Multigrid Solution of the Nonlinear Poisson-Boltzmann Equation and Calculation of Titration Curves

Himanshu Oberoi*‡ and Norma M. Allewell‡

‡Department of Biochemistry, University of Minnesota, St. Paul, Minnesota 55108; *Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, Connecticut 06457 USA

ABSTRACT Although knowledge of the pK_a values and charge states of individual residues is critical to understanding the role of electrostatic effects in protein structure and function, calculating these quantities is challenging because of the sensitivity of these parameters to the position and distribution of charges. Values for many different proteins which agree well with experimental results have been obtained with modified Tanford-Kirkwood theory in which the protein is modeled as a sphere (reviewed in Ref. 1); however, convergence is more difficult to achieve with finite difference methods, in which the protein is mapped onto a grid and derivatives of the potential function are calculated as differences between the values of the function at grid points (reviewed in Ref. 6). Multigrid methods, in which the size of the grid is varied from fine to coarse in several cycles, decrease computational time, increase rates of convergence, and improve agreement with experiment. Both the accuracy and computational advantage of the multigrid approach increase with grid size, because the time required to achieve a solution increases slowly with grid size. We have implemented a multigrid procedure for solving the nonlinear Poisson-Boltzmann equation, and, using lysozyme as a test case, compared calculations for several crystal forms, different refinement procedures, and different charge assignment schemes. The root mean square difference between calculated and experimental pK_a values for the crystal structure which yields best agreement with experiment (1LZT) is 1.1 pH units, with the differences in calculated and experimental pK values being less than 0.6 pH units for 16 out of 21 residues. The calculated titration curves of several residues are biphasic.

INTRODUCTION

Because electrostatic effects play a central role in protein structure and function and are at best only partially accessible through experiment, methods for modeling electrostatic interactions in crystal structures are of great interest. Both microscopic and macroscopic models have been developed and applied to a variety of systems. Recent reviews include Refs. 1–8.

Methods in which the protein is mapped onto a grid have the advantage of allowing the protein-solvent interface to be treated. Warwicker and Watson (9) showed that the method of finite differences, in which derivatives are approximated as differences between values at grid points, can be used to calculate the potential function of the grid. This approach has since been further developed and applied to a number of systems (cf. Refs. 6, 9, 10–16).

Extending the method of finite differences to calculating pK_a values and titration curves that can be compared with experimental results and used to develop models of protein function has proven challenging because of the sensitivity of these functions to errors in the estimated potential at titrating sites. Approaches that have been used to improve the accuracy of these estimates include *focusing*, in which fine grids on the order of 0.25 Å are used in the vicinity of titrating sites, *rotational averaging*, in which potentials are averaged over several orientations of the molecule relative to the grid (12), and a hybrid *statistical mechanical/Tanford-Roxby* approach (17). The multigrid method (18–23 and references therein) is a mathematically rigorous approach which increases substantially the speed and accuracy of finite difference calculations. Its use to solve the *linear* Poisson-Boltzmann equation has recently been reported (24). We have developed a self-consistent approach in which multigrid methods are used to solve the *nonlinear* Poisson-Boltzmann equation and to calculate potentials, pK_a values, and titration curves. The nonlinear Poisson-Boltzmann equation eliminates the approximations that invalidate the linearized form of the Poisson-Boltzmann equation at higher ionic strengths, but converges much less readily. This approach has been applied to crystal structures of lysozyme as a test case.

METHODS

Potentials

We use the nonlinear Poisson-Boltzmann equation (25)

$$\nabla \cdot [(\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r})\sinh[\phi(\mathbf{r})] + 4\pi\rho(\mathbf{r}) = 0 \quad (1)$$

where $\phi(\mathbf{r})$ is the electrostatic potential expressed in units of kT/e. It depends on $\epsilon(\mathbf{r})$, the dielectric constant; $\bar{\kappa}$, the modified Debye-Hückel parameter, and $\rho(\mathbf{r})$, the charge density function. The modified Debye-Hückel parameter is given by

$$\bar{\kappa} = \epsilon^{1/2}\kappa \quad (2)$$

where $1/\kappa$ is the Debye screening distance.

The multigrid algorithm

Multigrid methods in which partial differential equations (PDEs) are discretized into sets of equations on grids of different sizes are used to accelerate convergence of numerical solutions of the PDEs. The resulting system of discrete equations is solved at each grid level, an operation known

Received for publication 4 March 1993 and in final form 20 April 1993.

Address reprint requests to Norma M. Allewell.

0006-3495/93/07/48/08 $2.00

as smoothing the error. The residual from the solution at a given level is transferred to the next level where it is solved to provide the correction for the previous grid. Convergence is accelerated because the residual equations are solved on coarse grids, where the computational cost is negligible (18). Moreover, iterations at each level need not be carried to the point where convergence becomes slow (cf. Ref. 18).

There are two basic schemes for transferring information between grids. The most commonly used is the non-nested *V* or *W cycle*, which begins with the fine grid and proceeds through the coarser levels to correct low frequency errors before cycling back to the fine grid (Fig. 1). The second scheme, *nested iteration*, begins at the coarsest level and proceeds to successively finer grids, using a V cycle at each stage (cf. Refs. 18, 19, 22). We have used a slightly modified variant of the standard nested iteration technique described in Refs. 19 and 26 in which the number of iterations at each level is chosen adaptively during the calculations.

The robustness and efficiency of multigrid calculations are strongly influenced by the choice of a smoother (27). Commonly used smoothers are GS (Gauss-Seidel), SGS (Symmetrical Gauss-Seidel), LGS (Line Gauss-Seidel), ILU (incomplete LU decomposition) (22). Successive Overrelaxation (SOR) (28) is an accelerated form of Gauss-Seidel that has been widely used in finite difference calculations; however, it is not suitable as a smoother for multigrid, because it damps out the crucial high frequency correction. We have used nonlinear point Gauss-Seidel with natural ordering (19, 22).

