

Semi-mechanistic Partial Buffer Approach to Modeling pH, the Buffer Properties, and the Distribution of Ionic Species in Complex Solutions

DANIEL P. DOUGHERTY,[†] EDITH RAMOS DA CONCEICAO NETA,[‡]
 ROGER F. MCFEETERS,^{*,‡} SHARON R. LUBKIN,[§] AND FREDERICK BREIDT, JR.[‡]

Lyman Briggs School of Science and Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48825-1107, U.S. Department of Agriculture, Agricultural Research Service, and Department of Food Science, North Carolina State University, Raleigh, North Carolina 27695-7624, and Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205

In many biological science and food processing applications, it is very important to control or modify pH. However, the complex, unknown composition of biological media and foods often limits the utility of purely theoretical approaches to modeling pH and calculating the distributions of ionizable species. This paper provides general formulas and efficient algorithms for predicting the pH, titration, ionic species concentrations, buffer capacity, and ionic strength of buffer solutions containing both defined and undefined components. A flexible, semi-mechanistic, partial buffering (SMPB) approach is presented that uses local polynomial regression to model the buffering influence of complex or undefined components in a solution, while identified components of known concentration are modeled using expressions based on extensions of the standard acid–base theory. The SMPB method is implemented in a freeware package, *pHTools*, for use with Matlab. We validated the predictive accuracy of these methods by using strong acid titrations of cucumber slurries to predict the amount of a weak acid required to adjust pH to selected target values.

KEYWORDS: Buffer capacity; *Cucumis sativus*; vegetable; titration; pH prediction; Davies equation

INTRODUCTION

The ability of acids, bases, or amphoteric compounds, such as amino acids, to be substrates for enzymes, to act as signaling molecules, to move through cell membranes, or to function in other biological processes is dependent upon their being in the correct state of ionization. Therefore, it is often necessary to know the distribution of ionic species when doing microbiological and biochemical research with foods or other biological systems. A general approximation for the concentrations of ionic species, particularly in complex systems, may be obtained by calculations involving the pK_a values for all ionizable groups on all molecules, the pH, and ionic strength in the solution. These calculations can be done using the algebraic methods described by Butler and Cogley (1) when the concentrations of ionizable components are known. However, when experiments involve solutions with greater complexity than one or two components with single ionizable groups, the calculations rapidly become quite complex.

In addition to known components with ionizable groups, biological and food systems typically contain unknown components that contribute significantly to the buffer capacity. There are no general methods to calculate the distributions of ionic species of the known components in systems that also contain unknown components. Recent attempts to carry out such calculations are limited in scope. Horiuchi et al. (2) used a three-layer neural network with a back-propagation algorithm to model changes in product distribution as a function of pH in chemostat cultures. This method involved training the neural network during controlled step changes in pH. Once trained, the neural network can predict transient changes in product distribution for continuously varying pH. A limitation of this method is that training the neural network can be expensive, and alterations in protocol may require re-training to maintain an acceptable level of accuracy.

Van Vooren et al. (3) introduced a method called “automatic model building” to accomplish similar goals. This method attempts to infer the dissociation constants and concentrations of simple buffer components present in a complex buffer solution by estimating individual pK_a and concentrations of potential components from buffer capacity curves. A database of potential components is available. Limits can be placed on the number of components to be estimated to control the overall

* Corresponding author. Tel: (919)515-2990. Fax: (919)513-0180. E-mail: rfm@unity.ncsu.edu.

[†] Michigan State University.

[‡] USDA-ARS and Department of Food Science, North Carolina State University.

[§] Department of Mathematics, North Carolina State University.

complexity of the resulting model. A limitation of this method is that, since control of model complexity is not completely data-driven, overly complex models can result.

The desire to model buffer properties in a comprehensive yet robust manner is motivated by the need to test hypotheses in complex systems. For example, this laboratory recently proposed a new hypothesis for the ability of protonated species of organic acids to elicit sour taste (4). Development and testing of this hypothesis required the ability to determine the concentrations of the protonated and nonprotonated organic acid species of multiple mono-, di-, and tricarboxylic acids in solutions of different pH and ionic strength. It was also necessary to estimate the concentrations of multiple acid species in pickled cucumbers and sauerkraut to determine if this hypothesis could be applied to food systems. Calculation of protonated acid species in solutions with three organic acids that contained a total of four to seven ionizable carboxyl groups was required to test the hypothesis (5). Bjornsdottir et al. (6) investigated the effect of fully protonated mono-, di-, and tricarboxylic acids on the survival of the acid-tolerant pathogen *Escherichia coli* O157:H7 at constant pH and ionic strength in a noninhibitory gluconic acid buffer (7). The calculations required to determine the concentrations of protonated acid species for these investigations were done using a new approach to modeling complex buffer systems implemented as pHTools in MATLAB, a commonly used modeling software package.

The pHTools software implements and extends the algebraic approach originated by Van Slyke (8) and formalized and expanded by Butler and Cogley (1). These earlier investigators demonstrated that simple nonlinear equations could be used to describe ionic equilibria and, therefore, predict the titration curve and buffer capacity of mixtures of acids and bases, including polyvalent and amphoteric molecules. The Davies equation or a modified form of the equation have been shown to accurately adjust for the effects of ionic strength up to 0.5 (12) or 1.2 (13), temperature, and dielectric properties on the ionization constants of buffering molecules.

Although quite general, these approaches are not applicable to complex biological systems, such as foods that typically contain many components that contribute significantly to buffering but are not identified or are of unknown concentration. When dealing with the compositional complexity of biological materials, the diversity of ionizable molecules means there will be many dissociation constants in almost any situation. However, as a result of this complexity, titration curves of biological materials tend to become smooth and relatively linear. Advantage was taken of these characteristics of titrations curves of complex materials by using local polynomial regression (9) to model the partial contribution to a titration from components with unknown concentrations or unknown dissociation constants. The distributions of ionic species from components with known concentration and dissociation constants, whether naturally present or added, can then be estimated in biological systems. The use of generalized cross-validation in selection of the smoothing parameter of the local regression allows for automatic control of model complexity. A nearest-neighbor bandwidth is employed to account for unequally spaced points in the titration data. Iteratively, re-weighted least squares are used to down-weight the influence of outlier data points. Since this local polynomial regression approach has been combined with the theoretical approach to ionization and buffer calculations of Van Slyke (8) and Butler and Cogley (1) for components with known dissociation constants and total concentrations, this hybrid

approach is referred to as the semi-mechanistic partial buffer (SMPB) approach.

The objective of this paper is to describe in detail the mathematical basis of the pHTools software and then to demonstrate the use of the SMPB approach to predict the amounts of different acids required to adjust pH in cucumber slurries.

