PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Titration_DB: Storage and analysis of NMR-monitored protein pH titration curves

Damien Farrell,[1] Emanuel Sá Miranda,[1] Helen Webb,[1] Nikolaj Georgi,[1] Peter B. Crowley,[2] Lawrence P. McIntosh,[3,4,5] and Jens Erik Nielsen[1]*

[1] School of Biomolecular and Biomedical Science, Centre for Synthesis and Chemical Biology, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland

[2] School of Chemistry, National University of Ireland, Galway, University Road, Galway, Ireland

[3] Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

[4] Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

[5] Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

ABSTRACT

NMR-monitored pH titration experiments are routinely used to measure site-specific protein p*Ka* values. Accurate experimental p*Ka* values are essential in dissecting enzyme catalysis, in studying the pH-dependence of protein stability and ligand binding, in benchmarking p*Ka* prediction algorithms, and ultimately in understanding electrostatic effects in proteins. However, due to the complex ways in which pH-dependent electrostatic and structural changes manifest themselves in NMR spectra, reported apparent p*Ka* values are often dependent on the way that NMR pH-titration curves are analyzed. It is therefore important to retain the raw NMR spectroscopic data to allow for documentation and possible re-interpretation. We have constructed a database of primary NMR pH-titration data, which is accessible via a web interface. Here, we report statistics of the database contents and analyze the data with a global perspective to provide guidelines on best practice for fitting NMR titration curves. Titration_DB is available at http://enzyme.ucd.ie/Titration_DB.

Proteins 2010; 00:000–000.

Key words: protein p*Ka* values; protonation; chemical shift; database; NMR pH-titration; titration.

INTRODUCTION

Many biophysical characteristics of proteins are sensitive to pH, with the pH-dependent changes in protein stability and enzyme activity being of major importance for their function both *in vivo* and in biotechnological applications. Proteins characteristically display pH-dependent properties due to titration of amino acid side chains and termini, coupled with changes in the relative free energies of specific conformational states.[1] The catalytic activity of an enzyme, for example, is dependent on having a fraction of its population in a catalytically competent protonation state[2,3] and the pH-dependence of protein stability is a result of different titrational properties of its folded versus unfolded states.

The ionization of an isolated titratable group in a protein is characterized by the well-known sigmoid curve defined by the Henderson–Hasselbalch (HH) equation (1)

$$\mathrm{pH} = \mathrm{p}Ka + \frac{[\mathrm{A}^-]}{[\mathrm{HA}]} \quad (1)$$

The p*Ka* value of a HH titration is defined as $-\log(K_a)$, where K_a is the equilibrium constant for the acid dissociation reaction $HA + H_2O \leftrightarrows H_3O^+ + A^-$ (with molar concentrations generally replacing thermodynamic activities). The p*Ka* values of amino acids in random coil polypeptides are well known, but in the context of a folded protein, these p*Ka* values may be shifted substantially depending on local environmental factors.[4] The omnipresence and importance of titratable groups in proteins has led to a significant effort being invested in understanding the determinants of protein p*Ka* values. At the forefront of these efforts are experimental NMR-monitored pH-titration curves, which can provide site-specific insights into the ionization states of protein amino acid residues.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Science Foundation Ireland President of Ireland Young Researcher Award; Grant number: 04/YI1/M537; Grant sponsor: SFI Research Frontiers Award; Grant number: 08/RFP/BIC1140; Grant sponsor: Natural Sciences and Engineering Research Council of Canada.

*Correspondence to: Jens Erik Nielsen, School of Biomolecular and Biomedical Science, Centre for Synthesis and Chemical Biology, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: jens.nielsen@ucd.ie.

Received 27 May 2009; Revised 28 August 2009; Accepted 31 August 2009
Published online 11 September 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22611

Presently, numerous NMR-measured p*K*a values have been reported,[5] and benchmarking and optimization of structure-based p*K*a calculation algorithms on these datasets has led to a significant increase in our understanding of the factors that determine the p*K*a values of titratable groups in proteins. However, current datasets have been produced by a large number of different research groups using a variety of different NMR experiments and analysis methods. This has resulted in p*K*a values being determined with varying accuracy/precision and under a host of different conditions. Furthermore, there is no worldwide repository for the deposition of raw NMR pH-titration curves (i.e., chemical shift vs. pH) or derived p*K*a values, and there are no consistent, accepted standards on how to extract p*K*a values from NMR titration curves. This leads to problems for experimental researchers since they have to "invent" proper protocols for measuring and fitting NMR pH-titration curves, and it leads to problems for theoreticians because it is difficult to estimate the reliability of published p*K*a values without access to the original titration curves.

In this report, we present a database and a set of general guidelines aimed at remedying this situation. Titration_DB (http://enzyme.ucd.ie/titration_db) is a database of NMR monitored pH-titration curves extracted from primary literature and contributed by researchers worldwide. Titration_DB provides consistent data formats, automated fitting procedures, the ability to re-analyze NMR pH-titration curves with various models, and, most importantly, access to the primary experimental data in a simple file format. We will explain the motivation for constructing Titration_DB and discuss best practices measuring and fitting NMR titration curves, followed by an analysis of the 1929 titration curves presently in Titration_DB.

The need for a database of NMR pH-titration curves

The accurate measurement of p*K*a values are, as mentioned earlier, paramount for our understanding of protein stability,[6,7] enzyme catalysis,[8–10] protein conformational transitions,[11] and protein energetics in general.[2,12] Structure-based p*K*a calculations,[13–19] in particular, have benefited from the availability of large experimental data sets, and have developed to be able to predict p*K*a values with a somewhat satisfactory accuracy (±0.75 p*K*a units for the best methods). Significant progress in the field of protein electrostatics[20] has thus been made largely thanks to the availability of experimental data for benchmarking and calibration purposes. However, the size and quality of the experimental data sets could be improved significantly thus leading to the development of better theoretical methods. It is our hope that establishing a database of NMR monitored pH-titration curves will act as an effective interface between experimentalists and theoreticians to achieve this goal. Just as the Protein Data Bank[21] acts both as a repository and as quality filter for protein structures, we envisage Titration_DB to encourage experimentalists to adopt standard procedures when measuring and analyzing NMR pH-titration curves. This should facilitate the creation of uniform high-quality datasets of protein NMR titration curves and p*K*a values. The database will allow the benchmarking of theoretical algorithms, thus giving theoreticians a way of testing their programs against a consistent set of p*K*a values. We chose to populate Titration_DB with the NMR pH-titration curves themselves instead of simply collecting a dataset of published p*K*a values for four reasons:

1. To allow fitting of NMR pH-titration curves with various model to extract p*K*a values.
2. To allow for the analysis of the NMR titration curves to extract additional structural/electrostatic parameters.[3,6]
3. To allow for the proper representation of non-HH titration curves, which reflect coupled ionization equilibria.[2]
4. To provide statistics on the chemical shift changes observed for various NMR-detectable nuclei in proteins as a function of pH.

Measuring NMR pH-titration curves

Although the ionization equilibria of proteins can be examined by a multitude of biophysical techniques ranging from the kinetics of chemical modifications to X-ray/neutron crystallography, NMR-monitored pH-titration measurements remain the preferred method for experimentally determining the site-specific p*K*a values of individual protein residues.[22] The NMR chemical shift is an exquisitely sensitive measure of the chemical environment of the reporter nucleus, and thus can be exploited to monitor changes in the local structure, dynamics or the electric field of a specific residue in a protein. However, it is important to stress that one rarely observes an ionizable proton directly, but rather monitors nonlabile ^{1}H, ^{13}C, or ^{15}N nuclei whose chemical shifts are sensitive to the ionization state of a functional group (e.g., $^{13}C^{\gamma/\delta}$ of Asp/Glu, $^{15}N^{\delta 1/\varepsilon 2}$ and ^{1}H-$^{13}C^{\delta 2/\varepsilon 1}$ of His, $^{15}N^{\zeta}$ of Lys, $^{13}C^{\beta}$ of Cys, and $^{13}C^{\zeta}$ of Tyr). Furthermore, for reasons of spectral dispersion and sensitivity, pH titrations are typically recorded with two-dimensional NMR experiments using detection via a nonlabile ^{1}H nucleus that is assignable and ideally within a spin system close to or containing the ionizable functional group of interest. Such experiments include ^{1}H-^{13}C HSQC spectra for monitoring the titration of histidine side chains,[23,24] long range

^{1}H-^{15}N HSQC (or HMBC) spectra for determination of histidine ionization and tautomerization states,[2,3,24,25] H2(C)CO-type spectra for titrations of C-termini, aspartic acid, and glutamic acid carboxyls,[4,26,27] and H2(C)N-type spectra for N-termini, lysine and arginine side chains (although titrations of the latter, with a p*K*a > 12.5, have not been reported for proteins).[26,28] In the event that an ionizable group does not show a titration over the pH-range studied, one cannot determine its protonation state based on chemical shift information alone (except in the case of histidine, for which ^{15}N chemical shifts are highly diagnostic of charge state; see Ref. 3). Additional NMR experiments, such as the use of deuterium isotope shifts[4,29] or the measurement of $^{1}J_{NH}$ scalar couplings,[30] can help resolve this ambiguity.