Implementation

The set of equations that result from discretization of Eq. 1 on a uniform mesh of grid size h with dimension N^3 can be written as

$$L_h \mathbf{u}_h = F_h \tag{3}$$

where L is the coefficient matrix derived from the finite difference form of the Poisson-Boltzmann equation, $\mathbf{u}$ is a vector whose elements are the values of ϕ at the N^3 grid points, and F is derived from the source term containing the charge density ρ and the nonlinear terms from $\sinh[\phi(\mathbf{r})]$ (cf. Refs. 29 and 30).

Let $\tilde{u}_h$ be some approximate solution to Eq. 1, let the error in $\tilde{u}_h$ or the *correction* be v_h and the residual or *defect* be r_h. Then

$$\mathbf{u}_h = \tilde{\mathbf{u}}_h + v_h \tag{4}$$

and

$$L_h(\tilde{\mathbf{u}}_h + v_h) - L_h\tilde{\mathbf{u}}_h = F_h - L_h\tilde{\mathbf{u}}_h = -r_h. \tag{5}$$

The *residual* on a coarse mesh of size H (r_H) is defined by a *restriction* operator R that restricts r_h to the coarser grid

$$r_H = Rr_h \tag{6}$$

Similarly a *prolongation* operator P interpolates the *correction* to the finer grid

$$v_h = Pv_H \tag{7}$$

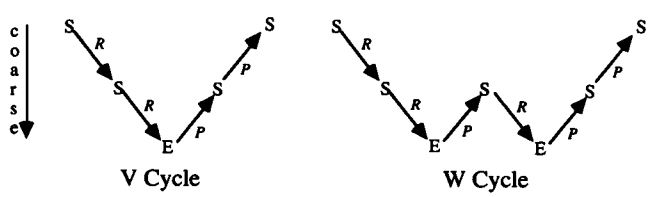

FIGURE 1 *V* and *W* multigrid cycles. Both begin at the top of the diagram with the finest grid. *S* denotes smoothing while *E* denotes an exact solution, routinely replaced by a two-grid iteration. Descending arrows denote restriction *R*; ascending arrows denote prolongation *P*. The *P* operator is a weighted bilinear interpolation, while the *R* operator is the adjoint to *P*.

The operation performed by *P* is a weighted trilinear interpolation of neighboring grid points. The restriction step *R* is a *straight injection* which fills each coarse grid point from the corresponding fine grid point.

Given a residual r_h from a fine grid, the residual equation for a coarser grid is

$$L_H\mathbf{u}_H - L_HR\tilde{\mathbf{u}}_h = -Rr_h. \tag{8}$$

The new value of the potential $\tilde{\mathbf{u}}_h$ on the fine grid is

$$\tilde{\mathbf{u}}_h^{new} = \tilde{\mathbf{u}}_h + P(\tilde{\mathbf{u}}_H - R\tilde{\mathbf{u}}_h). \tag{9}$$

The smoothing operator is applied at each grid level, and convergence is checked by calculating the Euclidian norm of the residual.

Free energies

As shown previously (31, 32), the *standard free energy change*, ΔG_i, for protonating a site on the protein is given by

$$\Delta G_i = \Delta G_{\text{int},i} + \Delta G_{\text{interact},i} \tag{10}$$

where $\Delta G_{\text{int},i}$ is the standard association free energy site i would have if other sites were electrically neutral, and $\Delta G_{\text{interact},i}$ is the free energy of interaction of the charge of this site with all other titrating groups on the protein. $\Delta G_{\text{int},i}$ can in turn be expressed as

$$\Delta G_{\text{int},i} = \Delta G_{\text{Born},i} + \Delta G_{\text{dipole},i} \tag{11}$$

where $\Delta G_{\text{Born},i}$ is the Born solvation energy or the work required to add a charge to the site, and $\Delta G_{\text{dipole},i}$ is the standard free energy of interaction of the charge at the site with the partial charges that represent the dipoles of the protein. The interaction term is

$$\Delta G_{\text{interact},i} = \sum_j Q_i W_{ij} \tag{12}$$

where Q_i is the charge at the titrating site and W_{ij} is the site-site coupling resulting from the potential produced at site $\mathbf{r}_j$ by a unit charge at $\mathbf{r}_i$, a second titrating site (31, 32). A thermodynamic cycle describing the behavior of the site with reference to a model compound allows the Born, dipole, and site-site terms to be partitioned (Fig. 2).

Titration curves

We have used two different procedures to calculate the fractional protonation of each site.

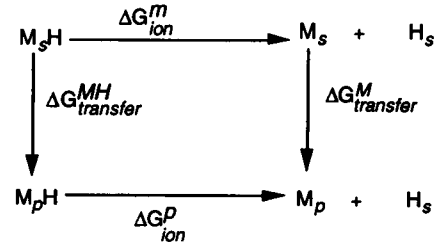

FIGURE 2 Thermodynamic cycle relating free energy changes associated with deprotonating a residue, M, in the protein (M_p) and in solution (M_s) (ΔG^p_{ion} and ΔG^m_{ion}, respectively). $\Delta G^{MH}_{transfer}$ and $\Delta G^{M}_{transfer}$ are the free energy changes associated with the desolvation of a residue in the protonated and deprotonated state, respectively. $\Delta\Delta G_{Born}$, the solvation contribution to the free energy change, is given by $\Delta G^{MH}_{transfer} - \Delta G^{M}_{transfer}$ with charges located only on titrating sites. $\Delta\Delta G_{dipole}$, the contribution of fixed charges, is obtained in a similar calculation in which only static or nontitrating charges are included. $\Delta G_{transfer}$ is the difference in the *electrostatic* free energy of the group in protein and in solution.