MODELING METHODS

Generalizing Moles of Base Added (C_b). The central quantity of interest in modeling the properties of a buffer is the moles per liter of base added (C_b) to a solution (Note: Acid addition gives a negative C_b). Of particular interest to modeling buffers are the following: (i) a titration curve determines how C_b varies as a function of pH, (ii) buffer capacity is the partial derivative of C_b with respect to pH (8), and (iii) the initial pH of a solution corresponds to the pH at which C_b is equal to zero. When the dissociation constant(s) and total concentration of a buffer component are known, C_b may be related to pH by considering the Henderson equation for the dissociation step(s), along with the mass and charge balances of the solution (1). For an acid with a single dissociable proton, the Henderson equation is

$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]} \quad (1)$$

where K_a is the dissociation constant, $[\text{H}^+]$ is the molar concentration of protons, $[\text{A}^-]$ is the molar concentration of acid anions, and $[\text{HA}]$ is the molar concentration of the protonated acid. A more general expression that accounts for activities of each of the molecules will be discussed below. In aqueous solution with C_b moles of base (BOH) per liter and C moles per liter of acid ($[\text{HA}] + [\text{A}^-]$), the aqueous charge balance is

$$[\text{B}^+] + [\text{H}^+] = [\text{OH}^-] + [\text{A}^-] \quad (2)$$

Combining these two expressions, it is found that

$$C_b = C \frac{K_a}{[\text{H}^+] + K_a} + \frac{K_w}{[\text{H}^+]} - [\text{H}^+] \quad (3)$$

where K_w is the ion-product "constant" of water (1.00×10^{-14} (mol/L)² at 25 °C, 1 atm, and 0 ionic strength). By similar arguments, C_b for an acid having two dissociable protons is (2)

$$C_b = C \frac{[\text{H}^+]K_{a1} + 2K_{a1}K_{a2}}{[\text{H}^+]^2 + [\text{H}^+]K_{a1} + K_{a1}K_{a2}} + \frac{K_w}{[\text{H}^+]} - [\text{H}^+] \quad (4)$$

For an acid containing an arbitrary number of acidic and basic sites of dissociation, it can be shown that the titration can be modeled as [Proof is by induction and is given in Dougherty (10).]:

$$C_b = C \frac{\sum_{i=1}^N [(n+1-i)[\text{H}^+]^{i-1} \prod_{j=1}^{N+1-i} K_{aj}]}{\sum_{i=1}^{N+1} [[\text{H}^+]^{(i-1)} \prod_{j=1}^{N+1-i} K_{aj}]} + \frac{K_w}{[\text{H}^+]} - [\text{H}^+] \quad (5)$$

where n represents the number of $\text{p}K_a$ (acidic sites of dissociation) and N represents the total number of sites of dissociation

(number of pK_a plus number of pK_b). In cases where K_b are required (basic sites of dissociation), the j th K_b must be converted to a K_a using the relation:

$$K_{aj} = \frac{K_w}{K_{bj}} \quad (6)$$

Note that in this formulation, just as in eqs 3 and 4, each term in the summation corresponds to the partial contribution of each ion of the buffer to the overall titration. The quantity $(K_w/[H^+] - [H^+])$ represents the contribution to the titration from the dissociation of water into hydroxyl ions and protons, and the value $(n + 1 - i)$ may take on negative values and accounts for the effects of the basic sites of dissociation on the charge balance.

A more general formulation is required to handle buffers that are associated with electrolytes (e.g., Ca^{2+} , Mg^{2+} , Cl^-). Such buffers are usually referred to as being the "conjugate acid" or "conjugate base" form, although this terminology is somewhat imprecise and can lead to confusion. Some common examples include monosodium phosphate (NaH_2PO_4) and calcium lactate ($Ca[lactate]_2$). The charge balance equation must account for these electrolytes or so-called "counterions." The consequence of this in terms of a model for C_b is the following:

$$C_b = C \frac{\sum_{i=1}^N [(n + 1 - i)[H^+]^{i-1} \prod_{j=1}^{N+1-i} K_{aj}]}{\sum_{i=1}^{N+1} [[H^+]^{(i-1)} \prod_{j=1}^{N+1-i} K_{aj}]} + \frac{K_w}{[H^+]} - [H^+] - C \sum_{k=1}^{n_c} v_k \quad (7)$$

where v_k is the signed charge on the k th counterion and n_c is the total number of counterions per formula unit of the buffer. Thus, if the counterion is positive, there is a reduction in moles of base added, and if it is negative, there is an increase in the moles of base added. For example, an acid in its conjugate base form having a positive counterion has a higher initial pH when put into solution, while the opposite is expected for bases in their conjugate acid form. In general, the association of an acid or base with electrolytes simply shifts the C_b curve vertically by an amount proportional to the sum of the signed charges of the electrolytes.

The advantage of this general formulation is that other general quantities of interest can be derived from it. Since C_b is derived on the basis of charge balance, it is easy to modify eq 7 to provide a general equation for the concentration of each species of the acid in the buffer solution. This is done by recognizing that the $(n + 1 - i)$ factor represents the charge of the i th ion in the summation. By setting this factor to 1, the concentration of the i th ion is obtained in the i th term of the summation. Thus, the concentration of the acid ion with charge $(n + 1 - i)$, which is denoted $C_{(n+1-i)}$, is given by

$$C_{(n+1-i)} = C \frac{[H^+]^{i-1} \prod_{j=1}^{n+1-i} K_{aj}}{\sum_{i=1}^{n+1} [[H^+]^{(i-1)} \prod_{j=1}^{n+1-i} K_{aj}]} \quad (8)$$

The concentration of the nonionic form of the acid is determined

by subtraction of the concentrations of all the ionic forms of the acid from the total concentration (C):

$$C_{(0)} = C - \sum_{i=1}^n C_{(n+1-i)} \quad (9)$$

The concentration of electrolyte is simply the product of C and the number formula units of the electrolyte.

A general formula for the ionic strength contribution from the acid is also easily obtained from eq 7. The ionic strength of a solution is defined (2) as

$$I = 1/2 \sum_{i=1}^m z_i^2 C_i \quad (10)$$

where z_i is the charge on the i th ion in the solution and m is the total number of ions in the solution. Therefore, if the term $(n + 1 - i)$ in eq 7 is squared and the summation is multiplied by 1/2, a general formula for ionic strength contributions from the acid and water is obtained:

$$I = 1/2 \left(\frac{C \sum_{i=1}^N [(n + 1 - i)^2 [H^+]^{i-1} \prod_{j=1}^{N+1-i} K_{aj}]}{\sum_{i=1}^{N+1} [[H^+]^{(i-1)} \prod_{j=1}^{N+1-i} K_{aj}]} + \left(\frac{K_w}{[H^+]} + [H^+] \right) + C \sum_{k=1}^{n_c} v_k^2 \right) \quad (11)$$

Each term in the summation of eq 11 represents the partial contribution of each ionic form of the buffer to the overall ionic strength as well as contributions from the n_c electrolytes. The uncharged form of the buffer (which may include zwitterions) is assumed to not contribute to the ionic strength and is ignored. When calculating the total ionic strength of a solution, it is important to not neglect the contribution of ions from the titrant. For example, if one assumes a +1, -1 electrolyte such as HCl or NaOH, then the ionic strength contribution from that titrant is equal to $C_b/2$. For more "general titrants" including 2-1-1 electrolytes, etc. or titrants containing complex or unknown components, the reader is referred to next section, which concerns generalized titrations.

