed stoichiometry, a thorough purity validation must be conducted. This can be done using mass spectrometry (MS), quantitative nuclear magnetic resonance spectroscopy (qNMR), high-performance liquid chromatography (HPLC) or elemental analysis.

How accurate are ITC results, what is the true error and which systematic errors exist?

As mentioned, a comparative study across 14 independent laboratories has investigated the simple one-to-one binding reaction between carbonic anhydrase II and

4-carboxybenzensulfonamide. This study gave a $\Delta H°$ of -43.5 ± 10.5 kJ mol^{-1} and a K_a of $1.00 \pm 0.22 \times 10^6$ M^{-1} [47]. The reported values suggest rather worrying uncertainties of more than 20 % in $\Delta H°$ as well as in K_a! For the determination of K_a, *c*-values of the isotherms below 20 were found to be the main source of such pronounced uncertainty. However, due to the logarithmic relationship between K_a and $\Delta G°$ (Eq. 1), uncertainties in the values of K_a are of minor influence for the calculation of accurate $\Delta G°$ values [29]. For the determination of $\Delta H°$, an accurate ligand concentration was found to be particularly critical. In a reanalysis of this study [70], the observed ligand concentration uncertainties were found to amount to about 10 %, while it was stated that based on all precision limiting steps, uncertainties of below 1 % in the ligand concentration could have been achieved for this reaction. If quantifiable, uncertainties in the ligand concentration should be stated together with other errors as a total uncertainty value [70]. However, the errors reported for ITC experiments are often simply taken from the nonlinear least squares fit of a model curve to the data points. Alternatively, standard deviations are given for multiply performed measurements, which state the repeatability of the measurement by one person, but not its reproducibility by independent persons and over independent laboratories [3]. Hence, the observed deviations of more than 20 % in the study must be considered as systematic errors which would otherwise have never been reported, and the errors detected by all independent laboratories would have been greatly underestimated. One way to discover systematic errors in $\Delta H°$ originating from deficiencies in the execution of the measurement is the use of enthalpy standards, which can also uncover uncertainties originating from the ITC instrumental setup itself, for instance from devices such as VP-ITC, ITC_{200}, and the Nano ITC-III. One proposed enthalpy standard reaction is the titration of 5 mM NaOH into the cell containing 0.5 mM HNO_3 at 25 °C [71]. The performance of chemical calibration has been suggested in addition to the routine electric calibration in order to avoid the occurrence of undetectable, systematic calibration errors of the ITC instrument and thus lowered accuracy [72]. Electrical calibration errors of 5 % were reported as not uncommon [73]. This is of particular importance if the determination of thermodynamic data on an absolute scale is intended. However, for ligands with affinities in the optimal range, especially for relative comparison, congeneric ligand series of well-characterized systems are expected to show deviations in $\Delta H°$ smaller than 1 kJ mol^{-1} [14]. This estimation is in reasonable agreement with reported achievable deviation of 1 % for K_a and $\Delta H°$ without the inclusion of systematic errors, and of 3 % with systematic errors consideration [32].

Recently, certain weaknesses in the commonly applied protocol for analyzing K_a (and thus $\Delta G°$ and $-T\Delta S°$) from the experimentally obtained enthalpies have been pointed out [7]. For the correct calculation of the equilibrium constant, all components involved in the equilibrium must be considered. In addition to protein–ligand interactions, this involves interactions between buffer components and protein molecules, as well as protein–protein interactions. Clearly, such interactions are almost never considered in the parameter determination. Furthermore, as mentioned in the beginning, activities of the solvated components must be taken rather than concentrations because the studied solutions are likely not ideal. Typically, the differences between concentrations and their real activities are considered negligible. However, the concentrations of especially weak binding ligands and protein solutions can differ significantly from their activities. For instance, the activity coefficient of protein molecules can be influenced by the applied buffer (buffer salt, pH, additives). The ligand solution activity can be influenced by partial insolubility or ligand aggregation, especially of hydrophobic compounds. One option for considering the possible influence of activities instead of concentrations is the implementation of ITC measurements over a protein concentration range and in different buffers [7]. This should show whether the recorded equilibrium constants are equal for every measurement. Strong concentration dependencies would suggest a necessity to determine the real protein and ligand activities, for instance via equilibrium dialysis or potentiometric titration [74]. We performed ITC titrations of the same protein–ligand binding reaction with different thermolysin concentrations between 50 and 300 μM (Fig. 5). As a result, the magnitude of $\Delta G°$, $\Delta H°$ and $-T\Delta S°$ significantly decreases with increasing protein concentration, whereas the relative difference between $\Delta H°$ and $-T\Delta S°$

