Biochemistry

 Volume 40, Number 12 March 27, 2001

New Concepts

A Novel View of pH Titration in Biomolecules†

Alexey Onufriev, David A. Case,* and G. Matthias Ullmann

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, TPC-15, La Jolla, California 92037

Received December 1, 2000; Revised Manuscript Received January 31, 2001

ABSTRACT: When individual titratable sites in a molecule interact with each other, their pH titration can be considerably more complex than that of an independent site described by the classical Henderson−Hasselbalch equation. We propose a novel framework that decomposes any complex titration behavior into simple standard components. The approach maps the set of N interacting sites in the molecule onto a set of N independent, noninteracting quasi-sites, each characterized by a pK'_a value. The titration curve of an individual site in the molecule is a weighted sum of Henderson−Hasselbalch curves corresponding to the quasi-sites. The total protonation curve is the unweighted sum of these Henderson−Hasselbalch curves. We show that pK'_a values correspond to deprotonation constants available from methods that can be used to assess total proton uptake or release, and establish their connection to protonation curves of individual residues obtained by NMR or infrared spectroscopy. The new framework is tested on a small molecule diethylenetriaminepentaacetate (DTPA) exhibiting nonmonotonic titration curves, where it gives an excellent fit to experimental data. We demonstrate that the titration curve of a site in a group of interacting sites can be accurately reconstructed, if titration curves of the other sites are known. The application of the new framework to the protein rubredoxin demonstrates its usefulness in calculating and interpreting complicated titration curves.

The protonation state of acidic and basic sites is of great importance for biomolecular function, and much effort has therefore been spent to understand the pH titration of biomolecules (*1*−*12*). Several methods exist for determining titration curves experimentally. Methods such as nuclear magnetic resonance or infrared spectroscopy can be used to determine the titration curves of individual sites. Other methods such as potentiometry, the assessment of proton uptake or release, or the measurement of the pH dependence of free energy changes describe the titration of the molecule as a whole. The two classes of methods provide different information but are complementary to each other.

The protonation probability $\langle x \rangle$ of an isolated site is given by eq 1 below, algebraically equivalent to the Henderson−Hasselbalch (HH)[1] equation describing sigmoidal standard titration curves (*12*):

$$\langle x \rangle = \frac{10^{pK_a - pH}}{1 + 10^{pK_a - pH}} \quad (1)$$

The pK_a value of an isolated titratable site is equal to the pH at which the protonation probability of this site is $^1/_2$.

† This work was supported by a postdoctoral fellowship of the Deutsche Forschungsgemeinschaft to G.M.U. (UL 174/1-1) and by NIH Grant GM 57513.

* To whom correspondence should be addressed. E-mail: case@scripps.edu. Phone: (858) 784-9768. Fax: (858) 784-8896.

[1] Abbreviations: DSR, decoupled sites representation; DTPA, diethylenetriaminepentaacetate; HH, Henderson−Hasselbalch.

10.1021/bi002740q CCC: $20.00
Published on Web 03/01/2001

However, titration curves of individual sites in biomolecules often have nonsigmoidal shapes due to interactions between sites within the molecule. By analogy to sigmoidal HH curves, the pH values at which the protonation probabilities of these sites are $^1/_2$, so-called p$K_{1/2}$ values, are often used to describe the titration behavior in such cases. However, these p$K_{1/2}$ values are not directly related to the free energy of proton uptake, as we will show later. In cases when the titration curves are not monotonic, p$K_{1/2}$ values can even be ill-defined (*13*, *14*).

In this paper, we describe a general framework that describes complicated titration behavior in a simple manner. We call this framework the decoupled sites representation (DSR). The DSR has some analogy to the normal-mode analysis in classical mechanics. Complicated titration curves of original interacting sites are represented as a weighted sum of HH curves describing decoupled quasi-sites. The total average protonation of the molecule exactly equals the sum of HH curves of the quasi-sites, even when the titration curves of original sites show a nonstandard shape. The DSR connects the titration of original sites and the energetics of protonation reactions. The DSR is not an approximation. If the number of proton binding sites is known, it provides a general description and model-free analysis of titration behavior of biomolecules.

The paper has the following structure. First, we give the derivation of the DSR using basic thermodynamic principles. Second, we apply the DSR to three different systems. A system of two interacting sites is used to explain the principles of this new framework. Then we apply the DSR to fit the experimental data of diethylenetriaminepentaacetate (DTPA), which exhibits highly unusual titration curves. We compare the DSR in this context to a different, widely used model (*4*) that assumes pairwise interactions among sites. Finally, we apply the DSR to the protein rubredoxin and demonstrate its usefulness in interpreting titration curves of biomolecules.