(*a*) The procedure of Tanford and Roxby (33) iterates to a set of self-consistent *partial* charges on titrating sites at a given pH. The fractional protonation state, θ_i, is estimated from the Henderson-Hasselbach equation

$$log \frac{\theta_i}{1 - \theta_i} = \mathrm{p}K_i - \mathrm{pH}. \quad (13)$$

The pK of site i is given by

$$pK_i = \mathrm{p}K_{\mathrm{adj},i} - \frac{1}{2.303kT} \sum_{j=1}^{N} W_{ij}(Q_j^0 + \theta_j) \quad (14)$$

where Q_j^0 is the charge of site j in the unprotonated state, and $\mathrm{p}K_{\mathrm{adj},i}$ is the $\mathrm{p}K_a$ of the residue in the protein with all other sites unprotonated and is given by

$$pK_{\mathrm{adj},i} = \mathrm{p}K_a - \frac{1}{2.303kT}(\Delta\Delta G_{\mathrm{Born}} + \Delta\Delta G_{\mathrm{dipole}}). \quad (15)$$

$\Delta\Delta G_{\mathrm{Born}}$ and $\Delta\Delta G$*dipole* are calculated as described in the legend of Fig. 2 and Ref. 34.

(*b*) The "reduced-site" approach of Bashford and Karplus (34) calculates the protonation state of a molecule with N titrating sites by taking a Boltzmann-weighted sum over all possible protonation states at each pH. The fractional protonation of site i is

$$\theta_i = \frac{\sum_{\{\mathbf{x}\}} x_i e^{-\beta\Delta G(\mathbf{x}) - \nu(\mathbf{x})2.303\,\mathrm{pH}}}{\sum_{\{\mathbf{x}\}} e^{-\beta\Delta G(\mathbf{x}) - \nu(\mathbf{x})2.303\,\mathrm{pH}}} \quad (16)$$

where $\{\mathbf{x}\}$ is an N-element protonation state vector whose elements $\mathbf{x}_i$ are zero or one depending on whether site i is unprotonated or protonated, respectively. $\nu(\mathbf{x})$ is the number of protons added in bringing the molecule to state $\mathbf{x}$. $\Delta G(\mathbf{x})$ is the standard free energy of the reaction $P + \nu(\mathbf{x})H \rightarrow PH_{\nu(\mathbf{x})}$, where $PH_{\nu(\mathbf{x})}$ is the protein in protonation state $\mathbf{x}$. $\{\mathbf{x}\}$ indicates the summation over all 2^N possible protonation states, and $\beta = 1/kT$ (31, 34).

Equation 15, though more rigorous than Eq. 12, is computationally expensive; in a protein with N titrating groups, the summation contains 2^N terms. The "reduced-site" approximation is based on the observation that at a given pH many of the 2^N sites are fully unprotonated, so the Boltzmann summation can be taken over a reduced set of sites. We find that for lysozyme the approach of Tanford and Roxby (33) yields results that are not significantly different from those obtained with the reduced site approach.

The protein model

The protein was modeled as a rigid object with a dielectric constant of 4; the solvent was assigned a dielectric constant of 80 and the ionic strength was set at 0.1 M. The boundary between solvent and the molecule and the ion exclusion shell were defined with probes of 1.4 Å (35) and 2.0 Å (16), respectively. Polar hydrogen positions were generated using the HBUILD facility of X-PLOR (36). For titrating groups the formal charge was placed on a single atom with the atom selected primarily on the basis of solvent accessibility and secondarily on proximity of nearby titrating sites. Calculations in which the charge was distributed over two or more atoms were also carried out; however, as discussed in Results, this procedure resulted in poorer agreement with experiment. Atomic radii and charges on nontitrating groups were taken from the CHARMM charge set (37) in X-PLOR, and the protein coordinates were obtained from the Protein Data Bank (PDB) (38).The final grid size of the whole protein was 181^3, corresponding to a resolution of 0.28 Å.

RESULTS

Comparison of multigrid and optimal successive overrelaxation (OSOR)

Recently an SOR application for the Poisson-Boltzmann equation termed optimal successive over-relaxation (OSOR), in which the optimal over-relaxation parameter was determined through an adaptive procedure, has been described (30). Fig. 3 shows two calculations carried out on a IBM RS6000 550 using our multi-gridding algorithm and an implementation of OSOR. The limiting fine and coarse grids used in the multigrid calculation were 111^3 and 3^3; the OSOR calculation used a grid of 111^3. A value of 10^{-4} for the norm of the residual is generally considered adequate. As shown, our algorithm achieves this value within approximately 10 s, while OSOR requires approximately 108 s. Discretization times are not included in these results. Our algorithm converges to a minimum of 10^{-11}, while the residuals calculated by OSOR do not converge within the time of the calculation. Multigrid is intrinsically more efficient than SOR; in addition, multigridding eliminates the need to determine the over-relaxation parameter required for OSOR (39) .

Timing considerations

Although multigrid methods have been demonstrated to be optimal and to solve general elliptic equation in O(N) operations, where N is the number of unknowns, the theory is not applicable to model problems with discontinuous coefficients (22). However, one can show empirically that the time spent is proportional to N, in this case the number of grid points, as illustrated in Fig. 4. Whereas solution time increases sharply with grid size for OSOR; it is a shallow function of grid size in multigridding, and the increased time is almost solely a function of setup time. Hence the advantage of the multigrid approach grows with increasing size of the grid.