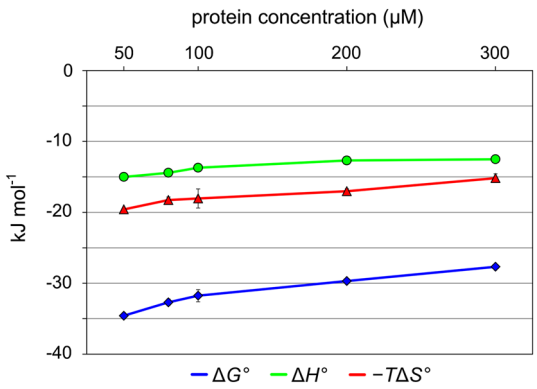

Fig. 5 Thermodynamic binding profiles of the same protein–ligand binding reaction measured at different thermolysin concentrations of 50, 80, 100, 200 and 300 μM. The chemical structure of the analyzed ligand is shown in Fig. 4b (crystal structure PDB code 5DPF). Experimental details are given in the supplementary material

remains constant. Accordingly, over the studied concentration range, the measured protein solutions are no ideal mixtures and their concentrations are not equal to their real activities. Thus, comparison of data on an absolute scale and from measurements based on different protein concentrations cannot be performed accurately without knowledge of the activity coefficients.

Further systematic errors can originate from numerous sources, including solvent evaporation during the measurement, adsorption of reactive components at the cell wall, mechanical effects (e.g. from the stirring of the syringe paddle), metal corrosion of the device [73], smaller volume of the sample cell than usually assumed [75], as well as the temperature dependency of the buffer pH [76]. These factors will not be discussed here in detail.

Comparison of available analysis software

For the analysis of raw data, several analysis programs are available. For instance, *Origin* 7 SR4 (OriginLab Corporation, Northampton, MA, USA) is useful for peak integration and model fitting. Alternatively, *NITPIC* [54] can be used for peak integration in combination with its companion program *SEDPHAT* [77] for model curve fitting. Another option is *AFFINImeter* (Software for Science Developments, Santiago de Compostela, A Coruña, Spain), a web-based tool for model fitting of integrated data. In our own experience, *Origin* gives comparable results to *NITPIC*/*SEDPHAT* for titrations showing strong heat peaks. However, for smaller peaks with poor signal-to-noise ratio or a less well-defined baseline, analysis can be tricky using *Origin*. Manual adjustment of baseline and integration limits is frequently required, and can easily induce undesired bias, especially in the hands of unexperienced users. We have found that the shape analysis and integration of heat peaks by *NITPIC* in combination with model fitting by *SEDPHAT* delivers the most unbiased, well-defined thermograms. The achievability of quality improved isotherms by *NITPIC* compared to *Origin* has also been described in literature [78]. For further improvement in data precision, *SEDPHAT* offers the combined analysis of several ITC isotherms ('global ITC'), and even offers the analysis in combination with data originating from other biophysical techniques ('global multi-method analysis') such as surface plasmon resonance (SPR) [79].