Before embarking on a project to measure pH-titration curves by NMR spectroscopy, a researcher has to choose the nucleus/nuclei for tracking pH-dependent chemical shifts, the order and range of pH values to measure, and the sample buffer, ionic strength, and temperature. All of these parameters can have a significant influence on the conclusions interpretable from the measured final titration curves, and it is therefore of high importance to make informed decisions before starting an extensive set of experiments. Most important is the choice of nucleus/nuclei to be monitored. As noted earlier, nuclei closest in covalent structure to an ionizable functional group generally give the most accurate p*K*a values because their pH-dependent chemical shift changes are dominated by the titration of that moiety. Nuclei more distant (through bonds or through space) from the site of ionization may yield titration curves that also track the ionization equilibria of other residues, thus complicating the analysis. If the nuclei are part of a conjugated system or if they are connected by a very polarizable bond (as in the case of the backbone amide N—H bond), the sensitivity of their chemical shifts to the electric field in the protein can become a determining factor.[31,32] In fact, this can be exploited to measure the local dielectric constants within a protein core.[33] In general, one should track the pH-dependences of all available peaks in a given NMR spectrum since the collective information often aids in assigning fitted p*K*a values to specific titratable groups. Titration curves measured for nuclei in nonionizable residues can also give valuable information on pH-dependent conformational changes[11] and on changes in local electric field.[31,32]

It is also of crucial importance to choose a range and order of pH values that will yield an accurate definition of the titration curve. For nuclei reporting a single, ideal HH-titration, as few as three judiciously chosen pH values (corresponding to a midpoint and two end points) could in principle be used to at least obtain an estimated p*K*a value. However, as a rule of thumb, the chemical shift should be recorded at every $\sim$0.2 pH units to allow for precise and accurate fitting of the p*K*a values and for detection of multiphasic titrations, including those indicative of coupled ionization equilibria. Since protonation equilibria typically occur in the fast exchange limit, this also allows one to confidently track the pH-dependent chemical shifts of nuclei that were assigned initially at one reference pH value. Furthermore, it is important to define the starting and ending plateau chemical shifts of a titration curve accurately, and therefore one should use the widest possible range of pH values under which a protein remains soluble and folded. The choice of buffer is also important, as is the variation of the ionic strength resulting from adjustments of pH. Of course, the stability or solubility of the protein will often limit the pH range or choice of sample conditions, and available spectrometer time may limit the number of spectra that can be obtained with acceptable signal-to-noise and resolution. In these cases, one must find an acceptable trade-off between practicality and the desired level of accuracy/precision in the final fitted p*K*a values. Finally, chemical shifts should be referenced against those of an internal, inert pH-independent compound, such as 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). Note that water (and hence the lock solvent, HDO) does shows small dependences on pH (-2 ppb/pH unit from pH 2–7) and ionic strength (-9 ppb/100 m*M* salt), and a strong sensitivity to temperature (-12 ppb/$^{\circ}$C).[34]

Care must also be taken to ensure that the NMR titration curves reflect ionization events within the protein state of interest, and not alternative pH-dependent changes, such as global unfolding, oligomer/ligand dissociation, or detrimental chemical modifications. This can be accomplished with appropriate controls, such as independently determining the pH-dependent stability of a protein, recording ^{1}H-^{15}N HSQC spectra during a titration experiment to monitor for conformational changes, and measuring repeat titration points to ensure reversibility.[35]

Once the pH-titration curves have been recorded, p*K*a values can be extracted using nonlinear fitting methods, and subsequently assigned to titratable groups in the protein. In the following sections, we will discuss these tasks separately.

Fitting NMR titration curves

In the case of titration curves that show only one transition with well-defined endpoint (or plateau) chemical shifts, it is straightforward to employ the HH equation to obtain a single p*K*a value. However, in the case of multiphasic titration curves, or when one or both of the end-point chemical shifts are poorly defined, the task of extracting accurate and meaningful p*K*a values becomes difficult, if not impossible. For example, Figure 1 shows a titration curve that can be fitted with models involving 2, 3, or 4 independently titrating groups [see Eqs. (6)–(8) in Materials and Methods]. In such cases, it is impor-

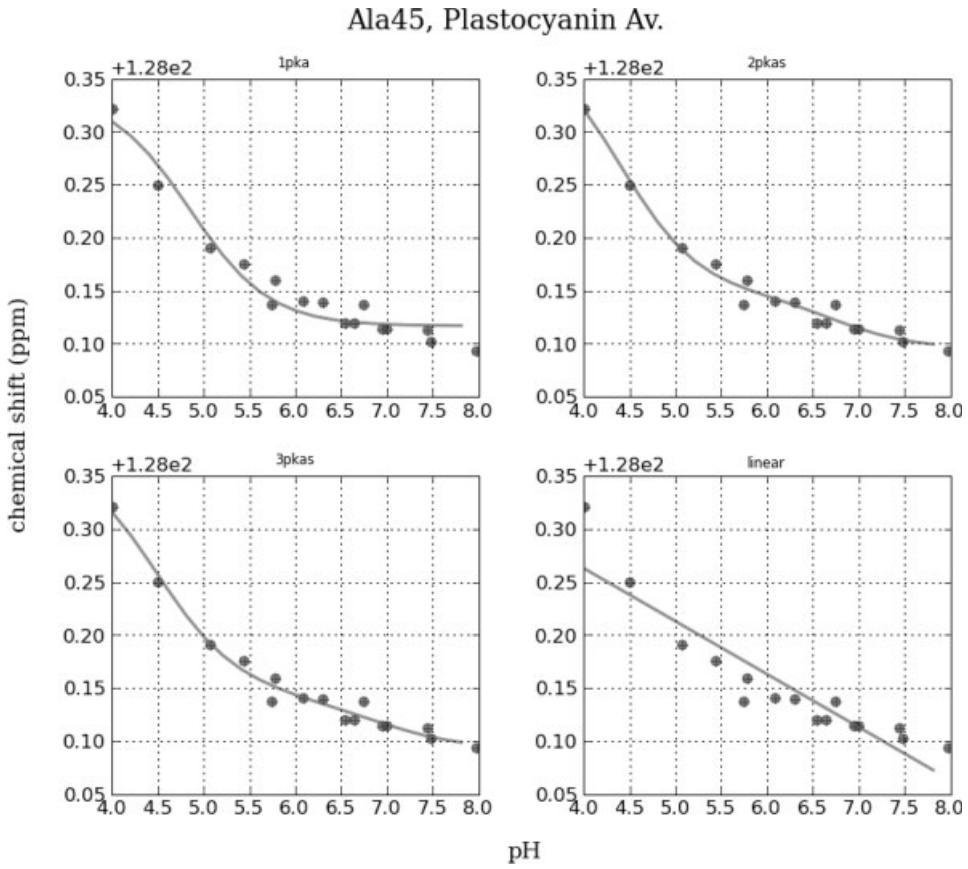

Figure 1

pH-dependent ^{15}N chemical shifts from Ala 45 of plastocyanin (A.v.). These non-HH shaped titration data can be fitted with multiple independent p*Ka* values. In this case, our procedure selects a model with 2 p*Ka*s. The obtained p*Ka* values may however be filtered out as unreliable in the later analyses.

tant to use a rigorous statistical procedure to decide how many statistically-significant p*Ka* values one can extract from a specific titration curve. In this work, we employ the *F*-test[36] because it is simple and easy to use in an automated fashion. However, it is pertinent to consider any additional available experimental information, such as p*Ka* values from independent measurements or expected chemical shift changes (sign and magnitude) based on reference random coil polypeptides, when selecting a best model. One may also simultaneously fit the titration curves of multiple nuclei, provided that they report the same ionization event.

An additional complication arises in the case of NMR titration curves for reporter nuclei that are a part of a strongly connected cluster of titratable groups. Such groups can display non-HH titrational behavior,[2,29,37] and in such cases, it is possible to fit the titration curves equally well with equations describing either independently titrating groups or coupled titrational behavior. As clearly discussed by Shrager et al.,[38] the NMR-monitored titration curves for even two coupled ionization equilibria (i.e., with four microscopic p*Ka* values) are experimentally under-determined. In such cases, there is no straightforward way to resolve this ambiguity, without additional experimental approaches, such as the use of mutants lacking combinations of the coupled residues, or possibly through fitting with a statistical-mechanical model.[2]

Assigning fitted p*Ka* values

The p*Ka* values extracted from a given NMR monitored pH-titration curve now must be assigned to titratable groups in the protein. If a curve is dominated by a single titration, and if the detected nucleus is within or adjacent to an ionizable moiety, then it is usually safe to assign the fit p*Ka* value to the specific residue containing that moiety. This conclusion is bolstered further if more than one nucleus with the same residue reports the iden-

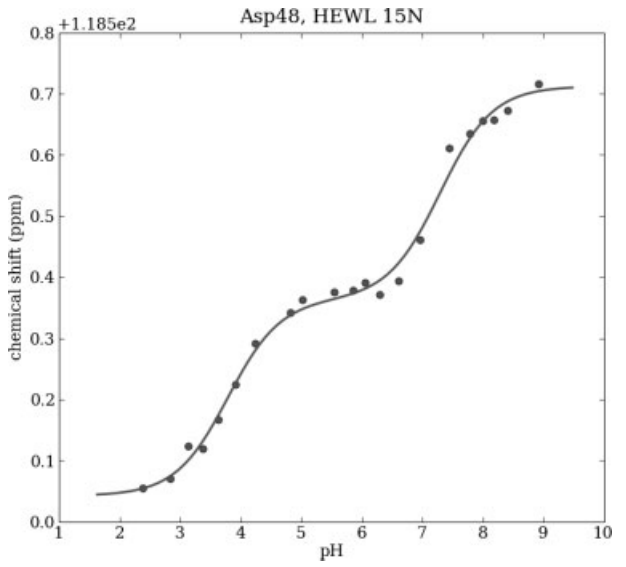

Figure 2

Asp48 $^{15}N^{H}$ titration curve in wt HEWL. The fit p*K*a values of 3.8 and 7.3 correspond to the ionization equilibria of Glu35 and Asp52, respectively.

tical titration. However, in many cases, it is not so simple to assign a fit p*K*a value to a specific ionization equilibrium, particularly if the detected nucleus is not part of a titratable residue. For example, Figure 2 shows the titration curve of the backbone $^{15}N^{H}$ of Asp48 in hen egg white lysozyme (HEWL). This curve contains two titrations with p*K*a values of 3.8 and 7.3, each displaying a chemical shift change of ~0.32 ppm. Using mutagenesis, Webb et al. (in preparation) were able to demonstrate that neither of these titrations originate from the carboxyl of Asp48 itself (which has a p*K*a value < 2.0), but instead they can be ascribed to the ionization of Asp 52 and Glu35, respectively. Similarly in *E. coli* reduced thioredoxin, Mössner et al.[39] were unable to unambiguously assign the p*K*a values extracted from the titration curves of Cys32 and Cys35 because of their multiphasic shapes and the ambiguous magnitudes of the associated chemical shift changes.