Generalizing Buffer Capacity. Along with the fact that $(d[H^+]/dpH) = -\ln(10)[H^+]$, eq 7 can be differentiated with respect to $[H^+]$ to give a formula for the buffer capacity, β , of a buffer containing N dissociable groups:

$$\beta = -\ln(10)[H^+] \left[\left(\frac{BA' - AB'}{B^2} \right) - \frac{K_w}{[H^+]^2} - 1 \right] \quad (12)$$

where

$$A = C \sum_{i=1}^N [(n + 1 - i)[H^+]^{i-1} \prod_{j=1}^{N+1-i} (K_{aj})] \quad (13)$$

$$B = \sum_{i=1}^{N+1} [[H^+]^{i-1} \prod_{j=1}^{N+1-i} (K_{aj})] \quad (14)$$

$$A' = C \sum_{i=2}^N [(n+1-i)(i-1)[H^+]^{i-2} \prod_{j=1}^{N+1-i} (K_{aj})] \quad (15)$$

$$B' = \sum_{i=2}^{N+1} [(i-1)[H^+]^{i-2} \prod_{j=1}^{N+1-i} (K_{aj})] \quad (16)$$

Note that terms corresponding to ions from neutral salt are absent since these are assumed to not vary as a function of pH.

The above approaches may also be extended to mixtures of buffers. Since the Bronsted-Lowry theory is assumed, each of the dissociation steps among the acids and bases in a solution is independent, except for the common ions (H^+ and OH^-). As a consequence, the contribution from each of the buffer compounds (each of which may have multiple acidic or basic dissociation sites) is summed with the contributions from the dissociation of water included only once.

Ionic Strength, Temperature, and Dielectric Corrections.

It is well-known that the dissociation relation in eq 1 is only an approximation (11). In fact, each of the ions has associated with it an activity coefficient, γ :

$$K_a = \frac{[H^+]\gamma_+[A^-]\gamma_-}{[HA]\gamma_0} \quad (17)$$

It is also true that $pH = [H^+]\gamma_+$, but this fact is often ignored since pH meters measure hydrogen ion activity. The activity coefficients depend primarily on the ionic strength, the temperature, and the dielectric properties of the solution. Other important factors include the charges of the ions involved and their atomic radii. Although theoretical developments have been undertaken that account for all of these factors, acceptable and sometimes more accurate estimates have been obtained using the Davies equation (12) or one of its modifications (1). This is especially true for low to moderate ionic strengths. The Davies equation provides accurate estimates of activity coefficients for a wide range of electrolyte solutions up to ionic strength 0.2–0.5 M (1, 11, 13). For the charged ion pairs in the numerator of eq 17, the geometric mean of the single ion activities is calculated by

$$-\log_{10}(\gamma_{\pm}) = A|z_+z_-|\left(\frac{\sqrt{I}}{1+\sqrt{I}} - bI\right) \quad (18)$$

where the z_* are the unsigned charges on the ions, $A = 1.825 \times 10^6 (\epsilon T)^{-3/2}$, ϵ is the dielectric constant of the solvent ($\epsilon = 78.3808$ for water (9)), T is the temperature in Kelvin, and b is the “salting-out” parameter. However, if a molecule is neutral (as in the denominator of eq 1) the activity is predicted by

$$-\log_{10}(\gamma_0) = -bI \quad (19)$$

Thus, we suggest that a tabulated pK_a value (pK_a^{tab}), which was measured under one set of conditions may be corrected to a pK_a under a new set of conditions (pK_a^{new}) using the formula:

$$pK_a^{\text{new}} = pK_a^{\text{tab}} - 2 \log_{10}\left(\frac{\gamma_{\pm}^{\text{new}}}{\gamma_{\pm}^{\text{tab}}}\right) + \log_{10}\left(\frac{\gamma_0^{\text{new}}}{\gamma_0^{\text{tab}}}\right) \quad (20)$$

Most tabulated pK_a and pK_b are estimates that have been extrapolated to zero ionic strength and so this formula simplifies to

$$pK_a^{\text{new}} = pK_a^{\text{tab}} - 2 \log_{10}(\gamma_{\pm}^{\text{new}}) + \log_{10}(\gamma_0^{\text{new}}) \quad (21)$$

For molecules with multiple pK_a , the denominator term of the Henderson equation is not always neutral. In such cases, eq 18 is usually replaced by an approximation of the form: $Az^2(\sqrt{I}/(1+\sqrt{I}) - bI)$ with the appropriate nonzero charge z .

Although the Davies equation can be used to calculate activity coefficients, it requires ionic strength as an input. The activity coefficients, however, affect the extent of dissociation in the buffer components, which in turn affects ionic strength. To achieve optimal activity coefficient predictions in such a coupled system, an iterative refinement procedure was employed. Convergence of activity coefficient estimation usually occurred after three to four iterations.

The Davies or Samson equation can be used to predict the effect of the dielectric properties of the solvent on the buffer properties. Unfortunately, there are currently no general mathematical models for predicting the dielectric constant of a buffer solution as a function of its composition. It is known, however, that the dielectric “constant” varies with temperature, density, concentration of ions, and physical properties of the ions in solution (14). The only general approach currently available is to utilize tabulated empirically measured values of dielectric constants for buffer solutions or to measure the dielectric constant directly. Datta et al. (15) tabulated dielectric properties from many different literature sources for many different foods. Nelson and Datta (14) gave dielectric properties of several fresh fruits and vegetables and milk products. However, different measurement techniques give somewhat different values for complex mixtures, and, although single pure compounds have been well-studied, the interactions of various food components are poorly understood (15).

Properties of Complex Buffers. The equations given in the preceding sections are applicable as long as the concentrations of all acids and bases and their dissociation equilibria are known. Unfortunately, these requirements are usually not met in food and biological applications due to the complexity and undefined nature of the systems. Although the buffer capacities of such solutions can be quite complex chemically, it is reasonable to assume that the relationship between C_b and pH should be smooth and continuous. While a titration can be used to determine the relationship between C_b and pH, the accuracy of pH measurements is typically limited to ± 0.01 units in common laboratory settings. It is also not possible to observe the effect of adding a small amount of base at every pH value for which a prediction of C_b is desired. A flexible regression procedure is, therefore, required that will provide a continuous differentiable prediction function for any C_b given pH.