Derivation of the Decoupled Sites Representation

Consider a macromolecule with N proton-binding sites in equilibrium with protons in the surrounding solution. Each site is capable of binding one proton, and its protonation state is specified by x_i, where $x_i = 1$ or 0 depending on whether site i is occupied or empty. The total number of possible protonation microstates of the molecule is 2^N, and a microstate can be conveniently represented by an N-dimensional vector $\vec{\mathbf{x}}^{\mathrm{k}} = (x_1^{\mathrm{k}}, x_2^{\mathrm{k}}, ..., x_N^{\mathrm{k}})$. The superscript k denotes the microstate. Each microstate is characterized by its free energy[2] G^{k} and the number of bound protons n ($=\sum_i^N x_i^{\mathrm{k}}$). Note that only two microstates are uniquely specified by the number of bound protons: the fully occupied state (1,1,...,1) with $n = N$ and the completely empty state (0,0,...,0) with $n = 0$. In what follows, we take the latter to be the reference state with a free energy G^0 of 0. The equilibrium properties of the system can be calculated from its grand canonical partition function, which can be written (*3*, *15*) as a power series (binding polynomial) in the ligand activity $\Lambda = 10^{-\mathrm{pH}}$:

$$Z = \sum_k^{2^N} \exp(-\beta G^{\mathrm{k}})\Lambda^n \tag{2}$$

where $\beta = (kT)^{-1}$. The equilibrium protonation of site i averaged over all microstates is given by

$$\langle x_i \rangle = \frac{1}{Z}\sum_k^{2^N} x_i^{\mathrm{k}} \exp(-\beta G^{\mathrm{k}})\Lambda^n \tag{3}$$

In the special case of noninteracting sites, the free energy of microstate $\vec{\mathbf{x}}^{\mathrm{k}}$ can be written as

$$G^{\mathrm{k}} = -\beta^{-1} \ln 10 \sum_i^N x_i^{\mathrm{k}}\, \mathrm{p}K_{\mathrm{a},i} \tag{4}$$

This additive property allows one to group terms in eq 2 and rewrite the partition function of the N-site system in the form of a product (*15*):

$$Z = \prod_i^N (1 + \Lambda 10^{\mathrm{p}K_{\mathrm{a},i}}) = \prod_i^N \left(1 + \frac{\Lambda}{K_i}\right) \tag{5}$$

where K_i ($=10^{-\mathrm{p}K_{\mathrm{a},i}}$) is the deprotonation constant of site i. The average protonation $\langle x_i \rangle$ of each site is then given by the HH titration curve

$$\langle x_i \rangle = \frac{10^{\mathrm{p}K_{\mathrm{a},i}-\mathrm{pH}}}{1 + 10^{\mathrm{p}K_{\mathrm{a},i}-\mathrm{pH}}} = \frac{\Lambda}{K_i + \Lambda} \tag{6}$$

and the total average protonation of the molecule is just the sum of individual HH curves:

$$\langle X \rangle = \sum_i^N \langle x_i \rangle = \sum_i^N \frac{10^{\mathrm{p}K_{\mathrm{a},i}-\mathrm{pH}}}{1 + 10^{\mathrm{p}K_{\mathrm{a},i}-\mathrm{pH}}} \tag{7}$$

When sites interact, the additive property (eq 4) does not hold, and $\langle x_i \rangle$ and $\langle X \rangle$ are no longer given by the simple formulas eqs 6 and 7. However, by analogy with normal-mode analysis in classical mechanics, one can expect to decouple the interacting sites and recover the simplicity of eqs 6 and 7 by an appropriate linear transformation. We are, therefore, seeking the linear transformation $\{\vec{\mathbf{x}}^{\mathrm{k}}\} \rightarrow \{\vec{\mathbf{y}}^{\mathrm{k}}\}$ that describes the occupancy of independent quasi-sites:

$$\langle x_i \rangle = \sum_j^N a_{ij} \langle y_j \rangle \tag{8}$$

Later in this section, we determine the elements of the transformation matrix $\{a_{ij}\}$ explicitly, which proves that this transformation is always possible. Out of the $N \times N$ entries of this matrix, only $(N-1)^2$ are independent, since the $\{a_{ij}\}$ values have to satisfy[3]

[2] G^{k} is the standard Gibbs free energy of microstate k when the proton activity equals 1.

[3] The total average number of protons occupying all of the original sites must equal the total average number of protons on new quasi-sites ($\sum_i^N \langle x_i \rangle = \sum_j^N \langle y_j \rangle$). Substituting $\langle x_i \rangle$ from eq 8 into the left-hand side of the above equality, we obtain $\sum_i^N \sum_j^N a_{ij} \langle y_j \rangle = \sum_j^N \langle y_j \rangle$ which is possible if and only if the first constraint in eq 9 holds. The other constraint is determined by considering the average protonation in the limit of extreme low pH: $\langle x_i \rangle = \langle y_j \rangle = 1$ for all i and j values.