In this work, we present a database and a set of methods aimed at automating the fitting and assignment of NMR titration curves. We present statistics on the fitted p*K*a values and associated pH-dependent chemical shift changes. Experimentalists and theoreticians can use this data to guide experiments and algorithm development, respectively. The database, Titration_DB, contains the primary titration data for each NMR reporter nucleus, and each titration curve is linked directly to the appropriate atom in a protein structure file. The original literature reference (where available) and link to the structure on the PDB are also stored. Further details are given in Materials and Methods section.

RESULTS AND DISCUSSION

Titration_DB currently holds 1929 individual NMR titration curves from 97 proteins, 31 of which are mutants. NMR titration curves were digitized manually from journal publications[11,27,29,32,37,39–78] or provided by contributors in spreadsheets. All titration curves were uploaded to Titration_DB using PEAT (Farrell et al., in preparation) as described in the Materials and Methods section. We fitted all titration curves to models for one to four independent titrating groups, and the best was selected by comparing progressively more complex models using an *F*-test with a *P*-value cutoff of 5%. The majority (55%) of the titration curves in the database are measured at protons, with ^{13}C and ^{15}N nuclei accounting for the remaining 15 and 30%, respectively (Table I).

The quality and completeness of the titration curves in Titration_DB varies considerably between proteins, with 44% of proteins having five or less measured titration curves. At the other extreme, for HEWL, beta-lactoglobulin, *Phormidium laminosum* (*P.l.*) plastocyanin and *Anabaena variabilis* (*A.v.*) plastocyanin, the database holds titration curves for 100% of the amide $^{1}H^{N}$ and ^{15}N nuclei. Only 927/1929 (48%) of the titration curves are recorded for atoms residing in titratable groups. This number changes to 705/840 (84%) if only the incomplete datasets are considered. The majority of titration curves originate from histidine (15%), aspartate (14%), and glutamate (13%) residues, with lysine (5%) accounting for a smaller fraction of experiments.

In total, the database contains 30,646 data points (a data point being a pH and corresponding chemical shift value). The average number of data points per titration curve is 16 (SD 5.4), the average span of the pH range is 5.1 (SD 1.3), and the average lowest pH value is 3.0 (SD 1.2). The average spacing between pH points is 0.3, although 40% of curves have an average pH spacing of less than 0.3. Only three titration curves were measured with an average pH spacing less than 0.2 pH units.

In the following, we analyze the contents of Titration_DB to extract information on:

1. The p*K*a values of titratable groups.
2. Chemical shift distributions per residue/nucleus.
3. Correlation of pH-dependent amide $^{1}H^{N}$-$^{15}N^{H}$ chemical shifts for four complete datasets.

Table I

Summary of Current Dataset and Fitting Models

	One p*K*a, two shifts	Two p*K*as, three shifts	Three p*K*as, four shifts	None (linear)	Total
^{1}H	497	346	45	169	1057
^{15}N	182	231	66	104	583
^{13}C	123	83	32	49	287
				Total	1927

pKa values of titratable groups

The main point of analyzing NMR monitored protein pH-titration curves is to extract and assign accurate p*K*a values. To automatically fit the titration curves to one or more p*K*a values, we apply the *F*-test-based approach outlined in the Materials and Methods section. To assign the fitted p*K*a values to titratable groups, we test the extracted p*K*a values against a number of criteria (see later). In this work, we have limited ourselves to analyzing the 908 titration curves measured at nuclei in titratable residues. A first step in the analysis was to filter the p*K*a values by a set of automatable reliability criteria. A similar approach has been adopted in a previous study.[42] We have done this by defining two terms: "reliable p*K*a value" and "primary p*K*a value."

A reliable p*K*a is defined with the following criteria for any curve fit to a model with one or more p*K*a values:

$$\Delta\delta_r \geq \text{cutoff}; \quad \frac{\Delta\delta_r}{\Delta\delta_{other}} \geq 1.2 \tag{2}$$

$$pKa_r - pKa_{other} \geq 0.2 \tag{3}$$

$$pKa_r - pKa_{model} \leq 3 \tag{4}$$

$$pKa_r - pH_{min} \geq 0.3; \quad pH_{max} - pKa_r \geq 0.3 \tag{5}$$

where pKa_r is the reliable p*K*a value, $\Delta\delta_r$ is the value of the chemical shift change associated with the p*K*a value and must be greater than the cutoff value for each nucleus. $\Delta\delta_{other}$ are the chemical shifts associated with the other p*K*a values in the same titration curve. pKa_{model} is the model p*K*a for the titratable residue considered, and pH_{min} and pH_{max} are the upper and lower end of the range of pH data.

In summary, reliable p*K*a values are associated with the largest chemical shift changes in a titration curve. Furthermore, a reliable p*K*a value must be at least 0.2 pH units distant from the other p*K*a values fitted to the titration curve, must not be shifted more than three units from the model p*K*a[79] value of the amino acid, and must be at least 0.3 pH units from the pH value of the terminal data points. The criteria specified by Eqs. (3) and (5) serve to filter out improperly fitted titration curves, Eqs. (2) and (5) are criteria that have been "conservatively" chosen to produce a set of reliable p*K*a values that contain very few false positives since this is important for producing reliable statistics (see later). When manually fitting NMR titration curves, the criteria specified above can of course be overruled if human judgment (i.e., consideration of additional experimental or theoretical data) produces a better global interpretation of the data.

We define *primary* p*K*a values simply as the subset of the reliable p*K*a values that originate from titration curves with only one titration. A primary p*K*a value represents the most reliable p*K*a value measurement for a titratable group since no other titrations influences the extraction of a p*K*a value from the NMR titration curve.

Further comments on criteria for selecting "reliable" p*K*a values:

- *Equation 2:* This is the sole p*K*a value associated with a chemical shift change ($\Delta\delta$) greater than the cutoff or, for a multiple p*K*a model, the p*K*a corresponding to the largest chemical shift change. $\Delta\delta$ cutoff values were chosen as being the lower limit of a recognizable titration for a given nucleus. This figure in part reflects the resolution of typical NMR spectra and hence the limitations in chemical shift measurements. Cutoff values of 0.08 ppm[42] and 0.07 ppm[80] have been previously suggested for ^{1}H. Although these values may vary with the sample (i.e., line widths) and NMR protocols, for the sake of automation, our chosen values for cutoff were 0.06 ppm for ^{1}H, 0.3 ppm for ^{15}N, and 0.2 ppm for ^{13}C.
  For a multiple p*K*a model, the associated $\Delta\delta$ is greater than the others by an arbitrary factor of 1.2. The rationale is that the p*K*a would then represent the primary titrational event detected for that residue.
- *Equation 3:* For a multiple p*K*a model, the p*K*a value is further than 0.2 units from any another p*K*a, so that they can reasonably be considered as distinct events.
- *Equation 4:* The p*K*a value is within 3 pH units of the unperturbed (model) p*K*a, and therefore represents a realistic value. However, residues with highly perturbed p*K*a values are indeed found in proteins and these omitted values can be flagged for later analysis. The model p*K*a values used here are those provided by Grimsley and coworkers[79] measured using alanine pentapeptides. These are very close to other standard values previously published[81,82] and the particular choice of values does not affect this filtering step.
- *Equation 5:* The p*K*a value is not too close to edge of the available range of sample pH conditions. This is important because the algorithm will attempt to fit a discontinuous titration, where the data points do not sufficiently define the entire event. This generates incorrectly large $\Delta\delta$ values.

In the present dataset, we analyze 908 titration curves and fit 2276 p*K*a values. In this set of p*K*a values, we identify 517 reliable p*K*a values with 343 of these being primary p*K*a values. The number of p*K*a values found, per residue type, along with their average values are shown in Table II. Supporting Information Tables 1–3 show a detailed list of all primary p*K*a values with the specification of protein name in the DB, protein residue number, atom name, and fitting error. These tables may be generated dynamically via the web interface at any time and will reflect the current data in the Titration_DB.

Table II
Average Extracted Δδ and Associated Average p*Ka* Values with Errors Broken Down by Atom/Nuclei for Titratable Residues

	^{1}H						^{15}N						^{13}C					
Residue	Atom	Avg. Δδ	Avg. Δδ err	Avg. p*Ka*	Avg. p*Ka* error	No. curves	Atom	Avg. Δδ	Avg. Δδ err	Avg. p*Ka*	Avg. p*Ka* error	No. curves	Atom	Avg. Δδ	Avg. Δδ err	Avg. p*Ka*	Avg. p*Ka* error	No. curves
His	$H^{\varepsilon 1}$	0.86	0.02	6.5	0.03	54	N	2.86	0.06	5.7	0.04	5	C^{γ}	2.6	0.08	3.8	0.05	2
	H	0.82	0.02	6.1	0.04	12	$N^{\varepsilon 2}$	2.78	0.06	6	0.04	3	$C^{\varepsilon 1}$	2.15	0.07	6	0.07	18
	$H^{\beta *}$	0.66	0.01	6.6	0.03	6												
	$H^{\delta 2}$	0.33	0.01	6.6	0.07	23												
	$H^{\delta 1}$	0.3	0.01	5.1	0.04	2												
Asp	H	0.32	0.01	3.3	0.08	22	N	1.32	0.12	4.3	0.25	4	C^{γ}	3.12	0.07	3.5	0.04	24
	$H^{\beta *}$	0.25	0.01	3.8	0.08	28							C^{β}	2.3	0.04	4.2	0.04	3
	H^{α}	0.12	0.01	3.2	0.12	5							C^{α}	1.43	0.06	8.5	0.07	2
Glu	H	0.51	0.01	4	0.07	28	N	0.67	0.02	5.7	0.09	4	C^{δ}	3.74	0.07	4.1	0.07	27
	$H^{\gamma *}$	0.24	0.01	4.3	0.07	33												
	$H^{\beta *}$	0.22	0.01	3.7	0.07	3												
	H^{α}	0.07	0.01	2.8	0.33	2												
Lys	H	0.41	0.02	8.8	0.07	13	N	0.65	0.03	5	0.14	7	C^{ε}	1.27	0.08	10.8	0.1	11

Nuclei for which there is one curve only are omitted. $H^{\beta *}$ in Asp refers to both $H^{\beta 2}/H^{\beta 3}$ because most of the reported shifts are not usually stereospecifically assigned. The same applies for $H^{\gamma *}$ in Glu.