Conclusion

For the estimation of the quality of data obtained by ITC, it is necessary to develop a basic understanding of the method itself. Then, under the assumption that sufficient experimental details can be extracted from the measurement protocol, judgement about accuracy and uncertainty of thermodynamic data can be drawn. Analysis of the shape and curvature of the ITC isotherm and of the stoichiometry n of the reaction provides information about data accuracy—the shape of the fitted model curve relative to the data points informs whether the chosen binding model is appropriate and how well the fit can actually be achieved, whereas the curvature (c-value) informs whether a reliable extraction of the thermodynamic binding parameters from the curve is possible. Curvatures that are too flat or rectangular can lead to inaccuracies in the parameter extraction and require to apply special ITC techniques, such as low c-value titrations or displacement titrations. The stoichiometry, only experimentally available for sigmoidal titration curves of a binding reaction, especially when comparable across a series of ligands, can be an indicator for the purity of both ligand and protein. Inaccuracies in the latter will likely affect the accuracy of the recorded thermodynamic parameters. A very important point is the dependence of the thermodynamic parameters on the applied measurement conditions, especially if comparison of data on an absolute scale is intended, which to our opinion is hardly possible to achieve. Nevertheless, this has been frequently done in literature, particularly to derive general rules about thermodynamic properties and optimization strategies in medicinal chemistry. A lot of care is needed in the interpretation to establish such correlations. Protonation reactions superimposed onto the actual binding event can strongly affect the measured enthalpic contribution to binding. If this is the case, the buffer effect must be corrected prior to data usage. A comparison of thermodynamic data including different, uncorrected heats of protonation will induce vast systematic errors, and artificial enthalpy–entropy compensation will arise from this lack of proper data correction. Trends can disappear in such arbitrarily correlated data. Furthermore, thermodynamic measurements have to be performed at the same temperature if mutual comparison is intended.

The best data quality can be achieved by using an experimental setup that is optimized with respect to the number of injections and the injection volume (resulting in strong heat signals and a sigmoidal curvature of the isotherm), the ratio between ligand and protein at the end of the titration (sufficient protein saturation) and the buffer conditions (small heat of dilution, experimentally determined heat of ionization). Usage of the same protein batch with unchanged concentrations across the entire experimental series and highly pure ligand, measurement at a constant temperature, and performing all steps with the same operator and ITC device are also important. If necessary, heats of ionization must be corrected. Considering the complexity of ITC experiments and the large variety of

possibly superimposed systematic effects, it is highly recommended to use ITC data only for a relative comparison within narrow congeneric compound series. In our eyes, only such evaluations make sense and can lead to relevant and reliable conclusions. We also believe that classifications of ligands as "enthalpic" or "entropic" binders should only be done as relative comparisons of closely matching pairs. In any case such relative classifications have to be limited to "*more* enthalpic" or "*more* entropic" in light of the fact that with increasing temperature protein–ligand binding becomes in general more enthalpy-driven and ITC experiments are usually performed at 25 °C and not at body temperature.

For the assessment of the data quality, we rely on detailed experimental protocols provided by the experimenter. They describe the measurement parameters, raw thermograms, report ITC isotherms, assessment of possibly superimposed ionization reactions, and prove ligand purity. Unfortunately, this is often not given, even though it should be self-evident to include such data in the publication or in the supplementary material. Accordingly, putative reviewers of paper submissions are prompted to request such information from the authors. Only this will enable others to validate whether the data are suitable and reliable enough for their purposes, for instance for a computational study.

Acknowledgments SK was kindly supported by the European Research Council (ERC) Advanced Grant No. 268145-DrugProfilBind awarded to GK. We want to thank Mahalia Lepage for thorough proof-reading of the manuscript.