Generalized Additive Modeling of Complex Buffers. In the previous section, the additivity of C_b was noted. This will be used to advantage here where a solution is assumed to contain both simple and complex buffer components. Let C_{bi} be the sum of the contributions from the simple and complex buffer components at pH_i . Thus, a titration data set consists of the ordered pairs (pH_i, C_{bi}) , $i \in \{1, 2, \dots, Q\}$, where Q is the total number of data points in the titration. It is assumed that there are M simple buffer components where the j th one has N_j dissociable groups, n_j acidic groups, and n_{cj} electrolytes per formula unit, and there is also a complex buffer component. Then the following model can describe the C_b relation:

$$C_b(\text{pH}_i) = \sum_{j=1}^M \left(\frac{C_j \sum_{k=1}^{N_j} [(n_j + 1 - k)[\text{H}^+]^{k-1} \prod_{l=1}^{N_j+1-k} K_{ajl}]}{\sum_{k=1}^{N_j+1} [[\text{H}^+]^{(k-1)} \prod_{l=1}^{N_j+1-k} K_{ajl}]} + C_j \sum_{k=1}^{n_{cj}} v_{kj} \right) + \frac{K_w}{[\text{H}^+]} - [\text{H}^+] + F(\text{pH}_i) + \epsilon_i \quad (22)$$

where $[\text{H}^+] = 10^{-\text{pH}_i}$, ϵ_i is a random error assumed to come from a Gaussian distribution with zero mean and unknown variance σ^2 , and $F(\text{pH}_i)$ is the contribution from the complex media in which the weak acids are dissolved. Since this model blends both theoretical and statistical models of partial buffer components, we have called it the semi-mechanistic partial buffer (SMPB) model. To fit this model to titration data, the following algorithm is employed. First, the simple buffer contributions (eq 7), which are assumed to be known without error, are subtracted from the observed C_{bi} . The result of this subtraction is the unexplained partial contribution to C_{bi} from the unknown components. Then, $F(\text{pH}_i)$ is regressed on the partial contribution to C_{bi} from the unknown components using local polynomial regression. This entails estimating a global smoothing parameter by generalized cross-validation. Briefly, as the smoothing parameter is decreased, $F(\text{pH}_i)$ tends to interpolate the Q data points and approaches a $(Q - 1)$ th order polynomial. For larger values of the smoothing parameter, $F(\text{pH}_i)$ smooths the data and approaches a p th order polynomial where p is the underlying order of polynomial model being used. Since a titration requires careful execution and may be sensitive to human error either in preparation of the titrants or in dispensing of the aliquots, it is desirable to have an automatic method for dealing with outliers. Thus, the final step in the algorithm utilizes iteratively reweighted least squares (IRWLS) to down-weight data corresponding to large residuals (16). However, during a titration it may be desirable or necessary to focus pH readings in regions of the titration curve, where the pH is changing relatively fast while sampling less densely in other regions. To account for unequal sampling, a nearest-neighbor bandwidth is used (9).

Buffer Capacity of a Complex Buffer. An estimate of the buffer capacity of a complex buffer is easily obtained from the least-squares estimate of C_b outlined above. Since at each pH a polynomial of degree p is fit to the data, the partial derivative of F with respect to pH evaluated at pH_i is estimated by

$$\frac{\partial \hat{F}}{\partial \text{pH}_i} = \hat{b}_{i,2} \quad (23)$$

where $\hat{b}_{1,2}$ is the second parameter (i.e., specifically the linear term in the polynomial) estimated at the point pH_i . This partial contribution to the buffer capacity from the complex components is then added to the contributions from the simple components and water (see eqs 12–16) to give an estimate of the total buffer capacity. Since an estimate of the first derivative of C_b is needed for calculating the buffer capacity, the order of the underlying polynomial model (p) must be at least 1 (linear).

Lower Bound for the Ionic Strength Contribution From Complex Buffers. Since there are ionic species in complex

buffers, the ionic strength will be greater than zero. A lower bound for the ionic strength contribution from a complex buffer can be estimated. Consider the charge balance of a solution of NaOH and C mol/L of mono-protic weak acid HA:

$$[\text{Na}^+] + [\text{H}^+] = [\text{OH}^-] + [\text{A}^-] \quad (24)$$

which is rearranged to give C_b :

$$[\text{Na}^+] = [\text{A}^-] + [\text{OH}^-] - [\text{H}^+] \quad (25)$$

$$C_b = \frac{CK_a}{K_a + [\text{H}^+]} + \frac{K_w}{[\text{H}^+]} - [\text{H}^+] \quad (26)$$

Comparing this to the ionic strength of the solution:

$$I = 0.5([\text{Na}^+] + [\text{OH}^-] + [\text{A}^-] + [\text{H}^+]) \quad (27)$$

it is noticed that $I = C_b + [\text{H}^+]$. Unfortunately, this relationship does not hold in the presence of buffers with multiple dissociable groups or salts with counterions with charges greater than 1 or -1 are present because of the squaring of charges in eq 11. However, from comparing eqs 7 and 11, it is clear that $I \geq (1/2)C_b + [\text{H}^+]$. Since this lower bound may be negative, we suggest $\max(0, (1/2)C_b + [\text{H}^+])$ as a lower bound for the ionic strength contribution from complex buffer components. We view this as a convenient improvement over the zero ionic strength assumption.

Generalized Titrations. Normally, titrations are performed using strong acids or strong bases. Because these titrants dissociate essentially completely, their contribution to C_b is easily calculated (see **Figure 1A**). In many instances, (e.g., food applications) weak acids or bases are used instead of strong acids or bases to adjust pH. To use data from such titrations in calculating buffer capacities, a more general approach to calculating C_b must be taken. When a weak acid or weak base is used as the titrant, the process of inferring a relationship between pH and C_b will be referred to as a generalized titration. The contribution to C_b from a weak acid or base will vary depending on the extent to which it dissociates and, thereby, depend on the pH and the ionic strength (see **Figure 1B**). For clarity, let $C_b(\text{pH})$ indicate C_b at a particular pH. If the partial contribution of the weak titrant to $C_b(\text{pH})$ is denoted by $C\tau$ (pH) then

$$C_b(\text{pH}) = -C\tau(\text{pH}) \quad (28)$$

That is, the moles of base added is the negative of the moles of base that would be required to bring the buffer τ to the pH in question. Similarly, for complex buffers containing more than one buffering compound τ_1, τ_2, \dots , the additivity of the partial C_b can be exploited and the following definition made:

$$C_b(\text{pH}) = -(C\tau_1(\text{pH}) + C\tau_2(\text{pH}) + \dots + C\tau_n(\text{pH})) \quad (29)$$