Precision of fitted p*Ka* values

The precision of the fitted p*Ka* values due to experimental uncertainty was assessed using a Monte Carlo approach. The experimentally measured points on the titration curves were perturbed with estimated random errors of ±0.1 for the pH values and the following for chemical shift: ^{1}H, ±0.03 ppm[83]; ^{15}N, ±0.1 ppm[32]; and ^{13}C, ±0.2 ppm). The fitting algorithm was rerun 50 times using the randomly perturbed values to obtain the standard deviation of the fit p*Ka* values. Note that, unlike most cases of error analysis for which the independent variable is assumed exact and only the dependent variable erroneous, in the case of NMR-monitored titrations, the sample pH measurements are likely to contain more systematic and random errors than the monitored chemical shifts.

Average fitted p*Ka* values

A comparison between our average fitted p*Ka* values for four ionizable groups (^{1}H dataset) to the recent work summarizing 541 experimentally derived p*Ka*s in folded proteins[84] is shown in Table III. All show close agreement, apart from lysine, for which we have excluded more than half the curves using our filtering criteria and therefore have only 14 p*Ka*s available for the average value. This indicates the much greater difficulty in fitting the lysine data (note the large standard deviation, 3.3 of the 14 p*Ka* values we were able to fit).

Distribution of $\Delta\delta_{pH}$

For each type of residue in a protein, the chemical shifts and pH-dependent shift changes measured at a given nucleus are, to a large extent, characteristic of the corresponding amino acid. Thus, the observable chemical shifts and chemical shift changes are generally found to be within restricted ranges. Here, we analyze the data in Titration_DB to establish the magnitude of chemical shift changes that are observed when a specific amino acid titrates. Such values are useful in the interpretation of NMR pH-titration curves to establish the origin of an observed titration (i.e., this will aid in assigning the observed p*Ka* value to a titratable group).

^{1}H chemical shift changes

The relatively larger data set for ^{1}H provides the most complete distribution and therefore gives the best impression of the range of observable chemical shift changes. The distribution of ^{1}H pH-dependent Δδ values for Asp $^{1}H^{N}/^{1}H^{\beta *}$ nuclei and Glu $^{1}H^{N}/^{1}H^{\gamma *}$ nuclei generally overlap and are found almost entirely in the 1–1.5 ppm range [Fig. 3(A,B)]. In contrast, the Δδ of histidine $^{1}H^{\delta 2}$ and $^{1}H^{\varepsilon 1}$ are clustered in two distinct distributions of 0–0.6 ppm and 0.6–1.2 ppm, respectively (Supporting Information Figure 1). There is insufficient data on the remaining nuclei to form a clear picture of the ranges of Δδ. This plot, and those for ^{15}N and ^{13}C data, can be generated in the web interface.

Table III
Average Fitted p*Ka*s for Four Tiratable Groups, ^{1}H Data Only

	Model p*Ka*[a]	Protein p*Ka*[b]	Average ^{1}H fitted p*Ka*[c]	No. of curves fitted
Asp	3.9	3.5 ± 1.2	3.4 ± 1.1	79
Glu	4.3	4.2 ± 0.9	4.2 ± 1.0	88
His	6.5	6.6 ± 1.0	6.4 ± 0.9	150
Lys	10.4	10.5 ± 1.1	9.0 ± 3.3	14

[a]From Ref. 79.
[b]From Ref. 84.
[c]Using all reliable p*Ka* values in this study.

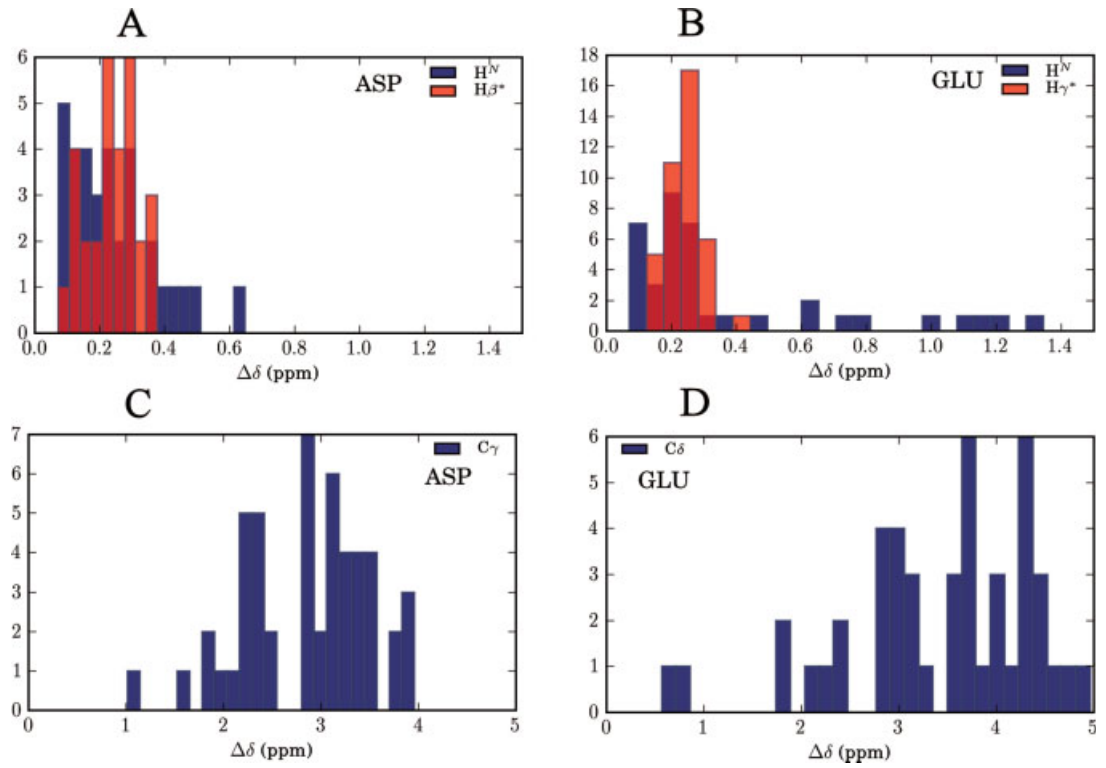

Figure 3

Distributions of Δδ associated with all reliable p*K*a values for selected ^{1}H Asp/Glu and ^{13}C Asp/Glu nuclei. (**A**) H^N and $H^{\beta*}$ nuclei in ^{1}H Asp, (**B**) H^N and $H^{\gamma*}$ nuclei in ^{1}H Glu, (**C**) C^{γ} nuclei for Asp ^{13}C data, and (**D**)^{13}C Glu C^{δ} nuclei. $H^{\beta*}$ and $H^{\gamma*}$ in Asp/Glu denote undifferentiated $H^{\beta2}/H^{\beta3}$ and $H^{\gamma2}/H^{\gamma3}$ assignments.

^{15}N and ^{13}C chemical shift changes

The distribution of Δδ for Asp $^{13}C^{\gamma}$ nuclei and Glu $^{13}C^{\delta}$ nuclei display similar ranges, with Asp being confined to a somewhat narrower range of 1–4 ppm [Fig. 3(C,D)]. Unfortunately, there is a relative lack of data for the pH-dependent ^{15}N and ^{13}C chemical shift changes in Titration_DB, and the histograms for these nuclei should therefore be considered very incomplete distributions. Similar plots can be generated dynamically from the web interface to provide statistics on any nucleus/amino acid pair, and thus allows for the capture of any future data that is added to Titration_DB.

Correlation between p*K*a values extracted from amide $^{1}H^{N}$ and $^{15}N^{H}$ nuclei

Because of the relative ease of assigning amide $^{1}H^{N}$ and ^{15}N chemical shifts, and of using ^{1}H-^{15}N HSQC spectra to monitor pH-titrations, several investigations have yielded experimental data of $^{1}H^{N}$ and $^{15}N^{H}$ chemical shifts versus pH for essentially every residue in a given protein. We have thus far retrieved complete amide $^{1}H^{N}$ and $^{15}N^{H}$ data sets for four proteins: Plastocyanin *A.v.*,[62] plastocyanin *P.l.*,[85] bovine beta lactoglobulin (βLG)[11] and HEWL.

Amides do not titrate over any accessible pH range; rather, their nuclei report pH-dependent changes due for example to inductive, electrostatic, or structural perturbations. Chemical shift changes measured at the backbone amide ^{15}N and $^{1}H^{N}$ nuclei of the same residue should, to some extent, reflect the same titrational event(s) on their ionizable side chains. However, this is not always the case due to the complicated propagation of the electrostatic field and the complex nature of structural relaxation. We therefore decided to investigate if p*K*a values extracted from $^{15}N^{H}$ and $^{1}H^{N}$ titration curves are indeed identical. Figure 4 shows plots of ^{15}N p*K*a values versus $^{1}H^{N}$ p*K*a values for the titratable residues in the four proteins with complete datasets. We observe a reasonable correlation (correlation coefficient = 0.86) between the reliable p*K*a values [i.e., the p*K*a values associated with the largest chemical shift change within a titration curve; Fig. 4(A)], thus demonstrating that if a reliable p*K*a value is found in both curves, then those p*K*a values are likely to be similar.