Lysozyme

We have carried out calculations for the structures of three crystal forms of lysozyme deposited in the Protein Data Bank (Table 1) (Hodsdon, J. M., G. M. Brown, L. C. Sieker, L. H. Jensen, 1985. Private communication. PDB entry 1LZT; and

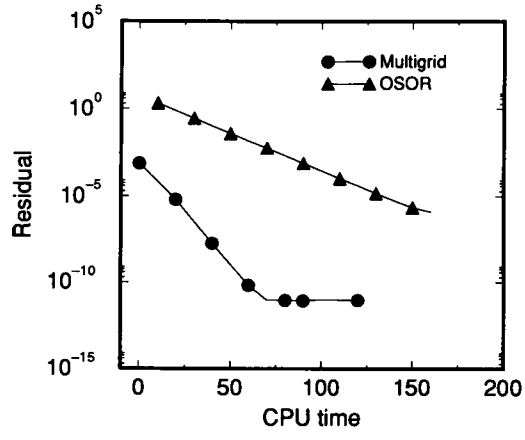

FIGURE 3 Comparison of rates of convergence with OSOR (▲) and one three-level Multigrid V cycle (●). The norm of the residual is defined as $(\Sigma_i r^{k^2})^{1/2}$ where r^k is the residual at the kth iteration (17). CPU time is in seconds. Discretization time is not included in these results.

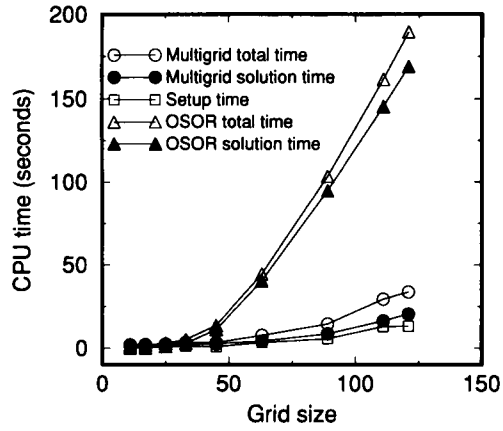

FIGURE 4 Comparison of OSOR (optimal successive overrelaxation) (△) and (▲) with Multigrid (○) and (●). Open symbols include setup time, while filled symbols represent solution time. Convergence was assumed when the relative residual changed less than 10^{-7}. The CPU time was obtained from the C clock subroutine. Discretization time on the finest grid (□) and the time required for setting up the coarse grids are included in the reported time to account for the overhead of setting up the problem on different grid levels in multigrid.

TABLE 1 Standard deviations and means of absolute differences between calculated and experimental pK values*

PDB entry	1LZT‡	1LYZ‡	1LYM‡
Unit cell	Triclinic	Tetragonal	Monoclinic
Resolution (Å)	1.9	2.0	2.5
Refinement	Blocked least-squares	Without refinement	Constrained least-squares
Standard deviation	1.14	2.97	5.85
Mean	0.65	2.17	3.39

* Experimental values from Refs. 31 and 44.
‡ PDB entries 1LZT (J. M. Hodsdon, G. M. Brown, L. C. Sieker, and J. H. Jensen. 1985. Private communication); 1LYZ (41) and 1LYM (42).

Refs. 41 and 42). The root mean square (RMS) deviation of all of the atoms in these structures is 1.8 Å, with the largest differences being in sidechain positions and orientations of carbonyl oxygens. Experimental pK_a values for 21 ionizable groups have been determined by Kurimatsu and Hamaguchi (43) and Dobson et al. (unpublished data; cited in Ref. 31).

We also compared results for five structures of tetragonal lysozyme, one derived without refinement and the other four using different refinement procedures (PDB entries 1–5 LYZ). Differences between these results are a measure of the effects of conformational differences independent of crystal packing. The variation in the calculated $pK_{1/2}$ values is smaller than between the three crystal forms (3.25 vs. 0.76).

Choice of ionizable atom

Fig. 5 compares agreement with experiment when the charge on a given sidechain is placed on a single atom rather than being distributed over two or more atoms. As the correlation coefficients indicate, agreement with experiment is best with the formal charge placed on a single atom, with the atom selected on the basis of solvent accessibility and, secondarily,

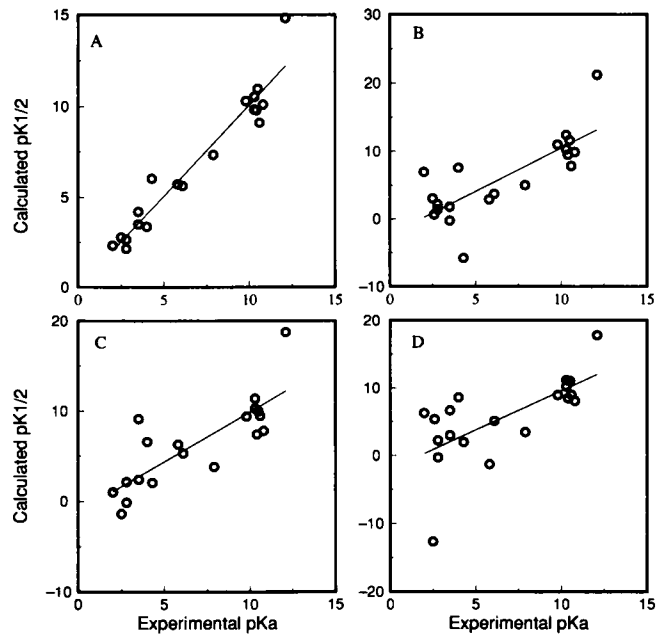

FIGURE 5 Effects of choice of charge scheme on the correlation of calculated $pK_{1/2}$ and experimental pK_a values. Since calculated pK values are pH-dependent, $pK_{1/2}$ is defined as the pH at which the group is half protonated. (*A*) 1LZT, charge placed on a single atom selected on basis of solvent accessibility and proximity to other charged groups, correlation coefficient 0.97. (*B*) 1LZT, charge distributed over two or more atoms, correlation coefficent 0.78. (*C*) 1LYZ, charge placed on a single atom selected on the basis of solvent accessibility and proximity to other charged groups, correlation coefficient 0.81. (*D*) 1LYZ, charge distributed over two or more atoms, correlation coefficent 0.67.

proximity of nearby titrating sites, when two atoms have equal solvent accessibility.