$$\sum_i^N a_{ij} = 1, \sum_j^N a_{ij} = 1 \tag{9}$$

Since the new quasi-sites are, by definition, noninteracting, their average protonations $\langle y_j \rangle$ are given by the HH curves (eq 6) with a set of new deprotonation constants K'_j that replace K_i. In the new basis, the partition function (eq 2) takes the form of eq 5:

$$Z = \prod_j^N \left(1 + \frac{\Lambda}{K'_j}\right) \tag{10}$$

As has been noted before, the binding polynomial can always be factorized (*3*, *16*), so we can rewrite eq 2 in the form of eq 10 where K'_j values are the negatives of the roots of the partition function (eq 2). When the number of binding sites N is greater than 2, there is no straightforward way to determine analytically the number of real roots of the binding polynomial (eq 2) (*3*). Our numerical analysis suggests that repulsion between ligands on different sites is a sufficient condition for all of the roots to be purely real. Conversely, an effective attraction[4] between the ligands may result in some of the roots being complex, which implies positive cooperativity between the sites, such that the binding of a ligand increases the affinity for subsequent ligand binding. The opposite effect, i.e., the decrease in the ligand binding affinity with the number of ligands already bound, is called anticooperativity. Although the mathematical formalism developed here applies verbatim to positive cooperative binding, we will consider this case in detail in a separate paper. Because protons are charged, they usually bind anticooperatively. Cooperativity is an unlikely scenario when protons are the only type of ligand, the situation with which we are concerned in this work. Note that since all the coefficients of the binding polynomial (eq 2) are positive, its real roots can only be negative. The corresponding K'_j values are therefore real positive numbers and may be regarded as deprotonation constants of quasi-sites. Since quasi-sites are noninteracting, the total average protonation is the sum of HH titration curves described by a set of new deprotonation constants K'_j. In this sense, it is meaningful to describe the N-site molecule in terms of N independent sites, each capable of taking only one proton. The free energy ΔG required to fully protonate the molecule[5] at a given pH can be expressed as a sum of contributions from each quasi-site (eq 11), where $pK'_{a,j} = -\log K'_j$

$$\Delta G = -\beta^{-1} \ln 10 \sum_j^N (pK'_{a,j} - pH) \tag{11}$$

The entries of the transformation matrix $\{a_{ij}\}$ can be determined from eq 8. We use the general expression (eq 3) for the average protonation of a quasi-site [$\langle y_j \rangle = (1/Z)\sum_k^{2^N} y_j^k \exp(-\beta G^k)\Lambda^n$] and the noninteracting form (eq 4) for the free energy of a microstate in terms of quasi-site occupancies ($G^k = -\beta^{-1} \ln 10 \sum_j^N y_j^k \, pK'_{a,j}$). Grouping together the terms corresponding to $n = 1, 2, \ldots$, total protons bound, we obtain

$$\langle y_j \rangle = \frac{1}{Z}\left(\Lambda 10^{pK'_{a,j}} + \Lambda^2 \sum_{k\neq j}^N 10^{pK'_{a,j}+pK'_{a,k}} + \Lambda^3 \sum_{\substack{k<m \\ k\neq j}}^N \sum_{m\neq j}^N 10^{pK'_{a,j}+pK'_{a,k}+pK'_{a,m}} + \ldots\right) \tag{12}$$

Substituting $\langle y_j \rangle$ from eq 12 into the right-hand side of eq 8, using expression eq 3 for $\langle x_i \rangle$ on the left-hand side, and equating coefficients in front of the same powers of Λ, we arrive at N systems of N linear equations for a_{ij}:

$$\begin{bmatrix} e^{-\beta G\{i\}} \\ \sum_{p\neq i}^N e^{-\beta G\{i,p\}} \\ \sum_{q<p,\, q\neq i}^N \sum_{p\neq i}^N e^{-\beta G\{i,p,q\}} \\ \vdots \end{bmatrix} = \begin{bmatrix} 10^{pK'_{a,1}} & 10^{pK'_{a,2}} & \cdots \\ \sum_{k\neq 1}^N 10^{pK'_{a,1}+pK'_{a,k}} & \sum_{k\neq 2}^N 10^{pK'_{a,2}+pK'_{a,k}} & \cdots \\ \sum_{k<m,\, k\neq 1}^N \sum_{m\neq 1}^N 10^{pK'_{a,1}+pK'_{a,k}+pK'_{a,m}} & \sum_{k<m,\, k\neq 2}^N \sum_{m\neq 2}^N 10^{pK'_{a,2}+pK'_{a,k}+pK'_{a,m}} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} a_{i1} \\ a_{i2} \\ a_{i3} \\ \vdots \end{bmatrix} \tag{13}$$

where we identify $G\{i,p,q,\ldots\}$ as the free energy of a microstate in which only original sites i ,p, q, ..., are protonated and others empty. There are a total of N systems (eq 13), one for every i value from 1 to N. Each of them has a unique[6] solution which we find numerically (*17*). As we have already mentioned, only $(N-1)^2$ out of N^2 matrix elements $\{a_{ij}\}$ are independent, and we use $2N-1$ constraints (eq 9) to check the numerical accuracy of the calculations. The resulting set of parameters $\{a_{ij}, K'_j\}$ completely defines the new quasi-sites and the transformation matrix to the original sites. The description of macromolecule titration in terms of quasi-sites is as general as the statistical mechanical description provided by eq 3.

Applications of the Decoupled Sites Representation

In the previous section, we have shown that for a molecule with N interacting titratable sites, the average protonation of a site and the total average protonation of the molecule are always given by

[4] Effective attraction may result from conformational transitions during titration, the presence of more then one type of ligand, or nonelectrostatic site−site interactions.

[5] Wyman showed that the free energy required to fully protonate the macromolecule at unit ligand activity $\Delta G = kT \ln \Lambda_m$, where Λ_m is the mean ligand activity (*24*). This connects quasi-site pK's with Λ_m: $-\ln 10 \sum_j^N pK'_{a,j} = \ln \Lambda_m$.