We also investigate the cases where a reliable p*K*a value is found in only one of the titration curves (i.e., either in

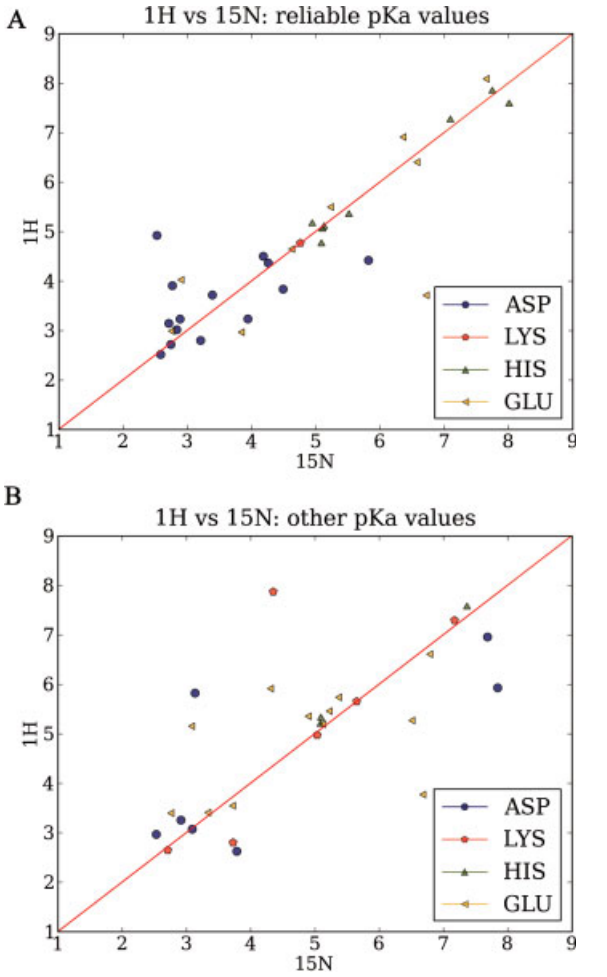

Figure 4

p*Ka* values for fits of pH-dependent amide ^{15}N versus $^{1}H^{N}$ chemical shifts for His, Asp, Glu, and Lys residues from four complete datasets. (**A**) shows the correlations between all reliable ^{15}N and $^{1}H^{N}$ p*Ka* values from the four proteins. The outliers represent genuine differences in ^{15}N versus $^{1}H^{N}$ curves. (**B**) illustrates why the p*Ka*s have been filtered for reliability. This plot includes points where either one of the ^{15}N versus $^{1}H^{N}$ p*Ka*s is not considered reliable and hence the large amount of scatter—we are not comparing two reliable p*Ka*s.

the $^{1}H^{N}$ or in the $^{15}N^{H}$ titration curve), by plotting the reliable p*Ka* value against the closest p*Ka* value in the titration curve without a reliable p*Ka* value [Fig. 4(B)]. The lack of an exact correlation between many ^{15}N and $^{1}H^{N}$ p*Ka* values in this plot demonstrates that directly bonded $^{15}N^{H}$ and $^{1}H^{N}$ nuclei do not always report the same ionization event, thus making it difficult to assign fit p*Ka* values to specific ionization equilibria.

The data points in Figure 4(A) may be broadly placed in three categories:

1. Closely correlated p*Ka* values for the $^{15}N^{H}$ and $^{1}H^{N}$ of a given amide, with essentially identical chemical shift pH dependency.
2. Correlated p*Ka* values, but with a Δδ of opposite sign.
3. Data from titration curves that cannot be fit to matching p*Ka* values.

In cases (i) and (ii), the bonded nuclei almost certainly reflect the same titrational event and we can extract the same p*Ka*(s), whereas the p*Ka* values belonging to case (iii) correspond to outliers in the plot where the two nuclei do not report the same titrational event. We have excluded nuclei for which there is no discernable Δδ in either nucleus.

To illustrate the details of the differences in $^{1}H^{N}$ and $^{15}N^{H}$ curves, we examine a number of titration curves of titratable residues. We examine two residues in plastocyanin because we have complete datasets for the closely related plastocyanins from *Anabaena variabilis* and *Phormidium laminosum*. Furthermore, we analyze the titration curves for Asp64 and Glu114 from β-lactoglobulin.

The titration curves for His61 in plastocyanin [Fig. 5(A,B)] are almost identical in the two species, however, the titration curves vary significantly depending on the type of nucleus. The proton titration curves are bell-shaped, whereas the nitrogen titration curve displays two titrations that both increase the chemical shift of the nucleus. The $^{15}N^{H}$ and $^{1}H^{N}$ titration curves of His61 are therefore good examples of case (ii) behavior, with the p*Ka* values being similar but where the second titrations have opposite sign of their associated chemical shift changes. It has previously been established that His61[62] occupies two different conformations with different hydrogen-bonding networks. In addition, the chemical shift is perturbed by interactions with the copper ion (9 Å away) and/or His92, which has a p*Ka* of 5.1 and 5.0 in plastocyanin *A.v.* and *P.l.*, respectively [Fig. 5(C,D)]. The titration with a low p*Ka* value in the titration curve of His61 [Fig. 5(A,B)] thus reflects the deprotonation of His92, whereas the higher p*Ka* value reflects the titration of His61 itself. Protonation of His92 is a well-studied attribute of the plastocyanin active site.[86,87] On protonation, His92 dissociates from the copper and the site rearranges to result in increased interactions between the copper and the remaining three ligands (His, Cys, and Met). The consequent structural and electrostatic changes are likely to be sensed by nuclei in the vicinity. For example, as seen earlier, the His61 nuclei may be sensitive to pH dependent changes in the active site as well as protonation of the His61 side chain.

Glu114 from βLG shows a $^{15}N^{H}$ curve with two Δδ (0.2, 0.45 ppm) with p*Ka* values of 3.3 and 6.7 in comparison with the $^{1}H^{N}$ curve, which is found to be monophasic by the *F*-test procedure and produces a single p*Ka* value of 3.7. (This p*Ka* value is actually close to 3.3, but since we are comparing reliable p*Ka*s, the point appears as an outlier). Glu114 is thus an example of case (ii) scenario where a p*Ka* value in one titration curve is observed in the other titration curve, but where there is multiple p*Ka*s in either curve [see Fig. 5(E)].

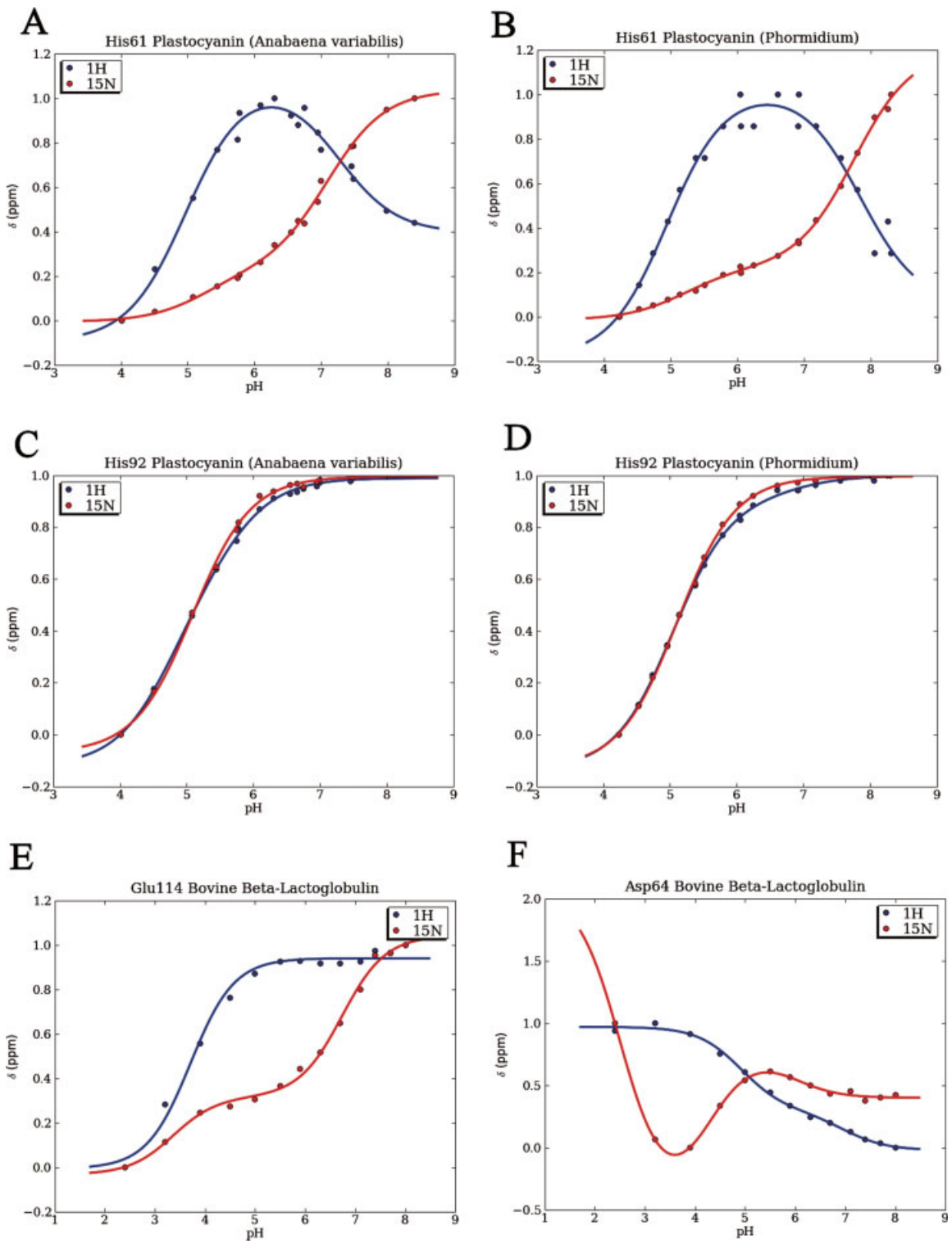

Figure 5

Selected corresponding titration curves of the pH-dependent $^{15}N^{H}$ and $^{1}H^{N}$ chemical shifts of the same residue. (**E**) and (**F**) are outliers in Figure 4(a). (**C**) and (**D**) are the curves for His92 from both plastocyanin from *Phormidium laminosum* (P.l.) and *Anabaena variabilis* (A.v.). (**A**) His61, A.v. plastocyanin; (**B**) His61, P.l. plastocyanin; (C) His92, A.v. plastocyanin; (D) His92, P.l. plastocyanin; (E) GLU114, βLG; and (F) ASP64, βLG.