References

1. Klebe G (2015) The use of thermodynamic and kinetic data in drug discovery: Decisive insight or increasing the puzzlement? ChemMedChem 10:229–231
2. Kramer C, Lewis R (2012) QSARs, data and error in the modern age of drug discovery. Curr Top Med Chem 12:1896–1902
3. Kramer C, Kalliokoski T, Gedeck P, Vulpetti A (2012) The experimental uncertainty of heterogeneous public K(i) data. J Med Chem 55:5165–5173
4. Kalliokoski T, Kramer C, Vulpetti A (2013) Quality issues with public domain chemogenomics data. Mol Inform 32:898–905
5. Wätzig H, Oltmann-Norden I, Steinicke F, Alhazmi HA, Nachbar M, El-Hady DA, Albishri HM, Baumann K, Exner T, Böckler FM, El Deeb S (2015) Data quality in drug discovery: the role of analytical performance in ligand binding assays. J Comput Aided Mol Des. doi:10.1007/s10822-015-9851-6
6. Tanford C (1978) The hydrophobic effect and the organization of living matter. Science 200:1012–1018
7. Pethica BA (2015) Misuse of thermodynamics in the interpretation of isothermal titration calorimetry data for ligand binding to proteins. Anal Biochem 472:21–29
8. Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem 179:131–137
9. Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 11:560–566
10. Thomson JA, Ladbury JE (2004) Part II, Isothermal titration calorimetry: a tutorial. In: Ladbury JE, Doyle ML (eds) Biocalorimetry 2: application of calorimtry in the biological sciences, chap 2. Wiley, Chichester
11. Perozzo R, Folkers G, Scapozza L (2004) Thermodynamics of protein–ligand interactions: history, presence, and future aspects. J Recept Signal Transduct Res 24:1–52
12. Freyer MW, Lewis EA (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. Methods Cell Biol 84:79–113
13. Martin SF, Clements JH (2013) Correlating structure and energetics in protein–ligand interactions: paradigms and paradoxes. Annu Rev Biochem 82:267–293
14. Klebe G (2015) Applying thermodynamic profiling in lead finding and optimization. Nat Rev Drug Discov 14:95–110
15. Biela A, Sielaff F, Terwesten F, Heine A, Steinmetzer T, Klebe G (2012) Ligand binding stepwise disrupts water network in thrombin: enthalpic and entropic changes reveal classical hydrophobic effect. J Med Chem 55:6094–6110
16. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? Drug Discov Today 13:869–874
17. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:23–27
18. Ferenczy GG, Keseru GM (2010) Thermodynamics guided lead discovery and optimization. Drug Discov Today 15:919–932
19. Rühmann E, Betz M, Fricke M, Heine A, Schäfer M, Klebe G (2015) Thermodynamic signatures of fragment binding: validation of direct versus displacement ITC titrations. Biochim Biophys Acta 1850:647–656
20. Ruehmann E, Betz M, Heine A, Klebe G (2015) Fragments can bind either more enthalpy or entropy-driven: crystal structures and residual hydration pattern suggest why. J Med Chem. doi:10.1021/acs.jmedchem.5b00812
21. Geschwindner S, Ulander J, Johansson P (2015) Ligand binding thermodynamics in drug discovery: Still a hot tip? J Med Chem. doi:10.1021/jm501511f
22. MicroCal LLC (2004) ITC data analysis in Origin® tutorial guide. MicroCal LLC, Northampton
23. Le VH, Buscaglia R, Chaires JB, Lewis EA (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. Anal Biochem 434:233–241
24. GE Healthcare Life Sciences (2012) Microcal ITC200 system user manual 29017607 AA. GE Healthcare Bio-Sciences AB, Uppsala
25. Cubrilovic D, Zenobi R (2013) Influence of dimehylsulfoxide on protein–ligand binding affinities. Anal Chem 85:2724–2730
26. Ghai R, Falconer RJ, Collins BM (2012) Applications of isothermal titration calorimetry in pure and applied research—survey of the literature from 2010. J Mol Recognit 25:32–52
27. Biswas T, Tsodikov OV (2010) An easy-to-use tool for planning and modeling a calorimetric titration. Anal Biochem 406:91–93
28. Broecker J, Vargas C, Keller S (2011) Revisiting the optimal c value for isothermal titration calorimetry. Anal Biochem 418:307–309
29. Turnbull WB (2011) Divided we fall? Studying low-affinity fragments of ligands by ITC. GE Healthcare Bio-Sciences AB, Uppsala
30. Turnbull WB, Daranas AH (2003) On the value of c: Can low affinity systems be studied by isothermal titration calorimetry? J Am Chem Soc 125:14859–14866