[6] Equation 13 has a unique solution providing the determinant of the matrix is not equal to zero. The determinant can become zero if $K'_{a,m} = K'_{a,k}$ for a pair of m and k values. The system is then degenerate, which means that some a_{ij} can be chosen arbitrarily subject to the constraints (eq 9). In practice, the degeneracy is lifted by site−site interactions.

$$\langle x_i \rangle = \sum_j^N a_{ij} \frac{10^{pK'_{a,j}-pH}}{1+10^{pK'_{a,j}-pH}} \quad (14)$$

$$\sum_i^N \langle x_i \rangle = \sum_j^N \frac{10^{pK'_{a,j}-pH}}{1+10^{pK'_{a,j}-pH}} \quad (15)$$

where $pK'_{a,j}$ values are the pK_a values of a set of N noninteracting quasi-sites connected to the original interacting sites by a linear transformation defined by $(N-1)^2$ independent coefficients $\{a_{ij}\}$. A set of pK'_a and a_{ij} can be used as independent parameters to interpret titration data in a model-free way.

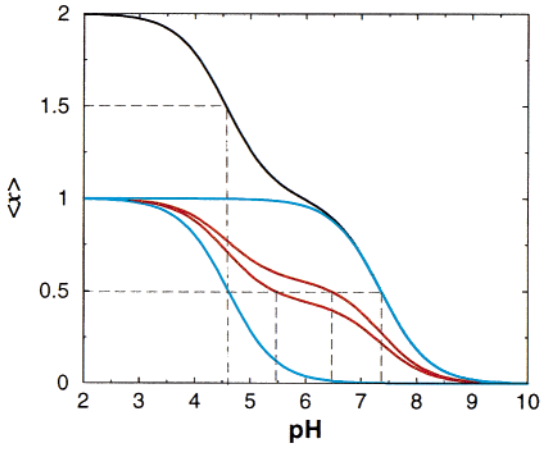
Two Interacting Sites. The principles of the DSR can be illustrated by a simple example. Suppose that the free energy G^k of a protonation state $\vec{\mathbf{x}}^k$ can be written in terms of contributions that arise from protonation and those arising from pairwise interactions between the titratable groups:

$$G^k = -\beta^{-1} \ln 10 \sum_i^N x_i^k pK_{a,i}^{intr} + \frac{1}{2}\sum_i^N \sum_{j\neq i}^N x_i^k x_j^k W_{ij} \quad (16)$$

where $pK_{a,i}^{intr}$ is the intrinsic pK_a value, i.e., the pK_a value this site would have if all other titratable sites were in the reference state, and W_{ij} is the interaction between the protonated forms of sites i and j. The free energy G^k in eq 16 can then be used in eq 3 to calculate the protonation probability of each site (*4*−*6*). When $N > 2$, eq 16 is *not* the most general model.

The interaction between just two sites is enough to cause complicated titration behavior (*18*, *19*). Let us assume that the pK_a^{intr} values of the sites are 7.0 and 7.1. They interact with an energy W_{12} of 3.0 kcal/mol, which is reasonable for strongly interacting sites in proteins. Figure 1 shows the titration behavior of this system. Application of the DSR in this case involves two steps. We first use the set of G^ks to construct the binding polynomial (eq 2) and find its roots to obtain the quasi-site $pK'_{a,i}$ values. We then solve the $N = 2$ sets of linear equations eq 13 to obtain $\{a_{ij}\}$. The system can be described by two quasi-sites having pK'_a values of 4.6 and 7.4. The total protonation curve exactly coincides with the sum of the two HH curves corresponding to the noninteracting quasi-sites, and we identify these quasi-site pK'_as with the apparent pK_a values. The latter can be measured experimentally, and are, to a good approximation, given by the inflection points of the total protonation curve. The $pK_{1/2}$ values of the original sites, 5.5 and 6.5, clearly do not correspond to the apparent pK'_as. The energetics of the system, described by the total protonation curve, is not well represented by these two $pK_{1/2}$ values, and the pK'_as of the quasi-sites can be used to calculate the free energy of protonation of the molecule at each pH (eq 11). The individual titration curves of the two original sites can be reconstructed from the titration curves of the quasi-sites via eq 14 using matrix $\{a_{ij}\}$ (Figure 1). For instance, when $pH = pK'_{a,2} = 7.4$, the first quasi-site ($pK'_{a,1} = 4.6$) is practically unoccupied, the second one is half-protonated, and therefore the probability of finding the proton on each of the original sites is $^1/_2 \times 0.56$ and $^1/_2 \times 0.44$, respectively.