Asp64 $^{15}N^{H}$ was found to be fit with three p*K*as (2.5, 4.2, 5.9) in contrast to the two p*K*as of 4.9 and 6.9 found in the $^{1}H^{N}$ fit. Both of these curves are highly perturbed from a HH shape. Curves for Asp64 are shown in Figure 5(F).

Note that one can combine $^{1}H^{N}$ and $^{15}N^{H}$ chemical shift changes, using an expression of the form

$$((\delta_{1H,pH} - \delta_{1H,ref})^2 + w(\delta_{15N,pH} - \delta_{15N,ref})^2)^{1/2}$$

(where w is a weighting factor and the reference chemical shifts are at a clear endpoint to avoid potential problems due to the loss of sign information).[88] Alternatively, one could fit the uncombined $^{1}H^{N}$ and $^{15}N^{H}$ titrations simultaneously to obtain common p*K*a values. Titration curves analyzed in this way might have an improved signal/noise ratio due to a larger range of pH-dependent chemical shift changes. However, the aforementioned analysis shows that $^{1}H^{N}$ and $^{15}N^{H}$ signals sometimes track different ionization events and that one of the curves may be incomplete/noisy. Thus combining or co-fitting curves might actually decrease the precision of the extracted p*K*a values. Furthermore, combining titration curves for residues with behavior similar to His61 in plastocyanin can lead to situations where one or more titrations cancel each other. Thus, one should always carefully inspect the two titration curves separately before combining the two into a single curve using the equation mentioned earlier.

The modified Hill equation

Finally, it should be mentioned that NMR titration curves often are fitted to the modified Hill equation

$$\delta_{obs} = \delta_p + \frac{\Delta\delta}{1 + 10^{n(pKa - pH)}},$$

where n is the Hill parameter, δ_{obs} is the observed chemical shift, δ_p is the chemical shift of the fully protonated form, and $\Delta\delta$ is the chemical shift change associated with the titration governed by the p*K*a value. In this study, we do not advocate the use of the Hill equation because a fit producing a Hill value different from 1 (i.e., a non-HH titration curve) indicates that the titration of the group is coupled to other titrational or structural events in the protein. In these cases, we therefore believe it appropriate to fit the titration curve using a model that explicitly accounts for these effects instead of "hiding" the effects in the Hill parameter. Furthermore, it is important to realize that straight lines can be fit very well with the modified Hill equation when the Hill parameter is low enough ($\leq$0.5). In these cases, it is questionable if the observed change in the chemical shift represents a real titrational event, or if the data essentially is noise or some inexplicable combination of titrational and structural events. In this article, we remove titration curves if the *F*-test gives the best fit to the linear model. Using the modified Hill equation instead of the linear model in the *F*-test framework leads to a large number of titration curves being fit best by the modified Hill equation. Thus, the Hill equation fits many nontitrations (straight lines) well, and additionally fits many curves with two or more titrational events, thus potentially causing a loss of valuable experimental information. In summary, we discourage the use of the modified Hill equation because it does not provide an explicit description of the physics underlying the observed deviation from the HH equation.

CONCLUSIONS

We have created a repository of NMR-monitored pH titration data for a large number of proteins, and made the raw NMR data available for download. The statistics and analyses presented in this article will serve as useful guidelines for interpreting NMR-monitored titration curves, and we have furthermore illustrated several aspects that are of importance when p*K*a values are extracted from such curves. All data that could be retrieved has been included in the DB and any omitted data will be gratefully accepted. We will continue to populate Titration_DB with experimental data, and we encourage experimental researchers to submit the current and newest raw NMR data to us using the contact information on the website. In summary, we hope that the repository will strengthen the interest in measuring p*K*a values using NMR, and thus lead to an improved understanding of pH-dependent effects in proteins and protein electrostatics in general.

MATERIALS AND METHODS

Data collection and storage

The data was retrieved either directly from the authors or by reconstructing published graphs using g3data.[89] In either case, data points are stored into text format and imported into Protein Engineering Analysis Tool (PEAT) (Farrell et al., in preparation), which contains a specialized component used to display and fit NMR titration curves. PEAT is written in the python (http://www.python.org) programming language and stores the datasets internally as a python specific data type (dictionary). The entire dataset is written to plain text files that form the basis of the PEAT project file. On the server side, the project file is kept on a subversion[90] server and when new data is uploaded, the versioning system creates a new "revision" of the database. This client-server system allows anyone with the PEAT client to download (i.e., "checking out" in subversion terminology) the latest version of the database and apply the same analysis and fitting tools that we have used. One consequence of this approach is that a record is kept of all the previous versions of the datasets at each revision and this log can also be viewed online. New records can be added to the database by anyone with write access in the same way. The PEAT software is available free for academic use at http://enzyme.ucd.ie/PEAT.

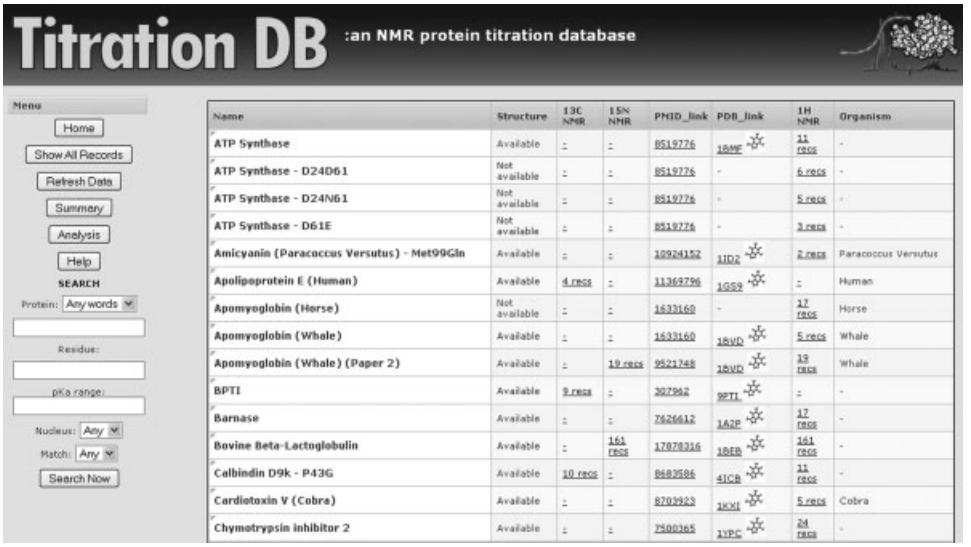

Figure 6

The Titration_DB web interface.

The following fields are stored for each record where available: Protein name, one column for each of the NMR nuclei (^{1}H, ^{15}N, and ^{13}C), link to original article, PDB structure coordinate file, and links to the PDB and organism name.

The following formatting was applied to the data:

- Protein names are supplied in the following form: name (organism)—mutation, for example, FNfn10 (Human)—D7K.
- PDB structure coordinate files for each protein or mutant are required for the purpose of our analysis. X-ray crystallography structures at the best resolution possible have been found for the majority of proteins in the database. In the cases, where data is available for mutants of a protein, the WHAT IF software[91] was used to model the mutated structure from the published wild type. For each record, the structures are stored along with the pH titration data.
- Assigned fields are provided for residue names and atoms in each dataset, allowing search algorithms and quick location of residue in the structure during analysis.
- Experimental conditions, when available, have been stored with each set of curves.

Fitting and models

Titration data was fit using an implementation of the Levenberg–Marquardt algorithm.[92] The choice of the appropriate number of apparent titration events (p*Ka* values) required to fit the experimental data to a modified HH equation was determined primarily by the *F*-test. An *F*-test is a standard statistical tool for comparing the fits of a given data set to two models with differing degrees of freedom.[93] This has previously been used for fitting of HH curves.[43] We use an iterative procedure, fitting a curve in turn with a linear fit and then with progressively more complex models (Models 1–3). For each model:

1. The sum of squares errors for current and next models are compared: if the next more complex model has a higher error (this is unusual, but can happen when the data are scattered around a straight line in a way that produces a lower sum of squared differences than the more complex model), stop and accept the current model; else proceed to step 2.
2. Perform *F*-test: if the *F*-test returns a *P*-value outside the confidence level of 95%, the more complex model is accepted.
3. The new model becomes the current model and we return to step 1 until the statistically best model is found.

If a linear fit of chemical shift versus pH is the best found, then the data are considered to be a nontitration. The results obtained depend on the *F*-test confidence level used (95% is typical) and it must be stressed that some fits will always be open to interpretation. In a small number of cases, curves were re-fit by hand and the best model selected visually, but generally the fits from this global method were accepted. This fitting procedure is implemented in the titration analysis module of our PEAT software.

Fitting models:

1. A single p*Ka* value, two chemical shifts:

$$\delta_{obs} = \delta_p + \frac{\Delta\delta_1}{1 + 10^{[pKa_1 - pH]}} \quad (6)$$

2. Two p*Ka* values, three chemical shifts:

$$\delta_{obs} = \delta_p + \frac{\Delta\delta_1}{1 + 10^{[pKa_1 - pH]}} + \frac{\Delta\delta_2}{1 + 10^{[pKa_2 - pH]}} \quad (7)$$

3. Three p*Ka* values, four chemical shifts:

$$\delta_{obs} = \delta_p + \frac{\Delta\delta_1}{1 + 10^{[pKa_1 - pH]}} + \frac{\Delta\delta_2}{1 + 10^{[pKa_2 - pH]}} + \frac{\Delta\delta_3}{1 + 10^{[pKa_3 - pH]}} \quad (8)$$

where δ_{obs} is the observed chemical shift at each pH, is the δ_p is the chemical shift of the fully protonated state with respect to the titration(s), and $\Delta\delta_i$ is the chemical shift change associated with the titration governed by pKa_i.

Web interface

The web interface is a CGI front end to our custom software for storage and analysis (Fig. 6). The primary purpose of which is to allow convenient access to experimental NMR titration data, with search functionality, links to structure, visualization of curves, and download links to data. The web interface is simply a front end using the same codebase that we have written to perform the analyses discussed previously. Thus, we can add analysis components to the web interface relatively easily. The data is publicly accessible and no software is required to view and use the web interface apart from a web browser. The current version of the web interface is available at http://enzyme.ucd.ie/titration_db.