That is, the number of moles of base added per liter is equal to the negative of the sum of the contributions from the weak acids or bases. In such a context, it is possible to use a complex buffer as a titrant. To illustrate these points, consider **Figure 1A**, which is a plot of the partial C_b of a 0.01 M NaOH solution over the pH range 2–12. Over this interval, the contribution to C_b remains a constant 0.01 M because NaOH is a strong base and its dissociation is essentially complete over this pH range. In contrast, consider **Figure 1B**, which gives the partial contributions to C_b from a 0.01 M solution of gluconic acid ($\text{p}K_a$ 3.6). At low pH, this weak acid is completely protonated and,

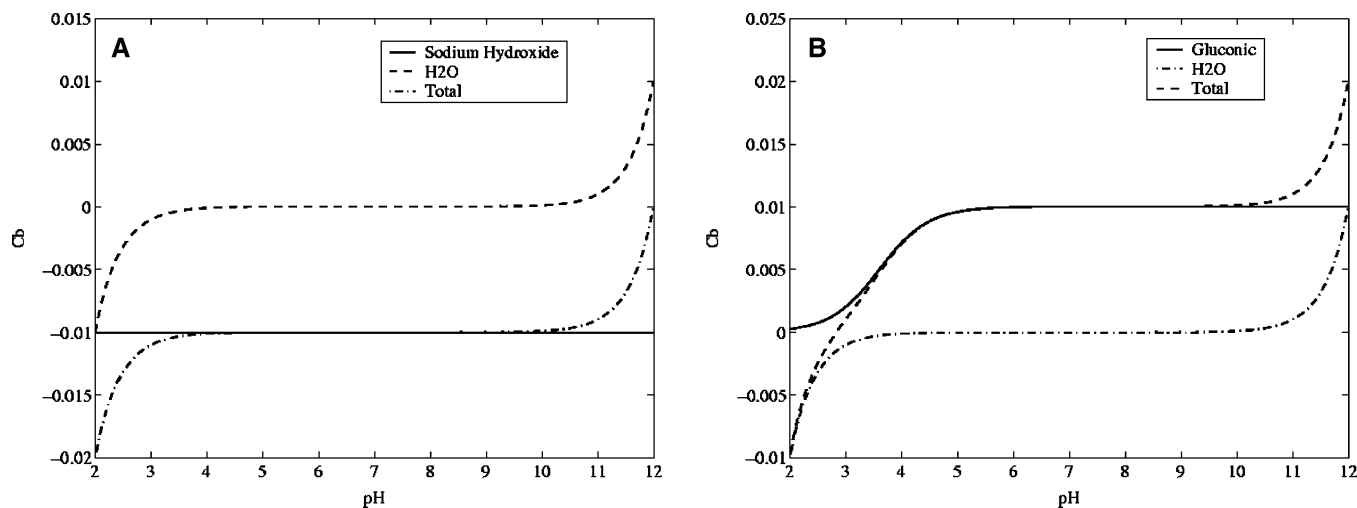


Figure 1. (A) Partial contributions to C_b from the components of a 0.01 M NaOH solution. The partial C_b (solid line) is nearly constant and represents the amount of the species available for titration. (B) Partial contributions to C_b derived from the components of a 0.01 M gluconic acid solution. The partial C_b (solid line) is the relevant quantity in a generalized titration since it represents the amount of the species available for titration.

therefore, makes no contribution to C_b . Conversely, at high pH all of the gluconic acid is dissociated and its contribution to C_b approaches an amount equivalent to the negative of its concentration (i.e., the conjugate base of a weak acid is a strong base). Clearly, to determine the moles of base available for titration, the dissociation of the weak titrant at the particular pH in question needs to be considered. Thus, for generalized titrations, determination of a mathematical relationship between C_b and pH requires a function which maps pH to C_b (an approach which at first may seem counter-intuitive to those acquainted only with strong acid or strong base titrants). It is therefore our convention to treat pH as the predictor variable and C_b as the response variable.

Initial pH Determination. Since C_b represents the moles of base added per liter of titrant as a function of pH, it is possible to use C_b to determine the initial pH of a simple or complex buffer. Consequently, determining the initial pH of a simple buffer requires finding the zero of a polynomial where the order of the polynomial is equal to the maximum number of dissociable groups of any molecule in the solution. For complex buffers, determination of the initial pH is similar and requires finding the zero of the additive model:

$$C_b(\text{pH}) = \sum_{j=1}^M \left(\frac{C_j \sum_{k=1}^{N_j} [(n_j + 1 - k)[\text{H}^+]^{k-1} \prod_{l=1}^{N_j+1-k} K_{ajl}]}{\sum_{k=1}^{N_j+1} [[\text{H}^+]^{(k-1)} \prod_{l=1}^{N_j+1-k} K_{ajl}]} + C_j \sum_{k_j=1}^{n_{c_j}} v_{k_j} \right) + \frac{K_w}{[\text{H}^+]} - [\text{H}^+] + F(\text{pH}) \quad (30)$$

In either case, numerical methods are generally required. Any standard scalar zero-finding algorithm is generally sufficient. We have found good success with the standard Nelder–Mead simplex method (17). However, since pH can theoretically take any value (0 to 14) and $C_b(\text{pH})$ can be somewhat flat near $\text{p}K_a$, we have implemented Nelder–Mead with optional upper and lower bounds on the pH.

pH Adjustment to Target Values. It is often desirable to determine the amount of a primary buffer (simple or complex) to add to a secondary buffer (simple or complex) to attain a target pH. The first approach one may think of is to use an iterative approach where an amount of the primary buffer is guessed and then the initial pH of the mixture is determined. If this pH is further from the target pH than some tolerance, an improved guess is generated. Upon iteration of this process one should arrive at the desired amount. Although such an approach will work, it is computationally wasteful since an additional optimization (initial pH determination) is required at each step. A better method is to find the amount of the primary buffer such that the C_b of the mixture when evaluated at the target pH is zero. This slight modification achieves the same result, but does not require a separate optimization at each iteration. Again, any standard scalar root-finding algorithm should suffice for carrying out such an optimization and we have had good results using the Nelder–Mead simplex method.