Comparison with experiment

Deviations of calculated $pK_{1/2}$ values from experiment are shown in Fig. 6. Differences between calculated and experimental values have the same sign for all three crystal forms. The best agreement with experiment is for the triclinic crystal structure (PDB entry 1LZT). The calculated pK values for 16 of the 21 groups (amino terminus, Asp-66, -87, -101, -119; Glu-7, -35; His-15; Lys-13, -97, -116; Tyr-20, -23; carboxyl terminus) for which experimental pK values are known (cf. Refs. 31 and 43) differ from the experimental values by less than 0.6 pH units. In general, agreement with experiment is best when the pK is close to the pH at which the crystal structure was determined. This is to be expected since the calculations do not take into account pH-induced changes in conformation. Other factors that may be important include crystal packing, bound solvent or ions, and conformational variability in solution.

Bashford and Karplus reported differences between calculated and experimental pK_a values for Tyr-53 of 9.2 and 11.2 for the tetragonal and triclinic forms, respectively (31). Our calculated value for Tyr-53 in the triclinic structure (1LZT) differs from experiment by only 2 pH units, although

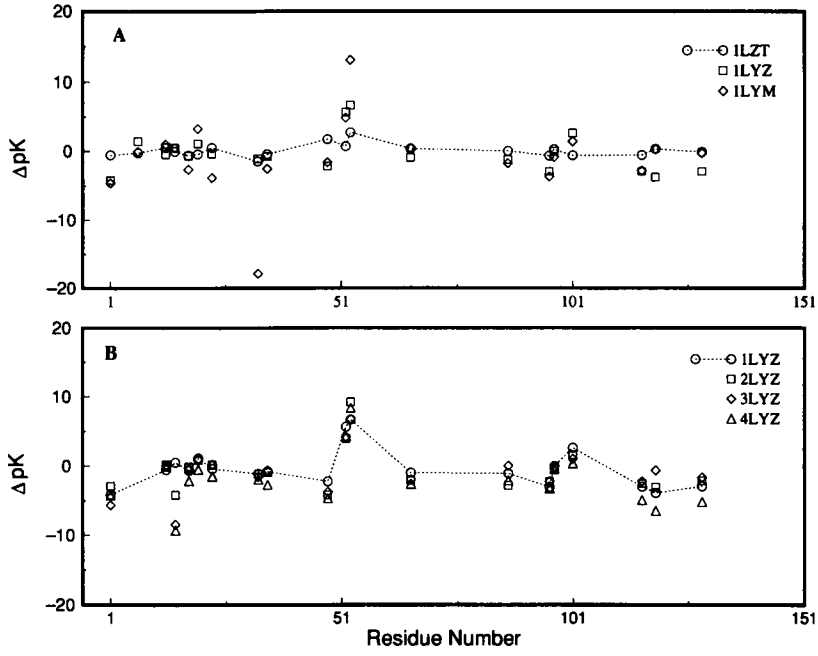

FIGURE 6 Differences between calculated and experimental $pK_{1/2}$ values at ionic strength 0.1 M. (*A*) Three crystal forms. (*B*) structures 1–5LYZ obtained with different refinement procedures. The dotted lines connect values for the structures which agree best with experiment (1LZT and 1LYZ). Residues in order are: amino terminus, Glu-7, Lys-13, His-15, Asp-18, Tyr-20, Tyr-23, Lys-33, Glu-35, Asp-48, Asp-52, Tyr-53, Asp-66, Asp-87, Lys-96, Lys-97, Asp-101, Lys-116, Asp-119, carboxyl terminus.

the differences are 6–13 pH units larger for the 1LYZ and 1LYM structures, respectively.

Glu-35

Glu-35 at the active site has been shown experimentally to have an anomalously high pK_a (6.1 compared to 4.4 for model compounds). Two explanations have been advanced. Blake et al. (44) attributed this anomalous pK_a to the location of Glu-35 in a highly nonpolar region formed by residues Ala-110, Gln-57, and Trp-108. Later Kurimatsu and Hamaguchi (43) focused on the electrostatic interaction of Glu-35 with Asp-52, 7 Å away. Our calculations indicate that the nonpolar local environment of Glu-35 is the major determinant. The calculated values of pK_{adj} (6.2 and 6.5 for 1LZT and 1LYZ crystals, respectively) are within 0.5 pH units of the experimentally determined pK_a value, while W_{ij} terms are small (0.6 and 0.9 for the tetragonal and triclinic forms, respectively). Hence the dominant terms appear to be the Born solvation and protein background terms, rather than the site-site interaction term.

His-15

The pK_a of His-15 has been determined as a function of ionic strength (45). These data are shown in Table 2 with calculated values for the 1LZT and 1LYZ data sets. There is good agreement between calculated and experimental values for the 1LZT data set.

Biphasic titration curves

The calculated titration curves for some residues are biphasic in certain crystal forms. Typical examples are shown in Fig. 7, where titration curves for Asp-52 (1LTZ) and Tyr-20 (1LTZ) are biphasic.