The interface allows the user to browse all the available records and perform searches. Searches can be currently performed by protein name/key word, residue, and p*Ka* range. When an individual residue (or sets of residues within one record) is selected, the web server will dynamically generate graphs from the stored data points and our fit data. Current fit model and p*Ka* values will also be displayed for each residue. Any combination of curves can be over-laid and displayed on a single graph or shown as separate graphs. The data may be normalized for easy visual comparison.

The data may currently be searched by:

- The p*Ka* values (or range of p*Ka* values) occurring as either single or double titrations.
- The residue type and atom type.
- The shape of titration curve (i.e., the model used and fit details).

Additional search parameters to be added:

- The environment of titratable group (surface accessibility, position in structure (helix, sheet, turn, coil), B-factor, density of charged groups, and crystal contacts).

Other functionality:

- Cross-reference with calculated p*Ka* values in pKD server.[94]
- Load selected curves into p*Ka* tool.[95]
- Ability to download experimental data and fit information for selected datasets as a text file.

ACKNOWLEDGMENTS

The authors thank Predrag Kukic and Kaare Teilum for insightful discussions. They also thank the following people, in particular, for their contribution of data to Titration_DB: J.J. Led, Matthias Hass, Sara Linse, Kazumasa Sakurai, Barbara Tynan-Connolly, Elisa Nogueira, and Carolyn Fitch. Chresten Søndergaard is gratefully acknowledged for implementing the Levenberg–Marquardt fitting algorithm, Tommy Carstensen for help with F-test statistics, and Michael Johnston for discussions on the web interface.

REFERENCES

1. Tanford C. Protein denaturation. C Theoretical models for the mechanism of denaturation. Adv Protein Chem 1970;24:1–95.
2. Søndergaard CR, McIntosh LP, Pollastri G, Nielsen JE. Determination of electrostatic interaction energies and protonation state populations in enzyme active sites. J Mol Biol 2008;376:269–287.
3. Nielsen JE. Analyzing enzymatic pH activity profiles and protein titration curves using structure-based pKa calculations and titration curve fitting. Methods Enzymol 2009;454:233–258.
4. Yamazaki T, Lee W, Revington M, Mattiello DL, Dahlquist FW, Arrowsmith CH, Kay LE. An HNCA pulse scheme for the backbone assignment of 15N,13C,2H-labeled proteins: application to a 37-kDa Trp repressor-DNA complex. J Am Chem Soc 2002;116:6464–6465.
5. Toseland CP, McSparron H, Davies MN, Flower DR. PPD v1.0—an integrated, web-accessible database of experimentally determined protein pKa values. Nucleic Acids Res 2006;34(database issue): D199–D203.
6. Pace CN, Alston RW, Shaw KL. Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. Protein Sci 2000;9:1395–1398.
7. Garcia-Moreno B, Chen LX, March KL, Gurd RS, Gurd FR. Electrostatic interactions in sperm whale myoglobin. Site specificity, roles in structural elements, and external electrostatic potential distributions. J Biol Chem 1985;260:14070–14082.
8. Fersht A. Enzyme structure and mechanism, Vol. 21. New York: W.H. Freeman; 1985. 475 p.
9. Joshi MD, Sidhu G, Nielsen JE, Brayer GD, Withers SG, McIntosh LP. Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. Biochemistry 2001;40:10115–10139.
10. Nielsen JE, McCammon JA. Calculating pKa values in enzyme active sites. Protein Sci 2003;12:1894–1901.
11. Sakurai K, Goto Y. Principal component analysis of the pH-dependent conformational transitions of bovine beta-lactoglobulin monitored by heteronuclear NMR. Proc Natl Acad Sci USA 2007;104: 15346–15351.
12. Warshel A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. J Biol Chem 1998;273: 27035–27038.
13. Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B. On the calculation of pKas in proteins. Proteins 1993;15:252–265.

14. Nielsen JE, Vriend G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. Proteins 2001;43:403–412.
15. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238:415–436.
16. Mehler EL, Guarnieri F. A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. Biophys J 1999;77:3–22.
17. van Vlijmen HW, Schaefer M, Karplus M. Improving the accuracy of protein pKa calculations: conformational averaging versus the average structure. Proteins 1998;33:145–158.
18. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. Proteins 2005;61:704–721.
19. Alexov EG, Gunner MR. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. Biophys J 1997;72:2075–2093.
20. Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. Biochim Biophys Acta 2006;1764:1647–1676.
21. Berman H, Henrick K, Nakamura H. Announcing the worldwide protein data bank. Nat Struct Biol 2003;10:980.
22. Popov K, Rönkkömäki H, Lajunen LHJ. Guidelines for NMR measurements for determination of high and low pKa values (IUPAC technical report). Pure Appl Chem 2006;78:663–675.
23. Yu L, Fesik SW. pH titration of the histidine residues of cyclophilin and FK506 binding protein in the absence and presence of immunosuppressant ligands. Biochim Biophys Acta 1994;1209:24–32.
24. Schubert M, Poon DK, Wicki J, Tarling CA, Kwan EM, Nielsen JE, Withers SG, McIntosh LP. Probing electrostatic interactions along the reaction pathway of a glycoside hydrolase: histidine characterization by NMR spectroscopy. Biochemistry 2007;46: 7383–7395.
25. Pelton JG, Torchia DA, Meadow ND, Roseman S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. Protein Sci 1993;2:543–558.
26. Yamazaki T, Yoshida M, Nagayama K. Complete assignments of magnetic resonances of ribonuclease H from *Escherichia coli* by double- and triple-resonance 2D and 3D NMR spectroscopies. Biochemistry 1993;32:5656–5669.
27. Oda Y, Yamazaki T, Nagayama K, Kanaya S, Kuroda Y, Nakamura H. Individual ionization constants of all the carboxyl groups in ribonuclease HI from *Escherichia coli* determined by NMR. Biochemistry 1994;33:5275–5284.
28. Andre I, Linse S, Mulder FA. Residue-specific pKa determination of lysine and arginine side chains by indirect ^{15}N and ^{13}C NMR spectroscopy: application to apo calmodulin. J Am Chem Soc 2007; 129:15805–15813.
29. Joshi MD, Hedberg A, McIntosh LP. Complete measurement of the pKa values of the carboxyl and imidazole groups in Bacillus circulans xylanase. Protein Sci 1997;6:2667–2670.
30. Poon DK, Schubert M, Au J, Okon M, Withers SG, McIntosh LP. Unambiguous determination of the ionization state of a glycoside hydrolase active site lysine by ^{1}H-^{15}N heteronuclear correlation spectroscopy. J Am Chem Soc 2006;128:15388–15389.
31. Buckingham AD. Chemical shift in the nuclear magnetic resonance spectra of molecules containing polar groups. Can J Chem 1960;38:300–307.
32. Hass MA, Jensen MR, Led JJ. Probing electric fields in proteins in solution by NMR spectroscopy. Proteins 2008;72:333–343.
33. Kukic P, Farrell D, Søndergaard C, Bjarnadottir U, Bradley J, Pollastri G, Nielsen JE. Improving the analysis of NMR spectra tracking pH-induced conformational changes: Removing artefacts of the electric field on the NMR chemical shift. Proteins (in press).
34. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 1995;6:135–140.
35. Tomlinson JH, Ullah S, Hansen PE, Williamson MP. Characterization of salt bridges to lysines in the protein G B1 domain. J Am Chem Soc 2009;131:4674–4684.
36. How the F test works to compare models. GraphPad Software, Inc. http://www.graphpad.com/help/prism5/prism5help.html?howtheftestworks.htm
37. Joshi MD, Sidhu G, Pot I, Brayer GD, Withers SG, McIntosh LP. Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J Mol Biol 2000;299:255–279.
38. Shrager RI, Cohen JS, Heller SR, Sachs DH, Schechter AN. Mathematical models for interacting groups in nuclear magnetic resonance titration curves. Biochemistry 1972;11:541–547.
39. Mossner E, Iwai H, Glockshuber R. Influence of the pK(a) value of the buried, active-site cysteine on the redox properties of thioredoxin-like oxidoreductases. FEBS Lett 2000;477:21–26.
40. Alexandrescu AT, Mills DA, Ulrich EL, Chinami M, Markley JL. NMR assignments of the four histidines of staphylococcal nuclease in native and denatured states. Biochemistry 1988;27:2158–2165.
41. Assadi-Porter FM, Fillingame RH. Proton-translocating carboxyl of subunit c of F1Fo H(+)-ATP synthase: the unique environment suggested by the pKa determined by ^{1}H NMR. Biochemistry 1995;34:16186–16193.
42. Baker WR, Kintanar A. Characterization of the pH titration shifts of ribonuclease A by one- and two-dimensional nuclear magnetic resonance spectroscopy. Arch Biochem Biophys 1996;327:189–199.
43. Betz M, Lohr F, Wienk H, Ruterjans H. Long-range nature of the interactions between titratable groups in *Bacillus agaradhaerens* family 11 xylanase: pH titration of *B. agaradhaerens* xylanase. Biochemistry 2004;43:5820–5831.
44. Bokman AM, Jimenez-Barbero J, Llinas M. ^{1}H NMR characterization of the urokinase kringle module. Structural, but not functional, relatedness to homologous domains. J Biol Chem 1993;268:13858–13868.
45. Bombarda E, Morellet N, Cherradi H, Spiess B, Bouaziz S, Grell E, Roques BP, Mely Y. Determination of the pK(a) of the four Zn^{2+}-coordinating residues of the distal finger motif of the HIV-1 nucleocapsid protein: consequences on the binding of Zn^{2+}. J Mol Biol 2001;310:659–672.
46. Cocco MJ, Kao YH, Phillips AT, Lecomte JT. Structural comparison of apomyoglobin and metaquomyoglobin: pH titration of histidines by NMR spectroscopy. Biochemistry 1992;31:6481–6491.
47. Cohen JS, Griffin JH, Schechter AN. Nuclear magnetic resonance titration curves of histidine ring protons. IV. The effects of phosphate and sulfate on ribonuclease. J Biol Chem 1973;248:4305–4310.
48. Consonni R, Arosio I, Belloni B, Fogolari F, Fusi P, Shehi E, Zetta L. Investigations of Sso7d catalytic residues by NMR titration shifts and electrostatic calculations. Biochemistry 2003;42:1421–1429.
49. Cosgrove MS, Loh SN, Ha JH, Levy HR. The catalytic mechanism of glucose 6-phosphate dehydrogenases: assignment and ^{1}H NMR spectroscopy pH titration of the catalytic histidine residue in the 109 kDa *Leuconostoc mesenteroides* enzyme. Biochemistry 2002;41:6939–6945.
50. Day RM, Thalhauser CJ, Sudmeier JL, Vincent MP, Torchilin EV, Sanford DG, Bachovchin CW, Bachovchin WW. Tautomerism, acid-base equilibria, and H-bonding of the six histidines in subtilisin BPN′ by NMR. Protein Sci 2003;12:794–810.
51. Geierstanger B, Jamin M, Volkman BF, Baldwin RL. Protonation behavior of histidine 24 and histidine 119 in forming the pH 4 folding intermediate of apomyoglobin. Biochemistry 1998;37:4254–4265.
52. Gooley PR, Keniry MA, Dimitrov RA, Marsh DE, Keizer DW, Gayler KR, Grant BR. The NMR solution structure and characterization of pH dependent chemical shifts of the beta-elicitin, cryptogein. J Biomol NMR 1998;12:523–534.
53. Kern G, Pelton J, Marqusee S, Kern D. Structural properties of the histidine-containing loop in HIV-1 RNase H. Biophys Chem 2002;96:285–291.
54. Kesvatera T, Jonsson B, Thulin E, Linse S. Measurement and modelling of sequence-specific pKa values of lysine residues in calbindin D9k. J Mol Biol 1996;259:828–839.