31. Hansen LD, Fellingham GW, Russell DJ (2011) Simultaneous determination of equilibrium constants and enthalpy changes by titration calorimetry: methods, instruments, and uncertainties. Anal Biochem 409:220–229
32. Tellinghuisen J (2012) Designing isothermal titration calorimetry experiments for the study of 1:1 binding: problems with the "standard protocol". Anal Biochem 424:211–220
33. Tellinghuisen J (2005) Optimizing experimental parameters in isothermal titration calorimetry. J Phys Chem B 109:20027–20035
34. Mizoue LS, Tellinghuisen J (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. Anal Biochem 326:125–127
35. Tellinghuisen J (2008) Isothermal titration calorimetry at very low c. Anal Biochem 373:395–397
36. Tellinghuisen J (2007) Optimizing experimental parameters in isothermal titration calorimetry: variable volume procedures. J Phys Chem B 111:11531–11537
37. Zhang YL, Zhang ZY (1998) Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B. Anal Biochem 261:139–148
38. Velazquez-Campoy A, Freire E (2006) Isothermal titration calorimetry to determine association constants for high-affinity ligands. Nat Protoc 1:186–191
39. Chodera JD, Mobley DL (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. Annu Rev Biophys 42:121–142
40. Cornish-Bowden A (2002) Enthalpy–entropy compensation: a phantom phenomenon. J Biosci 27:121–126
41. Dunitz JD (1995) Win some, lose some: enthalpy–entropy compensation in weak intermolecular interactions. Chem Biol 2:709–712
42. Olsson TSG, Ladbury JE, Pitt WR, Williams MA (2011) Extent of enthalpy–entropy compensation in protein–ligand interactions. Protein Sci 20:1607–1618
43. Reynolds CH, Holloway MK (2011) Thermodynamics of ligand binding and efficiency. ACS Med Chem Lett 2:433–437
44. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Res 35:198–201
45. Olsson TSG, Williams MA, Pitt WR, Ladbury JE (2008) The thermodynamics of protein–ligand interaction and solvation: insights for ligand design. J Mol Biol 384:1002–1017
46. Li L, Dantzer JJ, Nowacki J, O'Callaghan BJ, Meroueh SO (2008) PDBcal: a comprehensive dataset for receptor–ligand interactions with three-dimensional structures and binding thermodynamics from isothermal titration calorimetry. Chem Biol Drug Des 71:529–532
47. Myszka DG, Abdiche YN, Arisaka F, Byron O, Eisenstein E, Hensley P, Thomson JA, Lombardo CR, Schwarz F, Stafford W, Doyle ML (2003) The ABRF-MIRG'02 study: assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction. J Biomol Tech 14:247–269
48. Krimmer SG, Betz M, Heine A, Klebe G (2014) Methyl, ethyl, propyl, butyl: futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. ChemMedChem 9:833–846
49. Jelesarov I, Bosshard HR (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to invesitigate the energetics of biomolecular recognition. J Mol Recognit 12:3–18
50. Doyle ML, Louie GL, Dal Monte PR, Sokoloski TD (1995) Tight binding affinities determined from linkage to protons by titration calorimetry. Methods Enzymol 259:183–194
51. Baker BM, Murphy KP (1996) Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. Biophys J 71:2049–2055
52. Parker MH, Lunney EA, Ortwine DF, Pavlovsky AG, Humblet C, Brouillette CG (1999) Analysis of the binding of hydroxamic acid and carboxylic acid inhibitors to the stromelysin-1 (matrix metalloproteinase-3) catalytic domain by isothermal titration calorimetry. Biochemistry 38:13592–13601
53. Goldberg RN, Kishore N, Lennen RM (2002) Thermodynamic quantities for the ionization reaction of buffers. J Phys Chem Ref Data 31:231–370
54. Keller S, Vargas C, Zhao H, Piszczek G, Brautigam CA, Schuck P (2012) High-precision isothermal titration calorimetry with automated peak shape analysis. Anal Chem 84:5066–5073
55. Baum B, Muley L, Heine A, Smolinski M, Hangauer D, Klebe G (2009) Think twice: understanding the high potency of bis(phenyl)methane inhibitors of thrombin. J Mol Biol 391:552–564
56. Grüner S, Neeb M, Barandun LJ, Sielaff F, Hohn C, Kojima S, Steinmetzer T, Diederich F, Klebe G (2014) Impact of protein and ligand impurities on ITC-derived protein–ligand thermodynamics. Biochim Biophys Acta 1840:2843–2850
57. Boström M, Williams DRM, Ninham BW (2003) Specific ion effects: why the properties of lysozyme in salt solutions follow a Hofmeister series. Biophys J 85:686–694
58. Xie D, Gulnik S, Collins L, Gustchina E, Suvorov L, Erickson JW (1997) Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurement of the protonation states of the catalytic aspartic acid residues. Biochemistry 36:16166–16172
59. Chu AH, Turner BW, Ackers GK (1984) Effects of protons on the oxygenation-linked subunit assembly in human hemoglobin. Biochemistry 23:604–617
60. Biela A, Nasief NN, Betz M, Heine A, Hangauer D, Klebe G (2013) Dissecting the hydrophobic effect on the molecular level: the role of water, enthalpy, and entropy in ligand binding to thermolysin. Angew Chem Int Ed 52:1822–1828
61. Neeb M, Betz M, Heine A, Barandun LJ, Hohn C, Diederich F, Klebe G (2014) Beyond affinity: enthalpy–entropy factorization unravels complexity of a flat structure-activity relationship for inhibition of a tRNA-modifying enzyme. J Med Chem 57:5566–5578
62. Gibb CLD, Oertling EE, Velaga S, Gibb BC (2015) Thermodynamic profiles of salt effects on a host-guest system: new insight into the hofmeister effect. J Phys Chem B. doi:10.1021/acs.jpcb.5b01708
63. Gibb CLD, Gibb BC (2011) Anion binding to hydrophobic concavity is central to the salting-in effects of hofmeister chaotropes. J Am Chem Soc 133:7344–7347
64. Fox JM, Kang K, Sherman W, Héroux A, Sastry M, Baghbanzadeh M, Lockett MR, Whitesides GM (2015) Interactions between hofmeister anions and the binding pocket of a protein. J Am Chem Soc . doi:10.1021/jacs.5b00187
65. Spolar RS, Record MT (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science 26:777–784
66. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed 41:2644–2676
67. Horn JR, Russell D, Lewis EA, Murphy KP (2001) Van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: Are there significant discrepancies? Biochemistry 40:1774–1778
68. Mizoue LS, Tellinghuisen J (2004) Calorimetric vs. van't Hoff binding enthalpies from isothermal titration calorimetry: Ba2+-crown ether complexation. Biophys Chem 110:15–24
69. Pan AC, Borhani DW, Dror RO, Shaw DE (2013) Molecular determinants of drug-receptor binding kinetics. Drug Discov Today 18:667–673
70. Tellinghuisen J, Chodera JD (2011) Systematic errors in isothermal titration calorimetry: concentrations and baselines. Anal Biochem 414:297–299