TABLE 2 Comparison of calculated and experimental p*K* values of His-15 as a function of ionic strength

		$pK_{1/2}^{calc}$	
Ionic strength	pK_a^{exp}*	1LZT	1LYZ
M			
0.4	6.00 ± 0.05	6.1	7.3
0.21	5.84 ± 0.03	6.2	6.4
0.11	5.73 ± 0.04	5.5	6.3
0.059	5.66 ± 0.05	5.4	6.4

* Experimental values from Ref. 45.

Potential surface

Fig. 8, *A* and *B*, shows the potentials at a contour level 2 *kT/e* at the active site of lysozyme when the net charge on the protein is +8 and −8, respectively. The asymmetric distribution of charge is apparent. These figures were produced with the program DAMPS, written in this laboratory.

DISCUSSION

Multigridding

Although multigrid algorithms are the subject of intensive research and have been shown rigorously through model problems to be frequently the method of choice for solving partial differential equations, they are not yet in general use. Because a general approach for applying multigrid to a variety of problems has not been developed, very few robust implementations exist; those that are available are generally designed around a specific problem.

We have developed a method for the rapid calculation of titration curves of large macromolecules using a multigrid procedure and the nonlinear Poisson-Boltzmann equation.

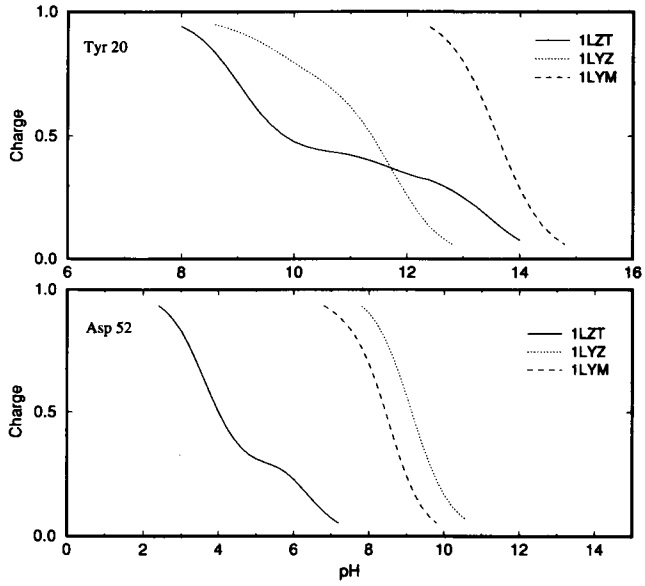

FIGURE 7 Representative biphasic titration curves. These calculations required approximately 3 h of CPU time on an IBM RISC 6000.

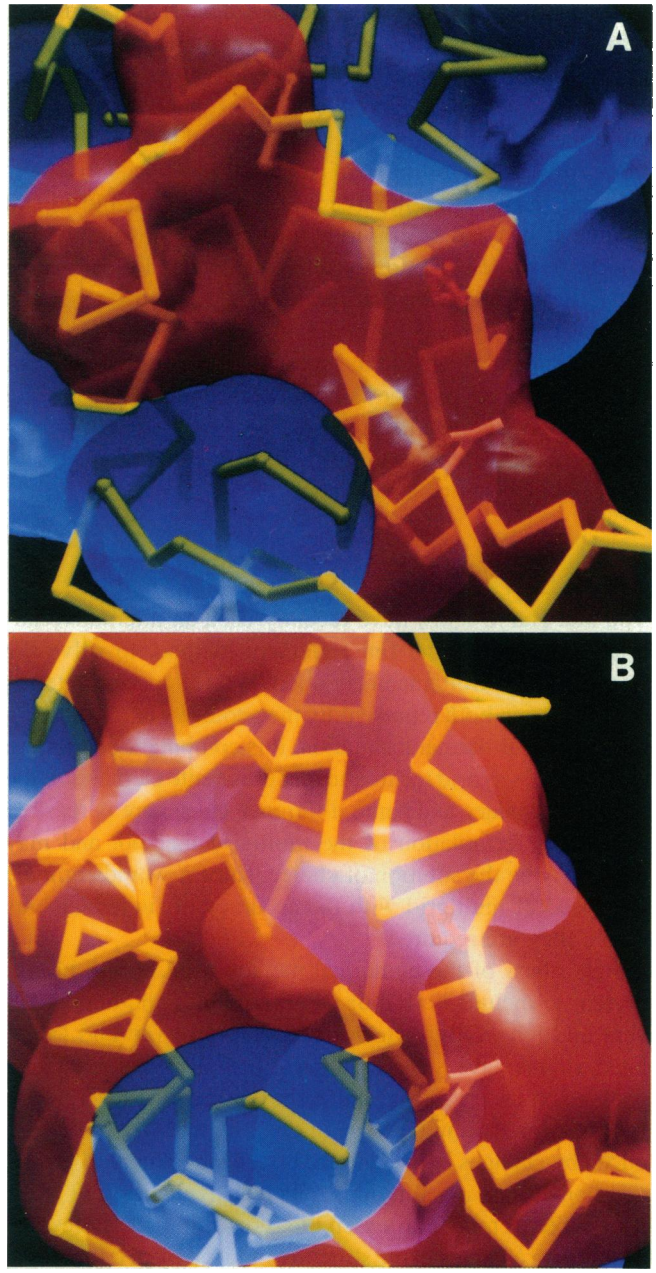

FIGURE 8 Potential surfaces of the active site of lysozyme at 2 *kT/e*. Positive potentials are in blue and negative potentials are in red. Net charge on protein +8 at pH 6.6 (*A*) and net charge −8 at pH 12 (*B*).