55. Koide A, Jordan MR, Horner SR, Batori V, Koide S. Stabilization of a fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface. Biochemistry 2001;40:10326–10333.
56. Kuhlman B, Luisi DL, Young P, Raleigh DP. pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. Biochemistry 1999;38:4896–4903.
57. Laurents DV, Huyghues-Despointes BM, Bruix M, Thurlkill RL, Schell D, Newsom S, Grimsley GR, Shaw KL, Trevino S, Rico M, Briggs JM, Antosiewicz JM, Scholtz JM, Pace CN. Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI = 3.5) and a basic variant (pI = 10.2). J Mol Biol 2003;325:1077–1092.
58. Lee KK, Fitch CA, Lecomte JT, Garcia-Moreno EB. Electrostatic effects in highly charged proteins: salt sensitivity of pKa values of histidines in staphylococcal nuclease. Biochemistry 2002;41:5656–5667.
59. Legler PM, Massiah MA, Mildvan AS. Mutational, kinetic, and NMR studies of the mechanism of *E. coli* GDP-mannose mannosyl hydrolase, an unusual Nudix enzyme. Biochemistry 2002;41:10834–10848.
60. Lund-Katz S, Wehrli S, Zaiou M, Newhouse Y, Weisgraber KH, Phillips MC. Effects of polymorphism on the microenvironment of the LDL receptor-binding region of human apoE. J Lipid Res 2001;42:894–901.
61. Marti DN, Bosshard HR. Electrostatic interactions in leucine zippers: thermodynamic analysis of the contributions of Glu and His residues and the effect of mutating salt bridges. J Mol Biol 2003;330:621–637.
62. Schmidt L, Christensen HE, Harris P. Structure of plastocyanin from the cyanobacterium *Anabaena variabilis*. Acta Crystallogr D Biol Crystallogr 2006;62(Pt 9):1022–1029.
63. Marti DN, Jelesarov I, Bosshard HR. Interhelical ion pairing in coiled coils: solution structure of a heterodimeric leucine zipper and determination of pKa values of Glu side chains. Biochemistry 2000;39:12804–12818.
64. Oliveberg M, Arcus VL, Fersht AR. pKA values of carboxyl groups in the native and denatured states of barnase: the pKA values of the denatured state are on average 0.4 units lower than those of model compounds. Biochemistry 1995;34:9424–9433.
65. Perez-Canadillas JM, Campos-Olivas R, Lacadena J, Martinez del Pozo A, Gavilanes JG, Santoro J, Rico M, Bruix M. Characterization of pKa values and titration shifts in the cytotoxic ribonuclease alpha-sarcin by NMR. Relationship between electrostatic interactions, structure, and catalytic function. Biochemistry 1998;37: 15865–15876.
66. Richarz R, Wuthrich K. High-field ^{13}C nuclear magnetic resonance studies at 90.5 MHz of the basic pancreatic trypsin inhibitor. Biochemistry 1978;17:2263–2269.
67. Sato K, Dennison C. Effect of histidine 6 protonation on the active site structure and electron-transfer capabilities of pseudoazurin from *Achromobacter cycloclastes*. Biochemistry 2002;41:120–130.
68. Schaller W, Robertson AD. pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. Biochemistry 1995;34:4714–4723.
69. Song J, Laskowski M, Jr, Qasim MA, Markley JL. Two conformational states of Turkey ovomucoid third domain at low pH: three-dimensional structures, internal dynamics, and interconversion kinetics and thermodynamics. Biochemistry 2003;42:6380–6391.
70. Sun X, Sun H, Ge R, Richter M, Woodworth RC, Mason AB, He QY. The low pKa value of iron-binding ligand Tyr188 and its implication in iron release and anion binding of human transferrin. FEBS Lett 2004;573:181–185.
71. Sundd M, Iverson N, Ibarra-Molero B, Sanchez-Ruiz JM, Robertson AD. Electrostatic interactions in ubiquitin: stabilization of carboxylates by lysine amino groups. Biochemistry 2002;41:7586–7596.
72. Sundd M, Robertson AD. Rearrangement of charge-charge interactions in variant ubiquitins as detected by double-mutant cycles and NMR. J Mol Biol 2003;332:927–936.
73. Szyperski T, Antuch W, Schick M, Betz A, Stone SR, Wuthrich K. Transient hydrogen bonds identified on the surface of the NMR solution structure of Hirudin. Biochemistry 1994;33:9303–9310.
74. Tan YJ, Oliveberg M, Davis B, Fersht AR. Perturbed pKA-values in the denatured states of proteins. J Mol Biol 1995;254:980–992.
75. Thomas CL, McKinnon E, Granger BL, Harms E, van Etten RL. Kinetic and spectroscopic studies of tritrichomonas foetus low-molecular weight phosphotyrosyl phosphatase. Hydrogen bond networks and electrostatic effects. Biochemistry 2002;41:15601–15609.
76. Tishmack PA, Bashford D, Harms E, van Etten RL. Use of ^{1}H NMR spectroscopy and computer simulations to analyze histidine pKa changes in a protein tyrosine phosphatase: experimental and theoretical determination of electrostatic properties in a small protein. Biochemistry 1997;36:11984–11994.
77. Wolff N, Deniau C, Letoffe S, Simenel C, Kumar V, Stojiljkovic I, Wandersman C, Delepierre M, Lecroisey A. Histidine pK(a) shifts and changes of tautomeric states induced by the binding of gallium-protoporphyrin IX in the hemophore HasA(SM). Protein Sci 2002;11:757–765.
78. Zhou Z, Swenson RP. Evaluation of the electrostatic effect of the 5′-phosphate of the flavin mononucleotide cofactor on the oxidation–reduction potentials of the flavodoxin from desulfovibrio vulgaris (Hildenborough). Biochemistry 1996;35:12443–12454.
79. Thurlkill RL, Grimsley GR, Scholtz JM, Pace CN. pK values of the ionizable groups of proteins. Protein Sci 2006;15:1214–1218.
80. Fogolari F, Ragona L, Licciardi S, Romagnoli S, Michelutti R, Ugolini R, Molinari H. Electrostatic properties of bovine beta-lactoglobulin. Proteins 2000;39:317–330.
81. Nozaki Y, Tanford C. Examination of titration behavior. Methods Enzymol 1967;11:715–734.
82. Gurd FR, Keim P, Glushko VG, Lawson PJ, Marshall RC, Nigen AM, Vigna RA. Carbon-13 nuclear magnetic resonance of some pentapeptides. Ann Arbor Science Publishers: Ann Arbor, MI; 1972. pp 45–49.
83. Bartik K, Redfield C, Dobson CM. Measurement of the individual pKa values of acidic residues of hen and turkey lysozymes by two-dimensional ^{1}H NMR. Biophys J 1994;66:1180–1184.
84. Grimsley GR, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci 2009;18:247–251.
85. Crowley PB. Ph.D. Thesis: Transient Protein Interactions of Photosynthetic Redox Partners. Leiden University; 2002.
86. Guss JM, Harrowell PR, Murata M, Norris VA, Freeman HC. Crystal structure analyses of reduced (CuI) poplar plastocyanin at six pH values. J Mol Biol 1986;192:361–387.
87. Yanagisawa S, Crowley PB, Firbank SJ, Lawler AT, Hunter DM, McFarlane W, Li C, Kohzuma T, Banfield MJ, Dennison C. Pi-interaction tuning of the active site properties of metalloproteins. J Am Chem Soc 2008;130:15420–15428.
88. Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig KP, Kalbitzer HR. Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. J Biomol NMR 2007;39:275–289..
89. Frantz J. g3data. http://www.frantz.fi/software/g3data.php[accessed: March 2007].
90. Tigris.org. Subversion. http://subversion.tigris.org/; 2002 [accessed on March 2007].
91. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56.
92. Levenberg K. Method for the solution of certain non-linear problems in least squares. Q Appl Math 1944;2:164–168.
93. Motulsky HJ, Ransnas LA. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. FASEB J 1987;1:365–374.
94. Tynan-Connolly BM, Nielsen JE. pKD: re-designing protein pKa values. Nucleic Acids Res 2006;34(web server issue):W48–W51.
95. Nielsen JE. Analysing the pH-dependent properties of proteins using pKa calculations. J Mol Graph Model 2007;25:691–699.