71. Baranauskiene L, Petrikaite V, Matuliene J, Matulis D (2009) Titration calorimetry standards and the precision of isothermal titration calorimetry data. Int J Mol Sci 10:2752–2762
72. Wadsö I (2000) Needs for standards in isothermal microcalorimetry. Thermochim Acta 347:73–77
73. Wadsö I, Wadsö L (2005) Systematic errors in isothermal micro- and nanocalorimetry. J Therm Anal Calorim 82:553–558
74. Hermans J, Barry L (2014) Equilibrium and kinetics of biological macromolecules. Wiley, Hoboken
75. Tellinghuisen J (2004) Volume errors in isothermal titration calorimetry. Anal Biochem 333:405–406
76. Good NE, Winget GD, Winter W, Connolly TN, Izawa S, Singh RMM (1966) Hydrogen ion buffers for biological research. Biochemistry 5:467–477
77. Houtman JCD, Brown PH, Bowden B, Yamaguchi H, Appella E, Samelson LE, Schuck P (2007) 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 Sci 16:30–42
78. Scheuermann TH, Brautigam CA (2015) High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC. Methods 76:87–98
79. Zhao H, Piszczek G, Schuck P (2015) SEDPHAT—a platform for global ITC analysis and global multi-method analysis of molecular interactions. Methods 76:137–148