Titration of DTPA. We apply the DSR to analyze titration of three nitrogen atoms in small molecule DTPA (Figure 2a) which is one of the simplest systems that exhibit strongly nonmonotonic titration. The experimentally measured apparent pK_a values (*20*) of the total titration curve of the system are 1.8, 2.5, 4.4, 8.8, and 10.4. At pH >13, DTPA is completely deprotonated. The chemical shift of the hydrogens bound to the carbons next to the nitrogens in DTPA has also been measured in the pH range from 3 to 13 (*14*). In agreement with experiments (*21*), we assume that only the nitrogen sites titrate in this pH range, that they are totally protonated at pH 3, and that the chemical shifts relate linearly to the protonation probabilities shown in Figure 2. Because the molecule is symmetric, the two terminal nitrogens cannot be distinguished and their titration is given by a single curve.



$$pK'_a = \begin{pmatrix} 4.6 \\ 7.4 \end{pmatrix} \quad \{a_{ij}\} = \begin{pmatrix} 0.56 & 0.44 \\ 0.44 & 0.56 \end{pmatrix}$$

FIGURE 1: Titration curves of two interacting sites. The two sites have pK_a^{intr} values of 7.0 and 7.1, and interact with $W_{12} = 3$ kcal/mol. The titration curves of the original sites are shown with red lines. Note that the $pK_{1/2}$ values of these curves (5.5 and 6.5) do not correspond to the two inflection points of the total protonation curve (black line). The decoupled sites representation gives two independent quasi-sites with pK'_a values of 4.6 and 7.4 (cyan lines). These numbers coincide with the apparent pK'_a values approximately given by the inflection points of the total protonation curve. The individual titration curves of the original interacting sites (red lines) can be reconstructed from the standard titration curves of the quasi-sites (cyan lines) via eq 14 using matrix $\{a_{ij}\}$. Both the red and the cyan curves sum to the same total titration curve (solid line).

Two different approaches can be used to fit the experimental data.

(1) We fit the experimental data for the individual sites to the DSR (eq 14). Because of the symmetry of the molecule and the conditions in eq 9, there are three $pK'_{a,i}$ values and two a_{ij} values available for simultaneous fitting. The remaining a_{ij} values can be obtained from eq 9 and from symmetry relations. The best fit gives the following values: $pK'_{a,1} = 4.6$, $pK'_{a,2} = 8.8$, $pK'_{a,3} = 10.2$, $a_{11} = 0.74$, and $a_{12} = -0.62$. The $pK'_{a,i}$ values agree well with the measured pK_a values. As one can see from the solid lines in Figure 2b, the DSR fits the experimental data excellently. Another choice would be to fit the total titration curve to eq 15 to obtain the $pK'_{a,i}$ values, and subsequently to fit the $\{a_{ij}\}$ using those $pK'_{a,i}$ values. This yields the following values: $pK'_{a,1} = 4.6$, $pK'_{a,2} = 9.0$, $pK'_{a,3} = 10.2$, $a_{11} = 0.74$, and $a_{12} = -0.68$. This fitting procedure is much easier and faster and still provides an excellent fit.

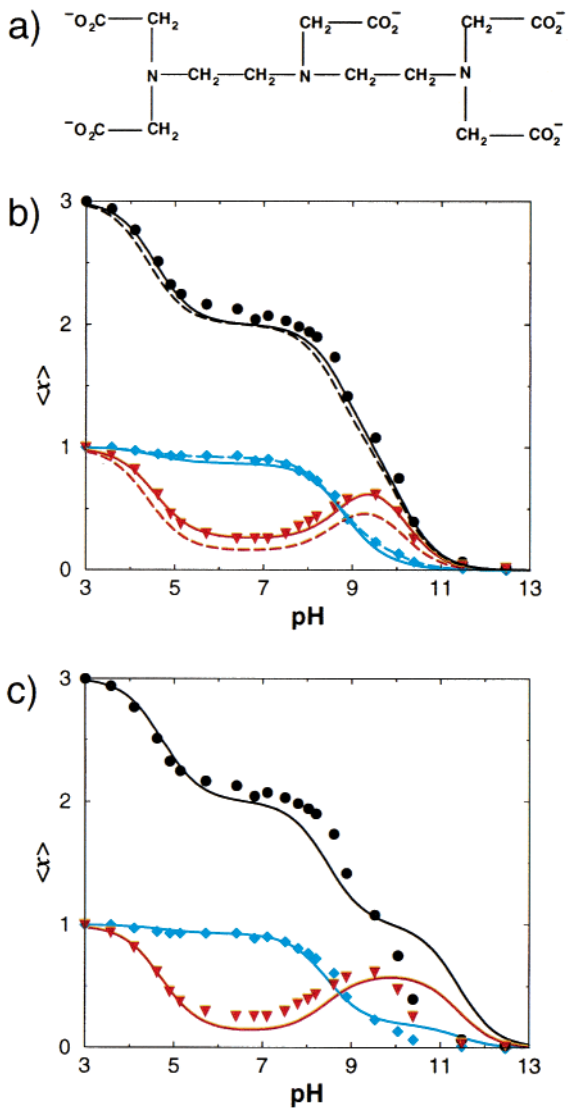


FIGURE 2: (a) Structural formula of diethylenetriaminepentaacetate (DTPA). (b) Fitting the titration curves of DTPA to the decoupled sites representation (eq 14). The experimental data are represented by discrete symbols, and the results of the DSR fit are given by lines. The colors represent the following protonation probabilities: red, middle nitrogen; cyan, terminal nitrogens; and black, total. Solid lines represent a simultaneous DSR fit to the two individual titration curves. For the long-dashed curves, only the data of the terminal nitrogens are used (cyan diamonds). The decoupled sites representation describes the titration of DTPA very well. (c) Fitting the titration curves of DTPA to eqs 3 and 16. The coloring is the same as for panel b. Although the model describes the general shape of the curves, it fails to reproduce the experimental data correctly over the entire range of pH values.