The predicted amounts of required primary buffer are sensitive to errors in titration. This is especially so when attempting to realize a target pH that is in a range where the titrant's buffer capacity is low. A contributing factor to this variability is the limited accuracy of pH meters. The accuracy of most laboratory pH meters is ± 0.01 pH units. A more conservative approach is to estimate the amounts required at $\pm \delta$ from the target pH, where δ is an acceptable measure of pH meter sensitivity. Such an approach allows a researcher, for example, to be fairly confident about adding the amount corresponding to the lower estimate and then carefully adjusting to the target pH by addition of buffer.

The methods developed above provide a transparent means of using the computational formulas given for buffers with known dissociation constants and extending them to complex buffers containing both known and undefined components. The Appendix describes a general database framework that allows application of these methods to a wide range of both types of buffers. In the next section, we validate these approaches in experimental setting.

MATERIALS AND METHODS

Standard titrations of fresh cucumber slurries were done with a strong acid, HCl. The amounts of four weak acids required to reach certain target pH values were calculated using the generalized titration

approach. Then the amounts of weak acids required to reach the target pH values were determined experimentally and compared to the predicted amounts of weak acids. Acids and NaCl were reagent grade from Sigma-Aldrich (St. Louis, MO). Solutions were prepared with deionized water. Four lots of size 2A (25–32 mm diameter) fresh cucumbers were supplied by a local commercial processor.

Cucumbers were blended for 2 min in a Waring blender (Conair Corporation, Torrington, CT). Slurries for titration were prepared by mixing blended cucumbers with an equal weight of water. Duplicate slurries (100 g each) from each of three lots of cucumbers were titrated with standardized 2.843 M HCl at $25 \pm 1^\circ\text{C}$. A slurry was constantly stirred on a magnetic stir plate as at least 25 aliquots of HCl solution were accurately dispensed into the slurry with motorized microliter pipets (Rainin Instruments, Oakland, CA). The pH of a slurry was measured and recorded before addition of HCl and then after addition of each aliquot of HCl. The pH was determined using an Accumet Research AR25 pH meter with an AccuFet solid-state electrode (Fisher Scientific, Atlanta, GA). The relative accuracy of the meter was ± 0.002 pH units. Standardization of the pH meter was done at 25°C with pH 4.00 ± 0.01 and pH 7.00 ± 0.01 standard buffers. Titrations were carried out starting at the natural pH of the cucumber slurry (approximately 6.4) and continued until the pH decreased to at least 2.6. Data points from all six HCl titrations (two slurries from each of three lots of cucumbers) were entered into pHTools , along with temperature, ionic strength, and, if known, dielectric constant. Titration data were fitted using local polynomial regression to obtain a single calibration titration curve to make predictions for pH adjustments with other acids.

The millimoles of acetic acid, malic acid, citric acid, and phosphoric acid required to achieve target pH values of 4.50, 4.00, 3.70, and 3.30 in cucumber slurry samples with water only and with 0.3 M NaCl were predicted using pHTools based upon the HCl calibration titration curve. Triplicate slurry samples from a lot of cucumbers different from the three lots used for HCl titration were prepared with and without the addition of 0.3 M NaCl. Solutions of 3.05 M acetic acid, 1.52 M malic acid, 1.01 M citric acid, and 2.21 M phosphoric acid were used as titrants. The predicted and experimental amounts of each acid required to hit the four target pH values were plotted as a function of pH.

The dielectric constant of cucumber slurries was determined with a digital network analyzer (HP 8753C, Agilent Technologies, Palo Alto, CA) with an open-ended coaxial probe (HP 85070B, Agilent). The instrument was calibrated with air using the procedure recommended by the manufacturer. The instrument was allowed to equilibrate for 2 h before measurements were made. Data were acquired with a PC-compatible computer equipped with a National Instruments GP-IB card (National Instruments, Austin, TX) and the HP 85070B v. 1.0 software provided with the instrument.

RESULTS AND DISCUSSION

Dielectric Constant of Cucumber Slurries. The dielectric constant for cucumber slurries was 78.01 ± 0.03 without addition of NaCl. When 0.3 M NaCl was added to cucumber slurries, the dielectric constant was 74.27 ± 0.94 . This compares to a dielectric constant of 78.38 for water (11).

Comparison of Predicted and Measured Amounts of Acids. Figure 2 shows the calibration curve of fresh cucumber slurry used to predict the amount of titrant required to acidify slurries in the presence and absence of 0.3 M NaCl to pH 4.5, 4.0, 3.7, and 3.3. Duplicate titrations were very close, but there were differences among lots of cucumbers as would be expected for different lots of fresh produce.

Figure 3 shows the experimental and predicted quantity of titrant required to reach various target pH values in cucumber slurries prepared without addition of salt by addition of deionized water to the blended cucumbers. Bars on the predicted curves show how much the predicted amount of acid changes with a ± 0.05 pH unit increase or decrease in the target pH. The error bars on the experimental curves show the standard deviation for pH adjustments on triplicate samples of cucumber

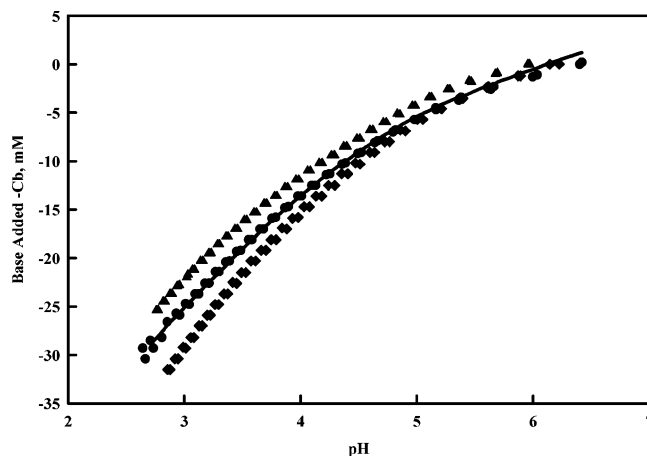


Figure 2. Titration curves of duplicate samples (same symbols) from three lots (different symbols) of cucumbers. Solid line is the average titration curve.

slurry. Predictions are reasonably close to the amount of acid required to reach the target pH for all four acids at all four target pH values. The largest deviations between pHTools predictions and the amounts of acid required experimentally were observed for phosphoric acid. Deviations were in the direction of underestimating the amount of phosphoric acid required and increased as pH decreased. The underestimation was 16% at pH 3.30.

The use of the Davies equation in pHTools is illustrated in Figure 4 to mathematically account for changes in ionic strength when salt, in addition to acids, was added to cucumber slurries. Addition of salt increases the dissociation of acids (lowers the $\text{p}K_a$ values) so that less acid is required to lower pH to designated target values. This can be seen in a comparison of Figures 3 and 4 for each acid used for pH adjustment. The pHTools calculations gave reasonable predictions for the amounts of acetic, malic, citric, and phosphoric acids required to reach the target pH levels of 4.5, 4.0, 3.7, and 3.3 when salt was added. This result indicates that the Davies equation reasonably accounts for ionic strength changes with a complex buffer system like a cucumber slurry. The errors in the predicted amount of acids required were in the direction of over estimating the amount of acid required to lower pH. There was an over-prediction of about 2 mM in the amount of citric acid required over the pH range of 3.1–4.5. For malic acid and phosphoric acid, deviations were larger at the higher pH values and became smaller at lower pH.