The multigrid method has a number of advantages. It converges rapidly, because iterations at each grid level are not carried beyond the stage of rapid convergence. It does not "saturate"; that is, the global error continues to decrease essentially without limit as the computation prcoeeds (18), because the use of multiple grids minimizes errors due to both discretization and truncation. In contrast to the "rapid" solvers that have been used to solve specific elliptic equations in $O(N \log N)$ operations where the coefficients in the equations are constant over most of the domain of the problem, multigrid methods solve general elliptic equation with nonconstant coefficients in $O(N)$ operations (22, 26). Because the limiting time is roughly the time taken for smoothing on the finest grid, the time required versus grid size is almost independent of grid size even when the time required for setting up the coarse grid coefficients is included. Furthermore, comparison of our results with those of Holst and Saied (24) indicates that the dependence of solution time on grid size is very similar for both the linear and nonlinear Poisson-Boltzmann equations. Hence this method provides the capability of using the nonlinear Poisson-Boltzmann equation to model large molecules and macromolecular assemblies while incorporating the effect of the protein-solvent interface.

The results for lysozyme presented above represent the best agreement between calculated and experimental pK values and titration curves reported to date. Fixed charges were neglected in calculations based on the Tanford-Kirkwood model (e.g., Ref. 1) and finite difference calculations of pK_a shifts resulting from site-directed mutations (14, 15). Errors in background terms were relatively large in previous calculations of titration curves for lysozyme by the finite difference method, because only site-site terms were calculated at high resolution (31, 46). Multigridding further increases the speed and accuracy of finite difference calculations because potentials are calculated at high resolution throughout the protein, reducing the errors in subsequent calculations of p$K_{1/2}$ values and individual site titration curves.

At the same time, these results demonstrate the sensitivity of calculated pK values to details of the structure. While the calculated pK values for data set 1LZT agree with experiment to within an RMS error of 1.1 pH units, the RMS error for the 1LYZ structure, which was solved at comparable resolution (2.0 Å vs. 1.95 Å) but with a different refinement algorithm is 2.9 pH units. There are also significant differences in the values calculated from structures that differ only

in the method used to refine the data. The 1LYM structure, which has the largest RMS error between calculated and experimental $pK_{1/2}$ values (5.80), is at relatively low resolution, and several residues, generally in loops, have high temperature factors. Accurate $pK_{1/2}$ values probably require structures solved at resolutions of at least 1.9 Å.

These results are reminiscent of those of Wendoloski and Matthew (47) and Northrup et al. (48), who examined the effects of protein flexibility on the calculated electrostatic properties and showed that families of structures generated by molecular dynamics simulations have substantial differences in local electrostatic interactions. The variations of 0.3–2.0 pK units that they observed for individual sites are analogous to the results presented here for lysozyme structures generated by using different refinement strategies to solve a single set of x-ray data.

Future prospects

You and Harvey (40) have recently used an alternative numerical approach, the finite element method, to solve the linear Poisson-Boltzmann equation. While superior in terms of accuracy in both representing the geometry and in solving the Poisson-Boltzmann equation, it is computationally more expensive. A hybrid technique that uses finite elements at the protein boundary and finite difference elsewhere is feasible within the formulation of the multigrid method, and may ultimately be the solution of choice, since it would combine the speed of multigrid with the accuracy of finite elements where they are required.

Regardless of the numerical method used, it will be important to include both molecular dynamics and bound solvent and ions. Discrepancies between observed and calculated values have several sources: the influence of crystal packing on the crystal structure, the existence of several conformers in solution, changes in the structure in solution induced by changes in pH, and the effects of "structural" water and bound ions in solution. Molecular dynamics will make it possible to model an ensemble of solution structures and to simulate changes in the structure induced by pH. Including bound solvent molecules and ions is equally important in terms of simulating changes in pK values induced by solvent and ion binding, although methods to do this well have yet to be developed.

We thank B. K. Lee and J. V. White for helpful comments that improved the quality of the presentation.

Supported by National Institutes of Health grant DK-17335 (to N. M. Allewell), IBM, and the University of Minnesota. The Minnesota Supercomputer Institute provided time on the Cray-2 for some of the calculations.