(2) As an alternative, we could apply the commonly used model based on eqs 3 and 16, where we vary $pK_{a,i}^{intr}$ and W_{ij} for the best fit. Because of the symmetry of the molecule, there are only two $pK_{a,i}^{intr}$ values and two W_{ij} values. We obtain the following values for the best fit: $pK_{a,1}^{intr} = pK_{a,2}^{intr} = 10.7$, $pK_{a,3}^{intr} = 11.2$, $W_{12} = 2.2$ kcal/mol, and $W_{13} = W_{23} = 4.4$ kcal/mol. The results are shown in Figure 2c. Although the general features of the titration curves can be represented by the model based on eqs 3 and 16, it fails to reproduce the experimental data correctly over the entire pH range.

Why does the DSR fit the experimental data better than the widely used model based on eq 16? Two explanations are possible. First, eq 16 relies on the assumption that the free energy of a protonation state can be decomposed into energy terms that account for the proton binding and for the pairwise interaction between the proton binding sites. Such a pairwise decomposition is not always possible, especially since the titrating nitrogens in DTPA are separated by only two carbons. Second, eq 16 assumes that the molecule adopts a single conformation throughout the whole pH range. The DSR does not use any particular assumptions for calculating the free energy of the microstates. It is thus general and useful for fitting experimental data in a "model-free" fashion.

Encouraged by the finding that the DSR fits the titration curves of DTPA very well, we go one step further. We take the seemingly less-structured titration curve of one of the two terminal nitrogens, i.e., the cyan curve in Figure 2b, and fit this curve to the DSR. We again obtain a fit yielding a similar parameter set as in the case when we use both titration curves: $pK'_{a,1} = 4.4$, $pK'_{a,2} = 8.7$, $pK'_{a,3} = 10.1$, $a_{11} = 0.84$, and $a_{12} = -0.48$. Using this fit, we are able to reconstruct the titration curve of the central nitrogen, i.e., the red curve in Figure 2b. The resulting curves are shown by the long-dashed curves in Figure 2b. These results demonstrate that it is possible to reconstruct the unknown titration curve of a site from the known curves of the sites that interact with the site of interest.

Titration of Rubredoxin. To illustrate the application of the DSR to proteins, we applied it to rubredoxin, which has 54 amino acids of which 20 are titratable. Because this protein is small and highly charged, all residues interact strongly with each other. We use eq 16, which relies on the decomposition of energy contributions into protonation energies and interaction energies, to compute free energies G^k of the microstates. This decomposition is possible in linear theories such as the linearized Poisson−Boltzmann equation (*22*). The pK_a^{intr} values and the interaction energies W_{ij} are estimated by solving the linearized Poisson−Boltzmann equation using the same parameters as in previous applications (*23*). We apply the DSR following the procedures used in the two-site example, i.e., first find the roots of the binding polynomial using the pK_a^{intr} values and the interaction energies W_{ij} and then compute the coefficients $\{a_{ij}\}$ by solving N sets of linear equations.

Figure 3 shows the titration of four residues of rubredoxin computed by the DSR. Whereas the shapes of the curves in Figure 3a are sigmoidal and resemble HH curves closely, the shapes of those in Figure 3b clearly deviate from HH curves. Table 1 lists the expansion coefficients a_{ij} used in eq 14 for constructing the titration curves using quasi-site pK'_a values. The N-terminus of the protein has only two quasi-sites contributing significantly to its titration. The pK'_a values of these two quasi-sites are similar, as are all the pK'_a values that have large contributions to the titration of Glu50. Therefore, the titrations in Figure 3a do not deviate much from HH curves. Since the contributing pK'_a values are not identical, the titration curves are somewhat flattened. On the contrary, the titration curves of Glu53 and Glu54 are clearly not of HH shape. Both residues have many quasi-sites with very different pK_a values contributing to their titration. For example, Glu53 has large contributions from quasi-sites with pK'_a values of ~5 and ~8. Accordingly, the titration curve of this residue has two distinctive steps, one