The combination of local polynomial modeling of the complex buffer components with the theoretical modeling of simple buffer components has been found to provide a general and robust method for predicting species concentration, buffer capacity, and carrying out target pH optimizations. Applications of this approach have already been seen by this laboratory in suggesting a novel hypothesis for the chemical basis of sour flavor (4, 5) and in studying the effect of fully protonated organic acid molecules on the mortality of *E. coli* O157:H7 in low pH solutions (6). In the future, it will be important to determine the effects of preservative acids on pathogen survival in food products as well as in defined buffer solutions. In addition to its use in calculation of the distribution of the different protonated forms of compounds with ionizable groups, pHTools may be useful for subtracting the buffer contribution of known components from the titration curves of biological materials in order that it could be determined how much of the buffer capacity is a result of unknown components in the system.

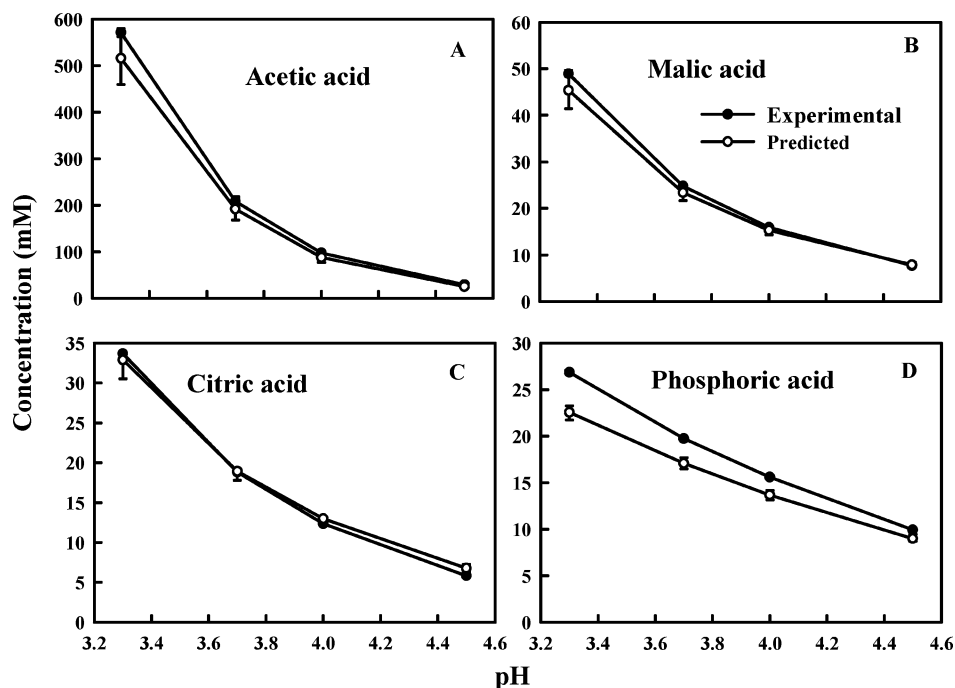


Figure 3. Comparison of experimental amounts of acids required to adjust 100 g of cucumber slurries without added NaCl to target pH values with the amounts of the acids predicted by pHTools . Error bars on the experimental data show the standard deviation of titration of duplicate samples. Error bars on the predicted values show the predicted amounts of acids ± 0.05 pH unit of the target pH.

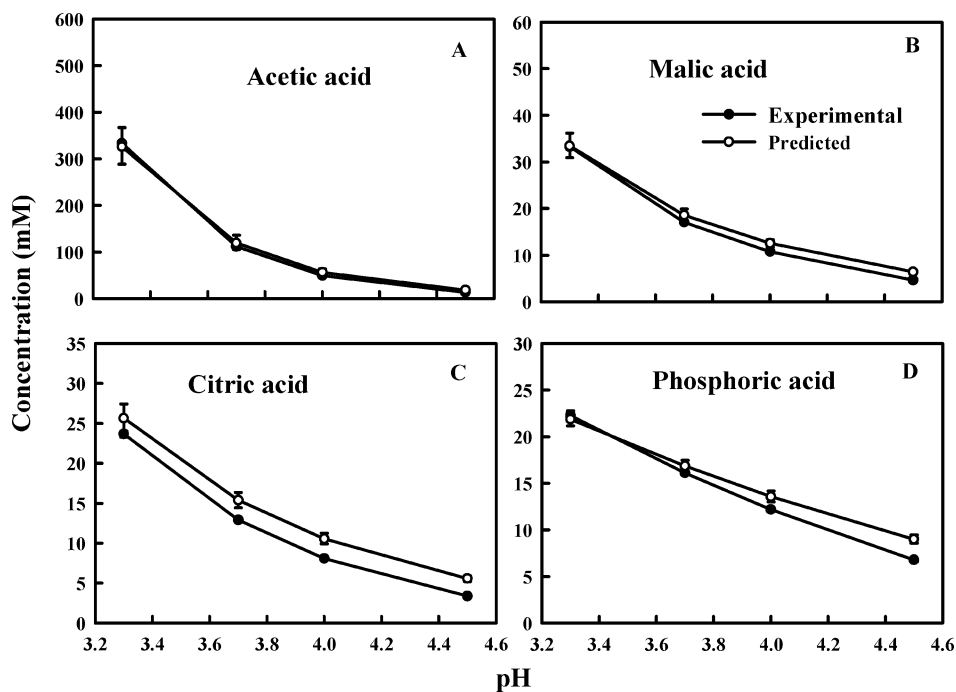


Figure 4. Comparison of experimental amounts of acids required to adjust 100 g of cucumber slurries with 0.3 M added NaCl to target pH values with the amounts of the acids predicted by pHTools . Error bars on the experimental data show the standard deviation of titration of duplicate samples. Error bars on the predicted values show the predicted amounts of acids ± 0.05 pH unit of the target pH.

The pHTools software for use in MATLAB may be downloaded at <http://www.mathworks.com/matlabcentral/fileexchange/index.jsp>. The software allows a user to build and share databases of titrations relevant to their research (see Appendix).