REFERENCES

1. Matthew, J. B., and F. R. N. Gurd. 1986. Enzyme structure: calculation of electrostatic interactions in proteins. *Methods Enzymol.* 130:413–436.
2. Warshel, A., and S. T. Russell. 1984. Calculation of electrostatic interactions in biological systems and in solution. *Quart. Rev. Biophys.* 17:283–422.
3. Matthew, J. B. 1985. pH dependent processes in proteins. *CRC Crit. Rev. Biochem.* 18:91–197.
4. Rogers, N. K. 1986. The modeling of electrostatic interactions in the function of globular proteins. *Prog. Biophys. Mol. Biol.* 48:37–66.
5. Harvey, S. C. 1989. Treatment of electrostatic effects in macromolecular modeling. *Proteins.* 5:78–92.
6. Sharp, K. A., and B. H. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332.
7. Davis, M. E., and J. A. McCammon. 1990. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 90:509–521.
8. Allewell, N. M., and H. Oberoi. 1991. Electrostatic effects in protein folding and function. *Methods Enzymol.* 202:3–19.
9. Warwicker, J., and H. C. Watson. 1982. Calculation of electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.* 155:53–62.
10. Klapper, I., R. Hagstrom, R. Fine, K. A. Sharp, and B. H. Honig. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins.* 1:47–59.
11. Davis, M. E., and J. A. McCammon. 1990. Calculating electrostatic forces from grid-calculated potentials. *J. Comp. Chem.* 11:401–409.
12. Gilson, M. K., K. A. Sharp, and B. H. Honig. 1988. Calculating electrostatic interactions in biomolecules: method and error assessment. *J. Comp. Chem.* 9:327–335.
13. Warwicker, J. 1986. Continuum dielectric modelling of the protein-solvent system, and calculation of the long range electrostatic field of the enzyme phosphoglycerate mutase. *J. Theor. Biol.* 121:199–210.
14. Gilson, M. K., and B. H. Honig. 1987. Calculation of electrostatic potentials in an enzyme active site. *Nature. (Lond.).* 330:84–86.
15. Sternberg, M. J. E., F. R. F. Hayes, A. J. Russell, P. G. Thomas, and A. R. Fersht. 1987. Prediction of electrostatic effects of engineering of protein charges. *Nature. (Lond.).* 330:86–88.
16. Gilson, M. K., and B. H. Honig. 1988. Energetics of charge-charge interactions in proteins. *Proteins.* 3:32–52.
17. Yang, A-S., M. R. Gunner, R. Sampogna, K. A. Sharp, and B. H. Honig. 1983. On the calculation of pK_as in proteins. *Proteins.* 15:252–265.
18. Brandt, A. 1977. Multi-level adaptive solutions to boundary-value problems. *Math. Comput.* 31:333–390.
19. Douglas, C. C. 1984. A multigrid optimal solver for elliptic boundary value problems: the finite difference case. *In* Advances in Computer Methods for Partial Differential Equations. V. R. Vichnevetsky and R. S. Stepleman, editors. IMACS. New Brunswick, NJ. 369–374.
20. Ortega J. M., and R. G. Voigt. 1985. Solution of partial differential equations on vector and parallel computers. *SIAM Rev.* 27:149–240.
21. Briggs, W. L. 1988. A multigrid tutorial. SIAM. Philadelphia, PA. 1–90.
22. Hackbusch, W. 1985. Multi-grid methods and applications. Springer-Verlag, Berlin. 1–377.
23. Hackbusch, W. 1988. A new approach to robust multi-grid solvers. *In* ICIAM: Proceedings of the First International Conference on Industrial and Applied Mathematics. SIAM. Philadelphia. 114–126.
24. Holst, M., and F. Saied. 1993. Multigrid solution of the Poisson-Boltzmann equation. *J. Comp. Chem.* 14:195–113.
25. McQuarrie, D. A. 1976. *Statistical mechanics.* Harper & Row, New York. 1–641.
26. Douglas, C. C., and J. Douglas. 1993. A unified convergence theory for abstract multigrid or multilevel algorithms, serial and parallel. *SIAM J. Numer. Anal.* 30:136–158.
27. Bramble, J. H., and J. E. Pasciak. 1992. The analysis of smoothers for multigrid algorithms. *Math. Comp.* 58:467–488.
28. Young, D. M. 1971. Iterative solution of large linear systems. Academic Press, New York. 1–570.
29. Jayaram, B., K. A. Sharp, and B. H. Honig. 1989. The electrostatic potential of B-DNA. *Biopolymers.* 28:975–993.
30. Nicholls, A., and B. H. Honig. 1991. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* 12:435–445.
31. Bashford, D., and M. Karplus. 1990. pK_as of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry.* 29:10219–10225.

32. Tanford, C., and J. G. Kirkwood. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79:5333–5339.

33. Tanford, C., and R. Roxby. 1972. Interpretation of protein titration curves. Application to lysozyme. *Biochemistry.* 11:2192–2198.

34. Bashford, D., and M. Karplus. 1991. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* 95:9556–9561.

35. Lee, B. K., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

36. Brunger, A. T. 1992. X-PLOR Manual, Version 3.0. Yale Univ. New Haven, Connecticut. 1–403.

37. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

38. Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyers, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

39. Davis, M. E., and J. A. McCammon. 1989. Solving the finite difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. *J. Comp. Chem.* 10:386–391.

40. You, T. J., and S. C. Harvey. 1983. A finite element approach to the electrostatics of macromolecules with arbitrary geometries. *J. Comp. Chem.* 14: 484–501.

41. Diamond, R. 1974. Real-space refinement of the structure of hen egg-white lysozyme. *J. Mol. Biol.* 82:371.

42. Rao, S. T., J. Hogle, and M. Sundaralingam. 1983. Studies of monoclinic hen egg white lysozyme. II. The refinement at 2.5 angstroms resolution - conformational variability between the two independent molecules. *Acta. Crystallogr.* C39:237.

43. Kurimatsu, S., and K. Hamaguchi. 1980. Analysis of acid-base titration curve of hen lysozyme. *J. Biochem.* 87:1215–1219.

44. Blake, C. C. F., L. N. Johnson, G. A. Mair, A. T. C. North, D. C. Philips, and V. R. Sarma. 1967. Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. R. Soc. Lond. Ser. B. Biol Sci.* 167:378–388.

45. Takahashi, T., H. Nakamura, and A. Wada. 1992. Electrostatic forces in two lysozymes: calculations and measurements of histidine pK_a values. *Biopolymers.* 32:897–909.

46. Beroza, P., D. R. Fredkin, M. Y. Okamura, and G. Feher. 1991. Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroids. Proc. Natl. Acad. Sci. USA.* 88:5804–5808.

47. Wendoloski, J. J., and J. B. Matthew. 1989. Molecular dynamics effects on protein electrostatics. *Proteins.* 5:313–321.

48. Northrup, S. H., T. G. Wensel, C. F. Meares, J. J. Wendoloski, and J. B. Matthew. 1990. Electrostatic field around cytochrome c: theory and energy transfer experiment. *Proc. Natl. Acad. Sci. USA.* 87:9503–9507.