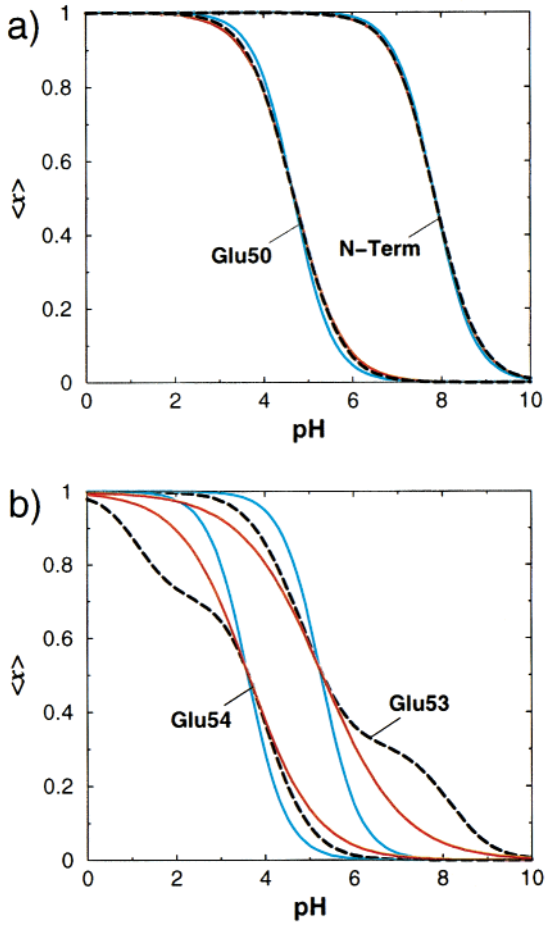


FIGURE 3: Titration curves of selected residues in rubredoxin representing two different types of titration behavior. The corresponding pK'_a values and expansion coefficients $\{a_{ij}\}$ are given in Table 1. The titration curves of the original sites are represented by black dashed lines. Their approximations by Henderson−Hasselbalch (eq 6) and Hill (eq 17) equations are represented by cyan and red solid lines, respectively. The corresponding $pK_{1/2}$ values and Hill coefficients are listed in Table 1. (a) Titration curves that have a sigmoidal shape but are somewhat flattened out compared to a standard Henderson−Hasselbalch curve and are closely approximated by the Hill equation. (b) Two-step titration curves. If quasi-sites with very different pK'_a values contribute to the titration of a real site, titration curves differ considerably from both the Henderson−Hasselbalch and Hill forms and exhibit multiple steps.

at pH ~5, the other at pH ~8. Glu54 has contributions from quasi-sites with pK'_a values of ~1 and ~4, and its titration curve also has two steps.

When a titration curve deviates from the HH form, it is often customary to describe it by the Hill equation (12), where n is the Hill coefficient reflecting the degree of cooperativity between the sites.

$$\langle x\rangle = \frac{10^{n(pK_{1/2}-pH)}}{1+10^{n(pK_{1/2}-pH)}} \qquad (17)$$

While the shape of the titration curves of Glu50 and the N-terminus in Figure 3a is virtually identical to that of the Hill equation ($n = 0.92$ and 0.82, respectively), the titration curves of Glu54 and Glu53 in Figure 3b are poorly approximated by it. The corresponding Hill coefficients, 0.49 and 0.57, are considerably less than 1, indicating strong anticooperative binding in this case. In situations such as this, one can clearly benefit from using the DSR for interpreting titration curves.

Table 1: Expansion Coefficients and pK'_a Values for the Decoupled Sites Representation of Four Representative Residues in Rubredoxin[a]

N-terminus $pK_{1/2} = 7.9$, $n = 0.92$		Glu50 $pK_{1/2} = 4.7$, $n = 0.82$		Glu53 $pK_{1/2} = 5.3$, $n = 0.49$		Glu54 $pK_{1/2} = 3.6$, $n = 0.57$	
a_{1j}	$pK'_{a,j}$	a_{2j}	$pK'_{a,j}$	a_{3j}	$pK'_{a,j}$	a_{4j}	$pK'_{a,j}$
0.774	7.7	0.176	5.1	0.263	5.1	0.514	3.9
0.219	8.4	0.173	4.5	0.196	8.4	0.297	1.1
		0.161	4.4	0.112	7.7	0.043	4.4
		0.126	4.2	0.080	3.9	0.043	4.2
		0.113	5.4	0.073	5.4	0.036	4.5
		0.098	4.7	0.071	4.2	0.027	4.7
		0.097	4.9	0.067	4.4	0.022	5.1
		0.046	3.9	0.049	4.9	0.017	4.9
		0.004	3.2	0.048	4.5		
				0.028	4.7		
				0.007	6.1		

[a] The $pK_{1/2}$ value and the Hill coefficient that are often used to describe titration curves are given at the top. We list only those $M \le Na_{ij}$ coefficients and pK'_a values that are required for reconstructing the titration curves within 1% error from the exact result. Thus, the protonation of a residue is approximated here by $\langle x_i\rangle \approx \sum_j^M a_{ij}[10^{pK'_{a,j}-pH}/(1+10^{pK'_{a,j}-pH})]$. To reconstruct the curves exactly, all N coefficients are needed.

These examples show that the DSR can help to analyze and rationalize the titration curves. Residues with titration curves that have a sigmoidal shape but are flattened compared to HH curves have contributions from several quasi-sites with about the same pK'_a values. The larger the deviation between the pK'_a values of the contributing quasi-sites is, the flatter the titration curve is. Residues with stepwise titration curves are described by quasi-sites with very different pK'_a values.