APPENDIX

Tableau Approach for Modeling Buffers. To cover the variety of Bronsted–Lowry-type buffers possible, a tableau approach was used to specify each acid and base. For purposes

of standardization and ease of implementation, we assume that the information in the tableau always refers to the buffer as it is found in its “native form” (i.e., the form in which it exists prior to be placed in aqueous solution). This is necessary to discriminate the various formulations available from chemical supply companies, most notably with regard to salts of conjugate acid and conjugate base formulations as described in **Table 1**.

The three components of the tableau approach are (i) PKA, (ii) PKB, and (iii) SALTTYPE. As the names suggest, the PKA

Table 1. Tableau Approach

	acetic	calcium lactate	cysteine	sodium bicarbonate	magnesium chloride
pK _a	[4.75]	[], [3.86], [3.86]	[1.8, 8.3]	[], [6.37, 10.31]	[], [], []
pK _b	[]	[], [], []	[3.2]	[], []	[], [], []
salt type	[]	[2], [-1], [-1]	[]	[1], [-1]	[2], [-1], [-1]

component stores the (typically) zero-extrapolated $-\log(K_a)$, the PKB component stores the zero-extrapolated $-\log(K_b)$, and the SALTTYPE component stores the charges associated with the ions that result upon solvation (i.e., acids or bases ionically bound to electrolytes, the electrolytes of salts, etc.). Although pK_a and pK_b may be interconverted, it is the convention of the tableau approach that pK_b are only used to represent basic dissociations (i.e., lead to conjugate acids more positively charged than the native form of the buffer). For a given buffer, each of these components is represented as a list of numbers separated by brackets ([]). The number of brackets in the PKA, PKB, and SALTTYPE components is, therefore, always equal. We illustrate the usage of this tableau approach with the following buffers: acetic acid, calcium lactate, cysteine, sodium bicarbonate and magnesium chloride.

The native form of acetic acid is not a salt, so it gets just 1 bracket and has a single dissociable proton with pK_a = 4.75. Lactic acid is an acid with a single dissociable proton with pK_a = 3.86. In this formulation, lactic acid is in its conjugate base form with two lactic acid anions being ionically bound to a single calcium cation. Cysteine is an amino acid with a carboxylic acid group with pK_a = 1.8, a basic amino group with pK_b = 3.2, and an acidic R-group (a thiol) with pK_a = 8.3. Interestingly, for intermediate pH values the amino group can take on a positive charge, while the carboxylic acid can take on a negative charge, and the thiol R-group is uncharged. Zwitterionic molecules carrying simultaneous positive and negative charges are easily handled by the tableau approach. Sodium bicarbonate is a conjugate base of carbonic acid (H₂CO₃) that is ionically bound to sodium and has two dissociable protons with pK_a = 6.37 and 10.31. Magnesium chloride is a salt. Once in aqueous solution the two chloride ions and the magnesium ion dissociate. None of the resulting ions have sites for acid or base dissociation.

It should be noted that buffers in the database may have all components left as empty brackets. In such cases, the buffer is typically a mixture of other defined buffers or may be associated with a local polynomial regression of experimental titration data (complex buffer).

LITERATURE CITED

- (1) Butler, J. N.; Cogley, D. R. *Ionic Equilibrium: Solubility and pH Calculations*; John Wiley and Sons: New York, 1998.
- (2) Horiuchi, J.; Kikuchi, S.; Kobayashi, M.; Kanno, T.; Shimizu, T. Modeling of pH response in continuous anaerobic acidogenesis by an artificial neural network. *Biochem. Eng. J.* **2001**, *9*, 199–204.

- (3) Van Vooren, L.; Van De Steene, M.; Ottoy, J.; Vanrolleghem, P. Automatic buffer capacity model building for the purpose of water quality monitoring. *Water Sci. Technol.* **2001**, *43*, 105–113.
- (4) Johanningsmeier, S. D.; McFeeters, R. F.; Drake, M. A. A hypothesis for the chemical basis for perception of sour taste. *J. Food Sci.* **2005**, *70*, R44–R48.
- (5) Da Conceicao Neta, E. R.; Johanningsmeier, S. J.; Drake, M. A.; McFeeters, R. F. A chemical basis for predicting sour taste intensity in acid solutions. *Abstract of Institute of Food Technologists Meeting*, 2006.
- (6) Bjornsdottir, K.; Breidt, F., Jr.; McFeeters, R. F. Protective effects of organic acids on survival of *Escherichia coli* 0157:H7 in acidic environments. *Appl. Environ. Microbiol.* **2006**, *72*, 660–664.
- (7) Breidt, F., Jr. A genomic study of *Leuconostoc mesenteroides* and the molecular ecology of sauerkraut fermentations. *J. Food Sci.* **2004**, *69*, 30–32.
- (8) Van Slyke, D. On the measurement of buffer values and on the relationship of buffer value to the dissociation constant of the buffer and the concentration and reaction of the buffer solution. *J. Am. Chem. Soc.* **1922**, *52*, 525–570.
- (9) Fan, J.; Gijbels, I. *Local Polynomial Modelling and its Applications*; Chapman and Hall: London, 1996.
- (10) Dougherty, D. P. Deterministic and semi-mechanistic approaches in predictive fermentation microbiology. Ph.D. Thesis, North Carolina State University, Raleigh, NC, 2002.
- (11) Pitzer, K. S. *Activity Coefficients in Electrolyte Solutions*; CRC Press: Boca Raton, FL, 1991.
- (12) Davies, P. T.; Tso, M. K. S. Procedures for reduced rank regression. *Appl. Stat. J. R. St. C* **1982**, *31*, 244–255.
- (13) Samson, E.; Lemaire, G.; Marchand, J.; Beaudoin, J. Modeling chemical activity effects in strong ionic solutions. *Comput. Mater. Sci.* **1999**, *15*, 285–294.
- (14) Nelson, S. O.; Datta, A. K. Dielectric properties of food materials and electric field interactions. In *Handbook of Microwave Technology for Food Applications*; Datta, A. K., Anantheswaran, R. C., Eds.; Marcel Dekker: New York, 2001.
- (15) Datta, A. K.; Sun, E.; Solis, A. Food dielectric property data and their composition-based prediction. In *Engineering Properties of Foods*, 2nd ed.; Rao, M. A., Rizvi, S. S. H., Eds.; Marcel Dekker: New York, 1995.
- (16) Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836.
- (17) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical recipes. Fortran: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1992; pp 402–406.

Received for review December 15, 2005. Revised manuscript received May 15, 2006. Accepted May 17, 2006. Paper No. FSR05-34 of the Journal Series of the Department of Food Science, North Carolina State University, Raleigh, NC. Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or North Carolina Agricultural Research Service, nor does it imply approval to the exclusion of other products that may be suitable. We acknowledge the support of the National Science Foundation Agreement No. 0112050 and a Dutton Fellowship from the Lyman Briggs School of Science.

JF0531508