Conclusions

We have introduced a novel framework, the decoupled sites representation (DSR), for calculating, describing, and interpreting complex titration curves of molecules. The framework represents the titration of a molecule with N interacting sites as the titration of a system of N decoupled quasi-sites. We show that the pK'_a values of the quasi-sites are directly related to the roots of the grand canonical partition function. The quasi-sites are independent, and their titration is therefore described by the Henderson−Hasselbalch (HH) equation. The total protonation of the molecule is always given by the sum of the HH curves corresponding to the quasi-sites. Hence, no matter how complicated the titration behavior of individual sites may be, methods that can be used to assess overall proton uptake or release always lead to a sum of HH curves (as long as protons do not bind cooperatively, i.e., the binding polynomial has only real roots). Therefore, by looking at the total titration curve, one cannot conclude if the original sites interact.

Furthermore, we find from our formalism that the titration curve of each original site can be represented by a weighted sum of the HH curves describing the quasi-sites. The DSR connects titration curves measured by NMR with titration curves measured by potentiometry as demonstrated in the case of DTPA. To discuss the energetics of chemical processes based on proton uptake, one should use pK'_a

values, not the $K_{1/2}$ values of titration curves of original sites as is often done in the literature. The DSR provides a connection between the titration of the whole molecule, the titration of its sites, and the energetics of the protonation process.

Analysis of the framework allows us to draw a number of general conclusions. If the titration curve of a certain site has an unusual shape, it must interact with at least one other site in the molecule. However, the converse is not true:. If a measured titration curve of a site has the HH form, one still *cannot* straightforwardly conclude that only a single site is responsible for this behavior as can be seen from our calculation on rubredoxin. Nonmonotonic titration curves, such as that of the central nitrogen in DTPA, indicate that this site is interacting with at least two more titratable sites. Even if titration curves of the original sites are nonmonotonic, the titration curve of the whole molecule is always monotonic since it is a sum of HH curves representing noninteracting quasi-sites.

In the applications, we have shown that the DSR is able to fit even very complicated titration curves. Moreover, we show that it is possible to reconstruct the unknown titration curve of a site from the known curves of the sites that interact with the site of interest. This is particularly interesting if the protonation behavior of one site is not easily accessible experimentally. The application of the DSR to the rubredoxin illustrates the use of the novel framework for discussing and interpreting complicated titration curves of biomolecules.

The DSR is not an approximation. It is as general as the most general thermodynamic description, but provides new insights into the titration behavior of biomolecules. This theory is generally applicable to binding and not restricted to protons as ligands.

ACKNOWLEDGMENT

We thank Veronika Lunyak for the help with interpreting the experimental data and Louis Noodleman, Donald Bashford, and Erik Zuiderweg for helpful discussions.

REFERENCES

1. Tanford, C., and Kirkwood, J. G. (1957) *J. Am. Chem. Soc. 79*, 5333−5347.
2. Garcia-Moreno, E. B. (1995) *Methods Enzymol. 240*, 512−538.
3. Wyman, J., and Gill, S. J. (1990) *Binding and linkage*, University Science Books, Mill Valley, CA.
4. Bashford, D., and Karplus, M. (1990) *Biochemistry 29*, 10219−10225.
5. Ullmann, G. M., and Knapp, E. W. (1999) *Eur. Biophys. J. 28*, 533−551.
6. Beroza, P., and Case, D. A. (1998) *Methods Enzymol. 295*, 170−189.
7. Gunner, M., and Alexov, E. (2000) *Biochim. Biophys. Acta 1458*, 63−87.
8. Warshel, A., and Papazyan, A. (1998) *Curr. Opin. Struct. Biol. 8*, 211−217.
9. Poland, D. (2000) *J. Chem. Phys. 113*, 4774−4784.
10. Madura, J. D., Davis, M. E., Gilson, M. K., Wade, R. C., Luty, B. A., and McCammon, J. A. (1994) *Rev. Comput. Chem. 5*, 229−267.
11. Briggs, J. M., and Antosiewicz, J. (1999) *Rev. Comput. Chem. 13*, 249−311.
12. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry. Part III. The Behavior of Biological Macromolecules*, W. H. Freeman and Co., New York.
13. Zuiderweg, E. R. P., van Beek, G. G. M., and De Bruin, S. H. (1979) *Eur. J. Biochem. 94*, 297−306.
14. Sudmeier, J. L., and Reilley, C. N. (1964) *Anal. Chem. 36*, 1698−1706.
15. Schellman, J. A. (1975) *Biopolymers 14*, 999−1018.
16. Briggs, W. E. (1983) *Biophys. Chem. 18*, 67−71.
17. Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992) *Numerical Recipes in C, Vol.* 2, Cambridge University Press, Cambridge, U.K.
18. Bashford, D., and Karplus, M. (1991) *J. Phys. Chem. 95*, 9557−9561.
19. Ackers, G. K., Shea, M. A., and Smith, F. R. (1983) *J. Mol. Biol. 170*, 223−242.
20. Chaberek, S., Frost, A. E., Doran, M. A., and Bicknell, N. J. (1959) *J. Inorg. Nucl. Chem. 11*, 184−196.
21. Letkeman, P. (1979) *J. Chem. Educ. 56*, 348−351.
22. Honig, B., and Nicholls, A. (1995) *Science 268*, 1144−1149.
23. Ullmann, G. M. (2000) *J. Phys. Chem. B 104*, 6293−6301.
24. Wyman, J. (1964) *Adv. Protein Chem. 18*, 223−286.

BI002740Q